

Improving b-jet identification and searching for additional Higgs bosons with ATLAS

Gregory Barbour

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Physics & Astronomy
University College London

March 29, 2022

I, Gregory Barbour, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

The ATLAS experiment at the LHC has completed its run-2 data collection at $\sqrt{s} = 13$ TeV. An upgrade phase has begun in preparation of the high luminosity LHC. Improvements to the algorithms underpinning data analysis in ATLAS are researched, with a particular focus on data science techniques and machine learning. The research carried out focused on b-tagging, the labelling of jets produced by b decays, and a search for new physics signals in ATLAS data. The use of neutral tracks in the fit of the b-decay topology in existing ATLAS algorithms was studied. No major performance boost was observed. Software development efforts were carried out to improve the design and extensibility of existing b-jet decay topology fitting algorithms. Research into an RNN based b-jet topological reconstruction algorithm was carried out. The optimization studies and performance evaluation techniques are summarized alongside key insights. Such a technique could serve next to existing b-tagging algorithms at ATLAS. Finally, a search was performed for heavier version of the Higgs boson using the 139 fb^{-1} ATLAS data. Such particles are motivated by several BSM theories. Expected upper limits on the strength of various signal models are given, and expected exclusion contours at 95% confidence level are drawn for the various parameters of the model.

Impact Statement

The field of particle physics explores the most fundamental equations of laws and nature. The experimental measurements made by the ATLAS detector are at the very cutting edge of human knowledge. This work includes an analysis of ATLAS data in a completely novel final state, which is well motivated as a signal for a new heavy particle with Higgs-like properties. The results of this analysis will directly contribute to help answer one of the greatest questions of particle physics: What lies beyond the Standard Model? Further impacts on the ATLAS data processing chain are also present in this work. These focus around efforts for the development of new and existing flavour tagging techniques. A particular focus on Machine Learning algorithms underpinned this research. The benefits outside of academia will derive mainly from the technologies developed during this effort. In particular, the expertise and algorithms developed in the field of AI are expected to be of significant importance to the future of the world economy.

Acknowledgements

I would first, and foremost, like to thank my supervisor Andreas for all his help and guidance throughout the past years. This work would not have been possible without his advice.

I would also like to thank Yanhui for all his help, and the members of the GHH analysis team, in particular Yue and Ke, who were always ready to give answers or guidance when needed. Many thanks also to the members of the ATLAS Flavour Tagging group for all of their help. It has been a pleasure working with all of you over the years.

My thanks to Tom, Chris and the many members of Faculty who guided and supported me during my internship. Thanks also to Davide and Alex, it was a pleasure working alongside you both.

Thank you to Mayumi and to all the friends I have made over the years at UCL, CERN and beyond. The journey would not be worth it without your companionship, your jokes, your support.

Finally, thank you to my family, Oliver, Julian and Laura and to my Mum and Dad for everything they have done and continue to do for me. I am so grateful and lucky to have you.

Contents

1	Introduction	27
2	Theory	29
2.1	Particle Physics at ATLAS	29
2.1.1	Interactions of the SM	30
2.1.2	Higgs Mechanism	33
2.1.3	Heavy Quark Phenomenology	38
2.2	Beyond the Standard Model	39
2.2.1	Problems with the SM	40
2.2.2	Examples of BSM	42
3	The ATLAS Detector at the LHC	46
3.1	The Large Hadron Collider (LHC)	46
3.2	The ATLAS Detector	47
3.2.1	Co-ordinate System	48
3.2.2	Inner Detector	49
3.2.3	Calorimeter System	51
3.2.4	Muon Spectrometer	53
3.2.5	Trigger	53
4	Data Processing in ATLAS	55
4.1	Object Reconstruction	55
4.1.1	Tracks	55
4.1.2	Jets	56

4.1.3	Large-R Jets	60
4.1.4	Electrons	63
4.1.5	Muons	65
4.1.6	Missing Transverse Momentum	67
4.1.7	Overlap Removal	68
4.2	Data Formats in ATLAS	69
4.2.1	Data Samples	69
4.2.2	MC Simulation in ATLAS	69
4.2.3	Derivations	70
5	Machine Learning	71
5.1	Supervised learning	72
5.1.1	Regression	72
5.1.2	Classification	72
5.1.3	Loss functions and Gradient Descent	73
5.1.4	Training	75
5.2	Models	75
5.2.1	Boosted Decision Trees	75
5.2.2	Neural Networks	78
5.2.3	Recurrent Neural Networks	79
5.2.4	Other Neural Networks	81
5.2.5	Variational Autoencoders	82
5.3	Working Points	84
6	b-Tagging	87
6.1	Physics Motivation	87
6.2	b decay properties	88
6.3	Overview of ATLAS Techniques	88
6.3.1	Impact Parameter based methods	89
6.3.2	RNNIP	90
6.3.3	Inclusive Secondary Vertex Reconstruction	92

6.3.4	Topological Reconstruction	93
6.3.5	Leptons	93
6.3.6	High-level Taggers	94
6.4	JetFitter	97
6.4.1	Track Selection	98
6.4.2	Vertex Finding	100
6.4.3	Fitting Procedure	101
6.5	Neutral Tracks	103
6.5.1	Neutral Particle Identification and Veto	104
6.5.2	Neutral Particle Reconstruction	107
6.6	Code Development	109
7	Jet Topology Fitting with an RNN	113
7.1	Overview	113
7.2	Motivation	114
7.3	Toy Jets	115
7.4	RNN Fitting	116
7.4.1	Training	117
7.4.2	Model Optimization	119
7.4.3	Best Model	127
7.5	Performance metrics	128
7.5.1	Individual Jet Performance	129
7.5.2	Vertex Identification	131
7.5.3	Impact of Individual Tracks	136
7.6	Outlook	138
8	Analysis	141
8.1	Physics Motivation	141
8.2	Analysis Overview	144
8.2.1	Strategy	144
8.2.2	Personal Contributions	145

8.3	Data and Simulated Samples	146
8.3.1	Simulated Signal Samples	146
8.3.2	Simulated Background Samples	147
8.4	Object Definitions and Selections	147
8.4.1	Primary Vertex and pile-up	147
8.4.2	Triggers	148
8.4.3	Electrons	149
8.4.4	Muons	149
8.4.5	Jets	150
8.4.6	Missing Transverse Momentum	151
8.5	Event Selection and Optimization	151
8.6	Background Estimation	153
8.6.1	Non-prompt Leptons	154
8.6.2	Charge mis-identification	157
8.6.3	Fake Leptons from Photons	160
8.6.4	Backgrounds with prompt leptons	162
8.7	Validation Region	164
8.8	Systematic Uncertainties	165
8.8.1	Experimental Systematics	166
8.8.2	Data-Driven Background Systematics	171
8.8.3	Theoretical Systematics	174
8.9	Statistical Analysis	176
8.9.1	Hypothesis Test for Exclusion Limits	177
8.9.2	Binning in the GHH Analysis	179
8.10	Results	179
8.10.1	Signal Acceptance and Efficiency	179
8.10.2	Normalization factors	180
8.10.3	Blinded Exclusion Fit	181
8.10.4	Systematic Uncertainties	183
8.10.5	Expected Results in the Inclusive Fit	184

8.10.6 Resolved and Boosted-only Sensitivities	187
8.11 Summary	194
9 Conclusion	196
Appendices	198
A Neutral Tracks in JetFitter	198
A.1 Overview	198
A.2 List of Modified Files	199
A.3 List of Identified Bugs	200
A.4 Variable Plots	200
B JetFitter Refactor	202
B.1 JetFitter Code Maps	203
C Error Propagation for a Single Track	207
D Analysis Framework and Triggers	209
D.1 Analysis Framework	209
D.2 Analysis Triggers	210
E Signal Samples	211
Bibliography	216

List of Figures

2.1	The particles of the SM	31
2.2	Weak interaction vertex with an electron and electron neutrino. . . .	34
2.3	A b quark decaying to c via the weak interaction. The b to c vertex picks up a factor V_{cb} from the CKM matrix.	34
2.4	The possible Higgs vertices from the SM. f is any massive fermion (leptons or quarks, but not neutrinos). V is any massive boson ($W^\pm$ or Z)	38
2.5	Feynman diagram of the first loop correction to the bare Higgs mass from a fermion f	41
2.6	Feynman diagram of the first loop correction to the bare Higgs mass from a superpartner $\tilde{f}$ of fermion f	43
3.1	CERN's accelerator complex	47
3.2	Schematic of the ATLAS detector	48
3.3	The barrel region of the inner detector.	51
4.1	A flow chart of the PFlow algorithm.	57
4.2	An example receiver operating characteristic (ROC) curve show- ing the light-jet rejection performance at given b-tagging efficiency working points for several ATLAS b-tagging algorithms.	60
4.3	The ATLAS data processing chain for real (left) and simulated (right) data.	69
5.1	An illustration of gradient descent for a one dimensional loss/cost function $J(w)$	74

- 5.2 An example of a decision tree. At each decision node a cut is applied on some feature x_i of the data, splitting the dataset. The final node (output leaf) labels the data as signal (S) or background (B). 76
- 5.3 A schematic of a fully-connect feed-forward neural network with three hidden layers. Information flows from left to right. The black lines denote connections between nodes. 78
- 5.4 A schematic of an RNN and its appearance when “unrolled”. Data at each timestep of the sequence is x_t , the RNN/LSTM cell is the box A. At each timestep the cell receives the hidden state from the previous timestep and the next data of the sequence. The output from the final cell is the RNN output. 80
- 5.5 A schematic of an autoencoder showing the encoder and decoder networks, and the latent encoding. 83
- 5.6 A schematic of a VAE. The input data enters on the left, the reconstructed data leaves on the right. The latents produced by the encoder correspond to means and variances of normal distributions. The z are sampled from these distributions and pass through the decoder. 84
- 5.7 An example ROC curve showing the TPR vs FPR for a logistic regression classifier (orange). Each point on the ROC curve corresponds to a different threshold value. The optimal classifier is in the top left corner. A random classifier (no skill) is represented by the blue dashed line. 86
- 6.1 The distributions of lifetime signed (a) transverse and (b) longitudinal impact parameters for b, c and light-jets in a $t\bar{t}$ sample. The b and c jets tends to have more tracks with large positive lifetime-signed impact parameters. The light jets also show an exponential tail in this direction due to tracks from K_S and Λ decays. The longitudinal impact parameter shows an additional component in the tail due to pile-up. 90

- 6.2 Schematic of RNNIP showing the network architecture, the inputs and the outputs. Note when used as an input to high-level taggers (MV2 and DL1), the RNNIP output is generally reduced to a 3-class p_b , p_c and p_{light} 92
- 6.3 The b-tagging efficiency vs (a) light-jet rejection and (b) c-jet rejection of MV2c10 (MV2) with the SMT discriminant added (MV2Mu) and the RNNIP outputs added (MV2MuRnn) evaluated on a $t\bar{t}$ sample. The ratio of all taggers to MV2c10 is also plotted. The addition of muon information does not improve the c-jet rejection above 50% efficiency, since leptons are produced b- and c-decays. 96
- 6.4 The b-tagging efficiency vs (a) light-jet rejection and (b) c-jet rejection of two high-level (MV2 and DL1) and three low-level (IP3D, SV1, JetFitter) taggers evaluated on $t\bar{t}$ events. The ratio of all taggers to MV2 is also plotted. 97
- 6.5 The b-tagging efficiency vs (a) light-jet rejection and (b) c-jet rejection of DL1, DL1r and DL1rmu evaluated on $t\bar{t}$ events. The ratio of all taggers to DL1 is also plotted. 98
- 6.6 The b decay topology showing the decay vertex of the b-hadron (B) and c-hadron (D), and the JetFitter assumption that these lie along the B flight axis. 99
- 6.7 A schematic of a b-decay chain topology within a jet, showing the variables defined in the JetFitter fitting procedure. 102
- 6.8 The effect on a MV2c10 of removing the tracks identified with neutral decay vertices. The b-jet efficiency is plotted against (a) light-jet rejection and (b) c-jet rejection. The black line is with the veto turned off, red dashed line is with the veto turned on. The ratio of the red curve divided by black is plotted. 105

- 6.9 The effect on a JetFitter based BDT (with the MV2 architecture) of removing the tracks identified with neutral decay vertices. The black line is with the veto turned off, red dashed line is with the veto turned on. The ratio of the red curve divided by black is plotted. . . . 106
- 6.10 A schematic view of the two charged particle tracks forming from a neutral particle decay. The neutral track is reconstructed from these, extrapolation back to its perigee is shown (closest approach point of the primary vertex). 107
- 6.11 The effect on a MV2c10 of using reconstructed K_S tracks in JetFitter fit. The b-jet efficiency is plotted against (a) light-jet rejection and (b) c-jet rejection. The black line is with neutral particles identified and vetoed, red dashed line is with the addition of reconstructed K_S tracks to the fit. The ratio of the red curve divided by black is plotted. 109
- 6.12 Map of the JetFitter code structure showing the four main modules. BTagSecVertexing is part of the main b-tagging algorithm and directs JetFitter to find its secondary vertices by calling InDetImprovedJetFitterVxFinder. The resulting decay chain topology is passed to JetFitterVariablesFactory to create the variables for high-level taggers. The flow of information is shown by arrows. Key methods within modules are shown. This map is not an exhaustive description but rather meant to show the important information only. 110
- 6.13 Map of the JetFitter code structure but with the InDetImprovedJetFitterVxFinder module, and in particular the track selection method doTheFinding shown in more detail. More expanded maps similar to this can be found in the appendix B. 111
- 6.14 Map of the refactored JetFitter code structure. InDetImprovedJetFitterVxFinder has been greatly simplified. Four new modules have been created. 111

- 7.1 A toy b-jet with noiseless pion tracks originating from the primary (red), secondary (blue) and tertiary (orange) vertices. The vertex positions are denoted by the black squares. The jet axis is the dashed blue line. 117
- 7.2 A toy b-jet with noise added pion tracks originating from the primary (red), secondary (blue) and tertiary (orange) vertices. The vertex positions are denoted by the black squares. The jet axis is the dashed blue line. Gaussian noise is added to the d_0, z_0, ϕ, θ (with $\sigma = 10^{-5}$) and q/P (a 0.1% error). 118
- 7.3 Histograms showing the number of tracks per jet for samples of 300,000 toy (a) b-jets, (b) c-jets and (c) light-jets. 119
- 7.4 Histograms showing the track momentum across 300,000 (a) b-jets, (b) c-jets and (c) light-jets. 120
- 7.5 Histogram of the secondary and tertiary vertex distances from the primary vertex in b-jets ((a) and (c)) and c-jets ((b)). 121
- 7.6 The training and validation loss (mean absolute error) versus the number of training epochs. 121
- 7.7 The distribution of the longitudinal impact parameter (z_0) in a toy jet sample of 3×10^5 b-jets, (a) shows the distribution of z_0 using 1000 equal spaced bins (b) shows the distribution of $|z_0|$ with a logarithmic binning. 124
- 7.8 The mean absolute and percentage errors across the six output values of each jet. 129
- 7.9 The secondary and tertiary vertex errors as measured by the distance between true and reconstructed vertices. The percentage vertex error is vertex error divided by the true vertex displacement. 130

- 7.10 An image of a jet showing the reconstructed vertices (cyan crosses), true vertices (black squares) and primary (red), secondary (blue) and tertiary (orange) tracks after noise was added. This jet had large vertex errors. The true secondary and primary vertices cannot be distinguished at this scale 131
- 7.11 An image of a jet showing the reconstructed vertices (cyan crosses), true vertices (black squares) and primary (red), secondary (blue) and tertiary (orange) tracks after noise was added. This jet had small vertex errors. 132
- 7.12 The change in the secondary and tertiary vertex error (in mm) when a track is removed from the jet. The removed track is indexed by its position in the jet after ordering by decreasing t_1 , e.g. removed track 1 has the highest value t_1 137
- 7.13 The evolution of the error in the secondary and tertiary vertex prediction as more tracks are seen. The value at N of tracks added = 0 is calculated on a jet with all null tracks. The value at N=k is calculated on a jet with the first k tracks (ordered by decreasing t_1) only, the rest are null. 138
- 8.1 The leading order cross sections of different production methods and final states as a function of heavy Higgs mass. Two models are compared, dim4 has $\rho_H = 0.05$, $f_W = f_{WW} = 0$ and dim4+6 has $\rho_H = 0.05$, $f_W = f_{WW} = 50$. (a) shows the cross-sections for the VBF and VH 2ℓ processes ($pp \rightarrow Hjj \rightarrow \ell\ell jjjj$ and $pp \rightarrow VH \rightarrow \ell\ell jjjj$). (b) compares the VBF 2ℓ process and the VBF 4ℓ process ($pp \rightarrow Hjj \rightarrow 4\ell + jj$). 142
- 8.2 Feynman diagram of the SS 2ℓ process for a generic heavy Higgs. . . 144
- 8.3 The ABCD method for the non-prompt fake background estimation. The fake factor is derived using the ratio of events in B and D, then applied to C to obtain the number of fake events in A. 154

- 8.4 The data/MC agreement in the control region B ee sub-channel. CF refers to the charge flip background. The background from non-prompt leptons is estimated using the p_T binned non-prompt electron fake factor applied to the control region D (ee sub-channel). These four distributions show the typical agreement in all the kinematic plots. 157
- 8.5 The data/MC agreement in the control region B $e\mu$ sub-channel. CF refers to the charge flip background. The background from non-prompt leptons is estimated using the $|\eta|$ binned non-prompt muon fake factor applied to the control region D ($e\mu$ sub-channel). These four distributions show the typical agreement in all the kinematic plots. 158
- 8.6 A few kinematic distributions in the Z mass validation region showing typical agreement. The charge flip contribution is estimated using the charge flip rate applied to the data in the opposite-sign lepton region instead of using the Z+jets MC. Only statistical uncertainties are shown. 160
- 8.7 The ABCD method for the photon conversion fake background estimation. The fake factor is derived using the ratio of events in B and D, then applied to C to obtain the number of fake events in A. . . 161
- 8.8 Data/MC comparison for four observables in the $Z\gamma$ ID+ID+ID control region. The $Z\gamma$ background is estimated using the photon conversion fake rate applied to the $Z\gamma \rightarrow \ell_{ID}\ell_{ID}\ell_{Anti-IBL}$ CR. Statistical uncertainties are included. 162
- 8.9 Data/MC plots in the boosted WZ CR of four variables showing typical agreement: missing Et (top left), number of jets (top right), effective mass (bottom left), invariant mass for three leptons (bottom right). A 0.93 normalization factor applied to the WZ events. Statistical and systematic uncertainties are added. 164

- 8.10 Data/MC plots in the resolved WZ CR of four variables showing typical agreement: missing E_t (top left), number of jets (top right), effective mass (bottom left), invariant mass for three leptons (bottom right). A 0.82 normalization factor applied to the WZ events. Statistical and systematic uncertainties are added. 165
- 8.11 Data/MC plots in the $ssWW$ CR of four variables showing typical agreement: missing E_t (top left), leading lepton p_t (top right), effective mass (bottom left), leading jet p_t (bottom right). WZ and $ssWW$ normalization factors are both applied and statistical and systematic uncertainties shown. 166
- 8.12 Various kinematic distributions in the boosted validation region (inclusive channel), showing the agreement of data with the estimated backgrounds. Systematic and statistical uncertainties are shown. The signal is shown stacked on top of the SM background (solid red block), the signal $\times 4$ is also shown unstacked (black line). . . . 167
- 8.13 Various kinematic distributions in the resolved validation region (inclusive channel), showing the agreement of data with the estimated backgrounds. Systematic and statistical uncertainties are shown. The signal is shown stacked on top of the SM background (solid red block), the signal $\times 4$ is also shown unstacked (black line). . . . 168
- 8.14 The product of signal acceptance and efficiency ($\mathcal{A} \times \epsilon$) for the inclusive, boosted-only and resolved-only signal regions. The quantity is displayed for a number of benchmark signals with values of m_H from 300 GeV to 2 TeV and two sets of the parameters (f_W, f_{WW}) : (1350,0) and (0,6200). 180
- 8.15 The (left) pre- and (right) post-fit distributions in the (top) boosted and (bottom) resolved validation regions. In the post-fit case, the normalization factors from the background only fit in the CR has been applied. The model is the SM background (b) plus signal (s). In this background-only fit no signal is present. 181

- 8.16 The pre- and post-fit distributions in the boosted SR from an exclusion fit using real data in the CRs and Asimov data with $\mu_{sig} = 0$ in the SRs. The model is the SM background (b) plus the GHH600fW1350fWW0 signal (s). 182
- 8.17 The pre- and post-fit distributions in the resolved SR from an exclusion fit using real data in the CRs and Asimov data with $\mu_{sig} = 0$ in the SRs. The model is the SM background (b) plus the GHH600fW1350fWW0 signal (s). 182
- 8.18 The fitted value of the nuisance parameters for the top 20 largest systematic pulls, normalization factors and statistical errors. These values are for the background-only fit. For the naming conventions, refer to table 8.13. 184
- 8.19 The fitted value of the nuisance parameters for the top 20 largest systematic pulls, normalization factors and statistical errors. These are the values from the exclusion fit using Asimov data ($\mu_{sig} = 0$) in the SR. For the naming conventions, refer to table 8.13. The γ parameters are the statistical uncertainties from the different bins. . . 185
- 8.20 The reduced correlation matrix of all nuisance parameters with an off-diagonal value greater than 0.1 (the remaining 170 nuisance parameters fell under this threshold). These are the values from the exclusion fit using Asimov data ($\mu_{sig} = 0$) in the SR. 186
- 8.21 The systematic ranking plot from the exclusion fit with Asimov data ($\mu_{sig} = 1$) in the SR and real data in the CR. The top 20 by impact nuisance parameters are shown. For the naming conventions, refer to table 8.13. 187
- 8.22 The expected upper limit of signal GHH600fW2130fWW0 is 0.34. The observed values are calculated using Asimov data in the signal region as the real data remains blinded. This model has a theoretical cross-section of 0.04731 pb. 188

- 8.23 The expected upper limit on signal strength for signal samples with 300, 600 and 900 GeV mass and a range of f_W and f_{WW} parameters. 189
- 8.24 The expected exclusion contours in the f_W vs f_{WW} plane for 300, 600 and 900 GeV heavy higgs. All parameter points outside of the contours are excluded. 190
- 8.25 The expected upper limit on the cross-section versus the heavy Higgs mass for a model with (top) $f_W = 1350$ and $f_{WW} = 0$ and (bottom) $f_W = 0$ and $f_{WW} = 6200$. The theoretical cross-section, calculated from simulation, is drawn in red. 191
- 8.26 The expected exclusion contours with heavy Higgs mass of 300 GeV, 600 GeV and 900 GeV for resolved-only fit. 192
- 8.27 The expected exclusion contours with heavy Higgs mass of 300 GeV, 600 GeV and 900 GeV for boosted-only fit. 193
- 8.28 The expected exclusion contours in the mass vs cross-section space for (top) $f_W = 1350$ and $f_{WW} = 0$ and (bottom) $f_W = 0$ and $f_{WW} = 6200$ for resolved-only fit. 194
- 8.29 The expected exclusion contours in the mass vs cross-section space for (top) $f_W = 1350$ and $f_{WW} = 0$ and (bottom) $f_W = 0$ and $f_{WW} = 6200$ for boosted-only fit. 195
- A.1 The effect on the number of tracks associated to secondary vertices in (a) light-jets (c) c-jets and (e) b-jets and the effect on the invariant mass of all tracks associated to secondary vertices in (b) light-jets (d) c-jets and (f) b-jets when using reconstructed K_S tracks in Jet-Fitter fit. The blue histogram is with no neutral tracks used in the fit (but neutral decay vertices are vetoed), the red is with the addition of K_S tracks to the fit. The ratio of the red histogram over blue histogram is plotted. 201

- B.1 Map of the JetFitter code structure showing the four main modules. BTagSecVertexing is part of the main b-tagging algorithm and directs JetFitter to find its secondary vertices by calling InDetImprovedJetFitterVxFinder. The resulting decay chain topology is passed to JetFitterVariablesFactory to create the variables for high-level taggers. The flow of information is shown by arrows. Key methods within modules are shown. This map is not an exhaustive description but rather meant to show the important information only. 203
- B.2 Map of the JetFitter code structure but with the InDetImprovedJetFitterVxFinder overloaded findSecVertex mehod shown in more detail. 204
- B.3 Map of the JetFitter code structure but with the InDetImprovedJetFitterVxFinder method doTheFit shown in more detail. 204
- B.4 Map of the JetFitter code structure but with the InDetImprovedJetFitterVxFinder module shown in more detail. 205
- B.5 Map of the JetFitter code structure but with the JetFitterRoutines module shown in more detail. 205
- B.6 Map of the refactored JetFitter code structure. InDetImprovedJetFitterVxFinder has been greatly simplified. Four new modules have been created. 206

List of Tables

2.1	The 19 free parameters of the SM and their current measured values	30
2.2	Fermion singlets/doublets and their properties. ℓ^- refers to the leptons e^- , μ^- , τ^- , ν_ℓ to the neutrinos ν_e , ν_μ , ν_τ , q_u to the up-like quarks u , c , t and q_d to the down-like quarks d , s , b	33
2.3	The branching ratios of the 125 GeV SM Higgs boson.	38
3.1	Summary of the components in the ATLAS ID. The z/R resolution refers to the resolution in z direction for barrel components, R direction for end cap components.	49
3.2	Designed resolution of the five track parameters used at ATLAS. The multiple scattering and intrinsic resolution are shown.	51
3.3	Energy resolutions obtained from test beam data. The HCAL is divided into Tile (in the barrel), HEC (end-cap) and FCAL (forward calorimeter). The first term is from stochastic noise whilst the second reflects local non-uniformities in the calorimeter response.	52
4.1	Signal electron definition used in this work.	66
4.2	Muon signal definition used in this work	67
6.1	The track variables used in IP3D and RNNIP	91
6.2	The variable inputs to the high-level taggers MV2 and DL1 (not including the SMT and RNNIP inputs). The JetFitter c-tagging variables are only used in DL1.	95
6.3	The optimised hyperparameters used in DL1.	97

6.4	The table of probabilities for merging combinations of vertices. The probability of merging vertex i and vertex j is P_{ij} assuming they form a new vertex along the b-hadron flight axis.	101
6.5	The requirements on vertex candidates and their constituent tracks to be identified as neutral particle decays. The Boolean <code>atLeastOneLoose</code> is true if at least one of the tracks has $ d_0 > 1\text{mm}$ and $ z_0 > 2\text{mm}$. r_T refers to the transverse radius of the vertex candidate. $S_{l_{sd}}$ is significance of the lifetime signed distance between the primary vertex and vertex candidate $l_{sd}/\sigma_{l_{sd}}$	104
6.6	The effect on the number of reconstructed JetFitter vertices of removing the tracks identified with neutral decay vertices. The percentage of different flavoured jets containing 0, 1, 2 or 3+ vertices is shown when the veto code is turned (a) off and (b) on. Results were produced on 130,000 selected jets from a $t\bar{t}$ sample.	106
7.1	The investigated model parameters, and the value of these in the baseline model.	120
7.2	The hyperparameters and their values in the the RNN model.	122
7.3	The final validation MSE and MAE of a model trained on 20,000 toy b-jets using a MSE loss function or MAE loss function.	123
7.4	Performance of the baseline model with modified track ordering. The performance metric is the mean absolute error on the validation set.	125
7.5	Performance of the baseline model with and without the addition of two new features (x_p, y_p) . Two track orderings were investigated. The performance metric is the mean absolute error on the validation set.	126
7.6	The resolution of track parameters in our toy jets for two different levels of noise (baseline and high noise) compared to the ATLAS design specifications.	127

7.7	The performance as measured by validation loss of the best RNN model trained on a sample of jets with the baseline noise or high noise.	127
7.8	The threshold errors calculated from toy jet samples for the high noise and baseline noise cases.	135
7.9	Performance of the RNN tagger for the baseline noise model using the derived performance metric. A vertex is considered identified if it lies within a certain threshold distance of a true vertex, given by $\sigma_{d_{min}}^{max}$ or $\sigma_{d_{min}}^{max/median}$. This threshold can be loosened to some integer multiple n . f_{vID} is the total fraction of vertices identified and $f_{j,kv}$ the fraction of jets with k identified vertices.	135
7.10	Performance of the RNN tagger for the high noise model using the derived performance metric. A vertex is identified if it lies within a certain threshold distance of a true vertex, given by $\sigma_{d_{min}}^{max}$ or $\sigma_{d_{min}}^{max/median}$. This threshold can be loosened to some integer multiple n . f_{vID} is the total fraction of vertices identified and $f_{j,kv}$ the fraction of jets with k identified vertices.	135
8.1	Expected signal and background yields in the signal regions of the three studied channels with 300 fb^{-1} [31].	144
8.2	The simulated background physics processes used in the analysis and the Matrix Element (ME) and Parton Shower (PS) generators used for each.	148
8.3	Summary of all the electron definitions used in this work. AntiID electrons must not pass the signal electron definitions.	149
8.4	Summary of all the muon definitions used in this work. AntiID muons must not pass the signal muon definitions.	150
8.5	Summary of the jet definitions used in this work.	150
8.6	The optimized selection criteria for the $SS2\ell$ signal region	152
8.7	The inclusive and p_T binned non-prompt fake rates for electrons and muons with statistical uncertainty included.	156

8.8	The inclusive and $ \eta $ binned non-prompt fake rates for electrons and muons with statistical uncertainty included.	156
8.9	The charge flip background yields in the inclusive boosted and resolved signal regions.	159
8.10	The selection cuts defining the WZ CR.	163
8.11	The selection cuts defining the $ssWW$ control region. Note not passing the boosted category means the event is vetoed if the leading reconstructed large-R jets has $p_T > 200$ GeV and $50 \text{ GeV} < M_J < 200$ GeV.	164
8.12	The validation region selection criteria.	166
8.13	Summary of the experimental systematic uncertainties.	172
8.14	The effect of applying a 1.1 or 0.9 scale factor to the prompt MC and charge flip backgrounds.	173
8.15	The effect of p_T on the photon conversion rate. Measured using two bins of equal statistical power.	174
8.16	The photon conversion rates derived using $Z\gamma$ or $W\gamma$ MC simulated samples.	174
8.17	Cross-section uncertainties for other background.	175
8.18	The normalization factors (and corresponding uncertainties) calculated from the background only fit of the M_{eff} distributions of data and MC samples.	180
8.19	The signal yields in the Boosted and Resolved SR for various different GHH mass samples.	187
C.1	The thresholds errors estimated from uniform samples across the domain of the function for $\sigma_{d_{min}}$	208
D.1	The single-lepton triggers used in this analysis	210
E.1	Simulated signal samples used in the analysis with $m_H = 300$ GeV.	212
E.2	Simulated signal samples used in the analysis with $m_H = 600$ GeV.	213
E.3	Simulated signal samples used in the analysis with $m_H = 900$ GeV.	214

E.4	Simulated signal samples used in the analysis for mass scan. . . .	215
-----	--	-----

Chapter 1

Introduction

The high energy particle collisions at the LHC offer physicists an unprecedented opportunity to study fundamental physics. Protons are collided at a 13 TeV centre of mass energy, on average a billion times per second. This big data challenge has led to the research and use of algorithmic tools such as Machine Learning at the LHC. The collision events at the LHC are best understood through the Standard Model of Particle Physics. Much of the research program of collaborations such as ATLAS is dedicated to investigating the predictions of this model and its limitations. ATLAS physicists search for new physics, experimental signatures that lie beyond the predictive power of the Standard Model.

This thesis covers the work of the author as a member of the ATLAS collaboration. Personal contributions are particularly found in chapters 6, 7 and 8. The earlier chapters review all necessary background material. The thesis is structured as follows.

Chapter 2 conducts a review of the particle physics theory relevant in this work. The Standard Model of particle physics is presented here, alongside theoretical extensions under research.

Chapter 3 describes the ATLAS detector and the LHC. This instrument was used to gather experimental data used in the work.

Chapter 4 focuses on the data processing chain of ATLAS. The physics objects used in ATLAS analysis are described alongside the format of observed data, and predicted data from Monte Carlo simulation.

Chapter 5 is a review of the machine learning tools and theory used in this work.

Chapter 6 reviews the b -tagging algorithms used in ATLAS, with the author's contributions to the JetFitter algorithm covered in sections 6.5 and 6.6.

Chapter 7 presents the author's research into topological b -tagging with recurrent neural networks.

Chapter 8 describes the search for generic heavy versions of the Higgs boson. The author was a member of the analysis team behind this work. Attention will be drawn towards the author's personal contributions to this work.

The work presented here and future outlook will be summarized in chapter 9. As part of the author's research program, a six month internship was carried out with industry partner Faculty. The author worked on the use of variational inference and its application to data privacy [1]. This work will not be presented in this thesis.

Chapter 2

Theory

2.1 Particle Physics at ATLAS

The Standard Model of Particle Physics (SM) is the theoretical framework underpinning experimental research at ATLAS [2, 3, 4]. It provides a description of three of the fundamental forces of nature: Electromagnetism, the Weak interaction and the Strong interaction. The SM is a relativistic quantum field theory with a $SU(3) \times SU(2) \times U(1)$ gauge symmetry. A Lagrangian density is written down constrained by these symmetries and the requirement of renormalizability. From here the dynamics and kinematics are determined, dependent only on 19 experimentally observed free parameters¹.

The phenomenology of the Standard Model is generally expressed in terms of fundamental particles. The fundamental fermions, having half-integer spin, comprise the leptons and quarks. Leptons can interact via the weak force and electromagnetism, whilst quarks additionally interact through the strong force. Leptons and quarks come in three mass generations. The fundamental bosons, having integer spin, are the force-carrying particles which mediate the strong, weak and electromagnetic interactions. The Higgs boson, the only scalar particle in the SM, is the most recently discovered particle [6]. It is responsible for the masses of gauge bosons through the Higgs Mechanism (see section 2.1.2). The particle content of

¹An additional 7 free parameters are usually added to the SM to account for neutrino masses (and the PMNS matrix). These are not relevant to this work and thus not covered here. For a more full description see [4].

Symbol	Description	Value
m_e	Electron mass	0.511 MeV
m_μ	Muon mass	105.7 MeV
m_τ	Tau mass	1.78 GeV
m_u	Up quark mass	1.9 MeV
m_d	Down quark mass	4.4 MeV
m_s	Strange quark mass	87 MeV
m_c	Charm quark mass	1.32 GeV
m_b	Bottom quark mass	4.24 GeV
m_t	Top quark mass	173.5 GeV
θ_{12}	CKM 12-mixing angle	13.1°
θ_{23}	CKM 23-mixing angle	2.4°
θ_{13}	CKM 13-mixing angle	0.2°
δ	CKM CP violation Phase	0.995
g'	U(1) gauge coupling	0.357
g	SU(2) gauge coupling	0.652
g_s	SU(3) gauge coupling	1.221
θ_{QCD}	QCD vacuum angle	≈ 0
v	Higgs vacuum expectation value	246 GeV
m_H	Higgs mass	125 GeV

Table 2.1: The 19 free parameters of the SM and their current measured values [5]

the SM is summarized in figure 2.1.

2.1.1 Interactions of the SM

2.1.1.1 The Strong Interaction

Quantum Chromodynamics (QCD) is the quantum field description of the Strong force. The mediator of the strong force is the gluon. Quarks and gluons carry a color charge that allow them to interact through the strong force, the strength of this interaction determined by the strong coupling constant, α_s . The value of α_s can be determined from QCD theory through a mathematical procedure known as the renormalization group. This leads to the phenomenon of the running of α_s , the value of the coupling strength is dependent on the momentum/length scale² at which it is probed. QCD exhibits asymptotic freedom [8], the interaction strength of the strong force decreases at high momentum (small distance) allowing perturbative calculations. Conversely, at low energy and large length scales the interaction

²The momentum and length scales are inversely related by quantum mechanics.

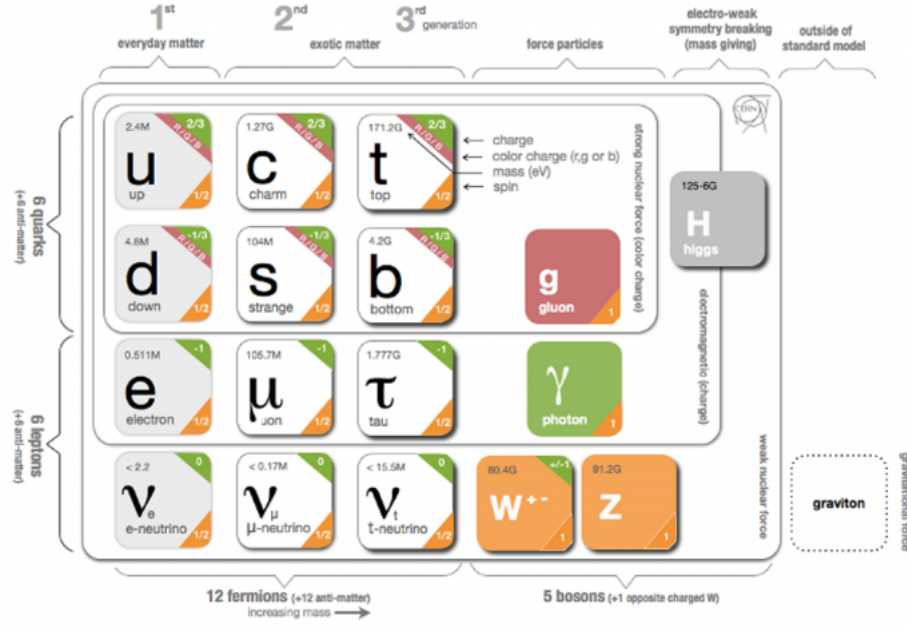


Figure 2.1: The particles of the SM [7]

becomes much stronger. This leads to the observed phenomenon of colour confinement: particles with a non-zero colour charge cannot exist independently. Quarks and gluons found in nature are confined to colourless hadrons. The two common forms are mesons, formed of a quark and anti-quark, and baryons, formed of either three quarks or three anti-quarks³.

Quarks can be produced as the final states of interactions in the LHC. The cross-sections of these interactions can be calculated perturbatively thanks to asymptotic freedom. The quarks/gluons produced in these events will be observed as jets, a collimated group of hadronic particles. These jets are produced by the hadronization/fragmentation of the quark or gluon. The hadronization process cannot be calculated directly from QCD theory, but is well studied experimentally. Several different models for the process exist, such as the Lund String Model [9] used by the Pythia event generator [10]. Experimental observation of hadronization shows differing behaviour for each quark and gluon [11]. In particular, the heavier b-quark carries a much larger mean fraction of energy into its final b-hadron than

³Three colours (rgb) and three anti-colours ($\bar{r}\bar{b}\bar{g}$) exist. A combination of all three colours or a colour and its anti-colour give a colourless object.

the lighter quarks. Top quarks decay weakly before hadronization occurs and can be treated as free particles [12].

2.1.1.2 The Electroweak Interaction

Electroweak theory is the quantum field description of electromagnetism and the weak interaction. Electromagnetism is the force governing interactions between all electrically charged particles, for a more detailed description see [2, 3, 4]. The weak interaction is responsible for the decay of fermions between generations. Its force carrying particles are the $W^\pm$ and Z bosons. The weak interaction allows quarks to change their flavour through weak decay. It is responsible for the decay of heavy leptons, such as the muon and tau, and hadrons⁴, through the decay of quarks.

The electroweak framework contains four fundamental fields under an $SU(2)_L \times U(1)_Y$ gauge symmetry. The fields are denoted $\vec{W}_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$ and B_μ . By Noether's theorem, the gauge symmetries entail two conserved quantities: the third component of weak isospin T_3 and weak hypercharge Y . After electroweak symmetry breaking we obtain the $W^\pm$ and Z boson fields, which are responsible for the weak force, and the photon or electromagnetic field A_μ (see section 2.1.2). In electroweak theory, quarks and leptons gain a weak isospin and hypercharge and are organised into left- and right-handed chiral doublets or singlets (see table 2.2). The electric charge $Q = T_3 + \frac{1}{2}Y$ becomes the conserved quantity after electroweak symmetry breaking via the Higgs mechanism (discussed in section 2.1.2).

The $W^\pm$ bosons interact with the fundamental fermions through the charged-current interaction. This allows a quark or lepton to transform into its doublet counterpart, for example a left-handed electron into an electron neutrino (see figure 2.2). The W boson is unstable and decays rapidly, the W decays and corresponding

⁴Hadrons can also decay through the strong or electromagnetic forces, those decaying exclusively by the weak force have longer lifetimes

Fermion Singlet/Doublet	chirality	Q	T_3	Y
$\begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix}_L$	Left	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} -1 \\ -1 \end{pmatrix}$
$\begin{pmatrix} q_u \\ q_d \end{pmatrix}_L$	Left	$\begin{pmatrix} +\frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} +\frac{1}{3} \\ +\frac{1}{3} \end{pmatrix}$
$(\ell^-)_R$	Right	(-1)	(0)	(-2)
$(q_u)_R$	Right	$(+\frac{2}{3})$	(0)	$(+\frac{4}{3})$
$(q_d)_R$	Right	$(-\frac{1}{3})$	(0)	$(-\frac{2}{3})$

Table 2.2: Fermion singlets/doublets and their properties. ℓ^- refers to the leptons e^- , μ^- , τ^- , ν_ℓ to the neutrinos ν_e , ν_μ , ν_τ , q_u to the up-like quarks u , c , t and q_d to the down-like quarks d , s , b

branching fractions are as follows [5]:

$$\text{Br}(W \rightarrow e \bar{\nu}_e) = 0.1046 \pm 0.0042 \pm 0.0014, \quad (2.1)$$

$$\text{Br}(W \rightarrow \mu \bar{\nu}_\mu) = 0.1050 \pm 0.0041 \pm 0.0012, \quad (2.2)$$

$$\text{Br}(W \rightarrow \tau \bar{\nu}_\tau) = 0.1075 \pm 0.0052 \pm 0.0021, \quad (2.3)$$

$$\text{Br}(W \rightarrow \text{hadrons}) = 0.6832 \pm 0.0061 \pm 0.0028. \quad (2.4)$$

The weak interaction with quarks is more complex. The weak eigenstates of quarks are mixtures of the mass eigenstates, as determined by the CKM matrix [5]:

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.0410 \pm 0.0014 \\ 0.0080 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.030 \end{bmatrix} \quad (2.5)$$

The weak decay vertex for quarks thus picks up an extra term from the CKM matrix. This can be seen from the Feynman diagram in figure 2.3. Quark decays are greatly affected by this as the transition rate picks up a factor of $|V_{ij}|^2$.

2.1.2 Higgs Mechanism

The Higgs Mechanism was introduced to the SM to explain the observed masses of the W and Z gauge bosons [13, 14, 15, 16, 17, 18]. In gauge field theory, simple

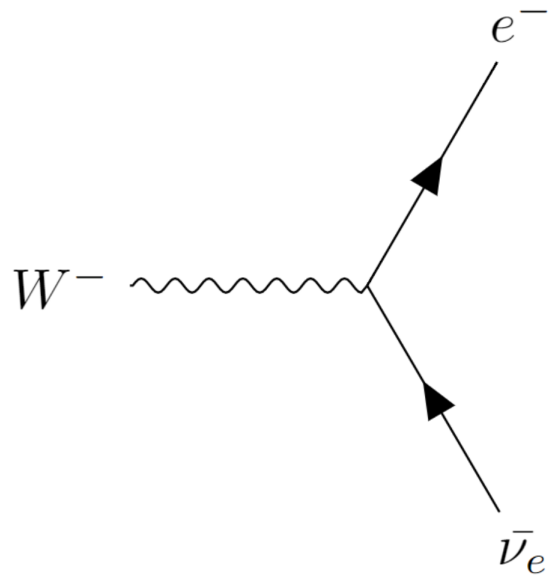


Figure 2.2: Weak interaction vertex with an electron and electron neutrino.

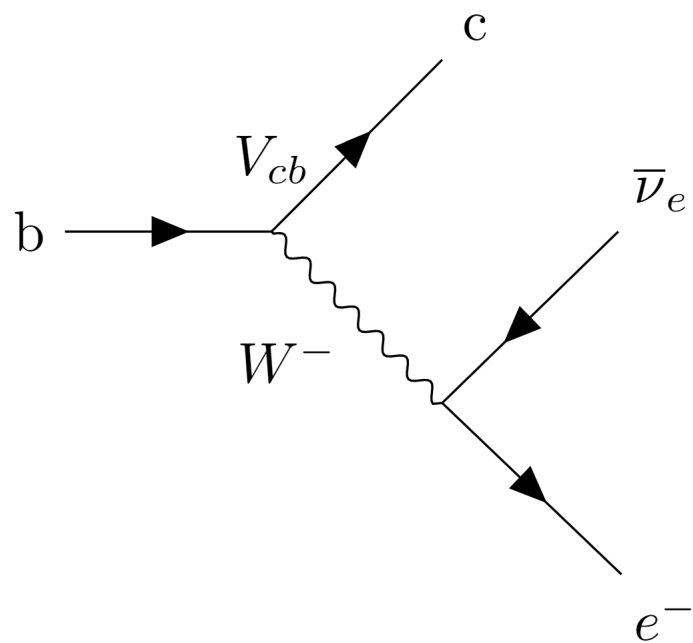


Figure 2.3: A b quark decaying to c via the weak interaction. The b to c vertex picks up a factor V_{cb} from the CKM matrix.

mass terms of the form:

$$\mathcal{L}_{mass} = -m^2 A^\mu A_\mu \quad (2.6)$$

are not allowed as these are not invariant under a gauge transformation (for example the $U(1)$ gauge transformation $A^\mu \rightarrow A^\mu - \partial^\mu \theta(x)$). Gauge invariance naively prevents us adding SM mass terms for the W , Z and chiral fermions. The solution is to introduce a new scalar field, the Higgs field, with a non-zero vacuum expectation value. This field couples to the gauge bosons and fermions giving them mass in a gauge-invariant manner.

The Higgs field, Φ is an $SU(2)$ doublet:

$$\Phi = \begin{pmatrix} \phi_a \\ \phi_b \end{pmatrix} \quad (2.7)$$

where the complex scalar fields ϕ_a and ϕ_b are written in terms of real scalar fields:

$$\phi_a = \phi_1 + i\phi_2, \phi_b = \phi_3 + i\phi_4 \quad (2.8)$$

The Lagrangian for the Higgs field is written as:

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) \quad (2.9)$$

with the Higgs Potential defined as:

$$V(\Phi) = \lambda (\Phi^\dagger \Phi)^2 + \mu^2 \Phi^\dagger \Phi \quad (2.10)$$

The covariant derivative D_μ is defined according to the gauge symmetry of the electroweak $SU(2)_L \times U(1)_Y$ sector⁵:

$$D_\mu = \partial_\mu + i\frac{g}{2}\sigma_j W_\mu^j + iY\frac{g'}{2}B_\mu \quad (2.11)$$

where the $W_\mu^j = (W_\mu^1, W_\mu^2, W_\mu^3)$ and B_μ are electroweak gauge fields (before sym-

⁵The Higgs field does not, as far as we know, directly couple to gluons.

metry breaking), Y is the weak hypercharge and σ_j are the Pauli matrices. Spontaneous symmetry breaking occurs when the Higgs potential has $\mu^2 < 0$. This leaves a non-zero vacuum expectation value corresponding to the minimum of the potential $V(\Phi)$. It is possible to expand the Higgs field in terms of a single scalar field, this is called the unitary gauge:

$$\Phi = \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (2.12)$$

Where $v^2 = \frac{\mu^2}{\lambda}$. Combining this with the Lagrangian in equation (2.9) and performing the electroweak symmetry breaking produces:

$$\begin{aligned} \mathcal{L}_\Phi = & \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{g^2}{4}(v+h)^2(W_\mu^+ W^{-\mu}) \\ & + \frac{1}{8}(g^2 + g'^2)(v+h)^2 Z_\mu Z^\mu + \frac{\mu^2}{2}(v+h)^2 - \frac{\lambda}{4}(v+h)^4 \end{aligned} \quad (2.13)$$

Where we have used the convention:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \quad (2.14)$$

$$Z_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gW_\mu^3 - g'B_\mu), \quad (2.15)$$

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(g'W_\mu^3 + gB_\mu) \quad (2.16)$$

Excitations of these fields give rise to their respective bosons. The $W_\mu^\pm$ and Z_μ fields give the $W^\pm$ and Z bosons, the electromagnetic field A_μ gives the photon and the Higgs field h gives the Higgs boson. We can read off the mass of the electroweak

bosons in terms of v :

$$m_W = \frac{1}{2}g v \quad (2.17)$$

$$m_Z = \frac{1}{2}v\sqrt{(g^2 + g'^2)} \quad (2.18)$$

$$m_\gamma = 0 \quad (2.19)$$

$$m_h = \sqrt{-2\mu^2} \quad (2.20)$$

2.1.2.1 Fermion Masses

To obtain the fermion masses, Yukawa coupling terms are added to the SM Lagrangian. The Yukawa coupling for a fermion f to the Higgs field is:

$$\mathcal{L}_{Yukawa} = -G_f(\bar{\Psi}_L\Phi\Psi_R + \bar{\Psi}_R\Phi\Psi_L) \quad (2.21)$$

where $\Psi_{L/R}$ are the left/right handed fermion isospin fields (doublet or singlet), Φ is the complex scalar Higgs field and G_f is the Yukawa coupling (determined experimentally). We can plug in the Higgs field from eq. (2.24) and the fermion doublets to obtain the mass terms. For example, the first lepton generation doublets $(\nu, e)_L$ and $(e)_R$ give:

$$\mathcal{L} = -\frac{G_e(v + h(x))}{\sqrt{2}}(\bar{e}e) \quad (2.22)$$

An additional term is required for the masses of up-like quarks:

$$\mathcal{L}_{Yukawa,up} = -G_f(\bar{\Psi}_L\tilde{\Phi}^c\Psi_R + \bar{\Psi}_R\tilde{\Phi}^c\Psi_L) \quad (2.23)$$

where

$$\tilde{\Phi}^c = \begin{pmatrix} v + h(x) \\ 0 \end{pmatrix} \quad (2.24)$$

The Yukawa coupling terms for each fermion are then determined by experimentally measuring v and the fermion mass⁶.

⁶The neutrinos require special consideration as the method for neutrino mass generation is not currently known. For an example of neutrino mass generation see [5].

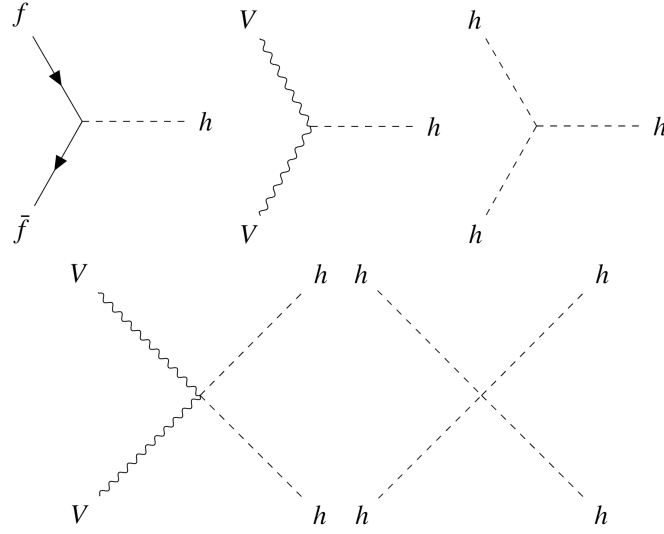


Figure 2.4: The possible Higgs vertices from the SM. f is any massive fermion (leptons or quarks, but not neutrinos). V is any massive boson ($W^\pm$ or Z)

Decay channel	Branching ratio	Rel. uncertainty
$H \rightarrow \gamma\gamma$	2.28×10^{-3}	+5.0% -4.9%
$H \rightarrow ZZ$	2.64×10^{-2}	+4.3% -4.1%
$H \rightarrow W^+W^-$	2.15×10^{-1}	+4.3% -4.2%
$H \rightarrow \tau^+\tau^-$	6.32×10^{-2}	+5.7% -5.7%
$H \rightarrow b\bar{b}$	5.77×10^{-1}	+3.2% -3.3%
$H \rightarrow Z\gamma$	1.54×10^{-3}	+9.0% -8.9%
$H \rightarrow \mu^+\mu^-$	2.19×10^{-4}	+6.0% -5.9%

Table 2.3: The branching ratios of the 125 GeV SM Higgs boson [5].

With the addition of the Yukawa couplings, the SM Lagrangian is complete. Perturbation theory can give us the Feynman vertices of the Higgs boson (shown in figure 2.4). The branching ratios of the most common Higgs decays are shown in table 2.3.

2.1.3 Heavy Quark Phenomenology

The phenomenology of heavy quarks is of particular importance in this thesis and is described here. For more detail on the phenomenology of other particles, such as

leptons, see any of the references [2, 3, 4].

The strong and weak interactions both have important effects on quark phenomenology as detailed in section 2.1.1. The two heaviest quarks, the top (t) and bottom (b), are of particular interest to scientists at the LHC. The b quark is the most common decay product of the Higgs boson and the top quark. The top quark, as the most massive fermion, is important in the searches for BSM physics at high energy scales.

The t quark is the heaviest known fermion. This leads to its extremely short lifetime (around 5×10^{-25} s). It decays via the weak force into a b quark, although the mixing of the CKM matrix means it can very rarely decay to an s or d quark. The branching ratio for $t \rightarrow W^+ b$ was measured by the CMS experiment at around 1.014 ± 0.003 (stat.) ± 0.032 (syst.) [19]. This gives the CKM matrix value of $|V_{tb}| \approx 1$, consistent with unitarity. At the LHC, top quarks are therefore expected to decay rapidly producing a b-jet and a W boson, which will further decay either leptonically or hadronically.

The b quark is the second heaviest fermion, albeit much less massive than the top. It might therefore be expected to decay rapidly, however the weak decay of b is suppressed as $|V_{ub}|^2 \ll |V_{cb}|^2 \ll 1$. Hadrons containing b quarks consequently have lifetimes of the order 10^{-12} s. Combined with their large mass (the b is four times heavier than a proton), this gives b hadrons a very distinctive signature in collider physics (see section 6). The b hadrons will mainly decay to hadrons containing c quarks [5], as $|V_{cb}|$ is an order of magnitude larger than $|V_{ub}|$. The weak decays of b hadrons not involving $b \rightarrow c$ are called rare decays. These have branching fractions with values around 10^{-5} [5]. The W produced in the b-decay then decays to hadrons or leptons, this latter is termed a semi-leptonic decay.

2.2 Beyond the Standard Model

There are numerous reasons to look for beyond Standard Model (BSM) physics. Arguments for BSM broadly fall into two categories. Firstly, the SM is limited to modelling the behaviour of three fundamental forces and particles at the small

scale, it is not a complete model of all physics. The second category of arguments are more philosophical, stemming from desires for naturalness and elegance.

2.2.1 Problems with the SM

2.2.1.1 Gravity

The SM provides no description of the force of gravity. Our current best theory of gravity is General Relativity. Attempts to express General Relativity as a quantum field theory have so far proven unsuccessful. It can, however, be expressed as an effective field theory up to a certain energy scale. The energy scale at which this effective field description breaks down is known as the Planck mass, $M_{Planck} \sim 10^{19}$ GeV.

2.2.1.2 Dark Matter

Cosmological observations of the universe suggest only 15% of all matter in the universe is formed from the SM particles. The other 85% of matter appears to only interact via gravity and is thus difficult to detect [20]. Evidence for dark matter comes from many cosmological observations such as gravitational lensing, galaxy rotation curves, the cosmic microwave background and many more phenomena [21]. The standard model of cosmology [22], which has general relativity as its mathematical basis, requires dark matter to fit current observations of the universe.

2.2.1.3 Dark Energy

The standard model of cosmology further predicts 68% of all energy in the universe is so called dark energy [22]. This energy is responsible for the observed accelerating expansion of the universe. In the mathematics of general relativity, it is characterized by the cosmological constant. This can be interpreted as the energy density of vacuum. The SM currently could offer an explanation through vacuum energy, quantum fields have a non-zero energy value at all points in space, however this gives a value almost 120 orders of magnitude off from the observed value of dark energy [23, 24].

2.2.1.4 Neutrino masses

The observations of neutrino oscillations suggest that neutrinos have a non-zero mass [25, 26]. To introduce a neutrino mass into the SM requires the addition of right-handed sterile neutrinos or majorana mass terms [5]. To date, no experimental evidence for either has been found.

2.2.1.5 Hierarchy Problem

The hierarchy problem relates to the observed large difference between the mass of the Higgs (~ 125 GeV) and the Planck Mass ($\sim 10^{19}$ GeV). The energy scale of the physics described by the SM is vastly different to the energy scale of quantum gravity. This leads to questions of naturalness in a theory where one force is vastly different to the rest [27]. It further leads to excessive fine-tuning.

The observed mass of the Higgs would be expected to have contributions at the order of the Planck scale. Consider a fermion, f , with a Yukawa coupling, G_f to the Higgs field. The Feynman diagram in figure 2.5 gives a quadratic order correction to the Higgs mass [27]:

$$\delta m_h^2 = -\frac{G_f^2}{32\pi^2} \Lambda_{UV}^2 \quad (2.25)$$

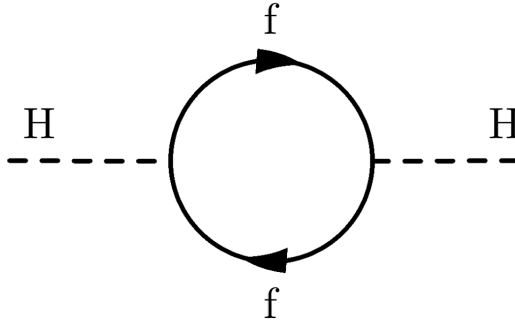


Figure 2.5: Feynman diagram of the first loop correction to the bare Higgs mass from a fermion f .

The negative sign in eq. (2.25) is due to the fermion spin-statistics. Every particle which couples to the Higgs contributes a loop correction. Bosons, such as the W and Z, contribute similar quadratic loop terms of order Λ_{UV}^2 but with a positive sign. If we take the UV cutoff to be the Planck mass, we find a huge fine-

tuning of the bare Higgs mass is required to produce the observed value of 125 GeV. This large fine-tuning is seen as unnatural, and as an argument for BSM.

2.2.1.6 The Structure of the SM

There are several things about the SM that still puzzle physicists. One is the existence of three, and only three, generations of fermions. The SM relies on 19 (26 with neutrino masses) free parameters which must be experimentally determined. Physicists would like to understand the origins of these values and the SM structure, and many BSM theories attempt to solve this.

2.2.2 Examples of BSM

There exist many examples of BSM theories produced to solved the experimental and theoretical challenges of the SM. Examples include SUSY [28], which can offer a solution to the hierarchy problem, or string theory [29], which attempts to reconcile gravity and quantum field theory. Some of these theories offer phenomenological predictions we can search for at the LHC. In experimental research, it is more common to start with a simplified model that targets a particular signature. This simplified model could be explained as a consequence of a more complete BSM model. An example of this is the generic heavy Higgs model in 2.2.2.2.

2.2.2.1 SUSY

Supersymmetry posits that every particle has a so called super partner with different spin statistics. They differ only by their spin quantum number until a spontaneously broken symmetry is introduced to allow different masses, such as in the minimal supersymmetric Standard Model (MSSM) [28]. Thus all fermions have a bosonic superpartner and vice versa. The new particles introduced by theories such as MSSM offer potential candidates for dark matter. As an example the neutralino predicted by MSSM is massive, stable and interacts only through the weak force. This would make it an ideal candidate for dark matter (as a Weakly-Interacting Massive Particle (WIMP)). Supersymmetric theories additionally offer a neat solution to the hierarchy problem. For example, consider the superpartner, $\tilde{f}$ of the fermion f in section 2.2.1.5. The loop correction to the Higgs mass comes from the Feynman diagram

in figure 2.6. This could cancel the mass correction term of figure 2.5 as it will have opposite spin statistics and thus the interaction term has an opposite sign (but equal couplings), solving the issues of fine-tuning⁷.

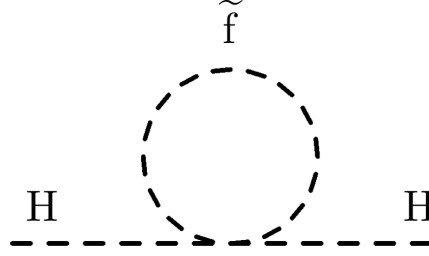


Figure 2.6: Feynman diagram of the first loop correction to the bare Higgs mass from a superpartner $\tilde{f}$ of fermion f

2.2.2.2 Extra Higgs Bosons

In many BSM models such as the MSSM or NMSSM⁸, multiple Higgs particles are predicted beyond the 125 GeV Higgs measured at the LHC [28]. It is possible to predict the phenomenology of a heavy version of the Higgs in a model-independent way [30, 31, 32]. Suppose there are multiple Higgs $SU(2)_L$ fields such as those in equation (2.7): $(\Phi_1, \Phi_2, \dots)$. Next, add the most generic multi-Higgs potential $V(\Phi_1, \Phi_2, \dots)$. These fields can mix to produce the mass eigenstates. Consider the lightest mass eigenstate, Φ_h , and the next-to-lightest mass eigenstate, Φ_H . The particle h would correspond to the 125 GeV particle observed at the LHC⁹. The particle H is the next lightest Higgs particle not yet discovered.

The most general Lagrangian will, in this case, give the following order 4

⁷This cancellation is not exact in models such as MSSM due to the different masses of superpartners. A residual fine-tuning is needed to the order of the $m_{\tilde{f}}$, this is known as the little hierarchy problem [27]

⁸next-to-minimal supersymmetric Standard Model

⁹In general, this doesn't have to be the case.

couplings to the vector bosons [30]:

$$\mathcal{L}_{hWW}^{(4)} = \rho_h g m_W h W^\mu W_\mu, \quad (2.26)$$

$$\mathcal{L}_{hZZ}^{(4)} = \rho_h \frac{g m_W}{2 \cos^2 \theta_W} h W^\mu W_\mu, \quad (2.27)$$

$$\mathcal{L}_{HWW}^{(4)} = \rho_H g m_W H W^\mu W_\mu, \quad (2.28)$$

$$\mathcal{L}_{HZZ}^{(4)} = \rho_H \frac{g m_W}{2 \cos^2 \theta_W} H W^\mu W_\mu \quad (2.29)$$

where θ_W is the weak mixing angle and m_W is the W boson mass. The scaling factors ρ_h and ρ_H arise as the vacuum expectation values (v.e.v.) and the couplings, g , are in general different from the SM values. If there are only two Higgs doublets, then it is possible to write these explicitly [32]:

$$\rho_h = \cos(\beta - \alpha) \quad (2.30)$$

$$\rho_H = \sin(\beta - \alpha) \quad (2.31)$$

Where β is defined from the v.e.v. of the two fields (v_1 and v_2):

$$\tan \beta = \left| \frac{\langle \Phi_2^0 \rangle}{\langle \Phi_1^0 \rangle} \right| \equiv \left| \frac{v_2}{v_1} \right| \quad (2.32)$$

and α is a mixing angle relating the Higgs doublets Φ_1, Φ_2 to the neutral scalar Higgs and heavy Higgs fields, h and H :

$$\begin{pmatrix} \sqrt{2} \text{Re}(\Phi_2^0) - v_2 \\ \sqrt{2} \text{Re}(\Phi_1^0) - v_1 \end{pmatrix} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} h \\ H \end{pmatrix} \quad (2.33)$$

For a SM-like light Higgs ρ_h must be close to 1.

Dimension-6 effective operators for the heavy Higgs can also be written to encapsulate physics at a higher energy scale [30]:

$$\mathcal{L}_{HVV}^{(6)} = \sum_n \frac{f_n}{\Lambda^2} \mathcal{O}_n \quad (2.34)$$

Where Λ is the scale up to which the effective field description is valid and $\mathcal{O}_n$ are

all the gauge invariant six-dimensional operators. The effective Lagrangian terms involving the heavy Higgs and the vector gauge bosons are found after electroweak symmetry breaking¹⁰:

$$\mathcal{L}_{HWW}^{(6)} = gm_W \frac{f_W}{2\Lambda^2} \left(W_{\mu\nu}^+ W^{-\mu} \partial^\nu H + h.c. \right) - gm_W \frac{f_{WW}}{\Lambda^2} W_{\mu\nu}^+ W^{-\mu\nu} H, \quad (2.35)$$

$$\mathcal{L}_{HZZ}^{(6)} = gm_W \frac{c^2 f_W + s^2 f_B}{2c^2 \Lambda^2} Z_{\mu\nu}^+ Z^{-\mu} \partial^\nu H - gm_W \frac{c^4 f_{WW} + s^4 f_{BB}}{2c^2 \Lambda^2} Z_{\mu\nu}^+ Z^{-\mu\nu} H \quad (2.36)$$

where $s = \sin \theta_W$ and $c = \cos \theta_W$. The terms f_W , f_{WW} , f_B , f_{BB} are coupling strengths to be determined experimentally. These follow from equation (2.34) with the subscripts relating to the fields W and B (see section 2.1.1.2 and ref. [30]). Terms of $O(s^2)$ and $O(s^4)$ will be small and could be neglected.

This generic heavy Higgs model leads to the possibility of an experimental signature in the ATLAS detector at the LHC. The addition of dimension-6 terms has interesting effects on phenomenology. It enhances the cross-section of certain production modes, in particular associated vector boson production [31]. This particular mode has not been searched for as of yet at the LHC. It has a particularly clean signal (low standard model backgrounds) in its same-sign two lepton final state: $HW^\pm \rightarrow W^\pm W^\mp W^\pm \rightarrow \ell^\pm \nu q \bar{q} \ell^\pm \nu$. By searching for such a signal with the ATLAS detector, evidence for an undiscovered heavy Higgs boson could be found. In the absence of such a signal, limits could be set on BSM model parameters such as f_W and f_{WW} . A search for such a generic heavy Higgs is presented in chapter 8.

¹⁰We ignore the triple gauge couplings as ATLAS and CMS have imposed strong constraints on these [30].

Chapter 3

The ATLAS Detector at the LHC

3.1 The Large Hadron Collider (LHC)

Particle accelerators allow us to study fundamental physics at high energies and extremely small scales. The Large Hadron Collider (LHC) [33] at the European Organisation for Nuclear Research (CERN) is the most powerful accelerator currently in operation. The LHC accelerates beams of protons along two 27km long circular tunnels using powerful superconducting magnets. These protons then collide at up to 13 TeV centre of mass energies, with the resulting events measured by three main detectors along the ring: ATLAS [34], CMS [35] and LHCb [36]. Several other detectors exist such as ALICE [37], which observes the collision of large ions also accelerated by the LHC.

The CERN accelerator complex is sketched in figure 3.1. Protons, from ionised Hydrogen gas, are first accelerated up to 50 MeV by Linac 2. These are then accelerated to 450 GeV through the Proton Synchrotron Booster (up to 1.5 GeV), Proton Synchrotron (25 GeV) and finally the Super Proton Synchrotron (450 GeV). The protons are then fed into the LHC, where they are accelerated up to 6.5 TeV. Two proton beams travelling in opposite directions then collide producing 13 TeV centre of mass energy events. The protons in beams travel in bunches of up to 1.1×10^{11} particles spaced 25 ns apart. Run 2 of the LHC finished in 2018 delivering 150 fb^{-1} of integrated luminosity¹ [38].

¹Integrated luminosity: $L = N_{evt}/\sigma$, where σ is the cross-section of a process, and N_{evt} is the number of events of this process

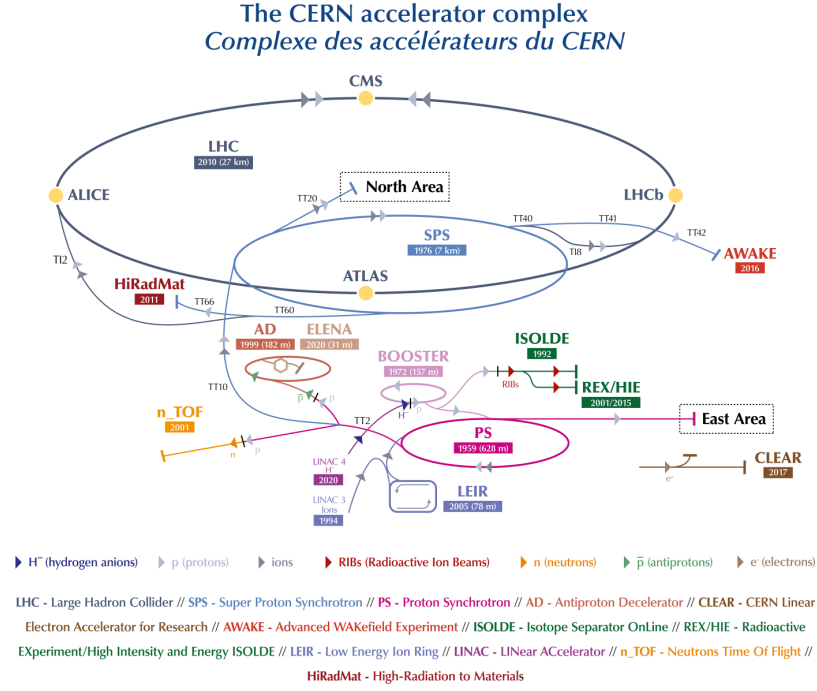


Figure 3.1: CERN's accelerator complex [42]

The large number of protons in a bunch can lead to multiple collision events per bunch crossing. This helps to boost the luminosity of the LHC but also leads to pile-up, as the measurement of a collision is affected by other collisions in the same bunch crossing (in-time pile-up) and other bunch crossings (out of time pile-up) [39].

The LHC is currently shut down in preparation for run 3 (delayed until spring 2022 due to COVID-19). After run 3, the LHC will be upgraded in preparation for the High-Luminosity LHC (HL-LHC) [40]. The HL-LHC will see ~ 200 collisions per bunch crossing for an expected $3000\text{--}4000 \text{ fb}^{-1}$ of integrated luminosity throughout its scheduled lifetime. In contrast, current run conditions have around $\sim 40 - 60$ collisions per bunch crossing [41]. Pile-up effects will consequently increase.

3.2 The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) detector [34] is a large cylindrical detector with the LHC beam pipe running along its central axis. It is a multi-purpose

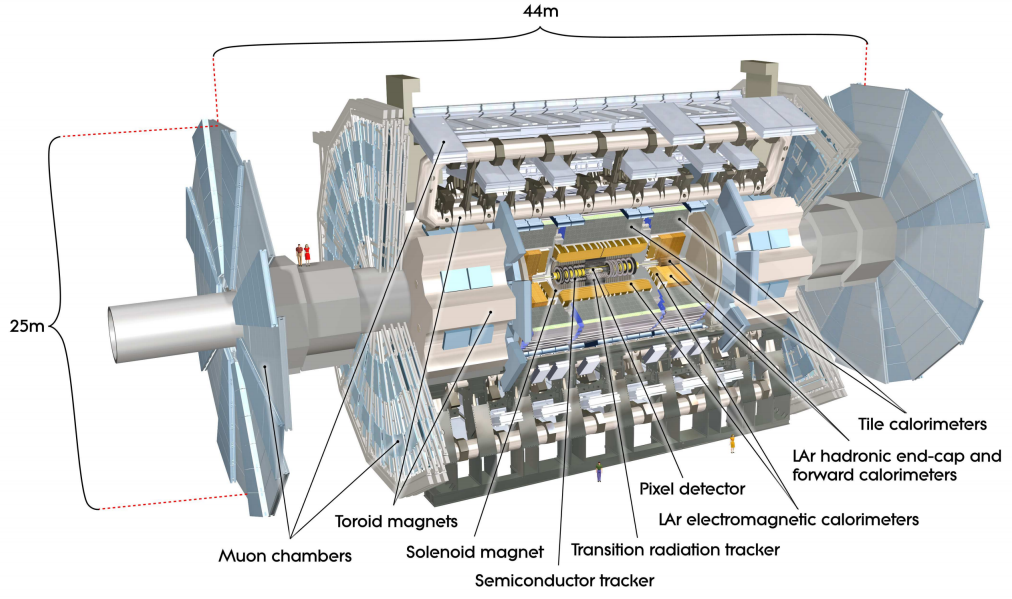


Figure 3.2: Schematic of the ATLAS detector [34]

detector and has many components to help study proton-proton collision events. The three main components are the sub-detectors: the Inner Detector, Calorimeter System and Muon Spectrometer. These form concentric rings around the beam-pipe in the barrel of the detector. Two end caps additionally provide extra coverage for forward physics. A schematic of the detector is shown in figure 3.2.

3.2.1 Co-ordinate System

A right-handed Cartesian coordinate system is used for ATLAS with the z -axis along the beam-pipe, the y -axis pointing up away from the earth's centre and the x -axis pointing towards the centre of the LHC ring. The azimuthal angle (ϕ) is defined starting from the x -axis going around the z -axis. The polar angle (θ) is the angle from the z -axis ($\theta = 0$ is along the z -axis). The radius $r = \sqrt{x^2 + y^2}$.

The momentum of a particle in Cartesian coordinates is:

$$\mathbf{p} = (p_x, p_y, p_z) \quad (3.1)$$

The transverse momentum is the projection of $\mathbf{p}$ onto the x - y plane: $p_T = |\mathbf{p}| \sin \theta$.

Component	Angular Coverage	Element Size	$R - \phi$ Resolution	z/R Resolution
IBL	$ \eta < 2.5$	$50 \times 250 \mu\text{m}^2$	$8 \mu\text{m}$	$40 \mu\text{m}$
Pixel (barrel)	$ \eta < 1.7$	$50 \times 400 \mu\text{m}^2$	$10 \mu\text{m}$	$115 \mu\text{m}$
Pixel (End Cap)	$1.7 < \eta < 2.5$	$50 \times 400 \mu\text{m}^2$	$10 \mu\text{m}$	$115 \mu\text{m}$
SCT (barrel)	$ \eta < 1.4$	$80 \mu\text{m}$	$17 \mu\text{m}$	$580 \mu\text{m}$
SCT (End Cap)	$1.4 < \eta < 2.5$	$80 \mu\text{m}$	$17 \mu\text{m}$	$580 \mu\text{m}$
TRT (barrel)	$ \eta < 0.7$	$4000 \mu\text{m}$	$130 \mu\text{m}$	
TRT (End Cap)	$0.7 < \eta < 2.0$	$4000 \mu\text{m}$	$130 \mu\text{m}$	

Table 3.1: Summary of the components in the ATLAS ID. The z/R resolution refers to the resolution in z direction for barrel components, R direction for end cap components. [34, 43]

A transformation of the polar angle known as pseudorapidity is often used:

$$\eta = -\ln \tan \frac{\theta}{2} \quad (3.2)$$

This quantity is useful as, in the massless limit, it is equal to the rapidity of a particle, $y = \frac{1}{2} \ln \left(\frac{E+p_z}{E-p_z} \right)$. Differences in rapidity are Lorentz invariant to longitudinal boosts. The value $\eta = 0$ corresponds to a particle travelling perpendicular to the beamline, whilst $\eta \rightarrow \infty$ corresponds to a particle travelling parallel (or anti-parallel) to the beamline. A useful distance parameter derived from pseudorapidity and azimuthal angle is often used:

$$\Delta R^2 = \Delta \eta^2 + \Delta \phi^2 \quad (3.3)$$

In the massless limit, where $\eta \approx y$, this quantity is Lorentz invariant to boosts along the beam direction.

3.2.2 Inner Detector

The inner detector (ID) is the collection of detectors closest to the beampipe. The ID provides position measurements of charged particles passing through it. The ID is placed inside a 2 T magnetic field causing charged particles to curve as they travel through it. The components of the ID going radially outward from the beampipe are: the insertable b-layer (IBL), the pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). The resolution and coverage of these components are summarized in table 3.1.

The IBL and pixel detector form four cylindrical pixel layers in the barrel region, and three disc-shaped pixel layers in each of the end caps. The IBL is the innermost pixel layer in the barrel at a radius of 30 mm. The three pixel detector layers in the barrel cover the space $50.5 < r < 122.5$ mm. The IBL and pixel detector, when its end cap layers are included, cover the region $|\eta| < 2.5$. The pixel detector and IBL provide high resolution position measurements close to the primary interaction point of proton-proton collision events. This is particularly important not only for track reconstruction but to identify tracks originating from b hadron decays (see chapter 6). The pixel sizes and achieved resolution are summarized in table 3.1.

The SCT consists of four layers of pairs of silicon microstrips. Each strip is ~ 120 mm in length and the pairs have a separation of $80 \mu\text{m}$ (strip pitch). A 3D position measurement is obtained by angling the strips by 40 mrad from their common normal. The achieved resolution is $10 \mu\text{m}$ in the transverse direction and $580 \mu\text{m}$ in the longitudinal (see table 3.1).

The TRT is composed of 4mm radius cylindrical gas drift tubes. The space between these tubes is filled with polymer fibres (barrel) and foils (end caps). Charged particles radiate photons as they pass through these materials, more specifically the boundary between materials. The intensity of the transition radiation depends on the velocity of the particle. This can be combined with a measurement of the particles momentum to deduce its mass. This helps in particle identification, e.g. the distinguishing of electrons from pions.

The inner detector has been designed to give a high resolution measurement of particle tracks. The design target for transverse impact parameter resolution is $10 \mu\text{m}$. A 2 T magnetic field is applied in the ID. This causes charge particles to curve allowing for a measurement of momentum. The ATLAS ID has a target track momentum resolution of:

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T \oplus 1\% \quad (3.4)$$

The first term encapsulates the momentum dependence of resolution, as high mo-

parameter	resolution (designed)
d_0	$140 \mu\text{m}/p_T \oplus 10\mu\text{m}$
$z_0 \sin \theta$	$209 \mu\text{m}/p_T \oplus 91\mu\text{m}$
ϕ_0	$39 \text{ mrad}/p_T \oplus 70\mu\text{rad}$
$\cot \theta$	$5.0/p_T \oplus 0.7$
q/p	$0.05\% p_T \oplus 1\%$

Table 3.2: Designed resolution of the five track parameters used at ATLAS. The multiple scattering resolution term dominates at low p_T and intrinsic resolution term at high p_T . [44]

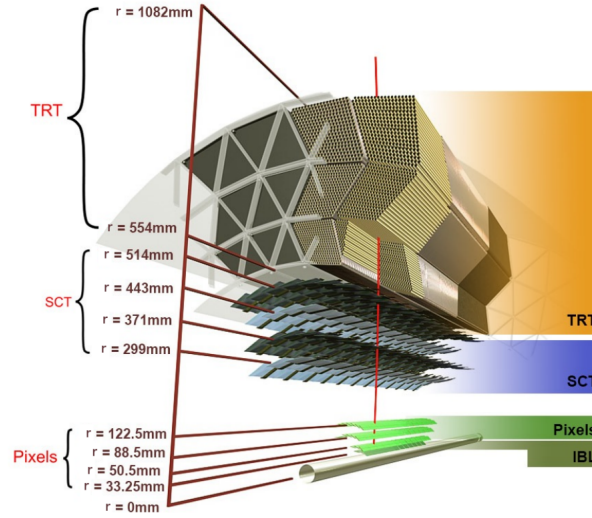


Figure 3.3: The barrel region of the inner detector. [45]

momentum tracks form almost straight lines [34]. The second term is due to multiple-scattering. The designed resolution of track parameters is summarized in table 3.2.

3.2.3 Calorimeter System

The calorimeter systems surround the inner detector. Their purpose is to measure the energy of particles produced in an event. There are two different calorimeter systems: The Electromagnetic Calorimeter (ECAL) which measure the energy of photons/electrons and the Hadronic (HCAL) calorimeter which measures the energy of hadrons². Both use layers of absorber material and active material. The absorber material reduces the energy of a particle as it passes through. The active material then detects this energy. The calorimeters are characterised by a critical

²Hadrons deposit energy in both the HCAL and ECAL

Calorimeter	σ_E/E	$\sqrt{E}$
EM (barrel and end-cap)	10%	0.7%
Tile and HEC	50%	3%
FCAL	100%	10%

Table 3.3: Energy resolutions obtained from test beam data. The HCAL is divided into Tile (in the barrel), HEC (end-cap) and FCAL (forward calorimeter). The first term is from stochastic noise whilst the second reflects local non-uniformities in the calorimeter response.

length defined by the distance through which the energy of a target particle (electron/photon or hadron) is reduced by $1/e$. For the ECAL this is the radiation length, X_0 . For the HCAL it is the characteristic depth, λ .

The energy resolution of these calorimeters is summarized in table 3.3. The ATLAS calorimeter system is non-compensating, the response to hadrons and electromagnetic particles is not the same. This is accounted for by calibration.

3.2.3.1 Electromagnetic Calorimeter (ECAL)

The ECAL is the innermost calorimeter system. It uses a lead absorber material³. The active material is liquid Argon (LAr). The ECAL is arranged with an accordion geometry to provide continuous coverage in ϕ . It is divided between the barrel component, covering $|\eta| < 1.475$, and two end cap components at either end of the detector covering $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$. The length of the ECAL is $X_0 > 22$ in the barrel and $X_0 > 24$ in the end caps.

3.2.3.2 Hadronic Calorimeter

The HCAL lies outside the ECAL. In the barrel region, it uses a steel absorber and scintillating tiles as active material (termed the tile calorimeter). The tile calorimeter is divided into the central and extended barrel sections covering $|\eta| < 1.0$ and $0.8 < |\eta| < 1.7$. It consists of three layers of thickness 1.5λ , 4.1λ and 1.8λ in the central region and 1.5λ , 2.6λ and 3.3λ in the extended barrel.

The HCAL uses LAr active material in the end-cap (HEC) and forward regions (FCAL). The HEC uses copper absorbers and is formed of two wheels per end-

³The EM coupling of a particle to an absorber is proportional to atomic number squared, hence lead has a short X_0 [5]

cap with depth of 12λ , covering the region $1.5 < |\eta| < 3.2$. The FCAL uses one layer of copper absorber and two of tungsten. It has a depth of 10λ and covers $3.1 < |\eta| < 4.9$.

3.2.4 Muon Spectrometer

The muon spectrometer (MS) is the outermost system of ATLAS. Muons are minimally ionising and have a large mass leading to reduced bremsstrahlung emission. This allows them to pass through most of the detector undisturbed, until they reach the MS. The MS contains layers of detectors that provide positional measurements. A toroidal magnetic field provided by magnets in the barrel and end caps permeates the MS. This causes the trajectory of muons to curve allowing for momentum measurement.

Four types of detector are used in the MS to measure the position of muons. Two of these are known as trigger muon chambers: Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs). These provide a position measurement with very short time delays (< 25 ns [46]). RPCs cover $|\eta| < 1.05$ and TGCs cover $1.05 < |\eta| < 2.4$. The other two muon detectors are monitored drift tubes, covering $|\eta| < 2.7$ and cathode strip chambers covering $2.0 < |\eta| < 2.7$.

The MS additionally provides measurements in the case of jet punch-through. This occurs when a non-negligible amount of energy from a high- p_T jet is deposited in the MS, having “punched through” the calorimeters.

3.2.5 Trigger

Proton collisions in the LHC occur at the bunch crossing rate of 40MHz. The trigger helps to reduce the rate of events to ~ 1 kHz, by identifying and recording only events of interest. The run 2 trigger is divided into two levels: the Level 1 hardware trigger (L1) and the software based High Level Trigger (HLT) [47].

The L1 trigger reduces the initial event readout rate from 40MHz to 100kHz. As such, it operates with very limited latency. This means no ID information is available. It only makes use of the calorimeter and MS detectors (the trigger muon chambers). Regions of interest (RoI) within the calorimeter and muon systems are

selected that may contain high p_T jets, leptons or photons. Global information is available to trigger on as well, such as missing transverse momentum. The events passing the L1 trigger then pass to the HLT.

The HLT reduces the event rate down to 1kHz. It is a software based trigger running on a computer farm. The online algorithms in the HLT reconstruct events similar to offline reconstruction algorithms⁴ but are designed for speed. This allows for an improvement in trigger efficiency.

The different triggers used in ATLAS make up the “trigger menu”. Analyses will have a trigger chain which details all the triggers used, an example of this can be found in chapter 8. Triggers can be prescaled if the output rates are too large. Only a fraction of the events passing the HLT are then kept. Thresholds on p_T offer an alternative way to reduce the trigger output to the desired rate. For instance, the HLT-j380 trigger requires at least one jet with a $p_T > 380$ GeV. Many analyses prefer to use p_T thresholds as prescaled triggers entail correction factors.

⁴see chapter 4

Chapter 4

Data Processing in ATLAS

4.1 Object Reconstruction

The measurements performed by the ATLAS detector are used to reconstruct physically meaningful objects. The tracking detectors provide the locations of hits, energy deposits as a charged particle passes through the detector. These hits are used to reconstruct particle tracks. The energy deposited in the calorimeters are formed into clusters with a specific location. These can be combined with tracks to deduce objects such as electrons or jets. The reconstructed objects used in this work are presented in this section.

4.1.1 Tracks

Charged particle trajectories are reconstructed from hits in the IBL, pixel layer and SCT (hits in the TRT are initially excluded) [45]. Single hits in the IBL or pixel layer, or two hits in the SCT, form 3D space-points. A track seed is created from three such points. More hits are then iteratively added to the track seed to form a track candidate. Track seeds are created from all combinations of three hits then grown assuming a helical track.

The resultant track candidates occasionally share hits. This issue is solved using a track score. The set of tracks with the highest combined score and no shared hits is selected. The track score is based on their p_T , number of associated hits, number of holes (missing expected hits) and overall χ^2 of the fit. Tracks with less than 7 hits or $p_T < 400$ MeV are removed.

The information from the TRT is added in a final step. Hits in the TRT consistent with track candidates are added and the tracks are refit.

4.1.1.1 Vertex reconstruction

A reconstructed vertex is the point at which multiple tracks meet. The primary vertex refers specifically to where the two protons in an event collide. Vertices may arise due to many other physical processes, such as heavy flavour hadrons decaying. These secondary vertices are covered in more detail in chapter 6. Due to the large number of pp collisions per bunch crossing, the vertexing procedure will identify multiple primary vertex candidates. The primary vertex with the highest sum of squared transverse momentum from its associated tracks is assumed to be the hard-scatter primary vertex.

The vertexing procedure starts by identifying a vertex seed position from the track collection. Tracks incompatible with this position are removed and the vertex position is refit. This procedure recurs iteratively until the vertex candidate passes certain quality requirements.

4.1.2 Jets

4.1.2.1 Reconstruction

The first step in jet formation is the construction of topological clusters (topocluster) in the calorimeters [48]. Energy deposits in calorimeter cells are given an energy significance S_{cell} :

$$S_{cell} = \frac{E_{cell}}{\sigma_{noise,cell}} \quad (4.1)$$

Where E_{cell} is the energy in the cell and $\sigma_{noise,cell}$ the background noise of the cell, both are measured calibrated to the electromagnetic (EM) energy scale. Topological cluster seeds are formed from cells with $S_{cell} > 4$. Neighbouring¹ seed cells are merged. Any neighbouring cells with $S_{cell} > 2$ are then merged to the cluster until only boundary cells with $0 < S_{cell} < 2$ are left. These final boundary cells are collected, completing the process.

¹Adjacent cells in the same sampling layer, or cells with partial overlap in the (η, ϕ) plane if in different sampling layers.

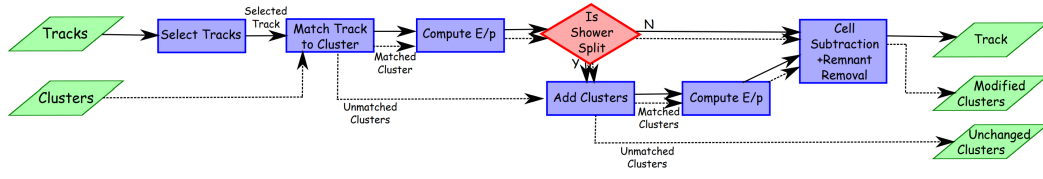


Figure 4.1: A flow chart of the PFlow algorithm [49].

The particle flow (PFlow) algorithm is a recent addition that combines ID track information with topoclusters [49]. This allows for better angular resolution, pile-up mitigation and increased resolution at low p_T , where the ID outperforms the calorimeters. The algorithm selects a number of tracks from the ID and matches them each to an individual topocluster. The expected energy in a topocluster is calculated from its position and the momentum of its matched track. The algorithm can then decide to add more topoclusters to the track/topocluster system by calculating the probability the particle deposited its energy in multiple topoclusters. The expected energy from the track is then subtracted cell-by-cell from the matched topoclusters. The remaining topocluster cells are removed if their energy is consistent with the expected shower fluctuations of the particle. A flow-chart of the algorithm can be seen in figure 4.1. The final output of PFlow is the tracks, the topological cells remaining after the expected energy removal and any unmodified topoclusters that were not matched to a track.

The *anti- k_t* algorithm [50] is used in ATLAS to reconstructs jets from input four-vector objects². These objects can be simply calorimeter topoclusters (e.g. in the `AntiKt4EMTopo` jet collections) or they can be combinations of the topoclusters and ID tracks as outputted by PFlow (the `AntiKt4PFlow` collection). The algorithm uses two distance measures:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (4.2)$$

$$d_{iB} = k_{ti}^{2p} \quad (4.3)$$

where k_{ti} is the p_T of object i and $p = -1$ for anti- k_t ($p = 1$ gives the k_t algorithm).

²This algorithm is infrared and collinear safe.

The distance R can be freely chosen, for jets in ATLAS the common choice is $R = 0.4$ (AntiKt4). The algorithm proceeds iteratively:

1. Compute d_{ij} and d_{iB} for all i, j .
2. If the smallest distance is d_{ij} , combine k_{ti} and k_{tj} .
3. Otherwise remove object i , having the smallest distance d_{iB} as a new jet.
4. Repeat until all objects are removed clustered.

4.1.2.2 Jet Calibration

The jet energy calculated by the ATLAS reconstruction differs from the true jet energy due to many effects: the calorimeter is non-compensating (its energy response to hadrons and electrons is different, see section 3.2.3), the impact of pile-up, detector inhomogeneities and other losses in the detector material. The following jet calibration procedure is applied to correct for these effects in PFlow jets [49]:

- The input tracks require $|z_0 \sin \theta| < 2$ mm, where z_0 is the longitudinal impact parameter with respect to the hard scatter primary vertex. This removes a large fraction of tracks from pile-up. The topocluster η and ϕ are recalculated with respect to the hard scatter primary vertex to conserve jet energy.
- An area-based pile-up correction is applied. An energy correction is subtracted from the jet. It is calculated by the product of the jet area A and the transverse energy density ρ . The density ρ is calculated using charged and neutral particle flow objects.
- A jet energy scale correction is applied to account for the differences in hadron energy reconstruction at different positions in the detector. The ATLAS calorimeter system is initially calibrated to the EM scale, such that the energy deposited by electrons and photons is the same at all positions. The reconstructed hadron energy will still change depending on where in the detector the shower develops. A correction is calculated using Monte Carlo (MC) simulated samples to account for this.

- Global sequential calibration accounts for differences due to the flavour of the originating parton and composition of hadrons created in fragmentation. This procedure uses three variables consecutively: the fraction of the jet energy from its constituent tracks, the fraction of jet energy in the third ECAL layer, and the fraction in the first Tile calorimeter layer.
- An in-situ calibration is then performed to account for the differences between the simulated samples used in the previous steps and data. This uses data samples with a jet balanced by a recoiling $Z \rightarrow \mu\mu$ decay, a photon, or multiple low p_T jets. Correction factors for data to MC are derived.

A similar calibration procedure is applied for `AntiKt4EMTopo` jets, more detail can be found in Refs. [51, 52, 53, 54].

Several important systematic uncertainties arise from the jet energy calibration, in particular uncertainties relating to jet energy scale (JES) and jet energy resolution (JER). The derivation of these are detailed in Refs. [55, 56]

4.1.2.3 Flavour tagging

An important aspect of jet studies in ATLAS is flavour tagging. This procedure attaches a label to jets denoting their origin, e.g. the flavour of quark producing the jet. A b-jet is a jet originating from the hadronic shower of a b quark. Flavour tagging is explored in more detail in chapter 6. The currently recommended b-tagging algorithm in ATLAS is DL1r. It is a neural network based algorithm with two supported working points with respect to b-tagging efficiency, the 85% working point and 77% working point. The working point is chosen to have a specific b-jet efficiency (percentage of true b-jets which will be tagged). The light jet rejection (one over the fraction of light jets incorrectly tagged as b-jets) decreases at higher working points (see figure 4.2). Jets passing one of these working points are labelled as b-jets.

4.1.2.4 Pile-up Mitigation

Low p_T jets from pile-up are rejected using the multivariate jet vertex tagger (JVT) discriminant. This uses information on the ID tracks associated to a jet to perform

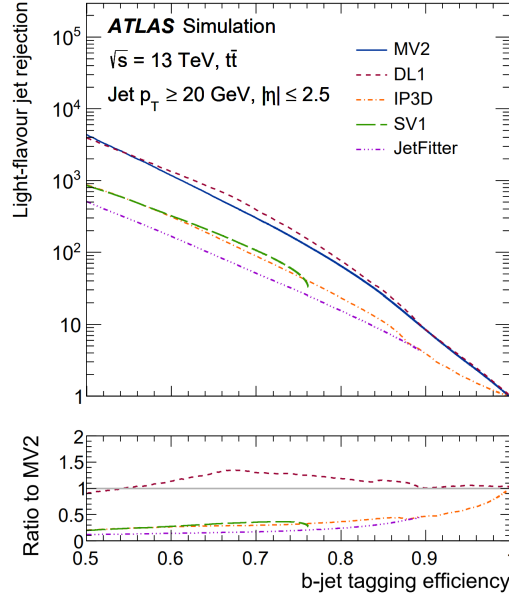


Figure 4.2: An example receiver operating characteristic (ROC) curve showing the light-jet rejection performance at given b-tagging efficiency working points for several ATLAS b-tagging algorithms.

pile-up jet rejection. A number of working points are implemented such as the *Medium* working point ($JVT > 0.2$). An area based pile-up correction is applied to the jets passing the JVT cut.

4.1.3 Large-R Jets

The ATLAS detector collected data for run-2 at a centre of mass energy of 13 TeV. At such high energy scales, the jets from a highly boosted particle decay can overlap. This makes the individual reconstruction of jets, using the algorithms of section 4.1.2, sub-optimal. Instead, a new object called a large-R jet (or fatjet) can be created.

4.1.3.1 Reconstruction

Large-R jets [57, 58] are reconstructed using slight modifications of the topological clustering and *anti- k_t* algorithms presented in section 4.1.2.1. Topological clusters are calibrated to the LCW (local cell weighting) scale instead of the EM scale. LCW accounts for the differences in response of the ECAL and HCAL, thus the topological clusters are corrected to the hadronic scale. The *anti- k_t* algorithm is used with a larger radius parameter, $R = 1.0$, to capture the decay products of the

two overlapping jets.

Large-R jets are much more sensitive to pile-up as they enclose a larger volume of the calorimeter. Pile-up can interfere with measurements of jet substructure variables, these are used in determining the origin of large-R jets. A procedure known as *grooming* (or *trimming*) is used to reduce these issues. In jet grooming, the constituents of a large-R jet are reclustered using the k_t algorithm with a radius parameter R_{sub} . Any sub-jets found by this procedure with a p_T less than a fraction f_{cut} of the large-R jets have their constituent objects removed. The current recommendation for large-R jets is to use $R_{sub} = 0.2$ and $f_{cut} = 0.05$ giving the `AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets` jet collection. This is found to give excellent robustness to pile-up conditions, up to 200 collisions per bunch crossing.

4.1.3.2 Calibration

Large-R jets are calibrated in a similar fashion to small-R jets (section 4.1.2.2). The calibration accounts for pile-up effects, corrects for the jet energy scale, and does in-situ validation. One extra step is added for large-R jets, calibration for the jet mass scale (JMS). This is important as large-R jets typically require a well defined mass to aid in their interpretation [58]. The calibration proceeds as follows:

- Pile-up mitigation is performed using jet grooming.
- JES calibration is performed, in the same way as small-R jets. MC simulated jets are used, with the reconstructed jets compared to the truth jets. This accounts for energy differences due to detector effects. The calibration is done binned by energy and η .
- JMS calibration is done in a similar way, by matching reconstructed jets to truth jets in simulation. The mass response is calibrated (parameterized by mass, η and p_T).
- In-situ calibration accounts for the differences between data and MC used in previous steps. For JES, data samples with a large-R jet balanced with a γ or

multiple low- p_T jets are used. For JMS, the forward folding method is used (this process uses the detector response to compare expected and observed measurements, see ref. [59]). This uses samples with hadronically decaying W bosons and top quarks.

Systematic uncertainties again arise for these calibrations analogously to the small-R jet case.

4.1.3.3 Boosted Jet Tagging

The major advantage of large-R jets is their interpretation as the decay product of boosted particles. The reconstructed jet mass and its rich substructure can be used to identify its origin. Several multivariate techniques have been used to tag large-R jets originating from W/Z boson decays, Higgs bosons, top quarks or other QCD signals. Such an algorithm used in this work is the W/Z tagger [60]. The parent particle of a large-R jet will decay to form a specific type of substructure. For instance, a high- p_T W boson decaying hadronically forms two jets. The large-R jet formed from their overlap will have “W-like” substructure.

The W/Z tagger uses three variables to tag large-R jets. The first is the reconstructed jet mass, this is the combined mass formed by adding together mass information from the tracker and the topocluster mass [61]. The energy correlation ratio, D_2 is defined as:

$$D_2 = E_{CF3} \left(\frac{E_{CF1}}{E_{CF2}} \right)^3 \quad (4.4)$$

where the energy correlation functions (E_{CF}):

$$E_{CF1} = \sum_i p_{T,i} \quad (4.5)$$

$$E_{CF2} = \sum_{ij} p_{T,i} p_{T,j} \Delta R_{ij} \quad (4.6)$$

$$E_{CF3} = \sum_{ijk} p_{T,i} p_{T,j} p_{T,k} \Delta R_{ij} \Delta R_{jk} \Delta R_{ki} \quad (4.7)$$

and i, j refer to sub-constituents of the fatjet. The interpretation of this variable is the ratio of “3-prongness” over “2-prongness”. Jets with a 2-prong substructure

(i.e. two main subjet components) will have a low value of D_2 , whilst jets with 1 or 3-prongs will have a high value. A vector boson (V) would be expected to be 2-pronged due to the decay $V \rightarrow q\bar{q}$. The final substructure variable is the number of ghost-associated³ tracks to the jet, n_{trk} . W/Z bosons are color-singlets, as opposed to quarks and gluons, hence the particle multiplicity of the W/Z bosons is nearly constant with p_T , whilst it changes quickly for quark/gluon jets.

Two working points are available for the W/Z tagger. These are the 50% and 80% working points, referring to the W/Z signal efficiency. The tagger signal and background efficiencies are calibrated with scale factors derived from data to MC comparisons [60].

4.1.4 Electrons

4.1.4.1 Reconstruction

Electron candidates are reconstructed by associating a track to a cluster of energy in the ECAL. The seed energy clusters in the ECAL are identified using a sliding window⁴. The total transverse energy in a seed cluster must exceed 2.5 GeV. Tracks are then associated with this cluster. In the main ATLAS track reconstruction, a pion hypothesis is used which assumes up to 30% momentum loss at every material layer intersection. An additional electron hypothesis, allowing for larger momentum loss, can be used to reconstruct tracks which failed to reach 7 hits in the pixel layers or SCT. This improves performance without interfering with the main track reconstruction. In the event of multiple tracks matching the cluster, a “primary” track is chosen by ranking the candidates based on the ΔR separation of the track and cluster, the number of pixel hits and whether a hit is present in the first silicon layer. Electron candidates with no associated tracks are considered to be photons.

Electrons are only reconstructed in the region $|\eta| < 2.47$, to avoid edge effects in extrapolating to the ECAL.

³Ghost-association is a robust way of matching tracks to a jet using its catchment area [62]

⁴Window size: 3×5 in units of 0.025×0.025 in $\eta \times \phi$ (corresponds to granularity of middle layer of ECAL)

4.1.4.2 Identification

The electron ID algorithm is used to distinguish between prompt⁵ electrons and background-like electrons. Background-like electrons are produced by hadronic jets or photon conversions.

The electron ID algorithm is a likelihood method based on variables relating to the electron cluster and track measurements. For a full description see [63]. Three working points are used, in order of increasing background rejection: *LooseLH*, *MediumLH* and *TightLH*. All three working points are defined with respect to a Likelihood-based discriminant. These are defined in table 1 of [63].

Additional requirements on η and track parameters are usually applied to reduce backgrounds. The region $1.37 < |\eta| < 1.52$ is excluded as it is the transition region between the barrel and end-cap ECAL. The large amount of material here causes an increase in background-like electrons from photon conversions. The transverse impact parameter significance must satisfy $|d_0|/\sigma_{d_0} < 5$, whilst the longitudinal impact parameter and polar angle of the track must satisfy $|z_0 \sin \theta| < 0.5$ mm. This ensures compatibility to the primary vertex, further reducing backgrounds.

4.1.4.3 Isolation

Electron isolation helps to further differentiate between prompt electrons, originating from the hard-scatter event, and non-prompt electrons, e.g. from heavy hadron decays. The prompt-lepton veto tagger (PLV) is employed to perform isolation. This BDT⁶ based algorithm makes use of the properties of track jets⁷ matched to an electron or muon [64, 65].

4.1.4.4 Charge Reconstruction

The electron charge is reconstructed from the curvature of the track in the ATLAS magnetic field. A BDT method known as ECIDS (Electron Charge ID selector) attempts to reject electrons with incorrectly reconstructed charge [66]. ECIDS is

⁵Refers to products originating from the primary vertex. Also known as signal-like.

⁶Boosted decision tree, see section 5.2.1.

⁷A jet with tracks in the ID associated to it.

also known as the charge mis-assignment veto BDT in [64]. It uses various input variables such as p_T , η , charge and d_0 . The variables are motivated by the causes of charge mis-identification, such as bremsstrahlung due to material interactions or the tracks being almost straight at high momentum. Only one working point is currently recommended by the egamma combined performance group, known as the loose working point. Scale factors are implemented for this working point to correct for differences between MC and data [66].

4.1.4.5 Rejection of Photon Conversions

Electron and photon signals can often be hard to distinguish. Photons can undergo conversion in the detector material: pair-production of an electron and positron $\gamma \rightarrow e^- e^+$. The produced electron could then be reconstructed, giving us a fake signal. An electron can also be faked if a track is found to point to a deposit in the ECAL from a photon.

Two variables exist to help reject electrons from converted photons and photons mis-reconstructed as electrons. The electron author (ambiguityType) variable tells if an electron has also been reconstructed as a photon [67]. The addAmbiguity variable has been newly implemented to offer further rejection of internal and material conversions [68].

4.1.4.6 Electron Definitions

Combinations of the different requirements set out previously are used to define several different types of electron in this thesis, see chapter 8. The definition of a signal electron used here is summarized in table 4.1.

4.1.5 Muons

4.1.5.1 Reconstruction

Muon tracks are reconstructed by combining information in the ID and MS [69]. Two Muon track definitions are used here:

- Combined (CB) muon tracks are formed from tracks in the MS extrapolated to match with a track in the ID. The Muon track in the MS is found by first

Requirement	Value
p_T	$> 20\text{GeV}$
η	$ \eta < 1.37 \& 1.52 < \eta < 2.47$
Identification	TightLH
$\frac{d_0}{\sigma_{d_0}}$	< 5
$ z_0 \sin \theta $	$< 0.5 \text{ mm}$
Object quality	author == 1, addAmbiguity ≤ 0
Isolation/PLV	PLVTight
ECIDS WP	Loose

Table 4.1: Signal electron definition used in this work.

identifying track segments using a Hough transform [70]. A track candidate can be formed by matching at least two segments⁸. Track segments shared between candidates are assigned to the most compatible track using an overlap removal procedure. The hits associated to this candidate are then reconstructed in a χ^2 fit. The MS track is then matched to an ID track, and the hits from both are refit together to form a combined track.

- Segment-tagged (ST) muons are formed if an ID track extrapolated to the MS matches at least one track segment. This allows for a higher muon reconstruction efficiency and is useful in regions where the MS lacks good coverage ($|\eta| < 0.1$).

4.1.5.2 Identification

To separate prompt muons from background ones (mainly from hadron decays), a set of requirements on different variables is defined. CB muons make use of the q/p significance⁹, the normalized χ^2 of the track fit and ρ ¹⁰. There are further requirements on the number of hits and holes in the ID and MS.

Several different working points are defined, each having its specific set of requirements. These are termed *loose*, *medium* and *tight* in decreasing order of reconstruction efficiency (increasing order of purity). An additional set of require-

⁸In the barrel-endcap transition region, one high-quality segment is enough.

⁹The absolute difference of the ratio of charge and momentum of the track measured in the ID and MS, divided by uncertainty.

¹⁰Absolute difference of p_T of the MS track and ID track, divided by the p_T of the combined track.

Requirement	Value
p_T	$> 20 \text{ GeV}$
η	$ \eta < 2.5$
Identification	TightLH
$\frac{d_0}{\sigma_{d_0}}$	< 3
$ z_0 \sin \theta $	$< 0.5 \text{ mm}$
Object quality	medium
Isolation/PLV	PLVTight

Table 4.2: Muon signal definition used in this work

ments, *high- p_T* , gives the best reconstruction efficiency for muons with high p_T ($> 100 \text{ GeV}$).

Additional requirements on kinematics are normally used to distinguish prompt muons from background ones, such as heavy flavour decays, and muons from pile-up. These include a requirement on $p_T > 20 \text{ GeV}$, $|d_0|/\sigma_{d_0} < 3$ and $|z_0 \sin \theta| < 0.5 \text{ mm}$. A geometrical acceptance term is usually required as well, $|\eta| < 2.5$.

4.1.5.3 Isolation

The PLV is used to isolate muons similarly to electrons. Two different working points exist, with the *PLVTight* offering the best rejection of muons from heavy flavour jets and other backgrounds [65].

4.1.5.4 Muon Definitions

Several different muon definitions will be used in this work, see chapter 8. The signal muon definition is summarized in table 4.2.

4.1.6 Missing Transverse Momentum

The total transverse momentum in an ATLAS event should be zero by conservation of momentum. When this sum is performed using the measurements made there will usually be some missing momentum. This can be due to detector effects or mis-measurement but is also a signal of particles invisible to the detector such as neutrinos. The missing transverse momentum, E_T^{miss} , is calculated as:

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss},e} + E_{x(y)}^{\text{miss},\gamma} + E_{x(y)}^{\text{miss},\tau} + E_{x(y)}^{\text{miss},jets} + E_{x(y)}^{\text{miss},\mu} + E_{x(y)}^{\text{miss},soft} \quad (4.8)$$

where each term, $E_{x(y)}^{miss,i}$, is the x or y component of the negative vector sum of the momentum of all objects of type i . The soft term takes into account all the p_T objects in the detector that were not reconstructed into baseline objects (electrons, muons, photons, taus, jets). This is done in the track-based soft term using the ID tracks not associated to baseline objects. The E_T^{miss} is the magnitude of the 2D vector given by the $E_{x(y)}^{miss}$ components [71].

4.1.7 Overlap Removal

The reconstruction algorithms are applied separately and may on occasion reconstruct a detector signal as two separate objects. An overlap removal procedure is used to mitigate against this. This ensures no accidental addition of extra particles or energy. The following procedure is known as the “heavy-flavour” recommendation:

- **Electron/muon overlap:** If an electron and muon lie within $\Delta R < 0.1$ of each other, the electron is removed. A muon can radiate a photon, which can then convert into an electron/positron pair.
- **Electron/Jet overlap:** The ECAL deposits of an electron could be reconstructed as a jet. However, an electron could be produced by heavy flavour decay in a jet. It is therefore recommended to remove the electron if the overlapping jet ($\Delta R < 0.4$) is b-tagged with DL1r (85%). If the jet is not b-tagged, the electron is removed if the electron-jet separation is $0.2 < \Delta R < 0.4$, and the jet is removed if the separation is $\Delta R < 0.2$.
- **Muon/Jet overlap:** The procedure is the same as for electrons but with one addition. In the event of a non-b-tagged jet with separation $\Delta R < 0.2$ from a muon, the muon will be discarded if the jet has at least three associated tracks, otherwise the jet is discarded. This behaviour is not typical of a prompt muon, it is likely light-flavour jet punch-through to the MS, or a muon from a hadronic decay in the jet.

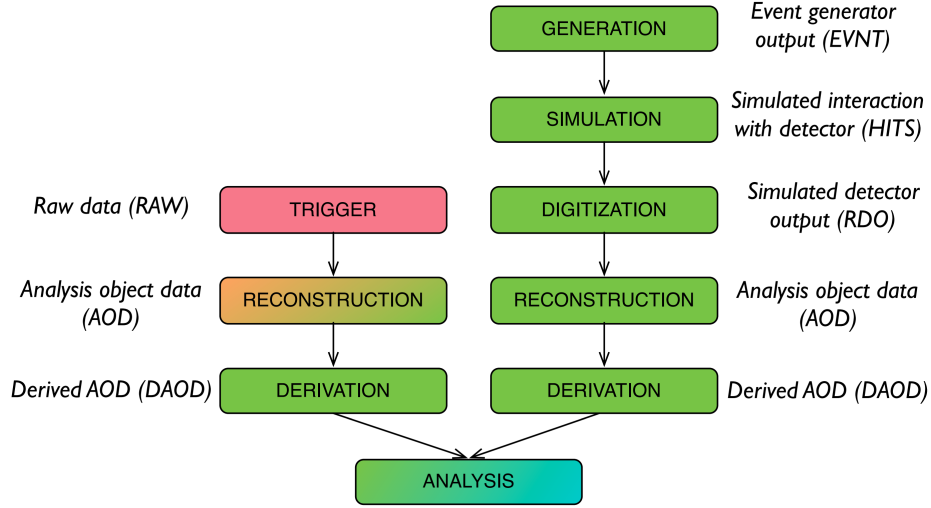


Figure 4.3: The ATLAS data processing chain for real (left) and simulated (right) data. [72]

4.2 Data Formats in ATLAS

The ATLAS data processing chain is responsible for producing data formats for physics analysis. Real and simulated data are presented to analysis groups in a format known as an analysis object data (AOD). The processing chain is summarized in figure 4.3.

4.2.1 Data Samples

ATLAS data samples are gathered over the course of runs. Run 2 data was collected between 2015-2018 at $\sqrt{s} = 13$ TeV. Three collections exist, the 2015-16, 2017 and 2018 datasets. The ATLAS reconstruction framework takes the RAW data outputted by the trigger and creates the physics objects described in section 4.1. The final output is an AOD, typically having order PB size.

4.2.2 MC Simulation in ATLAS

The SM gives us a theoretical framework to predict what we expect to see in ATLAS. Monte Carlo (MC) methods are used to simulate ATLAS events, starting from the hard scatter process [73]. The non-perturbative QCD processes such as hadronization and parton shower are simulated with programs such as PYTHIA [10] and Herwig [74]. It is common to interface amplitude generating programs to parton shower ones.

The end result of event generation is the final state particles expected in an ATLAS event. The next step is to simulate a detector response. GEANT 4 [75] is used to simulate the ATLAS detector response. The process of digitization is then responsible for turning this into the simulated detector output, a raw data object (RDO). Once an RDO is obtained, the exact same reconstruction algorithms used for real data are applied to create an AOD. MC simulated AODs are created for specific background or signal processes.

4.2.3 Derivations

A recent addition to the ATLAS data chain in run-2 is the derivations framework [76]. Derivations take the PB scale AODs and reduce them to TB sized DAODs. The reduction is achieved by keeping only the information necessary to a specific analysis, or a group of similar analyses. The removal of information happens in three separate ways:

- *Skimming*: Removal of events based on criteria relating to their features.
- *Thinning*: Removal of objects within an event based on criteria relating to object features.
- *Slimming*: Removal of variables of a certain object across all objects of that type and all events. Slimming is applied independent to the features of a particular object or event.

The derivations framework also has the ability to add information through *augmentation*. This can be adding entirely new objects created by the framework, or decorating existing objects (adding a new member variable to the object). An example decoration might be if a jet has been b-tagged by a certain algorithm. The derivations framework contains all the tools necessary for augmentation, such as the b-tagging algorithms.

Chapter 5

Machine Learning

The ATLAS experiment handles petabytes of data. The big data challenge entailed by this motivates the use of artificial intelligence (AI), and in particular machine learning, to aid ATLAS researchers. The large amount of data and simulated data ensures that AI techniques can be trained to achieve excellent performance. These techniques allow for complex non-linear modelling of a variety of physics processes. ML techniques are consequently widely used across the ATLAS physics program. For a review of ML techniques at the LHC, see reference [77].

ML algorithms are models trained on data to optimise performance on a defined task [78, 79, 80]. Performance is measured by a loss function (or cost function). The set task can be generic, but in this work is usually either classification or regression. Examples of classification tasks in ATLAS are flavour tagging or the categorization of events as signal-like or background-like during analysis work. Regression ML algorithms are used for instance in tracking [77]. ML is also used for tasks such as clustering or fast simulation of data, for instance using a variational auto-encoder (VAE).

The training procedures used in this work can be broadly categorized as supervised learning, these are detailed in section 5.1. Several different models are used in this work, these are covered in section 5.2.

5.1 Supervised learning

Supervised learning requires labelled training data, each datapoint is a pair $(\mathbf{X}, \mathbf{y})$ where $\mathbf{X}$ is the input features and $\mathbf{y}$ the output or label. These are generically vectors. The learning process creates a map of $\mathbf{X} \rightarrow \mathbf{y}$. The two main tasks set for supervised learning are regression and classification.

5.1.1 Regression

Regression tasks involve the prediction of a continuous output $\mathbf{y}$, termed the dependent variable(s), from input $\mathbf{X}$, termed the independent variable(s). The most commonly known is linear regression, where the relationship between $\mathbf{X}_i$ and a scalar output y_i is modelled as a linear combination:

$$\tilde{y}_i = \boldsymbol{\theta} \cdot \mathbf{X}_i + b \quad (5.1)$$

where $\tilde{y}_i$ is the predicted output, $(\mathbf{X}_i, y_i)$ are the observed values for the i – th data point, and the weights $\boldsymbol{\theta}$ and intercept b are to be determined such that some loss function $Q(\tilde{y}, y)$ is minimized. Commonly, the sum of differences squared is used: $Q(\tilde{y}, y) = \sum_i (\tilde{y}_i - y_i)^2$. Linear regression can be performed analytically but is limited in application. In the case of high-dimensional data or non-linear processes, ML algorithms such as neural networks can be used to perform regression.

5.1.2 Classification

A classification task involves the mapping of input $\mathbf{X}$ to a discrete categorical value (or vector of values) $\mathbf{y}$. The inputs in this context are often referred to as features, these need not be categorical themselves. A simple classifier model can be formed from an extension of linear regression called logistic regression [79]. Our label y is now a binary variable $(0, 1)$, corresponding to one of two categories (e.g. signal or background). The variable $\tilde{y}$ from Eq. (5.1) is now passed through a logistic function which maps to the interval $[0, 1]$:

$$y' = \sigma(\tilde{y}) = \frac{1}{1 + e^{-\tilde{y}}} \quad (5.2)$$

$\tilde{y}$ is now referred to as the latent variable. The category can be predicted from $\sigma(\tilde{y})$, for example the predicted label $y' = 0$ if $\sigma(\tilde{y}) < 0.5$. In the framework of ML, $\sigma(\tilde{y})$ is recast as a conditional probability:

$$\Pr(y = 1|\mathbf{X}, \boldsymbol{\theta}) = \sigma(\tilde{y}) \equiv \sigma(\mathbf{X}, \boldsymbol{\theta}) \quad (5.3)$$

We change the notation to make explicit that $\sigma(\mathbf{X}, \boldsymbol{\theta})$ is a function of $\mathbf{X}$ parametrized by some values $\boldsymbol{\theta}$. The loss function to be minimised will be the negative log likelihood. To construct the likelihood, we make the assumption that y is an independent Bernoulli distributed variable with parameter $p = \Pr(y = 1|\mathbf{X}, \boldsymbol{\theta})$ as defined above:

$$\Pr(y|\mathbf{X}, \boldsymbol{\theta}) = p^y(1 - p)^{1-y} \quad (5.4)$$

giving the likelihood

$$L(\boldsymbol{\theta}|\mathbf{X}, y) = \prod_i (\sigma(\mathbf{X}_i, \boldsymbol{\theta}))^{y_i} (1 - (\sigma(\mathbf{X}_i, \boldsymbol{\theta})))^{1-y_i} \quad (5.5)$$

where we have multiplied across all datapoints, i , in the dataset.

5.1.3 Loss functions and Gradient Descent

The learning procedure is formulated as the minimization of a loss function. Several algorithms exist for finding the minima of multivariate functions, the most commonly used is gradient descent [81]. This involves iteratively updating the parameters $\boldsymbol{\theta}$ using the negative gradient of the loss function:

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} - a \nabla Q(\boldsymbol{\theta}^{(n)}) \quad (5.6)$$

where Q is the loss function, $\boldsymbol{\theta}^{(n)}$ the parameters at the n – th step, ∇ is the gradient with respect to the parameters $\boldsymbol{\theta}$ and a is specified value known as the learning rate. The learning rate must be chosen to optimise the minimization procedure¹. The determination of ∇Q is often done by summing over all the observed data points

¹Too small and we risk getting stuck in a local minimum, too large and we risk missing the global minimum.

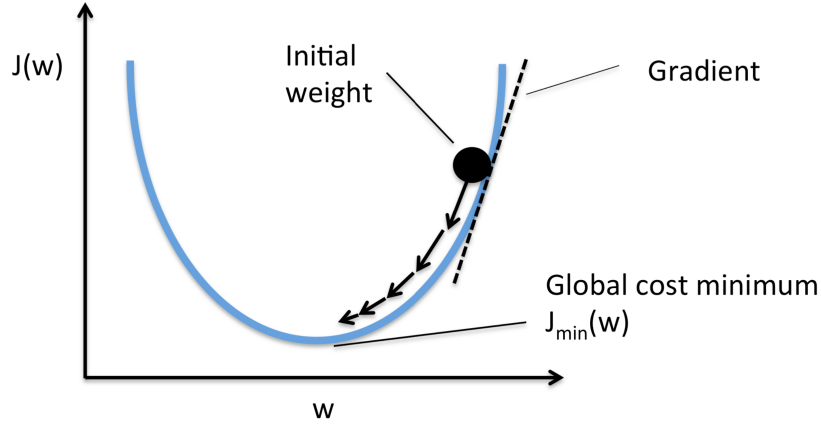


Figure 5.1: An illustration of gradient descent for a one dimensional loss/cost function $J(w)$ [82].

$\sum_i \nabla Q_i$ (batch gradient descent). For large datasets this may be computationally expensive, so methods such as stochastic gradient descent [81], which calculates the gradient with respect to a single randomly chosen datapoint, are used.

The choice of loss function is generally problem specific. Two loss functions have been introduced in the context of classification and regression. The loss function in linear regression problems can take many forms, but the most common is termed ordinary least squares:

$$Q(\boldsymbol{\theta}) = \sum_i (\tilde{y}_i(\boldsymbol{\theta}) - y_i)^2 \quad (5.7)$$

This is often chosen as it corresponds to the maximum likelihood estimator in the case where the measurements of y_i differ from the linear combination of $\mathbf{X}_i$ by the addition of Gaussian noise. The loss function in logistic regression was chosen as the negative log likelihood:

$$Q(\boldsymbol{\theta}) = -\log L(\boldsymbol{\theta}|\mathbf{X}, y) = -\sum_i^N y_i \log \sigma_i - \sum_i^N (1 - y_i) \log(1 - \sigma_i) \quad (5.8)$$

where N is the total number of datapoints in the dataset, and $\sigma_i \equiv \sigma(\mathbf{X}_i, \boldsymbol{\theta})$ is defined in equation (5.2)². This corresponds to maximum likelihood estimation for binary classification.

²This loss function corresponds to the binary cross-entropy between two distributions p and q .

5.1.4 Training

The training procedure is vital to achieving a useful ML model. A given base model is selected for a defined task. The loss function is then chosen with respect to that base model and learning proceeds, usually by minimising the loss with gradient descent. Gradient descent will iterate through the data (either stochastically or in batch method), every iteration through the training dataset is termed an epoch. An important aspect of all ML algorithms is ensuring we do not bias our results to the dataset we are using, this is often referred to as overfitting. To avoid overfitting, we generally split our dataset into two parts termed the training dataset and validation dataset. The model fit on the training dataset is evaluated on the validation dataset and training generally stops when we reach our best performance (or minimal loss) on the validation dataset. Finally, results are generally quoted on a third, held-out, test set. This has not been seen during the training to avoid any potential bias.

A method called k-fold cross validation [80] is often used at ATLAS. The dataset is split into k sets and a model is trained on all but one of these sets, with the hold-out set used for validation. This is repeated for all k sets. The model performance is then evaluated by the aggregation across each k-fold. 2-fold cross-validation is used in training most ML algorithms in ATLAS.

5.2 Models

Two models have already been introduced in the context of regression and classification tasks. These are the linear and logistic regression models in sections 5.1.1 and 5.1.2. This section summarizes the main models used in this work.

5.2.1 Boosted Decision Trees

5.2.1.1 Decision Trees

A decision tree is a predictive model that operates by making a series of cuts on variables to predict a label. In classification tasks, the label y is discrete and the variables are $\mathbf{X} = (x_1, x_2, \dots, x_n)$. At each node (or leaf) in the tree, a cut is applied on some variable x_i . A simple example is shown in figure 5.2. The final node in each branch predicts the label y .

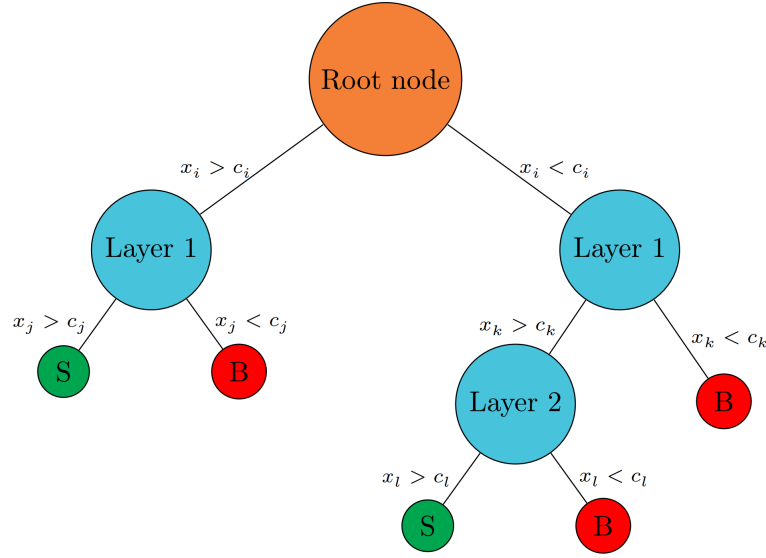


Figure 5.2: An example of a decision tree. At each decision node a cut is applied on some feature x_i of the data, splitting the dataset. The final node (output leaf) labels the data as signal (S) or background (B) [83].

Decision trees are generally constructed in a top-down manner. The optimal cut at each node is determined using some quantitative measure. An example is the Gini impurity, which gives a measure of the separation between categories in a node [84]. It is given by the probability that a datapoint that reaches a given node would be labelled incorrectly, assuming it were randomly labelled according to the distribution of labels in the set of points reaching that node:

$$G = \sum_i^M p_i(1 - p_i) \quad (5.9)$$

where M is the total number of classes and p_i is the probability of the i -th class. For example, if a node contains two classes A and B , split in a ratio 40:60 then $p_A = 0.4$ and $p_B = 0.6$. Tree branches continue to grow until the Gini impurity stops increasing i.e.:

$$G_{parent} - \sum G_{daughters} > 0 \quad (5.10)$$

Trees will also stop growing if the number of datapoints in a node goes below a certain threshold, or if the number nodes in a branch exceeds a certain depth.

5.2.1.2 Forests and Boosting

Decision trees alone tend to perform quite poorly at classification tasks, performing not much better than random guessing. They suffer greatly from problems of overfitting to data. Several methods exist to solve this problem, generally by using collections of decision trees called forests. In ATLAS, the most common method is boosting [84] using AdaBoost [85]. This constructs a forest called a boosted decision tree (BDT) in the following procedure: a decision tree is trained, the events in the training sample misclassified by the decision tree are reweighted to increase their importance, and a decision tree is trained on the new training sample. This procedure continues iteratively. A weighting term w_i is added to each term in the sum of equation (5.9). This is updated each iteration by:

$$w_i^{(n+1)} = w_i^{(n)} e^{\alpha_n} \quad (5.11)$$

where the score for the n th tree α_n is:

$$\alpha_n = \beta \log \left(\frac{1 - f_n}{f_n} \right) \quad (5.12)$$

$$f_n = \frac{\sum_{y_i \neq T_n(\mathbf{X}_i)} w_i}{\sum_i w_i} \quad (5.13)$$

where the sum on the numerator of eq. 5.13 is taken with respect to only the misclassified points, and $T_n(\mathbf{X}_i)$ refers to the prediction of the n -th tree on the i -th datapoint. The value β is the learning rate within the context of boosting. Once the training is finished, the predicted label of a point is taken by summing over the weighted predictions of each tree in the forest:

$$T(\mathbf{X}_i) = \sum_n^{N_{trees}} \alpha_n T_n(\mathbf{X}_i) \quad (5.14)$$

5.2.1.3 Examples

BDTs are used fairly ubiquitous in ATLAS. Several examples were mentioned in section 4.1. These include the prompt-lepton veto tagger in section 4.1.4.3 and

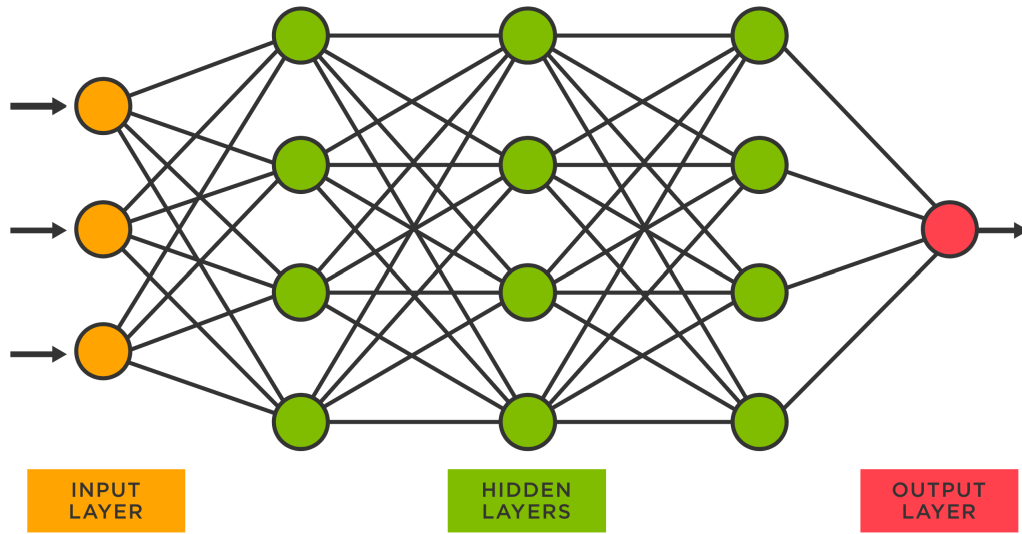


Figure 5.3: A schematic of a fully-connect feed-forward neural network with three hidden layers. Information flows from left to right. The black lines denote connections between nodes [86].

ECIDS in section 4.1.4.4. Another important BDT is the flavour tagging algorithm MV2, which attempts to classify jets by their flavour (either b, c or light). This is covered in section 6.3.6.2.

5.2.2 Neural Networks

A neural network is a model consisting of nodes connected by edges. A simple model is shown in figure 5.3. The nodes are arranged into layers. In a fully-connected model all nodes have connections to every node in the subsequent layer. The first layer of nodes is termed the input layer, nodes in this layer take the input data features and pass them on. Nodes in subsequent layers receive data from their connections to the previous layer then output a value through connections to the next layer. This value is determined by some function of the weighted sum of inputs. This function is termed the activation function and is nonlinear. Activation functions are commonly sigmoid functions such as the logistic function in eq. (5.2) [79].

Neural nets are learnt by adjusting the connection weights between nodes. In a supervised learning approach, the weights are modified to minimize the error in the

network's prediction as characterized by some loss function $Q(y', y)$, where y' is the predicted label and y the actual label. Learning proceeds through gradient descent with respect to the network weights:

$$w_{ij}^{(n+1)} = w_{ij}^{(n)} - a \frac{\partial Q}{\partial w_{ij}^{(n)}} \quad (5.15)$$

where $w_{ij}^{(n)}$ is the weight connecting the i and j – th nodes (at the n – th step). The partial derivative can be determined using backpropagation [79], which requires the activation function to be differentiable.

5.2.2.1 Deep Neural Nets

A deep neural network simply refers to a network with more than one hidden layers. The diagram in figure 5.3 is thus a deep neural net. The addition of hidden layers allows for far more complex problems to be learnt. These models can be applied to highly dimensional datasets with little feature engineering needed. They outperform more simple traditional methods. Deep neural nets have as such become extremely commonplace in many ML problems. An example in ATLAS is the b-tagging algorithm DL1r (see section 6.3.6.3). This trend began around 10 years ago with the emergence of a variety of procedures that enabled the rapid training of such models.

5.2.3 Recurrent Neural Networks

A recurrent neural network (RNN) is an extension of feed-forward networks for the purpose of processing ordered sequences of data. An RNN can make a prediction preserving the time-ordered nature of the data. The data input, $\mathbf{X}$, is an ordered sequence of scalars or vectors $(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_t)$, with the subscript denoting the timestep. The total length of the sequence, t , is free to vary. An example sequence, $\mathbf{X}$ might be a sentence with each $\mathbf{x}_i$ corresponding to a word.

An RNN is formed from a cell that recurs at every timestep in the sequence. A diagram can be seen in figure 5.4. At each timestep i the cell receives the input value $\mathbf{x}_i$ of the sequence and the output of the cell from the previous timestep (this

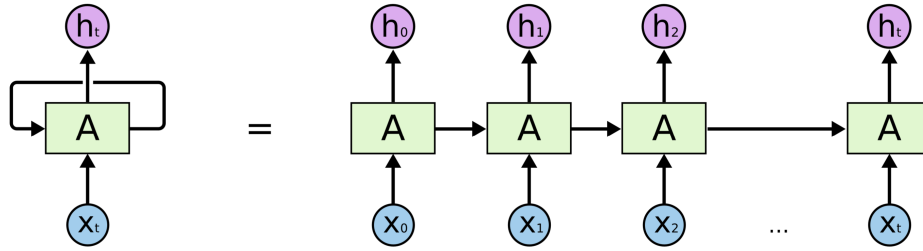


Figure 5.4: A schematic of an RNN and its appearance when “unrolled”. Data at each timestep of the sequence is x_t , the RNN/LSTM cell is the box A. At each timestep the cell receives the hidden state from the previous timestep and the next data of the sequence. The output from the final cell is the RNN output [87].

is termed the hidden state). The cell then produces an output for the next timestep. The output of the final cell is the final output of the RNN, y . In the simple example of a sentence, this output y could be the sentiment (positive or negative).

RNNs are trained using the same gradient descent techniques as feed-forward networks. The RNN cell in the simplest case could be a one-layer feed-forward neural network. Most RNNs in practice use a modification called a Long Short-Term Memory cell (LSTM). These are necessary to avoid the problem of vanishing gradients in backpropagation. In a simple RNN, vanishing gradients would cause the earliest entries in the input data sequence to have no impact on the final output³. LSTMs ensure information from the early part of the sequence is not forgotten. For a good description of how LSTM cells operate and the vanishing gradients problem see reference [88].

5.2.3.1 Examples

A simple example of sentence sentiment analysis was briefly presented above. This is a classification task on sequenced data, each sentence is a sequence of words. In general, RNNs can be used for classification or regression tasks.

An important application of RNNs is in b -tagging at ATLAS. The RNNIP [89] tagger classifies jets by the flavour of their parent quark (b , c , or l). It is trained on collections of jets. Each jet can be presented as an ordered sequence

³Consider a simple RNN processing a sequence of length t . This corresponds to a deep neural network with t layers. The nodes in the earliest layers of such a network will have very little impact on the final output due to vanishing gradients.

$\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_t)$ (note the length of this will vary from jet to jet). In RNNIP, the inputs $\mathbf{x}_i$ correspond to particle tracks ordered by the 2D distribution of transverse impact parameter significance of the tracks. Each track is a vector of five characteristic variables, $\mathbf{x}_i = (S_{d_0}, S_{z_0}, Category, p_T^{frac}, \Delta R(track, jet))$, for the definition of these see table 6.1 in chapter 6. The output of the final cell in the RNN is then fed to a simple feed-forward network to produce three output variables characterizing the probability of each jet flavour, $\mathbf{y} = (p_b, p_c, p_l)$. In training, performance is evaluated by comparing the output predictions, $\tilde{\mathbf{y}}$ to the true jet flavour $\mathbf{y}$ using a loss function called categorical cross-entropy:

$$Q = - \sum_i^N y_i \log \tilde{y}_i \quad (5.16)$$

where the values $\tilde{y}_i$ are the probabilities p_b, p_c, p_l . The model is trained on jets from MC simulations using gradient descent. RNNIP has many advantages over traditional b-tagging algorithms, this algorithm is covered in more detail in section 6.3.2.

A final example of an RNN is the model presented in chapter 7 of this work. The RNN is again used within the context of b-tagging but instead focuses on identifying the positions of secondary and tertiary decay vertices within jets. These form from the decay of the b- and c-hadrons. The model has a similar architecture to RNNIP, using LSTM cells and then a feed-forward network applied to the final LSTM output. Each jet is again an ordered sequence of tracks $\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_t)$. Each track is a vector of several variables (for example the track parameters). The key difference is the output of the model corresponds to the predicted positions of secondary and tertiary vertices in the jet, $\mathbf{y} = (x_s, y_s, z_s, x_t, y_t, z_t)$. As it is a regression task, an appropriate loss function such ordinary least squares (equation (5.7)) can be used.

5.2.4 Other Neural Networks

A variety of different neural network types have been developed over the years which have advantages in performing certain tasks. Convolutional neural networks

(CNNs) use convolution layers to scan across input data, analogous to sliding a window across an image. CNNs are particularly useful at image processing tasks. Graph neural networks (GNNs) [90] are a very recent development in the field of ML. These operate directly on graph representations of data, a structured form modelling relationships (edges) between objects (vertices). Vertices (or nodes) can be thought of as some feature vector, edges are generally some neural network connecting one vertex to another. The values of vertices and edges are learnt in the training of GNNs.

5.2.5 Variational Autoencoders

A variational autoencoder (VAE) is a machine learning model built for the task of data generation. Generating new realistic datapoints is a hot topic of research within ML, and many so-called generative models have been produced.

5.2.5.1 Autoencoders

The principle behind an autoencoder (AE) is dimensionality reduction. It is formed of two networks called an encoder and decoder, see figure 5.5. The encoder takes an original data vector $\mathbf{x}$ and reduces it to some smaller dimensional vector $\mathbf{z}$, called the latent. The decoder is then responsible for restoring the original data as well as possible from $\mathbf{z}$. This reconstruction, $\mathbf{x}'$ has the same dimensionality. Autoencoders are trained end-to-end by minimising some loss function representing the error between $\mathbf{x}'$ and $\mathbf{x}$.

5.2.5.2 Variational Latent Space

The decoder of an AE would at first seem a good candidate for a generative model. By passing random samples in the space $\mathbf{z}$ through the decoder we might hope to generate some new sensible looking data. However this requires the latent space ($\mathbf{z}$ space) to have some structure reflecting the distributions of $\mathbf{x}$. A vanilla AE will not necessarily learn such a space as it can just encode datapoints in a deterministic manner.

To create a decoder capable of generating data a VAE is constructed [92]. The latents $\mathbf{z}$ outputted by the encoder now must correspond to the mean and variance of

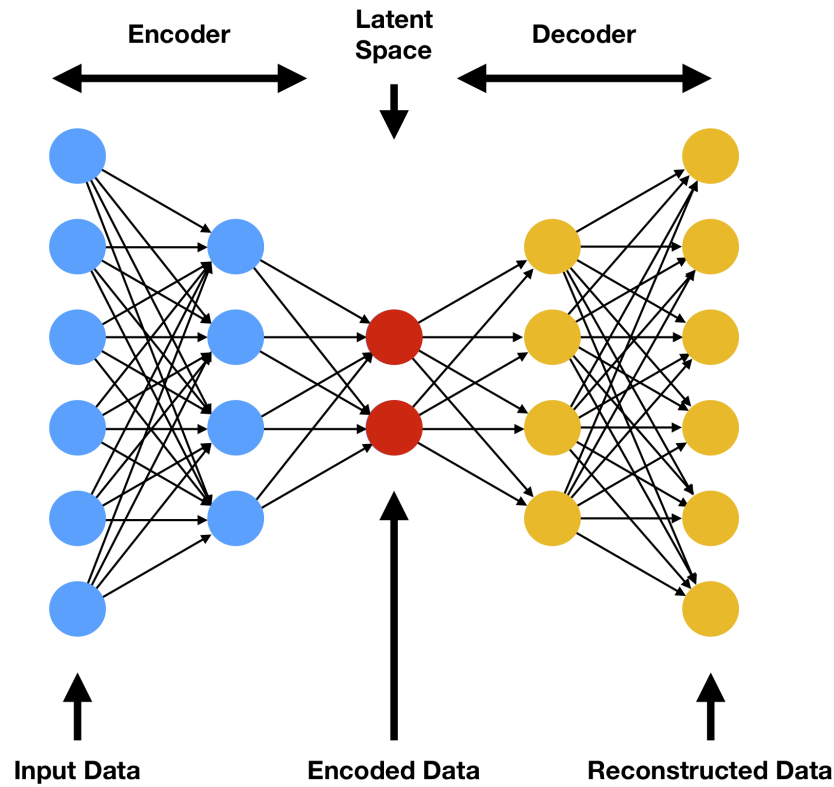


Figure 5.5: A schematic of an autoencoder showing the encoder and decoder networks, and the latent encoding [91].

a Gaussian distribution (in effect we double the number of values in z and take half to be means, and half to be variances). This enforces a regularization of the latent space such that the decoder will now form sensible data if we were to pass random samples from latent space through it.

The training of a VAE is different to an AE. The encoder produces means and variances that characterize normal distributions for each latent dimension. Points are then sampled from these distributions and the resultant vector is passed through the decoder. The loss function used to train this VAE in an end-to-end way gains a term responsible for regularizing the latent space. This function is derived from maximum likelihood estimation methods when applied to the proper mathematical formulation of the VAE model. A derivation of this can be found in [79] or [92].

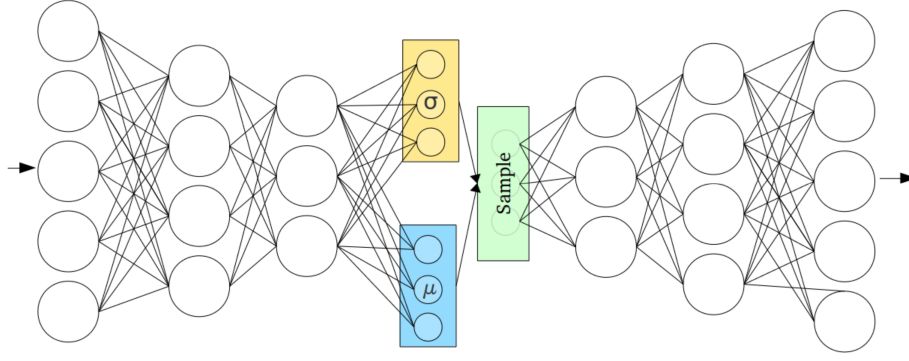


Figure 5.6: A schematic of a VAE. The input data enters on the left, the reconstructed data leaves on the right. The latents produced by the encoder correspond to means and variances of normal distributions. The z are sampled from these distributions and pass through the decoder [93].

5.3 Working Points

ML and other multivariate algorithms deployed in ATLAS will often have a number of different working points (or operating points) defined. In a classification task such as b-tagging, the output of an algorithm such as a BDT is usually chosen to be some continuous value as opposed to a discrete label. This value can be interpreted as a probability of a certain class. Consider an algorithm (BDT or neural net) attempting to classify events as signal (S) or background (B). The classifier outputs a value p for a given event. If $p > 0.5$, we label the event as S, and label it B otherwise. The value 0.5 is referred to as our threshold, the value p is referred to as the discriminant.

For a given threshold value, the following quantities are defined:

- True Positive Rate (TPR): The fraction of correctly labelled signal events. Also called the signal efficiency or sensitivity.
- True Negative Rate (TNR): The fraction of correctly labelled background events. Also called the background efficiency or specificity.
- False Positive Rate (FPR): The fraction of incorrectly labelled signal events. Also called the signal inefficiency. This is related to TNR by $FPR = 1 - TNR$.
- False Negative Rate (FNR): The fraction of incorrectly labelled background

events.

In ATLAS a further quantity known as background rejection is defined, this is the inverse of the background efficiency.

Modifying the threshold value will cause the signal and background efficiencies to shift. A tight threshold will lead to a very high TPR. Using the example above, we might require $p > 0.99$ (p is some measure of the probability an event is signal-like). Conversely a loose threshold will lead to lower TPR but a better TNR. The effect of modifying the threshold value is usually shown in a receiver operating characteristic (ROC) curve (figure 5.7). In ATLAS, ROC curves generally plot background rejection on the y-axis and signal efficiency on the x-axis, for instance in figure 4.2 where the signal event is a b-jet, and the background is a light-jet. This is done as the signal efficiency is generally the parameter we fix. The background rejection is preferred as it can be plotted on a log-scale and at ATLAS performance tends to be very high (background rejection can reach up to $\sim 10^5$ equivalent to a FPR of 0.99999).

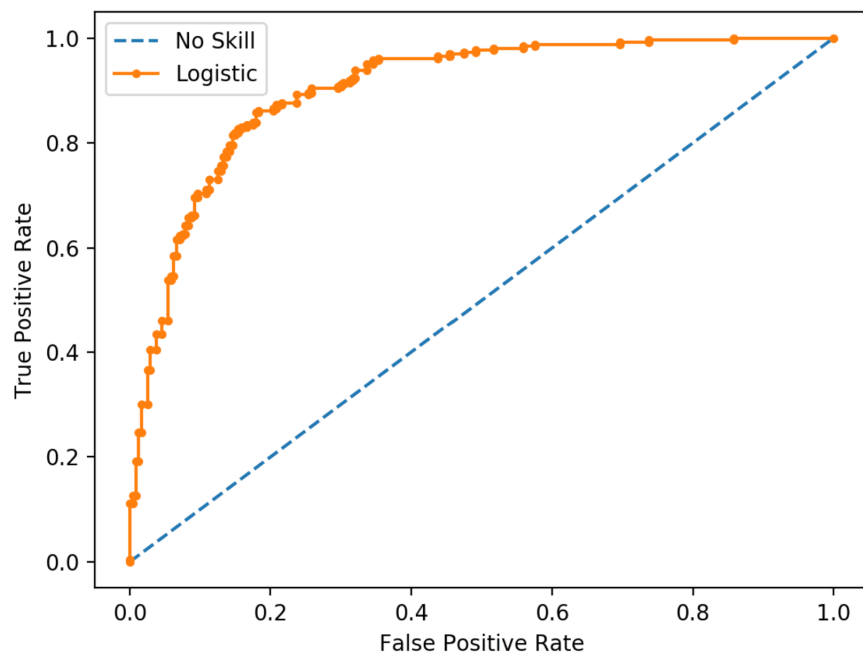


Figure 5.7: An example ROC curve showing the TPR vs FPR for a logistic regression classifier (orange). Each point on the ROC curve corresponds to a different threshold value. The optimal classifier is in the top left corner. A random classifier (no skill) is represented by the blue dashed line [94].

Chapter 6

b-Tagging

6.1 Physics Motivation

The identification of jets containing a b-quark, known as b-tagging, plays a key role in physics analysis at ATLAS. As one of the heaviest known fermions, it is important in searches for new physics at high energy scales and for testing the predictions of the SM. It is essential in studies of the two heaviest known particles, the Higgs and top quark. The identification of the jet origin is more broadly referred to as flavour tagging, as more than just the b flavour jets can be identified. For instance, many b-tagging algorithms are capable of discriminating between jets from c quarks, light quarks (u,d,s) and those forming from τ lepton decays¹.

The top quark decays before hadronising and can only be studied via its decay products. In practice, this will always be a b quark as $|V_{tb}| \sim 1$. Thus b-tagging is essential in studying the top. The top quark is of particular interest to particle physicists because of its large mass. It is thus likely to couple to high energy scale BSM physics, and gives physicists the opportunity to study a bare quark.

The Higgs boson decays preferentially to b-quarks. This is because it couples to mass, the b is the heaviest particle it can feasibly decay to. The branching fraction of $H \rightarrow b\bar{b}$ is around 58%, see table 2.3. Studies of the Higgs boson and its properties therefore often involve b-tagging.

¹The τ lepton has a mass of 1778.86 MeV allowing it to decay hadronically (around 65% of the time), leaving a hadronic jet signal in the detector.

6.2 *b* decay properties

Several important properties of *b*-quarks are exploited to perform *b*-tagging. A *b*-quark will hadronise to a *b*-hadron (around 91% of the time this is a *B* meson) [5]. This hadron may be in an excited state (and decay strongly), but the decay chain will always contain a weakly decaying *b*-hadron. These quasi-stable hadrons have longer lifetimes than most others ($\sim 10^{-12}$ s). They also have very large masses, around 5 times the mass of a proton. The most common kinematically possible decay for the *b* is to a *c*-quark, but this decay is suppressed by the small CKM matrix element $|V_{cb}|$ leading to longer lifetimes than expected. The fragmentation of *b*-quarks is hard, meaning the *b*-hadron generally carries a large fraction of the *b*-quark momentum. Combined with their long lifetime, this means *b*-hadrons typically travel several mm in the detector. A *secondary decay vertex* is thus resolvable in the *b*-jet, occurring where the *b*-hadron decays. The *b*-hadron will in most cases decay producing a hadron with a *c*-quark (*c*-hadron). These have only slightly shorter lifetimes than *b*-hadrons² ($\sim 10^{-13}$ s) and can similarly travel far enough in some instances for a *tertiary displaced vertex* to be resolved, at the position of the *c*-hadron decay. These similar properties of *c*-hadrons to *b*-hadrons means *c*-jets are often identified by flavour tagging algorithms alongside *b*-jets and light quark jets. The *b*-hadron decay also occurs with high charged track multiplicity, typically leaving around 5 stable charged products [95].

Leptons are an important signal in *b*-tagging. Around 42% of *b*-quark decay chains contain an electron or muon. These arise in the semi-leptonic decay of the *b*-hadron ($\sim 11\%$) and the *c*-hadron ($\sim 10\%$). The leptons are much less massive than the parent hadron and thus tend to have a large momentum perpendicular to the jet axis (as this approximates to the *b*- and *c*-hadron flight direction).

6.3 Overview of ATLAS Techniques

Several algorithms exist in ATLAS to perform *b*-tagging. These are broadly categorised as low-level and high-level *b*-taggers. Low-level *b*-taggers operate on indi-

²Whilst the decay $c \rightarrow s$ is not suppressed by the CKM matrix, the weak decay is still a slow process.

vidual jets and produce a number of output variables³ which can be used to classify jets by flavour (b, c or light). Light jets refer to jets forming from a light quark (u, d, s) or gluon. High-level b-taggers are multivariate algorithms which take the outputs of low-level algorithms as their inputs.

The jets used by the b-tagging algorithms are particle flow jets (AntiKt4EMPFLOW). AntiKt4EMTopo jets were used in the past. Tracks are associated to these jets using the ΔR between the jet axis and the track momentum. The jet axis is the vectorial sum of the cluster objects defining the jet. The track momenta are defined at the perigee: the position of closest approach along a track helix to the z-axis.

6.3.1 Impact Parameter based methods

Impact parameter (IP) methods exploit the displacement of tracks from secondary and tertiary vertices, i.e. the tracks from decay products of the b- and c-hadrons, to discriminate b, c and light jets [96]. The displaced tracks are differentiated from primary vertex tracks using the signed transverse and longitudinal impact parameter significances:

$$S_{d_0} = \text{sign}_{r\phi} \frac{d_0}{\sigma_{d_0}} \quad (6.1)$$

$$S_{z_0} = \text{sign}_z \frac{z_0}{\sigma_{z_0}} \quad (6.2)$$

The lifetime signs, sign_z and $\text{sign}_{r\phi}$, are defined as positive if the track crosses the jet direction in front of the primary vertex and negative if it crosses behind. Tracks from secondary and tertiary vertices are expected to have positive lifetime signs. Probability distribution functions (PDFs) are derived from the distributions of S_{d_0} and S_{z_0} (see figure 6.1) as for each jet flavour. These additionally depend on the track hit pattern to increase discriminating power: tracks are assigned a category according to the number of missing, expected and observed hits in the different layers of the ID [97]. This helps discriminate tracks by their origin (e.g. secondary or primary vertex) and quality. Two different methods are used: IP2D where the

³These express some characteristic of the jet, such as the number of tracks or its invariant mass.

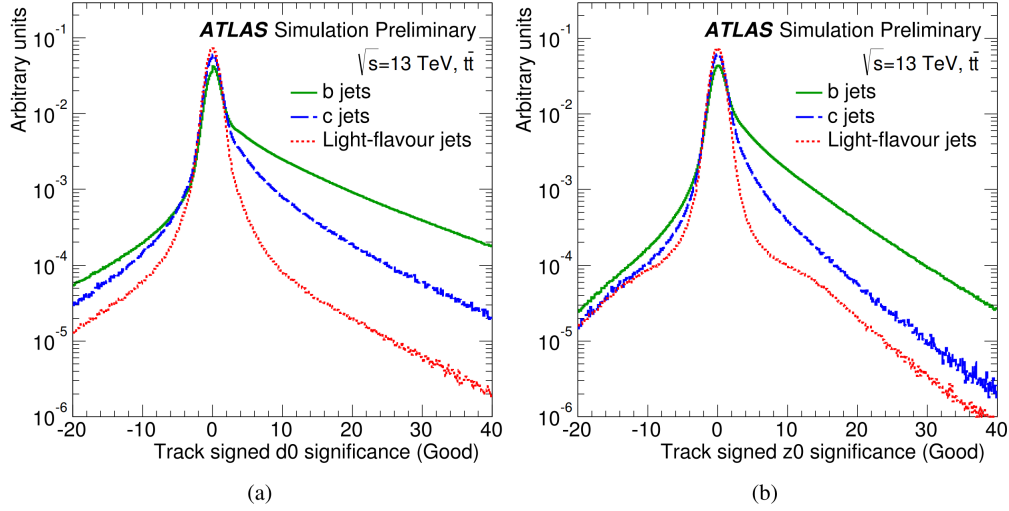


Figure 6.1: The distributions of lifetime signed (a) transverse and (b) longitudinal impact parameters for b, c and light-jets in a $t\bar{t}$ sample. The b and c jets tends to have more tracks with large positive lifetime-signed impact parameters. The light jets also show an exponential tail in this direction due to tracks from K_s and Λ decays. The longitudinal impact parameter (z_0) shows an additional bump in the light-jet tail due to pile-up [97].

PDF is a 1D function of S_{d_0} and IP3D where the PDF is a 2D function of S_{d_0} and S_{z_0} :

$$\text{IP2D} : f_i(IP_k) = f_i(S_{d_0}) \quad (6.3)$$

$$\text{IP3D} : f_i(IP_k) = f_i(S_{d_0}, S_{z_0}) \quad (6.4)$$

for a track k in a jet flavour i . Log-likelihood ratios (LLR) are defined to discriminate the three jet flavours (b, c and light):

$$\text{LLR}_{ij} = \log \left(\frac{\prod_{k=1}^N f_i(IP_k)}{\prod_{k=1}^N f_j(IP_k)} \right) \quad (6.5)$$

for a jet with N tracks, $i, j \in \{b, c, u\}$ where u refers to light flavour jets. The ratios LLR_{bu} , LLR_{bc} and LLR_{cu} are used as inputs to high-level taggers.

6.3.2 RNNIP

RNNIP is a new ML based tagger that uses an RNN with the jet's tracks as inputs [89]. This algorithm can rectify several issues with the more simple LLR based

Track Variable	Description
Variables for IP3D and RNNIP	
S_{d_0}	The lifetime signed transverse impact parameter significance (d_0/σ_{d_0})
S_{z_0}	The lifetime signed longitudinal impact parameter significance (z_0/σ_{z_0})
Category [97]	Track categorization based on number of observed, expected or missing hits in the ID. Tracks are organised based on impact parameter resolutions.
Extra variables for RNNIP	
p_T^{frac}	The fraction of the jet transverse momentum carried by the track, p_T^{track}/p_T^{jet}
$\Delta R(track, jet)$	The ΔR between the track and the jet axis

Table 6.1: The track variables used in IP3D and RNNIP [89].

methods. In the IP2D and IP3D algorithms, it is assumed that the track flavour probabilities in a jet are independent. The templates used to define the PDFs also require large sample sizes and are too computationally expensive to produce if additional kinematic dependencies are added. Furthermore, the tracks in jets are not independent of each other, but instead will be spatially and kinematically correlated. For example, tracks originating from the same b-decay are related to each other. An RNN can account for the track correlations and additional dependencies in a fast and efficient way.

The RNNIP architecture is shown in figure 6.2. It consists of two parts: An RNN formed from LSTM cells and a fully-connected feed-forward network. Tracks in a candidate jet are formed into an arbitrary length sequence. They are ordered based on $|S_{d_0}|$ and passed to the RNN. The tracks are a vector of the variables defined in table 6.1. The output of the RNN then passes through a feed-forward network which produces the final output probabilities for each flavour $p_b, p_c, p_\tau, p_{light}$. A softmax⁴ layer ensures these probabilities sum to 1.

⁴This simply applies the softmax function to the output vector. The vector of k real values is converted to a vector of k values between 0 and 1, which all sum to 1. These can thus be interpreted as probabilities. The final size of a component is proportional to its initial value.

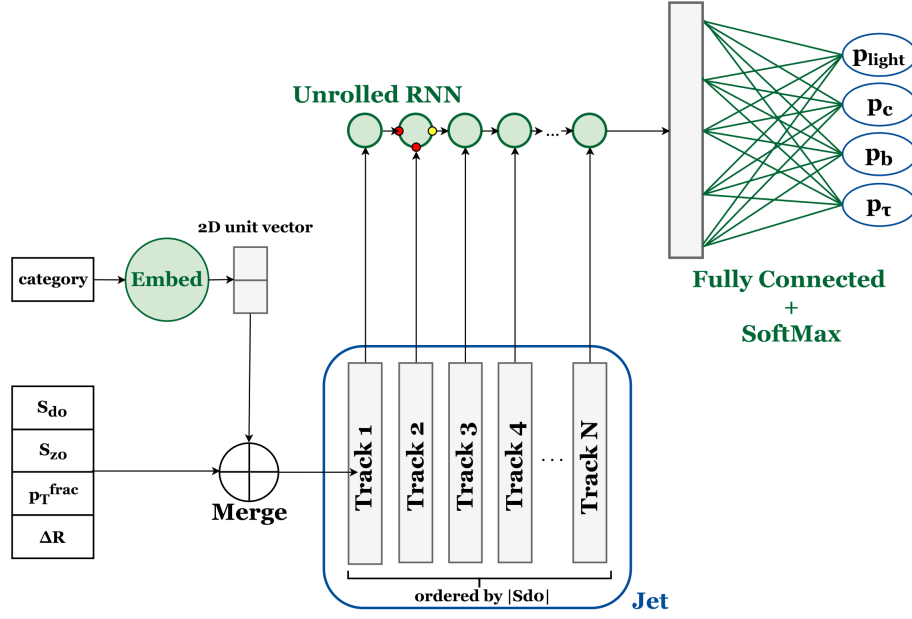


Figure 6.2: Schematic of RNNIP showing the network architecture, the inputs and the outputs. Note when used as an input to high-level taggers (MV2 and DL1), the RNNIP output is generally reduced to a 3-class p_b , p_c and p_{light} .

6.3.3 Inclusive Secondary Vertex Reconstruction

The displaced tracks from the secondary and tertiary vertices can be used to reconstruct an inclusive secondary vertex. This is done by the secondary vertex tagging algorithm SV1 [98]. The algorithm first reconstructs all possible two-track vertices from the selected tracks in the jet. It rejects those consistent with material interactions and neutral particle decays (such as the K_s and Λ). The tracks associated to remaining two-track vertices are then used to iteratively try and fit a secondary vertex. At each iteration, the least compatible track as measured by a χ^2 test of the track-to-vertex association is removed. This proceeds until the χ^2 of the vertex passes a fixed acceptance and the vertex mass is less than 6 GeV^5 , or until no more tracks are left.

The properties of the reconstructed secondary vertex are used by high-level taggers to discriminate b, c and light jets. These include, for example, the mass of the secondary vertex (assuming the pion mass tracks) and the number of tracks associated to it. The properties used are summarized in full in Table 6.2.

⁵This value is roughly related to the mass of b and c hadrons.

6.3.4 Topological Reconstruction

The JetFitter [99] algorithm exploits the full topology of the $b \rightarrow c$ decay chain to tag jets. It reconstructs both secondary and tertiary vertices using a Kalman Filter based algorithm. This algorithm is detailed in section 6.4. JetFitter assumes that the decay vertices all lie along the flight path of the b , this allows for the reconstruction of vertices from a single track. Several output variables from the fitted decay topology of a jet are used by high-level taggers, these are detailed in Table 6.2.

6.3.5 Leptons

The leptons produced in semi-leptonic decays of the b - and c -hadrons are a useful feature for b -tagging. The soft muon tagger (SMT) [96] exploits this fact, specifically looking at soft muons associated to jets. Due to the branching ratios, muons are expected in 20% of b -jets, and 10% of c -jets [100]. The muon is expected to have a large momentum perpendicular to the jet direction p_{Trel} . Combined muons reconstructed by ATLAS are associated with jets if the separation of the muon direction and jet axis $\Delta R < 0.4$. The SMT makes use of the ΔR , p_{Trel} and d_0 of the muon track alongside three variables quantifying the muon track quality:

- The scattering neighbour significance: $S = q \times \sum_i \frac{\delta\phi_{scat}^i}{\sigma_{\delta\phi_{scat}^i}}$ where q is the particle charge and $\delta\phi_{scat}^i$ the angular difference between two half-tracks starting or ending at the i -th pair of adjacent hits in the muon track. This quantifies how significant any kinks along the track may be.
- The momentum imbalance significance $M = \frac{p_{ID} - p_{MS}^{extr}}{\sigma_{E_{loss}}}$ where p_{ID} is the muon momentum measured in the ID, p_{MS}^{extr} is the momentum from the MS extrapolated to the vertex and $\sigma_{E_{loss}}$ is the uncertainty on the calorimeter measured energy loss.
- $R = \frac{(q/p)_{ID}}{(q/p)_{MS}}$ where $(q/p)_{ID(MS)}$ is the track curvature measured by the ID (or MS).

These variables are useful in separating muons from b - or c -hadron decays from muons originating in pion or kaon decays. The latter could occur in light-jets causing a fake signal of a b -jet.

These six variables can be passed directly to a high-level tagger, or combined into a single multivariate discriminant by an SMT BDT.

6.3.6 High-level Taggers

The optimal separation of b, c and light jets is achieved by high-level taggers. These are multivariate algorithms which combine the information produced by the low-level taggers. The two most important in ATLAS are MV2 and DL1.

6.3.6.1 Training Samples

The high-level taggers are ML algorithms and thus require labelled training data. This comes from MC simulation, the truth information is kept in MC samples to provide a truth label (b-, c- or light-jet). The samples are produced as a hybrid of $t\bar{t}$ and Z' events. The Z' events are important for jets with high p_T [101].

6.3.6.2 MV2

MV2 [96] is a BDT discriminant formed from the output variables summarized in table 6.2. The algorithm is trained using TMVA from the ROOT toolkit. As well as the outputs from low-level taggers, jet p_T and η enter as inputs to the BDT. Several different flavours of the MV2 algorithm exist. The default version, often called MV2c10, is trained on a hybrid sample where the background jets are 7% c-jets and 93% light jets⁶. These values were chosen to give the optimal balance between c-jet and light-jet rejection for the majority of ATLAS analyses [102]. The MV2 tagger uses only the inputs from IP2D/3D, SV1 and JetFitter. The recent addition of new low-level taggers has led to two new versions: MV2mu additionally uses the SMT BDT discriminant, and MV2MuRnn further adds the outputs of RNNIP (p_b, p_c, p_{light}). The performance of these are shown in figure 6.3. MV2 used to be the recommended b-tagger at ATLAS but has now been displaced by DL1.

6.3.6.3 DL1

DL1 [96] is a feed-forward neural network with multiple hidden layers. It is trained using Keras [103] (Theano backend) [104] with an Adam optimiser [105]. The

⁶Historically, this c-jet fraction was varied to produce taggers optimised for c-jet rejection (e.g. MV2c100 had a larger c-jet fraction) or light-jet rejection (e.g. MV2c00 where all background jets are light jets), these will not be considered in this work.

Input	Variable	Description
Kinematics	p_T	Jet p_T
	η	Jet $ \eta $
IP2D/IP3D	$\log(P_b/P_{\text{light}})$	Likelihood ratio between the b -jet and light-flavour jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio between the b - and c -jet hypotheses
	$\log(P_c/P_{\text{light}})$	Likelihood ratio between the c -jet and light-flavour jet hypotheses
SV1	$m(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion mass
	$f_E(\text{SV})$	Energy fraction of the tracks associated with the secondary vertex
	$N_{\text{TrkAtVtx}}(\text{SV})$	Number of tracks used in the secondary vertex
	$N_{2\text{TrkVtx}}(\text{SV})$	Number of two-track vertex candidates
	$L_{xy}(\text{SV})$	Transverse distance between the primary and secondary vertex
	$L_{xyz}(\text{SV})$	Distance between the primary and the secondary vertex
	$S_{xyz}(\text{SV})$	Distance between the primary and the secondary vertex divided by its uncertainty
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.
JETFITTER	$m(\text{JF})$	Invariant mass of tracks from displaced vertices
	$f_E(\text{JF})$	Energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{JF})$	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices
	$S_{xyz}(\text{JF})$	Significance of the average distance between PV and displaced vertices
	$N_{\text{TrkAtVtx}}(\text{JF})$	Number of tracks from multi-prong displaced vertices
	$N_{2\text{TrkVtx}}(\text{JF})$	Number of two-track vertex candidates (prior to decay chain fit)
	$N_{1\text{-trk vertices}}(\text{JF})$	Number of single-prong displaced vertices
	$N_{\geq 2\text{-trk vertices}}(\text{JF})$	Number of multi-prong displaced vertices
JETFITTER c -tagging	$L_{xyz}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Distance of 2^{nd} or 3^{rd} vertex from PV
	$L_{xy}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Transverse displacement of the 2^{nd} or 3^{rd} vertex
	$m_{\text{Trk}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Invariant mass of tracks associated with 2^{nd} or 3^{rd} vertex
	$E_{\text{Trk}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Energy fraction of the tracks associated with 2^{nd} or 3^{rd} vertex
	$f_E(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Fraction of charged jet energy in 2^{nd} or 3^{rd} vertex
	$N_{\text{TrkAtVtx}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Number of tracks associated with 2^{nd} or 3^{rd} vertex
	$Y_{\text{trk}}^{\text{min}}, Y_{\text{trk}}^{\text{max}}, Y_{\text{trk}}^{\text{avg}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Min., max. and avg. track rapidity of tracks at 2^{nd} or 3^{rd} vertex

Table 6.2: The variable inputs to the high-level taggers MV2 and DL1 (not including the SMT and RNNIP inputs). The JetFitter c -tagging variables are only used in DL1 [101].

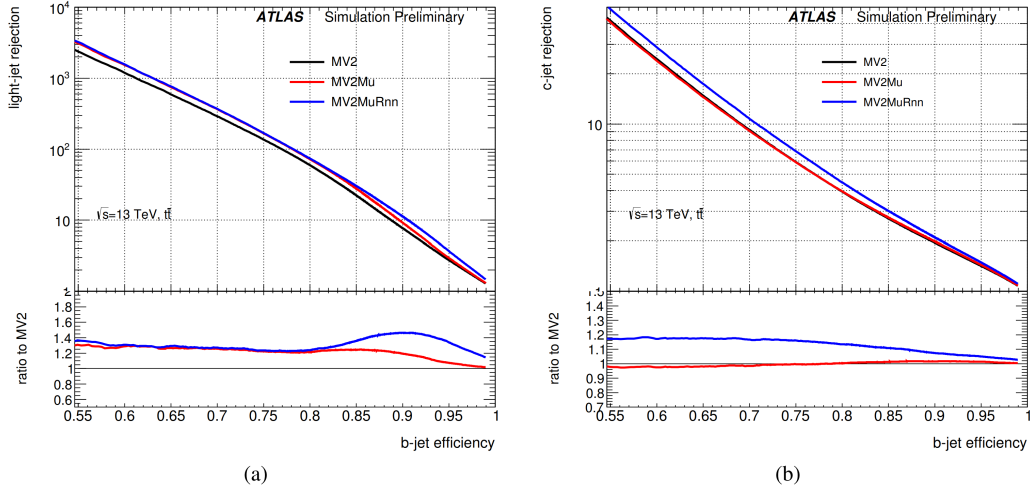


Figure 6.3: The b-tagging efficiency vs (a) light-jet rejection and (b) c-jet rejection of MV2c10 (MV2) with the SMT discriminant added (MV2Mu) and the RNNIP outputs added (MV2MuRnn) evaluated on a $t\bar{t}$ sample. The ratio of all taggers to MV2c10 is also plotted. The addition of muon information does not improve the c-jet rejection above 50% efficiency, since leptons are produced in both b- and c-decays [96].

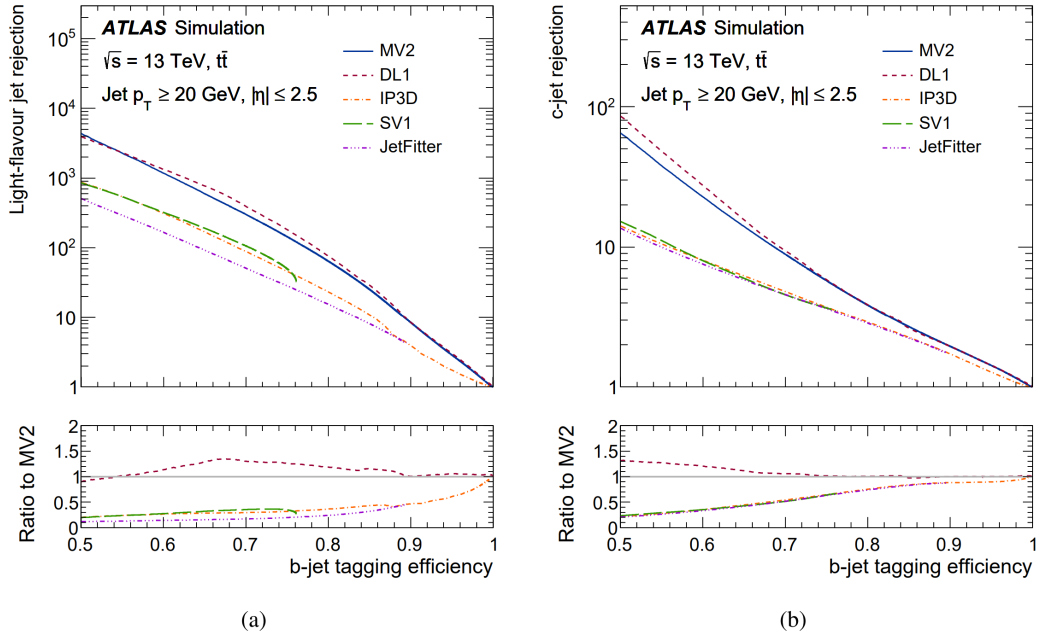
architecture is a mixture of feed-forward layers and Maxout layers [106]. Several hyperparameters such as network architecture, learning rate, number of epochs, etc. are tuned. The optimal parameters are shown in table 6.3. The inputs to DL1 are the same as for MV2 but additional JetFitter c-tagging variables are used (see table 6.2). All the inputs are summarized in Table 6.2. DL1 outputs three values corresponding to the probability of a b-jet, c-jet or light-jet (p_b, p_c, p_{light}). This allows DL1 to be used for b-tagging and c-tagging. The final b-tagging discriminant is defined as:

$$D_{DL1} = \ln \frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{light}} \quad (6.6)$$

where f_c is the fraction of c-jet backgrounds in the training sample. An optimal value of $f_c = 0.08$ is used. The performance of DL1 as compared to the other taggers is shown in figure 6.4, note for a fair comparison the extra inputs from RNNIP and SMT are not used for DL1 or MV2. DL1 is found to match or outperform MV2 everywhere.

DL1 can be upgraded similarly to MV2 to include inputs from the RNNIP and SMT taggers. DL1r includes the additional outputs of RNNIP (p_b, p_c, p_{light}) with

Hyperparameter	Value
Number of input variables	28
Number of hidden layers	8
Number of nodes (per layer)	[78,66,57,48,36,24,12,6]
Number of Maxout layers [position]	3 [1, 2, 6]
Number of parallel layers per Maxout layer	25
Number of training epochs	240
Learning Rate	0.0005
Training minibatch size	500

Table 6.3: The optimised hyperparameters used in DL1 [101].**Figure 6.4:** The b-tagging efficiency vs (a) light-jet rejection and (b) c-jet rejection of two high-level (MV2 and DL1) and three low-level (IP3D, SV1, JetFitter) taggers evaluated on $t\bar{t}$ events. The ratio of all taggers to MV2 is also plotted [101].

DL1rmu adding further the six variables from the SMT tagger. The performance of these taggers is shown in figure 6.5. The current recommended b-tagger for ATLAS analyses is DL1r.

6.4 JetFitter

JetFitter [108, 99] exploits the topology of the b-hadron decay (see figure 6.6). JetFitter makes the assumption that the b- and c-hadron decay vertices and the primary vertex lie on a straight line. This has several advantages: it allows for the reconstruc-

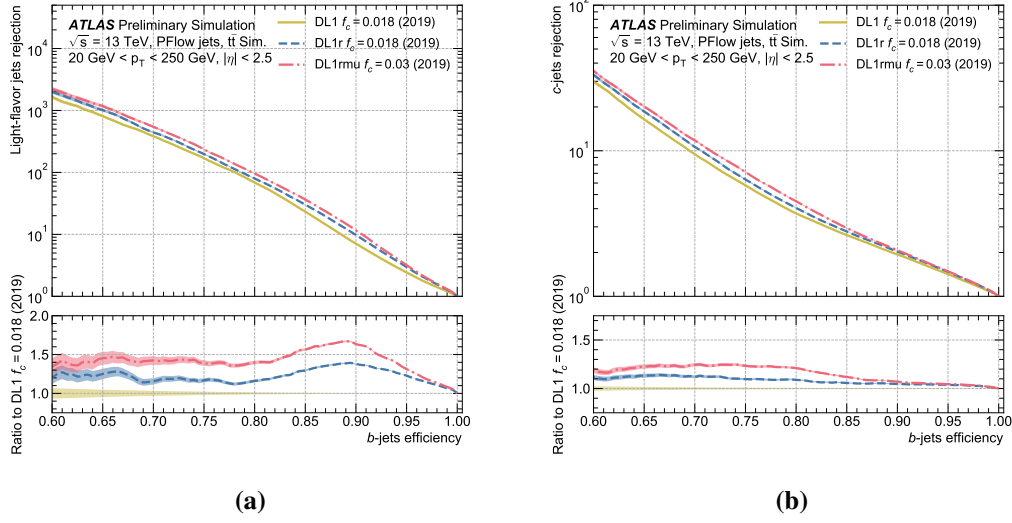


Figure 6.5: The b-tagging efficiency vs (a) light-jet rejection and (b) c-jet rejection of DL1, DL1r and DL1rmu evaluated on $t\bar{t}$ events. The ratio of all taggers to DL1 is also plotted [107].

tion of incomplete topologies (for instance where the only a single track for each vertex is found), it reduces the degrees of freedom of the fit increasing the chance of separating the vertices, and it increases the discrimination against light quark jets by specifying the compatibility to the given topology. The assumption is well physically motivated due to the hard b-quark fragmentation and the large masses of b- and c-hadrons. It is only in the transverse direction where this assumption can potentially cause bias as the track resolution can occasionally be better [108]. This is fixed by a looser cut on the χ^2 probability of the decay chain vertex. The straight line used is the b-hadron flight direction, approximated by the direction of the jet axis [108].

6.4.1 Track Selection

The tracks in a jet are associated during the clustering and formation of the jet. These in general contain tracks from the hadronization process (primary tracks), those from heavy hadron or neutral particle decays (secondary tracks) and tracks from other source such as pile-up. Most low-level b-tagging algorithms therefore apply a track selection procedure to remove jets that might obscure the flavour of the jet.

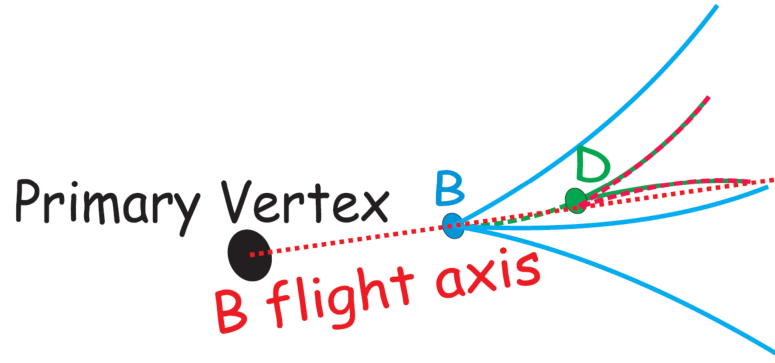


Figure 6.6: The b decay topology showing the decay vertex of the b-hadron (B) and c-hadron (D), and the JetFitter assumption that these lie along the B flight axis [108].

The JetFitter algorithm has a custom built track selection. Its aim is to remove all charged particle tracks not originating from b- or c-hadron decays. This includes primary tracks and tracks from other sources such as material interactions, neutral particle decays and pile-up.

Primary tracks are removed by cutting on the χ^2 compatibility to the primary vertex. This cut is dependent on the lifetime sign of the tracks. If the sign is positive the χ^2 probability must be below 10%. If it is negative, this cut drops to 5% as these tracks are less likely to come from the b- or c-hadron decays.

The remaining tracks (called secondary tracks) are paired up and examined as potential vertices. The two-track vertices are formed using a χ^2 vertexing procedure. The vertices failing cuts on χ^2 probability and lifetime-signed decay length significance (L/σ_L) are removed⁷. The resulting candidates are vetoed as hadronic interactions if the vertex position is in the boundaries between pixel layers [108]. Candidates can optionally be vetoed if they are compatible with photon conversion, K_s decays or Λ_0 decays. This is deduced by inspecting the invariant mass of the vertex candidate. Currently, this neutral particle veto is not applied, more detail on this is presented in section 6.5.

Tracks associated to the remaining vertex candidates only are used in a first fit of JetFitter (termed tracksToUseInFirstFit). These tracks are required to have

⁷accept vertices with ($P_\chi^2 > 3\%$ and $(L/\sigma_L) > 1.5$) or ($P_\chi^2 > 5\%$ and $(L/\sigma_L) > 1$)

$|d_0| < 3.5\text{mm}$, $|z_0| < 5\text{mm}$ and $p_T > 0.5\text{GeV}$. All combinations of two tracks are considered. Tracks can in principle be shared by multiple two-track vertices. If the formed two-track vertex is of good quality (as measured using a χ^2 test), and not consistent with a material interaction, its tracks enter the `tracksToUseInFirstFit`⁸.

The secondary tracks which did not pass the requirements above are used in a second fit if they pass certain quality requirements. These are termed `tracksToUseInSecondFit`. These tracks must fulfill $|d_0| < 1.5\text{mm}$, $|z_0| < 3\text{mm}$ and $p_T > 0.75\text{GeV}$. To further mitigate against pile-up, tracks with $|z_0 \sin \theta|/\sigma_{z_0} > 5$ and $|d_0|/\sigma_{d_0} < 2$ are not considered [99].

6.4.2 Vertex Finding

The vertex finding procedure is a two-stage process. The first stage takes the best secondary tracks (the `tracksToUseInFirstFit`) to fit a decay chain topology. A second stage then augments this topology using the remaining `tracksToUseInSecondFit`.

An initial jet topology is created assuming each track forms a single vertex along the b-hadron flight axis, which is taken as the jet axis. This is passed to the Kalman filter fitting procedure (see section 6.4.3). The resultant fitted topology then undergoes a clustering procedure:

1. Each vertex is assigned a χ^2 probability based on the compatibility of the tracks assigned to it and the compatibility of the vertex to the decay chain. If this probability is less than 0.1%, the vertex is removed.
2. All combinations of the remaining vertices are then formed and the χ^2 probability of the new vertex is calculated. This fills a table of probabilities as shown in table 6.4.
3. Any vertices (k) with a resultant χ^2 compatibility to the primary vertex $P_{kP} > 0.1\%$ are merged to the primary vertex.
4. The two vertices i and j forming the merged vertex with highest χ^2 score, P_{ij} , are merged. The entire topology is then refit with the Kalman filter.

⁸Note: Any track that is consistent with a material interaction two-track vertex will be hard vetoed out of the collection.

Vertex	1	2	3	...	Primary
1	X	P_{12}	P_{13}	...	P_{1P}
2	P_{21}	X	P_{23}	...	P_{2P}
3	...	...	...	...	P_{3P}
...	...	...	...	...	...
Primary	...	...	...	...	X

Table 6.4: The table of probabilities for merging combinations of vertices. The probability of merging vertex i and vertex j is P_{ij} assuming they form a new vertex along the b-hadron flight axis [108].

- Steps 2 and 3 are repeated until no more pairs can be formed with a χ^2 score above a certain threshold.

The threshold in step 4 depends on the reconstructed mass of the merged vertex, m_{sec}^{JF} . This is done to improve tertiary vertex reconstruction. The probability that a true vertex with mass $m_{sec}^{truth} > m_{sec}^{JF}$ is determined by comparing to truth secondary vertices in simulation. This value is then used to select the vertex merge threshold.

The second stage of vertex finding takes the returned topology from the first stage and adds the tracksToUseInSecondFit, again assuming they form single track vertices where they intersect the flight axis. The same fit and clustering procedure is then performed. The end result is a set of well defined (at least one-track) vertices forming the topology of the b decay. From this topology, the JetFitter output variables of table 6.2 can be deduced.

6.4.3 Fitting Procedure

JetFitter uses a Kalman Filter based fitting procedure [108]. The variables describing the decay chain to be updated are $\vec{d} = (x_{PV}, y_{PV}, z_{PV}, \phi, \theta, d_1, d_2, \dots, d_N)$, these are sketched in figure 6.7:

- The primary vertex coordinates (x_{PV}, y_{PV}, z_{PV}) (and covariance matrix) are obtained from the primary vertex finding.
- The flight direction of the b-hadron in spherical polar coordinates (ϕ, θ) is taken as the jet axis. Uncertainties are derived from the jet direction resolution and the difference between the b-hadron flight axis and jet direction in MC simulation [99, 108].

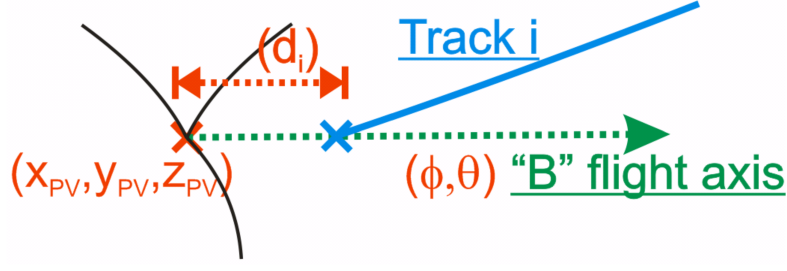


Figure 6.7: A schematic of a b-decay chain topology within a jet, showing the variables defined in the JetFitter fitting procedure [108].

- The distances along the flight axis of the single-track vertices are termed $d_1, \dots, d_N$. These are calculated by finding the intersection of each track with the flight axis using a Newton-based method [108].

At each step of the Kalman filter fit, the track parameters of an individual track are added and the vertex to be updated is specified. The Kalman Filter takes as its main input the measurement equation which relates the track helix parameters $\vec{q}$ to the vertex positions to be updated, and the momentum of the track evaluated at the vertex, $\vec{p}$. The measurement equation for a single vertex, position $\vec{x}$, is [109]:

$$\vec{q} = \vec{C} + A\vec{x} + B\vec{p} \quad (6.7)$$

where A and B are Jacobians, C is a constant vector. These encode the linearisation parameters. For JetFitter we replace $\vec{x}$ with the full decay configuration $\vec{d}$ giving:

$$\vec{q} = \vec{q}(\vec{d}, \vec{p}) \Big|_{\vec{d}=\vec{d}_0, \vec{p}=\vec{p}_0} + \frac{\partial \vec{q}(\vec{d}, \vec{p})}{\partial \vec{d}} \Big|_{\vec{d}=\vec{d}_0} (\vec{d} - \vec{d}_0) + \frac{\partial \vec{q}(\vec{d}, \vec{p})}{\partial \vec{p}} \Big|_{\vec{p}=\vec{p}_0} (\vec{p} - \vec{p}_0) \quad (6.8)$$

which is linearised to be used in the Kalman Filter by a first order Taylor expansion. The JetFitter algorithm is adapted from pre-existing ATLAS algorithms for single vertex finding [110] by noting that the $\vec{x}$ position of the i -th vertex can be written:

$$x = x_{PV} + d_i \sin \theta \cos \phi \quad (6.9)$$

$$y = y_{PV} + d_i \sin \theta \sin \phi \quad (6.10)$$

$$z = z_{PV} + d_i \cos \theta \quad (6.11)$$

The values from (6.7) and (6.8) are thus equivalent $C = \hat{C}$ and $B = \hat{B}$, only A must be modified $\hat{A} = A \frac{d\vec{x}(\vec{d})}{d\vec{d}} \bigg|_{\vec{d}_0}$ by some transformation matrix, see for instance reference [108]. This allows the JetFitter Kalman Filter re-use the code from the main ATLAS vertex finding algorithms.

The Kalman Filter procedure applies each candidate track iteratively to update the decay chain $\vec{d}$. The procedure will generally cycle through all the tracks several times before the fit converges due to the non-linearity of the measurement equation. At the end of each cycle, an updated $\vec{d}$ and covariance matrix are produced. Convergence is defined using the χ^2 probability of the fit, if after a new cycle $|\chi_{new}^2 - \chi_{old}^2| < 0.001$. The procedure is exactly the same when multiple tracks are associated to a single vertex, i , in that case they enter with the same value d_i .

6.5 Neutral Tracks

Secondary tracks form not only from heavy hadron decays, but from the decay of neutral particles. The K_s , for instance, has a similar lifetime to b-hadrons ($\sim 10^{-11}$) and thus will decay at similar displacements from the primary vertex. The K_s can be produced during the decay chain of the b-hadron.

JetFitter can optionally identify two-track vertices compatible with a neutral particle decay. The two charged tracks forming this vertex can then be removed from the jet candidate, as they do not therefore originate in the B- or C-Hadron decays. JetFitter used to go further and explicitly reconstruct the neutral particle tracks from the identified charged tracks. These new neutral track objects could then be used as inputs to JetFitter, so long as they formed from the b- or c-hadron decay. This can improve performance due to the increase in number of candidate tracks, as these can be used to better fit the secondary vertex position [108]. This functionality was lost with the upgrades to the new software base and event data model (EDM) in release 19 of Athena.

K_s	Λ_0	Conversion
opposite charged tracks	opposite charged tracks	opposite charged tracks
$ m(\pi^+, \pi^-) - m_{K_s} < 18 \text{ MeV}$	$ m(p, \pi^-) - m_{\Lambda_0} < 7 \text{ MeV}$	$m(e^+, e^-) < 30 \text{ MeV}$
$S_{lsd} > 4.5$	$S_{lsd} > 4.5$	$S_{lsd} > 0$
$P(\chi^2_{track1} + \chi^2_{track2}) < 10^{-4}$	$P(\chi^2_{track1} + \chi^2_{track2}) < 10^{-4}$	
$P(\chi^2_{track1(2)}) < 10^{-3}$	$P(\chi^2_{track1(2)}) < 10^{-3}$	
$r_T > 2 \text{ cm}$ OR atLeastOneLoose	$r_T > 2 \text{ cm}$ OR atLeastOneLoose	

Table 6.5: The requirements on vertex candidates and their constituent tracks to be identified as neutral particle decays. The Boolean atLeastOneLoose is true if at least one of the tracks has $|d_0| > 1 \text{ mm}$ and $|z_0| > 2 \text{ mm}$. r_T refers to the transverse radius of the vertex candidate. S_{lsd} is significance of the lifetime signed distance between the primary vertex and vertex candidate lsd/σ_{lsd} .

6.5.1 Neutral Particle Identification and Veto

The neutral particle decays of interest are conversions, K_s and Λ . Their two-particle decays are:

$$\gamma \rightarrow e^+ + e^- \quad (6.12)$$

$$K_s \rightarrow \pi^+ + \pi^- \quad (6.13)$$

$$\Lambda_0 \rightarrow p + \pi^- \quad (6.14)$$

All possible two-track combinations of secondary tracks are examined as potential neutral particle decay vertices (see section 6.4.1). The selections applied to identify these are detailed in table 6.5. This includes requirements on oppositely charged tracks and the invariant mass of the two-tracks, assuming the tracks are the decay products in equations (6.12)-(6.14). Additional cuts are applied on the χ^2 compatibility of the tracks to the primary vertex, the impact parameters of the tracks, the significance of the lifetime signed distance between the candidate vertex and primary vertex and the transverse displacement of the candidate vertex from the origin. More detail on these selections and their motivation can be found in [108]. The vertices identified as originating in neutral particle decays can be used in a track veto procedure. If this option is turned on, the two tracks belonging to the identified neutral vertex are removed from the fit.

The effect of the neutral candidate veto is investigated using an MC simulated

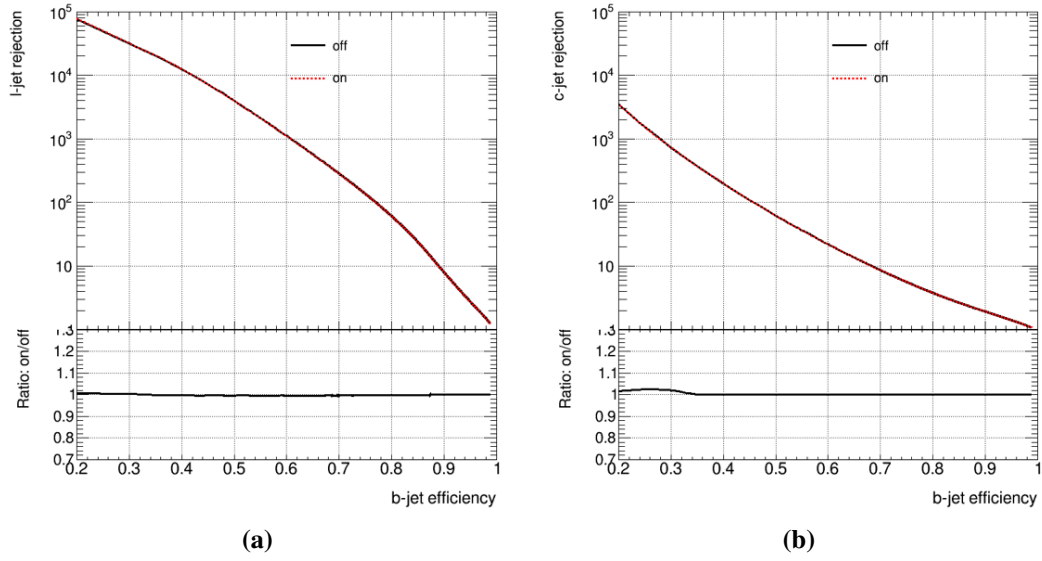


Figure 6.8: The effect on a MV2c10 of removing the tracks identified with neutral decay vertices. The b-jet efficiency is plotted against (a) light-jet rejection and (b) c-jet rejection. The black line is with the veto turned off, red dashed line is with the veto turned on. The ratio of the red curve divided by black is plotted.

$t\bar{t}$ sample⁹. The AntiKt4EMTopo jet collection is used. Neutral decay vertices were only identified in less than 2% of events. The result on b-tagging performance is shown in the MV2c10 ROC curves in figure 6.8. A small boost in c-jet rejection was seen at very low b-jet efficiency but performance was unchanged at the normal working points. No significant effect was seen for light-jet rejection at any efficiency. As such, the neutral candidate veto is not used by default. An old version of MV2 which uses only the JetFitter variables as inputs was also used to test the effect of the neutral candidate veto. The light-jet rejection ROC curve is shown in figure 6.9. A small performance boost was seen ($< 2\%$). This JetFitter only BDT is no longer maintained as it is not used by any analysis stream. It is, however, still useful as a diagnostic tool to see the impact of changes to JetFitter variables alone. The veto appears to slightly improve JetFitter, but the benefit does not transfer to the high-level tagger MV2. The extra information used by the high-level tagger is already sufficient.

The effect of the veto was also investigated by counting the number of recon-

⁹mc16_13TeV.410470.PhPy8EG_A14_ttbar_hdamp258p75_nonallhad.deriv.DAOD_FTAG1.e6337_s3126_r10201_p3582

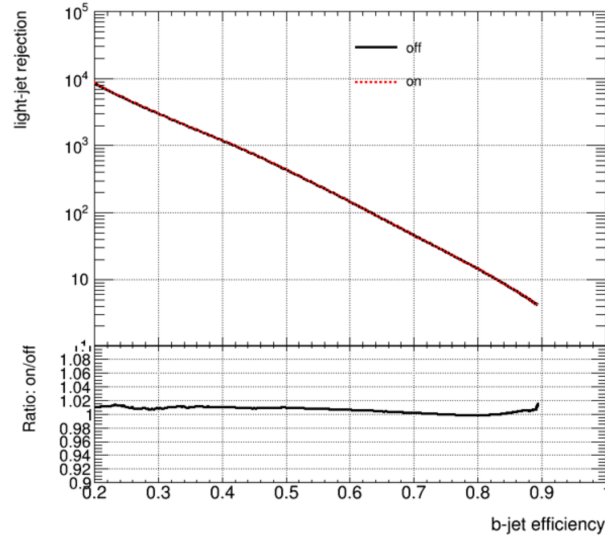


Figure 6.9: The effect on a JetFitter based BDT (with the MV2 architecture) of removing the tracks identified with neutral decay vertices. The black line is with the veto turned off, red dashed line is with the veto turned on. The ratio of the red curve divided by black is plotted.

N of JF Vtx	b	C	light	N of JF Vtx	b	c	light
0	17.6%	51.7%	90.0%	0	17.7%	51.9%	90.1%
1	49.2%	41.7%	8.9%	1	49.2%	41.7%	8.7%
2	25.3%	5.9%	1.1%	2	25.3%	5.8%	1.1%
3+	7.9%	0.8%	0.1%	3+	7.8%	0.7%	0.1%

(a) Veto Code Off
(b) Veto Code On

Table 6.6: The effect on the number of reconstructed JetFitter vertices of removing the tracks identified with neutral decay vertices. The percentage of different flavoured jets containing 0, 1, 2 or 3+ vertices is shown when the veto code is turned (a) off and (b) on. Results were produced on 130,000 selected jets from a $t\bar{t}$ sample.

structed vertices in different flavour jets. The jet flavour was identified using truth information in the MC sample, the number of JetFitter vertices was then counted. Additional selections on jets were applied: jet $p_T > 20$ GeV, jet $|\eta| < 2.5$, jet JVT parameter > 0.59 (to reduce pile-up), and the jet must not overlap with an electron¹⁰. The result on approximately 100,000 events is shown in table 6.6. Again, no significant changes were seen.

¹⁰A variable jet_aliveAfterOR must equal 1

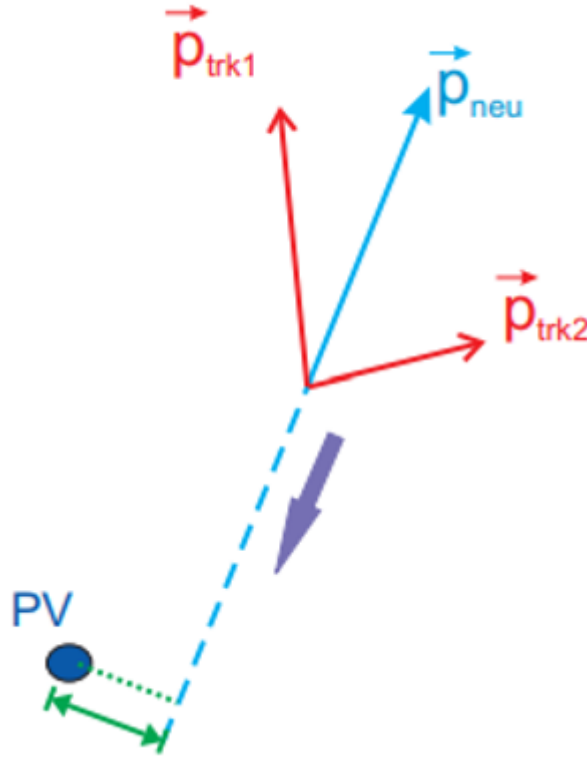


Figure 6.10: A schematic view of the two charged particle tracks forming from a neutral particle decay. The neutral track is reconstructed from these, extrapolation back to its perigee is shown (closest approach point of the primary vertex).

6.5.2 Neutral Particle Reconstruction

JetFitter historically went further than simply vetoing neutral decay vertex candidates. The corresponding neutral particle was reconstructed as a neutral track. This code functionality was lost in the recent upgrade of the ATLAS event data model (EDM). The re-created neutral particles could then be used in the JetFitter fitting procedure, entering as a new track for the first fit. This is physically motivated as neutral particles, such as the K_S or Λ , can be produced in the b decay chain. Re-creating neutral particles allows additional tracks from the b-decay to be used in the secondary and tertiary vertex reconstruction.

To restore this functionality, the old code was upgraded to be compatible with the new EDM. The code detail is described in appendix A. The neutral particle's track parameters are calculated from the track parameters of the two charged particles producing the vertex candidate. A schematic of this is shown in figure 6.10.

The neutral track momentum can be deduced from conservation of momentum. The track is assumed to pass through the vertex position, and is thus fully defined. To deduce the covariance matrix, correlations between the charged track parameters of the daughters and vertex position is needed. This is in a 9×9 covariance matrix denoted $\text{cov}[\vec{r}, \vec{p}_{trk1}, \vec{p}_{trk2}]$, where $\vec{r}$ is the vertex position. The correlations between the momenta of tracks 1 and 2 can be recovered by propagating the covariance matrices of the track momenta with the vertex position [108]:

$$\text{cov}[\vec{p}_{trk1}, \vec{p}_{trk2}] = \text{cov}[\vec{p}_{trk1}, \vec{r}] \cdot \text{cov}^{-1}[\vec{r}, \vec{r}] \cdot \text{cov}[\vec{r}, \vec{p}_{trk2}] \quad (6.15)$$

The covariance matrix for the neutral track $\text{cov}[\vec{r}, \vec{p}_{neutral}]$ is then obtained by error propagation. This 6D representation of the neutral track is then converted to the 5D perigee representation used at ATLAS ($d_0, z_0, \phi, \theta, 1/p$) (with the primary vertex as reference point) by noting the neutral track will travel in a straight line. It is entirely analogous to a charged track, except $1/p$ is used instead of q/p (for neutrals $q = 0$). The impact parameters are used to define a χ^2 compatibility to the primary vertex [108].

The recreated neutral track is now ready to be used in JetFitter. The created tracks are checked for their χ^2 compatibility to the primary vertex. Those with $\chi^2 < 36$ are flagged as primary vertex tracks and removed. The remaining K_s and Λ neutral tracks are available to be used in the fit (conversions are much more likely to be fakes).

The neutral reconstruction code was tested using the same $t\bar{t}$ sample as in section 6.5.1 and AntiKt4EMTopo jet collection. To measure the impact, the baseline version of JetFitter, with the neutral candidate veto switched on, was used. The neutral particle reconstruction is then added, and K_s tracks satisfying the above requirements were added to the fit¹¹. The impact on the performance of MV2c10 is shown in figure 6.11. Some benefit was seen at very low b-jet efficiencies, however no significant change was observed at the working points used in most ATLAS analyses ($> 60\%$). The impact on the individual JetFitter output variables is further

¹¹The addition of Λ tracks was tested, but this impact was small.

explored in appendix A.

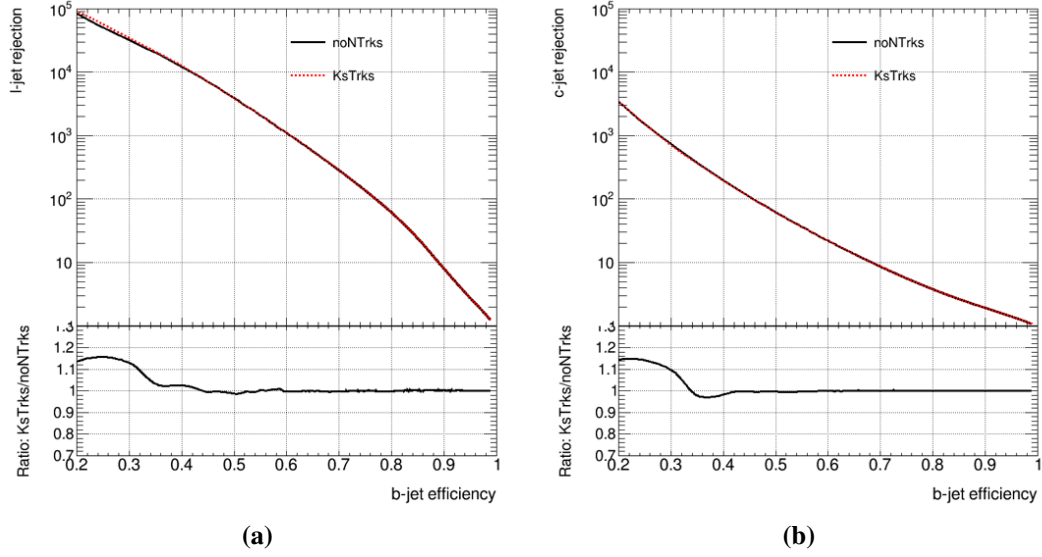


Figure 6.11: The effect on a MV2c10 of using reconstructed K_s tracks in JetFitter fit. The b-jet efficiency is plotted against (a) light-jet rejection and (b) c-jet rejection. The black line is with neutral particles identified and vetoed, red dashed line is with the addition of reconstructed K_s tracks to the fit. The ratio of the red curve divided by black is plotted.

6.6 Code Development

The efforts to restore neutral tracks to JetFitter uncovered a lot of issues with code readability and design. It was thus decided to complete an overhaul of the code structure of the JetFitter vertex finding procedure. The new design will maintain full backward compatibility while allowing easy extension, improvement and maintenance. A modular design makes it easier to replace certain aspects of the code, such as the track selection.

A map of the old code structure used by JetFitter is presented in figure 6.12. The vertex finding procedure and track selection are all contained within `InDetImprovedJetFitterVxFinder`. This file is thus extremely long, around 2000 lines of code. A more detailed view of this file can be seen in the second map, figure 6.13. A modularized version was proposed where the track selection, two-track vertex analysis (called V0s) and merge procedure form separate modules. The role of `InDetImprovedJetFitterVxFinder` is thus reduced to directing these procedures. The

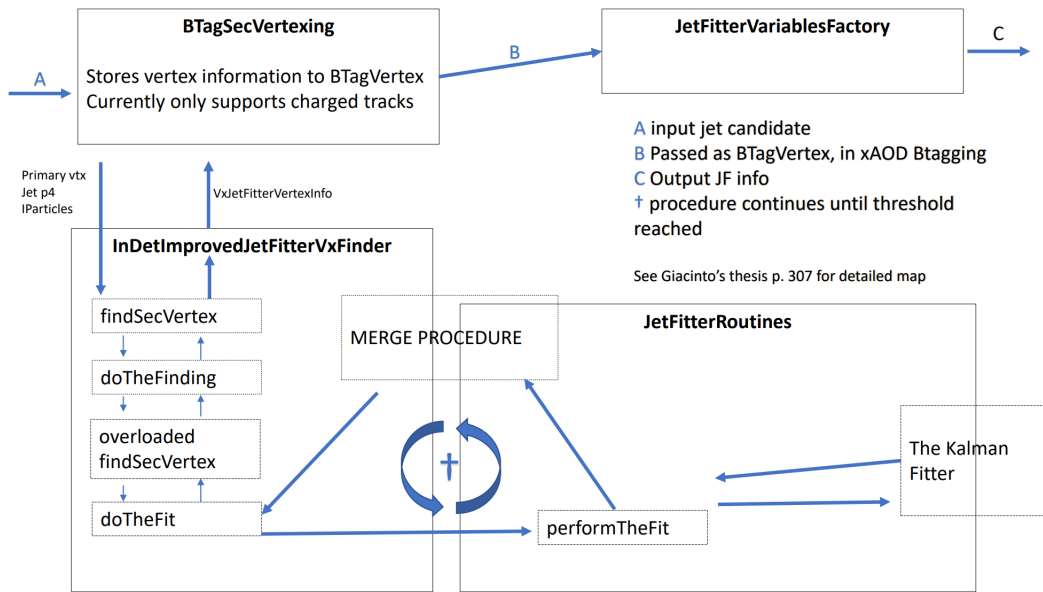


Figure 6.12: Map of the JetFitter code structure showing the four main modules. BTagSecVertexing is part of the main b-tagging algorithm and directs JetFitter to find its secondary vertices by calling InDetImprovedJetFitterVxFinder. The resulting decay chain topology is passed to JetFitterVariablesFactory to create the variables for high-level taggers. The flow of information is shown by arrows. Key methods within modules are shown. This map is not an exhaustive description but rather meant to show the important information only.

final refactored version of the code was under 300 lines of code. A map of the new code is shown in figure 6.14.

The refactor effort also added some improvements to the code. The use of smart pointers, new to C++ 11, was implemented. Some bugs to do with object ownership were uncovered and fixed. An explicit track ordering by p_T was implemented as the fitting procedure is dependent on track order. This ordering is the same as that used in the old JetFitter, but is now made explicit in the code.

With a refactored code, several avenues for future development open up. Each of the newly created modules could be replaced or upgraded. In recent years, ML algorithms have been increasingly favoured in particle physics and could serve as a good replacements. The JetFitter fitting procedure could be replaced a ML based algorithm, such as an RNN, this is explored in chapter 7. The two-track vertex finding and V0 identification procedures could be replaced by a GNN, for example see ref. [111, 112]. A GNN like this could potentially replace the entire vertex

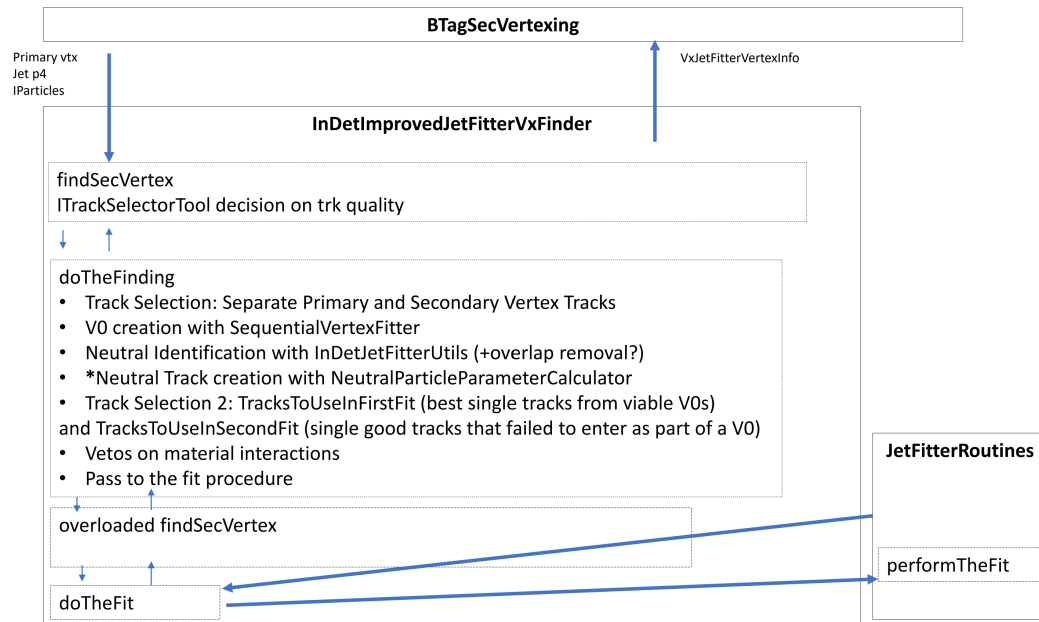


Figure 6.13: Map of the JetFitter code structure but with the InDetImprovedJetFitterVxFinder module, and in particular the track selection method doTheFinding shown in more detail. More expanded maps similar to this can be found in the appendix B.

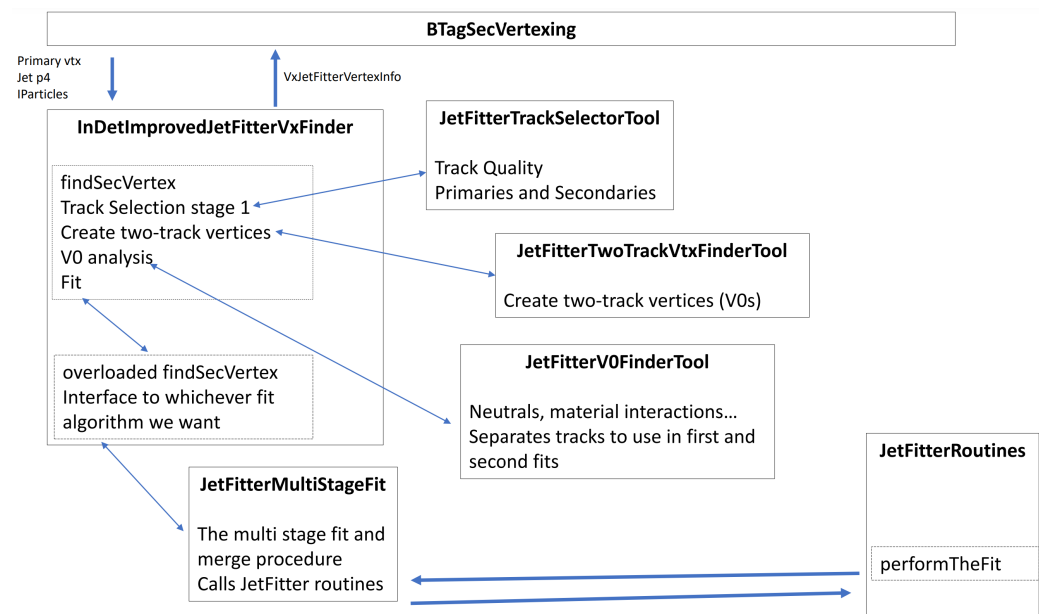


Figure 6.14: Map of the refactored JetFitter code structure. InDetImprovedJetFitterVxFinder has been greatly simplified. Four new modules have been created.

finding procedure. The end result of such a GNN would be a set of secondary vertices each with a number of tracks associated to them. This topology could then be fit by JetFitter's Kalman Filter, and the resultant JetFitter topology used to deduce the JetFitter output variables.

Some further examples of code development might be replacing the track selection with an external ATLAS tool. The track ordering procedure could also be investigated for possible gains. Cuts on m_{sec}^{JF} are currently used to determine when to merge vertices. These are based on results from simulation done in 2009 and could be re-optimized.

Chapter 7

Jet Topology Fitting with an RNN

7.1 Overview

This chapter presents work on an ML based topological jet fitter. The use of an RNN in fitting the positions of secondary and tertiary decay vertices in b-jets is explored, motivated in part by the JetFitter and RNNIP algorithms. The motivation is covered in more detail in section 7.2. The vertex fitting is a regression task. The model must predict the coordinates of the secondary and tertiary vertices, optimising a loss function based on the difference between the predicted and true vertex positions. The inputs to the model will be the ordered set of tracks in a jet, each track parameterized by some features. The training data is constructed from a simple model of jets. Each particle track in the jet is simulated alongside the decay vertices. The simple toy jet model is covered in detail in section 7.3. Jets of three different flavours (b , c and l) can be simulated, but only the b-jets are used.

The model is tasked with fitting the secondary and tertiary vertex positions in b-jets. With the scope of the model set and the training data generated, the detail of the model is covered in section 7.4. The training procedure and model architecture are laid out. Several experiments on optimising the model were performed and the results are quoted in section 7.4.2.

A major difficulty with the ML model is evaluating performance. The predictions of the model do not come with associated uncertainties. It is therefore not straightforward to interpret the errors on the predicted vertices with respect to the

true vertices. This problem is attacked in sections 7.5.1 and 7.5.2. A particular difficulty of this model is how to compare with the current b-tagging algorithms in ATLAS. The scope of the ATLAS models is not the same. They predict the jet flavour, not the secondary vertex positions. The exception is JetFitter, which uses the reconstructed vertices and their associated tracks to create the JetFitter variables. Some investigation into the association of tracks to vertices in the RNN model was therefore performed. The results of this are covered in section 7.5.3.

7.2 Motivation

The use of ML in ATLAS is an important part of the efforts to upgrade performance in advance of the high luminosity era. The increase in pile-up and data production rates brings new challenges that ML algorithms are well placed to solve. B-Tagging, both online and offline, is an area where ML algorithms have had considerable success. Two recent examples are DL1 [96] and RNNIP [89]. DL1 uses a deep neural net and is the current recommended tagger in ATLAS. RNNIP is a low-level tagger which uses an RNN (LSTM) to predict jet flavour. The choice of an RNN was motivated as they can account for the correlations between tracks in a jet. By expressing the tracks in this jet as a sequence, a considerable boost in performance over conventional IP based methods was obtained [89]. The RNN can further handle sequences of varying lengths, the number of tracks in a jet is not fixed.

A sequential treatment of tracks in a jet is not a new concept in b-tagging. The Kalman Filter procedure in JetFitter processes tracks in a sequence (ordered by p_T). RNNs, in particular LSTM, have been successfully used as replacements for Kalman Filters in several fields [113, 114, 115, 116]. The RNN offers several advantages over a Kalman Filter. Firstly, Kalman Filters operate under the assumption of linearity. In vertex fitting [110, 109] and JetFitter [108], this leads to the linearisation procedure (see section 6.4.3). An RNN makes no assumption of linearity and could thus be used to study inherently non-linear models such as vertexing. Kalman Filters also require an initial state with known covariances. JetFitter is able to deduce these up to certain approximations (e.g. the approximations of the flight

direction and single-track vertex positions in section 6.4.3). An RNN however can be initialized in a null state, such as in RNNIP [89]. There is thus motivation to investigate replacing the Kalman Filter fit procedure in JetFitter with an RNN.

7.3 Toy Jets

The training procedure for a ML model requires labelled training data. In ATLAS, these are produced using MC simulations based on SM theory and detector simulations. These MC simulated samples are complex and large. They contain a lot of information about ATLAS run conditions and underlying physics. As such, it was decided to investigate an RNN based jet fitting procedure on a simplified “toy jet” model. This allows much more control over key aspects such as noise, as well as a complete knowledge of the decay chain topology. If an RNN algorithm can do a good job at fitting a simplified model, it could then be extended to ATLAS simulated data.

The toy jets were created with minimal constraints from phenomenology. The jet is assigned an overall energy and a flavour (b, c or light). The jet is assumed to travel in the $\theta = \phi = \pi/4$ direction. This choice is arbitrary but will not limit the algorithm performance as the scenario under investigation is constructed to be spherically symmetric¹. The primary vertex is fixed at the origin, this is allowed from translational symmetry. The primary vertex particles, produced by the quark fragmentation, are then simulated. Finally, in the case of b and c jets, the heavy hadrons are propagated and decay to form the secondary and tertiary particles. In a light jet, all particles form at the primary vertex. The toy jet model uses only three particles: pions, b-hadrons and c-hadrons².

Primary vertex particles are formed from random fractions of the jet energy³. In light jets, the total jet energy is used. In b and c jets, a heavy hadron is created first and assigned a large constant fraction of the jet energy. This value is deter-

¹We assume no eta dependence on the detector resolution when adding noise, see section 7.4.2.6. In reality, there would be an η dependence to the noise due to the ATLAS detector geometry.

²The b- and c-hadrons do not correspond to real particles, instead they have a mass and lifetime roughly similar to the B and D mesons.

³The procedure ensures the overall jet direction is fixed.

mined from the observed fragmentation energies of b- and c-jets (80% for the b, 50% for the c) [95]. The heavy hadron is propagated some distance before decaying. This distance depends on the hadron momentum and lifetime (which follows an exponential distribution). The heavy hadron formed at the primary vertex is assumed to travel along the jet axis. The hadron then decays to two, three or four new particles. In the case of the b-hadron, one of these particles is a c-hadron which propagates further before decaying at a tertiary vertex. The jets resulting from this simple model thus contain the correct decay chain topology.

Once the particle content of the jet is simulated, a jet object is formed by converting all pions into tracks. For simplicity, the tracks are assumed to be straight lines. The tracks are parameterized using the same perigee system as the ATLAS tracks: $(d_0, z_0, \phi, \theta, q/p)$. Figure 7.1 shows an example toy b-jet with the tracks drawn as straight lines. To simulate the imperfect response of the detector, Gaussian noise can be added to each of the track parameters. The values of the noise is modified during experiments. Figure 7.2 shows the same example b-jet as before but with added noise.

The model above was used to create several training samples. The jet energy is uniformly distributed between 10 and 100 GeV. The distribution of total tracks in jets of each flavour category is plotted in figure 7.3. The momentum distribution of these tracks is plotted in figure 7.4. Similar distributions are seen in all three categories, although c-jets have on average lower track p_T (due to the fragmentation function). The secondary vertex positions for b and c-jets were similarly distributed, on average around 4mm (median was 2mm) from the origin. The tertiary vertices in b-jets had an average displacement of 7mm (median 5mm). These can be seen in figure 7.5, the distributions are exponential due to the heavy hadron lifetime distribution.

7.4 RNN Fitting

The toy jets are used to train an RNN designed to find the secondary and tertiary vertex positions. This is a supervised regression task.

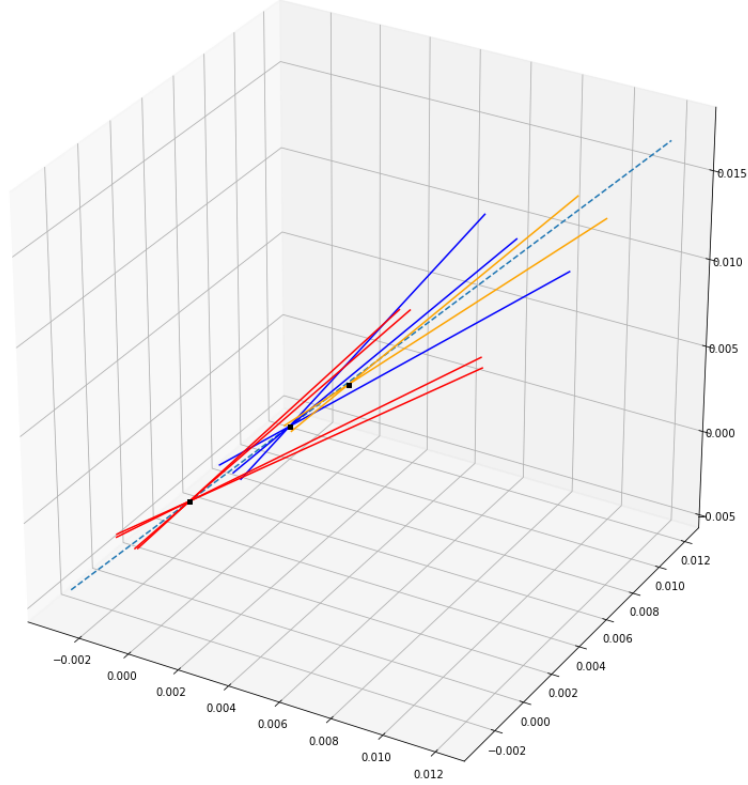


Figure 7.1: A toy b-jet with noiseless pion tracks originating from the primary (red), secondary (blue) and tertiary (orange) vertices. The vertex positions are denoted by the black squares. The jet axis is the dashed blue line.

The RNN architecture used is based on RNNIP. It is formed of an LSTM network, each cell has a single hidden layer, followed by a fully-connect feed forward network, with a single dense layer. The output of the RNN is six continuous variables corresponding to the Cartesian coordinates of the secondary and tertiary vertex, $(x_s, y_s, z_s, x_t, y_t, z_t)$ where the subscripts s, t denote secondary or tertiary vertex. Jets are processed by the RNN as ordered sequences of their tracks. The input features for each track can be freely chosen. As the number of tracks per jet varies, null tracks are added to ensure each jet has exactly 30 tracks. Initially, only b-jets were studied.

7.4.1 Training

The RNN model is trained using Keras [103] and Tensorflow [117] using the Adam optimizer [105]. The jets are labelled by the true positions of secondary and tertiary

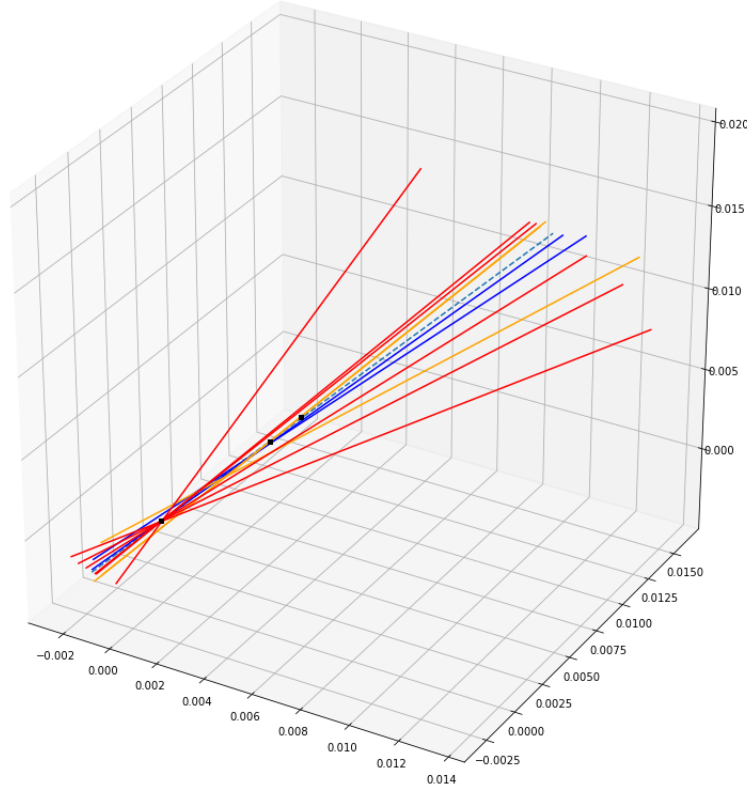


Figure 7.2: A toy b-jet with noise added pion tracks originating from the primary (red), secondary (blue) and tertiary (orange) vertices. The vertex positions are denoted by the black squares. The jet axis is the dashed blue line. Gaussian noise is added to the d_0, z_0, ϕ, θ (with $\sigma = 10^{-5}$) and q/P (a 0.1% error).

vertices⁴. The toy jet samples form a training set and held out test set as described in section 7.4.2.2. Regression loss functions, such as the mean squared and mean absolute error, are used. These are defined in section 7.4.2.1.

The models are trained for a maximum of 100 epochs⁵. An example loss curve is shown in figure 7.6 The model training set is split in a ratio 80:20 to form a validation set. Each experiment is repeated five times to measure an uncertainty on the final performance. The training split and initial network weights rely on a random seed, hence the same model trained again will not have the exact same performance.

⁴In the case of c- and light-jets, the missing vertex would be assigned the primary vertex position, (0,0,0).

⁵Note the training will early stop if validation loss does not decrease in 10 subsequent epochs.

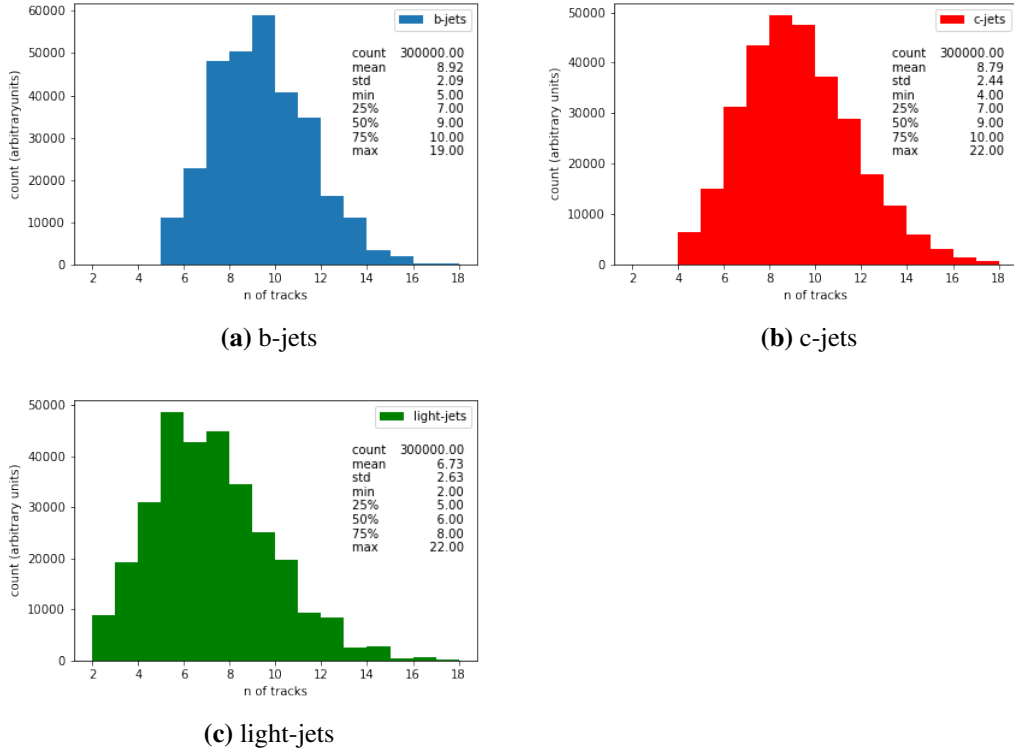


Figure 7.3: Histograms showing the number of tracks per jet for samples of 300,000 toy (a) b-jets, (b) c-jets and (c) light-jets.

7.4.2 Model Optimization

A number of experiments were carried out to create an optimal RNN based vertex fitter. The aspects of the model investigated are:

- Amount of Training Data
- Loss Function
- Feature scaling
- Track Ordering
- Feature Selection
- Noise

A baseline toy jet sample was created for these experiments with minimal (Gaussian) noise added: (d_0, z_0, ϕ, θ) take an absolute uncertainty with standard deviation

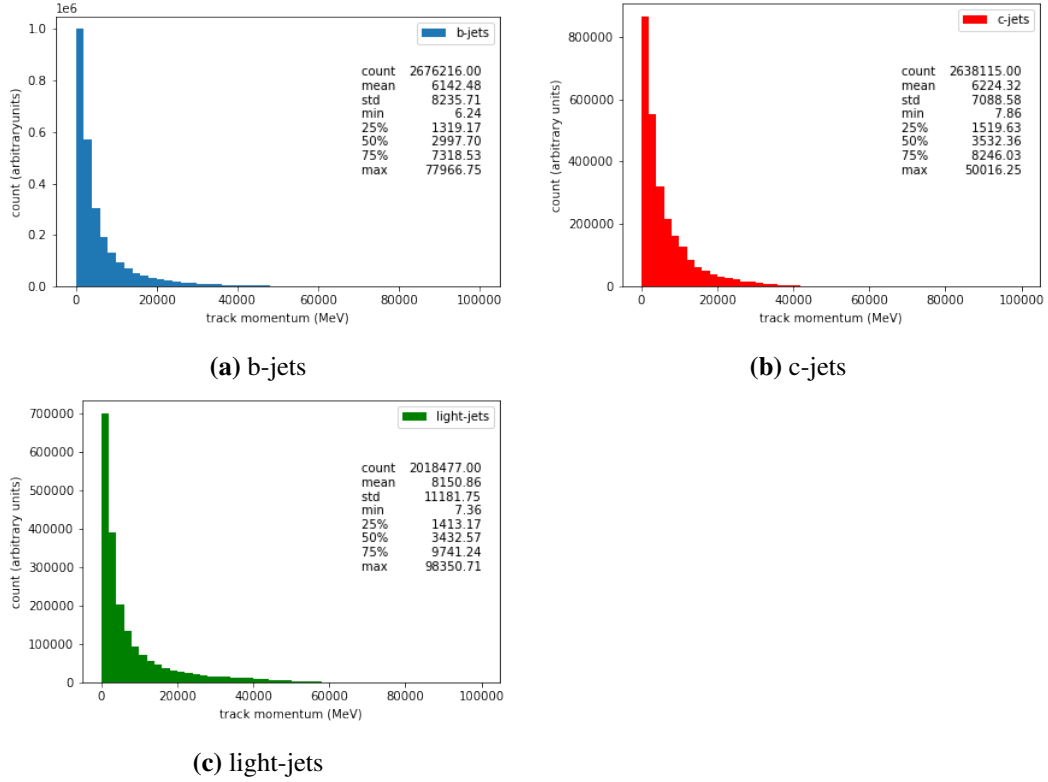


Figure 7.4: Histograms showing the track momentum across 300,000 (a) b-jets, (b) c-jets and (c) light-jets.

Experiment	Baseline Model Aspect
Training Set Size	20,000
Validation Split	0.2
Loss Function	MAE
Feature Scaling	MinMax for ϕ, θ , RobustScaler for $d_0, z_0, q/p$
Track Ordering	Random
Features	$d_0, z_0, \phi, \theta, q/p$

Table 7.1: The investigated model parameters, and the value of these in the baseline model.

of 10^{-5} (metres/radians), (q/p) is given a 0.1% uncertainty. A second sample was created when investigating the impact of noise in section 7.4.2.6. The results of each experiment are compared to a baseline model which has the aspects described in table 7.1.

The hyperparameters of the RNN model are the learning rate, batch size, and network size (number of nodes and hidden layers). An extensive hyperparameter tuning was not carried out due to limited time and computational resources. Some small experiments were carried out on each one individually using the RNNIP value

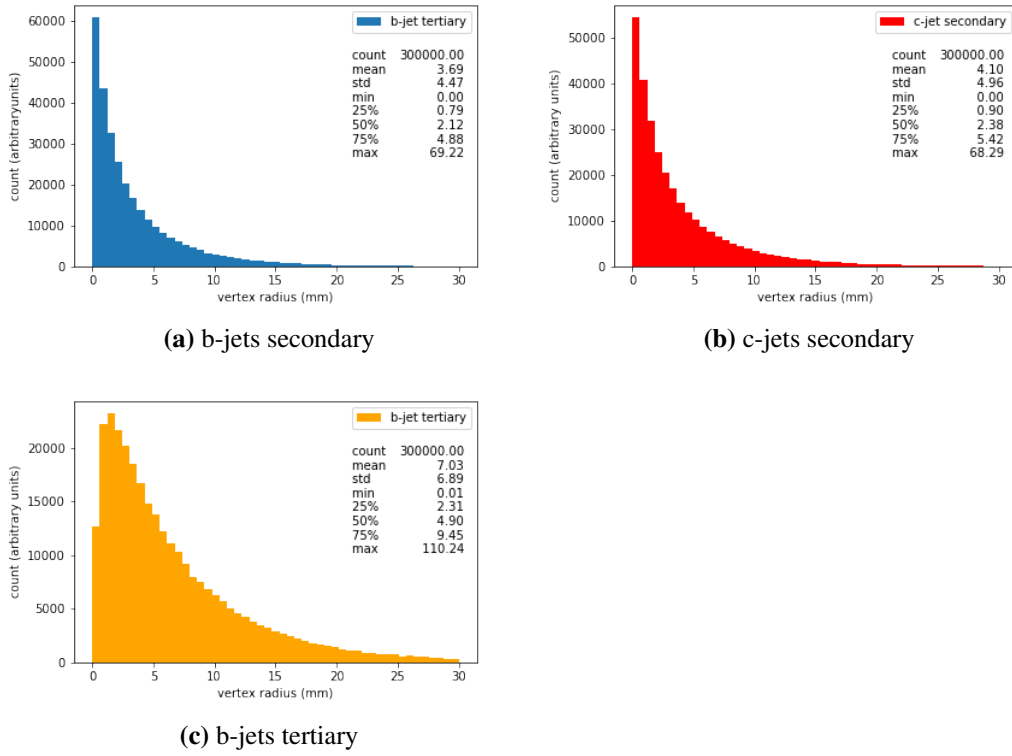


Figure 7.5: Histogram of the secondary and tertiary vertex distances from the primary vertex in b-jets ((a) and (c)) and c-jets ((b)).

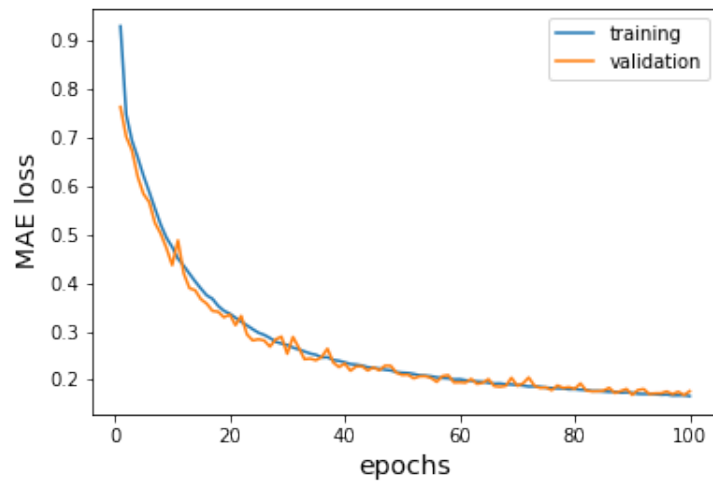


Figure 7.6: The training and validation loss (mean absolute error) versus the number of training epochs.

Hyperparameter	Value
Nodes in LSTM Layer	300
Nodes in Dense Layer	40
Learning Rate	0.0001
Batch Size	256

Table 7.2: The hyperparameters and their values in the the RNN model.

as an initial guess. The values used in all experiments are quoted in figure 7.2.

7.4.2.1 Loss function

A standard choice of loss function in regression problems is mean squared error (MSE). For our output $\vec{y} = (x_s, y_s, z_s, x_t, y_t, z_t)$ this is written:

$$L(\vec{y}, \vec{y}') = \frac{1}{n} \sum_i^n (\vec{y}_i - \vec{y}'_i)^2 \quad (7.1)$$

Where $\vec{y}_i$ is the true label and $\vec{y}'$ the predicted label of the i -th sample, and n the number of samples. Note the sum is over each of the 6 output values⁶. An alternative loss function, the mean absolute error (MAE), was found to outperform MSE. The MAE loss is written:

$$L(\vec{y}, \vec{y}') = \frac{1}{n} \sum_i^n |\vec{y}_i - \vec{y}'_i| \quad (7.2)$$

Again the sum is over all 6 output values⁷. The two metrics MSE and MAE are similar. Optimising either one during training will therefore to some extent also optimise the other. For instance, models trained to optimise the MAE loss achieved a similar final MSE (measured on the validation set) as those optimising directly the MSE loss. Table 7.3 shows the final MSE and MAE values measured on the validation set for two different models, one trained with an MAE loss the other with MSE loss. The MAE loss improves on performance for both metrics. This is likely because MAE is more robust to outliers in data. For this reason, the MAE is chosen

⁶As such it does not correspond to the sum of the L2 norm distances between true and reconstructed secondary and tertiary vertices. A loss function based on such a measure could be constructed but was not considered in our experiments.

⁷This is the sum of the L1 distance between true and reconstructed secondary, and true and reconstructed tertiary vertices.

Loss function	Final Validation MSE /mm ²	Final Validation MAE /mm
MSE Loss	1.82 ± 0.12	0.76 ± 0.02
MAE Loss	1.78 ± 0.16	0.67 ± 0.01

Table 7.3: The final validation MSE and MAE of a model trained on 20,000 toy b-jets using a MSE loss function or MAE loss function.

as the loss function.

7.4.2.2 Amount of training data

ML models perform better when more training data is used. However, this increases the training time. Therefore, most experiments were carried out with a smaller set of training data. The reduction in performance was investigated by training the baseline model on two training samples with different sizes. The model trained on 20k jets reached a final validation loss (MAE) of 0.67 ± 0.01 mm (early stopped after loss did not improve for 10 epochs). The same model trained on 280k jets achieved validation loss of 0.169 ± 0.002 mm after 100 epochs. The difference is significant, therefore the final optimized model is retrained with 280k jets. The improvements deduced while training with 20k jets are shown to improve the model trained on 280k jets as well (see section 7.4.3).

7.4.2.3 Feature Scaling

An important part of all ML tasks is data pre-processing. The goal is to present the data in the most optimal fashion to the neural network. As a general rule, features of equal importance should have similar orders of magnitude. Scaling involves projecting a feature x from its original domain to a new domain $[a, b]$. The simplest scaling process is minmax scaling:

$$x' = a + (b - a) \frac{x - \min(x)}{\max(x) - \min(x)} \quad (7.3)$$

The domain chosen in our work is $[a, b] = [-1, 1]$.

Another method of scaling is used which is more robust to outliers. This is good for features such as the longitudinal impact parameter z_0 ⁸, where a few outliers

⁸The RobustScaler on z_0 improved performance to 0.67 mm (0.80 mm with MinMax).

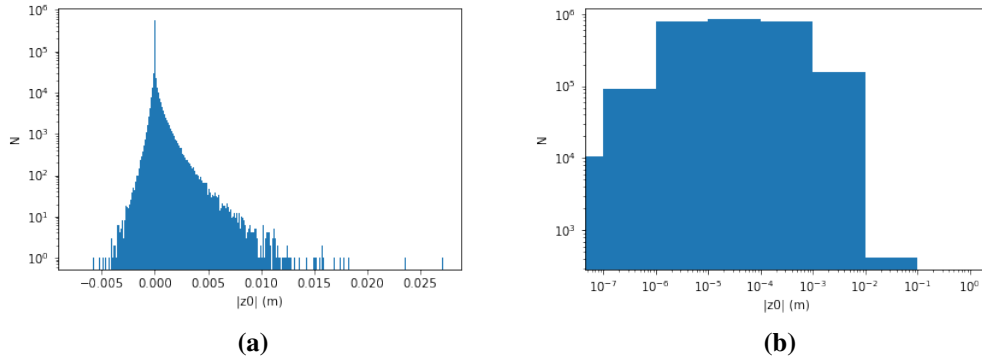


Figure 7.7: The distribution of the longitudinal impact parameter (z_0) in a toy jet sample of 3×10^5 b-jets, (a) shows the distribution of z_0 using 1000 equal spaced bins (b) shows the distribution of $|z_0|$ with a logarithmic binning.

are orders of magnitude larger than the majority of values as shown in figure 7.7. Robust scaling uses the interquartile range:

$$x' = \frac{x - \text{median}(x)}{Q3(x) - Q1(x)} \quad (7.4)$$

Where $Q3(x)$ and $Q1(x)$ refer to the 75th and 25th percentile values of the data.

Many experiments were performed with all combinations of scaling. The best scalings for each feature were deduced using a grid search, and are used in the baseline model. MinMax scaling is used for the track parameters ϕ and θ . Robust scaling is used for all other features ($d_0, z_0, (q/p), x_p, y_p$). The output vertex positions are measured in mm to be order unity.

7.4.2.4 Effect of Track Ordering

The order of tracks entering an RNN has an impact on performance. Several different orderings were tried. The results are presented in table 7.4, compared to the baseline which uses random ordering. In addition to three orderings based on the track parameters $d_0, z_0, q/p$, two additional derived orderings were tried: An ordering based on $r_0 = \sqrt{d_0^2 + z_0^2}$, and an ordering based on the closest point on the jet axis to the track. This latter ordering is the same as used in JetFitter. The distance along the jet axis can be derived by finding the closes points between the two skew

Ordering	Validation Loss (MAE) /mm
Random	0.67 ± 0.01
Increasing d_0	0.58 ± 0.01
Decreasing d_0	0.50 ± 0.02
Increasing $ z_0 $	0.59 ± 0.01
Decreasing $ z_0 $	0.48 ± 0.02
Increasing $ 1/p $	0.59 ± 0.01
Decreasing $ 1/p $	0.61 ± 0.02
Increasing r_0	0.58 ± 0.01
Decreasing r_0	0.50 ± 0.02
Increasing t_1	0.442 ± 0.006
Decreasing t_1	0.367 ± 0.008

Table 7.4: Performance of the baseline model with modified track ordering. The performance metric is the mean absolute error on the validation set.

lines:

$$\vec{L}_1 = t_1 \vec{d}_1 \quad (7.5)$$

$$\vec{L}_2 = \vec{r}_p + t_2 \vec{d}_2 \quad (7.6)$$

Where $\vec{L}_1$ is the jet axis and $\vec{d}_1$ is its unit direction vector, $\vec{L}_2$ is the equation of the track, $\vec{r}_p$ is the perigee position of the track and $\vec{d}_2$ is its unit direction vector. The distance along the jet axis can be written as:

$$t_1 = \frac{\vec{r}_p \cdot (\vec{d}_1 - (\vec{d}_1 \cdot \vec{d}_2) \vec{d}_2)}{1 - (\vec{d}_1 \cdot \vec{d}_2)^2} \quad (7.7)$$

where we have made use of the fact $|\vec{d}_1|^2 = |\vec{d}_2|^2 = 1$. The jet axis has been fixed to $\vec{d}_1 = \frac{1}{2}(1, 1, \sqrt{2})$ for all experiments⁹. This choice is arbitrary, but the model is spherically symmetric so will not change the results.

A significant improvement on the baseline can be seen for all orderings. The best ordering was decreasing values of t_1 . This ordering will approximately show the RNN tracks from the tertiary vertex first, then secondary and finish with primary vertex tracks. It is of note that ordering in this way is more successful than the reverse order, where primary tracks enter first.

⁹In polar coordinates $\theta = \phi = \frac{\pi}{4}$

Features	Ordering	Validation Loss (MAE) /mm
Baseline	Random	0.67 ± 0.01
Baseline	Decreasing t_1	0.367 ± 0.008
$+(x_p, y_p)$	Random	0.424 ± 0.009
$+(x_p, y_p)$	Decreasing t_1	0.299 ± 0.003

Table 7.5: Performance of the baseline model with and without the addition of two new features (x_p, y_p) . Two track orderings were investigated. The performance metric is the mean absolute error on the validation set.

7.4.2.5 Feature Selection

The five track parameters chosen in the baseline are not the only features available. The perigee point (x_p, y_p, z_p) can also be used (note that $z_p \equiv z_0$ and $d_0 = \sqrt{x_p^2 + y_p^2}$). The track parameters (ϕ, θ, d_0, z_0) alone do not uniquely determine a line in 3D, in ATLAS a reference position must also be defined (the perigee). Hence it was hypothesised that adding in (x_p, y_p) could aid the RNN performance. The results of this experiment are shown in table 7.5 using a random ordering and the best ordering from section 7.4.2.4. As was hypothesized, the additional track features help to improve performance from a MAE of 0.67 mm to 0.42 mm for the baseline model.

7.4.2.6 Effect of Noise

The effect of noise is an important aspect of any observational model. The baseline model added an absolute uncertainty $\sigma = 10^{-5}$ m/rad to the parameters (d_0, z_0, ϕ, θ) and a 0.1% error to (q/p) . This value ensures we are always equal or below the ATLAS design noise. The choice of absolute versus relative uncertainties is motivated by the ATLAS design detector resolution. The extra features (x_p, y_p) were then derived from the noised (d_0, z_0) assuming no angular measurement error. This noising procedure is based on the ATLAS measurement uncertainty see table 3.2 in section 3.2.2. The ATLAS uncertainties include a p_T dependent resolution term which for simplicity was ignored here. The baseline model did not use the exact values from table 3.2, instead it slightly underestimated the noise. The noise level in the baseline model is compared to the design resolution of ATLAS in table 7.6.

Parameter	ATLAS (design)	Baseline	High Noise
d_0	$10\mu\text{m}$	$10\mu\text{m}$	$100\mu\text{m}$
z_0	$91\mu\text{m}$	$10\mu\text{m}$	$100\mu\text{m}$
ϕ	$70\mu\text{rad}$	$10\mu\text{rad}$	$100\mu\text{rad}$
θ	$1000\mu\text{rad}$	$10\mu\text{rad}$	$1000\mu\text{rad}$
q/p	1%	0.1%	1%

Table 7.6: The resolution of track parameters in our toy jets for two different levels of noise (baseline and high noise) compared to the ATLAS design specifications. The values for the latter are calculated from [44] assuming tracks perpendicular to the beamline, with the exception of the θ resolution which is taken from [118]. The high noise model thus has a larger noise than would be expected at ATLAS. For simplicity, the absolute error for the baseline was set to $10\mu\text{m}$ for all parameters (thus all parameters are less or equal than ATLAS design noise).

	Baseline Noise	High Noise
Best model (20k)	$0.299 \pm 0.003 \text{ mm}$	$0.53 \pm 0.02 \text{ mm}$
Best model (280k)	$0.119 \pm 0.001 \text{ mm}$	$0.283 \pm 0.003 \text{ mm}$

Table 7.7: The performance as measured by validation loss of the best RNN model trained on a sample of jets with the baseline noise or high noise.

To investigate the effect of noise, a high noise training sample of toy jets was produced using the values in table 7.6. This was done by multiplying the standard deviation of noise added by a factor of 10, with the exception of the theta parameter which was increased by a factor of 100 to be greater than the ATLAS design resolution. The best RNN model was then trained on these and the results are quoted in table 7.7.

7.4.3 Best Model

The current best model is determined from the experiments above. The tracks are ordered in decreasing value of t_1 (the closest point on the jet axis to the track). Two extra features, the x_p and y_p perigee coordinates, are used. The scalings from section 7.4.2.3 are used. This best model is retrained using 280,000 training samples and achieves a validation MAE of $0.119 \pm 0.001 \text{ mm}$, compared to the baseline model's $0.169 \pm 0.002 \text{ mm}$. Using the high noise dataset the best model achieved $0.283 \pm 0.003 \text{ mm}$ compared to the baseline which achieved $0.356 \pm 0.003 \text{ mm}$.

7.5 Performance metrics

The MAE evaluated on the validation set was the main metric used to determine the best performing model. More performance metrics, calculated on a held out test set, are studied here. These metrics shed light on the factors affecting individual jet performance and how this algorithm performs in the context of ATLAS. A key goal is a metric that allows comparison of the RNN model with existing ATLAS algorithm.

ATLAS b-tagging algorithms such as SV1 or JetFitter, are not evaluated by the exact position of reconstructed vertices. This reflects their usage: They produce output variables for the high-level taggers, which in turn discriminate b-jets from other flavour jets. These output variables are calculated from reconstructed secondary vertices and their associated tracks. They rely on the vertices being an accurate representation of some true physical decay. Determining whether a reconstructed vertex represents a “true” decay vertex is more important than knowing its exact position. ATLAS algorithms determine this using the tracks associated to a vertex. In JetFitter, the χ^2 compatibility of reconstructed vertices to the associated tracks determines the vertex quality. Furthermore, the truth identity of the tracks can be used to determine if a vertex is accurate. In the RNN method, however, all the tracks are used to fit both vertices. It is difficult to identify which individual track the RNN used to fit which vertex.

The vertex reconstruction performance is investigated in three ways. First, some simple quantities are derived using the true vertex and reconstructed vertex positions in section 7.5.1. Secondly, a vertex identification measure from error propagation of the resolution uncertainties is investigated in section 7.5.2. Finally, some studies on the track association problem are shown in section 7.5.3. For the following, the best model trained on 280,000 training samples is used. The results quoted are evaluated on the held out 20,000 test samples.

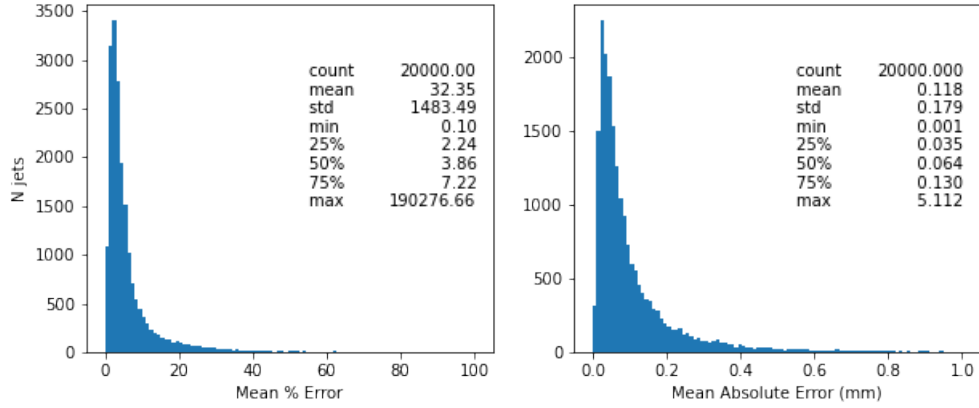


Figure 7.8: The mean absolute and percentage errors across the six output values of each jet.

7.5.1 Individual Jet Performance

The RNN performance can vary significantly for individual jets. The distributions of the mean absolute and mean percentage error¹⁰ per jet are shown in figure 7.8. The distributions have very long tails, in fact some values lie outside the plotted ranges (for instance one jet had a MAE of around 5.11 mm). The mean and median of these distributions are significantly different. The mean of the mean absolute error per jet is 0.118 mm (this is equivalent to the loss function value calculated on the test set) whilst the median value is around 0.064 mm. The mean of the mean percentage error per jet is around 32.35%, whilst the median is 3.86%.

The (L2) distance between the true and reconstructed secondary/tertiary vertex positions can be used as a measure of the vertexing performance¹¹. This is termed the *secondary/tertiary vertex error*. Figure 7.9 shows the distributions of these values in the test set. It is interesting to note the median tertiary vertex error (135 μm) is slightly higher than the secondary vertex error (76 μm). If, however, we consider the vertex error as a fraction of the true vertex displacement then the reverse trend is seen (this is termed the percentage vertex error). The median percentage secondary

¹⁰This is the sum of the absolute or percentage error of each output coordinate. The MAE can be thought of as the secondary L1 error plus the tertiary L1 error. A vertex L1 error is the L1 distance between its true and reconstructed position. The percentage error is taken as a fraction of the absolute (L1) displacement from 0.

¹¹Contrast this to the mean squared error loss function, which was summed over all 6 coordinates independently. The L2 distance is calculated for the secondary and tertiary vertices separately.

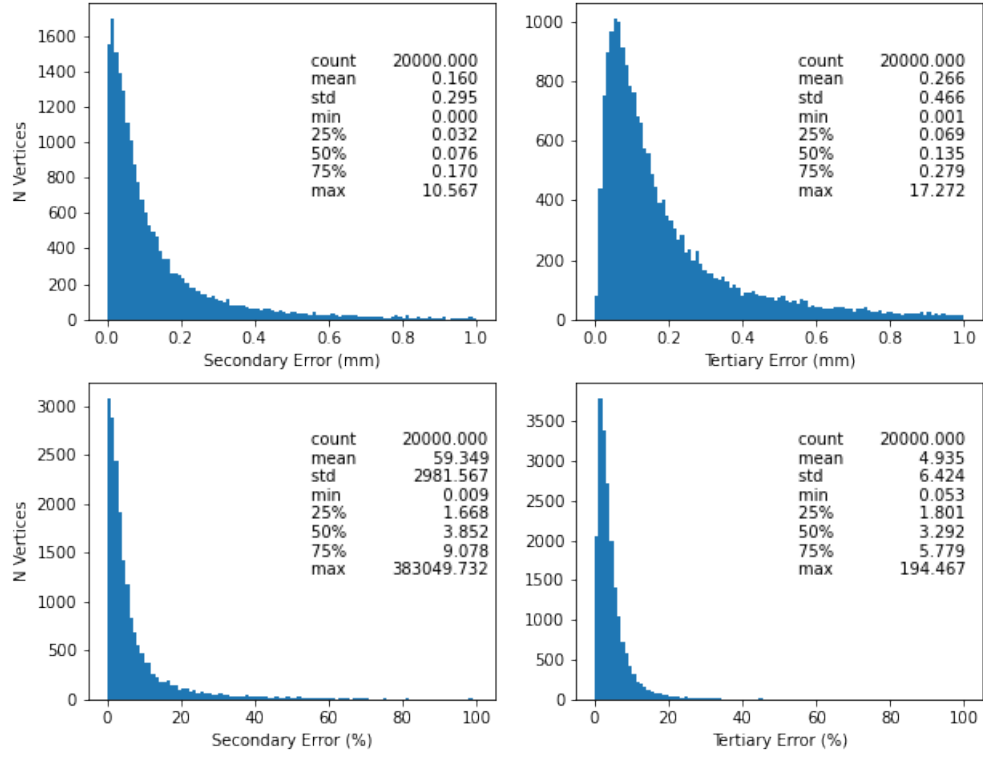


Figure 7.9: The secondary and tertiary vertex errors as measured by the distance between true and reconstructed vertices. The percentage vertex error is vertex error divided by the true vertex displacement.

vertex error is 3.9% and the median for the tertiary vertex is 3.3%.

These measures do not always give a reasonable estimate of jet performance. The percentage vertex errors can blow up when the absolute vertex displacement is small. One jet had a percentage secondary vertex error of 383049%. This particular jet had a true secondary vertex at a position $(3.8 \times 10^{-5}, 3.8 \times 10^{-5}, 5.4 \times 10^{-5})$ mm, a displacement of only $0.076 \mu\text{m}$ which is much less than our detector resolution. This jet is drawn in figure 7.10. Its absolute secondary vertex error is not much worse than other jets at around $290 \mu\text{m}$ (its tertiary error was $267 \mu\text{m}$). The percentage vertex error is clearly not a useful quantity at such scales. The jet in figure 7.11, had secondary and tertiary vertex percentage errors of less than 0.5% (absolute secondary error was $90 \mu\text{m}$ and tertiary was $11 \mu\text{m}$). Considerations of

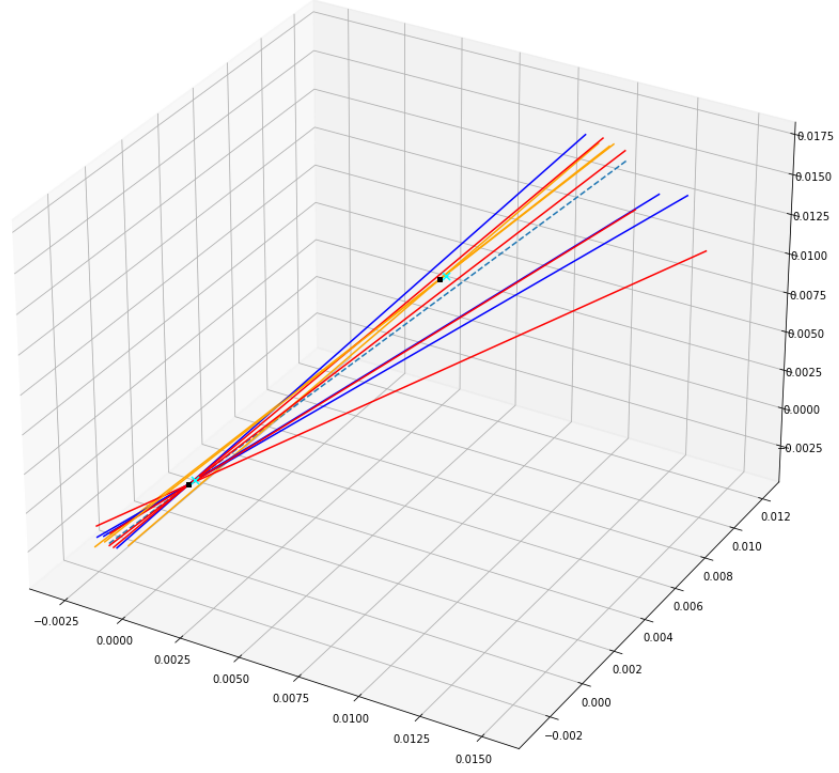


Figure 7.10: An image of a jet showing the reconstructed vertices (cyan crosses), true vertices (black squares) and primary (red), secondary (blue) and tertiary (orange) tracks after noise was added. This jet had large vertex errors. The true secondary and primary vertices cannot be distinguished at this scale

absolute or percentage error alone do not help determine the RNN performance.

7.5.2 Vertex Identification

The previous section covered simple measures of the errors on individual vertex reconstruction in a jet. Performance varied depending on the individual jet candidate. This is somewhat expected, the resolution uncertainties are randomly distributed as are the vertex displacements. In this section, a vertex identification is derived to

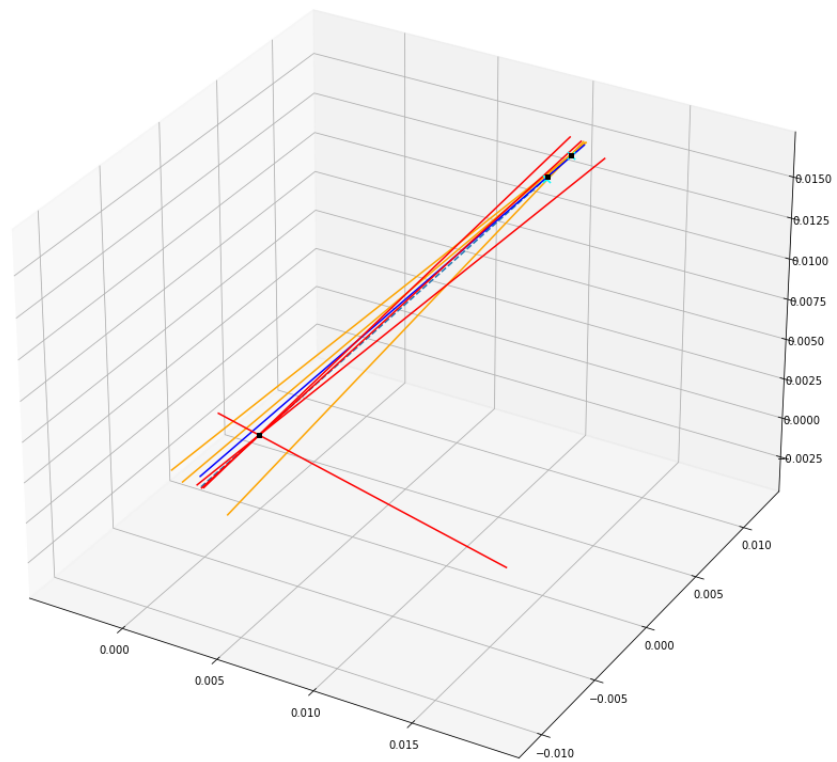


Figure 7.11: An image of a jet showing the reconstructed vertices (cyan crosses), true vertices (black squares) and primary (red), secondary (blue) and tertiary (orange) tracks after noise was added. This jet had small vertex errors.

account for these effects.

The secondary and tertiary vertex errors were introduced in section 7.5.1. A simple vertex identification method would be to place a threshold value on this error metric. If the error is less than this value, the vertex has been correctly identified. In principle, the vertex error must be due to the noise on the data. In the noiseless scenario, it is expected that all vertices could be exactly identified. This could in fact be done analytically, it reduces to finding the intersection of two lines in 3D. When noise is added, the tracks will generally no longer intersect the jet axis. The best analytical guess at a vertex position would be the point closest to all tracks belonging to that vertex. The error on this guess will vary depending on the noise added, this could be calculated using error propagation. This value can then be used as the threshold.

Consider the case of a single track and the jet axis. The two can be represented by equations (7.5) and (7.6). The shortest distance between the two skew lines is given by:

$$d_{min} = \left| \frac{\vec{r}_p \cdot (\vec{d}_1 \times \vec{d}_2)}{|\vec{d}_1 \times \vec{d}_2|} \right| \quad (7.8)$$

The perigee and direction vectors are:

$$\vec{r}_p = (d_0 \cos \phi_p, d_0 \sin \phi_p, z_0) \quad (7.9)$$

$$\vec{d}_1 = \frac{1}{2}(1, 1, \sqrt{2}) \quad (7.10)$$

$$\vec{d}_2 = (\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta) \quad (7.11)$$

where it is assumed that ϕ_p , the azimuthal angle of the perigee position, and the jet axis are constant. The shortest distance can thus be written as a function of four of the five track parameters: (d_0, z_0, ϕ, θ) :

$$d_{min} = |A(\phi, \theta)d_0 + B(\phi, \theta)z_0| \quad (7.12)$$

where:

$$A(\phi, \theta) = K(\phi, \theta)(\cos \phi_p(\cot \theta - \sqrt{2} \sin \phi) - \sin \phi_p(\cot \theta - \sqrt{2} \cos \phi)) \quad (7.13)$$

$$B(\phi, \theta) = K(\phi, \theta)(\sin \phi - \cos \phi) \quad (7.14)$$

$$K(\phi, \theta) = \frac{1}{\sqrt{(\cot \theta - \sqrt{2} \sin \phi)^2 + (\cot \theta - \sqrt{2} \cos \phi)^2 + (\sin \phi - \cos \phi)^2}} \quad (7.15)$$

This is a highly non-linear function. The error propagation formula for these four parameters is (assuming independent variables):

$$\sigma_{d_{min}}^2 = \sigma_{d_0}^2 \left(\frac{\partial d_{min}}{\partial d_0} \right)^2 + \sigma_{z_0}^2 \left(\frac{\partial d_{min}}{\partial z_0} \right)^2 + \sigma_{\phi}^2 \left(\frac{\partial d_{min}}{\partial \phi} \right)^2 + \sigma_{\theta}^2 \left(\frac{\partial d_{min}}{\partial \theta} \right)^2 \quad (7.16)$$

$$= \sigma_{d_0}^2 A^2 + \sigma_{z_0}^2 B^2 + \sigma_{\phi}^2 \left(d_0 \frac{\partial A}{\partial \phi} + z_0 \frac{\partial B}{\partial \phi} \right)^2 + \sigma_{\theta}^2 \left(d_0 \frac{\partial A}{\partial \theta} + z_0 \frac{\partial B}{\partial \theta} \right)^2 \quad (7.17)$$

This can be written in an analytical form. It is therefore possible to solve exactly for the one track case. Some special note should be taken of the partial derivative terms. These contain divergences in the domain $\theta \in [0, \pi]$, $\phi \in [0, 2\pi]$, $\phi_q \in [0, 2\pi]$. Intuitively, these relate to the case where the track and jet axis are almost parallel (or anti-parallel), e.g. at $\phi = \theta = \pi/4$. A small change in theta or phi can lead to a large change in d_{min} . This has significant consequences for highly boosted jets, where the tracks tend to align with the jet axis.

A similar analysis can be used to calculate the error for a vertex with multiple tracks. However, this requires prior knowledge of which tracks are used to fit a vertex. The RNN does not easily provide this information, see section 7.5.3. We instead fix a simple threshold value using the single track case. We could consider each term in equation 7.17 individually and take the maximum over the domain $\theta \in [0, \pi]$, $\phi \in [0, 2\pi]$, $\phi_q \in [0, 2\pi]$. This is explored in appendix C. Alternatively, we can just derive the values of $\sigma_{d_{min}}$ for the tracks in our toy jet samples.

The toy jet samples are used to calculate $\sigma_{d_{min}}$ for each secondary and tertiary track. This is done in the high and baseline noise samples separately. A threshold error is then calculated using the sample maximum or median values, termed $\sigma_{d_{min}}^{max}$ and $\sigma_{d_{min}}^{median}$. The results are presented in table 7.8 for the two noise regimes:

Threshold	Baseline	High Noise
$\sigma_{d_{min}}^{max}$	110 μm	2470 μm
$\sigma_{d_{min}}^{median}$	5.80 μm	58.0 μm

Table 7.8: The threshold errors calculated from toy jet samples for the high noise and baseline noise cases.

	$\sigma = \sigma_{d_{min}}^{max}$			$\sigma = \sigma_{d_{min}}^{median}$		
	n=1	n=2	n=5	n=1	n=2	n=5
$f_{vID} @ n\sigma$	55.0%	77.2%	93.7%	2.39%	5.51%	15.4%
$f_{j,2v} @ n\sigma$	34.6%	63.4%	88.9%	0.00%	0.145%	2.18%
$f_{j,1v} @ n\sigma$	40.7%	27.5%	9.46%	4.78%	10.8%	26.5%
$f_{j,0v} @ n\sigma$	24.7%	9.04%	1.62%	95.2%	89.1%	71.4%

Table 7.9: Performance of the RNN tagger for the baseline noise model using the derived performance metric. A vertex is considered identified if it lies within a certain threshold distance of a true vertex, given by $\sigma_{d_{min}}^{max}$ or $\sigma_{d_{min}}^{max/median}$. This threshold can be loosened to some integer multiple n . f_{vID} is the total fraction of vertices identified and $f_{j,kv}$ the fraction of jets with k identified vertices.

The thresholds in table 7.8 are now used to determine whether a vertex is correctly identified. Consider the displacement between true and reconstructed (secondary/tertiary) vertices d_{vtx} . If $d_{vtx} < n\sigma_{d_{min}}^{max/median}$ then the vertex is considered to have been correctly identified, where n is some integer (for instance $n = 2$ corresponds to the p-value of 0.05). By this metric, the fraction of correctly identified secondary and tertiary vertices and the number of jets with 0, 1 or 2 correctly identified secondary vertices is shown in table 7.9 for the baseline noise model. The same quantities are calculated for the high noise model in table 7.10.

	$\sigma = \sigma_{d_{min}}^{max}$			$\sigma = \sigma_{d_{min}}^{median}$		
	n=1	n=2	n=5	n=1	n=2	n=5
$f_{vID} @ n\sigma$	99.8%	100.0%	100.0%	32.4%	57.0%	83.7%
$f_{j,2v} @ n\sigma$	99.6%	99.9%	100.0%	12.5%	36.9%	73.1%
$f_{j,1v} @ n\sigma$	0.35%	0.07%	0.005%	39.7%	40.0%	21.3%
$f_{j,0v} @ n\sigma$	0.05%	0.015%	0%	47.8%	23.0%	5.66%

Table 7.10: Performance of the RNN tagger for the high noise model using the derived performance metric. A vertex is identified if it lies within a certain threshold distance of a true vertex, given by $\sigma_{d_{min}}^{max}$ or $\sigma_{d_{min}}^{max/median}$. This threshold can be loosened to some integer multiple n . f_{vID} is the total fraction of vertices identified and $f_{j,kv}$ the fraction of jets with k identified vertices.

The RNN performance at high noise appears to be better than the low noise (baseline), using both maximum and median methods. This occurs because the threshold values, $\sigma_{d_{min}}$, increase with the detector resolution uncertainties. We are, in effect, comparing our performance to a simple algorithm that predicts the closest point between the track and the jet axis. In the baseline noise case, the performance is poor by this metric. This suggests that the detector resolution is not the limiting factor here. The performance at high noise is, however, quite good. This is likely due to RNN making use of extra information in the dataset, rather than simply finding the closest point to all tracks.

This vertex identification metric is clearly imperfect. It firstly assumes a single track and the jet axis. It is also estimated from the sample distribution, not on an individual jet basis. Some interesting observations were nevertheless drawn.

7.5.3 Impact of Individual Tracks

An important aspect of ATLAS vertexing is the association of tracks to a vertex. This is used to assess the validity of the vertex, and further to calculate output variables for high-level taggers. Associating tracks to vertices is not straightforward with the RNN. The RNN uses all of the tracks to fit the vertex positions, the origin of a single track is unknown. To this end, the impact of individual tracks on the RNN prediction was studied.

If an individual track belongs to the secondary or tertiary vertex, removing it would increase the reconstruction error of that vertex. As such the RNN was run on the same jet repeatedly removing one track at a time. The impact of the removed track on the absolute and percentage error of the vertices (as measured by the distance between reconstructed and true) was plotted. The plot for the good jet from figure 7.11 is shown in figure 7.12.

From inspection of these plots, it is difficult to identify which tracks belong to the secondary or tertiary vertex. The truth information from the jet tells us the first 3 tracks are tertiary tracks, the fourth is a secondary track, and the rest are primary tracks. It is perhaps possible to identify the primary vertex tracks, these are the ones with the lowest impact and smallest value of t_1 (remember these tracks are ordered

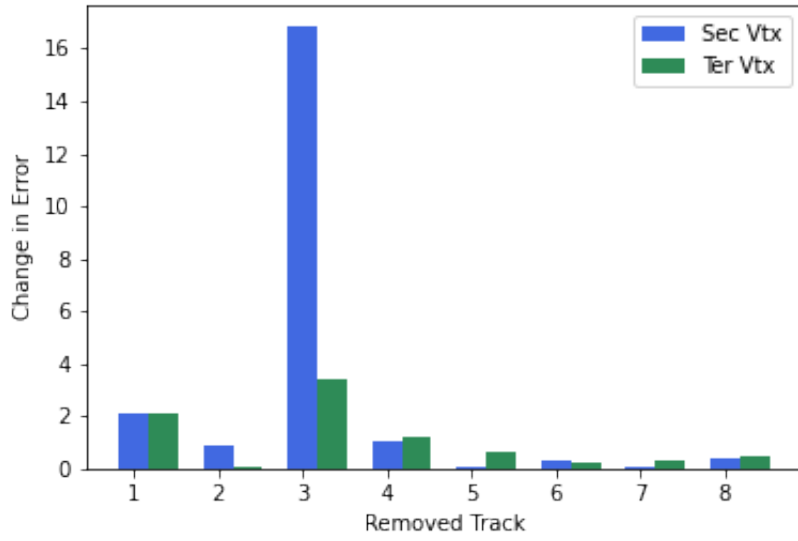


Figure 7.12: The change in the secondary and tertiary vertex error (in mm) when a track is removed from the jet. The removed track is indexed by its position in the jet after ordering by decreasing t_1 , e.g. removed track 1 has the highest value t_1 .

by decreasing t_1). However, some of the secondary/tertiary vertex tracks also have a fairly small impact. The 3rd track in this jet seems to be of particular importance for both vertices, despite this being a tertiary vertex track it greatly impacts the secondary vertex prediction. These plots show that information from all the tracks is being used by the RNN in vertex fitting, regardless of their origin.

A second measure of the track's impact is to plot the evolution of the vertex prediction as each track is added into the RNN. This was done by running the RNN multiple times, first on a null set, then with the first track added in, then the first and second tracks, and so on until all tracks are back in. The error is plotted as a function of the number of tracks added back. The plot for the same good jet as above is shown in figure 7.13.

It is again difficult to determine which tracks belong to the secondary or tertiary vertex. Some insight into the RNN working may be drawn as we know that tracks 1-3 belong to the tertiary vertex, track 4 is the secondary vertex track and the rest are primary tracks. The error only significantly drops when tracks from different vertices are present. The RNN seems to use the differences between tertiary,

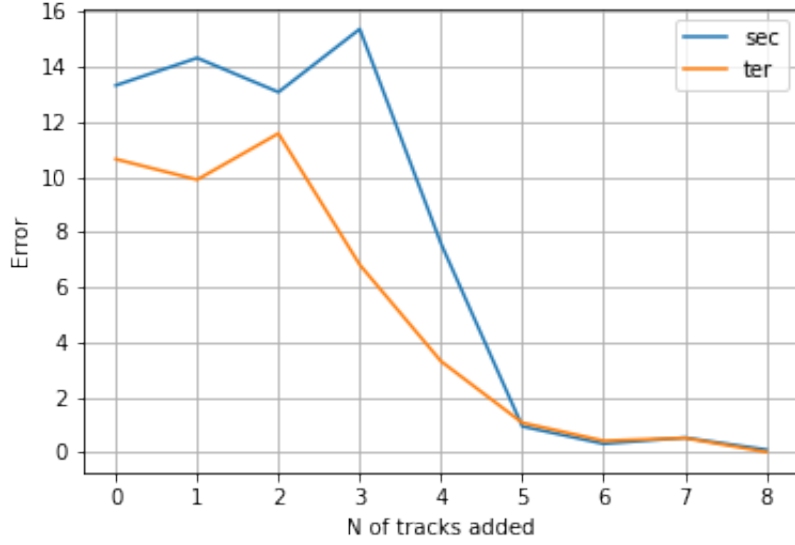


Figure 7.13: The evolution of the error in the secondary and tertiary vertex prediction as more tracks are seen. The value at N of tracks added = 0 is calculated on a jet with all null tracks. The value at $N=k$ is calculated on a jet with the first k tracks (ordered by decreasing t_1) only, the rest are null.

secondary and primary tracks in its calculation.

7.6 Outlook

An RNN algorithm was implemented to predict the positions of secondary and tertiary vertices in b-jets. The RNN performance was qualitatively inspected using jet drawings such as figures 7.11 and 7.10. A quantitative measure of performance was attempted using error propagation methods to set a threshold for the expected vertex error, the results of this analysis are in section 7.5.2. There remain several key disadvantages to the RNN method. It is not immediately comparable to the JetFitter algorithm as it does not associate tracks to vertices. Evaluating its performance is not straightforward as no uncertainty on the predicted values is given by the RNN. The RNN based vertex fitting method is therefore likely not a direct replacement for the current Kalman Filter based algorithms. Nevertheless, some interesting observations can be made.

Difficulties arose in evaluating performance. The vertex position errors (the secondary and tertiary vertex errors) were analysed in section 7.5.1, however these

were not very informative as they do not account for the noise added to the tracks. In section 7.5.2, an error propagation method was used to determine the reconstruction performance accounting for this noise. Multiple threshold values for vertex error, if reconstructed vertex had an error below this threshold it was considered to be correctly identified. The results vary significantly depending on which threshold value is used. But all of them shows that the RNN is robust to noise: the fraction of vertices correctly identified is higher in the high noise sample than the low noise case with every single threshold metric. The threshold metrics predicted from error propagation all grow faster with noise than the actual errors in the vertex predictions of the RNN. There is no clear choice as to which metric to use, as the method makes several assumptions and simplifications in determining them. Finding a better performance metric would be key to comparing the RNN to current vertex fitting methods.

Secondary vertex fitting at ATLAS is not generally interested in the exact position of the vertex, but rather whether the vertex is viable or not. In b-tagging, the presence of a secondary vertex is the main discriminating condition, not necessarily its exact position. To assess this, tracks are associated to the secondary vertex in the ATLAS algorithms such as JetFitter. The validity of this association determines the validity of the vertex. Track association with the RNN was explored in section 7.5.3. The RNN method does not currently offer a simple method of directly associating tracks to a vertex, so no direct comparison is possible. Some further experiments could be carried out to achieve this association. One method could be to calculate some measure of the compatibility between tracks and a vertex predicted by the RNN. This could use a χ^2 method, analogous to the track-vertex compatibility calculations in JetFitter (see section 6.4.2). But this is not straightforward as the RNN does not return a covariance matrix for the fit vertices. Other methods based solely on the RNN might also be explored, such as Shapley values [119].

Several experiments were performed to optimise the RNN tagger performance (as measured by validation loss). There remain several more that could be tried. Firstly, an extensive hyperparameter tuning was not carried out. Another experiment

could be to try adding the tracks in reverse order to the end of the sequence, this is modelled on the Kalman filter's smoothing procedure, which steps backwards through the track sequence after its initial fit. By allowing the RNN to see the tracks again, it might perform better. The vertex fitting problem could also be constrained to a 1D problem. This could be done in the same way as JetFitter, by constraining the vertices to lie on the jet axis. The vertex distance along this axis is then fit. The RNN method could also be combined with a Kalman Filter method, this might allow for additional use of the domain knowledge provided by the Kalman Filter. This might be done in several ways, for instance using KF-LSTM cells such as in [114], or by using the RNN to initialise the decay topology before passing to the RNN. Finally, experiments using all different jet flavours were not performed. It would be of interest to see how the RNN works in this case, and whether it offers good discrimination between b, c and light-jets.

Chapter 8

Analysis

8.1 Physics Motivation

The SM faces several challenges which motivate research into new physics. Issues such as the existence of dark matter and energy motivate the search for new particles and interactions. The current Higgs sector suffers from the hierarchy problem and excessive fine-tuning. Much of ATLAS analysis work is therefore dedicated to finding new physics.

Heavy versions of the Higgs boson are predicted in many BSM models, such as supersymmetric models. A generic heavy Higgs (GHH) model was covered in section 2.2.2.2. The dim-4 and 6 couplings to the vector bosons were derived in equations (2.26) - (2.29) and (2.35) - (2.36). The couplings to fermions were not covered, channels involving these are not searched for in this work. This is motivated by the lack of any observations in searches requiring gluon-gluon fusion (ggF) production¹ of the GHH, for example in [120, 121, 122, 123]. As such, any generic heavy Higgs is assumed to have small couplings to the fermion sector.

The SM Higgs is produced at the LHC predominantly through ggF. For the fermiphobic heavy Higgs under investigation this process is disfavoured. The next two most prominent production mechanisms are vector-boson fusion (VBF) and associated vector boson Higgs production (VH) [30]. Of the two, the cross section of VBF is about an order of magnitude higher if only dim-4 terms are considered.

¹This involves a fermion loop, and thus the GHH-fermion coupling.

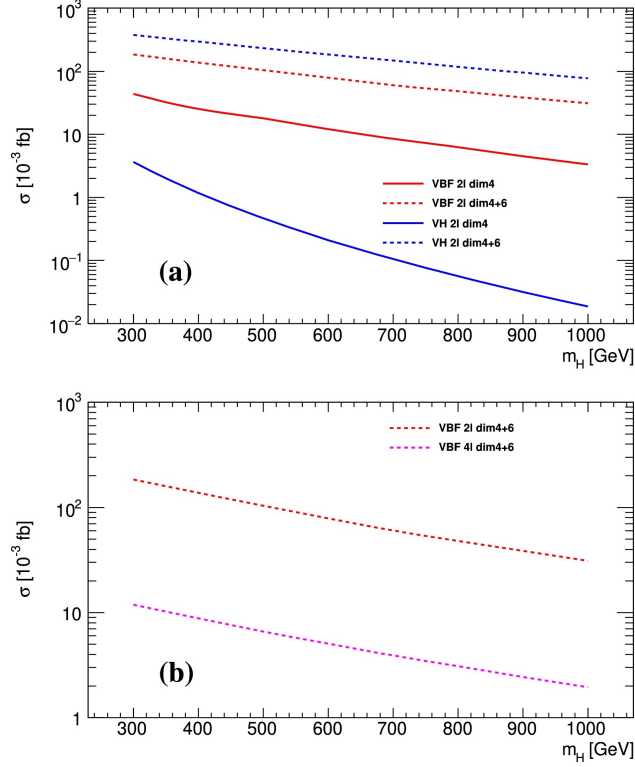


Figure 8.1: The leading order cross sections of different production methods and final states as a function of heavy Higgs mass. Two models are compared, dim4 has $\rho_H = 0.05$, $f_W = f_{WW} = 0$ and dim4+6 has $\rho_H = 0.05$, $f_W = f_{WW} = 50$. (a) shows the cross-sections for the VBF and VH 2ℓ processes ($pp \rightarrow Hjj \rightarrow \ell\ell jjjj$ and $pp \rightarrow VH \rightarrow \ell\ell jjjj$). (b) compares the VBF 2ℓ process and the VBF 4ℓ process ($pp \rightarrow Hjj \rightarrow 4\ell + jj$) [31].

When dim-6 operators are included, the VH and VBF production cross-sections are both enhanced. VH becomes the predominant mechanism, as shown in figure 8.1.

Final states containing leptons are of particular interest to suppress SM backgrounds. Single lepton channels, studied in more detail in [124] were not considered as they have low sensitivity due to large SM backgrounds (W +jets and $t\bar{t}$) [31]. This analysis instead requires two or more leptons. For a single Higgs produced by VBF, the two decay modes with best sensitivity are: $H \rightarrow ZZ \rightarrow \ell\ell jj$ and $H \rightarrow ZZ \rightarrow 4\ell$. The first of these has large SM backgrounds and the second does not have a large enough yield in the high momentum region to be detected [31]. For these reasons, VBF production mechanisms were not considered. This is further justified if dim-6 operators are present, as VH will have a large cross-section.

The VH final states, in the fermiphobic assumption, follow from the combi-

nations of the three vector bosons: $W^\pm ZZ$, $W^\pm W^\pm W^\mp$, $ZW^\pm W^\mp$ and ZZZ , where the last two bosons are produced by the Higgs. The W and Z can decay hadronically (reconstructed as two jets), or leptonically². Final states with τ leptons are not searched for as these decay rapidly, and do not offer a clean signal in the ATLAS detector. The key branching ratios are [5]:

$$\text{Br}(W \rightarrow e\nu_e) \approx \text{Br}(W \rightarrow \mu\nu_\mu) \approx 10\% \quad (8.1)$$

$$\text{Br}(W \rightarrow \text{hadrons}) \approx 70\% \quad (8.2)$$

$$\text{Br}(Z \rightarrow e^+e^-) \approx \text{Br}(Z \rightarrow \mu^+\mu^-) \approx 3.5\% \quad (8.3)$$

$$\text{Br}(Z \rightarrow \text{hadrons}) \approx 65\% \quad (8.4)$$

Final states with more than one leptonically decaying Z will have low signal yields so are not considered. The $WH \rightarrow W^\pm W^\pm W^\mp \rightarrow 3\ell 3\nu$ final state, with three semi-leptonic W decays, was similarly not considered. The remaining VH final states, containing at least two leptons, were studied in detail in [31]:

- $VH \rightarrow \ell^+\ell^- jjjj$: The OS2 ℓ (opposite-sign) channel. The two leptons originate from a Z decay and have the same flavour.
- $VH \rightarrow \ell^\pm \nu \ell^+ \ell^- jj$: The 3 ℓ channel. One of the lepton pairs originated in a Z decay.
- $VH \rightarrow \ell^\pm \nu \ell^\pm \nu jj$: The SS2 ℓ (same-sign) channel. This is formed from the triboson production process $WH \rightarrow W^\pm W^\pm W^\mp$.

The expected signal and background yields with 300 fb^{-1} of LHC data in the signal regions of these channels are quoted in table 8.1. The SS2 ℓ channel is the most sensitive due to the SM background process being very rare in proton proton collisions [31]. The Feynman diagram of this process is shown in figure 8.2.

²Ignoring the $Z \rightarrow \nu\nu$ decay.

	Signal	Z+QCD jets	Other
OS2 ℓ channel	4.0	41.6	3.5
3 ℓ channel	3.6	7.2	1.3
SS2 ℓ channel	11.4	4.1	3.1

Table 8.1: Expected signal and background yields in the signal regions of the three studied channels with 300 fb^{-1} [31].

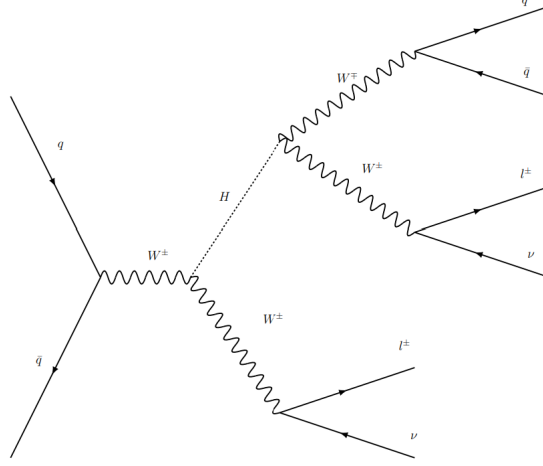


Figure 8.2: Feynman diagram of the SS2 ℓ process for a generic heavy Higgs.

8.2 Analysis Overview

8.2.1 Strategy

The final state studied has two same-sign leptons, two neutrinos and a hadronically decaying W . The H mass cannot be reconstructed due to the neutrinos. An effective mass observable, M_{eff} , is instead used. This is defined in equations (8.6) and (8.7). MC simulations are used to create samples for the studied signals and all relevant backgrounds. These are summarized in section 8.3.

To cover as much kinematic space as possible, the signal will fall into two separate categories. These are defined by the properties of the hadronically decaying W . In the low momentum region, the products of this boson can be resolved into two small-R jets, giving the resolved category. In the high momentum region, or boosted category, a single large-R jet is instead reconstructed. The two categories are orthogonal, if the leading large-R jet in an event can successfully pass object selections this event can only enter the boosted category. The two same-sign leptons

can be muons or electrons leading to four sub-channels termed ee , $e\mu$, μe and $\mu\mu$. The lepton listed first has the larger p_T , and is termed the leading lepton.

The basic object definitions used are summarized in section 8.4. The final signal regions were selected using an optimization procedure summarized in section 8.5. The main backgrounds of this analysis are described in section 8.6. The backgrounds are estimated using a variety of data-driven techniques and through the use of the simulated background samples. Control regions are defined to fit the normalization of key simulated background processes to data. A validation region is used to evaluate the background estimations as described in section 8.7. Systematic uncertainties and their treatment are covered in section 8.8. A statistical framework is used to test for the presence of a signal. This is described in section 8.9 and the results are presented in section 8.10.

8.2.2 Personal Contributions

The analysis work presented here is a product of the collaborative work of all the members of the GHH analysis team, and builds on the work of the wider ATLAS collaboration. I will therefore briefly bring attention to the parts of the analysis that are my own work.

- Experimental Systematics: I worked on the implementation, validation and documentation of the experimental systematics in various parts of the analysis framework.
- N-tuple production: I worked on the development of the `ntuplereader_ghh` framework and was responsible for ntuple production. I had a particular focus on implementing the systematic variations in this framework. I also helped with the application of the W tagger scale factors.
- Variable Plots: I worked within the plotting framework to add the systematic variations and produce the kinematic plots in analysis regions.
- The CxAOD Framework: I implemented, maintained and tested various aspects of the CxAOD Framework. I worked on various aspects such as: systematic variations, b-tagging (upgrade from MV2 to DL1), implementing the

W tagger (scale factors and systematics), the V gamma overlap removal tool and upgrading the framework to new Analysis Base.

- The Statistical Framework (HistFitter [125]): I worked on the implementation of systematics within our statistical framework, the production of pull and ranking plots to validate systematics and the formatting of these plots. I produced the upper limit results for the final signal grid and created plots to summarize these results. I also produced the exclusion contour plots. In addition to these responsibilities, I helped with the testing and running of the framework.
- Documentation: I was part of the team writing the internal note. I am the main contributor to the introduction and experimental systematics sections and as a native speaker helped edit the other sections.

8.3 Data and Simulated Samples

The analysis is performed on data samples gathered at the ATLAS detector during run 2 of the LHC (2015-2018). The data was collected during stable run conditions. The total integrated luminosity was 139fb^{-1} . Simulated signal and background samples are also used, these are described in more detail in the following sections. The ATLAS derivations framework is used to produce DxAOD format ntuples for the analysis, starting from the data or simulation xAODs. This is done by the ATLAS Standard Model group, more details can be found in chapter 4. These DxAODs are then further processed using an analysis specific version of the CxAODFramework. This gives the final ntuples used by the analysis. More detail on the analysis CxAODFramework is given in appendix D.

8.3.1 Simulated Signal Samples

The analysis considers the generic heavy Higgs model described in section 2.2.2.2. A range of parameters for the Higgs mass, $m_H \in [300, 2000]$ GeV, and coupling parameters $f_W \in [-2510, 2480]$ and $f_{WW} \in [-15000, 15000]$ are explored. These ranges are selected to span the expected exclusion limits determined from simula-

tion. No sensitivity is expected at higher mass values. These are defined in equations (2.35)-(2.36). The coupling parameters f_B and f_{BB} are set to 0 to neglect higher order terms. The scaling factors from equations (2.26)-(2.29) are fixed to $\rho_h = 1$ and $\rho_H = 0.05$. This is because the lighter Higgs must correspond to the boson already measured at the LHC, and $\rho_H + \rho_h \approx 1$. The full list of samples can be found in appendix E.

The samples are produced using MADGRAPH5_aMC@NLO [126] and FEYN-RULES [127] interfaced to PYTHIA 8.244 [10] for parton showering. Leading Order (LO) processes only are considered for the hard scatter corrected with K -factors. The K -factor, the ratio of the next-to-leading order (NLO) and LO cross-sections, is calculated at production level using the $pp \rightarrow VH$ process. It is measured to have a value of around 1.3.

8.3.2 Simulated Background Samples

The background samples are created using MC simulations and their detector response modelled using GEANT 4. The background sample types and the generator used to produce them are summarized in table 8.2. In all samples not generated with SHERPA, PYTHIA 8 is used for parton showering. Each background process can require multiple generated samples to be fully described. More detail on the background simulation can be found in the support note for this analysis [128].

8.4 Object Definitions and Selections

The objects reconstructed in the ATLAS detector were described in detail in chapter 4. This section describes the specific object criteria used in the heavy Higgs analysis.

8.4.1 Primary Vertex and pile-up

The primary vertex in an event is defined as the vertex with the largest $\sum p_T^2$ of associated tracks. Events in this analysis must have a primary vertex with at least three associated tracks with $p_T > 0.4$ GeV. Pile-up conditions at the LHC during run 2 differ for data gathered in 2015/16, 2017 and 2018. The MC simulated samples are thus generated three times to match each pile-up profile (these are termed the

Physics process	ME Generator	PS Generator
$WZ \rightarrow \ell\ell\ell\nu$	SHERPA 2.2.2	SHERPA 2.2.2
$WZ \rightarrow \ell\ell\ell\nu jj$	SHERPA 2.2.2	SHERPA 2.2.2
$WZ \rightarrow \ell\nu\nu\nu$	SHERPA 2.2.2	SHERPA 2.2.2
$ZZ \rightarrow \ell\ell\ell\ell$	SHERPA 2.2.2	SHERPA 2.2.2
$ZZ \rightarrow \ell\ell\ell\ell jj$	SHERPA 2.2.2	SHERPA 2.2.2
$t\bar{t}Z \rightarrow \ell\ell + X$	MADGRAPH5_aMC@NLO	PYTHIA 8
$t\bar{t}W \rightarrow \ell\nu + X$	MADGRAPH5_aMC@NLO	PYTHIA 8
tZ	MADGRAPH5_aMC@NLO	PYTHIA 8
$WWZ \rightarrow \ell\ell\ell\ell\nu\nu$	SHERPA 2.2.2	SHERPA 2.2.2
$ZZZ \rightarrow \ell\ell\ell\ell\nu\nu$	SHERPA 2.2.2	SHERPA 2.2.2
same-sign WW (ssWW)	SHERPA 2.2.2	SHERPA 2.2.2
SingleTop	POWHEG-BOX	PYTHIA 8
$WWW \rightarrow \ell\ell + X$	SHERPA 2.2.2	SHERPA 2.2.2
$WH \rightarrow WWW$	POWHEG-BOX	PYTHIA 8
$WWW \rightarrow \ell\ell\ell$	SHERPA 2.2.2	SHERPA 2.2.2
$t\bar{t}$	POWHEG-BOX	PYTHIA 8
$Z + j$	SHERPA 2.2.1	SHERPA 2.2.1
$W + j$	SHERPA 2.2.1	SHERPA 2.2.1
$W\gamma$	SHERPA 2.2.2	SHERPA 2.2.2
$Z\gamma$	SHERPA 2.2.2	SHERPA 2.2.2

Table 8.2: The simulated background physics processes used in the analysis and the Matrix Element (ME) and Parton Shower (PS) generators used for each.

MC16a, MC16d and MC16e campaigns respectively). Pile-up reweighting is used to match the pile-up profile of samples to data if these are different, for instance when the samples were generated before data-taking.

8.4.2 Triggers

ATLAS triggers were described in more detail in section 3.2.5. The triggers used in this analysis are single lepton triggers. Dilepton triggers were considered, but the gain in signal efficiency was small. Instead, a combination of single lepton trigger and a requirement on the leading $p_T > 27$ GeV and sub-leading lepton $p_T > 20$ GeV can sufficiently decrease fake lepton rates. Events must have at least one electron or muon matching a trigger object. The trigger object's p_T must be at least 1 GeV larger than the trigger threshold. This ensures correctly calibrated objects are always used. The trigger threshold negates the need for pre-scaling and is set to 26 GeV. Scale factors are used to correct the trigger efficiency in MC simulated samples to

data. For more detail on the analysis triggers see appendix D.

8.4.3 Electrons

Leptons are important objects in this analysis, as such good quality signal electrons are needed. The signal electron definition is shown in table 4.1 and again in table 8.3. For more detail on the definitions, see section 4.1.4. Other “electron” objects are also defined in this analysis to be used for background estimation, these are the veto, anti-IBL and antiID electrons. Veto electrons are used to veto events with more than two leptons. An event requires two signal leptons and no veto leptons. AntiID electrons are designed to contain more non-prompt electrons. These are used in the estimation of backgrounds from jets reconstructed as electrons. The antiID and signal electron collections are defined to be orthogonal. This is required as these electrons are used to define orthogonal control regions in section 8.6. Anti-IBL electrons are “photon-like” and are used in the estimation of backgrounds from photons mis-reconstructed as electrons. They require no hits in the IBL. This ensures orthogonality with signal and antiID electrons.

Summary of all electrons definitions				
	Signal	Anti-ID	Anti-IBL	Veto
	Reconstructed electron candidate			
p_T	$> 20 \text{ GeV}$			$> 7 \text{ GeV}$
$ \eta $	$ \eta < 1.37 \text{ \& } 1.52 < \eta < 2.47$			< 2.47
Identification	TightLH	MediumLH	TightLH no IBL	LooseLH
$\frac{d_0}{\sigma_{d_0}}$	< 5			—
$ z_0 \times \sin \theta $	$< 0.5 \text{ mm}$			—
Object quality	author == 1, addAmbiguity ≤ 0		—	no BADCLUSTER
Isolation	PLVTight	—	PLVTight	—
ECIDS WP	Loose	Loose	Loose	—
Further		not Signal		

Table 8.3: Summary of all the electron definitions used in this work. AntiID electrons must not pass the signal electron definitions.

8.4.4 Muons

Signal Muons are required to be high quality, these were described in table 4.2. More information on the requirements is found in section 4.1.5. Veto and AntiID

muons are also defined, analogously to the veto and AntiID electrons. All muon definitions are summarized in table 8.4.

Summary of all muon definitions			
	Signal	Anti-ID	Veto
	Reconstructed muon candidate		
p_T	$> 20 \text{ GeV}$		$> 4.5 \text{ GeV}$
$ \eta $	< 2.5		< 2.7
$\frac{d_0}{\sigma_{d_0}}$	< 3	< 10	—
$ z_0 \times \sin \theta $	$< 0.5 \text{ mm}$		—
Object quality	medium		loose
Isolation	PLVTight	!PLVTight	—
Further		not Signal	

Table 8.4: Summary of all the muon definitions used in this work. AntiID muons must not pass the signal muon definitions.

8.4.5 Jets

The analysis strategy looks at boosted and resolved kinematic regions. This requires the use of both small-R and large-R jets. The small-R jets are referred to using the symbol j , whilst large-R jets are represented as J . The jet definitions used are summarized in table 8.5. The b-tagging algorithm DL1r is used with its 85% working point as our signal region is looking for the products of a W boson decay, thus b-jets are vetoed. The W tagger is used on large-R jets. For more detail on the terms used see sections 4.1.2 and 4.1.3.

Summary of all jet definitions		
	small-R jet	large-R jet
Jet collection	AntiKt4EMPFLOW	AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets
p_T	$> 20 \text{ GeV}$	$> 200 \text{ GeV}$
$ \eta $	< 2.5	< 2.0
mass	—	(50, 200 GeV)
W/Z-tagger	—	80% WP
JVT	Medium	—
b-tagging	DL1r @ 85%	—

Table 8.5: Summary of the jet definitions used in this work.

8.4.6 Missing Transverse Momentum

The missing transverse momentum, sometimes referred to as MET or E_T^{miss} , is the negative vectorial sum of all reconstructed objects (see section 4.1.6). The veto muon and leptons defined in tables 8.3 and 8.4 are used in this calculation, alongside the jets defined in table 8.5. The “heavy-flavour” overlap removal algorithm described in section 4.1.7 is applied to these objects.

8.5 Event Selection and Optimization

The $SS2\ell$ channel is characterised by the $VH \rightarrow \ell^\pm \nu \ell^\pm \nu J(jj)$ decay. The signal events are therefore required to have 2 same-sign leptons, hadronic jets (either two reconstructed small-R jets, j , or a single large-R jet J) and E_T^{miss} . The two hadronic jets must be consistent with a W decay, this is done using mass cuts in the resolved case and the boosted W tagger in the boosted case. Additional cuts are used to reduce backgrounds. This includes a b-jet veto to reduce $t\bar{t}$ background and a veto on events with more than two leptons to reduce background from WZ processes. Our signal is expected to have boosted leptons, hence two cuts on the leading lepton p_T and combined lepton invariant mass ($M_{\ell\ell}$) are investigated. These can reduce backgrounds from $Z + jets$. The selections are optimized using a significance metric:

$$\sigma = \sqrt{\sum_{i \in \text{bins}} 2((S_i + B_i) \ln(1 + \frac{S_i}{B_i}) - S_i)} \quad (8.5)$$

S and B are the number of signal and background events in the effective mass distribution estimated from MC simulation. The following cuts were optimized:

- W tagger working point: The 80% working point is selected as it suppresses SM background.
- $M_{\ell\ell}$: $M_{\ell\ell} > 100$ GeV chosen to preserve most signals whilst reducing $Z+jets$ background
- Leading lepton p_T : $p_T(\ell_1) > 27$ GeV. Increasing this did not improve significance, likely due to the $M_{\ell\ell}$ cut. The correlation of this cut with the $M_{\ell\ell}$ cut

was investigated but did not change the optimal value.

- E_T^{miss} : The optimization procedure suggests $E_T^{\text{miss}} > 80$ GeV for the boosted category and $E_T^{\text{miss}} > 60$ GeV for the resolved category
- Small-R jet p_T : A 20 GeV cut was found to work best for both the leading and sub-leading small-R jet p_T .
- Invariant mass of the two leading small-R jets, M_{jj} , in the resolved category. The optimal cut was $50 \text{ GeV} < M_{jj} < 120 \text{ GeV}$ (a window cut around the W mass).

The optimization procedure scans each variable one at a time. Thus the optimization of later variables is dependent on the ones coming before. A number of signal samples are used in this procedure, scanning a range of heavy Higgs mass, f_W and f_{WW} values. The results are summarized in table 8.6.

Table 8.6: The optimized selection criteria for the $SS2\ell$ signal region

$SS2\ell$ ($l_1^\pm l_2^\pm = e^\pm e^\pm, e^\pm \mu^\pm, \mu^\pm e^\pm, \mu^\pm \mu^\pm$)	
Boosted	Resolved
Two same-sign leptons with $p_T > (27)20$ GeV	
3^{rd} lepton veto	
no b-jet @ 85% DL1r	
$M_{ll} > 100$ GeV	
Boosted Category	!Boosted Category
$E_T^{\text{miss}} > 80$ GeV	$E_T^{\text{miss}} > 60$ GeV
$N_J \geq 1$	$N_J \geq 2$
$p_T(J_1) > 200$ GeV	$p_T(j_1), p_T(j_2) > 20$ GeV
J_1 @ 80% W tagger	$M_{jj} \in (50, 110)$ GeV

The $SS2l$ region does not allow for direct reconstruction of the heavy Higgs mass. The best discriminating variable is instead the effective mass, M_{eff} :

$$\text{boosted category: } \sum_i p_T(\ell_i) + p_T(J_1) + E_T^{\text{miss}} \quad (8.6)$$

$$\text{resolved category: } \sum_i p_T(\ell_i) + p_T(j_1) + p_T(j_2) + E_T^{\text{miss}} \quad (8.7)$$

where $p_T(j_{1(2)})$ is the p_T of the leading(sub-leading) small-R jet, $p_T(\ell_{1(2)})$ is for the leading(sub-leading) leptons and $p_T(J_1)$ is the p_T of the leading large-R jet.

8.6 Background Estimation

The backgrounds of this analysis are all SM processes which mimic the SS2l signature (two same-sign leptons, two jets and E_T^{miss}). This section summarizes how each background process is controlled for. There are six main categories:

- The non-prompt background: Hadronic jets can be mis-reconstructed as leptons due to non-prompt leptons produced in the jet. If one or two of these are mistaken for signal leptons then many different processes cause a fake signal, including V +jets, $t\bar{t}$, single top or QCD multijet.
- The charge mis-identification background (misID): A prompt electron may radiate a photon (bremsstrahlung). At high momentum, where tracks are almost straight, this can cause mis-reconstruction of the track curvature and thus charge misID. Equally, the photon can pair-produce e^+e^- , the original electron could be mistaken for the emitted positron. This background mainly comes from $t\bar{t}$, $W^\pm W^\mp + \text{jets}$ and $Z/\gamma + \text{jets}$ processes.
- The photon fake process: In $V\gamma + \text{jets}$ processes the photon can be mis-reconstructed as an electron. The vector boson decay provides the second lepton.
- The WZ background: In the $WZ + \text{jets}$ process, one of the leptons from the Z decay may not be reconstructed (or in $ZZ + \text{jets}$ if two leptons are missing).
- Double parton scattering (DPS): Collisions between pairs of partons could produce $W\text{jets} + W/Z\text{jets}$ or $W(W/Z)$ dijet processes. This background was found to be negligible however.
- The same-sign W boson (ssWW) background. In the SM, this is due to same-sign W vector boson scattering. This background has QCD and EW components.

Data-driven techniques are used to estimate the non-prompt background as described in section 8.6.1, the charge misID background as described in 8.6.2 and the photon fake background as described in 8.6.3. Two control regions are used to establish normalization factors on the MC estimated backgrounds for WZ and ssWW.

8.6.1 Non-prompt Leptons

A non-prompt background fake rate is estimated in the signal region using an ABCD method (figure 8.3). This is a fully data-driven technique, MC simulation doesn't accurately model this process thus cannot be used. The non-isolated antiID leptons are used as an estimate of the non-prompt leptons. Non-prompt electron and muon inclusive fake rates are derived separately using the ee and $\mu\mu$ sub-channels. This is to avoid the ambiguity over deciding which lepton is the fake in the $e\mu$ and μe subchannels. Fake rates binned by p_T and $|\eta|$ are also derived.

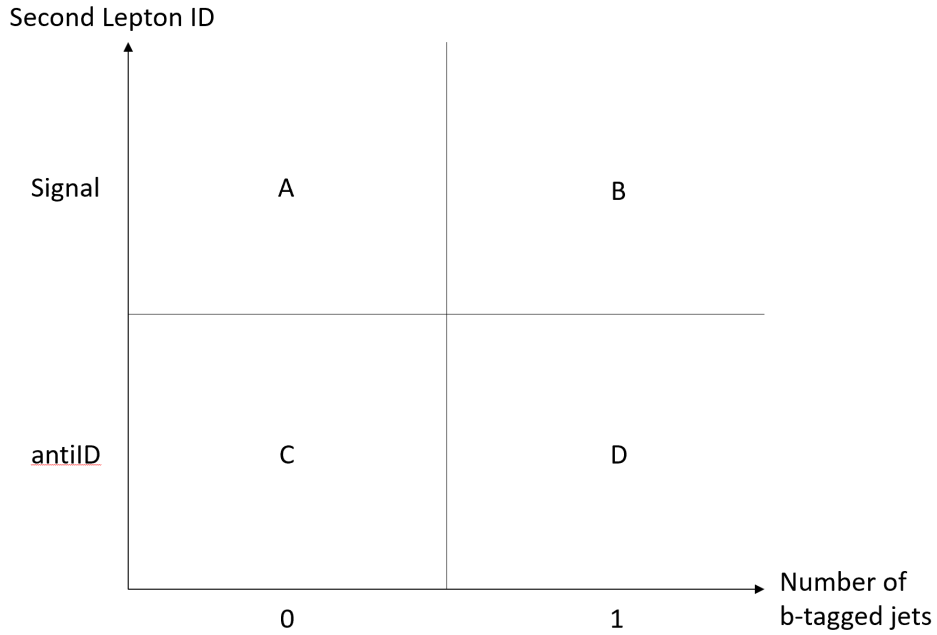


Figure 8.3: The ABCD method for the non-prompt fake background estimation. The fake factor is derived using the ratio of events in B and D, then applied to C to obtain the number of fake events in A.

The ABCD method assumes that the ratio of fake leptons in the region A to region C is equal to the ratio from region B to region D. Each region is defined to

be orthogonal³⁴. The region A is the signal region. The fake rate/factor is estimated using B and D, then applied to the region C to obtain the non-prompt background in region A. The regions A and C differ only by the second lepton, being either a signal (ID) or antiID lepton. The regions B and D differ from A and C by the requirement of a b-tagged jet to ensure orthogonality, and from each other by the ID requirement of the second lepton. Some other restrictions in B and D are loosened or removed entirely to increase the statistics:

- The W-tagger requirement and M_{jj} cut are removed. For this method, we do not separate into boosted and resolved categories.
- $M_{\ell\ell}$ cut is removed in all sub-channels, with the exception of the ee sub-channel where a cut is kept to remove the Z+jets background: $|M_{ee} - 90 \text{ GeV}| > 10 \text{ GeV}$.

An inclusive fake rate with respect to the ABCD method is derived separately for electrons and muons using the ee and $\mu\mu$ sub-channels. For the $e\mu$ and μe signal region sub-channels, the non-prompt background is estimated by applying either the inclusive electron or muon fake rate depending on the flavour of the antiID lepton in the $e\mu$ and μe region B sub-channel. Electron and muon fake rates are additionally calculated binned by p_T and $|\eta|$ by assuming the sub-leading lepton in the region B is fake. This is again done using the ee sub-channel for electrons but for muons the $e\mu$ sub-channel is used instead as studies with truth information show the sub-leading fake assumption holds better. The estimation for the remaining two sub-channels is done as before by use of the antiID lepton flavour. Non-prompt leptons originate mainly in jets due to photon conversions or heavy hadron decays. Photon conversions would be expected to increase with $|\eta|$ as more detector material is present whilst heavy hadron decays will generally give a more energetic (higher p_T) lepton. It is found that a p_T binning is best for the electron fake factor and an η binning for the muon fake factor. This performance is checked by applying

³No two regions share the same event.

⁴It is also implicitly assumed the cuts applied to ensure orthogonality are uncorrelated. Any bias resulting from this seems small as there is good agreement in the validation plots in section 8.7.

the derived fake factors to the control region B. The data/MC agreement is plotted in the region B using the non-prompt fake factor to estimate the background. The other background source are from the MC samples of processes that produce prompt leptons (ttV, WZ, ZZ, etc.) and the charge flip background (see section 8.6.2). As the non-prompt factor is derived using B and D, the normalizations of the data and MC will agree by construction but not the shape. The shape in the data/MC plot thus serves as a validation of the method. A good shape agreement was seen for all kinematic distributions in this region B, for example in figures 8.4 and 8.5. The inclusive, p_T binned and $|\eta|$ binned fake rates are shown in tables ?? and 8.8.

pT bin	electron	muon
Inclusive	0.048 ± 0.006	0.018 ± 0.001
$20 \text{ GeV} < p_T < 30 \text{ GeV}$	0.027 ± 0.005	0.015 ± 0.002
$30 \text{ GeV} < p_T < 40 \text{ GeV}$	0.051 ± 0.012	0.019 ± 0.004
$p_T > 40 \text{ GeV}$	0.139 ± 0.034	0.020 ± 0.005

Table 8.7: The inclusive and p_T binned non-prompt fake rates for electrons and muons with statistical uncertainty included.

$ \eta $ bin	electron	muon
Inclusive	0.048 ± 0.006	0.018 ± 0.001
$0 < \eta < 0.5$	0.049 ± 0.011	0.024 ± 0.004
$0.5 < \eta < 1.5$	0.046 ± 0.009	0.016 ± 0.003
$ \eta > 1.5$	0.050 ± 0.014	0.011 ± 0.003

Table 8.8: The inclusive and $|\eta|$ binned non-prompt fake rates for electrons and muons with statistical uncertainty included.

The composition of the non-prompt background is estimated from the MC simulations. There is reasonable agreement between the MC and the data in the plots of kinematic variables⁵, hence some estimates can be made. The derived background appears to be mainly composed of non-prompt leptons from heavy hadron decays, e.g. $t\bar{t}$ or $W + b/c$ jet events. In channels with electrons, the contribution from $W + \text{light jet}$ events is fairly large. No additional uncertainty is considered to account for this, as it would be much smaller than the statistical uncertainty on the derived

⁵Plots are not yet public.

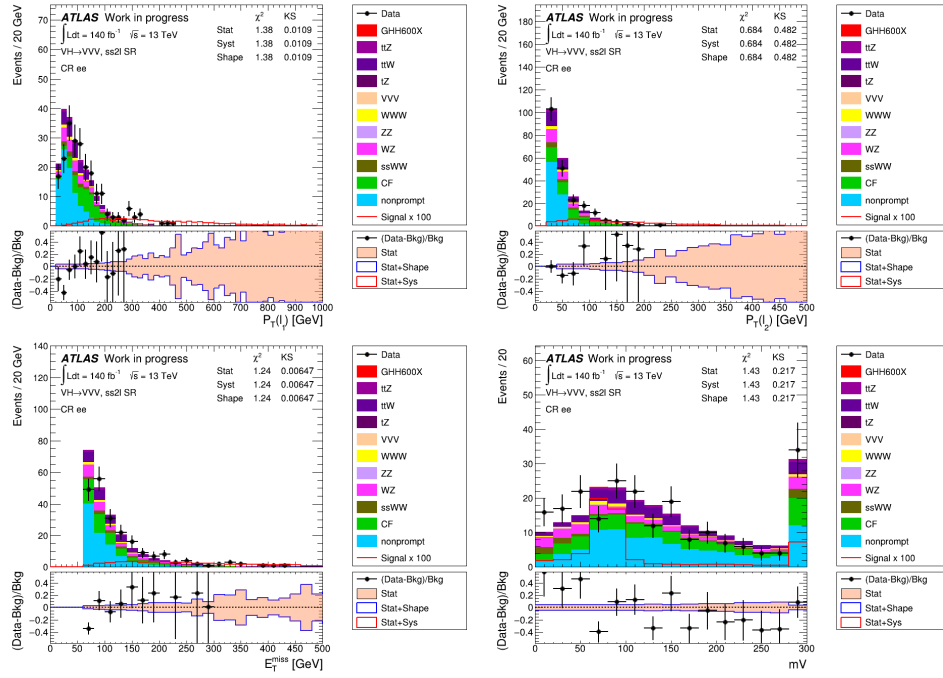


Figure 8.4: The data/MC agreement in the control region B ee sub-channel. CF refers to the charge flip background. The background from non-prompt leptons is estimated using the p_T binned non-prompt electron fake factor applied to the control region D (ee sub-channel). These four distributions show the typical agreement in all the kinematic plots.

non-prompt background. Furthermore, the good agreement in the validation region in figures 8.12 and 8.13 in section 8.7 suggests this assumption has not caused any issues. The non-prompt background is a relatively small contributor, thus any additional bias remains small.

8.6.2 Charge mis-identification

The charge mis-identification (misID) background stems from electrons that produce a photon through bremsstrahlung and track mis-reconstruction uncertainties at high p_T . The charge misID of muons is negligible in the regions considered by this analysis. The charge misID rate for electrons can be estimated using a data sample enriched in $Z \rightarrow e^+e^-$ events (known as the Zee control region). The region is split between events where the electrons have the same sign (SS) or opposite sign (OS). The region is further split into a mass window region and a sideband region. The selections are the same as for the ee signal region in table 8.6 but with the following changes:

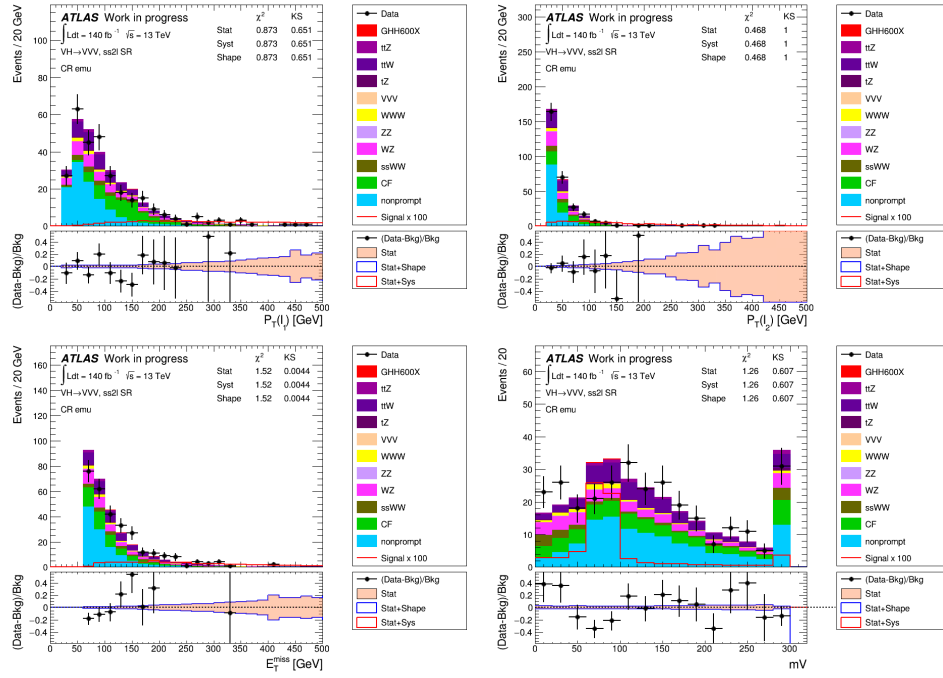


Figure 8.5: The data/MC agreement in the control region B $e\mu$ sub-channel. CF refers to the charge flip background. The background from non-prompt leptons is estimated using the $|\eta|$ binned non-prompt muon fake factor applied to the control region D ($e\mu$ sub-channel). These four distributions show the typical agreement in all the kinematic plots.

- Z mass window region: $75 \text{ GeV} < M_{ee} < 105 \text{ GeV}$
- Z mass sideband region: $60 \text{ GeV} < M_{ee} < 75 \text{ GeV}$ or $105 \text{ GeV} < M_{ee} < 120 \text{ GeV}$
- Jet requirements are removed.

The sideband regions are used to roughly estimate the non-Z background events. It is assumed that the data yield in the sideband regions is entirely from all non-Z SM backgrounds. As the sidebands have the same 30 GeV mass range, this yield is then roughly equal to the non-Z background in the Z mass window region. We thus can obtain the number of Z events in the mass window region by subtracting the data yield in the sideband regions. This is done for both SS and OS regions and is validated using MC simulation events. The following likelihood fit procedure is used to estimate the charge misID rates as a function of the electron p_T and η .

The charge flip rates of two electrons are defined as ϵ_i and ϵ_j . The probability distribution of observing N_{ss} same-sign ee events is a Poisson distribution:

$$P(N_{ss}|\epsilon_i, \epsilon_j, N_{os+ss}^{exp}) = \frac{(N_{ss}^{exp})^{N_{ss}} e^{-N_{ss}^{exp}}}{N_{ss}!}, \quad (8.8)$$

where the expected number of same-sign events, N_{ss}^{exp} , is defined as the probability that exactly one electron has its charge flipped multiplied by the expected total number of $Z \rightarrow ee$ events (of any sign combination), N_{os+ss}^{obs} :

$$N_{ss}^{exp} = N_{os+ss}^{exp} [(1 - \epsilon_i)\epsilon_j + (1 - \epsilon_j)\epsilon_i], \quad (8.9)$$

The number of same-sign and opposite-sign events are then observed in 4 p_T and 6 η regions, for a total of 24 separate bins. The likelihood function can then be defined:

$$L(\epsilon) = \prod_i \prod_j P(N_{ss,ij}^{obs}|\epsilon_i, \epsilon_j, N_{ss+os,ij}^{obs}), \quad (8.10)$$

where $i(j)$ corresponds to the $i(j)$ -th electron $p_T \times \eta$ bin. $N_{ss,ij}^{obs}$ is the number of observed same-sign events with the leading electron in the i -th bin and sub-leading electron in the j -th bin. The charge flip rates, ϵ_i for each bin are obtained by minimising the negative log-likelihood. The obtained rates are then used to estimate the charge flip background. The charge flip rates are found to be generally well below 0.1% in most of the $p_T \times \eta$ regions, except for the highest bin ($2.3 < |\eta| < 2.5$ and $130 < p_T < 1000$ GeV) where it is around 2%. The estimate charge flip background yields in the signal regions are shown in table 8.9.

	Boosted	Resolved
Charge Flip	0.43	8.56
All Backgrounds	29.28	166.30

Table 8.9: The charge flip background yields in the inclusive boosted and resolved signal regions.

A Z mass validation region is used to evaluate the performance of the charge flip background estimation. This uses all the ee signal region cuts except it removes the cut on E_T^{miss} and requires the dilepton mass:

- $80 \text{ GeV} < M_{ee} < 100 \text{ GeV}$

It is thus orthogonal to the signal region. The charge flip background in this region is estimated using the derived charge misID rates, with all other backgrounds estimated from MC simulation. Good agreement between the predicted backgrounds and data was seen in plots of the data/MC distributions in the Z mass validation region, as shown in figure 8.6.

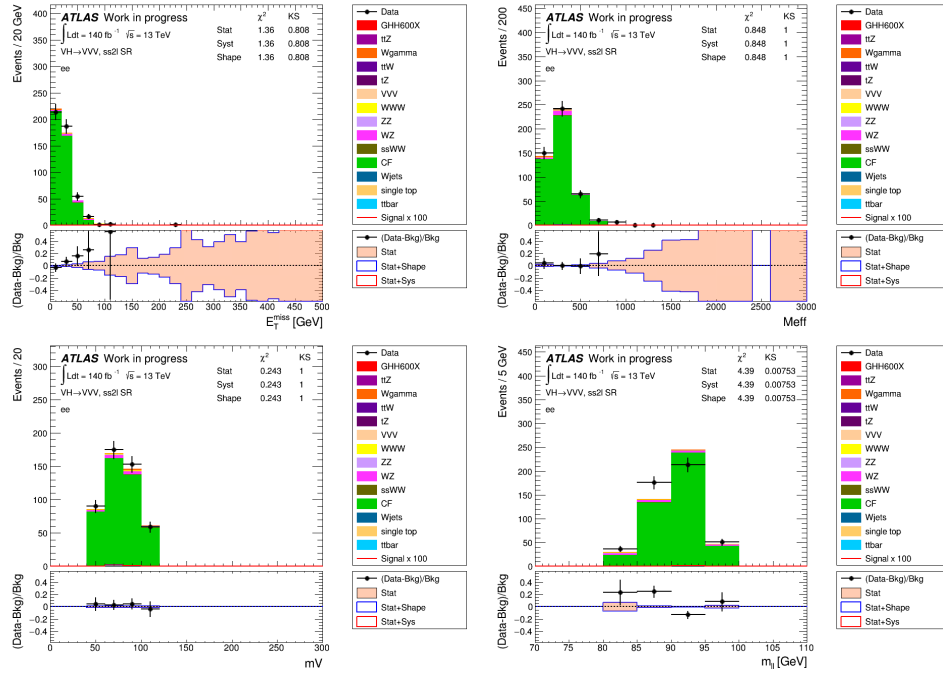


Figure 8.6: A few kinematic distributions in the Z mass validation region showing typical agreement. The charge flip contribution is estimated using the charge flip rate applied to the data in the opposite-sign lepton region instead of using the Z+jets MC. Only statistical uncertainties are shown.

8.6.3 Fake Leptons from Photons

Prompt photons, such as those produced in $V\gamma$ events, can be mis-reconstructed as electrons. This can happen when the photon converts to e^+e^- and one of these leptons is not reconstructed. When combined with the lepton or leptons from the V decay, the event will enter the signal region.

The photon fake rate, i.e. photon mis-reconstructed as electron, is estimated using another ABCD method similar to section 8.6.1. In this instance, the “A” (signal region) and “C” (termed the $W\gamma$ CR) regions are differentiated by the sub-

leading lepton being a signal or anti-IBL electron. The “B” ($Z\gamma \rightarrow \ell_{ID}\ell_{ID}\ell_{ID}$ CR) and “D” ($Z\gamma \rightarrow \ell_{ID}\ell_{ID}\ell_{Anti-IBL}$ CR) regions differ from “A” and “C” respectively by containing a third lepton. This is summarized in figure 8.7. Some additional changes are made for the $Z\gamma$ CR definitions:

- Two opposite-sign signal muons and an additional signal or anti-IBL electron. Only the $\mu^\pm\mu^\pm e^\mp$ sub-channel is used to calculate the photon fake rate. This is done to reduce the influence of charge-flip backgrounds.
- Veto events with more than 3 leptons (using veto leptons).
- $80 \text{ GeV} < M_{\ell\ell\ell} < 100 \text{ GeV}$, this replaces the $M_{\ell\ell}$ cut. This is needed to ensure orthogonality with the WZ control region in section 8.6.4.

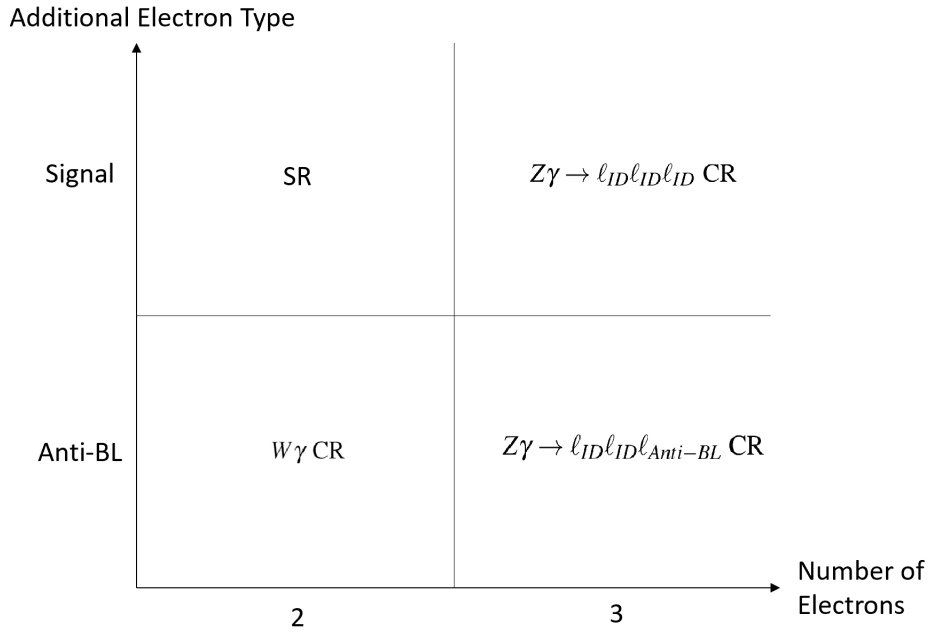


Figure 8.7: The ABCD method for the photon conversion fake background estimation. The fake factor is derived using the ratio of events in B and D, then applied to C to obtain the number of fake events in A.

The photon fake rate is then calculated analogously to the non-prompt fake rate, using the ratio of yield in the two $Z\gamma$ CRs and applying this to the data yield in the $W\gamma$ CR. The method is again validated using a closure test: plotting data/MC comparisons in the $Z\gamma \rightarrow \ell_{ID}\ell_{ID}\ell_{ID}$ CR and replacing the $Z\gamma$ MC sample with the

estimation made from the photon fake rate applied to the $Z\gamma \rightarrow \ell_{ID}\ell_{ID}\ell_{Anti-IBL}$ CR. The data/MC plots using the photon conversion estimated background showed good shape agreement, see figure 8.8. To produce these plots, the VGammaORTool was implemented in our CxAOD framework to remove overlapping MC events in $V\gamma + \text{jets}$ and $V + \text{jets}$. These two MC samples contain overlapping events, due to the production of final state photons from the leptons in $V + \text{jets}$ samples [129].

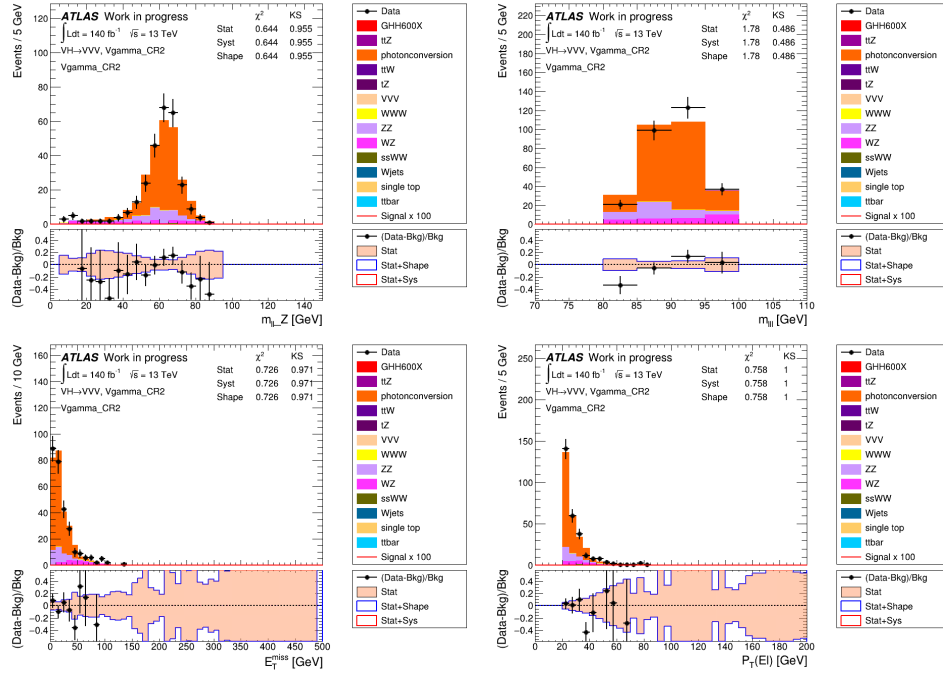


Figure 8.8: Data/MC comparison for four observables in the $Z\gamma$ ID+ID+ID control region. The $Z\gamma$ background is estimated using the photon conversion fake rate applied to the $Z\gamma \rightarrow \ell_{ID}\ell_{ID}\ell_{Anti-IBL}$ CR. Statistical uncertainties are included.

8.6.4 Backgrounds with prompt leptons

WZ Background The $WZ^{(*)} + \text{jets}$ process ($W^{\pm}Z \rightarrow \ell^{\pm}\nu\ell^{+}\ell^{-}$) can form a background in the event that one of the decay leptons is not identified. The MC simulated sample is used in the fit but first normalized to data using two normalization factors. These are derived by fitting to data in two separate WZ control regions. The Boosted and Resolved WZ CRs are three lepton regions⁶, other than that the SR cuts are used. In the eee and $\mu\mu\mu$ sub-channels, one lepton has the opposite sign

⁶Ensuring orthogonality to the SR.

of the other two. In the $ee\mu$ and $\mu\mu e$ sub-channels, the same flavour leptons have opposite sign. The WZ CRs are defined fully in table 8.10, and are found to be 95% pure⁷. The normalization factors are estimated from the background-only fit (see section 8.10) as 0.927 ± 0.067 in the boosted region and 0.824 ± 0.025 in the resolved region. Figure 8.9 and Figure 8.10 show a few typical variable distributions in the boosted and resolved WZ CRs, with the corresponding the WZ normalization factors applied. Good data MC agreement can be observed.

WZ CR	
Boosted	Resolved
3 leptons with $p_T > 27, 20, 20\text{GeV}$	
4^{th} lepton veto	
at least one pair of same-flavour opposite sign leptons	
$M_{lll} > 110\text{ GeV}$	
no b-jet @ 85% DL1r	
$E_T^{\text{miss}} > 40\text{ GeV}$	
Boosted Category	!Boosted Category
≥ 1 large-R jet	≥ 2 small-R jets
$p_T > 200\text{ GeV}$	$p_T(j_1), p_T(j_2) > 20\text{ GeV}$

Table 8.10: The selection cuts defining the WZ CR.

ssWW Background The background from the same-sign $W^\pm W^\pm$ vector boson scattering (VBS) process is again estimated using a normalization factor derived from a $ssWW$ CR. The $ssWW$ CR is defined in the same way as the resolved signal region⁸ in table 8.6 but with $M_{jj} > 200\text{ GeV}$. This ensures orthogonality and improves the $ssWW$ purity in the region. The full region definition is found in table 8.11. The normalization factor is estimated from the background-only fit (see section 8.10) as 1.441 ± 0.174 , which agrees with the value from recent measurements [130]. Figure 8.11 shows a few typical variable distributions in the $ssWW$ CR after application of the normalization factors for $ssWW$ and WZ. Good agreement between data and MC can be observed.

⁷More than 95% of events in the CRs come from the WZ process.

⁸There is no boosted $ssWW$ control region.

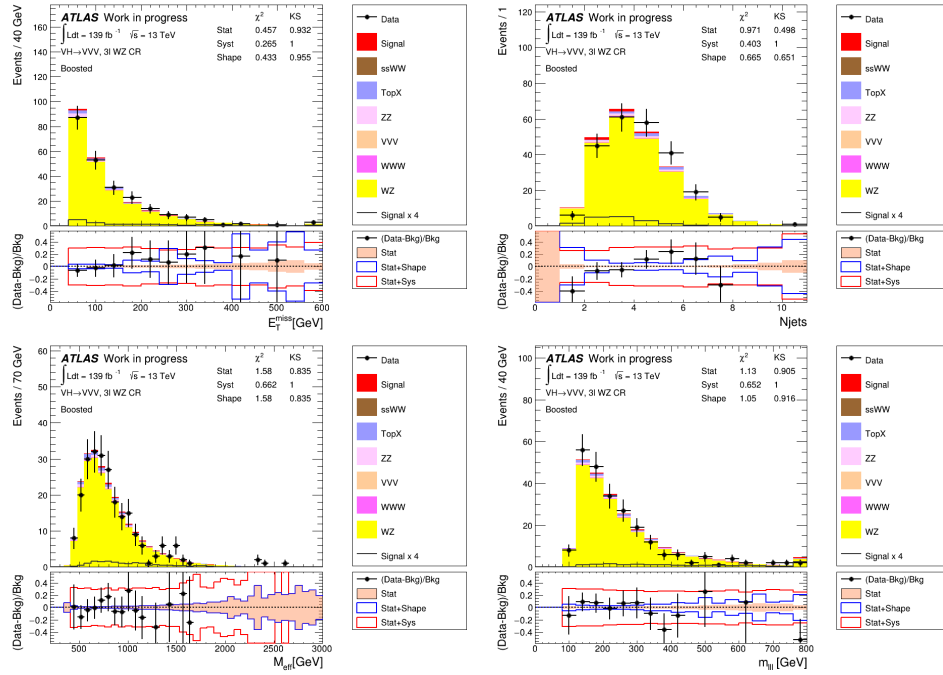


Figure 8.9: Data/MC plots in the boosted WZ CR of four variables showing typical agreement: missing Et (top left), number of jets (top right), effective mass (bottom left), invariant mass for three leptons (bottom right). A 0.93 normalization factor applied to the WZ events. Statistical and systematic uncertainties are added.

$ssWW$ CR
not pass boosted category definition
Two same-sign leptons with $p_T > (27)20$ GeV
3^{rd} lepton veto
no b-jet @ 85% DL1r
$M_{ll} > 100$ GeV
$E_T^{miss} > 40$ GeV
$N_j \geq 2$
$p_T(j_1), p_T(j_2) > 20$ GeV
$M_{jj} > 200$ GeV

Table 8.11: The selection cuts defining the $ssWW$ control region. Note not passing the boosted category means the event is vetoed if the leading reconstructed large-R jets has $p_T > 200$ GeV and $50 \text{ GeV} < M_J < 200$ GeV.

8.7 Validation Region

A validation region (VR) is designed in this analysis to provide data-to-simulation comparisons for validation of the background estimation methods. The data in the signal region remains blinded throughout these procedures. The VR is designed to

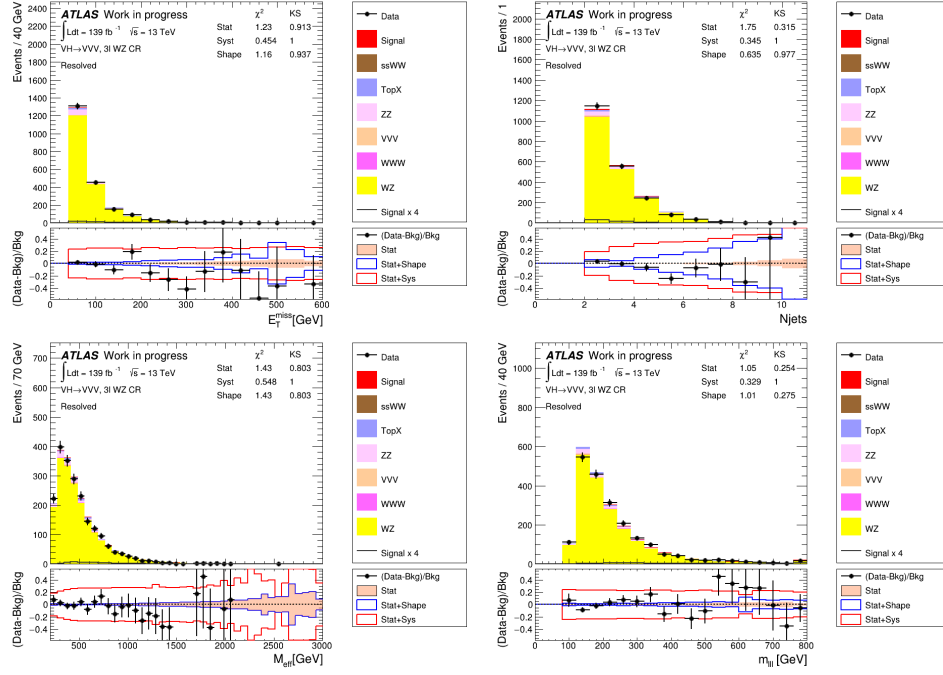


Figure 8.10: Data/MC plots in the resolved WZ CR of four variables showing typical agreement: missing E_t (top left), number of jets (top right), effective mass (bottom left), invariant mass for three leptons (bottom right). A 0.82 normalization factor applied to the WZ events. Statistical and systematic uncertainties are added.

contain minimal signal events and be orthogonal to the signal region. This is done by reversing the W tagger requirement in the boosted category and using the mass sideband region in the resolved category. Otherwise, all selections are the same as SR. The full selection criteria are shown in table 8.12. The agreement of the data and estimated backgrounds is shown in kinematics distribution plots for the boosted validation region (figure 8.12) and resolved validation region (figure 8.13). Reasonable agreement is seen. These distributions are made before the fit to data and thus do not include the normalization factors.

8.8 Systematic Uncertainties

The various systematic uncertainties are summarized in this section. These enter the statistical fit in section 8.9.

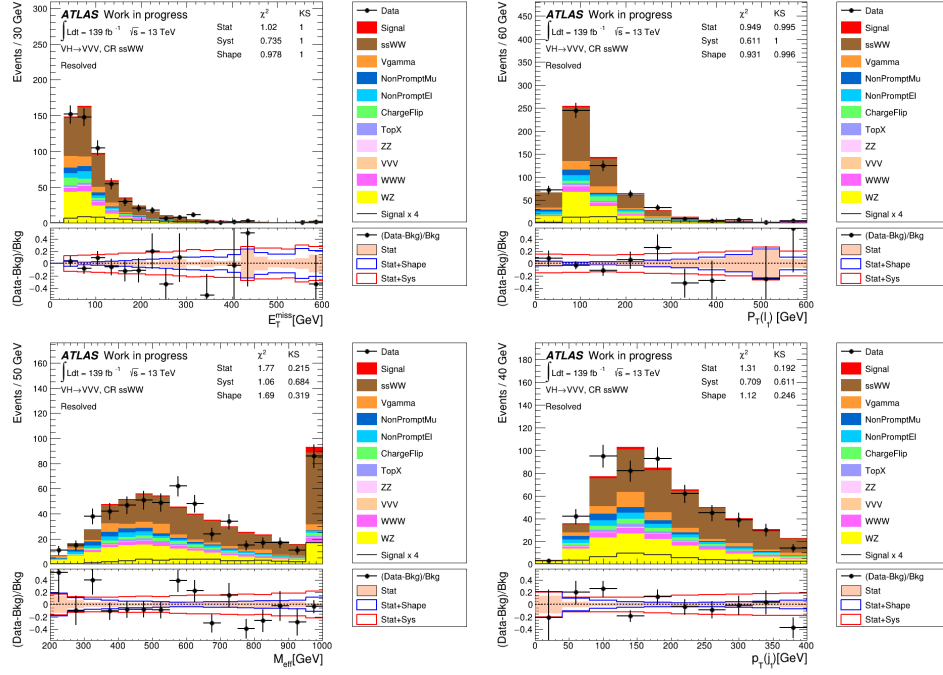


Figure 8.11: Data/MC plots in the $ssWW$ CR of four variables showing typical agreement: missing E_T (top left), leading lepton p_T (top right), effective mass (bottom left), leading jet p_T (bottom right). WZ and $ssWW$ normalization factors are both applied and statistical and systematic uncertainties shown.

$SS2\ell$ VR ($l_1^\pm l_2^\pm = e^\pm e^\pm, e^\pm \mu^\pm, \mu^\pm e^\pm, \mu^\pm \mu^\pm$)	
Boosted	Resolved
Two same-sign leptons with $p_T > (27)20$ GeV 3 rd lepton veto no b-jet @ 85% DL1r $M_{ll} > 100$ GeV	
Boosted Category	!Boosted Category
$E_T^{miss} > 80$ GeV	$E_T^{miss} > 60$ GeV
$N_J \geq 1$	$N_J \geq 2$
$p_T(J_1) > 200$ GeV	$p_T(j_1), p_T(j_2) > 20$ GeV
J_1 not @ 80% W tagger	$M_{jj} \leq 50$ GeV or 110 GeV $\geq M_{jj} \leq 200$ GeV

Table 8.12: The validation region selection criteria.

8.8.1 Experimental Systematics

This section covers the systematic uncertainties related to the experimental measurements of the ATLAS detector.

Luminosity Uncertainty The systematic uncertainty on the luminosity value for the 2015-2017 dataset is 1.7%. It is derived from the luminosity scale calibrations

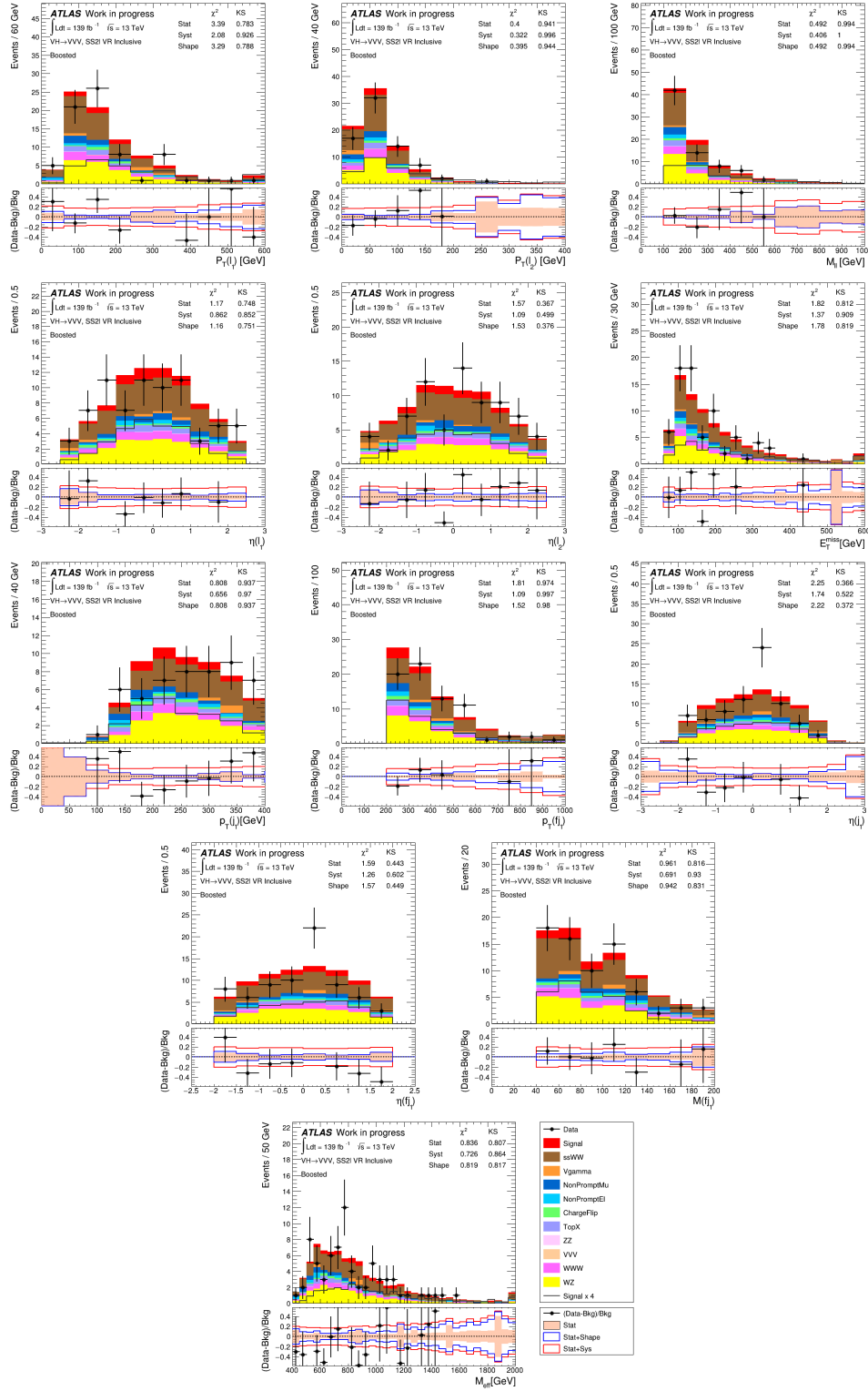


Figure 8.12: Various kinematic distributions in the boosted validation region (inclusive channel), showing the agreement of data with the estimated backgrounds. Systematic and statistical uncertainties are shown. The signal is shown stacked on top of the SM background (solid red block), the signal $\times 4$ is also shown unstacked (black line).

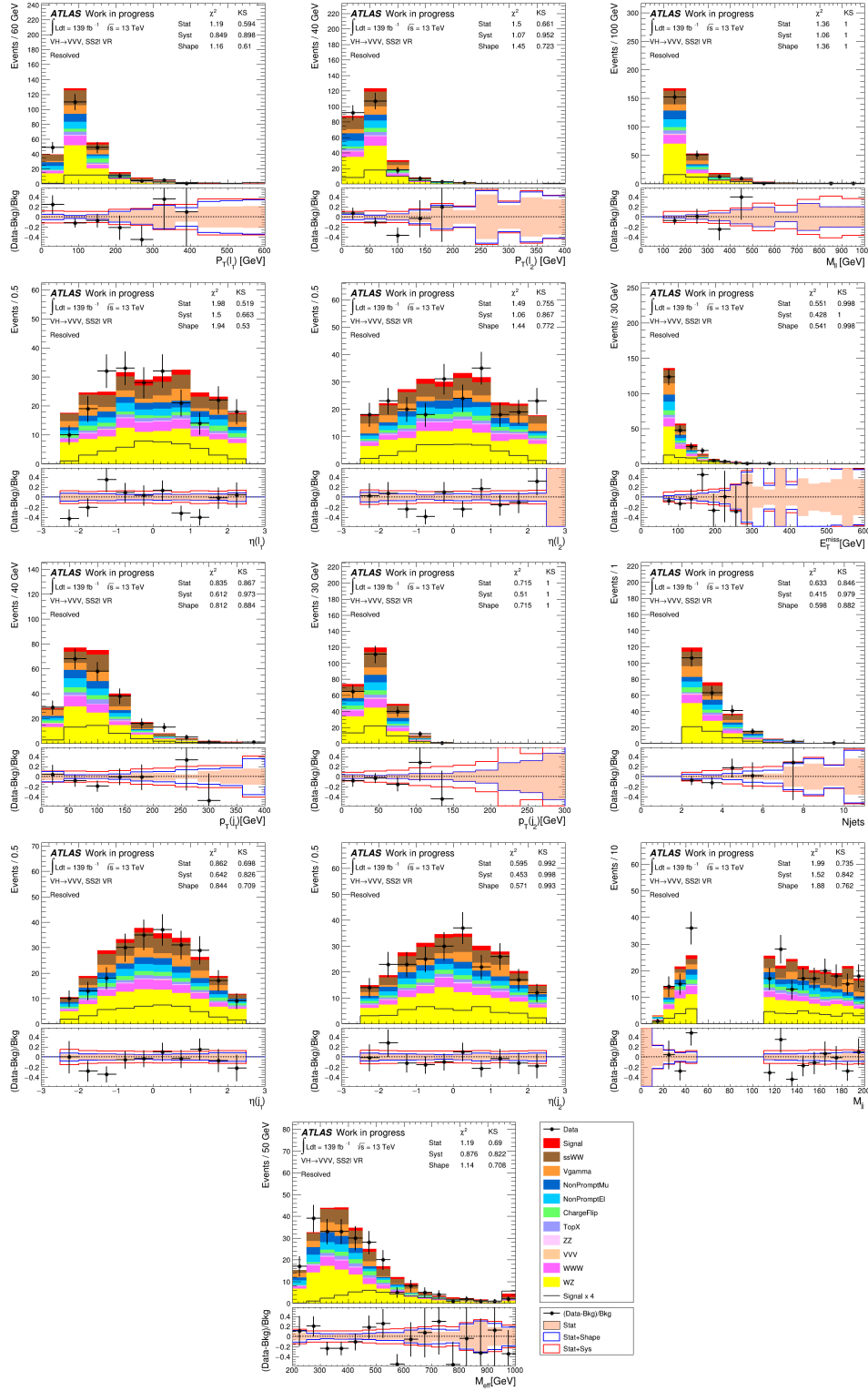


Figure 8.13: Various kinematic distributions in the resolved validation region (inclusive channel), showing the agreement of data with the estimated backgrounds. Systematic and statistical uncertainties are shown. The signal is shown stacked on top of the SM background (solid red block), the signal $\times 4$ is also shown unstacked (black line).

using xy beam-separation scans. The methodology follows that detailed in [131].

Pile-up Uncertainty An uncertainty is applied on the pile-up reweighting of MC samples. The recommendation for this comes from a dedicated ATLAS group.⁹

Muon Uncertainties The systematic uncertainties on reconstructed muons are implemented by the dedicated Muon Combined Performance (MCP) group¹⁰. A number of uncertainties are implemented for the different aspects of muon reconstruction:

- **Track Uncertainties:** These relate to variations from the smearing of ID and MS tracks used to create the combined muon track. These smearings are applied cover the uncertainties in the track p_T reconstruction and resolution.
- **Momentum Scale Uncertainty.**
- **Sagitta Uncertainties:** Variations to the charge dependent momentum scale due to corrections on the Z scale or residual charge-dependent bias.
- **Isolation Efficiency:** The PLV is applied to muons. Uncertainties applied on the MC-to-data scale factors of this tool.
- **Reconstruction Efficiency Uncertainty:** These apply to the muons passing the identification working points from [69].
- **Track to Vertex association efficiency:** Relating to association of muons tracks to vertices.
- **Trigger Efficiency Uncertainty:** Relating to the scale factors for muon trigger efficiency.

These are the recommendations for the ATLAS derivation framework (`AnalysisBase, 21.2.152`). The uncertainties were derived using data samples with large statistical power (e.g. $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu\mu$). The muon uncertainties do not generally have a large impact on our results¹¹.

⁹Internal documentation (ATLAS members only) is found at [ExtendedPileupReweighting](#).

¹⁰Internal documentation can be found [here](#)

¹¹Muon resolution tends to be very good at the momenta used in this analysis (p_T of order 100 GeV).

Electron Uncertainties The uncertainties on electron reconstruction performance are from the dedicated e/gamma combined performance group¹². The following uncertainties are provided:

- Reconstruction and Identification Efficiency: The uncertainties relating to the electron ID and reconstruction efficiencies. These are measured using events around the Z mass peak.
- Isolation Efficiency: The PLV is used to isolate electrons as with muons.
- Momentum scale and resolution uncertainties.
- Trigger efficiency uncertainties.

Small-R Jet Uncertainties The following uncertainties are considered for the small-R jet objects, including uncertainties on flavour tagging (as a b-jet veto is used). These are developed by the Jet/Etmiss and Flavour Tagging groups¹³. The uncertainties used are:

- Jet Energy Scale Uncertainties: A category reduction scheme is used to calculate the uncertainty in MC simulation. This is detailed in [55].
- Jet Energy Resolution Uncertainties: A scheme called SimpleJER is used [56]. This analysis does not have hard selections on jets.
- Flavour Tagging Efficiency Uncertainties: The b-jet tagging differences between MC simulations and data are corrected using scale factors. These are varied to obtain systematic uncertainties.
- JVT Efficiency: The JVT tool efficiency uncertainty is derived from a comparison of hard scatter jets in $Z \rightarrow \mu\mu$ data and simulation.

¹²e/gamma Combined Performance internal link.

¹³Internal links: [Jet/Etmiss](#) and [Flavour Tagging](#)

Large-R Jet Uncertainties The large-R jet systematic uncertainties are applied based on the recommendations of the ATLAS combined physics (CP) group. The following are used:

- **Fatjet Energy Resolution:** The uncertainty is obtained by applying a 2% smearing in absolute p_T . This follows the method used in ref. [132].
- **Fatjet Energy Scale:** These uncertainties relate to the in-situ calibration of the JES detailed in chapter 4. A category reduction scheme is used [133].
- **Boosted Jet Tagging:** The W tagging is performed using the W/Z tagger detailed in 4.1.3.3. These relate to the data-to-MC scale factors used for the tagger efficiency and inefficiency.
- **Fatjet Mass Scale and Resolution:** As mentioned in section 4.1.3.2, a well defined mass is required for large-R jets. Uncertainties relating to the mass scale and resolution are used.

Large-R jets are a recent addition to the main ATLAS reconstruction program, hence dedicated recommendations for all their uncertainties do not yet exist. The flavour tagging of large-R jets is not considered in this analysis, so their uncertainties are not used.

Missing Transverse Momentum Uncertainties The E_T^{miss} entails systematic uncertainties relating to the momenta of objects used in its calculation. The soft term, momenta of unassociated tracks, also entails an uncertainty. The dedicated Etmis group ¹⁴ provides tools to handle these uncertainties.

8.8.2 Data-Driven Background Systematics

The systematic uncertainties for the data-driven background estimation methods are presented here.

Non-Prompt Lepton Uncertainties The derivation of the non-prompt lepton rate (section 8.6.1) entails three systematic uncertainties. First, the statistical uncertainty

¹⁴Internal link: [Etmis](#)

Table 8.13: Summary of the experimental systematic uncertainties.

Systematic uncertainty name	Description
	Event
LUMI_2015_2018	uncertainty on total integrated luminosity
PRW_DATASF	uncertainty on pile-up modelling
	Electrons
EL_EFF_Trigger_TOTAL_1NPCOR_PLUS_UNCOR	trigger efficiency uncertainty
EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	reconstruction (reco.) efficiency uncertainty
EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR	identification (id.) efficiency uncertainty
EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR	isolation efficiency uncertainty
EG_SCALE_ALL	energy scale uncertainty
EG_RESOLUTION_ALL	energy resolution uncertainty
	Muons
MUON_EFF_RECO_STAT(SYS)	reconstruction and ID efficiency uncertainty for muons with $p_T > 15$ GeV
MUON_EFF_RECO_STAT(SYS)_LOWPT	reco. and ID efficiency uncertainty for muons with $p_T < 15$ GeV
MUON_EFF_ISO_STAT(SYS)	isolation efficiency uncertainty
MUON_TTVA_STAT(SYS)	track-to-vertex association efficiency uncertainty
MUON_SCALE	momentum scale uncertainty
MUON_SAGITTA_RHO(RESBIAS)	momentum scale uncertainty to cover charge-dependent local misalignment effects
MUON_ID(MS)	momentum resolution uncertainty of the inner detector (muon spectrometer)
MUON_EFF_TrigStat(Syst)	trigger efficiency uncertainty
	Small-R jets
JET_CR_BJES_Response	energy scale uncertainties for b -quark jets
JET_CR_EffectiveNP_Detector{1-2}	energy scale uncertainties due to in-situ calibration
JET_CR_EffectiveNP_Mixed{1-3}	energy scale uncertainties due to in-situ calibration
JET_CR_EffectiveNP_Modelling{1-4}	energy scale uncertainties due to in-situ calibration
JET_CR_EffectiveNP_Statistical{1-6}	energy scale uncertainties due to in-situ calibration
JET_CR_EtaIntercalibrationModelling	energy scale uncertainties to cover η -intercalibration non-closure
JET_CR_EtaIntercalibrationNonClosure_highE	energy scale uncertainties to cover η -intercalibration non-closure
JET_CR_EtaIntercalibrationNonClosure_negEta	energy scale uncertainties to cover η -intercalibration non-closure
JET_CR_EtaIntercalibrationNonClosure_posEta	energy scale uncertainties to cover η -intercalibration non-closure
JET_CR_EtaIntercalibration.TotalStat	energy scale uncertainties to cover η -intercalibration non-closure
JET_CR_Flavor.Composition(Flavor.Response)	energy scale uncertainty related to flavour composition (response)
JET_CR_PileUp_{OffsetMu(NPV), PtTerm, RhoTopology}	energy scale uncertainties due to pile-up effects
JET_CR_PunchTroughMC16	energy scale uncertainty for 'punch-through'
JET_CR_SingleParticle.HighPt	energy scale uncertainty for the behavior of high- p_T jets
JET_JvtEfficiency	JVT efficiency uncertainty
JET_CR_JER_DataVsMC_MC16	energy resolution uncertainty
JET_CR_JER_EffectiveNP_{1-6, 7restTerm}	energy resolution uncertainties
	Large-R jets
FATJET_Medium_JET_Comb_Baseline_Kin	energy and mass scale uncertainty due to basic data-simulation differences
FATJET_Medium_JET_Comb_Modelling_Kin	energy and mass scale uncertainty due to simulation differences
FATJET_Medium_JET_Comb_Tracking_Kin	energy and mass scale uncertainty on reference tracks
FATJET_Medium_JET_Comb_TotalStat_Kin	energy and mass scale uncertainty from stat. unc. on the measurement
FATJET_BJT	boosted jet tagger uncertainties
FATJET_JER	energy resolution uncertainty
FATJET_JMR	mass resolution uncertainty
	: VR track jets
FT_EFF_Eigen.B_{0-4}	b -tagging efficiency uncertainties for b jets
FT_EFF_Eigen.C_{0-3}	b -tagging efficiency uncertainties for c jets
FT_EFF_Eigen.Light_{0-3}	b -tagging efficiency uncertainties for light jets
FT_EFF_extrapolation	b -tagging efficiency uncertainty for high- p_T b -quark and jets
FT_EFF_extrapolation.from.charm	b -tagging efficiency uncertainty on τ -lepton jets
	E_T^{miss}
MET_SoftTrk_ResoPara(Perp)	'soft term'-related longitudinal(transverse) resolution uncertainty
MET_SoftTrk_Scale	'soft term'-related scale uncertainty
MET_JetTrk_Scale	scale uncertainty

on this derivation is considered. Second, the uncertainties on the prompt MC normalizations and charge flip background, as these backgrounds are subtracted from the data events in the non-prompt CRs. Third, the impact of b -tagging uncertainties on the non-prompt rate are considered, as the b -jet requirement is the key difference between these control regions with the signal region. There is no guarantee this is not correlated to probability of a non-prompt lepton.

The statistical uncertainty relates to the number of data events in the non-prompt control regions. Non-prompt electrons in particular suffer due to contam-

ination from many other processes. The muon non-prompt rate has a statistical uncertainty of 5.5%, whilst the electron rate has an uncertainty of 12.5%.

The uncertainty from the MC and charge flip backgrounds subtraction from the ID+ID non-prompt control region (control region B in fig. 8.3) is derived by applying scale factors. The non-prompt rates are rederived but a scale factor of 1.1 or 0.9 is applied to the prompt MC and charge flip backgrounds before these are subtracted. The fake rates derived in each case are shown in table 8.14. These differences are well covered by statistical uncertainty.

Scale Factor	electron	muon
1 (default)	0.048 ± 0.006	0.018 ± 0.001
1.1	0.042 ± 0.006	0.017 ± 0.001
0.9	0.051 ± 0.006	0.020 ± 0.001

Table 8.14: The effect of applying a 1.1 or 0.9 scale factor to the prompt MC and charge flip backgrounds.

To measure the impact of the b-tagging, the 85% working point is modified to the 77% working point and the fake factors re-derived. For the electron fake factor the rate changes from 0.048 ± 0.006 to 0.052 ± 0.008 , whilst for the muon fake rate it changes from 0.018 ± 0.001 to 0.021 ± 0.002 . The difference is within the statistical uncertainty.

Charge misID Uncertainties The systematics stemming from the derivation of the charge flip rate are considered. The charge flip rate is derived from a $Z \rightarrow ee$ sample in a Z mass window. The non-Z component of this is estimated from the sideband region and subtracted. The uncertainty is estimated in three ways.

The first way is to vary the Z mass window selections by 2 GeV and check the bias. The second way is by comparing with the rates derived using $Z + \text{jets}$ MC simulation. The third way is to subtract the MC predicted non-Z contribution instead of the data in sideband regions. Finally, the statistical uncertainty on the charge flip rate is considered.

Photon Fake Rate Uncertainties Four systematic uncertainties are defined for the derivation of the rate of photons mis-reconstructed as electrons. The first stems from the statistical uncertainty of the measured rate. The second uncertainty is

measured by calculating the photon conversion rate in two separate p_T bins (chosen to have roughly equal statistics). The results are shown in table 8.15, both are roughly consistent with the inclusive rate and no p_T dependence is observed. The third uncertainty is due to prompt MC normalization uncertainties used for the subtraction in the $Z\gamma \rightarrow \ell_{ID}\ell_{ID}\ell_{ID}$ control region. A 1.1 or 0.9 scale factor is applied and the rate re-derived, entirely analogously to the non-prompt uncertainty. Finally, an uncertainty due to the extrapolation of the misID rate measured in $Z\gamma$ events to a $W\gamma$ enriched region is measured. This is done by measuring a photon misID rate in $W\gamma$ and $Z\gamma$ MC and comparing the two. No strong difference was observed, see table 8.16.

p_T bin	photon conversion rate
Inclusive	0.043 ± 0.004
$p_T < 25 \text{ GeV}$	0.044 ± 0.004
$p_T > 25 \text{ GeV}$	0.041 ± 0.004

Table 8.15: The effect of p_T on the photon conversion rate. Measured using two bins of equal statistical power.

Sample	photon conversion rate
$Z\gamma$	0.16 ± 0.03
$W\gamma$	0.13 ± 0.04

Table 8.16: The photon conversion rates derived using $Z\gamma$ or $W\gamma$ MC simulated samples.

8.8.3 Theoretical Systematics

The MC simulated samples used in the fit entail theoretical systematic uncertainties. These arise from the various theory parameters used in simulating these samples. These uncertainties therefore relate to the production algorithm used for each sample. For the background samples that contribute only a small number of events, e.g. ZZ , VVV and TopX , a 15-20% cross-section uncertainty is considered in the statistical fit (see table 8.17). For the main background samples, the systematic uncertainty due to varying the parton shower, PDF and scale are considered following the recommendations of the Physics Modelling Group (PMG). Similar uncertainties

will in future be derived for the signal MC samples once alternative signal samples have been produced.

Background	Uncertainty (%)
TopX	15 [134, 135]
VVV	15 [136]
ZZ	20 [137]

Table 8.17: Cross-section uncertainties for other background.

The parton distribution function (PDF) models the internal structure of the proton, and thus affects how proton-proton events are simulated. PDF uncertainties follow from the recommendations of the PDF4LHC [138] working group. The parton shower can be modelled in different ways, as the hadronization process cannot be calculated perturbatively (see chapter 2). The scale relates to two parameters, the renormalization (μ_R) scale and factorization (μ_F) scale. These parameters are needed to solve problems with ultraviolet and infrared divergences when using perturbation theory to simulate events (to some order, e.g. LO or NLO). Variations in these two parameters are used to extract the scale uncertainties.

WWW and WH Theory Uncertainties The WWW process is split between the on-shell $W^\pm W^\mp W^\mp$ process simulated using SHERPA 2.2.2 and the $pp \rightarrow WH^* \rightarrow$ WWW process simulated using POWHEG-BOX. The scale uncertainty is derived by varying the (μ_R) and (μ_Z) scales and taking the envelope over the output histograms of observables. The PDF uncertainties are derived from similar envelopes produced by varying the PDF set used in the simulation. The PDF+ α_S error is also considered. The parton shower uncertainties are estimated using alternative samples produced with SHERPA for the WWW sample and HERWIG 7 for the WH sample. The ssWW uncertainties are estimated using the same approach as the WH samples. The WZ sample uses the same prescription as the WWW sample. Finally, ATLAS has recently measured the WWW production cross-section finding a combined signal strength of 1.61 ± 0.25 [139]. Due to the similarities between our signal region and the WWW same-sign two lepton signal region, an additional 60% uncertainty is applied to the WWW background.

ssWW and WZ Theory Uncertainties Special consideration is needed for the ssWW samples, which suffer from mis-modelling issues. The dijet mass, m_{jj} , is mis-modelled in the QCD samples for values $m_{jj} > 1$ TeV due to the shower activity and scale choice for NLO emissions. This range is, however, far removed from the signal region under consideration here. A color-flow issue in the electroweak ssWW samples also causes a significant reduction in predicted cross-sections. The impact on the normalization will be accounted for in this analysis by the dedicated ssWW normalization factor. An additional shape uncertainty is also derived to account for any further issues due to the mis-modelling.

8.9 Statistical Analysis

The data gathered by ATLAS is analysed by comparing to the various predicted signals and backgrounds. The latter are estimated using a mix of MC simulation and data-driven techniques, as described in sections 8.3 and 8.6. The fit of observed data to a given model prediction proceeds using maximum likelihood estimation. This is all implemented in a statistical framework known as HistFitter [125].

For an event counting analysis such as this, the probability of measuring a given number of events, N_{data}^i , in a bin, i , (belonging to whatever regions we fit over, e.g. signal and/or control regions) is best modelled as a Poisson distribution. The likelihood function can be written as [125]:

$$L(\mathbf{N}_{data}, \boldsymbol{\theta}^0 | \mu_{sig}, \boldsymbol{\mu}_p, \boldsymbol{\theta}) = \prod_i^{\text{bins}} P_{poiss}(N_{data}^i | \lambda_i(\mu_{sig}, \boldsymbol{\mu}_p, \boldsymbol{\theta}) \times C_{syst}(\boldsymbol{\theta}^0, \boldsymbol{\theta})) \quad (8.11)$$

where $\mathbf{N}_{data}$ is the input data and $\boldsymbol{\mu}_p$ are our normalization factors to be fit from the control regions. The nuisance parameters $\boldsymbol{\theta}$ continuously parameterise the systematic uncertainties, and are defined such that $\theta_i = 0$ returns the nominal template and $\theta_i = \pm 1$ return the +1 and -1 systematic variation templates¹⁵. These are assumed to be Gaussian¹⁶ distributed around the central values $\boldsymbol{\theta}^0$ (typically these are set to

¹⁵In general, these are asymmetric (the +1 and -1 are different). Symmetric uncertainties are also used, where only the up variations is provided. The down variation is inferred from the nominal and up variation.

¹⁶Log-normal probability distributions are used for normalization uncertainties, such as the nor-

0, but can be varied in toy experiments):

$$C_{syst}(\boldsymbol{\theta}^0, \boldsymbol{\theta}) = \prod_j^S G(\boldsymbol{\theta}_j^0 - \boldsymbol{\theta}_j) \quad (8.12)$$

where S is the total number of nuisance parameters. The Poisson distributions, P_{poiss} , are characterized by the frequency:

$$\lambda_i(\mu_{sig}, \boldsymbol{\mu}_p, \boldsymbol{\theta}) = \mu_{sig} s^i(\boldsymbol{\theta}, \boldsymbol{\mu}_p) + b^i(\boldsymbol{\theta}, \boldsymbol{\mu}_p) \quad (8.13)$$

where μ_{sig} is our signal strength (a positive scalar value), s^i is the expected yield of signal events in the bin i and b^i the expected yield of background events. If the value of $\mu_{sig} = 0$, then the signal component has been turned off. The signal yield s^i will come from the particular model under consideration (in our case, that is an MC simulated sample dependent on a particular combination of the three model parameters: m_H , f_W and f_{WW}).

The fit strategy employed in this analysis proceeds in two steps. A background-only fit is first performed to evaluate the normalization factors μ_p . The fitted values are tested by extrapolating to the validation region and examining the result. The background-only fit is performed only in the bins of the control regions and using $\mu_{sig} = 0$. If a satisfactory result is found, with no unexpected normalization factor values or nuisance parameter pulls, a model-dependent fit is then performed for each model point (combination of m_H , f_W and f_{WW}). This fit is performed in the signal and control regions whilst varying the value of μ_{sig} . In the case of a measured excess, this fit will characterize properties such as the signal strength. In the absence of any detected signal, exclusion limits can be drawn using a hypothesis test.

8.9.1 Hypothesis Test for Exclusion Limits

The hypothesis test used to establish 95% confidence level expected upper limits on the signal cross-section is the CL_s method [140]. This method begins by establish-

malization uncertainty on the WW sample, to ensure these are always positive.

ing a test-statistic:

$$q_\mu = -2 \ln \left(\frac{L(\mu_{sig}, \hat{\boldsymbol{\theta}}_{\mu_{sig}})}{L(\hat{\mu}_{sig}, \hat{\boldsymbol{\theta}})} \right), \quad (8.14)$$

This is known as the profile likelihood ratio and is generally used by the LHC collaborations [125]. The null hypothesis is that the true value of μ_{sig} corresponds to the tested value. The numerator of this expression, $L(\mu_{sig}, \hat{\boldsymbol{\theta}}_{\mu_{sig}})$ is the conditional likelihood maximised with respect to the free parameters $\hat{\boldsymbol{\theta}}_{\mu_{sig}}$ for a constant μ_{sig} . The denominator $L(\hat{\mu}_{sig}, \hat{\boldsymbol{\theta}})$ is the unconditional likelihood, maximised with respect to both $\hat{\boldsymbol{\theta}}$ and $\hat{\mu}_{sig}$ as free parameters. The major advantage of this test statistic is that, in the limit of large samples, its probability density functions (pdfs) can be calculated using approximate methods such as the asymptotic calculator in HistFitter [141].

Derivation of the 95% C.L. limits requires calculating the conditional pdf of the test-statistic given some hypothesis, $f(q_\mu | hyp)$. These can be calculated using approximate methods such as the asymptotic calculator or using Monte Carlo toy experiments. This is done for the background only hypothesis, B, and signal plus background hypothesis, S+B. Once the pdfs are known, a reference value of the test statistic q_μ^{ref} is used to calculate two values:

$$p_B = P(q_\mu < q_\mu^{\text{ref}} | B) = \int_{-\infty}^{q_\mu^{\text{ref}}} f(q_\mu | B) dq_\mu \quad (8.15)$$

$$p_{S+B} = P(q_\mu > q_\mu^{\text{ref}} | S+B) = \int_{q_\mu^{\text{ref}}}^{\infty} f(q_\mu | S+B) dq_\mu \quad (8.16)$$

The reference value, q_μ^{ref} , is either the best fit test value from the observed data (giving the observed limit), or the median of the B hypothesis pdf, $f(q_\mu | B)$ (giving the expected limit). The excluded μ_{sig} at 95% C.L is given by:

$$\text{CL}_s \equiv \frac{p_{S+B}}{1 - p_B} < 5\% \quad (8.17)$$

Exclusion contours can be given in the model phase space by computing the CL_s values for each model point and interpolating an exclusion surface.

8.9.2 Binning in the GHH Analysis

The GHH analysis is formed of two signal regions (boosted and resolved) alongside two control regions from which normalization factors are derived (the WZ and ssWW control regions). The histograms describing these regions (which are fit) are the effective mass distributions, M_{eff} . The histogram binning, is in general a free parameter. Fits in the CRs are done using a single bin. In the SRs, a brief study was performed to determine an optimal binning. A variety of binnings were examined, constrained only by having a minimum bin content of 6 events and maximum bin error of 10%. The expected upper limit for a small grid of signal models was calculated using all the different binnings. The optimal binning is chosen to have the smallest expected upper limit across the different signal samples. For the two SR the optimal binning was:

- Boosted: $(0, 600, 800, inf)$ GeV
- Resolved $(0, 300, 400, 500, 600, inf)$ GeV

8.10 Results

The initial results of the analysis fit are shown here. Unblinded results will be made public in the near future but for now only blinded results are shown (data in the signal regions is removed). As such only expected results derived from Asimov data¹⁷ are shown. The fit results are on the samples from the most recent analysis framework (r04-02).

8.10.1 Signal Acceptance and Efficiency

The signal acceptance is the number of signal events passing the phase space (p_T and η) cuts of the signal regions. The signal efficiency relates to all other cuts and reconstruction efficiencies that signal events must pass to enter the signal regions. The product of signal acceptance and efficiency gives us the fraction of the simulated events in a sample that will enter out signal regions. This quantity is calculated

¹⁷This is produced using the sum of the predicted background yields from the MC simulations. In some cases, such as the ranking plot, the predicted signal yield may be added.

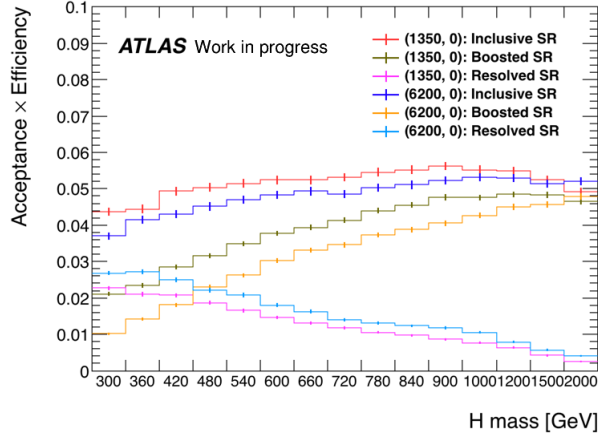


Figure 8.14: The product of signal acceptance and efficiency ($\mathcal{A} \times \epsilon$) for the inclusive, boosted-only and resolved-only signal regions. The quantity is displayed for a number of benchmark signals with values of m_H from 300 GeV to 2 TeV and two sets of the parameters (f_W, f_{WW}) : (1350, 0) and (0, 6200).

for several benchmark signals with a range of mass from 300 GeV to 2000 GeV for two value pairs of (f_W, f_{WW}) equal to (1350, 0) and (0, 6200). The result is plotted in figure 8.14.

8.10.2 Normalization factors

The background-only fit is performed to give a first value of the normalization factors $\mu_{WZResolved}$, $\mu_{WZBoosted}$ and μ_{ssWW} . The agreement between the observed data events and the predicted (background-only) events in the validation region is plotted to examine performance. This is first done plotting the predicted events with normalization factors set to 1 (pre-fit) and then with the best fit values (post-fit). For the r04-02 results, the best fit values of the normalization factors are quoted in table 8.18. The validation region pre- and post-fit plots are shown in figure 8.15.

MC process	scaling factor
$WZBoosted$	0.927 ± 0.067
$WZResolved$	0.824 ± 0.025
$ssWW$	1.441 ± 0.174

Table 8.18: The normalization factors (and corresponding uncertainties) calculated from the background only fit of the M_{eff} distributions of data and MC samples.

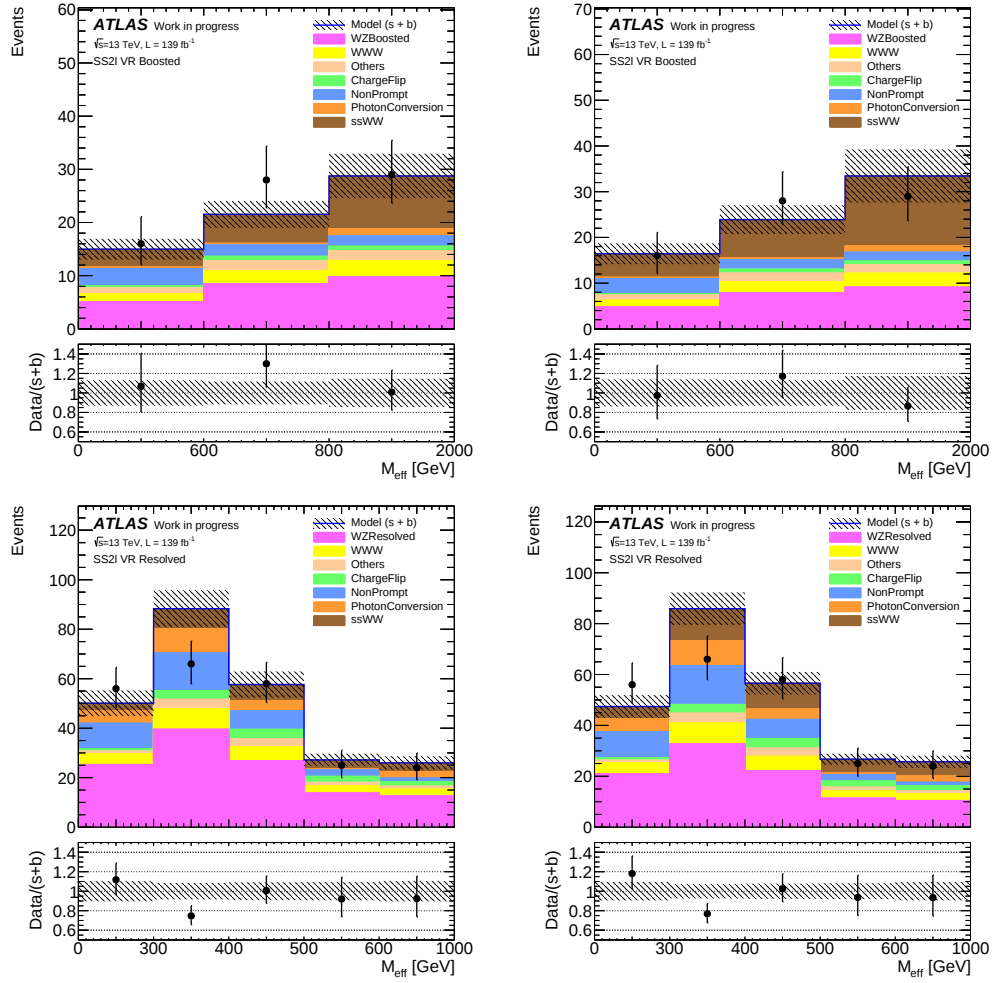


Figure 8.15: The (left) pre- and (right) post-fit distributions in the (top) boosted and (bottom) resolved validation regions. In the post-fit case, the normalization factors from the background only fit in the CR has been applied. The model is the SM background (b) plus signal (s). In this background-only fit no signal is present.

8.10.3 Blinded Exclusion Fit

The next step in the fitting procedure (the model-dependent exclusion fit) can be tested using an Asimov dataset. The real data in the SR remains blinded. Instead of running with no data in the SR, we use Asimov data where the observed data distribution is set equal to the background prediction ($\mu_{\text{sig}} = 0$). Note that real data is still used to fit the CRs. The pre- and post-fit distributions in the SR are shown in figures 8.16 and 8.17.

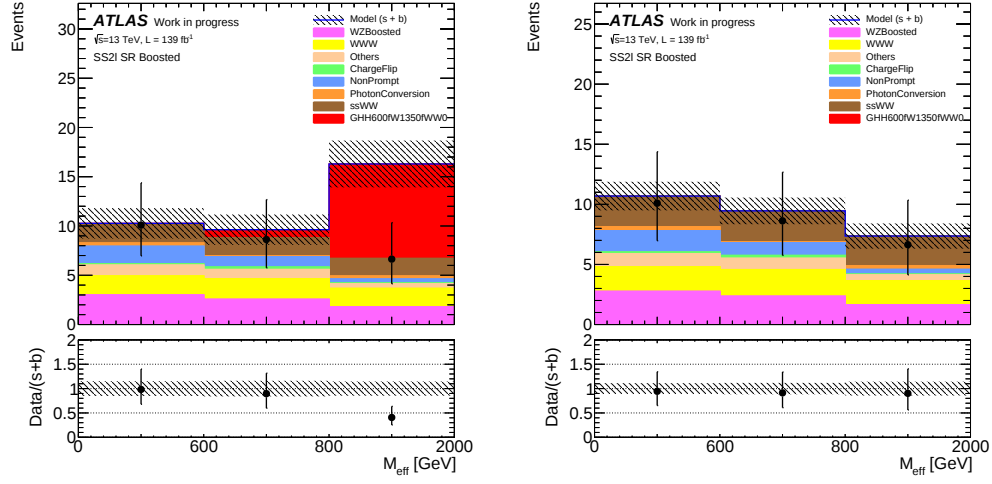


Figure 8.16: The pre- and post-fit distributions in the boosted SR from an exclusion fit using real data in the CRs and Asimov data with $\mu_{\text{sig}} = 0$ in the SRs. The model is the SM background (b) plus the GHH600fW1350fWW0 signal (s).

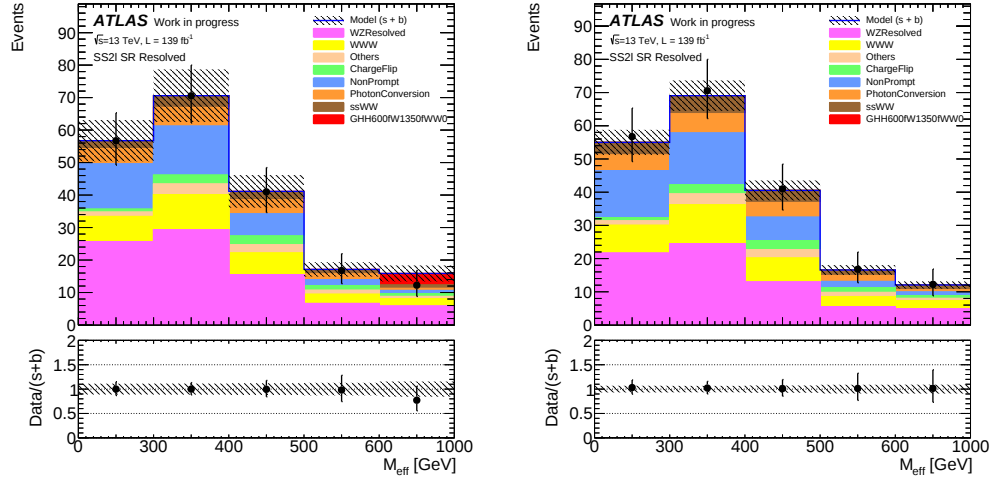


Figure 8.17: The pre- and post-fit distributions in the resolved SR from an exclusion fit using real data in the CRs and Asimov data with $\mu_{\text{sig}} = 0$ in the SRs. The model is the SM background (b) plus the GHH600fW1350fWW0 signal (s).

8.10.4 Systematic Uncertainties

The impact of the various systematic uncertainties can be assessed in two ways. The first is through so called pull plots. These show how the values of the fitted systematics θ compared to their central values θ^0 . A pull plot shows the value of the fitted nuisance parameters θ , in the plot these are termed α , as well as the other parameters such as the fitted normalization factors μ_p and the statistical uncertainties, termed γ . The pulls for the background-only fit are plotted in figure 8.18, this fit is only over the CR using real data in those regions (signal data blinded). The systematics in this fit are not pulling in any direction as expected. In previous fits, a single normalization factor for the WZ background was used, calculated from both boosted and resolved regions together, This led to the systematics all pulling in the positive direction, as the normalizations in these two regions are different. This motivated the switch to using two factors, one for the resolved region, $\mu_{WZResolved}$ and another for the boosted $\mu_{WZBoosted}$.

The pulls are also plotted for the exclusion fit with Asimov data in the SRs ($\mu_{sig} = 0$), and real data in CRs in figure 8.19. The correlation matrix of all the nuisance parameters can also be plotted. An example is in figure 8.20 where the nuisance parameters with off-diagonal correlations greater than 0.01 are plotted. There do not seem to be any large correlations between nuisance parameters, as is assumed in the fitting procedure [125].

A further way to visualise the impact of the systematics is through systematic ranking plots. In addition to the pulls, these plots show the impact of individual systematics on the fitted parameter of interest (μ_{sig}). The impact is calculated by varying one systematic parameter at a time to its plus one and minus one sigma values, and measuring the change in the fitted value of μ_{sig} (from its nominal fit value). The systematics are then ranked by their largest absolute impact. The top 20 nuisance parameters by impact are shown in the ranking plot in figure 8.21. For this plot, an exclusion fit is performed using Asimov data in the SR but using the expected background and signal yield (i.e. the Asimov data is calculated using $\mu_{sig} = 1$). The systematic uncertainties with the largest impact on the measured signal strength are

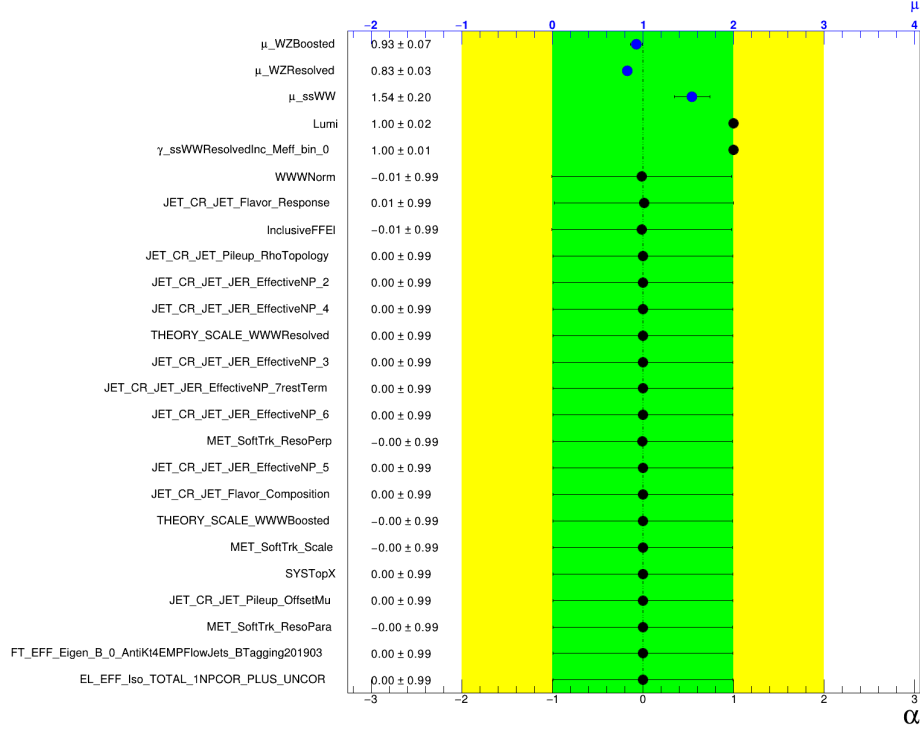


Figure 8.18: The fitted value of the nuisance parameters for the top 20 largest systematic pulls, normalization factors and statistical errors. These values are for the background-only fit. For the naming conventions, refer to table 8.13.

the PDF and scale theoretical uncertainties on the signal sample, followed by the WWW normalization uncertainty. The signal theoretical systematics directly impact the signal shape and normalization, it is thus expected that they rank highly. The WWW normalization uncertainty is very large (60%) to account for the recent WWW cross-section measurement value (see section 8.8.3). B-tagging uncertainties and the scale factor uncertainties for the W-tagger also feature prominently.

8.10.5 Expected Results in the Inclusive Fit

For each model point in our signal grid, the expected upper limit is calculated using the CLs method. The asymptotic calculator is used in this search. The signal strength is varied until the expected CLs falls below 0.05, the value at which this happens is the expected upper limit. The p value versus signal strength is

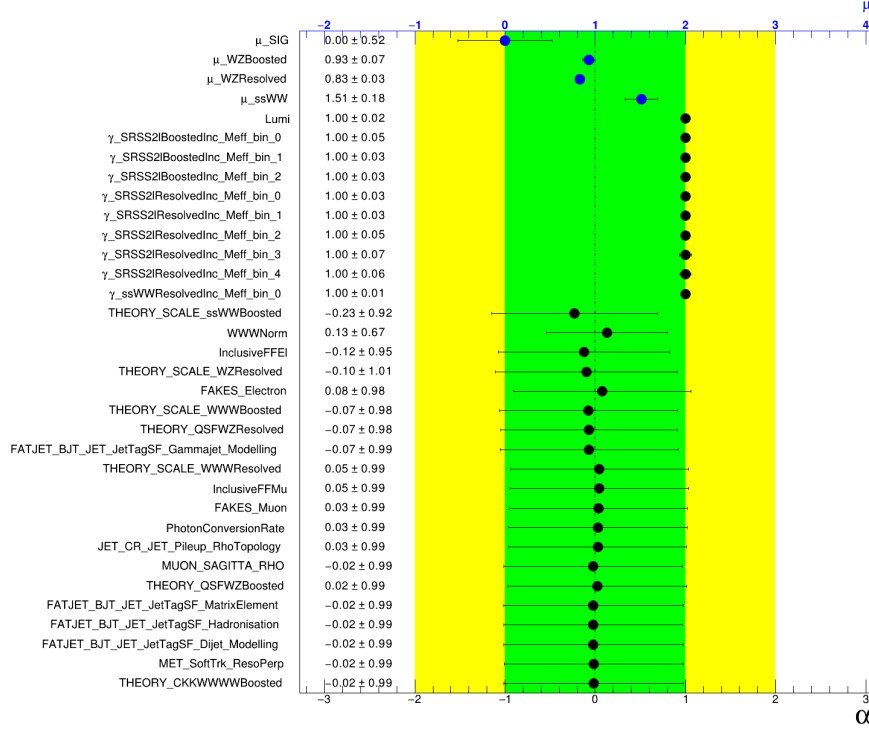


Figure 8.19: The fitted value of the nuisance parameters for the top 20 largest systematic pulls, normalization factors and statistical errors. These are the values from the exclusion fit using Asimov data ($\mu_{\text{sig}} = 0$) in the SR. For the naming conventions, refer to table 8.13. The γ parameters are the statistical uncertainties from the different bins.

plotted for a benchmark signal¹⁸ in figure 8.22. There are 172 different signal points which were studied. The results for these are shown in a graphical format. Figure 8.23 shows all the expected upper limit values for signals with a fixed $m_H = (300, 600, 900)$ GeV and varying values of f_W and f_{WW} .

Expected exclusion contours at 95% confidence level are also drawn. These are calculated for each signal point by running a hypothesis test that $\mu_{\text{sig}} = 1$ and storing the resulting p value. The results for each point can be interpolated in the plane defined by two model parameters, and contours then drawn where the surface is equal to 0.05. The interpolation is done in the (f_W, f_{WW}) plane using the signal points with the fixed mass values of $(300, 600, 900)$ GeV to produce the plots in

¹⁸The benchmark signal is termed GHH600fW2130fWW0: this corresponds to $m_H = 600$ GeV, $f_W = 2130$, $f_{WW} = 0$.

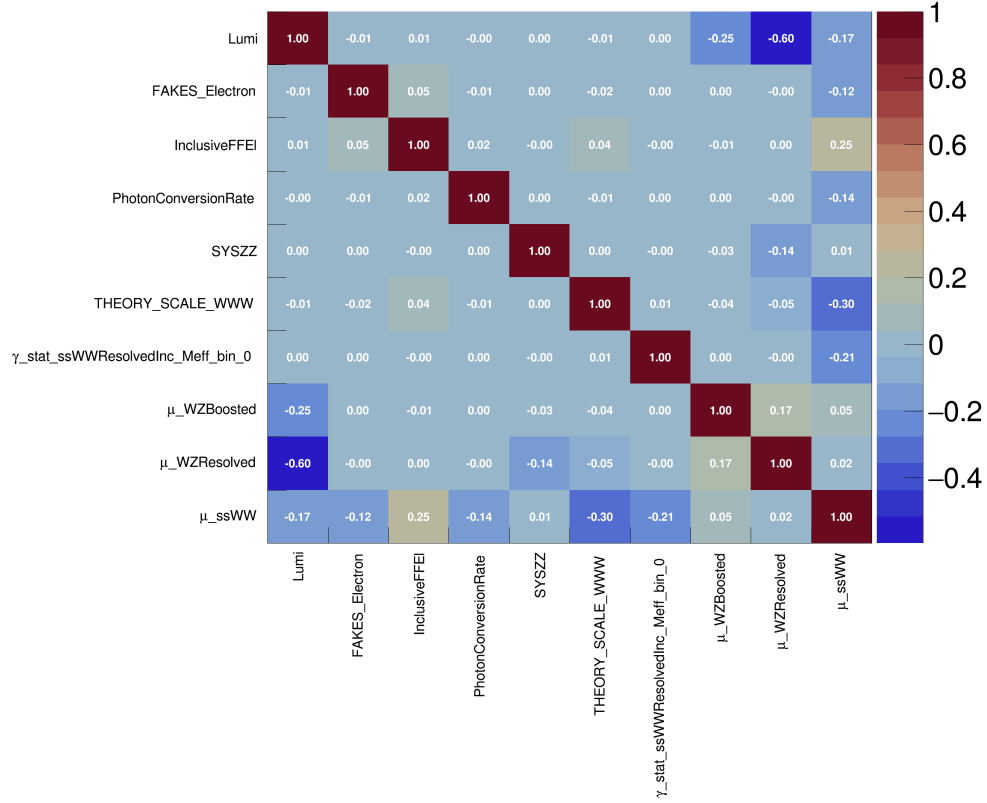


Figure 8.20: The reduced correlation matrix of all nuisance parameters with an off-diagonal value greater than 0.1 (the remaining 170 nuisance parameters fell under this threshold). These are the values from the exclusion fit using Asimov data ($\mu_{\text{sig}} = 0$) in the SR.

figure 8.24. These plots exhibit a radial symmetry, this is expected from the form of the Lagrangian in equation (2.35). The exclusion limits shift outward as mass increases, this is because for a fixed value of (f_w, f_{WW}) , the SR yields decrease as mass increases (see Table 8.19). The expected upper limit on the signal cross-section are compared to the theoretical cross-sections in figure 8.25. This was done for (f_w, f_{WW}) of (1350, 0) and (0, 6200) over a range of heavy Higgs mass values. For three of the very high mass points¹⁹ the fit does not converge, as such the plots do not extend to include these. This occurs due to the low signal yields in these high mass samples, see Table 8.19.

¹⁹GHH2000fW1350fWW0, GHH1500fW0fWW6200 and GHH2000fW0fWW6200.

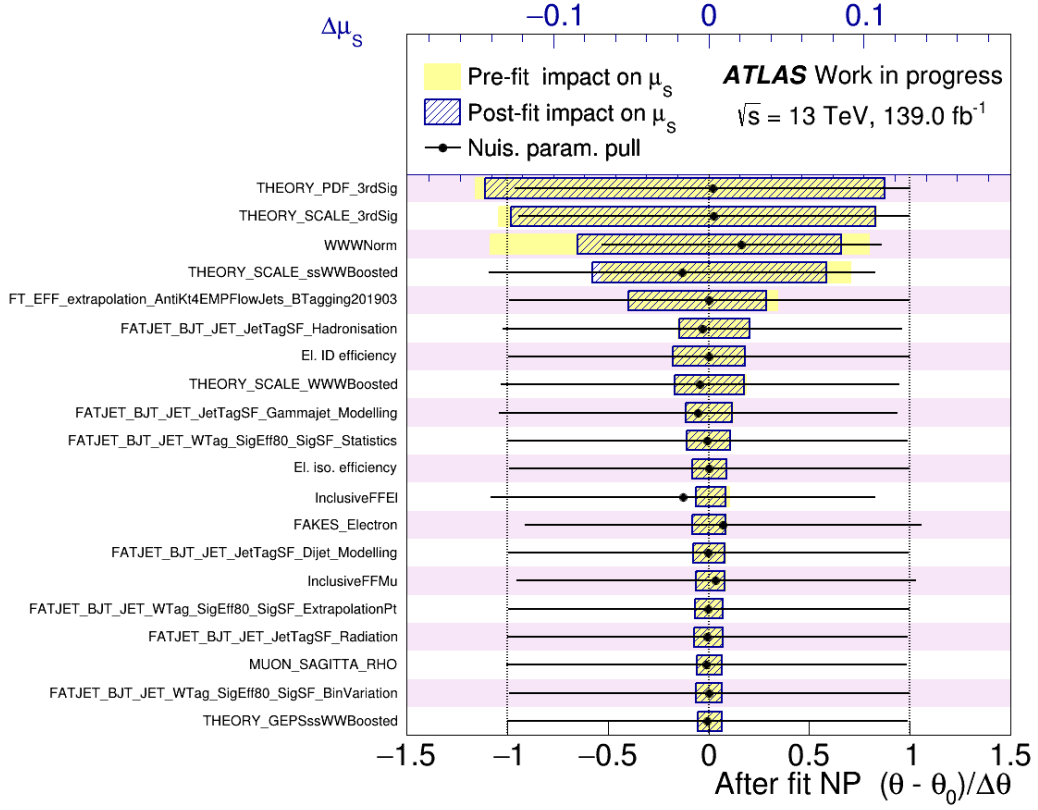


Figure 8.21: The systematic ranking plot from the exclusion fit with Asimov data ($\mu_{sig} = 1$) in the SR and real data in the CR. The top 20 by impact nuisance parameters are shown. For the naming conventions, refer to table 8.13.

Mass	(f_w, f_{WW})	Boosted SR Yield	Resolved SR Yield
300	(1350,0)	10.26	11.04
600	(1350,0)	10.60	4.13
900	(1350,0)	7.51	1.35
1500	(1350,0)	2.31	0.20
2000	(1350,0)	0.78	0.044
1500	(0,6200)	1.01	0.13
2000	(0,6200)	0.25	0.022

Table 8.19: The signal yields in the Boosted and Resolved SR for various different GHH mass samples.

8.10.6 Resolved and Boosted-only Sensitivities

The sensitivity of the resolved and boosted signal regions was explored by re-running the fit twice using first only the resolved SR, second only the boosted SR. Exclusion contours in the (f_w, f_{WW}) were created from the resolved-only and boosted-only fits and are shown in figures 8.26 and 8.27 respectively. The resolved-

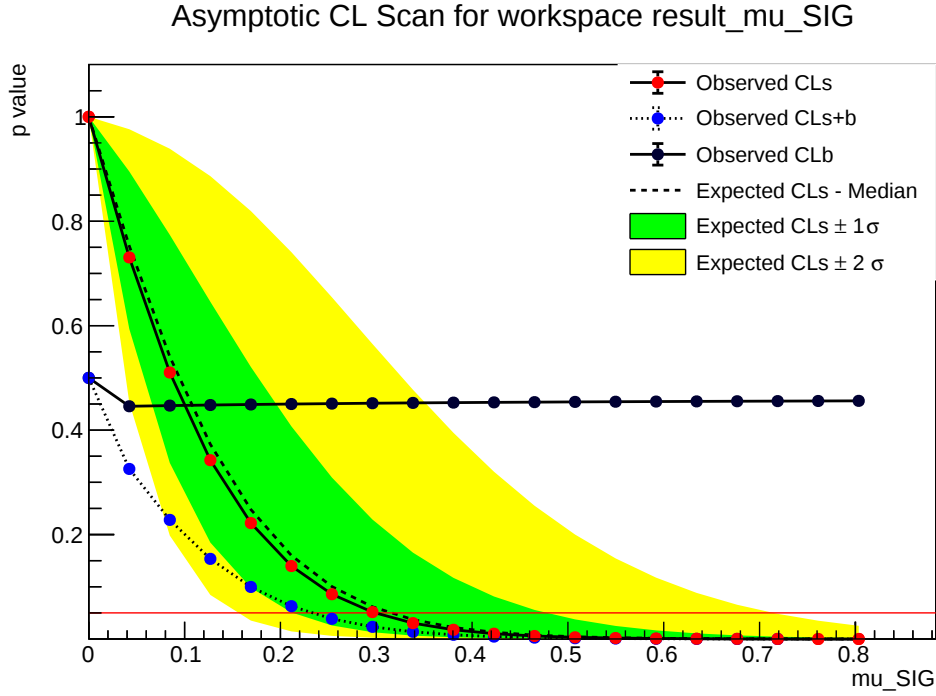


Figure 8.22: The expected upper limit of signal GHH600fW2130fWW0 is 0.34. The observed values are calculated using Asimov data in the signal region as the real data remains blinded. This model has a theoretical cross-section of 0.04731 pb.

only fit performs much worse than the inclusive fit, particularly for the higher GHH mass samples, as can be seen by the outward shift of the exclusion limits when compared to figure 8.24. The boosted-only fit in contrast appears to perform almost as well as the inclusive fit for the higher mass samples $m_H = 600$ and 900 GeV. This conclusion is further supported by plots of the expected upper limits on the signal cross-section against mass for (f_w, f_{WW}) equal to $(1350, 0)$ and $(0, 6200)$. These plots are shown for the resolved-only and boosted-only fits in figures 8.28 and 8.29 respectively. The resolved-only fit has a much lower sensitivity than the inclusive whilst the boosted-only fit performs similarly, as can be seen by comparison to figure 8.25. Nevertheless, some sensitivity is gained from the inclusion of the resolved channel at the lower values of m_H .

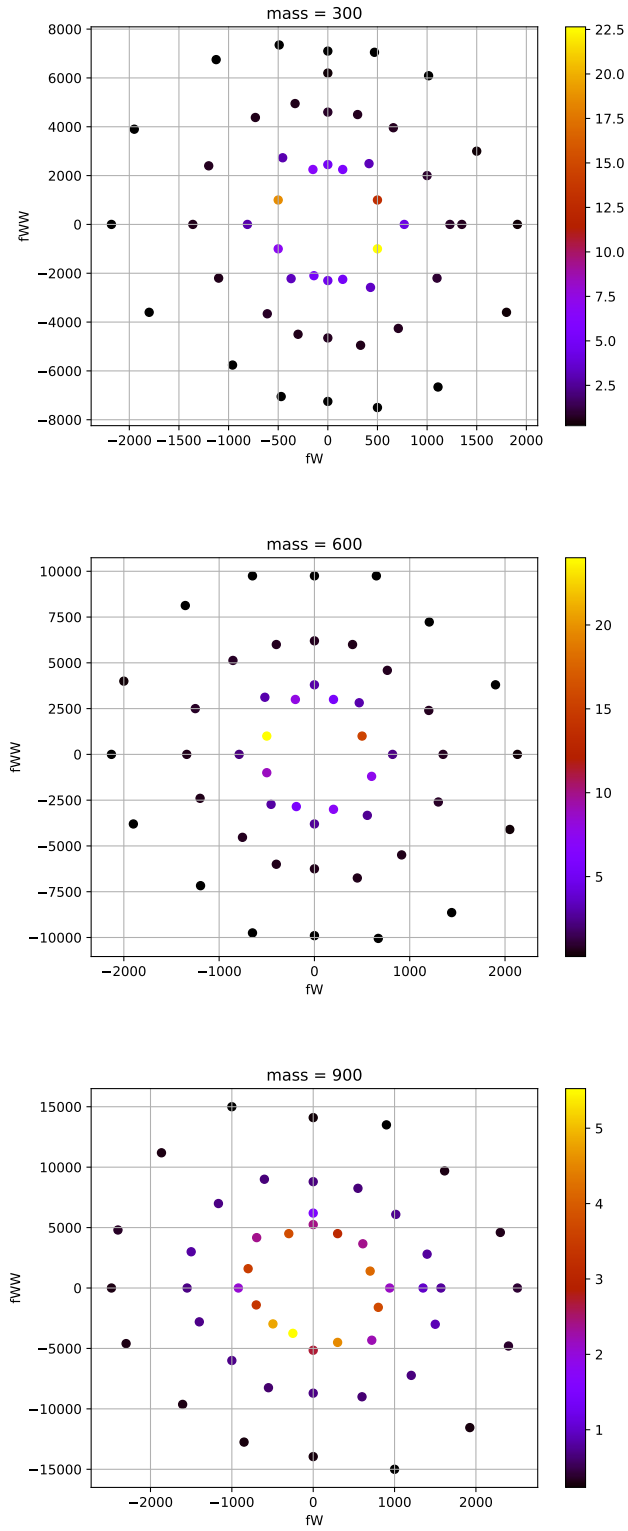


Figure 8.23: The expected upper limit on signal strength for signal samples with 300, 600 and 900 GeV mass and a range of f_W and f_{WW} parameters.

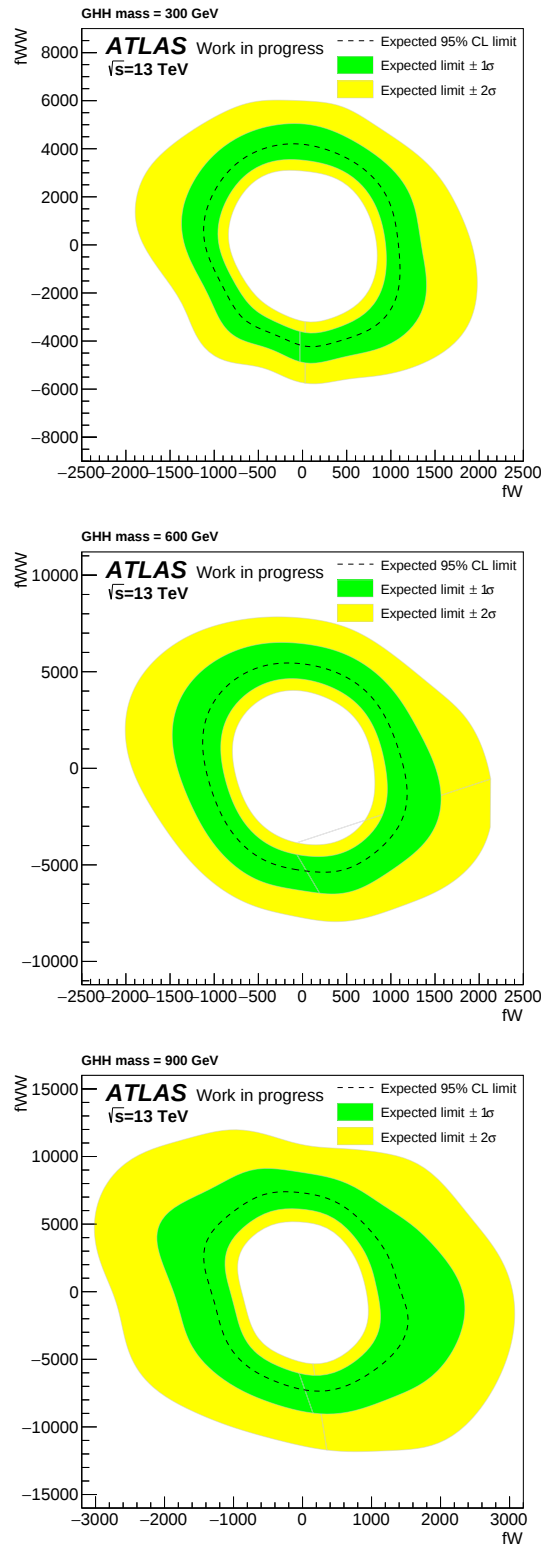


Figure 8.24: The expected exclusion contours in the fW vs fWW plane for 300, 600 and 900 GeV heavy higgs. All parameter points outside of the contours are excluded.

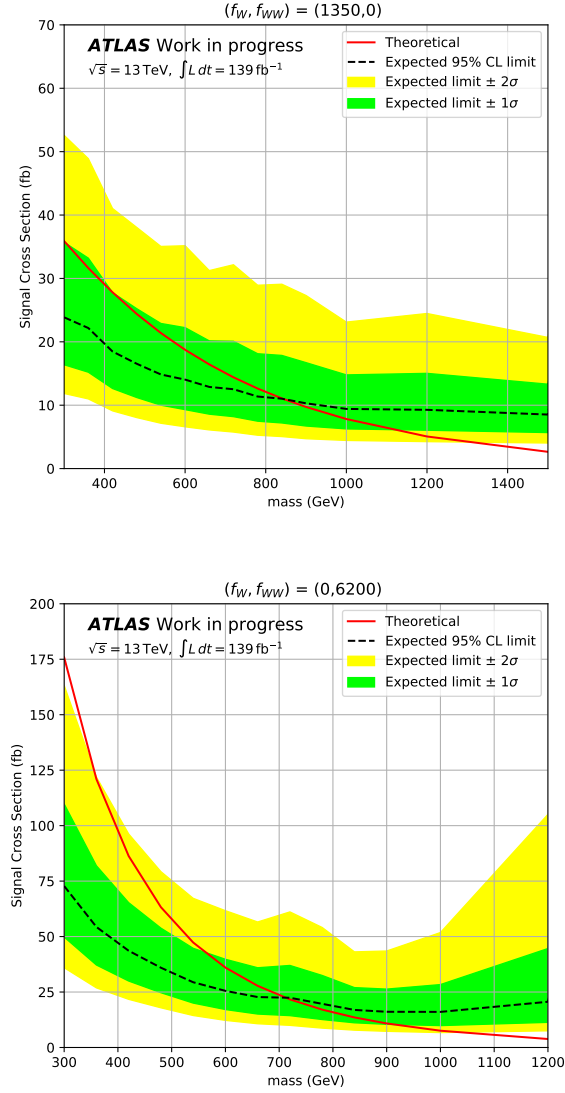


Figure 8.25: The expected upper limit on the cross-section versus the heavy Higgs mass for a model with (top) $f_W = 1350$ and $f_{WW} = 0$ and (bottom) $f_W = 0$ and $f_{WW} = 6200$. The theoretical cross-section, calculated from simulation, is drawn in red.

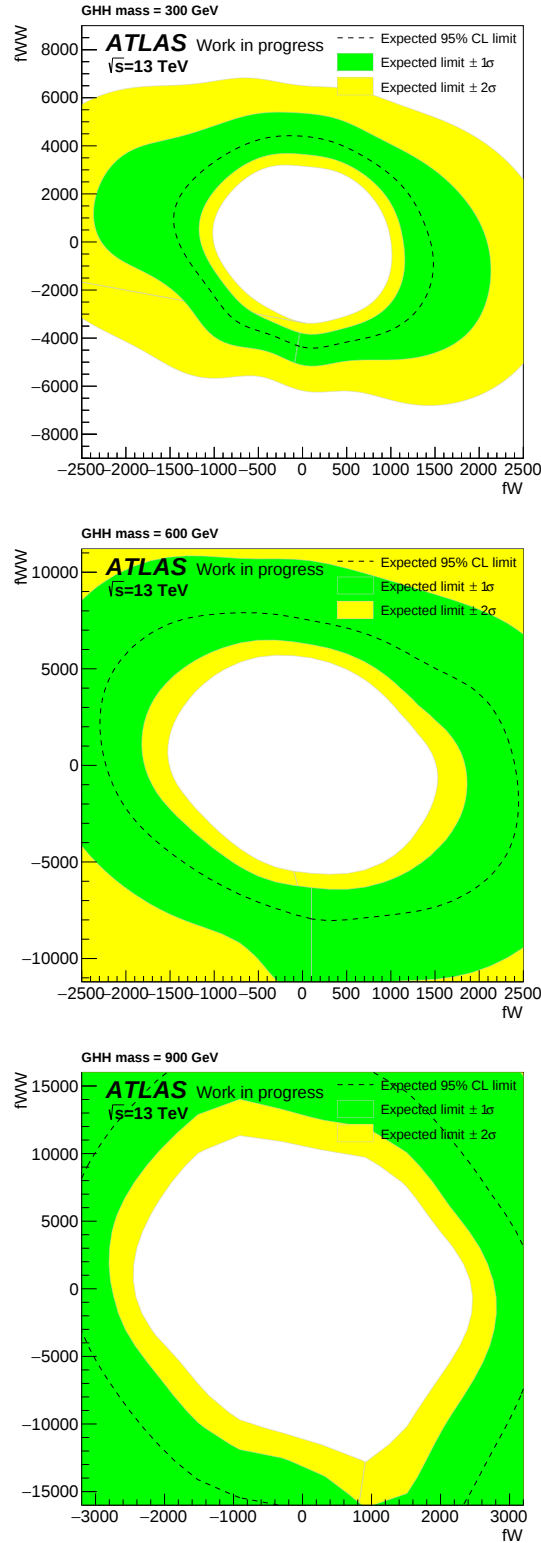


Figure 8.26: The expected exclusion contours with heavy Higgs mass of 300 GeV, 600 GeV and 900 GeV for resolved-only fit.

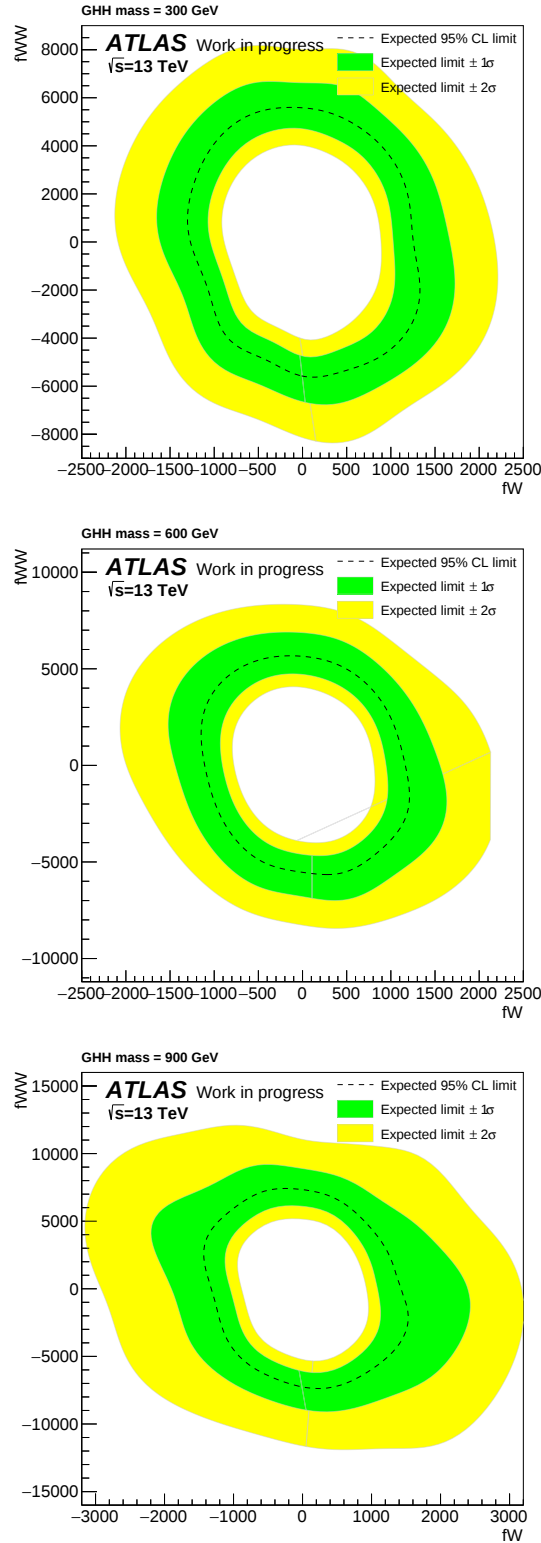


Figure 8.27: The expected exclusion contours with heavy Higgs mass of 300 GeV, 600 GeV and 900 GeV for boosted-only fit.

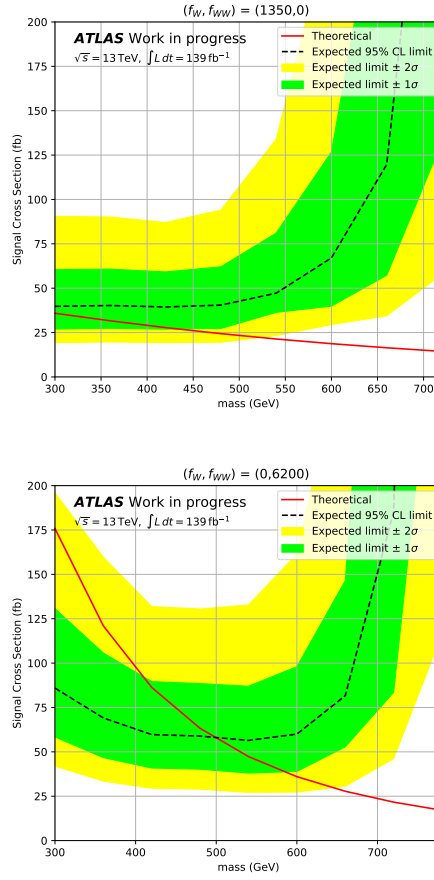


Figure 8.28: The expected exclusion contours in the mass vs cross-section space for (top) $f_W = 1350$ and $f_{WW} = 0$ and (bottom) $f_W = 0$ and $f_{WW} = 6200$ for resolved-only fit.

8.11 Summary

The GHH analysis searched for a fermion-phobic heavy Higgs in the same-sign two lepton channel. This is well motivated as a signal of heavy Higgs associated vector production. It is particularly sensitive in the case of an effective six-dimensional operator, that can be considered a stand-in for new physics at some heavier scale. The two same-sign lepton signal is associated with two resolved small-R jets or a single boosted large-R jet, covering a wide phase space. The backgrounds were estimated using a combination of data-driven techniques and MC simulation. Estimates for the upperlimit on the signal strength were calculated over a large phase space of three variable parameters: the Higgs mass, and the coupling strengths f_W and f_{WW} . Exclusion contours for these three parameter spaces are drawn. For the

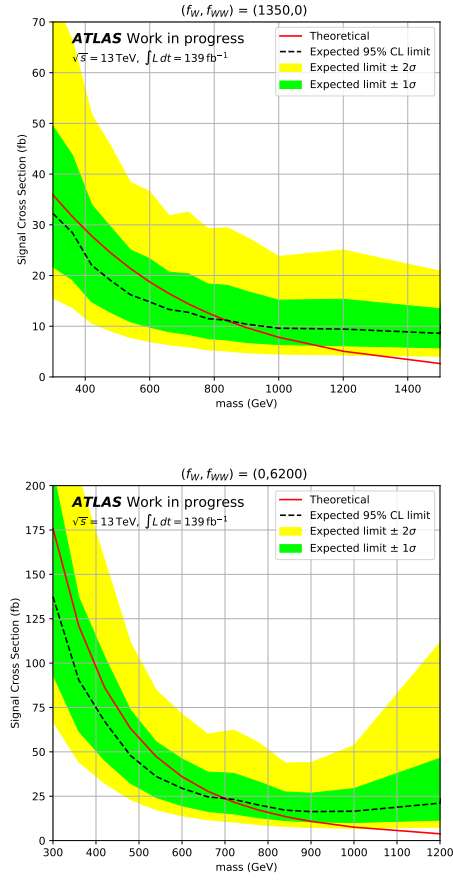


Figure 8.29: The expected exclusion contours in the mass vs cross-section space for (top) $f_W = 1350$ and $f_{WW} = 0$ and (bottom) $f_W = 0$ and $f_{WW} = 6200$ for boosted-only fit.

coupling parameters $(f_W, f_{WW}) = (0, 6200)$, the largest expected GHH mass that can be excluded at 95% C.L. is 700 GeV. For $(f_W, f_{WW}) = (1350, 0)$, this value is around 800 GeV.

Chapter 9

Conclusion

The ATLAS physics program aims to continue making ever more precise measurements of SM predictions and search for new BSM physics. The advent of the high-luminosity LHC brings not only the promise of increased statistics but greater challenges relating to the the processing of big data in high pile-up conditions. The current algorithmic tools, such as b-tagging, will require upgrades to keep pace in this new regime.

The continued development of the JetFitter algorithm was a key focus of this work. Neutral track veto and reconstruction algorithms were implemented with the aim of improved tagger efficiency. The impact of these was measured on the high-level b-tagger MV2. No major improvement to light-jet or c-jet efficiency was seen at the working points used by the collaboration. Further efforts were carried out in the refactoring and documentation of the JetFitter code. These will form an important part of future ATLAS software development, with implementation already pushed to the latest ATLAS code-base. This refactored code will greatly aid future code development efforts for JetFitter.

Research was carried out into a new Machine Learning version of the topological reconstruction used in JetFitter. The use of an RNN in secondary and tertiary vertex reconstruction was investigated on a toy jet model. The challenges of this RNN regression task were made apparent. Various performance measures were used to assess the success of this model. The RNN fitter showed a fairly good robustness to noise, it can particularly improve upon simple analytic methods at

higher noises. However, difficulties remained in comparing such a method to current ATLAS b-taggers. A major issue is the inability of an RNN to assign tracks to a particular vertex. The method nevertheless could be expanded in future. Further improvements could be made to the algorithm itself. The RNN could prove useful not as a stand-alone tool but in conjunction with other algorithms such as the Kalman Filter in JetFitter. More generally, other machine learning models, such as graph neural networks, could be considered instead of the RNN.

A search for generic heavy versions of the Higgs bosons was carried out. The motivations for such a model and the methods used to perform the search were presented. A likelihood fit was performed on data and simulated samples to produce expected upper limit estimates and 95% C.L. exclusion contours in the model parameter space. The analysis is currently awaiting approval to unblind the signal sample data, and is expected to be made public within the coming months. In the absence of a signal, this analysis will help to narrow down the phase space of potential new discoveries at the energy frontier.

Appendix A

Neutral Tracks in JetFitter

This appendix describes the ATLAS specific code used to reconstruct neutral tracks in JetFitter. As such a lot of terminology specific to the ATLAS EDM is used. This is intended more as a guide to future ATLAS developers attempting to understand the neutral track code.

A.1 Overview

Charged particle tracks are currently stored as `xAOD::TrackParticle` objects, the analogous neutral particle object is the `xAOD::NeutralParticle`. To create a NeutralParticle we must first calculate its track parameters and a reference position. The reference position will be the position of the two-track vertex candidate (known as a V0 candidate in JetFitter). The two charged tracks are used to determine the NeutralPerigee using the `NeutralParticleParameterCalculator` (found under `TrkVertexFitterUtils`). The `NeutralParticleParameterCalculator` required substantial overhaul to ensure compatibility with the new EDM. The neutral particle creation is called inside the main vertex finding package of JetFitter (`InDetImprovedJetFitterVxFinder`), around phase 5. The neutrals are then added to `tracksToUseInFirstFit` if they pass certain selection criteria.

The JetFitter fitting routines required some small modifications to handle neutral particles correctly. This happens at one point in `JetFitterRoutines` and in the `KalmanVertexOnJetAxisSmoother`. The rest of the fit can then operate smoothly as the new EDM interfaces well to the old EDM still used by JetFitter.

Major changes involving the storage of neutral particles used in the fit were made. JetFitter saves the fitted decay topology using BTagVertex objects, these objects contain links to the tracks used to fit to them. When a neutral track is present some modification is required. The objects themselves were modified, these are all found under xAOD/xAODBTagging. These objects are created in the BTagSecVertexing module, which required substantial modification for this. Finally the store-gate containers (NeutralParticleContainer) must be created in the StandAloneJetBTaggerAlg to ensure correct ownership.

The last set of changes are to the JetFitterVariablesFactory, where the neutral tracks are now used alongside charged tracks to create the JetFitter output variables.

A.2 List of Modified Files

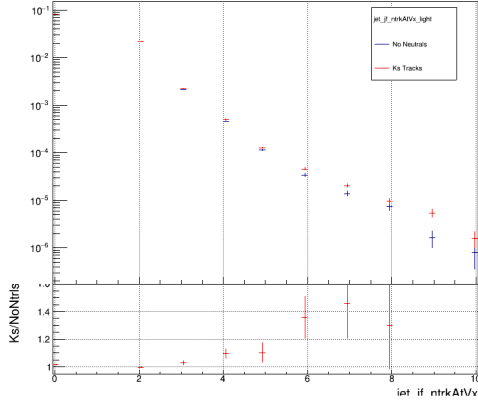
- InnerDetector/InDetRecTools/InDetSecVxFinderTool/*/
InDetImprovedJetFitterVxFinder*
- InnerDetector/InDetRecTools/InDetSecVxFinderTool/*/InDetJetFitterUtils*
- Event/xAOD/xAODBTagging/*/*BTagVertexAuxContainer_v1*
- Event/xAOD/xAODBTagging/*/*BTagVertex_v1*
- PhysicsAnalysis/JetTagging/JetTagAlgs/Btagging/*BTagSecVertexing*
- PhysicsAnalysis/JetTagging/JetTagAlgs/Btagging/*/
StandAloneJetBTaggerAlg*
- PhysicsAnalysis/JetTagging/JetTagTools/src/JetFitterVariablesFactory.cxx
- Tracking/TrkVertexFitter/TrkJetVxFitter/src/JetFitterRoutines.cxx
- Tracking/TrkVertexFitter/TrkJetVxFitter/src/
KalmanVertexOnJetAxisSmoother.cxx
- Tracking/TrkVertexFitter/TrkVertexFitterUtils/src/
NeutralParticleParameterCalculator.cxx

A.3 List of Identified Bugs

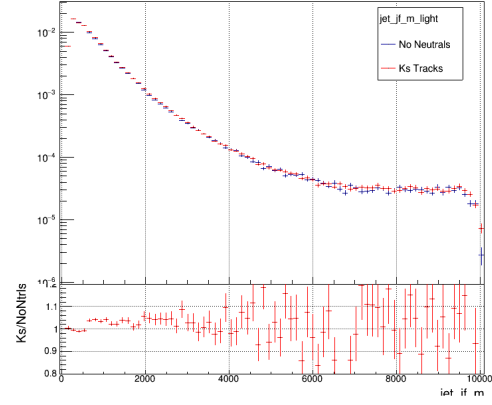
- KalmanVertexOnJetAxisSmoother: creating perigeeAtVertex not neutralPerigeeAtVertex for neutral tracks
- JetFitterVariablesFactory: Not storing mass of neutral (K_s) tracks
- JetFitterRoutines: Not correctly casting to NeutralPerigee
- BTagSecVertexing: Neutral track info storage separate to charged, not casting to NeutralPerigee
- VOs with negative eigenvalues (not fixed, just veto these candidates)

A.4 Variable Plots

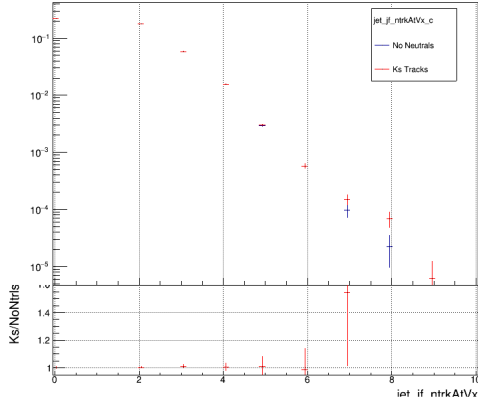
The following are plots of two of the output jetfitter variables for a simulated $t\bar{t}$ sample. The effect of including the reconstructed K_s tracks on the output variables can be seen in figure A.1 (assuming the K_s mass for all neutral tracks).



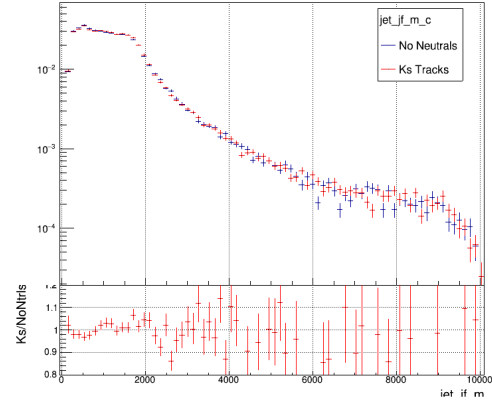
(a) light-jets



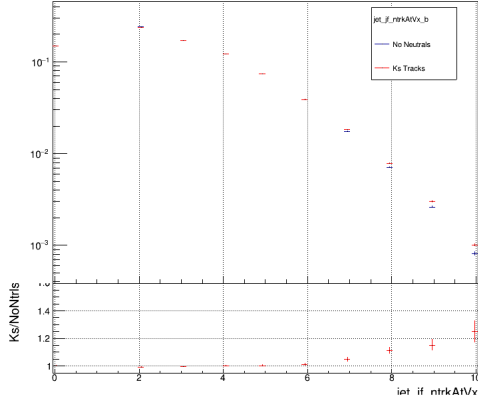
(b) light-jets



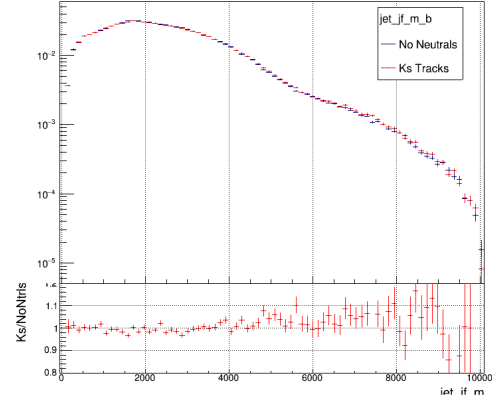
(c) c-jets



(d) c-jets



(e) b-jets



(f) b-jets

Figure A.1: The effect on the number of tracks associated to secondary vertices in (a) light-jets (c) c-jets and (e) b-jets and the effect on the invariant mass of all tracks associated to secondary vertices in (b) light-jets (d) c-jets and (f) b-jets when using reconstructed K_s tracks in JetFitter fit. The blue histogram is with no neutral tracks used in the fit (but neutral decay vertices are vetoed), the red is with the addition of K_s tracks to the fit. The ratio of the red histogram over blue histogram is plotted.

Appendix B

JetFitter Refactor

This appendix covers the technical detail relating to the refactoring of the JetFitter code. This refactor was carried out in release 21 of Athena. It was not merged to the latest version of release 21, but forms the basis for the release 22 of Athena currently used.

The main focus of the refactor was the `InDetImprovedJetFitterVxFinder` module. This file contains all the code relating to the track selection and vertex merging procedure of JetFitter. These are both long and complex procedures, hence the file is around 2000 lines of code. A single method called `doTheFinding` makes up the bulk of this. It is formed of nine phases. The first two separate primary and secondary vertex tracks. Phase 3 creates the two track vertices which are then analysed as potential neutral tracks in phases 4 and 5. Phase 6 creates the `tracksToUseInFirstFit` from the good two-track vertices and phase 7 finds the `tracksToUseInSecondFit`. Phase 8 passes to the vertex fitting and merging procedure and phase 9 is just storing the result.

The entire track selection procedure is very long (phases 1-7 above). It can broadly be divided into the separation of primaries, the creation of two-track vertices, and the analysis of these V0s to create the `tracksToUseInFirst/SecondFit`. The refactor thus aimed to create three separate modules for these processes: `JetFitterTrackSelectorTool`, `JetFitterTwoTrackVtxFinderTool`, `JetFitterV0FinderTool`. Finally, it was decided that to separate the fit and merge procedure out into its own module `JetFitterMultiStageFit`. This could then be replaced by some other fitting

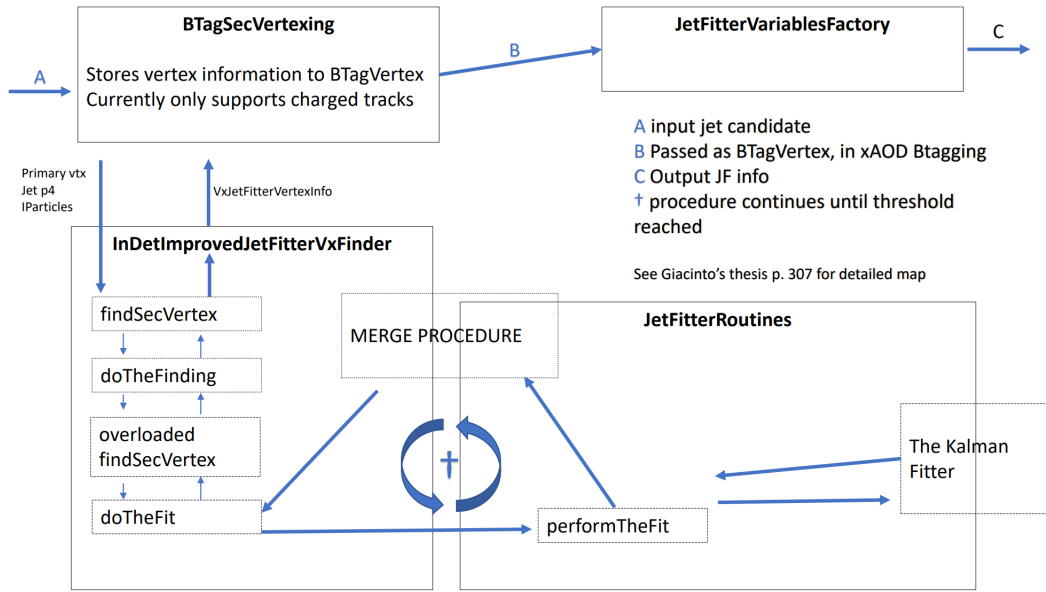


Figure B.1: Map of the JetFitter code structure showing the four main modules. BTagSecVertexing is part of the main b-tagging algorithm and directs JetFitter to find its secondary vertices by calling InDetImprovedJetFitterVxFinder. The resulting decay chain topology is passed to JetFitterVariablesFactory to create the variables for high-level taggers. The flow of information is shown by arrows. Key methods within modules are shown. This map is not an exhaustive description but rather meant to show the important information only.

algorithm in the future.

The final InDetImprovedJetFitterVxFinder file becomes much simpler to understand. It is now just directing the track selection and vertex fitting procedures, interfacing to the appropriate modules in order.

The refactor effort also added some improvements to the code. The use of smart pointers, new to C++ 11, was implemented. Some bugs to do with object ownership were uncovered and fixed. An explicit track ordering by p_T was implemented as the fitting procedure is dependent on track order. This ordering is the same as that used in the old JetFitter, but is now made explicit in the code.

B.1 JetFitter Code Maps

The code maps in figures B.1 - B.5 describe the JetFitter code before the refactoring takes place. The final refactored code is mapped in figure ??.

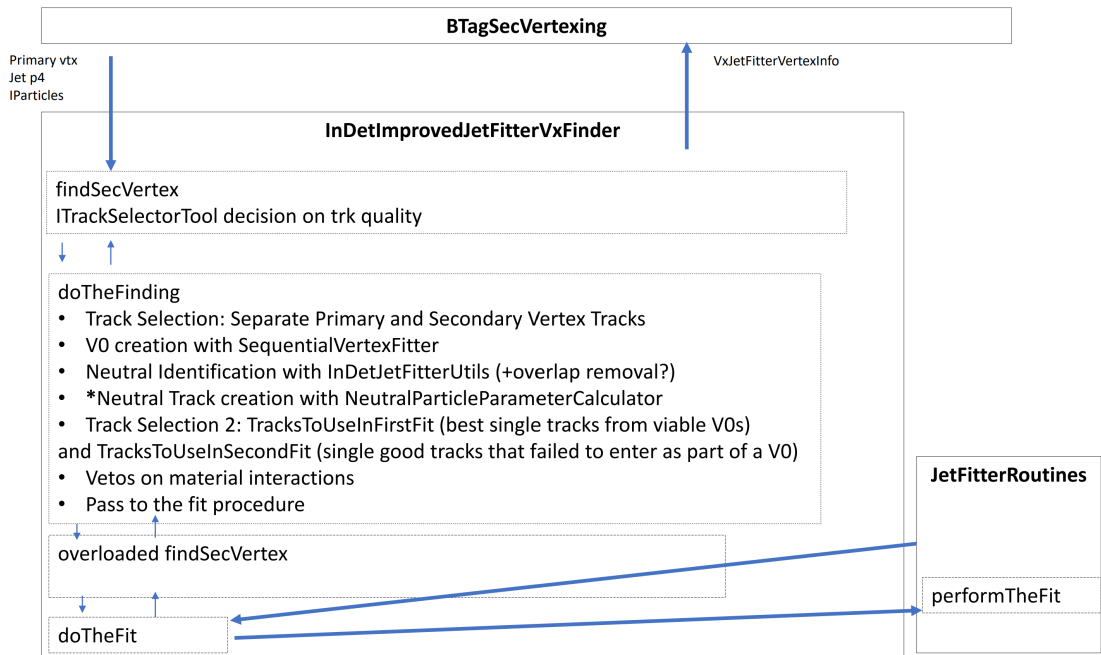


Figure B.2: Map of the JetFitter code structure but with the `InDetImprovedJetFitterVxFinder` overloaded `findSecVertex` method shown in more detail.

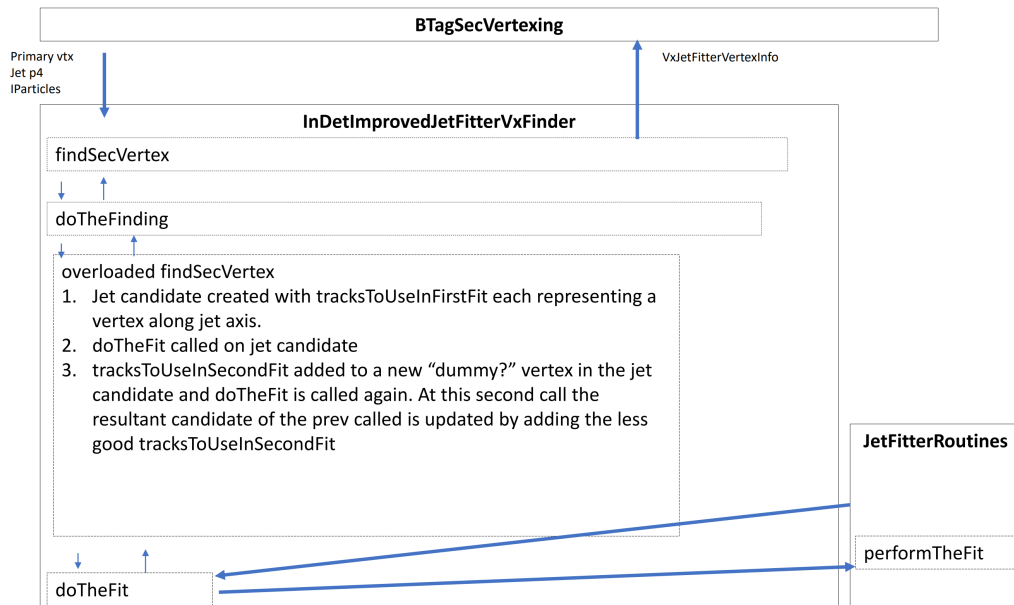


Figure B.3: Map of the JetFitter code structure but with the `InDetImprovedJetFitterVxFinder` method `doTheFit` shown in more detail.

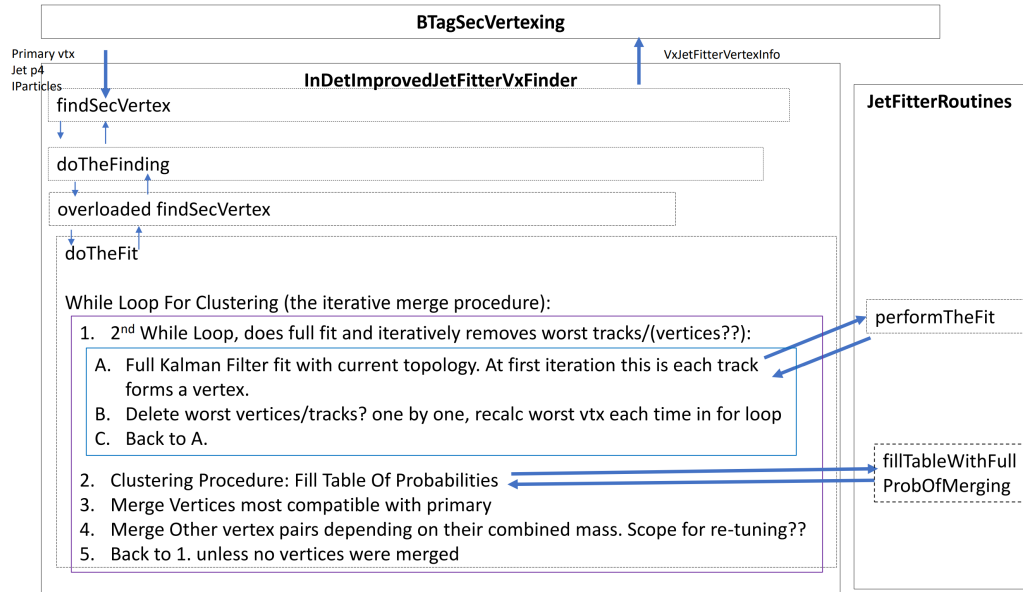


Figure B.4: Map of the JetFitter code structure but with the InDetImprovedJetFitterVxFinder module shown in more detail.

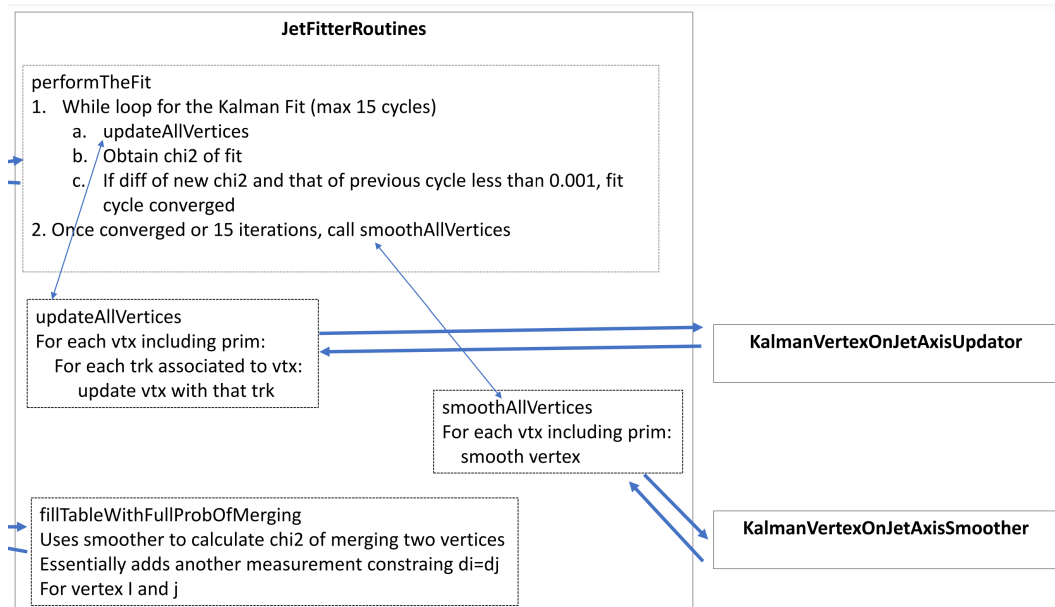


Figure B.5: Map of the JetFitter code structure but with the JetFitterRoutines module shown in more detail.

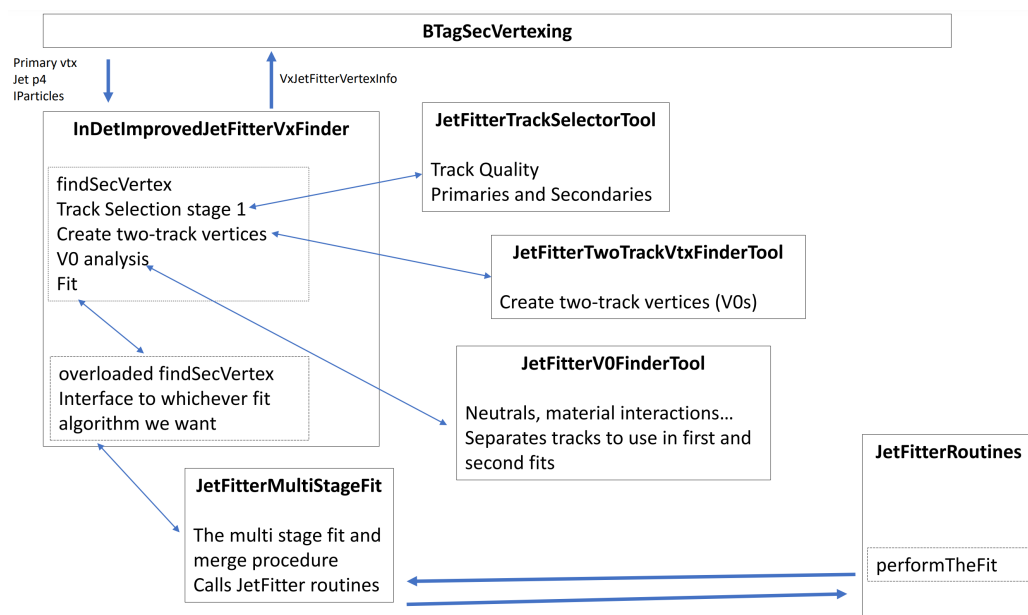


Figure B.6: Map of the refactored JetFitter code structure. **InDetImprovedJetFitterVxFinder** has been greatly simplified. Four new modules have been created.

Appendix C

Error Propagation for a Single Track

To fix a simple measure of the maximum $\sigma_{d_{min}}$, we could consider each term individually and take the maximum over the domains $\theta \in [0, \pi]$, $\phi \in [0, 2\pi]$, $\phi_q \in [0, 2\pi]$:

$$\max(A) = \frac{\sqrt{(2)}}{2} \quad (C.1)$$

$$\max(B) = 1 \quad (C.2)$$

However, problems arise with the partial derivative terms. These contain divergences in the domain of interest. Intuitively, these relate to the case where the track and jet axis are almost parallel (or anti-parallel). A small change in theta or phi can lead to a large change in d_{min} . If these divergences are ignored, it is still possible to get some measure of the maximum value by sampling sparsely in the (ϕ, θ, ϕ_p) space and calculating the derivatives at each sample point. For each angle, 100 equally spaced values are sampled forming a 3D grid with 10,000 total points. The derivatives are then calculated using numpy's gradient method. The sample maximums are:

$$\max_s\left(\frac{\partial A}{\partial \phi}\right) = 15.5 \quad (C.3)$$

$$\max_s\left(\frac{\partial A}{\partial \theta}\right) = 22.0 \quad (C.4)$$

$$\max_s\left(\frac{\partial B}{\partial \phi}\right) = 3.70 \quad (C.5)$$

$$\max_s\left(\frac{\partial B}{\partial \theta}\right) = 21.9 \quad (C.6)$$

	Baseline	High Noise
$\sigma_{d_{min}}^{max}$	21 μm	1540 μm
$\sigma_{d_{min}}^{median}$	5.16 μm	51.6 μm

Table C.1: The thresholds errors estimated from uniform samples across the domain of the function for $\sigma_{d_{min}}$.

The maximum values can change extremely quickly depending on how fine a sampling is used. To get around this, a different measure based on the sample median (of the absolute value) could be used. 50% of the sampled points have an absolute value less than this number. The sample medians are calculated as:

$$median_s(A) = 0.354 \quad (C.7)$$

$$median_s(B) = 0.376 \quad (C.8)$$

$$median_s(|\frac{\partial A}{\partial \phi}|) = 0.193 \quad (C.9)$$

$$median_s(|\frac{\partial A}{\partial \theta}|) = 0.507 \quad (C.10)$$

$$median_s(|\frac{\partial B}{\partial \phi}|) = 0.155 \quad (C.11)$$

$$median_s(|\frac{\partial B}{\partial \theta}|) = 0.427 \quad (C.12)$$

The formula for $\sigma_{d_{min}}$ now depends only on two parameters: (d_0, z_0) . To get to a final measure dependent only on the noise values added, we estimate these using distributions of (d_0, z_0) for secondary and tertiary vertex tracks from our 300,000 toy jets. The maximum and median values are given below:

$$max_s(d_0) = 42,600 \mu m \quad (C.13)$$

$$max_s(z_0) = 27,100 \mu m \quad (C.14)$$

$$median_s(d_0) = 312 \mu m \quad (C.15)$$

$$median_s(z_0) = 28.9 \mu m \quad (C.16)$$

The error $\sigma_{d_{min}}$ is then calculated using the max and median values derived to create $\sigma_{d_{min}}^{max}$ and $\sigma_{d_{min}}^{median}$. The results are presented in table 7.8 for the two noise regimes:

Appendix D

Analysis Framework and Triggers

D.1 Analysis Framework

The GHH analysis chain covers all the processing of the data starting from the input DxAODs and finishing with the final result. The chain is composed of several key frameworks as follows:

1. The CxAOD Framework: Comprised of two main parts, called CxAODMaker and CxAODReader, a GHH specific version of the CxAOD Framework was developed. The CxAOD Framework is broadly responsible for applying the various calibrations recommended by the ATLAS CP groups, reducing the input DxAODs to a much more manageable CxAOD (calibrated xAOD) object. A size reduction is achieved from the order TB DxAODs to the order GB CxAODs. This is achieved by removing all information not necessary for the GHH analysis. The CxAODs are produced by the CxAODMaker. The CxAODReader can then be used to produce analysis ntuples in an easily readable format. Documentation for the CxAOD Framework can be found [here](#).
2. The n-tuple Reader: The analysis specific `ntuplereader_ghh` transforms the analysis ntuples outputted from CxAODReader into the kinematic histograms of the various analysis regions. It applies all the various cuts and outputs the histograms with each systematic variation applied.
3. The plotting tool: The `PlottingTool_GHH` framework generates kinematic

	Electron Trigger	Muon Trigger
2015	HLT_e24_lhemedium_L1EM20VH HLT_e60_lhemedium HLT_e120_lhloose	HLT_mu20_iloose_L1MU15 HLT_mu50
2016 - 2018	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhemedium_nod0 HLT_e140_lhloose_nod0 HLT_e300_etcut	HLT_mu26_ivarmedium HLT_mu50

Table D.1: The single-lepton triggers used in this analysis

variable plots in the various analysis regions. These are used to compare the various estimated backgrounds to data and validate these methods.

4. The statistical framework: An analysis specific framework is developed that uses HistFitter [125] to perform the statistical fit. This takes the histograms produced by ntuplereader_ghh and outputs the various results shown in section 8.10.

D.2 Analysis Triggers

The single lepton triggers used are summarized in table D.1. The single-lepton triggers are defined in the tech report [142].

Appendix E

Signal Samples

The various signal samples used in this analysis are summarized here. The 172 samples found across tables E.1 to E.4 are the current signal samples in use to produce the results.

Table E.1: Simulated signal samples used in the analysis with $m_H = 300$ GeV.

MC Type	DSID	m_H [GeV]	f_W	f_{WW}
GHH300fW1230fWW0	505229	300	1230	0
GHH300fW770fWW0	505230	300	770	0
GHH300fW1910fWW0	505231	300	1910	0
GHH300fW660fWW3960	505232	300	660	3960
GHH300fW415fWW2490	505233	300	415	2490
GHH300fW1015fWW6090	505234	300	1015	6090
GHH300fW0fWW4600	505235	300	0	4600
GHH300fW0fWW2450	505236	300	0	2450
GHH300fW0fWW7100	505237	300	0	7100
GHH300fWm730fWW4380	505238	300	-730	4380
GHH300fWm455fWW2730	505239	300	-455	2730
GHH300fWm1125fWW6750	505240	300	-1125	6750
GHH300fWm1360fWW0	505241	300	-1360	0
GHH300fWm810fWW0	505242	300	-810	0
GHH300fWm2180fWW0	505243	300	-2180	0
GHH300fWm610fWWm3660	505244	300	-610	-3660
GHH300fWm370fWWm2220	505245	300	-370	-2220
GHH300fWm960fWWm5760	505246	300	-960	-5760
GHH300fW0fWWm4650	505247	300	0	-4650
GHH300fW0fWWm2300	505248	300	0	-2300
GHH300fW0fWWm7250	505249	300	0	-7250
GHH300fW710fWWm4260	505250	300	710	-4260
GHH300fW430fWWm2580	505251	300	430	-2580
GHH300fW1110fWWm6660	505252	300	1110	-6660
GHH300fW1000fWW2000	505253	300	1000	2000
GHH300fW500fWW1000	505254	300	500	1000
GHH300fW1500fWW3000	505255	300	1500	3000
GHH300fW300fWW4500	505256	300	300	4500
GHH300fW150fWW2250	505257	300	150	2250
GHH300fW470fWW7050	505258	300	470	7050
GHH300fWm330fWW4950	505259	300	-330	4950
GHH300fWm150fWW2250	505260	300	-150	2250
GHH300fWm490fWW7350	505261	300	-490	7350
GHH300fWm1200fWW2400	505262	300	-1200	2400
GHH300fWm500fWW1000	505263	300	-500	1000
GHH300fWm1950fWW3900	505264	300	-1950	3900
GHH300fWm1100fWWm2200	505265	300	-1100	-2200
GHH300fWm500fWWm1000	505266	300	-500	-1000
GHH300fWm1800fWWm3600	505267	300	-1800	-3600
GHH300fWm300fWWm4500	505268	300	-300	-4500
GHH300fWm140fWWm2100	505269	300	-140	-2100
GHH300fWm470fWWm7050	505270	300	-470	-7050
GHH300fW330fWWm4950	505271	300	330	-4950
GHH300fW150fWWm2250	505272	300	150	-2250
GHH300fW500fWWm7500	505273	300	500	-7500
GHH300fW1100fWWm2200	505274	300	1100	-2200
GHH300fW500fWWm1000	505275	300	500	-1000
GHH300fW1800fWWm3600	505276	300	1800	-3600
GHH300fW1350fWW0	505373	300	1350	0
GHH300fW0fWW6200	505383	300	0	6200

Table E.2: Simulated signal samples used in the analysis with $m_H = 600$ GeV.

MC Type	DSID	m_H [GeV]	f_W	f_{WW}
GHH600fW1350fWW0	505277	600	1350	0
GHH600fW820fWW0	505278	600	820	0
GHH600fW2130fWW0	505279	600	2130	0
GHH600fW765fWW4590	505280	600	765	4590
GHH600fW470fWW2820	505281	600	470	2820
GHH600fW1205fWW7230	505282	600	1205	7230
GHH600fW0fWW6200	505283	600	0	6200
GHH600fW0fWW3800	505284	600	0	3800
GHH600fW0fWW9750	505285	600	0	9750
GHH600fWm855fWW5130	505286	600	-855	5130
GHH600fWm520fWW3120	505287	600	-520	3120
GHH600fWm1355fWW8130	505288	600	-1355	8130
GHH600fWm1340fWW0	505289	600	-1340	0
GHH600fWm790fWW0	505290	600	-790	0
GHH600fWm2130fWW0	505291	600	-2130	0
GHH600fWm755fWWm4530	505292	600	-755	-4530
GHH600fWm455fWWm2730	505293	600	-455	-2730
GHH600fWm1195fWWm7170	505294	600	-1195	-7170
GHH600fW0fWWm6250	505295	600	0	-6250
GHH600fW0fWWm3800	505296	600	0	-3800
GHH600fW0fWWm9900	505297	600	0	-9900
GHH600fW915fWWm5490	505298	600	915	-5490
GHH600fW555fWWm3330	505299	600	555	-3330
GHH600fW1440fWWm8640	505300	600	1440	-8640
GHH600fW1200fWW2400	505301	600	1200	2400
GHH600fW500fWW1000	505302	600	500	1000
GHH600fW1900fWW3800	505303	600	1900	3800
GHH600fW400fWW6000	505304	600	400	6000
GHH600fW200fWW3000	505305	600	200	3000
GHH600fW650fWW9750	505306	600	650	9750
GHH600fWm400fWW6000	505307	600	-400	6000
GHH600fWm200fWW3000	505308	600	-200	3000
GHH600fWm650fWW9750	505309	600	-650	9750
GHH600fWm1250fWW2500	505310	600	-1250	2500
GHH600fWm500fWW1000	505311	600	-500	1000
GHH600fWm2000fWW4000	505312	600	-2000	4000
GHH600fWm1200fWWm2400	505313	600	-1200	-2400
GHH600fWm500fWWm1000	505314	600	-500	-1000
GHH600fWm1900fWWm3800	505315	600	-1900	-3800
GHH600fWm400fWWm6000	505316	600	-400	-6000
GHH600fWm190fWWm2850	505317	600	-190	-2850
GHH600fWm650fWWm9750	505318	600	-650	-9750
GHH600fW450fWWm6750	505319	600	450	-6750
GHH600fW200fWWm3000	505320	600	200	-3000
GHH600fW670fWWm10050	505321	600	670	-10050
GHH600fW1300fWWm2600	505322	600	1300	-2600
GHH600fW600fWWm1200	505323	600	600	-1200
GHH600fW2050fWWm4100	505324	600	2050	-4100

Table E.3: Simulated signal samples used in the analysis with $m_H = 900$ GeV.

MC Type	DSID	m_H [GeV]	f_W	f_{WW}
GHH900fW1570fWW0	505325	900	1570	0
GHH900fW940fWW0	505326	900	940	0
GHH900fW2510fWW0	505327	900	2510	0
GHH900fW1015fWW6090	505328	900	1015	6090
GHH900fW610fWW3660	505329	900	610	3660
GHH900fW1615fWW9690	505330	900	1615	9690
GHH900fW0fWW8800	505331	900	0	8800
GHH900fW0fWW5250	505332	900	0	5250
GHH900fW0fWW14100	505333	900	0	14100
GHH900fWm1165fWW6990	505334	900	-1165	6990
GHH900fWm695fWW4170	505335	900	-695	4170
GHH900fWm1865fWW11190	505336	900	-1865	11190
GHH900fWm1550fWW0	505337	900	-1550	0
GHH900fWm920fWW0	505338	900	-920	0
GHH900fWm2480fWW0	505339	900	-2480	0
GHH900fWm1000fWWm6000	505340	900	-1000	-6000
GHH900fWm495fWWm2970	505341	900	-495	-2970
GHH900fWm1605fWWm9630	505342	900	-1605	-9630
GHH900fW0fWWm8700	505343	900	0	-8700
GHH900fW0fWWm5150	505344	900	0	-5150
GHH900fW0fWWm13950	505345	900	0	-13950
GHH900fW1205fWWm7230	505346	900	1205	-7230
GHH900fW720fWWm4320	505347	900	720	-4320
GHH900fW1925fWWm11550	505348	900	1925	-11550
GHH900fW1400fWW2800	505349	900	1400	2800
GHH900fW700fWW1400	505350	900	700	1400
GHH900fW2300fWW4600	505351	900	2300	4600
GHH900fW550fWW8250	505352	900	550	8250
GHH900fW300fWW4500	505353	900	300	4500
GHH900fW900fWW13500	505354	900	900	13500
GHH900fWm600fWW9000	505355	900	-600	9000
GHH900fWm300fWW4500	505356	900	-300	4500
GHH900fWm1000fWW15000	505357	900	-1000	15000
GHH900fWm1500fWW3000	505358	900	-1500	3000
GHH900fWm800fWW1600	505359	900	-800	1600
GHH900fWm2400fWW4800	505360	900	-2400	4800
GHH900fWm1400fWWm2800	505361	900	-1400	-2800
GHH900fWm700fWWm1400	505362	900	-700	-1400
GHH900fWm2300fWWm4600	505363	900	-2300	-4600
GHH900fWm550fWWm8250	505364	900	-550	-8250
GHH900fWm250fWWm3750	505365	900	-250	-3750
GHH900fWm850fWWm12750	505366	900	-850	-12750
GHH900fW600fWWm9000	505367	900	600	-9000
GHH900fW300fWWm4500	505368	900	300	-4500
GHH900fW1000fWWm15000	505369	900	1000	-15000
GHH900fW1500fWWm3000	505370	900	1500	-3000
GHH900fW800fWWm1600	505371	900	800	-1600
GHH900fW2400fWWm4800	505372	900	2400	-4800
GHH900fW1350fWW0	505382	900	1350	0
GHH900fW0fWW6200	505392	900	0	6200

Table E.4: Simulated signal samples used in the analysis for mass scan.

MC Type	DSID	m_H [GeV]	f_W	f_{WW}
GHH300fW0fWW4600	505235	300	0	4600
GHH300fW0fWW2450	505236	300	0	2450
GHH300fW0fWW7100	505237	300	0	7100
GHH300fW0fWWm4650	505247	300	0	-4650
GHH300fW0fWWm2300	505248	300	0	-2300
GHH300fW0fWWm7250	505249	300	0	-7250
GHH600fW1350fWW0	505277	600	1350	0
GHH600fW0fWW6200	505283	600	0	6200
GHH600fW0fWW3800	505284	600	0	3800
GHH600fW0fWW9750	505285	600	0	9750
GHH600fW0fWWm6250	505295	600	0	-6250
GHH600fW0fWWm3800	505296	600	0	-3800
GHH600fW0fWWm9900	505297	600	0	-9900
GHH900fW0fWW8800	505331	900	0	8800
GHH900fW0fWW5250	505332	900	0	5250
GHH900fW0fWW14100	505333	900	0	14100
GHH900fW0fWWm8700	505343	900	0	-8700
GHH900fW0fWWm5150	505344	900	0	-5150
GHH900fW0fWWm13950	505345	900	0	-13950
GHH300fW1350fWW0	505373	300	1350	0
GHH360fW1350fWW0	505374	360	1350	0
GHH420fW1350fWW0	505375	420	1350	0
GHH480fW1350fWW0	505376	480	1350	0
GHH540fW1350fWW0	505377	540	1350	0
GHH660fW1350fWW0	505378	660	1350	0
GHH720fW1350fWW0	505379	720	1350	0
GHH780fW1350fWW0	505380	780	1350	0
GHH840fW1350fWW0	505381	840	1350	0
GHH900fW1350fWW0	505382	900	1350	0
GHH300fW0fWW6200	505383	300	0	6200
GHH360fW0fWW6200	505384	360	0	6200
GHH420fW0fWW6200	505385	420	0	6200
GHH480fW0fWW6200	505386	480	0	6200
GHH540fW0fWW6200	505387	540	0	6200
GHH660fW0fWW6200	505388	660	0	6200
GHH720fW0fWW6200	505389	720	0	6200
GHH780fW0fWW6200	505390	780	0	6200
GHH840fW0fWW6200	505391	840	0	6200
GHH900fW0fWW6200	505392	900	0	6200
GHH1000fW1350fWW0	507657	1000	1350	0
GHH1200fW1350fWW0	507658	1200	1350	0
GHH1500fW1350fWW0	507659	1500	1350	0
GHH2000fW1350fWW0	507660	2000	1350	0
GHH1000fW0fWW6200	507661	1000	0	6200
GHH1200fW0fWW6200	507662	1200	0	6200
GHH1500fW0fWW6200	507663	1500	0	6200
GHH2000fW0fWW6200	507664	2000	0	6200

Bibliography

- [1] Alex Mansbridge, Gregory Barbour, Davide Piras, Christopher Frye, Ilya Feige, and David Barber, *Learning to Noise: Application-Agnostic Data Sharing with Local Differential Privacy*, In preparation for submission, 2021, arXiv: 2010.12464 (cit. on p. 28).
- [2] Matthew D. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge University Press, 2014, ISBN: 978-1-107-03473-0, 978-1-107-03473-0 (cit. on pp. 29, 32, 39).
- [3] Francis Halzen and Alan Martin, *Quarks & Leptons: An introductory course in modern particle physics*, John Wiley & Sons, 1984 (cit. on pp. 29, 32, 39).
- [4] Philip G Ratcliffe, *An Introduction to Elementary Particle Phenomenology*, 2053-2563, IOP Publishing, 2014, ISBN: 978-0-7503-1072-7, URL: <http://dx.doi.org/10.1088/978-0-7503-1072-7> (cit. on pp. 29, 32, 39).
- [5] P.A. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01 (cit. on pp. 30, 33, 37–39, 41, 52, 88, 143).
- [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*, Physics Letters B **716** (2012) 1, ISSN: 0370-2693, URL: <https://www.sciencedirect.com/science/article/pii/S037026931200857X> (cit. on p. 29).

- [7] David Galbraith and Carsten Burgard, 2013, URL:
<http://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model/> (cit. on p. 31).
- [8] David J. Gross and Frank Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (26 1973) 1343, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.30.1343> (cit. on p. 30).
- [9] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *Parton fragmentation and string dynamics*, Physics Reports **97** (1983) 31, ISSN: 0370-1573, URL: <https://www.sciencedirect.com/science/article/pii/0370157383900807> (cit. on p. 31).
- [10] Torbjörn Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159, ISSN: 0010-4655, URL: <http://dx.doi.org/10.1016/j.cpc.2015.01.024> (cit. on pp. 31, 69, 147).
- [11] S. Aid et al., *A study of the fragmentation of quarks in e^-p collisions at HERA*, Nuclear Physics B **445** (1995) 3, ISSN: 0550-3213, URL: <https://www.sciencedirect.com/science/article/pii/055032139591599H> (cit. on p. 31).
- [12] V. M. Abazov et al., *Evidence for production of single top quarks*, Phys. Rev. D **78** (1 2008) 012005, URL: <https://link.aps.org/doi/10.1103/PhysRevD.78.012005> (cit. on p. 32).
- [13] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on p. 33).

- [14] Peter W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508> (cit. on p. 33).
- [15] P.W. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12** (1964) 132, ISSN: 0031-9163, URL: <https://www.sciencedirect.com/science/article/pii/0031916364911369> (cit. on p. 33).
- [16] Peter W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (4 1966) 1156, URL: <https://link.aps.org/doi/10.1103/PhysRev.145.1156> (cit. on p. 33).
- [17] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (5 1967) 1554, URL: <https://link.aps.org/doi/10.1103/PhysRev.155.1554> (cit. on p. 33).
- [18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585> (cit. on p. 33).
- [19] V. Khachatryan et al., *Measurement of the ratio $B(t \rightarrow Wb)/B(t \rightarrow Wq)$ in pp collisions at $\sqrt{s} = 8$ TeV*, Physics Letters B **736** (2014) 33, ISSN: 0370-2693, URL: <http://dx.doi.org/10.1016/j.physletb.2014.06.076> (cit. on p. 39).
- [20] Michael E. Peskin, *Dark Matter and Particle Physics*, Journal of the Physical Society of Japan **76** (2007) 111017, ISSN: 1347-4073, URL: <http://dx.doi.org/10.1143/JPSJ.76.111017> (cit. on p. 40).

- [21] Matts Roos, *Dark Matter: The evidence from astronomy, astrophysics and cosmology*, 2010, arXiv: 1001.0316 [astro-ph.CO] (cit. on p. 40).
- [22] P. A. R. Ade et al., *Planck2013 results. I. Overview of products and scientific results*, *Astronomy Astrophysics* **571** (2014) A1, ISSN: 1432-0746, URL:
<http://dx.doi.org/10.1051/0004-6361/201321529>
(cit. on p. 40).
- [23] Ronald J. Adler, Brendan Casey, and Ovid C. Jacob, *Vacuum catastrophe: An elementary exposition of the cosmological constant problem*, *American Journal of Physics* **63** (1995) 620 (cit. on p. 40).
- [24] Sean M. Carroll, *The Cosmological Constant*, *Living Reviews in Relativity* **4** (2001), ISSN: 1433-8351, URL:
<http://dx.doi.org/10.12942/lrr-2001-1> (cit. on p. 40).
- [25] Y. Fukuda et al., *Evidence for Oscillation of Atmospheric Neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562, ISSN: 1079-7114, URL:
<http://dx.doi.org/10.1103/PhysRevLett.81.1562>
(cit. on p. 41).
- [26] Q. R. Ahmad et al., *Measurement of the Rate of $\nu_e + d \rightarrow p + p + e$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **87** (2001), ISSN: 1079-7114, URL:
<http://dx.doi.org/10.1103/PhysRevLett.87.071301>
(cit. on p. 41).
- [27] Csaba Csáki and Philip Tanedo, “Beyond the Standard Model”, 2013 *European School of High-Energy Physics*, 2015, arXiv: 1602.04228 [hep-ph] (cit. on pp. 41, 43).
- [28] Stephen P. Martin, *A SUPERSYMMETRY PRIMER*, *Advanced Series on Directions in High Energy Physics* (1998) 1, ISSN: 1793-1339, URL:
http://dx.doi.org/10.1142/9789812839657_0001 (cit. on pp. 42, 43).

- [29] Katrin Becker, Melanie Becker, and John H. Schwarz, *String Theory and M-Theory: A Modern Introduction*, Cambridge University Press, 2006 (cit. on p. 42).
- [30] Yu-Ping Kuang, Hong-Yu Ren, and Ling-Hao Xia, *Model-independent probe of anomalous heavy neutral Higgs bosons at the LHC*, Phys. Rev. D **90** (2014), ISSN: 1550-2368, URL: <http://dx.doi.org/10.1103/PhysRevD.90.115002> (cit. on pp. 43–45, 141).
- [31] Xin Chen et al., *Search for a generic heavy Higgs at the LHC*, Physics Letters B **804** (2020) 135358, ISSN: 0370-2693, URL: <https://www.sciencedirect.com/science/article/pii/S0370269320301623> (cit. on pp. 43, 45, 142–144).
- [32] Nathaniel Craig, Jamison Galloway, and Scott Thomas, *Searching for Signs of the Second Higgs Doublet*, 2013, arXiv: 1305.2424 [hep-ph] (cit. on pp. 43, 44).
- [33] Lyndon Evans and Philip Bryant, *LHC Machine*, JINST **3** (2008) S08001, URL: <https://doi.org/10.1088/1748-0221/3/08/s08001> (cit. on p. 46).
- [34] The ATLAS Collaboration, G Aad, et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003, URL: <https://doi.org/10.1088/1748-0221/3/08/s08003> (cit. on pp. 46–49, 51).
- [35] The CMS Collaboration, S Chatrchyan, et al., *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004, URL: <https://doi.org/10.1088/1748-0221/3/08/s08004> (cit. on p. 46).
- [36] The LHCb Collaboration, A Augusto Alves, et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005, URL:

- <https://doi.org/10.1088/1748-0221/3/08/s08005>
(cit. on p. 46).
- [37] The ALICE Collaboration, K Aamodt, et al., *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002, URL:
<https://doi.org/10.1088/1748-0221/3/08/s08002>
(cit. on p. 46).
- [38] ATLAS Collaboration, *ATLAS Experiment - Public Results*, URL:
<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/>
(cit. on p. 46).
- [39] Zachary Marshall and, *Simulation of Pile-up in the ATLAS Experiment*, Journal of Physics: Conference Series **513** (2014) 022024, URL:
<https://doi.org/10.1088/1742-6596/513/2/022024>
(cit. on p. 47).
- [40] G Apollinari, I Béjar Alonso, O Brüning, M Lamont, and L Rossi, *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*, CERN Yellow Reports: Monographs, CERN, 2015, URL:
<https://cds.cern.ch/record/2116337> (cit. on p. 47).
- [41] Corinne Pralavorio, *Record luminosity: Well done LHC*, 2017, URL:
<https://home.cern/news/news/accelerators/record-luminosity-well-done-lhc> (cit. on p. 47).
- [42] Fabienne Marcastel, *CERN's Accelerator Complex. La chaîne des accélérateurs du CERN*, (2013), General Photo, URL:
<https://cds.cern.ch/record/1621583> (cit. on p. 47).
- [43] Giulia Ripellino, *The alignment of the ATLAS Inner Detector in Run 2*, tech. rep., CERN, 2016, URL:
<https://cds.cern.ch/record/2213441> (cit. on p. 49).
- [44] Wolfgang Lukas, “ATLAS Inner Tracking Detectors”, *cds.cern.ch*, ICHEP, 2014, URL:

<https://cds.cern.ch/record/1746744/files/ATL-PHYS-SLIDE-2014-503.pdf> (cit. on pp. 51, 127).

- [45] M. Aaboud et al., *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017), ISSN: 1434-6052, URL:
<http://dx.doi.org/10.1140/epjc/s10052-017-5225-7>
(cit. on pp. 51, 55).
- [46] Georges Aad et al., *Performance of the ATLAS RPC detector and Level-1 muon barrel trigger at $\sqrt{s} = 13$ TeV*, JINST **16** (2021) P07029. 63 p, arXiv: 2103.01029, URL:
<http://cds.cern.ch/record/2753039> (cit. on p. 53).
- [47] M. Aaboud et al., *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017), ISSN: 1434-6052, URL:
<http://dx.doi.org/10.1140/epjc/s10052-017-4852-3>
(cit. on p. 53).
- [48] G. Aad et al., *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017), ISSN: 1434-6052, URL:
<http://dx.doi.org/10.1140/epjc/s10052-017-5004-5>
(cit. on p. 56).
- [49] M. Aaboud et al., *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017), ISSN: 1434-6052, URL:
<http://dx.doi.org/10.1140/epjc/s10052-017-5031-2>
(cit. on pp. 57, 58).
- [50] Matteo Cacciari, Gavin P Salam, and Gregory Soyez, *The anti-kt jet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063, ISSN: 1029-8479, URL:

<http://dx.doi.org/10.1088/1126-6708/2008/04/063>
(cit. on p. 57).

- [51] *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s} = 8$ TeV*, tech. rep., CERN, 2015, URL: <https://cds.cern.ch/record/2008678> (cit. on p. 59).
- [52] *Jet global sequential corrections with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV*, tech. rep., CERN, 2015, URL: <https://cds.cern.ch/record/2001682> (cit. on p. 59).
- [53] *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*, tech. rep., CERN, 2015, URL: <https://cds.cern.ch/record/2044941> (cit. on p. 59).
- [54] *Determination of the jet energy scale and resolution at ATLAS using Z/γ -jet events in data at $\sqrt{s} = 8$ TeV*, tech. rep., CERN, 2015, URL: <https://cds.cern.ch/record/2059846> (cit. on p. 59).
- [55] M. Aaboud et al., *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **96** (7 2017) 072002, URL: <https://link.aps.org/doi/10.1103/PhysRevD.96.072002> (cit. on pp. 59, 170).
- [56] G. Aad et al., *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. C **73** (2013), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-013-2306-0> (cit. on pp. 59, 170).
- [57] *Identification of boosted, hadronically-decaying W and Z bosons in $\sqrt{s} = 13$ TeV Monte Carlo Simulations for ATLAS*, tech. rep., CERN, 2015, URL: <https://cds.cern.ch/record/2041461> (cit. on p. 60).

- [58] Schramm, Steven, *ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz boosted Objects*, EPJ Web Conf. **182** (2018) 02113, URL: <https://doi.org/10.1051/epjconf/201818202113> (cit. on pp. 60, 61).
- [59] *Measurement of the ATLAS Detector Jet Mass Response using Forward Folding with 80 fb^{-1} of $\sqrt{s} = 13\text{ TeV}$ pp data*, tech. rep., CERN, 2020, URL: <https://cds.cern.ch/record/2724442> (cit. on p. 62).
- [60] *Boosted hadronic vector boson and top quark tagging with ATLAS using Run 2 data*, tech. rep., CERN, 2020, URL: <http://cds.cern.ch/record/2724149> (cit. on pp. 62, 63).
- [61] *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, tech. rep., CERN, 2016, URL: <https://cds.cern.ch/record/2200211> (cit. on p. 62).
- [62] G. Aad et al., *Performance of jet substructure techniques for large- R jets in proton-proton collisions at $\sqrt{s} = 7\text{ TeV}$ using the ATLAS detector*, Journal of High Energy Physics **2013** (2013), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP09\(2013\)076](http://dx.doi.org/10.1007/JHEP09(2013)076) (cit. on p. 63).
- [63] *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, tech. rep., CERN, 2016, URL: <https://cds.cern.ch/record/2157687> (cit. on p. 64).
- [64] M. Aaboud et al., *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. D **97** (7 2018) 072003, URL: <https://link.aps.org/doi/10.1103/PhysRevD.97.072003> (cit. on pp. 64, 65).
- [65] *PromptLeptonTaggerIFF Twiki*, 2018, URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/PromptLeptonTaggerIFF> (cit. on pp. 64, 67).

- [66] Christos Anastopoulos et al., *Electron identification and efficiency measurements in 2017 data*, tech. rep., CERN, 2019, URL:
<https://cds.cern.ch/record/2652163> (cit. on pp. 64, 65).
- [67] EGamma Working Group, *Electron and Photon Selection and Identification for Run2*, 2021, URL:
<https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaIdentificationRun2> (cit. on p. 65).
- [68] Henri Bachacou and Jean-Baptiste De Vivie, 2019, URL:
<https://indico.cern.ch/event/862748/contributions/3635039/attachments/1946503/3235131/conversionsIFF2019.11.18.pdf> (cit. on p. 65).
- [69] G. Aad et al., *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016), ISSN: 1434-6052, URL:
<http://dx.doi.org/10.1140/epjc/s10052-016-4120-y> (cit. on pp. 65, 169).
- [70] Richard O Duda and Peter E Hart, *Use of the Hough transformation to detect lines and curves in pictures*, Communications of the ACM **15** (1972) 11 (cit. on p. 66).
- [71] E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC p-p collisions, tech. rep., CERN, 2018, URL:
<https://cds.cern.ch/record/2625233> (cit. on p. 68).
- [72] James Catmore, *The ATLAS data processing chain: from collisions to papers*, 2016, URL: https://indico.cern.ch/event/472469/contributions/1982677/attachments/1220934/1785823/intro_slides.pdf (cit. on p. 69).
- [73] Andy Buckley et al., *General-purpose event generators for LHC physics*, Physics Reports **504** (2011) 145, ISSN: 0370-1573, URL:

<https://www.sciencedirect.com/science/article/pii/S0370157311000846> (cit. on p. 69).

- [74] Stefan Gieseke, Frashër Loshaj, and Patrick Kirchgaeßer, *Soft and diffractive scattering with the cluster model in Herwig*, Eur. Phys. J. C **77** (2017), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-017-4727-7> (cit. on p. 69).
- [75] S. Agostinelli, J. Allison, K. Amako, et al., *Geant4 — a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688> (cit. on p. 70).
- [76] James Catmore et al., *A new petabyte-scale data derivation framework for ATLAS*, J. Phys. Conf. Ser. **664** (2015) 072007 (cit. on p. 70).
- [77] Dan Guest, Kyle Cranmer, and Daniel Whiteson, *Deep Learning and Its Application to LHC Physics*, Annual Review of Nuclear and Particle Science **68** (2018) 161, ISSN: 1545-4134, URL: <http://dx.doi.org/10.1146/annurev-nucl-101917-021019> (cit. on p. 71).
- [78] Tom M. Mitchell, *Machine Learning*, McGraw-Hill, 1997, ISBN: 978-0-07-042807-2 (cit. on p. 71).
- [79] Kevin P. Murphy, *Probabilistic Machine Learning: An introduction*, MIT Press, 2022, URL: probml.ai (cit. on pp. 71, 72, 78, 79, 83).
- [80] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed., Prentice Hall, 2010 (cit. on pp. 71, 75).
- [81] Sebastian Ruder, *An overview of gradient descent optimization algorithms*, 2017, arXiv: 1609.04747 [cs.LG] (cit. on pp. 73, 74).

- [82] Sebastian Raschka, *Gradient descent and stochastic gradient descent*, 2020, URL: http://rasbt.github.io/mlxtend/user_guide/general_concepts/gradient-optimization/ (cit. on p. 74).
- [83] Andrew Bell, “b-Tagging and Evidence for the Standard Model $H \rightarrow bb$ Decay with the ATLAS Experiment”, PhD thesis: UCL, 2018, URL: <https://discovery.ucl.ac.uk/id/eprint/10043736> (cit. on p. 76).
- [84] Andreas Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, (2007), arXiv: physics/0703039 (cit. on pp. 76, 77).
- [85] Yoav Freund and Robert E Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences **55** (1997) 119, ISSN: 0022-0000, URL: <https://www.sciencedirect.com/science/article/pii/S002200009791504X> (cit. on p. 77).
- [86] *What is a neural network?*, 2021, URL: <https://www.tibco.com/reference-center/what-is-a-neural-network> (cit. on p. 78).
- [87] Christopher Olah, *Understanding LSTM networks*, 2015, URL: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (cit. on p. 80).
- [88] Sepp Hochreiter and Jürgen Schmidhuber, *Long Short-term Memory*, Neural computation **9** (1997) 1735 (cit. on p. 80).
- [89] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, tech. rep., CERN, 2017, URL: <https://cds.cern.ch/record/2255226> (cit. on pp. 80, 90, 91, 114, 115).
- [90] Jie Zhou et al., *Graph Neural Networks: A Review of Methods and Applications*, 2021, arXiv: 1812.08434 [cs.LG] (cit. on p. 82).

- [91] Steven Flores, *Variational autoencoders are beautiful*, 2019, URL: <https://www.compthree.com/blog/autoencoder/> (cit. on p. 83).
- [92] Diederik P Kingma and Max Welling, *Auto-Encoding Variational Bayes*, 2014, arXiv: 1312.6114 [stat.ML] (cit. on pp. 82, 83).
- [93] Irhum Shafkat, *Intuitively understanding variational autoencoders*, 2018, URL: <https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf> (cit. on p. 84).
- [94] Jason Brownlee, *How to use ROC curves and PRECISION-RECALL curves for classification in Python*, 2021, URL: <https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/> (cit. on p. 86).
- [95] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high p_T jets*, tech. rep., CERN, 2014, URL: <https://cds.cern.ch/record/1709132> (cit. on pp. 88, 116).
- [96] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run*, tech. rep., CERN, 2017, URL: <https://cds.cern.ch/record/2273281> (cit. on pp. 89, 93, 94, 96, 114).
- [97] ATLAS Collaboration, *Expected performance of the ATLAS b -tagging algorithms in Run-2*, tech. rep., CERN, 2015, URL: <https://cds.cern.ch/record/2037697> (cit. on pp. 89–91).
- [98] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, tech. rep., CERN, 2017, URL: <https://cds.cern.ch/record/2270366> (cit. on p. 92).

- [99] ATLAS Collaboration, *Topological b -hadron decay reconstruction and identification of b -jets with the JetFitter package in the ATLAS experiment at the LHC*, tech. rep., CERN, 2018, URL:
<https://cds.cern.ch/record/2645405> (cit. on pp. 93, 97, 100, 101).
- [100] ATLAS Collaboration, *Performance of b -jet identification in the ATLAS experiment*, JINST **11** (2016) P04008, ISSN: 1748-0221, URL:
<http://dx.doi.org/10.1088/1748-0221/11/04/P04008>
(cit. on p. 93).
- [101] G. Aad et al., *ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019), ISSN: 1434-6052, URL:
<http://dx.doi.org/10.1140/epjc/s10052-019-7450-8>
(cit. on pp. 94, 95, 97).
- [102] *Optimisation of the ATLAS b -tagging performance for the 2016 LHC Run*, tech. rep., CERN, 2016, URL:
<https://cds.cern.ch/record/2160731> (cit. on p. 94).
- [103] Francois Chollet et al., *Keras*, 2015, URL:
<https://github.com/fchollet/keras> (cit. on pp. 94, 117).
- [104] The Theano Development Team, Rami Al-Rfou, Guillaume Alain, et al., *Theano: A Python framework for fast computation of mathematical expressions*, 2016, arXiv: 1605.02688 [cs.SC] (cit. on p. 94).
- [105] Diederik P. Kingma and Jimmy Ba, *Adam: A Method for Stochastic Optimization*, 2017, arXiv: 1412.6980 [cs.LG] (cit. on pp. 94, 117).
- [106] Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio, *Maxout Networks*, 2013, arXiv: 1302.4389 [stat.ML] (cit. on p. 96).

- [107] ATLAS Collaboration, *Expected performance of the 2019 ATLAS b-taggers*, (2019), URL: <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/> (cit. on p. 98).
- [108] Giacinto Piacquadio, “Identification of b-jets and investigation of the discovery potential of a Higgs boson in the $WH \rightarrow l\nu b\bar{b}$ channel with the ATLAS experiment”, Presented on 11 Jan 2010, 2010, URL: <https://cds.cern.ch/record/1243771> (cit. on pp. 97–99, 101–104, 108, 114).
- [109] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874> (cit. on pp. 102, 114).
- [110] M. Aaboud et al., *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. C **77** (2017), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-017-4887-5> (cit. on pp. 102, 114).
- [111] Jonathan Shlomi et al., *Secondary vertex finding in jets with neural networks*, Eur. Phys. J. C **81** (2021), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-021-09342-y> (cit. on p. 110).
- [112] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant, *Graph neural networks in particle physics*, Machine Learning: Science and Technology **2** (2021) 021001, ISSN: 2632-2153, URL: <http://dx.doi.org/10.1088/2632-2153/abbf9a> (cit. on p. 110).

- [113] Philip Zhuang Dan Iter Jonathan Kuck, “Target Tracking with Kalman Filtering”, 2016 (cit. on p. 114).
- [114] Huseyin Coskun, Felix Achilles, Robert DiPietro, Nassir Navab, and Federico Tombari, *Long Short-Term Memory Kalman Filters: Recurrent Neural Estimators for Pose Regularization*, 2017, arXiv: 1708.01885 [cs.CV] (cit. on pp. 114, 140).
- [115] Anton Milan, Seyed Hamid Rezatofighi, Anthony Dick, Ian Reid, and Konrad Schindler, *Online Multi-Target Tracking Using Recurrent Neural Networks*, 2016, arXiv: 1604.03635 [cs.CV] (cit. on p. 114).
- [116] Steffen Jung, Isabel Schlangen, and Alexander Charlish, “Sequential Monte Carlo Filtering with Long Short-Term Memory Prediction”, *2019 22th International Conference on Information Fusion (FUSION)*, 2019, URL: <https://ieeexplore.ieee.org/document/9011258> (cit. on p. 114).
- [117] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from [tensorflow.org](https://www.tensorflow.org), 2015, URL: <https://www.tensorflow.org/> (cit. on p. 117).
- [118] Erika Garutti, *Vertex and Tracking detectors*, 2012, URL: https://www.desy.de/~garutti/LECTURES/ParticleDetectorSS12/L9_Tracking.pdf (cit. on p. 127).
- [119] Scott M Lundberg and Su-In Lee, “A Unified Approach to Interpreting Model Predictions”, *Advances in Neural Information Processing Systems 30*, ed. by I. Guyon et al., Curran Associates, Inc., 2017 4765, URL: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf> (cit. on p. 139).

- [120] M. Aaboud et al., *Search for heavy ZZ resonances in the $\ell^+\ell^-\ell^+\ell^-$ and $\ell^+\ell^-\nu\bar{\nu}$ final states using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-018-5686-3> (cit. on p. 141).
- [121] V. Khachatryan et al., *Search for a Higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons*, Journal of High Energy Physics **2015** (2015), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP10\(2015\)144](http://dx.doi.org/10.1007/JHEP10(2015)144) (cit. on p. 141).
- [122] M. Aaboud et al., *Combination of searches for heavy resonances decaying into bosonic and leptonic final states using 36 fb¹ of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **98** (2018), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.98.052008> (cit. on p. 141).
- [123] A. M. Sirunyan et al., *Search for a heavy Higgs boson decaying to a pair of W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Journal of High Energy Physics **2020** (2020), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP03\(2020\)034](http://dx.doi.org/10.1007/JHEP03(2020)034) (cit. on p. 141).
- [124] Yu-Ping Kuang, Hong-Yu Ren, and Ling-Hao Xia, *Further investigation of the model-independent probe of heavy neutral Higgs bosons at LHC Run 2*, Chinese Physics C **40** (2016) 023101, URL: <https://doi.org/10.1088/1674-1137/40/2/023101> (cit. on p. 142).
- [125] M. Baak, G. J. Besjes, D. Côté, A. Koutsman, J. Lorenz, and D. Short, *HistFitter software framework for statistical data analysis*, Eur. Phys. J. C **75** (2015), ISSN: 1434-6052, URL:

<http://dx.doi.org/10.1140/epjc/s10052-015-3327-7>
(cit. on pp. 146, 176, 178, 183, 210).

- [126] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer, *MadGraph 5: going beyond*, Journal of High Energy Physics **2011** (2011), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP06\(2011\)128](http://dx.doi.org/10.1007/JHEP06(2011)128) (cit. on p. 147).
- [127] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks, *FeynRules 2.0 — A complete toolbox for tree-level phenomenology*, Computer Physics Communications **185** (2014) 2250, ISSN: 0010-4655, URL: <https://www.sciencedirect.com/science/article/pii/S0010465514001350> (cit. on p. 147).
- [128] Boping Chen et al., *Search for Generic Heavy Higgs Boson Using 13 TeV pp Collision Data at ATLAS*, tech. rep., CERN, 2020, URL: <https://cds.cern.ch/record/2742567> (cit. on p. 147).
- [129] Johannes Krause and Frank Siegert, *NLO QCD predictions for $Z + \gamma + jets$ production with Sherpa*, Eur. Phys. J. C **78** (2018) 161, arXiv: 1708.06283 [hep-ph] (cit. on p. 162).
- [130] M. Aaboud, G. Aad, et al., *Observation of Electroweak Production of a Same-Sign W Boson Pair in Association with Two Jets in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, Phys. Rev. Lett. **123** (16 2019) 161801, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.123.161801> (cit. on p. 163).
- [131] M. Aaboud et al., *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 125, ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-016-4466-1> (cit. on p. 169).

- [132] ATLAS Collaboration, *Search for charginos and neutralinos in final states with two boosted hadronically decaying bosons and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Submitted to Phys. Rev. D, 2021, arXiv: 2108.07586 [hep-ex] (cit. on p. 171).
- [133] M. Aaboud et al., *In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector*, Eur. Phys. J. C **79** (2019), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6632-8> (cit. on p. 171).
- [134] G. Aad et al., *Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Journal of High Energy Physics **2020** (2020), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP07\(2020\)124](http://dx.doi.org/10.1007/JHEP07(2020)124) (cit. on p. 175).
- [135] Morad Aaboud et al., *Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **99** (2019) 072009, arXiv: 1901.03584 [hep-ex] (cit. on p. 175).
- [136] Georges Aad et al., *Evidence for the production of three massive vector bosons with the ATLAS detector*, Phys. Lett. B **798** (2019) 134913, arXiv: 1903.10415 [hep-ex] (cit. on p. 175).
- [137] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, **2/2017** (2016), arXiv: 1610.07922 [hep-ph] (cit. on p. 175).
- [138] Jon Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. **G43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 175).

- [139] *Observation of WW production in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector*, tech. rep., CERN, 2021, URL:
<https://cds.cern.ch/record/2779173> (cit. on p. 175).
- [140] A L Read, *Presentation of search results: the CLs technique*, Journal of Physics G: Nuclear and Particle Physics **28** (2002) 2693, URL:
<https://doi.org/10.1088/0954-3899/28/10/313> (cit. on p. 177).
- [141] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011), ISSN: 1434-6052, URL:
<http://dx.doi.org/10.1140/epjc/s10052-011-1554-0> (cit. on p. 178).
- [142] *Trigger menu in 2018*, tech. rep., All figures including auxiliary figures are available at
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-DAQ-PUB-2019-001>: CERN, 2019, URL:
<https://cds.cern.ch/record/2693402> (cit. on p. 210).