

Abstract

Machine Learning for Tau Leptons and the Search for the Associated Production of a Higgs Boson with a Vector Boson, with the Higgs Boson Decaying to a Tau Pair at ATLAS

Emma Grace Castiglia

2022

After the discovery of the Higgs boson in 2012, many particle physicists have focused on measuring its various production methods and decay modes. One channel that has previously only been measured at well below the observation threshold is the associated production of a Higgs boson with a vector boson, where the Higgs boson decays to a pair of taus. In this thesis, I describe an ongoing search for that channel using data from proton-proton collisions taken during Run 2 at the Large Hadron Collider with the ATLAS detector (2015-2018), with a recorded integrated luminosity of 139 fb^{-1} at a center of mass energy $\sqrt{s} = 13\text{TeV}$. In the analysis, the W boson must decay to a light lepton (e/μ) and a neutrino, while the Z boson is required to decay to a pair of electrons or muons. At least one of the taus from the Higgs boson must decay hadronically. The background from misidentified electrons, muons, and taus is modeled by a data-driven Fake Factor method. The remaining backgrounds are primarily from dibosons, and these are modeled with Monte Carlo simulation.

Machine learning has become an integral tool across many experiments in particle physics, including the ATLAS experiment. In this thesis, I describe how neural networks and decision trees are used to improve tau identification and the calibration of the tau energy, as well as ongoing studies for future improvements. Machine learning algorithms are also instrumental in separating out a complicated signal from a very similar background. In the Higgs boson analysis detailed in this thesis, the score from a neural network trained to target the diboson background is used as the final discriminating variable.

Machine Learning for Tau Leptons and the Search for the Associated Production of a Higgs
Boson with a Vector Boson, with the Higgs Boson Decaying to a Tau Pair at ATLAS

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
In Candidacy for the Degree of
Doctor of Philosophy

By
Emma Grace Castiglia

Dissertation Director: Professor Sarah Demers

May 2022

©2022 – EMMA GRACE CASTIGLIA
ALL RIGHTS RESERVED.

Contents

ACKNOWLEDGEMENTS	ix
LISTING OF FIGURES	9
LISTING OF TABLES	11
o INTRODUCTION	12
I THEORY INTRODUCTION	14
1.1 The Standard Model	15
1.1.1 Fermions	15
1.1.2 Bosons	17
1.1.3 Forces	17
1.2 Interactions between Particles	18
1.2.1 Conservation Laws	18
1.2.2 Feynman Diagrams for Fundamental Interactions	18
1.3 Lagrangian formalism of the Standard Model	20
1.3.1 The Dirac Interaction	20
1.3.2 Introduction to Gauge Invariance	21
1.3.3 Quantum Electrodynamics (QED)	22
1.3.4 Quantum Chromodynamics (QCD)	23
1.3.5 Weak Interaction	24
1.3.6 Electroweak Unification	25
1.4 Standard Model Lagrangian	26
1.5 Higgs boson	27
1.5.1 Electroweak Symmetry Breaking	27
1.5.2 Higgs Mechanism	28
1.5.3 Higgs boson and Mass	30
1.5.4 Higgs boson discovery	32
1.5.5 Higgs boson Production Mechanisms	32
1.5.6 Higgs boson decays	34
1.6 Limits of the Standard Model	37
2 LHC AND THE ATLAS DETECTOR	39
2.1 Large Hadron Collider and its Predecessors	39
2.1.1 Conventions at Particle Colliders	41
2.1.2 LHC Construction and Operation	43
2.1.3 Accelerating protons at the LHC	46
2.1.4 Run 2 of the LHC	48
2.2 ATLAS Detector	49
2.2.1 Inner Detector	52
2.2.2 Calorimeters	55

2.2.3	Muon Spectrometer	58
2.2.4	Magnet System	60
2.3	Trigger and Data Acquisition (TDAQ)	61
2.4	CERN Computing Grid	63
3	SIMULATION AND COMPUTATION FOR PARTICLE PHYSICS EVENTS AT THE ATLAS EXPERIMENT	65
3.1	Monte Carlo Simulation	67
3.1.1	Event Generation	67
3.1.1.1	QCD and PDFs in Proton Collisions	69
3.1.1.2	Parton Showers	70
3.1.2	Detector Simulation	72
3.1.3	Event Digitization	72
3.2	Particle Reconstruction	73
3.2.1	Tracking	73
3.2.2	Vertexing	76
3.2.3	Calorimeter Cluster Reconstruction	77
3.3	Particle Identification	79
3.3.1	Electrons and Photons	79
3.3.1.1	Electron Identification	81
3.3.1.2	Photon Identification	83
3.3.1.3	Energy Calibration	83
3.3.2	Muons	84
3.3.3	Jets	87
3.3.4	Missing Transverse Energy	91
4	INTRODUCTION TO MACHINE LEARNING	93
4.1	Introduction to Machine Learning	93
4.1.1	Boosted Decision Trees	98
4.1.1.1	Boosted Regression Trees	100
4.1.2	Neural Networks	101
4.1.2.1	Recurrent Neural Networks	105
4.2	ML in ATLAS	106
5	MACHINE LEARNING FOR TAU LEPTONS	109
5.1	Introduction to Taus	109
5.2	Tau Properties	109
5.3	Tau Reconstruction	111
5.3.1	Reconstruction in the Tau Trigger	112
5.4	Tau Identification	113
5.4.1	BDT for Tau Identification	113
5.4.2	RNN for Tau Identification	113
5.5	Tau Energy Scale	115
5.6	Machine Learning Improvements to the Tau Energy Scale	117
5.6.1	Boosted Regression Tree Architecture for the Tau Energy Scale	118
5.6.2	Input Variable Definitions	119
5.6.3	FTK Track Studies	122
5.6.4	Simulated HLT Tau Trigger Track Studies for MVATES	125

5.6.4.1	Tuning of Hyperparameters	128
5.6.5	Results of MVATES with Trigger-Level Track Variables	130
5.7	Lorentz Equivariant Neural Network for tau Tagging	130
5.7.1	What does equivariance mean?	131
5.7.2	Data Processing	132
5.7.3	Architecture of Network	133
5.7.4	Initial Results	135
5.7.5	Comparison to Deep Sets Architecture	135
5.8	Concluding Thoughts on Machine Learning for Taus	136
6	$VH, H \rightarrow \tau\tau$ ANALYSIS	137
6.1	Motivation	138
6.2	Four Signal Categories	140
6.2.1	$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$	140
6.2.2	$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$	141
6.2.3	$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$	141
6.2.4	$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$	142
6.3	Data and Monte Carlo	142
6.4	Object Criteria	144
6.5	Overlap Removal	146
6.6	Mass Reconstruction	147
6.6.1	Missing Mass Calculator (MMC)	147
6.6.2	Late-Projected Transverse Mass (M_{2T})	148
6.7	Backgrounds	150
7	FAKE RATE CALCULATIONS	151
7.1	Introduction to the Fake Factor Method	151
7.1.1	Fake Factor Method Derivation	152
7.2	Z + jets Selection	156
7.2.1	$Z \rightarrow \mu\mu$ Selection	157
7.2.2	$Z \rightarrow ee$ Selection	157
7.2.3	Diboson Subtraction	158
7.3	Fake Rate Measurements	159
7.3.1	Electron Fake Rate	159
7.3.2	Muon Fake Rate	161
7.3.3	Tau Fake Rate	165
8	ANALYSIS OPTIMIZATION	172
8.1	Selection Cuts	173
8.1.1	$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$	173
8.1.2	$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$	177
8.1.3	$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$	179
8.1.4	$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$	181
8.2	Optimization of Signal Selections Cuts	183
8.2.1	$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$	184
8.2.2	$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$	184
8.2.3	$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$	188
8.2.4	$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$	189

8.3	Neural Network Targeting Dibosons	190
8.3.1	$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$ Neural Network	194
8.3.1.1	WHlephad_ee Neural Network	194
8.3.1.2	WHlephad_emu Neural Network	197
8.3.1.3	WHlephad_mumu Neural Network	199
8.3.2	$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ Neural Network	203
8.3.3	$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$ Neural Network	204
8.3.4	$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$ Neural Network	206
9	ANALYSIS SENSITIVITY PROJECTIONS	208
9.1	Run 1 Results	208
9.2	Hypothesis Testing	210
9.2.1	Likelihood Function	211
9.2.2	Test Statistics	212
9.3	Systematics	214
9.3.1	Experimental Uncertainties	214
9.3.2	Data-Driven Fake Uncertainties	216
9.3.3	Theory Uncertainties	216
9.4	Asimov Fit Results	220
10	FINAL REMARKS	222
APPENDIX A	AUTHOR'S INDIVIDUAL CONTRIBUTIONS	224
A.1	$VH, H \rightarrow \tau\tau$ Analysis	224
A.1.1	Related Presentations	225
A.2	Machine Learning for Taus	225
A.2.1	Related Presentations	225
A.3	Outreach and Leadership	226
A.4	Teaching	226
REFERENCES		236

TO ALL ASPIRING WOMEN PHYSICISTS

Acknowledgments

THERE ARE FAR TOO MANY PEOPLE TO THANK IN A SHORT AMOUNT OF SPACE, BUT I WILL DO MY BEST.

First, and foremost, I would like to thank my family. Thank you to my parents for providing everything I needed to get to where I am and for always pushing me to do my best and take on each challenge I faced. I appreciate you always editing the writing work I sent you, not complaining too much when I arrived home with multiple loads of laundry, and for making the effort to try to understand what I am currently working on. Thank you to my sister for her constant support and for the friendly competition. Please know I am so proud of everything you have achieved these last few years. Thank you to my Grandad for inspiring a love of math and problem solving when I was young. Thank you to my Grandma, who I wish could be here to see me achieve this goal. I miss you and our games of Oh Hell and Rummikub over a cup of tea. Thank you to my extended family for putting up with me attempting to explain my physics research at the holiday dinner table and for not getting too upset when I slam the ball in table tennis.

To my physics cohort, we did it! I was so excited to attend the defenses before mine and look forward to cheering the rest of you on. I am so happy to have played (read yelled during) so many games of Avalon and Werewolf with you both in the grad school office and at our respective apartments. I enjoyed our trivia team and karaoke nights at Gryphon's, Sunday Family Dinners, and all you can eat sushi trips. I know that when I look back at my time in graduate school, our friendships will be what I am most grateful for.

To the physics girls: Emily, Kelly, and Luna, thank you for the outlet shopping trips, apple picking, cookie nights, nights at wine cave, venting sessions, dog walks, and constant support for the last six years. To my extended New Haven family: Amy, Caitie, Hannah, Holly, Jacob, Jeremy, Jessica, Jules, Paul, Samantha, Will. Thank you for the runs, the skiing trips, the hikes up East Rock, the dread games, the ice cream trips, the restaurant week meals, the fancy cocktails, the weeknight dinners, and so much more.

To my undergraduate friends, thank you for allowing me to escape grad school and come visit you and for all the career advice and referrals (h/t Alice, Alita, Emilee, and Megan). Thank you to "Lace is a Lifestyle" for the constant companionship as we took on various graduate programs and jobs across the country.

To Alec for being my pandemic companion and for always making sure I was fed when I got too busy to cook for myself. Thank you for believing in me and pushing me to try new things, from lifting weights to smoking brisket.

To the women in the department, thank you for letting me run Women in Physics+ for so many years. I know it is in good hands going forward and I hope it will continue to provide as much value to others as it did for me.

Last, but definitely not least, thank you to my research group. Giani, Mariel, and Savannah, we conquered the xTauFramework, the derivation and implementation of the fake backgrounds, and TREx-Fitter together. I will miss our chats over coffee in the astro lounge, working through problems on the white board, and our table of treats and legos/puzzles.

Thank you to Professors Helen Caines and Ian Moulton for serving on my thesis committee and thank you to both of my advisors during my time here at Yale. Thank you to Professor Keith Baker, who set me off on my path of using machine learning in particle physics. Thank you to Sarah Demers for teaching

me how to be a researcher in physics and for allowing me to take risks and pursue interests outside of (and within) particle physics research. Also, thank you for instilling the importance of chocolate during research meetings. It was definitely needed during the writing of this thesis.

“I have dreamt in my life, dreams that have stayed with me ever after, and changed my ideas; they have gone through and through me, like wine through water, and altered the color of my mind. And this is one: I’m going to tell it – but take care not to smile at any part of it.”

Emily Bronte, *Wuthering Heights*

Listing of figures

1.1	Standard Model Table formulation with three generations of quarks and leptons, the Gauge bosons, and Higgs boson along with their mass, charge, and spin. [1]	16
1.2	The main SM interaction vertices for QED, QCD, and the Weak force. The coupling strength is denoted by α [2]	19
1.3	The Standard Model Lagrangian mug sold at CERN.	27
1.4	An illustrative example of spontaneous symmetry breaking for a complex scalar field potential $V(\phi)$, with (a) showing solutions for $\mu^2 > 0$ and (b) for $\mu^2 < 0$. [2]	28
1.5	The main production modes for the SM Higgs boson as a function of the Higgs boson mass at $\sqrt{s} = 14\text{TeV}$. [3]	33
1.6	Feynman diagrams for the four main Higgs boson production mechanisms at the LHC. [4]	34
1.7	Relative frequency of Higgs boson branching ratios with respect to the mass of the Higgs boson. [5]	35
1.8	Four lepton invariant mass of selected signal of $H \rightarrow ZZ \rightarrow 4l$ and expected backgrounds in the 2012 ATLAS measurement. [6]	36
1.9	Summary of SM production cross section measurements and their agreement with theory. [7]	37
2.1	Distribution of mean interactions per crossing weighted by luminosity for the full Run 2 pp collision data set of the ATLAS detector. This includes all data taken during stable beams. [8]	44
2.2	Aerial view of CERN and Geneva, with the LHC tunnel in yellow, and the four main detectors in white, with the French/Swiss border shown with the white dashed line. [9] . .	45
2.3	Cross section of the LHC beam line showing location of quadrupole magnet bars, alignment target, vacuum vessel, non-magnetic collars, and dual beams. [10]	46
2.4	A schematic of the entire LHC linear and circular accelerator complex. Protons are injected into Linac 2 and then passed to the Booster, PS, SPS, and finally the LHC ring where they are accelerated to energies of 6.5TeV. [11]	47
2.5	Timeline of running the LHC created in 2016 before there were delays to Run 3 due to covid 19. [12]	48
2.6	Total integrated luminosity of Run 2 delivered by the LHC (green), recorded by ATLAS (yellow), and data collected that can be used for physics (blue) (left). Total projected luminosity for the HL-LHC (right). [13]	49
2.7	Schematic of the ATLAS detector with the four major subsystems and people for scale. [14].	50
2.8	Coordinate axes of the ATLAS detector showing the z-axis as the beam line, the x-axis pointing towards center of the LHC, and the polar angles. [15]	50
2.9	Diagram of pseudorapidity (η) for various values of polar angle θ with respect to the beam line. [15]	51
2.10	Schematic of a slice of the ATLAS detector, demonstrating where in the calorimeters various types of particles deposit their energy. [16]	52

2.1.1	Cut-away diagram of the ATLAS Inner Detector with barrel and end-cap portions of the Pixel Detector, SCT, and TRT detectors. [14]	53
2.1.2	Diagram of the TRT, SCT, and pixel subsystems and their locations in r , z , and η in the ATLAS Inner Detector. [14]	54
2.1.3	Cut-away of the ATLAS calorimeter showing both the liquid Argon (LAr) and Tile systems of the EM and hadronic calorimeters. [14]	55
2.1.4	Diagram of the EM calorimeter and its accordion structure, showing the resolution in ϕ and η (left) Diagram of the barrel cryostat for the LAr EM Calorimeter with the presampler and feedthroughs (right). [17]	56
2.1.5	Schematic of tile calorimeter module with alternating lead and scintillating layers normal to the beamline and with the wavelength shifting fiber readouts parallel to the beamline. [18]	58
2.1.6	Cut away view of the Muon Spectrometer with the 4 main detector components and magnets. [14]	59
2.1.7	Transverse view of the Muon Spectrometer with the Resistive Plate Chambers and Monitored Drift Tubes in the barrel portion. [19]	60
2.1.8	ATLAS Magnet System. [20]	61
2.1.9	Event Display from ATLAS in 2015 at the start of Run 2. A view directly at the beam line axis, shows the number of charged particle trajectories (curved lines) in the tracking detectors (left). A view facing perpendicular to the beam line axis (transversing along it) showing the same event is in the bottom right. There are a number of particle interactions coming from separate vertices. Some of these vertices should correspond to lower energy collisions or “pileup” events that are background to more interesting particle analyses. [21]	62
2.2.0	ATLAS TDAQ system in Run 2 with the main L1 components boxed in grey and the general flow of detector information to the HLT. The now defunct FTK is also included.	63
2.2.1	A photo of the CERN Data Center with some of its 230000 processor cores and 15000 servers.	64
3.1	Schematic of the data flow for both raw data and simulation. The top row on the left shows the process for data taken by the ATLAS detector and the bottom row on the left shows the process for simulated data. At the stage of reconstruction and interpretation, shown on the right, the data and simulation can be directly compared with each other and combined in various ways in different physics analyses. [22]	66
3.2	Sherpa collision simulation of a $t\bar{t}h$ event. The hard scattering modeled by matrix elements is in red (large red circle), with the smaller red circles representing the top quarks and Higgs boson. The parton showers producing bremsstrahlung are in blue, the purple parts show the secondary interaction, the green is highlighting the hadronization of QCD partons using fragmentation models, the darker green is simulations of unstable decays, and the QED bremsstrahlung (photon radiation) is in yellow. [23]	68
3.3	Proton PDFs for energy scales $Q^2 = 10\text{GeV}^2$ (left) and $Q^2 = 10^4\text{GeV}^2$ (right) with 68% confidence-level error bands. [24]	71
3.4	Sequence of steps in the ambiguity solver including track scoring, track requirement applications, and the NN for merged clusters. [25]	75
3.5	Efficiency of track finding as a function of p_T for tracks from a ρ , 3 and 5 prong τ , and a B^0 . [25]	75
3.6	The five track perigee parameter definitions.	77

3.7	Topo-clusters for simulated dijet event (2010) outlined in black. No pileup was included in this simulation. The white cells are negative signal cells and the grey cells have no signal but are completely between topo-clusters. [26]	78
3.8	Flow diagrams for the reconstruction algorithms of both electrons (left-side) and photons (right-side) showing the similarities and differences between them. [27]	80
3.9	Flow of the ambiguity algorithm to resolve superclusters reconstructed as both electrons and photons. It includes information from the trackers and potential conversion vertices. [27]	81
3.10	Path of an electron (red line) through the ATLAS detector where it passes through the inner tracking detectors and enters the EM calorimeter. The dashed red line shows the path of a photon produced during the electron's interaction with the detector. [28]	82
3.11	Reconstruction efficiency for the cluster, track, cluster and track, and electron as a function of E_T (left). Electron Identification efficiencies for the three working points as a function of E_T in a $Z \rightarrow ee$ sample (right). [27]	82
3.12	Efficiency of Tight photon identification for both converted (left) and unconverted background photons (right) from jets in a $Z \rightarrow ll + jets$ sample for $E_T < 25\text{GeV}$ and from a filtered dijet sample for $E_T > 25\text{GeV}$. The bottom panel is the ratio of E_T dependent and independent algorithms. [27]	84
3.13	Schematic showing the four main types of muons from the various reconstruction algorithms. [29]	86
3.14	Muon reconstruction and identification efficiency for the three working points as a function of muon p_T in $J/\Psi \rightarrow \mu\mu$ events (left) and as a function of η for $Z \rightarrow \mu\mu$ events with $p_T > 10\text{GeV}$ (right). [30]	86
3.15	Four main jet clustering algorithms performed on a Herwig generated parton level event. The various colors represent the different shapes of the jets. [31]	88
3.16	The steps of the Jet Energy Calibration. [32]	90
4.1	The tradeoff between underfitting and overfitting a model when comparing loss and model complexity.	96
4.2	The tradeoff between bias and variance in a model, with high variance related to overfitting and high bias related to underfitting.	97
4.3	Schematic of a boosted decision tree. A determination is made at each of the nodes regarding whether the values are greater than or less than the optimal split point. [33]	99
4.4	Schematic of a boosted regression tree. Each new tree is trained to learn, and thus reduce, the error left from the previous trees. [34]	101
4.5	Schematic of a perceptron with three input values (x_i) and a single output. [35]	102
4.6	Single layer neural network and a deep neural network with multiple hidden layers.	103
4.7	Various activation functions used in the neurons of a neural network.	103
4.8	Comparison of Adam (red) with gradient descent (black arrows) at arriving of minima of quadratic loss functions. [36]	105
4.9	Simple RNN on the left and the RNN unfolded on the right with the connections between input-hidden, hidden-hidden, and hidden-output shown. [36]	106
4.10	LSTM cell with the incoming new information $x(t)$, previously hidden information, and the forget, input, and output gates. Adapted from [37]	107

5.1	Example Feynman diagrams for the two main decay modes of the tau lepton. Both decay modes include a tau neutrino to preserve the tau flavor in the decay. The leptonic decay (on the left) includes a W decaying to a lepton and an anti-neutrino of the same flavor. The hadronic decay (on the right) includes a W coupling to quarks which hadronize into a variety of hadrons.	110
5.2	On the left is the typical signature of a QCD jet with many non-isolated tracks and on the right is the typical signature of a hadronic tau decay with its highly collimated one to three charged tracks that are contained in the core region. [38]	112
5.3	Rejection power for background quark- and gluon-initiated jets as a function of true $\tau_{had-vis}$ efficiency. The BDT performance is shown with the dashed lines in red for 1-prong and in blue for 3-prong, while the RNN is shown with solid lines. The solid dots represent the four working points: Tight, Medium, Loose, and Very Loose. [39]	114
5.4	Diagram of the RNN architecture for tau identification with the three separate branches for low-level track and cluster variables and high-level variables. [39]	115
5.5	Illustration of a hadronic tau decay highlighting some of the effects, such as underlying event and pileup, that make energy calibration tricky for taus.	116
5.6	Energy resolution of the baseline and BRT calibrations for taus as a function of tau momentum. The BRT calibration improves upon the Baseline calibration for taus with p_T below 250GeV. [40]	117
5.7	The left plot shows tau momentum distributions highlighting the difference between Truth p_T , the HLT Calibrated p_T (detector p_T), the total track p_T , and the average track p_T . The target value of regression for the BRT is the ratio of Truth p_T /Detector p_T . The right plot shows the weighted transverse momentum distributions for the leading three tau tracks, weighted by detector p_T . They are set to -1 if there are fewer than three tracks. As seen in the left plot, detector p_T and sum of tau track p_T are not the same, which leads to the weighted ratio of the leading track p_T to sometimes be greater than 1.	122
5.8	Comparison of the offset of the calibrated tau momentum from the truth momentum of three different trainings of the BRT for the MVATES weight: no FTK (only the HLT Calorimeter variables), three FTK variables (adding in three of the track variables: total track p_T , average track p_T , and nTracks), and using six FTK variables (adding in the weighted track p_T for the leading three taus).	123
5.9	Comparison of the offset of the calibrated tau p_T from the truth p_T of five different trainings of the BRT for MVATES: no FTK (only the HLT calorimeter variables), three FTK no split (adding in three of the track variables: total track p_T , average track p_T , and nTracks), and using six FTK no split variables (including the weighted track p_T), with training on 1- and 3-prong taus separately with two FTK variables added (adding in the total and average track p_T), and using five FTK variables (adding in the weighted track p_T for the leading three taus).	124
5.10	Ratio of calibrated momentum over truth momentum for weights trained on the even and odd (arbitrarily set) event numbers and then applied to both the input training set and the other unseen dataset. The comparable performance suggests that I am not over-training.	126
5.11	Correlation plot made in TMVA that shows which variables used in the training of the BRT supply similar information with regards to the target value for the regression. The values in yellow and blue are the ones most correlated and anti-correlated, respectively.	126
5.12	Comparison of performance of MVATES calibration when removing more or less correlated variables as described in section 5.6.4 with keeping all the variables. The best performance was the list of variables removing only 4 variables (without correlated variables.)	127

5.13	The left plot shows a comparison of performance of MVATES calibration using the BRT settings on the all tracks variable input list and changing the value of the MaxDepth while keeping NTrees constant. The right is a comparison of MVATES performance when changing the value of the NTrees while keeping MaxDepth at four or five. The 2018 weights are used as a benchmark	129
5.14	Comparison of performance of MVATES calibration with and without track variables and after tuning of the hyperparameters and input variable list.	130
5.15	Weighted distribution of jet p_T for all background samples, the legend refers to the p_T bin for the sample, as defined in table 5.5.	134
6.1	Cross section times branching ratio for all of the main production mechanisms and decays included in the current Run 2 data ($\sqrt{s} = 13\text{TeV}$) measurements of the Higgs boson, normalized to their SM predictions. This list is missing three primary Higgs boson production and decay combinations, including $VH, H \rightarrow \tau\tau$. The black error bars show the total uncertainties, which take into account both the systematic (in blue) and statistical (in yellow) contributions.	138
6.2	Leading order Feynman Diagram for Associated Production of the Higgs boson with a vector boson (W or Z).	139
6.3	Feynman diagram for the WH production mode where the Higgs boson decays to a pair of taus with one tau decaying hadronically and the other decaying leptonically.	141
6.4	Feynman diagram for the WH production mode where the Higgs boson decays to two hadronically decaying taus.	141
6.5	Feynman diagram for the ZH production mode where the Higgs boson decays to a hadronically decaying tau and a leptonically decaying tau.	142
6.6	Feynman diagram for the ZH production mode where the Higgs boson decays to two hadronically decaying taus.	142
6.7	Illustration of parent particles in an event and their associated visible and invisible products. The upstream visible momentum is there to include all visible products not part of one of the parent sets. [41]	149
7.1	Reconstructed Z mass peak for 2017 data and ATLAS Monte Carlo simulation for the $Z \rightarrow \mu\mu$ region.	158
7.2	Fake rate values for all four years of data in bins of p_T (left) and $ \eta $ (right) and their statistical uncertainties.	160
7.3	Fake Rates for electrons incorporating the MC truth subtraction from the data counts for selected and candidate electrons.	160
7.4	Fake rate values for muons for all four years of data in bins of p_T (left) and $ \eta $ (right) and their statistical uncertainties.	162
7.5	Fake rates for muons incorporating the MC truth subtraction from the data counts for selected and all candidate muons.	163
7.6	Fake rate values for taus for all four years of data in bins of p_T for 1-prong taus (left) and 3-prong taus (right) and their statistical uncertainties.	166
7.7	Fake rate values for taus for all four years of data combined in bins the two $ \eta $ bins for 1-prong and 3-prong taus and their statistical uncertainties (very small on this plot).	166
7.8	Fake rate values in the barrel region ($ \eta < 1.37$) for taus from all four years of data combined in bins for jet-width and p_T of the selected/anti-selected tau candidate.	167
7.9	Fake rate values in the end-cap region ($ \eta > 1.37$) for taus from all four years of data combined in bins for jet-width and p_T of the selected/anti-selected tau candidate.	168

8.1	General flow of data and simulation through our analysis. Everything goes through preselection and then each channel has its own signal and validation regions.	173
8.2	Some kinematic distributions in the WHlephad channel at preselection. The structure in the upper left hand plot is a result of the p_T binning in the tau fake rate. Some mismodeling can be seen in the lower right-hand plot of leading electron/muon p_T . However, these mismodeling effects are not seen in the validation region plots.	175
8.3	Some kinematic distributions in the $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ channel at preselection.	178
8.4	Some kinematic distributions in the ZHlephad channel at preselection.	180
8.5	Some kinematic distributions in the $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$ channel at preselection.	183
8.6	Distribution of the sum of light lepton and tau p_T for the signal MC (weighted by 10,000) and the same-sign fakes background for the WHlephad channel (left). The plot on the right shows the amount of signal MC and fake background remaining, and the significance, for each potential cut value on the sum of lepton p_T . This was used to determine where to place the cut in the WHlephad channel while maintaining high signal efficiency.	185
8.7	Distribution of the sum of light lepton and tau p_T for the signal MC (weighted by 100) and the non truth matched MC background, including W +jets, $Z \rightarrow ll$, $Z \rightarrow \tau\tau$ and other MC contributions for the WHlephad channel (left). The plot on the right shows the amount of signal MC and non-truth matched MC background remaining, and the significance, for each potential cut value on the sum of lepton p_T . This was used to determine where to place the cut in the WHlephad channel while maintaining high signal efficiency.	185
8.8	Distribution of the di-tau ΔR variable for the signal MC (weighted by 200) and the same-sign background used to determine the optimal cut window for the WHhadhad channel.	186
8.9	Distribution of the transverse mass variable for the signal MC (weighted by 200) and the same-sign background as an estimation of the contribution from the background due to mis-identified objects.	186
8.10	Distribution of the sum of tau p_T for the WHhadhad signal MC (weighted by 200) compared to the same-sign background to find the optimal cut placement (left). Distribution of the sum of tau p_T for the WHhadhad signal MC (weighted by 200) compared to the non truth matched Monte Carlo (as a background estimate) to find the optimal cut placement (right).	187
8.11	The plot on the left shows the amount of signal MC and same-sign background (due to misidentified jets) remaining, and the significance, for each potential cut value on the sum of tau p_T . This was used to determine where to place the cut in the WHhadhad channel while maintaining high signal efficiency. The plot on the right shows the amount of signal MC and background estimated from the non-truth matched MC, and the significance, for each potential cut value on the sum of tau p_T . This was used to determine where to place the cut in the WHhadhad channel while maintaining high signal efficiency.	187
8.12	Dilepton mass of our signal MC (weighted by 10) compared to the main background contributions. This distribution was used to determine the dilepton mass cut in the ZHlephad channel.	188

8.13	Distribution of the sum of tau p_T in the ZHlephad channel for the signal MC (weighted by 100) and the same-sign fakes contribution (left). The plot on the right shows the amount of signal MC and background from misreconstructed jets remaining, and the significance, for each potential cut value on the sum of the tau's p_T . This was used to determine where to place the cut in the ZHlephad channel while maintaining high signal efficiency. . . .	189
8.14	Dilepton mass of our signal MC (weighted by ten) compared to the main backgrounds modeled by the simulation. This distribution was used to determine the dilepton mass cut in the ZHhadhad channel.	190
8.15	Distribution of the sum of tau p_T in the ZHhadhad channel for the signal MC (weighted by 100) and the same-sign fakes contribution (left). The plot on the right shows the amount of signal MC and fake background remaining, and the significance, for each potential cut value on the sum of tau p_T . This was used to determine where to place the cut in the ZHhadhad channel while maintaining high signal efficiency.	190
8.16	Normalized distribution of MMC for the ZHlephad channel for signal and diboson Monte Carlo after signal selection (left) and of M_{2T} for WHlephad showing the challenge of using the Higgs boson mass reconstructions for signal separation.	191
8.17	Schematic of the neural network architecture with the input kinematics, rotational layers, three hidden fully-connected layers and the final output score.	192
8.18	Neural network score for the original WHlephad channel when applied to all three final states ('ee' - di-electron, 'emu' - electron & muon, 'mumu' - di-muon) simultaneously with clear signal and background separation (left) and on just the WHlephad_ee, di-electron, channel where the background (dibosons) has a flattened distribution and some of the separation has disappeared (right).	194
8.19	ROC curves for the training, test, and validation set for the WHlephad channel with the di-electron final state. The similar performance for all three samples demonstrates there is not significant over-training occurring.	195
8.20	Neural network score distribution at preselection for the WHlephad_ee (di-electron channel) test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).	195
8.21	The top left plot shows the ROC curves for different numbers of nodes in the three hidden layers, with 40 nodes having the best test set performance. The top right shows the improvement to the AUC when including in the new variables for training the WHlephad_ee network. The bottom left plot shows the effect of removing the rotational layers at the start of the network. The test performance is better when the layers are included. The bottom right plot shows the performance of two optimizers: SGD (our default) and Adam on the training and test sets.	196
8.22	Normalized distribution of M_{2T} for signal and diboson MC for full signal selection (left) and for the events passing a NN score cut of 0.5 (right) for the WHlephad_ee category.	197
8.23	ROC curves for the training, test, and validation set for the WHlephad channel with the electron/muon final state. There is a bit of over-training occurring, but test set performance is still comparable to our other models.	198
8.24	Neural network score distribution at preselection for the WHlephad_emu test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).	199

8.25	The top left plot shows the ROC curves for the baseline model and the optimized model with an abbreviated variable list for the WHlephad_emu channel. The top right shows the similar performance of the model when trained with different numbers of hidden layers. The bottom left plot shows the minimal effect of varying the number of nodes in the three layer model. The bottom right plot shows the similar performance of the model with various learning rates.	200
8.26	Normalized distribution of M_{2T} for signal and diboson MC for full signal selection (left) and for the events passing a NN score cut of 0.5 in the electron-muon channel of WHlephad (right).	200
8.27	Comparison of the baseline one-layer model and optimized three-layer model (right). ROC curves for the training, test, and validation set for the WHlephad channel with the di-muon final state (right) after adding in more discriminating variables.	201
8.28	NN scores on unweighted (left) and weighted (right) distributions of signal and diboson Monte Carlo for the di-muon WHlephad channel. While the signal distribution peaks at 1, the background distribution for NN score is flat.	201
8.29	ROC curves for the training, test, and validation set for the WHlephad channel with the di-muon final state after the same-sign lepton selection is performed. Lower training set sizes lead to the noise seen in the curves.	202
8.30	Neural network score distribution at preselection for the WHlephad_mumu test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).	202
8.31	Normalized distribution of M_{2T} for signal and diboson MC for full signal selection (left) and for the events passing a NN score cut of 0.5 for the di-muon WHlephad channel (right).	203
8.32	Neural network score distribution at preselection for the WHhadhad test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).	204
8.33	The ROC curves for the training and test sets for the ZHlephad neural networks trained on the full data set and the selected mass window.	205
8.34	Neural network score distribution at preselection (normalized and unweighted) for the ZHlephad test set on the full mass range (left) and neural network score for the mass window training set for signal and diboson MC after preselection (right).	205
8.35	Neural network score distribution at preselection for the ZHlephad test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).	206
8.36	Neural network score distribution at preselection for the ZHhadhad test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).	207
9.1	Plots of the NN score distribution in the ZHlephad with the up, down, and nominal variations of the tau energy scale uncertainties. The left plot shows the shape variation in the signal MC and the right shows the shape variation in the simulation of the ZZ background.	215
9.2	Plots of the NN score distribution in the WHlephad di-electron channel (left) and WHhadhad channel (right) with the up, down, and nominal variations of the fake MC subtraction on the background component from misidentification of jets.	216
9.3	Distribution of neural network score with two bins in the same-sign tau control region for the $ZH, H \rightarrow \tau_{had}\tau_{had}$ channel with ratio plot of data/background. This is used to calculate the closure uncertainty for the fake background contribution.	217

9.4	Plots of the NN score distribution in the WHhadhad channel (left) and ZHhadhad channel (right) with the up, down, and nominal variations of the theory uncertainty from varying μ_R and μ_F on the contribution from the signal MC.	218
-----	---	-----

Listing of tables

1.1	Relative strength and mediators with their properties for the four fundamental forces . . .	17
5.1	Branching ratios of some of the most common tau lepton decays. [4]	111
5.2	TMVA hyperparameters and their definitions that are used in training the BRT to perform MVATES. The initial value for each parameter is included in the final column.	119
5.3	Variables used in the training of the BRT to improve the Tau Energy Scale broken up into HLT Calorimeter Variables and the new Track Trigger Variables. The final column indicates whether they are used in the final, optimized training of the BRT.	121
5.4	Ranking of the 16 input training variables (or features) based on their importance in determining the final predicted value.	128
5.5	Relevant values for weighting the background samples, which are dijet samples with various binning in p_T with bin one having low p_T and bin seven having the highest p_T jets, and the signal sample for training the Lorentz equivariant neural network if more than one background sample is being used.	134
6.1	Monte Carlo generators used to produce the main simulated processes incorporated into the $VH, H \rightarrow \tau\tau$ analysis in Run 2, including the physics process, MC generator and the model of the underlying event with hadronization and parton showering (UEPS), the corresponding PDF set, and the perturbative order in QCD to which the cross-section has been calculated (LO = Leading order, NLO = Next-to-leading order, NNLO = Next-to-next-to-leading order).	144
7.1	Event types (using truth-level information), with their various possible reconstructions, and their probability to enter signal region. “F” designates a fake object, “T” designates a true object. “S” represents an object (which could be a fake or true object) that passes the selection criteria. “A” represents an object that fails the selection criteria. The right-hand column provides the probability for each event type. The numerical subscripts keep track of the individual objects that are considered.	155
7.2	Electron fake rates for each bin in p_T and $ \eta $ with their statistical and systematic uncertainties ($FR \pm \sigma_{stat} \pm \sigma_{syst}$).	161
7.3	Muon fake rates for each bin in p_T and $ \eta $ with their statistical and systematic uncertainties ($FR \pm \sigma_{stat} \pm \sigma_{syst}$)	164
7.4	Tau fake rates for each bin in p_T and jet-width with their statistical and systematic uncertainties ($FR \pm \sigma_{stat} \pm \sigma_{syst}$ for 1-prong taus in the barrel region, with $MET > 20\text{GeV}$).	169
7.5	Tau Fake Rates for each bin in p_T and jet-width with their statistical and systematic uncertainties ($FR \pm \sigma_{stat} \pm \sigma_{syst}$ for 1-prong taus in the barrel region, with $MET < 20\text{GeV}$).	169
7.6	Tau Fake Rates for each bin in p_T and jet-width with their statistical and systematic uncertainties ($FR \pm \sigma_{stat} \pm \sigma_{syst}$ for 3-prong taus in the barrel region, with $MET > 20\text{GeV}$).	170
7.7	Tau Fake Rates for each bin in p_T and jet-width with their statistical and systematic uncertainties ($FR \pm \sigma_{stat} \pm \sigma_{syst}$ for 3-prong taus in the barrel region, with $MET < 20\text{GeV}$).	170

8.1	Triggers used in the Run 2 analysis. This includes single muon, single electron, di-muon, di-electron, and e-mu. The trigger names use a convention where HLT stands for the high-level trigger and the number after mu or e is the online p_T threshold for the object. If there are two objects in the trigger than both online p_T thresholds are included in the name. If there is also an online identification working point used, i.e. loose or vlose (very loose), that is also included in the name. The final column of the table states the p_T threshold we applied additionally to the trigger selection in our offline processing.	174
8.2	Preselection for $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$ for the signal MC, main backgrounds, and data	174
8.3	Signal selection for $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$ for the signal MC and main backgrounds	176
8.4	Preselection yields for the $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ channel in the signal MC, main backgrounds, and data.	177
8.5	Signal region yields for the $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ channel in the signal MC and main backgrounds.	178
8.6	Preselection for $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$	180
8.7	Signal selection for $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$ with the signal MC and main backgrounds.	181
8.8	Preselection for $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$	182
8.9	Signal Region Cutoffs for $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$	182
8.10	Input variables for the neural network included in all channels and then for the specific channels.	193
9.1	Results from the Run 1 analysis for $VH, H \rightarrow \tau\tau$, including final observed counts (broken down into signal, diboson, and fake contributions) and the observed significance. [42]	209
9.2	Predictions for signal counts for each of the channels in the Run 2 analysis. The SM Prediction will serve as a higher bound as it does not take into account our analysis acceptance. We need to further investigate the unexpectedly low Run 2 Signal MC prediction for ZH-lephad.	210
9.3	Total cross-section uncertainties due to missing higher orders (QCD scale variations) from [43].	219
9.4	Total cross-section uncertainties due to PDFs and α_s from [43]	219

“It does not do to dwell on dreams and forget to live.”

J.K. Rowling, *Harry Potter and the Philosopher’s Stone*

O

Introduction

WE AS HUMANS HAVE ALWAYS SOUGHT TO LEARN MORE ABOUT THE WORLD WE LIVE IN. The fundamental questions of what exactly is the world around us made of and why are we here have remained ever present. While the latter may never be fully answered, the first is being tackled at numerous particle physics experiments throughout the world.

The Higgs boson, colloquially referred to as the ‘God particle’ because it is connected to how particles have mass, was discovered just before I started at the University of Chicago. I became involved in the ATLAS experiment at the end of my second year of undergrad. I was excited to join an international collaboration of thousands of scientists working together using one of the largest experimental apparatuses and computing systems in the world. The technical and engineering feat that is the ATLAS experiment is hard to fully fathom even after visiting the cavern 80 meters below the ground. The experiment has a physical grandeur that I have only felt matched by the Notre Dame Cathedral and the Sagrada Familia.

Collaborating with thousands of physicists throughout the world, requires a way of communicating effectively without being physically present. When the rest of the world went virtual in 2020, we were

ready, having already done most of our international meetings over video conference.

Another big change during my time on ATLAS is the incorporation of machine learning into both the running of the detector and the software used to analyze the data. Machine learning has helped improve our sensitivity and efficiency. It has been integrated into our trigger system to be able to select and store events at a 40 MHz frequency. It has helped us access signal events that would have been overwhelmed by background if we used our standard methods of selection. I believe machine learning's vital importance has shown us that communication and knowledge sharing between industry and academia is necessary for our success at making future discoveries.

The analysis of the associated production for the Higgs boson with a vector boson decaying to two tau particles is important to measure for many reasons. Because we are striving to build a complete picture of the Higgs boson, measuring the currently poorly-measured but well-motivated channels is essential. This channel is tricky to measure due to its two main backgrounds. Furthermore, the Higgs boson was originally discovered via its coupling to bosons. The tau is the heaviest fermion, making it a great candidate for studying fermionic couplings to the Higgs boson. Taus, themselves, are exciting particles to study. Taus decay almost immediately, making their identification and reconstruction a particular challenge. This challenge is one that is quite suitable for machine learning to solve.

In order to set the stage for my work and the main Higgs boson analysis in this thesis, I will first introduce the Standard Model of particle physics and how the Higgs boson fits in. Then I will explain how the ATLAS detector works and how we use software to reconstruct the raw data we collect into particles. I will then introduce machine learning and explain a few models used in both industry and particle physics research. I dedicate an entire chapter to the tau lepton and I will cover both its reconstruction and then my use of machine learning to build upon that. Afterwards, I will introduce the main analysis I have worked on: $VH, H \rightarrow \tau\tau$ with our four final state channels and mass reconstruction techniques. I will take you on a deep dive of how we calculate our fake background contribution and then through how I optimized our signal selection cuts and tuned a neural network to separate out our signal from our diboson backgrounds. Finally, I will show you the current state of our analysis and expected sensitivity projections.

*“Dust is only a name for what happens
when matter begins to understand itself.”*

— Philip Pullman, *The Amber Spyglass*

1

Theory Introduction

FOR THE PAST CENTURY, PHYSICISTS HAVE BEEN BUILDING A DICTIONARY FOR THE SUBATOMIC PARTICLES THAT THE UNIVERSE IS COMPOSED OF, SIMILAR TO THE PERIODIC TABLE OF ELEMENTS. This “dictionary” is known as the Standard Model and it attempts to be a unified picture of the particles and the forces between them, which are mediated by the exchange of particles. At this moment, experimental results continue to inform and expand our knowledge of the Standard Model. The last 50 years of experiments has led to a number of particle discoveries including the W and Z bosons, the top quark, which is the most massive known fundamental particle, and most recently, the Higgs boson, which was discovered at CERN in 2012. The Higgs boson discovery seemingly completed the Standard Model.

Before we dive into the development of a theory for how these fundamental particles interact, it is helpful to take a step back. All the matter that the human eye can see interacts via the laws of classical mechanics - think Newton’s $F = ma$. Very small particles, far too small for the human eye, interact via the theory of quantum mechanics, think Heisenberg’s uncertainty principle $\Delta x \Delta p \geq \frac{\hbar}{2}$. The physics for extremely fast objects, travelling close to the speed of light and described by relativistic mechanics, is gov-

erned by Einstein's theory of special relativity - $E = mc^2$. In order to advance the frontiers of knowledge, scientists have often decided to accelerate particles to extremely fast speeds. To combine this very fast, very small physics, one looks to relativistic quantum field theory.

The next few sections bring together ideas from a variety of sources, which the reader may reference for a deeper, mathematical understanding ([2], [44], [45], [46], and [47]).

1.1 THE STANDARD MODEL

The Standard Model (SM) describes both the fundamental particles and the interactions between them. The general SM structure is shown in figure 1.1. The SM includes two main types of particles: bosons and fermions. Within the fermions, there are quarks and leptons, which come in three generations (shown from left to right). The first generation are believed to be elementary particles meaning they are not made up of any other particles. In addition, for the quarks and charged leptons, there are two copies of each first generation particle which differ only in their mass. These are the second and third generation particles which are more massive and unstable. They quickly decay into other particles in the SM. On the right side of figure 1.1 are the bosons, or the force carriers; these will be described in detail in section 1.1.2. The other main component of the Standard Model are the forces through which the particles interact. The strong, weak, and electromagnetic forces are included in the Standard Model.

1.1.1 FERMIONS

Fermions in the SM are the spin $\frac{1}{2}$ particles. All fundamental SM fermions experience the weak force. They are the building blocks of the matter we see in the universe. There are two main categories of fermions: *leptons* and *quarks*. The leptons can exist as free particles and are known as electrons (e), muons (μ), taus (τ), and their corresponding neutrinos: (ν_e , ν_μ , and ν_τ). The electron, muon, and tau all have a charge of e , while the neutrinos are chargeless. Since the neutrinos do not have charge, they do not experience the electromagnetic force, while the other leptons do. Neutrinos also have an extremely small mass that is not accounted for in the SM. All six of these leptons also have an anti-particle, with the opposite charge and lepton number, resulting in a total of 12 leptons.

Quarks are different from fermions as they also carry a color charge in addition to a fractional electric charge ($\pm\frac{1}{3}, \pm\frac{2}{3}$). Thus quarks also interact via the strong force. There are six types of quarks: up (u),

Standard Model of Elementary Particles

	three generations of matter (fermions)			interactions / force carriers (bosons)	
	I	II	III		
mass	$\approx 2.2 \text{ MeV}/c^2$	$\approx 1.28 \text{ GeV}/c^2$	$\approx 173.1 \text{ GeV}/c^2$	0	$\approx 124.97 \text{ GeV}/c^2$
charge	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	0
spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
QUARKS	u up	c charm	t top	g gluon	H higgs
	d down	s strange	b bottom	γ photon	
	e electron	μ muon	τ tau	Z Z boson	
LEPTONS	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson	

Figure 1.1: Standard Model Table formulation with three generations of quarks and leptons, the Gauge bosons, and Higgs boson along with their mass, charge, and spin. [1]

charm (c), top (t), down (d), strange (s), and bottom (b). Each quark can come in one of three colors: blue, green, and red. Quarks must follow the law of color confinement, which requires that their interactions remain color neutral as all naturally occurring particles are colorless (defined as having total amount of each color being zero or all three colors existing equally). Thus there are two main bound states of quarks: baryons ($qq\bar{q}$) and mesons ($qqq, \bar{q}\bar{q}\bar{q}$). The well known proton (uud) and neutron (ddu) are bound states of the lightest quarks (u and d) with a charge of $+1$ and 0 respectively. In total, there are 36 quarks when including the six types, three colors, and their antiparticles.

The currently known stable, fundamental particles in the universe include the electron, up quark, and down quark.

1.1.2 BOSONS

There are two main types of fundamental bosons based on their spin: *Vector bosons* ($S=1$) and *Scalar bosons* ($S=0$). Bosons are the particles that mediate the forces described in section 1.1.3, otherwise known as the “force carriers”. The photon (γ) and gluon (g) are massless. The photon mediates the electromagnetic forces and the gluon mediates the strong force. Meanwhile, the $W^\pm$ and Z^0 are massive and mediate the weak force. The γ , g , $W^\pm$, and Z^0 make up the vector bosons, while the Higgs boson is the sole scalar boson. The Higgs boson is not a force carrier like the vector bosons, but it is associated with the mechanism for giving particles mass.

1.1.3 FORCES

There are four fundamental forces through which the elementary particles interact (e.g. exchange of bosons between fermions). These four forces are the strong, electromagnetic, weak and gravitational force. Three of these forces is described by a quantum field theory. There is still no quantum field theory of gravity, and at the moment, the graviton is the theorized force carrier for the gravitational force. Table 1.1 describes the forces’ relative strengths and mediators, with the caveat that the strength of these forces changes with the scale of the interaction.

Force	Strength	Mediator Boson	Spin	Mass (GeV)
Strong	1	Gluon g	1	0
Electromagnetic	10^{-3}	Photon γ	1	0
Weak	10^{-8}	$W^\pm$ and Z	1, 1	80.4, 91.2
Gravity	10^{-42}	graviton?	2	0

Table 1.1: Relative strength and mediators with their properties for the four fundamental forces

The strongest force is aptly named the *strong* force. The strong force is what binds together quarks in the neutron and proton, and then again binds those neutrons and protons together in a nuclei. *Electromagnetic* interactions are responsible for most interactions between charged fermions, such as the transfer of momentum from one electron to another. The well known rule that opposite charges attract and like charges repel is from the electromagnetic force. The *weak* force is responsible for particle decays and processes such as nuclear β -decay and nuclear fusion. The final force is *gravity* (what attracts massive objects to each other), and it is so much weaker than the other forces that it is challenging for us to investigate in

our current collider experiments. Gravity is not currently incorporated in the SM.

1.2 INTERACTIONS BETWEEN PARTICLES

Section 1.1.3 introduced the main forces and stated that fermions interact through bosonic force mediators. In this section, some other conservation laws that particle interactions and decays follow are introduced, and the interactions of the fundamental forces are further described.

1.2.1 CONSERVATION LAWS

Based on our current understanding of the laws of nature and knowledge gleaned from experiment, there are several conservation laws particle interactions follow:

- **Energy** in the form of Potential, Kinetic, Mechanical etc. must be conserved during an interaction, such as a particle decay. This means more massive particles will decay into less massive particles, as allowed based on other conservation laws.
- **Momentum** of particles in the initial state must equal the momentum of the final-state particles. This rule is used in particle collisions to reconstruct events from a common interaction point.
- **Electric charge** of particles must be the same total sum before and after an interaction.
- **Color Charge** must be conserved during a strong interaction (weak and electromagnetic interactions do not affect color). Color charge can be carried by quarks and/or gluons emerging from an interaction (see section 1.2.2).
- **Lepton number** must be conserved. This means that there have to be the same number and type of leptons going in and out of an interaction with anti-leptons adding negative counts to the total number of leptons. Lepton number is not affected in electromagnetic interactions as those do not change the particle type and leptons do not interact via the strong force.
- **Baryon number** is also conserved, i.e. if a quark goes in, a quark comes out. Similar to when we count leptons in a system, we count anti-quarks as negative contributions to the total. Each quark contributes $\frac{1}{3}$ the baryon number.

1.2.2 FEYNMAN DIAGRAMS FOR FUNDAMENTAL INTERACTIONS

These conservation laws and their allowed particle interactions are easily viewed through the formalism known as Feynman diagrams. These visual descriptions of particles show ingoing (left-most) and outgoing

(right-most) particles and include internal lines that are the propagators that mediate the interaction. Since the propagators are completely internal to the process, they need not have the rest mass of a real, free particle. When this occurs they are referred to as virtual particles. The time direction goes from left to right and the arrows indicate whether there is a particle (goes with time) or anti-particle (goes backwards with respect to time). Thus a positron going forwards is the same as an electron going backwards. Each particle has a different line style: fermions (e^- , q) are straight lines with arrows, gluons are curly lines, and W/Z bosons and photons (γ) are wavy lines.

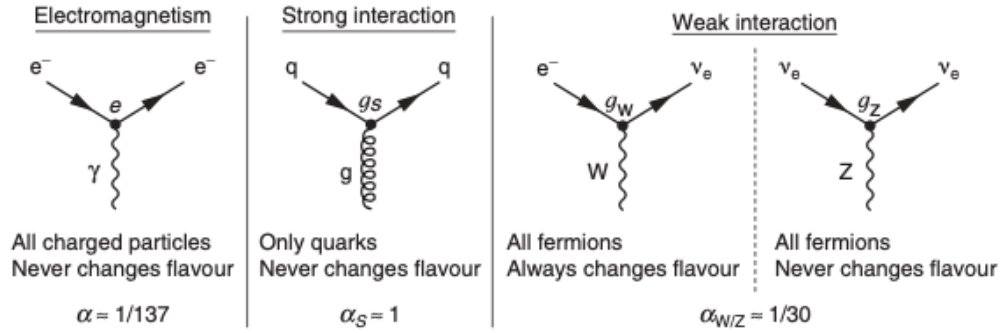


Figure 1.2: The main SM interaction vertices for QED, QCD, and the Weak force. The coupling strength is denoted by α [2]

The Standard Model interaction vertices in figure 1.2 show the couplings of gauge bosons to fermions (quarks and leptons). Particles can only couple to a force mediating boson if they have the charge of the interaction. One way to compare the relative strengths of each force interaction is to look at the coupling strength g . For electromagnetic interactions (QED) the coupling strength is given by the charge of an electron, $g_{QED} = e$. Each vertex shown in figure 1.2 contributes a factor of g to the matrix element of the scattering process. Conventionally, we use the associated dimensionless constant, $\alpha \propto g^2$, in place of the coupling constant. α is also known as the fine structure constant, where in the case of electromagnetic interactions: $\alpha_{EM} = \frac{e^2}{4\pi\epsilon_0\hbar c} = \frac{1}{137}$. Spin-orbit splitting in atomic spectra is defined by the fine structure constant. Now, we have a factor of α for each vertex included in the quantum-mechanical probability for the interaction. The QCD interaction is stronger than the QED with $\alpha_s = 1$ and the weak interaction is in between QED and QCD, with $\alpha_W = \frac{1}{30}$. The W boson's large mass suppresses the strength of the weak force. Another way to express α_w is $\frac{\alpha}{(M_w/m_p)^2}$, which directly shows the diminishing in coupling strength based on the W boson mass.

Because the coupling constant α is added for each vertex in a Feynman diagram, when there are vertices, an interaction is less likely to occur. Feynman diagrams with internal loops represent “higher order” interactions.

1.3 LAGRANGIAN FORMALISM OF THE STANDARD MODEL

In the next few subsections, I work through the construction of the SM Lagrangian from the fundamental forces and introduce the need for the Higgs mechanism and resulting Higgs boson. This mathematical formalism is the result of a combination of mathematical theory, experimental results, and our understanding of the universe.

1.3.1 THE DIRAC INTERACTION

Lagrangians are used to describe the interactions of particles and fields. Lagrangians sum up the energy in a system and can describe all the dynamics of the system. In classical mechanics, they are generally written as $L = T - V$ where T is the kinetic energy and V is the potential energy of the system. When moving to field theory, the Lagrangian becomes a function of fields and their derivatives:

$$L = \int L(\phi, \partial_\mu \phi) dx^\mu. \quad (1.1)$$

One can then describe the motion of the relativistic fields using the Euler-Lagrange equation:

$$\partial_\mu \left(\frac{\partial L}{\partial(\partial_\mu \phi)} \right) - \frac{\partial L}{\partial \phi} = 0. \quad (1.2)$$

The Lagrangians for the forces described in the following sections originate with the Dirac equation. The Dirac equation is a wave equation explaining the interactions of fermions that solved the problems of negative probability densities in the Klein-Gordon equation ($(\partial^\mu \partial_\mu + m^2)\psi = 0$), which was an initial attempt to have a Lorentz invariant relativistic theory of quantum mechanics. The Dirac equation is first order in both space and time derivatives and upholds Einstein’s energy-momentum relationship ($E^2 = \mathbf{p}^2 + m^2$). In its covariant form, the Dirac equation for fermions with mass m states:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \quad (1.3)$$

and μ is the Lorentz index of the four-vector and the γ -matrices in Dirac-Pauli representation are:

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \text{ and } \gamma^k = \begin{pmatrix} 0 & \sigma_k \\ -\sigma_k & 0 \end{pmatrix} \quad (1.4)$$

$$\text{where } \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \text{ and } \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (1.5)$$

These gamma matrices satisfy the anti-commutation relations, such that $\{\gamma^i, \gamma^j\} = -2\delta^{ij}$. The solutions to the Dirac equation are free-particle plane waves of the form:

$$\psi(\mathbf{x}, t) = u(E, \mathbf{p})e^{i(\mathbf{p} \cdot \mathbf{x} - Et)} \quad (1.6)$$

where ψ obeys the Dirac equation and $u(E, \mathbf{p})$ is a four-component Dirac spinor. The solutions for the Dirac equation are the relativistic QM description of fundamental spin- $\frac{1}{2}$ particles and the solutions for $E < 0$ are representative of antiparticles. The Lagrangian for the Dirac interaction describing ψ is:

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi. \quad (1.7)$$

Using the continuity equation $\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0 = \partial_\mu j^\mu$, one can express the four-vector current as $j^\mu = \bar{\psi}\gamma^\mu\psi$, where $\bar{\psi} \equiv \psi^\dagger\gamma^0$ is the Dirac adjoint spinor. The current becoming $j^\mu = \bar{\psi}\gamma^\mu\psi$ comes from the internal symmetry of $\psi \rightarrow e^{-i\alpha}\psi$. We can also express the probability density as $\bar{\psi}\psi$. Furthermore, Noether's Theorem states that any symmetry of the Lagrangian implies the existence of a conserved current. Enforcing gauge invariance for the various forces will add terms to this Lagrangian.

1.3.2 INTRODUCTION TO GAUGE INVARIANCE

A clear way to see gauge invariance at play is to look at its use in electromagnetism. Classical electromagnetism is defined by Maxwell's equations:

$$\nabla \cdot \mathbf{E} = \rho, \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \nabla \cdot \mathbf{B} = 0, \text{ and } \nabla \times \mathbf{B} = \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}. \quad (1.8)$$

Because charge is conserved in all frames, $j^\mu = (\rho, \mathbf{J})$ is a four vector and $\partial_\mu j^\mu = 0$. $\mathbf{B}$ and $\mathbf{E}$ can then be expressed in terms of a four-vector potential $A^\mu = (\phi, \mathbf{A})$ and then Maxwell's equations can be condensed into $\partial_\mu F^{\mu\nu} = j^\nu$. $F^{\mu\nu}$ is known as the Maxwell field strength tensor and is defined as

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix} \quad (1.9)$$

The $\mathbf{E}$ and $\mathbf{B}$ fields are unchanged by this gauge transformation: $\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla\chi$ and $\phi \rightarrow \phi' - \frac{\partial\chi}{\partial t}$. Here $\chi(\mathbf{x}, t)$ can be any finite differentiable function. Standardly, the Lorenz gauge ($\partial^\mu A_\mu = 0$) is chosen as it simplifies the covariant formulation. The fields and thus Maxwell's equations remain unchanged after this transformation, so they have a gauge symmetry.

In particle physics a gauge theory is a field theory whose Lagrangian is invariant under local transformation between Lie groups (differential manifolds) and there exists a group of these transformations, or symmetries, of the field that leave the physics invariant. Global gauge symmetries must be invariant under a transformation at all points in spacetime. The gauge symmetries also have redundancies or extra degrees of freedom. Local gauge symmetries are not invariant at all points in space and a specific basis must be defined to describe them. Two gauge fields are then related by a transformation on this coordinate basis. It is also key to note that with local gauge invariance the derivatives acting on the field also act on the local phase and we pick up an extra term, necessitating the need for a covariant derivative.

1.3.3 QUANTUM ELECTRODYNAMICS (QED)

The quantum field theory for the electromagnetic interaction is Quantum Electrodynamics (QED). We can obtain the field theory of QED directly from the free fermion Lagrangian by requiring local phase invariance. This interaction is mediated by the photon, and the free photon field is $A_\mu = \varepsilon_\mu^{(\lambda)} e^{i(p \cdot x - Et)}$ where ε is a four-vector and λ denotes the polarization of the field. The polarisation of the photon is transverse to its direction of motion. To get the interaction between a free particle and the photon (EM field), we add an additional gauge invariance by including spinor phase transformations $\psi \rightarrow e^{iq\chi(x)}$. Now, the Dirac equation is modified by introducing the covariant derivative: $\partial_\mu \rightarrow \partial_\mu + iqA_\mu$, where

$A_\mu = (\phi, -\mathbf{A})$. This introduces an extra vector field and to compensate we need A_μ to transform under local gauge transformation by $A_\mu \rightarrow A_\mu + \partial_\mu \chi$ where the derivative cancels the extra term. A_μ is a field corresponding to the massless gauge boson, in this case the photon. A modified Dirac equation including an interaction term $(q\gamma^\mu A_\mu \psi)$ incorporates this change: $i\gamma^\mu (\partial_\mu + iqA_\mu)\psi - m\psi = 0$, where $D_\mu \rightarrow \partial_\mu + iqA_\mu$.

Now the Lagrangian with the incorporation of QED has become:

$$\begin{aligned}\mathcal{L}_{\text{QED}} &= \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \\ &= \bar{\psi}i\gamma^\mu \partial_\mu \psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - q\bar{\psi}\gamma^\mu A_\mu \psi - m\bar{\psi}\psi\end{aligned}\tag{1.10}$$

where $F^{\mu\nu}$ is from equation 1.9. In summary, QED follows a local U(1) gauge symmetry. It is mediated by a massless photon, which is the generator of that symmetry, and electric charge is conserved.

1.3.4 QUANTUM CHROMODYNAMICS (QCD)

Quantum Chromodynamics (QCD) is the QFT description of the strong force. Its underlying symmetry is invariant under SU(3) local phase transformations of the form $\psi(x) \rightarrow \psi'(x) = e^{ig_s \alpha(x) \hat{\mathbf{T}}} \psi(x)$. The eight generators of the SU(3) symmetry group are $\hat{T}$, where $T^a = \frac{1}{2}\lambda^a$ in relation to the Gell-Mann matrices. The Gell-Mann matrices are all 3x3, meaning ψ must now have three new degrees of freedom. By convention, these are called color charges that come in flavors of red, green, and blue. Color charge must be conserved in QCD.

The local gauge invariance for QCD requires the introduction of eight fields: $G_\mu^a(x)$. The covariant derivative for this transformation is $D_\mu \rightarrow \partial_\mu + ig_s G_\mu^a T^a$, where g_s is the coupling strength of the gluon field. The Dirac equation now becomes $i\gamma^\mu (\partial_\mu + ig_s G_\mu^a T^a)\psi - m\psi = 0$. The gluon fields will then transform as $G_\mu^k \rightarrow G_\mu^{k'} = G_\mu^k - \partial_\mu \alpha_k - g_s f_{ijk} \alpha_i G_\mu^j$, where f_{ijk} are structure constants of SU(3) and they follow the commutation relation $[\lambda_i, \lambda_j] = 2if_{ijk}\lambda_k$. The result of them not commuting is the additional term in the transformed fields, and this describes gluon self-interactions. Similar to the electromagnetic field strength tensor, $G_{\mu\nu}^a$ is the gauge invariant gluon field strength tensor with $G_{\mu\nu}^a = \partial^\mu G_\nu^a - \partial^\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c$ where G_μ^a are the gluon fields. The QCD Lagrangian is then:

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_i(i\gamma_\mu \partial_\mu \delta_{ij} - m\delta_{ij})\psi_j - g_s \bar{\psi} \gamma_\mu (T_a)_{ij} \psi G_\mu^a - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}\tag{1.11}$$

All free particles are colorless. The carrier of the strong force, gluons, must group other particles together to form colorless hadrons. If the symmetry was under $U(3)$ and not $SU(3)$ there would be a ninth gluon ($\frac{1}{\sqrt{3}}(r\bar{r} + g\bar{g} + b\bar{b})$) but this would be colorless and would imply the existence of an infinite-range strong force. Evidently, this is not the world we live in, confirming the $SU(3)$ symmetry of QCD. Like in QED and its massless photon, the gluons must also be massless to preserve gauge invariance. Furthermore, QCD is a non-Abelian gauge theory due to the self-interaction term in the fields, while QED is Abelian and the photon cannot self-interact.

1.3.5 WEAK INTERACTION

The weak interaction is a charged current interaction that is mediated by charged $W^\pm$ bosons. Unlike in QED and QCD, parity (spatial inversion through the origin) is not conserved in weak interactions. This was confirmed experimentally with the nuclear β -decay of cobalt-60 when electrons were measured to be preferentially emitted in one direction. Both QED and QCD have four-vector currents ($j^\mu = \bar{u}(p')\gamma^\mu u(p)$), while the general form of the weak interaction current can be written as $\bar{u}(p')\Gamma u(p)$ and Γ is 4×4 matrix of products of Dirac γ -matrices. Based on experimental evidence, the weak charged current for $W^\pm$ exchanges is a V-A (vector minus axial current) with vertex factor of $\frac{-ig_W}{\sqrt{2}} \frac{1}{2} \gamma^\mu (1 - \gamma^5)$.

The $(1 - \gamma^5)$ term is key for understanding the handedness of fermions. To review, helicity states whether the particle's spin is aligned with its direction of motion and chirality is determined by whether a particle transforms in a right-handed or left-handed way. For massless particles, chirality is the same as helicity, but it is not the same for massive particles like the fermions. The right- and left-handed chiral projection operators are $P_R = \frac{1}{2}(1 + \gamma^5)$ and $P_L = \frac{1}{2}(1 - \gamma^5)$. Spinors $u(p)$ can be written as a sum of right- and left-handed states. However, when looking at the interaction between chiral states and the weak vertex, only the left-handed spinors have non-zero currents. Thus only left-handed chiral particle states and right-handed chiral anti-particle states participate in the charged weak interaction.

While QED is a $U(1)$ symmetry and QCD is a $SU(3)$ symmetry, the weak force is described by an $SU(2)_L$ symmetry, as right-handed particles and left-handed antiparticles are unaffected by $SU(2)$. The phase transformations for $SU(2)$ look like: $\phi(x) \rightarrow \phi(x)' = e^{ig_w \alpha(x) \mathbf{T}} \phi(x)$, where $\alpha(x)$ refers to the local phase at each point in space time and $\mathbf{T} = \frac{1}{2} \sigma$ are the three generators of the group. Since the generators are 2×2 matrices, one can write ϕ in terms of two components known as the weak isospin doublet which

differ by one charge unit from each other, i.e. $\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}$ and $\begin{pmatrix} u \\ d' \end{pmatrix}$. This local gauge invariance is maintained by the introduction of three gauge bosons: W_μ^k . Two of these relate to the physical $W^\pm$ boson by $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^{(1)} \mp iW_\mu^{(2)})$. The third field is an independent, weak neutral current.

1.3.6 ELECTROWEAK UNIFICATION

You may now be asking, is the Z boson this third of the three gauge bosons in the local gauge invariance for the weak interaction? Experimentally we know that the Z boson couples with both right- and left-handed particles. Another, option for this weak neutral current is the photon. It was Glashow, Weinberg, and Salam (GWS) that proposed a unified picture incorporating both the Z boson and photon. They introduced the idea of hypercharge, $Y = 2(Q - I_W^{(3)})$ and promoted the $U(1)$ symmetry of QED to a $U(1)_Y$ symmetry. This is done with a gauge transformation of the form: $\psi(x) \rightarrow \psi'(x) = \hat{U}(x)\psi(x) = e^{-ig'\frac{Y}{2}\xi(x)}\psi(x)$, which introduces a new gauge field B_μ that couples to the new weak hypercharge. The covariant derivative becomes $D_\mu = \partial_\mu - i\frac{g'}{2}YB_\mu - i\frac{g}{2}T_jW_\mu^j$, where Y is the hypercharge and T_j are the isospin components. The electroweak Lagrangian taking into account the difference between right- and left-handed fermions is then:

$$\begin{aligned} \mathcal{L}_{EW} &= -\frac{1}{4}W_j^{\mu\nu}W_{\mu\nu}^j - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + L_{CC} + L_{NC} \\ &= -\frac{1}{4}W_j^{\mu\nu}W_{\mu\nu}^j + \bar{\psi}_L\gamma^\mu(\partial_\mu - i\frac{g'}{2}YB_\mu - i\frac{g}{2}T_jW_\mu^j)\psi_L + \bar{\psi}_R\gamma^\mu(\partial_\mu - i\frac{g'}{2}YB_\mu)\psi_R \end{aligned} \quad (1.12)$$

Now that we have this new gauge field B_μ , γ and the Z boson can be written as a linear combination of the B_μ and the neutral $W_\mu^{(3)}$:

$$A_\mu = +B_\mu\cos\theta_W + W_\mu^{(3)}\sin\theta_W \quad (1.13)$$

$$Z_\mu = -B_\mu\sin\theta_W + W_\mu^{(3)}\cos\theta_W \quad (1.14)$$

where θ_W is the weak mixing angle, which can be expressed in relation to the electric charge by $e = g_W\sin\theta_W$. θ_W has been experimentally determined, and using that value, one can determine how the Z and $W^\pm$ will couple to other particles. This combines both the charged weak interactions (CC) of the

$W^\pm$ to fermions of the same generation and quarks of any generation (though mixed generation couplings are much weaker as shown by the CKM matrix), and the neutral weak interaction (NC) of the Z boson that couples fermions to themselves.

There are no terms involving fermion mass in the Lagrangian for the EW theory and in order for gauge invariance to hold, the bosons must all be massless; problems we will address later. While electroweak unification completes the picture of QED and the weak interaction, it leaves several questions open: Why do the $W^\pm$ and Z have large mass while the photon and gluons do not and when is this spontaneous symmetry broken?

1.4 STANDARD MODEL LAGRANGIAN

Combining the above theories of electroweak unification and QCD interactions, with their coupling constants remaining separate, we can arrive at the full symmetry of the Standard Model: $SU(3) \times SU(2) \times U(1)$. You may now think we are done explaining the SM, however the $SU(2)_L$ gauge transformation and its introduction of a mass term for fermions breaks gauge invariance. Lucky for us, the introduction of a Higgs field from the Brout-Englert-Higgs mechanism solves this problem. We will discuss this and $\mathcal{L}_{Higgs}$ in detail in section 1.5. Including this new term, the full standard model Lagrangian can be written succinctly as:

$$\begin{aligned}\mathcal{L}_{SM} &= \mathcal{L}_{gauge\ fields} + \mathcal{L}_{fermions} + \mathcal{L}_{Higgs} \\ &= -\frac{1}{4} \sum_{gauge} F_{\mu\nu}^a F^{a\mu\nu} - \sum_f \bar{f} \gamma^\mu D_\mu f + \mathcal{L}_{Yukawa} - \frac{1}{2} (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \mu^2 (\phi^\dagger \phi) + \lambda (\phi^\dagger \phi)^2\end{aligned}\tag{1.15}$$

where $F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c$, and A_μ is one of gauge fields, g is its coupling parameter, D_μ is the covariant derivative combining the changes for $U(1)_Y$, $SU(2)_L$, and $SU(3)$, and $\mathcal{L}_{Yukawa}$ refers to the fermion fields. Reducing the notation, this is succinct enough to fit on a coffee mug (figure 1.3)!

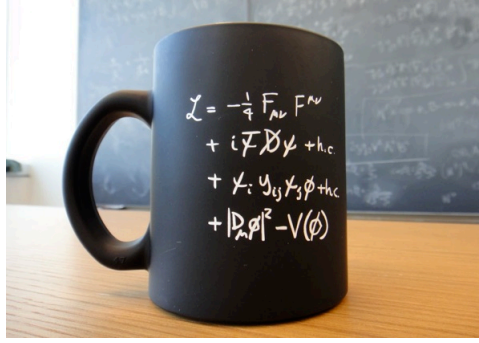


Figure 1.3: The Standard Model Lagrangian mug sold at CERN.

1.5 HIGGS BOSON

The Higgs mechanism combines both the Higgs field, which gives other particles mass, and the Higgs boson, a spin 0, neutral, massive particle. The spontaneous symmetry breaking of the vacuum state of the $\mathcal{L}_{SM}$ creates a massless field and three massive gauge bosons that solves both problems of massless force mediators and gauge invariance with the fermion mass terms. This reduces the symmetry of the electroweak model from $SU(2)_L \times U(1)_Y$ to $U(1)_{EM}$.

1.5.1 ELECTROWEAK SYMMETRY BREAKING

To begin to understand spontaneous symmetry breaking, let's look at the Lagrangian for a scalar field (you may see this resembles the $\mathcal{L}_{Higgs}$ in equation 1.15):

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}\mu^2 \phi^2 - \frac{1}{4}\lambda \phi^4 \quad (1.16)$$

where the first term represents the kinetic energy and the ϕ^2 term is the mass, and the ϕ^4 term is the self-interactions of the scalar field. To find the lowest energy state, we want to find the minimum of $V(\phi)$ using $\frac{\partial V}{\partial \phi} = 0$. There are no restrictions on the sign of μ^2 , while λ must be greater than zero. Thus there are two possible scenarios:

- $\mu^2 > 0$: $\phi = 0$ and vacuum expectation value ν is non-zero
- $\mu^2 < 0$: $\phi = \pm \nu = \pm \sqrt{\frac{-\mu^2}{\lambda}}$, giving two degenerate vacuum states

In the second case ($\mu^2 < 0$), the act of choosing a vacuum state is the spontaneous symmetry breaking. The next step is to look at excitations around the the vacuum. Introducing $\phi(x) = v + \eta(x)$ to the $\mathcal{L}$ in

1.16 as ϕ , creates a mass term of the form $\lambda\nu^2\eta^2$, where the mass can then be solved as $m_\eta = \sqrt{2\lambda\nu^2} = \sqrt{-2\mu^2}$. Thus this spontaneous symmetry breaking shows that $\mathcal{L}$ describes a massive scalar field.

Now, the Higgs mechanism comes from the spontaneous symmetry breaking of a *complex* scalar field. The field ϕ becomes $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$, and the resulting Lagrangian is

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^*(\partial^\mu\phi) - \frac{1}{2}\mu^2\phi^*\phi - \frac{1}{4}\lambda(\phi^*\phi)^2 \quad (1.17)$$

which is really just a slight change of notation from equation 1.16. Again, when finding the minima of $V(\phi)$, both $\mu^2 < 0$ and $\mu^2 > 0$ solutions are considered. For $\mu^2 > 0$, again we have a stable equilibrium if $\phi = 0$, and for $\mu^2 < 0$, there are infinitely many solutions corresponding to $\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda}$. Figure 1.4(a) shows the potential for a complex scalar field ϕ with $\mu^2 > 0$ where $SU(2)$ symmetry is preserved and there is just one stable solution for minimizing the potential. Figure 1.4(b), also known as the Mexican hat potential, is the potential of a scalar field for $\mu^2 < 0$ when $SU(2)$ symmetry is broken and there are infinitely many possible minimizing solutions, represented by the circle at the minima of the potential.

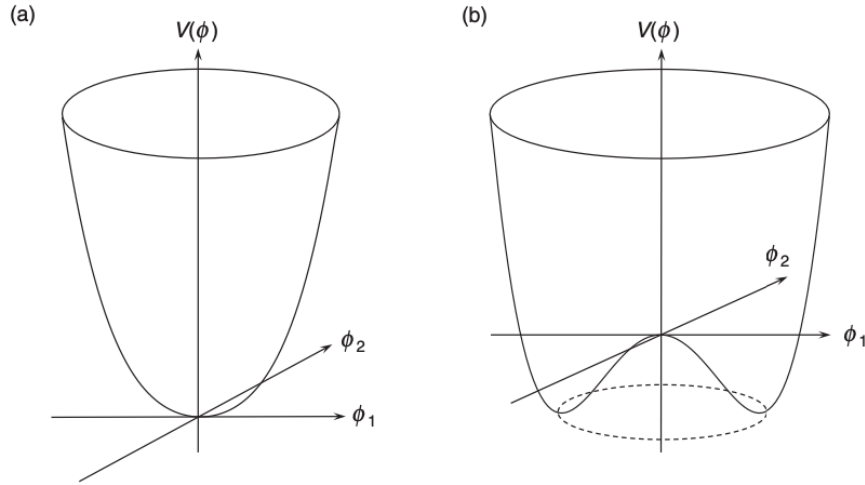


Figure 1.4: An illustrative example of spontaneous symmetry breaking for a complex scalar field potential $V(\phi)$, with (a) showing solutions for $\mu^2 > 0$ and (b) for $\mu^2 < 0$. [2]

1.5.2 HIGGS MECHANISM

$\mathcal{L}_{SM}$ is required to be invariant under local gauge transformations. However, as it is written this complex ϕ is not invariant under a $U(1)$ local gauge transformation. This local gauge invariance can be recovered (as

it was in QED) with the introduction of a covariant derivative: $\partial_\mu \rightarrow D_\mu = \partial_\mu + igB_\mu$. Incorporating the required transformation of the new gauge field $B_\mu \rightarrow B'_\mu = B_\mu - \partial_\mu \chi(x)$, the Lagrangian becomes:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + (D_\mu\phi)^*(D^\mu\phi) - \mu^2\phi^2 - \lambda\phi^4 \quad (1.18)$$

where $F^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu$ is the kinetic term. B must also be a massless gauge field or its mass term would break gauge invariance. This Lagrangian has a degenerate vacuum state for $\mu^2 < 0$ and including the perturbations around the vacuum state $\phi(x) = \frac{1}{\sqrt{2}}(\nu + \eta(x) + i\xi(x))$ results in a Lagrangian that includes a massive scalar field η , massless Goldstone boson ξ , and a mass term for the massless gauge field B . Success!.... or rather not quite.

This Lagrangian includes a direct coupling between the Goldstone field and gauge field implying a spin-1 gauge field can interact with a spin-0 field, which is not possible. A clever choice of the gauge transformation $B_\mu \rightarrow B'_\mu = B_\mu + \frac{1}{g\nu}\partial_\mu\xi(x)$ removes the Goldstone field from the resulting Lagrangian. This is known as the *Unitary gauge* and makes the complex scalar field $\phi(x)$ real:

$$\phi(x) \equiv \frac{1}{\sqrt{2}}(\nu + h(x)) \quad (1.19)$$

with $h(x)$ as the Higgs field. The degree of freedom from the Goldstone boson is now related to the polarization of the massive gauge potential. Using this gauge, the Higgs Lagrangian becomes:

$$\mathcal{L}_H = \frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \lambda\nu^2 h^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2\nu^2 B_\mu B^\mu + g^2\nu B_\mu B^\mu h + \frac{1}{2}g^2 B_\mu B^\mu h^2 - \lambda\nu h^3 - \frac{1}{4}\lambda h^4. \quad (1.20)$$

The first two terms represent the massive scalar Higgs field, the third and fourth represent the massive gauge boson, and last two terms show h self-interactions. The other terms are the interactions between h and B . Looking at the mass terms, we can get the mass of the Higgs boson as $m_H = \sqrt{2\lambda}\nu$ and the mass of the gauge boson as $m_b = g\nu$.

1.5.3 HIGGS BOSON AND MASS

The Higgs mechanism needs a neutral scalar field in order to create the masses of the electroweak gauge bosons. The Higgs field can be written in a weak isospin doublet as $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$. Following the steps in the previous section, we can find a non-zero vacuum expectation value of $\langle \phi \rangle = \begin{pmatrix} 0 \\ \frac{\nu}{\sqrt{2}} \end{pmatrix}$. After spontaneous symmetry breaking, the Higgs boson itself can be written in a doublet, like the left handed spinors for electroweak theory:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} \quad (1.21)$$

Using $Y = 2(Q - I_W^{(3)})$, and the fact that the lower component of the Higgs doublet is neutral, the overall Higgs doublet has $Y = 1$.

We can now determine the gauge boson masses by using the covariant derivative for the electroweak interaction:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_w \mathbf{T} \cdot \mathbf{W}_\mu + ig' \frac{Y}{2} B_\mu \quad (1.22)$$

and look at the $(D_\mu \phi)^\dagger (D^\mu \phi)$ terms in the Lagrangian. The mass terms are related to quadratic powers of the gauge boson fields and reduce to:

$$\frac{1}{8} g_w^2 \nu^2 (W_\mu^{(1)} W^{(1)\mu} + W_\mu^{(2)} W^{(2)\mu}) + \frac{1}{8} \nu^2 (g_W W_\mu^{(3)} - g' B_\mu)^2. \quad (1.23)$$

From these, we can get that the mass of the W boson is $m_w = \frac{1}{2} g_W \nu$, corresponding to the $W^{(1)}$ and $W^{(2)}$ spin-1 fields. To get the mass of the $W^{(3)}$ field (and then the mass of the Z boson and photon), we look at the terms quadratic in $W^{(3)}$ after the institution of D_μ from equation 1.22. After some algebra, these terms look like:

$$\frac{\nu^2}{8} \begin{pmatrix} W_\mu^{(3)} & B_\mu \end{pmatrix} \begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix} \begin{pmatrix} W^{(3)\mu} \\ B^\mu \end{pmatrix}. \quad (1.24)$$

Once the Lagrangian is written in the diagonal basis, this reduces to:

$$\frac{1}{2} \begin{pmatrix} A_\mu & Z_\mu \end{pmatrix} \begin{pmatrix} m_A^2 & 0 \\ 0 & m_Z^2 \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix}. \quad (1.25)$$

and we can see that the mass of the photon is $m_A = 0$ and the mass of the Z is $m_Z = \frac{1}{2}\nu\sqrt{g_W^2 + g'^2}$. Through the Higgs mechanism, the Z boson now has mass, while the photon is still massless. Combining the definitions of A_μ and Z_μ in equations 1.13 and 1.14, and the coupling ratio of $U(1)_Y$ and $SU(2)_L$ gauge symmetries ($\frac{g'}{g_W} = \tan\theta_W$), we can get a relation between m_W and m_Z via:

$$\frac{m_W}{m_Z} = \cos\theta_W. \quad (1.26)$$

Some physicists like to describe the $W^\pm$ and Z acquiring mass as these fields obtaining longitudinal field components by “eating” the Goldstone boson degrees of freedom. Using experimental measurements, physicists have approximated the value of the vacuum expectation value of the Higgs field, v , as $246\text{GeV}/c^2$, which is known as the electroweak scale. Finally, we can get the couplings of the Higgs boson to the gauge bosons by looking at the VVh and $VVhh$ terms in the Lagrangian where $V = W^\pm, Z$.

The final piece of *this* puzzle is the mass of the fermions, which is included in $\mathcal{L}_{yukawa}$. Unlike gauge boson masses that come from “eating” the gauge degrees of freedom, fermion masses come from interaction terms in the $\mathcal{L}_{SM}$ between the Higgs field and the fermions. The left-handed and right-handed particles act differently as they are placed in $SU(2)$ doublets and singlets respectively. Interaction terms that follow $SU(2)_L \times U(1)_Y$ symmetry look like $-g_f(\bar{L}\phi R + \bar{R}\phi^\dagger L)$. For example, the doublet for the tau (a key particle in the main analysis for this thesis) is:

$$\mathcal{L}_\tau = -g_\tau \left[\begin{pmatrix} \bar{\nu}_\tau & \bar{\tau} \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tau_R + \bar{\tau}_R \begin{pmatrix} \phi^{+*} & \phi^{0*} \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \right] \quad (1.27)$$

Choosing the unitary gauge, we can reduce this expression and get a term with the correct form for fermion masses, with a Yukawa coupling of $g_f = \sqrt{2}\frac{m_f}{v}$.

To handle the $SU(2)$ singlets, which includes giving the top quark mass, we construct the conjugate

doublet $\phi_c = \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix}$. This gives us a mass term of the form $-\frac{g_t}{\sqrt{2}}\nu(\bar{t}_L t_R + \bar{t}_R t_L)$, where the Yukawa coupling again (for any fermion) is $g_f = \sqrt{2}\frac{m_f}{\nu}$. The Yukawa couplings are free parameters that can be constrained experimentally.

1.5.4 HIGGS BOSON DISCOVERY

While it may now seem that the Higgs boson was the obvious solution to the local gauge invariance problem, it was initially one of several plausible solutions [48]. In 2012, the Higgs boson was discovered at CERN in both the ATLAS and the CMS detectors. ATLAS used the channels $H \rightarrow ZZ^{(*)} \rightarrow 4l$, $H \rightarrow \gamma\gamma$, and $H \rightarrow WW^{(*)} \rightarrow e\nu\mu\nu$ combined with the previous channels: $H \rightarrow ZZ^{(*)}$, $H \rightarrow WW^{(*)}$, $H \rightarrow b\bar{b}$, and $H \rightarrow \tau\tau^{(*)}$ [6]. The combined channels gave a discovery significance of 5.9 standard deviations. Meanwhile, CMS used combined measurements from the $\gamma\gamma$, ZZ , W^+W^- , $b\bar{b}$, and $\tau^+\tau^-$ to get a mass measurement with a local significance of 5.0 standard deviations [49]. This all but confirmed the Higgs boson as the solution particle physicists were looking for and led to the next few years at CERN focusing on fully categorizing the Standard Model Higgs boson.

1.5.5 HIGGS BOSON PRODUCTION MECHANISMS

There are four main Higgs boson production modes that can be measured at the LHC. All four have now been measured and agree with Standard Model predictions. Figure 1.5 shows the production mode cross sections depending on the Higgs mass at $\sqrt{s} = 14\text{TeV}$ center of mass energy for the LHC.

Gluon-gluon fusion (ggF): The most common production mode for the Higgs boson is gluon-gluon fusion ($gg \rightarrow H + X$). This interaction is mediated by the exchange of a virtual top quark, as shown in Figure 1.6 (a). This is required as the Higgs boson must couple to mass, and gluons are massless. Lighter quarks can also contribute to this process, but the massive top quark is the most dominant contribution and the others are suppressed relative to m_q^2 [4]. With a center of mass energy ($\sqrt{s}$) of 13TeV, the cross section for ggF is $48.6^{+4.6\%}_{-6.7\%}\text{pb}$ for a 125GeV Higgs boson. This amounts to approximately 88% of the total Higgs bosons produced at the LHC [50].

Vector Boson Fusion (VBF): The production mode with the second largest cross section at the LHC is Vector Boson Fusion (VBF). This production mode occurs when two quarks are scattered, along with an

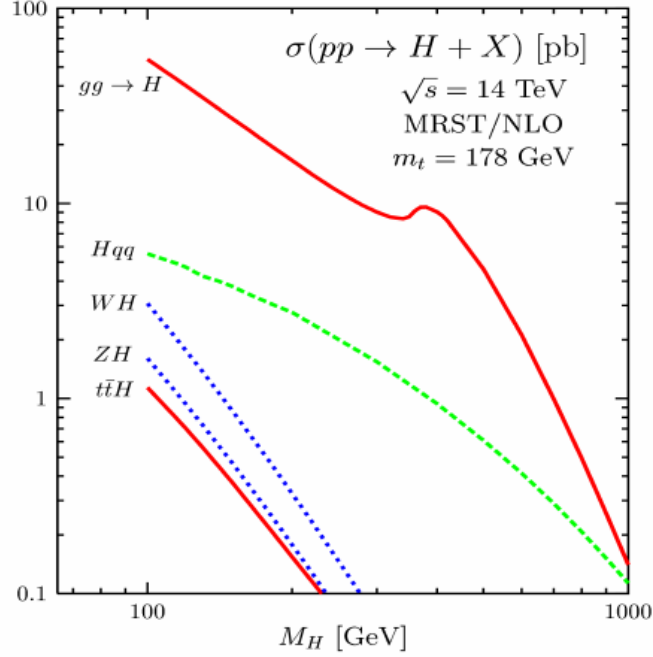


Figure 1.5: The main production modes for the SM Higgs boson as a function of the Higgs boson mass at $\sqrt{s} = 14\text{TeV}$. [3]

exchange of a W or Z boson, with the Higgs boson then radiating off of the W or Z boson ($qq \rightarrow qqH$) [4]. The W/Z is represented as the internal squiggly line in figure 1.6 (b). This signature is categorized by two hard jets in the forward or backward regions of the detector. The cross section for this process at $\sqrt{s} = 13\text{TeV}$ is $3.78^{+2.2\%}_{-2.2\%}\text{pb}$, accounting for $\sim 7\%$ of Higgs boson production [50].

Both ggF and VBF production methods were used in the Higgs discovery [6].

Associated Production with a Vector Boson (VH): The third most common production mode at the LHC is associated production with a vector boson (VH) and is the focus of the main analysis of this thesis. The process occurs when two quarks interact and form a W or Z boson which then radiates the Higgs boson ($pp \rightarrow VH + X$, with $V = W^\pm, Z$). Sometimes this process is referred to as Higgsstrahlung and is depicted in figure 1.6 (c). The cross section for WH is higher than for ZH due to the lower mass of the W boson compared to the Z boson. The cross section value for WH is $1.37^{+2.6\%}_{-2.6\%}\text{pb}$ and for ZH is $0.88^{+4.1\%}_{-3.5\%}\text{pb}$ at $\sqrt{s} = 13\text{TeV}$ [50]. Thus, VH is about 4% of the total Higgs boson production at the LHC.

Top Fusion: The final main contributor to Higgs boson production at the LHC is top fusion ($pp \rightarrow t\bar{t}H$). The process is gluon-gluon initiated and each of the initial quarks produces a pair of top quarks.

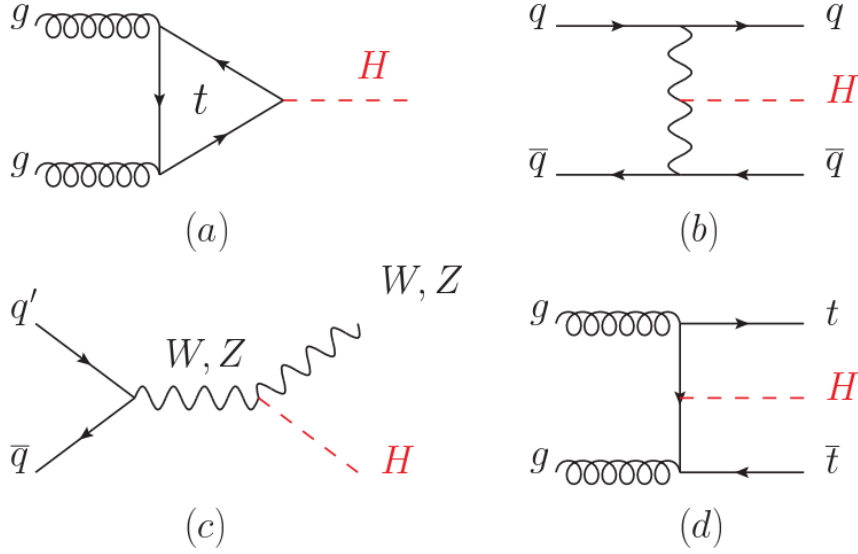


Figure 1.6: Feynman diagrams for the four main Higgs boson production mechanisms at the LHC. [4]

These top quarks can then fuse and produce a Higgs boson. Due to the high mass of the top quarks, this production mechanism requires a lot of energy, resulting in a low cross section value of $0.50^{+6.8\%}_{-9.9\%}$ pb at $\sqrt{s} = 13\text{TeV}$, or less than 1% of total Higgs boson production [50]. The process is shown in 1.6 (d).

1.5.6 HIGGS BOSON DECAYS

After being produced in one of the four main ways at the LHC, the Higgs boson almost immediately decays into other standard model particles. The Higgs boson has an extremely short lifetime of 1^{-22}s . The main decay mechanisms are shown relative to the Higgs boson mass in figure 1.7. Since the Higgs boson couples to mass it can decay to all massive particles except neutrinos, assuming the decay follows the conservation laws mentioned in 1.2.1. The Higgs boson couples more strongly to particles with larger mass.

Decays to massive fermions: The most common decay process $H \rightarrow b\bar{b}$ is also one of the hardest decay processes to observe in the detector due to the overwhelming QCD background. It was first measured in 2018 using the VH production channel that allowed for additional lepton(s) to trigger on [51]. Due to the difficult nature of measuring the $H \rightarrow b\bar{b}$ channel, the first measurement of the Yukawa coupling, or how the Higgs boson couples to fermions, was first seen through the $H \rightarrow \tau^+\tau^-$ decay channel. The

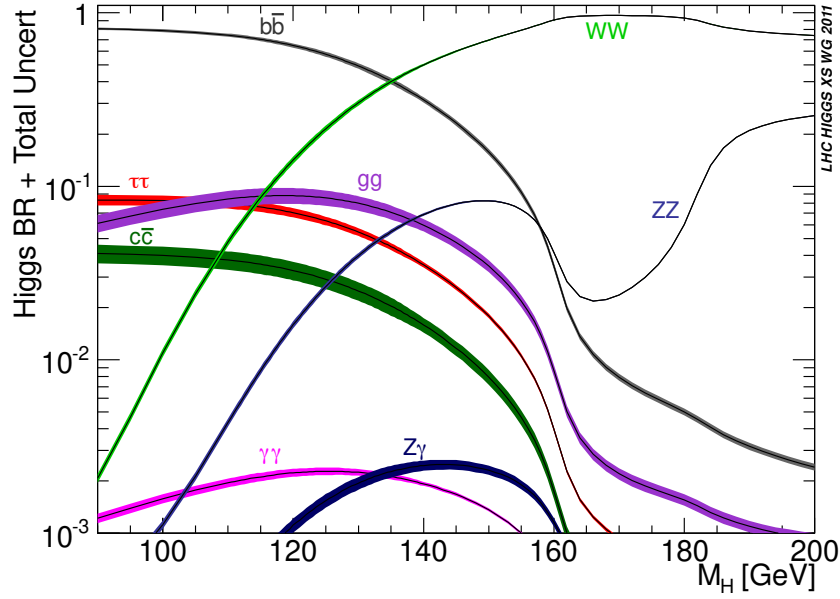


Figure 1.7: Relative frequency of Higgs boson branching ratios with respect to the mass of the Higgs boson. [5]

current ATLAS measurement of $H \rightarrow \tau^+ \tau^-$ combines VBF and ggF production [52]. The next most common Higgs boson to lepton coupling is $H \rightarrow \mu^+ \mu^-$. In 2020, ATLAS obtained a 2σ measurement on this rare channel [53].

The second most abundant Higgs boson to quark/anti-quark decay is $H \rightarrow c\bar{c}$. As of yet, there is no official observation, however CMS and ATLAS have set recent limits on the measurement [54] [55]. It is believed that decays to lighter quarks and lighter leptons (electrons) will be too rare for observation at the LHC.

Decays to massive bosons: The largest branching ratio for Higgs boson to other massive bosons, is via $H \rightarrow W^+ W^-$. It was first measured in ATLAS in the VBF and ggF production modes in 2015 with an excess of 6.1σ over the background [56]. Despite the $H \rightarrow ZZ$ branching ratio being much lower than $H \rightarrow W^+ W^-$, it was one of the channels used in the initial discovery measurement. The full decay process $H \rightarrow ZZ \rightarrow ll$ is known as the “golden channel” because of its clean, easy to see final state of four leptons. Figure 1.8 shows the clear signal of a Higgs boson above the background using this channel.

Decays to massless particles: One might assume that because the Higgs boson couples to mass, it is impossible for it to decay to massless particles, such as the photon. However, through the incorporation of an intermediate loop of virtual quarks or massive bosons, it is possible. $H \rightarrow gg$ is the most likely

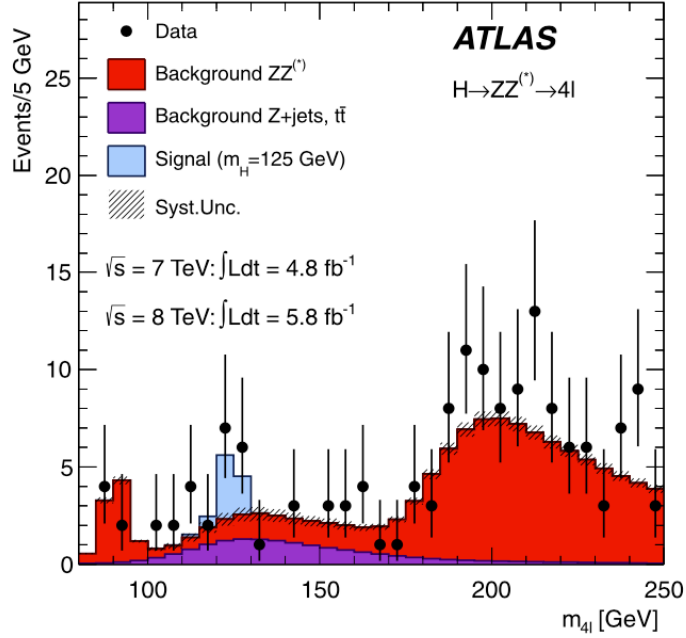


Figure 1.8: Four lepton invariant mass of selected signal of $H \rightarrow ZZ \rightarrow 4l$ and expected backgrounds in the 2012 ATLAS measurement. [6]

Higgs boson to massless particle decay with an expected BR of 8.5% [57]. Because gluons carry color, this process is (usually) mediated by a (top) quark loop, which requires a lot of energy. However, $H \rightarrow \gamma\gamma$ was another one of the key channels for the original Higgs boson discovery in 2012 [6]. Despite a BR of 0.22%, it was able to be seen due to the detector's ability to identify and reconstruct photons with great precision.

Goals of HL-LHC: Currently, the measurements of the Higgs boson decays and production measurements are confirming it as the SM Higgs. There are still some unmeasured channels that experimentalists hope to measure during future runs of the LHC, such as $H \rightarrow \mu^+\mu^-$ and $H \rightarrow Z\gamma$. In addition, the higher energy of the HL-LHC should allow for a 3σ combined measurement from ATLAS and CMS for di-Higgs production. Lastly, with the ability to conduct precision measurements, there is hope for sensitivity for Beyond the Standard Model Physics [58].

1.6 LIMITS OF THE STANDARD MODEL

Despite the current success of the Standard Model depicted by the consistency of data/theory measurements in figure 1.9, and the affirming discovery of the Higgs boson in 2012, there are still many unanswered questions in particle physics.

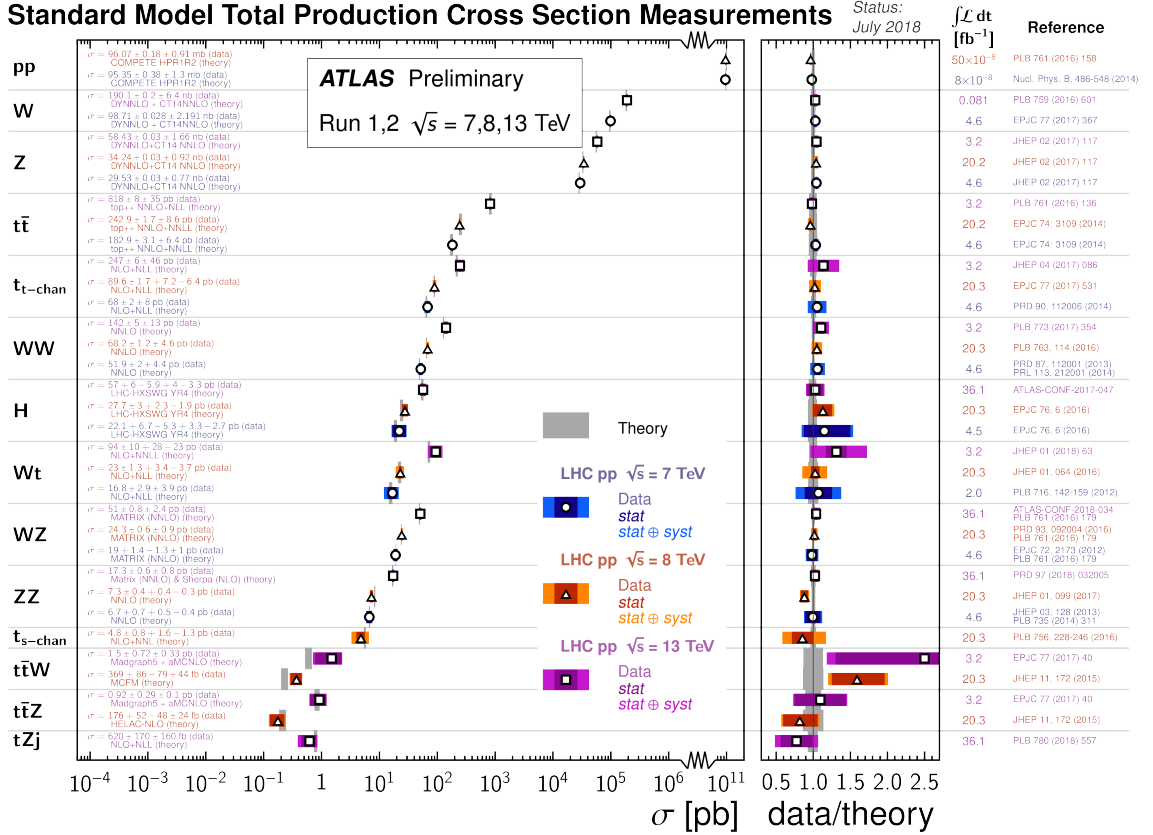


Figure 1.9: Summary of SM production cross section measurements and their agreement with theory. [7]

Dark Matter (DM) and Dark Energy (DE)

The standard model of particle physics describes the matter that we can see, however there is experimental evidence from astrophysics experiments that show that nearly 95% of the universe is made up of dark matter and dark energy that humans have not directly detected. The Hubble Telescope looked at the merger of two clusters and based on our knowledge of gravity's strength and gravitational lensing, scientists concluded with 8σ certainty that the majority of the mass of the system is unseen [59]. Other evidence of dark matter comes from the Cosmic Microwave Background and the patterns in it that cannot

be explained by gravity and photons alone [60]. WIMPs (weakly interacting massive particles) and axions are two possible dark matter particle candidates. Experiments, such as HAYSTAC at Yale, are looking for direct detection confirmation of axions [61]. Furthermore, the universe we live in is expanding at an accelerating rate and this could potentially be caused by the existence of dark energy.

Neutrino Mass and Oscillations

Experimental evidence that neutrinos oscillate, such as the results from Super-Kamiokande, require that neutrinos have mass, which goes against Standard Model predictions [62]. Furthermore, neutrino mass ordering can be determined from these oscillation measurements [63]. Despite experimentally seeing these results, scientists do not know how neutrinos get mass and why their masses are so small compared to other SM particles.

Matter-Antimatter Asymmetry

According to the Standard Model, matter and anti-matter should exist equally in the universe, with a slight imbalance allowed for by CP violation. Our understanding of the Big Bang includes the knowledge it created matter and antimatter in equal amounts. However, our universe is dominantly made out of matter, experimentally confirmed by astrophysical measurements. CP violation alone cannot explain this phenomenon.

Charged Lepton Flavor Violation

Recent experimental results at LHCb show evidence of charged lepton flavor violation. One of the conservation laws included in the Standard Model is that lepton flavor is conserved in interactions. LHCb saw hints of the breaking of lepton universality in beauty-quark decays [64]. Mu2e, an intensity frontier experiment, is looking for evidence of another charged lepton flavor violating decay of a muon decaying directly to an electron.

Anomalous Magnetic Moment of the Muon

Recent results from Muon g-2, grouped with a previous measurement at Brookhaven National Lab, show another conflict between Standard Model prediction and experimental measurement. In their high precision measurement of the anomalous magnetic moment of the muon, they found 4.2σ significance difference between their experimental result and theory [65]. While there is some discussion of whether the theory is correct, as this difference is several decimal places deep, it is an interesting example of the current methods used to determine whether beyond the Standard Model physics exists.

*“Silly things do cease to be silly
if they are done by sensible people in an impudent way.”*

Jane Austen, *Emma*

2

LHC and the ATLAS Detector

THE LARGE HADRON COLLIDER (LHC) AT CERN (CONSEIL EUROPÉEN POUR LA RECHERCHE NUCLÉAIRE) IS A MAGNIFICENT FEAT OF ENGINEERING MADE POSSIBLE BY A LARGE INTERNATIONAL COLLABORATION. CERN was founded in 1954 and is located along the French Swiss border, just outside the city of Geneva. It was created after the end of World War II to unite the scientists of Europe to push science forward for non-military purposes [66]. Today, there are 23 Member States of CERN, ten Associate Member States, and three countries with Observer Status, of which the USA is one. There are over 18,000 employees that work for CERN [67]. There are currently four main experiments located on the LHC ring to observe particle collisions: ATLAS, CMS, ALICE, and LHCb. The analysis documented in this thesis uses data taken by the ATLAS experiment during Run 2 (2015-2018) of the LHC.

2.1 LARGE HADRON COLLIDER AND ITS PREDECESSORS

In order to see new fundamental particles, scientists need to create a huge amount of energy in a very small amount of space. When choosing the type of collider to build, there are several options to consider: will

it be a fixed target or colliding beams experiment, will it be a linear or circular collider, will you accelerate composite or fundamental particles? Questions that guide these choices are: “what energy levels do I need?”, “what type of particles do I want to look at?”, and “how much money and/or space do I have?”.

Fixed target experiments work by accelerating particles, either in a linear or circular collider, and then hitting a fixed (non-moving) target with the particle beam. In these collisions, the energy of the collision is just the energy of the beam. When running a colliding beam experiment, two separate beams of particles collide together, which means you can reach double the center of mass energy (see section 2.1.1). Needing to reach a high level of energy in a minimal amount of space without super strong electric fields would lead one to choosing a circular collider.

Linear particle accelerators use oscillating electric fields to propel charged particles forward to then collide with a fixed target or with an oncoming beam. Linear colliders are used today in radiation therapy, in addition to physics research. Circular colliders work by accelerating two separate beams of particles in a circle, either in the same beam or parallel beams, and colliding them together at set interaction points to reach a higher center of mass energy. Both circular and linear colliders can reach high levels of energy. However, when running a linear collider, you only get one chance to accelerate the particles travelling down the beamline requiring extremely high electric fields and/or a very long tunnel. Furthermore, linear colliders lose the particles that do not interact. In a circular tunnel, we are able to recycle those particles and collide them again. A benefit of circular colliders is their ability to continually accelerate the same particles to higher and higher energies while they travel in the same circle. A downside to consider is the constantly changing magnetic fields needed to keep the particles travelling in a circle. Larger circles can smooth out the turns but are then more expensive to construct.

Deciding which type of particle to collide, composite or fundamental, ties in to choosing the shape of the collider. Fundamental particles, like electrons, radiate a lot of their energy due to bremsstrahlung. This means that while a circular collider can continually accelerate electrons, the electrons are also continually losing energy. On the other hand, a benefit of colliding fundamental particles is a more precise knowledge of the center of mass energy of the collision. This allows scientists to measure total missing energy instead of missing transverse energy.

The last component of the decision making comes from the cost and location of the collider. Some countries only have space for a linear or circular collider depending on their geography. Also, having a

collider fully in one country, rather than straddling a border, can be a cause of concern for international collaborations who want open access to the data and results. Lastly, cost and energy use is a growing concern and there is concerted effort to reach the highest energy with the lowest amount of money and operational costs. There are ongoing discussions now on whether to follow up the LHC with a linear collider (CLIC) or a circular collider (FCC). Both designs could collide electrons or positrons in order to study different physics from what we are currently measuring at the LHC.

Construction of the LHC, a circular collider that accelerates protons, at CERN was approved in 1994 and it was to be built in two stages with an initial 20 year commitment as a succession to the Large Electron Positron (LEP) collider. In fact, the LHC was built in the same tunnel as LEP and the work to create the LHC began well before LEP was finished running.

LEP was operated from 1989 to 2000 and was a particle, anti-particle collider. LEP eventually reached an energy of 209GeV and was able to produce the recently discovered W and Z bosons, confirm that there were three generations of matter particles, and set both lower (excluding $m_H < 114\text{GeV}$) and upper mass limits ($m_H < 222\text{GeV}$) on the Higgs boson [68]. The Tevatron at Fermilab had its Run II in between LEP and the LHC, from 2001 to 2011, and it reached center of mass energies of 1TeV colliding protons and anti-protons [69]. The Tevatron further confirmed the possible existence of the Higgs boson but was unable to claim a discovery prior to the start of the LHC. The Tevatron was able to make a 2.9σ measurement that the Higgs boson mass was between 115 and 135GeV just two days before the Higgs boson discovery from ATLAS and CMS [70]. Based on the limits set for the Higgs boson prior to the start of the LHC [12], but lack of discovery by both LEP and the Tevatron, researchers believed that the LHC needed to run at $\mathcal{O}(10)\text{TeV}$ center of mass energy to discover the Higgs boson and consequently the origin of mass. Scientists also hoped the LHC could provide answers to questions such as, “what kind of unification may exist beyond the standard model?” and “is there a deeper reason for gauge symmetry?” [71].

2.1.1 CONVENTIONS AT PARTICLE COLLIDERS

As a quick aside, let’s discuss the way energy is measured at particle colliders. The main unit used is eV, or electron volts, which is the amount of kinetic energy an electron gains after being accelerated through an electric potential difference of 1 Volt. This is also equivalent to $1.602 \times 10^{-19} J$. TeV stands for Tera

electron volt or 1 trillion electron volts. This may seem like a lot of energy, however 14TeV relates to $14 \times 10^{12} \times 1.602 \times 10^{-19} = 22.4 \times 10^{-7} J$ [72]. A benefit of colliding beam experiments is the energy increase achieved when moving from a fixed target experiment with a single beam to two colliding beams. This setup doubles the energy, which is measured as $\sqrt{s} = \text{center of mass energy}$. Each beam during Run 2 of the LHC was 6.5TeV, leading to a $\sqrt{s}$ of 13TeV. There are two main ways to think about energy of particles: mass and distance. $E = mc^2$ gives us mass in terms of energy, and the Compton wavelength $\lambda = \frac{hc}{E}$ gives a length scale in terms of inverse energy. So more energy allows us to see more massive and smaller particles! In order to create a certain type of particle, our collision energy must match or exceed that particle's rest mass.

Another key concept of particle colliders is *Luminosity*, which is a count of how many collisions will take place in a collider. There is *instantaneous luminosity* which gives the instantaneous rate of collisions, or how frequently we observe events within the collider. Assuming the particles in the proton bunches are identical in their transverse profile and follow a 2D Gaussian distribution along their axes perpendicular to the beamline, instantaneous luminosity is calculated as:

$$\mathcal{L} = f * \frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \quad (2.1)$$

where f is bunch crossing frequency, n_1, n_2 are number of particles contained in each bunch, and σ_x, σ_y are the transverse beam widths [73]. $\mathcal{L}$ has units $\frac{1}{cm^2 s}$ or $\frac{1}{time*area}$. To determine how many particle collisions will happen over a certain period of time, or how much data is taken in that time period, we calculate the *integrated luminosity*: $L = \int \mathcal{L} dt$. This is usually expressed as the inverse of cross section in fb^{-1} (inverse femtobarns). The use of barns as a measure of area comes from the phrase “couldn’t hit the broad side of a barn” (something is so big it would be difficult to miss) and the size of a uranium nucleus ($10^{-24} cm^2$). The barn was invented during WWII when developing nuclear weapons. A femtobarn is 10^{-15} barns and is used for cross section measurements in particle physics to turn large numbers into more manageable units [74]. Putting this all together, the number of expected events for a given time running the LHC is:

$$N_{exp} = \sigma_{exp} \times \int \mathcal{L}(t) dt. \quad (2.2)$$

There are several ways to move to higher luminosity in the circular collider. One way is to decrease the

bunch crossing time (or increase f in equation 2.1). In Run 1, we ran with 50ns bunch spacing. In Run 2, this was decreased to 25ns. This is currently at the edge of experimental feasibility given our detector readout times. Alternatively, you can increase the number of protons in a bunch that are then collided (increase n_i in equation 2.1). Finally, with tighter focusing of the beams, resulting in more densely packed colliding proton bunches, luminosity will also increase.

One of the consequences of moving to higher luminosity if you have increased the number of protons in a bunch is additional *pileup*. Pileup is a measure of the number of interactions per crossing. High levels of pileup can cause the detector to record many events simultaneously that need to then be disentangled from each other so that the individual interactions can be studied. Both simultaneous (in-time pileup) and near-time occurring (out-of-time pileup) collisions add to event reconstruction difficulties. Pileup levels seen in Run 2 and expected in Run 3 can be handled by the detector. However, increasing pileup can increase trigger rates and detector occupancy, which can then increase the amount of bandwidth needed to read the data from the detector and increase the space required to store events. A greater level of sophistication might be required in reconstruction algorithms and analysis designs in order to cope with the higher background levels. This can perhaps be done with the incorporation of more advanced machine learning algorithms in each step of our data processing.

2.1.2 LHC CONSTRUCTION AND OPERATION

The LHC was built by over 10,000 scientists and engineers from more than 100 countries. Construction began in 1994 and the first LHC Conceptual Design Report was published in 1995 [75] [76]. The very first proton beam was in 2008, but an accident occurred resulting in significant damage. The running of the LHC was then shut down until repairs could be made. Official data taking of the first proton collisions started in 2009 [72].

The LHC collides protons 100 meters underground in a tunnel with a circumference of 26.7km under the France/Switzerland border near Geneva. During Run 2, each beam has an energy of 6.5TeV making the center of mass energy 13TeV. Protons are accelerated to 0.999999991 times the speed of light at top energy of $\sqrt{s} = 14\text{TeV}$ in the LHC [72]. The bunches collide every 25ns and there was an average pileup of 33.7 interactions per crossing in Run 2. Figure 2.1 shows the mean interactions per crossing (pileup) during each year of Run 2 and all years combined. The protons travel in bunches and there are

about 2800 bunches in the beam at once. This all leads to about a billion collisions per second occurring in the LHC. The protons are designed to collide at four main interaction points in each of the detectors as seen in figure 2.2.

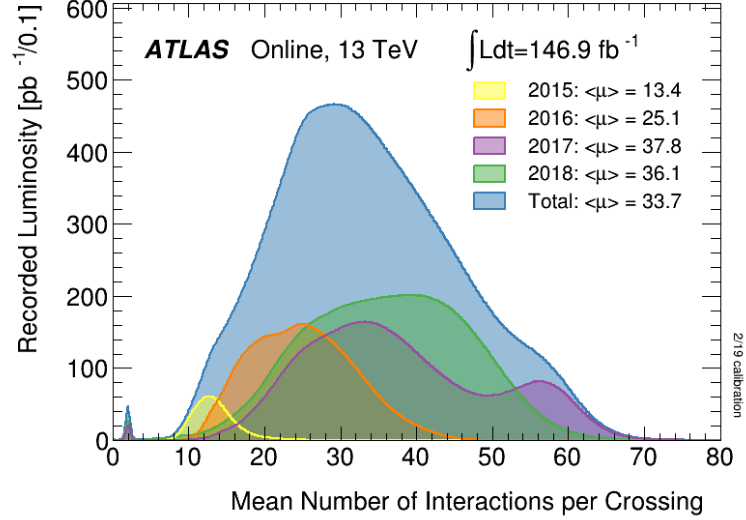


Figure 2.1: Distribution of mean interactions per crossing weighted by luminosity for the full Run 2 pp collision data set of the ATLAS detector. This includes all data taken during stable beams. [8]

The LHC is kept at vacuum and superconducting temperatures in order for the particles to travel at close to the speed of light and operate the magnets that bend the beam path. There are actually three vacuum systems at the LHC: one for the beams, one for the cryomagnets, and one for the liquid helium that is used to cool the magnets [76]. Vacuum is necessary in the beam so that the bunches only collide with each other and not other particles. The beam pipes are lowered to a pressure of $\sim 10^{-13}$ atm, which is close to the pressure on the surface of the moon. The beam pipe has two sections under different temperatures: the arc beams are maintained at 1.9K and the straight sections are kept at room temperature. This warming area allows for “bakeout” at 300°C which maintains the vacuum purity when performed at regular intervals. The insulating vacuum for the helium and magnets takes up 15,000 cubic meters of volume and reaches a pressure of 10^{-6} mbar.

There are 9600 magnets in the LHC that control the particles’ trajectories. Taking advantage of the basic physics law that charged particles change directions in the presence of a magnetic field, we are able to steer the protons to travel in a circle and then collide at the four interaction points. This is done with



Figure 2.2: Aerial view of CERN and Geneva, with the LHC tunnel in yellow, and the four main detectors in white, with the French/Swiss border shown with the white dashed line. [9]

a combination of dipoles, quadrupoles, sextupoles, octupoles, etc [72]. The lattice magnets - dipoles and quadrupole magnets, keep the particles on the right path. The 1232 main dipoles have a magnetic field of 8.3T, more than $100000 \times$ Earth's magnetic field, and are each 15m long and weigh 35 tons. These dipole magnets are used to bend the path of the particles along the path of the LHC tunnels. The 319 quadrupoles, which are 5 – 7 meters long, focus the beam, keeping the particles in dense bunches. Next, there are magnets in the insertion area right before collision in the detectors. Three quadrupoles are combined together to make an inner triplet. There are two inner triplets at each of the four detectors and they narrow the beam in preparation for collisions. Dipole magnets then separate the beams after collision and set them back on their usual beam path. A final system of magnets, known as the kicker magnets, that will deflect the beam in the abort gaps into absorption material outside the main ring [76].

In order to fit into the 3.7m radius LEP tunnel, there could not be two separate beamlines with all the necessary supporting services and room for personnel. To solve this issue, we chose a “two-in-one” design

that is space saving but causes the beams to be magnetically coupled [10]. This is shown in figure 2.3. The beam lines themselves are 3cm in diameter. In order to keep the two beamlines from being torn apart due to their opposing magnetic fields, non-magnetic steel collars are critical to the design.

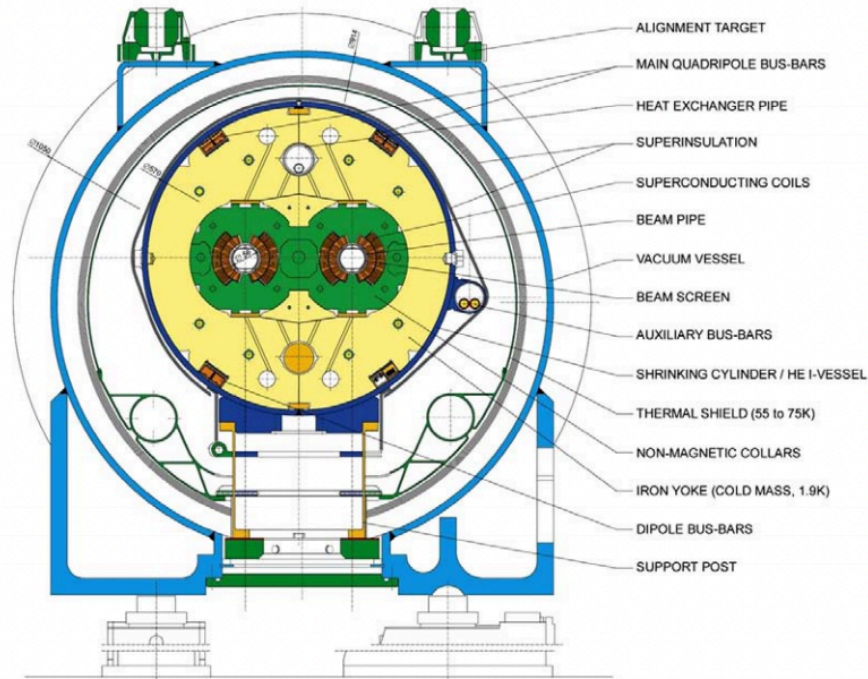


Figure 2.3: Cross section of the LHC beam line showing location of quadrupole magnet bars, alignment target, vacuum vessel, non-magnetic collars, and dual beams. [10]

2.1.3 ACCELERATING PROTONS AT THE LHC

In the previous section, we detailed how the protons are kept circulating in the main LHC ring, but what does it take to get the protons to circulate at a speed allowing them to travel the whole 27km circle 11,000 times per second? Protons are isolated in a system called a duoplasmatron (metal cylinder) [77]. Hydrogen gas is stored inside until it is extracted and exposed to electric fields that are applied to strip the electrons from the protons. The particles are then accelerated by a 90kV supply from an RF quadrupole and inserted into the LINAC system at about 4000km/s. The protons are then sped up by a series of linear and circular accelerators as shown in figure 2.4 [78]. In Linac 2, the protons are accelerated to 50MeV, then to 1.4GeV in the Proton Synchrotron Booster, which is composed of four stacked synchrotronic rings. In the Proton

Synchrotron (PS) the protons are accelerated to 26GeV. The PS has a circumference of 628m and it creates 81 bunch packets 25ns apart [79]. The protons are then fed to the Super Proton Synchrotron (SPS) which accelerates them to 450GeV in its 7km ring with 744 dipole magnets to bend the beam. Both the PS and SPS also provide particle beams to experiments located on or near their rings. The SPS injects the proton bunches in opposite directions into the LHC beamline and it takes just over four minutes to fill the final beam. When the protons are injected into the LHC there are on order of 10^{11} protons in each bunch at the start [72]. It takes about 20 minutes to accelerate the protons from 450GeV to 6.5TeV in the LHC.

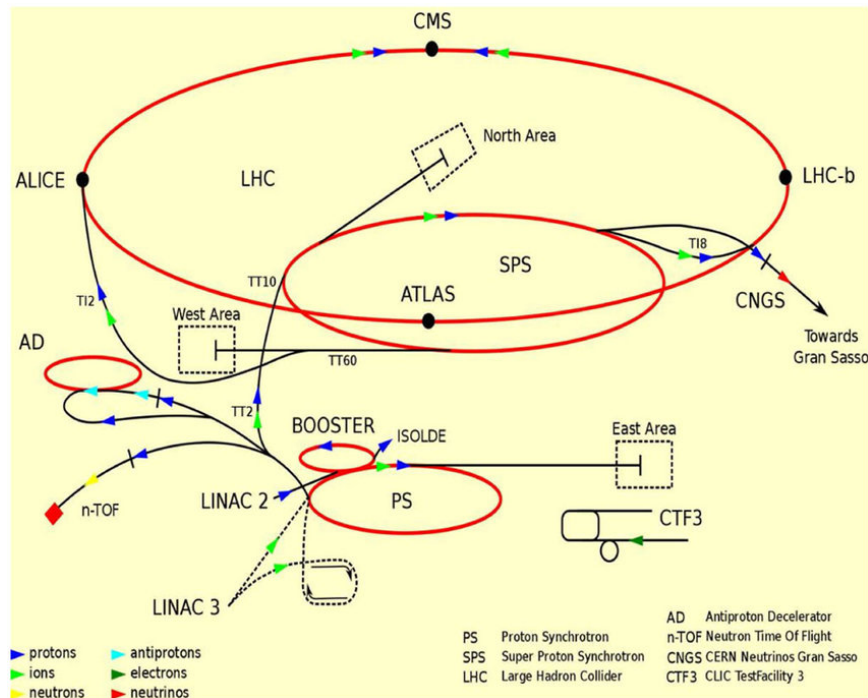


Figure 2.4: A schematic of the entire LHC linear and circular accelerator complex. Protons are injected into Linac 2 and then passed to the Booster, PS, SPS, and finally the LHC ring where they are accelerated to energies of 6.5TeV. [11]

To accelerate the bunches further in the LHC, radio frequency (RF) cavities are employed. There are 16 RF cavities on the LHC, each housed in one of four cryogenic chambers to allow them to operate in a superconducting state. The RF cavities are metallic chambers that house an electromagnetic field. This electromagnetic field is oscillated at a frequency of 400MHz. The proton bunches are constrained by this oscillating RF field as the frequency determines the boundaries for these proton bunches (RF buckets). Protons outside of these bunches (or not travelling at the final speed) will be accelerated or decelerated to match. This leads to a bunching of protons around the synchronous proton.

With the help of the magnet system explained earlier, the proton bunches are made to collide at the four main interaction points, where each of the four main detectors are located. ATLAS is located at point 1, right near the main campus of CERN, while CMS is on the opposite side at point 5. There are also special runs of the LHC where heavy ions (lead nuclei) are collided at lower energies, such as $\sqrt{s_{NN}} = 5.02\text{TeV}$. These runs enable scientists to study other types of physics phenomena, such as quark-gluon plasma.

2.1.4 RUN 2 OF THE LHC

The LHC program consists of a series of multi-year runs and periods of long shutdowns to upgrade the accelerator to reach higher energies and the detectors to deliver physics results, as detailed in figure 2.5. The data for this thesis comes from Run 2, which occurred from 2015-2018 and had an integrated luminosity of 13TeV with 139fb^{-1} of good data. The total data delivered by the LHC in Run 2 is shown in green on the left of figure 2.6, as well as the amount ATLAS recorded in yellow and the amount that was good for physics in blue.

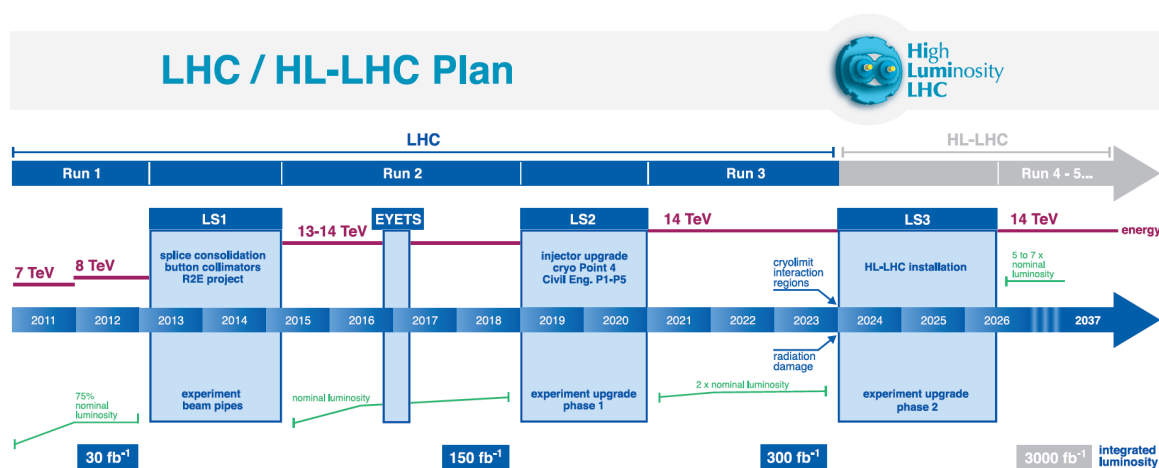


Figure 2.5: Timeline of running the LHC created in 2016 before there were delays to Run 3 due to covid 19. [12]

Run 2 was followed by an originally planned two year shutdown, that extended to three years during the pandemic, in order to increase beam brightness. Run 3, which will begin in early 2022, will extend over three or four years and is expected to deliver 300fb^{-1} of data at $13 - 14\text{TeV}$ [12]. After Run 3, another long shut down will allow for upgrades to both detectors and the beamline to handle the increase

in luminosity planned for the HL-LHC. Currently, there are two major runs planned for the HL-LHC. The total projected luminosity for the HL-LHC, a dramatic increase over current recorded luminosity, is shown on the right of figure 2.6. Discussions are ongoing to determine what to do after the HL-LHC. It is a pivotal time for particle physics as we decide how and what we want to look for next!

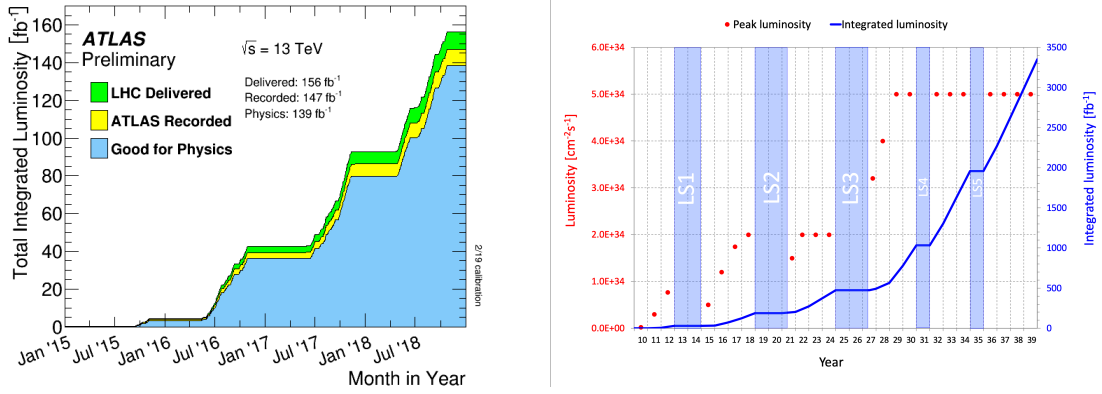


Figure 2.6: Total integrated luminosity of Run 2 delivered by the LHC (green), recorded by ATLAS (yellow), and data collected that can be used for physics (blue) (left). Total projected luminosity for the HL-LHC (right). [13]

2.2 ATLAS DETECTOR

The ATLAS experiment (A Toroidal LHC ApparatuS) is one of four main detectors located on the Large Hadron Collider (LHC) beamline.

The ATLAS detector itself weighs 7000 tons, is 44m long and 25m high, and located 100m underground [80]. The detector has a forward/backward symmetric cylindrical geometry. The barrel region has cylindrical surfaces around the beam line and the End-Cap has surfaces perpendicular to the beamline. There are four major components of the ATLAS detector: the Inner Detector, the Electromagnetic and Hadronic Calorimeters, the Muon Spectrometer, and the Magnet System. These are used to measure the particles created in the collisions. The first three components are located in concentric cylinders around the beam line, while the magnets encompass the inner detector and the muon system to bend the charged particles. There are over a billion particle interactions every second in the ATLAS detector and only 0.0001% of events are saved via the Trigger and Data Acquisition (TDAQ) system and available for further study. A schematic of the ATLAS detector with people for scale is shown in Figure 2.7.

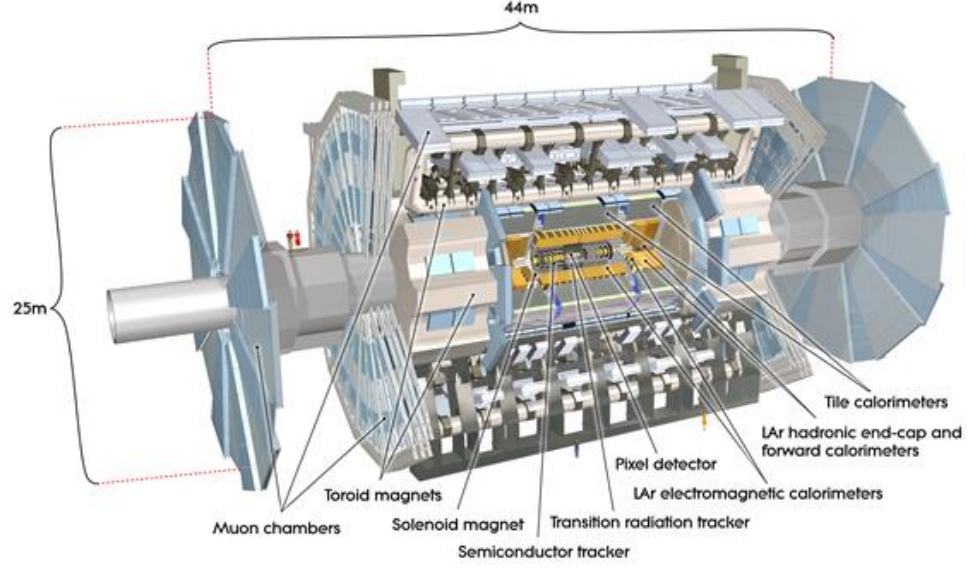


Figure 2.7: Schematic of the ATLAS detector with the four major subsystems and people for scale. [14].

For the coordinate system of ATLAS, the z -axis is defined as the beam line, and the x - and y -axes are the transverse plane, with the x -axis pointing towards the LHC ring center and the y -axis in the upwards direction. θ is the polar axis from the beam line and ϕ is the azimuthal axis around the beamline [81]. The interaction point is at the center of the detector. The reference coordinate axes are shown in figure 2.8.

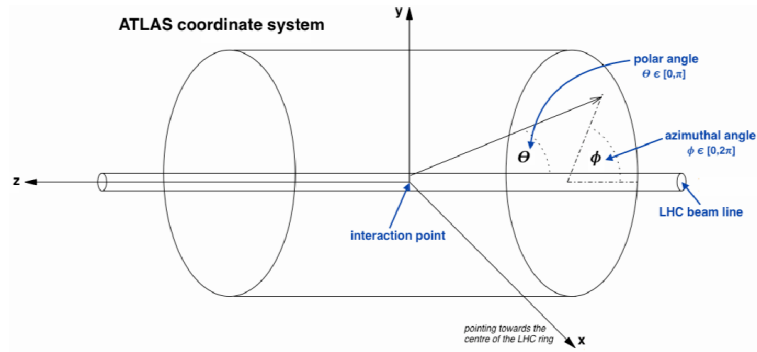


Figure 2.8: Coordinate axes of the ATLAS detector showing the z -axis as the beam line, the x -axis pointing towards center of the LHC, and the polar angles. [15]

Another way to measure particle location in the detector is with η (pseudorapidity). η is a function of θ where

$$\eta = -\ln \tan(\theta/2) \quad (2.3)$$

as shown in figure 2.9. The values of η range from negative infinity to positive infinity if aligned directly with the beam line. A value of $\eta = 0$ means directly perpendicular to the beam line. Another useful distance measurement is ΔR , which is a measure of angular separation and is calculated as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (2.4)$$

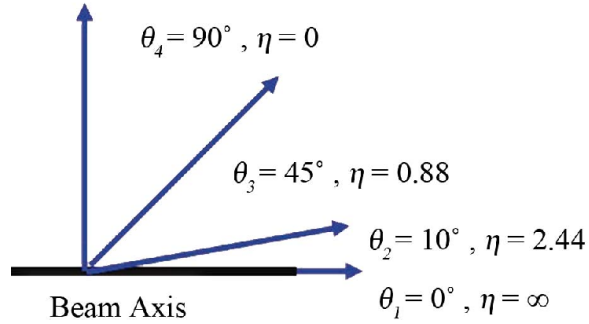


Figure 2.9: Diagram of pseudorapidity (η) for various values of polar angle θ with respect to the beam line. [15]

Before diving into each sub-component of the ATLAS detector, let's detail the general path a particle will follow through the detector depending on its type. Figure 2.10 shows the general paths of several representative particles. All particles within the geometric acceptance will pass through the Inner Detector, however only the charged particles will have tracks reconstructed. Many particles will then interact in the electromagnetic calorimeter, with electrons and photons typically depositing the bulk of their energy within its volume. Energetic hadrons will typically move on through to the hadronic calorimeter. Pions are the main particle measured in the hadronic calorimeter. Muons will continue on to the Muon Spectrometer at the outermost portion of the ATLAS detector. Unfortunately, the neutral and almost massless neutrino will pass through undetected, complicating our attempts to fully reconstruct the particles that were created within the detector.

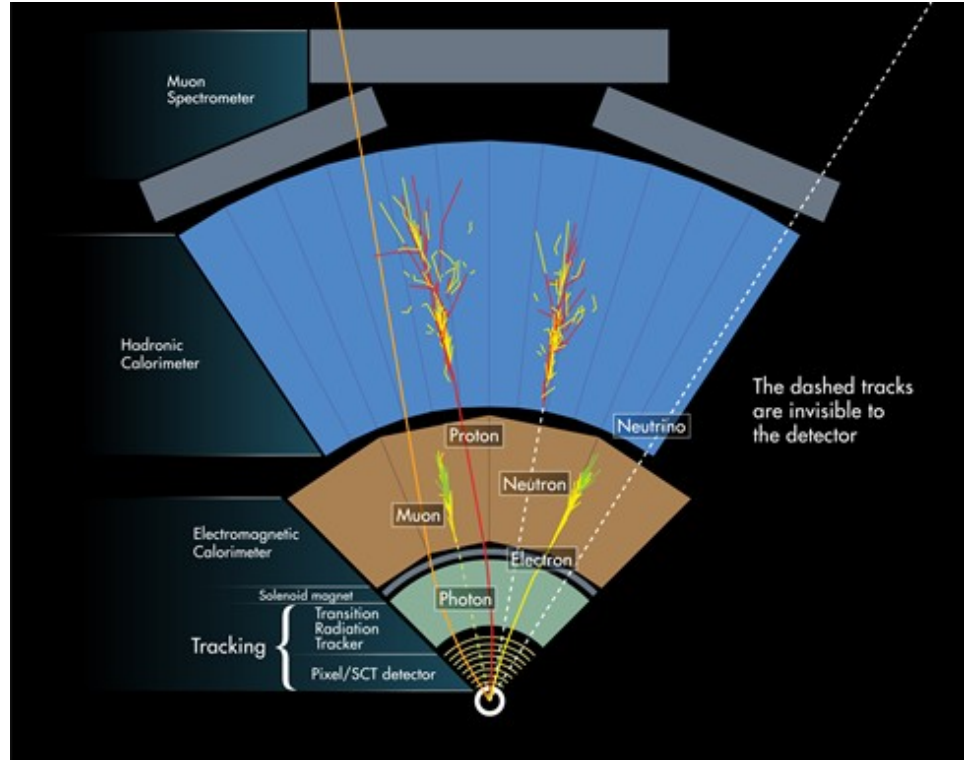


Figure 2.10: Schematic of a slice of the ATLAS detector, demonstrating where in the calorimeters various types of particles deposit their energy. [16]

2.2.1 INNER DETECTOR

The innermost portion of the ATLAS detector is aptly named the Inner Detector (ID) as shown in figure 2.11. The ID is 7m long and has a radius of 1.15m, with a 2T magnetic field applied to enable measurements of particle momenta and charge [81]. The ID covers a wide region for particle tracking and reconstruction within a range of $|\eta| < 2.5$ [82]. There are three main parts of the Inner Detector: the Pixel Detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT). The goal of the ID is to measure the charge, momentum, and trajectories of particles.

Pixel Detector

The first part of the ID is the pixel detector, which initially had 80 million readout channels to help detect short-lived particles, like the b -hadrons, by providing excellent resolution close to the beam pipe to enable the detection of secondary vertices. The pixel detector has four barrel layers and three pixel disks in each end-cap and spans the radial region of 30 – 150mm from the beam line [82]. The pixel size is $50(r - \phi) \times$

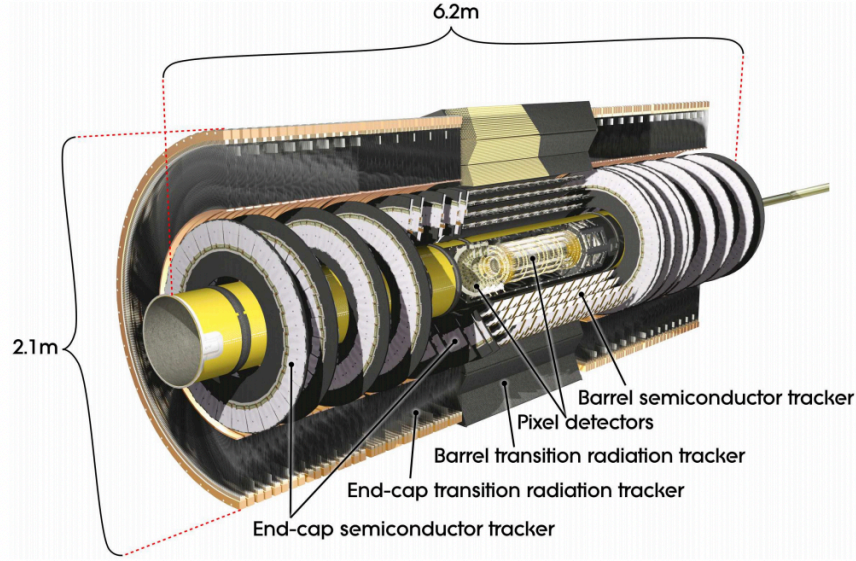


Figure 2.11: Cut-away diagram of the ATLAS Inner Detector with barrel and end-cap portions of the Pixel Detector, SCT, and TRT detectors. [14]

$400(z) \mu m^2$, with a spatial resolution of $14 \times 115 \mu m^2$.

At the beginning of Run 2, the Insertable B-layer (IBL) was added. This is now the innermost layer of the ATLAS detector and is located at a radial distance of 3.3cm from the beamline. The IBL covers a region of $|\eta| < 3.03$ and contains 12 million silicon pixels with a size of $50 \mu m(r - \phi) \times 250 \mu m(z)$. This layer improves both tracking readouts of charged particles and reconstruction of the secondary vertex. This is key for the identification of jets from short-lived particle decays, such as the decay of the τ -lepton. The IBL also helped mitigate damage to the previous inner layer during Run 1 and to the entire pixel system in Run 2. The IBL also provided robust b-tagging performance even with module failures in various layers [83].

Semiconductor Tracker

The middle portion of the ID is the Semiconductor Tracker (SCT), which is made of silicon micro-strips for pattern recognition. The SCT has four cylindrical double-sided barrel layers and nine end-cap layers with over six million readout strips. The silicon strips have a diode with a p-n junction and an applied bias voltage. The signal is read out by the Front End Electronics with a position accuracy of $17 \mu m$ per layer transverse to the strips in $(r - \phi)$ and $580 \mu m$ in z . There is a goal of eight SCT location measurements per particle track. These eight precise measurements allow for momentum, impact parameter, and vertex

position measurement.

Transition Radiation Tracker

The outermost layer of the Inner Detector is the Transition Radiation Tracker (TRT) which is made of 4mm diameter straw-tube detectors with 250,000 0.39m-long straws in the end-caps and 50,000 1.44m-long straws in the barrel. The straw tubes are filled with 70% xenon gas, 27% CO₂ and 3% O₂ and are plated in gold. The particle tracks are read out because the passing particle ionizes the gas, causing free electrons to move to the wires and produce a signal. The TRT has 350,000 read-out channels and is able to provide particle type differentiation between electron and pions due to the difference in high threshold hits ($> 6\text{keV}$) in the tracker [84].

The locations and layout of the many layers of the Inner Detector including the TRT are shown in figure 2.12.

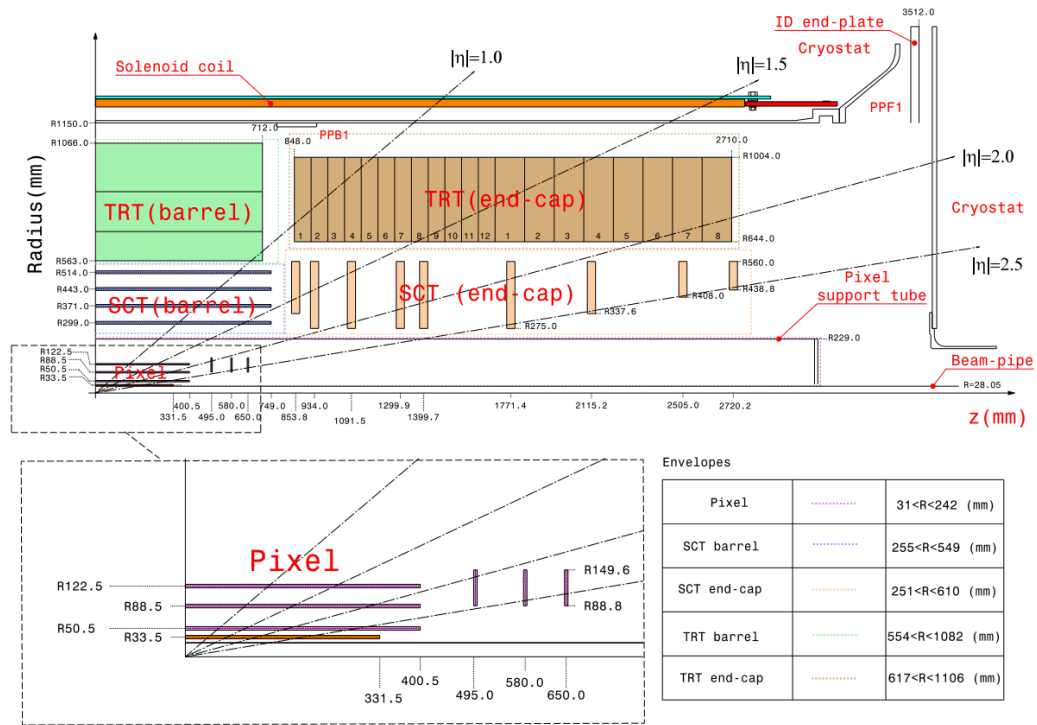


Figure 2.12: Diagram of the TRT, SCT, and pixel subsystems and their locations in r , z , and η in the ATLAS Inner Detector. [14]

2.2.2 CALORIMETERS

The calorimeters measure the energy of both charged and neutral particles. The alternating material layers of the calorimeters are chosen to stop the particles and measure their energy. The electromagnetic (EM) calorimeters are designed to measure the energy of electrons and photons as they pass through, whereas hadronic calorimeters are designed to measure the energy of hadrons as many pass through the EM calorimeters without depositing most of their energy. The EM calorimeter is directly outside of the Inner Detector, and the hadronic calorimeter is outside of that as shown in figure 2.13. The EM barrel calorimeter matches the η coverage of the ID while two calorimeters together cover a range of $|\eta| < 4.9$ [14].

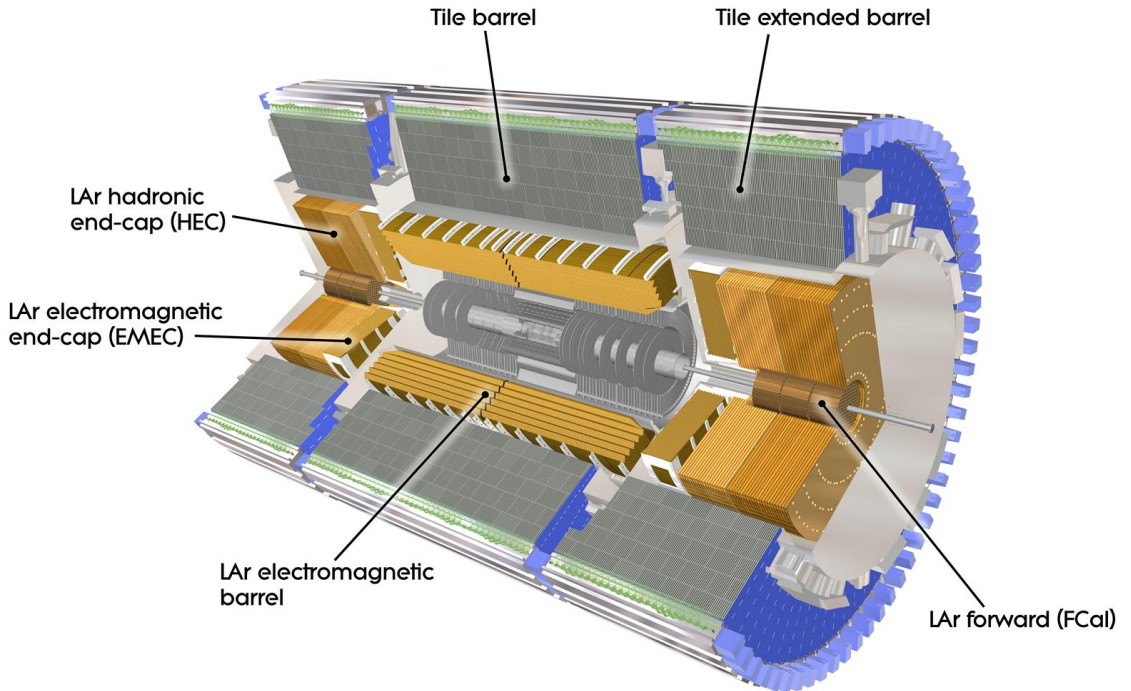


Figure 2.13: Cut-away of the ATLAS calorimeter showing both the liquid Argon (LAr) and Tile systems of the EM and hadronic calorimeters. [14]

Liquid Argon (LAr) Calorimeter

The EM calorimeter is made up of liquid argon and lead plates. The lead is there to absorb the particle's energy and generate an electromagnetic shower through the process of bremsstrahlung. The liquid argon

that is between the plates is ionized by the passing shower particles. This process is repeated several times through the calorimeter layers in order to fully collect the energy from the incoming particle. The LAr calorimeter is 6.4m long and 0.53m thick. The calorimeter thickness is greater than 22 radiation lengths (X_0) in the barrel and 24 X_0 in the end-caps; deep enough to collect most of the energy from electrons and photons [14]. The calorimeter has three layers of accordion-shaped lead plates as shown in the left-hand side of figure 2.14. The barrel covers the region of $|\eta| < 1.475$ and the end-caps cover $1.375 < |\eta| < 3.2$ causing a small gap with lower energy resolution at around $|\eta| = 1.5$ [17]. The barrel has two identical half barrel segments and the end cap has two coaxial wheels. The accordion shape allows for complete ϕ symmetry with a readout cell size of 0.025 in ϕ in the inner layers and without major cracks in the azimuthal angle. The calorimeter is kept in a cryostat at -184°C and vacuum sealed cables bring the 200,000 readout channels through cold-to-warm feedthroughs.

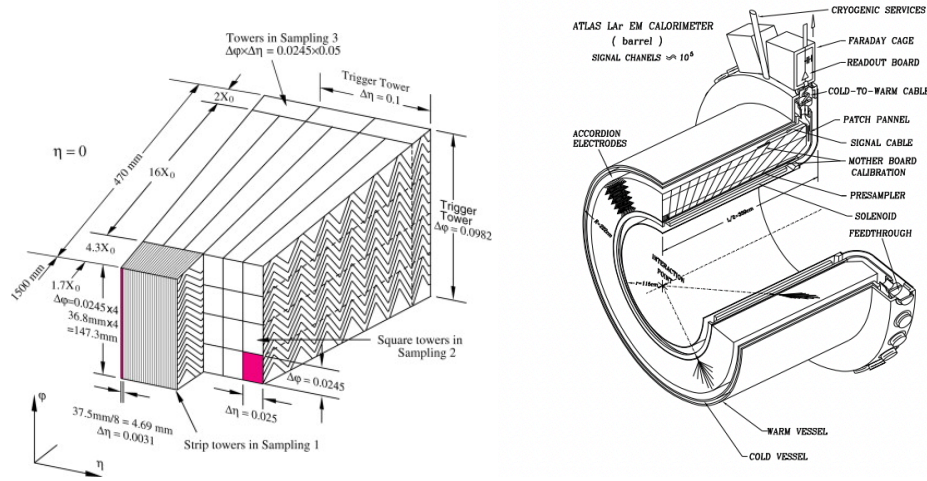


Figure 2.14: Diagram of the EM calorimeter and its accordion structure, showing the resolution in ϕ and η (left) Diagram of the barrel cryostat for the LAr EM Calorimeter with the presampler and feedthroughs (right). [17]

A presampler is also included to account for the energy lost by photons and electrons upstream of the calorimeters [14]. This allows us to have higher granularity position and direction measurements for the incoming shower. In the barrel and end-cap regions, the presampler feature is a 1cm or 5mm LAr active layer perpendicular or parallel to the beam line. The presampler in the barrel region is shown on the right of figure 2.14. In between the barrel and end-cap is a scintillator layer to capture needed energy measurements for jets.[17]

Hadronic Calorimeter

On the outside of the EM calorimeter sits the hadronic calorimeter with a central barrel that is 5.6m long and two extended barrels that are 2.6m long. The goal of the hadronic calorimeter is to measure energy while stopping hadrons from plowing into the muon spectrometer. There are 420,000 plastic scintillator tiles to measure the interactions in 5182 cells, intermixed with steel plates to stop the particles [80]. A key particle typically measured in the hadronic calorimeter for tau-lepton analyses are charged pions.

There are three components to the hadronic calorimeter: the tile calorimeter, LAr hadronic end-cap calorimeter (HEC) and the LAr forward calorimeter (FCAL). The central barrel of the calorimeter covers the region $|\eta| < 1.0$ and the outer barrels cover $0.8 < |\eta| < 1.7$. Each barrel has 64 modules and there are three readout layers; the inner two have a segmentation of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ and the outer one has a segmentation of $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ [85]. The extended barrel gives a total detector thickness of 9.7 nuclear interaction lengths (λ), allowing for precise measurements of high-energy jets. The scintillators are read out by 9500 photomultiplier tubes via wavelength shifting fibres. Figure 2.15 shows a schematic of a tile calorimeter cell with the alternating steel and trapezoidal scintillating tiles going out radially and perpendicular to the beamline, and the photomultiplier tube and fibres for read-outs travelling parallel [18].

In the HEC, there are two independent wheel end-caps, which use liquid argon like the EM calorimeter and cover $1.5 < |\eta| < 3.2$. There are 64 total modules in the End-cap calorimeter, with 32 modules in each wheel. The inner wheel has 25mm copper plates and the outer one has 50mm copper plates. In between the copper plates are 8.5mm LAr gaps that serve as scintillating material. Inside the gap are three parallel electrodes, making four drift spaces. The central one is the readout electrode and the other two serve as high voltage carriers.

Finally, there is the FCAL, which has to deal with high amounts of radiation due to its proximity with the beamline as it covers the η region of $3.2 < |\eta| < 4.9$. The front face is also 4.7m from the interaction point. This calorimeter requires a high-density design due to its recessed face with respect to the EM calorimeter face. There are three modules in the ECAL, one copper and two tungsten, which cover about 9.5 active interaction lengths with its dense design [17]. Each calorimeter section is a grounded metal matrix with channels filled with concentric rods at high voltage and grounded tubes. This matrix setup allows for excellent control of the gaps, which are as small as $250\mu m$ in the first section. In total,

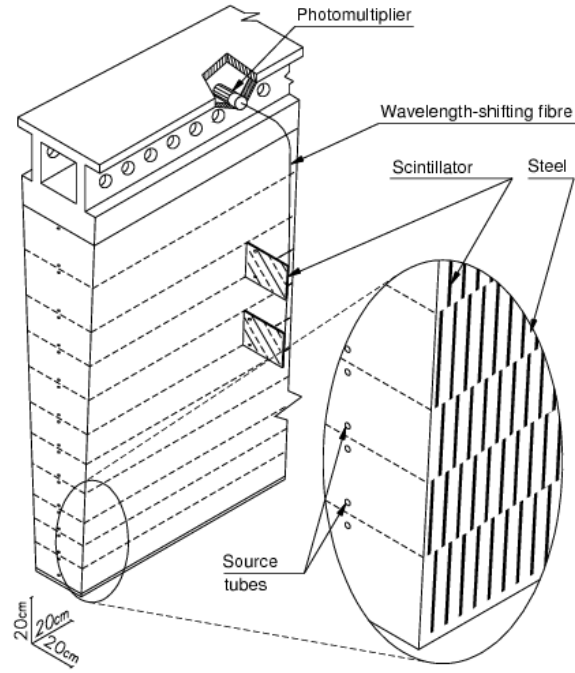


Figure 2.15: Schematic of tile calorimeter module with alternating lead and scintillating layers normal to the beamline and with the wavelength shifting fiber readouts parallel to the beamline. [18]

there are 3485 readout channels for the FCAL.

2.2.3 MUON SPECTROMETER

Since muons are minimally ionizing particles in the energy ranges of interest at the LHC, they do not deposit the majority of their energy in the calorimeters. Thus we need additional information to identify them. ATLAS has a special component, the Muon Spectrometer, to measure muons' tracks and momenta. This is the outermost portion of the ATLAS detector. There are 4000 muon chambers where magnetic deflection is used to measure the particle's trajectory. The large barrel toroid magnet bends muon tracks in the $|\eta| < 1.4$ region and the end-cap magnets bend tracks in $1.6 < |\eta| < 2.7$ region [81]. In between ($1.4 < |\eta| < 1.6$), the muon tracks are bent by a combination of the fields from the barrel and end-cap. Tracks are measured in all regions of the muon spectrometer. In the barrel region, chambers in three cylindrical layers around the beam axis measure tracks. In the end-cap and transition region the three stations of chambers are installed vertically.

Four types of gaseous radiation detectors are used in the Muon Spectrometer: the MDT, CSC, RPC,

and TGC. These can be seen in figure 2.16. These detection chambers are arranged in projective towers with alignment rays optically connecting them. There are two classes of detector types: *precision* and *trigger* chambers.[19] The *precision* measurements are conducted by the Monitored Drift Tubes (MDT) that measure the curve of tracks for most of the η -range and the Cathode Strip Chambers (CSC) that measure coordinates at the ends of the detector [80]. The MDT have 1171 chambers with 354,240 tubes for a single-wire resolution of $80\mu m$. Each 30mm diameter tube is made of aluminum and contains $50\mu m$ diameter central wire. The tubes are filled with 93%Ar and 7%CO₂. The MDT chambers are constructed from multi-layer pairs of three or four monolayers of tubes to allow for precise reconstruction. The CSC are multi-wire proportional chambers with segmented cathodes and have 70,000 channels with a hit resolution of $60\mu m$. The hit coordinate is measured by reading out the avalanche formed on the anode wire by the induced charge on the segmented cathode. The CSC are in 2×4 layers and are filled with a mixture of 30%Ar, 50%CO₂, and 20%CF₄ [81].

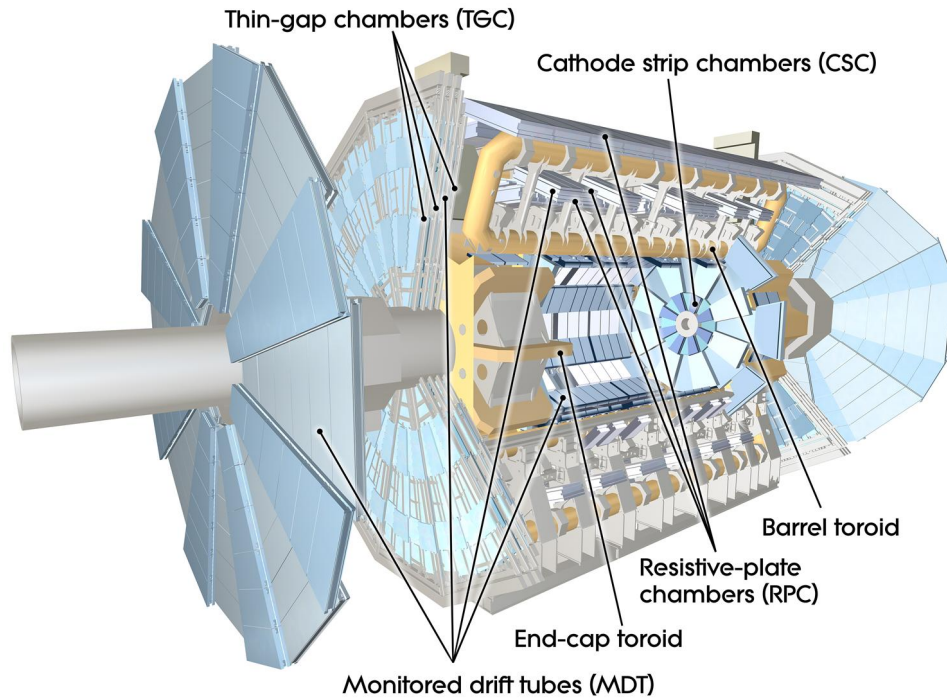


Figure 2.16: Cut away view of the Muon Spectrometer with the 4 main detector components and magnets. [14]

The *trigger* chambers, used for coordinate measurement in the bending plane, are the Thin Gap Chambers (TPC) and the Resistive Plate Chambers (RPC). The three stations of RPC have 380,000 chan-

nels, cover $3659m^2$ and are located in the barrel. In the end cap, the three stations of TGC have 440,000 channels and provide the trigger. The RPC applies a uniform electric field of 5000V/mm that amplifies the ionization electrons. Each RPC has a 2mm gas gap between two 2mm thick Bakelite parallel resistive plates. The signal is read out by two orthogonal series of pick-up strips: η strips are parallel and ϕ strips are orthogonal to the MDT. The TGC is a multi-wire proportional chamber with a wire pitch larger than the cathode-anode distance. The gas is a CO_2 and $n - C_5H_{12}$ mixture and with the high electric field of 3.1kV and small wire distance, the TGC is able to have excellent time resolution. The presence of a muon in the spectrometer can be matched to the tracks/energy deposits left in the inner portions of the detector to confirm particle identification.

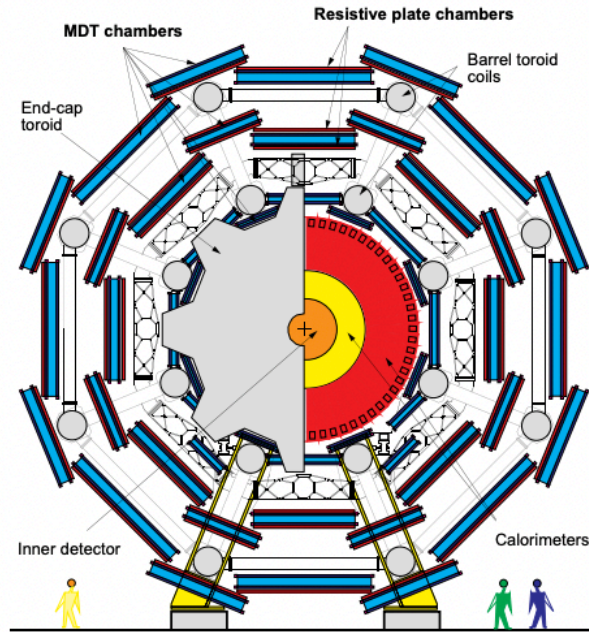


Figure 2.17: Transverse view of the Muon Spectrometer with the Resistive Plate Chambers and Monitored Drift Tubes in the barrel portion. [19]

2.2.4 MAGNET SYSTEM

Magnets are used to apply fields that bend the charged particles in order to determine their charge and measure their momenta. The bending of tracks is caused by the Lorentz force: $F = qE + q(v \times B)$. There are three main components to the magnet system: Barrel Toroid, End-cap Toroid and Central Solenoid Magnet. Figure 2.18 shows the three magnet parts in relation to the cylindrical geometry of ATLAS. The

toroidal magnets are used for muon momentum measurements and tracking. The toroids are both made up of a series of eight coils in a radial or symmetric symmetry around the beam axis. The barrel toroid is 25.3m long, with an outer diameter of 20.1m and can supply a 0.5T magnetic field. The End-Cap Toroids are each 5m in axial length, with a 10.7m outer diameter, and have a 1T magnetic field [14]. The Central Solenoid magnet is 5.2m long with a 2.4m diameter and supplies a 2T magnetic field to bend charged particles in the Inner Detector for momentum measurements [80]. Due to the Central Solenoid's position right in front of the EM calorimeter, the Central Solenoid and LAr calorimeter share a vacuum wall to eliminate material usage [81].

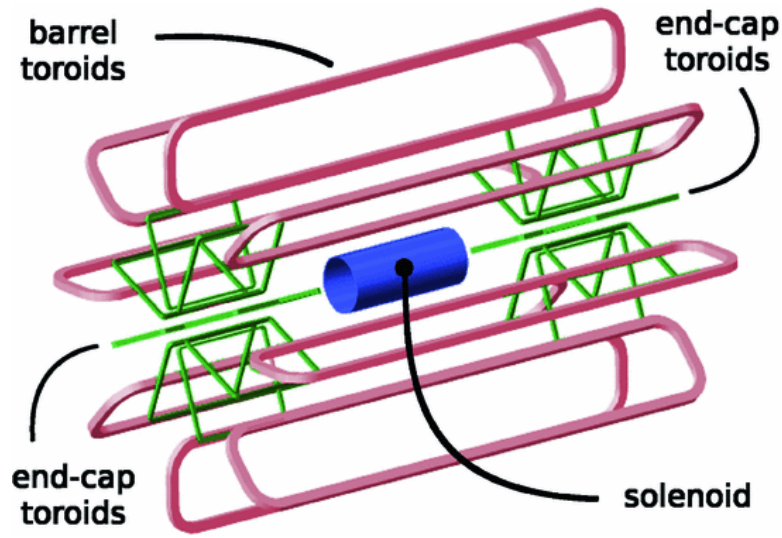


Figure 2.18: ATLAS Magnet System. [20]

2.3 TRIGGER AND DATA ACQUISITION (TDAQ)

During Run 2, the LHC delivered collisions at 40MHz, with particles colliding every 25ns, and an average of 33.7 interactions per bunch. This means ATLAS observes a combined data volume of 60 million megabytes per second as events are 1 – 2MB in size on average. It should be clear that there are many orders of magnitude more events occurring at ATLAS than can both be stored and that are useful for new analyses. In addition to the challenge of high volumes of data, each event is complicated due to pileup and the large number of particles moving through the detector. Figure 2.19 demonstrates the complexity of a single event from 2015, showing both the number of particles/jets emitted and the number of event vertices along the beam line occurring at once. ATLAS needed a fast acting, accurate trigger system in

order to save the events of interest within the bandwidth budget for Run 2. This was accomplished with the combination of a Level 1 (L1) hardware trigger system and High Level Trigger (HLT) software based trigger system and a trigger menu of about 1500 event selections. ATLAS data was then saved at an average rate of 1kHz [86].

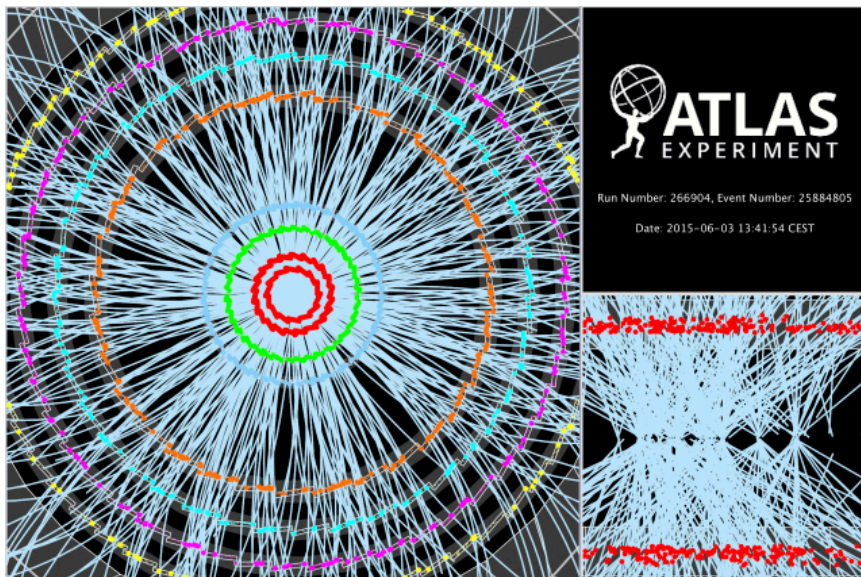


Figure 2.19: Event Display from ATLAS in 2015 at the start of Run 2. A view directly at the beam line axis, shows the number of charged particle trajectories (curved lines) in the tracking detectors (**left**). A view facing perpendicular to the beam line axis (transversing along it) showing the same event is in the bottom right. There are a number of particle interactions coming from separate vertices. Some of these vertices should correspond to lower energy collisions or “pileup” events that are background to more interesting particle analyses. [21]

There are three different levels of the ATLAS trigger system: a hardware based L1, a software based L2, and the software based event filter. The software based L2 and event filter are effectively merged together and considered one step known as the High Level Trigger. The L1 trigger takes in information from the muon spectrometer and calorimeters and finds Regions-of-Interest (ROIs) in η and ϕ for the event. The L1 Calo trigger reads analogue output from the calorimeter and identifies electron, photon, tau-lepton, and jet candidates, as well as an initial guess for the missing energy. The L1 Muon trigger takes hits from the RPC and TGC systems. This information is fed into the Central Trigger Processor, which adds a dead time to meet read-out latency constraints. There are $2.2\mu s$ available to make a L1 trigger decision and the L1 trigger can keep up to 100,000 events per second.

The first part of the High Level Trigger (HLT) is the L2 trigger. The L2 has $\sim 75ms$ to make a deci-

sion and the aim of the trigger is to select e, γ, μ, τ , jets, and MET, among other particles/objects. It uses 40,000 CPU cores to execute algorithms similar to offline reconstruction around those ROIs identified in L1. The Event Filter is the second stage of the HLT. It has one second per event to make a decision and its software closely matches the offline software structure. It uses the most recent data for the magnetic field map and calibration and alignment information to adapt the algorithms. The HLT eventually selects ~ 1000 events per second that are then further analyzed with the Offline Software.

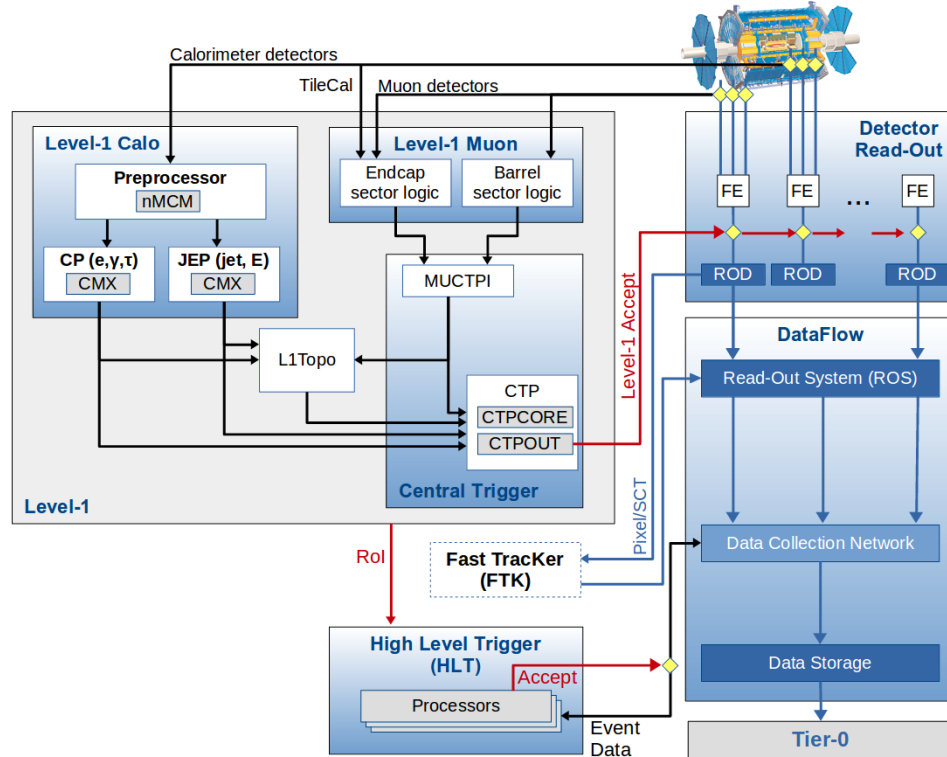


Figure 2.20: ATLAS TDAQ system in Run 2 with the main L1 components boxed in grey and the general flow of detector information to the HLT. The now defunct FTK is also included.

2.4 CERN COMPUTING GRID

In order to store and analyze the particle collision events, CERN needs a massive computing system and the ability to share data throughout the world. To do this, CERN has its own computing center (CERN Data Center shown in figure 2.21) and is also part of the Worldwide LHC Computing Grid (WLCG). The WLCG is made up of 900000 computing cores in 170 sites in 42 different countries. This center runs two million tasks per day and allows data to be easily accessed by any one in the world with a global transfer rate

of 60GB/s. The Grid is organized in a Tier system of four levels, with Tier 0 being the CERN Data Centre, which processes all of the LHC data originally. In each year of Run 2, the CERN Data Centre recorded 30 Petabytes of data and 100 petabytes are permanently archived on tape. Tier 1 has 13 computing centers around the globe and they both store and reprocess the data and distribute it to various Tier 2s. Tier 2 data centers are typically hosted at universities and help with production and reconstruction of simulated events. These are the main computing power for analysis-specific tasks. Tier 3s are the local clusters in a department (or even a personal PC) that access the other parts of the Grid. Data sharing between scientists has come a long way since CERN was the birth place of the world wide web in 1989!



Figure 2.21: A photo of the CERN Data Center with some of its 230000 processor cores and 15000 servers.

“They say that a person’s personality is the sum of their experiences. But that isn’t true, at least not entirely, because if our past was all that defined us, we’d never be able to put up with ourselves. We need to be allowed to convince ourselves that we’re more than the mistakes we made yesterday. That we are all of our next choices, too, all of our tomorrows.”

Fredrik Backman, *Anxious People*

3

Simulation and Computation for Particle Physics Events at the ATLAS Experiment

IN THE PREVIOUS CHAPTER, I DESCRIBED HOW ATLAS IS DESIGNED TO DETECT PARTICLES, BUT HOW DO WE KNOW WHAT WE ARE LOOKING FOR? How do we actually make sense of those hits in the inner detector and energy deposits in the calorimeter? In order to design analyses and interpret our data, we need to first simulate the proton-proton (pp) collisions and potential resulting physics processes so that we understand what specific events will look like. This way we can get a really good idea of what a specific event will look like in our detector and can then try to find it in the data. To do the simulations, physicists incorporate knowledge about SM interactions and decays along with behaviours expected in new theorized models into event generators. From these generated events one can build models for both background-only processes and background-plus-signal hypotheses that can be simulated in the ATLAS detector.

Many experiments looking for beyond-the-standard model physics will add the theory expectation

for their new type of physics on top of SM predictions so that they can rule out or discover the theory with the data. For our $VH, H \rightarrow \tau\tau$ analysis, we are looking for a yet-unobserved SM process, so our background-only model is SM minus our process, and our signal model incorporates the SM process of interest. To complete our analysis, we will compare the data taken from Run 2 with our background plus VH process simulated model. There is a complication here, as one of our main backgrounds is modeled by a data-driven matrix method that will be described in chapter 7.

In order to make sense of those raw read-outs from the detector, we put our data through a complex reconstruction process that assembles particle candidates and event-level quantities. Simulated data also goes through this reconstruction process so that we can compare the data and simulated results, and even use the simulation to help us test our detector's ability to correctly reconstruct particles. The reconstruction process varies for each of our main particles: electrons, muons, taus, jets, etc. Figure 3.1 shows a general flow for both data and simulation through reconstruction and analysis on the LHC Worldwide Computing Grid described in 2.4.

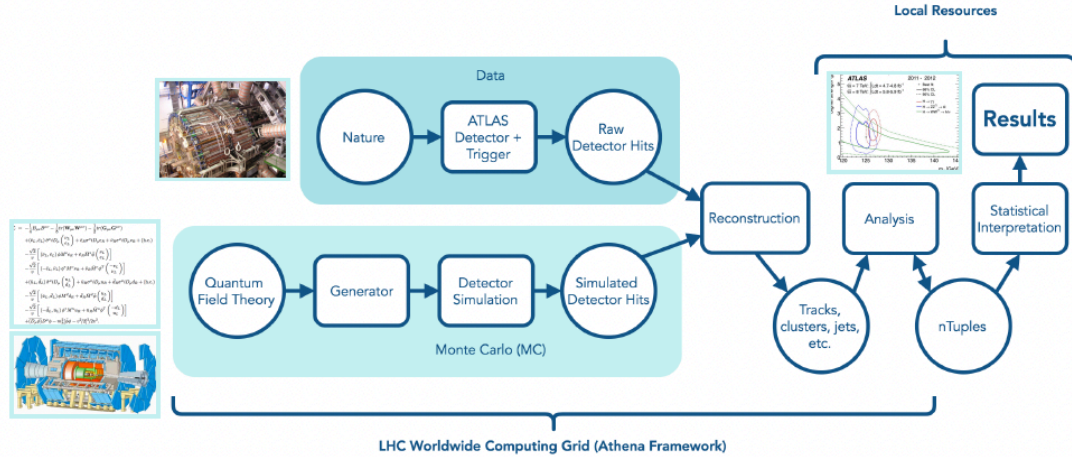


Figure 3.1: Schematic of the data flow for both raw data and simulation. The top row on the left shows the process for data taken by the ATLAS detector and the bottom row on the left shows the process for simulated data. At the stage of reconstruction and interpretation, shown on the right, the data and simulation can be directly compared with each other and combined in various ways in different physics analyses. [22]

3.1 MONTE CARLO SIMULATION

Simulation is critical for the continued success of our experiments. In addition to enabling tests of various physics models in our analyses, simulation is used for checking alignment and operation of the detector, seeing how upgrades will impact our results and whether they are worthwhile and feasible, and to determine what potential new physics we can access. In an ideal world, we would have infinite compute and time resources and be able to model precisely what any physics process of interest would look like in our detector, but this is unfeasible. We can, however, work with comparisons between data and simulation output to understand which approximations are safe for us to include in the simulation. A excellent simulation will run in a reasonable time and be indistinguishable to scientists from the real data we record.

The entire ATLAS simulation program is incorporated into the ATLAS software framework, Athena, and utilizes the Geant4[87] simulation toolkit. There are also some algorithms specifically called “Fast Simulation” that are designed to keep computing times down. Athena uses Python as an interpreter to load C++ algorithms and objects.

There are three main steps in creating the Monte Carlo (MC) simulations of ATLAS data:

1. Generation of event and immediate decays
2. Simulation of detector and interactions of particles
3. Digitization of event

I will describe these three steps in detail in the following sections.

3.1.1 EVENT GENERATION

The first step, event generation, varies in complexity based on the physics being generated. There are two main types of event generators: those that create just the matrix element and those that combine matrix element generation with parton showering. The generators, which are written by theorists, are run within the Athena framework. Though many were originally written in FORTRAN, most have a current version written in C++. Examples of commonly used generators are Pythia[88], Herwig[89], and Sherpa[90]. The generators support broad ranges of physics processes, both SM and beyond, that the user can configure. They produce the range of expected outcomes in terms of the kinematics of generated particles for a given

energy. Some processes, such as those within electroweak physics like W and Z boson production and decay, are calculable, well tested, and well understood. Other processes, like many within QCD, are not exactly calculable and require perturbation theory. The parton showering and hadronization process, in which quarks and gluons are grouped into known hadrons, is particularly challenging to model precisely. The steps of event generation can be seen in the Sherpa collision simulation of a $t\bar{t}h$ event in figure 3.2.

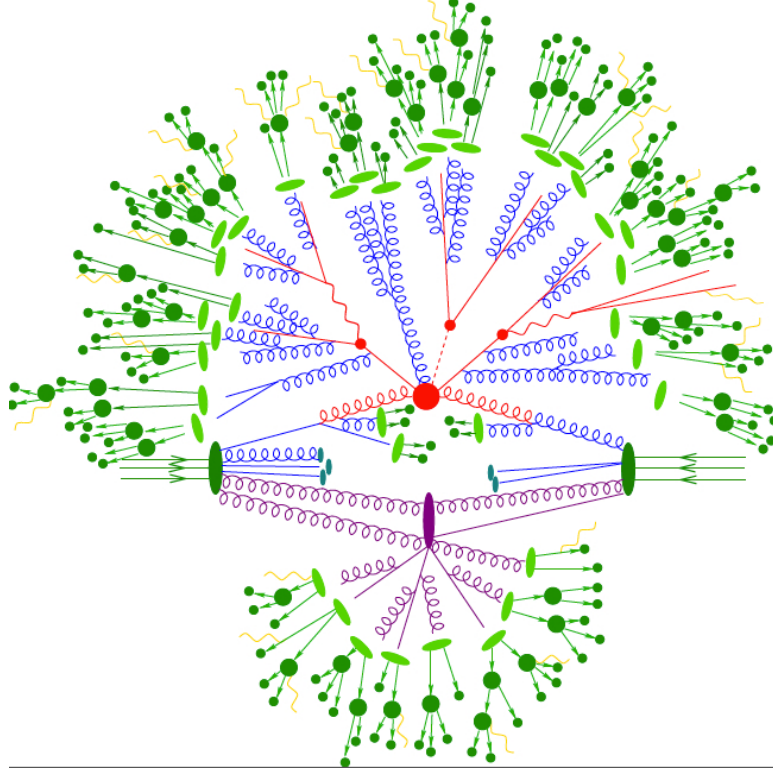


Figure 3.2: Sherpa collision simulation of a $t\bar{t}h$ event. The hard scattering modeled by matrix elements is in red (large red circle), with the smaller red circles representing the top quarks and Higgs boson. The parton showers producing bremsstrahlung are in blue, the purple parts show the secondary interaction, the green is highlighting the hadronization of QCD partons using fragmentation models, the darker green is simulations of unstable decays, and the QED bremsstrahlung (photon radiation) is in yellow. [23]

Events can be filtered at generation time so that only events with a specific property or type of particle are saved. The generation step is fairly quick to compute and the event generator outputs a list of all particles that come out of the proton-proton collision. The list is then reduced to all the stable particles. “Stable” is typically defined as $c\tau > 10mm$, or those particles with a lifetime of at least one hundredth of a billionth of a second. The output particles are stored in C++ classes in the HepMC format [91], which also contains a tree-like structure that includes both the stable particles and many particles along the decay

chains of the processes involved.

The trick for this step is to be able to model quantum processes with a classical decay chain. Pythia and Herwig attempt to solve this by first starting with the hard scattering process to lowest order in QCD and then add in QCD and QED radiation in the shower approximation. Their showering phenomenological model is then used to combine quarks and gluons into hadrons. Sherpa, meanwhile, uses the CKKW duplicate removal prescription to match fixed-order QCD matrix elements with QCD showers. It then uses a Pythia hadronization algorithm.

3.1.1.1 QCD AND PDFs IN PROTON COLLISIONS

Since protons are not a single point object, but a group of three quarks in a sea of virtual QCD particles, modelling their interactions requires extra care. The collision of protons can be split into two main interactions: a hard perturbative component and a soft non-perturbative component. The hard component, or hard scatter, is computed with the SM Lagrangian. Meanwhile, the soft component needs to take into account the complicated structure of the proton.

A key component in these calculations is the relation between the strong coupling constant α_s and Q^2 . α_s is a free parameter in the QCD Lagrangian and determines the interaction strength of gluons and quarks. It depends on the momentum transfer, Q . For pp collisions at ATLAS, we care about the relationship at high Q^2 (short distances, high energy) and our uncertainty regarding the value of α_s factors into our systematic uncertainties in measurements. pQCD (perturbative QCD) is used to describe these high energy processes where asymptotic freedom (confinement) applies. We know that α_s coupling strength gets weak at high Q^2 and strong at low Q^2 (or large distances). As an illustrative limit, if all the momentum of the proton is carried in a single parton (very low Q^2), then we can model the behavior of the proton as a point-like particle.

To understand hadronization and quark confinement, we need to know the behaviour of α_s at low Q^2 (long distances, proton mass scale). The cross-over between perturbative and non-perturbative effects can be set by the QCD scale parameter Λ_{QCD} . Λ_{QCD} is also associated with the quark confinement scale and hadron masses. Non-perturbative QCD occurs at the onset of the confinement scale at low Λ_{QCD} . When $Q = \Lambda$ there is a divergence in α_s (Landau Pole) and the perturbative methods no longer hold.

A renormalization mass scale μ is introduced here in order to compute the pQCD expressions. MC

generators then use Parton Distribution Functions (PDFs) which model the amount of momentum carried by each parton to calculate the cross-sections of these QCD collisions. μ is usually broken down into two terms: the *factorization scale*, μ_F , which determines whether phenomena are grouped into PDFs or into the hard-scatter, and the *renormalization scale*, μ_R , which defines the ultra-violet cutoff scale. These two parameters are typically varied up and down to determine systematic uncertainties. The factorization formula for scattering process at hadron colliders ($ab \rightarrow n$ or $pp \rightarrow X$) is [92]:

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) d\hat{\sigma}_{ab \rightarrow n}(\mu_F, \mu_R) \quad (3.1)$$

where $\hat{\sigma}_{ab \rightarrow n}$ is the parton-level cross section and $f_a^h(x, \mu)$ are the PDFs depending on momentum fraction x of parton a relative to parent hadron h and factorization scale μ_F . This can be rewritten in terms of the matrix element $\mathcal{M}_{ab \rightarrow n}$ which is calculated as a sum of possible Feynman diagrams as:

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_n f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) \times \frac{1}{2\hat{s}} |\mathcal{M}_{ab \rightarrow n}|^2(\Phi_n; \mu_F, \mu_R) \quad (3.2)$$

where Φ_n is the final-state phase space.

PDFs cannot be explicitly calculated, but are instead measured experimentally via fits to pp collision data. These measurements are done at various Q^2 scales then scaled to other values and experimental conditions via the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [93]. Proton PDFs measured using 14TeV data from the LHC for various Q^2 are shown in figure 3.3.

3.1.1.2 PARTON SHOWERS

The partons involved in our pp collisions can also lose energy through radiation of other particles, like gluons and quark-anti quark pairs. Parton showering is caused by an evolution of scales, from high scales and hard scattering processes to low scales and confinement of partons. This Markov process is simulated by Monte Carlo techniques and is calculated with a forward evolution of outgoing particles and backwards evolution of incoming particles. There are two main types of parton showers based on the particles involved. If the showers are linked to incoming particles they are called *initial state radiation*. If the

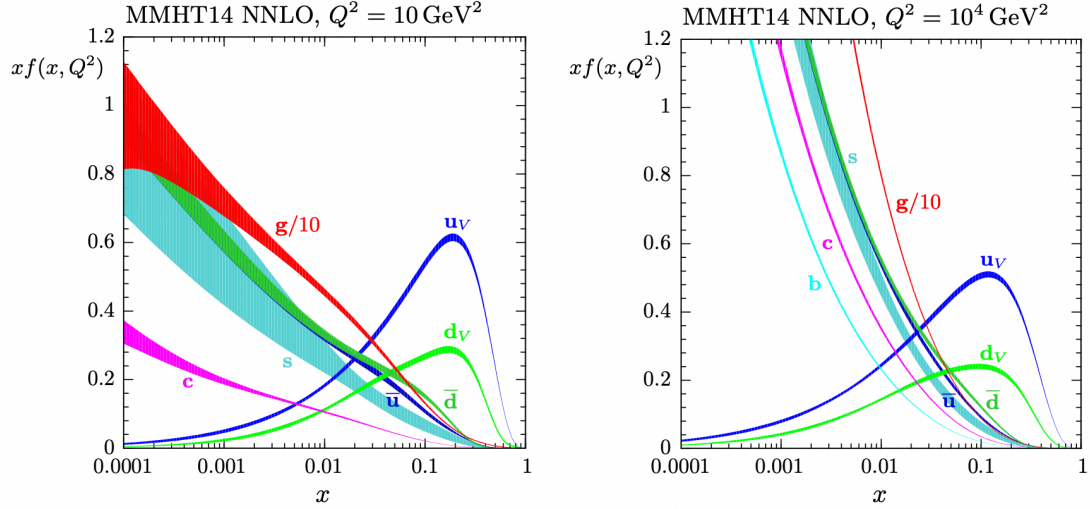


Figure 3.3: Proton PDFs for energy scales $Q^2 = 10 \text{ GeV}^2$ (left) and $Q^2 = 10^4 \text{ GeV}^2$ (right) with 68% confidence-level error bands. [24]

showers are matched to outgoing particles they are defined as *final state radiation*.

Parton showers can also be differentiated based on the two main types of showering: collinear showers that follow the main vertex of outgoing particles, or soft showers. This becomes an important consideration in the jet reconstruction algorithms described in section 3.3.3. In the collinear approximation for a hard process with cross section σ_0 that produces partons of flavor i , we can calculate the cross section for that process to radiate a parton j with momentum fraction z as [92]:

$$d\sigma \approx \sigma_0 \sum_{\text{partons}, i} \frac{\alpha_S}{2\pi} \frac{d\theta^2}{\theta^2} dz P_{ji}(z, \phi) d\phi \quad (3.3)$$

where ϕ is azimuthal angle, θ is the opening angle between partons j and i , and P_{ji} are flavor- and spin-dependent functions. Some of the splittings they correspond to are $q \rightarrow qq$ (P_{qq}), $q \rightarrow gq$ (P_{gq}), $g \rightarrow gg$ (P_{gg}), $g \rightarrow q\bar{q}$ ($P_{q\bar{q}}$). In the MC generators, equation 3.3 is repeatedly applied to get arbitrarily many parton splittings and final state particles.

Soft showers come from the emission of a soft gluon and should be calculated as if the gluon is emitted from the scattering process as a whole, and not an individual parton. However, this can be approximated with the collinear parton shower algorithm by setting the opening angle as the evolution scale.

3.1.2 DETECTOR SIMULATION

The second step in event simulation for the ATLAS detector is simulating the way the particles created during event generation will interact with the detector. This is mainly handled by the Geant4 toolkit. For this step, we need an accurate depiction of the detector including all materials, cables, empty spaces, and the detector’s magnetic field map. Incorporating all of those thousands of cables, tiles, and wires described in section 2.2, would be too computationally expensive, so some simplifications are made that do not have significant impacts on the results. These simplifications will need to be re-evaluated as we move to higher energies when some approximations that were initially determined to have minimal impact could become more problematic.

Next, we include all the physics models of the processes that occur in the detector, such as the photoelectric effect, bremsstrahlung, Compton scattering, particle decays and other interactions. Again, we cannot implement detailed descriptions of every process. We must choose the most important to implement in detail and then use the speediest way to produce results that most closely resemble the expected output. The simulation of pp collisions in our detector is therefore an optimization problem on several fronts. There is active interest in the ATLAS collaboration regarding applying machine learning to improve simulation.

The final step in detector simulation is deciding which parts of the detector interaction we need to save. This involves teaching Geant4 what are the most important or sensitive parts of ATLAS. The output information is then saved in two ways. The generator-level information, also known as “truth” information, is saved for detector calibration and checking our understanding of the physics model. The truth information also includes the links between the generator-level particles and the reconstructed objects, such as their associated tracks and energy clusters. We also save the detector outputs consisting of the energy deposits, times, and locations of these interactions in the detector for the next step of simulation.

3.1.3 EVENT DIGITIZATION

The last major step in event simulation is event digitization, which takes the simulation output and translates it into a format that is consistent with raw data outputs from the detector. This involves changing energies into voltages and currents for each respective detector portion and also incorporating known de-

detector effects, such as gaps and dead material, that may vary these values. This step, which is generally steered by Python code, results in the production of raw data objects (RDOs) and simulated data objects (SDOs). The RDO should match detector output and the SDO should include truth information.

Effects from “pileup” are accounted for during the digitization step by pre-simulating extra pp collisions and overlaying them. This is also the point at which we add in additional backgrounds, like cosmic rays, long-lived particles from other beam-crossings, beam halo effects, and backgrounds due to the drift of ionized gas in the beam pipe [94].

3.2 PARTICLE RECONSTRUCTION

After we collect raw events in our trigger and data acquisition system described in section 2.3 or simulate data as demonstrated in section 3.1, we need to reconstruct the particle hits and energy deposit information into particle candidates. These will then be fed into dedicated particle identification algorithms as described in section 3.3. The goal of particle reconstruction is to group together the detector-specific quantities so that identification algorithms can be applied. There are three main steps in this process: tracking, vertexing, and calorimeter cluster reconstruction.

3.2.1 TRACKING

The goal of the tracking algorithms is to reconstruct the charged particle trajectories. This is done with a series of algorithms that apply filters, pattern matching, and clustering of hits from the tracking detectors [95], [25]. The combining of pattern finding and track fitting into a single process is known as the ATLAS New Tracking (NEWT) approach [96]. The tracking in the Inner Detector is a two step process: *inside-out* track reconstruction and *outside-in* tracking.

The first step in the *inside-out* algorithm is applying a connected component analysis to group adjacent energy deposits above a certain threshold in the pixels and strip sensors of our detector. Three-dimensional measurements, referred to as *space-points*, are made from these hit clusters. In the innermost pixel layers, we already have three-dimensional coordinates due to the granularity of the detector tiles. In the SCT layer, we need both sides of the back-to-back strips to get a 3D measurement for the space-points. Since the TRT straws are 2D objects, we have coordinates for space-points in the $(r - \phi)$ direction in the barrel and in

the $(\phi - z)$ plane in the end-caps. The lower spatial resolution in the TRT is offset by the high number of hits.

Once we have the space-points we can begin iterative track finding. To start we need a track seed, which is a group of three space-points. There are various levels of final track quality that depend on which sub-detector the seed originates in. Therefore, we start by considering the seeds with SCT-only hits. We then look at pixel-only seeds before finally considering seeds made by combining together the SCT and pixel hits. A combinatorial Kalman filter [97] is then used to create track trajectories starting from those seeds and adding in possible space-points from the rest of the Inner Detector. There may be more than one track candidate per seed if there are multiple compatible trajectories. Furthermore, a hit can be associated to more than one track, there can be incomplete tracks, and tracks can originate from more than one particle. These potential complications require an ambiguity solver to best assign hits to tracks and therefore optimize the tracking performance.

First, we determine a track score for each track based on quality criteria such as the number of clusters, the number of holes, the p_T of the track (higher momentum is better), and the χ^2 of track fit. There are both single-track and merged-track clusters, and these can be labeled if running on MC with truth information. Clusters cannot belong to more than two tracks and a track cannot have more than one shared cluster. We expect some shared clusters between track due to the high density environment, but if the clusters that are shared are not identified as merged clusters, we have evidence that they may have been mis-assigned. Tracks must also meet a basic set of standards including: $p_T > 400\text{MeV}$, $|\eta| < 2.5$ (the ATLAS detector does not have good coverage outside this range), at least seven hits in the SCT and pixel layers (up to 12 hits are possible), and no more than one hole in the pixel and two holes in the pixel and SCT combined, where a “hole” means that no hit was found in a particular location even though one was expected to be there based on the track trajectory. To aid in solving ambiguity, a neural network was developed that has 90% efficiency at resolving clusters shared by two particles, and 80% efficiency at resolving clusters shared among three charged particles. The general sequence of the ambiguity solver is shown in figure 3.4.

After the ambiguity solver, tracks are extended out to the TRT and added to the final track collection. As you can see in figure 3.5, track reconstruction efficiency is very high for most p_T values.

In order to reconstruct tracks that come from secondary vertices and/or processes like bremsstrahlung

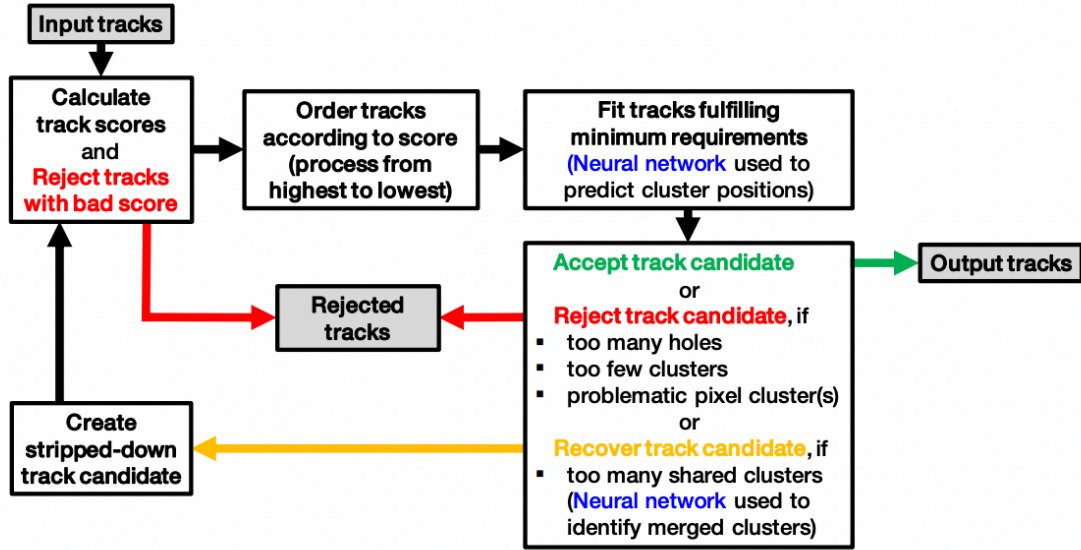


Figure 3.4: Sequence of steps in the ambiguity solver including track scoring, track requirement applications, and the NN for merged clusters. [25]

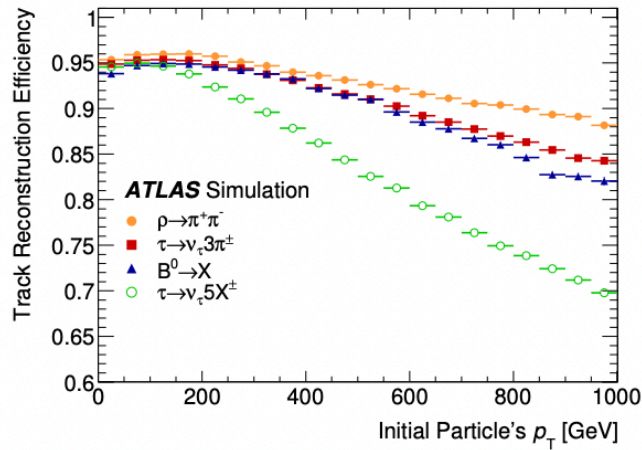


Figure 3.5: Efficiency of track finding as a function of p_T for tracks from a ρ , 3 and 5 prong τ , and a B^0 . [25]

and photon conversion, the second stage of the track finding algorithm - the *outside-in* portion is run. This starts with TRT hit sequences and moves inward to the silicon detectors.

3.2.2 VERTEXING

Tracks passing the initial track-finding algorithms are then used to help find the primary-interaction vertices [98] [99]. The *iterative vertex finding* approach starts by selecting a seed for the vertex by finding the position along the z-axis (beamline) with the highest track density. Then the *adaptive vertex fitting* algorithm, which is a χ^2 algorithm that takes the input seed and tracks to find the true vertex position, is run. Each input track is given a weight that measures its compatibility with the current vertex position. The vertex position is then recalculated using the updated weights for the tracks, and so on. The more compatible a track is with the vertex, the less pull on the vertex position the track will have, and consequently it will be weighted less.

Once this process is concluded, tracks that are incompatible with the final vertex position, determined as being further than 7σ away will be removed from that candidate and put into a pool of unmatched tracks to find other vertices. This vertex finding procedure continues until all tracks are matched to a vertex or no new vertices can be created. All vertices must have at least two tracks associated with them. After this process, track vertex association working points are assigned. The three levels are Loose, Nominal, and Tight. These working points reflect a trade-off between efficiently rejecting background tracks and efficiently keeping the tracks that truly match the vertex.

To determine the position resolution of a vertex, the *split-vertex* method (SVM) is used, in which the tracks associated to the vertex are ranked based on their p_T . The vertex fit is then applied to the even- and odd-numbered tracks separately and the difference in final vertex position gives the resolution. Primary vertices generally have a resolution of about $30\mu m$ [98]. As pileup increases, vertex reconstruction efficiency decreases. This is mainly due to merging of adjacent vertices, which becomes a noticeable affect when pileup goes above 40 overlapping interactions.

At the end of the track fitting and vertex finding process, we can define a track by its track (*perigee*) parameters that are based on its primary vertex: $(d_0, z_0, \phi, \theta, q/p)$ as shown in figure 3.6. d_0 and z_0 are the transverse and longitudinal impact parameters, which measure the distance between the track (perigee) and the primary vertex. θ is the polar angle and ϕ is the azimuthal angle of the track from the vertex. q/p is the charge momentum ratio defined by the curvature of the track. This ratio defines the helical structure of the track.

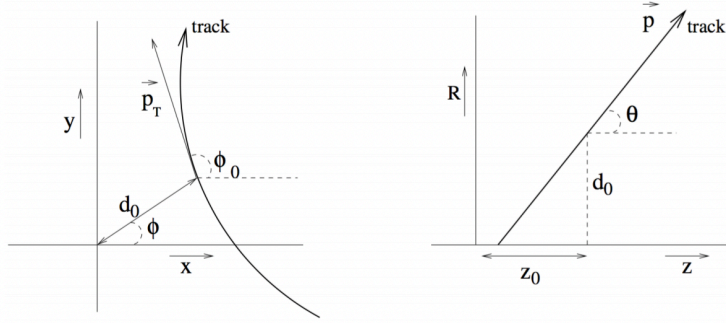


Figure 3.6: The five track perigee parameter definitions.

3.2.3 CALORIMETER CLUSTER RECONSTRUCTION

The other main detector readout that needs to be partially reconstructed before particle-specific identification algorithms are applied is the calorimeter energy deposits, which need to be clustered together. There are two main types of clustering algorithms: *sliding window* and *topological*.

The sliding window algorithm creates clusters from both the EM and hadronic calorimeter in a three-step process: tower building, pre-cluster finding, and cluster filling. To begin, the calorimeter is divided into a grid of $N_\eta \times N_\phi$ squares of size $\Delta\eta \times \Delta\phi$. The energy inside each square is summed into a “tower” energy. This algorithm saves the CaloTower information in addition to the individual calorimeter cell contributions.

The next step is seed finding, in which a window of size $N_\eta^{window} \times N_\phi^{window}$ is slid across the towers and any time the energy in that window is a local maximum above a set energy threshold, a pre-cluster (seed) is created. The location of the center of the pre-cluster is set using energy-weighted η and ϕ centers. Duplicate pre-clusters are removed, with the higher energy one being saved [100]. The final cluster is built by adding all cells in a window of $N_\eta^{cluster} \times N_\phi^{cluster}$ around the seed together. These clusters are key for electron and photon (egamma) identification.

The goal of topological clustering is to group neighboring cells that have an energy above the expected background level into proto-clusters. There are different energy levels that need to be met for a seed to be defined, and then for growth and boundary cells to be included in the final cluster. These energy require-

ments are determined by the cell signal significance:

$$\varsigma_{cell}^{EM} = \frac{E_{cell}^{EM}}{\sigma_{noise,cell}^{EM}} \quad (3.4)$$

For a cell to be a seed, the absolute value of the cell significance must be greater than four, to be a growth cluster it must be greater than two, and to be the boundary it needs to be greater than zero [26]. Cells are continually added to the proto-cluster as long as the energy requirements are met, and proto-clusters can be merged together. On the other-hand if a merged cluster has two isolated energy maxima, that is cause for cluster splitting with the cells between the maxima split relative to their energies. A final clustered dijet event is shown in figure 3.7, which visualizes how topo-clusters are formed.

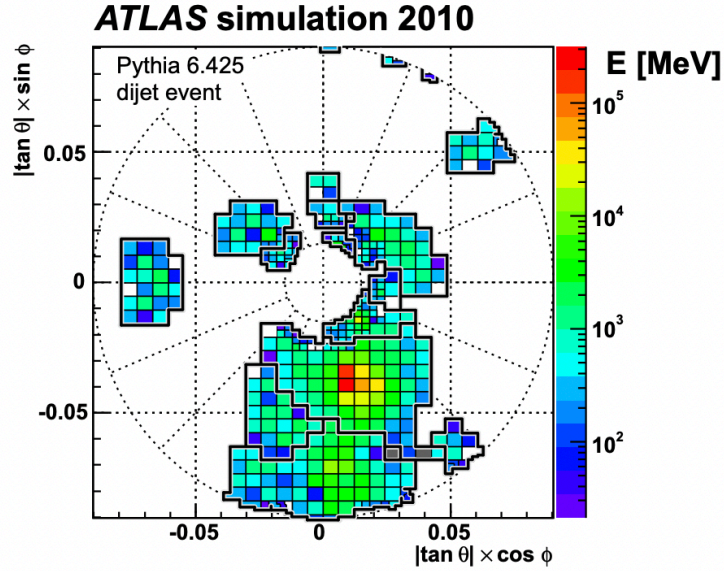


Figure 3.7: Topo-clusters for simulated dijet event (2010) outlined in black. No pileup was included in this simulation. The white cells are negative signal cells and the grey cells have no signal but are completely between topo-clusters. [26]

There are also negative energy cells, which are usually a result of pileup or electronic noise. Including these negative energy cells (and potentially full clusters) can improve detector performance by cancelling out the positive fluctuations caused by pileup. They also help with noise suppression by cancelling local fluctuations. Ignoring negative energy cells would cause a positive bias in our cluster energies.

These calorimeter clusters are then used as inputs to particle identification algorithms. During the particle identification step, their final energy may be tweaked to compensate for other particle-specific

physical effects.

3.3 PARTICLE IDENTIFICATION

Now that we have a collection of interesting tracks and clusters for each event, we can begin identifying which physical objects were present! There is a specific procedure for each type of final state object and the missing energy in the event. The algorithms run offline, meaning there is more time to perform complicated algorithms that can recalculate more precise tracks and clusters than in the original reconstruction. Since taus are critical to our analysis, their identification algorithms will be discussed in detail later on in section 5.

3.3.1 ELECTRONS AND PHOTONS

During Run 2, electron and photon clusters in the calorimeter were recreated offline with a more precise sliding window clustering algorithm that has dynamic, variable sized *superclusters*. Electrons are objects that have a supercluster matched to a track (or tracks) in the Inner Detector. A converted photon is a calorimeter cluster matched to a track from a conversion vertex and an unconverted photon is a cluster unmatched to either an electron track or conversion vertex. Tracks are reconstructed as described earlier in section 3.2.1. The superclusters are formed from topocluster seeds as described in section 3.2.3. The full pathway to reconstruction for electrons and photons is outlined in figure 3.8

Cluster Reconstruction Algorithms

The baseline reconstruction for calorimeter clusters is adjusted as follows for electrons and photons. To start, the towers used in the sliding window are of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ (matching the granularity of the second layer of the LAr electromagnetic calorimeter). The energy deposited in the tower is calculated by summing over the first, second, and third layers, and the presampler if $|\eta| < 1.8$. To be considered a topocluster seed this energy must exceed 2.5GeV when summed over a 3×5 grid window. Now, to rebuild the clusters for electrons and converted photons, the sliding window is set at 3×7 in the barrel and 5×5 in the end-cap. For unconverted photons, the sliding window in the barrel is reduced to 3×5 [101]. Cells in those windows are added to the cluster if their energy is above a set threshold. The efficiency of this reconstruction ranges from 64% at $E_T = 4.5\text{GeV}$ to 99% at $>15\text{GeV}$ [28].

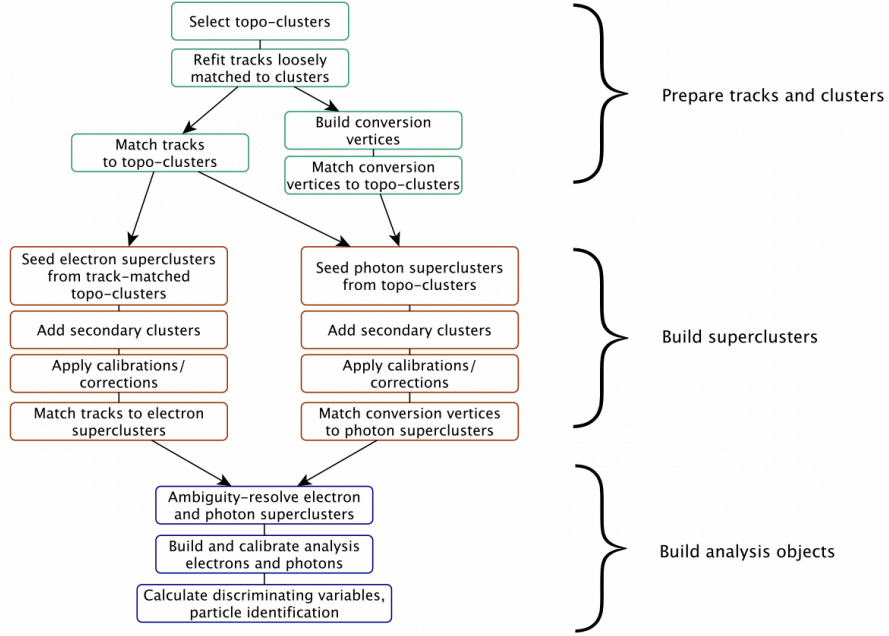


Figure 3.8: Flow diagrams for the reconstruction algorithms of both electrons (left-side) and photons (right-side) showing the similarities and differences between them. [27]

The supercluster reconstruction works separately for both electrons and photons. First, the list of EM topoclusters from the baseline reconstruction is ordered in terms of E_T and then those with a $E_T > 1\text{GeV}$ for electrons and $E_T > 1.5\text{GeV}$ for photons are selected to be a supercluster seed. An electron seed additionally requires that a track is matched to it with at least four hits in the silicon tracker. Once a seed is selected, a search for satellite clusters begins. In order to be selected as a satellite, the cluster must be within a window of $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ around the barycenter of the seed [27]. We add these satellite clusters to capture energy from EM showers. Electron clusters also allow satellites within an extended window of $\Delta\eta \times \Delta\phi = 0.125 \times 0.300$ if the satellite's best-matched track is also the best-matched track for the seed cluster.

To ensure that an object is not reconstructed as both a photon and an electron an ambiguity algorithm is applied. This is particularly important because superclusters for photons have no track requirement applied and are created separately from electrons, so there is risk of overlap. An overview of the ambiguity algorithm is provided in figure 3.9.

Figure 3.10 shows the path of an electron through the detector and the path of the photon produced

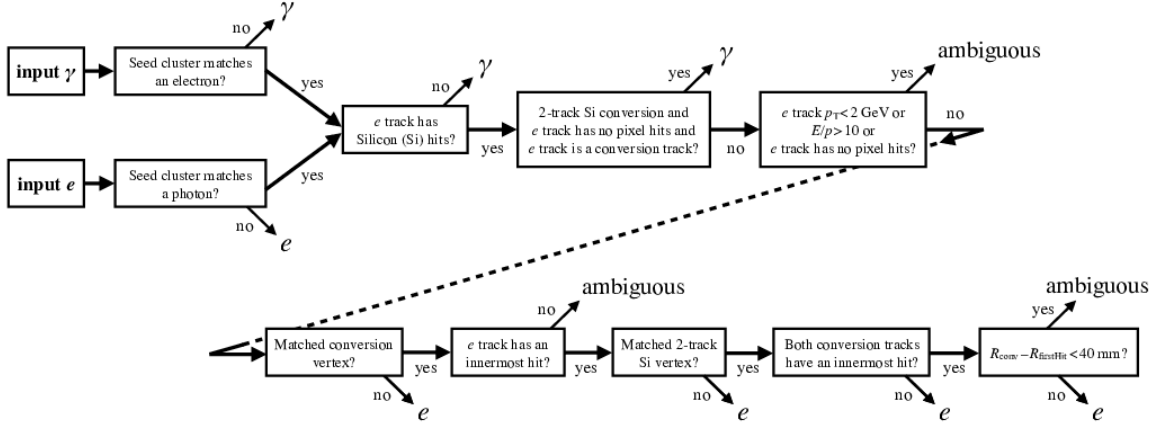


Figure 3.9: Flow of the ambiguity algorithm to resolve superclusters reconstructed as both electrons and photons. It includes information from the trackers and potential conversion vertices. [27]

by its interaction with the detector. The efficiency for electron reconstruction is shown on the left of figure 3.11 for a range of E_T values.

3.3.1.1 ELECTRON IDENTIFICATION

Once a matched track and cluster are identified as a potential electron candidate, they go through a further identification process to separate them from other jet signatures and photons. A data-driven likelihood discriminator is used to set four working points for electron identification. The inputs to the likelihood include track information such as the number of hits in each layer, impact parameter significance, and momentum loss, cluster variables such as the ratio of energy deposited in different layers, and track-cluster variables such as the ratio of cluster energy to track momentum. The full list can be found on page 18 of reference [28]. The four operating points are VeryLoose, Loose, Medium, and Tight, based on increasing

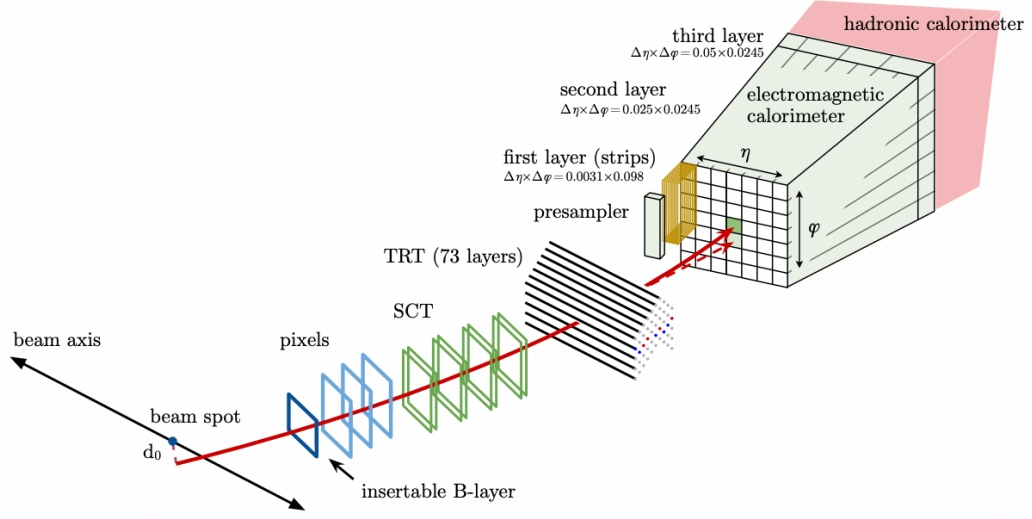


Figure 3.10: Path of an electron (red line) through the ATLAS detector where it passes through the inner tracking detectors and enters the EM calorimeter. The dashed red line shows the path of a photon produced during the electron's interaction with the detector. [28]

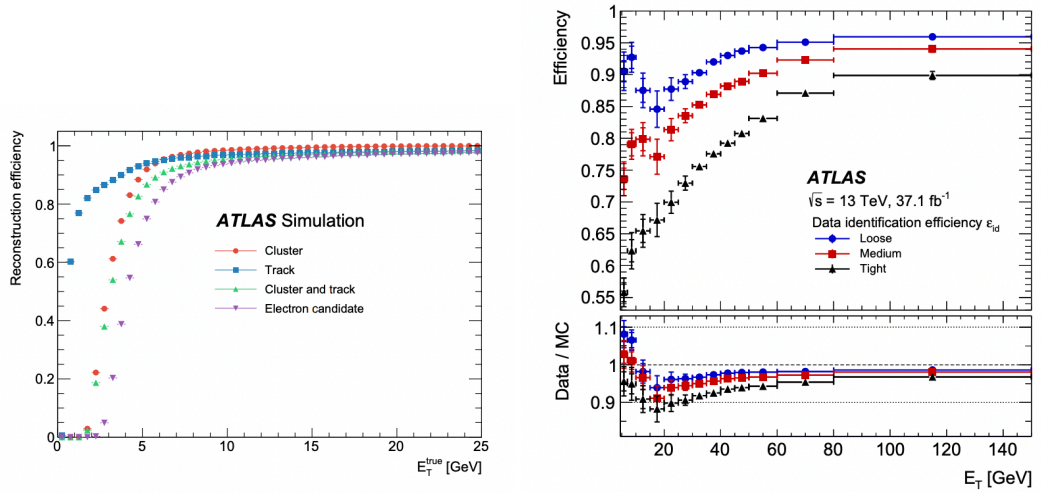


Figure 3.11: Reconstruction efficiency for the cluster, track, cluster and track, and electron as a function of E_T (left). Electron Identification efficiencies for the three working points as a function of E_T in a $Z \rightarrow ee$ sample (right). [27]

thresholds in the likelihood discriminant. They are defined in nine bins of $|\eta|$ and 12 bins of p_T [27]. Loose, Medium, and Tight electrons all have at least seven track hits, with at least two in the pixel detector. Medium and Tight electrons additionally require one hit in the innermost pixel layer. The efficiencies for

the various working points as a function of E_T for a $Z \rightarrow ee$ sample are shown on the right hand side of figure 3.11.

Additionally, there are isolation working points that can be applied to reject electrons with other objects in a cone of ΔR around their cluster position for calorimeter-based isolation or a distance of R_{max} from their track for track-based isolation. For electrons with $E_T > 10\text{GeV}$, isolation efficiencies range from 90-99%. A charge misidentification Boosted Decision Tree algorithm is also applied to make sure the charge is correctly reconstructed. Wrong charge calculations can occur from incorrect measurements of track curvature or from a mis-matched track usually due to bremsstrahlung with multiple tracks occurring very close together. The BDT algorithm has an efficiency greater than 95% for all values of E_T and from 90-99% for all values in η [27].

For our analysis, we require electrons to be Tight identified and isolated.

3.3.1.2 PHOTON IDENTIFICATION

Photon identification uses a cut-based selection algorithm. Shower shape variables are used as input. Information from the first layer of the EM calo is particularly important in rejecting π^0 s. 20-65% of photons will convert in the Inner Detector depending on their position relative to $|\eta|$, with higher $|\eta|$ meaning that the photons pass through more material and are therefore more likely to convert [27]. Converted photons will have a track matched to a cluster, while unconverted photons will just have a cluster. Remember, photons are electrically neutral particles so they will not leave a track in the tracker unless they convert. Photons in the crack region ($1.37 < |\eta| < 1.52$) or with an $|\eta| > 2.37$ are vetoed. The three levels of photon identification are Loose, Medium, and Tight. The working points are optimized separately in bins of $|\eta|$. The Tight identification is also optimized in bins of E_T by an E_T dependent TMVA algorithm. The efficiencies for both converted and unconverted photons from jets are shown in figure 3.12. There are also three levels of photon isolation based on calorimeter and track isolation variables.

3.3.1.3 ENERGY CALIBRATION

The energy contained in the reconstructed electron and photon clusters is not equal to the actual energy deposited by those particles and thus needs to be calibrated. The calibration process begins by estimating the true energy from the energy deposits in the calorimeter with a multivariate regression algorithm that

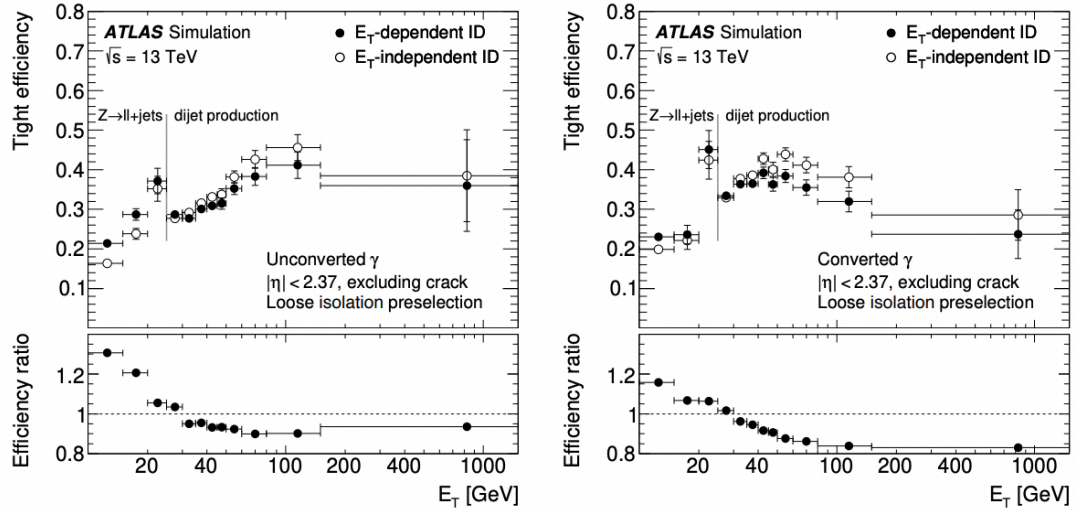


Figure 3.12: Efficiency of Tight photon identification for both converted (left) and unconverted background photons (right) from jets in a $Z \rightarrow ll + jets$ sample for $E_T < 25\text{GeV}$ and from a filtered dijet sample for $E_T > 25\text{GeV}$. The bottom panel is the ratio of E_T dependent and independent algorithms. [27]

has been trained on simulation. Then there is an adjustment applied based on the relative energy scales of the various layers of the calorimeter. This adjustment was determined by studying muon energy deposits and electron showers. Then detector effects, such as non-uniformity of the calorimeter cell response, are added to the energy calibration. Finally there is an overall energy scale adjustment that was derived by looking at the difference between simulation and data in $Z \rightarrow e^+e^-$ samples.

The energy calibration performs quite well, with a systematic uncertainty due to the energy calibration of 0.3-0.8% in electrons with $E_T = 10\text{GeV}$ and 0.03-0.2% in electrons with $E_T = 45\text{GeV}$. For photons with $E_T = 60\text{GeV}$, the systematic uncertainty from energy calibration is 0.25 to 1% [102].

3.3.2 MUONS

Muons are minimally-ionizing particles in our energy range of interest and therefore deposit very little of their energy in the EM and hadronic calorimeters, instead reaching the Muon Spectrometer (MS), which provides additional information for their reconstruction and identification. The tracks in the Inner Detector (ID) are created using the algorithms detailed in section 3.2.1. The tracking in the spectrometer is performed using an outside-in approach. The straight line fits are performed on the hits found in the MDT

to create segments and a combinatorial approach is applied in the CSC. The segments are connected together based on hit multiplicity and fit quality to form tracks. All tracks (except those in the barrel-endcap transition region) need at least two matching segments to be considered muon track candidates. A χ^2 fit is applied on the hits in the track and a cut is applied to those that have the worst agreement with the fit. Finally, a combined ID-MS reconstruction is applied [103].

There are four main types of muons depending on the subdetectors involved in reconstruction:

- **Combined (CB) Muons** are muons created by a global refit on hits from both the ID and MS. They are mostly reconstructed using an outside-in procedure starting from a track in the MS, extrapolated inward, and then matched to a track in the ID.
- **Segment-tagged (ST) muons** are muons that only pass through one chamber in the MS. A track in the ID is classified as a muon if it is matched to at least one segment in the MDT or CSC.
- **Calorimeter-tagged (CT) muons** are the lowest purity muons and are created from a fitted ID track and a single energy deposit in the calorimeter that matches the shape of a minimally-ionizing particle.
- **Extrapolated (ME) muons** are muons reconstructed only from tracks in the MS that are loosely compatible with the interaction point. They must traverse two (or three in the forward region) layers of the MS. ME muons are included to boost acceptance of muons in the forward region ($2.5 < |\eta| < 2.7$), which is not covered by the ID.

Figure 3.13 shows the tracks in the ID and MS of the four main reconstructed muon types. Overlaps are resolved in the order they are presented above.

Muon identification is then performed by applying additional quality requirements to the tracks. This is especially important to differentiate prompt muons from non-prompt muons that occur from in-flight decays of charged hadrons and usually have a ‘kink’ in their track path. This leads to non-prompt muon tracks having incompatible momentum measurements in the ID and MS. The main variables used in muon identification are:

- **q/p significance:** absolute value of the difference between the ratio of the charge and momentum of muon tracks in the ID and MS divided by sum in quadrature of their uncertainties
- ρ' : difference between p_T measurements in ID and MS divided by total p_T
- **normalized χ^2 of fit**

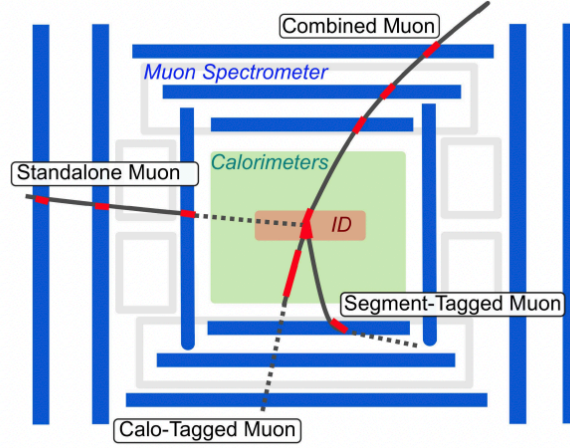


Figure 3.13: Schematic showing the four main types of muons from the various reconstruction algorithms. [29]

There are four identification working points: Loose, Medium, Tight, and High- p_T . Tight muons are used in this analysis and are defined as CB muons with hits in at least two stations in the MS, normalized $\chi^2 < 8$, with a two-dimensional cut in q/p and ρ' as a function of p_T for muons with $p_T < 20\text{ GeV}$. The reconstruction efficiency for Tight, Medium, and Loose muons is shown as a function of both p_T and η in figure 3.14.

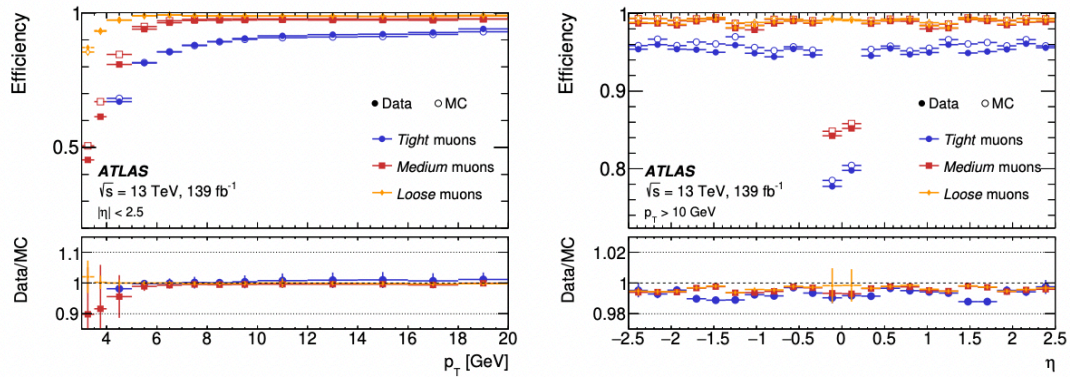


Figure 3.14: Muon reconstruction and identification efficiency for the three working points as a function of muon p_T in $J/\Psi \rightarrow \mu\mu$ events (left) and as a function of η for $Z \rightarrow \mu\mu$ events with $p_T > 10\text{ GeV}$ (right). [30]

For our analysis we also require that the muons are isolated.

3.3.3 JETS

Jets are not actually a fundamental particle, but rather a collimated spray of energy resulting from the fragmentation and hadronization of partons after the pp collisions. Jets are the dominant outcome of pp collisions and they become trickier to resolve as more pileup increases both the number of jets and their energy deposits. However, jets provide a crucial link between the colorless particles we observe in the detector and the underlying QCD physics. Accurate reconstruction allows us to study both the Higgs boson and other short-lived particles produced in these collisions. Jets also allow physicists to study the quark-gluon plasma, especially during the heavy ion collisions portion of the LHC Runs. By studying how jets are modified while travelling through the plasma, physicists can learn more about the plasma itself.

The hadronic decays of tau leptons are a special type of jet. This analysis relies on accurate jet reconstruction to differentiate between actual tau particles and background jets.

Jet Clustering

There are two main types of jet clustering algorithms: cone based and sequential clustering. Jet algorithms are also classified based on their adherence to two physics-derived properties; these are *infrared* and *collinear safety* (IRC). Infrared safe (IR) means the reconstructed jet is invariant to additional soft radiation and collinear safe means that the jet stays the same if we substitute one hard parton for two collinear partons with the same momentum. The various clustering algorithms also have to account for the trade-off in cluster size. A larger jet cluster would contain all the energy from soft radiation, but it also could contain extra energy from pileup and the underlying event.

The inputs to these clustering algorithms are the topo-clusters as formed by the algorithm described in section 3.2.3 with seeds requiring energy deposits above 4σ and neighbor cells with energy deposits above 2σ . The hadronic calorimeter jets go through an initial energy re-weighting known as local cell weighting to account for the different detector response to hadronic and electromagnetic particles [104]. The electromagnetic jets are calibrated by the EM scale. The mass of the jet topoclusters is set to zero as well [105]. Jets can also be constructed from tracks, which have low dependence on pileup, but are limited to $|\eta| < 2.5$ due to Inner Detector coverage.

First, we will cover *cone* algorithms, which were created to reflect the cone-like shape of a jet and result in clusters in $(\eta - \phi)$ space with rigid circular boundaries. These were used early on in the running of

ATLAS because they are quick and easy to compute, but they are not IRC safe. Two common cone algorithms are *IC-PR* and *SIS Cone* [106]. IC-PR is an iterative cone with a progressive removal algorithm that starts by taking the largest p_T cell as the seed and then selecting a cone of radius R around it. Then a trial jet axis is calculated by summing up the cells in that R cone. If that axis matches the seed it is considered a stable jet, otherwise it iterates on the process using the trial axis as the new seed. Once converted, the jet is removed and added to the list of final objects and the algorithm moves on to the next seed. Unfortunately this algorithm is collinear unsafe. The IRC-safe cone algorithm is SIS cone, which checks for particle overlaps within $2R$ of the initial jet and either splits or merges those jets that overlap. Figure 3.15 shows the outcome of SIScone (bottom left), along with the sequential clustering algorithms, where an example of a split cluster is in the top right corner of the plot.

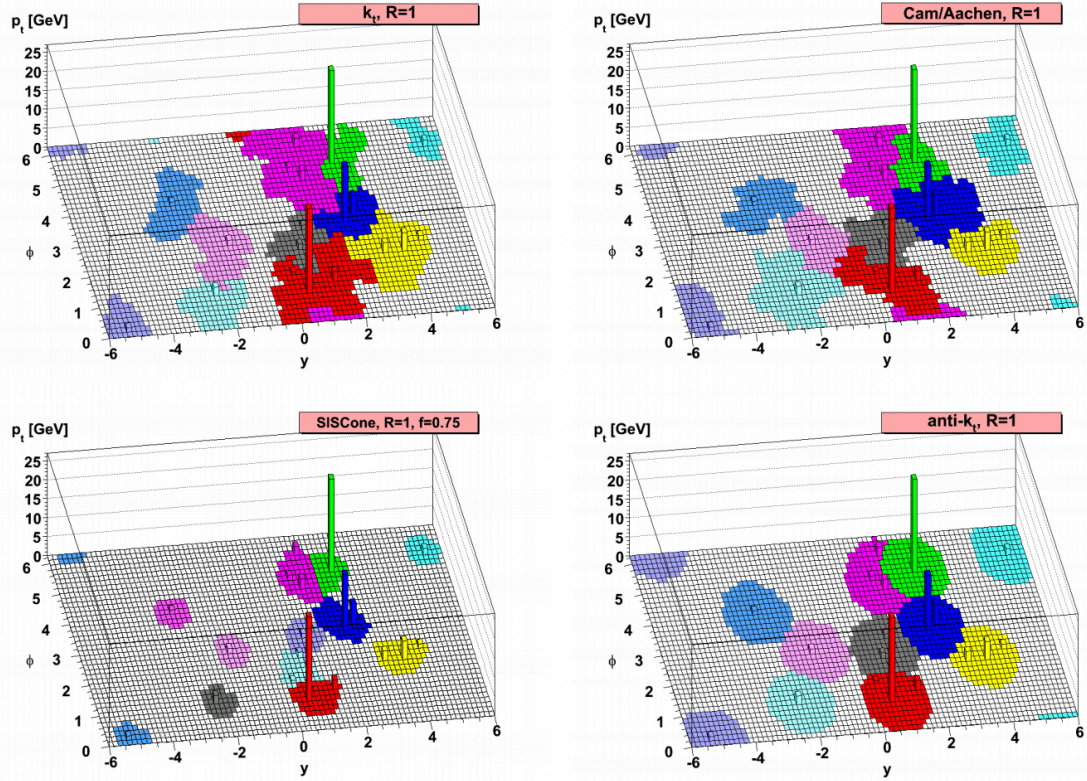


Figure 3.15: Four main jet clustering algorithms performed on a Herwig generated parton level event. The various colors represent the different shapes of the jets. [31]

Sequential clustering is different from the cone-shaped algorithm since it groups jets based on positions in momentum space. Originally, these algorithms were too computationally unwieldy to run but FastJet

[107] algorithms have made it feasible to use them in analyses. Sequential clustering uses two distance variables: the distance between two particles (d_{ij}) and the distance between the particle and the beam (d_{iB}). It finds the minimum of these two distances. If the minimum is the distance between jets it merges those two jets. If the distance between the jet and the beam is the minimum, then it removes that jet and adds it to list of final objects in the event. This process is repeated until all particles are part of a jet with distance between another jet greater than R , or by reaching a maximum number of jets per event. The various sequential algorithms follow this procedure using different definitions for d_{ij} and d_{iB} .

The main algorithm used by ATLAS in Run 2 for jet clustering is *anti- k_t* [31]. For anti- k_t , $d_{ij} = \min(\frac{1}{p_{ti}^2}, \frac{1}{p_{tj}^2}) \times \frac{R_{ij}^2}{R}$ and $d_{iB} = \frac{1}{p_{ti}^2}$. This algorithm is good for resolving jets, but not for studying jet substructure in quark-gluon plasma studies. It also tends to cluster the hard particles first. Another sequential algorithm is Cambridge Aachen (C/A), where $d_{ij} = \frac{R_{ij}^2}{R}$ and $d_{iB} = 1$. This algorithm has no momentum dependence and declusters particles well, but is susceptible to contamination from the underlying event and pileup [106]. The jet clusters formed for an example simulated event by C/A and anti- k_t are shown in figure 3.15 (top right and bottom right). In ATLAS, we run anti- k_t with $R = 0.4$ and C/A with $R = 1.2$ [105].

Jet Energy Calibration

As in previous objects described, we need to calibrate the energy measured in our detector in order to more closely approximate the actual energy of the particle. There are several steps in the jet energy scale (JES) calibration that is applied to EM jets. These corrections all affect the four-momentum of the jet. The first step is the origin correction that ensures that the four momentum of the jet points to the hard scatter origin and not just the center of the detector. Then there is the pileup correction, which has two steps. It first applies a correction of $p_T^{corr} = p_T - A\rho$ where A is area of jet and ρ is pileup p_T density, and then it removes the dependence of the jet energy on the number of vertices. Then there is the absolute correction of the detector response, which is based on MC simulation. Next, a global sequential correction is applied, which removes flavor dependence by looking at variables such as the number of tracks and the amount of energy deposited in each calorimeter. Finally, a residual in situ calibration is applied, which accounts for the difference between data and MC based on simulation studies. This final step is only applied to data. After performing this calibration, the jet energy was measured with a precision of 1% in Run 2 [104]. The

full process of the jet energy calibration algorithm can be seen in figure 3.16.

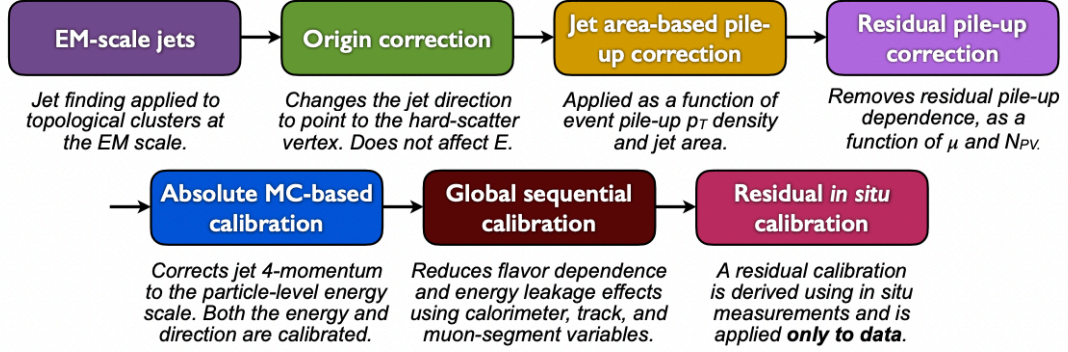


Figure 3.16: The steps of the Jet Energy Calibration. [32]

Pileup Jet Suppression

In order to reduce pileup in our events, we need to remove jets that are not associated to hard-scatter vertices and are therefore simply adding noise to our event. A multivariate algorithm, known as the Jet Vertex Tagger (JVT), was trained to differentiate between hard-scatter and pileup jets [105]. ATLAS also uses the Jet Vertex Fractions (JVF) variable, which is defined as the fraction of the momenta of tracks matched to the jet that are associated with the hard-scatter vertex:

$$JVF(jet_i, PV_j) = \frac{\sum_k p_T(track_k^{jet_i}, PV_j)}{\sum_n \sum_l p_T(track_l^{jet_i}, PV_n)} \quad (3.5)$$

where k runs over all tracks from PV_j matched to jet_i , n over all primary vertices, and l over all tracks from PV_n matched to jet_i . This value is only calculated for tracks with $p_T > 500\text{MeV}$. A $JVF = -1$ means a calorimeter jet with no associated tracks, $JVF = 0$ means that all tracks are from pileup vertices, and $JVF = 1$ means that all tracks come from the hard scatter. The recommended tight cut is $|JVF| > 0.5$ [108].

Jet Flavor Tagging

The final step in jet reconstruction and identification is flavor tagging. The algorithm for b-jet tagging is described in detail in [109]. While our analysis does not require b-jets, some channels do include a specific veto.

3.3.4 MISSING TRANSVERSE ENERGY

Not all particles leave a trace in our detector. Examples of non-interacting particles include SM particles, such as neutrinos, and theorized particles like possible dark matter candidates. We can calculate the amount of energy these kinds of particles must have carried away by carefully imposing conservation of momentum. The colliding protons should have no momentum in the transverse plane, thus the missing transverse energy should be equal to the opposite of the sum of measured transverse energy:

$$\begin{aligned} E_T^{miss} &= -E_T^{tot} \\ E_T^{miss} &= \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \end{aligned} \quad (3.6)$$

Conceptually this seems simple, but in practice it requires extremely careful determination of the energies of all the objects that interact in the detector, including the “soft term” deposits that may not have been clustered with identified objects. To do this, ATLAS uses a Track Soft Term algorithm. This algorithm, which relies heavily on the output of the reconstruction, uses tracks for the calculation of the energy of the soft terms. These soft-term energies are then added to the energies from reconstructed objects, including electrons, photons, muons, taus, and jets. There also has to be a careful rejection of pileup jets with the jet vertex tagging algorithm in order to remove any potential energy imbalance due to pileup as opposed to the hard interaction.

Because the MET calculation relies on reconstructed objects, we need to apply overlap removal to ensure that we do not double-count an object by considering it under multiple labels. Objects with high identification purity are given priority in the overlap removal. The order of precedence is electrons, photons, taus, jets, and unused tracks [110]. Muons do not generally overlap with the other particles as they are usually constructed from tracks in the Inner Detector and Muon Spectrometer. The calculation of missing transverse energy (MET) terms is therefore:

$$E_{x(y)}^{miss} = - \sum_{i \in \{hard\ objects\}} p_{x(y),i} - \sum_{j \in \{soft\ signals\}} p_{x(y),j} \quad (3.7)$$

, i.e. the negative vector sum of transverse momenta of reconstructed physics objects. The MET calculation needs to be adaptable for each analysis because changing the particle identification criteria can result

in needing to change the order of overlap removal.

“The trouble about trying to make yourself stupider than you really are is that you very often succeed.”

C.S. Lewis, *The Magician’s Nephew*

4

Introduction to Machine Learning

SEVERAL OF THE PROJECTS I UNDERTOOK THAT ARE DOCUMENTED IN THIS THESIS INVOLVE MACHINE LEARNING. This chapter serves as a brief introduction to machine learning, and it includes a short primer on the algorithms discussed in later chapters that are then placed in the broader context of a range of topics in machine learning that are used across the ATLAS collaboration.

4.1 INTRODUCTION TO MACHINE LEARNING

At its simplest definition, machine learning (ML) algorithms are algorithms that learn from the data. In other words, as the algorithm gains experience (i.e. is fed the training data over and over), its performance on the task at hand will improve [36]. ML algorithms are able to find patterns and/or interpret data without being given explicit instructions.

The theory of ML, and many common ML algorithms, have been around for a long time. Only recently have we had the computing power to make many of these ML algorithms feasible to train. The perceptron, invented in 1958, was inspired by the workings of a human neuron [11]. Ten years later, the

nearest neighbor algorithm was created, which is an example of an early implementation of pattern recognition [112]. There was then a stretch of time known as the “AI winter” during which people doubted the feasibility of ML because computers were not up to the computational power the algorithms required to be trained. By the 1990s computers had become more powerful and there was a shift of focus to data-driven approaches in which algorithms were trained on large datasets. Deep Learning did not take off until the 2010s as it required vast improvements in computing power. There is a current worry that we may have reached another plateau with respect to computing feasibility.

Some of the tasks that can be done with ML are:

- Classification: determine which category an input belongs to, for example separating a signal from background or recognizing an image
- Regression: predict a continuous variable distribution based on the input data or find the relationship between two values, for example calibrating the measured energy to better resemble the actual energy
- Translation: convert one sequence of symbols or text to another
- Anomaly Detection: flag events or objects that are unusual
- Synthesis: generate new events that are similar to existing ones, for example through Monte Carlo simulation
- Denoising: clean up data with the removal of spurious signals, such as cross-talk or other electronics-generated detector effects

In order to train an algorithm for these tasks we need a sufficiently large data set for the task at hand. There are two main types of learning: *supervised* and *unsupervised*. Supervised learning is when the training set is labeled, either with the correct classification or the value we are aiming for with the regression. Thus the data members contain a label of their target value. Unsupervised learning occurs when there are no truth labels on the data set. We therefore want the algorithm to learn some useful properties of the data. Often this can occur in the form of clustering, denoising, or reducing dimensionality in the data [36]. In this thesis, I only consider supervised learning, so the remainder of the chapter will assume input data is labeled unless stated otherwise.

Before we begin training an algorithm, we usually split our input data into a training set and a test set. The model learns on the training set and is evaluated on the test set. Sometimes, there is a third division

made from the input data set, known as the validation set. The validation set can be used to direct/evaluate the ongoing training, while leaving the test set for a final evaluation once the model is fully trained. The general process for training a model is to select the algorithm and architecture (number of nodes/layers) desired and initialize the parameters with random numbers. The algorithm then classifies or predicts the target value using those random initializations, determines how wrong their prediction is, and self adjusts to reduce the difference between the prediction and the target value. A validation set can be useful to prevent overfitting to the training data, a process known as “over-training”.

Performance is quantified with a loss function, which measures how well that algorithm models the data. The ML algorithm tries to minimize this value. Common loss functions for regression are the Mean Squared Error (MSE), where

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (4.1)$$

and the Mean Absolute Error (MAE), where

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}. \quad (4.2)$$

Here n is the training set size, $\hat{y}$ is the model guess, and y is the true value. Standard loss functions for classification are Cross-Entropy Loss when there are just two categories, where

$$L = -(y_i(\log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i))), \quad (4.3)$$

and the Negative Logarithmic Likelihood, where

$$L = -\frac{1}{n} \sum_{i=1}^n \log(\hat{y}_i). \quad (4.4)$$

Here n refers to the number of observables.

While the goal of training is to minimize the loss as much as possible, our results need to be generalizable to data not encountered in the training process. There is a known trade-off between overfitting and underfitting (see figure 4.1) and an optimal point we want our model to reach. If we do not optimize enough to our training set, then we will under-fit and perform poorly both on seen and unseen data. How-

ever, if we train too much on the input data, the model will perform even more poorly on the unseen data. We can use the validation loss as a proxy for the generalization loss and stop training our model when the validation loss stops decreasing.

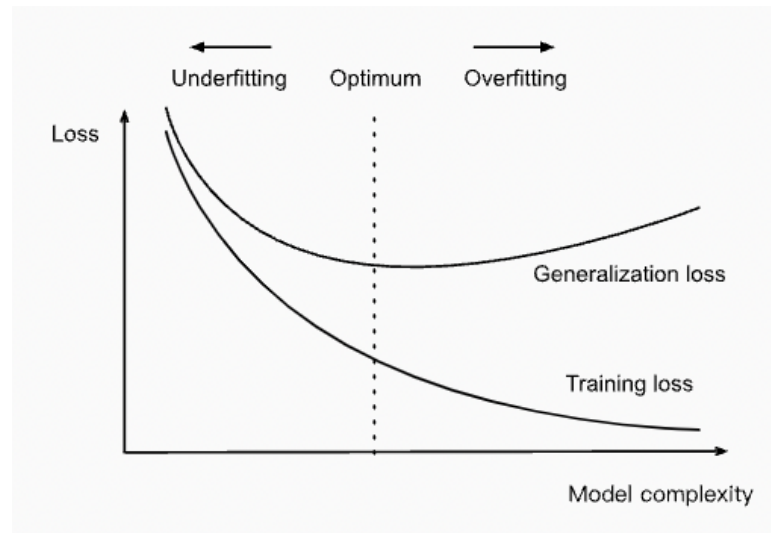


Figure 4.1: The tradeoff between underfitting and overfitting a model when comparing loss and model complexity.

We must also consider the bias and variance trade-off. Bias is a measure of how far off the average prediction of the model is from the correct value. Variance is a measure of the spread of the predictions, or the flexibility of the model. If we overfit the model we will have a high variance as the results depend a lot on the input. Meanwhile, if we do not train the model enough, we will have high bias where the results do not average around the true value we are looking for. The various combinations of bias and variance are shown in figure 4.2. An ideal model will have low bias and low variance, but that can be very hard to achieve. Neither bias nor variance can be fully eliminated as both contribute to the total error of the problem, with $Error = Bias^2 + Variance + Irreducible\ Error$.

There are several methods to combat over-training in addition to early stopping and lowering the bias and variance in the model. One can reduce the complexity of the ML framework in a random fashion in a process known as drop out. In the example of neural networks, which rely on an architecture of connected nodes, drop out could be employed by randomly removing some fraction of the nodes. There is also regularization, which is a modification to the algorithm that reduces its generalization error and not the training error. Regularization can be done in the form of weight decay or added terms to the loss

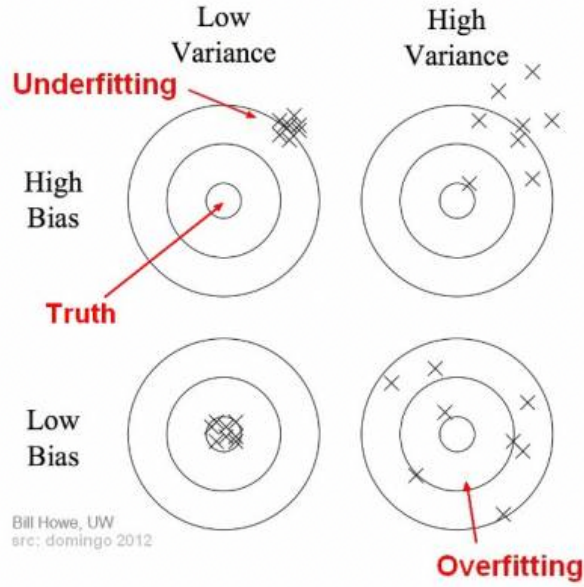


Figure 4.2: The tradeoff between bias and variance in a model, with high variance related to overfitting and high bias related to underfitting.

function that encourage a shrinking of the coefficients in the model [113].

The main types of regularization are Ridge and Lasso. Ridge adds a penalty to the cost term that is related to the squares of the coefficients of the model. A more flexible, though also prone to overfitting, model will have larger weights and this Ridge penalty ensures those weights stay small. The implementation looks like:

$$Cost = Cost + \lambda \sum_{i=1}^n w_i^2 \quad (4.5)$$

where n is the number of weights w to which we apply our penalty λ . The value of λ can be tuned to find the optimal performance, however a too-high value of λ will result in the weights all going to zero and the model not learning anything.

Lasso regression only penalizes the high-value weights and has the ability to send some weights to zero. This can lead to better model interpretability as we remove less important predictors. In this way, it performs a version of feature selection. Lasso regularization is performed by modifying the cost to be:

$$Cost = Cost + \lambda \sum_{i=1}^n |w_i| \quad (4.6)$$

Regularization helps us reduce the variance of the model without raising the bias by much.

Besides the value of the loss function for the training and validation sets, there are many other ways to evaluate the performance of a machine learning algorithm. There is **accuracy**, which, with classification problems, is a measure of how often the model predicted correctly. **Error rate** measures the opposite effect by stating how often we are wrong. In addition, we can look at the **confusion matrix**, which details the number of true positives, false negatives, false positives, and true negatives. This can give you insight into how the model is thinking. Also, depending on what you are doing, you may be more concerned about false negatives than false positives, such as in a model for predicting illnesses. We also look at **receiver operating characteristic (ROC) curves**. In ROC curves, the true positive rate can be seen versus the false positive rate for a set (or various) threshold(s). The higher the area under the curve (AUC), the better the performance of the model.

There are a number of things that we can do to improve model performance. First, we can train on high-quality and high-statistics datasets. We can also tune the settings of the model, known as the hyperparameters. These include the value of the regularization, the learning rate (how fast we take steps in the “right” direction to lower the loss), the optimizer that we use, how many events we train on at once, and how many epochs of training we do. These hyperparameters can be tuned in a grid search by varying values by hand or in a random, automated way.

4.1.1 BOOSTED DECISION TREES

One of the important algorithms used in high energy physics, and in ML in general, is the decision tree. Decision trees are constructed to categorize input events, for example into either “signal-like” or “background-like” distinctions. At its simplest implementation, it is a single tree as depicted in figure 4.3. All input data events are equally weighted and they are labeled with either the classification category or target regression value, depending on the context. The input data also have a number of features that the algorithm will use to try to classify the data or predict a value.

To simplify the description, I will first focus on how a tree is trained for classification. Each of the nodes in the tree enables a decision based on one variable. A scan of potential cut values is undertaken to find the value that best separate the two (or more) classes. The goal of these splits is to increase the purity of event types on the nodes as much as possible. We measure purity with either the Gini index, which is the probability of misclassifying an observation, or entropy, which is a measure of the randomness in the

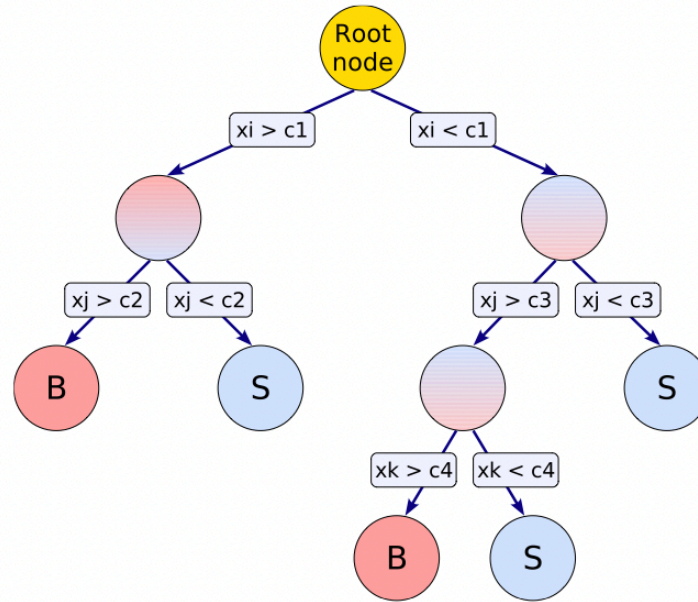


Figure 4.3: Schematic of a boosted decision tree. A determination is made at each of the nodes regarding whether the values are greater than or less than the optimal split point. [33]

data. The lower the Gini Index (or impurity), the better the split. With regards to entropy, the higher the entropy value is, the more randomness in the data. We also want to lower the entropy with each split. The decision tree will keep splitting until it reaches one of the stopping criteria. These criteria include the max depth of the tree, the minimum number of samples per split, and the minimum impurity decrease. These stopping criteria help reduce overfitting in the tree. The final split groupings are known as the leaves of the tree. The tree is then used to classify data by sending new samples through the decisions and seeing which final split they end up in. The average value or mode of the training data that was in the final leaf node is then used as the final classification.

There are two common improvements to the decision tree algorithm that turn it into an ensemble method: *bagging* and *boosting*. Bagging, or bootstrap aggregation, is used to reduce the variance. This is done by creating several subsets of the training data randomly, with replacement. Each subset is then used to train a different tree and the prediction of all trees is averaged for the final result. This can be extended into a random forest, where you retrain trees with both a random subset of the data and a subset of the input features. The number of features selected for each tree to train with is a hyperparameter of the model known as the *maxfeature*. The forest is then a series of parallel estimators. Bagging and random forests

are good when you are missing data, and they handle high dimensional data well, but they are not great for regression tasks.

Boosting trains new trees based on the result of the previous tree. We combine the individual trees in a sequential way to make them more powerful. This means we make a strong learner from a group of weak learners, where each addition helps lower the loss of the model. One way to implement boosting, known as AdaBoost, is to re-weight the training data going into each additional tree by weighting the events that were misclassified higher than those that were correctly classified [114]. This helps the new tree focus on the errors of the former tree.

Another common method is gradient boosting, which combines the methods of boosting and gradient descent [115]. Each tree is trying to minimize the loss from the prior trees combined. To do this we need to optimize with a differentiable loss function. Then we fit each new tree to the residuals of the previous tree, i.e. the residuals become the new target values. For example, if $f_1(x)$ should be y , the residual is then $y - f_1(x)$. Then our second tree is trying to target $f_2(x) = y - f_1(x)$. The learning rate (η) sets how fast we will move in the right direction. The combined equation then becomes results of the first tree + η *results of the second tree. When using the differentiable loss function, the pseudo-residual for the i th training example at tree m is:

$$r_{i,m} = - \frac{\partial L(y, f_{m-1}(x))}{\partial f_{m-1}(x)} \Big|_{x=x_i, y=y_i} \quad (4.7)$$

where x is the input and y is the prediction. This way each tree brings us a little bit closer to the optimal answer based on the training set. To avoid over-training, the learning rate can be reduced so that you learn more slowly and the number of trees can be set to stop the training process.

4.1.1.1 BOOSTED REGRESSION TREES

Boosted Regression Trees (BRT) are a specific type of boosted decision tree, with the goal of determining a continuous variable rather than a classification label. The node splits are therefore designed to reduce the variance in the target values in each leaf or branch. The average value of the target variables in each leaf is used as the predicted value. The ensemble learns sequentially through a form of functional gradient descent. After the first tree, we see how far away we are from the target values and we train the sequen-

tial tree to learn that difference, known as the residual. The final prediction then looks something like, $\text{Prediction} = \text{First Tree Guess} + \eta^*(\text{Second Tree Guess}) + \eta^*(\text{Third Tree Guess}) + \dots$ with the catch that Second Tree Guess is trying to solve for the residual (e.g. $\text{target} - \text{First Tree Guess}$) and Third Tree Guess is solving $\text{target} - \text{First Tree Guess} + \eta^* \text{Second Tree Guess}$. The First Tree is optimized to reduce the usual loss function. This process is illustrated in figure 4.4, in which a series of trees are trained to deliver a final prediction.

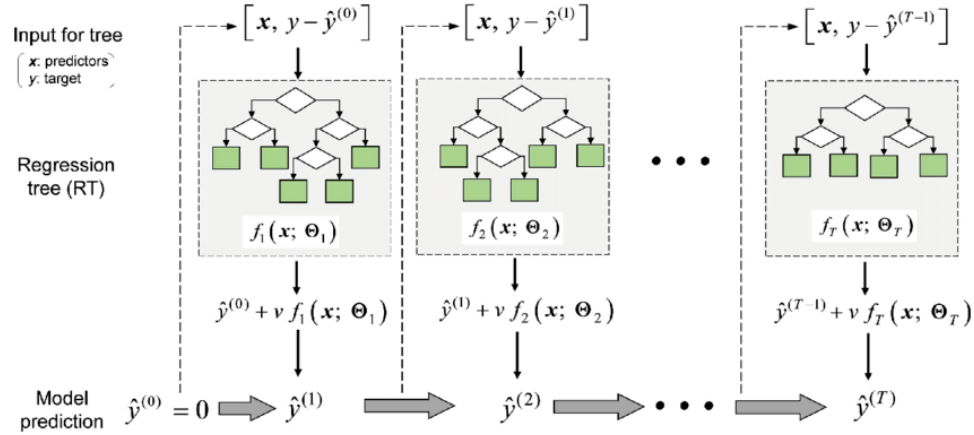


Figure 4.4: Schematic of a boosted regression tree. Each new tree is trained to learn, and thus reduce, the error left from the previous trees. [34]

Another adjustment we can make is to use a stochastic boosted regression tree. This is done by using a random subset of the data to fit each additional tree. This means that models undergoing this training will not be identical, unless they undergo the same sequence of training data selections by setting the same random seed.

A single BDT or BRT is fairly straightforward to interpret because it is essentially a series of selection-cut optimizations. While some of this interpretability is removed when using the ensemble method, the trees can still straightforwardly be constructed and trained, and will perform well with minimal optimization. However, they are prone to over-training, particularly when there is a large number of input variables. Thus feature selection (or feature engineering) is key when training a BDT.

4.1.2 NEURAL NETWORKS

Another common algorithm seen in high energy physics is the neural network. Neural networks are a connection of hidden nodes, known as perceptrons or artificial neurons, that have weights and biases as-

sociated to them. The goal of ML is to learn the optimal values of these weights and biases. There are input layers and output layers with middle, or hidden, layers between them. The input layer contains the features for training and the labels for the target value. The output layer gives the result of the network.

The basis of these networks is the perceptron, shown in figure 4.5 [35]. The perceptron takes in several inputs (x_i). Each input has an associated weight w_i that reflects its importance with respect to the output value. The perceptron also has a threshold that needs to be passed in order for the neuron to fire and send information out. The perceptron returns a value of zero or one according to whether $\sum_j w_j x_j \leq threshold$. The weights and threshold for the neuron can be varied to deliver different results.

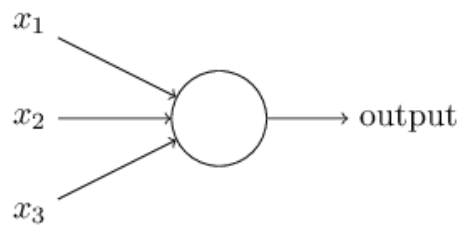


Figure 4.5: Schematic of a perceptron with three input values (x_i) and a single output. [35]

A neural network is constructed from a series of connected perceptrons. Each perceptron can also have a bias associated with it that can help it reach its threshold, in which case the equation becomes $\sum_j w_j x_j + b \geq threshold \rightarrow 1$. When the perceptrons are connected together in many layers, each individual perceptron will have multiple outputs that connect to all the perceptrons in the next layer. Each of these connections will now have a weight associated to it. Figure 4.6 shows a single hidden layer neural network on the left and a deep neural network on the right with multiple hidden layers of connected perceptrons or neurons. The model learns by making small adjustments in the weights (for connections) and biases (of neurons) to minimize the loss function. The initial values for all the weights and biases are usually randomly set.

Deep neural networks are able to find hidden connections, or abstract representations of data. This allows them to approximate the function we are trying to solve. According to the Universal Approximation Theorem, a single hidden layer neural network can approximate any other function. However, this may necessitate an extremely large hidden layer. Non-linearity is an essential feature of a network for that network to successfully approximate functions. To ensure non-linearity in our neural networks, we must

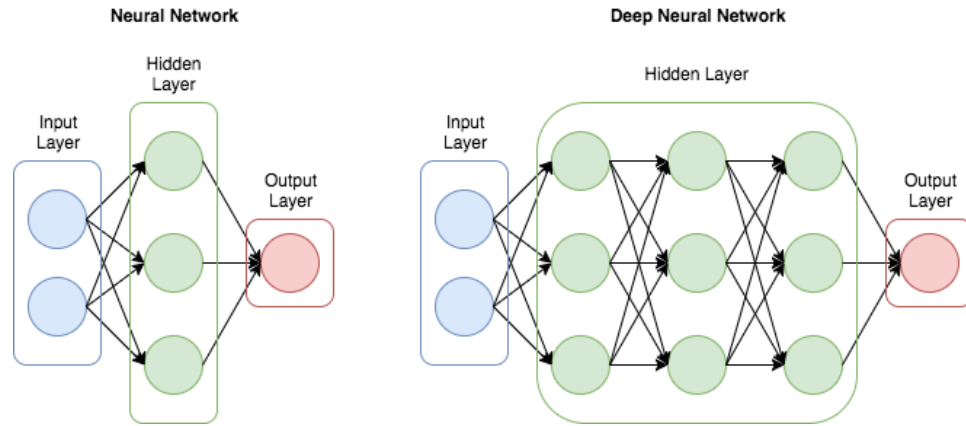


Figure 4.6: Single layer neural network and a deep neural network with multiple hidden layers.

adjust the threshold equation of the neurons shown above to having non-linear activation functions by inserting $(w * x + b)$ rather than returning that value; from now on I will refer to this input as x .

The most common choice of activation functions for the internal nodes of a network is the rectified linear unit (ReLU). In the implementation of ReLU, the maximum of the input x and zero is chosen as the output of the neuron, i.e. $ReLU(x) = \max(x, 0)$. This means that only positive elements are kept. The sigmoid activation is typically used in the output layer for binary classification problems. While it has the complication of not being centered around zero and potentially leading to a vanishing gradient, it does put the outputs on a scale of zero to one. The equation for the sigmoid activation is $sigmoid(x) = \frac{1}{1+e^{-x}}$. The other common activation function re-centers to zero and resembles a linear function nearby zero. This is the tanh function, which is popular in recurrent neural networks. The equation for tanh is $tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. The three activation functions and their shapes can be seen in figure 4.7.

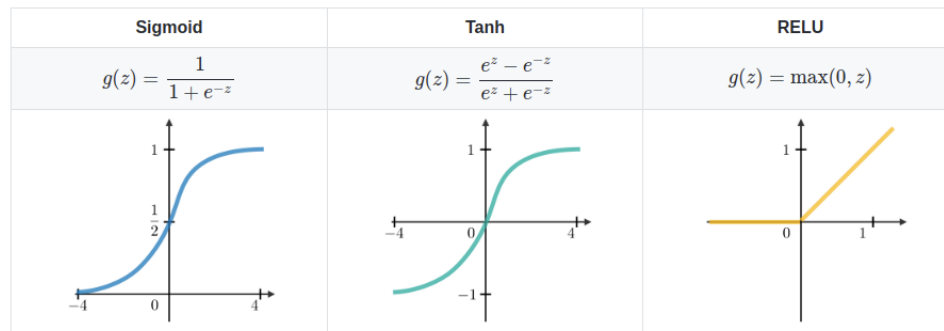


Figure 4.7: Various activation functions used in the neurons of a neural network.

In order for the model to learn to approximate a function, it needs to update the initially randomly

set weights and biases to the optimal choices. Similar to the decision tree algorithms, the model is fed the training data, processes it through its network of trainable parameters (weights and biases), and outputs a decision. This travelling through the network is known as *forward propagation*. The same loss functions are then used to determine how far away the output was from the target value or true classification. Once the loss has been calculated, the process goes backwards to adjust the trainable parameters via a method known as *backpropagation*. The goal of this backward process is to reduce the loss function.

There are two parts to the process of learning: backpropagation for efficient calculation of the gradients and optimization of training the network with the gradients. For backpropagation we take the gradient of the loss function with respect to all the weights and biases. We then update the weights and biases at each layer (l), with the learning rate η included to adjust the speed at which we update, by:

$$\begin{aligned} w^{(l)} &= w^{(l)} - \eta \times \frac{\partial L}{\partial w^{(l)}} \\ b^{(l)} &= b^{(l)} - \eta \times \frac{\partial L}{\partial b^{(l)}} \end{aligned} \tag{4.8}$$

This is the algorithm for gradient descent. The vanishing gradient problem occurs when many of these values go to zero and thus the model cannot learn anything. Smart choices of activation functions can help avoid this problem.

Gradient descent is easy to implement and understand, but it may get trapped at local minima that look like optimal solutions compared to neighbors but are not global solutions. Learning can also take a while as weights are only updated after the whole training set is seen. A solution to the computational problem is to perform mini-batch gradient descent. In mini-batch gradient descent, smaller batches of data are digested at a time, with updates to the weights/biases after each batch is fed through based on the average result of each batch. While this method is computationally faster, there can be variance in the results, and the algorithm can still get stuck in a local minima. This optimization method is known as Stochastic Gradient Descent (SGD).

Adding momentum to the optimizer allows us to accelerate the learning and hopefully avoid local minima. To implement this we add an additional hyperparameter, β , that sets how quickly the previous gradient contributions can decay. SGD with momentum calculates the moving average of the gradient ($m_t = \beta m_{t-1} + (1 - \beta) \nabla(\theta_{t-1})$ where θ s are the trainable parameters) and includes that in

updating the parameters (i.e. $w = w - \eta m_t$). Another method for including momentum is Adam, or Adaptive Momentum Estimation [116]. This includes the second moment v_t in addition to m_t , where $v_t = \beta_2 v_{t-1} + (1 - \beta_2) \nabla(\theta_{t-1})^2$. We bias-correct the m_t and v_t , i.e. $\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$. Then the parameters are updated with $\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$, where ϵ is there to ensure no division by zero. Figure 4.8 shows how Adam reaches the global minima faster than the standard gradient descent algorithm.

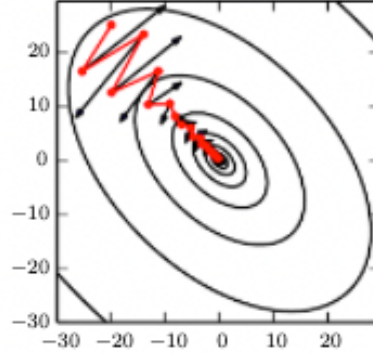


Figure 4.8: Comparison of Adam (red) with gradient descent (black arrows) at arriving of minima of quadratic loss functions. [36]

There are many neural networks that build upon this architecture, including convolutional neural networks, however I will focus on recurrent neural networks as they appear in chapter 5.

4.1.2.1 RECURRENT NEURAL NETWORKS

Recurrent neural networks (RNN) are designed to process sequential data. They have special gates that allow them to keep information about proceeding inputs to influence current decisions. RNNs can handle variable length data and data for which the order matters. This is useful for time series predictions, track finding in a detector, and speech translation. In figure 4.9 a diagram of a basic RNN can be found on the left. The right hand side shows the RNN unfolded. The RNN performs the same actions for every entry in the sequence of data. The RNN maps the input x to an output o . The middle node h is the hidden state which calculates $h(t) = f(U * x^{(t)} + W * h^{(t-1)} + b)$, where f is the chosen activation function and U is the weight matrix for the input to hidden connections, W is the weight matrix for hidden-to-hidden connections, and b is the bias. The output is then $o(t) = c + Vh^{(t)}$, where V is the weight matrix between the hidden state and the output and c is the bias term. To get a final prediction/result, the $o(t)$ is

put through a final activation function, such as softmax [36].

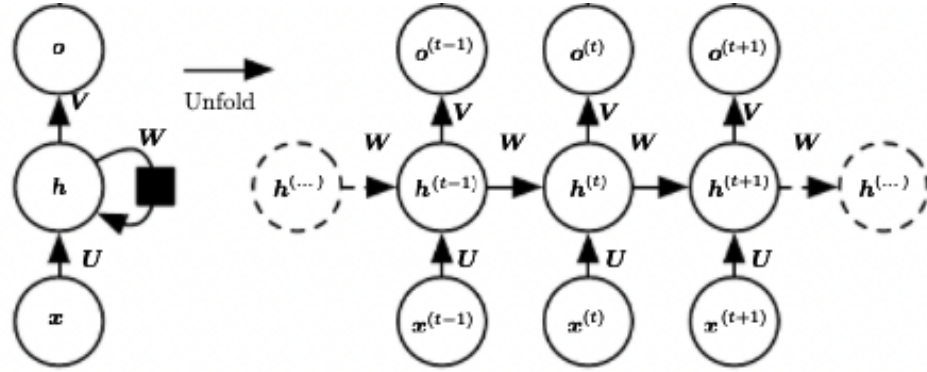


Figure 4.9: Simple RNN on the left and the RNN unfolded on the right with the connections between input-hidden, hidden-hidden, and hidden-output shown. [36]

One problem with RNNs is their short term memory; they forget data they see early on in a sequence. When doing backpropagation, RNNs also suffer from the vanishing gradient problem so those earlier layers may stop learning. LSTM, or Long Short Term Memory, cells help to solve this problem as they allow the cell to keep or forget information. The cell state is where the data passes through and can be considered the memory of the network. The information on the cell state of the LSTM can be added to, or removed by, gates. Sigmoids are used to decide whether to keep, consistent with an output of one, or forget, consistent with an output of zero, previously seen data. The forget gate takes information from the current input and previous hidden state and chooses whether to forget or keep that information. The input gate updates the cell state by deciding what is important to add. Finally, the output gate chooses the next hidden state [37]. The LSTM cell is shown in figure 4.10

4.2 ML IN ATLAS

There are a number of places where ML is being used in the ATLAS experiment at large. One common place is in physics analyses to discriminate between the signal and the background. For example, a Run 2 dijet resonance search used weakly supervised ML to do anomaly detection to try to find new physics over a Standard Model background [117]. Another search for a Higgs boson decaying into a Z boson and light resonance in two lepton plus jet events used multivariate discriminants for event selection. Multi-layer perceptrons (MLPs) were trained to estimate the mass of the light resonance and to classify signal. The

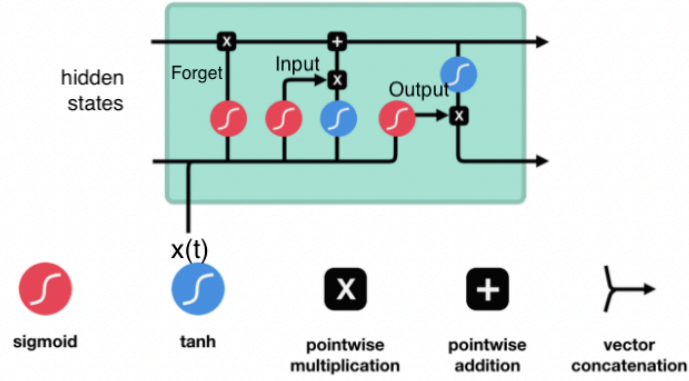


Figure 4.10: LSTM cell with the incoming new information $x(t)$, previously hidden information, and the forget, input, and output gates. Adapted from [37]

classification MLP was trained on kinematic information from the tracker and used a sigmoidal activation function [118].

In my first year of graduate school, I worked with Professor Keith Baker’s group to look at the impact of using ML for signal discrimination. We were working on a search for Higgs boson decays to ZZ^* to four-lepton states, specifically, the $H \rightarrow ZZ_d \rightarrow llll$ channel. For this search, the Higgs boson was used as a portal to the dark sector [119]. The process could occur when the Higgs boson is produced by gluon-gluon fusion and then couples to a Z_d , a vector particle with spin 1 and negative parity. The Z_d will mix with the Z^* , off-shell Z , and decay to leptons. This targets a mediator of the theorized $U(1)_d$ gauge symmetry, a possible extension of the SM. I looked at using boosted decision trees to help separate out signal from background instead of the standard selection cut based methods. For an initial study, I trained a BDT using TMVA on the final state radiation masses of the two Z bosons, the four leptons, the rapidity and ϕ of the leptons, and the number of jets. In the final iteration of the analysis, BDTs were used on simulation to help distinguish particular Higgs boson production processes from the background using variables like lepton p_T , $\Delta_R(j, 4l)$, and η_j and they improved cross section measurements by 15-35% [119].

The biggest contributor to the computation budget of ATLAS is the simulation of Monte Carlo datasets. ML efforts to both better simulate data and speed up the simulation are ongoing. One ML proposal, known as CaloGAN, was started by Yale physicists [120]. They used generative adversarial networks to create fast simulations of the electromagnetic calorimeters.

Machine learning can also be used for tagging particles in our detector. Work done by Yale postdoc Chase Shimmin showed the possibility of using adversarial neural networks to tag jets based on their sub-structure [121]. ML is also being used in the trigger systems. There is an ongoing push to try to convert deep neural networks into a format that can be run on FPGAs. This also requires a reduction in model size, latency, and energy use while maintaining performance. Scientists have developed a workflow known as QKeras+hls4ml to translate deep learning models to be run on FPGAs [122].

Physicists are also reaching out to the ML community at-large to help solve computational problems. The Tracking ML challenge [123] was a Kaggle challenge that made LHC data public with the goal of constructing particle tracks. The results of the accuracy phase were then presented at NeurIPS 2018.

Machine learning is being incorporated into all aspects of the particle physics data processing pipeline: from simulation, to triggering and tracking, to final data analysis.

*"It was good, and nothing good is truly lost. It stays part of
a person, becomes part of their character."*

Rosamunde Pilcher, *The Shell Seekers*

5

Machine Learning for Tau Leptons

5.1 INTRODUCTION TO TAUS

THE ANALYSIS OF $VH, H \rightarrow \tau\tau$ THAT I WILL PRESENT IN THE FOLLOWING CHAPTERS RELIES ON OUR ACCURATE RECONSTRUCTION AND IDENTIFICATION OF TAU LEPTONS. Taus are a tricky object to fully reconstruct as they decay almost immediately, have neutrinos in their final state, and are massive enough to decay to quarks. Machine learning algorithms have given critical boosts in performance to our standard reconstruction algorithms, allowing us to make more precise measurements and improve our sensitivities in analyses involving taus.

5.2 TAU PROPERTIES

Taus are third generation fermions. They have a charge of $-e$ and have a mass of 1777MeV . That means the tau is ~ 17 times more massive than a muon and ~ 3500 times more massive than an electron. While the tau's higher mass makes it desirable for the study of the Higgs boson coupling to fermions at the AT-

LAS detector, the tau is an unstable particle that will decay almost immediately. It has the short lifetime of only 2.9×10^{-13} s. This means that its decay length, calculated as $d = \text{velocity} * \text{time} = \beta c \gamma \tau_\tau$, is approximately $87 \mu m$, where τ_τ is its lifetime and we have taken into account relativistic effects. Consequently, almost all taus will decay inside the LHC beamline without ever reaching the ATLAS detector. We therefore have to identify taus based on their decay products.

There are two main modes of decay for taus: leptonic and hadronic decays. Example Feynman diagrams for both decay processes are shown in figure 5.1. Leptonic tau decays ($\tau \rightarrow \nu_\tau l \bar{\nu}_l$ where l refers to an electron or muon) occur about 35% of the time. In this decay process, the tau decays into a tau neutrino, to preserve the tau lepton flavor contribution, and into a lepton and anti-neutrino of the same flavor. We identify leptonic decays by the lepton and missing energy in the detector. This relies on the electron, muon, and MET reconstruction described in chapter 3.

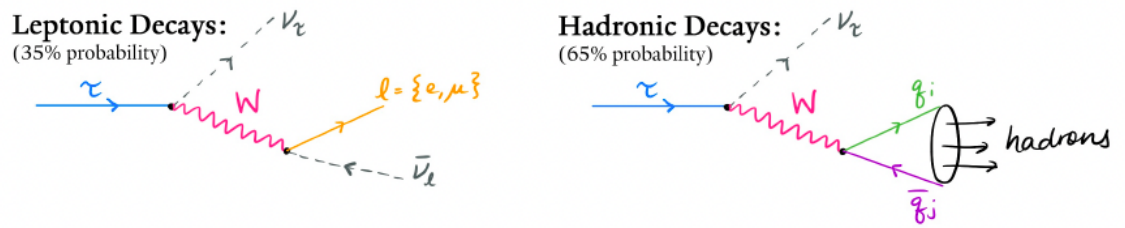


Figure 5.1: Example Feynman diagrams for the two main decay modes of the tau lepton. Both decay modes include a tau neutrino to preserve the tau flavor in the decay. The leptonic decay (on the left) includes a W decaying to a lepton and an anti-neutrino of the same flavor. The hadronic decay (on the right) includes a W coupling to quarks which hadronize into a variety of hadrons.

About 65% of tau decays are hadronic tau decays. This means the tau decays into a collection of particles usually comprised of charged and neutral hadrons. The hadronic decay of the tau also includes a tau neutrino in order to preserve tau lepton flavor. In addition, the hadronic decay of the tau usually includes one or three charged hadrons, though this number can be five, seven, etc, and a number of neutral hadrons. We therefore classify hadronic decay modes into two main types: 1-prong and 3-prong based on the number of charged hadrons. 72% of hadronic decays are 1-prong and about 22% of hadronic decays are 3-prong. About 68% of hadronic decays include at least one neutral pion. The most common type of hadron produced in tau decays is a pion because pions are mesons that combine quarks from the same first generation (i.e. up and down). Kaons, another possible decay product of taus, are mesons that are made

Decay	Prongness	Final State	Branching Ratio (%)
Leptonic	1-prong	$\nu_\tau e \nu_e$	17.8
		$\nu_\tau \mu \nu_\mu$	17.4
Hadronic	1-prong	$\pi^- \pi^0 \nu_\tau$	25.5
		$\pi^0 \nu_\tau$	10.8
		$\pi^- \pi^0 \pi^0 \nu_\tau$	9.3
		$\pi^- \pi^0 \pi^0 \pi^0 \nu_\tau$	1.0
		$\pi^- \pi^- \pi^+ \nu_\tau$	9.0
	3-prong	$\pi^- \pi^- \pi^+ \pi^0 \nu_\tau$	2.7

Table 5.1: Branching ratios of some of the most common tau lepton decays. [4]

up of quark/anti-quark pairs from different generations, such as strange and up or down.

The main decay modes of the tau and their branching ratios are included in table 5.1 [4].

5.3 TAU RECONSTRUCTION

When discussing tau reconstruction in this thesis, I will only be referring to the hadronic decays of taus. Furthermore, we only look at the decay products that are visible to our detector, which excludes the neutrinos. Tau decays form a type of jet and thus their reconstruction begins with the standard jet reconstruction algorithms described in section 3.3.3. Taus are seeded by jets that are reconstructed with the anti- k_T algorithm with $R=0.4$. The TopoClusters for the tau are additionally calibrated using a local hadronic calibration. The jets must also have $p_T > 10\text{GeV}$, $|\eta| < 2.5$, and not be in the crack region of the detector in order to be considered as a seed for tau reconstruction [40].

A track is then associated to the tau-jet if it has a $\Delta R < 0.2$ from the tau-jet center and it passes a series of requirements. The tracks must have $p_T > 1\text{GeV}$ and the distance from the track to the tau vertex in the transverse plane ($|d_0|$) is $< 1\text{mm}$. The track also needs to have a $|\Delta z_0 \sin(\theta)| < 1.5\text{mm}$, where z_0 is closest approach in the longitudinal axis [40].

Tracks are then split into two main designations: *core* and *isolation*. Core tracks are within a ΔR of 0.2 of the center of the tau candidate and isolation tracks are in the range of $0.2 < \Delta R < 0.4$. The tau vertex is set by selecting the vertex of the tracks associated with the tau-jet that has the largest fraction of the momentum of the tau. The direction (η, ϕ) of the tau is set by the vectorial sum of TopoClusters within ΔR of 0.2 of the jet barycenter. Finally, the tau mass is set to zero making the transverse energy E_T equal to the momentum p_T . This is a safe approximation to make for our analysis because we only look at taus

with a visible energy greater than 20GeV, which is already much higher than the 1.7GeV tau mass.

5.3.1 RECONSTRUCTION IN THE TAU TRIGGER

The tau trigger performs a similar reconstruction to the offline reconstruction but due to its time constraint of $2.5\mu s$ in the hardware trigger (L1) and $0.35s$ in the software high-level trigger (HLT), the algorithms are simplified and the accuracy reduced. The main goal of the trigger is to separate hadronic taus from QCD jets: quark- or gluon- initiated jets from non-tau QCD processes. The key properties we can use to discriminate taus from jets are their low charged-track multiplicity, the isolation of their calorimeter cluster, and the narrowness of their jet [124]. These characteristics are highlighted in figure 5.2.



Figure 5.2: On the left is the typical signature of a QCD jet with many non-isolated tracks and on the right is the typical signature of a hadronic tau decay with its highly collimated one to three charged tracks that are contained in the core region. [38]

The L1 trigger looks at calorimeter towers with a granularity $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ to identify regions of interest based on the sum of energy in 2×2 groupings of these towers. Then in the HLT phase, we look at LC-calibrated TopoClusters in a cone of $\Delta R = 0.2$ around the L1 seed. There is a very basic Tau Energy Scale correction applied in this trigger stage, which is inclusive for both 1-prong and 3-prong taus and has a less precise pileup correction than in the full process described in section 5.5 as it uses the average number of interactions per crossing instead of a more accurate calculation.

The calorimeter cluster selection is followed up by the Fast Track Finder (FTF) algorithm that first finds a leading track within a ΔR of 0.1 of the cluster center with $|z| < 225$ mm. The FTF is then broadened to the isolation region where it will find all tracks with a $\Delta R < 0.4$ from the core and at $|z| < 10$ mm with respect to the leading track. Finally, a cut is applied based on track multiplicities to distinguish from jets, with a selection of one or three tracks in the core region and one or no tracks in the isolation region.

5.4 TAU IDENTIFICATION

The main purpose of the tau identification algorithms is to reject the background from quark- and gluon-initiated jets. At the start of Run 2 we were using a Boosted Decision Tree (BDT) algorithm that was then upgraded to a Recurrent Neural Network (RNN) algorithm. The identification algorithms set working points (Loose, Medium, Tight) for the hadronic taus based on their target efficiencies. In addition to the tau-specific identification algorithms, there is also a specific electron rejection (eVeto) where any 1-prong tau candidates within a ΔR of 0.4 of a “Very Loose” electron are rejected.

5.4.1 BDT FOR TAU IDENTIFICATION

For the beginning of Run 2 there was a BDT algorithm for tau identification. It was trained and applied separately for 1- and 3-prong taus. The training sample for the signal was composed of $Z \rightarrow \tau\tau$ events and the background sample consisted of simulated dijet events. The training events had a p_T between 10 and 360 GeV, and at least one reconstructed $\tau_{had-vis}$ candidate passing trigger and full online selection. Input variables included the energy fraction in the central region of the tau, the leading track momentum, and the track radius [125]. Some variables were only considered for 1- or 3-prong taus, such as the track mass and maximum ΔR . These input variables were corrected so that their distribution mean was constant as a function of pileup to remove efficiency dependence on pileup. The Tight working point had 0.6(0.45) target efficiency for 1-prong(3-prong) taus, the Medium working point had 0.75(0.6), and Loose had 0.85(0.75) [40]. The rejection power for the BDT is shown in the dashed lines in figure 5.3.

5.4.2 RNN FOR TAU IDENTIFICATION

Later on in Run 2, ATLAS moved to using a RNN algorithm for $\tau_{had-vis}$ identification that can be applied offline to the full Run 2 dataset. We use the RNN working points for taus in the final analysis for the $VH, H \rightarrow \tau\tau$ study. The input variables to the RNN are mix of low-level inputs for tracks and clusters and high-level observables similar to the ones used in the BDT tau identification algorithm. These low-level input variables include cluster moments, track momentum, track impact parameters, angular distances, number of track hits, and the transverse energy of the clusters. The high-level variables include the central energy fraction, transverse flight path significance, maximum track ΔR , and the ratio of EM

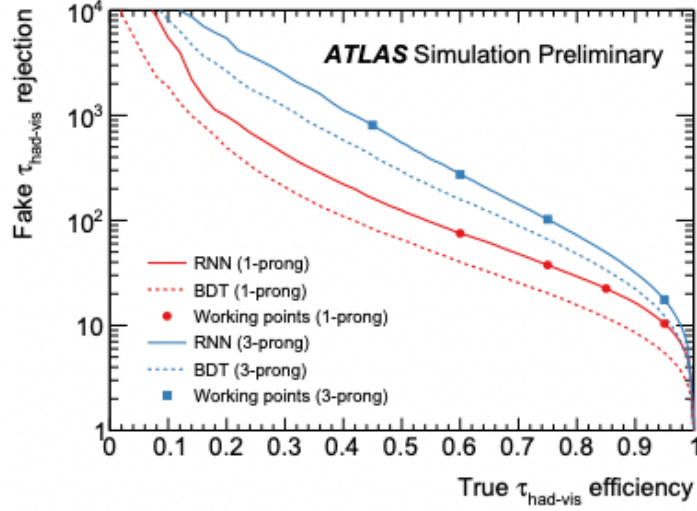


Figure 5.3: Rejection power for background quark- and gluon-initiated jets as a function of true $\tau_{had-vis}$ efficiency. The BDT performance is shown with the dashed lines in red for 1-prong and in blue for 3-prong, while the RNN is shown with solid lines. The solid dots represent the four working points: Tight, Medium, Loose, and Very Loose. [39]

energy and track momentum [39].

As one can think of objects in the ATLAS detector as an ordered sequence in terms of their p_T , RNNs can be applied successfully to our data. The training inputs for the RNN come from a $Z/\gamma^* \rightarrow \tau\tau$ sample that is split up and trained separately on 1-prong and 3-prong taus. There is a p_T re-weighting to remove the p_T -dependence bias because the dijet background sample has a softer distribution than the signal. The sequential inputs to the RNN are the first ten, or fewer, leading tracks in p_T and six leading clusters in E_T . The network architecture is detailed in figure 5.4. The three dense layers for the high-level inputs have 128, 128, and 16 nodes, respectively. The shared dense layers have 32 nodes. The ReLU function is applied after all fully connected layers except the last, which uses a hyperbolic tangent, and the LSTM layers use the sigmoid activation function.

As can be seen in figure 5.3, the RNN is able to achieve the same working point efficiencies with almost double the background rejection power of the BDT.

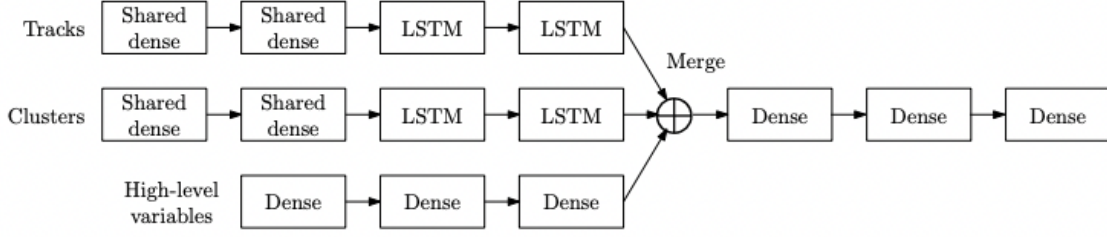


Figure 5.4: Diagram of the RNN architecture for tau identification with the three separate branches for low-level track and cluster variables and high-level variables. [39]

5.5 TAU ENERGY SCALE

After a tau candidate has been identified in the ATLAS detector, its measured energy needs to be calibrated in order to have an accurate estimate of the true value. This is due to both detector performance and tau decay kinematics. Detector issues include the presence of uninstrumented material, the incomplete coverage of the detector, and the non-uniformity of the detector. After the first LC (Local Calibration) scale has been applied to all jet-like objects, ATLAS applies an additional energy scale to correct the measured tau lepton energy. The tau decay makes it particularly tricky to calibrate the energy of the tau. Taus decay after such a short period of time that the neutrino has already been produced before the decay products reach the ATLAS detector.

A typical method of improving energy calibration is to replace the energy measured in the calorimeter that is associated with deposits from charged particles with information from the more precise tracking detectors. This is complicated in the case of tau decays due to the possible presence of neutral pions, which can deposit energy that overlaps with the charged-particle deposits in the calorimeter. These neutral particles do not have associated tracks to aid in the final energy calculation. Also, when identifying a tau in the detector, the tau candidate is restricted to a small cone around its vertex. Thus the reconstructed tau energy may be an underestimation due to energy associated with the tau that was deposited outside of the cone and left out of the calculation. There is also interference in the calorimeter due to both the underlying event and pileup in the detector that can affect the energy measured. Figure 5.5 illustrates some of the complications we need to consider when calculating the final energy of a tau.

In order to have an accurately reconstructed tau energy, we need to calibrate the tau energy obtained from the calorimeters. There are two calibrations used at ATLAS; one is the baseline calibration and the

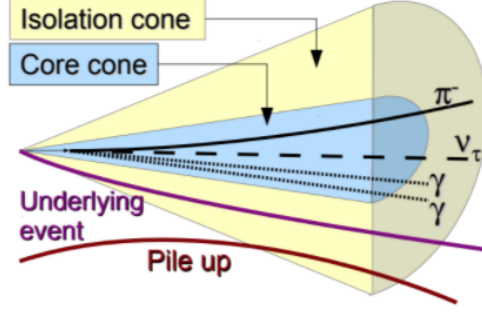


Figure 5.5: Illustration of a hadronic tau decay highlighting some of the effects, such as underlying event and pileup, that make energy calibration tricky for taus.

other is the Boosted Regression Tree (BRT) calibration. With the baseline method, one can calculate the newly calibrated energy from the measured energy:

$$E_{calib} = \frac{E_{LC} - E_{pileup}}{\mathcal{R}(E_{LC} - E_{pileup}, |\eta|, n_p)} \quad (5.1)$$

where E_{LC} is the LC-calibrated sum of energy in the core region, which then has the energy contribution from the multiple interactions per bunch crossing (E_{pileup}) subtracted [40]. E_{pileup} is a function of the number of primary vertices (N_{PV}), $|\eta|$, and the number of prongs n_p . E_{pileup} is calculated as:

$$E_{pileup}(N_{PV}, |\eta|, n_p) = A(|\eta|, n_p) \times (N_{PV} - \langle N_{PV} \rangle) \quad (5.2)$$

where A is a linear coefficient. $\mathcal{R}$ is the detector response calibration, which is the Gaussian mean of the $(E_{LC} - E_{pileup})/E_{true}^{vis}$ distribution, that is used to scale the tau energy. E_{true}^{vis} accounts for the energy from the generated tau decay products and final state radiation, but not the energy from the (invisible) neutrinos. The LC-calibrated tau momenta is then scaled to the true visible momentum in bins of number of prongs (1- and 3-prong taus) and η .

Currently the baseline calibration works well for high p_T taus but performs poorly at low p_T when compared to truth MC values of the energy. The BRT method combines the baseline calculation and information from the Tau Particle Flow (TPF) which calculates the charged pion momentum in the tracking system of the ATLAS detector. Tau candidates that pass the medium BDT identification point from $Z \rightarrow \tau\tau$ samples are used for training the BRT. There are two figures of merit for evaluating the cali-

brated momentum: resolution (half of the 68% interval of the ratio of calibrated momentum to true momentum) and non-closure (offset from unity). Variables used in the initial training for the ATLAS BRT method include pileup, cluster variables like shower depth and cluster moments for both the charged and neutral decay products, number of tracks for the charged pions, energy of the pions, and ratio of their momenta [40]. The Tau Energy Scale (TES) finds a shift factor α , so that the tau energy is recalculated as $E_T \rightarrow (1 + \alpha)E_T$. The BRT makes a demonstrated improvement in energy resolution of the tau lepton over the baseline calibration for a range of tau p_T momenta as shown in figure 5.6.

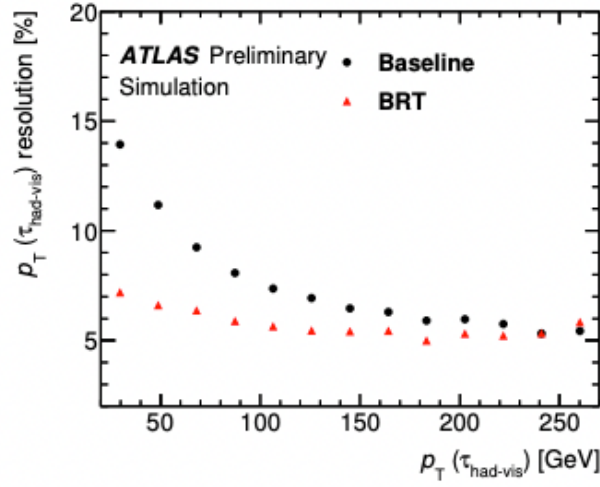


Figure 5.6: Energy resolution of the baseline and BRT calibrations for taus as a function of tau momentum. The BRT calibration improves upon the Baseline calibration for taus with p_T below 250GeV. [40]

5.6 MACHINE LEARNING IMPROVEMENTS TO THE TAU ENERGY SCALE

The energy scale determination (MVATES) uses information from the ATLAS calorimeter and tracker and applies a scale factor derived from a Boosted Regression Tree (BRT). There is a particular challenge at the trigger level, where there are ongoing studies to attempt to improve MVATES for taus by including track information into the training of the BRT. If a more accurate energy value for taus can then be used in the trigger decision, we will be better able to select the taus we want for physics analysis within the allocated bandwidth.

Initially, in order improve the MVATES resolution for the trigger, in particular for the low momentum

taus, I studied tracks from the planned FTK (Fast TracKer). At the time, FTK was a potential addition to the ATLAS tracking system for Run 3. The track variables in the initial study came from simulated trigger output incorporating FTK tracks. With these FTK tracks added to the list of BRT input variables, I performed studies such as determining which track variables helped improve performance and whether it was best to make the training prong-dependent by training on 1- and 3-prong taus separately and together. I quickly determined that the trigger track information had too high of a p_T cutoff to allow for good resolution for low p_T tracks. With the eventual cancellation of the FTK project, the trigger tracks were then switched to be standard simulated high-level trigger (HLT) tracks, which could give important guidance for the expected performance in the HL-LHC era when ATLAS will have a hardware-based trigger. I carried out additional studies with a focus on determining the importance of various tracking-related variables and the tuning of hyperparameters.

5.6.1 BOOSTED REGRESSION TREE ARCHITECTURE FOR THE TAU ENERGY SCALE

In order to determine scale factors to use in the calibration of the tau energy, I trained a Boosted Regression Tree (BRT) using TMVA (Toolkit for MultiVariate Data Analysis with ROOT)[33]. In this study, the regression tree will contain a series of right/left decisions to find the best weight to apply to obtain a target value of 1 for the ratio between the true tau momentum and the reconstructed tau momentum. As a reminder, boosting means that instead of training only one tree, many trees are trained and combined into a single regressor using a weighted average.

The hyperparameters to be specified in MVATES, along with their definitions are in table 5.2. These are important to tune to get the best performance. For example, reducing the depth of the trees helps limit overtraining, or heavily biasing the output weights based on the input events. Incorporating boosting improves the stability of the training result, which can fluctuate when training with decision trees. Introducing bagging, where a random subset of the data is fed into each tree, helps stabilize the response by essentially smearing the results. MVATES uses a gradient boost, which tries to minimize the difference between the model response and the true value by adjusting the node splits and parameters. The difference is usually quantified using a loss function. When performing a gradient boost, a Huber loss function is

used:

$$L(F, y) = \begin{cases} \frac{1}{2}(y - F(x))^2 & |y - F| \leq \delta \\ \delta(|y - F| - \delta/2) & |y - F| > \delta \end{cases} \quad (5.3)$$

which is adjusted using a shrinkage parameter, or learning rate, set to the standard value of 0.1 [33].

Hyperparameter	Definition	Initial Setting
NTrees	Number of trees trained to get final answer	2000
NCuts	Precision of cut values tested at each node split	200
BoostType	Boosting algorithm	Grad
Shrinkage	Learning rate for the boost	0.1
UseBaggedBoost	Whether to include bagging when training	True
Bagged Sample Fraction	Size of bagged events relative to total sample	0.5
MaxDepth	Number of layers in a tree	5

Table 5.2: TMVA hyperparameters and their definitions that are used in training the BRT to perform MVATES. The initial value for each parameter is included in the final column.

When I trained the BRT, I split the input dataset in half into a training and validation set, based on whether the arbitrary event number set during simulation was even or odd. The training set is used to make the decision trees and the validation set is used to calculate the test error after applying the trained regression. The performance on the validation set is then used to update the trees until a set performance threshold is met.

The initial training of the BRT used the input parameter settings shown in table 5.2. These are the parameters used in all trainings until the hyperparameter tuning is checked in section 5.6.4.1. The benchmark for determining performance of my model is the performance of the weights trained previously (in 2018) without trigger-level track information.

5.6.2 INPUT VARIABLE DEFINITIONS

For training MVATES, I input tau decay events with reconstruction- and truth-level quantities. Both HLT Calorimeter variables and track trigger variables are included. A BRT cannot be trained on an input that has variable length. This results in a challenge for training on tau candidates because taus can decay into 1- or 3-track final states. Thus the tau-track-related vectors, such as momentum, can therefore be variable in length. In order to use track information, one must convert these tracking-related numbers into single-valued inputs, such as weighted values, or include null inputs to keep the variable length the same.

The input values for training the BRT are listed in table 5.3. This includes some variables that were used in the initial training, but did not end up helping performance. They were not used in the final training because it is preferable to have fewer input features. The variables used in the final optimal training have a check-mark next to them in the right-most column of table 5.3. All of the HLT calorimeter values are single-valued per each event so they can be used directly in the BRT. This is not true for most of the trigger-track variables. Most taus should have one or three tracks, and sometimes five for rarer decays, but due to detector reconstruction effects there are sometimes zero, two, four, and six tracks associated to the tau. In order to have the same size inputs to the BRT for all events, default values of -1 for specific track momenta variables are used in order to create a single-valued entry for every variable for each event. This value of -1 does provide information to the BRT by indicating that a limited number of tracks existed for a given event, which in turn leads to a different predicted scale factor.

The weighted track p_T are set to default values of -1 if there are fewer than three tracks in the event. Furthermore, values for leading track weighted p_T peak roughly at one if there is only one track in the event, but this is not exact since the summed total track momentum does not always equal the detector axis momentum. Furthermore, the HLT-calibrated and tau-trigger-track momentum are not the same, which leads to the weighted ratio of the leading track p_T to sometimes be greater than 1. Track 1+2+3 weighted p_T average is not included in the the FTK studies documented in section 5.6.3.

Variable Name	Definition	Used
<i>HLT Calorimeter Variables</i>		
μ	Pileup	✓
$\langle \rho \rangle$	Cluster first moment in energy density (E/V) where E is energy of cluster and V is volume	✓
λ_{center}	Shower depth at cluster centroid	✓
$\langle \lambda^2 \rangle$	Second moment in λ , which is the distance of a calorimeter cell from shower center along shower axis	✓
$f_{presampler}$	Fraction of cluster energy deposited in the presampler	✓
Lead Cluster Fraction	Fraction of energy contained in highest energy cluster	✓
Upsilon Cluster	Function of tau polarization	✓
P_{EM}	Classification probability of cluster to be EM-like	✓
Detector p_T	LC calibrated four momentum of the tau	✓
Detector η	LC calibrated η of the tau	✓
<i>Track Trigger Variables</i>		
n_{Tracks}	Number of tracks associated to the tau	✓
Total Track p_T	Sum of track momenta from all tracks in the tau	✓
Average Track p_T	Total Track p_T divided by n_{Tracks}	✓
$Track_i p_T$ for $i=1-5$	p_T for each of the first five tracks	
$Track_i p_T / \text{Detector } p_T$	Leading three tracks' p_T weighted by detector p_T	✓
$Track_i p_T / \text{Total Track } p_T$	Leading three tracks' p_T weighted by Total Track p_T	
$\sum_{i=1}^{i=3} p_{T,i} / \text{Detector } p_T$	Sum of three leading track p_T weighted by detector p_T	
$\sum_{i=1}^{i=3} p_{T,i} / (\text{Total Track } p_T)$	Sum of three leading track p_T weighted by Total Track p_T	✓

Table 5.3: Variables used in the training of the BRT to improve the Tau Energy Scale broken up into HLT Calorimeter Variables and the new Track Trigger Variables. The final column indicates whether they are used in the final, optimized training of the BRT.

Sample input distributions of the track variables used in the training of the BRT with the simulated tau trigger tracks are included in Figure 5.7.

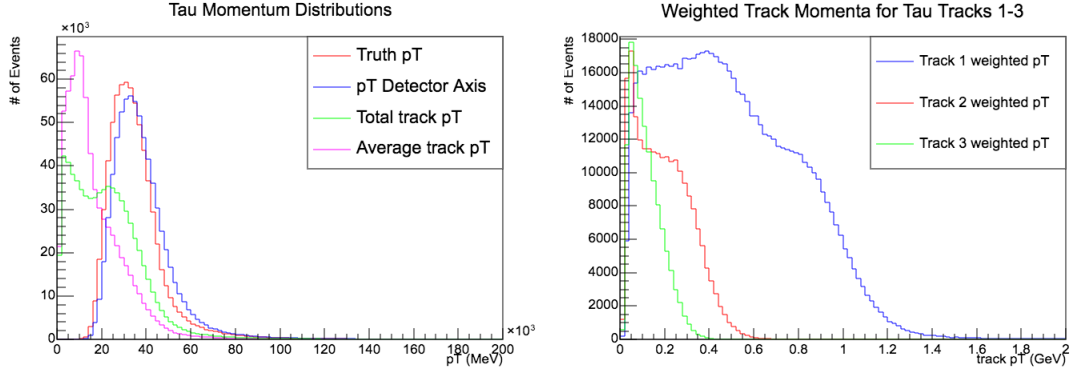


Figure 5.7: The left plot shows tau momentum distributions highlighting the difference between Truth p_T , the HLT Calibrated p_T (detector p_T), the total track p_T , and the average track p_T . The target value of regression for the BRT is the ratio of Truth p_T /Detector p_T . The right plot shows the weighted transverse momentum distributions for the leading three tau tracks, weighted by detector p_T . They are set to -1 if there are fewer than three tracks. As seen in the left plot, detector p_T and sum of tau track p_T are not the same, which leads to the weighted ratio of the leading track p_T to sometimes be greater than 1.

5.6.3 FTK TRACK STUDIES

For the FTK studies, the input track trigger values were from a simulated FTK trigger with a 12GeV p_T threshold. The goal of using these tracks was to see if adding track information to MVATES at the trigger level could improve upon the energy calibration for potential use in the trigger decision to keep or reject events with taus. The two main tasks looked at to improve the performance of MVATES were (1) adding in track variables to the BRT training and (2) training on 1- and 3-prong taus either independently or simultaneously. I trained on an input $Z \rightarrow \tau\tau$ sample. The figure of merit used was the offset of the ratio of truth and detector p_T from one.

Including Track Trigger Variables

To begin the studies, weights were trained using the various input variables listed in table 5.3. First, only the HLT Calorimeter variables were used to train the weights. Then three of the FTK track variables were added: total track p_T , average track p_T , and number of tracks. Then six of the FTK variables were used: the initial three and the three weighted by individual track p_T . In order to see the performance of

the weights, the MVATES was applied to a $Z \rightarrow \tau\tau$ sample. Figure 5.8 shows that the addition of the track variables improved performance for low momentum taus. Poorer performance for high momentum taus with the addition of track variables could have been due to overtraining with the higher number of input variables.

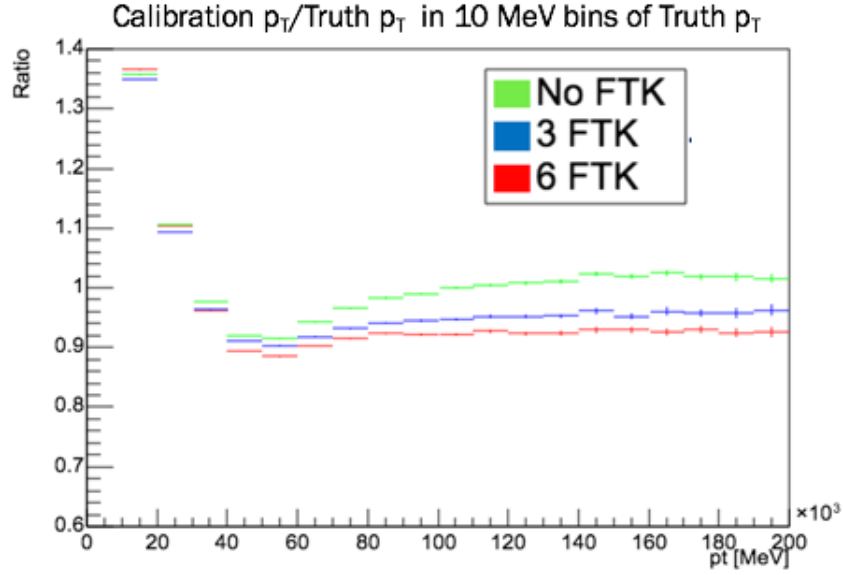


Figure 5.8: Comparison of the offset of the calibrated tau momentum from the truth momentum of three different trainings of the BRT for the MVATES weight: no FTK (only the HLT Calorimeter variables), three FTK variables (adding in three of the track variables: total track p_T , average track p_T , and nTracks), and using six FTK variables (adding in the weighted track p_T for the leading three taus).

1- and 3-Prong Split for Training the BRT

Then the performance of splitting the training of MVATES on 1- and 3-prong taus separately versus together was evaluated. When training independently, an additional requirement was added to the TMVA training to only train on either taus with less than two tracks or taus with greater than two tracks (since the taus are not perfectly reconstructed as 1- and 3-prong taus). The effect of splitting up the training was also tested using different combinations of input track variables. The HLT Calorimeter only variables (no FTK) cannot be split by track number since that information is not available. When the taus are split up by 1- and 3-prong (or single track and multi-track) the variable nTracks is not included in the training since that information was already used to create the track multiplicity split. As shown in figure 5.9, the HLT calorimeter variable only training (no FTK) performs best for high momentum taus as the offset values

are closest to the goal value of one. However, the use of all track variables (5 FTK) with the split on 1- and 3-prong taus performs best for low momentum taus.

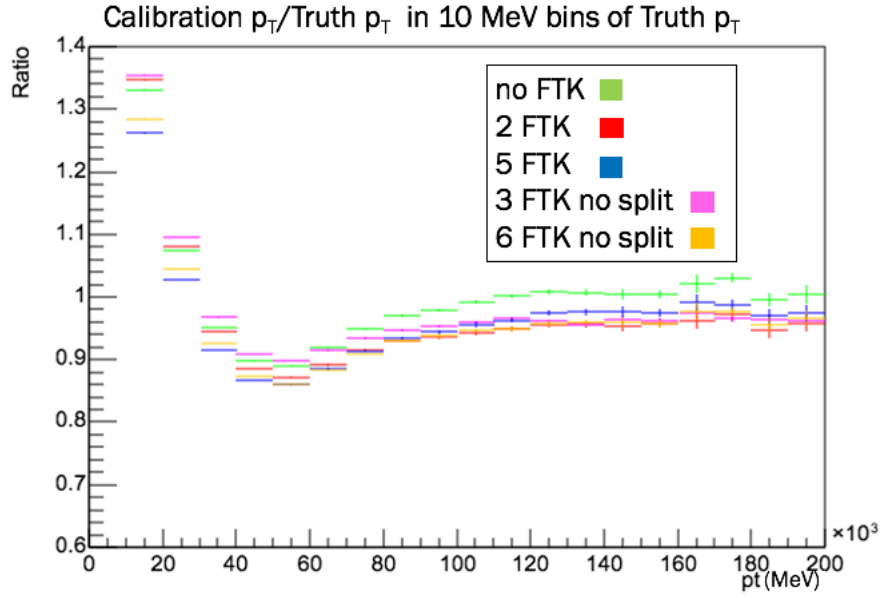


Figure 5.9: Comparison of the offset of the calibrated tau p_T from the truth p_T of five different trainings of the BRT for MVATES: no FTK (only the HLT calorimeter variables), three FTK no split (adding in three of the track variables: total track p_T , average track p_T , and nTracks), and using six FTK no split variables (including the weighted track p_T), with training on 1- and 3-prong taus separately with two FTK variables added (adding in the total and average track p_T), and using five FTK variables (adding in the weighted track p_T for the leading three taus).

Results for Energy Calibration using FTK Tau Tracks

Including FTK track information in the training of the BRT improved energy calibration performance for low momentum taus, but not high momentum taus. Furthermore, the improvement for low momentum taus was incremental and the offset of the ratio of calibrated/truth momentum was still very far from one. This was believed to be caused by the momentum cutoff for the tau used in the simulated FTK trigger. At the time of this study, there were no simulated FTK triggers without a tau momentum cutoff available. More importantly, FTK was cancelled in its entirety and thus will not be included in Run 3. However, track information will be used in the HL-LHC trigger. I decided it was worthwhile to continue attempting to add track information to MVATES using simulated tau trigger information in order to explore possible improvements for later runs of ATLAS.

5.6.4 SIMULATED HLT TAU TRIGGER TRACK STUDIES FOR MVATES

For the rest of the studies, I used a standard simulated high-level trigger to include tau track information in MVATES. My training input was a simulated $Z \rightarrow \tau\tau$ sample with ~ 1 million events in both the training and validation sets. The studies detailed in this section include checking for over-training, feature selection, and training based on track multiplicity.

Check for Over-training of the BRT

A worry when training a BRT is over-training. This happens when you bias the BRT too much on the input dataset so that it performs extremely well on its training set but has poor performance on new events. There are many quick ways to test this, such as looking at the validation error. TMVA already splits the input validation set into a training and test set that is meant to help prevent over-training. However, since we apply the weights in a second step in MVATES, a further division is made. Half of the weights (the odd events) are used as input to the BRT, and the other half (the even events) are then used in the actual application and evaluation of MVATES. To determine if I was biasing the training on the odd event numbers, I trained on both the odd and even events separately and then applied those weights to both the training set and the unseen half (i.e. weights trained on even events were applied separately to both even and odd events). If the weights performed significantly better on the division they were trained on than the other half of the dataset, that was evidence for over-training. Figure 5.10 shows the performance of MVATES through the ratio of the MVATES calibrated momentum over true momentum of the taus for four different output datasets: weights trained on even events (Evaluation set) applied to both themselves and the odd events (Train set) and weights trained on odd events (Train set) applied to both the even events (Evaluation set) and themselves (Train set). The essentially identical performance of all four calibrated datasets suggested that there was not over-training occurring.

Feature Selection using Correlated Variables and Variable Rank

Another way to determine the best final input list of variables for the training of the BRT is to run the training with many variables and then look at the correlation and variable ranking that TMVA provides. A correlation plot was obtained (shown in figure 5.11) that demonstrated which variables were highly correlated (in yellow) or anti-correlated (in blue). Variables that were independent of each other are in

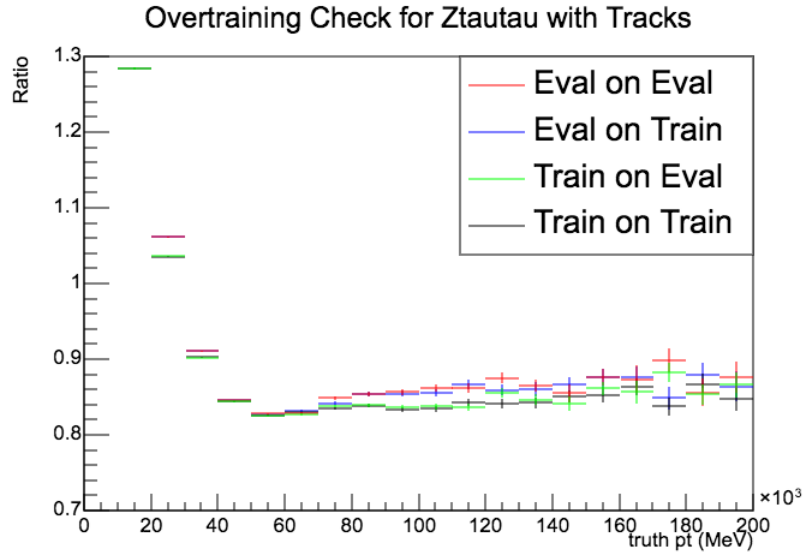


Figure 5.10: Ratio of calibrated momentum over truth momentum for weights trained on the even and odd (arbitrarily set) event numbers and then applied to both the input training set and the other unseen dataset. The comparable performance suggests that I am not over-training.

green. From this table, one can see that First Energy Density and Center Lambda were anti-correlated, Upsilon Cluster and EM Probability were anti-correlated, track one p_T and average track p_T were very correlated, total track p_T and average track p_T were very correlated, track 3 p_T and number of tracks were semi-correlated, and track two and track three p_T were also semi-correlated.

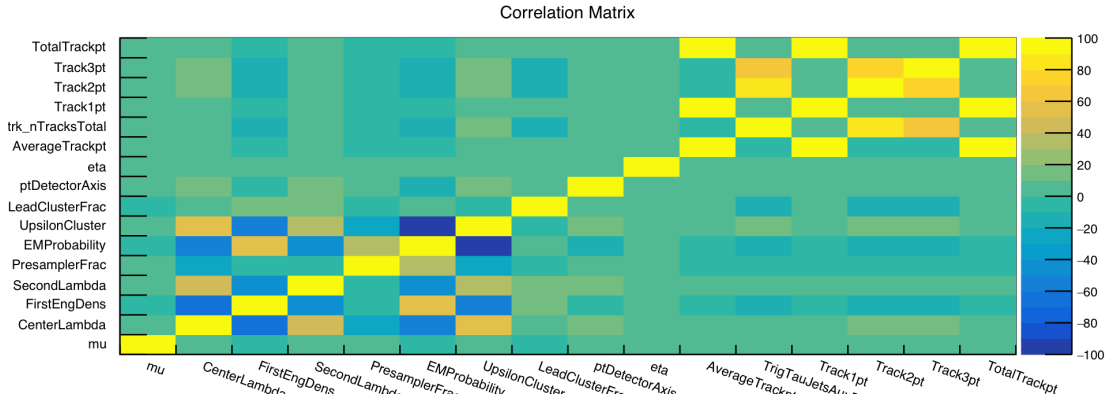


Figure 5.11: Correlation plot made in TMVA that shows which variables used in the training of the BRT supply similar information with regards to the target value for the regression. The values in yellow and blue are the ones most correlated and anti-correlated, respectively.

There were two attempts at removing the correlated variables:

1. *Without correlated variables*: removed Center Lambda, EM probability, Average Track p_T , and

Track 3 p_T

2. *With fewer correlated variables*: removed Center Lambda, EM probability, Track 1 p_T , Total Track p_T , nTracks, and Track 2 p_T

The MVATES calibration performance of these two reduced variable lists in relation to keeping all variables is shown in figure 5.12. The best performing weights are trained using the grouping *without correlated variables* for taus with transverse momentum greater than 40GeV.

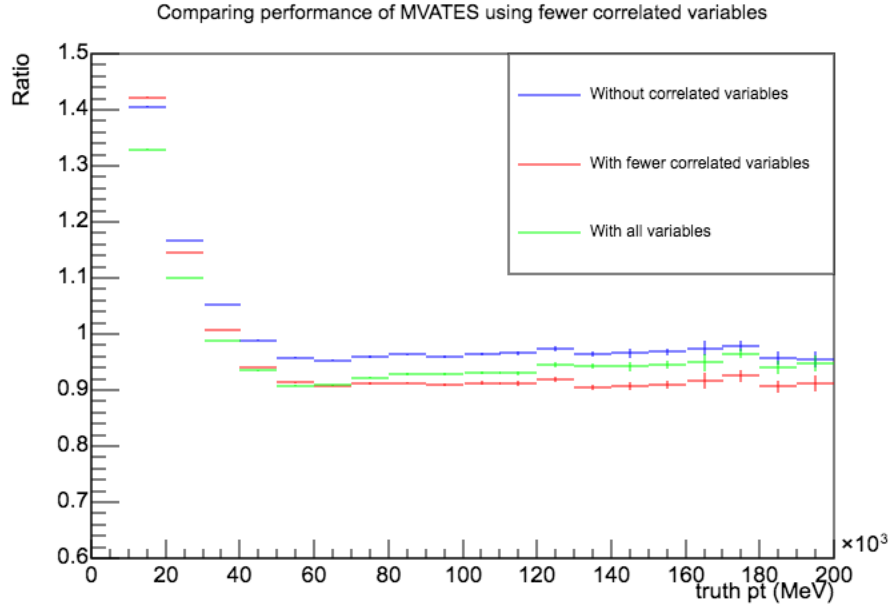


Figure 5.12: Comparison of performance of MVATES calibration when removing more or less correlated variables as described in section 5.6.4 with keeping all the variables. The best performance was the list of variables removing only 4 variables (without correlated variables.)

Another way of choosing which variables to keep is to look at their ranking of correlation with the target (truth p_T /detector p_T). The list of ranked variables is shown in table 5.4.

The values that signify how many tracks a tau has – subleading and track three weighted p_T and number of tracks – are highly ranked with respect to the target value. A training of the BRT was made using the top ten ranked variables, however removing variables like this severely hurt performance of the MVATES.

Ranking of Input Training Features	
1. Subleading track weighted p_T	9. P_{EM}
2. n_{Tracks}	10. λ_{center}
3. Detector p_T	11. μ
4. Track three weighted p_T	12. Track one Weighted p_T
5. $\langle \rho \rangle$	13. Total Track p_T
6. Lead Cluster Fraction	14. Average Track p_T
7. $f_{preseamler}$	15. $\langle \lambda^2 \rangle$
8. Upsilon Cluster	16. Detector η

Table 5.4: Ranking of the 16 input training variables (or features) based on their importance in determining the final predicted value.

Training on 1- and 3-Prong Taus

As I did when using the FTK tracks, I investigated whether training on inclusive taus or training on 1- and 3-prong taus separately would provide better calibration results. The new input variable list to the BRT includes the sum of track 1, 2, and 3 p_T weighted by total track p_T . When training using the simulated HLT track-trigger, training separately on 1- and 3-prong taus hurts MVATES performance compared to training on the taus inclusively.

5.6.4.1 TUNING OF HYPERPARAMETERS

After optimizing the input training set and the variables used, one can still tune the hyperparameters of TMVA to improve the performance of the weights. This includes both the type of loss function and structural parameters for the trees. I included a benchmark for these studies in my performance comparison, which are labeled the 2018 weights. These 2018 weights were trained without using track-level trigger variables.

NTrees and MaxDepth

Since the performance for both the removing correlated variables and using all track variables input lists for the BRT performed similarly well, I tuned the hyperparameters of both trainings. The values for:

- NTrees: which is the number of trees that is trained and averaged
- MaxDepth: the number of decision layers in each tree

were optimized via a grid search as there is currently no algorithmic way to do a random search of parameters in TMVA for regression.

As a demonstration, I will walk through the tuning of parameters for the all track variable list. First I varied the MaxDepth parameter, while the number of trees was kept constant. I show on the left in figure 5.13 that the performance using a depth of four or five is very similar, while shortening the trees to a depth of three caused the calibrated momentum to be too high.

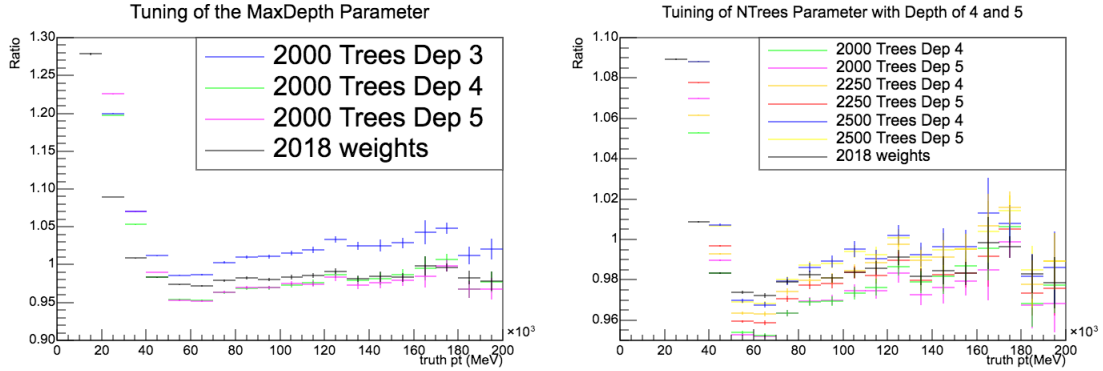


Figure 5.13: The left plot shows a comparison of performance of MVATES calibration using the BRT settings on the all tracks variable input list and changing the value of the MaxDepth while keeping NTrees constant. The right is a comparison of MVATES performance when changing the value of the NTrees while keeping MaxDepth at four or five. The 2018 weights are used as a benchmark

Next, I tuned the value of NTrees by varying its value from 2000 to 2500. Previous studies showed that going below 2000 and above 2500 hurt the performance of the BRT weights on the evaluation set. The right of figure 5.13 shows the offset from a ratio of one of calibrated momentum over truth momentum for all of the different weights trained with the various hyperparameters. The performance of the weights trained with 2500 Trees and either depth of four or five outperform the weights trained in 2018 on high momentum taus (for $p_T > 40\text{GeV}$).

For the removing correlated variables, the initial hyperparameter values of MaxDepth of five and 2000 trees performed best when checking a range of other values. These settings were used for the final comparison.

5.6.5 RESULTS OF MVATES WITH TRIGGER-LEVEL TRACK VARIABLES

Figure 5.14 sums up the main results and goal of this study. When adding in track variables to the training of MVATES, calibration performance improves for taus with p_T greater than 30GeV as compared to the weights trained in 2018 without track variables. Then with optimized tuning of the TMVA training, the ratio for calibrated p_T over truth p_T is near 1 for taus with p_T greater than 40GeV. This suggests that including trigger-level track information in the training of MVATES and then using this updated MVATES in the trigger decision for events with taus could be a benefit in future runs of the ATLAS detector (once track information can be included in the trigger decision).

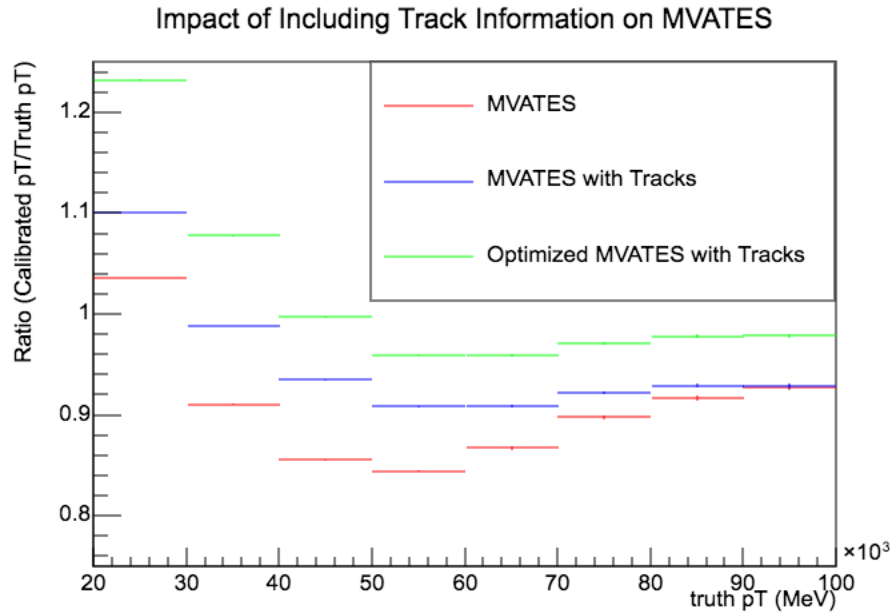


Figure 5.14: Comparison of performance of MVATES calibration with and without track variables and after tuning of the hyperparameters and input variable list.

5.7 LORENTZ EQUIVARIANT NEURAL NETWORK FOR TAU TAGGING

As shown earlier, the tau identification algorithm started at the beginning of Run 2 with a boosted decision tree algorithm and was upgraded to a better performing recurrent neural network. I spent some time investigating a potential next improvement to the tau ID that would identify taus using a Lorentz Equivariant neural network. Simply put, this means that the algorithm will classify a boosted and non-boosted tau as the same object without having to be trained on a wide variety of differently boosted taus.

5.7.1 WHAT DOES EQUIVARIANCE MEAN?

The notion of equivariance comes from both symmetry and covariance. In special relativity, equations can be written in a covariant format so that they look the same for every observer even though different observers will experience things differently. In particle physics we have built an entire Standard Model based on symmetry groups.

In Deep Learning, symmetry is usually thought of with respect to translations and rotations. Convolutional neural networks have become popular in part due to their inherent equivariance under translations. This means that convolutional layers need fewer parameters because you can use the same weights on different parts of the images and capture the same patterns [126]. If, for example, we are classifying images of a dog, the neural network will know that a dog shifted to the right in the image is still a dog.

We can also enforce equivariance by building structure into the neural network so that it respects symmetries. An example to illustrate that is forcing the network to react the same way to an object that is first transformed and then convolved versus one that is convolved and then transformed. In mathematical terms, that means:

$$f \circ \phi_g^X = \phi_g^Y \circ f \quad (5.4)$$

where f is a function mapping spaces X and Y ($f : X \rightarrow Y$) and ϕ is a transformation.

The advantages of enforcing equivariance in a neural network architecture include:

- increased data efficiency as you can share weight values
- decreased computational complexity as there are fewer trainable parameters
- disentanglement of orientation and presence in the latent space representation
- increased prospects of interpretability
- potential for more improvement when combined with data augmentation

The major disadvantage of equivariant architectures is that they require knowing the representation theory of the underlying group. However, for our task of identifying taus, we have the underlying Lorentz group symmetry, making tau tagging an ideal candidate for an equivariant architecture! Here the underlying energy-momentum vector of the particle depends on the inertial frame of the observer. We can transform between these frames in a way described by the Lorentz group $O(1, 3)$. This group combines spatial

rotations and Lorentz boosts. Enforcing equivariance allows us to train a neural network that will be able to identify variably boosted taus as the same object without having to train on a large dataset containing all those variations of the same object.

5.7.2 DATA PROCESSING

Input Samples

The original input samples are seven MC dijet samples with various jet momenta ranges for the background and a $\gamma^* \rightarrow \tau\tau$ sample for the signal. These are the same samples used to train the RNN tau identification. They are run through a standard ATLAS package that creates a common data format that can be customized for various tau projects, such as the tau identification and energy scale studies [127]. This processing step reduces the number of events for some files.

Data Formatting

In order to get the data into the right format for training, we first had to flatten the data files and then convert them to hdf5 files. Flattening is what takes the multi-particle per event information and breaks it down into a data entry per particle instead of per event. This was done using the same scripts the Tau Identification group uses [128]. During the flattening of the file, we chose to save a single 1- or 3-prong tau per event with the highest momenta. This removes some of our background jets and a handful of signal taus, but makes event weighting easier to implement. During the conversion to hdf5 format, we further select only reconstructed taus with a $p_T > 20\text{GeV}$ (and $p_T < 60\text{GeV}$ for initial studies). We remove any taus in the crack region or with an $|\eta| > 2.5$. Then these files are split between train, validation, and test sets with 50% in training and 25% each for validation and testing. All even event numbers are in the training set and then the odd numbered events are split evenly between test and validation datasets.

Once a single tau per event has been saved that meets the basic requirements, we then convert the data into the correct format for the Lorentz equivariant network. The network takes in the four momenta of each object (track and cluster) in the jet, and any scalar variables we want to include, like mass and a label to distinguish tracks from clusters. We convert p_T , η , and ϕ into (E, p_x, p_y, p_z) for the first ten tracks and six clusters in the reconstructed tau. Each object is then labeled as a track, cluster, or set to a default value of zero if fewer than ten tracks and/or six clusters. Input taus are also labeled as signal ('1') or background

(‘0’). The number of objects (tracks and clusters) per tau is also saved. In addition, we include a particle weight that is used in calculating the loss and performance of the network during training.

Weighting the Background

For the current studies, we are just using one signal sample and the lowest p_T dijet sample. To simplify matters, we train on the first 25, 000 taus in each sample (and 13, 000 are included for validation and test sets), so we do not need to use class weights between the signal and background inputs. The signal objects are weighted by their MC event weight and the background events are weighted by (Object MC event weight)/(Mean of all MC event weights in sample).

When we include more background samples, we will then need a two-step weighting procedure. We will first need to weight all of the backgrounds relative to each other and then weight all of the background so that its total relative importance matches that of the signal. The background event weighting is to even out the numbers of low p_T taus with the artificially high number of taus in the simulated samples at high p_T with very low MC event weights. Each background sample (with varying DSID) contains a different range of potential p_T values. The weighting equation for background is to divide the MC event weight by the Luminosity for that DSID. Luminosity is calculated by:

$$L = \frac{N}{\sigma * eff * eff_{proc}} \quad (5.5)$$

where N is the number of taus in the sample, σ is the cross-section of that DSID (in pb), eff is generator efficiency, and eff_{proc} is the processing efficiency (% taus remaining after our data processing pipeline). The input values for this equation for all background files and the signal file are included in table 5.5.

Once these background weights are applied, we obtain a smooth distribution in jet p_T for all background jets, as displayed in figure 5.15.

5.7.3 ARCHITECTURE OF NETWORK

The architecture used is detailed in [129] and the code can be found here: [130]. The input layer takes in the four-momenta of N_{obj} particles for the reconstructed tau and associated scalars (mass, label). For our inputs, N_{obj} cannot be greater than 16. The input is then a set of vectors in $(T^{(0,0)})^{\oplus \tau_0} \oplus T^{(1,1)}$ representation of the Lorentz space. τ_0 is the number of scalars we put in, such as object label. $T^{(1,1)}$ is

DSID	N_{events}	σ (pb)	eff	eff_{proc}
Dijets p_T bin 1	98896	$78.45e9$	0.00067	0.6539
Dijets p_T bin 2	145185	$2.43e9$	0.0003	0.9679
Dijets p_T bin 3	149545	$2.65e7$	0.0003	0.9969
Dijets p_T bin 4	149930	$2.55e5$	0.0005	0.999
Dijets p_T bin 5	149979	4550	0.0009	0.999
Dijets p_T bin 6	149994	258	0.0009	0.999
Dijets p_T bin 7	150000	16.2	0.0004	1.0
$\gamma^* \rightarrow \tau\tau$ Signal	126274	0.00322	1	0.6314

Table 5.5: Relevant values for weighting the background samples, which are dijet samples with various binning in p_T with bin one having low p_T and bin seven having the highest p_T jets, and the signal sample for training the Lorentz equivariant neural network if more than one background sample is being used.

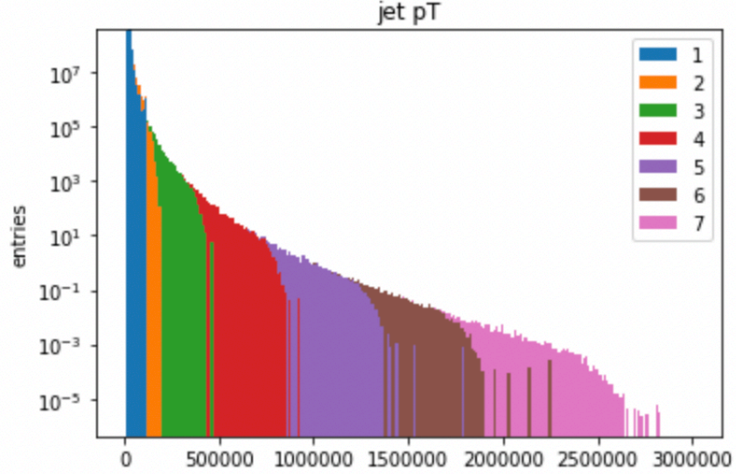


Figure 5.15: Weighted distribution of jet p_T for all background samples, the legend refers to the p_T bin for the sample, as defined in table 5.5.

our p_μ vector: (E, p_x, p_y, p_z) . The input layer is fully-connected and creates $N_{ch}^{(0)}$ vectors, where N_{ch} is one of the tunable parameters, and 0 refers to the input layer. The input layer shares weights between all inputs, but acts on each one individually.

The input layer is followed up with N_{obj} activations, $\mathcal{F}_i^{(0)}$, where i goes from 1, ..., N_{obj} . These activations live in the $(T^{(0,0)}) \oplus T^{(1,1)^{\oplus N_{ch}^{(0)}}}$ space. This is followed by an update layer:

$$\mathcal{F}_i^{(p+1)} = \mathcal{L}_{CG}(\mathcal{F}^{(p)})_i = W \cdot \left(\mathcal{F}_i^{(p)} \oplus CG[\mathcal{F}_i^{(p)}]^{\otimes 2} \oplus CG \left[\sum_j f(p_{ij}^2) p_{ij} \otimes \mathcal{F}_j^{(p)} \right] \right) \quad (5.6)$$

The next layer takes in the previous layer $\mathcal{F}_i^{(p)}$ and calculates the direct sum of that layer and a self-interaction term of the input layer (previous layer in network). The term in square brackets includes a learnable scalar function f that weights the interactions of the i 'th particle with other particles. This is applied to p_{ij}^2 , where $p_{ij} = p_i - p_j$ is the pairwise difference. The sum over j ensures permutation invariance. The two terms with direct products go through a Clebsch-Gordan (CG) decomposition. The number of terms saved in that representation is also a tunable parameter of the network. Finally, W is the equivariant learnable operator of the network. It is actually a sparse matrix made up of smaller matrices on its diagonal which only connect vectors of the same weight.

This CG layer is then followed up by as many MLP layers as desired. These are applied to the scalar inputs and the weights are shared across all N_{obj} nodes. The final output layer projects onto Lorentz-invariant outputs and produces two scalar weights to be used in classification: $w_{out} = W_{out} * \left(\sum_i \mathcal{F}_i^{(N_{CG})} \right)_{(0,0)}$.

5.7.4 INITIAL RESULTS

For the initial training, we ran on 25,000 signal and 25,000 of the low p_T dijet background events. We tried varying parameters such as number of CG layers, number of channels, learning rate, and batch size. After ~ 200 epochs, the model either crashed due to too much memory consumption from the tensor calculations or settled on achieving an AUC of 0.8, accuracy of 70%, and background rejection of 48. The model is successfully learning, but it is not yet tuned to match the efficiency of the current RNN Tau ID. Furthermore, it is not the only interesting potential future direction for the Tau ID.

5.7.5 COMPARISON TO DEEP SETS ARCHITECTURE

We also trained a Deep Sets architecture on the same signal and background files, with 25,000 events from each sample for training and 13,000 each for validation.

The network begins by taking in the four momenta as a vector and the label for the clusters and tracks. The labels are put through an embedding layer and then concatenated with the momenta vector. Then there are three dense fully connected layers with 255 nodes. Then we construct a mask to kill zero-padded particles (those with less than 16 clusters and tracks). Then we sum over non-zero particles by multiplying with that mask. This is followed up with a dense layer with 512 nodes and final dense layer with two, one for signal and one for background. There are a total of 153,376 trainable parameters in this model.

We chose Adam for the optimizer and categorical cross entropy for the loss function. The final layer used the softmax activation function, while the previous dense layers used ReLu. After ten epochs with a learning rate of $1e - 5$, we achieved a training loss of 0.255, training accuracy of 0.893, validation loss of 0.2745, validation accuracy of 0.890, a validation background rejection at 30% signal of 142, and a validation AUC of 0.943. Based on our initial studies, this model performs well at tagging taus, but it requires a large number of trainable parameters.

5.8 CONCLUDING THOUGHTS ON MACHINE LEARNING FOR TAUS

As you can see there are many ongoing studies that aim to improve our reconstruction and identification of taus, both online at the trigger level and in our offline environment. This will become an even higher priority as we head to higher center of mass energies and increased interaction rates in future runs of ATLAS. We are exploring both incorporating new detector information into our models, such as introducing track information at an earlier stage of the trigger to more precisely reconstruct the tau energy, and designing more sophisticated models that can better identify taus. We saw a significant jump in performance with the switch to an RNN from a BDT algorithm for identification, and there is promise seen in other ML methods, such as Lorentz equivariant neural networks and Deep Sets architectures. It is quite an exciting time for this intersection of particle physics and ML research.

*“Autumn leaves don’t fall, they fly. They take their time
and wander on this their only chance to soar.”*

Delia Owens, *Where the Crawdads Sing*

6

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

ONE OF THE MAIN GOALS OF THE LHC, AFTER DISCOVERING THE HIGGS BOSON, IS TO DETERMINE HOW WELL THE NEW HIGGS BOSON PARTICLE FITS THE STANDARD MODEL THEORY. The Standard Model provides predictions for the cross sections and branching ratios of the Higgs boson. With the data we collect at ATLAS, we can test these predictions. $VH, H \rightarrow \tau\tau$ is the shorthand for the associated production of a Higgs boson with a vector boson (W or Z), with the Higgs boson decaying to two taus. $VH, H \rightarrow \tau\tau$ is one of the three most accessible channels that is not part of the combined Run 2 Higgs boson cross section measurements (see figure 6.1) [131]. Measuring $H \rightarrow \tau\tau$ decays could give us insight into how the Higgs boson is responsible for lepton masses via the Yukawa couplings. As the most massive lepton, the tau is the best chance for studying Higgs boson and lepton interactions at ATLAS. By combining the $H \rightarrow \tau\tau$ decay mode with VH production, the events can be selected via an electron or muon trigger with minimal bias on the Higgs boson, which could prove useful for future studies of $H \rightarrow \tau\tau$ decays.

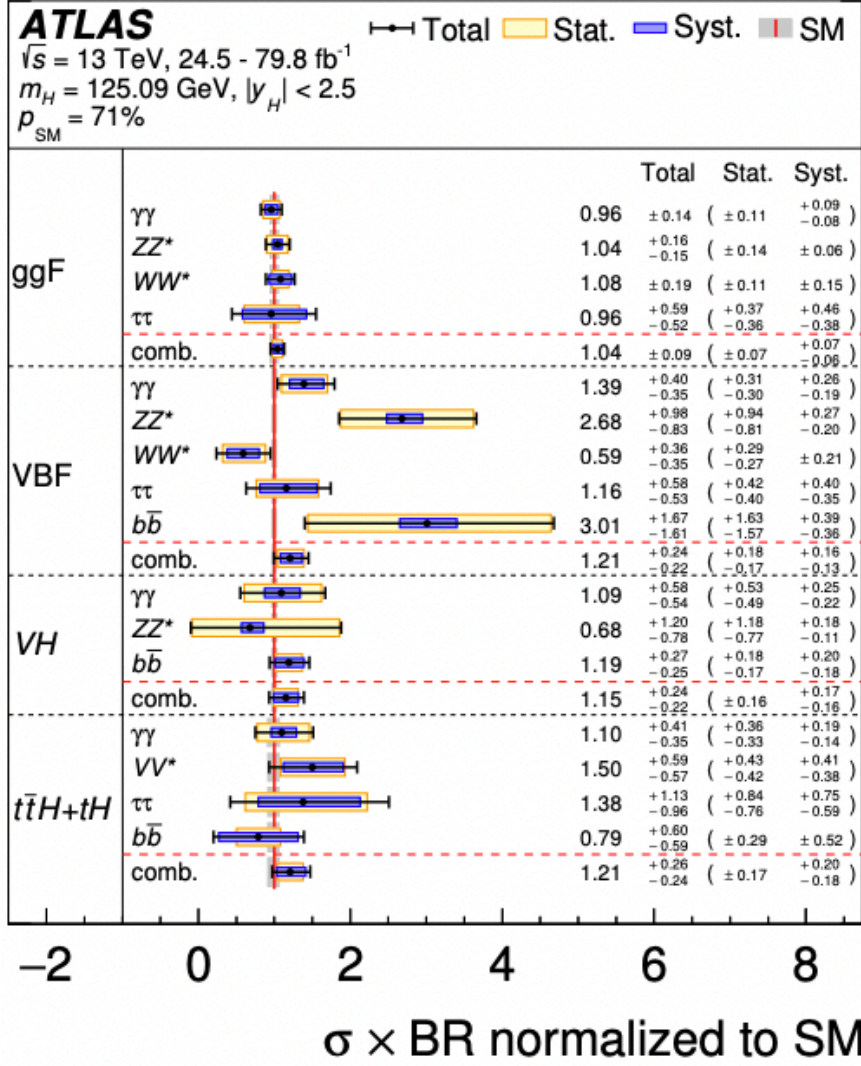


Figure 6.1: Cross section times branching ratio for all of the main production mechanisms and decays included in the current Run 2 data ($\sqrt{s} = 13\text{TeV}$) measurements of the Higgs boson, normalized to their SM predictions. This list is missing three primary Higgs boson production and decay combinations, including VH , $H \rightarrow \tau\tau$. The black error bars show the total uncertainties, which take into account both the systematic (in blue) and statistical (in yellow) contributions.

6.1 MOTIVATION

Besides the theoretical motivation for this channel, there is clear experimental evidence to suggest we should be able to measure it with the full ATLAS Run 2 dataset. Assuming the Standard Model Higgs boson cross section at $\sqrt{s} = 13\text{TeV}$, 4.08% of Higgs boson production is via VH (associated production with a vector boson). Furthermore, $H \rightarrow \tau\tau$ has been observed with $> 5\sigma$ from the ggF and VBF

Higgs boson production modes [132]. Due to the relatively large mass of the tau with respect to other fundamental particles, the branching ratio of $H \rightarrow \tau\tau$ is 6.32% [133].

This Run 2 analysis builds on a simpler implementation that was conducted with Run 1 data. The Run 1 study looked at 20.3 fb^{-1} of integrated luminosity at 8TeV. They achieved an observed combined signal strength of $\sigma/\sigma_{SM} = 2.3 \pm 1.6$, and set upper limits on cross sections for each category [42]. In Run 2, we have a much higher total integrated luminosity (from 20.3 fb^{-1} to 139 fb^{-1}) and an increased center of mass energy (from $\sqrt{s} = 8\text{GeV}$ to $\sqrt{s} = 13\text{GeV}$). A higher center of mass energy gives us more opportunities to produce a Higgs boson and more data allows us to reduce our statistical uncertainties. A recent CMS analysis that used machine learning techniques on 2016 and 2017 data (a subset of Run 2), measured $\sigma_{incl}\mathcal{B}(H \rightarrow \tau\tau) = 2.56 \pm 0.48(\text{stat}) \pm 0.34(\text{syst}) \text{ pb}$ [134].

For this process to occur in ATLAS, the W or Z boson is produced from a quark quark pair. The Higgs boson is then radiated from the W or Z boson. The cross-section for WH is higher than for ZH because the W boson is approximately 10GeV less massive than the Z boson. As illustrated in the leading order Feynman diagram shown in Figure 6.2, both a Higgs boson and a vector boson emerge from the interaction.

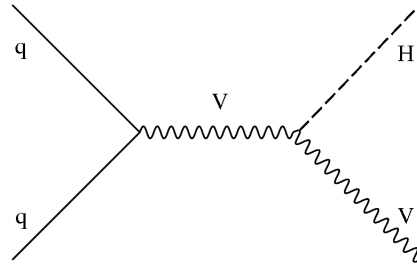


Figure 6.2: Leading order Feynman Diagram for Associated Production of the Higgs boson with a vector boson (W or Z).

$VH, H \rightarrow \tau\tau$ is an excellent channel for studying Higgs boson physics for a few reasons. We search for events in which a Z or W boson decay produces light leptons (e or μ) and light leptons have very high trigger efficiencies in the ATLAS detector. About 21.6% of the time the W boson decays to an electron/muon and a neutrino, and about 6.7% of the time the Z boson decays to pairs of electrons or muons. Additionally, some fraction of the W or Z boson decays to tau leptons can also produce a light lepton with enough energy to meet our selection cuts. The $H \rightarrow \tau\tau$ decay suffers from a high background

from quark- or gluon-initiated jets, because jets are produced at much higher rates than taus, and some fraction of these jets will trick the tau identification algorithm and be misclassified as taus. Our additional constraint that the associated vector boson decay produces light leptons is important because it allows for a reduction in this QCD-initiated background. However, the light-lepton requirement does lower the cross-section for the measurement, with the hadronically decaying W and Z bosons not contributing to the signal.

In addition, we require that at least one of the two taus decay hadronically so that we stay orthogonal to other ATLAS Higgs boson analyses, such as $H \rightarrow ZZ^* \rightarrow lll$, to enable a future combination. Note that our requirement of the Z or W to decay leptonically is particularly important in the categories where both taus from the Higgs boson decay hadronically, so that there is still a muon or electron to trigger on. Furthermore, for the ZH channels, we can reduce backgrounds by requiring a reconstructed Z boson mass from the dilepton (ee or $\mu\mu$) pair. This gives us a clean handle on those events. Unfortunately, due to the multiple sources of neutrinos in the $WH, H \rightarrow \tau\tau$, we cannot fully reconstruct the W boson mass.

The four $H \rightarrow \tau\tau$ channels studied in this analysis are:

$$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$$

$$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$$

$$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$$

$$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$$

where τ_{had} refers to a hadronically decaying tau and τ_{lep} refers to a leptonically decaying tau, which results in either a e or μ in our final state.

6.2 FOUR SIGNAL CATEGORIES

6.2.1 $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$

The signature for this channel will be two light leptons (one from the tau decay and one from the W boson decay) and a hadronic tau. We associate the lowest p_T light lepton with the Higgs boson. This should be the most abundant decay channel in the analysis because a Higgs boson decay to $\tau_{lep}\tau_{had}$ occurs 46% of

the time, but the restriction of the same-sign leptons lowers the number of final state events by about 50%. The leptons are required to be the same-sign in order to reduce background from neutral bosons, such as the $Z \rightarrow ee$ or $Z \rightarrow \mu\mu$ decays.

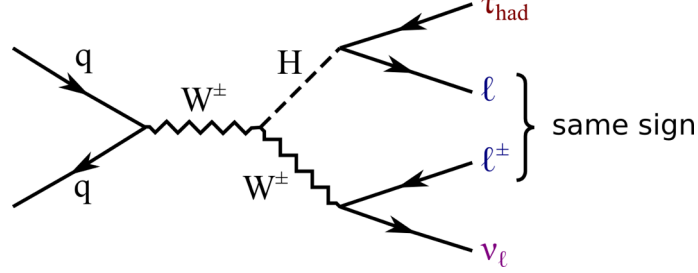


Figure 6.3: Feynman diagram for the WH production mode where the Higgs boson decays to a pair of taus with one tau decaying hadronically and the other decaying leptonically.

6.2.2 $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$

In this analysis channel the Higgs boson decays to two hadronically decaying taus, and the W boson decays to a light lepton and neutrino. The final state is therefore several hadrons, most often pions, from the hadronic tau decays, an electron or muon, and three neutrinos. We require that the two hadronic taus be of opposite sign because the Higgs boson is a neutral particle. The single lepton in the event is associated with the W boson.

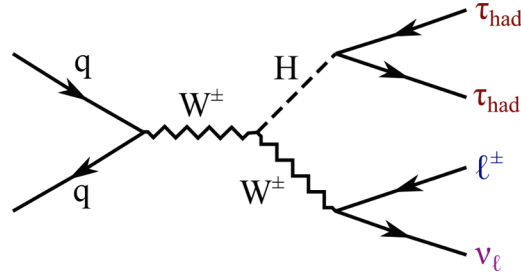


Figure 6.4: Feynman diagram for the WH production mode where the Higgs boson decays to two hadronically decaying taus.

6.2.3 $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$

This category occurs when the Z boson decays to two light leptons and the Higgs boson decays to one hadronic and one leptonic tau. The two leptons that are same-flavor and opposite-sign with a mass closest

to the Z mass, 91 GeV, are associated with the Z boson. The hadronic tau and remaining light lepton are required to be of opposite-sign.

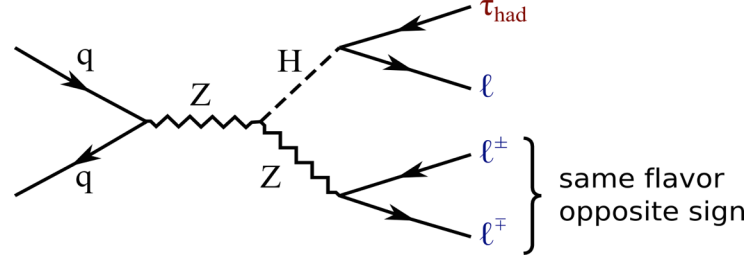


Figure 6.5: Feynman diagram for the ZH production mode where the Higgs boson decays to a hadronically decaying tau and a leptonically decaying tau.

6.2.4 $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$

The final category in the analysis occurs when the Z boson decays to two same-flavor opposite-sign light leptons, i.e. two muons or two electrons, and the Higgs boson decays to two hadronic taus. A benefit of this channel is the ability to reconstruct a Z boson from the two opposite-sign leptons. We also require the two hadronic taus to be of opposite-sign.

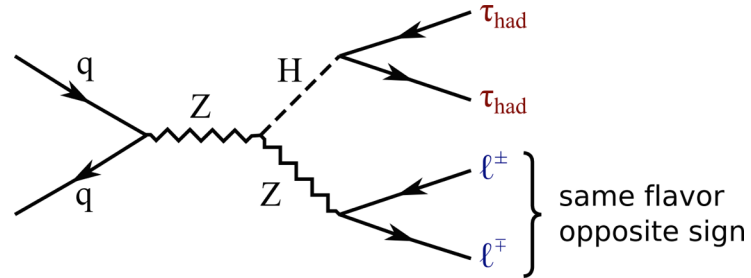


Figure 6.6: Feynman diagram for the ZH production mode where the Higgs boson decays to two hadronically decaying taus.

6.3 DATA AND MONTE CARLO

The data, and corresponding simulation, is from Run 2 of the ATLAS experiment, which corresponds to data taken in 2015 through 2018. Only data that was deemed good for physics (i.e. passes event selection to ensure that the detector was in good working order) is used in the analysis. This provides a total of 139 fb^{-1} of data, with $\approx 3fb^{-1}$ from 2015, $\approx 33fb^{-1}$ from 2016, $\approx 43fb^{-1}$ from 2017, and $\approx 58fb^{-1}$

from 2018. For our analysis, both the data and MC is from two different derivations: HIGG4D1 for the ZH categories and fake factor calculations and HIGG4D2 for the WH categories.

Event Cleaning

All events from 2015 and 2016 with excessive noise in the electromagnetic calorimeter are removed. The cut was developed based on the number of high- p_T clusters with poor quality in that region of the detector.

Monte Carlo Simulation Samples

Simulated datasets for the $VH, H \rightarrow \tau\tau$ signal as well as all background samples were created using various Monte Carlo (MC) generators, as described in Table 6.1. Samples of Monte Carlo (MC) simulated events are used to optimize the event selection, and to model the signal and several of the background processes. The samples were produced with the ATLAS simulation infrastructure using the full detector simulation performed by the GEANT4 [87] toolkit.

The Higgs boson mass was set to $m_H = 125\text{GeV}$. The associated Higgs boson production modes (WH, ZH) and the ttH production mechanisms are considered individually, while the others Higgs boson production mechanisms are grouped together in our “other” backgrounds. The cross-sections of all Higgs boson production processes were fixed to the SM predictions [133]. Other background processes involve electroweak production of W/Z bosons via VBF, Drell–Yan production of W/Z in association with jet(s) as well as dibosons, single top-quark and top-quark pair ($t\bar{t}$) production. The generators and parton shower models used to simulate different processes are summarized in Table 6.1.

Data Processing Pipeline

After the derivations are made, we use a tau-specific suite of software to apply standard calibrations, or corrections, to reconstructed objects. We also use this framework to implement our event selection requirements. The output format is a ROOT tree, or nTuple.

For our analysis, we then use uproot [135] to convert the variable-length vectors in the ROOT file to set-length Pandas DataFrames [136] stored in a HDF5 binary file format. These dataframes are then run through our internal Python-based analysis framework.

Process	MC Generator + UEPS	PDF Set	Perturbative Order
$W \rightarrow l\nu, H \rightarrow \tau\tau$	POWHEG+PYTHIA8	NNPDF30NLO	NLO
$Z \rightarrow ll, H \rightarrow \tau\tau$	POWHEG+PYTHIA8	NNPDF30NLO	NLO
$ggF H \rightarrow \tau\tau$	POWHEG+PYTHIA8	NNPDF30NNLO	NNLO
VBF $H \rightarrow \tau\tau$	POWHEG+PYTHIA8	NNPDF30NLO	NLO
$ttH, H \rightarrow \tau\tau$	POWHEG+PYTHIA8	NNPDF23LO	LO
Diboson	Sherpa 2.2.2 and 2.2.1	NNPDF30NNLO	NNLO
Triboson	Sherpa 2.2.2	NNPDF30NNLO	NNLO
W +jets	Sherpa 2.2.1	NNPDF30NNLO	NNLO
Z +jets	Sherpa 2.2.1	NNPDF30NNLO	NNLO
t	POWHEG+PYTHIA8	NNPDF30NLO	NLO
$t\bar{t}$	POWHEG+PYTHIA8	NNPDF30NLO	NLO

Table 6.1: Monte Carlo generators used to produce the main simulated processes incorporated into the $VH, H \rightarrow \tau\tau$ analysis in Run 2, including the physics process, MC generator and the model of the underlying event with hadronization and parton showering (UEPS), the corresponding PDF set, and the perturbative order in QCD to which the cross-section has been calculated (LO = Leading order, NLO = Next-to-leading order, NNLO = Next-to-next-to-leading order).

6.4 OBJECT CRITERIA

Beyond the basic requirements for each object to be reconstructed as in Chapter 3 and Chapter 5 for taus, we apply additional requirements on physics objects in our analysis.

Muons

All muons in our analysis must have a $p_T > 9\text{GeV}$ and have $|\eta| < 2.5$, after being reconstructed from hits in the Inner Detector and Muon Spectrometer. Muons with $p_T > 2\text{GeV}$ and passing the Loose identification working point are part of the overlap removal process. For a muon to be one of the final state leptons in our analysis it must additionally pass the Tight identification working point, corresponding to a $\sim 90 - 92\%$ efficiency. It must also pass the track isolation criteria:

$$p_T^{varcone30} < 0.06 * p_T^\mu \quad (6.1)$$

which means that for muons with a transverse momentum of less than 50GeV , the scalar sum of the muon p_T in a cone of $\Delta R = \min(10\text{GeV}/p_T^\mu, 0.3)$ is less than 6% of the total muon p_T . If $p_T^\mu > 50\text{GeV}$, then a radius of $\Delta R < 0.2$ is used. Muons that are not used to satisfy signal requirements, but to estimate

backgrounds from misidentification, may fail one, or both, of the identification and isolation working points.

Electrons

Electrons must have a $p_T > 13\text{GeV}$, an $|\eta| < 2.47$, and have a matching track in the Inner Detector and corresponding cluster in the EM calorimeter. They must also not be in the crack region of the detector: $1.37 \leq |\eta| \leq 1.52$. Finally, to be included in the overlap removal process, the electron candidate must also pass the Very Loose Likelihood working point for electron identification. For the electron to be considered part of our signal selection, it must also pass the Tight identification working point and an isolation requirement. The Tight identification is 65% efficient for electrons with $p_T = 15\text{GeV}$ and 90% efficient at 100GeV . The isolation requirement has both a track and calorimeter isolation requirement. For the track requirement,

$$p_T^{iso}(\Delta R^{var} < 0.3)p_T < 0.15 \quad (6.2)$$

meaning that good-quality tracks within a cone of a p_T -dependent ΔR radius around the electron must have a scalar sum of p_T that is less than 15% of the total electron p_T . ΔR^{var} ranges from a maximum of 0.2 and then decreases by $10\text{GeV}/p_T$, hence the p_T -dependent nature. For the calorimeter isolation, we require:

$$E_T^{iso}(\Delta R < 0.2)/E_T < 0.2. \quad (6.3)$$

The calorimeter topo-clusters in a $\Delta R = 0.2$ around the electron must have a summed transverse energy of less than 20% of the total electron E_T . Electrons that are not used to satisfy signal requirements, but to estimate backgrounds from misidentification, may fail one, or both, of the Tight identification and isolation working points.

Taus

The taus used in this analysis are seeded by anti- k_T jets with $R = 0.4$. In addition, they must have $p_T > 20\text{GeV}$, an $|\eta| < 2.5$, and not be in the crack region ($1.37 \leq |\eta| \leq 1.52$). We also require the taus to have a charge of $|q| = 1$ and either one or three reconstructed tracks. For our baseline requirement, we check that taus do not fall into the bottom 1% of tau RNN ID scores from a distribution flattened vs. pileup and p_T . We also apply a dedicated muon veto criteria and select taus that pass a Loose working point (0.05)

in the electron veto (eBDT). Finally, for a tau to pass our signal criteria it must pass the Medium identification working point. Taus that are not used to satisfy signal requirements, but to estimate backgrounds from misidentification, may be used in our fake factor calculations.

Jets

While jets are not one of our final state signal particles, they are important to include in our overlap removal process and when rejecting b -jets. Like taus, jets are reconstructed from seeds from the anti- k_T algorithm with $R=0.4$. Jets must have a $p_T > 20\text{GeV}$ and an $|\eta| < 4.5$. Jets must also pass basic cleaning criteria to be included in the overlap removal. For the analysis, we apply an 85% efficient jet b -tagging working point, in channels where events with b -jets are vetoed.

6.5 OVERLAP REMOVAL

Sometimes in reconstructing particles, ATLAS can use the same tracks or calorimeter clusters, or tracks and clusters that fall in a small ΔR region, to reconstruct two different types of particles. Since we have three or four selected objects, depending on the analysis category, we need to make sure that they do not overlap such that we are tagging one object as two distinct particles. In order to separate out these particles, an ordered process for removal is applied. The detailed sequence is below, but it can be summarized as a preference for electrons and muons over taus, and then taus over jets.

The overlap removal order is to choose:

1. Electrons over taus (within $\Delta R < 0.2$)
2. Muons over taus (within $\Delta R < 0.2$)
3. Electrons over calo-tagged muons if they share track(s) in the Inner Detector
4. Remaining muons over electrons if they share track(s) in the Inner Detector
5. Electrons over jets (within $\Delta R < 0.2$)
6. Non-pileup jets over electrons (within $\Delta R < 0.4$)
7. Muons over jets (if less than three tracks and within $\Delta R < 0.2$)
8. Non-pileup jets over muons (within $\Delta R < 0.4$)
9. Hadronic taus over jet (within $\Delta R < 0.2$)

6.6 MASS RECONSTRUCTION

Due to the presence of neutrinos in the decays of our W boson and tau leptons, we cannot directly reconstruct the mass of the Higgs boson from the visible decay products. As a reminder, decays of the tau lepton via a hadronic channel ($\tau \rightarrow \nu_\tau + \text{hadrons}$) or leptonic channel ($\tau \rightarrow \nu_\tau + l + \bar{\nu}_l$) produce at least one neutrino. Since ATLAS cannot detect the neutrinos directly they are accounted for in our missing transverse energy (E_T^{miss}) calculation. The crucial complication here is the fact that E_T^{miss} can be measured at the event level, but not the object level. We therefore cannot simply calculate the missing energy associated with each parent particle; in our case, the W and Higgs boson.

In order to recover some information regarding the Higgs boson mass, for the ZH channels we use a reconstruction method known as the Missing Mass Calculator (MMC) and for the WH channels that have that additional neutrino from the W boson decay, we use the last-projected transverse mass (M_{2T}).

6.6.1 MISSING MASS CALCULATOR (MMC)

The Missing Mass Calculator (MMC) is used for the two ZH categories. The algorithm scans possible possible neutrino momenta from the di-tau decay to reconstruct the Higgs boson mass [137]. Developed as an improvement to the collinear mass approximation, the MMC algorithm was crafted to reconstruct the true mass of the parent particle. The calculation assumes that there are no extra sources of neutrinos (or missing energy) in the event other than from the Higgs boson’s tau decays. The calculator scans for both the tau’s visible mass and the x and y-components of E_T^{miss} for the event. There are four equations that constrain the unknown variables:

$$\begin{aligned}
 \cancel{E_{T_x}} &= p_{\text{mis}_1} \sin\theta_{\text{mis}_1} \cos\phi_{\text{mis}_1} + p_{\text{mis}_2} \sin\theta_{\text{mis}_2} \cos\phi_{\text{mis}_2} \\
 \cancel{E_{T_y}} &= p_{\text{mis}_1} \sin\theta_{\text{mis}_1} \sin\phi_{\text{mis}_1} + p_{\text{mis}_2} \sin\theta_{\text{mis}_2} \sin\phi_{\text{mis}_2} \\
 M_{\tau_1}^2 &= m_{\text{mis}_1}^2 + m_{\text{vis}_1}^2 + 2\sqrt{p_{\text{vis}_1}^2 + m_{\text{vis}_1}^2} \sqrt{p_{\text{mis}_1}^2 + m_{\text{mis}_1}^2} - 2p_{\text{vis}_1} p_{\text{mis}_1} \cos\Delta\theta_{v m_1} \\
 M_{\tau_2}^2 &= m_{\text{mis}_2}^2 + m_{\text{vis}_2}^2 + 2\sqrt{p_{\text{vis}_2}^2 + m_{\text{vis}_2}^2} \sqrt{p_{\text{mis}_2}^2 + m_{\text{mis}_2}^2} - 2p_{\text{vis}_2} p_{\text{mis}_2} \cos\Delta\theta_{v m_2}
 \end{aligned} \tag{6.4}$$

where $\cancel{E_{T_x}}$ is the x-component of the event’s missing transverse energy; $\cancel{E_{T_y}}$ is the y-component of the event’s missing transverse energy; the subscript *vis* refers to the visible τ decay products’ momenta, mass, polar (θ) angle, and azimuthal (ϕ) angle. The *mis* subscript refers to the “missing”, unknown, com-

ponents of the invisible neutrinos [137]. M_τ is the τ mass and $\Delta\theta_{vm1,2}$ is the angle between the visible and missing momentum vectors for the tau-lepton decay. If a tau has decayed hadronically, the corresponding value of m_{mis} is set to zero.

Because there are more unknown variables than equations, the system is underconstrained. There are six unknown variables when both taus decay via the hadronic channel, which are the x -, y -, and z -components of the invisible momenta from the neutrinos. There is a seventh unknown variable, the invariant mass of the neutrino, when there is a leptonic tau decay. Thus additional kinematic information is included to better determine the di-tau mass, such as $\Delta R = \sqrt{(\eta_{vis} - \eta_{mis})^2 + (\phi_{vis} - \phi_{mis})^2}$. The algorithm searches through possible neutrino momenta until the most-likely di-tau mass is calculated.

The better constrained case, with just six unknown variables, is when both taus decay hadronically and $m_{mis} = 0$, i.e. the $Z \rightarrow ll, H \rightarrow \tau_{had}\tau_{had}$ channel for our analysis. We can then solve the equations above explicitly for any point in $(\phi_{mis1}, \phi_{mis2})$ space. This means we will conduct a grid search over the kinematically-allowed ϕ_{mis} pairs, and calculate fully defined momentum vectors. We then weight each of the calculated $M_{\tau\tau}$ by $\mathcal{P}(\Delta R_1, p_{\tau_1}) \times \mathcal{P}(\Delta R_2, p_{\tau_2})$.

For the $Z \rightarrow ll, H \rightarrow \tau_{lep}\tau_{had}$ category, we have an additional unknown due to the extra neutrino from the leptonic tau decay. Now, we have to perform the scan over $(\phi_{mis1}, \phi_{mis2}, m_{mis1})$ including all kinematically allowed values of m_{mis1} .

6.6.2 LATE-PROJECTED TRANSVERSE MASS (M_{2T})

The MMC algorithm requires that all of the event missing energy comes from the neutrinos from the Higgs boson decay, but in our WH categories there is an extra neutrino from the W decay. We will then use the M_{2T} , late-projected transverse mass, algorithm for the WH categories. M_{2T} sets a lower bound on the mass of the parent particles. For our signal events, M_{2T} is bounded above by the invariant Higgs boson mass. The term “late-projection” refers to the order of the summing and projecting operations. Here, we first sum the Lorentz vectors and then project them into the transverse plane. The plane transverse to the beam is chosen for the projection because the incoming momentum in the transverse direction is known to be zero. M_{2T} is a specific case of a more general early-partitioned, late-projected invariant mass algorithm. [41].

We can divide the particles in our events into two categories: visible and invisible. *Visible* particles can

be measured in our detector, such as electrons and muons, and *invisible* particles are neutrinos, or other objects, that leave no direct signal. Each parent particle can have visible and invisible particles associated with its decay chain. This division of particles is shown in figure 6.7. The transverse momentum of each group of visible and invisible daughters is calculated individually.

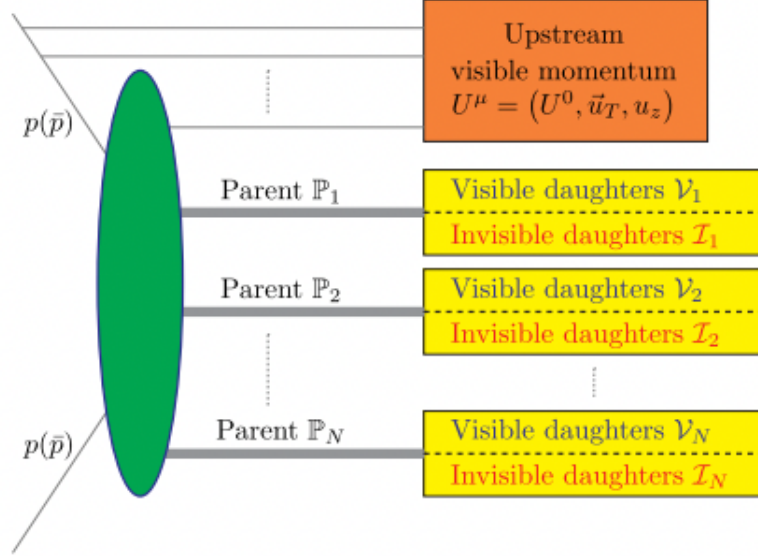


Figure 6.7: Illustration of parent particles in an event and their associated visible and invisible products. The upstream visible momentum is there to include all visible products not part of one of the parent sets. [41]

For the calculation, each τ has only one considered invisible object in the final state, meaning the two neutrinos from leptonic tau decays are treated as if they were one neutrino. This means that the W boson and the Higgs boson each have two (one and three)-dimensional four vectors, one for the visible and one for the invisible components. These are then combined into the late-projected transverse mass of parent particle a , where a can refer to either the Higgs boson or W boson,

$$\mathcal{M}_{aT} = \sqrt{g_{\alpha\beta}(\mathbf{p}_{aT}^\alpha + \mathbf{q}_{at}^\alpha)(\mathbf{p}_{aT}^\beta + \mathbf{q}_{at}^\beta)} \quad (6.5)$$

and $g_{\alpha\beta}$ is the Minkowski metric ($\text{diag}(1, -1, -1, -1)$), $\mathbf{p}_{aT}$ is the visible transverse energy-momentum vector, and $\mathbf{q}_{at}$ is the invisible transverse energy-momentum vector [41].

Generally, you then choose the largest parent mass and minimize it by considering all possible combi-

nations of invisible momenta. For our purposes, that is the Higgs boson mass. We require that the sum of the invisible transverse momenta equals the missing transverse momenta of the event ($\sum_i \mathbf{q}_{iT} = \mathbf{p}_T^{miss}$). In addition, the sum of the invariant mass of the lepton and neutrino associated to the W boson must be constrained to be near the W mass (80.4 GeV).

M_{2T} is then calculated by:

$$M_{2T} = \min_{\sum \bar{q}_{iT} = \bar{p}_T^{miss}} [\max_a [\mathcal{M}_{aT}]], \quad (6.6)$$

where a refers the Higgs boson mass. The minimization is done in a nine-dimensional space (three neutrinos with three missing momentum components) for $WH, H \rightarrow \tau_h \tau_h$. As a reminder, we initially have a 12-dimensional space for $WH, H \rightarrow \tau_l \tau_h$ with the addition of the extra neutrino from the leptonic tau decay. However, we reduce the dimensionality of the problem by requiring that the neutrino momentum is collinear with visible tau decay products.

6.7 BACKGROUNDS

We divide our backgrounds into two categories which we name “reducible” and “irreducible”. *Reducible* backgrounds are those that we can easily separate out with a series of selection cuts. The reducible backgrounds do not contain the final state objects we are looking for, so they enter our signal due to misidentification. This misidentification typically occurs when we mistakenly tag quark- or gluon-initiated jets as electrons, muons, or taus. These backgrounds are difficult to accurately model in the simulation so we therefore estimate the backgrounds with a data-driven technique described in Chapter 7.

The *irreducible* background is made up of physics processes with the same final state objects as our signal categories. These are difficult to disentangle from our signal because the objects are all correctly reconstructed and identified. The primary contributors to the irreducible background are diboson events such as WZ and ZZ , when the Z decays to two tau leptons, but there are also events originating from top decays and other Higgs boson decays. We use the Monte Carlo simulation to estimate the contribution from these backgrounds. For Run 2, we trained a dedicated neural network to separate out the diboson backgrounds (see section 8.3).

“Leisure, the problem with the modern condition was the dearth of leisure. And finally, it took a force of nature to interrupt our routines. We just wanted to hit the reset button. We just wanted to feel flush with time to do things of no quantifiable value, our hopeful side pursuits like writing or drawing or something, something other than what we did for money.”

Ling Ma, *Severance*

7

Fake Rate Calculations

7.1 INTRODUCTION TO THE FAKE FACTOR METHOD

IN ORDER TO ESTIMATE OUR BACKGROUND FROM MISIDENTIFIED OBJECTS WE EMPLOY A DATA-DRIVEN METHOD KNOWN AS THE FAKE FACTOR, WHERE “FAKE” REFERS TO OBJECT MISIDENTIFICATION. The rate of misidentification, known as the “fake rate”, is calculated in a region distinct from, but closely resembling, the signal region. These fakes can be mis-identified jets, photon conversions, and other mis-reconstructed objects in the detector. The fake rate calculation measures how often background objects can enter into our signal region. To determine the fake rate, we take a sample of background objects and apply the same identification and isolation criteria we use in our signal selection. The denominator for the fake rate calculation is all objects that are candidates for an identification (or isolation) algorithm. The numerator is the subset of those objects that pass identification (and isolation). From these fake rates (FR), we get a fake factor (f):

$$f = \frac{FR}{1 - FR} \quad (7.1)$$

which is a ratio of background objects that pass identification requirements to background objects that fail identification requirements. This fake factor is what we use as an extrapolation factor to scale events in a region adjacent to our signal region to estimate the fake contribution in our signal region as described below.

The fake factor method is complicated in our analysis because we have tau, muon, and electron identifications that can be faked, and up to three (four) particles that can be misidentified in our WH (ZH) signal regions.

7.1.1 FAKE FACTOR METHOD DERIVATION

To derive the fake factor method, we first need to introduce some useful notation. T stands for true object (an actual electron, muon, or tau), which will have an efficiency ϵ of being successfully reconstructed and passing selection criteria (identification and/or isolation). F stands for fake object (the mis-reconstructed jets), that will have an efficiency FR (fake rate) of passing selection criteria. Objects that pass selection will be denoted with S and those that pass the fake criteria are denoted with A . We can then state the the number of selected objects (N_S) is:

$$N_S = N_{SF} + N_{ST} = FR * N_F + \epsilon N_T \quad (7.2)$$

While this is a data-driven method, we take into account MC-based estimates of contamination from other physics processes in our various measurement regions. This means that we will subtract the true MC that ends up in our anti-enriched region. Taking this into account, and relying on the relationship between fake rates FR and the fake factor f , the number of selected objects can also be written as:

$$N_S = f * N_A^{data} - f * N_{AT}^{MC} + N_{ST}^{MC} \quad (7.3)$$

To get the estimate of the number of objects that make up our fake background, we need to think about the ways objects can enter our signal region. To start with, let's consider some of the possible combinations of events with fake objects. Those combinations include:

- $F_1 \rightarrow S_1^F$ if the fake object passes selection or A_1^F if the fake object fails selection

- $F_1 F_2 \rightarrow S_1^F S_2^F$ if two objects both pass selection requirements or $S_1^F A_2^F$ or $A_1^F S_2^F$ if one object passes and the other fails or $A_1^F A_2^F$ if both fail selection requirements
- $T_1 F_2 F_3 \rightarrow S_1^T S_2^F A_3^F$ or $S_1^T A_2^F S_3^F$ where the true object is selected and one of the fake objects passes selection or $A_1^T S_2^F S_3^F$ where the true object does not pass selection and the fake objects do or $S_1^T S_2^F S_3^F$ where all three are selected or $A_1^T S_2^F A_3^F$ or $A_1^T A_2^F S_3^F$ where the true object does not pass selection and one of the fake objects is selected or $S_1^T A_2^F A_3^F$ where the true object is selected and neither of the fake objects pass or $A_1^T A_2^F A_3^F$ where all three objects fail selection requirements
- etc.

As can be seen from the list above, if we have one fake object it may either pass (S) or fail (A) those identification and/or isolation criteria. When we have two fake objects, they can either both pass or fail, or have one pass and the other fail. Further combinations can be constructed, expanding in the way demonstrated above.

The number of objects considered depends on the number of selected objects needed for a given analysis channel. Now the total number of event combinations comes from applying the probability of selecting (FR_i) and/or anti-selecting ($1 - FR_i$) the true and fake objects and is:

- $N_{S_1^F} = FR_1 * N_{F_1}$
- $N_{S_1^F A_2^F} = FR_1 * (1 - FR_2) N_{F_1 F_2}$
- $N_{A_1^F S_2^F} = FR_2 * (1 - FR_1) N_{F_1 F_2}$
- etc.

We need to solve for N_{F_1} , $N_{F_1 F_2}$, which we know also appear in the events in the fake enriched region:

- $N_{A_1^F} = (1 - FR_1) * N_{F_1}$
- $N_{A_1^F A_2^F} = (1 - FR_1) * (1 - FR_2) N_{F_1 F_2}$
- etc.

We can invert these expressions and use $f = \frac{FR}{1-FR}$ to see that:

- $N_{\bar{S}_1} = f_1 * N_{A_1}$
- $N_{\bar{S}_1 A_2} = f_1 * N_{A_1 A_2}$
- $N_{A_1 \bar{S}_2} = f_2 * N_{A_1 A_2}$

- etc.

where $N_{\bar{S}}$ is the number of objects that enter the signal region from the fake factor extrapolation. The order of these objects does matter as the fake rate depends on object kinematics and the specific type of object.

One Object Case

While our analysis channels require three or four identified objects for the WH and ZH channels respectively, we will carry out the derivation of the method on the one-object case first for simplicity. If we are considering the situation where there is only one selected object in the final state to first order, this enters as $F_1 \rightarrow S_1^F$ or A_1^F . We can get the number of selected events from: $N_{S_1^F} = FR_1 * N_{F_1}$ and $N_{A_1^F} = (1 - FR_1) * N_{F_1}$. We can solve for N_{F_1} from the second equation: $N_{F_1} = \frac{N_{A_1^F}}{(1 - FR_1)}$. Plugging that into the equation for $N_{S_1^F}$ and using $f = \frac{FR}{1 - FR}$, we get the result that:

$$N_{\bar{S}} = f * N_A \quad (7.4)$$

where N_A is number of anti-selected objects, and $N_{\bar{S}}$ is the number of objects that enter the signal region from the fake factor extrapolation. This equation is a simplified expression for a possibly infinite sum of combinations from which one object is selected for the event.

Two Object Case

We can also derive the two object case. Once again, we will think about the type of events that can enter into the signal region via the fake factor method. In the two object case, at least one object must be a fake (otherwise it would be a signal event). There can be many more than two fake candidates, but no more than two selected objects. In this way, a single event can enter our fake background in more than one way with a different combination of fakes. Possible ways to get objects from fakes and their probabilities are shown in table 7.1. From these we can determine the number of events entering the signal region. These combinations gets more complicated as more terms are added, however, the individual probability for each term to be selected decreases.

The possible first-order two object cases are F_1T , TF_2 , and F_1F_2 . The number of each possible

Event Type	Reconstruction	Probability
$F_1 T$	SS, AS	FR_1
$T F_2$	SS, SA	FR_2
$F_1 F_2$	SS, SA, AS, AA	$FR_1 * FR_2$
...	...	...

Table 7.1: Event types (using truth-level information), with their various possible reconstructions, and their probability to enter signal region. “F” designates a fake object, “T” designates a true object. “S” represents an object (which could be a fake or true object) that passes the selection criteria. “A” represents an object that fails the selection criteria. The right-hand column provides the probability for each event type. The numerical subscripts keep track of the individual objects that are considered.

reconstruction (with the A and S) is calculated by summing the times it appears:

$$\begin{aligned}
N_{SS} &= FR_1 * N_{F_1 T} + FR_2 * N_{T F_2} + FR_1 * FR_2 * N_{F_1 F_2} \\
N_{AS} &= (1 - FR_1) * N_{F_1 T} + FR_2 * (1 - FR_1) * N_{F_1 F_2} \\
N_{SA} &= (1 - FR_2) * N_{T F_2} + FR_1 * (1 - FR_2) * N_{F_1 F_2} \\
N_{AA} &= (1 - FR_1)(1 - FR_2) * N_{F_1 F_2}
\end{aligned} \tag{7.5}$$

The terms that enter into our signal region are those with two selected objects (N_{SS}). We can then solve for the necessary N_{ij} terms with the other equations. We can invert the final equation and get:

$$N_{F_1 F_2} = \frac{N_{AA}}{(1 - FR_1)(1 - FR_2)} \tag{7.6}$$

and we can plug that in to the middle two equations and get:

$$\begin{aligned}
N_{F_1 T} &= \frac{N_{AS}}{(1 - FR_1)} - \frac{FR_2 * N_{AA}}{(1 - FR_1)(1 - FR_2)} \\
N_{T F_2} &= \frac{N_{SA}}{(1 - FR_2)} - \frac{FR_1 * N_{AA}}{(1 - FR_1)(1 - FR_2)}
\end{aligned} \tag{7.7}$$

and then you can plug in $f_i = \frac{FR_i}{1 - FR_i}$ and reduce to:

$$\begin{aligned}
N_{F_1 T} &= \frac{N_{AS}}{(1 - FR_1)} - \frac{f_2 * N_{AA}}{1 - FR_1} \\
N_{T F_2} &= \frac{N_{SA}}{(1 - FR_2)} - \frac{f_1 * N_{AA}}{1 - FR_2}
\end{aligned} \tag{7.8}$$

From here we plug in equations 7.8 into the first line of equation 7.5, and use our definition of the fake factor f and get:

$$N_{SS} = \frac{FR_1 * N_{AS}}{(1 - FR_1)} + \frac{FR_2 * N_{SA}}{(1 - FR_2)} + \frac{FR_1 FR_2 * N_{AA}}{(1 - FR_1)(1 - FR_2)} - \frac{f_2 FR_1 * N_{AA}}{1 - FR_1} - \frac{f_1 FR_2 * N_{AA}}{1 - FR_2}$$

$$N_{SS} = f_1 N_{AS} + f_2 N_{SA} - f_1 f_2 N_{AA}$$
(7.9)

where the second line is our two object fake factor.

Three and Four Object Case

For our $VH, H \rightarrow \tau\tau$ analysis, all of our final states have three or four final state objects. The derivation for these terms continues in much the same way and results in:

$$N_{SSS} = f_1 N_{ASS} + f_2 N_{SAS} + f_3 N_{SSA}$$

$$- f_1 f_2 N_{AAS} - f_2 f_3 N_{SAA} - f_1 f_3 N_{ASA}$$

$$+ f_1 f_2 f_3 N_{AAA}$$
(7.10)

for the three object case and:

$$N_{SSSS} = f_1 N_{ASSS} + f_2 N_{SASS} + f_3 N_{SSAS} + f_4 N_{SSSA}$$

$$- f_1 f_2 N_{AASS} - f_2 f_3 N_{SAAS} - f_1 f_3 N_{ASAS} - f_2 f_4 N_{SASA} - f_1 f_4 N_{ASSA} - f_3 f_4 N_{SSAA}$$

$$+ f_1 f_2 f_3 N_{AAAS} + f_2 f_3 f_4 N_{SAAA} + f_1 f_3 f_4 N_{ASAA} + f_1 f_2 f_4 N_{AASA}$$

$$- f_1 f_2 f_3 f_4 N_{AAAA}$$
(7.11)

Hopefully you can see the repeating pattern where the sign flips as we add more antiselected objects.

7.2 Z + JETS SELECTION

We use two regions that are orthogonal to our signal region, known as fake factor measurement regions, to calculate our fake rates. These are $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$. These regions have similar kinematic properties

to and fake composition as our signal region. ATLAS has good muon and electron reconstruction and the ability to reconstruct the Z boson mass from reconstructed objects gives us a high-purity selection of these events. Given the cross sections of relevant physics processes, any extra identified object (electron, muon, or tau) in the event is likely a mis-reconstructed jet. We run these selections over both data and MC, though only the data and truth matched MC is used for the fake rate calculation. The rest of the MC is used to understand the purity of our $Z + jets$ selection and in our closure tests.

7.2.1 $Z \rightarrow \mu\mu$ SELECTION

The selection for the $Z \rightarrow \mu\mu + jets$ events is designed to select those events while also providing events that are kinematically similar to those selected by our analysis signal requirements. First, all medium identified muons with a $p_T > 9\text{GeV}$ and an $|\eta|$ within the fiducial region are added to the candidate list for selected and anti-selected muons. Then, all events with a pair of opposite sign muons with a dilepton mass within 10GeV of the Z mass (91GeV) are kept. If an event has more than one pair of muons meeting this criteria, the pair with the mass closest to 91GeV is chosen as the selected muon pair. The extra muon(s) in the event may be candidates for fake objects. Either the single- or di-muon trigger must have been activated. The single muon trigger p_T threshold is 26GeV , and our offline single muon threshold set 2GeV higher, at $p_T > 28\text{GeV}$. The di-muon p_T threshold is symmetric, at 14GeV for both muon candidates, and our offline di-muon p_T threshold is set at $p_T > 16\text{GeV}$. Next, both selected muons must pass a p_T -dependent tight isolation requirement to ensure that they are not objects within jets.

To illustrate our modeling of these events, the Z mass peak for 2017 data and ATLAS Monte Carlo simulation are shown in Figure 7.1. It is clear that these distributions consist of $Z \rightarrow \mu\mu + jets$ events.

7.2.2 $Z \rightarrow ee$ SELECTION

The selection for the $Z \rightarrow ee + jets$ events parallels the cuts made in the $Z \rightarrow \mu\mu$ selection. First, all medium identified electrons with a $p_T > 13\text{GeV}$, an $|\eta|$ within the fiducial region and that are not in the crack region are added to the candidate list. Then, all events with a pair of opposite-sign electrons with a dilepton mass within 10GeV of the Z peak (91GeV) are kept. If an event has more than one pair of electrons meeting this criteria, the pair with the reconstructed mass closest to 91GeV is chosen as the selected electrons. Extra electron(s) are kept for calculation in the electron fake rate. Either the single-

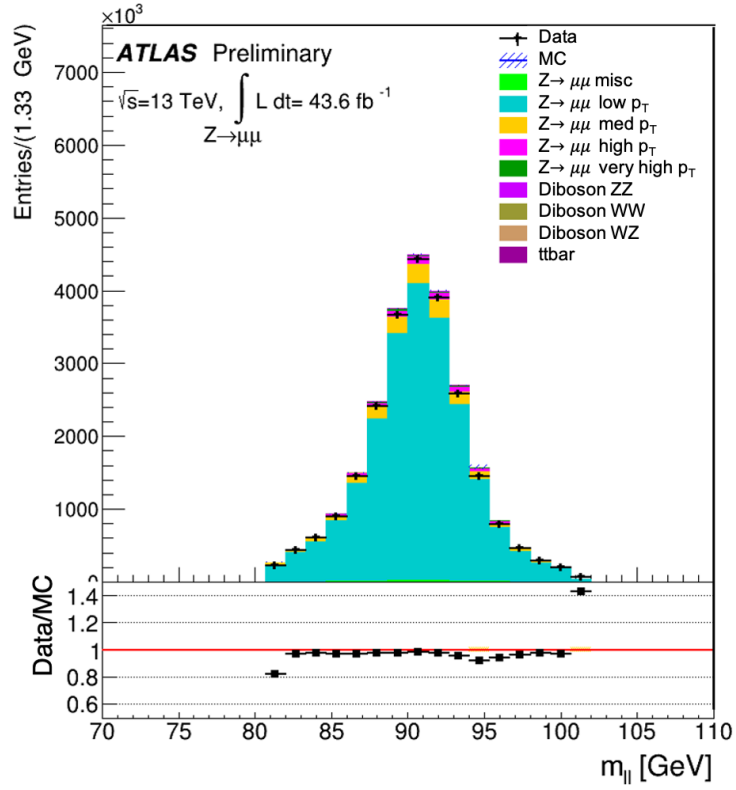


Figure 7.1: Reconstructed Z mass peak for 2017 data and ATLAS Monte Carlo simulation for the $Z \rightarrow \mu\mu$ region.

or di-electron trigger must have been activated. The single-electron trigger p_T threshold is 26GeV, and our offline single-electron threshold set 2GeV higher, at $p_T > 28\text{GeV}$. The di-electron p_T threshold is symmetric, at 17GeV for both electron candidates, and our offline di-electron p_T threshold is set at $p_T > 19\text{GeV}$. Both electrons must pass calorimeter-based and track-based isolation requirements.

7.2.3 DIBOSON SUBTRACTION

Diboson events, such as WZ and ZZ , can contaminate our fake rate calculations with true leptons unless we take additional steps to target these events. This diboson contamination in the fake rates is reduced in two main ways:

- **Targeted Selection Cuts:** Two cuts were added to the fake lepton calculation for muons and electrons to remove WZ background. The additional cuts require that the missing transverse energy is greater than 60GeV and the transverse mass between the third lepton and the missing transverse energy is less than 40GeV.

- **Monte Carlo Subtraction:** The requirements for selected and anti-selected objects as detailed below are also applied to MC simulation of relevant physics processes. We can therefore use the simulation to estimate the surviving contamination from diboson events, and subtract that estimate before calculating the fake rates.

7.3 FAKE RATE MEASUREMENTS

Critical inputs to the data-driven background estimation method are the fake rates for electrons, muons and taus. These measurements are done in $Z \rightarrow \mu\mu + \text{jets}$ events and $Z \rightarrow ee + \text{jets}$ events, where these jets are candidates for faking selected electrons, muons and taus. The $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ initial selections ensure that the fake rates are measured in events that are similar to those that are selected by our signal requirements.

7.3.1 ELECTRON FAKE RATE

Objects mis-identified as electrons can come from a variety of sources including γ conversions, light meson decays, and mis-identified jets. The fake rates for electrons are calculated in regions enriched in $Z \rightarrow ll + \text{jets}$ as described above. Any reconstructed electron with a $p_T > 13\text{GeV}$, not in the η crack region, and passing loose identification will be counted as a candidate electron. A “selected electron” will additionally pass both tight identification and isolation requirements. As noted above, the truth matched Monte Carlo component is subtracted from both the selected and candidate counts. The fake rate (FR) is therefore defined as

$$FR_e = \frac{N_{\text{SelectedElectrons}} - N_{\text{TruthMatchedSelected}}}{N_{\text{CandidateElectrons}} - N_{\text{TruthMatchedCandidate}}}$$

From FR_e the probabilities for a fake electron to pass selection can be found, these are denoted:

$$f_e = \frac{FR_e}{1 + FR_e} = \frac{\text{SelectedElectrons}}{\text{Anti} - \text{SelectedElectrons}}$$

where Anti-Selected electrons pass loose identification and $p_T > 13\text{GeV}$, but fail Tight identification and/or Isolation. The electron fake rates are calculated in three regions of $|\eta|$ and three regions of p_T . The

$|\eta|$ bins are divided into $0 < |\eta| \leq 0.8$, $0.8 < |\eta| \leq 1.37$, and $|\eta| > 1.37$. The three p_T regions are $13 - 15\text{GeV}$, $15 - 20\text{GeV}$, and $> 20\text{GeV}$. While the fake rates were calculated individually for the four years of data taking, they were essentially consistent within statistical uncertainties, so results from the four years were combined for the final measurement. Figure 7.2 shows the fake rates for each year binned by p_T (left) and $|\eta|$ (right).

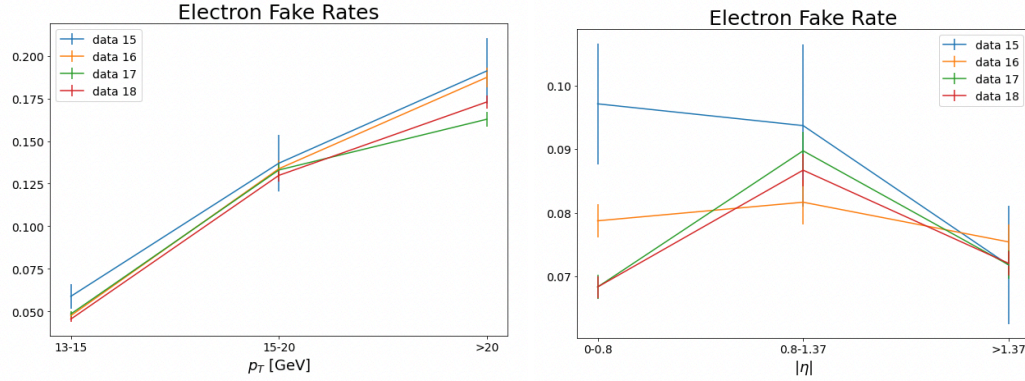


Figure 7.2: Fake rate values for all four years of data in bins of p_T (left) and $|\eta|$ (right) and their statistical uncertainties.

The electron fake rates after Monte Carlo subtraction for all four years combined are shown in 7.3.

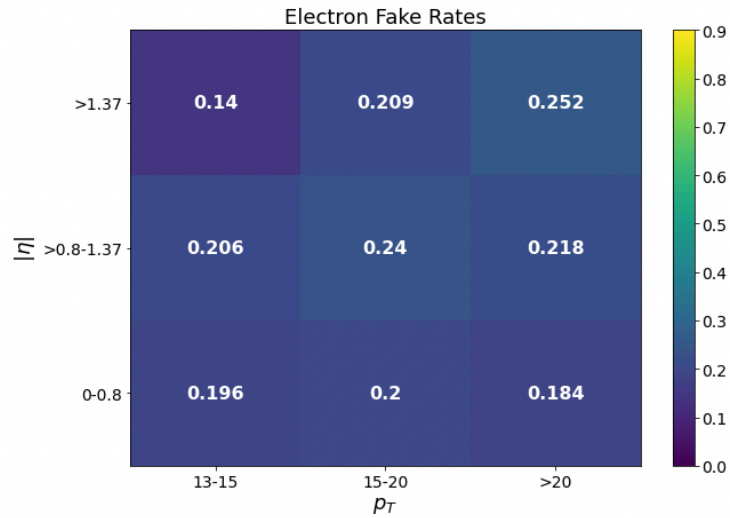


Figure 7.3: Fake Rates for electrons incorporating the MC truth subtraction from the data counts for selected and candidate electrons.

The statistical uncertainty is calculated for the electron fake rates (and the muon and tau fake rates as

	$13 \leq p_T \leq 15$	$15 \leq p_T \leq 20$	$p_T \geq 25$
$0 \leq \eta \leq 0.8$	$0.1962 \pm 0.0042 \pm 0.0009$	$0.2002 \pm 0.0042 \pm 0.0012$	$0.1844 \pm 0.0045 \pm 0.0125$
$0.8 \leq \eta \leq 1.37$	$0.2065 \pm 0.0051 \pm 0.0018$	$0.2405 \pm 0.0058 \pm 0.0048$	$0.2178 \pm 0.0061 \pm 0.0079$
$ \eta \geq 1.37$	$0.1401 \pm 0.0033 \pm 0.0015$	$0.2093 \pm 0.0047 \pm 0.0066$	$0.2517 \pm 0.006 \pm 0.0116$

Table 7.2: Electron fake rates for each bin in p_T and $|\eta|$ with their statistical and systematic uncertainties ($FR \pm \sigma_{stat} \pm \sigma_{syst}$).

well) by:

$$\sigma_{stat} = \frac{1}{N_{all}} \sqrt{\left(1 - \frac{N_{sel}}{N_{all}}\right) N_{sel}} \quad (7.12)$$

where N_{all} is the count of electron candidates in data minus those in MC and N_{sel} is the count of selected electrons in data minus those in MC.

The systematic uncertainty comes from varying the MC contribution to the calculated fake rates and is defined as:

$$\sigma_{syst} = |FR_{up} - FR_{down}| \quad (7.13)$$

where

$$FR_{up} = \frac{N_{sel}^{data} - (N_{sel}^{MC} + \sqrt{MC_{weights}^{sel}})}{N_{all}^{data} - (N_{all}^{MC} - \sqrt{MC_{weights}^{all}})} \quad (7.14)$$

and

$$FR_{down} = \frac{N_{sel}^{data} - (N_{sel}^{MC} - \sqrt{MC_{weights}^{sel}})}{N_{all}^{data} - (N_{all}^{MC} + \sqrt{MC_{weights}^{all}})} \quad (7.15)$$

7.3.2 MUON FAKE RATE

All reconstructed muons with $p_T > 9\text{GeV}$ and that pass loose identification are considered candidate muons. The fake rate for muons is calculated in regions enriched in $Z \rightarrow ll + jets$ as described above.

The fake rate after Monte Carlo subtraction is defined as

$$FR_{\mu} = \frac{N_{Iso-TightMuons} - N_{Iso-TightTruthMatched}}{N_{LooseMuons} - N_{TruthMatchedLoose}}$$

Selected muons pass tight identification and isolation requirements, whereas candidate muons just need to pass loose identification and can fail one or both of the selected criteria. From FR_{μ} the probabilities

for a fake muon to pass selection can be found, these are denoted:

$$f_{\mu} = \frac{FR_{\mu}}{1 + FR_{\mu}} = \frac{SelectedMuons}{Anti - SelectedMuons}$$

Anti-Selected muons pass Loose identification but fail Tight identification and/or isolation. The Fake Rates are calculated in three regions of $|\eta|$ and four regions of p_T . The $|\eta|$ bins are divided into bins of $0 < |\eta| \leq 0.8$, $0.8 < |\eta| \leq 1.37$, and $|\eta| > 1.37$. The p_T ranges are $9 - 10\text{GeV}$, $10 - 15\text{GeV}$, $15 - 20\text{GeV}$, and $> 20\text{GeV}$.

Similarly to electrons, the fake rates were initially calculated for the four years of data taking, but with the lower statistics in each year individually, especially in 2015, we combined them into a single fake rate. Figure 7.4 shows the fake rates for each year binned by p_T (left) and $|\eta|$ (right). Again, they align reasonably well within just statistical uncertainties, showing that we can safely combine them into one fake rate for all four years.

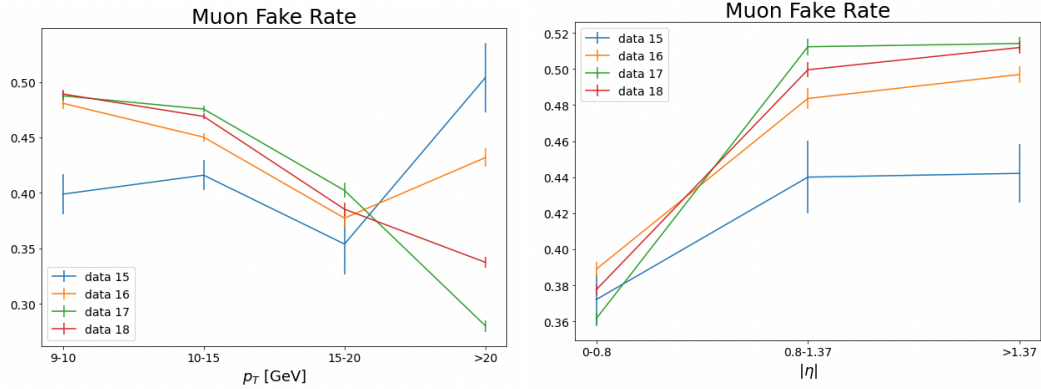


Figure 7.4: Fake rate values for muons for all four years of data in bins of p_T (left) and $|\eta|$ (right) and their statistical uncertainties.

The muon fake rates after Monte Carlo subtraction for all four years combined are shown in figure 7.5. The fake rates and their corresponding statistical and systematic uncertainties are shown in table 7.3.

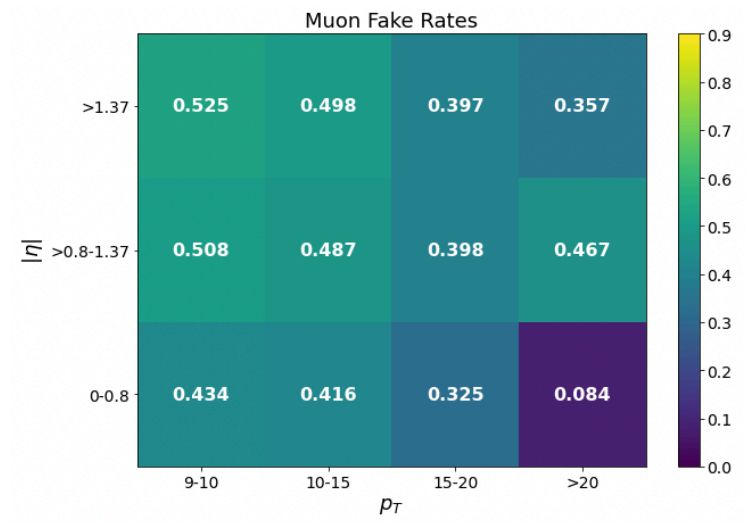


Figure 7.5: Fake rates for muons incorporating the MC truth subtraction from the data counts for selected and all candidate muons.

	$9 \leq p_T \leq 10$	$10 \leq p_T \leq 15$	$15 \leq p_T \leq 20$	$p_T \geq 25$
$0 \leq \eta \leq 0.8$	$0.4338 \pm 0.0037 \pm 0.0010$	$0.4163 \pm 0.0028 \pm 0.0014$	$0.3248 \pm 0.0058 \pm 0.0087$	$0.0838 \pm 0.0026 \pm 0.0108$
$0.8 \leq \eta \leq 1.37$	$0.5081 \pm 0.0049 \pm 0.0012$	$0.4868 \pm 0.0038 \pm 0.0024$	$0.3977 \pm 0.0084 \pm 0.0109$	$0.4669 \pm 0.0119 \pm 0.0554$
$ \eta \geq 1.37$	$0.5251 \pm 0.004 \pm 0.001$	$0.4982 \pm 0.0031 \pm 0.0021$	$0.3973 \pm 0.0068 \pm 0.0103$	$0.3568 \pm 0.0101 \pm 0.0686$

Table 7.3: Muon fake rates for each bin in p_T and $|\eta|$ with their statistical and systematic uncertainties ($\text{FR} \pm \sigma_{\text{stat}} \pm \sigma_{\text{syst}}$)

7.3.3 TAU FAKE RATE

The region in data in which the tau fake rates are calculated obeys the same kinematic selections of the $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ fake regions plus an additional requirement of at least one reconstructed tau with a transformed RNN Jet Score > 0.01 . To be a candidate tau, the reconstructed tau must additionally have a $p_T > 20\text{GeV}$, a charge of $\pm 1e$, 1- or 3-prongs, and not be in the Eta Crack Region. All taus passing this series of cuts are referred to as candidate taus.

From this point, all candidate taus are further classified as either *selected* (pass the medium tau identification working point) or *anti-selected* (do not pass the medium tau identification working point). We bin the tau fake rate based on p_T and jet-width, which is a function of the spread of the tau-jet in its associated calorimeter Topocluster, with higher values meaning a wider jet. These selected and anti-selected taus are then divided into six regions for jet-width ($0 - 0.05$, $0.05 - 0.07$, $0.07 - 0.09$, $0.09 - 0.11$, $0.11 - 0.15$, and > 0.15) and six regions in p_T ($20 - 25\text{GeV}$, $25 - 30\text{GeV}$, $30 - 35\text{GeV}$, $35 - 40\text{GeV}$, $40 - 60\text{GeV}$, and $> 60\text{GeV}$) for 1-prong taus and the same six regions in jet-width and just four regions in p_T ($20 - 25\text{GeV}$, $25 - 30\text{GeV}$, $30 - 40\text{GeV}$, and $> 40\text{GeV}$) for 3-prong taus.

In addition to the 1- and 3-prong split for tau fake rates, there is also a division based on the MET (missing transverse energy) in the event. The fake rates are further divided into regions with varying amount of MET: $\text{MET} < 20\text{GeV}$ or $\text{MET} > 20\text{GeV}$. Originally we split on $|\eta|$ and not on jet-width but found that the fake rates depended more on jet-width and thus achieved a better agreement between Monte Carlo and data when we parametrized our fake rates by jet-width. To include the fact that there is some $|\eta|$ dependence in the fake rates, we also separate the fake rate calculations into barrel ($|\eta| < 1.37$) and end-cap ($|\eta| > 1.37$) regions.

The equations to calculate these are the same as for muons and electrons:

$$FR_\tau = \frac{N_{\text{MediumTaus}} - N_{\text{Truth-MatchedMediumTaus}}}{N_{\text{TauCandidates}} - N_{\text{TruthMatchedTaus}}}$$

The fake factor to be applied to data is then calculated as:

$$f_\tau = \frac{FR_\tau}{1 + FR_\tau} = \frac{\text{SelectedTaus}}{\text{Anti} - \text{SelectedTaus}}$$

Figure 7.6 shows the validity of grouping the data from all four years into a single calculation for either 1- or 3-prong taus (regardless of their $|\eta|$ value) for each of our MET groupings, $< 20\text{GeV}$ and $> 20\text{GeV}$. A main goal of combining data from multiple years is to reduce the statistical uncertainty in our measurement.

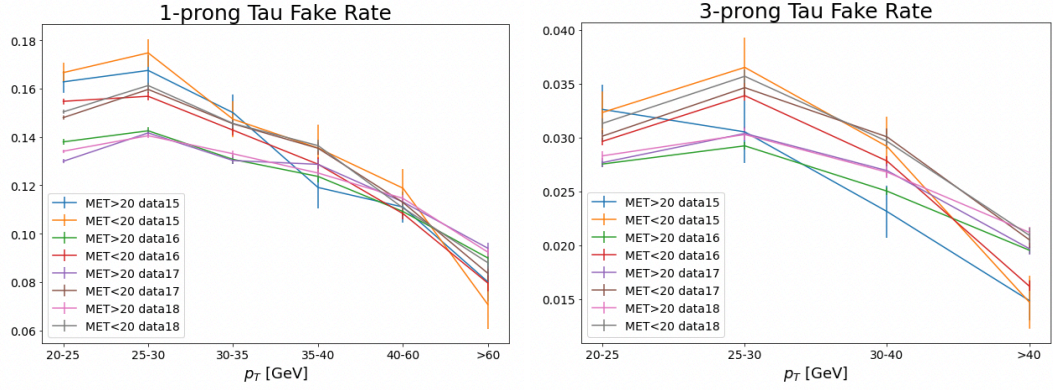


Figure 7.6: Fake rate values for taus for all four years of data in bins of p_T for 1-prong taus (left) and 3-prong taus (right) and their statistical uncertainties.

As can be seen in figure 7.7, there is not a strong relationship between the tau fake rate and the tau $|\eta|$. However, there is a difference larger than the statistical error for the fake rates calculated in the end-cap and barrel, validating our split in $|\eta|$ for fake rate calculation.

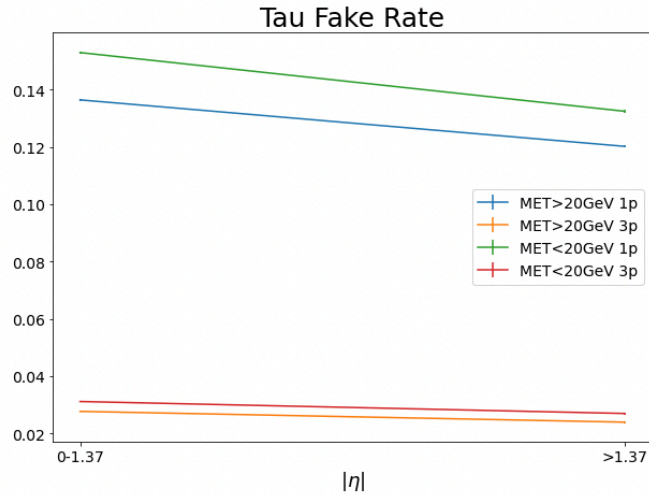


Figure 7.7: Fake rate values for taus for all four years of data combined in bins the two $|\eta|$ bins for 1-prong and 3-prong taus and their statistical uncertainties (very small on this plot).

The final tau fake rates after Monte Carlo subtraction (without systematic and statistical uncertainties)

are shown in figure 7.8 for the barrel portion ($|\eta| < 1.37$) and in figure 7.9 for the end-cap portion of the detector ($|\eta| > 1.37$). The tau fake rates with both statistical and systematic uncertainties for just the barrel region, as a quantitative example, are shown in tables 7.4 - 7.7.

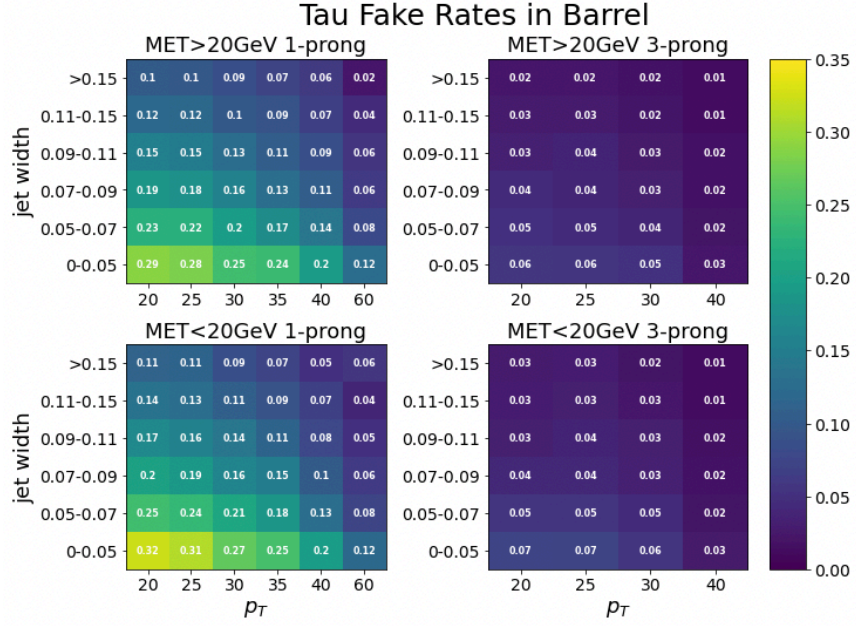


Figure 7.8: Fake rate values in the barrel region ($|\eta| < 1.37$) for taus from all four years of data combined in bins for jet-width and p_T of the selected/anti-selected tau candidate.

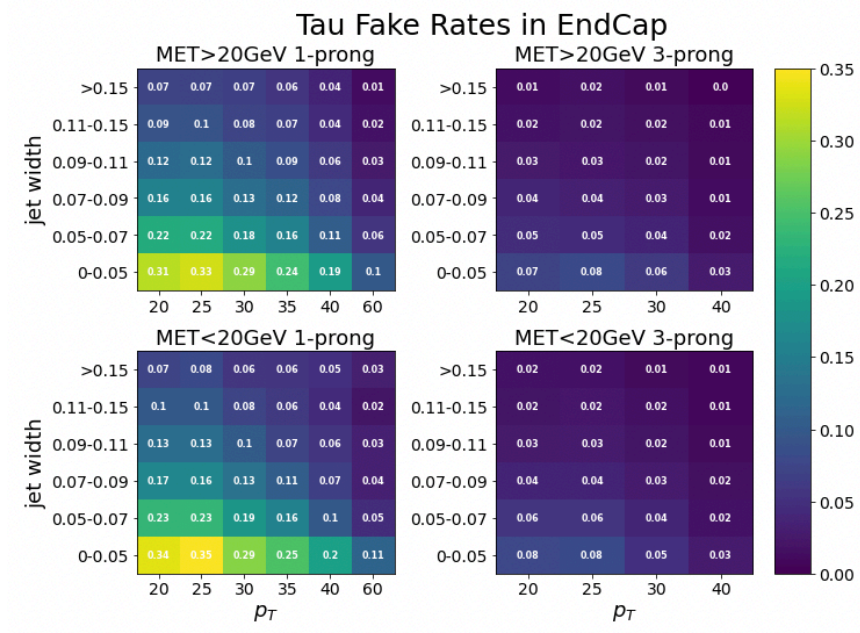


Figure 7.9: Fake rate values in the end-cap region ($|\eta| > 1.37$) for taus from all four years of data combined in bins for jet-width and p_T of the selected/anti-selected tau candidate.

Central Region, MET > 20	$20 \leq p_T \leq 25$	$25 \leq p_T \leq 30$	$30 < p_T \leq 35$	$35 < p_T \leq 40$	$40 < p_T \leq 60$	$p_T > 60$
$0 \leq jet - width \leq 0.05$	$0.2887 \pm 0.0049 \pm 0.0003$	$0.2824 \pm 0.0038 \pm 0.0003$	$0.2436 \pm 0.0063 \pm 0.0006$	$0.2419 \pm 0.0069 \pm 0.0007$	$0.1964 \pm 0.0037 \pm 0.0005$	$0.1249 \pm 0.0027 \pm 0.0008$
$0.05 \leq jet - width \leq 0.07$	$0.2267 \pm 0.0032 \pm 0.0002$	$0.2235 \pm 0.0038 \pm 0.0002$	$0.1981 \pm 0.0040 \pm 0.0003$	$0.1727 \pm 0.0043 \pm 0.0004$	$0.1386 \pm 0.0026 \pm 0.0003$	$0.0802 \pm 0.0026 \pm 0.0003$
$0.07 \leq jet - width \leq 0.09$	$0.1888 \pm 0.0022 \pm 0.0001$	$0.1775 \pm 0.0026 \pm 0.0002$	$0.1559 \pm 0.0029 \pm 0.0002$	$0.130 \pm 0.0032 \pm 0.0002$	$0.1077 \pm 0.0022 \pm 0.0002$	$0.0618 \pm 0.0029 \pm 0.0003$
$0.09 \leq jet - width \leq 0.11$	$0.1543 \pm 0.0016 \pm 0.00007$	$0.1464 \pm 0.0026 \pm 0.0001$	$0.1251 \pm 0.0024 \pm 0.0002$	$0.1088 \pm 0.0029 \pm 0.0002$	$0.0885 \pm 0.0023 \pm 0.0002$	$0.0555 \pm 0.0041 \pm 0.0004$
$0.11 \leq jet - width \leq 0.15$	$0.1222 \pm 0.0010 \pm 0.00004$	$0.1177 \pm 0.0014 \pm 0.00009$	$0.0993 \pm 0.0019 \pm 0.0001$	$0.0935 \pm 0.0026 \pm 0.0002$	$0.0696 \pm 0.0022 \pm 0.0002$	$0.0433 \pm 0.0051 \pm 0.0005$
$jet - width > 0.15$	$0.1031 \pm 0.0011 \pm 0.00004$	$0.0981 \pm 0.0018 \pm 0.0001$	$0.0867 \pm 0.0018 \pm 0.0001$	$0.0736 \pm 0.0043 \pm 0.0003$	$0.0603 \pm 0.0046 \pm 0.0003$	$0.0243 \pm 0.0097 \pm 0.0011$

Table 7.4: Tau fake rates for each bin in p_T and jet-width with their statistical and systematic uncertainties ($FR \pm \sigma_{stat} \pm \sigma_{syst}$ for 1-prong taus in the barrel region, with MET > 20 GeV).

Central Region, MET < 20	$20 \leq p_T \leq 25$	$25 \leq p_T \leq 30$	$30 < p_T \leq 35$	$35 < p_T \leq 40$	$40 < p_T \leq 60$	$p_T > 60$
$0 \leq jet - width \leq 0.05$	$0.3228 \pm 0.0045 \pm 0.00009$	$0.3107 \pm 0.0035 \pm 0.0002$	$0.2660 \pm 0.0063 \pm 0.0002$	$0.2496 \pm 0.0071 \pm 0.0004$	$0.1977 \pm 0.0042 \pm 0.0003$	$0.1206 \pm 0.0037 \pm 0.0003$
$0.05 \leq jet - width \leq 0.07$	$0.2457 \pm 0.0031 \pm 0.00007$	$0.2361 \pm 0.0038 \pm 0.0001$	$0.2121 \pm 0.0050 \pm 0.0002$	$0.1833 \pm 0.0050 \pm 0.0002$	$0.1311 \pm 0.0031 \pm 0.0002$	$0.0757 \pm 0.0035 \pm 0.0002$
$0.07 \leq jet - width \leq 0.09$	$0.2013 \pm 0.0022 \pm 0.00004$	$0.1907 \pm 0.0027 \pm 0.00006$	$0.1607 \pm 0.0032 \pm 0.0001$	$0.1486 \pm 0.0039 \pm 0.0001$	$0.1036 \pm 0.0027 \pm 0.0002$	$0.0628 \pm 0.0041 \pm 0.0003$
$0.09 \leq jet - width \leq 0.11$	$0.1692 \pm 0.0017 \pm 0.00003$	$0.1584 \pm 0.0023 \pm 0.00005$	$0.1383 \pm 0.0029 \pm 0.00009$	$0.1133 \pm 0.0029 \pm 0.00009$	$0.0767 \pm 0.0028 \pm 0.0001$	$0.0510 \pm 0.0054 \pm 0.0003$
$0.11 \leq jet - width \leq 0.15$	$0.1355 \pm 0.0011 \pm 0.00002$	$0.1309 \pm 0.0017 \pm 0.00003$	$0.1098 \pm 0.0023 \pm 0.00006$	$0.0909 \pm 0.0032 \pm 0.0001$	$0.0698 \pm 0.0030 \pm 0.0001$	$0.0395 \pm 0.0069 \pm 0.0004$
$jet - width > 0.15$	$0.1120 \pm 0.0013 \pm 0.00002$	$0.1130 \pm 0.0024 \pm 0.00004$	$0.0885 \pm 0.0038 \pm 0.00008$	$0.0741 \pm 0.0059 \pm 0.0002$	$0.0519 \pm 0.0060 \pm 0.0002$	$0.0583 \pm 0.0215 \pm 0.0009$

Table 7.5: Tau Fake Rates for each bin in p_T and jet-width with their statistical and systematic uncertainties ($FR \pm \sigma_{stat} \pm \sigma_{syst}$ for 1-prong taus in the barrel region, with MET < 20 GeV).

Central Region, MET > 20	$20 \leq p_T \leq 25$	$25 \leq p_T \leq 30$	$30 < p_T \leq 40$	$p_T > 40$
$0 \leq jet - width \leq 0.05$	$0.0583 \pm 0.0045 \pm 0.0002$	$0.0574 \pm 0.0050 \pm 0.0003$	$0.0532 \pm 0.0035 \pm 0.0003$	$0.0252 \pm 0.0011 \pm 0.0002$
$0.05 \leq jet - width \leq 0.07$	$0.0519 \pm 0.0026 \pm 0.0001$	$0.0463 \pm 0.0027 \pm 0.0002$	$0.0373 \pm 0.0018 \pm 0.0001$	$0.0212 \pm 0.0018 \pm 0.0001$
$0.07 \leq jet - width \leq 0.09$	$0.0433 \pm 0.0016 \pm 0.00006$	$0.0398 \pm 0.0016 \pm 0.00008$	$0.0322 \pm 0.0011 \pm 0.00008$	$0.0190 \pm 0.0007 \pm 0.00009$
$0.09 \leq jet - width \leq 0.11$	$0.0343 \pm 0.0010 \pm 0.00004$	$0.0361 \pm 0.0012 \pm 0.00005$	$0.0265 \pm 0.0009 \pm 0.00006$	$0.0184 \pm 0.0041 \pm 0.0004$
$0.11 \leq jet - width \leq 0.15$	$0.0283 \pm 0.0010 \pm 0.00002$	$0.0284 \pm 0.0007 \pm 0.00003$	$0.0219 \pm 0.0006 \pm 0.00004$	$0.0145 \pm 0.0008 \pm 0.00008$
$jet - width > 0.15$	$0.0244 \pm 0.0005 \pm 0.00002$	$0.0238 \pm 0.0076 \pm 0.00003$	$0.0204 \pm 0.0010 \pm 0.00006$	$0.0126 \pm 0.0015 \pm 0.0001$

Table 7.6: Tau Fake Rates for each bin in p_T and jet-width with their statistical and systematic uncertainties (FR $\pm \sigma_{stat} \pm \sigma_{sys}$ for 3-prong taus in the barrel region, with MET > 20 GeV).

Central Region, MET < 20	$20 \leq p_T \leq 25$	$25 \leq p_T \leq 30$	$30 < p_T \leq 40$	$p_T > 40$
$0 \leq jet - width \leq 0.05$	$0.0743 \pm 0.0046 \pm 0.00005$	$0.0728 \pm 0.0050 \pm 0.0002$	$0.0569 \pm 0.0037 \pm 0.0001$	$0.0269 \pm 0.0016 \pm 0.00015$
$0.05 \leq jet - width \leq 0.07$	$0.0537 \pm 0.0025 \pm 0.00003$	$0.0540 \pm 0.0027 \pm 0.00005$	$0.0455 \pm 0.0021 \pm 0.00009$	$0.0211 \pm 0.0012 \pm 0.00008$
$0.07 \leq jet - width \leq 0.09$	$0.0414 \pm 0.0014 \pm 0.00002$	$0.0416 \pm 0.0016 \pm 0.00003$	$0.0347 \pm 0.0013 \pm 0.00004$	$0.0204 \pm 0.0011 \pm 0.00006$
$0.09 \leq jet - width \leq 0.11$	$0.0335 \pm 0.0010 \pm 0.00002$	$0.0365 \pm 0.0012 \pm 0.00002$	$0.0316 \pm 0.0011 \pm 0.00003$	$0.0175 \pm 0.0011 \pm 0.00006$
$0.11 \leq jet - width \leq 0.15$	$0.0300 \pm 0.0006 \pm 0.000007$	$0.0336 \pm 0.0007 \pm 0.00001$	$0.0252 \pm 0.0008 \pm 0.00003$	$0.0148 \pm 0.0011 \pm 0.00005$
$jet - width > 0.15$	$0.0276 \pm 0.0006 \pm 0.000009$	$0.0294 \pm 0.0010 \pm 0.00002$	$0.0217 \pm 0.0012 \pm 0.00004$	$0.0128 \pm 0.0021 \pm 0.00009$

Table 7.7: Tau Fake Rates for each bin in p_T and jet-width with their statistical and systematic uncertainties (FR $\pm \sigma_{stat} \pm \sigma_{sys}$ for 3-prong taus in the barrel region, with MET < 20 GeV).

As can be seen both in the tables and figures for the tau fake rates, the highest fake rates are for 1-prong taus with low p_T and low jet-width. This makes sense as these will be the easiest type of tau for a jet to fake. Furthermore, the 3-prong fake rates are consistently lower than the 1-prong fake rates which verifies our understanding that taus with three clear tracks in the detector are harder to fake.

*“C’est le temps que tu as perdu pour ta rose qui fait ta rose
si importante... Tu deviens responsable pour toujours de ce
que tu as apprivoisé. Tu es responsable de ta rose.”*

Antoine de Saint-Exupéry, *Le Petit Prince*



Analysis Optimization

OUR ANALYSIS STRATEGY COMBINES SELECTION CUTS (DECISIONS BASED ON KINEMATIC VARIABLES) WITH A NEURAL NETWORK TO DISTINGUISH BETWEEN OUR SIGNAL AND BACKGROUNDS. There are three categories of selection for each of our four main channels. The first category is “preselection”. Preselection requirements provide some minimal level of rejection and categorize the data and simulation into our four orthogonal channels. These requirements include trigger selections and object identification and isolation working points. After preselection, our “signal selection” and “validation region” requirements further divide the events. The signal selection is designed to reduce the backgrounds while preserving as much of our signal as possible. These selections include mass reconstruction cuts and more targeted kinematic cuts. At the moment we are still blinded in the analysis, which means that we cannot look at data in the signal region, but our optimization is guided by the signal simulation. Other selections are made to define validation regions. These regions are orthogonal to our signal region and are either designed to provide similar kinematics to our signal regions or to select other physics processes. We also use the validation regions to validate the modeling of our fake background component. Since the validation selections are

orthogonal to the signal, we do not need to blind these regions. The full flow of data through our analysis is shown in figure 8.1.

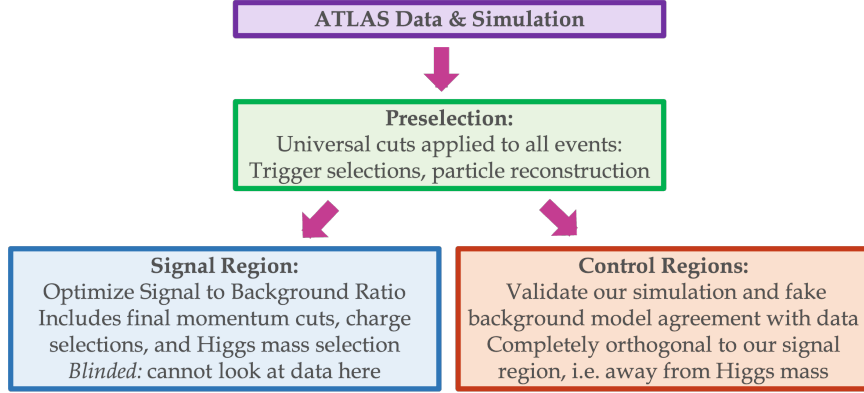


Figure 8.1: General flow of data and simulation through our analysis. Everything goes through preselection and then each channel has its own signal and validation regions.

8.1 SELECTION CUTS

Before an event can even be considered for our analysis, it must be selected by a trigger algorithm. Because we require that the Z/W boson decay to light leptons, we can rely on electron and muon triggers to select our events. There are five main triggers used in this analysis: single electron, single muon, di-electron, di-muon, and combined electron-muon triggers. Table 8.1 lists these triggers. We include an additional p_T requirement for the light lepton (electron/muon) that fires the trigger to reduce uncertainties by requiring our events to be on the plateau of the trigger efficiency curves. The object p_T trigger thresholds change as a function of year of data-taking as the instantaneous luminosity increased over the course of Run 2.

8.1.1 $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$

Preselection

The preselection begins after the initial trigger for electron and muons. The electron(s) and muon(s) are required to be isolated and they must pass the Tight identification working point. Furthermore, we need a tau-lepton that passes the Medium identification requirement. We also apply a b -jet veto to suppress the contribution from the top background. Table 8.2 shows the result of the preselection cutflow

Trigger Type	Trigger	Year	p_T Threshold (GeV)
single muon	HLT_mu20_iloose_L1MU15	2015	21
	HLT_mu50	2015-2018	52.5
	HLT_mu26_ivarmedium	2016-2018	27.3
single electron	HLT_e24_lhmedium_L1EM20VH	2015	25
	HLT_e60_lhmedium_nod0	2015-2018	61
	HLT_e120_lhloose		121
	HLT_e140_lhloose_nod0		141
	HLT_e26_lhtight_nod0_ivarloose	2016-2018	27
di-muon	HLT_2mu14	2015-2018	14.7, 14.7
	HLT_mu18_mu8noL1	2015	18.9, 8.4
	HLT_mu20_mu8noL1	2015-2018	21.4, 8.4/8.14
	HLT_mu22_mu8noL1	2015-2018	23.1, 8.4/8.14
di-electron	HLT_2e12_lhloose_L12EM10VH	2015	13, 13
	HLT_2e17_lhvloose_nod0	2015-2018	18, 18
	HLT_2e17_lhvloose_nod0_L12EM15VH		18, 18
electron-muon	HLT_e17_lhloose_mu14	2015-2018	18, 14.7
	HLT_e17_lhloose_nod0_mu14		18, 14.7

Table 8.1: Triggers used in the Run 2 analysis. This includes single muon, single electron, di-muon, di-electron, and e-mu. The trigger names use a convention where HLT stands for the high-level trigger and the number after mu or e is the online p_T threshold for the object. If there are two objects in the trigger than both online p_T thresholds are included in the name. If there is also an online identification working point used, i.e. loose or vlose (very loose), that is also included in the name. The final column of the table states the p_T threshold we applied additionally to the trigger selection in our offline processing.

for the data, the main backgrounds, and the MC signal expectations. The agreement between the data, contribution from mis-identified objects, and MC for several kinematic variables is shown in Figure 8.2.

Preselection Cuts	N_{obs}	N_{sig}	N_{fakes}	N_{WZ}	N_{ZZ}	N_{top}	N_{VVV}	N_{ttH}
1 Truth τ	-	473	-	36,393	125,432	1,414,416	262	2794
2 Truth e/μ	-	473	-	26,942	107	1,634	167	818
Trigger	94,215,328	309	391,567	16,232	52	1,015	126	534
Object ID	431354	106	391,567	4,209	7.7	246	43.6	118
Object Isolation	388726	101	384,028	3,986	7.1	227	42.1	106
b-jet Veto	357,964	95.8	351,952	3689	5.9	37.1	39.1	16.7

Table 8.2: Preselection for $W \rightarrow \mu\nu/e\nu$, $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$ for the signal MC, main backgrounds, and data

$W \rightarrow \mu\nu/e\nu$, $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$ Signal Region

For our signal region selection in WHlephad, we differentiate by final state leptons. For all final states,

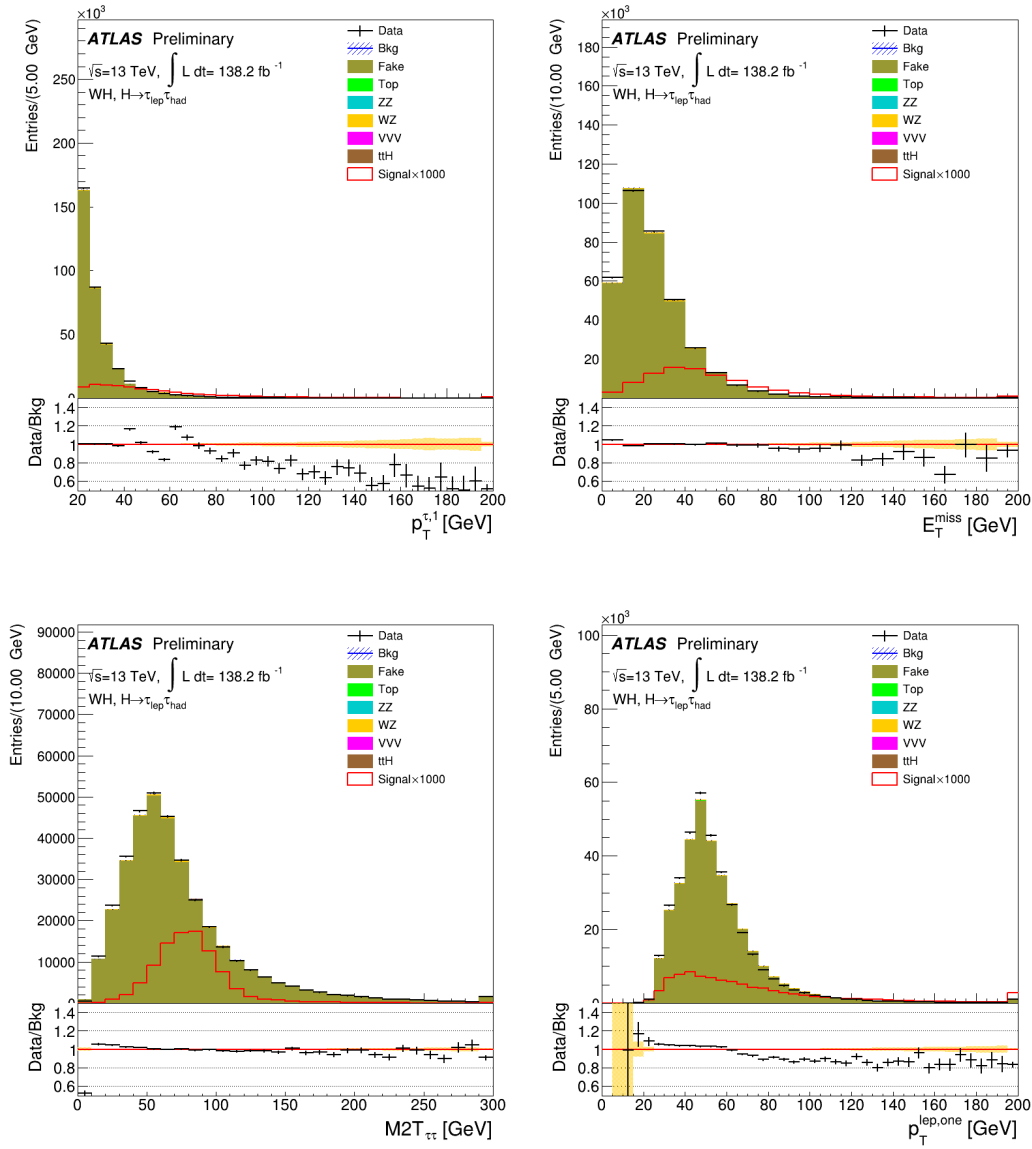


Figure 8.2: Some kinematic distributions in the WHlepHAD channel at preselection. The structure in the upper left hand plot is a result of the p_T binning in the tau fake rate. Some mismodeling can be seen in the lower right-hand plot of leading electron/muon p_T . However, these mismodeling effects are not seen in the validation region plots.

we first apply a same-sign (same electric charge) lepton requirement to reduce background from neutral boson decays, such as $Z \rightarrow ll$. The tau and electron/muon associated to the Higgs boson must have opposite-sign (opposite electric charge) as they originate from the neutral Higgs boson. For the di-electron final state, we additionally require that the mass of the two electrons be between 80 and 100 GeV. The next cut is on the sum of the transverse momentum from the tau and the two light leptons (> 90 GeV). Finally, we apply

a mass window selection to select for events with M_{2T} between 60 and 130 GeV, which helps hone in on our signal region. Table 8.3 shows the resulting signal and background yields. The contribution from ZZ backgrounds is eliminated with the same-sign lepton cut.

Signal Selection	N_{sig}	N_{fakes}	N_{WZ}	N_{top}	N_{VVV}	N_{ttH}
SS Leptons	46.3	4,758	529	8.4	13.2	7.1
OS Lep/Tau	45.8	3,117	516	8.2	13	7.0
mll(ee) in [80.0, 100.0]	44.4	2,656	498	7.9	12.7	6.7
Sum of object's $p_T > 90\text{GeV}$	41.1	1,330	410	7.4	12.2	6.3
M_{2T} in [60.0, 130.0]	33.4	723	273	3.6	6.5	4.2

Table 8.3: Signal selection for $W \rightarrow \mu\nu/e\nu$, $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$ for the signal MC and main backgrounds

$Z \rightarrow \tau\tau$ Validation Region

The $Z \rightarrow \tau\tau$ region requires the electrons or muons to have the same-sign and opposite-flavor so we can select a Z boson decaying to tau pairs, where the taus decay leptonically. The region also includes the additional constraint that the collinear mass of the leptons lie between 60 and 120 GeV.

$t\bar{t}$ Validation Region

For the $t\bar{t}$ region, there is a reversal of the b -jet veto and instead, a requirement of at least one tagged b -jet in the event.

All Same Sign Validation Region

For the all same-sign region, we include a requirement on the dilepton mass window in the ee final state. We also require the electrons and muons to have the same-sign to reduce contributions from $Z + jets$. Finally, the requirement for the Higgs boson decay products to be the same-sign forces the region to be orthogonal to the signal region.

8.1.2 $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$

Preselection

As there is only one electron or muon in the final state, the triggers for this channel are the single muon and single electron triggers. The preselection cuts are applied after the initial electron/muon and two taus requirement. The electron/muon is required to be isolated and tightly identified. The taus are required to pass the Medium identification requirement. Finally, we apply a b -jet veto to reject background events. Table 8.4 shows the preselection cuts for this analysis category for our signal, main backgrounds, and the data. Figure 8.3 shows the agreement between data, contribution from the fake factor model, and MC.

Preselection Cuts	N_{obs}	N_{sig}	N_{fakes}	N_{WZ}	N_{ZZ}	N_{top}	N_{VVV}	N_{ttH}
2 Truth τ	-	501	-	13,917	6,095	9,852	85	933
1 Truth e/μ	-	501	-	13,701	20.7	569	53.9	683
Trigger	91,919,431	294	17,102	7,710	6.8	340	34.9	400
Object ID	18,939	71.9	17,102	790	0.1	37.7	4.7	48.9
Object Isolation	18,486	71.6	17,098	787	0.1	37.4	4.7	48.5
b -jet Veto	15,336	68.6	13,904	727	0.08	5.94	4.3	7.8

Table 8.4: Preselection yields for the $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ channel in the signal MC, main backgrounds, and data.

$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ Signal Region

The signal region selection requires that the taus be of opposite-sign as they originate from a neutral Higgs boson. A requirement on $\Delta R(\tau_h, \tau_h) \in (0.8, 2.8)$ is applied to reduce the fake component. In order to reduce the $Z \rightarrow \tau\tau$ background, a cut on the transverse mass between the missing energy and the light-lepton is made. To target the W +jets and Z +jets backgrounds, a requirement that the sum of the taus' transverse momentum exceed 90GeV is made. Finally, we apply a mass window selection to select for events with M_{2T} between 80 and 130GeV, which helps hone in on our signal region. The selection cuts are shown in Table 8.5.

$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ Validation Regions

There are more validation regions for the WHhadhad channel than for the other channels because it is the highest rate channel with very low purity due to having only one selected electron or muon. Thus, the backgrounds are challenging and further validation is required.

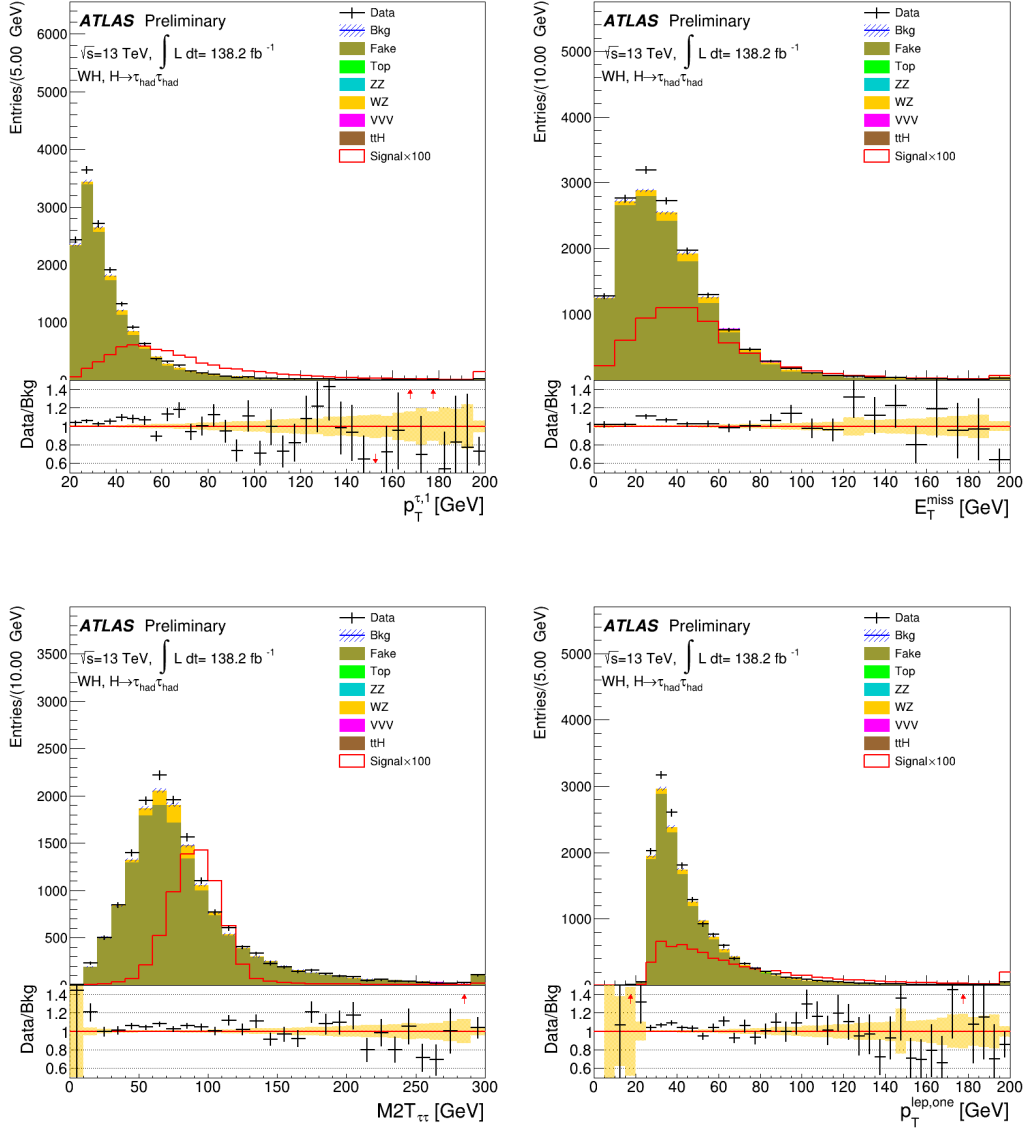


Figure 8.3: Some kinematic distributions in the $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ channel at pre-selection.

Signal Selection	N_{sig}	N_{fakes}	N_{WZ}	N_{ZZ}	N_{top}	N_{VVV}	N_{ttH}
OS Taus	64.6	7,080	686	0.06	5.2	3.0	7.0
$\Delta R(\tau_{had}, \tau_{had}) \in (0.8, 2.8)$	50.1	4,077	515	0.03	3.5	1.9	5.6
$m_T(lep, MET) > 20 \text{ GeV}$	46.3	3,391	480	0.02	2.9	1.8	5.0
$\sum p_T(\tau_{had}) > 90 \text{ GeV}$	29.2	280	184	0.01	1.8	1.1	3.4
$M_{2T} \in (80.0, 130.0) \text{ GeV}$	24.4	164	80.4	0.01	0.7	0.5	2.5

Table 8.5: Signal region yields for the $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ channel in the signal MC and main backgrounds.

$Z \rightarrow \tau\tau$ Validation Region

The $Z \rightarrow \tau\tau$ region includes an opposite-sign tau requirement, as well as an upper bound on the $m_T(l, MET)$ of 40GeV to reject background contamination from W decays. We also bound M_{2T} with 60GeV to be orthogonal to the signal region.

$t\bar{t}$ Validation Region

This validation region requires a b -jet in the final state.

W+jets Validation Region

In this region, the lower bound on the transverse mass between lepton and MET is increased to 60GeV in order to reduce the background from Z +jets. Also, the taus are required to be the same-sign.

Same-Sign Taus Validation Region

The same-sign taus validation region has an additional requirement that the taus be of the same-sign. The W+jets validation region is a subset of the same-sign taus validation region.

M_{2T} Sideband Validation Region

The M_{2T} sideband region uses all of the same cuts as the signal region but flips the M_{2T} window selection such that it must be less than 80GeV or higher than 130GeV. This cut completely excludes the signal region, but otherwise has similar kinematics, and thus this region is useful for validating the fake background model.

8.1.3 $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$

$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$ Preselection

The three light leptons (electrons and muons) must pass Tight identification and isolation requirements. The tau must pass the Medium identification requirement. The selection cut of $MMC > 0$ ensures the MMC algorithm has converged to a value. We also require two of the light leptons to be the decay products of a neutral Z boson. Table 8.6 shows the weighted events after each step of the preselection. Figure 8.4 shows the agreement at preselection between the data, background from mis-identified jets (fakes), and MC.

Preselection Cuts	N_{obs}	N_{sig}	N_{fakes}	N_{ZZ}	N_{VVV}	N_{Top}	N_{ttH}
1 Truth τ	-	117	-	5491	146	252,833	1038
3 Truth e/μ	-	117	-	3278	35	473	136
Trigger	11,822,458	102	764	2,523	31	339	109
Object ID	2130	22.8	764	209	6.8	32.7	16.5
Object Isolation	1139	21.1	764	185	6.3	27.6	14.1
2 OS, SF Leptons	934	20.8	734	184	6.0	26.0	10.7
$MMC > 0$	908	20.4	712	181	4.8	22.3	9.3

Table 8.6: Preselection for $Z \rightarrow \mu\mu/ee$, $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$.

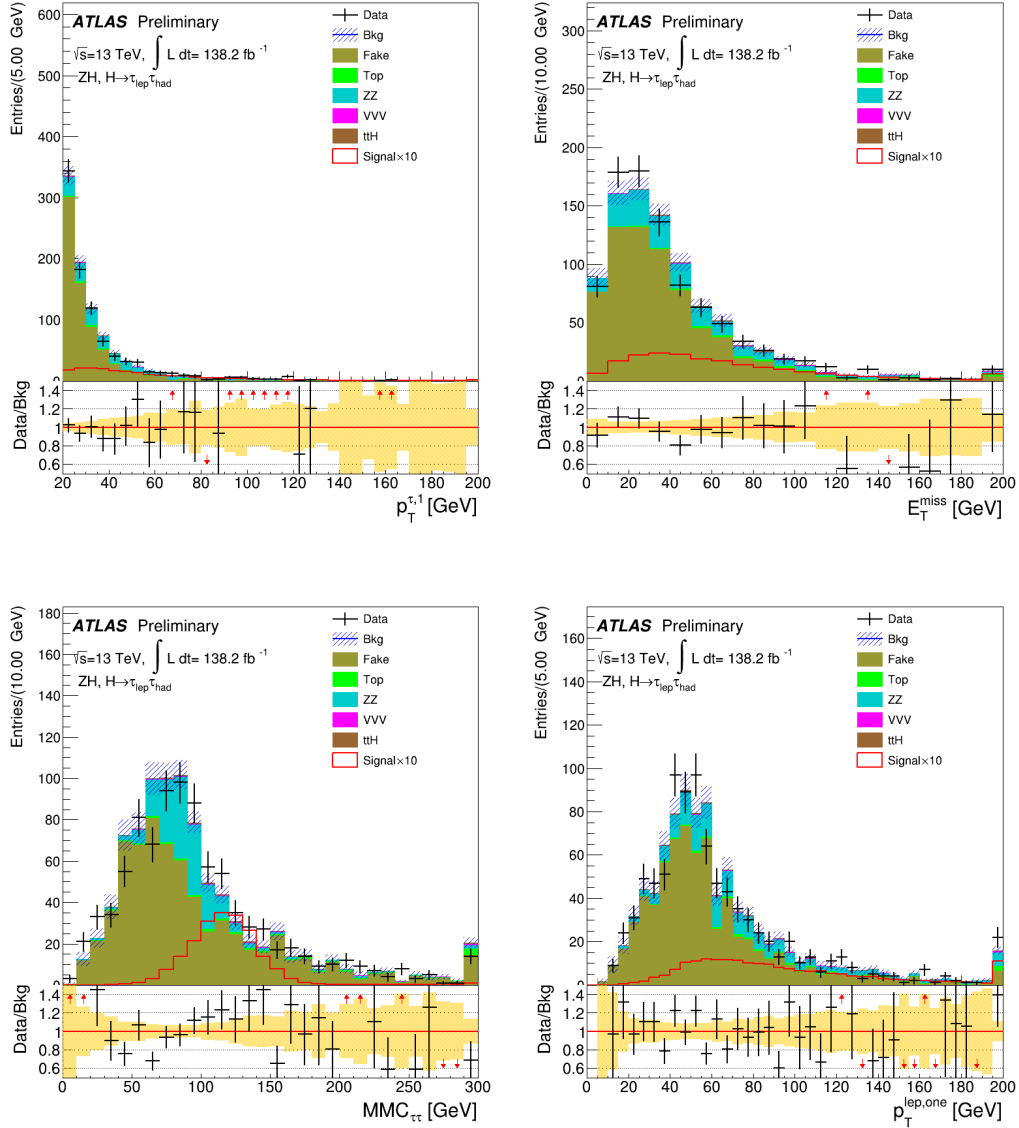


Figure 8.4: Some kinematic distributions in the ZHlep had channel at preselection.

$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$ **Signal Region**

We include a selection cut on the Higgs boson transverse momentum to reduce background from Z +jets. The Higgs boson decay products must also be of the opposite-sign in order to reconstruct the neutral Higgs boson. While the two light leptons with p_T closest to 91GeV are associated with the Z boson, their dilepton mass must also be between 81 and 101GeV for the event to be included in the analysis and to suppress contributions from the fake background. We also select events in a range of values for MMC that include the Higgs boson mass (125GeV) to hone in on our signal.

Signal Selection	N_{sig}	N_{fakes}	N_{ZZ}	N_{VVV}	N_{Top}	N_{ttH}
OS tau and electron/muon	20.2	387	179	4.7	21.8	9.1
Z mass selection	12.6	204	113	1.6	7.9	1.2
Sum Tau $p_T > 60$ GeV	9.6	30	55	1.3	6.0	1.0
MMC in [100.0, 170.0]	7.7	12	17.8	0.4	2.3	0.4

Table 8.7: Signal selection for $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$ with the signal MC and main backgrounds.

$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$ **Validation Regions**

Same-Sign Taus Validation Region

This region adds the requirement that the taus must be of the same-sign.

MMC Sideband Validation Region

This region uses all the signal region cuts and also inverts the MMC selection, such that $MMC < 100$ or $MMC > 170$ to be orthogonal to the signal region for fake background model validation.

8.1.4 $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$

This channel begins by requiring exactly two tightly identified electrons or muons and two taus. The taus must pass the Medium identification working point. The two highest p_T electrons/muons of opposite-sign and same-flavor are associated to the Z boson.

$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$ **Preselection**

The preselection starts with the trigger requirement on the light leptons. Then the existence of a same-flavor, opposite-sign light lepton pair is required. The MMC must have converged to a value. Table 8.8

shows the weighted cutflows for the preselection and figure 8.5 shows the agreement between MC, the data, and the fake component.

Preselection Cuts	N_{obs}	N_{sig}	N_{fakes}	N_{top}	N_{ZZ}	N_{VVV}	N_{ttH}
2 Truth τ	-	116	-	1808	2846.1	23.1	287
2 Truth e/μ	-	116	-	127	2816.6	17.7	156
Trigger	11,636,417	97	671	93	2145.0	14.5	117
Object ID	1147	18.1	671	7.4	149.4	1.5	10.6
Lepton Isolation	924	17.6	671	7.0	142.3	1.4	9.9
2 OS, SF Leptons	844	16.6	671	4.0	138.6	0.8	4.4
MMC > 0	832	16.2	659	3.2	136.3	0.7	3.7

Table 8.8: Preselection for $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$.

$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$ Signal Region

The signal region additionally requires that the taus from the Higgs boson are of opposite-sign to reconstruct the neutral Higgs boson. The invariant dilepton mass must be within 20GeV of the Z mass ($71 < m_{ll} < 111\text{GeV}$) to reduce the contribution from fake events. The sum of the momentum from the taus must also be greater than 75GeV to further suppress the contribution from the background due to fakes. We also select events in a range of values for MMC that include the Higgs boson mass (125GeV) to hone in on our signal. Table 8.9 shows the signal region cutflows applied to data and MC.

Signal Selection	N_{sig}	N_{fakes}	N_{top}	N_{ZZ}	N_{VVV}	N_{ttH}
OS taus	16.0	342	3.2	134	0.7	3.6
Z mass window	15.0	261	1.5	119	0.4	1.0
Sum Tau $p_T > 75\text{GeV}$	12.2	40	1.1	60.8	0.3	0.7
MMC in [100.0,180.0]	11.0	17	0.4	26.5	0.1	0.5

Table 8.9: Signal Region Cutflows for $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$.

$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$ Validation Regions

Same-Sign Taus Validation Region

In this region, we require the two taus to be of the same-sign.

MMC Sideband Validation Region

The MMC Sideband region has a requirement that the MMC be either $< 80\text{GeV}$ or $> 160\text{GeV}$. This is orthogonal to the signal region and allows us to validate the modelling of the background from mis-identified objects.

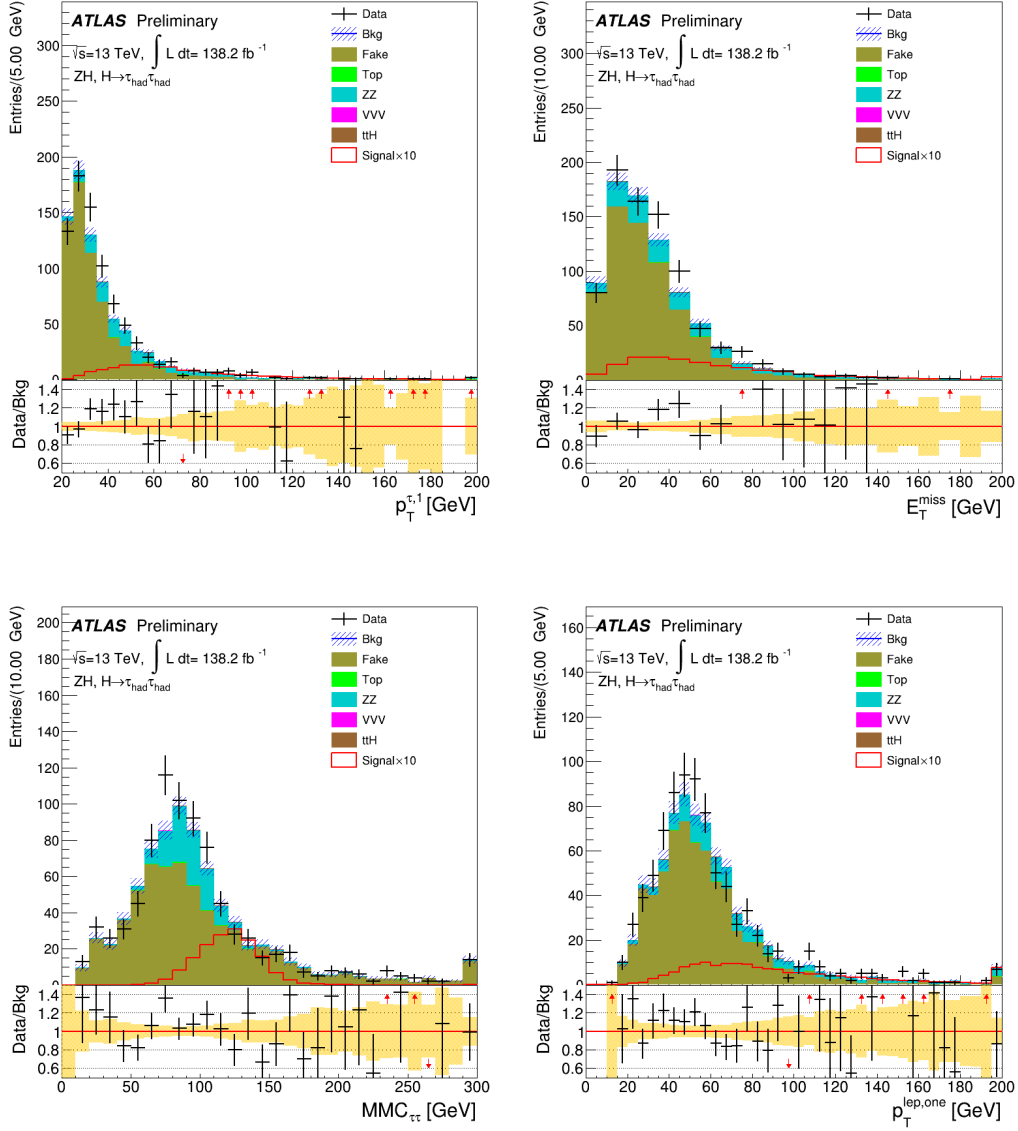


Figure 8.5: Some kinematic distributions in the $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$ channel at pre-selection.

8.2 OPTIMIZATION OF SIGNAL SELECTIONS CUTS

For each of the flexible signal selection cuts, such as the sum of object p_T , dilepton mass windows, and transverse mass cuts, the final cut value was determined by finding the optimal signal to background ratio while keeping the signal efficiency as high as possible. Most of these selection cuts are targeting our backgrounds due to misidentification, which cannot be directly modeled without unblinding. Thus, validation regions for the fake backgrounds such as same-sign taus, and non-truth matched simulation were looked at

to best approximate the expected distributions for the background that arises from misidentification. This means we only look at simulated events that actually contain the true objects, and not mis-reconstructed objects, we want in our four categories. However, we can remove those truth-matching selections to use our simulation to give us a different estimate for how misidentified objects can enter into our signal region.

8.2.1 $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$

The signal selection cut that can be optimized for the WHlep_{had} category is the sum of object p_T . First, it was determined that summing all three final state objects - the tau and lepton from the Higgs and the lepton from the W , gave better signal and background separation than just summing the p_T of the two electrons and/or muons. To optimize this cut, signal MC with all the previous signal selection criteria was compared to same-sign tau events, which are dominated by backgrounds from misidentification, and Monte Carlo channels without the needed number of true electrons, muons and taus that nonetheless pass the selection cuts on reconstructed objects (non-truth matched MC). We anticipate a significant background from $W + jets$ and $Z + jets$ events in both of these background selections. In our preselection, we apply truth-matching requirements to our MC contributions.

First, comparing signal MC to same-sign tau fakes, we see this slight peak separation on the left-hand side of figure 8.6. By looking at the corresponding significance and signal efficiency at each cut level, I found that placing the cut at 90GeV gives a gain in significance without much lost in signal efficiency as we keep 90% of our signal (right-hand side of figure 8.6).

The same variable distribution shape and significance study was looked at for the MC signal and non-truth matched MC. Here, one can see that a cut at greater than $\sim 85\text{GeV}$ will eliminate the $W + jets$ peak (left side of figure 8.7). Varying the cut from 40 – 120 GeV, one can see that a cut value of 90GeV will keep signal efficiency above 80% while reducing the background shown in non-truth matched MC (right side of figure 8.7). We chose 90GeV as our final cut placement as it is also an optimal cut point when comparing signal to the same-sign taus background.

8.2.2 $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$

There were several selection cuts that could have been varied in the WHhad_{had} signal category: ΔR between taus, transverse mass between lepton and MET, and the sum of the taus' p_T . After looking at the

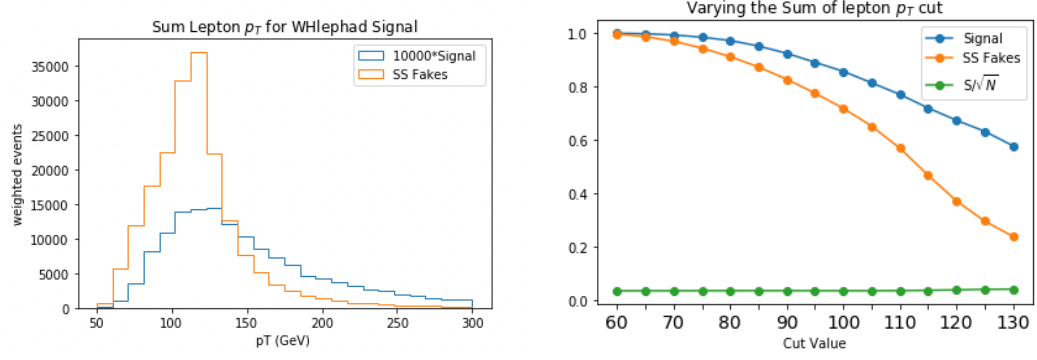


Figure 8.6: Distribution of the sum of light lepton and tau p_T for the signal MC (weighted by 10,000) and the same-sign fakes background for the WHlephad channel (left). The plot on the right shows the amount of signal MC and fake background remaining, and the significance, for each potential cut value on the sum of lepton p_T . This was used to determine where to place the cut in the WHlephad channel while maintaining high signal efficiency.

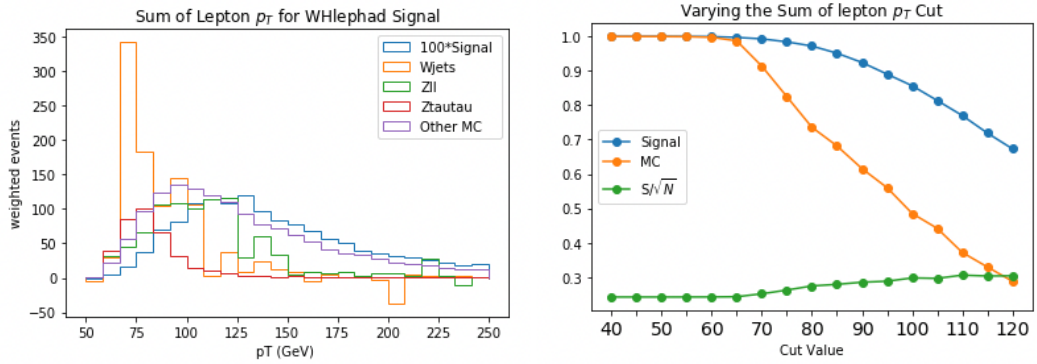


Figure 8.7: Distribution of the sum of light lepton and tau p_T for the signal MC (weighted by 100) and the non truth matched MC background, including W +jets, $Z \rightarrow ll$, $Z \rightarrow \tau\tau$ and other MC contributions for the WHlephad channel (left). The plot on the right shows the amount of signal MC and non-truth matched MC background remaining, and the significance, for each potential cut value on the sum of lepton p_T . This was used to determine where to place the cut in the WHlephad channel while maintaining high signal efficiency.

distributions of these variables in both our signal region and our same-sign taus region that is dominated by mis-identified (fake) objects, which can be seen in figures 8.8 and 8.9, I determined that the ΔR cut should remain between 0.8 and 2.8, and $m_T > 20\text{GeV}$. These cut placements are able to remove fake background contribution with minimal to no signal removal.

The remaining selection cut that can be tweaked in the WHhadhad signal selection, is the sum of

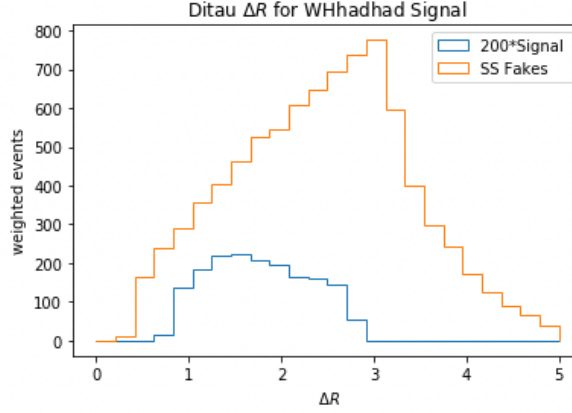


Figure 8.8: Distribution of the di-tau ΔR variable for the signal MC (weighted by 200) and the same-sign background used to determine the optimal cut window for the WHhadhad channel.

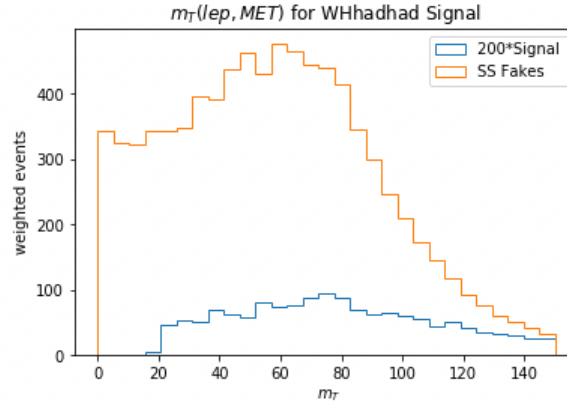


Figure 8.9: Distribution of the transverse mass variable for the signal MC (weighted by 200) and the same-sign background as an estimation of the contribution from the background due to mis-identified objects.

the p_T of the taus in the event. Both same-sign validation region as an estimation for the contribution of the background from misidentified jets and non-truth matched Monte Carlo distributions were looked at in order to determine optimal cut placement. The relevant distributions of the tau p_T sum are shown in figure 8.10. The signal peaks at around 100GeV, which is at a higher p_T value than the peaks in both same-sign fakes and non-truth matched Monte Carlo.

The questions that remain are (1) how much signal are we willing to lose and (2) how will the expected background in the signal region relate to the two distributions chosen? The two backgrounds distributions

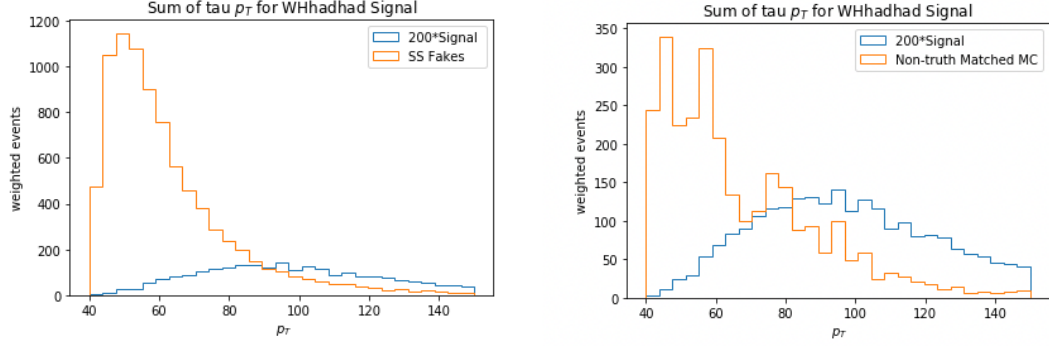


Figure 8.10: Distribution of the sum of tau p_T for the WHhadhad signal MC (weighted by 200) compared to the same-sign background to find the optimal cut placement (left). Distribution of the sum of tau p_T for the WHhadhad signal MC (weighted by 200) compared to the non truth matched Monte Carlo (as a background estimate) to find the optimal cut placement (right).

suggest slightly different optimal cut placements (see figure 8.11). The requirement was placed at 90GeV, which removes approximately 90% of the background while keeping 60% of the signal.

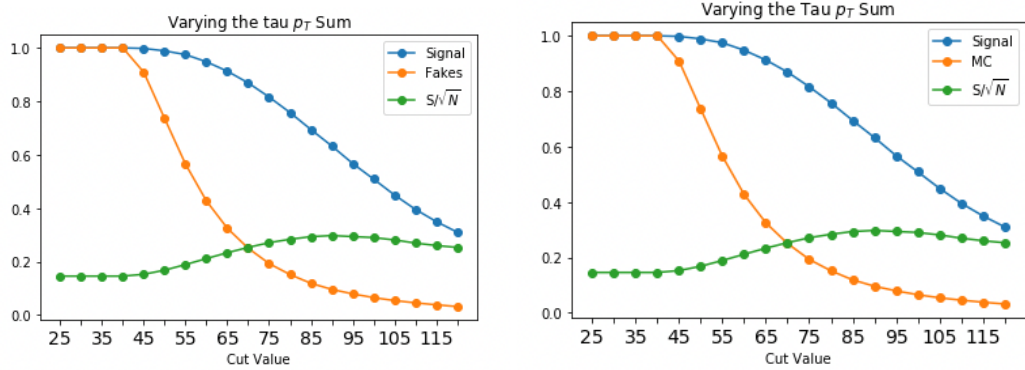


Figure 8.11: The plot on the left shows the amount of signal MC and same-sign background (due to misidentified jets) remaining, and the significance, for each potential cut value on the sum of tau p_T . This was used to determine where to place the cut in the WHhadhad channel while maintaining high signal efficiency. The plot on the right shows the amount of signal MC and background estimated from the non-truth matched MC, and the significance, for each potential cut value on the sum of tau p_T . This was used to determine where to place the cut in the WHhadhad channel while maintaining high signal efficiency.

8.2.3 $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$

For this channel, we are able to tune the range of the dilepton mass selection for the Z peak in order to optimize the amount of signal retained compared to the background. In Run 1, a window of 70–110 GeV was used. As can be seen in figure 8.12, the dilepton mass window can be narrowed to 81–101 GeV and still maintain high signal efficiency.

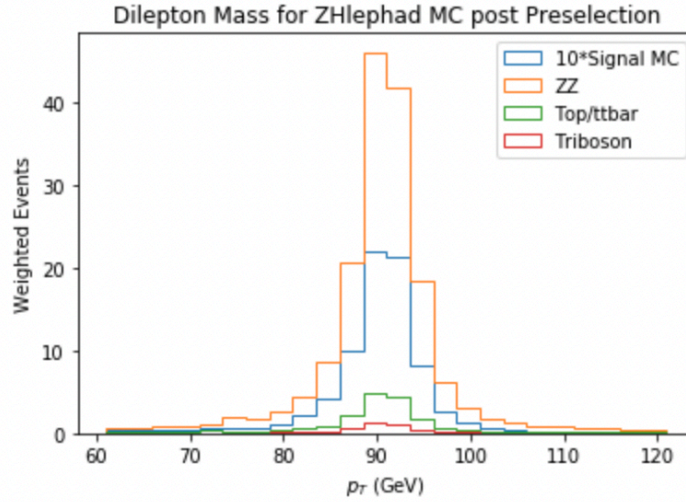


Figure 8.12: Dilepton mass of our signal MC (weighted by 10) compared to the main background contributions. This distribution was used to determine the dilepton mass cut in the ZHlepHad channel.

The next cut to develop was the sum of the p_T of the taus in the event. This cut is meant to reduce backgrounds from events where jets have been misidentified as taus. The cut was then optimized against the p_T distribution of same-sign tau fakes, which, while orthogonal to the signal region, should have a similar p_T spectrum of jets as expected in the signal region. For the signal MC looked at, all the previous preselection and signal selection criteria were applied. Based on both the two distributions and signal vs fake efficiency and significance (figure 8.13) the optimal cut placement while retaining 80% of our signal is 60 GeV.

The background contribution from ZZ events is quite high after the previous signal selection cuts. A similar study was performed where the signal distribution was compared to the simulated ZZ distribution for the sum of the p_T of the taus, but the kinematics in ZZ and ZH are similar enough that there is little separation gained in these variables. However, a dedicated neural network was trained to distinguish ZZ

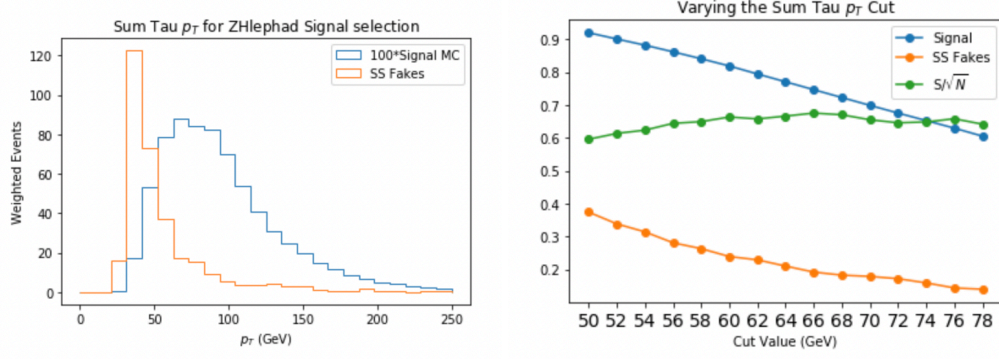


Figure 8.13: Distribution of the sum of tau p_T in the ZHlephad channel for the signal MC (weighted by 100) and the same-sign fakes contribution (left). The plot on the right shows the amount of signal MC and background from misreconstructed jets remaining, and the significance, for each potential cut value on the sum of the tau's p_T . This was used to determine where to place the cut in the ZHlephad channel while maintaining high signal efficiency.

events from the ZHlephad signal.

8.2.4 $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$

In the ZHhadhad channel we apply a mass requirement window around the di-muon or di-electron reconstructed mass. A wide window of 71–111 GeV is chosen, centered on the Z boson mass. As shown in Figure 8.14, this selection is extremely efficient in terms of keeping the signal. As expected, the selection does not reject the background from diboson (ZZ) events; that background is instead targeted with a neural network. The mass window selection does reduce the background from fakes, as can be seen in table 8.9.

The next cut developed was on the sum of the p_T of the taus in the event. This cut is meant to reduce backgrounds from events where jets have been misidentified as taus. The cut was optimized using the same-sign tau validation region which, though it is orthogonal to the signal region, should have a similar p_T spectrum of jets as is expected in the signal region. For the signal MC looked at, all the previous preselection and signal selection criteria were applied. In order to retain a signal efficiency of 80%, the sum p_T requirement was placed at 75 GeV. The signal and background efficiencies, as well as significance, are provided in figure 8.15.

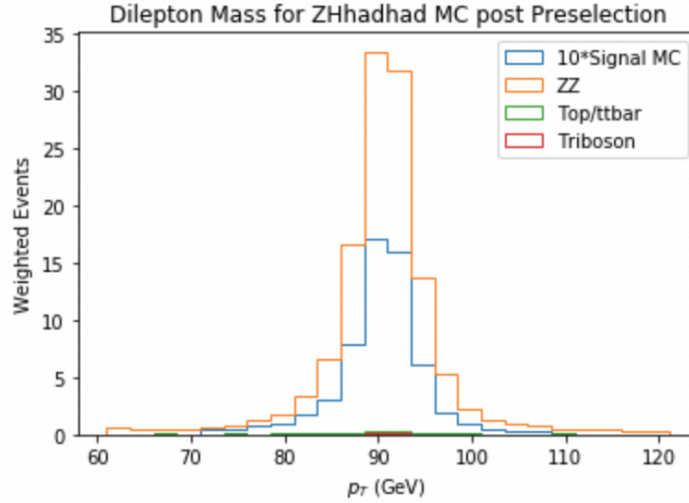


Figure 8.14: Dilepton mass of our signal MC (weighted by ten) compared to the main backgrounds modeled by the simulation. This distribution was used to determine the dilepton mass cut in the ZHhadhad channel.

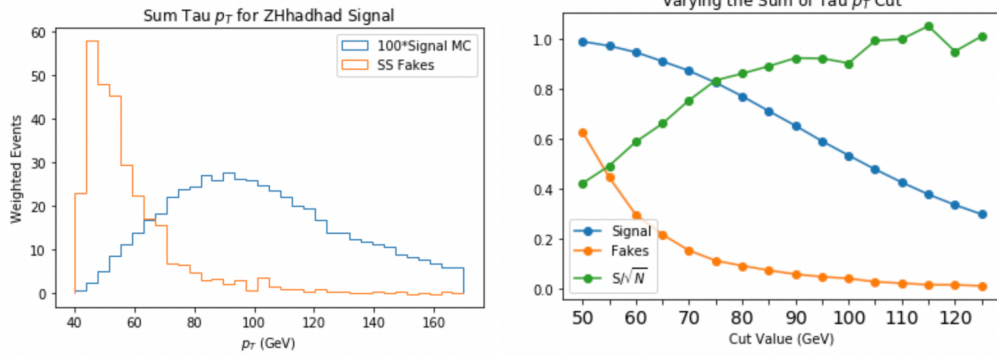


Figure 8.15: Distribution of the sum of tau p_T in the ZHhadhad channel for the signal MC (weighted by 100) and the same-sign fakes contribution (left). The plot on the right shows the amount of signal MC and fake background remaining, and the significance, for each potential cut value on the sum of tau p_T . This was used to determine where to place the cut in the ZHhadhad channel while maintaining high signal efficiency .

8.3 NEURAL NETWORK TARGETING DIBOSONS

In Run 1, the reconstructions of the Higgs boson mass via M_{2T} and MMC were used to separate signal from background as the final discriminating variable. As can be seen in figure 8.16 separating out signal from diboson background (modeled here with Monte Carlo) using MMC or M_{2T} is tricky. There is no strong separation between the signal and the background. In Run 2, we trained a neural network to

separate out the signal from the background with the goal of enhancing our sensitivity. We chose a neural network over a boosted decision tree as it allows for correlation of variables, is better adapted to training with lower statistics, and we can include layers to enforce physics properties, i.e. rotational invariance.

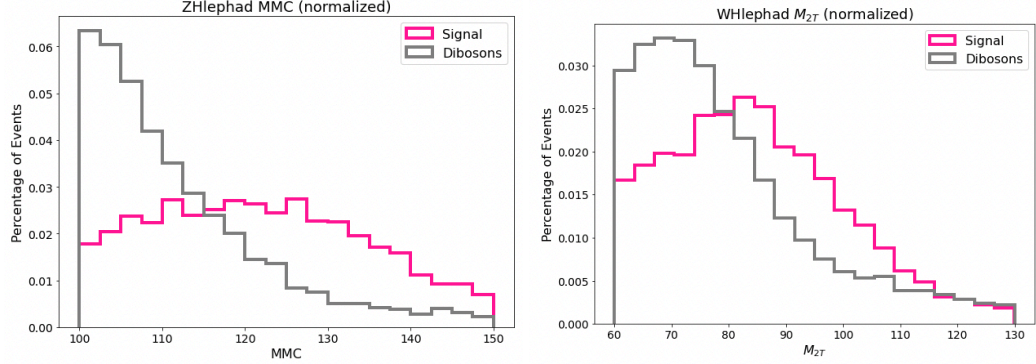


Figure 8.16: Normalized distribution of MMC for the ZHlephad channel for signal and diboson Monte Carlo after signal selection (left) and of M_{2T} for WHlephad showing the challenge of using the Higgs boson mass reconstructions for signal separation.

The general optimized architecture for our network begins with two layers enforcing rotational invariance and encoding. The first rotational layer adds a *global ϕ offset* to the ϕ input variables. This helps the network learn that events rotated with respect to ϕ do not alter the physics of the event for classification purposes. This layer is only included when training and is not used in the application of the neural network in our selection. The second angular layer does an *angular encoding* or parametrization of ϕ into $\sin(\phi)$ and $\cos(\phi)$ to signal the continuous nature of ϕ at the break from 2π to 0 and enforce that $\phi + 2\pi n \rightarrow \phi \forall n \in \mathbb{Z}$.

The rotational layers are followed up by three fully-connected hidden layers. The number of layers included in the network was also optimized during training. The general guideline we followed was to have ten times more training events than trainable parameters in the network. We tuned the number of nodes in each layer (and number of layers) to generally follow this rule. The activation function for the hidden layers was ReLU. The loss function was binary cross entropy, the learning rate was 0.0005, and the optimizer was stochastic gradient descent (SGD) with momentum = 0.9. The final layer was a single output node using the sigmoid activation function that gives a score between 0 (for background) and 1 (for signal). The full architecture for the network is shown in figure 8.17

For training, we apply the preselection cuts to the signal and diboson MC (with the exception of the

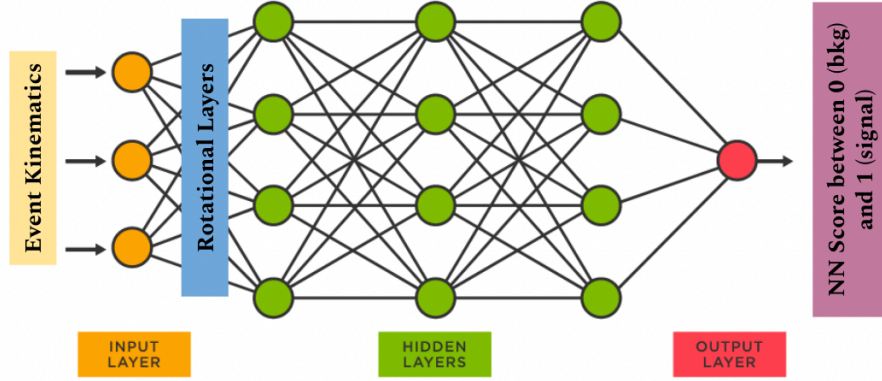


Figure 8.17: Schematic of the neural network architecture with the input kinematics, rotational layers, three hidden fully-connected layers and the final output score.

b -jet veto in the WH channels). We then separate into a training, validation, and test set with a 70/15/15% split, respectively. In the training set, we drop all events with a negative MC weight, but keep those events in our test and validation sets. We train on unweighted, with respect to MC weights, events. MC weights are typically used to enable data and MC to be directly compared, but their inclusion here can negatively impact the training of the neural network. However, we do apply a class weighting on the events to even out the importance of the signal and background events. We have many more background training events so without class weighting a network could achieve decent accuracy by classifying all events as background.

The input variables are a mix of low-level kinematic information for the final state particles and higher-level event information. I chose the input training variables that had the greatest separation power between signal and background. The input variables included in training for all channels and the additional ones for each specific channel are listed in table 8.10.

A baseline neural network was created for all channels consisting of just one fully-connected hidden layer with ten nodes using the ReLU activation function with an output layer of one node using the sigmoid activation. This baseline network was used to compare the performance of the more optimized neural network architectures.

All channels	ZHhadhad	ZHlepbad	WHhadhad	WHlepbad_ee	WHlepbad_emu	WHlepbad_mumu
Leading τ # Prongs	Subleading τ # Prongs	Subleading lepton p_T	Subleading τ # Prongs	Subleading lepton p_T	Subleading lepton p_T	Subleading lepton p_T
Leading τ p_T	Subleading τ p_T	Subleading lepton η	Subleading τp_T	Subleading lepton η	Subleading lepton η	Subleading lepton η
Leading τ η	Subleading τ η	Subleading lepton ϕ	Subleading $\tau \eta$	Subleading lepton ϕ	Subleading lepton ϕ	Subleading lepton ϕ
Leading τ ϕ	Subleading τ ϕ	Higgs lepton p_T	Subleading $\tau \phi$	Dilep $\Delta\eta$	Dilep $\Delta\eta$	Dilep $\Delta\eta$
$\Delta R(\tau, lep)$	Subleading lepton p_T	Higgs lepton η		Leading τ jet-width	Leading τ jet-width	Leading τ jet-width
$\Delta R(lepton s)$	Subleading lepton η	Higgs lepton ϕ		Ditau p_T sum	Ditau mass	Dilep ΔR
Leading lepton p_T	Subleading lepton ϕ			Ditau mass	# of jets	Ditau Mass
Leading lepton η				Ditau $\Delta\eta$	Ditau $\Delta\eta$	Ditau $\Delta\eta$
Leading lepton ϕ				Dilep $\Delta\phi$	Sum All p_T	Ditau ΔR
MET p_T				$\Delta\phi(\tau, MET)$	$\Delta\phi(\tau, MET)$	Sum of object p_T
MET ϕ						Dilep $\Delta\phi$
Dilepton mass						

Table 8.10: Input variables for the neural network included in all channels and then for the specific channels.

8.3.1 $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$ NEURAL NETWORK

Originally, we just trained four neural networks, one for each of the channels. When we separated the WHlep_{had} category into three final states based on the flavor of the light leptons, the WHlep_{had} neural network did not perform well. This can be seen in figure 8.18 where the left hand side shows the former neural network performance on all three categories (‘ee’ - di-electron, ‘emu’ - electron & muon, ‘mumu’ - di-muon) combined on the left and the di-electron channel alone on the right. When the old score was applied to all three categories at once there was clean signal/background separation, but that power was lost once the neural network score was applied to all three channels at once.

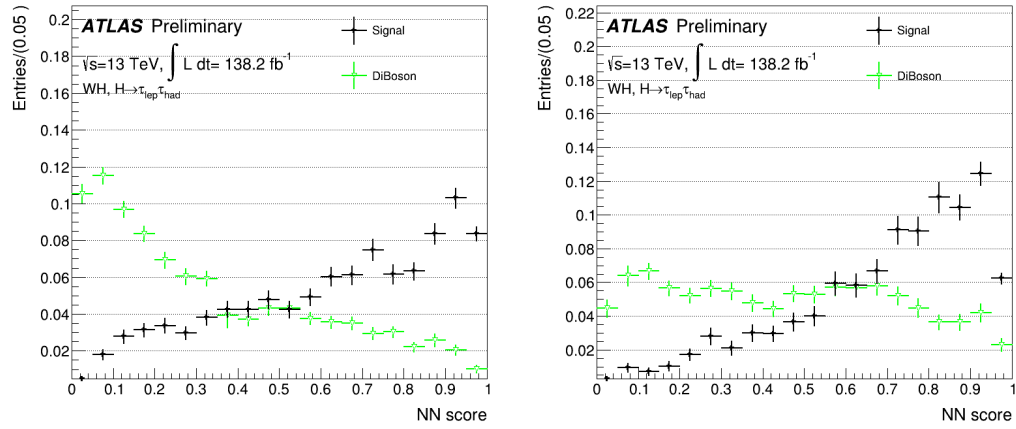


Figure 8.18: Neural network score for the original WHlep_{had} channel when applied to all three final states (‘ee’ - di-electron, ‘emu’ - electron & muon, ‘mumu’ - di-muon) simultaneously with clear signal and background separation (left) and on just the WHlep_{had_ee}, di-electron, channel where the background (dibosons) has a flattened distribution and some of the separation has disappeared (right).

8.3.1.1 WHLEPHAD_ee NEURAL NETWORK

The training set for the di-electron final state of the WHlep_{had} channel had 5,722 signal and 42,081 background events. We weighted the signal 7.4 times more than background. For the final optimized model, we had 40 nodes in the three hidden layers and used the optimizer SGD. The test AUC was 0.867. There was no demonstrated over-training as shown in the similar performance of the test, train, and validation sets and their ROC curves (see figure 8.19). The validation set had an AUC of 0.870 and the training set had an AUC of 0.884.

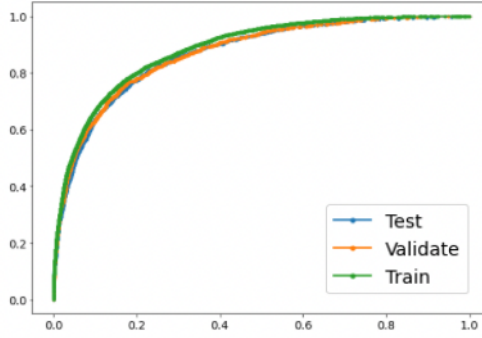


Figure 8.19: ROC curves for the training, test, and validation set for the WHlephad channel with the di-electron final state. The similar performance for all three samples demonstrates there is not significant over-training occurring.

Figure 8.20 shows the performance of the neural network at preselection on the test set without weights to correct for MC cross-sections on the left. The plot on the right shows the final performance of the neural network score on the signal and diboson MC that passes the full signal selection. The separation in the unweighted, with respect to MC weights, and normalized preselection test set is very clear between signal and background. In the final signal selection, the background distribution has the triangular shape we are aiming for, with the peak at a NN score of 0. The signal is less peaked than the background, but the majority of signal events are classified as more signal-like (NN score > 0.5).

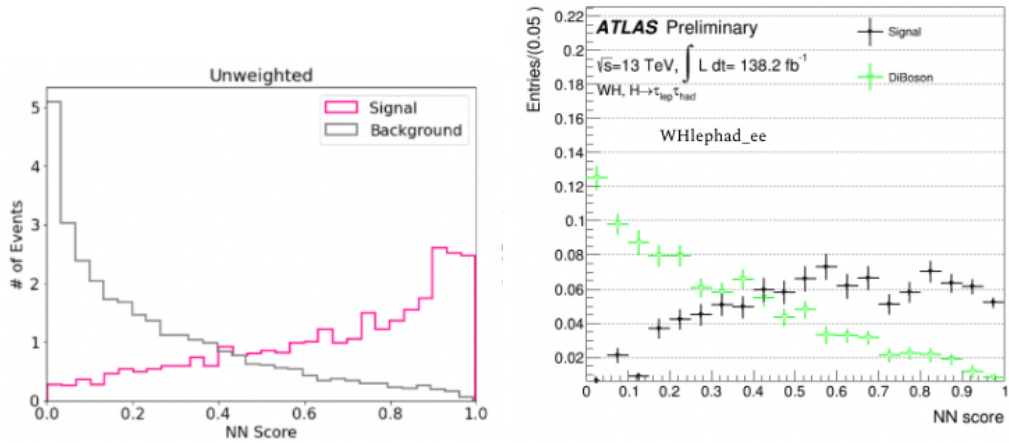


Figure 8.20: Neural network score distribution at preselection for the WHlephad_ee (di-electron channel) test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).

There were many studies performed to tune the network. First, we varied the number of nodes in

the three hidden layers, while sticking to the general guideline to have ten times more training events than trainable parameters. Trainable parameters are the weights connecting the layers/nodes and the biases at each node. Thus adding in more nodes increases the number of trainable parameters. It was seen that 40 nodes in the three hidden layers had the best AUC score on the test set (top left of figure 8.21).

Then I checked whether including more variables aimed at increasing the discrimination power of the network actually improved performance. I saw that adding higher-level variables, like the sum of the tau p_T and $\Delta\phi$ between the tau and MET improved the AUC score (as seen in the top right). I also confirmed that adding in the rotational layers improved performance as seen in the bottom left of figure 8.21. Finally, I changed the optimizer of the network to Adam[116], but performance was slightly worse than when training with SGD as the optimizer.

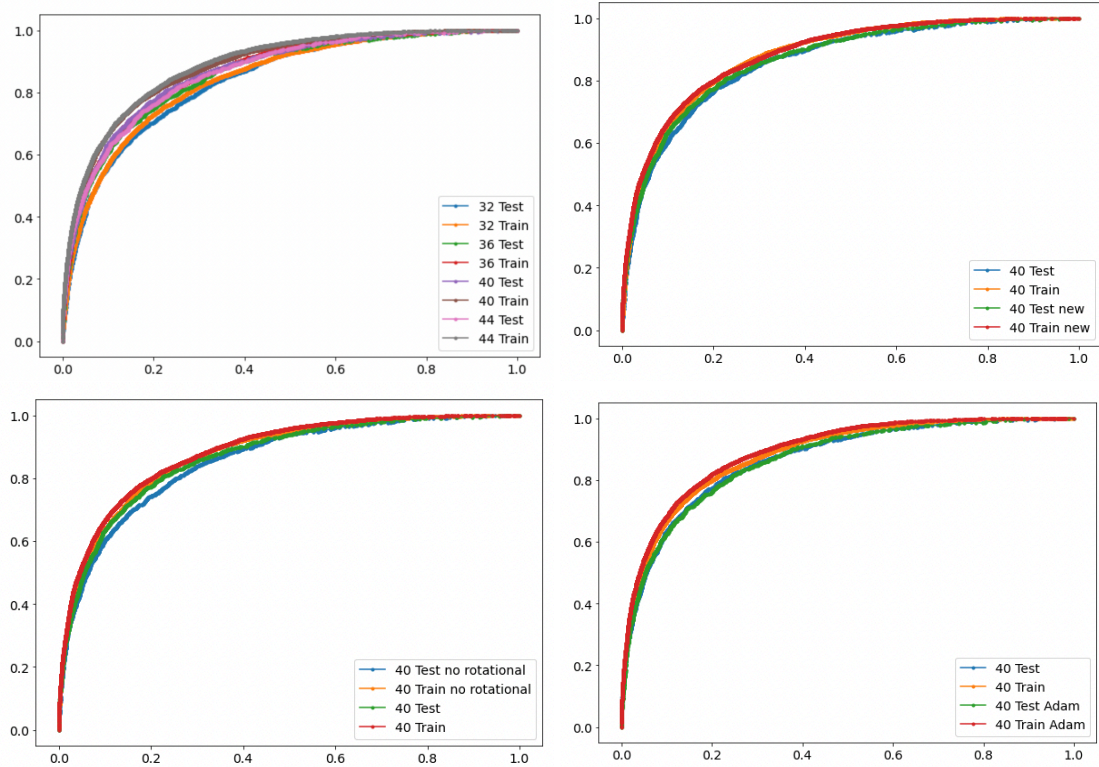


Figure 8.21: The top left plot shows the ROC curves for different numbers of nodes in the three hidden layers, with 40 nodes having the best test set performance. The top right shows the improvement to the AUC when including in the new variables for training the WHlephad_ee network. The bottom left plot shows the effect of removing the rotational layers at the start of the network. The test performance is better when the layers are included. The bottom right plot shows the performance of two optimizers: SGD (our default) and Adam on the training and test sets.

As a check to see whether the neural network was learning something beyond the Higgs boson mass discrimination M_{2T} to separate signal from background events, I did a test where I arbitrarily applied a 0.5 NN score cut, i.e. selecting the more signal-like events. I checked to see whether the M_{2T} distribution changed, or was shaped, by the neural network. If there was a dramatic change or if the distributions for signal and diboson ended up overlapping more, then the neural network could be relying on M_{2T} to separate the signal and background. However, as seen in figure 8.22, the normalized distributions of M_{2T} before and after a NN cut look very similar. This demonstrates that the neural network is learning something other than M_{2T} to separate signal from diboson background and thus is enhancing the separation power of our analysis.

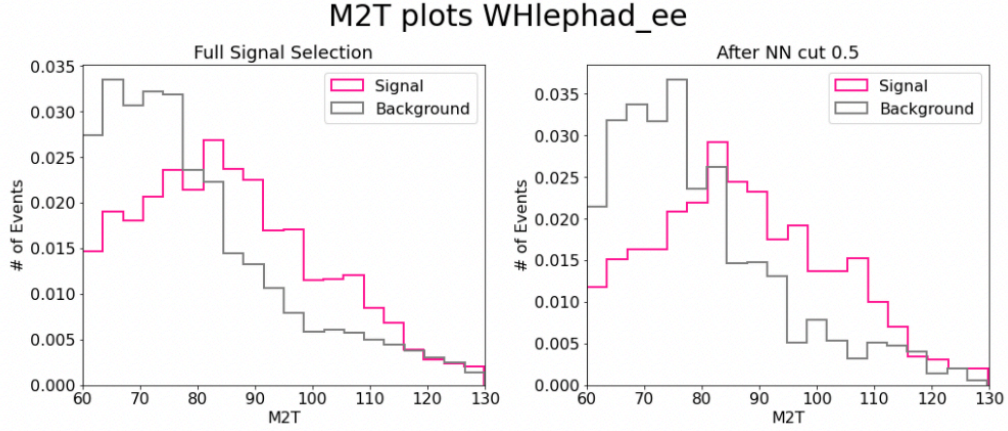


Figure 8.22: Normalized distribution of M_{2T} for signal and diboson MC for full signal selection (left) and for the events passing a NN score cut of 0.5 (right) for the WHlephad_ee category.

8.3.1.2 WHLEPHAD_EMU NEURAL NETWORK

The training set for the electron-muon final state of the WHlephad channel had 16,732 signal and 69,407 background events. We weighted the signal 4.1 times more than background. For the final optimized model, we had 58 nodes in the three hidden layers and used the optimizer SGD. The test AUC was 0.869. There was a little over-training as shown in the better performance of the training set as opposed to the validation and test sets based on their ROC curves (see figure 8.23). The validation set had an AUC of 0.871 and the training set had an AUC of 0.895. To try to remove some of the over-training effect, I doubled the validation size to 30% and used that to help tune the network, but that did not do much to improve the over-training. I also shortened the patience of the network. Patience is how long the model

will keep training after no improvement in the validation loss. A shorter patience should stop the model from over-fitting, however no major change to the difference between test and training set performance when using various levels of patience was seen. While there was slightly more over-training seen in this channel as compared to the other channels, it is not something to worry too much about based on the performance we see in our signal region. The test AUC is higher than in other channels, as well.

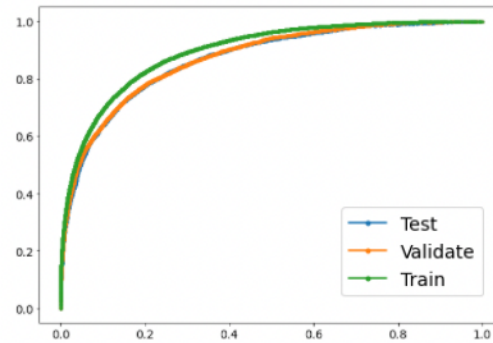


Figure 8.23: ROC curves for the training, test, and validation set for the WHlephad channel with the electron/muon final state. There is a bit of over-training occurring, but test set performance is still comparable to our other models.

Figure 8.24 shows the performance of the neural network at preselection on the test set without weights to correct for MC cross-sections on the left. The plot on the right shows the final performance of the neural network score on the signal and diboson MC that passes the full signal selection. The separation in the unweighted and normalized preselection test set is very clear between signal and background. After signal selection, the neural network score still cleanly separates signal from background with the dibosons peaking at 0 and the signal MC peaking at 1.

There were many studies performed to tune this network as well. First, I compared the performance of the neural network using an abbreviated list of training variables to the baseline one-hidden layer model (top left of figure 8.25). The optimized model outperforms the baseline model as shown by the AUC and ROC curve. Then after adding in more variables to improve discriminating power, I varied the number of hidden layers. I reduced the number of nodes in each layer so that the total number of trainable parameters stayed within our guideline. I found that the performance did not change much with the various numbers of layers and I implemented three layers in my final network.

After deciding on three layers, I varied the number of nodes in those three layers. I found that 58

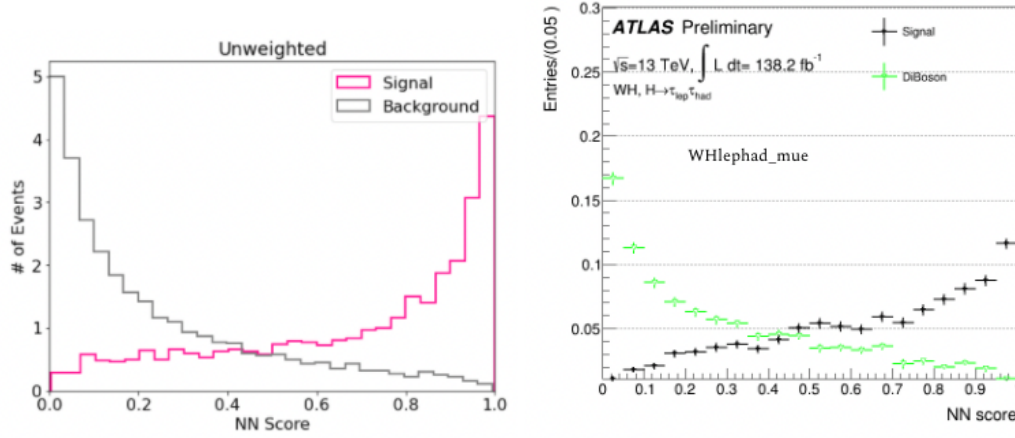


Figure 8.24: Neural network score distribution at preselection for the WHlephad_emu test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).

nodes in the three layers gave the highest AUC value. Finally, I varied the learning rate and found that $5e-4$ was the optimal one (bottom right of figure 8.25).

As a check to see whether the neural network was learning something beyond the Higgs boson mass reconstruction variable M_{2T} , I did a test where I arbitrarily applied a 0.5 NN score cut, i.e. selecting the more signal-like events. I checked to see whether the M_{2T} distribution changed, or was shaped, by the neural network. If there was a dramatic change or if the distributions for signal and diboson ended up overlapping more, then the neural network could have been primarily using M_{2T} to separate the signal and background. However, as seen in figure 8.26, the normalized distributions of M_{2T} before and after a NN cut look similar, though the NN cut did remove a few of the low M_{2T} background events. This demonstrates that the neural network is learning something other than M_{2T} to separate and thus is enhancing our separation power.

8.3.1.3 WHLEPHAD_MUMU NEURAL NETWORK

The training set for the initial training of the mumu final state of the WHlephad channel had 9,441 signal and 208,597 background events. We weighted the signal 22.1 times more than background. For the optimized model, we had 96 nodes in the three hidden layers and used the optimizer SGD. The test AUC was 0.939. The optimized model, before additional variables were added, performs much better than the baseline one-layer model (left side of figure 8.27).

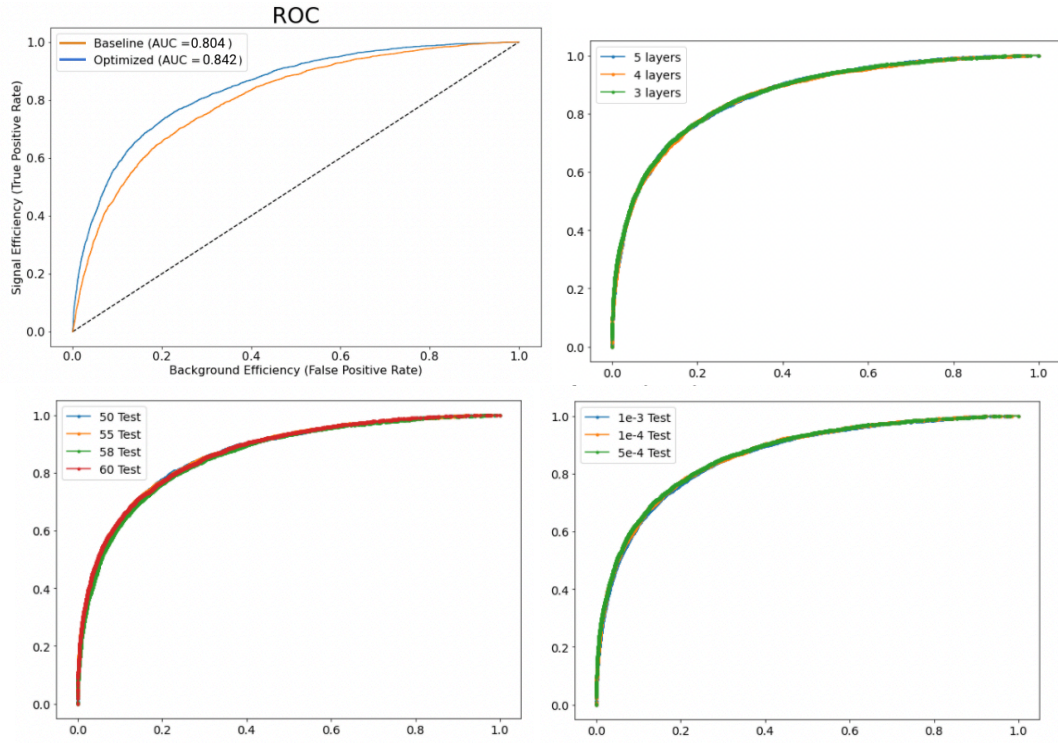


Figure 8.25: The top left plot shows the ROC curves for the baseline model and the optimized model with an abbreviated variable list for the WHlephad_emu channel. The top right shows the similar performance of the model when trained with different numbers of hidden layers. The bottom left plot shows the minimal effect of varying the number of nodes in the three layer model. The bottom right plot shows the similar performance of the model with various learning rates.

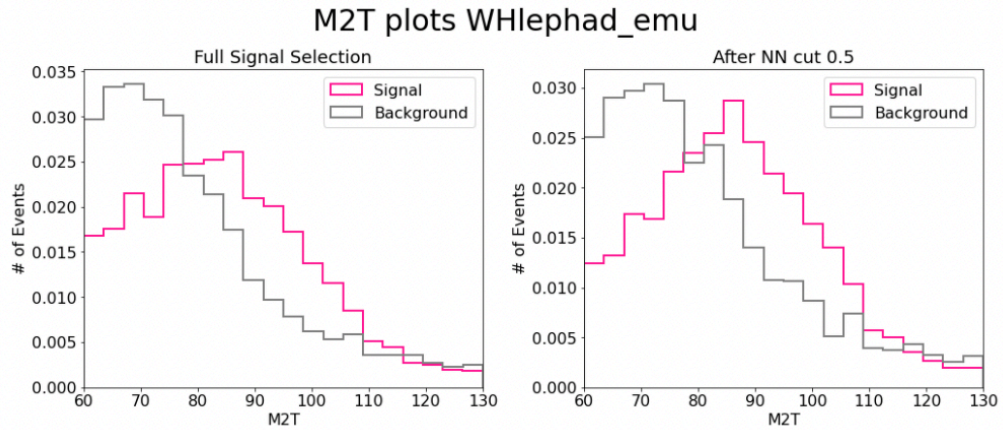


Figure 8.26: Normalized distribution of M_{2T} for signal and diboson MC for full signal selection (left) and for the events passing a NN score cut of 0.5 in the electron-muon channel of WHlephad (right).

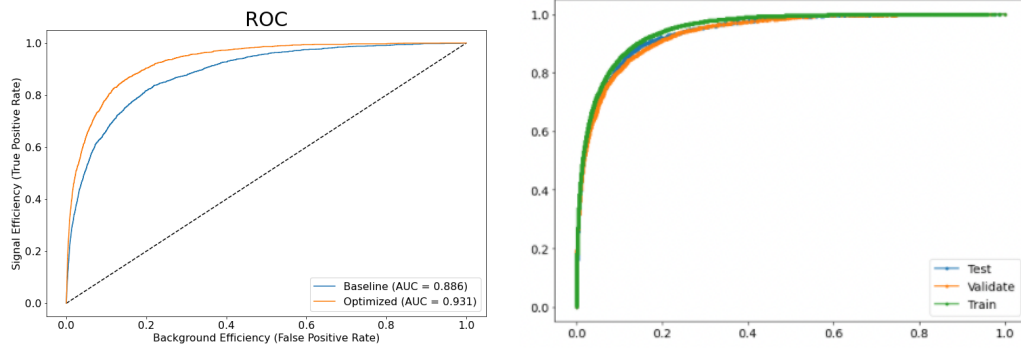


Figure 8.27: Comparison of the baseline one-layer model and optimized three-layer model (right). ROC curves for the training, test, and validation set for the WHlephad channel with the di-muon final state (right) after adding in more discriminating variables.

The performance at preselection was good for this model, however, once the full signal selection was applied, the separation power was not as good. The background distribution is flat for the NN score as shown in figure 8.28. To solve this, I added an additional cut to the training set that requires same-sign leptons. By including this in training and evaluation, the model needs to learn other things to separate signal from background.

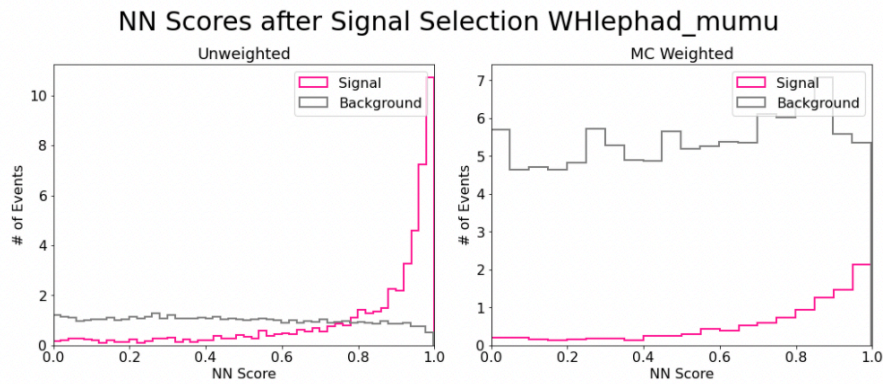


Figure 8.28: NN scores on unweighted (left) and weighted (right) distributions of signal and diboson Monte Carlo for the di-muon WHlephad channel. While the signal distribution peaks at 1, the background distribution for NN score is flat.

After applying the same-sign leptons selection, we have 4,474 signal and 18,388 background events. We weight the signal 4.1 times more than the background. The optimized model has 26 nodes per layer. We get a final test AUC of 0.863 and the three ROC curves for train, test, and validation can be seen in figure 8.29.

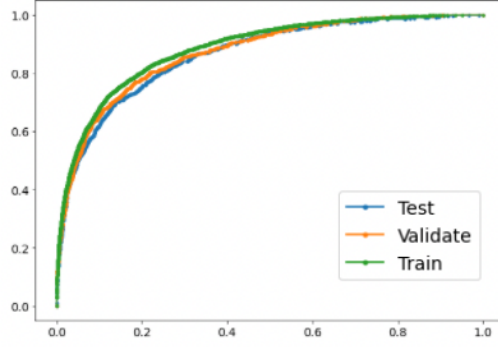


Figure 8.29: ROC curves for the training, test, and validation set for the WHlephad channel with the di-muon final state after the same-sign lepton selection is performed. Lower training set sizes lead to the noise seen in the curves.

Figure 8.30 shows the performance of the neural network at preselection on the test set without weights to correct for MC cross sections on the left. The plot on the right shows the final performance of the neural network score on the signal and diboson MC that passes the full signal selection. The separation in the unweighted and normalized preselection test set is very clear between signal and background. After signal selection, the neural network score still cleanly separates signal from background with the dibosons peaking at 0 and signal peaking at 1.

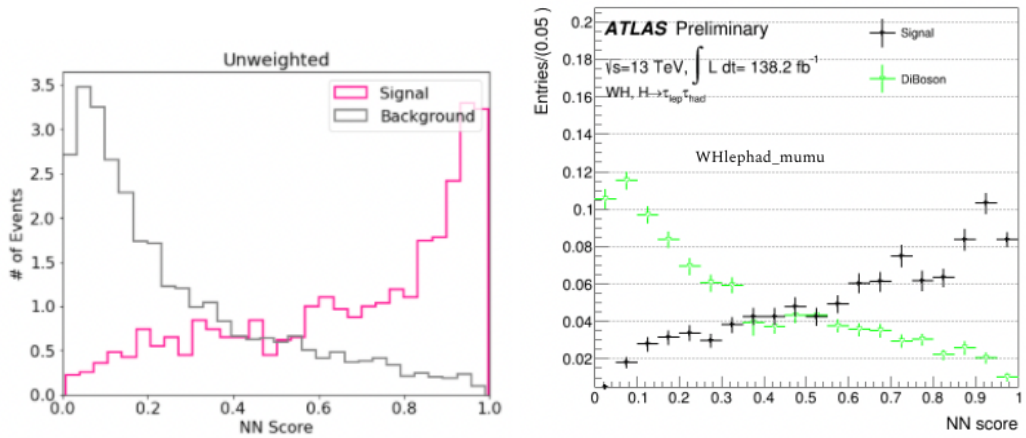


Figure 8.30: Neural network score distribution at preselection for the WHlephad_mumu test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).

As a check to see whether the neural network was learning something beyond the Higgs boson mass discrimination variable M_{2T} , I did a test where I arbitrarily applied a 0.5 NN score cut, i.e. selecting

the more signal-like events. I checked to see whether the M_{2T} distribution changed, or was shaped, by the neural network. If there was not a dramatic change in the distributions, then the NN is able to tease discriminating information out beyond M_{2T} to separate the signal and background. As seen in figure 8.26, the normalized distributions of M_{2T} before and after a NN cut look very similar. This demonstrates that the neural network is learning something other than M_{2T} to separate signal from the diboson background and thus is enhancing our separation power.

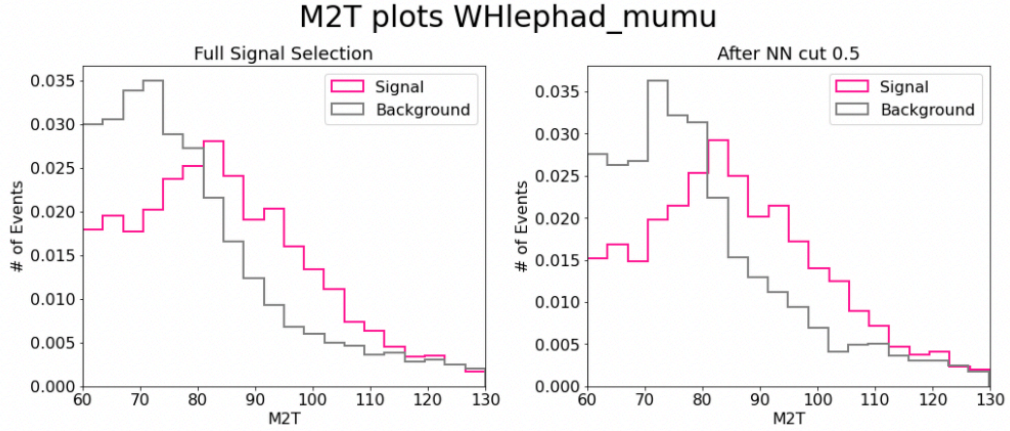


Figure 8.31: Normalized distribution of M_{2T} for signal and diboson MC for full signal selection (left) and for the events passing a NN score cut of 0.5 for the di-muon WHlephad channel (right).

8.3.2 $W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$ NEURAL NETWORK

The WHhadhad channel NN was trained by (at the time) Yale graduate student Mariel Pettee. The training set had 19,873 signal events and 41,955 diboson events. Class weights of 2.1 were used to weight the signal more than the background during the training. The optimized network had 128 nodes in each of the three hidden layers. The network achieved an AUC (area under curve) score of 0.877.

Figure 8.32 shows the performance of the neural network at preselection on the test set without weights to correct for the MC cross sections on the left. The plot on the right shows the final performance of the neural network score on the signal and diboson MC that pass the full signal selection. This also shows a nice separation between signal and background with both peaking in their respective sides, 0 for background and 1 for signal. While the signal does not peak exactly at 1, it is still peaking on the correct side and very few signal events are classified as being background like (< 0.5 NN score).

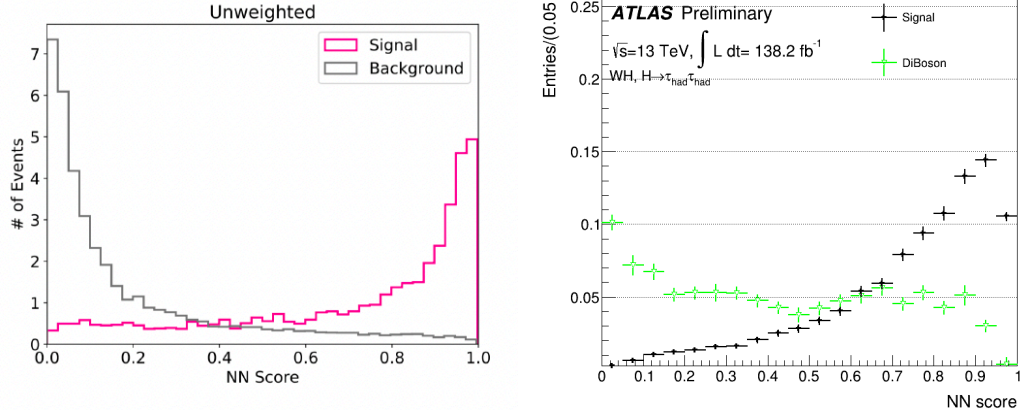


Figure 8.32: Neural network score distribution at preselection for the WHhadhad test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).

8.3.3 $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$ NEURAL NETWORK

I retrained the network for ZHlephad so that we could investigate the impact of training when removing the events that end up in the overflow bin for the dilepton mass variable. We do not cut on dilepton mass in the preselection cutflows that are used to train the network, but do remove those errant events in our signal selection. Two networks were optimized concurrently, one with the entire dilepton mass window and one with a mass window of $20 - 140\text{GeV}$. With the mass window applied we had 15,834 signal and 60,377 background events in the training set. We weighted the signal 3.8 times more than the background and aimed for 7,621 trainable parameters. For the full (no mass selection) training set, we had 20,210 signal and 71,287 background events. We weighted the signal 3.5 times more than the background and aimed for 9,149 trainable parameters. The optimized model for the mass window had 55 nodes per layer and the full selection has 58 nodes per layer.

To compare the two models' performances, we can look at the AUC for their train and test sets. Figure 8.33 shows that the test and train performance of the mass window is slightly better than the full selection. The test AUC was 0.842 for the full data set and 0.863 for the mass window selection. However, if we look at the performance of the model at separating signal from background in our final selection, the opposite result emerges. The full window has a mean NN score for the signal of 0.813 and 0.410 for the background, while the mass window selection has an average NN score of 0.804 for the signal and 0.420 for the background. The mass window does a worse job of separating the signal from background as seen

with the closer average scores and the distributions shown in figure 8.34

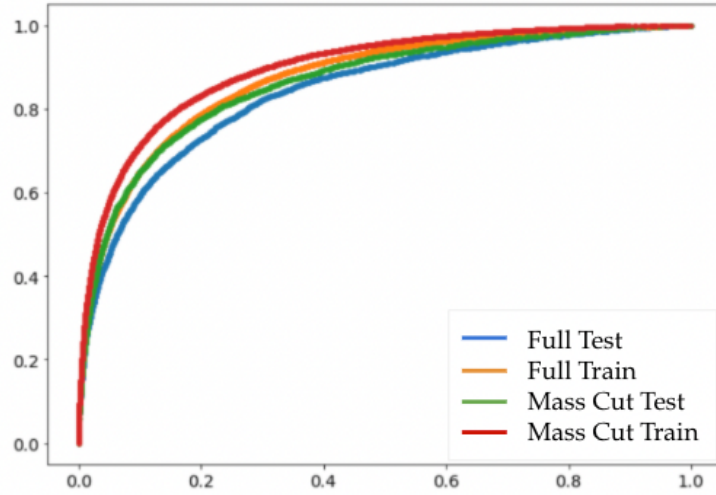


Figure 8.33: The ROC curves for the training and test sets for the ZHlephad neural networks trained on the full data set and the selected mass window.

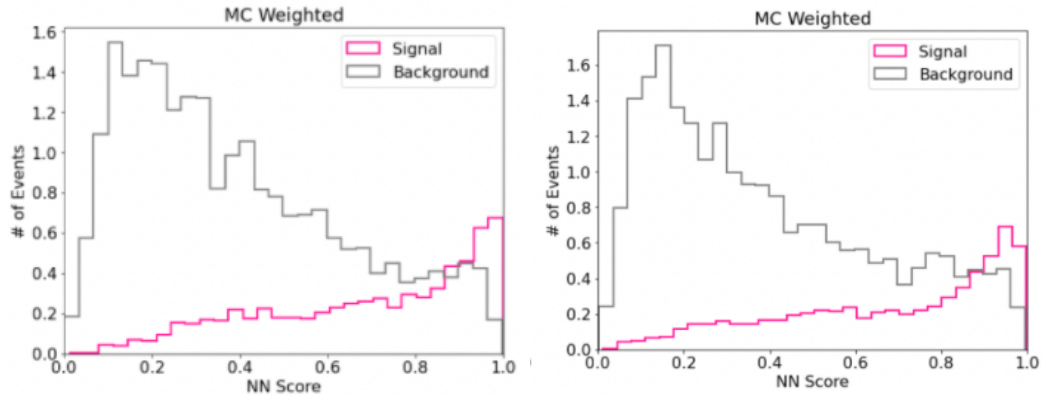


Figure 8.34: Neural network score distribution at preselection (normalized and unweighted) for the ZHlephad test set on the full mass range (left) and neural network score for the mass window training set for signal and diboson MC after preselection (right).

Figure 8.35 shows the performance of the neural network at preselection on the test set without weights to correct for MC cross sections on the left. The plot on the right shows the final performance of the neural network score on the signal and diboson MC that passes the full signal selection. This also shows a nice separation between signal and background with both peaking in their respective sides, 0 for background and 1 for signal. While the background peaks at around 0.2 and not 0, it is still peaking on the correct side. Meanwhile, the signal peaks at 1 very cleanly and dwarfs the background in that region.

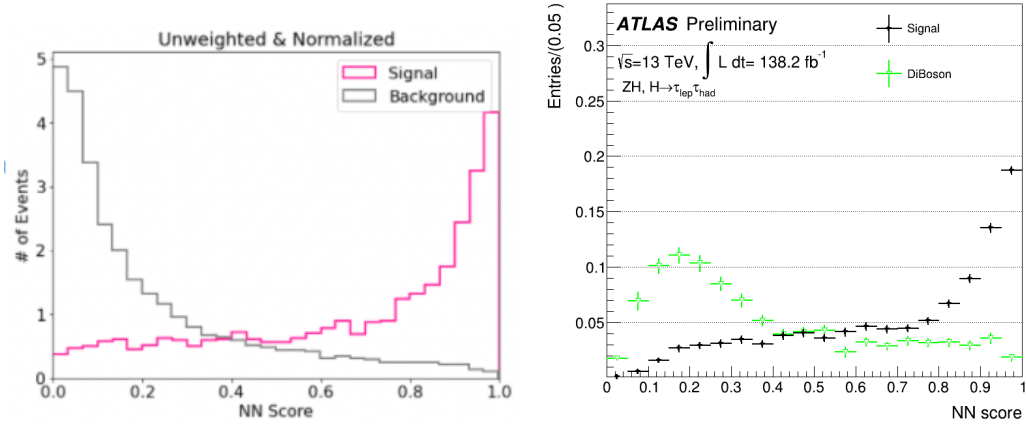


Figure 8.35: Neural network score distribution at preselection for the ZHlep_{had} test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).

8.3.4 $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$ NEURAL NETWORK

The ZHhadhad channel NN was trained by (at the time) Yale graduate student Mariel Pettee. The training set had 15,215 signal events and 51,508 diboson events. The signal was weighted 3.4 times more than the background during the training. The optimized network had 50 nodes in each of the three hidden layers. The network achieved an AUC (area under curve) score of 0.899.

Figure 8.36 shows the performance of the neural network at preselection on the test set without weights to correct for MC cross sections on the left. The plot on the right shows the final performance of the neural network score on the signal and diboson MC that pass the full signal selection. This shows a nice separation between signal and background with both peaking in their respective sides, 0 for background and 1 for signal.

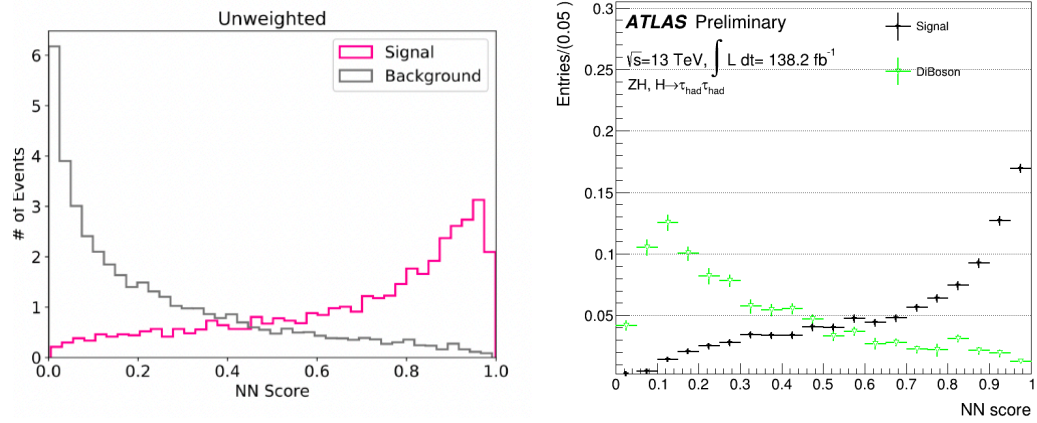


Figure 8.36: Neural network score distribution at preselection for the ZHhadhad test set, not weighted by MC weights (left) and neural network score for full signal and diboson MC after signal selection weighted by all scale factors (right).

“I want to do something splendid...something heroic or wonderful that won’t be forgotten after I’m dead. I don’t know what, but I’m on the watch for it and mean to astonish you all someday.”

“I’d rather take coffee than compliments just now.”

Louisa May Alcott *Little Women*

9

Analysis Sensitivity Projections

WHILE THE ANALYSIS FOR THE $VH, H \rightarrow \tau\tau$ MEASUREMENT IS STILL ONGOING, WE HAVE A FEW WAYS TO PREDICT WHAT WE EXPECT TO SEE WITH THE DATA FROM RUN 2. The search from Run 1 had an observed combined signal strength of $\sigma/\sigma_{SM} = 2.3 \pm 1.6$, and set upper limits on cross sections for each category [42]. We can extrapolate from the Run 1 results to create a Run 2 prediction. In addition, we can estimate an upper bound on the number of events produced in our channels using the SM cross-sections and branching ratios for our events. Finally, we can use the Run 2 simulation to estimate the number of events that pass our selection, and Asimov data sets to make sensitivity predictions.

9.1 RUN 1 RESULTS

Table 9.1 shows the total observed counts in Run 1 data, and then how that breaks down in to signal, background from mis-identified objects (fakes), and dibosons categories. The final column is the observed significance. The WH channels had the highest yield, but they also had the most background. The highest significance channel was the $Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$.

Channel	Observed	Signal	Fakes	Diboson	Significance
$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$	35	1.95 ± 0.05	13.1 ± 1.3	13.54 ± 0.35	0.44σ
$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$	33	1.84 ± 0.04	28.1 ± 2.4	7.4 ± 1.2	0.60σ
$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$	24	1.14 ± 0.03	17.1 ± 1.5	7.28 ± 0.16	0.29σ
$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$	7	0.64 ± 0.02	4.7 ± 1.2	2.09 ± 0.09	1.38σ

Table 9.1: Results from the Run 1 analysis for $VH, H \rightarrow \tau\tau$, including final observed counts (broken down into signal, diboson, and fake contributions) and the observed significance. [42]

For a simple estimate of what we would expect to see in Run 2 for our signal, we can scale up from the observed signal counts from Run 1. The luminosity increased from $20.3fb^{-1}$ in Run 1 to $139fb^{-1}$ in Run 2, giving a scale factor of 6.85. The increase in center of mass energy to 13TeV from 8TeV also increases the cross-section for Higgs boson production. The cross-section for a 125GeV Higgs boson produced via WH , increased to 1.373pb from 0.7065, leading to a scale factor of 1.94 for the WH channels. Similarly, the ZH production mode cross section increased to 0.8839pb from 0.4208, leading to a scale factor of 2.1 for the ZH channels. Thus, we would expect to see 13.29 times more WH events (25.9 for $WH_{lep}had$ and 24.4 for $WH_{had}had$), and 14.39 times more ZH events (16.4 for $ZH_{lep}had$ and 9.2 for $ZH_{had}had$) than in Run 1. This approximation does not account for different trigger efficiencies, the removal of the opposite-flavor lepton restriction in the $WH_{lep}had$ channel, and other improvements to the analysis including the neural network targeting dibosons.

Another way we can estimate the potential reach of our analysis is to produce an estimate of produced events using Standard Model predictions of cross-sections and branching ratios. This serves as a sort of upper bound since it would assume we perfectly select every expected signal event in our detector, while our actual event acceptance is on the order of a percent. The equation is:

$$N_{VH} = \mathcal{L}_{int} * \sigma_{VH} * BR(V \rightarrow leptons) * BR(H \rightarrow \tau\tau) \quad (9.1)$$

where $\mathcal{L}$, the integrated Luminosity is $139fb^{-1}$ and the σ_{VH} are 0.8839pb for ZH and 1.373pb for WH . The branching ratio of $Z \rightarrow ll$ is 0.0336 for each of the ee and $\mu\mu$ channels (0.067 combined), and the branching ratio of $W \rightarrow l\nu$ is 0.105 for either $e\nu$ or $\mu\nu$ (0.201 combined). It is possible for events to enter our analysis from the leptonic decays of the W and Z to taus where the taus either go on to decay to leptons or could be tagged as hadronic taus in the events. However, these objects have lower p_T than the

prompt objects so should not affect my final prediction values too much. Finally, the branching ratio of $H \rightarrow \tau_{had}\tau_{had}$ is $0.4225 * 0.06272$ and $H \rightarrow \tau_{lep}\tau_{had}$ is $0.455 * 0.06272$. Putting all of this together, we get estimates of 1198 events for WHlep τ_{had} , 1120 events for WHhad τ_{had} , 23.5 events for ZHlep τ_{had} , and 22 events for ZHhad τ_{had} .

The third way I will show for predicting the Run 2 expectations is using the signal MC. Table 9.2 shows the predictions using the projection from Run 1, the Standard Model theory and Run 2 Luminosity, and the Run 2 MC. As you can see, the SM model is an upper bound and the Run 1 projection should serve as a lower bound that our Run 2 MC is in between, with the exception of the prediction for ZHlep τ_{had} that needs further investigation.

Channel	Run 1 Projection	Total SM Events	Run 2 Signal MC
$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{lep}\tau_{had}$	25.9	1144	33.44 ± 0.42
$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{had}\tau_{had}$	24.4	1062	24.36 ± 0.34
$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{lep}\tau_{had}$	16.4	236	$7.66^* \pm 0.11$
$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{had}\tau_{had}$	9.2	219	10.98 ± 0.14

Table 9.2: Predictions for signal counts for each of the channels in the Run 2 analysis. The SM Prediction will serve as a higher bound as it does not take into account our analysis acceptance. We need to further investigate the unexpectedly low Run 2 Signal MC prediction for ZHlep τ_{had} .

9.2 HYPOTHESIS TESTING

The goal of this analysis is to observe the $VH, H \rightarrow \tau\tau$ channel. To do this we need to conduct an experiment with a null hypothesis (no Higgs boson in this channel) and also run our analysis with the SM hypothesis for our signal. We can think of our expected events n as $n = \mu s + b$ where μ is the signal strength, s is the number of signal events, and b is the number of background events. The null hypothesis, which we designate with H_0 , refers to the background-only hypothesis. This null hypothesis means that we expect to see all SM physics processes with the exception of our signal. Here μ , the signal strength, is equal to 0 and our expected event count is simply the number of background events b . Our hypothesis, H_μ , is then where $\mu = 1$, so our expected events should then be $s_{SM} + b$. Further background information on these topics and for the following sections can be found in [138] and [139].

To test our hypothesis, we use a binned histogram of the NN score in each of our signal channels. The

number of counts (n_i) in each bin (i) is expected to be:

$$E[n_i] = \mu s_i + b_i$$

where the SM prediction, $\mu = 1$, and the null, $\mu = 0$, are tested by varying values of μ until we have the best fit for our data.

As we are doing a counting experiment with very large N (total events) and very small p (probability of signal events) over a fixed window of time with a known probability of an event occurring, which are independent of one another, we can use Poisson statistics to get our probability for the expected number of events:

$$Pois(n_i, E[n_i]) = \frac{E[n_i]^{n_i} e^{-E[n_i]}}{n_i!}. \quad (9.2)$$

9.2.1 LIKELIHOOD FUNCTION

The procedure we will follow in this analysis is to calculate a frequentist significance using a likelihood ratio as a test statistic. This likelihood will depend on the nuisance parameters in addition to the rate of the signal and background processes.

If x , the observable, is frequentist in nature, replicating the experiment many times leads to different values of x . This creates a probability density function:

$$\int f(x) dx = 1.$$

Now, if we have discrete quantities, like the counts of events, this integral is replaced by a sum. In our experiment, we need to look at $f(x|\alpha)$, where α is the hypothesis and x is the set of measurements. $\alpha(\mu, \theta)$ then depends on both our parameter of interest μ and nuisance parameters θ . We can then use this α to calculate our likelihood function $\mathcal{L}(x|\alpha)$. To recap, the probability density function is a function of x given a fixed α , and the likelihood function is a function of α given fixed x .

Since we are doing a binned likelihood, the final likelihood function is constructed from the individual probabilities for each bin: $\mathcal{L} = \prod_{i=1}^N \mathcal{L}_i$. We can take our Poisson probability equation from above and

incorporate our dependence on both μ and θ to get a final expression for our binned likelihood:

$$\mathcal{L}(\mu, \theta) = \prod_{i=1}^N \frac{(\mu s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-(\mu s_i(\theta) + b_i(\theta))} \quad (9.3)$$

The nuisance parameters for our analysis include systematic uncertainties from theory predictions and working point efficiencies for identification, energy scales, triggers, etc. These will be detailed in section 9.3.

We calculate this likelihood for all six of our orthogonal channels individually. The final overall likelihood is obtained by multiplying the individual likelihoods. By performing a maximization of the likelihood function, we determine the best estimates for μ and θ . This is usually done via minimizing the negative log-likelihood.

We then define a profile likelihood ratio (λ), that is only dependent on μ . We can obtain an approximate dependence on θ by profiling the nuisance parameters and defining them in terms of the parameter of interest. The Neyman-Pearson lemma states that the likelihood:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\mu, \hat{\theta})} \quad (9.4)$$

is the optimal discriminator for testing hypothesis H_μ vs. H_0 . $\hat{\hat{\theta}}$ denotes the conditional maximum-likelihood estimators of the nuisance parameters, otherwise known as the value for the θ s that maximizes $\mathcal{L}$ for the specified value of μ . The denominator is the unconditional likelihood, where $\mathcal{L}$ is maximized via a scan of both μ and θ .

$\lambda(\mu)$ can vary from 1, meaning there is maximum compatibility between the data measurement and the hypothesis H_μ , and 0 meaning no compatibility.

9.2.2 TEST STATISTICS

The test statistics needed for a statistical test of compatibility are defined from $\lambda(\mu)$. A test statistic is then defined as:

$$t_\mu = -2 \ln \lambda(\mu) \quad (9.5)$$

where higher values of t_μ represent more incompatibility between data and hypothesis. A p-value is calculated from t_μ by:

$$p_\mu = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu) dt_\mu \quad (9.6)$$

where $f(t_\mu|\mu)$ is the probability density function of t_μ with μ as the assumed signal strength. This helps us quantify the level of disagreement between the hypothesis and the data. Smaller p-values are greater evidence against the corresponding hypothesis.

The test statistic for making a discovery comes from $\lambda(\mu)$ as well. Here we need to reject the null hypothesis of $\mu = 0$. The test statistic is defined as:

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \quad (9.7)$$

where $\lambda(0)$ is the profile likelihood ratio for $\mu = 0$. Because a discovery will only be claimed if we have more events than the background hypothesis predicts (i.e. $\hat{\mu} > 0$), we set q_0 to 0 for $\hat{\mu} < 0$. As the number of events measured above expected background increases, the value of q_0 also increases, which corresponds to a higher level of incompatibility between the data and the null hypothesis.

We can get a corresponding p-value for this test statistic:

$$p_0 = \int_{q_{0,obs}}^{\infty} f(q_0|0) dq_0 \quad (9.8)$$

where $f(q_0|0)$ is the distribution of test statistics. According to Wilks' Theorem, for large statistics data samples (even on the $\mathcal{O}(10)$ events in a sample), we know the distribution of the test-statistic: a χ^2 distribution with a parabolic shape around the minimum. This means we can directly calculate the significance p_0 and get an uncertainty on μ [140].

We can then convert this p-value into an equivalent significance,

$$Z = \Phi^{-1}(1 - p) \quad (9.9)$$

which is a Gaussian distributed variable with Φ^{-1} as the quantile of the Gaussian distribution. Z repre-

sents the number of standard deviations away our observed signal strength is from the null hypothesis. In particle physics, we consider a Z of 5 to represent a “discovery” (p-value around 2.87×10^{-7}) and a Z of 3 to signify an observation (p-value of 0.03).

9.3 SYSTEMATICS

In this section I discuss some of the systematic uncertainties considered in our final fit result.

9.3.1 EXPERIMENTAL UNCERTAINTIES

The uncertainties for experimental quantities are estimated by following the recommendations provided by dedicated groups within ATLAS that study the identification of various Standard Model particles. Each uncertainty is varied within the recommended $\pm 1\sigma$. The uncertainty on the integrated luminosity for all four years combined is $\pm 1.7\%$ [141].

Muon and Electron Uncertainties

The estimate of the values for the muon uncertainties are described in [103] and the electron uncertainties are in [27]. They are mainly calculated via tag-and-probe techniques that use Z bosons decaying to electron or muon pairs in order to select unbiased electrons or muons from the data. These values include shifts based on the identification algorithm, energy scale calculations, reconstruction efficiencies, isolation efficiency, and trigger uncertainties.

Hadronic Tau Uncertainties

The tau uncertainty estimates are calculated as described in [40]. The uncertainties associated to the tau identification and energy scale are found with the tag-and-probe technique that uses Z bosons decaying to tau pairs in order to select unbiased taus from the data. The algorithm’s performance is evaluated in a data sample enriched in $Z \rightarrow \tau_\mu \tau_{had}$, where one tau decays leptonically to a muon and the other decays hadronically. The events are then triggered by the muon, the tag particle, and contain the tau, the probe particle. The identification efficiency is calculated by performing a template-fit based on track multiplicity. The main background for the tau signal is quark- and gluon-initiated jets that are misidentified as taus. The ratio of signal events before and after the identification criteria are applied is used to calculate the efficiency

of each working point. The energy scale systematic uncertainty is derived from a χ^2 fit of the visible mass in the tag-and-probe events and from the difference between the energy and momenta measurements. The effect of the systematic uncertainty for the tau energy scale on the shapes of the signal and ZZ MC for the ZHlephad channel are shown in figure 9.1. The uncertainty from the eVeto is calculated in a $e - \tau$ tag-and-probe analysis by comparing the efficiencies for the eBDT identification algorithm.

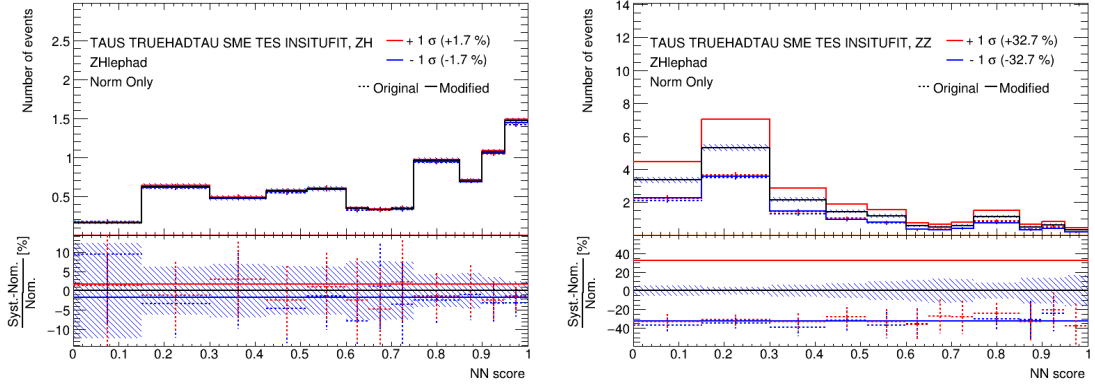


Figure 9.1: Plots of the NN score distribution in the ZHlephad with the up, down, and nominal variations of the tau energy scale uncertainties. The left plot shows the shape variation in the signal MC and the right shows the shape variation in the simulation of the ZZ background.

Jet and b-tagging Uncertainties

Uncertainties for jets come from recommendations by the jet performance group in [142]. These include uncertainties for energy scale and energy resolution, along with jet flavor and vertex tagging. While we do not select jets directly in our analysis, the jet resolution, and energy scale are included in the calculations for missing transverse energy. There is also a dedicated uncertainty for b-tagging efficiency related to the various flavors of jets.

Missing Transverse Energy Uncertainties

Missing transverse energy (MET) is calculated using clustered energy in the form of jets, electrons, muons, and taus, which is designated as the “hard-terms”, and unclustered energy, which is designated as the “soft-term”. We include uncertainties related to the soft-term of MET based on calculations described in [143]. Additionally, the energy scale and resolution of the objects included in the MET calculation, such as electrons, muons, taus, and jets, impact the final value for the hard component of the MET calculation.

9.3.2 DATA-DRIVEN FAKE UNCERTAINTIES

The uncertainties related to misidentified objects (the fakes) include a statistical uncertainty and an uncertainty related to varying the MC subtraction, as described in section 7.3. The effect of the MC subtraction systematic uncertainty is shown for the WHhadhad and WHlep had di-electron channels in figure 9.2. Additionally, we derived a closure uncertainty from the same-sign taus region for all channels except WHlep had, for which we derived the uncertainty in the all-same-sign region. We chose these regions as they have very high purity of fakes and strongly resemble the signal region, while being orthogonal due to the same-sign requirement. We create just two bins of the neural network score to reduce statistical fluctuations, where we have 80% of the signal in the high-side bin. The ratio of the data and background counts in these two bins is used to create the 1σ closure uncertainty. Figure 9.3 shows the distribution of the neural network score in two bins in the same-sign tau control region for the ZHhadhad channel and the ratio plot used to calculate the fake-closure uncertainty.

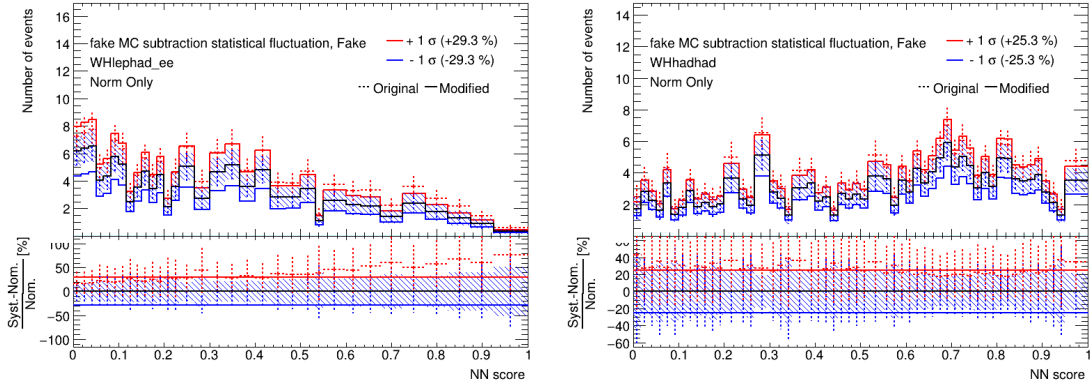


Figure 9.2: Plots of the NN score distribution in the WHlep had di-electron channel (left) and WHhadhad channel (right) with the up, down, and nominal variations of the fake MC subtraction on the background component from misidentification of jets.

9.3.3 THEORY UNCERTAINTIES

Theory uncertainties, arising from uncertainties in values used in our MC modeling and theories, such as α_S (fine-structure constant) and the parton distribution function (PDF) used in simulation, are relevant to include in both our signal and background modelling.

Signal and Higgs Boson Background Uncertainties

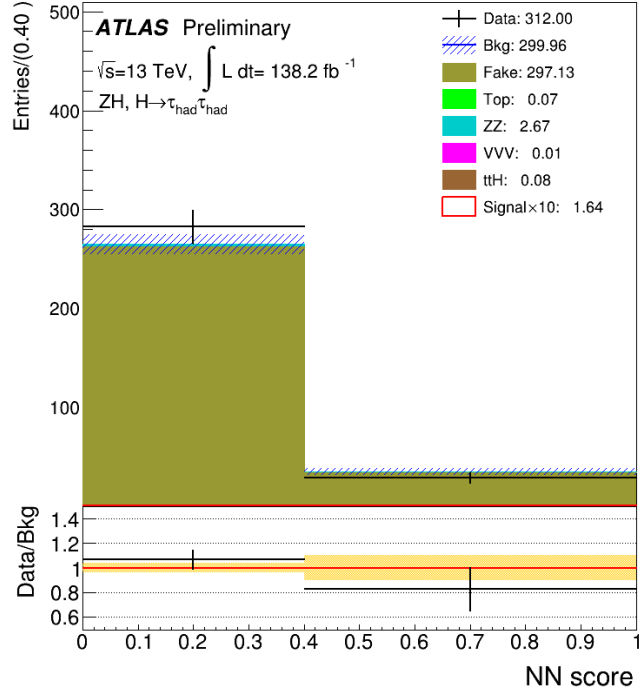


Figure 9.3: Distribution of neural network score with two bins in the same-sign tau control region for the $ZH, H \rightarrow \tau_{had}\tau_{had}$ channel with ratio plot of data/background. This is used to calculate the closure uncertainty for the fake background contribution.

The main contributions to our uncertainties in measuring the Higgs boson components in our signal are the QCD scale uncertainty, non-perturbative components such as the underlying-event, and the uncertainty on experimental input parameters such as α_S and PDFs. In our fit model, we are counting the number of events in each bin of our neural network score distribution. This means our uncertainty comes from the variance of the events (n_{tot}) in each bin:

$$n_{tot} = \sigma_{tot} \times \mathcal{L} \times A \times f_{bin} \quad (9.10)$$

where σ_{tot} is the total cross-section, $\mathcal{L}$ is the Luminosity, A is the acceptance of the total sample in the signal region, and f_{bin} is the fraction of events in a certain bin in the final distribution.

The uncertainties for the cross-section are provided by the LHC Higgs Cross Section Working Group [43]. One of the components of the cross-section uncertainty is the QCD scale uncertainty. We have uncertainty here due to missing higher order terms in our simulations. These uncertainties will be larger for

the ZH channels due to the presence of the one-loop $gg \rightarrow ZH$ contribution. The QCD uncertainties are estimated by varying the input values for the factorization scale (μ_F) and renormalisation scale (μ_R), first introduced in section 3.1.1.1. The default values are set to the Higgs boson mass, $\mu_R = \mu_F = 1$, and these are varied independently such that $0.5 < \mu_R, \mu_F < 2.0$. We vary these values in the MC and measure the impact on the signal acceptance by setting the systematic uncertainty as the greatest deviation from the default values. For the VH and $t\bar{t}H$ processes, the MC samples are generated with Powheg+Pythia8 at next-to-leading order (NLO) with the internal weights of Powheg being used for the scale variations.

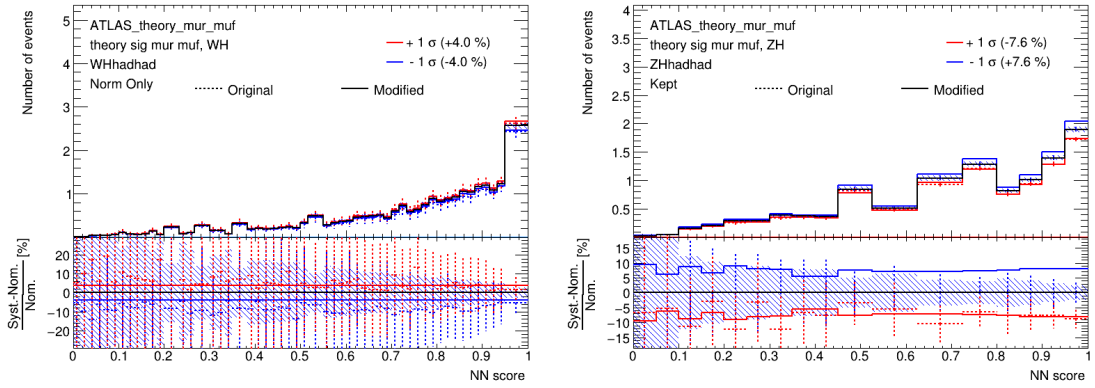


Figure 9.4: Plots of the NN score distribution in the WHadhad channel (left) and ZHhadhad channel (right) with the up, down, and nominal variations of the theory uncertainty from varying μ_R and μ_F on the contribution from the signal MC.

The other main component to the cross-section uncertainty is from the PDFs, which are not known precisely. Because we are scattering protons, we are actually colliding their constituents (quarks, gluons) that carry an unknown proportion of the proton momenta. As a reminder, PDFs predict the probability of gluons and quarks participating in the hard in-elastic scatter and they depend on the momentum transfer Q and the fraction x of the proton momentum the quark or gluon carries. The uncertainties again come from the LHC Higgs Cross Section Working Group. They are calculated by using three different PDF sets, known as CT10, NNPDF, and MSTW. CT10 is set as the nominal distribution, and the uncertainty is the quadrature sum of the differences between the nominal set and the variation sets, while also incorporating two variations of α_S . The uncertainties due to the QCD scale and PDF in our Higgs boson contributions are shown in tables 9.3 and 9.4.

The final uncertainty in our signal comes from the contributions of the underlying-events and hadronization. At ATLAS, high transverse momentum is a key component of our physics events of interest, however,

Production Process	+ QCD scale	- QCD scale
WH	+0.5%	-0.7%
ZH	+3.8%	-3.1%
$t\bar{t}H$	+5.8%	-9.2%

Table 9.3: Total cross-section uncertainties due to missing higher orders (QCD scale variations) from [43].

Production Process	PDF	α_s	PDF + α_s
WH	$\pm 1.7\%$	$\pm 0.9\%$	$\pm 1.9\%$
ZH	$\pm 1.3\%$	$\pm 0.9\%$	$\pm 1.6\%$
$t\bar{t}H$	$\pm 3.0\%$	$\pm 2.0\%$	$\pm 3.6\%$

Table 9.4: Total cross-section uncertainties due to PDFs and α_s from [43]

we also have soft interactions occurring in our detector. These can be from sources such as initial and final state radiating and remnants from the beam. In addition, there can be multi-parton interactions that overlap. These contributions are hard to model in MC, and thus lead to uncertainty in the MET resolution and momentum calculation for the Higgs boson. We also chose a certain model for the parton shower in our simulation, which leads to potential uncertainties in our signal. Both the uncertainty from the underlying-event and hadronization are derived by generating signal samples with the same matrix-element generator but different parton-shower models.

Background Uncertainties

For our diboson backgrounds, uncertainties similarly come from the renormalization and factorization scales (μ_R, μ_F) and α_S , which are varied in the same way as described in the signal uncertainties, except using the event weights provided by the SHERPA generator. We also look at the PDF-alternating value by comparing diboson acceptance using our nominal PDF set to the other PDF sets. Finally, we include a contribution to the systematic uncertainty from the PDF central value. This is evaluated from 100 replicas of event-weights using our nominal PDF set in SHERPA. The mean is used for our central value and the standard deviation becomes the uncertainty.

9.4 ASIMOV FIT RESULTS

The statistical analysis for this measurement is performed using an ATLAS software package that runs a binned likelihood fit. The likelihood is constructed taking into account the product of Poisson probability terms for each bin of the neural network score distribution in each analysis channel. The binned likelihood function $\mathcal{L}(\mu, \theta)$ depends on the parameter of interest (POI), μ , and the set of nuisance parameters (NP), θ . The package allows us to perform histogram smoothing, uncertainty breakdowns, nuisance parameter symmetrization and pruning, and creates many post-fit plots.

For our search the signal regions include:

- $W \rightarrow e/\mu\nu, H \rightarrow \tau_{had}\tau_{had}$,
- $W \rightarrow e\nu, H \rightarrow \tau_e\tau_{had}$,
- $W \rightarrow e/\mu\nu, H \rightarrow \tau_{\mu/e}\tau_{had}$,
- $W \rightarrow \mu\nu, H \rightarrow \tau_{\mu}\tau_{had}$,
- $Z \rightarrow ee/\mu\mu, H \rightarrow \tau_{had}\tau_{had}$,
- $Z \rightarrow ee/\mu\mu, H \rightarrow \tau_{lep}\tau_{had}$,

The backgrounds include:

- Events with misidentified objects (Fakes)
- Events with top quarks
- Tribosons
- Dibosons (primarily WZ and ZZ)
- ttH

Binning Optimization

I chose bin widths that maximized the number of bins, to provide as much information as possible to the fit, while simultaneously staying within reasonable statistical uncertainties in each bin. I used the provided gamma values as a measure of MC statistical uncertainty for each bin. The gamma values are defined from the Poisson constraint term and are expected to have a value of 1. However, they can vary

from 1 and this fitted error reflects the relative statistical uncertainty in that specific bin. The likelihood fit automatically dropped any gamma contributing a statistical uncertainty of less than 1%. To handle the other gammas, I combined bins with their neighbors in order to get the associated gamma values for each resulting bin between 0.8 and 1.2. This statistical uncertainty is important to reduce because the sample yields in each bin will eventually be scaled by the associated gamma value.

Once the value of all the gamma nuisance parameters is brought below a 20% deviation from 1 by merging neighboring bins, one can perform a second stage of binning optimization to smooth out the discontinuities in the NN distribution, as described below.

Symmetrization and Smoothing

Smoothing and symmetrization procedures remove fluctuations caused by our limited MC sample sizes. For the fake closure, fake MC subtraction, α_S theory, jet, muon, tau, and electron systematic uncertainties a two-sided symmetrization is performed. This is done bin-by-bin and is calculated as the (up-down)/2 variation for the up variation and then mirrored for the down variation. For each of these systematics, histograms are provided with the up and the down variations, respectively. For the μ_R , μ_F and PDF theory uncertainties a one-sided symmetrization is performed. Here, we only provide one upwards fluctuation histogram and the down variation is added as a mirrored version. Some of the systematics are combined using an “envelope method”, which combines sets of related uncertainties into a single nuisance parameter in the fit after the symmetrization has been performed for the individual components.

To smooth, we rebin the distribution until relative bin-to-bin MC statistical uncertainties are below a tolerance that we define. We can repeat this process by halving the tolerance in cases where there are a smaller number of variations than a defined maximum. In a final smoothing step we remove artificially flat uncertainties.

Pruning

All bin-by-bin statistical uncertainties (gammas) on samples with less than 1% relative contribution are removed from consideration. In addition, systematic uncertainties that impact either the shape or normalization of distributions with a contribution of less than 1% are removed. These systematics prunings are done per region and per sample.

“After all, tomorrow is another day!”

Margaret Mitchell, *Gone with the Wind*

10

Final Remarks

WE ARE AT AN EXCITING JUNCTURE IN THIS ANALYSIS. At the moment, we are finishing up our internal documentation and I am performing a Likelihood fit with an Asimov data set so we can estimate the sensitivity for our analysis. I will look at the sensitivity we can obtain with all six channels combined and with each one separately. We will also estimate the sensitivity we would achieve if we did not use our neural network, which I helped train, in order to explicitly demonstrate its separation power.

We were recently assigned an internal ATLAS review board, which is a crucial step toward unblinding the data in our signal region, which we plan to do this spring. That means we will finally get to look at data in our signal region and perform a final likelihood fit to determine our measurement sensitivity. Our research group is targeting a paper publication for the fall of 2022. I am looking forward to the continued contributions I will make, as an analysis contact and paper editor, before my graduation in May.

While I am, at least temporarily, stepping away from the field of particle physics, I see a bright experimental future ahead. The HL-LHC is set to run until the late 2030s meaning there is much more data to be collected and analyzed. The higher luminosity will allow for more precise measurements of the proper-

ties of the Higgs boson and access to potential beyond-the-standard-model physics, such as charged lepton flavor violation and CP violation of the Higgs boson.

Perhaps even more exciting to think about is what will happen after the HL-LHC. The Snowmass process is currently ongoing, during which the US particle physics community, working with other communities around the world, will decide what to build next. Will we have another circular collider in Europe (the FCC) and will it collide protons or electrons? Will there be a linear collider built in Japan? There are recent suggestions for a muon collider at Fermilab, the home of muon experiments Mu2e and Muon $g-2$, which would be a different way to access new and interesting physics here in the US. The final decision will need to involve thoughts about political considerations, budget constraints, environmental impacts, and what physics is of most interest to the community.

Regardless of which path is chosen, machine learning and increased computational power will be a large component of ensuring the data collected can be analyzed efficiently and effectively. I am eagerly waiting to see where the field goes while I look to my own next steps contributing to advancements in AI and computing.



Author's Individual Contributions

A.1 $VH, H \rightarrow \tau\tau$ ANALYSIS

- I helped produce simulation and data samples for the analysis with the xTauFramework
- I assisted in processing the input ROOT files for all the simulation and data into pandas dataframes
- I calculated fake rates for all three final state objects using data from all four years, incorporated the MC subtraction, and determined the associated statistical and systematic uncertainties
- I established our analysis code in python and laid the groundwork for our preselections, validation regions, and signal selections.
- I optimized the signal selections for all four categories to maximize signal to background ratios while keeping signal efficiency high
- I re-optimized the Neural Network after key changes were made in our category definitions
- I tuned the binning and inclusion of systematics in our final likelihood fit

A.1.1 RELATED PRESENTATIONS

- **Machine Learning for Tau Identification and the Associated Production of the Higgs Boson**, Fermilab Research Seminar, Virtual Jan 2022
- **MVA in $V(\text{lep})H$** , Tau and HLeptons Workshop, Virtual Dec 2021
- **Status of the VH Analysis**, Tau and HLeptons Workshop, Virtual Sep 2020
- **Associated Production of a Higgs boson with a vector boson, with the Higgs boson decaying to two tau leptons**, Yale Wright Laboratory Weak Interaction Discussion Group Nov 2019

A.2 MACHINE LEARNING FOR TAUS

- I incorporated trigger-level track variables in the training of the Boosted Regression Tree that is used to calibrate tau energy in MVATES
- I demonstrated a proof of principle that including trigger-level track information improves performance of MVATES for possible use in the trigger in future runs of ATLAS
- I processed signal (tau) and background (jet) simulation samples into a weighted format for use in the Lorentz Equivariant Network
- I got initial results from using the Lorentz equivariant network and compared them to training with a Deep Sets architecture

A.2.1 RELATED PRESENTATIONS

- **Improvements to the Tau Energy Scale (MVATES) at the Trigger Level**, ATLAS Trigger Meeting, Jan 2020
- **Energy Calibration in the ATLAS Tau Trigger**, American Physical Society April Meeting, April 2020
- **Tau Energy Scale at the ATLAS Detector**, USLUA Lightning Talk, Rice University, Oct 2019

A.3 OUTREACH AND LEADERSHIP

- I ran the Women in Physics+ group at Yale for four years and led the group in advocacy efforts, established an Allies group, and helped recruit more women to Yale
- I co-chaired the Equity in the Job Search Symposium and led the planning of three symposia focused on increasing gender equity in STEM careers
- I was an organizer for the Yale Physics Professional Development Organization and planned a year of virtual seminars, workshops, and networking opportunities to prepare graduate students and post-docs for both the academic and non-academic job search
- I participated in two annual trips to DC to meet with congressional staff and advocate for continued federal investments in basic science research
- I volunteered at multiple Girl's Science Investigations and Yale Physics Olympics
- I was a graduate affiliate with Pierson College

A.4 TEACHING

- Physics 526, Introduction to Elementary Particle Physics: Spring 2021
- Physics 165/166, Instructional Physics Labs for Non-Majors: Fall 2017, Spring 2018, Summer 2020, and Fall 2020
- Physics 118/ENAS 220/MUSI 187, Physics of Music and Instrument Design: Summer 2020
- Physics 200/201, Fundamentals of Physics: Fall 2018, Spring 2019, and Fall 2020
- Physics 206, Physics laboratory course for physics majors: Fall 2016 and Spring 2017

References

- [1] Wikipedia Commons. Standard model of elementary particles. June 2006.
- [2] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, New York, 2013.
- [3] Bruce Mellado. Prospects for higgs boson searches at the large hadron collider. *Pramana*, 72:15–22, January 2009.
- [4] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, 2018.
- [5] Mariotti C. Passarino G. Dittmaier, S. and Tanaka R. Handbook of lhc higgs cross sections: 1. inclusive observables. 2011. CERN-2011-002.
- [6] G. Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [7] The ATLAS Collaboration. Summary plots from the ATLAS Standard Model Group. <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/>. Last accessed Spring 2021.
- [8] G. Aad et al. Atlas data quality operations and performance for 2015–2018 data-taking. *Journal of Instrumentation*, 15(04):P04003–P04003, Apr 2020.
- [9] CERN Image Gallery. <https://home.cern/resources/image/cern/views-cern-images-gallery>. Last accessed 1 Dec 2021.
- [10] Lyndon R Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001. 164 p, 2008. This report is an abridged version of the LHC Design Report (CERN-2004-003).
- [11] Naeem Tahir, J. Sancho, Alexander Shutov, Rüdiger Schmidt, and A. Piriz. Impact of high energy high intensity proton beams on targets: Case studies for super proton synchrotron and large hadron collider. *Physical Review Special Topics - Accelerators and Beams*, 15, 05 2012.
- [12] Andreas Hoecker. Physics at the LHC Run-2 and Beyond. pages 153–212. 61 p, Nov 2016. Lecture notes from the 2016 European School of High-Energy Physics, 15-28 June 2016, Skeikampen, Norway (61 pages, 56 figures).
- [13] "The ATLAS Collaboration". Public Results: Luminosity Run 2. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [14] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [15] Sebastian Murk. Trigger algorithms and monte carlo event generation for dijet searches in atlas and beyond. 2017. Master's Thesis, Lund University.
- [16] Joao Pequeno and Paul Schaffner. How ATLAS detects particles: diagram of particle paths in the detector. <https://cds.cern.ch/record/1505342>, Jan 2013.

- [17] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1996. CERN-LHCC-96-041.
- [18] A. Henriques. The atlas tile calorimeter. In *2015 4th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA)*, pages 1–7, 2015.
- [19] *ATLAS muon spectrometer: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1997. CERN-LHCC-97-022.
- [20] Jeremiah Jet Goodson. *Search for supersymmetry in states with large missing transverse momentum and three leptons including a Z-boson*. PhD thesis, State University of New York at Stony Brook, 2012.
- [21] Marumi Kado Andreas Hoecker. Setting off to new energy horizons. <https://atlas.cern/updates/news/setting-new-energy-horizons>, 2015.
- [22] Michela Paganini. Machine learning solutions for high energy physics: Applications to electromagnetic shower generation, flavor tagging, and the search for di-higgs production, 2019. PhD Thesis, Yale University.
- [23] T Gleisberg, S Höche, F Krauss, M Schönherr, S Schumann, F Siegert, and J Winter. Event generation with sherpa 1.1. *Journal of High Energy Physics*, (02):007, Feb 2009.
- [24] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the lhc era: Mmht 2014 pdfs. *The European Physical Journal C*, 75(5), May 2015.
- [25] M. Aaboud et al. Performance of the atlas track reconstruction algorithms in dense environments in lhc run 2. *The European Physical Journal C*, 77(10), Oct 2017.
- [26] G. Aad et al. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J.*, C77:490, 2017.
- [27] Electron and photon performance measurements with the atlas detector using the 2015–2017 lhc proton-proton collision data. 14.
- [28] M. Aaboud et al. Electron reconstruction and identification in the atlas experiment using the 2015 and 2016 lhc proton–proton collision data at $\sqrt{s} = 13\text{TeV}$. *The European Physical Journal C*, 79(8), Aug 2019.
- [29] Sebastien Rettie. Muon identification and performance in the ATLAS experiment. Technical report, CERN, Geneva, Jun 2018. ATL-PHYS-PROC-2018-052.
- [30] G. Aad et al. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13\text{ TeV}$. *Eur. Phys. J. C*, 81(7):578, 2021.
- [31] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008.
- [32] M. Aaboud et al. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s}=13\text{ tev}$ with the atlas detector. *Physical Review D*, 96(7), Oct 2017.
- [33] A. Hoecker et al. Tmva 4: Toolkit for multivariate data analysis with root: Users guide, March 2017. arXiv:physics/0703039.

- [34] Mao-Xin Wang, Duruo Huang, Gang Wang, and Dian-Qing Li. Ss-xgboost: A machine learning framework for predicting newmark sliding displacements of slopes. *Journal of Geotechnical and Geoenvironmental Engineering*, 146:04020074, 06 2020.
- [35] Michael A. Nielsen. *Neural Networks Deep Learning*. Determination Press, 2015. <http://neuralnetworksanddeeplearning.com/faq.html>, Last accessed Jan 2022.
- [36] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>, Last Accessed Jan 2022.
- [37] Michael Phi. Illustrated guide to lstm's and gru's: A step by step explanation. *Towards Data Science*, 2018. "<https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>", Last accessed Jan 2022.
- [38] Geert-Jan Besjes. Performance of the ATLAS Tau Trigger in Run 2. Technical report, CERN, Geneva, Nov 2016. ATL-DAQ-PROC-2016-027.
- [39] Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment. Technical report, CERN, Geneva, Aug 2019. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-033>.
- [40] Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, May 2017. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2017-029>.
- [41] 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.
- [42] G. Aad et al. Search for the standard model higgs boson produced in association with a vector boson and decaying into a tau pair in pp collisions at $\sqrt{s}=8\text{TeV}$ with the atlas detector. *Physical Review D*, 93(9), May 2016.
- [43] 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>.
- [44] David J Griffiths. *Introduction to Elementary Particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008.
- [45] Francis Halzen and Alan D. Martin. *QUARKS AND LEPTONS: An Introductory Course in Modern Particle Physics*. John Wiley and Sons, New York, USA, 1984.
- [46] Donald H. Perkins. *Introduction to High Energy Physics, 4th Edition*. Cambridge University Press, United Kingdom, 2000.
- [47] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

- [48] Christophe Grojean. Alternatives to the standard model higgs boson. <https://indico.cern.ch/event/122285/>, April 2011.
- [49] 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.
- [50] M. Tanabashi et al. Status of higgs boson physics. *Phys. Rev. D*, 98:030001, 2018.
- [51] M. Aaboud et. al. Observation of H to $b\bar{b}$ decays and VH production with the ATLAS detector. *Physics Letters B*, 786:59–86, 2018.
- [52] M. Aaboud et al. Cross-section measurements of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at $\sqrt{s} = 13\text{TeV}$ with the ATLAS detector. *Physical Review D*, 99(7), Apr 2019.
- [53] G. Aad et al. A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector. *Physics Letters B*, 812:135980, Jan 2021.
- [54] Search for the standard model Higgs boson decaying to charm quarks. Technical Report CMS-PAS-HIG-18-031, CERN, Geneva, 2019.
- [55] Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays to charm quarks with the ATLAS detector. Technical report, CERN, Geneva, Jun 2021. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-021>.
- [56] G. Aad et al. Observation and measurement of higgs boson decays to WW^* with the atlas detector. *Phys. Rev. D*, 92:012006, Jul 2015.
- [57] G. Aad et al. Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment. *The European Physical Journal C*, 76(1):6, 2016.
- [58] R. Abdul Khalek et al. The physics potential of HL-LHC. Summary of the findings of the Workshop on “The physics of HL-LHC, and perspectives on HE-LHC.
- [59] Douglas Clowe, Maruša Bradač, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal*, 648(2):L109–L113, Aug 2006.
- [60] Katherine Freese. Status of dark matter in the universe. *International Journal of Modern Physics D*, 26(06):1730012, Mar 2017.
- [61] K.M. Backes et al. A quantum enhanced search for dark matter axions. *Nature*, 590:238–242, 2021.
- [62] Y. Fukuda, T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, H. Ishino, Y. Itow, T. Kajita, J. Kameda, S. Kasuga, et al. Evidence for oscillation of atmospheric neutrinos. *Physical Review Letters*, 81(8):1562–1567, Aug 1998.
- [63] Pablo F. de Salas, Stefano Gariazzo, Olga Mena, Christoph A. Ternes, and Mariam Tórtola. Neutrino mass ordering from oscillations and beyond: 2018 status and future prospects. *Frontiers in Astronomy and Space Sciences*, 5, Oct 2018.

- [64] R. Aaij et al. Test of lepton universality in beauty-quark decays, 2021. CERN-EP-2021-042.
- [65] B. Abi et al. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr 2021.
- [66] The History of CERN. <https://timeline.web.cern.ch/timeline-header/89>, Last Accessed Feb 2022.
- [67] CERN. Our member state. <https://home.cern/about/who-we-are/our-governance/member-states>, 2019.
- [68] S N Ganguli. The story of large electron positron collider. <https://www.ias.ac.in/article/fulltext/reso/007/11/0018-0034>, November 2002.
- [69] Stephen Holmes, Ronald S Moore, and Vladimir Shiltsev. Overview of the tevatron collider complex: goals, operations and performance. *Journal of Instrumentation*, 6(08):T08001, Aug 2011.
- [70] Ristori L. Bernardi G., Denisov D. and Roser R. Tevatron scientists announce their final results on the higgs particle. *Journal of Instrumentation*, July 2012. <https://news.fnal.gov/2012/07/tevatron-scientists-announce-final-results-higgs-particle/>.
- [71] CERN. *Proceedings of the ECFA-CERN Workshop : Large Hadron Collider in the LEP tunnel: Lausanne, Switzerland 21 - 27 Mar 1984. ECFA-CERN Workshop on Large Hadron Collider in the LEP tunnel*, Geneva, 1984. CERN.
- [72] LHC Guide. CERN-Brochure-2017-002-Eng, Mar 2017.
- [73] P.A.Zyla and others (Particle Data Group). Prog. theor. exp. phys. <https://doi.org/10.1093/ptep/ptaa104>, August 2020.
- [74] Sarah Charley. What is luminosity? <https://www.symmetrymagazine.org/article/what-is-luminosity>, February 2002.
- [75] CERN. The large hadron collider. <https://home.cern/science/accelerators/large-hadron-collider>, 2021. Last accessed 1 Dec 2021.
- [76] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report v.1: the LHC Main Ring*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. CERN-2004-003-V-1.
- [77] R Scrivens, M Kronberger, D Küchler, J Lettry, C Mastrostefano, O Midttun, M O’Neil, H Pereira, and C Schmitzer. Overview of the status and developments on primary ion sources at CERN. CERN-ATS-2011-172.
- [78] Michael Benedikt, Paul Collier, V Mertens, John Poole, and Karlheinz Schindl. *LHC Design Report v.3: the LHC Injector Chain*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. CERN-2004-003-V-3.
- [79] Xabier Cid Vidal and Ramon Cid Manzano. Taking a closer look at the lhc. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/1.home. Last accessed Nov 2021.
- [80] The ATLAS Experiment. Detector and Technology. <https://atlas.cern/discover/detector>. Last accessed Nov 2021.

- [81] The ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 1*. Technical design report. ATLAS. CERN, Geneva, 1999. CERN-LHCC-99-014 (Vol. 1).
- [82] 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.
- [83] F. Hugging. The ATLAS Pixel Insertable B-Layer (IBL). *Nucl. Instrum. Meth.*, A650:45–49, 2011.
- [84] Bartosz Mindur. ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider. *Nucl. Instrum. Methods Phys. Res.*, A 845 (2017), 257-261.
- [85] Michaela Mlynarikova. Performance of the ATLAS hadronic Tile calorimeter. In *5th Large Hadron Collider Physics Conference*, 8 2017.
- [86] The ATLAS collaboration. Operation of the atlas trigger system in run 2. *Journal of Instrumentation*, 15(10):P10004, Oct 2020.
- [87] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.
- [88] Torbjörn Sjöstrand et al. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.
- [89] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008.
- [90] Enrico Bothmann, Gurpreet Singh Chahal, Stefan Höche, Johannes Krause, Frank Krauss, Silvan Kuttimalai, Sebastian Liebschner, Davide Napoletano, Marek Schönherr, Holger Schulz, et al. Event generation with sherpa 2.2. *SciPost Physics*, 7(3), Sep 2019.
- [91] M Dobbs and J B Hansen. The HepMC C++ Monte Carlo Event Record for High Energy Physics. Technical report, CERN, Geneva, Jun 2000. ATL-SOFT-2000-001, revised version number 1 submitted on 2001-02-27.
- [92] Andy Buckley, Jonathan Butterworth, Stefan Gieseke, David Grellscheid, Stefan Höche, Hendrik Hoeth, Frank Krauss, Leif Lönnblad, Emily Nurse, Peter Richardson, et al. General-purpose event generators for lhc physics. *Physics Reports*, 504(5):145–233, Jul 2011.
- [93] Alan D. Martin. Proton structure, partons, qcd, dglap and beyond, 2008. *Acta Phys.Polon.B39:2025-2062,2008*.
- [94] Zach Marshall. Defending your life (parts 1,2,3). <https://atlas.cern/updates/blog/defending-your-life-part-3>, October 2014.
- [95] V Lacuesta. Track and vertex reconstruction in the ATLAS experiment. *Journal of Instrumentation*, 8(02):C02035–C02035, Feb 2013.
- [96] T Cornelissen, M Elsing, S Fleischmann, W Liebig, E Moyse, and A Salzburger. Concepts, Design and Implementation of the ATLAS New Tracking (NEWT). Technical report, CERN, Geneva, Mar 2007. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2007-007>.
- [97] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987.

- [98] Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment. Technical report, CERN, Geneva, Jul 2010. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2010-069>.
- [99] M. Aaboud et al. Reconstruction of primary vertices at the atlas experiment in run 1 proton-proton collisions at the lh. *The European Physical Journal C*, 77(5), May 2017.
- [100] 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.
- [101] G. Aad et al. Electron and photon energy calibration with the atlas detector using lh. run 1 data. *The European Physical Journal C*, 74(10), Oct 2014.
- [102] M. Aaboud et al. Electron and photon energy calibration with the atlas detector using 2015–2016 lh. proton-proton collision data. *Journal of Instrumentation*, 14(03):P03017–P03017, Mar 2019.
- [103] G. Aad et al. Muon reconstruction performance of the atlas detector in proton-proton collision data at $\sqrt{s} = 13$ tev. *The European Physical Journal C*, 76(5), May 2016.
- [104] Aliaksei Hrynevich. ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2. May 2017. JINST 12 C06038.
- [105] P Berta. ATLAS jet and missing-ET reconstruction, calibration, and performance. Technical report, CERN, Geneva, Oct 2014. Nuclear and Particle Physics Proceedings 273–275 (2016) 1121–1126.
- [106] Ryan Atkin. Review of jet reconstruction algorithms. *Journal of Physics: Conference Series*, 645:012008, oct 2015.
- [107] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. Fastjet user manual. *The European Physical Journal C*, 72(3), Mar 2012.
- [108] Pile-up subtraction and suppression for jets in ATLAS. Technical report, CERN, Geneva, Aug 2013. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2013-083>.
- [109] Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run. Technical report, CERN, Geneva, Jul 2017. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-013>.
- [110] M. Aaboud et al. Performance of missing transverse momentum reconstruction with the atlas detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 78(11), Nov 2018.
- [111] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65 6:386–408, 1958.
- [112] Bernard Marr. A short history of machine learning – every manager should read. *Forbes Magazine*, 2016.

- [113] Prashant Gupta. Regularization in machine learning. *Towards Data Science*, 2017.
- [114] Y. Freund and R. E. Schapire. A Decision Theoretic Generalization of On-Line Learning and an Application to Boosting. In Paul M. B. Vitányi, editor, *Second European Conference on Computational Learning Theory (EuroCOLT-95)*, pages 23–37, 1995.
- [115] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.
- [116] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. arXiv:1412.6980.
- [117] G. Aad et al. Dijet resonance search with weak supervision using $\sqrt{s} = 13\text{TeV}$ pp collisions in the atlas detector. *Physical Review Letters*, 125(13), Sep 2020.
- [118] G. Aad et al. Search for higgs boson decays into a z boson and a light hadronically decaying resonance using $\sqrt{s} = 13\text{TeV}$ pp collision data from the atlas detector. *Physical Review Letters*, 125(22), Nov 2020.
- [119] M. Aaboud et al. Measurement of the higgs boson coupling properties in the $h \rightarrow ZZ^*4l$ decay channel at $\sqrt{s} = 13\text{TeV}$ with the atlas detector. *Journal of High Energy Physics*, 2018(3), Mar 2018.
- [120] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Calogan: Simulating 3d high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Phys. Rev. D*, 97:014021, Jan 2018.
- [121] Chase Shimmin, Peter Sadowski, Pierre Baldi, Edison Weik, Daniel Whiteson, Edward Goul, and Andreas Søgaard. Decorrelated jet substructure tagging using adversarial neural networks. *Physical Review D*, 96(7), Oct 2017.
- [122] Claudionor N. Coelho, Aki Kuusela, Shan Li, Hao Zhuang, Jennifer Ngadiuba, Thea Klæboe Aarrestad, Vladimir Loncar, Maurizio Pierini, Adrian Alan Pol, and Sioni Summers. Automatic heterogeneous quantization of deep neural networks for low-latency inference on the edge for particle detectors. *Nature Machine Intelligence*, 3(8):675–686, Jun 2021.
- [123] Sabrina Amrouche, Laurent Basara, Paolo Calafiura, Victor Estrade, Steven Farrell, Diogo R. Ferreira, Liam Finnie, Nicole Finnie, Cécile Germain, Vladimir Vava Gligorov, et al. The tracking machine learning challenge: Accuracy phase. *The Springer Series on Challenges in Machine Learning*, page 231–264, Nov 2019.
- [124] K. Benslama, C. Belanger-Champagne, M. Bosman, R. Brenner, P. Casado, Z. Czychula, M. Dam, S. Demers, S. Farrington, O. Igonkina, A. Kalinowski, N. Kanaya, C. Osuna, E. Perez, E. Ptacek, A. Reinsch, A. Saavedra, A. Sfyrla, A. Sopczak, D. Strom, E. Torrence, S. Tsuno, V. Vorwerk, A. Watson, and S. Xella. Tau trigger at the atlas experiment, 2008. arXiv:0810.0465.
- [125] The ATLAS Tau Trigger in Run 2. Technical report, CERN, Geneva, Jul 2017. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2017-061>.
- [126] Taco S. Cohen and Max Welling. Group equivariant convolutional networks, 2016. Proceedings of the International Conference on Machine Learning (ICML), 2016.

- [127] Dirk Duschinger. Thor. <https://gitlab.cern.ch/atlas-perf-tau/THOR/blob/master/README.rst>, 2021. Last accessed 26 Jan 2022.
- [128] ATLAS RNN Tau ID Group. rnn-tauid. <https://test-rnn-tauid-docs.web.cern.ch/test-rnn-tauid-docs/>, 2021. Last accessed 26 Jan 2022.
- [129] Alexander Bogatskiy, Brandon Anderson, Jan T. Offermann, Marwah Roussi, David W. Miller, and Risi Kondor. Lorentz group equivariant neural network for particle physics, 2020. arXiv:2006.04780.
- [130] Jan Offerman and David Miller. <https://github.com/fizisist/LorentzGroupNetwork>, 2020.
- [131] 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.
- [132] 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, CERN, Geneva, Jun 2018. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2018-021>.
- [133] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 7 2013. CERN-2013-004.
- [134] CMS Collaboration. Measurement of Higgs boson production and decay to the $\tau\tau$ final state. 2019. CMS-PAS-HIG-18-032.
- [135] Jim Pivarski, Pratyush Das, Chris Burr, Dmitri Smirnov, Matthew Feickert, Tamas Gal, Luke Kreczko, Nicholas Smith, Noah Biederbeck, Oksana Shadura, Mason Proffitt, Bkrikler, Hans Dembinski, Henry Schreiner, Jonas Rembser, Marcel R., Chao Gu, Jonas Rübenach, Michele Peresano, and Ruggero Turra. scikit-hep/uproot: 3.12.0, July 2020.
- [136] Jeff Reback, jbrockmendel, Wes McKinney, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, Simon Hawkins, gyoung, Matthew Roeschke, Sinhrks, Adam Klein, Terji Petersen, Jeff Tratner, Chang She, William Ayd, Patrick Hoefler, Shahar Naveh, Marc Garcia, Jeremy Schendel, Andy Hayden, Daniel Saxton, JHM Darbyshire, Richard Shadrach, Marco Edward Gorelli, Fangchen Li, Matthew Zeitlin, Vytautas Jancauskas, Ali McMaster, Pietro Battiston, and Skipper Seabold. pandas-dev/pandas: Pandas 1.3.4, October 2021.
- [137] 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.
- [138] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), Feb 2011.
- [139] Kyle Cranmer. Practical statistics for the LHC, 2015. CERN-2014-003, pp. 267 - 308, arXiv:1503.07622.
- [140] Michele Pinamonti. Statistical methods at atlas and cms, 2018. Higgs Toppings Workshop, Benasque, <https://indico.cern.ch/event/727396/contributions/3021899/>

[attachments/1657532/2654085/Statistical_methods_at_ATLAS_and_CMS_2.pdf](#)".

- [141] Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. Technical report, CERN, Geneva, Jun 2019. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-021>.
- [142] M. et. al. Aaboud. Jet reconstruction and performance using particle flow with the atlas detector. *The European Physical Journal C*, 77(7), Jul 2017.
- [143] E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC p-p collisions. Technical report, CERN, Geneva, Jun 2018. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2018-023>.