Transforming flavor tagging and its application to Higgs to charm measurements at the ATLAS experiment

by

Johannes Michael Wagner

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Heather Gray, Chair
Professor Gabriel Orebi Gann
Professor Lawrence Hall

Summer 2026

Transforming flavor tagging and its application to Higgs to charm measurements at the ATLAS experiment

Copyright 2026
by
Johannes Michael Wagner

Abstract

Transforming flavor tagging and its application to Higgs to charm measurements at the ATLAS experiment

by

Johannes Michael Wagner

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Heather Gray, Chair

Modern particle physics mostly concerns itself with the extension and precision measurement of the Standard Model. The Higgs boson, which was discovered in 2012 by the ATLAS and CMS experiments, represents the newest component of this theory. It couples to fermions through the Yukawa interaction of the Higgs mechanism. Some Yukawa couplings, such as that between the Higgs boson and bottom quark, have been studied thoroughly over the past 14 years and show good agreement with theoretical predictions. Others are far less explored, requiring large quantities of data and precise reconstruction strategies to isolate them. Among these, the Higgs to charm quark coupling represents the next frontier of precision Higgs measurements that could offer hints towards other open questions beyond the Standard Model, such as the fermion mass hierarchy and potential modifications to the Higgs mechanism that generate neutrino masses.

This thesis presents a simultaneous measurement of the bottom and charm Yukawa couplings in the VH production channel using 140 fb^{-1} of run 2 ATLAS data. It provides the most sensitive observed upper limit on $\mu_{VHc\bar{c}}$ at present at a 95% CL of 11.5 $\times$ SM predictions. For $VH(\rightarrow b\bar{b})$, a series of results is presented using simplified template cross sections (STXS). When combined and interpreted through the Kappa framework, the observed limit is $|\kappa_c/\kappa_b| < 3.6$. In addition to VH, a reference measurement using the VZ process is also presented including the first observation of $VZ(\rightarrow c\bar{c})$.

The improvement of $VH(\rightarrow c\bar{c})$ measurements necessitates efficient identification of charm-jets, which is facilitated by the development of new Transformer-based algorithms. Separate models are shown for the case of low momentum and high momentum Higgs boson decays via `GN2` and `GN2X` which improve background rejections in both cases by factors of more than two. Each algorithm also performs explicit secondary vertex (SV) reconstructions, which are contrasted with previous methods. Finally, upgraded `GN3` and `GN3X` models are presented

and the former is trained to distinguish jet charge alongside jet flavor which could further assist in event selections for VH analyses.

Both the prior $VH(\to b\bar{b}/c\bar{c})$ analysis and the advances in jet flavor identification are leveraged to build a new analysis strategy for $VH(\to c\bar{c})$ in the high momentum regime, which has never been done in ATLAS. The multivariate analysis (MVA) algorithm from the previous result is replaced with a neural network approach whose performance is studied. Preliminary expected sensitivities are presented based on Asimov data using ATLAS run 2 simulation. These results indicate the feasibility of this analysis strategy and serve as a blueprint for a future measurement.

Contents

Contents **i**

List of Figures **iv**

List of Tables **x**

1 Introduction **1**

2 Theoretical Motivation **4**
2.1 Fundamentals of the Standard Model 5
2.1.1 Particles of the Standard Model 5
2.1.2 Mathematical Formalism 7
2.1.3 Quantum Electrodynamics (QED) 7
2.1.4 Quantum Chromodynamics (QCD) 8
2.2 The Electroweak Sector 9
2.2.1 The Charged Current 9
2.2.2 CP Violation and the CKM Matrix 10
2.2.3 Neutrino Oscillations 11
2.2.4 Electroweak Unification 12
2.3 The Higgs Boson 12
2.3.1 Scalar Fields and Symmetry Breaking 13
2.3.2 Particle Masses and Higgs Interactions 14
2.3.3 Higgs Boson Phenomenology 16
2.4 Collider Physics 19
2.4.1 Basics of Hadron Collisions 20
2.4.2 Final States and Hadronization 21
2.4.3 Monte Carlo Simulations 22

3 Experimental Apparatus **24**
3.1 The Large Hadron Collider 24
3.2 The ATLAS Detector 28
3.2.1 ATLAS Coordinate System 29

3.2.2 Inner Tracking Detector . . . 30
3.2.3 Electromagnetic and Hadronic Calorimeters . . . 32
3.2.4 Muon Spectrometer . . . 34
3.2.5 Magnet System . . . 35
3.2.6 Trigger and DAQ . . . 36

4 Event Reconstruction 38
4.1 Fundamental Objects . . . 38
4.1.1 Tracks, Clusters and Vertices . . . 39
4.1.2 Electrons and Photons . . . 43
4.1.3 Muons . . . 45
4.1.4 Particle Jets . . . 47
4.1.5 Hadronic Taus . . . 51
4.1.6 Missing Energy . . . 53
4.2 Jet Flavor Identification . . . 54
4.2.1 Basics of Flavor Tagging . . . 54
4.2.2 Traditional Flavor Tagging in ATLAS . . . 56
4.2.3 Flavor Tagging Calibrations . . . 62

5 Measurement of $VH(\to b\bar{b}/c\bar{c})$ using Run 2 Data 64
5.1 Analysis Overview . . . 65
5.1.1 Analysis Backgrounds . . . 67
5.1.2 Analysis Samples . . . 70
5.1.3 Event Selection . . . 71
5.1.4 Event Categorization . . . 76
5.2 Modeling and Uncertainties . . . 83
5.2.1 Experimental Systematics . . . 84
5.2.2 Modeling Systematics . . . 84
5.3 Fit Description . . . 97
5.3.1 Fit Overview . . . 98
5.3.2 Fit Development and Diagnostics . . . 100
5.4 Results . . . 107
5.4.1 $VZ(\to b\bar{b}/c\bar{c})$ Measurement Results . . . 107
5.4.2 $VH(\to b\bar{b}/c\bar{c})$ Measurement Results . . . 110
5.4.3 Final Fit Model Validation . . . 116
5.5 Conclusions . . . 118

6 Advances in Flavor Tagging using Machine Learning 123
6.1 `SVGNN`: Vertex Reconstruction with Graph Attention Networks . . . 124
6.1.1 Description of the `SVGNN` Architecture . . . 124
6.1.2 Training of `SVGNN` . . . 126
6.1.3 `SVGNN` Performance . . . 128

6.2 GN2: Transformer-based Jet Flavor Identification . . . 130
6.2.1 The Architecture of GN2 and Auxiliary Tasks . . . 131
6.2.2 GN2 Performance . . . 136
6.2.3 Edge Features in GN2 . . . 144
6.3 GN3: Jet Charge Identification . . . 148
6.3.1 The Architecture of GN3 and Jet Charge Identification . . . 149
6.3.2 GN3 Performance . . . 157
6.4 GN2X and GN3X: Extension to the Boosted Regime . . . 166
6.4.1 GN2X Architecture and Performance . . . 167
6.4.2 GN3X Architecture and Performance . . . 172
6.5 Conclusions . . . 174

7 $VH(\to c\bar{c})$ in run 3 and beyond 176
7.1 Feasibility studies on $VH(\to c\bar{c})$ with GN2 and GN2X . . . 177
7.2 Exploratory measurement of $VH(\to c\bar{c})$ in the boosted regime . . . 180
7.2.1 Analysis Overview . . . 180
7.2.2 Neural network based MVA Approach . . . 184
7.2.3 Preliminary Fit Results . . . 191
7.3 Future Work in $VH(\to c\bar{c})$ . . . 198

Bibliography 199

List of Figures

2.1 A visual diagram of all Standard Model particles from [175]. 6
2.2 Visual representation of the Higgs potential (with $\lambda = 2$) for positive $\mu^2 = 10$ (left) and negative $\mu^2 = -10$ (right). 14
2.3 Feynman diagrams for the most common Higgs production modes: (a) ggF, (b) VBF, (c) VH and (d) $t\bar{t}H$ from [27]. 17
2.4 Predicted cross sections for different Higgs production modes in hadron collisions at $\sqrt{s} = 14$ TeV as a function of m_H from [171]. 17
2.5 Predicted branching ratios for Higgs boson decays as a function of m_H from [171]. 18
2.6 Summary of theoretical and experimental agreement for various Yukawa couplings from [78]. 19
2.7 Visual representation of a $t\bar{t}$ production event as simulated by a Monte Carlo generator from [188]. 23

3.1 Schematic overview of the CERN accelerator complex including main synchrotron accelerators LHC, SPS and PS from [20]. 25
3.2 Plots showing data quality in the ATLAS experiment for (a) run 2 and (b) run 3 up to end of 2025 from [100] 27
3.3 Luminosity vs. mean number of interactions for runs 1, 2 and 3 (up to September 17th 2025) as recorded by ATLAS from [101]. 27
3.4 Diagram showing all components of the ATLAS detector in its original configuration from [38]. 28
3.5 Sketch showing how different particles interact with detector components from [73]. Dashed tracks are invisible to the detector. 29
3.6 Diagram showing the orientation of the coordinate system used in ATLAS as well as relevant kinematic variables from [181]. 30
3.7 Schematic of the ATLAS inner detector and its components for run 3 from [98]. 31
3.8 Schematic of the ATLAS calorimeters and their components for run 3 from [98]. 33
3.9 Schematic of the ATLAS muon spectrometer for run 3 from [98]. 35
3.10 Schematic of the ATLAS magnet system from [209]. 36
3.11 Diagram showing the core components of the ATLAS Trigger and DAQ for run 3 from [98]. 37

4.1 Detector signature of a $W(\to \mu\nu)H(\to b\bar{b})$ event candidate showing reconstructed tracks and calorimeter energy deposits from [56]. 39
4.2 Diagram visualizing track parameters used in ATLAS from [149]. 40
4.3 Plots showing (a) d_0 and (b) z_0 resolution as a function of η for final tracks, track candidates and track seeds in run 3 reconstruction from [109] 43
4.4 Visualization of jet clustering in k_T and anti-k_T algorithms in (y, ϕ) plane adapted from [25]. 48
4.5 Diagram visualizing a bottom and two light-flavor jets from [180]. 56
4.6 Plots showing distribution of track (a) d_0 and (b) $z_0 \sin\theta$ for different jet flavors in a simulated run 3 $t\bar{t}$ sample 57
4.7 Plots showing background rejections vs. (a) bottom- and (b) charm-jet efficiency for `DL1r`, `DL1` and `MV2c10` from [90] 61

5.1 Visualization of the three $V(\to ll)H(\to b\bar{b}/c\bar{c})$ analysis regimes under consideration from [72]. 66
5.2 Feynman diagrams showing semi-leptonic $VH(\to q\bar{q})$ decays in the (a) zero, (b) one and (c) two lepton channels. 67
5.3 Lowest order Feynman diagrams for (a) $qq \to VH$ and (b) $gg \to ZH$ processes. 68
5.4 Lowest order Feynman diagrams for V+jets processes. 69
5.5 Lowest order Feynman diagrams for gluon induced (a) t-channel, (b) u-channel and (c) s-channel $t\bar{t}$ production. Quark induced $t\bar{t}$ diagrams are similar. 70
5.6 Lowest order Feynman diagrams for (a) s-channel, (b) t-channel and (c) Wt-channel single-top production. 70
5.7 Lowest order Feynman diagrams for (a) t-channel, (b) u-channel and (c) s-channel VV processes. 71
5.8 Graphics detailing the PCFT scheme employed in the resolved regions of the $VH(\to b\bar{b}/c\bar{c})$ analysis showing PCFT categories (a) and resulting regions (b) from [104]. 75
5.9 Overview of all $VH(\to b\bar{b}/c\bar{c})$ analysis regions from [104]. 77
5.10 Histograms showing MC events passing boosted $VH(\to b\bar{b})$, $VH(\to c\bar{c})$ or both selections in $VH(\to b\bar{b})$ (a) and $VH(\to c\bar{c})$ signal samples (b). 78
5.11 Prefit composition of SRs and CRs in 150 GeV $\leq p_T^V \leq$ 250 GeV resolved regions for zero lepton (a), one lepton (b) and two lepton (c) and all p_T^V and lepton regions for boosted (d). 80
5.12 Comparison of MVA output distributions from testing and training data for BDT_{VH} in the two lepton channel for the 4+jet, $p_T^V \geq 150$ GeV training in resolved $VH \to b\bar{b}$ (a), the 3+jet, $p_T^V \geq 150$ GeV training in $VH \to c\bar{c}$ (b) and the inclusive training in boosted $VH \to b\bar{b}$ (c) 83
5.13 Example visualization of NP pull differences between "nominal" and "test" fit model. Systematics are ranked by $|\hat{\theta}_{\mathrm{test}} - \hat{\theta}_{\mathrm{nominal}}|/\sigma_{\hat{\theta}_{\mathrm{nominal}}}$, highlighting the largest pull differences between the two fit models. 102

5.14 Figures showing pulls of systematic uncertainties related to W+jets (a) and Z+jets (b) backgrounds in $VH(\to c\bar{c})$ test fit comparing m_{j1j2} with VH MVA binning. Uncharacteristically small constraints on shape uncertainties indicate possible modeling issues. 103
5.15 Figures showing pulls of flavor tagging related systematic uncertainties for five leading charm (a) and tau (b) NPs in a $VH(\to b\bar{b}/c\bar{c})$ test fit comparing three configurations excluding tau components, correlating them with charm components or including both but keeping them separate. 105
5.16 Figures showing ranking of systematic uncertainties based on $\Delta\mu$ impact on measured signal strengths for the final $VZ(\to b\bar{b})$ (a) and $VZ(\to c\bar{c})$ (b) fit alongside pulls and constraints for selected NPs from [72]. 108
5.17 Figures showing observed values and uncertainties of $\mu_{VZb\bar{b}}$ (a) and $\mu_{VZc\bar{c}}$ (b) decorrelated in the WZ and ZZ processes from [104]. 110
5.18 Figures showing observed values and uncertainties of $\mu_{VZb\bar{b}}$ (a) and $\mu_{VZc\bar{c}}$ (b) decorrelated in the zero, one and two lepton channels from [104]. 110
5.19 Contours showing observed and expected confidence intervals for $\mu_{VHb\bar{b}}$ and $\mu_{VHc\bar{c}}$ from [104]. 111
5.20 Figures showing Higgs S/B weighted event distributions across BB (a), C_TC (b) and C_TN (c) analysis regions. 112
5.21 Figures showing ranking of systematic uncertainties based on $\Delta\mu$ impact on measured signal strengths for $VH(\to b\bar{b})$ (a) and $VH(\to c\bar{c})$ (b) alongside pulls and constraints for selected NPs from [104]. 113
5.22 Figures showing observed values and uncertainties of $\mu_{VHb\bar{b}}$ (a) and $\mu_{VHc\bar{c}}$ (b) decorrelated in the WZ and ZZ processes from [104]. 115
5.23 Figures showing observed values and uncertainties of $\mu_{VHb\bar{b}}$ (a) and $\mu_{VHc\bar{c}}$ (b) decorrelated in the zero, one and two lepton channels from [104]. 115
5.24 Visualization of expected and observed limits for $\mu_{VHc\bar{c}}$ in fit decorrelated by lepton regions. 116
5.25 Figures showing observed values and uncertainties of $\mu_{VHb\bar{b}}$ in the 10 (a) and 13 (b) $VH(\to b\bar{b})$ POI STXS split from [104]. 117
5.26 Figures showing post-fit distributions for all $N_{\text{jet}} = 2$ signal regions in the resolved $VH(\to b\bar{b})$ regime arranged horizontally by p_T^V category and vertically by lepton channel from [104]. 119
5.27 Figures showing post-fit distributions for all $N_{\text{jet}} = 2$ C_TC signal regions in the $VH(\to b\bar{b})$ regime arranged horizontally by p_T^V category and vertically by lepton channel from [104]. 120
5.28 Figures showing post-fit distributions for all $N_{\text{jet}} = 2$ C_TN signal regions in the $VH(\to c\bar{c})$ regime arranged horizontally by p_T^V category and vertically by lepton channel from [104]. 121
5.29 Figures showing post-fit distributions for all signal regions in the boosted $VH(\to b\bar{b})$ regime arranged horizontally by p_T^V category and vertically by lepton channel from [104]. 122

6.1 Visualization of the architecture of the SVGNN algorithm. Here, nodes represent individual tracks and edges represent relationships between them. Information originally associated with each node is transformed into edge feature sets, which are classified to identify SV connections. 125
6.2 Plots showing distribution of z_0 of tracks passing and failing cuts for SVGNN training originating from (a) heavy flavor particle decays and (b) other sources . . . 128
6.3 Plots showing vertex reconstruction efficiency in SVGNN and SV1 as a function of PV-SV separation (L_{xy}) for (a) bottom-jets and (b) charm-jets with SV's 129
6.4 Plot showing vertex reconstruction fake rate in SVGNN and SV1 as a function of jet p_T in jets without SV's. 130
6.5 Plots showing vertex reconstruction efficiency vs. fake rate for SVGNN and SV1. . 131
6.6 Diagram showing the different components of the GN2 algorithm from [110]. . . . 132
6.7 Diagram showing "Pre-LN" Transformer architecture as implemented in GN2 based on [220]. 134
6.8 Plots showing background rejections vs. bottom-jet efficiency for DL1d and GN2 evaluated on (a) $t\bar{t}$ and (b) Z' from [110]. 138
6.9 Plots showing background rejections vs. charm-jet efficiency for DL1d and GN2 evaluated on (a) $t\bar{t}$ and (b) Z' from [110]. 139
6.10 Plots showing background rejections vs. bottom-jet efficiency for GN2, GN2NoAux and GN2Lep evaluated on (a) $t\bar{t}$ and (b) Z' from [110]. 140
6.11 Plots showing (a) bottom-jet efficiency, (b) charm-jet rejection and (c) light-jet rejection vs. p_T at 77% bottom-jet efficiency including MC statistical uncertainty bands for GN2 evaluated on different alternate $t\bar{t}$ samples from [110]. 141
6.12 Plots showing (a) charm-jet and (b) light-jet rejection vs. bottom-jet efficiency in MC and calibrated in data (with corresponding uncertainties) for GN2 and DL1r in semi-leptonic $t\bar{t}$ events from [110]. 142
6.13 Plots showing m_{SV} in bottom-jets with the (a) inclusive and (b) exclusive vertexing strategy for MC truth, GN2, SV1 and JetFitter (exclusive only) in semi-leptonic $t\bar{t}$ events. 143
6.14 Plots showing m_{SV} in charm-jets with the (a) inclusive and (b) exclusive vertexing strategy for MC truth, GN2, SV1 and JetFitter (exclusive only) in semi-leptonic $t\bar{t}$ events. 143
6.15 Plots showing total fraction of (a) bottom- and (b) charm-jets with (correctly) identified secondary vertices vs. p_T after enforcing vertex quality criteria using inclusive vertexing in semi-leptonic $t\bar{t}$ events. 145
6.16 Plot showing total fraction of light-jets with (falsely) reconstructed secondary vertices vs. p_T using inclusive vertexing in semi-leptonic $t\bar{t}$ events. 145
6.17 Plots showing track association efficiency for SVs in (a) bottom- and (b) charm-jets vs. p_T using inclusive vertexing in semi-leptonic $t\bar{t}$ events. 146
6.18 Plots showing track association purity for SVs in (a) bottom- and (b) charm-jets vs. p_T using inclusive vertexing in semi-leptonic $t\bar{t}$ events. 146

6.19 Plots showing background rejections vs. (a) bottom-jet and (b) charm-jet efficiency for GN2 and GN2Edge evaluated on $t\bar{t}$ jets. 149
6.20 Plots showing background rejections vs. (a) bottom-jet and (b) charm-jet efficiency for GN2 and GN2Edge trained on 10% of the nominal GN2 training data evaluated on $t\bar{t}$ jets. 150
6.21 Plots showing background rejections vs. (a) bottom-jet and (b) charm-jet efficiency for GN2 and GN2Edge trained on 1% of the nominal GN2 training data evaluated on $t\bar{t}$ jets. 151
6.22 Diagram showing the different components of the GN3 algorithm from [208]. . . . 154
6.23 Plots showing background rejections vs. bottom-jet efficiency for GN2, GN3NoCharge and GN3 evaluated on (a) $t\bar{t}$ and (b) Z'. 158
6.24 Plots showing background rejections vs. charm-jet efficiency for GN2, GN3NoCharge and GN3 evaluated on (a) $t\bar{t}$ and (b) Z'. 159
6.25 Histograms showing $\mathcal{D}_b^{\text{charge}}$ for GN3 evaluated on true bottom-jets in (a) $t\bar{t}$ and (b) Z'. 160
6.26 Histograms showing $\mathcal{D}_c^{\text{charge}}$ for GN3 evaluated on true charm-jets in (a) $t\bar{t}$ and (b) Z'. 160
6.27 Histograms showing $\mathcal{D}_b^{\text{charge}}$ for GN3 evaluated on bottom-jets passing (a,b) 70%, (c,d) 77% and (e,f) 85% ϵ_b^{jet} WPs for (a,c,e) $t\bar{t}$ and (b,d,f) Z'. All WPs are defined respective to the $t\bar{t}$ sample and MC event weights are applied. 162
6.28 Histograms showing $\mathcal{D}_c^{\text{charge}}$ for GN3 evaluated on charm-jets passing (a,b) 10%, (c,d) 30% and (e,f) 50% ϵ_c^{jet} WPs for (a,c,e) $t\bar{t}$ and (b,d,f) Z'. All WPs are defined respective to the $t\bar{t}$ sample and MC event weights are applied. 163
6.29 Plots showing $\bar{q}$ inefficiency vs. q efficiency evaluated on (a) $t\bar{t}$ and (b) Z'. . . . 164
6.30 Plots showing q efficiency variations vs. jet p_T at a fixed global $\epsilon_q^{\text{charge}} = 50\%$ with binomial uncerta on (a) $t\bar{t}$ and (b) Z'. 164
6.31 Plots showing $\bar{q}$ efficiency variations vs. jet p_T at a fixed global $\epsilon_q^{\text{charge}} = 50\%$ with binomial uncertainties on (a) $t\bar{t}$ and (b) Z'. 165
6.32 Occurrence frequency of different hadron species in (a) bottom-jets and (b) charm-jets across the $t\bar{t}$ sample. 165
6.33 Plots showing $\bar{q}$ inefficiency vs. q efficiency split by meson species for (a) bottom-jets and (b) charm-jets. All efficiency and inefficiency values are calculated respective to jets containing the meson species in question only. 166
6.34 Histogram showing $\mathcal{D}_c^{\text{charge}}$ for true charm-jets containing $D^{*\pm}$ mesons split by decay mode. Each distribution is normalized individually to emphasize shape differences. 167
6.35 Plots showing background rejections vs. (a) $H(\to b\bar{b})$ and (b) $H(\to c\bar{c})$ efficiency for GN2X, the DNN-based Xbb algorithm (for the former only) and the GN2-based strategy from [95]. 170
6.36 Histograms showing m_J distributions for $H(\to b\bar{b})$ and multi-jet before and after a GN2X based selection from [95]. 171

6.37 Plots showing (a) background rejections vs. $H(\to b\bar{b})$ and $Z(\to b\bar{b})$ efficiency for GN2X and (b) bin-by-bin variations of $H(\to b\bar{b})$ and $Z(\to b\bar{b})$ efficiencies as a function of m_J. 171
6.38 Plots showing background rejections vs. (a) $H(\to b\bar{b})$ and (b) $H(\to c\bar{c})$ efficiency for GN3X and GN2Xτ from [108]. 175

7.1 Diagram showing individually binned regions considered in the $VH(\to c\bar{c})$ feasibility study. 178
7.2 Plots showing (a) total and (b) relative cumulative cross sections for the main Higgs production modes as a function of the applied p_T cut from [14]. 181
7.3 Overview of all $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ analysis regions in the partial run 3 boosted measurement. 184
7.4 Prefit composition of SRs and CRs across the $VH(\to b\bar{b})$ (a) and $VH(\to c\bar{c})$ (b) regimes split by lepton channel and p_T^V region. 185
7.5 ROC curves comparing $H \to c\bar{c}$ signal efficiency vs. background rejection for binary classification MVA models in the (a) zero, (b) one and (c) two lepton channels. 190
7.6 ROC curves comparing $H \to c\bar{c}$ signal efficiency vs. background rejection for multi-class classification MVA models in the (a) zero, (b) one and (c) two lepton channels. 192
7.7 ROC curves comparing $H \to c\bar{c}$ signal efficiency vs. background rejection for DNN and Transformer "TwoHead" MVA models in the (a) zero, (b) one and (c) two lepton channels. 193
7.8 Diagrams showing 95% confidence level upper limits for $\mu_{VHc\bar{c}}$ by lepton channel and in an Asimov $VH(\to b\bar{b}/c\bar{c})$ fit using (a) m_J or (b) the MVA discriminant as a discriminating variable. 194
7.9 Diagrams showing uncertainties on floating normalization factors and POIs in an Asimov $VH(\to b\bar{b}/c\bar{c})$ fit using (a) m_J or (b) the MVA discriminant as a binning variable. 195
7.10 Figures showing post-fit Asimov distributions for all signal regions in the boosted $VH(\to b\bar{b})$ regime arranged horizontally by p_T^V category and vertically by lepton channel. 196
7.11 Figures showing post-fit Asimov distributions for all signal regions in the boosted $VH(\to c\bar{c})$ regime arranged horizontally by p_T^V category and vertically by lepton channel. 197

List of Tables

2.1 Summary of Higgs decay modes with associated branching ratios based on Standard Model predictions from [190]. 20

3.1 Overview of run periods of the LHC between 2008 and 2026 with luminosity delivered to the ATLAS experiment for each run period from [101]. 26

4.1 Proper decay lengths $c\tau$ for common B and D mesons. 55
4.2 Summary of all input variables for `DL1r` adapted from [90]. 60

5.1 Monte Carlo samples used in the $VH(\rightarrow b\bar{b}/c\bar{c})$ analysis adapted from [104]. This table summarizes generators used for matrix element calculation as well as parton shower (PS) and hadronization modeling. The specific parton distribution function (PDF) and underlying event (UE) tune are also given for each sample. 72
5.2 Lepton selection criteria used in the $VH(\rightarrow b\bar{b}/c\bar{c})$ analysis. 73
5.3 Input variables used in MVA trainings adapted from [104]. 82
5.4 Summary of experimental systematics related to leptons and MET used in the analysis adapted from [72]. 85
5.5 Summary of experimental systematics related to small-R and large-R jets used in the analysis adapted from [72]. 86
5.6 Summary of MC generators and systematic variations considered for all analysis samples, adapted from [72]. 89
5.7 Normalization and acceptance uncertainties for V+jets backgrounds adapted from [72]. 92
5.8 Normalization and acceptance uncertainties for Top backgrounds adapted from [72]. 95
5.9 Normalization and acceptance uncertainties for VZ backgrounds adapted from [72]. 96
5.10 Detailed breakdown of NP contributions to errors on VZ signal strengths adapted from [72]. 109
5.11 Detailed breakdown of NP contributions to errors on VH signal strengths adapted from [72]. 114

6.1 Summary of input variables for each track in `SVGNN`. 126

6.2 Detailed summary of SVGNN architecture with layer types and sizes. 127
6.3 Monte Carlo samples used for SVGNN training. 128
6.4 Summary of input variables for each track in GN2 adapted from [82]. 133
6.5 Track selection used for inputs to GN2 [110]. 133
6.6 Truth origin categories for track classification auxiliary task adapted from [82]. . 135
6.7 Detailed summary of GN2 architecture with layer types and sizes. 136
6.8 Monte Carlo samples used for GN2 training (nominal only) and performance evaluation (nominal and alternative) [110]. 137
6.9 Overview of GN2Edge architecture modifications relative to GN2 as shown in table 6.7. 148
6.10 Summary of input variables for each constituent in GN3 adapted from [107, 208]. 153
6.11 Track selection used for inputs to GN3 adapted from [107]. 153
6.12 Detailed summary of GN3 architecture with layer types and sizes. 156
6.13 Detailed summary of GN2X architecture with layer types and sizes. 168
6.14 Monte Carlo samples used for GN2X training and performance evaluation. 169
6.15 Detailed summary of GN3X architecture with layer types and sizes. 173
6.16 Track selection used for inputs to GN3X [108]. 173

7.1 Summary of $VH(\rightarrow c\bar{c})$ feasibility study results extrapolated from run 2 to run 3 with GN2 and GN2X based jet flavor identification selections. 179
7.2 Monte Carlo samples used in the boosted $VH(\rightarrow c\bar{c})$ analysis. 182
7.3 Detailed summary of DNN and Transformer MVA architectures for the boosted $VH(\rightarrow c\bar{c}$ measurement including layer types and sizes. 187
7.4 List of input features used for the analysis MVA studies in the run 3 boosted $VH(\rightarrow c\bar{c})$ analysis for the DNN and Transformer approaches. 188

Acknowledgments

First and foremost, I want to acknowledge the people without whom this work would not have been possible, starting with my advisor, Prof. Heather Gray, who provided me with so much guidance throughout this PhD. I'm very grateful you chose to take me on as one of your first students, and I greatly appreciate all of the things you've taught me in our many, many conversations. I also want to acknowledge and thank my other two thesis committee members, Prof. Gabriel Orebi Gann and Prof. Lawrence Hall, for all their efforts in helping me across the finish line.

More than anyone, I also want to give thanks to my beautiful wife, Amari, who has been my rock and shaped my thinking about both my research and the world. Thank you for the countless hours of listening to me rant about physics over the years and for always being there and supporting me; I love you. My parents, Doris and Michael Wagner, have always emphasized the value of hard work and dedication and have supported me throughout, both financially and emotionally. My brother, Stefan Wagner, who has been one of my closest friends my whole life and is always open to interesting conversations, whether it's about science or video games. My grandmas, Annemarie Eßbüchl and Christine Wagner, my uncle, Rupert Eßbüchl, and my late grandfathers, Rupert Eßbüchl and Johann Wagner, who raised me alongside my parents and have always been there for me. Thank you for welcoming me home with open arms every time I land in Vienna and supporting me from afar, as well as being the pillars among which the whole family gathers. Thank you also to my entire extended family across Lower Austria for all the support over the years as well as my new family in Kentucky. My parents-in-law, Serina and William Russell Sr., who embraced me from the beginning alongside my little brother, William Russell Jr., and sister, Chamique Cowan, as well as my uncle-brothers (as Amari likes to say), Joel and Melvin Cowan, Mai-Ling Garcia and baby Sir Cowan. I'm truly thankful you all have become family!

I also want to thank my entire circle of friends who have been there consistently throughout the years. Thank you to Maxi Mayr, who has been one of my oldest and closest friends since high school, for always standing by me despite the distance and helping me become the person I am today. You are truly an inspiration, and I'm glad we have been able to keep up our friendship over time and space. The same thing goes for Nikita Stepura, my lost Ukrainian brother, and his wife, Yulia Tarasenko. Both of you have inspired me personally and professionally, and I cherish the many long conversations we've had — always in a different city, but always over cherry liqour. My friends in Philadelphia, particularly Aaron Patel, Matt Brower, Zach Jung, Alejandro Castañeda, and Shannon Vo, who routinely make me miss my undergraduate years.

Last but definitely not least, I want to thank the family I have built in my years at Berkeley. Andres Franco Valiente, for teaching me about physics, health, and everything in between; your sense of humor and love for life is unmatched, and I'll miss our daily gym sessions. Linus Hyunseong Kim, for going on more dinner dates with me than anyone with the possible exception of my wife; catching up with you has always been a weekly highlight

during my time in the Bay. Malte Schwarz, for providing some cultural grounding and consistently entertaining commentary in a place so far away from home. Chad Harper, for all the insightful conversations we've had and for being one of the most versatile thinkers I know. Matt Harris, who made my time at CERN particularly fun through kebab dinners, gym sessions, and impromptu summit hikes. Special thanks also to the Berkeley RSF community, where I spent most of my time outside of my home and the lab and met dozens of great people who I have no doubt will accomplish amazing things in life.

I'm also thankful to Prof. Marjorie Shapiro who, despite her retirement, has remained an integral piece of many research projects and a great mentor. Thank you as well to Tomohiro Yamazaki for introducing me to ATLAS research in general and flavor tagging in particular; your influence has shaped the direction of my research greatly. Thank you also to Maria Mironova and Miha Muskinja, who taught me almost everything I know about likelihood fits and acted as invaluable guides in the more chaotic phases of the legacy analysis. This also goes for Louis-Guillaume Gagnon, who has been the most constant point of reference through my time at Berkeley for everything from machine learning to physics measurements; I would not have gotten to this point without your guidance. Thank you as well to Irina Ene and Robin Junwen Xiong, whom I got to know very well during the beginning years of the PhD. I learned a lot from you two and thoroughly enjoyed working alongside you. The same goes for Hadley Santana Queiroz, Stella Zhirong Zhang, and Liam Foster. I learned a lot from you three and I know you will accomplish great things in the future. I'm also looking forward to reading your theses one day.

Throughout my Ph.D., I have also had the privilege of being a mentor to some amazing undergraduate students, particularly Max Zhao, Nishank Gite, Isabella Deutsch, and Yanjie Chen. Seeing each of you grow as scientists has been incredibly rewarding, and I wish each of you all the best. The Berkeley ATLAS group as a whole deserves special thanks as well, particularly group leaders Kevin Einsweiler and Simone Pagan Griso, for all the amazing infrastructure they have built around the research to support all of us. I also want to thank the physics departments at UC Berkeley as well as Drexel University, particularly Prof. Naoko Kurahashi-Neilson, Prof. Russell Neilson, Prof. Michelle Dolinski, and Prof. David Goldberg, for instilling in me a passion for physics and research.

Perhaps the thing I am most grateful for during my time at Berkeley was having the privilege to work with one of the most fiercely intelligent and passionate communities the world will ever know at CERN. Both as an institution and a physical office complex, CERN truly is a special place, and the time I was able to spend there was invaluable. There are too many people to thank both in the ATLAS experiment and the broader CERN community, but in particular Sam Van Stroud, Francesco Armando di Bello, Nicole Hartman, Alexander Froch, Alvaro Lopez Solis, Andrea Milici, Romain Bouquet, Weitao Wang, Elisabeth Schopf, Yanhui Ma, and Hannah Arnold need to be mentioned by name. Thank all of you for the guidance over the years — none of these projects would have come to fruition without you!

An meine Familie in Österreich möchte ich auch noch einige Worte in meiner Muttersprache richten. Danke für alles, was ihr über die Jahre für mich getan habt, um meine persönliche Entwicklung und Ausbildung zu ermöglichen und zu unterstützen. Vor allem

will ich mich noch einmal bei meinen Eltern, meinen Omas Annemarie und Christine und meinem Onkel Rupert bedanken, die mich und meinen Bruder gemeinsam aufgezogen haben und für uns alle möglichen Türen geoffnet haben. Ich wäre ohne euch niemals so weit gekommen, und ich bin euch unendlich dankbar. Ich hoffe auch, dass meine Opas Rupert und Johann, obwohl sie nicht mehr unter uns sind, von oben herab wissen, wie sehr ich sie schätze. Ich weiß, die Distanz war nicht immer einfach (weder für mich noch für euch), aber ich bin froh, dass wir es immer wieder geschafft haben, zusammenzukommen. Trotz dem Ozean zwischen uns werde ich stets bei euch sein, egal wo es mich hinverschlägt.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 2146752 and the WATCHEP program sponsored by the U.S. Department of Energy, Office of Science under Award No. DE-SC-0023527. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the U.S. Department of Energy.

Chapter 1

Introduction

This thesis titled "Transforming flavor tagging and its application to Higgs to charm measurements at the ATLAS experiment" is intended to summarize results achieved as part of my PhD research performed between 2021 and 2026. It focuses on two main areas: jet flavor identification and measurements of the Higgs to charm coupling via $VH(\rightarrow c\bar{c})$ decays. The common core threaded throughout both topics is the use of modern machine learning models such as Transformers, which enable many of the observed improvements. This work is built on discoveries made by thousands of physicists over hundreds of years and much of it was performed in direct collaboration with other members of the ATLAS experiment. Emphasis is placed on areas with significant individual contributions and all figures made without personal involvement are cited in their captions.

The first three chapters provide an introduction to particle physics, the ATLAS experiment and event reconstruction techniques. Chapter 2 contains detailed explanations of the Standard Model (SM), the Higgs boson and the theoretical aspects of hadron collider physics which provide the theoretical foundation of this research. Chapter 3 describes the experimental apparatus, including both the ATLAS detector and the Large Hadron Collider (LHC) with which the analyzed data is collected. Chapter 4 details how collision events recorded in the ATLAS detector are reconstructed, starting from individual particles to larger scale structures such as jets. An overview of the process of jet flavor identification is also given. These three chapters are introductory and contain no personal contributions.

Chapter 5 summarizes one of the main results of this thesis: a measurement of the $VH(\rightarrow b\bar{b})$ and $VH(\rightarrow c\bar{c})$ processes using 140 fb^{-1} of ATLAS run 2 data recorded at $\sqrt{s} = 13$ TeV. This measurement was performed in a large analysis team with 89 individual contributors and was published in the "Journal of High Energy Physics" in 2025 [104]. The analysis methodology overall is summarized followed by an in depth discussion of the likelihood-based fit model and a summary of the results. The $VH(\rightarrow b\bar{b})$ signal strength was measured as $\mu_{\mathrm{VHb\bar{b}}} = 0.92^{+0.16}_{-0.15}$, which is consistent with Standard Model predictions and the presently most stringent upper limit has been placed on the $VH(\rightarrow c\bar{c})$ process at $11.5 \times \mathrm{SM}$ ($10.6 \times \mathrm{SM}$) observed (expected). My contributions centered on the development of the fit model, in particular for the measurement of $VH(\rightarrow c\bar{c})$ as well as the diboson reference

measurement. Numerous studies testing various configurations of signal and control regions were performed to optimize measurement sensitivity. A significant amount of time was also spent developing correlation schemes for the hundreds of systematic uncertainties across analysis regions. Due to specific contributions in understanding differences between the VH and VZ fit models, I also played a significant role in the unblinding process of the latter and presented the initial results to the wider collaboration.

Chapter 6 describes the second set of results from this thesis, which are related to a series of developments in jet flavor identification using modern machine learning techniques. Multiple algorithms based on Graph Neural Networks (GNNs) and Transformers are presented for the reconstruction of secondary decay vertices in jets as well as the identification of bottom- and charm-jets and high momentum Higgs decays. The first of these is `SVGNN`, which originates from an individual project to find secondary vertices in bottom- and charm-jets. While these results are unpublished, they informed the transition to work on the next generation of ATLAS flavor identification algorithms `GN2` (used in the low momentum regime to identify single quark jets) and `GN2X` (used in the high momentum regime to identify collimated resonance decays). My main contributions for these two algorithms include design and implementation of the secondary vertex reconstruction and performance evaluation pipeline, model architecture studies including the addition of relational edge features as well as model training, software development for the `salt` [214] and `pUma` [213] packages and direct contributions to multiple publications. Results for `GN2` were published in "Nature Communications" in 2026 [110] and show improvements in background rejections of factors of 3-7 for bottom-jet and 2-6 for charm-jet identification compared to the typical performance of its predecessor. Results for `GN2X` were made public in 2023, prior to the `GN2` publication, in the form of an ATLAS note [95], showcasing the first ATLAS algorithm to directly identify high momentum $H \to c\bar{c}$ decays in addition to improvements of a factor of 2 in $H \to b\bar{b}$ identification. A separate publication detailing the `salt` framework was published in the "Journal of Open Source Software" in 2025 [13]. I also worked on the development of the successor to `GN2` called `GN3`, where I designed components that became the first charm-jet charge identification algorithm in ATLAS. These results were published as an ATLAS note in 2026 [208] for which I served as a main editor. Finally, while I was not personally involved in the development of the `GN2X` successor algorithm `GN3X`, a summary is also given as it provides necessary context for the following chapter.

Chapter 7 brings together results from the previous two chapters to provide a blueprint for the next measurement of $VH(\to c\bar{c})$ using ATLAS run 3 data taken at $\sqrt{s} = 13.6$ TeV by leveraging algorithms such as `GN2`, `GN3`, `GN2X` and `GN3X`. A series of feasibility studies are presented based on the previous $VH(\to b\bar{b}/c\bar{c})$ measurement using approximate flavor identification efficiencies to estimate sensitivity improvements expected from the use of upgraded flavor identification algorithms. From there, an exploratory analysis of a brand new measurement of $VH(\to c\bar{c})$ in the high momentum regime is detailed and preliminary sensitivity estimates from Asimov fits on run 2 MC are presented. An Asimov sensitivity without considering systematic uncertainties of $7.4 \times$ SM for $\mu_{\mathrm{VHc\bar{c}}}$ is expected. This sets the foundation for a standalone high momentum $VH(\to c\bar{c})$ analysis with combined run 2 and run

3 data to surpass the sensitivity of the most recent public measurement focused on the low momentum regime. All contributions in this chapter are thus far unpublished, with personal involvement most concentrated in the development of the MVA strategy as well as sample production and framework development.

Aside from the work summarized above, I directly contributed to detector operations at CERN where I performed over 20 shifts on the data quality (DQ) desk in the ATLAS control room. My authorship qualification task for the experiment, which involved the investigation of the impact of tracking uncertainties on flavor tagging performance with the `DL1r` algorithm, is also not included in this thesis.

Chapter 2

Theoretical Motivation

Particle physics is the study of the most fundamental building blocks of the universe. It seeks to understand nature from the bottom up, concerning itself with the interactions of the smallest constituents of physical reality. Philosophically, the idea that matter consists of indivisible units dates back to the "atomos" postulated by Ancient Greek thinker Democritus. This concept was central to the early developments in particle physics when the movement of atoms was thought to be governed by classical descriptions of mechanics and electrodynamics. In 1897, this picture slowly began to change with the discovery of the electron by J.J. Thomson. Subsequent discoveries of the proton and neutron definitively showed that atoms are not fundamental themselves but rather made up of smaller constituents. Soon it became clear that classical physics was not fit to describe the motion and interactions of these particles. This led to the development of quantum mechanics and quantum electrodynamics, which provide the foundation of modern particle physics. A rapid succession of breakthroughs in the middle and latter half of the 20th century culminated in the discovery that protons and neutrons themselves were not indivisible, but rather made up of smaller particles called quarks. A full description of how these most fundamental particles interact is provided in the Standard Model (SM), which summarizes our present understanding of particle physics. The most recent contribution to this theory, the Higgs mechanism, is responsible for the addition of particle masses and predicts the existence of a new particle: the Higgs boson. While it was discovered in 2012, the Higgs boson is still at the center of many unsolved mysteries surrounding the SM.

This chapter serves as a brief explanation of the more pertinent aspects of the SM as it relates to the Higgs boson in general and its couplings to bottom and charm quarks in particular. After elucidating the basic foundations of the theory (sections 2.1 and 2.2), a more detailed exploration of the Higgs mechanism and electroweak symmetry breaking is given (section 2.3). This is followed by an overview of hadron collider physics (section 2.4), which provides the experimental foundation for Higgs research in the modern age. This chapter is structurally inspired by *Modern Particle Physics by Mark Thomson* [211].

2.1 Fundamentals of the Standard Model

The Standard Model (SM) [139, 219, 137, 143, 186, 201, 202] is a quantum field theory describing all known fundamental particles and their interactions through 3 out of the 4 known fundamental forces: the electromagnetic (EM) force, the weak nuclear force and the strong nuclear force. It combines our understanding of quantum mechanics and special relativity, providing a comprehensive but incomplete description of particle physics due to its incompatibility with general relativity and thus gravity. The Standard Model is defined by its gauge symmetry group $SU(3) \times SU(2) \times U(1)$, which describes a series of local mathematical transformations that leave a system of particles invariant. This symmetry is deeply linked with the appearance of the three forces incorporated in the SM. In particular, $SU(3)$ is associated with the strong nuclear force while $SU(2) \times U(1)$ describes the unified electroweak interaction. The latter symmetry is spontaneously broken by the Higgs mechanism, leading to the separate manifestation of the weak and EM forces in nature. This mechanism provides the theoretical foundation for Higgs boson research and will be built up throughout this chapter. The focus will be placed on the symmetries, which are at the core of all particle interactions within the SM. The particles themselves will be discussed first to contextualize their interactions.

2.1.1 Particles of the Standard Model

The particles of the SM are broadly grouped into two categories based on their spin: fermions have half-integer spin and obey Fermi-Dirac statistics [132], while bosons have integer spin and obey Bose-Einstein statistics [17]. All fundamental fermions in the SM have spin-$1/2$ and make up the matter particles, which are further categorized as either quarks or leptons. SM fermions come in 3 generations, each of which consists of one up-type quark ($q = \frac{2}{3}$), one down-type quark ($q = -\frac{1}{3}$), one charged lepton l ($q = -1$) and one neutrino ν ($q = 0$). Particles in successive generations are identical in their characteristics with the exception of their masses. This makes second and third generation fermions inherently unstable as they tend to decay into their lighter counterparts. Most ordinary matter in the universe existing in bound states is thus made up of up quarks, down quarks and electrons rather than charm quarks, strange quarks and muons or top quarks, bottom quarks and taus. Different types of quarks, leptons and neutrinos are commonly referred to as "flavors".

In addition to electromagnetic charge, quarks also have a property called color charge associated with the $SU(3)$ symmetry, which enables them to interact via the strong force. Leptons do not possess color charge, so they can only interact through the EM[1] or weak forces. The latter is the only fundamental interaction that all known particles[2] can experience and is related to the weak isospin I_3. Alongside the weak hypercharge Y_W, I_3 is the charge associated with the combined $SU(2) \times U(1)$ symmetry corresponding to the electroweak interaction. These two quantities are related to EM charge as $Q = I_3 + \frac{1}{2}Y_W$. After symmetry

[1] Only charged particles interact electromagnetically, so this does not apply to neutrinos.

[2] This statement comes with a slight caveat discussed in section 2.2.1.

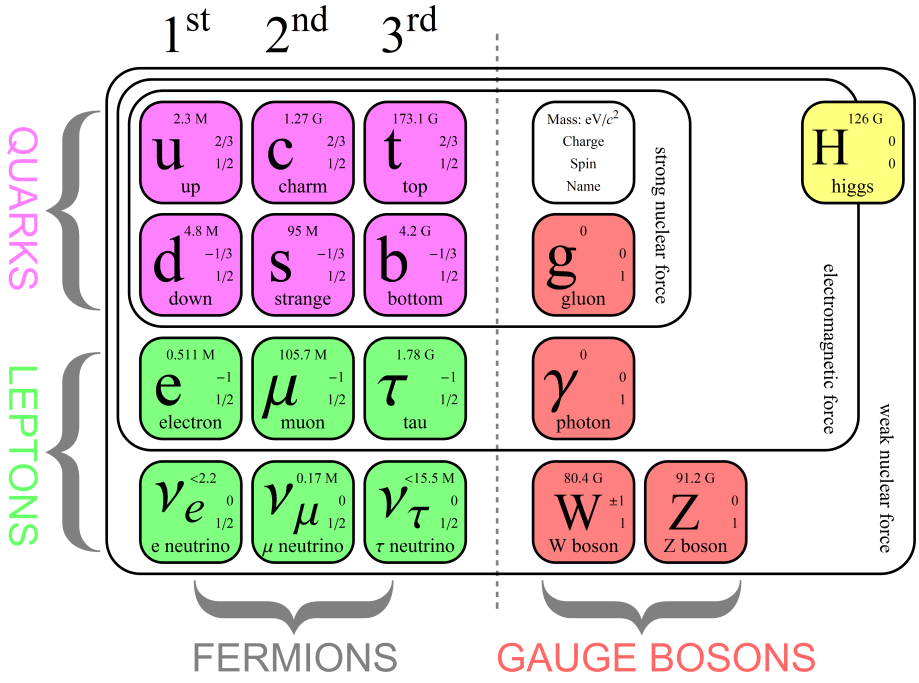


Figure 2.1: A visual diagram of all Standard Model particles from [175].

breaking, only Q remains conserved universally. Finally, each fermion has an anti-particle counterpart with identical characteristics, but opposite EM charge. Since neutrinos are charge neutral, their anti-particles are distinguished by their allowed interactions. Currently, it is not known whether neutrinos are their own anti-particles (i.e. Majorana neutrinos) or if anti-neutrinos are truly fundamentally distinct (i.e. Dirac neutrinos) [5].

Bosons in the SM are split into four vector bosons (with spin $s = 1$) and one scalar boson (with spin $s = 0$): the Higgs. Vector bosons are force mediators generated by the symmetry groups of the SM. Gluons are associated with the strong force and come in 8 configurations of color charge related to the 8 generators of $SU(3)$. Due to the aforementioned symmetry breaking of $SU(2) \times U(1)$ leading to the separation of the weak and EM forces, the photon and W and Z bosons do not map directly onto the generators of this group. In practice, the photon is associated with the EM interaction while the massive W and Z bosons propagate the weak interaction. The observation that the mediators of the weak interaction are not massless led to the introduction of a corrective mechanism to the SM which is the subject of section 2.3.1. This is the theoretical origin of the final particle in the SM, the Higgs boson. The Higgs is deeply connected to the existence of particle masses and is currently still the least understood aspect of the Standard Model. A visual summary of all SM particles and their properties is presented in Fig. 2.1.

2.1.2 Mathematical Formalism

The significance of the Higgs boson within the SM becomes apparent through the mathematical formalism of the theory. As a quantum field theory, the fundamental objects of the SM are quantum fields represented by spinors [29] whose excitations correspond to observable particles. The SM Lagrangian $\mathcal{L}_{SM}$ is defined by its underlying symmetries, which consist of a global combined charge conjugation, parity and time reversal (CPT) symmetry in addition to the $SU(3) \times SU(2) \times U(1)$ local gauge symmetries discussed previously. Charge conjugation (exchange of particles with anti-particles), parity (reversal of space) and time reversal do not always hold individually within the SM, but their combination must always remain unbroken.

Mathematically, fermions are represented by Dirac spinors ψ which have 4 independent components. These can be decomposed into two orthogonal two-component spinors $\psi_L = \frac{1}{2}(1-\gamma_5)\psi$ and $\psi_R = \frac{1}{2}(1+\gamma_5)\psi$ which represent chiral eigenstates that are related to each other via a parity transformation. This particular decomposition is important because these eigenstates (left-handed and right-handed spinors) are treated differently within some SM interactions leading to violation of parity-symmetry discussed in section 2.2.2. Each chiral spinor has a spin-up and a spin-down component. In the free particle case, a fermion field satisfies the Dirac equation given by

$$(i\partial\!\!\!/ - m)\psi = 0 \tag{2.1}$$

where $\partial\!\!\!/ = \gamma^\mu \partial_\mu$. Given $\bar{\psi} = \psi^\dagger \gamma_0$, the free particle Dirac Lagrangian consists of a kinetic term and a mass term

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi \tag{2.2}$$

This simple picture is complicated significantly by the presence of boson fields, which are responsible for introducing particle interactions to the SM. These are described by the theories of Quantum Electrodynamics (QED) [133, 192, 136] and Quantum Chromodynamics (QCD) [137, 143, 186], electroweak theory [139, 219] and the Higgs mechanism [151, 134], which are all related to a particular gauge symmetry or the breaking thereof.

2.1.3 Quantum Electrodynamics (QED)

Quantum electrodynamics (QED) is a quantization of traditional electrodynamics that works at the scale of the smallest particles such as quarks and leptons. Electromagnetism is one of the oldest components of classical physics and describes the macroscopic phenomena of electricity and magnetism via the electromagnetic force. QED is concerned with the underlying interactions that give rise to these phenomena. It is a quantum theory that postulates the SM Lagrangian is invariant under a $U(1)$ local phase transformation of the form

$$\psi(x) \to \psi'(x) = e^{iq\chi(x)}\psi(x) \tag{2.3}$$

This symmetry is enforced with the introduction of the covariant derivative $D_\mu = \partial_\mu + iqA_\mu$, where $A_\mu \to A'_\mu = A_\mu - \partial_\mu \chi(x)$. The Dirac Lagrangian is thus left invariant under the $U(1)$ transformation with the introduction of a new field A_μ, which is coupled via the electric charge q (and represents the photon). The choice of $\chi(x)$ is arbitrary and leads to the ability to define different gauges. This allows for some redundancy in the field description of electrodynamics, which is familiar from the classical theory. In practice, the $U(1)$ symmetry manifests via the exchange of photons between charged particles, described in full by the QED Lagrangian

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\gamma^\mu \partial_\mu \psi - q\bar{\psi}\gamma^\mu A_\mu \psi - m\bar{\psi}\psi \quad (2.4)$$

Here $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength tensor that represents the kinetic term of the photon, which describes the propagation of electromagnetic fields. Application of the covariant derivative results in both the standard kinetic term for fermions as well as a fermion-photon interaction term which is only present when $q \neq 0$, as expected. This is the reason photons only interact with particles that carry EM charge.

2.1.4 Quantum Chromodynamics (QCD)

The theory of QED outlines the basic principles behind quantum field theories with local gauge symmetries. Quantum Chromodynamics (or QCD), which defines the strong nuclear interaction, is a more complex example of a gauge symmetric quantum field theory. It has no classical counterpart and is defined by a $SU(3)$ symmetry enforcing invariance under transformations of the form

$$\psi(x) \to \psi'(x) = \exp[ig_S \boldsymbol{\alpha}(x) \cdot \hat{\boldsymbol{T}}]\psi(x) \quad (2.5)$$

Here $\hat{\boldsymbol{T}} = \{T^a\}$ are the 8 generators of the $SU(3)$ group and $\boldsymbol{\alpha}(x)$ is a set of 8 arbitrary functions. Each generator of $SU(3)$ is represented by a 3x3 matrix, requiring the addition of three degrees of freedom to the description of the fermionic wavefunction $\psi(x)$. The orthogonal states in this new three-dimensional space are the three color eigenstates: red, green and blue. Because these color states define the gauge symmetry, it is commonly referred to as $SU(3)_C$ in the context of QCD. All quarks carry a color charge while antiquarks carry an anticolor charge, and the combination of a matching color and anticolor charge (e.g. blue and anti-blue) is colorless. Strong interactions occur via the exchange of gluons, which carry one color and one anticolor charge. There are 8 gluons total, corresponding to the 8 generators of $SU(3)$. The analogy between QCD and QED can be seen directly in the QCD Lagrangian

$$\mathcal{L}_{QCD} = -\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a + i\bar{\psi}_i\gamma^\mu\partial_\mu\delta_{ij}\psi_j + g\bar{\psi}_i\gamma^\mu(T_a)_{ij}\mathcal{A}^a_\mu\psi_j - m\bar{\psi}_i\delta_{ij}\psi_j \quad (2.6)$$

The gluon exchange term, which contains the gluon fields $\mathcal{A}^a_\mu$, is the only term that can mix different color components of the wavefunction ψ. The gluon field strength tensor $G^a_{\mu\nu} = \partial_\mu \mathcal{A}^a_\nu - \partial_\nu \mathcal{A}^a_\mu + g f^{abc} \mathcal{A}^b_\mu \mathcal{A}^c_\nu$ replaces the electromagnetic field strength tensor to form the kinetic term of the fields.

A unique aspect of QCD is the requirement that free particles exist in colorless states. This is called color confinement and prevents quarks from existing outside of bound states such as baryons (qqq), such as protons and neutrons or mesons ($q\bar{q}$). While an analytical proof of this mechanism does not exist, it can be understood through the fact that the energy stored in the field between two separated quarks grows linearly with the separation distance $V(r) \sim \kappa r$ (with $\kappa \sim 1$ GeV/fm). This stands in stark contrast to the $\frac{1}{r}$ potential in electrodynamics which allows for the existence of free electric charge. Color confinement has wide-ranging implications for collider physics, which is discussed in section 2.4.

2.2 The Electroweak Sector

The theories of Quantum Electrodynamics and Quantum Chromodynamics share most of their structure, up to their unique symmetry groups that give rise to their respective exchange particles and conserved quantities. The third kind of interaction described by the SM, the weak interaction, follows a similar blueprint with some important differences. Among its peculiarities are its three massive exchange bosons $W^\pm$ and Z, which stand in contrast to the massless gauge bosons in QED and QCD. Because of the EM charge carried by the W bosons, weak interactions consist of charged-current interactions mediated by $W^\pm$ as well as neutral-current interactions mediated by Z. It also has the unique tendency to violate the combined symmetries of charge conjugation and parity (CP), which does not occur in QED and is allowed but constrained to be very rare in QCD. The weak interaction is fundamentally linked to the electromagnetic interaction through the manifestation of a combined $SU(2) \times U(1)$ symmetry in the SM. This symmetry in particular, or rather the breaking thereof, is key to understanding the role of the Higgs boson.

2.2.1 The Charged Current

The charged-current weak interaction occurs via the exchange of charged $W^\pm$ bosons and was first studied in nuclear beta decay, where its parity violating properties were observed. Examining beta decays in an applied magnetic field showed a preferred direction for the emission of the electron relative to the B field, suggesting that the weak interaction maximally violates parity. This stands in stark contrast to both QED and QCD, whose Lagrangians have interaction terms with Lorentz covariant currents of the form $j^\mu = \bar{u}(p')\gamma^\mu u(p)$, which transform as four-vectors and are thereby inherently parity conserving.

For this reason, the charged current of the weak interaction cannot take the same mathematical form as the QED and QCD currents. Assuming Lorentz covariance, the SM does not inherently prevent interactions with currents that transform as scalars, pseudoscalars, axial

vectors or tensors. Since in each case the rank of the object would correspond to the spin of a hypothetical exchange boson, the spin of the W boson suggests an axial vector current of the form $\bar{\psi}\gamma^\mu\gamma^5\phi$. In reality, parity violation requires a current with both vector and axial vector components. For the weak interaction this current is $j^\mu = \frac{g_W}{\sqrt{2}}\bar{u}(p')\frac{1}{2}\gamma^\mu(1-\gamma^5)u(p)$, which represents a maximally parity violating V-A (i.e. vector - axial vector) interaction. Notably, this term contains the left-handed chiral projection operator, which eliminates right-handed components of particle wavefunctions. As such, the charged-current weak interaction can only occur between left-handed particles (or right-handed anti-particles). Even though all fundamental quarks and leptons are able to interact weakly, only their left-handed components do so in practice.

In the language of fundamental symmetries, this interaction is associated with an invariance under local $SU(2)$ phase transformations for left-handed particles (and right-handed anti-particles) only. This is denoted $SU(2)_L$ and takes the form

$$\psi(x) \rightarrow \psi'(x) = \exp[ig_W\boldsymbol{\alpha}(x)\cdot\hat{\boldsymbol{T}}]\psi(x) \tag{2.7}$$

where $\hat{\boldsymbol{T}} = \frac{1}{2}\boldsymbol{\sigma}$ are the three generators of the SU(2) group (i.e. the Pauli spin matrices). The new interaction terms in the Lagrangian originating in this symmetry are $ig_W\frac{1}{2}\sigma_k\gamma^\mu W^k_\mu\psi_L$, and the physical $W^\pm$ bosons are linear combinations of $W^{(1)}$ and $W^{(2)}$ given by $W^\pm_\mu = \frac{1}{\sqrt{2}}(W^{(1)}_\mu \mp iW^{(2)}_\mu)$. The $W^{(3)}$ is related to the Z boson which facilitates the neutral-current interaction. Due to the 2x2 nature of the generators of $SU(2)$, ψ_L consists of two components making up a doublet (analogous to the color triplet in QCD). The two states of this doublet represent different physical particles separated by one unit of charge. They are connected via the conservation of weak isospin I_3 and the exchange of a $W^\pm$ boson allows the system to transition between these two states. This means that left-handed particles are represented as

$$\begin{pmatrix}\nu_l\\ l^-\end{pmatrix}_L, \begin{pmatrix}u\\ d'\end{pmatrix}_L \tag{2.8}$$

while right-handed particles are placed in $SU(2)_L$ invariant weak isospin singlets, thereby linking left-handed neutrinos with charged leptons and up-type with down-type quarks via the charged current weak interaction. The d' in the above description does not represent a pure down quark state but rather a combination of down-type quark states with marginal contributions from strange and bottom quarks.

2.2.2 CP Violation and the CKM Matrix

The details of how the charged-current weak interaction manifests differ in practice between leptons and quarks. In leptons, $W^\pm$ exchange allows for interactions between charged leptons and their corresponding neutrinos. Lepton flavor needs to be conserved in these interactions, so electron neutrinos only interact with electrons, muon neutrinos with muons and tau neutrinos with taus. The strength of each is identical, captured in a concept called

"lepton universality" in weak interactions. While parity and charge conjugation are maximally violated as individual symmetries due to the left-handedness of the weak interaction, the leptonic charged-current conserves CP symmetry.

Because of this, CP violation in weak interactions must come from the quark sector. In particular, attempts to extend lepton universality into a universal weak coupling for quarks and leptons proved unfruitful, as differences in interaction strengths between leptonic and hadronic weak interactions were observed. This observation also extended to different kinds of hadronic interactions, leading to the formulation of the Cabibbo hypothesis, which states that the mass eigenstates of quarks are not equal to their weak eigenstates. A universal coupling therefore exists when considering the weak eigenstates q', which are related to the mass eigenstates q via the Cabibbo-Kobayashi-Maskawa (CKM) matrix [24, 165]

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.9}$$

The nine elements of this matrix can be parametrized via three rotation angles and six complex phases because of unitarity conditions. Five of these phases are unphysical and can be absorbed into relative redefinitions of the six quark fields, leaving a single physical complex phase. The misalignment of quark eigenstates leads to observable instances of CP violation in weak interactions involving quarks, which is one of the main identified sources of CP violation within the SM. The CKM matrix itself has been studied thoroughly via experiments, with the most recent constraints showing off-diagonal elements to be relatively small in general [146]. This is particularly true for the top quark which almost exclusively decays to bottom quarks rather than strange or down quarks.

2.2.3 Neutrino Oscillations

CP violation is also possible in the lepton sector through the phenomenon of neutrino oscillations [15], which is analogous to the CKM mechanism for quarks. Neutrinos are produced via charged-current weak interactions in association with charged leptons in one of three flavor eigenstates ν_e, ν_μ and ν_τ according to lepton flavor conservation. In a similar fashion to the Cabibbo hypothesis for quarks, they also have three distinct mass eigenstates ν_1, ν_2 and ν_3 that do not map directly onto their flavor states. The exact relationship between these two sets of eigenstates is defined via a unitary matrix called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [178]. Because the time evolution operator in quantum mechanics is unique to a given physical mass eigenstate, each mass component in the original flavor eigenstate evolves with a different phase, leading to oscillations between neutrino flavor states. A neutrino that is produced in the pure ν_e flavor eigenstate can later interact as a ν_μ state with a likelihood determined by the corresponding matrix coefficient. While this does not automatically imply CP violation, the presence of an irreducible complex phase in the PMNS matrix (analogous to the CKM phase) can lead to an asymmetry between the

neutrino and antineutrino oscillation probabilities and thus leptonic CP violation. If neutrinos are their own antiparticles, two additional complex CP-violating phases are introduced due to the inability to rephase self-conjugate fields. The PMNS matrix is constrained to be much less diagonal than the CKM matrix, making neutrino flavor oscillations a frequently occurring phenomenon [15]. The actual neutrino masses as well as the mass hierarchy of the eigenstates are poorly constrained and present an active area of research.

2.2.4 Electroweak Unification

The local $SU(2)_L$ symmetry of the weak interaction has three generators, implying three conserved currents and thus three exchange particles. The first two currents j_+ and j_- correspond to the $W^\pm$ bosons which mediate the charged-current weak interaction. The third current, j_3, is related to the neutral-current interactions mediated by the Z boson. Because of the empirical observation that Z boson exchange occurs between particles regardless of handedness, this correspondence cannot be exact. The current j_3, like j_1 and j_2, applies to left-handed particles and right-handed anti-particles only. This complicates the previous description of the weak interaction significantly, requiring a modification of the presumed symmetry. In practice, the fundamental $U(1)$ symmetry of electromagnetism needs to be integrated with the $SU(2)_L$ symmetry of the weak interaction. Thus, the $U(1)$ symmetry cannot couple to electric charge as discussed earlier, but rather a different quantity called weak hypercharge Y. The resulting interaction term $g'\frac{Y}{2}\gamma^\mu B_\mu\psi$ is analogous to QED with an altered coupling and replacement of the photon field A_μ by B_μ. The actual photon and Z boson fields are

$$A_\mu = B_\mu \cos\theta_W + W_\mu^{(3)} \sin\theta_W \tag{2.10}$$

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^{(3)} \cos\theta_W \tag{2.11}$$

which are linear combinations of the $U(1)_Y$ field and the neutral $SU(2)_L$ field. The weak mixing angle θ_W characterizes the strength of the mixing. The electron charge is related to the $U(1)_Y$ and $SU(2)_L$ couplings as well as the mixing angle by $e = g_W \sin\theta_W = g' \cos\theta_W$. Putting all of this together results in a theory with a combined electroweak symmetry given by $SU(2)_L \times U(1)_Y$. While this provides a unified description of electromagnetism and weak sector physics, a significant issue remains: the structure of this theory does not allow for the existence of particle masses. Solving this problem requires the addition of a new mechanism and a new particle: the Higgs.

2.3 The Higgs Boson

The previously discussed symmetries of the Standard Model require that each term in the Lagrangian $\mathcal{L}_{SM}$ of the complete theory be invariant under $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$

local gauge transformations. The QED, QCD and weak components of $\mathcal{L}_{SM}$ are constructed with these symmetries in mind, but only contain kinematic terms and interaction terms between fermions and exchange bosons. Particle masses need to be added to the Lagrangian explicitly via self-interaction terms, which have the form $\frac{1}{2}m_V^2 V_\mu V^\mu$ for bosons and $m_f \bar{f} f = m_f(\bar{f}_R f_L + \bar{f}_L f_R)$ for fermions. These extra terms are fundamentally incompatible with the prescribed symmetries of the SM since they are not invariant under $SU(2)_L \times U(1)_Y$ transformations by default. This means without further modifications, the Standard Model is incompatible with the existence of massive particles, which stands in direct conflict with our understanding of the physical universe. Rectifying this necessitates the addition of a mechanism which spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry alongside a new scalar particle.

2.3.1 Scalar Fields and Symmetry Breaking

The mechanism that allows for particle mass terms in the Lagrangian of the SM is known as spontaneous symmetry breaking [151, 134] and occurs in the electroweak sector of the theory. The foundation of this mechanism is the introduction of a new complex scalar field ϕ (the Higgs field) with a Lagrangian of the form $(D_\mu\phi)^*(D^\mu\phi) - V(\phi)$ where $V(\phi) = \mu^2\phi^2 + \lambda\phi^4$ is the Higgs potential and D_μ represents the covariant derivative[3]. For a single complex scalar field $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$, the potential expands to

$$V(\phi) = \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) + \frac{1}{4}\lambda(\phi_1^2 + \phi_2^2)^2 \tag{2.12}$$

where $\lambda > 0$ is enforced to ensure the potential has a finite minimum. The choice of μ then determines the shape of the potential, with $\mu^2 > 0$ resulting in a minimum at 0 and $\mu^2 < 0$ in a set of minima defined by $\phi_1^2 + \phi_2^2 = -\frac{\mu^2}{\lambda}$, as visualized in Fig. 2.2. In the case where $\mu^2 < 0$, this potential requires a physical choice for the minimum (vacuum) state which spontaneously breaks the rotational symmetry in the (ϕ_1, ϕ_2) plane. The minimum energy of the system, also known as the vacuum expectation value (or VEV), is at $v = \sqrt{\frac{-\mu^2}{\lambda}}$ for such a theory.

In the SM, the scalar field is a complex $SU(2)_L$ doublet described by

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.13}$$

which has an infinite number of minima where $\phi^\dagger\phi = \frac{v^2}{2}$ given $\mu^2 < 0$. In order for the photon to remain massless, the vacuum expectation value is only non-zero for the ϕ^0 component of the doublet. The field can be expanded around the minimum in unitary gauge as

[3]The covariant derivative enforces gauge invariance as in the QED and QCD Lagrangians.

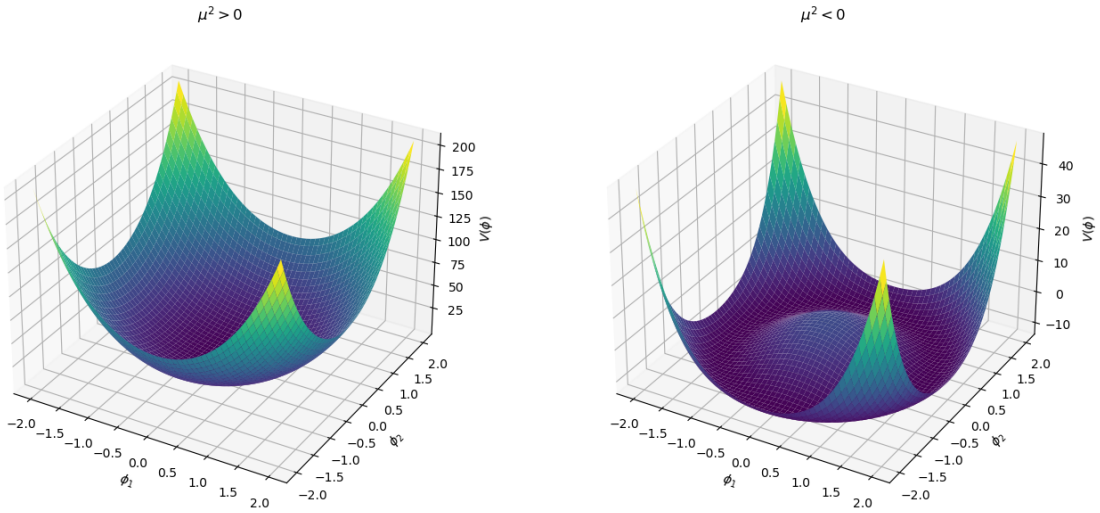


Figure 2.2: Visual representation of the Higgs potential (with $\lambda = 2$) for positive $\mu^2 = 10$ (left) and negative $\mu^2 = -10$ (right).

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.14}$$

where $h(x)$ directly represents the physical Higgs field. An equivalent, but more general description of this expansion around the VEV requires including an additional perturbing field around the minimum of ϕ^+ and then removing non-physical fields via gauge transformations afterwards. In the end, the symmetry breaking of this new scalar potential immediately results in the introduction of weak boson mass terms in the Lagrangian as well as the Higgs boson itself.

2.3.2 Particle Masses and Higgs Interactions

The new terms can be seen directly when expanding the kinematic term of the scalar field Lagrangian around the VEV as

$$\begin{aligned}(D_\mu\phi)^\dagger(D^\mu\phi) =& \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{8}g_W^2(W_\mu^{(1)} + iW_\mu^{(2)})(W^{(1)\mu} - iW^{(2)\mu})(v+h)^2 + \\ & \frac{1}{8}(g_W W_\mu^{(3)} - g'B_\mu)(g_W W^{(3)\mu} - g'B^\mu)(v+h)^2 \\ =& \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{4}g_W^2 W_\mu^- W^{+\mu}(v+h)^2 + \frac{1}{8}(g_W^2 + g'^2)Z_\mu Z^\mu (v+h)^2\end{aligned} \tag{2.15}$$

where the physical particles are written as combinations of the underlying fields $W^\pm = \frac{1}{\sqrt{2}}(W^{(1)} \mp iW^{(2)})$ and $Z = (g_W^2 + g'^2)^{-\frac{1}{2}}(g_W W^{(3)} - g'B)$ as discussed in section 2.2. In this

expression, the v^2 terms represent the weak boson masses (i.e. self-interaction terms) while the vh and h^2 terms describe the couplings (i.e. interaction strengths) between two weak bosons and one or two Higgs bosons, respectively. From the Lagrangian, the weak boson masses are thus $m_W = \frac{1}{2} g_W v$ and $m_Z = \frac{1}{2} v \sqrt{g_W^2 + g'^2}$. As expected, there is no mass term for the photon and it does not interact with the Higgs directly. Given the orthogonality requirement to the Z, the photon is represented as $A = (g_W^2 + g'^2)^{-\frac{1}{2}}(g' W^{(3)} + g_W B)$. The potential term of the scalar field Lagrangian can be expanded in the same fashion as

$$V(\phi) = \frac{1}{2}\mu^2(v+h)^2 + \frac{1}{4}\lambda(v+h)^4 = \lambda v^2 h^2 + \lambda v h^3 + \frac{1}{4}\lambda h^4 - \frac{1}{4}\lambda v^4 \tag{2.16}$$

where the h^2 terms define the Higgs mass $m_H = \sqrt{2\lambda} v$. Self-interactions between three or four Higgs bosons are governed by the h^3 and h^4 terms, which define the Higgs self-couplings.

Fermion masses can be introduced to the SM through similar means, but do not directly arise from spontaneous symmetry breaking explicitly. Gauge invariant fermion mass terms take the form $-g_f(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L)$ and $g_f(\bar{\psi}_L \phi_C \psi_R + \bar{\psi}_R \phi_C^\dagger \psi_L)$ where $\phi_C = -i\sigma_2 \phi^*$ is the conjugate Higgs doublet. Both of these expressions satisfy the $SU(2)_L \times U(1)_Y$ symmetry of the Standard Model as required. Since the VEV is only non-zero for the lower component of the scalar $SU(2)$ doublet (i.e. ϕ^0), the first expression generates the charged lepton and down-type quark terms while the second is responsible for the up-type quark terms, brought together in the fermionic part of the SM Lagrangian as

$$\mathcal{L}_f = -\frac{g_f}{\sqrt{2}} v(\bar{f}_L f_R + \bar{f}_R f_L) - \frac{g_f}{\sqrt{2}} h(\bar{f}_L f_R + \bar{f}_R f_L) \tag{2.17}$$

Because there are three generations of quarks and leptons, the couplings g_f arise from 3×3 complex matrices in generation space which are diagonalized to yield the physical fermion mass eigenvalues and generate the CKM matrix in the quark sector and the PMNS matrix in the lepton sector.

Neutrinos are still unaccounted for in this description due to their apparent fixed left-handedness. If right-handed neutrinos do exist and they are Dirac particles, no modification to the Yukawa mechanism is required to justify the existence of neutrino masses. This scenario leads to significantly smaller Yukawa couplings for neutrinos than all other fermions, leaving open questions about the reasons for these differences. If neutrinos are instead Majorana particles, their tiny masses can be generated naturally via additional BSM mechanisms [179, 177] which are still an open mystery. This possibility is especially pertinent given the difference in boson and fermion mass generation already present in the SM. For all other fermions, the Yukawa coupling g_f determines the coupling strength between a particular fermion species and the Higgs. While not completely fixed by theory, in the SM the relationship $g_f = \sqrt{2}\frac{m_f}{v}$ holds. Various BSM models predict modifications to this relationship for some fermions, which could have a significant impact on the mechanisms of fermionic mass generation in general [147, 28]. Deviations in Yukawa couplings are measurable via changes in interaction rates between the Higgs and different fermion species, motivating the need to perform precision measurements of these processes.

The Standard Model is a mathematically robust and broadly experimentally tested theory, but many open questions remain. In its minimal formulation it has 19 free parameters that are not fixed by theoretical predictions. Of these, 13 are related to the quark Yukawa sector (generating 6 quark mass eigenvalues, 3 CKM mixing angles and 1 CKM phase), 3 are the gauge coupling strengths, 2 describe the Higgs potential and 1 specifies the strong CP-violating parameter. When including non-zero neutrino masses, an additional minimum of 7 free parameters are introduced (3 masses, 3 oscillation angles and at least 1 oscillation phase). All 20 flavor parameters across the quark and lepton sectors stem directly from Yukawa interactions, making the fermion mass-generation mechanism a central focus of modern precision studies.

2.3.3 Higgs Boson Phenomenology

The physical Higgs boson is a measurable consequence of the Higgs mechanism and provides a natural pathway to its study. The SM predicts the Higgs to be a neutral scalar particle with no color charge and even parity, but does not inherently provide information about its mass. Its predicted mean lifetime of 1.56×10^{-22} s is very short and its decay width of $\Gamma_H = 4.1$ MeV very narrow, making it difficult to measure directly [6]. Higgs bosons can be most easily produced in particle collisions, assuming the accessible energy is at least equal to the rest mass of the Higgs itself. In this context, there are 4 common modes of Higgs production: gluon-gluon fusion (ggF), vector boson fusion (VBF), associated production with a vector boson or "Higgs Strahlung" (VH) and associated production with a quark pair, most commonly top quarks due to their large masses and corresponding Yukawa couplings ($t\bar{t}H$). Feynman diagrams for these processes are shown in Fig. 2.3. The cross section σ for each production mode depends on both the center of mass energy $\sqrt{s}$ achieved during the collision as well as the actual mass of the Higgs boson itself, as shown in Fig. 2.4. This represents the probability that a given process will occur.

Measurements of Higgs boson properties are generally performed by observing its decay products due to its short lifetime. Branching ratios are used to describe the fraction of times a particular decay mode j occurs as $BR(j) = \frac{\Gamma_j}{\Gamma}$ where $\Gamma = \frac{1}{\tau}$ is the total decay rate and is related to its overall lifetime. Variations in theoretical predictions for the most common decay branching ratios of the Higgs are shown in Fig. 2.5 for different Higgs masses. Since the Higgs does not couple directly to photons or gluons, decays involving these particles occur through a virtual quark loop making them comparatively uncommon. The coupling of the Higgs to different particles is typically proportional to their masses, leading to larger branching ratios for decays to heavier particles[4].

The Higgs boson was discovered in 2012 at $m_H \sim 125 - 126$ GeV through observations in the uncommon $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to 4l$ decay channels since the unique signatures of these decays provide high signal purity [40, 113]. Since then many of its properties have been

[4]This assumes the rest mass of the decay products is equal to or lesser than the rest mass of the Higgs boson itself, otherwise the decays are suppressed.

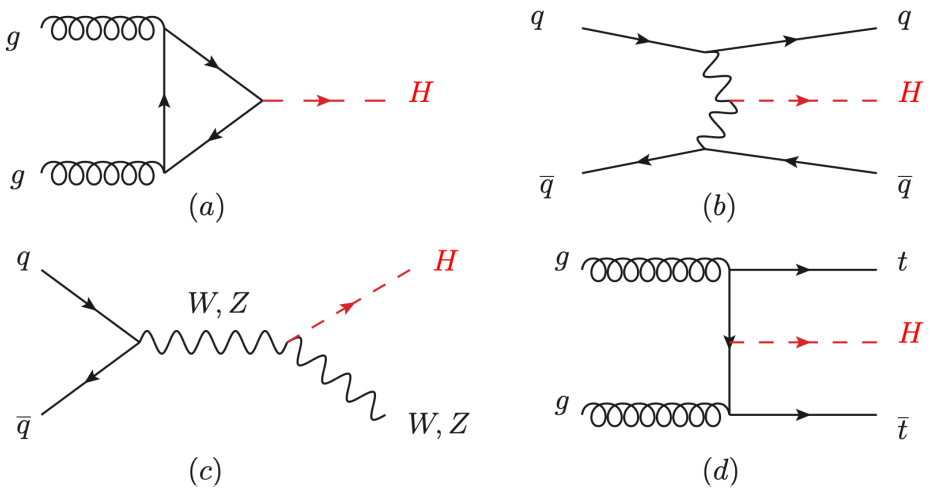


Figure 2.3: Feynman diagrams for the most common Higgs production modes: (a) ggF, (b) VBF, (c) VH and (d) $t\bar{t}H$ from [27].

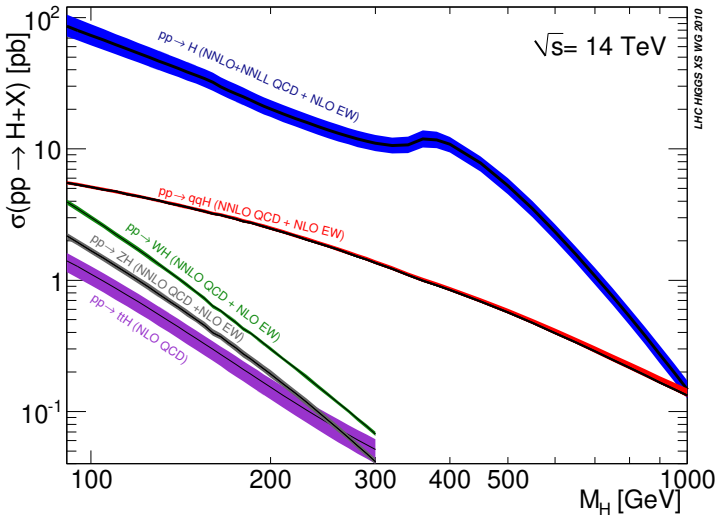


Figure 2.4: Predicted cross sections for different Higgs production modes in hadron collisions at $\sqrt{s} = 14$ TeV as a function of m_H from [171].

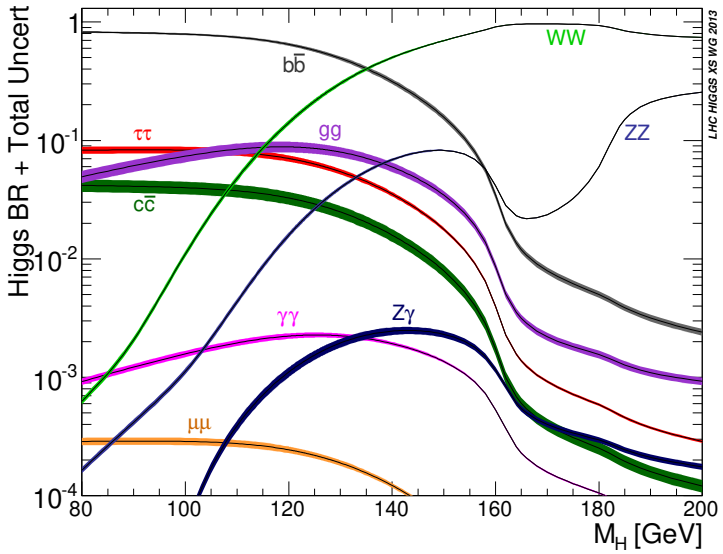


Figure 2.5: Predicted branching ratios for Higgs boson decays as a function of m_H from [171].

measured with increasing accuracy. Its mass has been constrained to $m_H = 125.22 \pm 0.14$ GeV [93, 94, 92] and its width to $\Gamma_H = 3.2^{+2.4}_{-1.7}$ MeV [210]. Various Yukawa couplings have also been constrained through direct measurements by fitting the signal strength

$$\mu_{i\to H\to f} = \frac{\sigma(i \to H) \times BR(H \to f)}{(\sigma(i \to H) \times BR(H \to f))_{SM}} \tag{2.18}$$

of a particular combination of Higgs production and decay processes. Large deviations from the theory expectation of $\mu_{i\to H\to f} = 1$ indicate the possible presence of new physics beyond the SM. Contributions from modified production cross sections and decay branching ratios can be isolated via the κ-framework [190], which makes their impact on the Yukawa couplings explicit. This is done by defining a set of constants κ_X which relate to the signal strength as $\mu_{i\to H\to f} = \frac{\kappa_i^2 \kappa_f^2}{\kappa_H^2}$. These constants represent multiplicative factors for the various Yukawa couplings where $\kappa_H = \sum_j \kappa_j^2 \frac{\Gamma_j^{SM}}{\Gamma_H^{SM}}$ relates to the Higgs decay width directly. Measurements of third generation quark, second and third generation lepton and vector boson couplings show excellent agreement with SM predictions as shown in Fig. 2.6. These couplings are well constrained due to either being relevant in Higgs production (such as top), having high decay branching ratios (such as bottom) or providing clear measurement signatures (such as μ). Yukawa couplings for first and second generation quarks are much less understood due to their low predicted branching ratios summarized in table 2.1 and indistinct signatures compared to other decays, which are detailed in chapter 4.2.

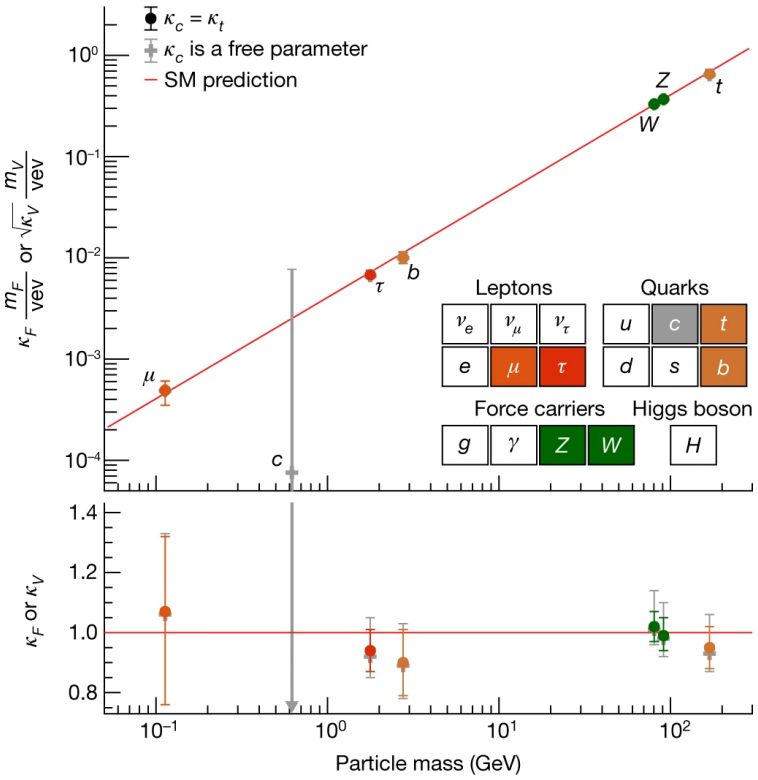


Figure 2.6: Summary of theoretical and experimental agreement for various Yukawa couplings from [78].

2.4 Collider Physics

The study of Higgs boson properties requires generating particle collisions with enough energy to produce them at statistically significant rates. This is best accomplished in the environment of a high energy particle collider, which facilitates particle interactions in a vacuum. The higher the collision center of mass energy provided by the collider, the larger the number of accessible interactions for study. To produce Higgs bosons and study their decays, this energy needs to minimally be equal to the rest mass of the Higgs. Both leptons (usually electrons or positrons) and hadrons (usually protons or anti-protons) can be accelerated in collider environments, with unique advantages and disadvantages in either case. Electrons provide clean interaction signatures since they are fundamental particles themselves, but lose

Decay channel	Branching ratio [%]
$H \to b\bar{b}$	57.5 ± 1.9
$H \to W^+W^-$	21.6 ± 0.9
$H \to gg$	8.56 ± 0.86
$H \to \tau^+\tau^-$	6.30 ± 0.36
$H \to c\bar{c}$	2.90 ± 0.35
$H \to ZZ$	2.67 ± 0.11
$H \to \gamma\gamma$	0.228 ± 0.011
$H \to Z\gamma$	0.155 ± 0.014
$H \to \mu^+\mu^-$	0.022 ± 0.001

Table 2.1: Summary of Higgs decay modes with associated branching ratios based on Standard Model predictions from [190].

energy rapidly during acceleration due to synchrotron radiation and bremsstrahlung and do not allow for probing QCD interactions to the same degree. Protons are easier to accelerate, but also result in messier collisions due to their constituent nature. The Higgs boson was discovered with a circular proton-proton collider at a center of mass energy of $\sqrt{s} = 7 - 8$ TeV. This collider accelerates two beams of protons simultaneously to maximize the available collision energy and is detailed in section 3.1.

2.4.1 Basics of Hadron Collisions

The square of the center of mass energy of a beam collider $s = (\sum_{i=1}^{2} E_i)^2 - (\sum_{i=1}^{2} \mathbf{p}_i)^2$ is one of its most important characteristics. In hadron colliders, this energy does not correspond directly to the collision energy since "hard scatter"[5] collisions occur between constituents called "partons" (i.e. quarks and gluons) rather than the hadrons themselves. This has three tangible consequences. The first is that the unknown parton momentum fraction in each collision needs to be accounted for via so-called parton density functions (PDFs) that get factored into calculations. The center of mass energy thus acts as an upper bound for the parton collision energy, which determines accessible final states. Secondly, event signatures are much messier in hadronic collisions compared to leptonic ones due to color confinement in QCD. In practice this is due to the fact that parton collisions impact other partons in their respective hadronic bound states. Finally, the beam energy requirement to make new particles is significantly lower compared to leptonic colliders due to the large amount of additional accessible energy when bound states are broken apart.

[5]In this context, "hard" refers to processes that occur at high enough energies to break apart the colliding hadrons while "soft" generally refers to lower energy radiation.

After a collision occurs, a detector is required to record the signatures of particles contained in the final state, which are used to reconstruct information about the underlying physical process. In essence, particle detectors are counting devices built to understand the frequency of a given process relative to expectations based on theory. The cross section σ, which represents the probability that a particular process will take place, is used to quantify this. Aside from $\sqrt{s}$, the second important quantity associated with the collider itself is its instantaneous luminosity $L = \frac{1}{\sigma}\frac{dN}{dt}$ which relates the cross section of a process to the expected occurrence rate. It provides a process independent measure of the intensity of particle collisions per unit area per unit time. The total integrated luminosity L_{int} (integrated over time) quantifies the total number of collision events recorded. Since beam accelerators collide many particles at once in so-called "bunches" there are generally multiple collisions occurring at a time. Of these, only a small subset are energetic enough to cause the proton to break up and thus interesting enough to study. These hard scatter interactions stand in contrast to the surrounding "pile-up" interactions, which need to be filtered. Increasing the instantaneous luminosity of a collider also increases the average number of pile-up interactions, requiring the use of more sophisticated filtering techniques to mitigate their impact.

In hard scatter collisions, the interacting particles are the fundamental partons rather than the bound protons which results in an additional layer of complexity. Given an interaction $pp \to X$, the total cross section is a combination of all possible parton pair cross sections. This combination is dependent on the proton PDFs (parton distribution functions) $f_{q/g}(x)$ which describe the probability of a particular type of parton holding a fraction x of the total proton momentum. The resulting combination of the soft long-range dynamics described in the proton PDFs and the instantaneous hard scatter parton collision of interest is a complicated QCD calculation. The factorization theorem within QCD [125] allows for the separate treatment of these two processes since the hard scatter collision can be treated perturbatively, leading to

$$d\sigma_{pp\to X} \sim \sum_{a,b} \int dx_{p1} f_a(x_{p1}, \mu_F) \int dx_{p2} f_b(x_{p2}, \mu_F) d\sigma_{ab\to X} \tag{2.19}$$

Here μ_F is a parameter called the factorization scale, corresponding to the energy cutoff where perturbative calculations can be performed. The underlying reason for this is the "running" of the QCD coupling α_S, which causes it to become small at high energies (in contrast to QED). The renormalization scale μ_R determines at which energy the QCD coupling is evaluated. While the cross sections $\sigma_{ab\to X}$ can be calculated using QCD, the proton PDFs need to be determined empirically via experiments [10].

2.4.2 Final States and Hadronization

The process $pp \to X$ inclusively describes all of the physics happening at the stage of the hard scatter collision to produce a collection of particles denoted X. These dynamics are

generally well understood mathematically and mostly calculable using QCD techniques. In many cases, X contains a relevant subset of unstable particles with short lifetimes τ, which means it does not generally correspond to the final state particles observed in the detector. An additional complexity comes from color confinement in QCD, which forces individual partons into bound states. This imposes extra constraints on the possible final states of any given collision event through a process called hadronization, which ensures partons involved in high-energy collisions end up in new bound states via the production of quark-antiquark pairs. QCD factorization allows for the separation of the hadronization process from the underlying collision event, making these calculations tractable. While hadronization is not entirely understood, it originates in the linear potential $V(\mathbf{r}) \sim \kappa\mathbf{r}$ (where $\kappa \sim 1\ GeV/fm$) experienced by quarks or gluons outside of bound states. As the distance between bound quarks increases due to the momentum imparted by a collision event, it eventually becomes energetically favorable to form new bound states via the spontaneous creation of new quark-antiquark pairs. This process continues until the partons within each bound state have a low enough energy to remain bound, thereby creating a number of new hadrons. In practice, these hadrons manifest as collimated sprays of particles originating from a common point, which are called particle jets. Jets can have different features depending on the flavor of their originating quark, which can be used to isolate certain final states as detailed in chapter 4.2.

2.4.3 Monte Carlo Simulations

Hadron collisions are very high dimensional events due to the large number of final state particles created in the combination of hard scatter collision, pile-up and initial- and final state radiation (ISR and FSR)[6]. Modelling their structure in a particle detector based on theory expectations requires the use of Monte Carlo (MC) techniques [161] to generate simulated collision events. Simulations are based on existing knowledge of and expectations from Standard Model physics or other underlying theories to be tested. MC event generation occurs in multiple steps, taking advantage of QCD factorization. It uses random sampling to generate a realistic distribution of events that can be compared against data to test an underlying theory.

The core of MC generators concerns the cross sections of hard scatter parton collisions outlined in equation 2.19, which are not analytically tractable. Using numerical techniques these can be estimated to some precision, usually leading order (LO), next-to-leading order (NLO) or next-to-next-to leading order (NNLO) in QCD. While these calculations are exact at infinite order, there are theoretical uncertainties associated with cutting them off at finite precision, which represents a trade-off born from high compute requirements. Some higher order corrections are approximated via parton showering algorithms which add additional parton contributions from ISR and FSR to the collision event. Once all of the final state partons are generated, hadronization algorithms are used to develop the final structure of

[6]ISR and FSR corresponds to low energy QCD and EM radiation happening before or after the hard scatter collision respectively.

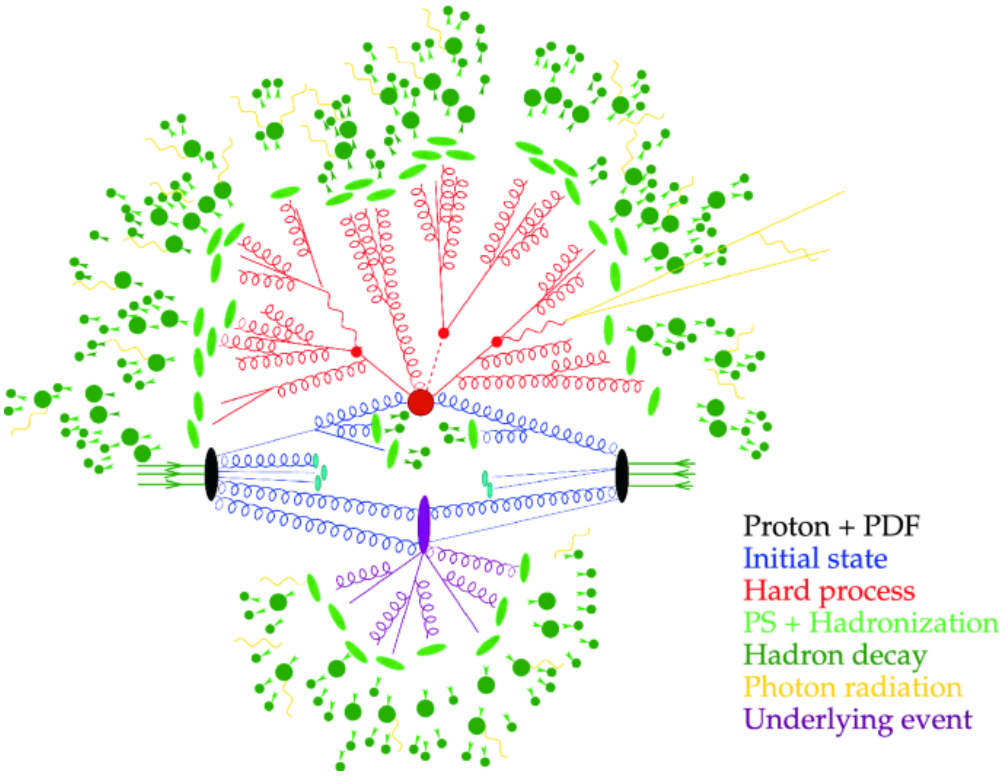


Figure 2.7: Visual representation of a $t\bar{t}$ production event as simulated by a Monte Carlo generator from [188].

the hard scatter collision. These algorithms are based on phenomenological models rather than derived from first principles and are not exact. As a further step, the "underlying event", which consists of soft and semi-hard activity from beam remnants and multiple parton interactions outside of the primary collision, is also generated. Pile-up from additional proton-proton collisions is modeled separately. All of the final state particles are then passed through a detector simulation to approximate the expected readout. A graphical summary of this pipeline is shown in Fig. 2.7.

Chapter 3

Experimental Apparatus

While the underlying theory provides the necessary motivation to pursue the study of particle physics in general and the Higgs boson in particular, an experimental approach is required to ensure it correctly describes physical reality. In the early history of the field, small detectors were sufficient to make great strides in the discovery of new particles and their interactions. Signatures were clear enough and backgrounds insignificant enough that modest amounts of data could lead to statistically significant results. Given the high rest mass of the Higgs boson at around 125 GeV and the rarity of its production, its study requires much more sophisticated tools. Its discovery necessitated the construction of the largest particle collider ever built with the Large Hadron Collider (LHC) alongside two general purpose detectors: ATLAS and CMS. Both experiments have a raw data flow of around one petabyte (PB) per second from 40 million collision events per second [33] with the ultimate aim of measuring a variety of quantities related to the fundamental structure of the universe, both through the Higgs boson as well as other particles in the Standard Model and beyond.

This chapter seeks to explain the technical details of the LHC and the ATLAS detector. An operational overview of the LHC and the CERN complex is given (section 3.1) followed by detailed descriptions of the various ATLAS detector components such as tracking detectors, calorimeters and muon detectors as well as the magnet and data acquisition (DAQ) systems (section 3.2). Emphasis is placed on the inner tracker, which is particularly important for measurements involving final state quarks, such as $H \to c\bar{c}$.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [21] is the largest proton-proton collider in the world, located at the CERN[1] laboratory near Geneva, Switzerland. It is a circular particle accelerator with a total circumference of 26.7 km used to bring two beams of protons to energies in the TeV range. Each beam consists of evenly spaced bunches (each containing billions

[1] "Organisation européenne pour la recherche nucléaire" or in its english translation "European Organization for Nuclear Research"

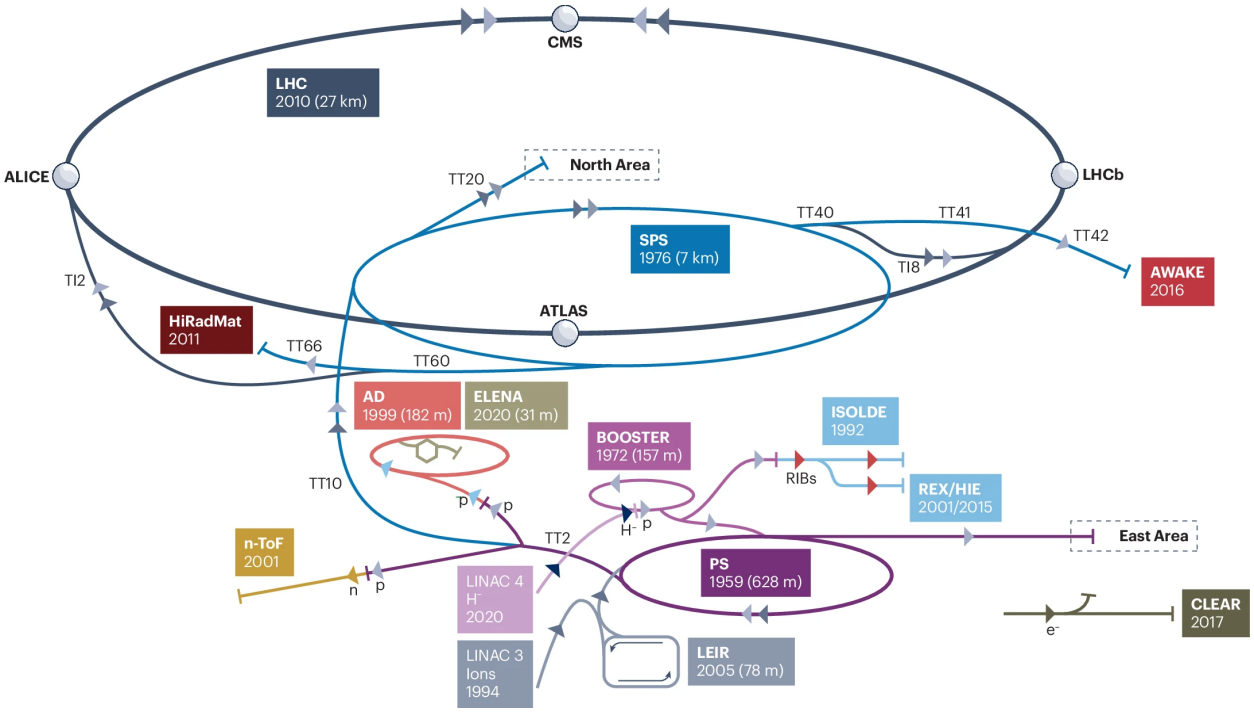


Figure 3.1: Schematic overview of the CERN accelerator complex including main synchrotron accelerators LHC, SPS and PS from [20].

of protons) with a minimum spacing of 25 ns [9], which are collided at 4 interaction points around the ring at rates of up to 40 MHz. In addition to protons, the LHC is also capable of accelerating and colliding heavy ions such as lead. Each interaction point is the site of a detector belonging to one of the 4 major LHC experiments: ATLAS, CMS, LHCb and ALICE. ATLAS [38] and CMS [112] are general purpose experiments designed to measure a wide variety of physical processes related to the Standard Model and BSM theories including measurements of the Higgs boson. The latter two have a more singularly dedicated focus: LHCb [122] was built to investigate CP violation and rare decays of B hadrons and ALICE [34] is dedicated to studying the quark-gluon plasma via heavy ion collisions.

Achieving the desired collision energies requires the use of multiple components of the CERN accelerator complex detailed in Fig. 3.1, which consists of linear and smaller circular colliders - many of which were operated as independent colliders in the past. The acceleration is facilitated by an intricate magnet system consisting of dipoles to guide protons beams along the circular path and quadrupoles to collimate and focus the beams as well as bring them together at the interaction points. Higher order multipole magnets are used for smaller field corrections. The acceleration process starts with H^- ions which are accelerated to 160 MeV by a linear accelerator called Linac4 [217]. From there the ions are injected into the circular Proton Synchrotron Booster (PSB), which strips both electrons from the Hydrogen

Run Number	Years	Center of mass energy	Integrated luminosity
1	2009-2013	$7-8$ TeV	26.4 fb^{-1}
2	2015-2018	13 TeV	147 fb^{-1}
3	2022-2026	13.6 TeV	300 fb^{-1} (up to and including 2025)

Table 3.1: Overview of run periods of the LHC between 2008 and 2026 with luminosity delivered to the ATLAS experiment for each run period from [101].

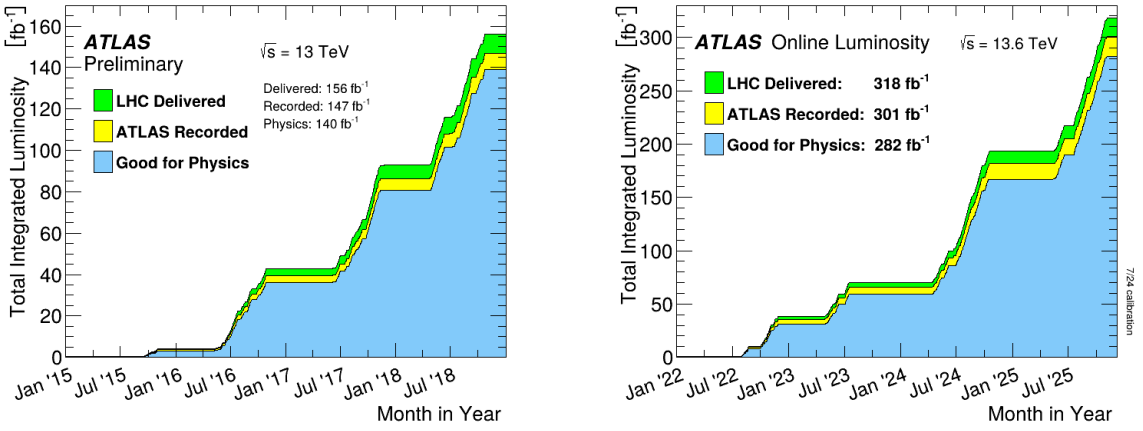
ions and creates a pure proton beam with an energy of 2 GeV [182]. Afterwards the beam is accelerated to 26 GeV by the Proton Synchrotron (PS) [1] and 450 GeV by the Super Proton Synchrotron (SPS) [2] before finally being injected into the LHC where it is accelerated until it reaches its final collision energy. The advantage of this modular approach lies in the ability to produce precise and stable guiding magnetic field configurations for protons within specific energy ranges.

The LHC has been operational since 2008, with the first proton-proton collisions occurring in 2009 at a center of mass energy of 0.9 TeV. By 2010 an energy of 7 TeV was reached, marking the start of the first data taking period. Operation has been split into three distinct run periods denoted runs 1, 2 and 3 with shutdowns in between for maintenance and accelerator as well as detector upgrades. Details on the timeline of these runs, alongside achieved energy and luminosity is shown in table 3.1. Due to the much lower integrated luminosity of run 1, this research focuses on data collected during runs 2 and 3.

The ATLAS detector records around 95% of luminosity delivered by the LHC and around 85-90% of recorded events are useable for physics purposes in runs 2 and 3, as shown in Fig. 3.2. Discrepancies between delivered and recorded luminosities arise from inefficiencies in data acquisition such as limits in the event processing rates. Recorded data is marked as "good for physics" assuming there are no major issues with detector subsystems or other concerns related to data quality. These two run periods differ mainly by a slight increase in center of mass energy from 13 TeV to 13.6 TeV as well as an increase in the mean number of interactions per bunch crossing[2] from $<\mu>=34$ to $<\mu>=55$ as recorded by ATLAS and pictured in Fig. 3.3. This difference results in a higher instantaneous luminosity for run 3, increasing total data volume as well as complexity of each recorded collision event due to excess pile-up.

A luminosity upgrade of the LHC is currently in preparation with first collisions expected in mid 2030 after significant upgrades to the LHC magnet system are completed. This High Luminosity LHC (HL-LHC) [183] is projected to increase the maximum instantaneous luminosity from the current 2×10^{34} cm^{-2}s^{-1} to 7.5×10^{34} cm^{-2}s^{-1}. This will result in a much larger number of average collisions per event at $<\mu>=200$.

[2]Collisions per bunch crossing refers to the number of protons that interact during the collision of two proton bunches, which are functionally simultaneous interactions.



(a) (b)

Figure 3.2: Plots showing data quality in the ATLAS experiment for (a) run 2 and (b) run 3 up to end of 2025 from [100]

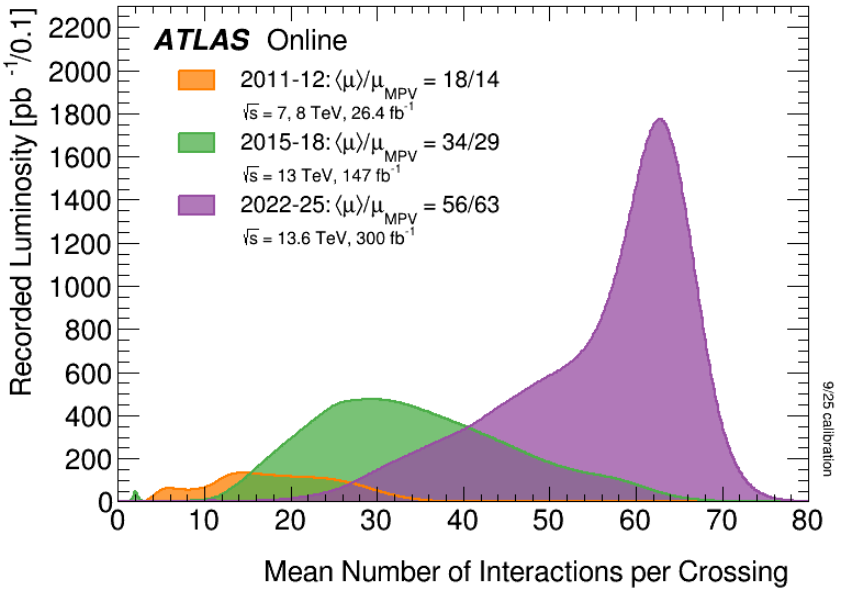


Figure 3.3: Luminosity vs. mean number of interactions for runs 1, 2 and 3 (up to September 17th 2025) as recorded by ATLAS from [101].

3.2 The ATLAS Detector

ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) [38] is the largest of the four main detectors around the LHC ring. It is a general purpose detector designed to record the proton-proton collision signatures generated in the LHC without explicit focus on any specific process or particle type. This multi-purpose approach necessitates a highly modular detector. It is around 46 meters long with a diameter of 25 meters and weighs 7000 tons. The detector, visualized in its original configuration in Fig. 3.4, consists of four main active components: an inner tracking detector, two calorimeters and a muon spectrometer. A series of upgrades have since been completed on all subsystems which are detailed in the corresponding sections.

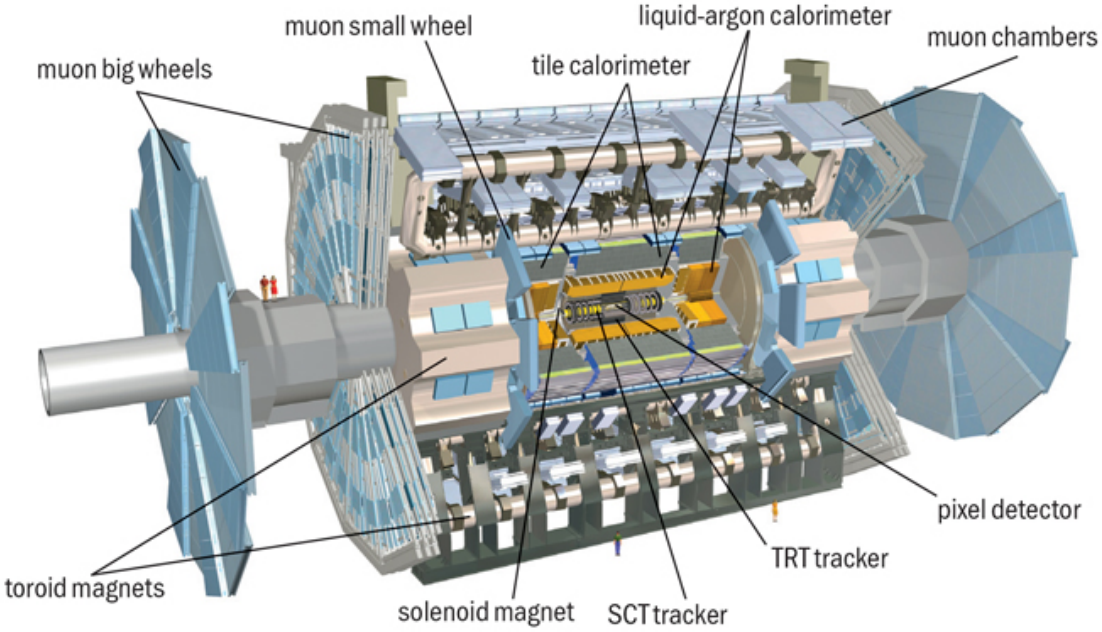


Figure 3.4: Diagram showing all components of the ATLAS detector in its original configuration from [38].

Each detector component interacts differently with different types of particles, as summarized in Fig. 3.5 and further detailed in section 4.1. The inner tracking detector [203, 204] is located closest to the collision point and is primarily used for precise trajectory measurements of charged particles. The two components of the calorimeter system [35, 36] surround the inner detector and are designed for precise energy measurements. An inner electromagnetic calorimeter measures the energies of electrons and photons and an outer hadronic calorimeter measures the energies of charged and neutral hadrons. Finally, a muon spectrometer [37] is located at the outside of the detector and is used to identify highly penetrating muons. Accurate measurements of particle momenta are possible because of a magnet system which

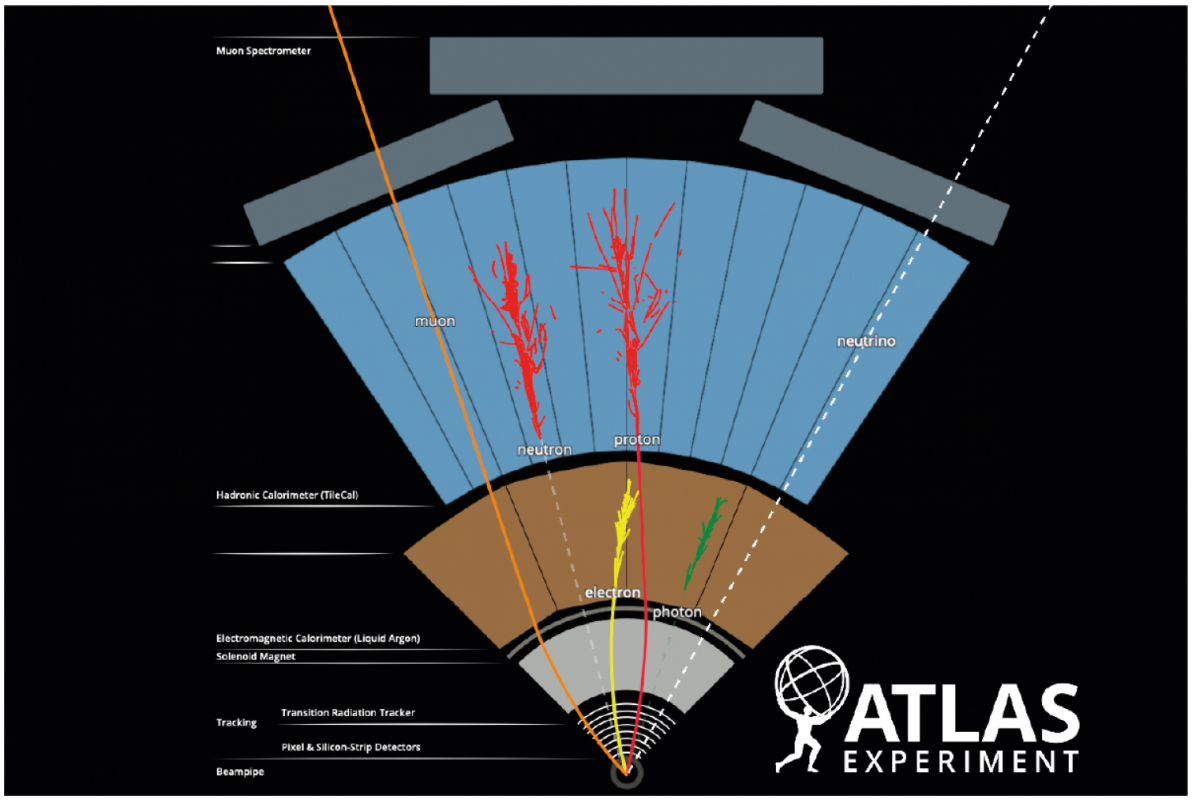


Figure 3.5: Sketch showing how different particles interact with detector components from [73]. Dashed tracks are invisible to the detector.

curves the trajectories of charged particles traveling through the detector. The ATLAS detector has undergone a series of upgrades performed during long shutdown periods between runs, which are detailed where relevant [98]. Past upgrades consist of the "Phase-0" upgrade implemented between runs 1 and 2 with a focus on the inner tracking detector and "Phase-I" upgrades implemented between runs 2 and 3 focused on the calorimeter system and muon spectrometer. A future Phase-II upgrade seeks to replace the inner tracking detector entirely with the fully Silicon-based "ITk" [140] in preparation for run 4 of the HL-LHC.

3.2.1 ATLAS Coordinate System

ATLAS uses a standardized coordinate system to describe both the detector and relevant physical objects. By convention, it is centered at the collision point, with the beam line representing the z-axis pointing towards the direction of the counterclockwise circulating beam, the x -axis pointing towards the center of the LHC and the y-axis pointing towards the sky, as seen in Fig. 3.6. The three important coordinates related to particle trajectories in this spherical coordinate system are the transverse momentum p_T, the azimuthal angle

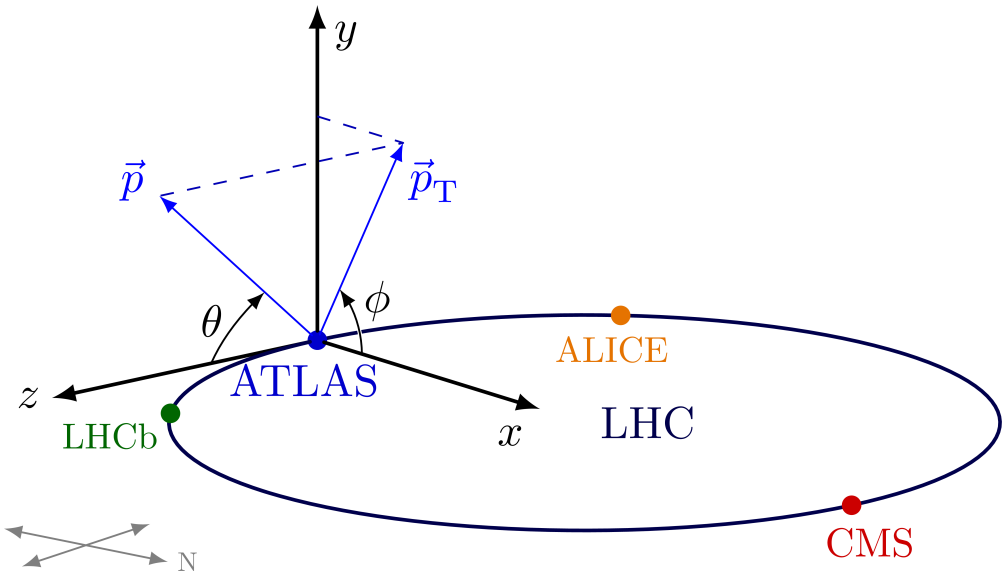


Figure 3.6: Diagram showing the orientation of the coordinate system used in ATLAS as well as relevant kinematic variables from [181].

ϕ and the pseudorapidity η. The first of these is the projection of the momentum vector of a particle onto the transverse xy-plane, which also defines the azimuthal angle measured starting from the positive x-axis. The final coordinate η is a measure of the longitudinal component[3] of a particle trajectory and is related to the polar angle θ by $\eta = -\ln(\tan(\theta/2))$. By definition, $\eta = 0$ if the trajectory is entirely transverse and $\eta \to \infty$ as it approaches the beamline. This coordinate representation is particularly useful in collision experiments. The use of transverse momentum allows for the isolation of momentum contributions from particle interactions. In particular, particles involved in hard scatter collisions frequently have high p_T. The preference of η compared to θ stems from the Lorentz invariance of $\Delta\eta$ under boosts along the z-axis for highly relativistic (i.e. massless) particles.

3.2.2 Inner Tracking Detector

The core of the ATLAS detector [38] is the inner tracking detector (ID), which is primarily responsible for precisely measuring the trajectories of charged particles by recording currents from electrons freed via the ionization of detector material. It is a combination of three different sub-systems: a pixel detector on the inside surrounded by a semiconductor tracker (SCT) and finally a transition radiation tracker (TRT) on the outside as shown in Fig. 3.7. The pixel detector and SCT are built from silicon wafers while the TRT is made up of

[3]This is frequently referred to as the "forward component"

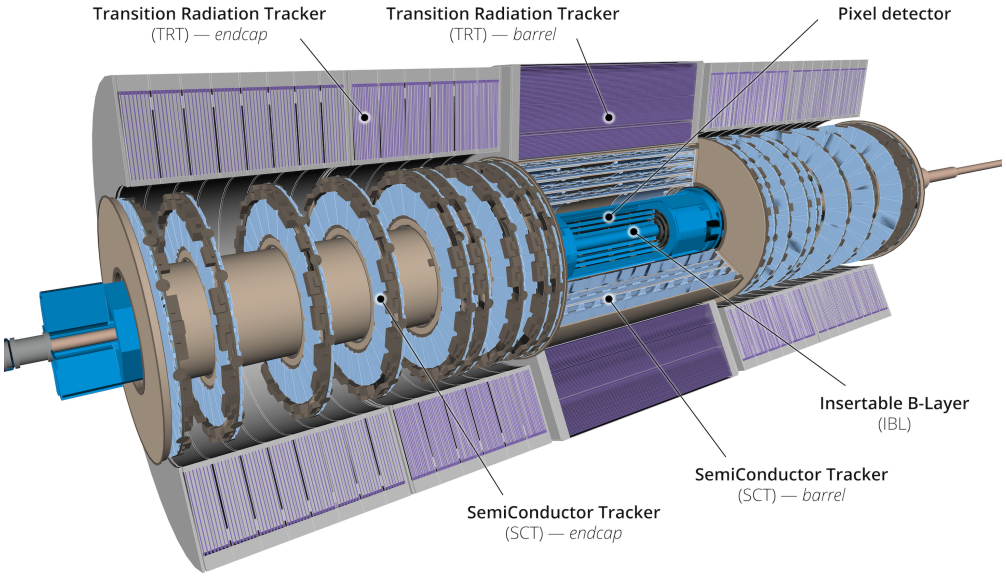


Figure 3.7: Schematic of the ATLAS inner detector and its components for run 3 from [98].

drift tubes filled primarily with a Xenon-based gas mixture[4]. Each subsequent part of the ID has larger elements and thus provides increasingly less granular position measurements, particularly along the z-axis. Geometrically the ID consists of barrel and endcap components where the barrel is a cylinder surrounding the beam and the endcaps are disks placed on either side of the barrel designed to detect forward particles with a total coverage of $|\eta| < 2.5$.

The pixel detector [185] is silicon-based and made up of 92 million individual sensor elements (pixels) arranged in 4 layers for the barrel component (1736 modules total) as well as 3 disk layers for each endcap (288 modules total). The innermost layer, called the insertable B-layer [60], was inserted prior to run 2 of the LHC and is located a distance of 33 mm from the beam pipe. It has the smallest individual pixel elements at 50×250 μm^2, allowing for increased position resolution closest to the beam. The remaining layers have pixels of size 50×400 μm^2, resulting in an overall spatial resolution of $\sim 8\mu$m in r/ϕ and $\sim 75\mu$m in z. All pixel sensors have a thickness of 250μm and provide measurements of particle charge and energy loss in addition to positional information. Individual sensor elements are not arranged end to end but rather partially overlap to ensure complete ϕ coverage of each layer. The final layer of the pixel detector is located at 122.5 mm from the beam. Due to the large amount of radiation exposure of the innermost detector layers, cumulative radiation damage leads to degradation of individual elements over time, which is modeled in detector simulations [69]. Due to its high granularity, the pixel detector provides

[4]Due to leaks, parts of the detector were switched to an Argon-based mixture to minimize losses of Xenon.

excellent precision for particle track reconstruction and drives the sensitivity of jet flavor identification as discussed in section 4.2.1.

Surrounding the pixel detector is the semiconductor tracker (SCT) [42], which is similar in design and functionality, but contains larger silicon-based detector elements (strips). It is made up of 4 barrel layers (2112 modules total) and 9 disk layers for each endcap (1976 modules total) starting at 299 mm from the beam for the first barrel layer and ending at 514 mm for the last. In total, the SCT contains over 6 million strips with a pitch[5] of 80μm and a length of 12 cm. To combat the low measurement resolution along the length of each strip, pairs of strips are rotated relative to each other at a stereo angle of 40 mrad on either side of a module for each layer.

The final sub-system of the ID is the transition radiation tracker (TRT) [52], which differs quite significantly in design from the other two components. It consists of 300,000 gas filled drift tubes (or straws) with a thin wire at the center, each with a diameter of 4 mm. There are 50,000 straws of 144 cm around the length of the barrel arranged parallel to the direction of the beam and extending outward to 1082 mm from the beam line. A further 250,000 straws of 39 cm pointing outward radially from the beam are located in the two endcaps. In addition to ionization, charged particles passing through the TRT also emit transition radiation from crossing a boundary between two mediums with different dielectric constants. Since the amount of emitted radiation is inversely proportional to the mass of the particle crossing the boundary, the TRT can be used to discriminate between electrons and heavier charged particles such as pions. This gives it a unique advantage over the other ID sub-systems, although it comes at the cost of decreased precision along the longitudinal direction of the straws, distinguishing only between positive and negative z measurements.

The combination of the three inner detector sub-systems is mainly responsible for performing precise measurements of particle trajectories and momenta while minimally affecting their energies. This is accomplished by the presence of a total of 100,000,000 finely segmented individual readout channels spread across pixels, strips and straws. A typical particle passing through the barrel component of the detector will cross 4 pixel layers providing 4 measurements (or "hits"), 4 double-sided SCT layers providing 8 hits and around 35-40 TRT straws [52]. Overall, the inner detector is designed to have a momentum resolution of at least

$$\sigma_{pT}/p_T = (0.05\%)p_T \oplus 1\% \tag{3.1}$$

where $\oplus$ denotes an addition in quadrature. Resolution generally worsens with increased particle momentum when considering the inner detector only due to the comparatively less pronounced curvature of high momentum tracks.

3.2.3 Electromagnetic and Hadronic Calorimeters

Surrounding the inner tracking detector of ATLAS is an extensive calorimeter system made up of both hadronic and electromagnetic sampling calorimeters (HCAL and ECAL) as shown

[5]Pitch in this context refers to the distance between the center of adjacent detector elements

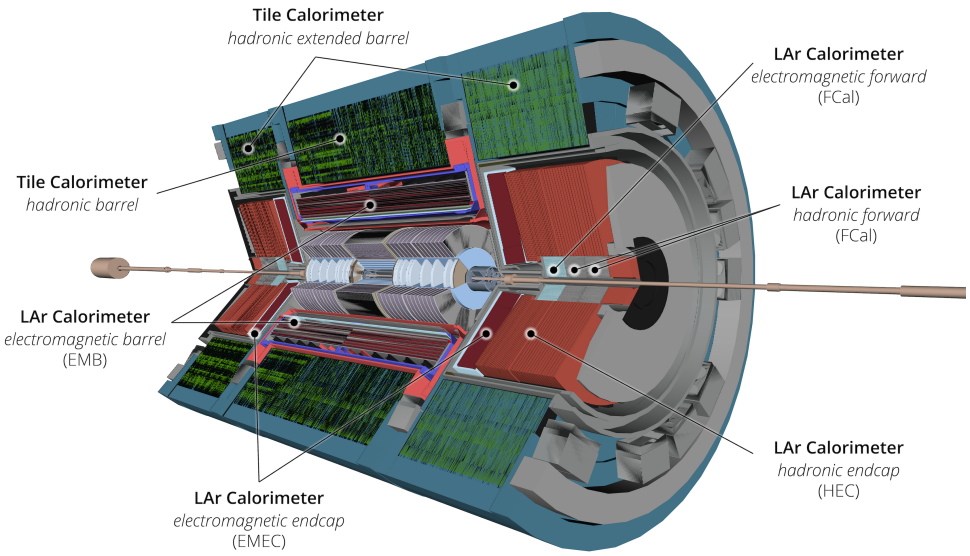


Figure 3.8: Schematic of the ATLAS calorimeters and their components for run 3 from [98].

in Fig. 3.8. The purpose of these components is to stop incoming particles, making them deposit their energy to facilitate precise energy measurements as well as particle identification for photons, electrons and hadrons (both charged and neutral). To accomplish this, sampling calorimeter elements are made up of alternating layers of an absorbing material and an active material. The absorber is dense, triggering interactions with incoming particles which leads to particle showers[6]. Electromagnetic showers proceed via repeated instances of Bremsstrahlung and e^+e^- pair creation triggered by a single photon, electron or positron. Hadronic showers start with a break up of the incoming hadron (both charged and neutral), evolve into further interactions and usually also contain an electromagnetic component. Each section of active material then captures information about the energy of a shower via ionization or scintillation[7] from charged particle interactions. Ideally, this process repeats through the layers of the calorimeter until all of the energy is deposited. Very high energy particles are sometimes not stopped completely, leading to breakthrough which reduces the accuracy of the energy measurement. The ATLAS calorimeter system uses two different technologies: a Liquid Argon (LAr) calorimeter on the inside and a Tile calorimeter on the outside.

The inner LAr calorimeter system has both electromagnetic and hadronic components (as well as a separate forward calorimeter). Each of these uses Liquid Argon as the active material with different absorbers. The electromagnetic calorimeter covers the barrel starting at a radius of 1.15m up to around 2.22m as well as the interior section of the endcaps covering

[6]A particle shower is a cascade of particles produced in an interaction with detector material
[7]These parts of the calorimeter are functionally identical to a tracking detector

$1.37 < |\eta| < 3.2$ and uses lead to stop photons and electrons up to a radiation length of $24X_0$[8] at $\eta = 0$. The hadronic and forward components of the LAr calorimeter are restricted to the endcaps only and use copper and tungsten as absorbers respectively to capture forward hadrons up to $|\eta| < 4.9$[9]. While hadrons lose part of their energy in the electromagnetic calorimeter, they are generally not stopped completely. Because of this, the tile calorimeter, which is made out of steel and plastic scintillator, surrounds the LAr components, providing a full hadronic calorimeter outside of the LAr endcaps. The energy resolution is designed for each component as

$$(\sigma_E/E)_{\text{EM}} = 10\%/\sqrt{E} \oplus 0.7\% \tag{3.2}$$

$$(\sigma_E/E)_{\text{Had}} = 50\%/\sqrt{E} \oplus 3\% \tag{3.3}$$

$$(\sigma_E/E)_{\text{Fwd}} = 100\%/\sqrt{E} \oplus 10\% \tag{3.4}$$

The first term represents stochastic contributions from shower development, which is reduced for high energy showers where more secondary particles are produced, thereby reducing statistical fluctuations from sampling. The second term is a constant that accounts for detector effects and is independent of the shower energy. The ATLAS calorimeter system is non-compensating, which means its energy response is different for electromagnetic and hadronic showers. In particular, the energy of hadronic showers is comparatively underestimated which is corrected through dedicated calibration procedures as discussed in section 4.1.1.

3.2.4 Muon Spectrometer

The only remaining charged particles that reliably penetrate the electromagnetic and hadronic calorimeters are muons. Muons do not interact significantly with detector material due to their much lower rates of energy loss via Bremsstrahlung, allowing them to keep comparatively undisturbed trajectories as they travel outward. This property alongside their relatively long lifetimes of 2.20×10^{-6} s ensures they are "detector-stable" and mostly don't decay within the detector at energies commonly seen in LHC collisions. As such, the outermost part of the ATLAS detector is a muon spectrometer (MS) used to identify muons and measure their momenta with coverage up to $|\eta| < 2.7$. Similar to the ID and calorimeter system, the MS also consists of a barrel and two endcaps as shown in Fig. 3.9.

The barrel is split into three concentric stations made up of Monitored Drift Tubes (MDT) as well as Resistive Plate Chambers (RPC) in the middle and outer station only. MDTs are tubes filled with a gas mixture which provide precision momentum measurements

[8] X_0 denotes the average length an electron needs to travel in material to reduce its energy by $1/e$

[9] More specifically, the hadronic endcaps cover approximately a range of $1.5 < |\eta| < 3.2$ while the forward calorimeter covers $3.1 < |\eta| < 4.9$

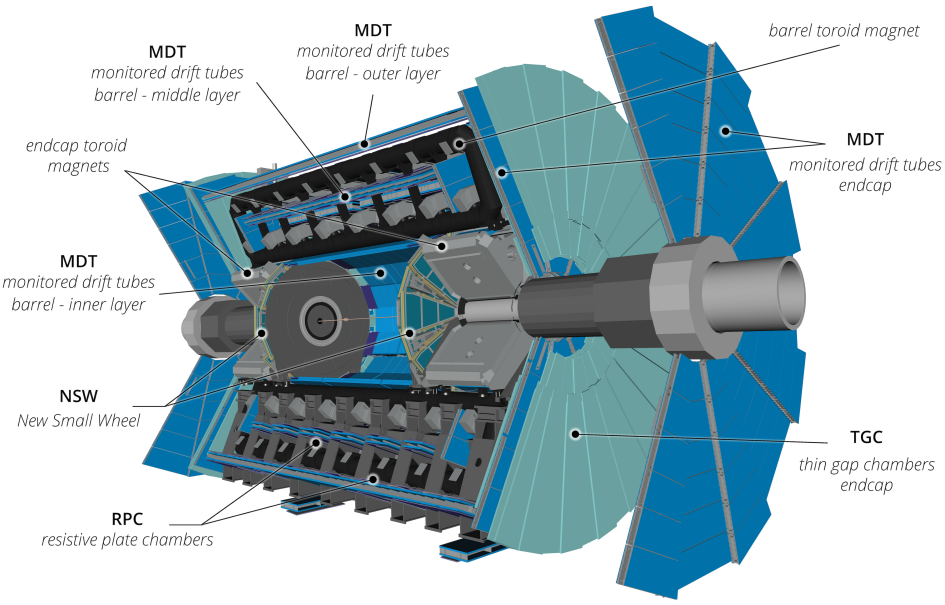


Figure 3.9: Schematic of the ATLAS muon spectrometer for run 3 from [98].

by measuring ionizing radiation (similar to the ID). There are 380,000 MDTs in the MS overall with an expected momentum resolution of

$$\sigma_{p_T}/p_T = 10\% \tag{3.5}$$

RPCs use similar technologies but are optimized for speed rather than precision, allowing for quick muon identification which is utilized in the Trigger system described in section 3.2.6. The endcaps of the MS are made up of 3 wheels each. Outer wheels contain MDTs only while the middle wheels contain both MDTs and Thin Gap Chambers (TGC), which are also optimized for speed. The inner (small) wheel was the subject of a major upgrade between runs 2 and 3 [98]. In its run 2 configuration it consisted of MDTs, TGCs and Cathode Strip Chambers (CSC) which did not provide enough precision to deal with the increased number of interactions per event in run 3 conditions. To combat this they were entirely replaced by the New Small Wheels (NSW) which use Small-Strip Thin Gap Chambers (sTGC) and micro-mesh gaseous structure (Micromegas) detectors to achieve high speed muon identification with better momentum resolution than the run 2 wheels.

3.2.5 Magnet System

Making measurements of charged particle momenta relies on identifying the curvature of their trajectory in an applied magnetic field. Stronger fields provide more curved trajectories which improve momentum resolution, especially for high p_T particles. In addition to the magnet

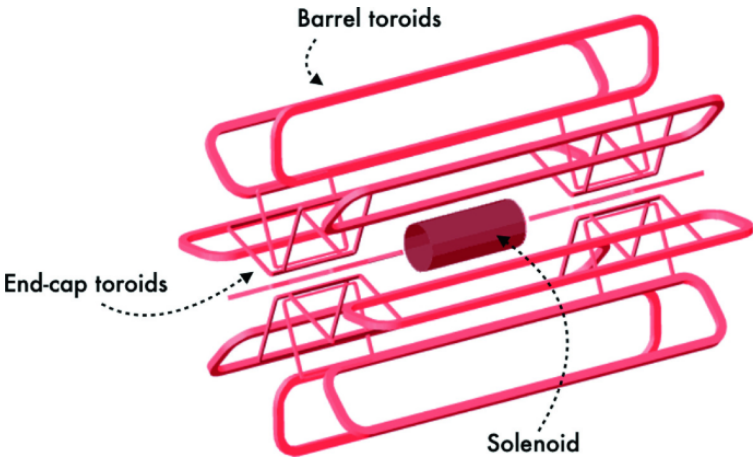


Figure 3.10: Schematic of the ATLAS magnet system from [209].

system present in the LHC, which is responsible for the acceleration of the proton beams, the ATLAS detector has its own integrated magnet system as shown in Fig. 3.10. It consists of a solenoidal magnet around the ID as well as three toroidal magnet systems made up of 8 coils each (one around the barrel portion of the detector and one around each of the end-caps). The solenoid applies a uniform $2T$ magnetic field along the direction of the beam, causing charged particles traveling from the collision point to follow helical trajectories with a radius of curvature determined by their transverse momentum p_T inside the ID. The toroids provide a circular magnetic field up to $3.5T$ in the outer parts of the detector used to deflect the trajectories of muons for more accurate momentum measurements in the MS.

3.2.6 Trigger and DAQ

The LHC provided bunch crossing rate of up to 40 MHz floods the ATLAS detector with data in excess of what can realistically be recorded and stored. Much of this data comes from collision events likely to contain little of interest to a physicist on the ATLAS experiment[10]. To combat this, decisions need to be made about whether individual events are worth recording and storing, which is the responsibility of the ATLAS Trigger and Data Acquisition (DAQ) systems outlined in Fig. 3.11. The Trigger system consists of the sequentially applied hardware-based "Level-1" (L1) trigger and the software-based "High Level Trigger" (HLT). Detector signatures are first filtered by the L1 trigger before being read out and stored temporarily in the "DataFlow" network where they are processed by the HLT and either discarded or moved to permanent storage. The combined effect of all experimental triggers reduces the frequency of recorded events from 40 MHz to around 1 kHz in run 2, and 3 kHz in run 3 [184] with write speeds of around 6 GB/s.

[10]For example, events where no "hard scatter" collision takes place (i.e. pile-up events).

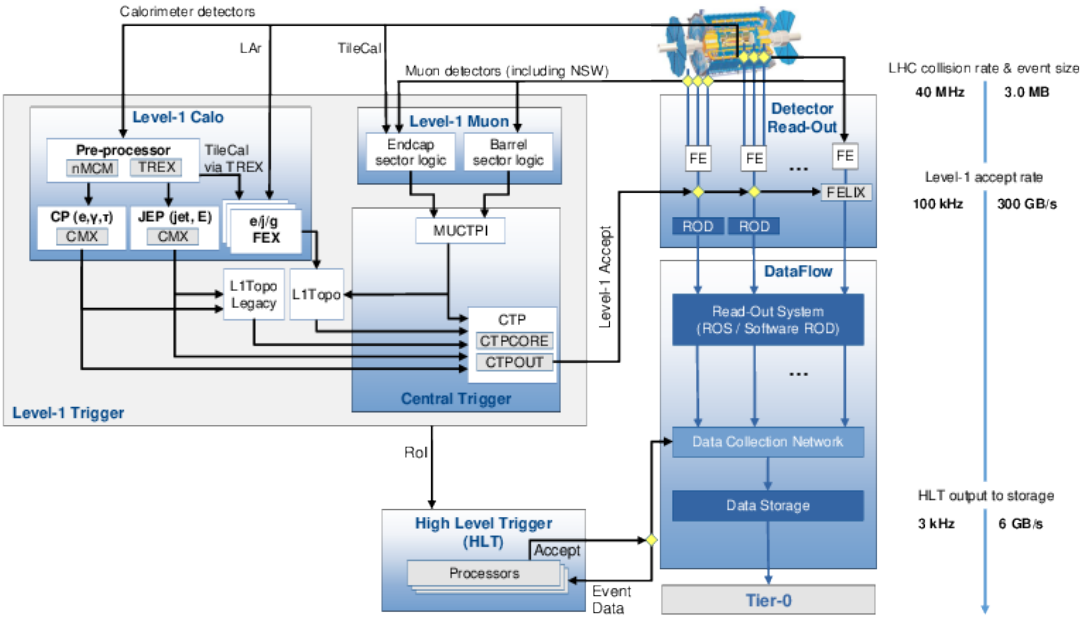


Figure 3.11: Diagram showing the core components of the ATLAS Trigger and DAQ for run 3 from [98].

The L1 trigger filters events by identifying large energy deposits in the calorimeter (L1-Calo) as well as high p_T muons in the fast-response muon system (L1-Muon) consisting of the RPCs and TGCs in the muon spectrometer. For run 3, the granularity of calorimeter information passed to L1-Calo was increased significantly via a shift from analog to digital processing to account for the expected increase in energy deposits from additional pile-up. L1-Muon was also augmented through an increase in the number of modules factored in as well as the replacement of the small wheel, which reduces the number of fake muon triggers by around a factor of 2 [99]. The Central Trigger Processor (CTP) records events based on both individual hardware triggers as well as their combination via computation of the pairwise invariant masses of muons and energy clusters in the calorimeter (L1-Topo). Events passing L1 undergo a lightweight version of the object reconstruction chain (described in detail in the next section) and get passed through the HLT, where a set of more complex criteria are applied to reduce the number of events further while covering a large number of different physics interests.

Chapter 4

Event Reconstruction

The ATLAS detector records electrical signals across millions of channels that correspond to an underlying set of physical processes originating in proton-proton collisions. In practice, these measurements are analogous in structure to digital images on a three-dimensional grid with heterogeneous pixel dimensions. Because of the inherent complexity of each measurement, a robust set of event reconstruction algorithms is necessary to make inferences about the underlying physics characterizing a recorded event. Reconstruction occurs in multiple discrete steps at different scales. Individual tracks, interaction vertices and calorimeter energy deposits are the foundation for the identification of particles such as muons, electrons and photons as well as hadronizing quarks in the form of particle jets, which can be classified based on flavor. These objects provide a common foundation for analyses of complex physical processes in ATLAS.

This chapter describes a set of common event reconstruction algorithms used in the ATLAS experiment, with an emphasis placed on those relevant for measurements of $H \to c\bar{c}$ decays in the VH production mode. After discussing fundamental objects and particle identification for leptons and photons (section 4.1) a detailed overview of jet reconstruction and flavor identification, also called "flavor tagging", is given (section 4.2).

4.1 Fundamental Objects

Raw detector measurements are turned into objects with physical meaning and interpretable characteristics through an elaborate event reconstruction sequence. This represents the first step in moving from an abstract collection of individual energy deposits as shown in Fig. 4.1 to a concrete measurement of processes such as $H \to b\bar{b}$. At the core of this are charged particle trajectories, which are created by combining individual measurements in segments of the inner detector into tracks with associated momenta and angular orientations. Calorimeter deposits are also clustered separately to isolate individual particle contributions. With this fundamental information, certain particles such as electrons, photons and muons can be directly identified and particle jets created by the hadronization of quarks can be recon-

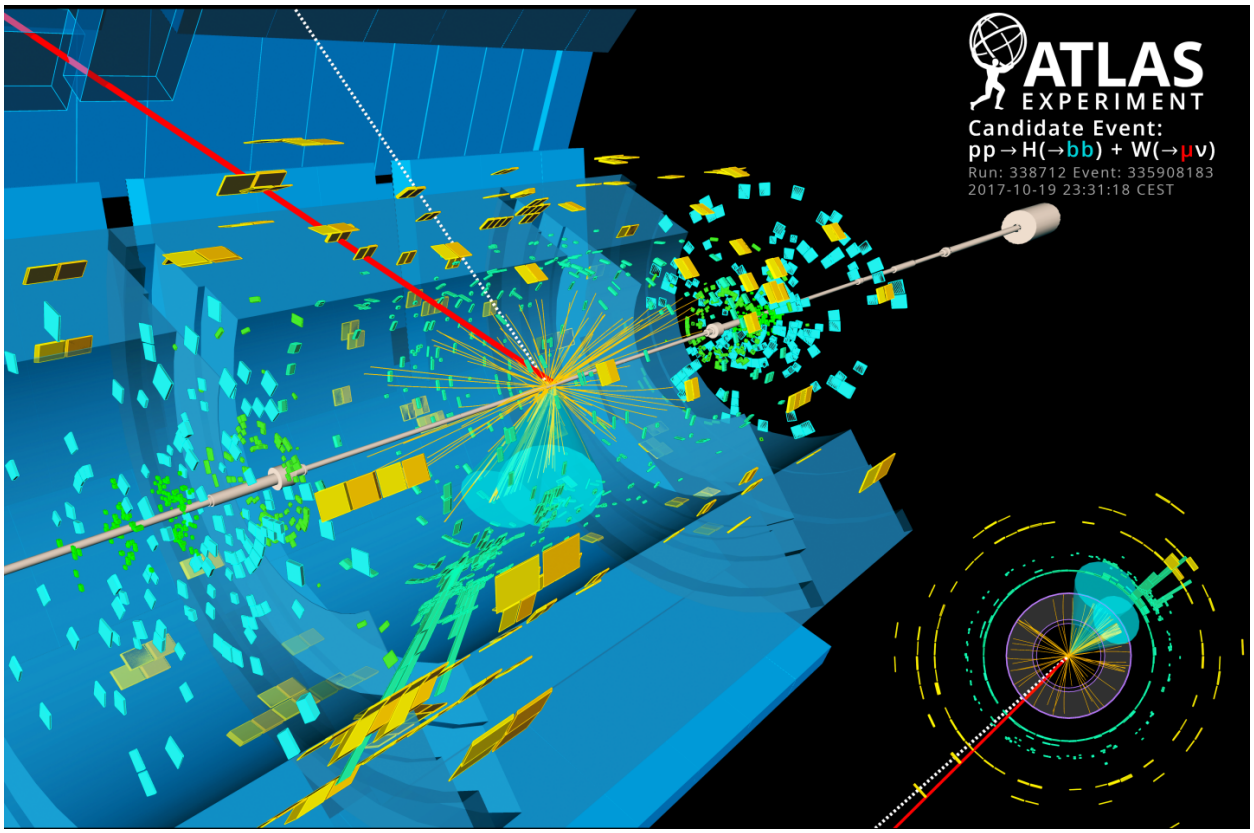


Figure 4.1: Detector signature of a $W(\rightarrow \mu\nu)H(\rightarrow b\bar{b})$ event candidate showing reconstructed tracks and calorimeter energy deposits from [56].

structed. Development of reconstruction algorithms relies on Monte Carlo (MC) simulated events, which can differ from observed data in unpredictable ways. Calibration procedures are thus applied to simulated reconstructed objects to account for measured differences and bring parity between performance in data and MC. This section outlines all aspects of the object reconstruction chain to give a complete picture of the low level physics objects relevant to measurements of physical processes in the ATLAS detector.

4.1.1 Tracks, Clusters and Vertices

The first step in the ATLAS reconstruction chain is identifying trajectories of particles passing through the inner detector [126]. The goal is the identification of a set of parameters given by $(d_0, z_0, \phi, \theta, \frac{q}{p})$ and visualized in Fig. 4.2 from individual energy deposits ("hits") in the detector for each particle passing through. The transverse and longitudinal impact parameters (IP) d_0 and z_0 are defined to be the coordinates of the point of closest approach between a track and the primary interaction point when the track is fully extrapolated in

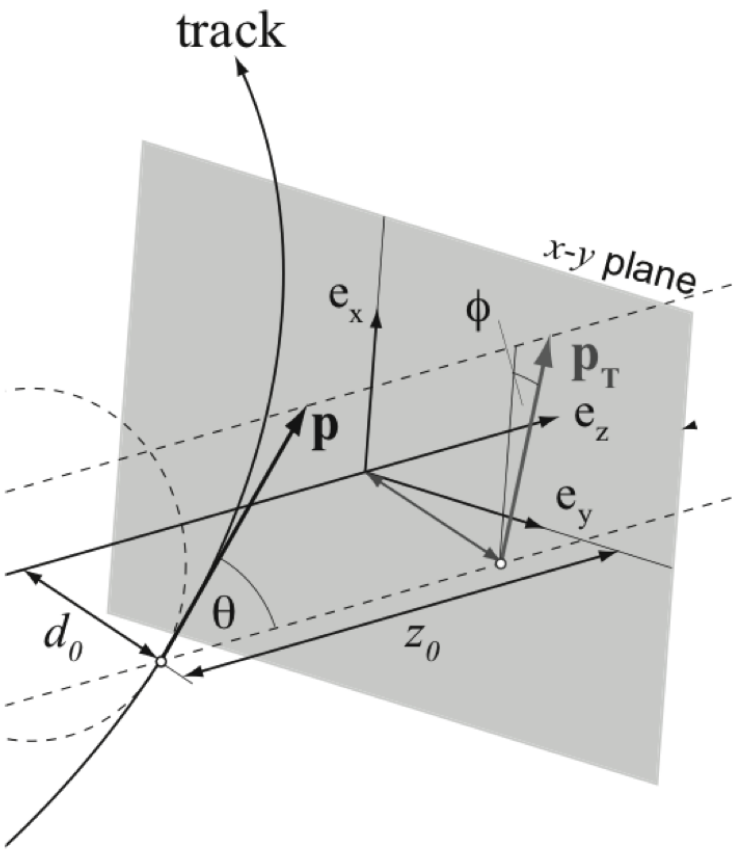


Figure 4.2: Diagram visualizing track parameters used in ATLAS from [149].

both directions. This functions as a measure of whether a track is likely to have originated from the primary interaction point (small IP) or from some later decay or interaction or pile-up (large IP).

Track Reconstruction

Tracking begins with a process called spacepoint formation, which clusters adjacent hits reaching some energy threshold in each layer of the pixel and SCT detectors. These three-dimensional clusters are transformed into global coordinates representing individual intersection points between particle trajectories and detector material. Spacepoints in subsequent layers are then grouped into batches of 3 to create track seeds. This is preferentially done in the SCT due to the lower density of spacepoints, but pixel seeds are also considered. Each seed is required to meet specific quality criteria based on rough estimates of track parameters to reduce fake tracks and computational overhead. The remaining seeds are then extended in both directions via a track finding algorithm based on the combinatorial

Kalman filter (CKF) [138]. This algorithm repeatedly adds compatible clusters on adjacent layers to a track candidate and adapts the estimated trajectory by taking into account detector resolution effects. In the case where multiple compatible spacepoints are found, diverging track candidates are created and extended separately. Additionally, a procedure called Bremsstrahlung recovery allows seeds that fail the track finding step entirely to be considered with a decreased quality threshold. This procedure is triggered when a failed seed is matched to an interesting region with significant energy deposits in the calorimeter, which helps recover performance for electron tracks radiating photons through Bremsstrahlung as they traverse the detector.

Once the full set of track candidates is established via the CKF, an ambiguity resolution stage attempts to remove all fake and duplicate tracks by enforcing a more stringent set of quality criteria. Tracks with a significant amount of overlapping (shared) hits are compared against each other and only the highest quality track is retained. Due to the size of each pixel compared to the separation between tracks, only a limited number of shared hits are allowed. A dedicated algorithm is used to perform cluster splitting [41], which aims to split clusters consisting of multiple particles into their individual components. After ambiguity resolution, a χ^2 fit is performed on the remaining candidates for a final track parameter estimate. Where applicable, an attempt is made to extend track candidates made from pixel and SCT hits into the TRT via another Kalman filter based algorithm before ambiguity resolution is applied. These extensions are only retained if the overall χ^2 fit of a track does not worsen with the additional TRT hits. Aside from this forward tracking procedure, a secondary back-tracking pass is done to increase reconstruction efficiency of particles produced further from the beam. This procedure is mostly identical to forward tracking, but starts by searching for hit segments in the TRT aligned with regions containing significant energy deposits in the calorimeter. Seeds are then created in layers of the pixel and SCT detectors consistent with these TRT segments using two spacepoints only.

Track Selection

While the overall approach to tracking is very similar between the run 2 and run 3 reconstruction, various speed ups and other improvements were introduced to deal with the denser tracking environment from increased pile-up under run 3 conditions [87]. This includes changes to seed and track quality requirements at multiple steps of the chain as well as earlier stopping for bad candidates. In order to balance efficiency of reconstructing true charged particle tracks and minimize the presence of fake tracks, ATLAS introduced two sets of cuts called working points (WP). These are denoted *loose* and *tight primary*, where at $\eta = 0$, the *loose* selection results in a successful reconstruction of around 90% of charged particle tracks with around 1% fake tracks [43, 109]. The *tight primary* selection sacrifices some efficiency for a total of around 85% to provide a lower fake rate of less than 0.1%. Overall, improvements in the reconstruction chain result in similar tracking performance across run 2 and run 3 despite the increased presence of pile-up.

Topo-clusters

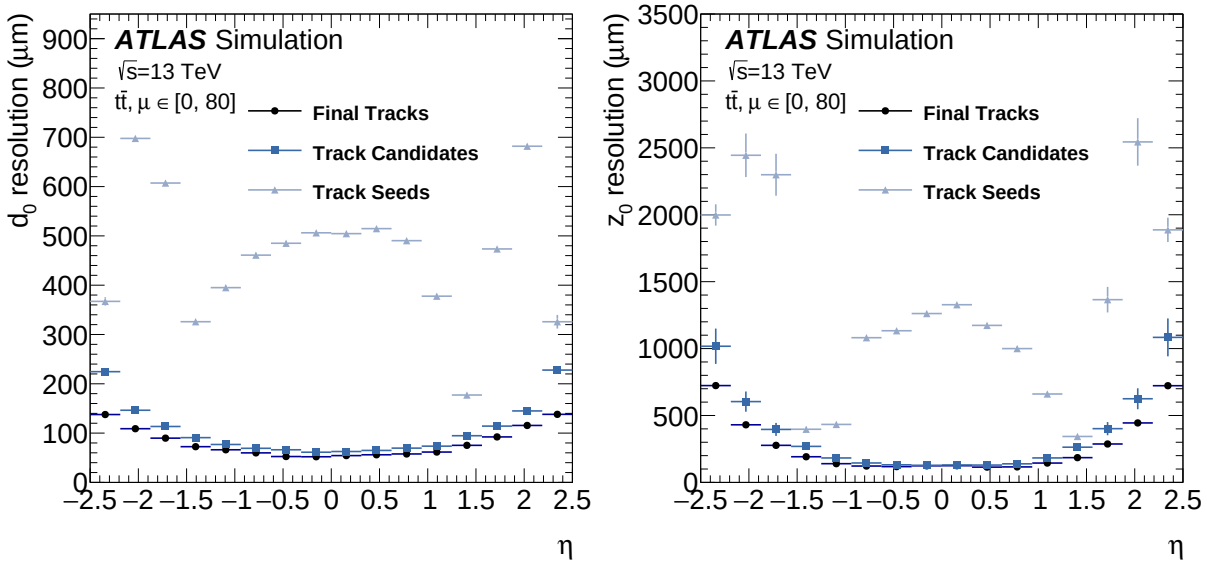
The incorporation of calorimeter information into higher-level object reconstruction involves clustering deposits across individual cells to identify particle signatures in a similar fashion to tracking. While details of this process differ between objects, topo-clusters [55, 97] provide the foundation. These clusters are built using a dedicated algorithm that uses all cells with $E/\sigma_E > 4$ as seeds and iteratively adds neighboring cells with $E/\sigma_E > 2$ before adding all perimeter cells with $E/\sigma_E > 0$. If a single topo-cluster resulting from this algorithm has multiple separable energy peaks, it is then split to reflect the presence of multiple particle contributions. To correct the energy scale of individual topo-clusters, a local cell-weighting (LCW) calibration scheme is applied [55]. This procedure corrects both hadronic and electromagnetic components simultaneously by weighting each cluster by its contributions. First, the non-compensating calorimeter response is corrected in the hadronic component followed by additional corrections for signal losses due to cells falsely identified as calorimeter noise as well as dead detector material. All corrections are based on single particle MC samples.

For many practical applications tracks and topo-clusters are matched to each other via dedicated algorithms such as "Particle Flow" (PFlow) [51], which was introduced during run 2. After these objects are combined, the resulting information can be used to subtract individual energy contributions from charged particles to each topo-cluster, which allows for the isolation of neutral components.

Primary Vertices

The presence of pile-up interactions leads to the reconstruction of excess tracks and calorimeter clusters that are not relevant to the interaction of interest identified by the trigger system. To mitigate the impact of these additional objects in selections enforced by physics analyses, a primary event vertex corresponding to the location of the hard scatter collision is identified via a vertex finding and fitting algorithm that differs slightly between run 2 and run 3 [53, 63]. In run 2, the process begins by creating an initial vertex seed centered at the beam-spot in the transverse plane and the mode of the z-position of all tracks passing kinematic cuts. All tracks within a certain impact parameter window of this seed are then considered in a Kalman filter based fit that iteratively associates nearby tracks and recomputes the vertex position based on χ^2 compatibility. Once this minimization converges, all tracks with a sufficient compatibility to the final vertex position are removed from the pool and the process is repeated. The run 3 version of the algorithm uses a Gaussian density function for seeding and allows each track to be associated with multiple vertices in the fit via a calculated weight. After multiple iterations of the fit, each track ends up primarily associated with a single vertex only. In both cases after fit convergence the vertex with the largest $\sum^{n_{\text{tracks}}} p_T^2$ is marked as the primary vertex.

The final position of the primary vertex is used to recalculate track impact parameters, which allows for removal of tracks likely coming from pile-up interactions. IPs also play a large role in identifying particles from secondary decays, which is of great importance to



(a) (b)

Figure 4.3: Plots showing (a) d_0 and (b) z_0 resolution as a function of η for final tracks, track candidates and track seeds in run 3 reconstruction from [109]

b and c hadron identification discussed in section 4.2. This makes IP resolution centrally important quantity in tracking. As shown in Fig. 4.3 it is at the level of around 100 μm for both d_0 and z_0 across most of the η spectrum in run 3 reconstruction. Tracks, calorimeter clusters and vertices are the foundations for higher level object reconstruction which relates underlying detector signatures to concrete physical particles, such as photons, leptons and quarks.

4.1.2 Electrons and Photons

Electrons ($e^{\pm}$) and photons (γ) are some of the most simple fundamental particles to identify in the ATLAS detector due to their unique interaction signatures in the EM calorimeter. A dedicated $e^{\pm}/\gamma$ reconstruction chain, which is built on top of tracks and calorimeter clusters, exists to find and isolate these signatures [96, 67, 64].

Reconstruction

Electron and photon reconstruction begins by selecting topo-clusters with an EM scale energy above 400 MeV and a ratio of EM to total cluster energy of at least 50%. From here the two algorithms proceed somewhat independently. Electrons leave tracks in the ID while photons only do so if they undergo $\gamma \to e^+e^-$ conversion (i.e. converted photons) before they enter

the calorimeter. As such, different track to cluster matching procedures exist to cover the three cases of electrons, converted and unconverted photons [64].

Before tracks are matched to clusters[1], they are refitted using a Gaussian Sum Filter (GSF) [39] algorithm, which models non-linear energy losses from Bremsstrahlung. Matching then considers tracks and clusters with $|\Delta\eta| < 0.05$ and $-0.10 < q * (\phi_{track} - \phi_{cluster}) < 0.05$ as electron candidates. If multiple tracks fulfill these criteria, tracks with hits in the pixel detector and a better ΔR match to the cluster are preferred. Converted photon candidates are identified by matching clusters to conversion vertices, built by considering either two tracks with opposite charges or single tracks without hits in the innermost detector layers. In either case, relevant tracks must have signatures in the TRT consistent with electrons and vertex candidates must likely be from a massless particle. In case of multiple matches, two track vertices with hits in the Silicon detectors (pixel or SCT) are preferenced. Remaining unmatched topo-clusters are considered unconverted photon candidates.

To account for radiation from Bremsstrahlung, a separate cluster matching procedure is performed for all electron and photon candidates, grouping individual topo-clusters into superclusters. Starting from the cluster with the highest transverse energy E_T, electron seed clusters are required to have $E_T > 1$ GeV and be matched to a track with at least 4 hits in the Silicon detectors. Photon seed clusters have no vertex matching requirement, but must have $E_T > 1.5$ GeV. Surrounding (satellite) clusters are then associated to the seed cluster if they are within a window of $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$, which is extended to $\Delta\eta \times \Delta\phi = 0.125 \times 0.300$ for electrons in cases where the best matched track of the seed cluster is also the best match for the satellite cluster. For converted photons, clusters where the best matched track belongs to the matched conversion vertex of the seed cluster are also added. Once the superclusters are built, relevant variables related to energy and shower shape are recomputed and the track and conversion vertex matching procedure is repeated. Reconstruction strategies for electrons and photons are kept consistent between run 2 and run 3 with only minor differences [206].

Selection, Isolation and Calibration

In order to improve the purity of reconstructed electron and photon objects, a selection procedure is applied to all candidates. For electrons, this is intended to distinguish prompt electrons from hadronic objects, converted photons and electrons originating from secondary decays. A likelihood discriminant based on variables describing the properties of the electron track, the associated calorimeter clusters and their compatibility is built for each candidate. This discriminant is used to define a series of WPs based on electron reconstruction efficiency in MC simulations in runs 2 and 3. They are referred to as *loose*, *medium* and *tight* with efficiencies of 93%, 88% and 80% at $E_T = 40$ GeV [64]. For photons, *loose* and *tight* WPs are defined based on direct cuts rather than a combined discriminant in an otherwise similar procedure. The loose selection uses shower shape variables from the second EM

[1]For electron reconstruction, tracks are required to have at least four hits in the silicon detector.

calorimeter layer and the hadronic calorimeter energy fraction while the tight selection adds requirements for the first EM calorimeter layer. These cuts are intended to separate prompt photons from those originating in hadronic decays with overall similar efficiencies to their electron counterparts [67, 64].

In addition to selection requirements, all electrons and photons have associated isolation variables which allow individual analysis groups in ATLAS to apply cuts based on spatial isolation of electrons or photons in the calorimeter and ID. Calorimeter isolation is defined as the sum of E_T of topo-clusters within $\Delta R < 0.2$ of an electron or photon cluster after subtracting an estimated contribution from pile-up. Similarly, the track isolation is defined to be the sum of p_T of tracks within a cone of the electron track or photon cluster direction using either a fixed or variable cone size (i.e. shrinking with increasing p_T). Separate *loose* and *tight* electron and photon isolation WPs are derived based on these two variables in MC simulations in a similar fashion to the isolation WPs [96].

To correct for differences in selection and isolation efficiency WPs in data and MC simulation, the impacts of all WP based cuts are compared between the two. In particular, scale factors representing the ratio of data efficiency to MC efficiency are derived in bins of photon/electron E_T and η for each WP. For electrons this is done via "tag-and-probe" methods by selecting $Z \to e^+e^-$ and $J/\Psi \to e^+e^-$ events, resulting in scale factors deviating from unity by less than 5-7% for all bins across both selection and isolation WPs in run 2[96]. Corresponding uncertainties are generally less than 1% for identification and up to 5% for isolation. For photons, scale factors and corresponding uncertainties are derived in $Z \to e^+e^-\gamma$ events and are similarly within 5% of unity with uncertainties of up to 7%. Run 3 results are generally consistent with observations from run 2 [206].

4.1.3 Muons

Muons ($\mu^\pm$) are easily identifiable by their detector signatures for different reasons than their lighter $e^\pm$ counterparts. Their nature as quasi minimum-ionizing particles makes the muon spectrometer at the outer edge of the ATLAS detector the most valuable tool for muon identification. Due to their long lifetimes and low rates of energy loss via Bremsstrahlung they commonly leave traces in all detector subsystems.

Reconstruction

While tracks in the muon spectrometer are the most obvious sign for the presence of muons, their reconstruction considers signatures across all parts of the detector. To this end there are five different algorithms focusing on different subsystem combinations to identify five types of muons: combined (CB), inside-out combined (IO), muon-spectrometer extrapolated (ME) or standalone, calorimeter-tagged (CT) and segment-tagged (ST) [76]. In general, muon reconstruction starts with the search for continuous track segments in each sector of the MS. This differs from standard track reconstruction due to the sparser tracking environment and is done via Hough transform on the individual detector hits for every sector [156].

Reconstructed track segments are then matched across sectors based on their estimated IPs and parabolic trajectories which estimate the path of a muon in the applied magnetic field. Finally, a χ^2 fit is performed to build track candidates which are used as a basis to remove outlier hits and add unmatched hits along the fitted trajectory. Ambiguity resolution ensures that only the highest quality track is retained in cases where there is significant hit overlap.

CB muons are reconstructed by matching tracks in the ID and the MS while factoring in energy loss in the calorimeter and performing a combined fit. At $|\eta| > 2.5$ the requirements are loosened to allow for shorter track segments in the Silicon detectors to be matched to tracks in the MS. IO muon reconstruction uses the same information, but performs inside-out association by starting from ID tracks and searching for at least 3 associated hits in the MS rather than a fully reconstructed muon track segment. When an ID track satisfies tight angular matching with at least one muon track segment it is classified as an ST muon and adopts its parameters directly from the track fit in the ID. Remaining tracks in the muon system are considered ME muons and are extrapolated back to the interaction point directly without further matching. CT muons do not rely on information from the MS at all but are rather reconstructed by extrapolating tracks from the ID to the calorimeter and checking for signatures consistent with minimum-ionizing particles. For these, track parameters are also adopted directly from the ID fit. For each reconstruction method a minimum muon p_T cut of 2 GeV is applied, except for a raised threshold of 5 GeV for CT muons given the larger amount of backgrounds that are otherwise vetoed by matched MS signatures. As with electrons, muon reconstruction strategies are consistent between run 2 and run 3 [105]

Selection, Isolation and Calibration

After all muon candidates are reconstructed, selection criteria are applied based on the number of hits in ID subsystems and MS sectors for each track, the compatibility of the track segments and overall fit properties. As with photons and electrons, a set of *loose*, *medium* and *tight* selection WPs are defined based on these criteria. These WPs are identified based on their prompt muon reconstruction efficiency and hadron misidentification rate, allowing individual analyses to balance the two as needed. There are also separate WPs defined for high-p_T and low-p_T muon identification to cover more specific use cases. In order to reject non-prompt muons originating from secondary decays of hadrons, a cut on the transverse and longitudinal impact parameters is enforced as $|d_0|/\sigma(d_0) < 3$ and $|z_0|\sin\theta < 0.5$ mm. Overall muon reconstruction efficiencies are stable around 90-99% for *loose* WP and 80-90% for *tight* across most of the η spectrum, except for the central $|\eta| < 0.2$ region due to a physical gap in the muon spectrometer around $|\eta| < 0.1$. Corresponding false identification rates are generally below 1% and uncertainties on both quantities below 0.1%. Reconstruction results are comparable for runs 2 and 3 [76, 105].

Further isolation of prompt muons is enforced via two isolation variables: track-based isolation and calorimeter-based isolation. This serves mainly to further reduce the number of reconstructed muons originating from hadron decays and mirrors the approach taken for photons and electrons. In particular, track-based isolation is calculated by summing the p_T of

all tracks within a cone of a particular size around a muon track. Calorimeter-based isolation is defined to be the sum of E_T deposited in the calorimeter within a cone of the extrapolated muon track while removing the expected muon contribution. A series of isolation WPs are defined based on either track-isolation only or a combination of both variables, using different cone sizes and cuts to optimize prompt muon retention and non-prompt muon rejection in a similar fashion to photons and electrons.

Finally, a calibration procedure is applied to correct differences in muon reconstruction efficiency across data and MC simulations using $Z \to \mu^+\mu^-$ and $J/\Psi \to \mu^+\mu^-$ events. SFs are derived for all selection WPs in bins of (η, ϕ) and in (p_T, η) for isolation WPs. They are generally within 1% of unity and have associated uncertainties of up to a few percent.

4.1.4 Particle Jets

The identification of quarks is of significant interest to measurements containing hadronic final states in ATLAS. They can originate both from the hard scatter collision itself as well as the decay of heavy resonances (such as W, Z or Higgs bosons). Due to the hadronization process described in section 2.4.2, energetic quarks (and gluons) manifest as conical sprays of particle tracks originating from a single vertex rather than having individually resolvable trajectories. These collections of tracks are reconstructed as particle jets, which also have differing characteristics depending on the flavor of their underlying quark. A generalized jet reconstruction procedure is applied before flavor differences are taken into account via jet flavor identification as discussed in section 4.2.

Jet Reconstruction

Due to the complexity of hadronization and QCD radiation, defining the extent of a particle jet signature is not as straightforward as defining the extent of a muon or an electron. The definition of a jet is in general dependent on the algorithm used to reconstruct it, which also requires a choice of constituent objects such as particle tracks, topo-clusters or some combination of both [88]. Truth jets are also considered by using the underlying detector-stable (i.e. $c\tau > 10$ mm, where τ is the proper lifetime) Monte Carlo simulated particles directly. These can be used to quantify the impact of detector response and pile-up on reconstructed jet energies and momenta. There are different jet reconstruction strategies for different physics purposes in ATLAS, each based on an algorithm called anti-k_t [25] which belongs to the generalized k_t[2] family of algorithms. All variations of generalized k_t iteratively cluster jet constituents based on two variables

$$d_{ij} = min(k_{Ti}^{2k}, p_{Tj}^{2k}) \frac{\Delta_{ij}^2}{R^2} \tag{4.1}$$

$$d_{iB} = p_{Ti}^{2k} \tag{4.2}$$

[2] k_t is another name for the transverse momentum p_T

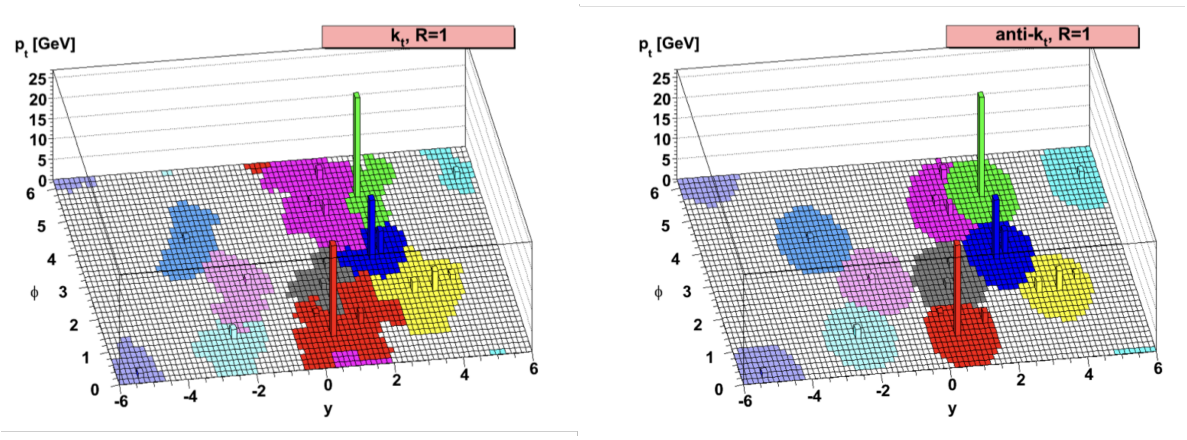


Figure 4.4: Visualization of jet clustering in k_T and anti-k_T algorithms in (y, ϕ) plane adapted from [25].

Here, d_{ij} is a distance measure between two constituents i and j defined by $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ where $y = \frac{1}{2}ln(\frac{E+p_z}{E-p_z})$ is the rapidity[3] of a particle. Similarly, d_{iB} is a distance measure between a constituent i and the beam. Finally, R and k are parameters of the algorithm, with R determining the approximate radius of the desired jet and k setting the relative importance of particle energy versus spatial separation between particles in the clustering order. The generalized k_t algorithm computes d_{ij} for all pairs and d_{iB} for all constituents and picks the smallest distance. If the minimum is d_{ij}, constituents i and j are clustered together and their combined axis, momentum and energy is computed. If the minimum d_{iB} then constituent i is considered a jet and removed from the pile of constituents under consideration. In anti-k_T, the clustering parameter $k = -1$ leads to preferential association of high momentum tracks to other tracks before low momentum tracks are associated with each other. For this reason, anti-k_T is inherently less susceptible to soft radiation than regular k_T with $k = 1$ where low momentum tracks are associated to each other first, which results in large fluctuations of the position of the jet axis at each step of the algorithm. A visualization of the final outcomes of k_t and anti-k_t clustering in an event with both hard and soft radiation present can be seen in Fig. 4.4. A third version of the generalized k_t algorithm with $k = 0$ called Cambridge-Aachen is also used in specific contexts. This removes all p_T dependence of the clustering and performs association by angular separation only.

There are two common jet reconstruction pipelines used in ATLAS for two different kinds of physics applications. Small-R jet reconstruction aims to isolate jets coming from the hadronization of individual quarks (usually in the primary collision) through application of anti-k_t with $R = 0.4$ [51]. Large-R jet reconstruction is tailored to identify jets coming from quark pairs usually originating in decays of heavy resonances at high momentum (or

[3]Rapidity is equivalent to pseudorapidity as $v \to c$ or as particle mass becomes negligible which is generally a fair approximation at LHC energies

"boost") and uses a larger radius parameter of $R = 1.0$ to encapsulate both quarks in a single jet object [77]. In both cases, only tracks within a certain IP range of the primary vertex are considered for pile-up mitigation purposes.

The fundamental objects used in small-R jet reconstruction are "Particle Flow Objects" built using the "Particle Flow" (PFlow) algorithm [51], which combines information from the tracking detector and calorimeters by matching isolated tracks with $p_T < 100$ GeV to one or multiple topo-clusters via their angular trajectories. An estimate for the calorimeter energy deposition of each track is calculated, which is then subtracted from matched topo-clusters. All clusters remaining after energy subtraction as well as unmatched clusters are retained as neutral PFlow objects. Both charged and neutral PFlow objects are used in jet reconstruction. To reduce contributions from pile-up to neutral constituents, an area-based energy subtraction is performed in the (η, ϕ) plane for both small-R and large-R jets after anti-k_t is applied. This is calculated based on the overall energy density of the event as well as the area of the jet in the given plane. Afterwards, a second residual correction is applied based on the total number of vertices and the average number of pile-up interactions in the event.

In contrast, large-R jet reconstruction uses input objects built with algorithms optimized for the presence of high momentum tracks and denser detector environments. These are Track-CaloClusters (TCCs) [55] and Unified Flow Objects (UFOs). TCCs are built without the energy estimation and subtraction corrections present in PFlow, which are less reliable at high p_T due to the larger amount of overlap between particle contributions in the calorimeter. Instead each track is assigned the energy of its matched topo-cluster. If there are multiple matched tracks, the energy is split based on relative track p_T. Neutral TCCs are only derived from unmatched clusters. UFO reconstruction combines aspects of PFlow objects and TCCs by applying energy estimation and subtraction to isolated tracks with $p_T < 40$ GeV in an identical fashion to PFlow first and then treating the remaining matched track and topo-cluster combinations with the TCC algorithm. After anti-k_t is applied, a separate grooming procedure aims to explicitly remove soft radiation from pile-up by re-clustering each jet using either k_t or Cambridge-Aachen and iteratively removing soft tracks in the clustering chain based on algorithm specific criteria. Examples of grooming algorithms in ATLAS are jet trimming [167] and SoftDrop [170].

To suppress jets from pile-up, an algorithm called the jet-vertex-tagger (JVT) [48] is applied after jet reconstruction for jets with $|\eta| < 2.5$. This is a likelihood-based discriminant built using variables derived from the kinematics of tracks within a jet in relation to their associated vertices. A separate version of this algorithm called forward JVT (fJVT) was developed to increase the rejection of pile-up jets in the high-$|\eta|$ regime [65]. For run 3, JVT was replaced by a Deep Neural Network (DNN) [198] based algorithm called NNJVT which improves pile-up rejection across all hard scatter efficiencies [207].

During run 2, large-R jet reconstruction was supplemented with additional reconstruction of smaller subjets within large-R jets meant to capture the individual hadron decays. Subjets were reconstructed with tracks only due to the difficulty of separating calorimeter contributions in dense environments and using a variable-radius (VR) parameter in anti-k_T

based on the p_T of the corresponding large-R jet as $R_{\text{eff}} = \rho/p_T$ bounded by some R_{max} and R_{min} [166]. These jets were used in addition to or instead of the large-R jets themselves depending on the application.

Jet Calibration

After the application of the anti-k_t algorithm in small-R and large-R jet reconstruction, a calibration procedure is applied to ensure consistency between truth and reconstructed jets as well as between data and MC simulations. For the first part, truth jets are matched to reconstructed jets to quantify the difference in jet energy scale (JES) and jet η for small-R and large-R jets as well as jet mass scale (JMS) for large-R jets only. These differences are calculated in bins of p_T and η and applied to the jet four-vector as multiplicative factors with corresponding uncertainties to account for detector response effects. In small-R jets, deviations from unity are less than 5% [74] in each bin while in large-R jets they are around 5-10% in jet energy and η and up to 50% in mass for extreme cases due to the larger impact of soft radiation [66]. To account for remaining discrepancies a separate procedure called the "Global Sequential Calibration" is applied to small-R jets. This is intended to correct for differences between quark and gluon initiated jets as well different responses to hadrons, leptons and photons and high energy jets not completely captured by the calorimeter. A series of corrective factors are derived based on six observables related to jet structure, which are then sequentially applied to the jet energy scale in bins of η and p_T without changing the average jet energy.

The final step in JES calibration for both small-R and large-R jets is the application of a series of *in-situ* techniques [44] comparing jets to calibrated reference objects in both data and MC. These are intended to correct for differences between real data and simulation by balancing energy and momentum between a jet and a well understood recoiling object. Multiplicative factors derived via double ratios of jet and reference object p_T in data and MC are applied to data only. The first technique is η-intercalibration which uses well measured central jets with $|\eta| < 0.8$ to correct the relative energy scale of forward jets with $0.8 < |\eta| < 4.5$ using dijet events. Next, a similar process is repeated to understand the absolute energy scale by comparing the central jets to recoiling photons or Z bosons with well measured momenta in Z/γ+jet events. Finally, multijet events are used to compare the relative energy scale between central jets with high p_T and systems of two or more recoiling lower p_T jets. The combined differences between data and MC in these three calibrations is found to be at most 3% on the double ratio with uncertainties of up to 2% [74]. A full set of 125 systematic uncertainties is provided for the jet energy scale (JES) through this process for run 2 which can be propagated to analyses. In run 3 the same procedures are applied and additional non-closure uncertainties are derived to account for observed differences, which are generally less than 1-3% [207].

While the JES corrects the average energy response, the jet energy resolution (JER) accounts for statistical fluctuations in that response. Differences in resolution between data

and MC are applied as additional uncertainties. The JER can be parameterized as a function of p_T as

$$\frac{\sigma(p_T)}{p_T} = \frac{N}{p_T} \oplus \frac{S}{\sqrt{p_T}} \oplus C \tag{4.3}$$

where N is a noise term from electronic noise and pile-up, S is a stochastic term from statistical fluctuations in the amount of energy deposited and C is a constant term capturing effects of passive detector material and uneven calorimeter response. To measure JER a momentum balance is performed on a dijet sample, similar to the η-intercalibration technique but without η restrictions on the second jet. A second technique to directly estimate the noise term is performed by sampling the energy of pile-up contributions in the calorimeter. Both techniques are combined to estimate JER differences between data and MC which are propagated as systematic uncertainties. In particular, in cases where the resolution of simulation surpasses data, Gaussian smearing is applied to MC [74]. Differences in JER between run 2 and run 3 are negligible for both small-R and large-R jets and are also provided as additional non-closure uncertainties [207].

For large-R jets, which are primarily used to identify heavy resonances, JMS and jet mass resolution (JMR) are also calibrated after the JES calibration is applied using two separate *in-situ* techniques. The JMR calibration is separate from JER due to the greater importance of large-R jet mass and its increased dependence on the angular distribution of track constituents compared to jet energy. The first method is called R_{trk} and quantifies the difference between jet momenta constructed using only the calorimeter or the inner tracker in dijet events [66]. Given the well understood differences of track momenta in the inner detector between data and MC, this method allows for the isolation of differences between neutral components in the calorimeter by considering relative ratios. A second technique called forward folding is applied to a selected sample containing $t\bar{t} \to W^+bW^-\bar{b}$ with one hadronic and one leptonic W decay [66]. The mass peaks in simulated samples are then smeared and shifted until they match observations in data. Both techniques result in double ratios within 1% of unity for JMS across the large-R jet p_T spectrum with JMR uncertainties of up to 3% [66]. Run 2 to run 3 non-closure uncertainties are comparatively larger for JMS and JMR than JES and JER but also remain at the few percent level [207].

4.1.5 Hadronic Taus

Compared to their lighter electron and muon counterparts, tau leptons are more difficult to identify due to their short lifetimes of 2.9×10^{-13} s causing them to decay before traveling a significant distance in the detector. Because of this they are usually observed through their decay products, which consist of a tau neutrino and either a combination of a lighter lepton and corresponding neutrino ($\sim$35% of the time) or one or more hadrons ($\sim$65% of the time) [146]. In the former case, electron and muon reconstruction algorithms are utilized as described in sections 4.1.2 and 4.1.3. Hadronic tau lepton decays most commonly produce either one or three charged pions, leading to different signatures in so called one-prong and

three-prong hadronic tau decays. More rarely, a pair of neutral pions is produced, which decay to photons. This often makes hadronic tau decays difficult to distinguish from quark and gluon jets as well as electrons. All hadronic tau decays are reconstructed as particle jets and classified in a similar but not identical fashion to jet flavor tagging as discussed in chapter 4.2. After a series of candidate selection steps, they are classified using two kinds of identification algorithms described below.

Reconstruction

Hadronic tau reconstruction begins with a standard small-R jet reconstruction procedure using topo-clusters as inputs followed by a set of kinematic cuts. Since tau leptons usually travel a short distance before decaying, a modified tau vertex association algorithm (TJVA) is applied to associate tau candidate jets with candidate vertices and kinematic features are recomputed relative to the updated vertex position. Slightly different reconstruction approaches are followed in run 2 [45] and run 3 [89]. In run 2, cuts are used to determine candidates for one-pronged and three-pronged decays directly. In run 3 tracks associated to candidate jets are classified by their low level features using a Recurrent Neural Network (RNN) based algorithm. This algorithm groups all tracks into 4 categories: tracks from taus (direct decay products), tracks from photon conversions, isolation tracks (coming from quark or gluon jets) or fake tracks. Candidate jets that have either one or three tracks that likely originate from a tau decay are kept as one-prong and three-prong tau candidates respectively.

Identification and Calibration

To further distinguish tau candidates from background quark or gluon jets as well as electrons, two separate identification algorithms are applied on a collection of track, cluster and jet level features. This is done with Binary Decision Trees (BDTs) in run 2 and RNNs in run 3. In each case, they are trained separately for one-prong and three-prong decays and separate WPs are derived based on respective identification efficiencies. The first is built to separate hadronic taus from quark or gluon jets originating in dijet events. A series of four τ_{had} identification WPs *tight*, *medium*, *loose* and *very loose* are defined. The *medium* WP provides 75% (60%) tau identification efficiency with a background misidentification rate of 2% (<1%) for one-prong (three-prong) decays with the run 3 RNN strategy. Three-prong WPs are in general taken to be tighter to reduce the absolute background contamination. Candidates passing the *medium* τ_{had} identification WP are then used to train a separate BDT or RNN to veto electrons from a $Z \to e^+e^-$ sample. Three (one) separate electron veto WPs are defined in run 3 (run 2) for both decay types. Given a signal efficiency of 95% in run 3 reconstruction, electron background contamination is <1% in both cases. Finally, a "Deep Sets" [221] neural network is used to identify the decay mode of the tau lepton candidate and calculate its 4-momentum, also replacing a BDT-based approach in run 2.

Systematic uncertainties related to tau identification and the electron veto are derived using different simulated $Z/\gamma* \to \tau^+\tau^-$ samples.

The energy scale of reconstructed tau leptons is calibrated via MC-based methods [45, 89] after subtraction of contributions from pile-up. In run 2, the tau momentum is computed from LCW corrected topo-clusters only. For run 3 an additional estimate is derived from "Particle Flow" objects [51] where topo-cluster contributions from charged particles are replaced with matched track momenta. The two energy scale estimates are then combined via a weighted average. Energy scale corrections are derived from the true tau momenta in MC, and in both cases further refined with a Boosted Regression Tree (BRT) targeting the ratio of true to reconstructed tau momentum using shower shape and other substructure variables as additional input. Uncertainties on reconstructed tau energies range from 2-6% for run 2 [45] and 4-6% for run 3 [89].

4.1.6 Missing Energy

Previously described object reconstruction techniques provide a comprehensive overview for all particle types expected to be visible in the ATLAS detector. Particles that do not interact with detector material, such as neutrinos, are not accounted for thus far. While they cannot be directly reconstructed given their missing interaction signatures, their existence can be inferred by the presence of missing transverse energy E_T^{miss} (MET) or equivalently missing transverse momentum p_T^{miss}. Identifying p_T^{miss} requires a full set of reconstructed and calibrated objects associated to the primary hard scatter vertex in each event. This includes hard contributions from electrons, muons, photons and jets as well as soft contributions from unmatched tracks combined as

$$\mathbf{E}_T^{miss} = -\sum_{\substack{\text{selected}\\\text{electrons}}} \mathbf{p}_T^e - \sum_{\substack{\text{accepted}\\\text{photons}}} \mathbf{p}_T^\gamma - \sum_{\substack{\text{accepted}\\\tau\text{ leptons}}} \mathbf{p}_T^\tau - \sum_{\substack{\text{selected}\\\text{muons}}} \mathbf{p}_T^\mu - \sum_{\substack{\text{accepted}\\\text{jets}}} \mathbf{p}_T^{jet} - \sum_{\substack{\text{unused}\\\text{tracks}}} \mathbf{p}_T^{track} \quad (4.4)$$

An explicit ambiguity resolution and rejection procedure is applied to each individual contribution in order to avoid double counting partially overlapping objects. Similar procedures are followed in run 2 [59] and run 3 [106]. Exact object reconstruction requirements vary but are built on the different WP definitions detailed throughout this section. Calibration of the E_T^{miss} calculation occurs in data and MC comparisons across three different event topologies. A sample of $Z \to \mu^+\mu^-$ decays (i.e. decays without true missing energy) is used to quantify the impact of object energy resolution effects on the MET calculation. A separate $W \to l\nu$ sample allows for comparisons of lepton energy and MET to measure the resolution of E_T^{miss} magnitude and direction. Finally, a sample of $t\bar{t}$ events with at least one leptonic W decay is used to quantify missing energy differences in data and MC in environments with high jet multiplicity. Systematic uncertainties are provided for both the E_T^{miss} scale and resolution based on a combination of these results.

4.2 Jet Flavor Identification

Particle jets represent some of the most complex fundamental objects in the event reconstruction chain of the ATLAS experiment and originate from the hadronization of quarks or gluons during the break-up of protons or the decays of heavy resonances (such as a H, $W^{\pm}$ or Z). Different flavors of quarks produce jets with slightly different characteristics, which can be identified with dedicated algorithms. This identification step ("jet flavor tagging") is separated from the fundamental jet reconstruction process discussed in section 4.1.4. Since measurements of the quark Yukawa couplings rely on knowledge of quark flavor, jet flavor identification is a crucial component of studying $H \to b\bar{b}$ and $H \to c\bar{c}$ decays. In the case of high p_T (boosted) jets, decays from heavy resonances are collimated enough that both quark hadronization processes are usually captured in the same large-R jet, which alters the dynamics of jet flavor identification to identifying resonance decays inclusively rather than the individual quark flavors in particular. In both low and high p_T cases, the general approach and physical intuition remains the same.

4.2.1 Basics of Flavor Tagging

The process of hadronization affects all quark flavors in a similar fashion with the exception of top quarks, which predominantly decay as $t \to W^+b$ before hadronizing due to their large mass and short lifetime. All other quarks undergo the same process involving the creation of new quark-antiquark pairs from gluon splitting $g \to q\bar{q}$ forming bound states with the originating quark in the form of new hadrons. Production rates for different flavors depend on the gluon energy, but are much larger for u, d and s (collectively referred to as "light") quarks than for b and c (or "heavy") quarks at the LHC on average. In particular, at a gluon energy equal to the Z mass, the predicted fractions of $g \to c\bar{c}$ and $g \to b\bar{b}$ are around 1% and 0.1% respectively [193]. This means that most hadrons created in the hadronization process are detector-stable pions and kaons which contain only light quarks. The presence of hadrons containing heavy quarks, such as B and D mesons, indicates a particle jet likely originates from a heavy flavor (i.e. bottom or charm) quark with some minor background contamination from gluon splitting. Heavy flavor hadrons are not detector-stable and generally decay some short distance from the primary collision vertex with proper decay lengths of common B and D mesons shown in table 4.1. This leads to bottom and charm hadrons travelling distances on the order of millimeters in the detector before decaying when taking relativistic effects into account, creating a separated secondary decay vertex (SV) in the jet.

While bottom and charm hadrons both decay some distance from the primary vertex, the separation between PV and SV is larger on average for bottom jets due to the longer lifetimes and decay lengths of bottom hadrons. The reason for this apparent inverted relationship between quark mass and stability is the lower rate of the CKM suppressed cross-generational decay of bound b quarks (mostly $b \to cW^-$) compared to the dominant intra-generational decay of bound c quarks ($c \to sW^+$). This leads to bottom jets in particular containing tracks whose closest approach to the collision point, denoted by their longitudinal and transverse

Particle	Quark content	Proper decay length $c\tau$
B^+	$u\bar{b}$	4.91×10^{-4} m
B^0	$d\bar{b}$	4.55×10^{-4} m
B_s^0	$s\bar{b}$	4.54×10^{-4} m
D^0	$c\bar{u}$	1.20×10^{-4} m
D^+	$c\bar{d}$	3.12×10^{-4} m
D_s^+	$c\bar{s}$	1.50×10^{-4} m

Table 4.1: Proper decay lengths $c\tau$ for common B and D mesons.

impact parameters (IP) z_0 and d_0 and their decay distance L_{xy}, is comparatively large as seen in Fig. 4.5. This is true to a lesser extent for charm jets due to the shorter decay lengths of charm hadrons. On average, secondary decay vertices in bottom jets also contain more tracks than those in charm jets due to the dominant two step decay of bottom quarks via an intermediate $b \to cW^-$ decay. The following $c \to sW^+$ decay can sometimes be reconstructed as a separated tertiary vertex, which presents another powerful identifier for bottom jets.

Since the extent and composition of a jet is inherently dependent on the algorithm used to reconstruct it, there are different techniques to identify true jet flavor in MC simulations. Particles that contain heavy flavor quarks are associated to jets with either geometric (delta-R) or ghost association. The former relies on measuring the distance between the jet axis and all heavy flavor particle trajectories in the (η, ϕ) plane. Particles that are within a specific radius of the jet axis are considered matched to the jet. Ghost association works by re-clustering all jet constituents together with modified versions of heavy flavor particles that carry zero momentum, but identical trajectories (i.e. "ghost particles"). This momentum rescaling prevents the added particles from influencing the clustering process or the resulting jet kinematics. Any ghost particles that end up within the jet after clustering are considered matched. Jet flavor is defined in each case by the matched heavy flavor particles: jets with at least one ≥ 5 GeV weakly decaying b hadron are bottom-jets, jets with no such b hadrons and at least one c hadron matching criteria are charm-jets and the rest are light-jets. When tau-jets are also considered, they correspond to jets with at least one hadronically decaying tau lepton. In general, bottom hadrons take precedent and the labeling scheme holds regardless of whether jets are initiated by quarks or gluons. Distributions of track IPs showcasing differences between jet flavors in a simulated $t\bar{t}$ sample based on ghost associated labels can be seen in Fig. 4.6.

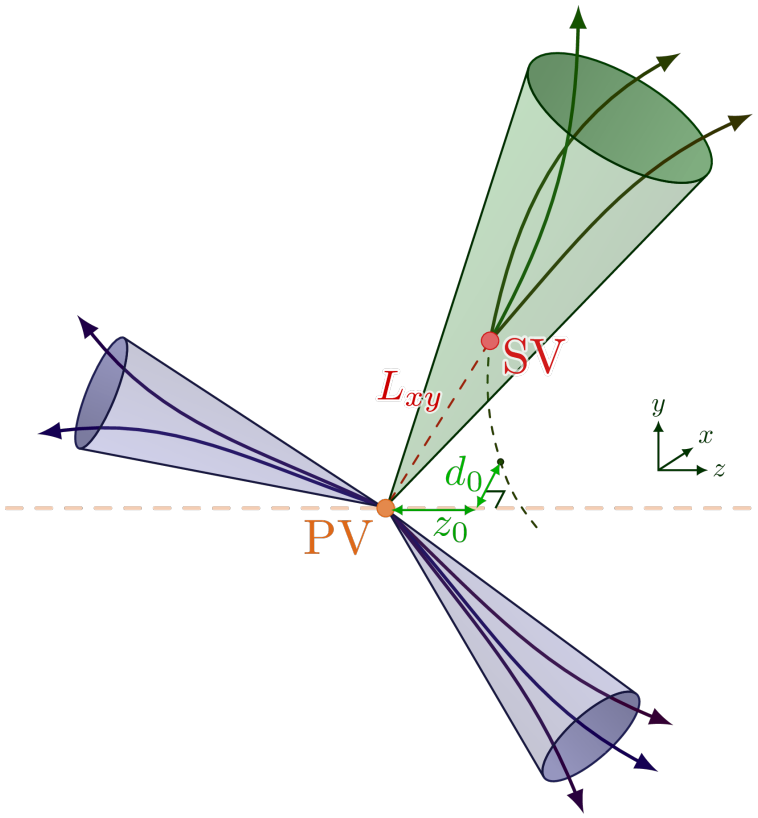
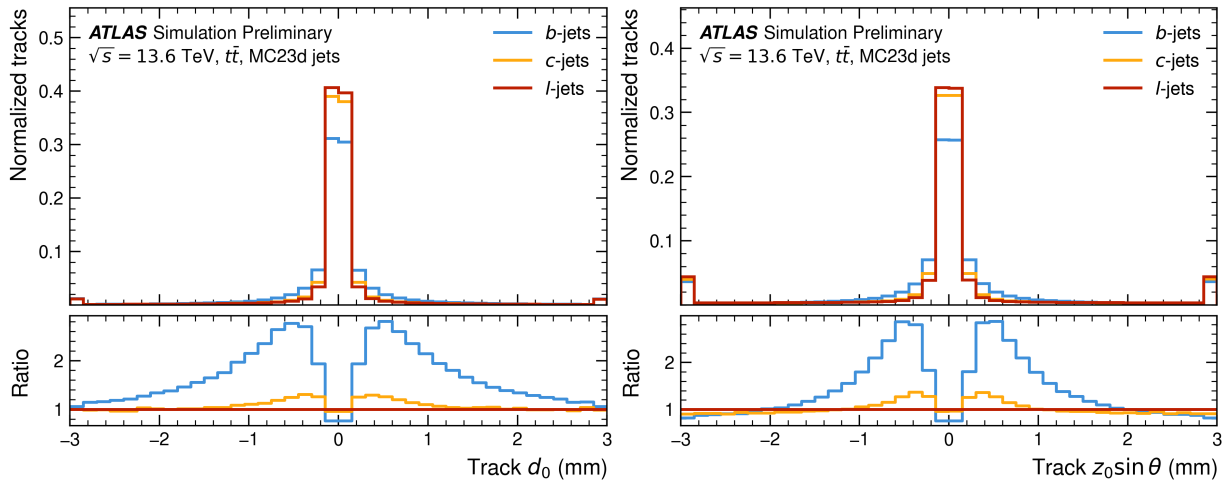


Figure 4.5: Diagram visualizing a bottom and two light-flavor jets from [180].

4.2.2 Traditional Flavor Tagging in ATLAS

ATLAS uses a suite of algorithms to identify small-R jets by flavor as bottom, charm or light jets. These algorithms traditionally come in two forms: low level algorithms that either explicitly reconstruct secondary decay vertices or classify jets based on their track IP distributions and high level algorithms that combine jet kinematic information with output from low level algorithms to build a combined set of classification scores. During run 2, the main algorithm for classifying jet flavor was a Deep Neural Network (DNN) [198] called `DL1r`, which combines information from the lower level IP-based algorithms `IP2D`, `IP3D` and `RNNIP` and SV reconstruction algorithms `SV1` and `JetFitter` [90]. `DL1r` is an evolution of `DL1`, which does not contain information from `RNNIP`. Both are successors to `MV2c10`, which is similar but based on a "Boosted Decision Tree" (BDT) architecture. Most flavor tagging algorithms in ATLAS output a set of probabilities p_u (light), p_b (bottom) and p_c (charm) for each corresponding jet flavor, which are combined into discriminants for bottom- and charm-jet identification as



(a) (b)

Figure 4.6: Plots showing distribution of track (a) d_0 and (b) $z_0 \sin\theta$ for different jet flavors in a simulated run 3 $t\bar{t}$ sample

$$D_{b/c} = \log\left(\frac{p_{b/c}}{f_{c/b}p_{c/b} + (1 - f_{c/b})p_u}\right) \tag{4.5}$$

These discriminants are used to place flavor tagging selections on individual jets defined through underlying signal efficiencies and background rejections on MC jets via so-called "working points" (WPs). Signal efficiency refers to the fraction of signal jets (i.e. bottom or charm) remaining after a WP cut on the discriminant is performed while the background rejections refer to the inverse of the fraction of jets of each background remaining, also called the mistag rate. The f_x constants are tunable parameters that allow for balancing the rejections of different backgrounds. These are generally defined separately for each algorithm and for bottom and charm identification.

Impact parameter based algorithms

Impact parameter based algorithms rely on the transverse and longitudinal track IP distributions of jets to implicitly search for separated secondary decay vertices. The transverse IP d_0 of a track is defined as the distance of closest approach between the track and the PV in the (r, ϕ) plane and its longitudinal IP z_0 is given by the z distance to the PV at that point. IPs are signed based on the location of the point of closest approach along the jet axis. Positive IPs indicate the point is located in front of the PV along the jet axis, suggesting a possible secondary decay vertex. Negative IPs denote tracks whose closest approach is behind the PV and are more representative of backgrounds such as material interactions or

pile-up. This information has historically been exploited by three dedicated algorithms in ATLAS: `IP2D`, `IP3D` and `RNNIP` [90].

IP2D and **IP3D** [47] are two variations of a likelihood-based algorithm used to identify the flavor of jets by discriminating based on track IP distributions. `IP2D` only uses the transverse IP significance d_0/σ_{d_0} as a discriminator, while `IP3D` uses both the transverse and longitudinal IP significance $z_0 \sin\theta/\sigma_{z_0 \sin\theta}$. Probability density functions are derived from MC simulation for each jet flavor as a function of these variables across a series of regions defined by hit patterns in different detector layers. Three log likelihood ratios are then built by summing probability ratios of all tracks in a jet to distinguish bottom-jets from light-jets as $\sum_{i=1}^{N} \log(p_b/p_u)$ as well as bottom-jets from charm-jets and charm-jets from light-jets in an analogous fashion. This approach treats all tracks as entirely independent and does not consider relevant correlations, which is an inherently flawed assumption due to the nature of the physical processes underlying jet formation.

RNNIP [50] is a Recurrent Neural Network (RNN) based extension of the traditional IP algorithms IP2D and IP3D built on an architecture called Long Short-Term Memory [152]. The main advantage of this architecture is its ability to make predictions on time-series data by taking correlations between elements into account. `RNNIP` is able to model correlations between individual tracks by considering each track an element in a sequence ordered by their transverse IP significance $|d_0/\sigma_{d_0}|$. It uses the same information as IP3D for each track with both longitudinal and transverse IP significances provided as input features alongside a variable that categorizes tracks based on their hit patterns. Two new variables are also included for each track: the momentum fraction p_T^{track}/p_T^{jet} and the angular separation between the track and jet axis ΔR in the (η, ϕ) plane. Four output classes predict probabilities that a jet is a bottom-, charm-, or light-jet and are combined into a discriminant defined by equation 4.5.

Secondary vertex reconstruction algorithms

While track IP values can be used to implicitly search for separated secondary decay vertices in jets, there are also algorithms built to explicitly find and reconstruct them. SVs are generally within a few mm of the PV, which makes their identification challenging because of the limited IP resolution of the ATLAS detector shown in Fig. 4.3. There are two main approaches to explicit SV reconstruction depending on whether tertiary vertices present from bottom to charm decays are reconstructed separately or merged into a single SV. Another issue in the identification of bottom and charm vertices is the presence of vertices from long-lived particle decays (i.e. Λ^0 and K_S) and electron conversions, which need to be accounted for. ATLAS uses two different algorithms for SV reconstruction with different design philosophies: `SV1` and `JetFitter`.

SV1 [54] is an SV reconstruction algorithm that is built to reconstruct a single displaced vertex per jet. This means that tertiary vertices are either ignored if they are separated by a resolvable distance from the SV or merged if they are close. The algorithm first builds plausible pairs of two-track vertices based on the physical distance of the two tracks. After

some mass-based cuts to reject vertex candidates likely originating from photon conversions or long-lived particle decays, the remaining two-track vertices are combined. A fit is then performed on all tracks to fit a single SV. Tracks with large χ^2 values are iteratively removed until a threshold is reached. Using the reconstructed SV, a set of 4 variables is calculated: the invariant mass of the SV, the energy fraction of tracks in the SV compared to all tracks in the jet, the number of two-track vertices and the ΔR between the jet axis and the PV-SV vector. These four variables are used to derive joint PDFs for the mass and energy fraction as well as individual PDFs for the number of two-track vertices and ΔR for both bottom and light jets using MC simulation. The likelihood ratios are used as a discriminant between the two jet flavors.

JetFitter [61] aims to inclusively reconstruct the full bottom to charm hadron decay chain using a Kalman filter [138] based approach with the underlying assumption that the two decay vertices are aligned on the original bottom hadron flight path. The first stage of `JetFitter` is similar to `SV1` and consists of a search of possible two-track vertices fulfilling stringent cuts used to reject tracks originating from the PV, photon conversions, long-lived particle decays or material interactions. The decay chain of the remaining tracks is described by a set of variables given by the PV position x_{PV}, y_{PV} and z_{PV}, the azimuthal and polar angles ϕ and θ describing the bottom hadron flight axis as well as the set of distances d_i of each vertex along the flight path. This set of variables is initialized using the reconstructed PV position, the jet axis for ϕ and θ and an estimates for the intersections between the trajectory of track i and the jet axis for d_i. These variables are iteratively updated with a modified Kalman filter until the algorithm converges. The final result is a full reconstruction of the bottom hadron decay chain.

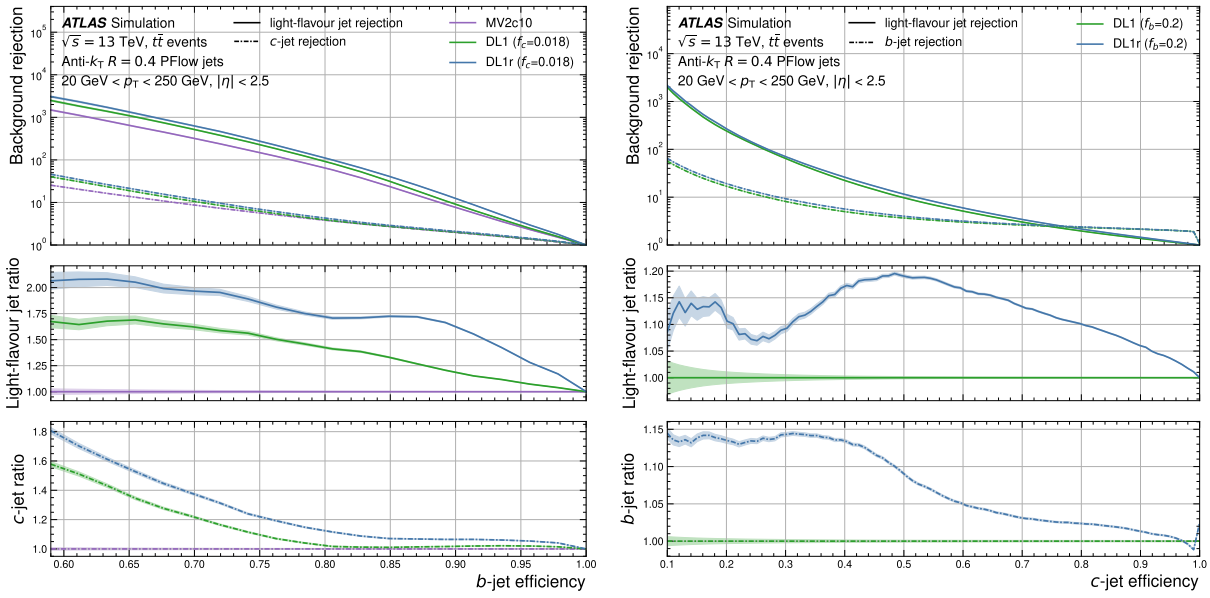
High level flavor tagging

Low level IP based and SV reconstruction algorithms are useful on an individual basis and provide physical intuition for decision making in jet flavor identification. The highly specialized nature of each algorithm can be exploited by considering a combined discriminant built from the low level outputs of each. In ATLAS, this discriminant is constructed with `DL1r`, which is based on a DNN architecture. It uses a set of 31 input features taken from `IP2D`, `IP3D`, `RNNIP`, `SV1` and `JetFitter` detailed in table 4.2 and outputs a set of three flavor probabilities for each jet: p_u, p_b and p_c. Discriminants for flavor identification D_b and D_c are constructed with these outputs as described above for bottom and charm jet identification respectively.

Performance of DL1r is quantified in terms of signal efficiencies and background rejections and is shown in Fig. 4.7 for a run 2 (i.e. $\sqrt{s} = 13$ TeV) simulated low momentum $t\bar{t}$ sample compared to both `DL1` and `MV2c10`. `DL1r` achieves a background rejection of ~ 10 for charm- and ~ 700 for light-jets at a 70% bottom-jet identification efficiency. This presents a factor of 1.4 and 2 improvement compared to `MV2c10`. Charm-jet identification is more difficult with corresponding bottom- and light-jet rejections of ~ 10 and ~ 70 at a charm-jet efficiency of 30%. `MV2c10` lacks charm-jet identification capabilities, so no comparisons are made.

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

Table 4.2: Summary of all input variables for DL1r adapted from [90].



(a) (b)

Figure 4.7: Plots showing background rejections vs. (a) bottom- and (b) charm-jet efficiency for `DL1r`, `DL1` and `MV2c10` from [90]

Classification of large-R jets from heavy resonances makes use of similar high level algorithms but is approached differently [71]. The main backgrounds for heavy resonance decays such as $H \to q\bar{q}$ and $Z \to q\bar{q}$ are identified directly. These are mostly collimated top quark decays $t \to bW^+$ and high energy QCD radiation $g \to q\bar{q}$. Signal categories are denoted $X \to b\bar{b}$ or $X \to c\bar{c}$ and generally target Higgs decays in particular. During run 2, a specialized high level tagger was built for boosted $H \to b\bar{b}$ identification based on a DNN architecture. It is based on the application of `DL1r` to variable-radius (VR) track jets reconstructed within large-R jets. The inputs to the $X \to b\bar{b}$ tagger consist of 11 features made up of large-R jet p_T and η as well as the three `DL1r` output scores for up to three VR track jets. Its outputs are class probabilities corresponding to $H \to b\bar{b}$, Top and QCD. A separate discriminant is defined in an analogous fashion to small-R jet identification as

$$D_{Hbb} = \frac{p_{Hbb}}{f_{Top}p_{Top} + (1 - f_{Top})p_{QCD}} \tag{4.6}$$

Though this algorithm provides the foundation for developments discussed in chapter 6, it is disfavored compared to direct application of `DL1r` to VR track jets in many cases. This is due to the necessity of a dedicated calibration procedure which is more complete for small-R jet flavor identification as discussed in the following section.

4.2.3 Flavor Tagging Calibrations

In order to perform a physical measurement on real data with a jet flavor identification algorithm that was trained on MC simulations, a dedicated set of calibration procedures is developed to quantify performance differences between the two. These differences result in additional systematic uncertainties distinct from the more fundamental object associated uncertainties discussed in section 4.1. So called "Scale Factors" (SFs) quantify absolute differences between signal efficiencies and background rejections in MC and data and are applied directly to MC templates after a flavor tagging selection is applied. In general, the techniques by which SFs and their uncertainties are derived vary by the signal or background type but are kept consistent between different algorithms. Broadly, each technique relies on isolating a representative sample pure in a given signal or background category and then extrapolating data and MC differences to the whole phase space, accounting for contamination from backgrounds. There are three distinct calibration procedures used for `DL1r`, focused on bottom, charm and light jets respectively. Each can be applied to extract either the signal efficiency or background mistag rate depending on the overarching goal.

Calibration of bottom jets

A representative sample of bottom jets can most easily be isolated within dileptonic $t\bar{t}$ events, where both W bosons from $t \to bW^-$ decay leptonically. This selection strategy is applied to calculate the difference between bottom jet identification efficiencies in MC and data [62]. To ensure that the correct bottom jets are selected, they are chosen based on a minimization of the combined invariant mass with each lepton. Fit regions are then defined based on these two sets of invariant masses m_{j1,l_i} and m_{j2,l_j}. The region where both masses are low is used as the signal region while three other regions where one or both masses are large control light jet backgrounds. After a statistical likelihood fit, SFs on the bottom jet efficiencies are extracted. These deviate from unity by less than 5% for the 70% bottom identification WP of `DL1r`, indicating good agreement between data and MC performance. Systematic uncertainties on these SFs are dominated by top modelling contributions and the jet energy scale (JES). They are generally under 2% for most p_T bins with increases up to 8% at very low p_T.

Calibration of charm jets

The strategy for charm jet calibrations relies on a selection of semileptonic $t\bar{t}$ events, where one W boson decays leptonically and one decays hadronically [83]. These events are isolated by requiring exactly four jets and a single lepton as well as some MET from the neutrino. A likelihood based reconstruction is used to match each jet to either of the two top decays or the hadronic W decay using only the invariant masses without any flavor identification information. After this association, nearly half of all selected events have a single charm jet associated with the W decay from $W^+ \to c\bar{s}$ or $W^- \to \bar{c}s$. A likelihood based fit is used to extract the charm jet efficiency or mistag rate from this signal by applying previously

computed bottom and light jet SFs to the respective background populations. In `DL1r`, charm jet mistag rates for bottom jet identification are found to be within 5% of unity for most bottom efficiency WPs with uncertainties generally around 5% as well. Lower efficiency WPs and low p_T jets show slightly larger deviations and corresponding uncertainties of up to 20% each. Dominant systematic uncertainties are due to top mismodelling and light jet SF calculations.

Calibration of light jets

Light jets are the most difficult to isolate and thus follow an entirely different calibration approach to bottom and charm jets which does not rely on event selection tricks [91]. Light jets are most easily distinguished from heavy flavor jets by the relative absence of large positive track IPs. The sign of the IP denotes whether the projection of the point of closest approach of the track to the PV relative to the jet axis lies behind the PV (negative IP) or in front of the PV (positive IP). Track IP distributions are approximately symmetric around zero for light jets, but highly asymmetric for bottom and charm jets. The negative part of each distribution is almost identical, which is exploited to calibrate light jets in the "negative tag" method. This method uses a "flipped" flavor tagging algorithm that inverts the signs of all track IPs as well as the decay length before evaluation. Flipping the IPs removes most of the discriminating factors between bottom, charm and light jets and ensures the resulting discriminant distribution closely resembles that of a pure light jet sample (regardless of the actual composition). Light jet SFs are extracted from this distribution in a Z+jets[4] sample using a likelihood based fit. Light jet mistag rate SFs for bottom jet identification are shown to be close to unity for `DL1r` with uncertainties of around 10-20%. Dominant systematics include extrapolation uncertainties between the nominal and "flipped" algorithms as well as MC modelling.

[4] Z+jets refers to a Z boson with additional jets from gluon radiation.

Chapter 5

Measurement of $VH(\rightarrow b\bar{b}/c\bar{c})$ using Run 2 Data

The Higgs Yukawa couplings define how the Higgs boson interacts with different fermions through the underlying mass generation mechanism of the Standard Model as described in detail in section 2.3. Some, such as the bottom Yukawa coupling, are well constrained to within a few percent of their theoretical expectation. Others remain poorly constrained, leaving room for deviations from SM predicted interaction rates that could indicate the presence of BSM physics and bring answers to questions surrounding the quark mass hierarchy. Among these, the charm Yukawa coupling presents the next frontier and its study is enabled by the ongoing improvements in charm jet identification at the ATLAS experiment detailed in section 4.2 and chapter 6.

Building on the object reconstruction techniques described in chapter 4, measuring the bottom and charm Yukawa couplings with ATLAS data presents several fundamental challenges. The main difficulty in these measurements lies in the indistinct signatures and low rate of $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ decays in isolation as well as high background contamination. As detailed in section 2.3.3, there are four common Higgs production modes with unique signatures observed at the LHC. Of these, "Higgs Strahlung" (VH) is considered the most sensitive to measurements of the bottom and charm Yukawa couplings due to its comparatively less background dominated environment relative to the leading production modes (ggF and VBF). The isolation of VH production events with subsequent Higgs to bottom or charm decays is best achieved by studying instances where the associated vector bosons decay to leptons, which are easily identifiable in the ATLAS detector. To suppress backgrounds further, performant jet flavor identification algorithms are employed. As a result, constraints can be placed on both bottom and charm Yukawa couplings. This can be achieved by measuring the signal strength μ of the $VH(\rightarrow b\bar{b})$ and $VH(\rightarrow c\bar{c})$ processes.

Historically, the first evidence for the Higgs to bottom quark coupling at the LHC was found in 2017 via the VH channel by ATLAS [49] and CMS [114] using a combination of run 1 and early run 2 data, though similar results were also achieved at the Tevatron 5 years earlier [124]. The first definitive observations of the coupling were made at the LHC a year

later in 2018 through a combination of results from different production modes dominated by VH [58, 115]. The Higgs to charm coupling has not been measured to the same degree of accuracy with the first direct search on a partial run 2 dataset published by CMS in 2020 [116] showing an observed (expected) upper limit of 70 (37) times Standard Model predictions which was reduced by subsequent full run 2 measurements by ATLAS to 26 (31) times Standard Model [81] in 2022 and by CMS to 14 (7.6) times Standard Model in 2023 [119]. The Higgs to bottom coupling has also been measured in other Higgs production modes (i.e. ggF, VBF and $t\bar{t}H$), either directly or through inclusive measurements which have provided less sensitive results in comparison [121, 103, 120, 117, 79].

This chapter presents a particular measurement of the Higgs to bottom and Higgs to charm couplings via the $V(\to ll)H(\to b\bar{b}/c\bar{c})$ process using ATLAS run 2 data. An overview of the analysis selection and main backgrounds is given (section 5.1) followed by a detailed description of systematic uncertainties (section 5.2) and the likelihood based fit model (section 5.3). To validate the fit model, a separate but analogous measurement of the similar $VZ(\to b\bar{b})$ and $VZ(\to c\bar{c})$ processes is also performed. Finally, results are presented for VH and VZ measurements and constraints on the bottom and charm Yukawa couplings are placed (section 5.4).

5.1 Analysis Overview

The run 2 "legacy"[1] $V(\to ll)H(\to b\bar{b}/c\bar{c})$ analysis is based on 140 fb^{-1} of ATLAS data collected between 2015 and 2018 at a center of mass energy of $\sqrt{s} = 13$ TeV. It supersedes three previous measurements of vector boson associated Higgs production decaying into $b\bar{b}$ in the low [68] and high momentum [75] regimes and $c\bar{c}$ in the low momentum [81] regime using run 2 ATLAS data[2] through a combined analysis strategy. This approach allows for a common description of systematic uncertainties between $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ analysis regions, which are separated by jet flavor identification. In addition to the combined analysis description, the main improvements respective to past results are the addition of DNN-based jet flavor identification via DL1r and the shift from using the Higgs candidate mass as a fit variable to a "multivariate analysis" (MVA) approach.

The low momentum ("resolved") and high momentum ("boosted") regimes are characterized by the p_T of the associated leptonically decaying vector boson (denoted p_T^V), which is highly correlated with the momentum of the recoiling Higgs candidate. Higher momentum Higgs decays generally result in more collimated decay products given the proportionally larger share of momentum originating from their parent particle. Because of this, resolved Higgs candidate reconstruction relies on two individual small-R jets while boosted reconstruction aims to capture the whole Higgs decay in a single large-R jet as discussed in section

[1] The term legacy is used to denote that this analysis is a re-examination of the previously collected run 2 dataset.

[2] The previous iteration of the high momentum $VH(\to b\bar{b})$ measurement uses a partial run 2 dataset only.

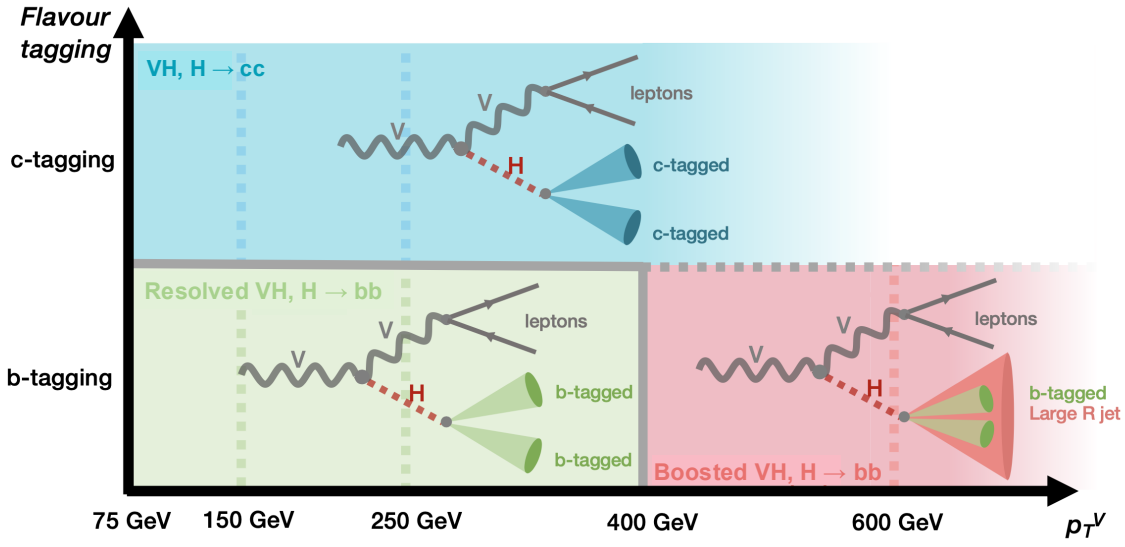


Figure 5.1: Visualization of the three $V(\to ll)H(\to b\bar{b}/c\bar{c})$ analysis regimes under consideration from [72].

4.1.4. The boundary between the two regimes is drawn at $p_T^V = 400$ GeV for $VH(\to b\bar{b})$. No explicit boosted $VH(\to c\bar{c})$ regime exists, but events above the $p_T^V = 400$ GeV threshold are included regardless using a resolved reconstruction strategy. The analysis strategies of the three components are visualized in Fig. 5.1.

Each analysis regime encompasses both W and Z boson associated production, which have significantly different physical signatures due to their different leptonic decay modes, visualized with Feynman diagrams in Fig. 5.2. Neutral Z bosons decay leptonically around 30% of the time, of which one third are $Z \to \ell^+\ell^-$ and two thirds are $Z \to \nu_\ell\bar{\nu}_\ell$ decays, resulting in the presence of either zero or two easily identifiable charged leptons in the detector. Charged W bosons always produce one charged lepton and a neutrino as $W^+ \to \ell^+\nu_\ell$ or $W^- \to \ell^-\bar{\nu}_\ell$ in the approximately 33% of cases where they decay leptonically. Current theory expectations for inclusive Higgs production cross sections are low for each process at ~ 29.82 fb for $Z(\to \ell^+\ell^-)H$, ~ 154.09 fb for $W^\pm \to \ell^\pm\nu_\ell$ and ~ 177.62 fb for $Z \to \nu_\ell\bar{\nu}_\ell$ at $\sqrt{s} = 13$ TeV and $m_H = 125$ GeV [129].

All three cases are considered in this analysis, with charged leptons being reconstructed explicitly and neutrinos identified indirectly via missing energy (MET) as detailed in section 4.1. Of the possible leptonic vector boson decay modes, only $Z \to \tau^+\tau^-$ is deliberately not treated as signal due to frequent hadronic decays of tau-leptons that do not provide clear leptonic signatures. This excludes one third of $Z \to \ell^+\ell^-$ events given that lepton

universality requires the rates of decay to electrons, muons and taus to be equal. The rest of the leptonic decay modes of vector bosons are categorized by the presence of charged leptons and placed in the zero, one or two lepton channels, which have different selection requirements.

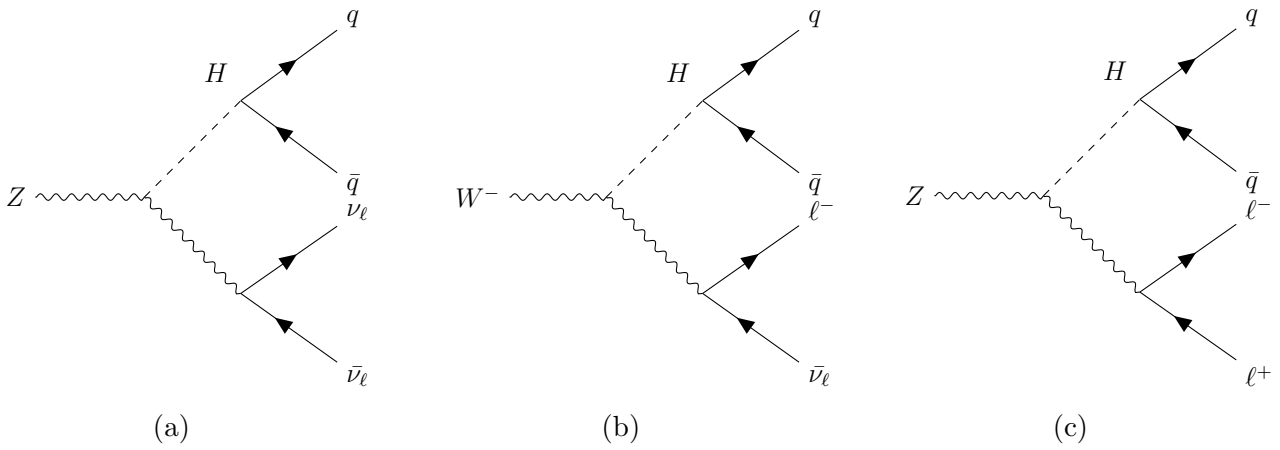


Figure 5.2: Feynman diagrams showing semi-leptonic $VH(\rightarrow q\bar{q})$ decays in the (a) zero, (b) one and (c) two lepton channels.

The parameters of interest are the signal strengths $\mu_{VHb\bar{b}}$ and $\mu_{VHc\bar{c}}$ which are related to the Higgs to bottom and charm Yukawa coupling constants using the κ-framework discussed in section 2.3.3. Validation measurements are performed targeting the $VZ(\rightarrow b\bar{b})$ and $VZ(\rightarrow c\bar{c})$ processes, which have similar signatures given their identical final states across the analysis phase space but larger predicted yields and different invariant masses. In this case the parameters of interest are replaced by $\mu_{VZb\bar{b}}$ and $\mu_{VZc\bar{c}}$, but the overall analysis strategy remains mostly unchanged. Given the relatively high expected sensitivity of the $VH(\rightarrow b\bar{b})$ measurement, the "Simplified Template Cross Sections" (STXS) framework (Stage 1.2) [16] is used to derive particle-level production cross sections in bins of p_T^V and jet multiplicity separately for the WH and ZH production modes. Bin boundaries are defined at the generator level inclusively in the Higgs decay, thereby minimizing theory dependence. This allows for a more granular treatment of theory uncertainties across the analysis phase space as well as easier combinations with other measurements due to shared bin definitions. Stage 1.2 considers 15 separate fiducial cross-section parameters for this analysis consisting of 7 for the WH process and 8 for ZH across 5 p_T^V intervals and 2 jet multiplicity categories[3].

5.1.1 Analysis Backgrounds

The VH target signature of the analysis consists of a leptonic V decay and a hadronic H decay resulting in a final state containing two leptons and a minimum of two quark initiated

[3]Jet multiplicity categories are only split for some p_T^V bins given statistical limitations.

jets identified as bottom or charm jets through flavor tagging. Semi-leptonic VH processes can be initiated by quarks or gluons as indicated in the lowest order Feynman diagrams of Fig. 5.3. A variety of background processes mimic this signature across different regions of the analysis phase space. These are characterized as V+jets, top quark (consisting of $t\bar{t}$, Wt and single-top), diboson and multi-jet in decreasing order of expected yield and are detailed below. The multi-jet (i.e. QCD radiation) background encompasses jets produced by gluon splitting (i.e. $g \to gg$ and $g \to q\bar{q}$) and is found to be negligible across most of the analysis phase space despite its high cross section since it does not present a significant source of leptons. It is only considered in the one lepton channel where it is a minor background after the analysis event selection is applied as detailed in section 5.1.3.

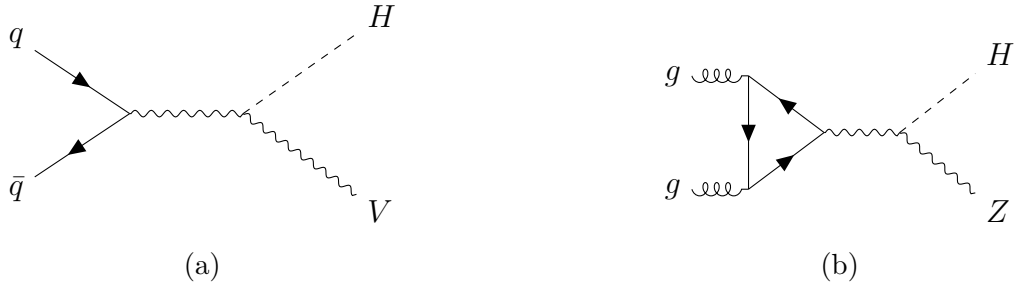


Figure 5.3: Lowest order Feynman diagrams for (a) $qq \to VH$ and (b) $gg \to ZH$ processes.

V+jets

The V+jets background category encompasses events containing vector bosons produced in association with jets from multi-jet radiation (i.e. gluon splitting) which can appear similar to hadronic Higgs decays as shown in Fig. 5.4. This can be split into W+jets, which is significant in the zero and one lepton channels and Z+jets, which is significant in the zero and two lepton channels. Both processes have very high cross sections predicted at ~ 20.08 nb for W+jets and ~ 1.89 nb for Z+jets with subsequent leptonic vector boson decays at $\sqrt{s} = 13$ TeV [46]. Due to jet multiplicity and flavor requirements, the majority of these events do not enter the analysis phase space, but V+jets remains the dominant background overall. The presence of W+jets events in the zero lepton channel can be attributed to leptons missed during reconstruction as well as hadronically decaying τ-leptons misidentified as jets. Because of this, events with hadronic τ decays identified according to section 4.1.5 are migrated to the one lepton channel. Both V+jets categories are further subdivided based on jet flavor as "V+hf" (heavy flavor) for two bottom or two charm jets, "V+lf" (light flavor) for two light jets and "V+mf" (mixed flavor) for any other combination. Generally, "V+hf" is a dominant background for $VH(\to b\bar{b})$, while the others are mostly important in the $VH(\to c\bar{c})$ regime given flavor requirements.

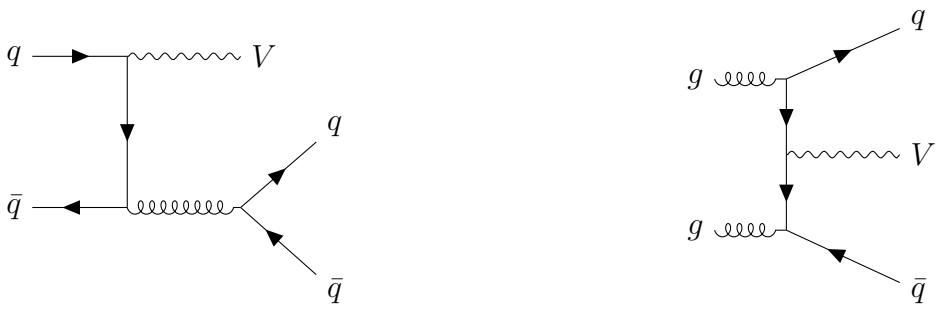


Figure 5.4: Lowest order Feynman diagrams for V+jets processes.

Top

Backgrounds involving top quarks fall into two main categories: single-top and top pair production ($t\bar{t}$), which enter the analysis phase space due to the dominant $t \to W^+b$ decay mode followed by leptonic or hadronic W boson decays. Single-top processes are separated by production channel as s-channel, t-channel and Wt-channel and are generally minor backgrounds with predicted inclusive cross sections of ~ 11.03 pb [172], ~ 214.2 pb [26, 145] and ~ 79.5 pb [162] respectively at $\sqrt{s} = 13$ TeV. Many of these events are rejected given the difficulty of fulfilling the lepton and jet requirements simultaneously in a single top quark decay, which generally either contain two leptons and one quark jet or three quark jets and no leptons depending on the W decay. Top pair production is a much more significant background in many channels due to both its larger predicted inclusive cross section of ~ 833.9 pb [144] at the same energy as well as a larger number of possible final states matching the VH signal signature. Lowest order Feynman diagrams for $t\bar{t}$ and single-top processes are shown in Fig. 5.5 and Fig. 5.6 respectively.

Due to structural similarities between $t\bar{t}$ and single-top Wt, these two are combined into a single category denoted "Top" for this analysis, which is split by jet flavor as "Top($b\bar{b}$)" and "Top(bq/qq)" and is dominated by $t\bar{t}$. These events most frequently pass selections when one W decays leptonically and the other decays hadronically, which makes them appear predominantly in the zero and one lepton channel. "Top($b\bar{b}$)" is a particularly important background in the $VH(\to b\bar{b})$ regime given the guaranteed presence of bottom jets in top quark decays. "Top(bq/qq)" is more relevant in the $VH(\to c\bar{c})$ regime with other quark flavors coming from hadronically decaying W bosons. The s-channel of the single-top background is negligible across the analysis phase space and the t-channel is a minor background in the zero and one lepton channels of $VH(\to b\bar{b})$.

Diboson

Diboson production encompasses WW, ZZ and WZ and presents nearly identically to the VH signature in cases where one vector boson decays leptonically and one decays hadroni-

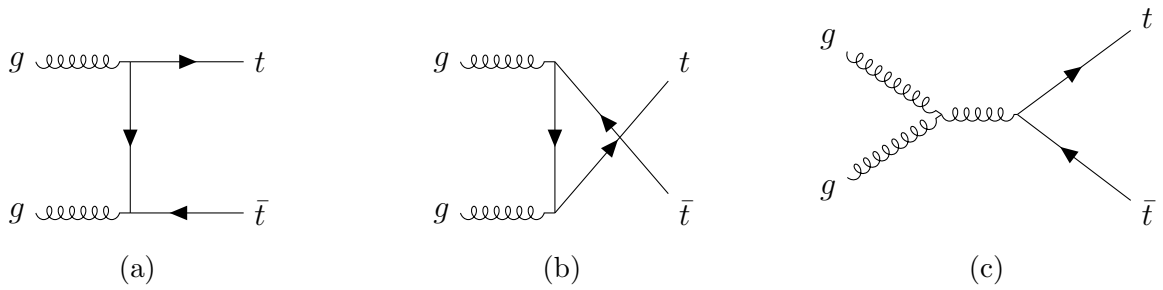


Figure 5.5: Lowest order Feynman diagrams for gluon induced (a) t-channel, (b) u-channel and (c) s-channel $t\bar{t}$ production. Quark induced $t\bar{t}$ diagrams are similar.

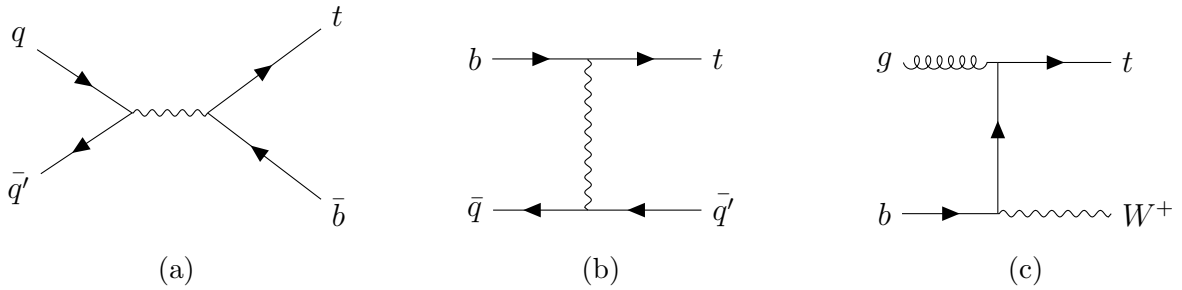


Figure 5.6: Lowest order Feynman diagrams for (a) s-channel, (b) t-channel and (c) Wt-channel single-top production.

cally. Events are grouped to indicate the hadronically decaying Higgs candidate as VW and VZ. Predicted cross sections are comparatively lower for these than other backgrounds at ~ 118.8 pb for WW [118], ~ 16.91 pb for ZZ [30] and ~ 51.11 pb for WZ [142] production at $\sqrt{s} = 13$ TeV which are further suppressed by the vector boson branching ratios. Lowest order Feynman diagrams visualizing these processes are shown in Fig. 5.7. In the $VH(\to b\bar{b})$ regime, WZ and ZZ are respectively dominant in the one and two lepton channels while the zero lepton channel contains equal contributions of both. In $VH(\to c\bar{c})$, WW is dominant in one lepton, WZ in two lepton and both contribute in zero lepton. Overall diboson events are present across the entire analysis phase space given their inherent similarities to VH. The $VZ(\to b\bar{b})$ and $VZ(\to c\bar{c})$ components used to validate the VH fit model in a separate measurement of $\mu_{VZb\bar{b}}$ and $\mu_{VZc\bar{c}}$ are split from the other diboson contributions and treated separately.

5.1.2 Analysis Samples

All signal and major background processes are modelled with MC simulations based on Standard Model predictions using combinations of `Powheg` [3], `Pythia` [196] and `Sherpa` [18]

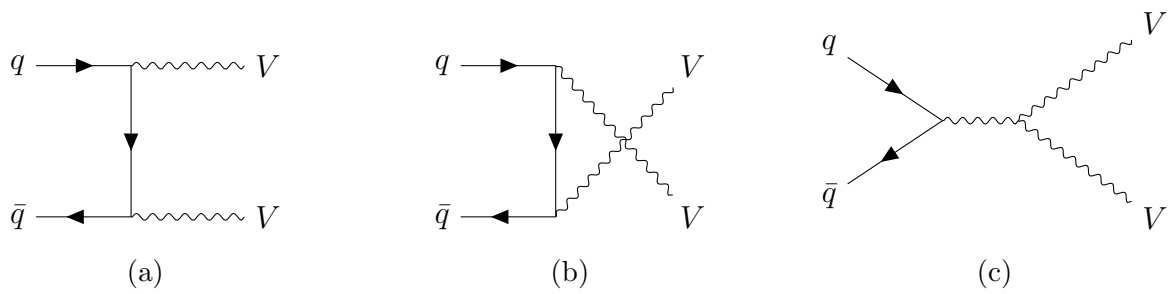


Figure 5.7: Lowest order Feynman diagrams for (a) t-channel, (b) u-channel and (c) s-channel VV processes.

as detailed in table 5.1. For some samples, different generators are used to calculate matrix elements for the hard scatter collisions and subsequent parton shower and hadronization processes as described in section 2.4.3. Since `Powheg` computes the matrix elements for the hard scatter collision only, the `Powheg`+`Pythia` combination is favored for processes with simple hard-scatter collisions where NLO accuracy is the priority. In events with signatures that contain QCD radiation, such as V+jets, `Sherpa` is preferred due to its better handing of different jet multiplicities. `Geant4` [135] is used to simulate detector response in all cases. The Higgs mass is set at 125 GeV with a branching ratio to $b\bar{b}$ of 58% [104]. For smaller backgrounds such as top quark production in the two lepton channel of $VH(\rightarrow b\bar{b})$ and multi-jet production, data-driven techniques are used for estimation. Background modeling is discussed in detail in section 5.2.2.

5.1.3 Event Selection

The core of $VH(\rightarrow b\bar{b}/c\bar{c})$ selection lies in reconstructing and identifying leptonically decaying vector bosons and hadronically decaying Higgs candidates within a pp collision event. In order to accomplish this a set of lepton and jet flavor identification criteria are defined, as detailed in this section.

Vector Boson Reconstruction

The identification of leptonically decaying vector bosons in VH events relies on the presence of electrons, muons, hadronic taus or neutrinos in the form of E_T^{miss} (MET) which are reconstructed as described in section 4.1. A series of selection criteria are detailed for all lepton candidates under consideration in table 5.2. Three sets of cuts are defined for both electrons and muons, with "VH Loose" representing a loose cut to reject possible additional lepton candidates or poorly reconstructed leptons. The "WH Signal" and "ZH Signal" cuts are comparatively tight and designed to retain high quality lepton candidates for the WH and ZH signal. Since neutrinos cannot be observed in the detector directly, the measurement

Process	Matrix element generator	Matrix element PDF	PS and hadronization	UE tune	Cross-section order
Signal					
$qq \to VH$	Powheg Box v2 [3] + GoSam [128] + MiNLO [148]	PDF4LHC15NLO [23]	Pythia 8.2 [196]	AZNLO [205]	NNLO(QCD) +NLO(EW)
$gg \to ZH$	Powheg Box v2	PDF4LHC15NLO	Pythia 8.2	AZNLO	NLO+NLL
Top quark					
$t\bar{t}$	Powheg Box v2	NNPDF3.0NLO [12]	Pythia 8.2	A14 [205]	NNLO+NNLL
single-top (s channel)	Powheg Box v2	NNPDF3.0NLO	Pythia 8.2	A14	NLO
single-top (t channel)	Powheg Box v2	NNPDF3.0NLO	Pythia 8.2	A14	NNLO
Wt	Powheg Box v2	NNPDF3.0NLO	Pythia 8.2	A14	Approx. NNLO+NNLL
V+jets					
V+jets	Sherpa 2.2.11	NNPDF3.0NNLO	Sherpa 2.2.11	Default	NNLO
Diboson					
$qq \to VV$	Sherpa 2.2.11	NNPDF3.0NNLO	Sherpa 2.2.11	Default	NLO
$gg \to VV$	Sherpa 2.2.2	NNPDF3.0NNLO	Sherpa 2.2.2	Default	NLO

Table 5.1: Monte Carlo samples used in the $VH(\to b\bar{b}/c\bar{c})$ analysis adapted from [104]. This table summarizes generators used for matrix element calculation as well as parton shower (PS) and hadronization modeling. The specific parton distribution function (PDF) and underlying event (UE) tune are also given for each sample.

of a significant amount of MET is taken as an indirect indication for their presence. These cuts are set to retain events with high energy neutrinos only.

Events in the zero lepton channel should be consistent with the presence of neutrinos through a $Z \to \nu_\ell\nu_\ell$ decay and therefore must contain no "VH Loose" electrons or muons and have $E_T^{\text{miss}} > 150$ GeV in the resolved and $E_T^{\text{miss}} > 200$ GeV in the boosted regime. In this channel, MET is taken to be equivalent to p_T^V. If a hadronic τ is reconstructed, an additional cut $m_T^W > 10$ GeV on the transverse invariant mass of the W[4] is applied in the resolved regime to remove contamination from fake τ candidates. Events with hadronic τ candidates reconstructed in the zero lepton channel are migrated to the one lepton channel.

In the one lepton channel, one charged lepton and one neutrino is expected via a $W \to \ell\nu_\ell$ decay, so events must contain at least one "WH Signal" lepton and no additional "VH Loose" leptons. Further cuts on MET are applied to reduce the multi-jet background for events containing an electron such that $E_T^{\text{miss}} > 30$ GeV in the resolved and $E_T^{\text{miss}} > 50$ GeV in the boosted regime. The W candidate momentum is calculated through the vector sum of the lepton and MET and must be larger than 75 GeV. A veto is also applied on

[4]In this case the W four vector is reconstructed from the hadronic τ decay and MET

events reconstructed in the one lepton channel with hadronic τ leptons present to reduce $t\bar{t}$ backgrounds with $W \to \tau\nu_\tau$ decays. This does not apply to events with hadronic τ leptons migrated from the zero lepton channel.

The two lepton channel requires the presence of two "VH Loose" leptons of the same flavor to indicate a $Z \to \ell^+\ell^-$ decay, with at least one also satisfying the "ZH Signal" classification. These cuts are relatively looser than in the one lepton channel given the combinatorial difficulties of finding two high quality leptons, which would result in a significant stat reduction. Opposite charge requirements are also enforced for muons, but not for electrons due to the higher rate of charge misidentification. The invariant mass of the two lepton system must be consistent with a Z boson where $81 < m_{\ell\ell} < 101$ GeV in the resolved and $66 < m_{\ell\ell} < 116$ GeV in the boosted regime given the comparatively worse mass resolution at high momentum.

ℓ	**Selection**	p_T	$\lvert\eta\rvert$	d_0/σ_{d_0}	$\lvert\Delta z_0 \sin\theta\rvert$	n_{trk}	ID	Isolation
e	VH Loose	> 7 GeV	< 2.47	< 5	< 0.5 mm	-	LH Loose	Loose_VarRad
	ZH Signal	> 27 GeV	< 2.47	< 5	< 0.5 mm	-	LH Loose	Loose_VarRad
	WH Signal	> 27 GeV	< 2.47	< 5	< 0.5 mm	-	LH Tight	HighPtCaloOnly
μ	VH Loose	> 7 GeV	< 2.7	< 3	< 0.5 mm	-	Loose	Loose_VarRad
	ZH Signal	> 27 GeV	< 2.5	< 3	< 0.5 mm	-	Loose	Loose_VarRad
	WH Signal	> 25 GeV[5]	< 2.5	< 3	< 0.5 mm	-	Medium	HighPtCaloOnly
τ	-	> 20 GeV	< 2.5	-	-	1 or 3	Loose	-

Table 5.2: Lepton selection criteria used in the $VH(\to b\bar{b}/c\bar{c})$ analysis.

Higgs Candidate Selection

While the vector boson reconstruction serves to identify possible Higgs production events, the Higgs candidate is identified through its decays to pairs of bottom or charm quarks for the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ measurements. To isolate these types of events, selections are placed on Higgs candidates based on the structure of the decay as well as the flavor of the resulting jets. In the resolved regime, Higgs candidates are defined by the presence of two small-R jets reconstructed using PFlow objects as described in section 4.1.4. In the boosted regime, a single large-R jet encompassing the whole Higgs decay is built from calorimeter clusters given the high level of collimation between individual quark jets in high momentum decays. To isolate the component quarks, variable radius (VR) jets are reconstructed within the boundary of each large-R jet from particle tracks only, which are more easily separable than calorimeter clusters but provide significantly worse energy resolution. These have a p_T dependent radius parameter defined by $R = \rho/p_T$ (where $\rho = 30$ GeV), are bounded by

[5]The cut is placed at > 27 GeV in cases where $p_T^V < 150$ GeV.

$R_{max} = 0.4$ and $R_{min} = 0.02$ and ghost associated to large-R jets. Small-R jets are defined as candidate signal jets if $p_T > 20$ GeV and $|\eta| < 2.5$. Large-R jets must fulfill $p_T > 250$ GeV and $|\eta| < 2.0$ while track jets have requirements of $p_T > 10$ GeV and $|\eta| < 2.5$. These cuts are set given the inherent degradation of jet reconstruction quality at high $|\eta|$ and low p_T. Final signal jet selection in both regimes then relies on jet flavor identification techniques, which differ between the two but are based on the `DL1r` algorithm detailed in section 4.2.2.

In the resolved regime, the simultaneous measurement of $VH \to b\bar{b}$ and $VH \to c\bar{c}$ requires orthogonal bottom and charm jet identification which is accomplished via a pseudo-continuous flavor tagging (PCFT) scheme. This scheme is visualized in Fig. 5.8a and results in four distinct designations based on the `DL1r` bottom and charm tagging discriminants: b-tagged for jets passing the 70% bottom WP, c-tight for jets passing the 25% charm WP but failing the 70% bottom WP, c-loose for jets passing the 45% charm WP but failing both previous selections and untagged for jets failing all other selections. Events are then split into regions based on the designation of the jets within as detailed in Fig. 5.8b. In each event, two jets are chosen as the Higgs candidate. The BB category requires exactly two b-tagged jets, which corresponds to likely $VH \to b\bar{b}$ signal events. In the zero and one lepton channels, there is an additional requirement that BB events do not contain additional c-tight jets to limit the presence of top backgrounds. Instead, zero and one lepton events with at least one b-tagged and one c-tight jet are categorized as BC_T, which is designed to isolate top events. In BC_T regions, the highest p_T jets of each tag designation are chosen as signal jets. All other regions require zero b-tagged jets and are defined by the two highest priority jets based on tag designation going from c-tight (C_T) to c-loose (C_L) to untagged (N). The C_TC_T and C_TC_L categories are merged as C_TC and capture likely $VH \to c\bar{c}$ signal events. C_TN and C_LN are also defined and are used to isolate other backgrounds as described in section 5.1.4.

In the boosted regime, a simplified scheme is followed due to the lack of a dedicated $VH \to c\bar{c}$ measurement in this regime and the lower total number of expected events. The large-R jet with the highest p_T and at least two ghost associated track jets is chosen as the Higgs candidate in each event. The three leading p_T track jets associated to the Higgs candidate large-R jet are then classified via `DL1r`. A looser 85% bottom identification WP is used due to lower statistics at high p_T. If exactly two out of three track jets pass this WP, the jet is considered b-tagged, otherwise it is untagged. In all cases, the momentum and mass of the Higgs candidate must satisfy $p_T^{H_{cand}} > 45$ GeV (only relevant in resolved) and $m_{H_{cand}} > 50$ GeV respectively.

Truth Tagging

Due to the low charm-jet identification efficiency, direct cuts on discriminant scores through the defined PCFT bins result in a significant loss of MC statistics across various analysis regions. This strategy is called "direct tagging" and removes events that do not pass flavor tagging requirements from the analysis completely. To maximize the statistical power of available MC samples, an alternate "truth tagging" strategy is employed, which retains

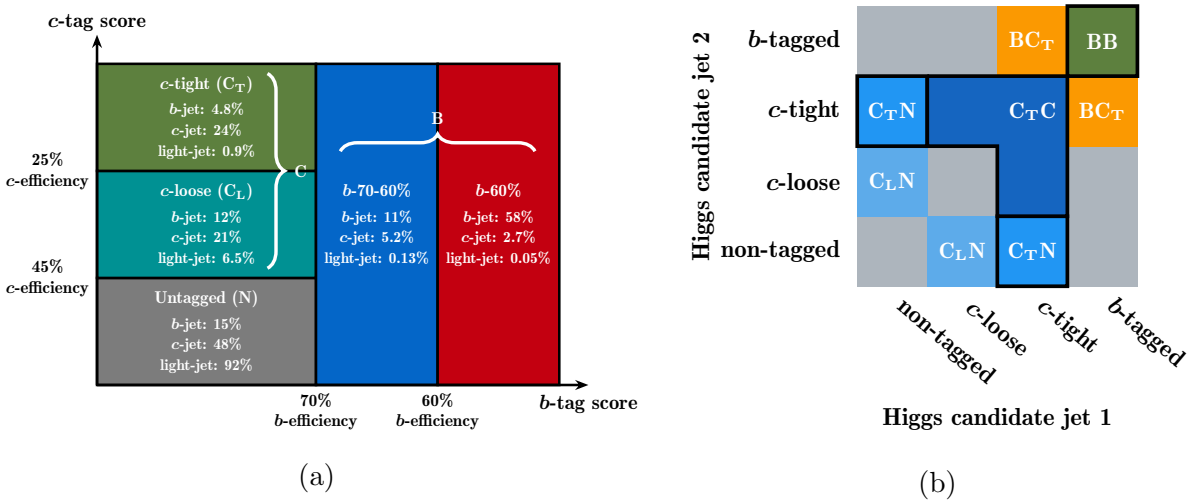


Figure 5.8: Graphics detailing the PCFT scheme employed in the resolved regions of the $VH(\to b\bar{b}/c\bar{c})$ analysis showing PCFT categories (a) and resulting regions (b) from [104].

events that do not pass flavor identification requirements by using a weighted combination of all events based on the expected flavor identification efficiency. Each individual weight is computed through a product of tagging efficiencies ϵ_j for all jets in an event that pass flavor tagging cuts and tagging inefficiencies $1 - \epsilon_j$ for all jets that do not pass these cuts. All individual ϵ_j values are derived using a Graph Neural Network (GNN) based approach [8], which takes into consideration the relationships between jets by using their associated kinematic information as inputs. In the resolved regime one GNN is trained for each sample while in the boosted regime one GNN is trained for each generator. The specific tagging strategy varies per sample and region between direct tagging, truth tagging and a mixed "hybrid tagging" approach where bottom jets are tagged directly while truth tagging is employed for all other jets in an event. The decision which approach to take is based on available MC statistics as well as agreement (closure) between truth tagged and direct tagged MC distributions for each sample. To avoid potential non-closure issues, a cautious approach is taken to the use of GNN truth tagging in this analysis. In resolved $VH \to b\bar{b}$, the direct tagging approach is used for all samples except V+jets and the s- and t-channels of single-top, where hybrid tagging is employed instead. Similarly, in $VH \to c\bar{c}$ these same samples employ truth tagging for 8% of all MC events, which are used to populate $C_T C_T$, while the other 92% as well as all other samples use direct tagging. In boosted $VH \to b\bar{b}$, truth tagging is applied to all samples except for signal and diboson.

Additional Selections and Corrections

To combat background contamination from multi-jet events, other selections are applied in the zero and one lepton channels. In both cases, the scalar sum of all jet transverse

momenta in an event S_T must exceed 120 GeV when two jets are present and 150 GeV in the case of three or more jets. In the zero lepton channel, additional angular cuts are applied requiring $|\Delta\phi(j_1, j_2)| < 140°$ in the resolved regime as well as $|\Delta\phi(E_T^{miss}, H_{cand})| > 120°$ and $\min[|\Delta\phi(E_T^{miss}, j_i)|] > 20°$[6] in general. In each case j_i are the selected Higgs candidate small-R or track jets. After event selection, a series of corrections are applied to Higgs candidate jets to improve their momentum resolution based on the presence of muons, final state radiation (FSR) and missing energy from neutrinos in bottom or charm jets [86]. Applied cuts reduce multi-jet yields to make up between 0.1% and 5.5% of backgrounds across the zero and one lepton channels [80].

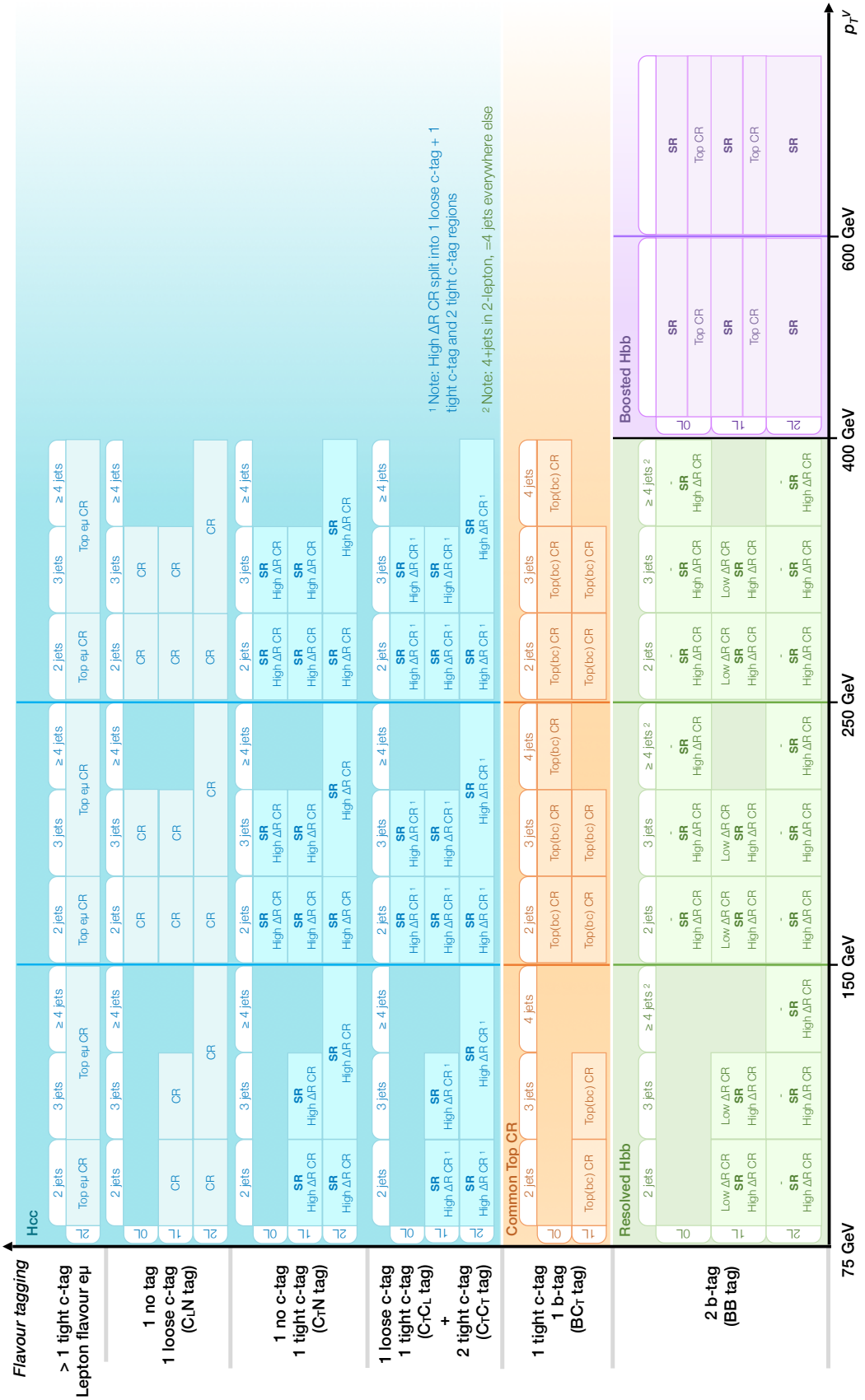
5.1.4 Event Categorization

Selected events are split into orthogonal signal regions (SR) and control regions (CR). SRs are designed to be as pure as possible in signal $VH \rightarrow b\bar{b}$ or $VH \rightarrow c\bar{c}$ events while CRs are used to understand and constrain remaining backgrounds. Multiple different types of CRs are used to model all dominant backgrounds across the analysis phasespace. A broad overview of this set-up is presented in Fig. 5.9. Distinct signal regions exist for $VH(\rightarrow b\bar{b})$ and $VH(\rightarrow c\bar{c})$ based on jet flavor tagging, with the former consisting of BB and the latter two sets of C_TC and C_TN regions. Different CRs exist for a variety of tag variations, including BB, C_TC, C_TN, C_LN and BC_T. All signal and control regions are split by lepton channel, p_T^V and jet multiplicity to increase analysis sensitivity and perform STXS [16] measurements.

Common Analysis Region Definitions

The p_T^V ranges are kept identical in the two resolved analysis regimes as $75 < p_T^V < 150$ GeV (in one and two lepton channels only), $150 < p_T^V < 250$ GeV and $250 < p_T^V < 400$ GeV. The lowest p_T^V region is not considered in the zero lepton channel due to overwhelming background contamination when reducing the cut on MET (i.e. p_T^V). In the $c\bar{c}$ regions, events with $p_T^V > 400$ GeV are merged into $250 < p_T^V < 400$ GeV regions the while for $b\bar{b}$ there are two additional regions in the boosted regime with $400 < p_T^V < 600$ GeV and $p_T^V > 600$ GeV. When statistical uncertainties allow, events with different jet multiplicities are generally separated into distinct regions. Because of this, jet multiplicity splits vary slightly across the resolved analysis phase but can consist of 2-jet, 3-jet and 4-jet categories including signal jets. The 4-jet category only exists in the zero and two lepton $VH(\rightarrow b\bar{b})$ channels. In the latter, events with more than four jets are also included because of the lower $t\bar{t}$ contamination in this channel, which is too large in the other lepton channels. No 4-jet region exist in the $VH(\rightarrow c\bar{c})$ space, but events with more than three jets are similarly included in the two lepton channel. In the boosted regime, regions are not split based on jet multiplicity at all. The analysis p_T^V and jet multiplicity splits are kept consistent between

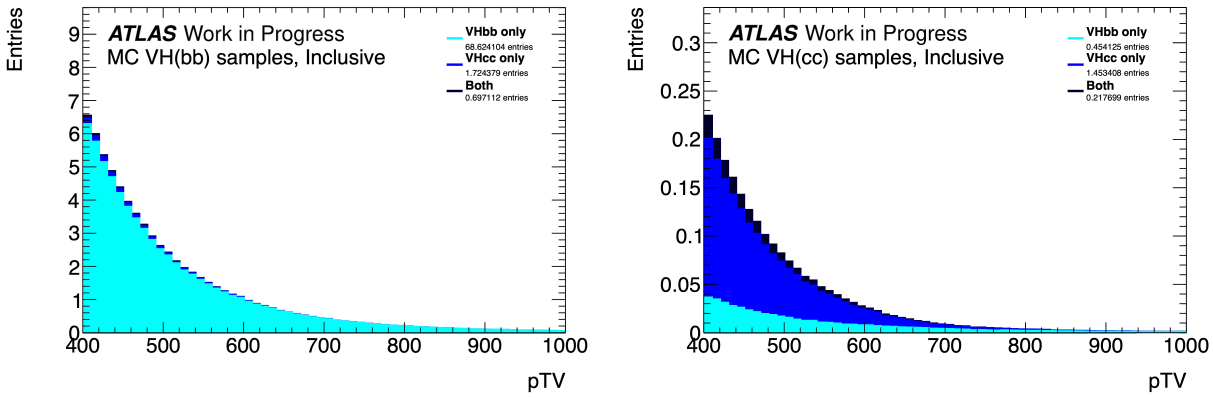
[6] In the boosted regime as well as cases where there are three or more jets in resolved this threshold is raised to 30°.


Flavour tagging
> 1 tight c-tag Lepton flavour eμ
1 no tag 1 loose c-tag (CLN tag)
1 no c-tag 1 tight c-tag (CTN tag)
1 loose c-tag 1 tight c-tag (CTCL tag) + 2 tight c-tag (CTCT tag)
1 tight c-tag 1 b-tag (BCT tag)
2 b-tag (BB tag)
Hcc
Common Top CR
Resolved Hbb
Boosted Hbb
SR
CR
Top eμ CR
Top(bc) CR
Top CR
High ΔR CR
Low ΔR CR
1 Note: High ΔR CR split into 1 loose c-tag + 1 tight c-tag and 2 tight c-tag regions
2 Note: 4+jets in 2-lepton, =4 jets everywhere else
75 GeV
150 GeV
250 GeV
400 GeV
600 GeV
p_T^V

SRs and CRs, of which there are six types in total: three each in resolved $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ (including one type present in both) and one in boosted $VH(\to b\bar{b})$.

Overlap between Regions

All analysis regions are completely orthogonal except for a small overlap in definitions between $b\bar{b}$ and $c\bar{c}$ regions containing events above $p_T^V > 400$ GeV, where no explicit exclusivity between a bottom-jet identification and charm-jet identification based selection is enforced. Because of this, an overlap constituting less than 1% of total MC yields is present in each affected region. As shown in Fig. 5.10, this translates to around 1% of $VH(\to b\bar{b})$ and around 10% of $VH(\to c\bar{c})$ signal events with $p_T^V > 400$ GeV passing both the $c\bar{c}$ and boosted $b\bar{b}$ selections simultaneously. While 10% is a significant fraction, there are no dedicated boosted $VH(\to c\bar{c})$ regions and events with $p_T^V > 400$ GeV are instead included in $p_T^V > 250$ GeV regions where they represent less than 1% of total events. Due to its miniscule size, no further treatment of the overlap is needed in the analysis.



(a) (b)

Figure 5.10: Histograms showing MC events passing boosted $VH(\to b\bar{b})$, $VH(\to c\bar{c})$ or both selections in $VH(\to b\bar{b})$ (a) and $VH(\to c\bar{c})$ signal samples (b).

Overview of Control Regions

The most numerous CRs are so-called "High ΔR Control Regions" (CRHigh), which exist for every resolved $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ SR to constrain heavy and mixed flavor V+jets and $t\bar{t}$ backgrounds. These CRs have an identical selection to their corresponding SRs but contain Higgs candidates with a higher angular separation between the two constituent jets $\Delta R(j_1, j_2)$ than is expected for a Higgs decay. The cut placed on this variable varies for each

region with p_T^V and jet multiplicity but is tuned such that 95% and 85% of signal events end up below the CRHigh thresholds for 2-jet and 3+-jet events respectively. An additional flat cut is placed for both CRHigh and SR at an upper bound of $\Delta R(j_1, j_2) = \pi$ to keep the two regions kinematically similar. Separate CRHigh regions are defined for C_TC_T and C_TC_L to better constrain the heavy and mixed flavor components of the dominant backgrounds. A set of "Low ΔR Control Regions" (CRLow) are also defined for one lepton SRs in resolved $VH(\to b\bar{b})$ only. These regions are specifically designed to further constrain the W+hf background by separating events with lower than expected $\Delta R(j_1, j_2)$ values. As in the CRHigh regions, the cut is dependent on jet multiplicity and p_T^V and is defined to keep 90% of all diboson events in the SRs, thereby not affecting signal purity. In the zero and two lepton channels as well as for $VH(\to c\bar{c})$, no cut is applied and affected events are kept in the SRs.

Additional constraints on the top quark background in resolved $VH(\to b\bar{b})$ are placed via dedicated "Top(bc) Control Regions", which target primarily mixed flavor states with one bottom jet and one charm or light jet. These regions are kept orthogonal to all SRs via a different flavor tagging requirement of BC_T and exist for the zero and one lepton channels only. In the two lepton channel, the top quark background can be constrained significantly by enforcing lepton flavor requirements to remove mixed elements. In $VH(\to c\bar{c})$ these events are put in dedicated "Top($e\mu$) Control Regions", which are merged for the C_TC and C_TN regions. Top(e/μ) CRs do not exist in resolved $VH(\to b\bar{b})$ and the events are instead used directly for a data-driven estimate of the $t\bar{t}$ background as described in section 5.1.1. The final set of CRs employed in the resolved regime for $VH(\to c\bar{c})$ are the "V+light Control Regions" intended to control the light flavor background of V+jets. These consist of events falling into the C_LN category for the zero, one and two lepton channels.

The boosted $VH(\to b\bar{b})$ regime only has a single set of "Top Control Regions" (TopCR) in the zero and one lepton channels. The two lepton channel does not have a dedicated TopCR given the negligible amount of top events passing selections. TopCRs require one additional track-jet identified as coming from a bottom quark via the 85% WP outside of the Higgs candidate large-R jet. Otherwise they are kept identical to the SRs in the boosted regime. A detailed breakdown of the composition of SR and CR selections is given in Fig. 5.11 for selected p_T^V regions across all jet multiplicities, lepton channels and analysis regimes.

In each region, events are separated into discrete bins in the form of histograms with respect to a particular variable. Signal regions are generally binned using a so-called multivariate analysis (MVA) discriminant built to separate signal and background events using a combination of other variables. The MVA algorithm used in this analysis is described in detail in the following section. In other regions, fit variables are chosen based on impact on various systematic uncertainties as discussed in sections 5.2 and 5.3. They consist mostly of the di-jet or large-R jet mass $m_{j_1j_2}$ or m_J and p_T^V. When information related to background distribution shapes is not relevant, single-bin regions are used to constrain background yields only.

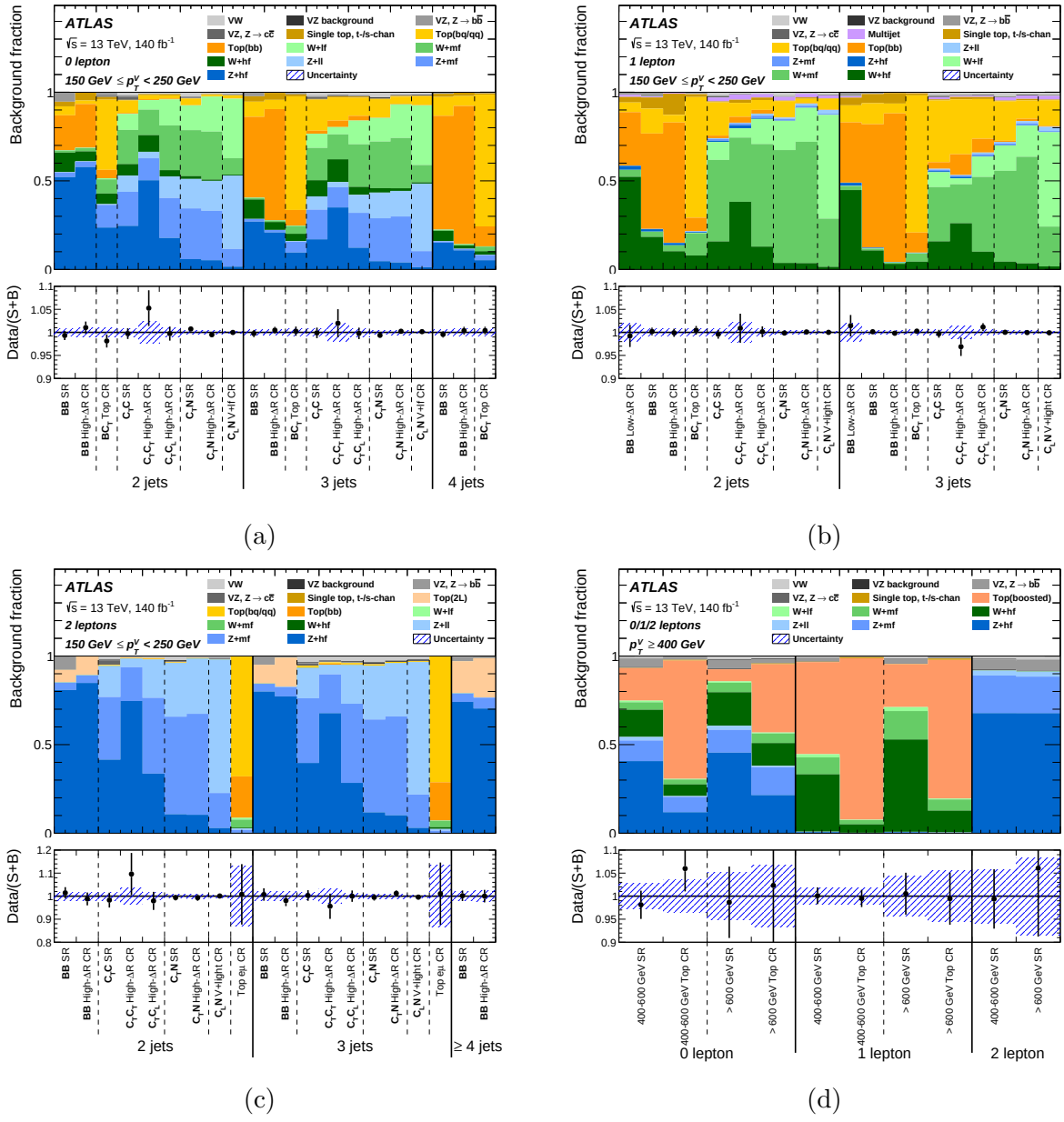


Figure 5.11: Prefit composition of SRs and CRs in 150 GeV $\leq p_T^V \leq$ 250 GeV resolved regions for zero lepton (a), one lepton (b) and two lepton (c) and all p_T^V and lepton regions for boosted (d).

Multivariate Analysis (MVA)

Analysis sensitivity is determined primarily by the degree of separation between VH signal and background events across the analysis SRs. Single variables, such as the Higgs candidate

mass m_H, can sometimes be used to effectively isolate signal events, particularly if the SR is not dominated by backgrounds. In the case of the heavily background dominated search for the $VH \to b\bar{b}$ and $VH \to c\bar{c}$ processes, combining multiple variables into a discriminant proves significantly more effective than a single variable fit. To this end, a series of "Boosted Decision Trees" (BDTs) based on the "AdaBoost" and "Gradient boosting" algorithms as implemented in TMVA [153] are used to identify optimal discriminants and maximize analysis sensitivity [85]. Input variables are detailed in table 5.3 and differ slightly by lepton channel and between the boosted and resolved regimes due to the differences in lepton selection and jet reconstruction respectively. Variables mostly consist of the kinematics of the H, W and Z candidates and their constituent leptons ℓ, large-R jets J and small-R jets or track jets j_i. In particular, $\text{bin}_{\text{DL1r}}(j_i)$ refers to the PCFT bin a particular jet falls into, S_T is the scalar sum of the p_T of the Higgs and leptons and the "Color" variable is a ratio of angular differences between small-R jets within a single large-R jet denoted by $\frac{\theta^2_{j_1,j_2}+\theta^2_{j_2,j_3}}{\theta^2_{j_1,j_2}}$ [85]. The latter is specifically designed to emphasize differences between quark and gluon initiated jets based on color splitting [22].

Trainings are performed individually for $VH \to b\bar{b}$ and $VH \to c\bar{c}$, which define the respective signal categories. Each BDT is trained to separate these signal events from all other backgrounds, excluding $VH \to c\bar{c}$ backgrounds in $VH \to b\bar{b}$ regions and vice versa. This is done using a combination of MC simulation from SRs and CRHigh regions. A second set of BDTs denoted BDT_{VZ} is trained for the diboson reference measurement in a similar fashion. These BDTs use $VZ \to b\bar{b}$ and $VZ \to c\bar{c}$ events as signal but are otherwise identical to the nominal set of BDT_{VH}. MVA trainings are further split in each signal category and lepton channel by p_T^V and jet multiplicity. In the resolved regime, BDTs for all jet multiplicity regions are trained separately while p_T^V trainings are split by $75\text{GeV} \leq p_T^V \leq 150\text{GeV}$ and $p_T^V > 150$ GeV[7] where relevant. In the boosted regime, a single BDT is trained per lepton channel. In addition to BDT_{VH} and BDT_{VZ}, a third set of BDTs denoted $\text{BDT}_{\text{CRLow}}$ is trained on events with exactly two bottom jets in CRLow regions. This training is intended to separate the W+jets background in these regions from the $t\bar{t}$ background, which is constrained well through other CRs. Through this approach, a number of bins pure in the difficult to isolate $W+b\bar{b}$ background are generated. Separate trainings are performed for events with $75\text{GeV} \leq p_T^V \leq 150\text{GeV}$ and $p_T^V > 150$ GeV, but are inclusive in jet multiplicity. Performance of select BDT_{VH} trainings is shown in Fig. 5.12 for all three analysis regimes in the two lepton channel.

In all cases, the application of a "single" BDT implies the training of two separate instances on 50% of all events in a given slice of the dataset. This k-folding approach with $k = 2$ ensures unbiased predictions for MC by separating events used for training from those that are evaluated. In data, half of all events are evaluated by each BDT training chosen at random. The resulting MVA output distributions are finely binned and often entirely dominated by backgrounds in all but the most signal rich bins. To improve analysis

[7]This also includes events with $p_T^V > 400$ GeV in $b\bar{b}$ regions, which are formally part of the boosted regime.

Variable	Resolved			Boosted		
	0 lepton	1 lepton	2 lepton	0 lepton	1 lepton	2 lepton
m_H	✓	✓	✓	✓	✓	✓
$m_{j_1 j_2 j_3}$	✓	✓	✓			
$p_T^{j_1}$	✓	✓	✓	✓	✓	✓
$p_T^{j_2}$	✓	✓	✓	✓	✓	✓
$p_T^{j_3}$				✓	✓	✓
$\sum p_T^{j_i}, i > 2$	✓	✓	✓			
$\mathrm{bin}_{\mathrm{DL1r}}(j_1)$	✓	✓	✓	✓	✓	✓
$\mathrm{bin}_{\mathrm{DL1r}}(j_2)$	✓	✓	✓	✓	✓	✓
p_T^V	$\equiv E_T^{\mathrm{miss}}$	✓	✓	$\equiv E_T^{\mathrm{miss}}$	✓	✓
E_T^{miss}	✓	✓		✓	✓	
$E_T^{\mathrm{miss}}/\sqrt{S_T}$			✓			
$\lvert\Delta\phi(V,H)\rvert$	✓	✓	✓	✓	✓	✓
$\lvert\Delta y(V,H)\rvert$		✓	✓		✓	✓
$\Delta R(j_1, j_2)$	✓	✓	✓	✓	✓	✓
$\min[\Delta R(j_i, j_1 \mathrm{or} j_2)], i > 2$	✓	✓				
N(track-jets in J)				✓	✓	✓
N(add. small-R jets)				✓	✓	✓
Color				✓	✓	✓
$\lvert\Delta\eta(j_1, j_2)\rvert$	✓					
$p_T^H + E_T^{\mathrm{miss}}$	✓					
m_T^W		✓				
m_{top}		✓				
$\min[\Delta\phi(\ell, j_1 \mathrm{or} j_2)]$		✓				
p_T^ℓ					✓	
$(p_T^\ell - E_T^{\mathrm{miss}})/p_T^V$					✓	
$m_{\ell\ell}$			✓			
$\cos\theta^*(\ell^-, V)$			✓			✓

Table 5.3: Input variables used in MVA trainings adapted from [104].

sensitivity by reducing overall statistical uncertainty while preserving relative signal purity in high scoring bins, a non-linear transformation is applied to the initial BDT score distribution. This transformation is described by $Z = z_s \frac{n_s}{N_s} + z_b \frac{n_b}{N_b}$ where $n_{s/b}$ are the number of signal and background events in a given bin, $N_{s/b}$ are the total number of signal and background events and $z_{s/b}$ are tunable parameters that approximately correspond to the number of signal and background rich bins desired in the final distribution. The transformation merges bins starting from the high end of the score distribution until $Z > 1$, at which point a new

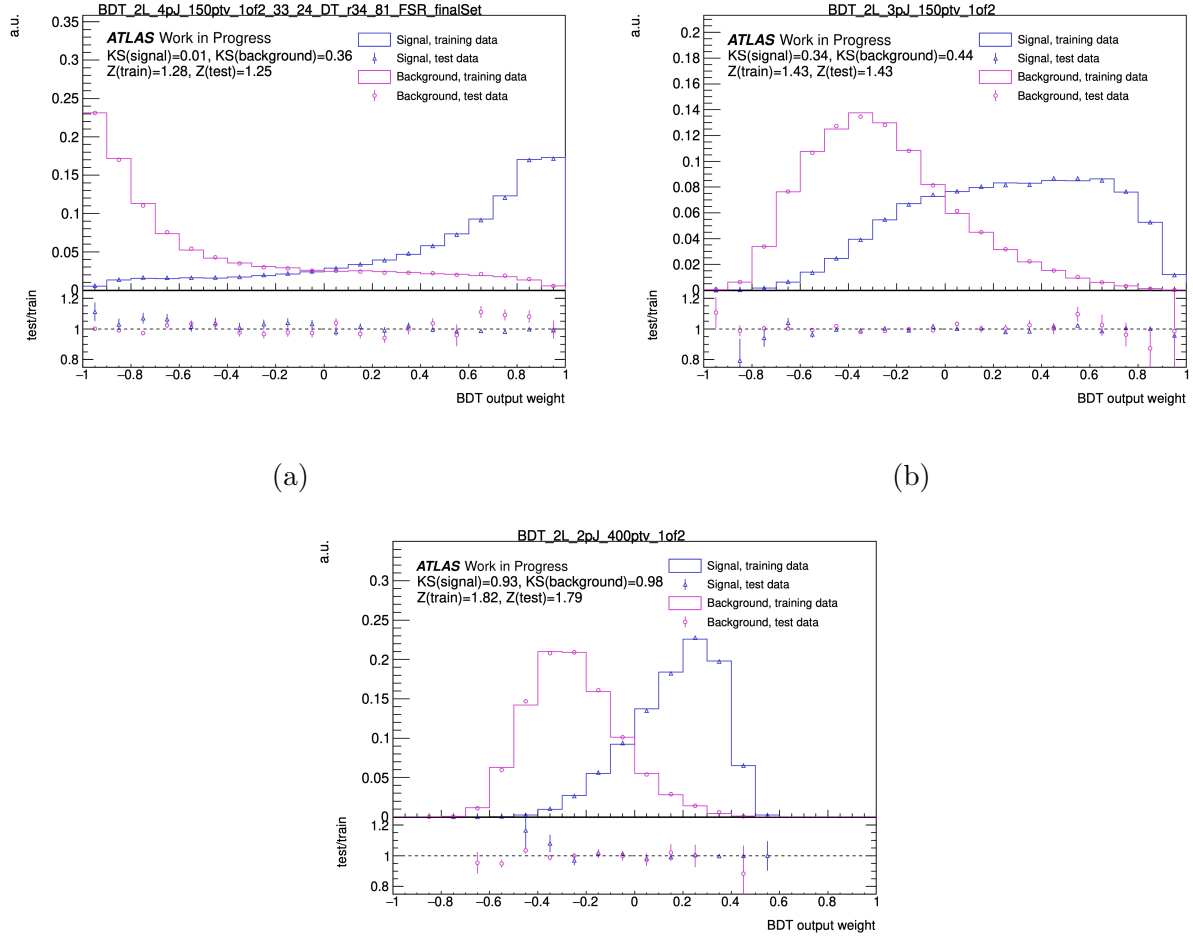


Figure 5.12: Comparison of MVA output distributions from testing and training data for BDT_{VH} in the two lepton channel for the 4+jet, $p_T^V \geq 150$ GeV training in resolved $VH \to b\bar{b}$ (a), the 3+jet, $p_T^V \geq 150$ GeV training in $VH \to c\bar{c}$ (b) and the inclusive training in boosted $VH \to b\bar{b}$ (c)

bin boundary is drawn. Parameters z_s and z_b are optimized independently by region [85].

5.2 Modeling and Uncertainties

The simultaneous measurement of $VH \to b\bar{b}$ and $VH \to c\bar{c}$ necessitates the consideration of a large number of systematic uncertainties, which are frequently correlated between different analysis regions and regimes. In particular, the resolved $VH \to b\bar{b}$ and $VH \to c\bar{c}$

regimes allow for a more correlated treatment of uncertainties, while boosted $VH \to b\bar{b}$ is mostly kept separate due to the vastly different phase space under consideration. Systematics can broadly be split into two categories: experimental systematics and modeling systematics. Experimental systematics are uncertainties related to detector measurements and object reconstruction. These uncertainties depend on the types of objects relevant to the analysis selection for each event and are applicable regardless of the underlying physical process. Modeling systematics are process dependent and originate from imperfections in MC modeling often related to fundamental theoretical uncertainties, finite precision of theoretical calculations or Monte Carlo generator mismodeling. The impact of each systematic uncertainty is quantified by evaluating upward and downward variations from the nominal template for every bin across each sample and analysis region. These variations can affect both the normalization and shape of a sample. Derived uncertainties are used as priors in the fit, which allows for deviations from the nominal model through Gaussian constraint terms as described in section 5.3.

5.2.1 Experimental Systematics

All objects relevant in the event selection and categorization process have associated uncertainties. These are standardized and quantified via the object reconstruction pipelines detailed in chapter 4 for small-R and large-R jets, electrons, muons, MET and hadronic taus by the relevant combined performance (CP) groups. Lepton systematics are mostly related to differences in identification and isolation efficiencies between data and MC as well as energy and momentum resolution effects. Jet uncertainties are separate for small-R and large-R jets and consist of energy and mass scale and resolution uncertainties as well as differences in flavor identification efficiency. Two additional event level systematics stem from uncertainties in pile-up modeling as well as the measurement of the luminosity. All experimental systematics are propagated through the event reconstruction and selection chain. A full list of relevant experimental systematics is given in tables 5.4 and 5.5.

5.2.2 Modeling Systematics

The general approach taken to signal and background modeling follows a common strategy for each process, which are all treated individually. These systematics are mostly derived by considering alternate MC simulated samples generated with individually varying parameters based on underlying theoretical uncertainties. In addition, some variations are applied to the nominal sample directly via event weights rather than the generation of an entirely new sample. Major analysis backgrounds have associated floating normalization uncertainties which allow the fit to freely tune overall background yields based on observations in data to reduce dependence on MC modelling. All other uncertainties have pre-calculated values based on differences observed in nominal and alternative MC samples. Most modeling systematics (aside from floating normalizations) fall into one of three categories: overall acceptance uncertainties, relative acceptance uncertainties and shape uncertainties.

Systematic uncertainty name	Description	Categories
	Event	
LUMI_2015_2018	Uncertainty on total integrated luminosity	all
PRW_DATASF	Uncertainty on pile-up modeling	all
	Electrons	
EL_EFF_Trigger_TOTAL	Trigger efficiency uncertainty	all
EL_EFF_Reco_TOTAL	Reconstruction (reco.) efficiency uncertainty	all
EL_EFF_ID_TOTAL	Identification (id.) efficiency uncertainty	all
EL_EFF_Iso_TOTAL	Isolation efficiency uncertainty	all
EG_SCALE_ALL	Energy scale uncertainty	all
EG_RESOLUTION_ALL	Energy resolution uncertainty	all
	Muons	
MUON_EFF_RECO_STAT(SYS)	Reconstruction and ID efficiency uncertainty for muons with $p_T > 15$ GeV	all
MUON_EFF_RECO_STAT(SYS)_LOWPT	Reco. and ID efficiency uncertainty for muons with $p_T \leq 15$ GeV	all
MUON_EFF_ISO_STAT(SYS)	Isolation efficiency uncertainty	all
MUON_EFF_TTVA_STAT	Track-to-vertex association efficiency uncertainty	all
MUON_SCALE	Momentum scale uncertainty	all
MUON_SAGITTA_DATASTAT(RESBIAS)	Momentum scale uncertainty to cover charge-dependent local misalignment effects	all
MUON_CB	Combined momentum resolution uncertainty of the inner detector (muon spectrometer)	all
MUON_EFF_TrigStat(Syst)Uncertainty	Trigger efficiency uncertainty	all
	Taus	
TAUS_TRUEHADTAU_EFF_RECO_TOTAL	Reconstruction efficiency	all
TAUS_TRUEHADTAU_EFF_ELEOLR_TOTAL	Electron veto efficiency for taus that truth matched to an electron	all
TAUS_TRUEHADTAU_EFF_RNNID_*	RNN ID efficiency	all
TAUS_TRUEHADTAU_SME_TES_*	In-situ Tau Energy Scale correction	all
TAUS_TRUEELECTRON_EFF_ELEBDT_*	Electron veto efficiency SF	all

Table 5.4: Summary of experimental systematics related to leptons and MET used in the analysis adapted from [72].

- **Overall acceptance uncertainties** describe the total yield differences between nominal and alternative MC samples across all analysis regions for smaller backgrounds without floating normalization uncertainties. These are determined via a ratio of total yields between samples.

- **Relative acceptance uncertainties** quantify relative yield discrepancies between analysis regions or processes covered by a single overall normalization factor. This can occur when underlying differences between nominal and alternative MC samples lead to a migration of events from one region to another at the reconstruction level. They are calculated for each sample via double ratios of acceptances in regions A and B as

$$\sigma_{\text{rel.acc.}}(A,B) = \frac{n_B^{\text{alternative}}}{n_B^{\text{nominal}}} \Big/ \frac{n_A^{\text{alternative}}}{n_A^{\text{nominal}}} - 1 \tag{5.1}$$

  where A constitutes the region which is purer in the particular background under examination. The derived double ratio explicitly prevents double counting the overall normalizations and is applied to region B. This is also denoted as an extrapolation uncertainty from region A to region B. In cases where relative acceptance uncertainties are necessary between multiple regions A, B and C for a single background, they are computed starting from the purest region A as $\sigma_{\text{rel.acc.}}(A, B+C)$ followed by $\sigma_{\text{rel.acc.}}(B,C)$. The final relative acceptance uncertainty for each region is a sum in quadrature of uncertainties from all variations under consideration.

Systematic uncertainty name	Description	Categories
	Small-R jets	
JET_BJES_Response	Energy scale uncertainties for b-jets	all
JET_EffectiveNP_Detector{1–2}	Energy scale uncertainties due to in-situ calibration	resolved
JET_EffectiveNP_Mixed{1–3}	Energy scale uncertainties due to in-situ calibration	resolved
JET_EffectiveNP_Modeling{1–4}	Energy scale uncertainties due to in-situ calibration	resolved
JET_EffectiveNP_Statistical{1–6}	Energy scale uncertainties due to in-situ calibration	resolved
JET_EtaIntercal_Modeling(TotalStat)	Energy scale uncertainties to cover η-intercalibration non-closure	resolved
JET_EtaIntercal_NonClosure_2018data	Energy scale uncertainties to cover non-closure with 2018 data	resolved
JET_Flav_Comp(Flavor_Response)	Energy scale uncertainty related to flavor composition (response)	resolved
JET_PU_{OffsetMu(NPV),PtTerm,RhoTopology}	Energy scale uncertainties due to pile-up effects	resolved
JET_PunchThroughMC16	Energy scale uncertainty for "punch-through"	resolved
JET_SingleParticle_HighPt	Energy scale uncertainty for the behavior of high-p_T jets	resolved
JET_JvtEfficiency	JVT efficiency uncertainty	resolved
JET_fJvtEfficiency	fJVT efficiency uncertainty	resolved
JET_JER_DataVsMC_MC16	Energy resolution uncertainty	resolved
JET_JER_EffectiveNP_{1–11,12restTerm}	Energy resolution uncertainties	resolved
	b-tagging: PFlow jets	
FT_EFF_PFlow_Eigen_B_{0–44}	b-tagging efficiency uncertainties for b-jets	resolved
FT_EFF_PFlow_Eigen_C_{0–19}	b-tagging efficiency uncertainties for c-jets	resolved
FT_EFF_PFlow_Eigen_Light_{0–35}	b-tagging efficiency uncertainties for light jets	resolved
FT_EFF_PFlow_Eigen_T_{0–19}	b-tagging efficiency uncertainties for τ-jets	resolved
FT_EFF_PFlow_extrapolation_pt	b-tagging efficiency uncertainty for high-p_T jets	resolved
	Large-R jets	
FJ_CR_JET_EffectiveNP_R10_Detector{1–2}	Energy and mass scale uncertainty due to in-situ calibration	boosted
FJ_CR_JET_EffectiveNP_R10_Mixed{1–4}	Energy and mass scale uncertainty due to in-situ calibration	boosted
FJ_CR_JET_EffectiveNP_R10_Modeling{1–4}	Energy and mass scale uncertainty due to in-situ calibration	boosted
FJ_CR_JET_EffectiveNP_R10_Statistical{1–6}	Energy and mass scale uncertainty due to in-situ calibration	boosted
FJ_CR_JET_EtaIntercal_Modeling	Energy and mass scale uncertainties to cover η-intercalibration non-closure	boosted
FJ_CR_JET_EtaIntercal_NonClosure_2018data	Energy and mass scale uncertainties to cover non-closure with 2018 data	boosted
FJ_CR_JET_R10_TotalStat	Energy and mass scale uncertainties to cover non-closure	boosted
FJ_CR_JET_Flav_Comp(Flavor_Response)	Energy and mass scale uncertainty related to flavor composition (response)	boosted
FJ_CR_JET_SingleParticle_HighPt	Energy and mass scale uncertainty for the behavior of high-p_T jets	boosted
FJ_CR_JET_LargeR_TopologyUncertainty_top (V)	Energy and mass scale uncertainty on topology	boosted
FJ_CR_JET_JMS_FF_*	Mass scale uncertainty due to basic data-simulation differences	boosted
FJ_CR_JET_JMS_Rtrk_*	Mass scale uncertainty on reference tracks	boosted
FJ_CR_JET_JMS_Topology_QCD	Mass scale uncertainty due to modeling	boosted
FJ_JER_DataVsMC_R10_MC16	Energy resolution uncertainty	boosted
FJ_JER_dijet_R10_*	Energy resolution uncertainty	boosted
FJ_JER_AllOthers	Energy resolution uncertainty	boosted
FJ_JMR_COMB_*	Mass resolution uncertainty	boosted
	b-tagging: VR track jets	
FT_EFF_VR_Eigen_B_{0–44}	b-tagging efficiency uncertainties for b-jets	boosted
FT_EFF_VR_Eigen_C_{0–19}	b-tagging efficiency uncertainties for c-jets	boosted
FT_EFF_VR_Eigen_Light_{0–35}	b-tagging efficiency uncertainties for light jets	boosted
FT_EFF_VR_Eigen_T_{0–19}	b-tagging efficiency uncertainties for τ-jets	boosted
FT_EFF_VR_extrapolation_pt	b-tagging efficiency uncertainty for high-p_T jets	boosted

Table 5.5: Summary of experimental systematics related to small-R and large-R jets used in the analysis adapted from [72].

- **Shape uncertainties** describe bin by bin differences between distributions normalized to the same overall yield. Due to a lack of available MC statistics for some alternative samples, a series of Deep Neural Networks (DNN) [198] are trained to distinguish nominal and alternative MC samples for affected variations on an event by event basis. The resulting scores are then used to reweight nominal samples, which smooths out statistical fluctuations present in the corresponding alternative samples. This approach is called the "Calibrated Likelihood Ratio Estimator" (CARL) [80] and is used to compensate for low event numbers where necessary. Separate trainings are performed for each background, lepton channel, analysis regime and variation. CARL uses the same input variables as the analysis MVA discussed in section 5.1.4 and outputs a single score giving the probability of a particular event being drawn from the alternative distribution rather than the nominal.

Relevant theoretical uncertainties differ by process, but include various corrections to Monte Carlo (MC) simulations discussed in sections 2.4.1, 2.4.2 and 2.4.3. Some represent corrections applied to the nominal samples used in the analysis, which are propagated through event reweighting. Others require the use of alternate MC generators to account for more fundamental differences. The following variations are considered for most backgrounds in decreasing order of impact.

- **Matrix element, parton shower and underlying event variations** are accounted for through alternate generators and represent the largest theoretical uncertainties. Depending on the nominal generator used to generate a given sample, uncertainties from parton shower and underlying event computations are either considered alone via an alternate showering algorithm or together with matrix element variations by using a different generator altogether. These variations require separate MC simulations entirely, leading to frequent issues with available MC statistics. For this reason, CARL is generally used to model shape differences from generator variations.

- **QCD scale variations** affect the renormalization and factorization scales μ_R and μ_F, which set the value of the QCD coupling α_S and the energy scale at which QCD calculations can be done in perturbation theory respectively. These are varied in tandem by factors of 0.5 and 2 compared to their nominal values and applied as weight variations to the nominal MC generator. A total of six variations are considered where μ_R and μ_F are not allowed to deviate in opposite directions compared to the nominal values.

- **Initial and final state radiation variations** are modifications to the total amount of initial and final state radiation via dedicated parameters across a sample and are applied directly to the nominal generator. They modify the total number of radiating quarks and gluons before and after the hard-scatter interaction respectively.

- **Electroweak corrections** are applied to correct predictions to NLO. Uncertainties are estimated via different combination schemes for interactions between NLO electroweak (EW) and NLO QCD corrections to process cross sections [84], which become particularly relevant at high p_T. In the `Sherpa` generator the default scheme is additive, given by

$$\sigma^{NLO\ add.}_{QCD+EW} = \sigma^{LO} + \Delta\sigma_{QCD} + \Delta\sigma_{EW} \tag{5.2}$$

  The alternate multiplicative and exponentiated combination schemes are also considered via reweighting due to higher precision when corrections become large. These are given respectively by

$$\sigma^{NLO\ mult.}_{QCD+EW} = (\sigma^{LO} + \Delta\sigma_{QCD})(1 + \Delta\sigma_{EW}) \tag{5.3}$$

$$\sigma^{NLO\ exp.}_{QCD+EW} = (\sigma^{LO} + \Delta\sigma_{QCD})e^{\Delta\sigma_{EW}} \tag{5.4}$$

- `Sherpa` **shape corrections** are applied to samples whose nominal generator is `Sherpa` 2.2.11 given observed distortions of the Higgs mass peak and a softer V+jets p_T^V spectrum respective to `Sherpa` 2.2.1 which was used for previous measurements [68]. These differences partially stem from changes in the heavy flavor fragmentation function across the two versions of `Sherpa` and are found to not be covered by other uncertainties associated with theory parameters. Because of the possibility of MC mismodeling in the recommended version of the generator given better data agreement in the p_T^V spectrum of `Sherpa` 2.2.1 than `Sherpa` 2.2.11, additional uncertainties are introduced which correct the shapes of the p_T^V distributions in the latter to match to expectations from the former. These allow the fit to smoothly interpolate between the nominal generator version and the prior one. Despite these challenges, `Sherpa` 2.2.11 is chosen as the nominal generator due to the larger number of available MC events as well as other improvements at the generator level [84].

- **PDF+α_S variations** are combined modifications of the parton distribution functions (PDF) of a particular sample alongside the strong coupling constant α_S. PDFs describe the momentum fractions of bound state quarks and are experimentally determined, resulting in associated uncertainties in the distribution shape that need to be taken into account. They are included by varying the PDF curves using their Hessian eigenvectors as well as simultaneous modification of α_S. These uncertainties are small and their impact is thus mostly negligible.

A full list of generators and systematic variations for each sample is presented in table 5.6. To derive both overall and relative acceptance uncertainties, only the largest variation for each source of uncertainty is considered with some exceptions detailed where relevant.

More detailed descriptions of the systematics model for signal and each background category are given below.

Sample	Nominal Generator	Alternative Generator	Systematic Effects
$VH(\to c\bar{c})$	`Powheg` + `Pythia` 8	`Powheg` + `Herwig` 7	μ_R, μ_F, ISR/FSR, PDF,
			electroweak corrections
$VH(\to b\bar{b})$	`Powheg` + `Pythia` 8	`Powheg` + `Herwig` 7	μ_R, μ_F, ISR/FSR, PDF,
			electroweak corrections
Diboson	`Sherpa` 2.2.11	`Powheg` +`Pythia` 8, `Sherpa` 2.2.1	μ_R, μ_F, PDF,
			electroweak corrections
$t\bar{t}$ and	`Powheg` +`Pythia` 8	`Powheg` +`Herwig` 7,	ISR/FSR,
single-top		`MadGraph` 5_aMC@NLO+`Pythia` 8	DS/DR (single-top Wt)
V+jets	`Sherpa` 2.2.11	`MadGraph` FxFx, `Sherpa` 2.2.1	μ_R, μ_F, PDF,
			electroweak corrections

Table 5.6: Summary of MC generators and systematic variations considered for all analysis samples, adapted from [72].

VH Modeling

VH signal modeling is considered separately for the $qq \to WH$, $gg \to ZH$ and $qq \to ZH$ production processes with a consistent approach between the three analysis regimes [57]. Lowest order Feynman diagrams of these processes are shown in Fig. 5.3. The analysis parameters of interest are the overall signal normalization factors $\mu_{VHb\bar{b}}$ and $\mu_{VHc\bar{c}}$, which are freely adjustable by the fit. Uncertainties for the VH signal production are correlated across regimes while those affecting Higgs decay modes are kept separate between $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$. The dominant uncertainties come from comparing QCD scale variations using the Stewart-Tackmann (ST) method [200], which allows for taking multiple variations into account simultaneously to give a more conservative uncertainty estimate for the signal. Overall and relative acceptance uncertainties between p_T^V and N_{jet} bins are considered across the STXS regions and additional acceptance uncertainties are derived for each STXS bin with 30 PDF and 2 α_S variations. Electroweak corrections mostly impact the p_T^V spectrum of the signal and are thus kept as shape uncertainties only. Parton shower and underlying event differences are modeled with the alternate showering algorithm `Herwig` 7. Distributions are compared directly without the use of CARL to derive combined normalization and shape uncertainties. This is also done separately for ISR and FSR variations. Finally, uncertainties related to the Higgs branching ratios totaling 1.61% for $b\bar{b}$ and $^{+5.53\%}_{-1.99\%}$ for $c\bar{c}$ are considered [72]. These are dominated by uncertainties around the values of bottom and charm quark masses and the strong coupling α_S.

V+jets Modeling

The V+jets category includes both W+jets and Z+jets events which are modeled individually. Feynman diagrams showing these processes at the lowest order are presented in Fig. 5.4. V+jets modeling is also considered separately for the 6 flavor composition groupings detailed in section 5.1.1 because of the broadly similar kinematics of each category: $V+hf$ (heavy flavor), $V+mf$ (mixed flavor) and $V+lf$ (light flavor) for each W and Z. Some uncertainties are applied to the flavor components individually as detailed below. Variations in μ_F and μ_R, PDF+α_S and electroweak combination scheme alongside two alternate generators are considered for each. ME, PS and UE variations are characterized with `MadGraph` FxFx [4] while `Sherpa` 2.2.1 is used to account for p_T^V shape discrepancies as with diboson backgrounds. While the approach to modeling is overall similar between the W+jets and Z+jets backgrounds, all modeling systematics are decorrelated between the two. Uncertainties are only applied to W+jets in the zero and one lepton regions and to Z+jets in the zero and two lepton regions.

Normalization uncertainties are not applied explicitly to the V+jets backgrounds. Instead, normalization factors are floating separately for each of the six flavor composition groupings as well as p_T^V and N_{jet} bins in the resolved regime. This approach is possible given the size of each background and the constraining power of CRHigh (for $V+hf$ and $V+mf$) and V+light (for $V+lf$) CRs in the one (for W) and two (for Z) lepton channels. The only exception to this setup is the correlation of W+jets normalizations between the 3-jet and 4-jet regions in the zero lepton channel of resolved $VH(\to b\bar{b})$ since the 4-jet bin does not have a dedicated control region. In the boosted regime, the $W+hf$ and $W+mf$ as well as $Z+hf$ and $Z+mf$ normalizations are correlated and floated together. The latter is also correlated across p_T^V bins. The $V+lf$ normalization is fixed due to the smaller prevalence of these backgrounds.

Relative acceptance uncertainties are calculated between lepton channels, signal and control regions and flavor composition categories. Separate uncertainties are derived for the resolved and boosted regimes, but not within the resolved regime. In general, resolved acceptance uncertainties for $V+mf$ and $V+lf$ are calculated in the $VH(\to c\bar{c})$ regime while those for $V+hf$ are mostly calculated in resolved $VH(\to b\bar{b})$. The full list of acceptance uncertainties is described below. Approximate values for all normalization and relative acceptance uncertainties are given in table 5.7.

- Lepton channel extrapolation uncertainties are calculated for extrapolation in the one lepton channel for W+jets and the two lepton channel for Z+jets and applied to the zero lepton channel in both cases across all analysis regimes.

- Signal and control region extrapolation uncertainties are set up to account for yield differences between CRLow, CRHigh, Top(bc) and V+light control regions and signal regions in the resolved regime. The extrapolation scheme depends on the particular background, but generally Top(bc) and V+light control regions are merged with signal

regions to calculate uncertainties between SR, CRLow and CRHigh due to their similar background profiles. Uncertainties in the resolved regime are propagated as follows:

  - SR+CRLow $\to$ CRHigh, then CRLow $\to$ SR for $W + hf$
  - CRHigh $\to$ SR+CRLow, then CRLow $\to$ SR for $W + mf$
  - SR $\to$ CRHigh for $W + lf$
  - CRHigh $\to$ SR for $Z + hf$ and $Z + mf$
  - SR $\to$ CRHigh for $Z + lf$

  In the boosted regime, uncertainties are extrapolated from the Top CRs to SRs.

- Flavor composition extrapolation uncertainties account for remaining differences within the flavor categories of the $V + hf$, $V + mf$ and $V + lf$ backgrounds. They are propagated from the larger to the smaller flavor component since the larger component determines the overall normalization and consist of the following [80]:

  - Wcc/Wbb for $W + hf$
  - Zcc/Zbb for $Z + hf$
  - Wbc/Wcl, Wbl/Wcl, $Wc\tau/Wcl$ and $Wb\tau/Wcl$ for $W + mf$
  - Zbc/Zcl, Zbl/Zcl for $Z + mf$
  - $Wl\tau/Wll$ and $W\tau\tau/Wll$ for $W + lf$

  Due to the shared normalization for $V+hf$ and $V+mf$ in the boosted regime, there are additional boosted $W + hf/W + mf$ and $Z + hf/Z + mf$ extrapolation uncertainties.

- N_{jet} and p_T^V acceptance uncertainties consist of a single uncertainty for $W + hf$ that is calculated in the 3-jet and applied to the 4-jet region for the zero lepton channel in resolved $VH(\to b\bar{b})$ as well as a single p_T^V extrapolation uncertainty for the combination of $Z + hf$ and $Z + mf$ in the boosted regime.

Shape uncertainties for V+jets are derived from the variations listed above, where the maximum variation is considered for each. In the case of the electroweak corrections, the multiplicative scheme is chosen to define the uncertainties. CARL is used to quantify deviations between the nominal `Sherpa` 2.2.11 and `MadGraph` generators. For differences between `Sherpa` 2.2.11 and 2.2.1, uncertainties are propagated from the nominal p_T^V spectrum to the other kinematic observables. All other shape systematics are derived directly from the difference between default and reweighted nominal distributions.

Uncertainties	resolved $VH \to b\bar{b}, c\bar{c}$	boosted $VH \to b\bar{b}$
Z+jets		
$Z + hf$ normalization	Floating	Floating (with $Z + mf$)
$Z + mf$ normalization	Floating	Floating (with $Z + hf$)
$Z + lf$ normalization	Floating	35%
$Z + hf$ flavor composition acc.	8% to 12%	6%
$Z + mf$ flavor composition acc.	4% to 10%	6%
CRHigh/Low $\leftrightarrow$ SR acceptance	5% to 30%	-
Top CR $\leftrightarrow$ SR acceptance	-	15% to 25%
$2 \to 0$ lepton acceptance	2% to 10%	3%
p_T^V acceptance	-	15%
W+jets		
$W + hf$ normalization	Floating	Floating (with $W + mf$)
$W + mf$ normalization	Floating	Floating (with $W + hf$)
$W + lf$ normalization	Floating	38%
$W + hf$ flavor composition acc.	4% to 25%	11%
$W + mf$ flavor composition acc.	8% to 29%	9% to 15%
$W + lf$ flavor composition acc.	9%	-
CRHigh/Low $\leftrightarrow$ SR acceptance	2% to 50%	-
Top CR $\leftrightarrow$ SR acceptance	-	16% to 27%
$1 \to 0$ lepton acceptance	3% to 30%	20%
p_T^V acceptance	-	3%

Table 5.7: Normalization and acceptance uncertainties for V+jets backgrounds adapted from [72].

Top Modeling

As described in section 5.1.1, the top quark backgrounds are broadly split into "Top", which encompasses $t\bar{t}$ and Wt, and single-top, which consists of the s- and t-production channels as detailed in Figs. 5.5 and Fig. 5.6. The "Top" category is then further split based on the jet flavor composition of the Higgs candidate as "Top(bb)" and "Top(bq/qq)". The overall modeling approach for top backgrounds differs between the lepton channels and analysis regimes, especially in the two lepton channel where the background is less prevalent. In particular, the resolved $VH(\to b\bar{b})$ two lepton channel uses a data-driven estimate to predict top quark backgrounds based on events that pass the SR selection but contain leptons of different flavor. These events are isolated to create an almost completely pure top sample and an extrapolation uncertainty of 0.8% is applied to the resulting yield. In the boosted $VH(\to b\bar{b})$ regime the top background is negligible and a 20% uncertainty is applied to the base MC prediction. Finally, in the $VH(\to c\bar{c})$ regime, the combined $t\bar{t}$ and Wt background is constrained by the Top($e\mu$) CRs through 5 floating normalization factors. The s-channel background is given a cross-section uncertainty of 4.6% across all regimes and

lepton channels.

For Top(bb), Top(bq/qq) and the t-channel of single-top in the zero and one lepton channels a more sophisticated modeling strategy based on alternative samples is deployed. Two different alternate MC generators are used for each single-top t-channel and Wt-channel as well as $t\bar{t}$ processes, which are all nominally generated using the `Powheg` + `Pythia` 8 generator. One of these is `MadGraph` 5_aMC@NLO + `Pythia` 8 which models variations in matrix element calculations while the other, `Powheg` + `Herwig` 7, is used to vary the parton shower and underlying event algorithms. Variations in ISR and FSR are also separately considered for each process. Finally, for the single-top Wt samples, diagram interference effects with $t\bar{t}$ need to be accounted for since these processes were generated separately. The nominal Wt sample uses a diagram subtraction (DS) scheme where the interference is accounted for at the level of cross-section calculations while the alternate sample uses a diagram removal (DR) scheme that removes resonant diagrams at the amplitude level. A summary of the resulting uncertainties is given below.

Normalization uncertainties are not applied to Top backgrounds. Floating normalization factors for the Top(bb) and Top(bq/qq) components are decorrelated in p_T^V and N_{jet} categories across all regimes. These are constrained by the CRHigh, Top(bc) and boosted TopCR control regions. The single-top t-channel normalization is not floating, but calculated from available variations and correlated in N_{jet} and p_T^V.

Relative acceptance uncertainties are calculated separately for Top(bb), Top(bq/qq) and single-top t-channel backgrounds between lepton channels, N_{jet} and p_T^V regions, signal and control regions as well as between samples and jet flavor categories. The full list is described below and summarized in table 5.8.

- Lepton channel extrapolation uncertainties are extrapolated from the one lepton channel to the zero lepton channel for the Top(bb) and Top(bb/qq) backgrounds in the resolved and boosted regimes as well as the single-top t-channel in the resolved regime only.

- Signal and control region extrapolation uncertainties follow different extrapolation schemes for the three components. As in the V+jets background, Top(bc) and V+light control regions are merged with signal regions for the purpose of calculating the extrapolation uncertainties between SR, CRHigh and CRLow. These are applied as follows in the resolved regime:

  - CRHigh $\to$ CRLow+SR, then CRLow $\to$ SR for Top(bb)
  - CRLow+SR $\to$ CRHigh, then SR $\to$ CRLow for Top(bq/qq)
  - SR $\to$ CRLow+CRHigh, then CRHigh $\to$ CRLow for single-top t-channel in the lowest p_T^V region of the one lepton channel
  - CRHigh $\to$ SR+CRLow, then SR $\to$ CRLow for single-top t-channel in all other regions

In the boosted regime, extrapolation is done from Top CRs to SRs for Top(bb) and Top(bb/qq).

- N_{jet} acceptance uncertainties are applied for the single-top t-channel backgrounds in the resolved regime from 3-jet to 2-jet and 4-jet regions.
- p_T^V acceptance uncertainties are calculated for single-top t-channel backgrounds in the resolved regime only. Since this background is more concentrated at low p_T^V these uncertainties are propagated from low to high p_T^V regions.
- Flavor composition extrapolation uncertainties are calculated for the Top(bq/qq) background only and focus on the main Top(bc) component. These uncertainties are propagated from the bc to the bl and qq components.
- $t\bar{t}$ vs. Wt acceptance uncertainties are calculated from the more dominant $t\bar{t}$ background and applied to Wt given their combined floating normalization in all analysis regimes.

Shape uncertainties for $t\bar{t}$, single-top Wt and single-top t-channel processes follow a common modeling approach between the boosted and resolved analysis regimes, using all variations discussed above. CARL is used to derive shape uncertainties for the two alternate generator samples. Models are trained inclusively in N_{jet} and flavor composition but separately in lepton channel for each sample. For Wt, the DR variation is used as a nominal training sample given the lack of available DS samples with ME and PS variations. The DS-DR uncertainty is also separately applied as a normalization plus shape uncertainty for the single-top Wt background without the use of CARL. Finally, ISR and FSR shape uncertainties are applied as symmetrized shape only uncertainties.

VZ Modeling

Diboson backgrounds are subdominant across all analysis regions but have a very similar profile as the VH signal. They consist of the WW, WZ (with one leptonic and one hadronic V decay) and ZZ processes whose lowest order Feynman diagrams are shown in Fig. 5.7. The $VZ(\to b\bar{b})$ and $VZ(\to c\bar{c})$ components are modeled independently from each other but merged across production modes and their normalization factors $\mu_{VZb\bar{b}}$ and $\mu_{VZc\bar{c}}$ are floated freely in the diboson reference fit. All other diboson components from WW, WZ with hadronic Z decays, ZW with leptonic Z decays and ZZ to anything other than $b\bar{b}$ and $c\bar{c}$ are considered backgrounds in both the VZ and VH analysis scenarios. These are modeled separately for each production mode, but merged across decay modes. The main sources of modeling uncertainties for diboson backgrounds come from QCD scale, PDF and α_S variations as well as electroweak corrections and imapcts from two alternate generators compared to the nominal `Sherpa` 2.2.11. In particular, `Powheg` + `Pythia` 8 models ME, PS and UE differences while `Sherpa` 2.2.1 accounts for potential p_T^V mismodeling.

Uncertainties	resolved $VH \to b\bar{b}, c\bar{c}$	boosted $VH \to b\bar{b}$
Top (0- and 1-lepton)		
Top($b\bar{b}$) normalization	Floating	Floating (with others)
Top(others) normalization	Floating	Floating (with $b\bar{b}$)
Flavor composition acceptance	5% to 10%	-
$1 \to 0$ lepton acceptance	2% to 8%	6% to 17%
CRHigh/Low $\leftrightarrow$ SR acceptance	2% to 10%	-
Top CR $\leftrightarrow$ SR acceptance	-	14% to 19%
$Wt/t\bar{t}$ ratio	12% to 48%	22% to 36%
Top (2-lepton)		
Normalization	Floating ($c\bar{c}$), 0.08% ($b\bar{b}$)	20%
Single-top s-/t-channel		
Normalization	4.6% (s), 17% (t)	4.6% (s), 10% (t)
N_{jet} acceptance	15%	-
p_T^V acceptance	7% to 15%	-
$1 \to 0$ lepton acceptance	6%	-

Table 5.8: Normalization and acceptance uncertainties for Top backgrounds adapted from [72].

Normalization uncertainties are applied to all diboson backgrounds given their relatively minor prevalence and lack of dedicated control region. The sole exception to this are the $VZ(\to b\bar{b})$ and $VZ(\to c\bar{c})$ normalizations which are floating in the VZ fit only. All variations are taken into account to derive priors for normalization factors.

Acceptance uncertainties are calculated separately for WW, WZ, ZW and ZZ across lepton channels, N_{jet} and p_T^V categories as well as between signal and control regions. Extrapolations occur from the one and two lepton to the zero lepton channel, signal to control regions, low jet multiplicity to high jet multiplicity and intermediate to low and high p_T^V regions. For the signal components $VZ(\to b\bar{b})$ and $VZ(\to c\bar{c})$, QCD scale variations are excluded from N_{jet} and p_T^V acceptance uncertainties and treated separately via the ST method [200]. All uncertainties are kept identical between the resolved $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ regimes. Signal components are calculated in the former while background components are calculated in the latter due to the more favorable event statistics in both cases. The same general approach is valid for the boosted regime. A summary of normalization and acceptance uncertainties is given in table 5.9.

Shape uncertainties are calculated from the same variations described above. CARL is used to quantify shape differences between `Sherpa` 2.2.11. and `Powheg` + `Pythia` 8 while all other systematics use distribution differences directly. The latter are made up of the maximum QCD scale and PDF+α_S variations as well as the multiplicative electroweak combination scheme.

Uncertainties	resolved $VH \to b\bar{b}, c\bar{c}$	boosted $VH \to b\bar{b}$
Diboson		
$WW/ZZ/WZ$ normalization	16% / 17% / 19%	16% / 17% / 27%
$gg \to VV$ normalization	30%	30%
Lepton channel acceptance	2 to 23%	7 to 10%
N_{jet} acceptance	10% to 42%	-
p_T^V acceptance	3% to 16%	7% to 40%
SR $\to$ CR acceptance	6% to 50%	-
STXS acceptance	-37.4% to 32.0%	-22% to 10.6%

Table 5.9: Normalization and acceptance uncertainties for VZ backgrounds adapted from [72].

Multi-jet Modeling

The smallest analysis background under consideration is QCD multi-jet production that passes the analysis selection. This includes events with leptons from bottom or charm hadron decays as well as fake leptons in conjunction with jets. Multi-jet backgrounds are very common overall but heavily suppressed due to the lepton selection. They are also inherently difficult to model using MC simulations due to the challenges of soft QCD computations and the combinatorical complexities of modeling large numbers of jets. For these reasons, multi-jet events are instead estimated using data-driven techniques. Multi-jet backgrounds are only considered in the one lepton channels of the resolved analysis regime as they were found to be negligible across the rest of the phase space. They are most prominent in low p_T^V regions.

Multi-jet background templates are extracted using a template fit with dedicated "mjCR" (multi-jet control regions). These are enriched control regions with an analogous selection to analysis SRs except for an inversion on the tight lepton isolation requirement and a requirement of only a single b-tagged jet in the $VH(\to b\bar{b})$ regime. The fit is performed separately for each tag and N_{jet} region as well as for electrons and muons and low p_T^V (< 150 GeV) and medium to high p_T^V (> 150 GeV). The reconstructed W boson mass m_T^W is used as a fit variable to discriminate multi-jet events from other backgrounds and floating normalization factors are applied separately for top, multi-jet and W+jets. The multi-jet template normalization is then extracted for each region by subtracting the backgrounds in SRs from data in mjCRs.

Systematics associated with the multi-jet background consist of a single combined normalization and two shape uncertainties in the one lepton channel. The latter come from the extrapolation between mjCRs and SRs and the incorporation of a separate background normalization factor derived from the fit before performing data subtraction. Both of these systematics have normalization components, which are combined with uncertainties from three other sources. The first also considers $m_T^W < 20$ GeV events for the low p_T^V fits only.

The second adds $E_T^{miss} < 30$ GeV events to the medium to high p_T^V electron fit. Finally, a series of fits is performed with an alternative variable $\Delta\phi(\ell, E_T^{miss})$ that also separates multi-jet from other backgrounds. This results in a combined normalization uncertainty of $20\% - 100\%$ for the multi-jet background.

5.3 Fit Description

The process of measuring the signal strengths $\mu_{VHb\bar{b}}$ and $\mu_{VHc\bar{c}}$ relies on a statistical procedure called maximum likelihood estimation (MLE) [127]. A likelihood function is maximized with respect to the "parameters of interest" (POIs) μ as well as sets of "nuisance parameters" (NPs) $\vec{\theta}$, $\vec{\tau}$ and $\vec{\gamma}$ which represent different kinds of uncertainties in the fit. In particular, θ parameters represent systematic uncertainties with known priors and τ parameters represent floating background normalization factors with no priors. The latter are constrained through background enriched CRs in the fit as described in section 5.2.2. Finally, γ parameters represent statistical uncertainties related to MC event counts respective to each bin. The full likelihood function is given by

$$\mathcal{L}(\vec{\mu}, \vec{\theta}, \vec{\gamma}, \vec{\tau}) = \prod_{i \in \text{bins}} \text{Pois}(N_i | \vec{\mu} s_i(\vec{\theta}) + b_i(\vec{\theta}, \vec{\gamma}, \vec{\tau})) \times \prod_{\theta \in \vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2} \times \prod_{i \in \text{bins}} \text{Gauss}(\gamma_i | 1, \tau_i) \quad (5.5)$$

which consists of a Poisson term for each bin and Gaussian terms for each θ and γ parameter. Maximizing the likelihood entails maximizing agreement of data and MC by tuning parameters within constraints from priors enforced via the Gaussian terms. In practice this is done by minimizing the negative logarithm of the likelihood function (i.e. the "negative log likelihood"), which is computationally more straightforward and implemented via the `MINUIT` [159] algorithm in `RooFit` [216]. This process generates "best fit" values for all POIs and NPs. NPs that deviate from their initial values in the fit are considered "pulled" while those that show decreased variance after fitting are "constrained". NP pulls are defined as the difference between the "best fit" and initial values divided by their initial uncertainty $\frac{\hat{\theta} - \theta_0}{\sigma_{\theta_0}}$, where $\hat{\theta}$ represents the best fit value of a particular NP. Post-fit parameter variances and correlations are calculated with the covariance matrix

$$V_{ij} = \left[\frac{\partial^2(-\ln\mathcal{L})}{\partial\theta_i \partial\theta_j} \right]^{-1} \quad (5.6)$$

evaluated at the "best fit" values. This defines the curvature of the likelihood function at its minimum and thereby the impact of each parameter variation on the likelihood value.

To test a particular $\vec{\mu}$ hypothesis, such as SM agreement with $\mu_{VHb\bar{b}} = \mu_{VHc\bar{c}} = 1$, the likelihood function is minimized both globally and with fixed values of $\vec{\mu}$ corresponding to the hypothesis to be tested. The likelihood ratio

$$\lambda(\vec{\mu}) = \frac{\mathcal{L}(\vec{\mu}_H, \hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\gamma}}}, \hat{\hat{\vec{\tau}}})}{\mathcal{L}(\hat{\vec{\mu}}, \hat{\vec{\theta}}, \hat{\vec{\gamma}}, \hat{\vec{\tau}})} \tag{5.7}$$

is used for a hypothesis test, where $\left\{\hat{\vec{\mu}}, \hat{\vec{\theta}}, \hat{\vec{\gamma}}, \hat{\vec{\tau}}\right\}$ are parameter values that maximize the likelihood in general and $\left\{\hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\gamma}}}, \hat{\hat{\vec{\tau}}}\right\}$ are values that maximize the likelihood assuming $\vec{\mu_H}$. The relevant test statistic is

$$q_H = \begin{cases} -2\ln\lambda(\vec{\mu_H}) & \hat{\mu} \geq \mu_H \\ 0 & \hat{\mu} < \mu_H \end{cases} \tag{5.8}$$

which is related to the p-value for a hypothesis H as

$$p_H = \int_{q_{H,\text{obs}}}^{\infty} f(q_H|H) dq_H \tag{5.9}$$

where $q_{H,\text{obs}}$ is the test statistic observed in data assuming H and $f(q_H|H)$ is the probability density function of q_H, which is well approximated by a χ^2 distribution in the large sample limit. The number of Gaussian standard deviations corresponding to a given p_H value under a hypothesis H is $Z = \Phi^{-1}(1 - p_H)$. This technique is used to calculate the significance of a measurement of the $VH \to b\bar{b}$ process using the background only p-value derived from the test statistic q_0 to reject the hypothesis $\vec{\mu} = 0$. For $VH \to c\bar{c}$ the sensitivity is small enough that a 95% upper confidence level (CL) is also calculated. This is done via a modified frequentist CL_S method [187] that uses a test statistic

$$q_\mu^{\text{CL}_\text{S}} = -2\ln\frac{\mathcal{L}(\vec{\mu} = x, \hat{\hat{\vec{\theta}}}(x), \hat{\hat{\vec{\gamma}}}(x), \hat{\hat{\vec{\tau}}}(x))}{\mathcal{L}(\hat{\vec{\mu}}, \hat{\vec{\theta}}, \hat{\vec{\gamma}}, \hat{\vec{\tau}})} \tag{5.10}$$

where x is tuned such that the resulting p-value is 0.05.

5.3.1 Fit Overview

The analysis fit model consists of a total of 21 signal and 39 control regions in resolved $VH(\to b\bar{b})$, 32 signal and 54 control regions in $VH(\to c\bar{c})$ and 6 signal and 4 control regions in boosted $VH(\to b\bar{b})$ respectively, as shown in Fig. 5.9. Control regions are detailed in section 5.1.4 and consist of CRHigh, CRLow and V+light control regions to constrain the V+jets backgrounds as well as Top(bc), Top($e\mu$) and boosted TopCR control regions for the top backgrounds. All regions are binned in variables which minimize fit complexity and the impact of bin-by-bin statistical uncertainties while maximizing the control of systematic uncertainties described in section 5.2.2. The former suggests maximizing the number of events in each bin, while the latter requires fine enough bins to show the underlying distribution

shapes and constrain shape systematics. This balance is optimized for each region based on overall impact of binning on statistical and systematic uncertainties. Some control regions are kept as single-bin and therefore do not contain any background shape information. Regions binned in p_T^V are used to understand mismodeling of the p_T^V shape in `Sherpa` 2.2.11.

- Signal regions are binned in the MVA_{VH} and MVA_{VZ} discriminants for the VH and VZ measurements respectively.
- CRLow regions are binned in the dedicated $\text{MVA}_{\text{CRLow}}$ discriminant to isolate W+jets events.
- CRHigh regions are single-bin in the zero lepton channel and binned in p_T^V for BB-tag regions and the dijet mass $m_{j_1 j_2}$ for $C_T C$-tag regions in the one and two lepton channels. Use of the MVA discriminant for binning CRHigh regions was also investigated, but not pursued due to potential mismodeling issues from p_T^V shape differences between data and MC.
- Top(bc) and boosted TopCR regions are binned in the di-jet mass $m_{j_1 j_2}$ and large-R jet mass m_J respectively.
- Top($e\mu$) regions are single-bin.
- V+light regions are single-bin in the zero lepton channel and binned in p_T^V in the one and two lepton channels.

The main VH fit POIs are the signal strengths $\mu_{VHb\bar{b}}$ and $\mu_{VHc\bar{c}}$ describing the ratio between observation and Standard Model expectations for the rates of the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ processes across the analysis phase space. In the diboson reference fit, these are replaced by the analogous signal strength parameters $\mu_{VZb\bar{b}}$ and $\mu_{VZc\bar{c}}$. Aside from the principal two POI fits, additional fits with decorrelated signal strength parameters across different categories are considered, which is detailed in section 5.4.

Systematic uncertainties added as fit NPs are pre-processed to ensure fit stability. Shape uncertainties are "smoothed" by combining multiple adjacent bins until the statistical uncertainty decreases below 5%. This prevents the fit from overfitting distribution shapes based on statistical fluctuations by applying shape uncertainties from the combined bins to each individual component bin. NPs are also "pruned" (i.e. removed) if their impact is less than 0.5% on the yield of each bin in all affected background templates to reduce the total number of redundant systematics. Systematics for small backgrounds making up less than 2% of the signal yield in sensitive regions (with signal $> 2\%$ of total yield) or 0.5% of the background yield in non-sensitive regions (with signal $< 2\%$ of total yield) are also removed entirely. After pruning around 2000 NPs remain, of which around 1400 are γ and the rest are mostly θ parameters.

5.3.2 Fit Development and Diagnostics

The starting point for the design of a coherent fit model for the full run 2 $VH(\to b\bar{b}/c\bar{c})$ measurement lies in the fit models of prior measurements in each individual analysis regime [68, 75, 81]. Given the complexity of the combined fit, significant amounts of additional development were required to ensure proper implementation of background modeling and systematics propagation. This development began on individual standalone $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ fits, before moving to studies on the combined fit. A near final version of the fit model was used to initially measure $\mu_{VZb\bar{b}}$ and $\mu_{VZc\bar{c}}$ in data with only minor tuning thereafter. Three types of fits were performed during development:

- **Asimov fits** are performed using an "Asimov" dataset, which is constructed directly from underlying MC truth in a way that "Asimov data" and MC distributions match exactly. In this fit, POIs and NPs remain at their default values after likelihood maximization and show no pulls. The purpose of Asimov fits lies in deriving baseline estimates for POI limits and significances as well as NP variances and correlations to gauge the sensitivity of an analysis based strictly on expectations from MC. These estimates cannot take MC mismodeling into account, which makes them more useful in the initial stages of fit model development.
- **Conditional data fits** use the complete observed dataset. POIs are fixed to the values taken from MC, which corresponds to Standard Model expectations leading to default values of 1 for all POIs in this analysis. This allows for fit model examination (including NP pulls) without revealing "best fit" POI values in a way that could bias development to match a particular hypothesis about the observed data.
- **Unconditional data fits** are identical to conditional data fits, but with unconstrained POIs. These fits are used in the final stages of fit development with obscured POI results to ensure consistent NP pulls and correlations in both cases. Given the small signal contributions across the $VH(\to b\bar{b}/c\bar{c})$ phase space, unconditional fits with masked POIs pose minimal risk for measurement biases since the relationships between shifts in NPs and changes in the POIs are not easily visible. Once fit model development is complete, unconditional data fits are used to extract results for the VZ and VH fits.

During the development process, data in signal rich bins are "blinded" in non-Asimov fits which obscures the total number of data events in affected bins to avoid biasing fit model development to match a particular hypothesis. For $m_{j_1j_2}$ or m_J binned regions this includes the Higgs mass window between $80-140$ GeV. In SRs using the MVA discriminant, bins are blinded starting from the most signal-like part of the distribution until bins containing 60% of total MC VH signal events in the region are covered. In regions using other variables (i.e. single-bin and p_T^V), all bins where the MC signal to background ratio is more than 10% are blinded.

The sensitivity of a fit model is determined by its resulting "expected" significance or limit for a given POI. This is calculated using an Asimov dataset with NPs set to either

nominal values (i.e. "pre-fit expected") or values obtained from a conditional data fit (i.e. "post-fit expected"). The "observed" significance or limit of a measurement is then obtained from the unconditional data fit.

During fit model tuning, the main factors under consideration are fit simplicity and flexibility, expected POI sensitivities and NP pulls, constraints and correlations. In general, a good fit model provides a coherent description of systematic uncertainties without introducing excess degrees of freedom for the likelihood minimizer to exploit. This means that common systematic uncertainties affecting different regions or regimes are fully correlated as much as possible. If too many degrees of freedom are removed via linking systematics, the minimizer can compensate by pulling other NPs, resulting in an unstable fit model. Unexpectedly large NP pulls and constraints or spurious correlations are indicators of such issues and require investigation. When testing alternate fit configurations, visualizing the largest NP differences between the two configurations at a glance as shown in Fig. 5.13 can help diagnose potential issues. Optimization of the fit model also involves maximizing the expected measurement sensitivity by finetuning signal and control regions in terms of both their fundamental definitions as well as details such as binning. A selection of fit studies performed during the course of model development are presented below. An emphasis is placed on the development of the $VH(\to c\bar{c})$ fit given its higher complexity compared to $VH(\to b\bar{b})$ as well as the larger number of modifications made in this regime compared to previous results.

Conditional $VH(\to c\bar{c})$ fit studies

The fit model for $VH(\to c\bar{c})$ was first optimized independently for the $\mu_{VHc\bar{c}}$ measurement without consideration of the $VH(\to b\bar{b})$ fit, which shares a lot of common systematics. Because of the large amount of common NPs to be correlated and resulting lower constraining power of uncertainties, measurement sensitivity in this standalone fit was expected to be relatively lower than in the final combined fit. Nevertheless, a variety of decisions regarding the $VH(\to c\bar{c})$ fit model could be made independently. A selection of decisions made at this stage is presented below.

- "Low ΔR Control Regions" were investigated for the $VH(\to c\bar{c})$ fit as single-bin regions to constrain the normalization of the V+jets backgrounds. Their presence was found to have no impact on background modeling systematics or fit sensitivity. This is likely due to their similar composition as the signal regions, which are heavily background dominant themselves. Events falling into these CRs were instead merged into corresponding SRs rather than excluded. This had no visible negative impact on the systematics description but was done to increase data and MC statistics per bin.

- Different binning options for C_TC-tag "High ΔR Control Regions" were considered for fit simplification. In particular, tests were performed with decorrelated C_TC_T- and C_TC_L-tag regions using either m_{j1j2}, single-bin configurations or applying the VH MVA. It was found that the m_{j1j2} approach provided the fit with relevant information

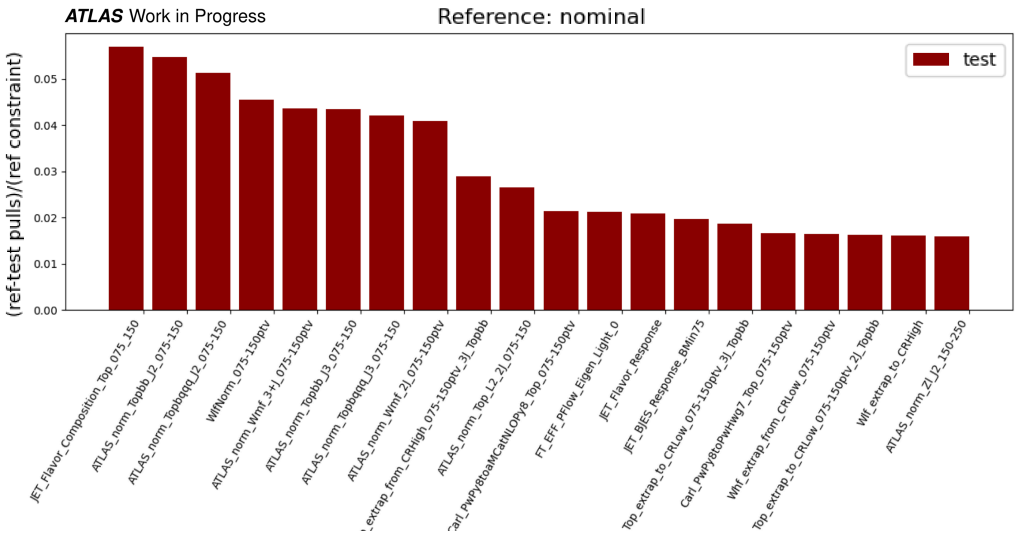
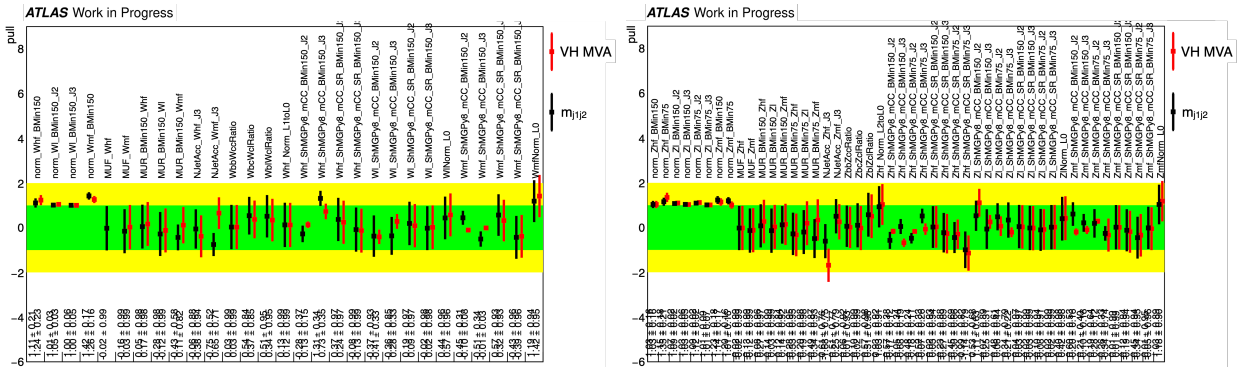


Figure 5.13: Example visualization of NP pull differences between "nominal" and "test" fit model. Systematics are ranked by $|\hat{\theta}_{\text{test}} - \hat{\theta}_{\text{nominal}}|/\sigma_{\hat{\theta}_{\text{nominal}}}$, highlighting the largest pull differences between the two fit models.

on V+jets distribution shapes which allowed it to constrain corresponding shape NPs in the one and two lepton channels. This information was lost in the single-bin approach. The VH MVA strategy lead to even larger constraints on V+jet background shape uncertainties, which were unjustifiably small as shown in Fig. 5.14. Because of the risk of MVA shape mismodeling this strategy was abandoned. The final fit configuration uses m_{j1j2} as a variable in the one and two lepton channels while keeping the zero lepton channel regions single-bin.

- Originally, C_TC_T- and C_TC_L-tag regions were kept separate in the fit due to their different compositions, with C_TC_T containing more $V + hf$ and C_TC_L more $V + mf$ and $V + lf$ backgrounds. Decorrelating the signal strength $\mu_{VHc\bar{c}}$ between the two revealed relatively similar expected limits for both, with slightly higher sensitivity in



(a) (b)

Figure 5.14: Figures showing pulls of systematic uncertainties related to W+jets (a) and Z+jets (b) backgrounds in $VH(\to c\bar{c})$ test fit comparing m_{j1j2} with VH MVA binning. Uncharacteristically small constraints on shape uncertainties indicate possible modeling issues.

C_TC_T where the post-fit expected limit is $\sim 17\%$ lower than in C_TC_L. Various merging strategies for the two tag regions were examined, including asymmetric merging in SRs or CRs only. No significant issues with background modeling were observed in any configuration, suggesting adequate fit flexibility to constrain $V+hf$, $V+mf$ and $V+lf$ separately in any case. Because of this and the relatively minor differences in SR sensitivity between C_TC_T and C_TC_L the two regions are combined as C_TC in the final fit to simplify the model.

- In the initial $VH(\to c\bar{c})$ fit configuration inherited from the previous standalone analysis [81], only two distinct sets of p_T^V regions were considered: $75\text{GeV} \leq p_T^V \leq 150\text{GeV}$ and $p_T^V > 150$ GeV. A split of the latter into $150\text{GeV} < p_T^V \leq 250\text{GeV}$ and $p_T^V > 250$ GeV consistent with resolved $VH(\to b\bar{b})$ boundaries was tested. While high p_T^V regions have lower event numbers overall, they also contain a greater proportion of signal events, leading to similar expected sensitivities for decorrelated POIs in $150\text{GeV} < p_T^V \leq 250\text{GeV}$ and $p_T^V > 250$ GeV and a much lower sensitivity in $75\text{GeV} \leq p_T^V \leq 150\text{GeV}$. This split is found to increase the sensitivity of the fit evidenced by a $\sim 4\%$ decrease in the post-fit expected upper limit for the combined POI, which justified the change.

Conditional $VH(\to b\bar{b}/c\bar{c})$ fit studies

After initial development on the standalone $VH(\to c\bar{c})$ fit, tuning of correlation schemes between systematics across regimes required consideration of a simultaneous $VH(\to b\bar{b}/c\bar{c})$

fit model. This presents a significant increase in complexity compared to either standalone fit, making issues harder to diagnose. The broad fit structure was mostly fixed via standalone fits in each regime, bringing the focus of combined fit studies primarily towards studying the impact of NP correlations. A subset of results is discussed below.

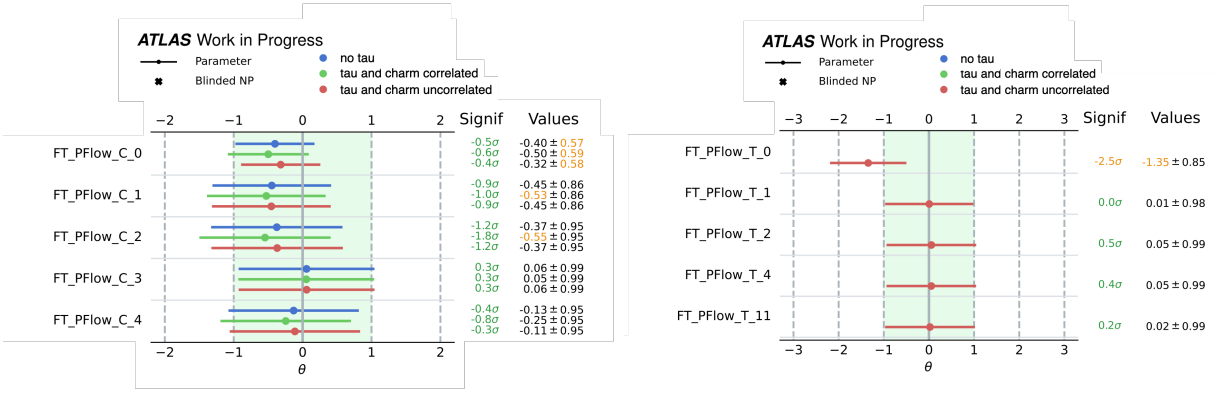
- $V + hf$ backgrounds have different compositions in the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ regimes, consisting mostly of $V + b\bar{b}$ in the former and $V + c\bar{c}$ in the latter. Given the similar underlying physics of these two processes, tests were conducted to see if $V+b\bar{b}$ and $V+c\bar{c}$ normalizations could be combined into $V+hf$ factors with additional acceptance uncertainties. These correlations did not lead to shifts in other NPs, but decreased the post-fit expected upper limit for $\mu_{VHc\bar{c}}$ by more than 5% in the two lepton channel due to constraints from $VH(\to b\bar{b})$ regions, leading to their adoption.

- Experimental systematic uncertainties are implemented according to prescriptions described in section 4.1. Due to low presence of hadronic τ decays in the analysis phase space, the corresponding flavor tagging uncertainties were correlated between τ- and charm-jets. The largest of these uncertainties showed a significant pull in the combined fit. As a relatively large combined normalization and shape systematic, this presented a concern due to the importance of c-jet identification in the analysis. Decorrelating the τ and charm components allowed for the isolation of this pull to the proportionally smaller τ-component while restoring the charm component further towards its nominal value, as shown in Fig. 5.15. Floating normalization factors for the $W + hf$ and Top(bb) backgrounds decreased to compensate for the reduced normalization effect from the charm uncertainty, leading to a more coherent uncertainty description.

- Top($e\mu$) control regions were first introduced in the combined fit to constrain top backgrounds in the two lepton channel. A 40% normalization uncertainty for the top quark background was replaced with separate floating normalization uncertainties across each N_{jet} and p_T^V region, except for the highest p_T^V region which was correlated in N_{jet} due to large uncertainties on individual normalizations. This addition did not impact other parts of the fit model, but brought reasonable constraints to top normalization factors which were found to be within 10% of unity (with standard deviations of up to 20%) for all but the highest p_T^V bins.

- Given the broadly similar calibration procedures between small-R and large-R jets described in section 4.1, a correlation scheme between associated jet NPs was taken into consideration. This scheme considered correlating all systematic uncertainties originating from the same type of calibration procedure between the two jet categories. However, due to the inexact nature of correspondences between systematics and the risk of skewing mostly tiny large-R jet uncertainties based on their small-R counterparts, this strategy was not adopted for the final model.

- Extrapolation uncertainties between signal regions and CRHigh or CRLow for the W+jets components $W+hf$, $W+mf$ and $W+lf$ dominant in the one lepton channel



(a) (b)

Figure 5.15: Figures showing pulls of flavor tagging related systematic uncertainties for five leading charm (a) and tau (b) NPs in a $VH(\to b\bar{b}/c\bar{c})$ test fit comparing three configurations excluding tau components, correlating them with charm components or including both but keeping them separate.

were initially considered independently for N_{jet} and p_T^V regions. Similar behavior of these systematics across regions motivated a correlation scheme that merged all aforementioned uncertainties for the $W+lf$ background and merged those for $W+hf$ and $W+mf$ across N_{jet} regions. The simplified fit model included 10 individual systematic uncertainties for this category compared to the initial 24. Related correlation schemes were adopted for other control region extrapolation uncertainties such as between signal and Top($e\mu$) control regions.

Conditional $VZ(\to b\bar{b}/c\bar{c})$ fit studies

After many iterations of conditional VH fits, the resulting fit model was tested on conditional VZ fits before the first unconditional data fit, or "unblinding" of the diboson results. Since the VZ fit serves as a reference measurement, this process is used to validate the fit model before "unblinding" the VH signal. The fit model is kept entirely consistent between the two measurements, except for the definition of analysis POIs and the signal target in MVA trainings. Given these fundamental similarities, the VH and VZ conditional fits are expected to produce similar results across the board, especially regarding NP modeling.

Initial translation of the VH fit model into the VZ realm produced notable discrepancies in floating normalization systematics in the zero and one lepton channels between the two fits. The largest differences were observed for the W+jets backgrounds in the one lepton channel. Background shape uncertainties and some experimental systematics related to

flavor tagging were also affected. Overall background yields were observed to be shifted towards heavy and mixed flavor contributions for W+jets by up to 35% in VH compared to VZ, indicating a fit model with excess flexibility or lacking constraints from the one lepton CRLow regions. Correspondingly, other background yields (most frequently light flavor contributions for W+jets) were consistently lower by up to 20% in the former fit model. The origin of the differences was identified to lie in the differing MVA distributions across analysis signal regions between VH and VZ. In particular, the W+jets background is shifted towards the center of the distribution in VZ SRs, causing some unexpected shifts in NPs. Constraints from CRLow regions were not powerful enough to account for these discrepancies in the fit model under observation. These differences were addressed with the following studies and changes to the fit model:

- While the underlying MVA trainings are fundamentally different between the VH and VZ signals, efforts were made to harmonize the binning strategy between the two by applying the same transformation algorithm to both and using the same number of bins. This reduced observed background yield and NP discrepancies by a few percent but did not resolve them entirely.

- An attempt was made to improve the constraining power of CRLow regions on the W+jets backgrounds by increasing the total number of bins in these regions, which had little impact on the fit overall and was thus not adopted. An opposite study reducing CRLow to single-bin regions worsened observed fit differences between VH and VZ across all regions with yield discrepancies of up to 50%. This indicates that shape information from CRLow plays an important role in constraining the W+jets background as expected and that the nominal fit configuration utilizes this information effectively.

- To examine the hypothesis of excess degrees of freedom in the fit model, a test fit with all normalization effects removed from combined shape and normalization NPs covering backgrounds with associated floating normalizations was done. This test was not intended as an alternate fit configuration due to the possible exclusion of acceptance effects between regions not covered by floating normalizations. Results showed a significant reduction in fit differences between VH and VZ fits, with yields remaining within 5% of each other in both fits, suggesting that the source of these differences is an underconstrained fit model.

To properly address issues regarding fit complexity, the final fit model was simplified relative to its previous iteration via correlation of various minor W+jets related shape and extrapolation uncertainties across regions. This reduction in complexity removed most discrepancies between VH and VZ fit results and provided a consistent basis to derive final analysis measurement results in both regimes.

5.4 Results

Final fit results from the run 2 "legacy" $VH(\to b\bar{b}/c\bar{c})$ analysis are presented below. Signal strengths for both VH and VZ processes are given and the final fit model is discussed. The only differences between the two fits are related to the MVA used to separate signal and background as well as the choice of POIs themselves. Experimental agreement with Standard Model physics would be given by a result of $\mu_{\mathrm{VHb\bar{b}}} = \mu_{\mathrm{VHc\bar{c}}} = \mu_{\mathrm{VZb\bar{b}}} = \mu_{\mathrm{VZc\bar{c}}} = 1$. Results are also shown for fits with additional decorrelated POIs using the same underlying fit model. Fits with 2+2 POIs splitting WH and ZH (or WZ and ZZ) as well as 3+3 POIs splitting each lepton channel are performed for both VH and VZ analyses. For VH, the STXS scheme (Stage 1.2) [16] is used to further decorrelate the $VH(\to b\bar{b})$ POI based on production channel, p_T^V and N_{jet}. A 10+1 POI fit separates the corresponding signal strength for each production channel and p_T^V region, while a 13+1 POI fit further separates all resolved $VH(\to b\bar{b})$ ZH POIs by $N_{\mathrm{jet}} = 0$ and $N_{\mathrm{jet}} \geq 1$. In addition to signal strengths, statistical significances or 95% confidence level upper limits are calculated for all POIs depending on measurement sensitivity. In general, limits are given for $VH(\to c\bar{c})$ due to the much lower expected analysis sensitivity to this process. All results quote "observed" and "post-fit expected" values for each POI and the impact of systematic uncertainties on signal strength is quantified via NP "rankings". An interpretation of measurement results is also presented using the κ-framework discussed in section 2.3.3.

5.4.1 $VZ(\to b\bar{b}/c\bar{c})$ Measurement Results

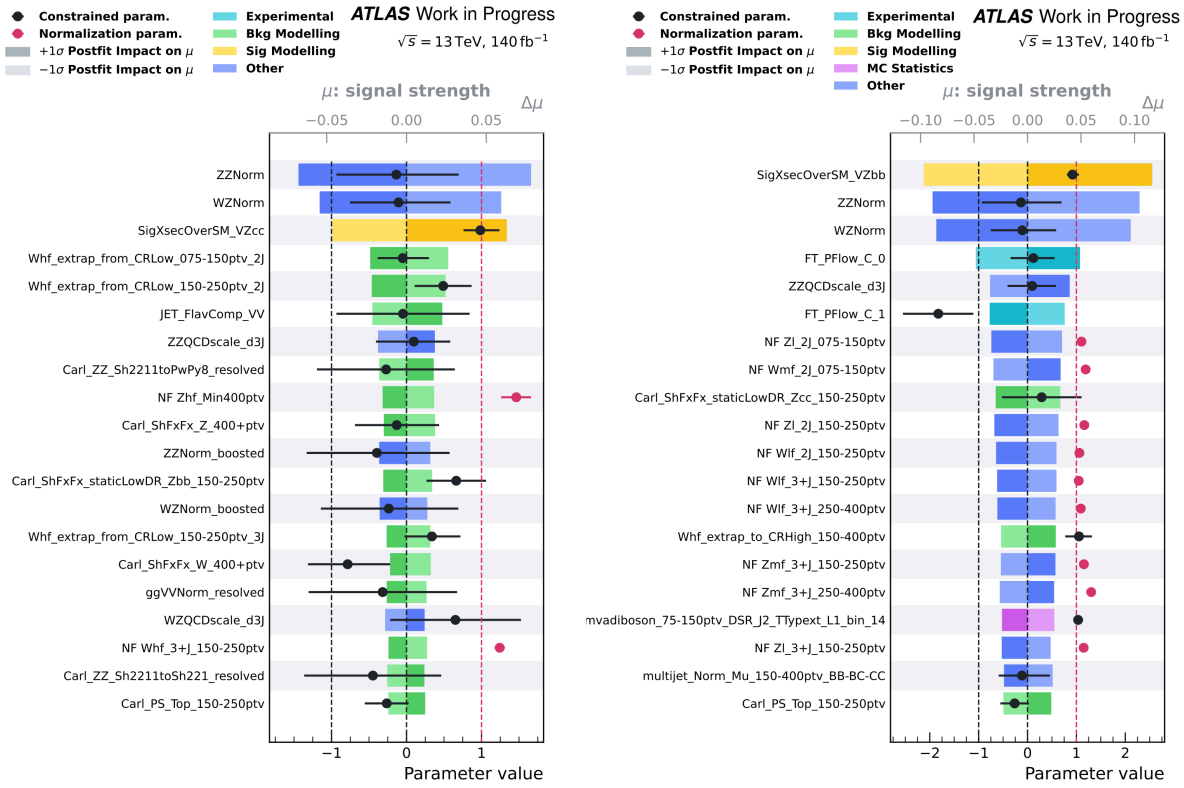
For the main combined $VZ(\to b\bar{b}/c\bar{c})$ reference measurement, the expected and observed significances using the final fit configuration are $> 10\sigma$ for $VZ(\to b\bar{b})$ and 5.3σ and 5.2σ respectively for $VZ(\to c\bar{c})$. This is beyond the nominal 5σ threshold and thus constitutes a discovery of the $VZ(\to c\bar{c})$ process. The corresponding signal strengths are

$$\mu_{\mathrm{VZb\bar{b}}} = 0.91^{+0.13}_{-0.11} = 0.91^{+0.05}_{-0.05}(\mathrm{stat.})^{+0.12}_{-0.10}(\mathrm{syst.}) \tag{5.11}$$

$$\mu_{\mathrm{VZc\bar{c}}} = 0.97^{+0.25}_{-0.22} = 0.97^{+0.13}_{-0.13}(\mathrm{stat.})^{+0.22}_{-0.18}(\mathrm{syst.}) \tag{5.12}$$

which are consistent with Standard Model expectations within uncertainty. These one dimensional intervals represent two separate fits where a single parameter is treated as the POI in each. Uncertainties are split by contribution as statistical, which includes floating normalizations (i.e. τ), data and MC (i.e. γ) statistical uncertainties and systematic, which includes all shape and normalization systematics with priors (i.e. θ).

To understand the direct impact of individual NP variations on POIs and thus generate a NP "ranking", two fits are performed per NP. In each of these, a single NP is fixed to either the 1σ upward or downward variation of its "best fit" value and the resulting POI value allows us to quantify the impact $\Delta\mu$ of the NP in question on the POI. This is shown for the VZ fits in Fig. 5.16. In addition to individual rankings, the collective impact of



(a) (b)

Figure 5.16: Figures showing ranking of systematic uncertainties based on $\Delta\mu$ impact on measured signal strengths for the final $VZ(\to b\bar{b})$ (a) and $VZ(\to c\bar{c})$ (b) fit alongside pulls and constraints for selected NPs from [72].

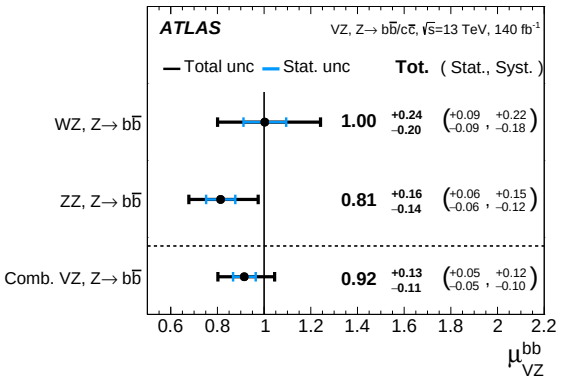
systematic categories is evaluated. This is achieved by performing fits where all parameters within a given category are fixed to their best-fit values. The resulting uncertainty on the POI is then compared to the total uncertainty from the nominal fit and subtracted in quadrature to isolate contributions from the NP group in question. This procedure yields the NP "breakdown" presented in table 5.10.

Dominant uncertainties in the measurement come from signal and background modeling, particularly related to diboson and V+jets processes. The largest overall contributions are WW, WZ and signal normalizations. Data statistical uncertainties also play a large role, albeit to a lesser extent than systematics. Experimental uncertainties are subdominant compared to both modeling and data statistics but are dominated by jet related uncertainties. Systematics from jet flavor identification are also non-negligible in the $VZ(\to c\bar{c})$ regime.

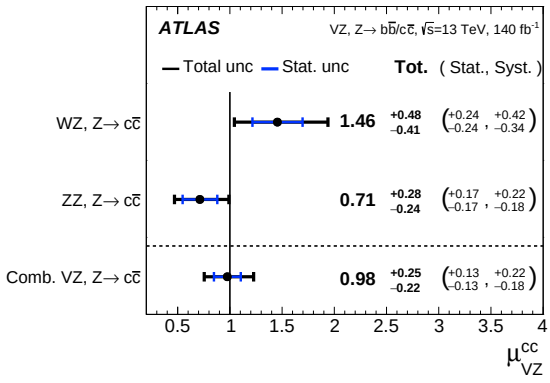
Additional fits with decorrelated POIs give similar results to the main 1+1 POI fit. Signal strengths for a 2+2 POI fit decorrelated in WW and WZ are shown in Fig. 5.17 while those for a 3+3 POI fit decorrelated in lepton channel are shown in Fig. 5.18. All values are within

Nuisance parameters	Impact on $\mu_{VZb\bar{b}}$ error		Impact on $\mu_{VZc\bar{c}}$ error	
Total	+0.131 / -0.114	±0.123	+0.255 / -0.224	±0.239
Data statistical	+0.049 / -0.049	±0.049	+0.129 / -0.128	±0.129
Data statistical only	+0.041 / -0.041	±0.041	+0.120 / -0.119	±0.119
Floating normalizations	+0.044 / -0.034	±0.039	+0.085 / -0.066	±0.076
$VZ(\to c\bar{c}/b\bar{b})$ normalization	+0.061 / -0.048	±0.055	+0.128 / -0.086	±0.107
Top($e\mu$) CR statistical	+0.006 / -0.006	±0.006	+0.002 / -0.004	±0.003
Systematic	+0.122 / -0.103	±0.113	+0.220 / -0.183	±0.202
MC statistical	+0.023 / -0.019	±0.021	+0.076 / -0.068	±0.072
VH modeling	+0.001 / -0.001	±0.001	+0.002 / -0.001	±0.001
Background modeling	+0.118 / -0.099	±0.108	+0.199 / -0.160	±0.179
multi-jet	+0.013 / -0.012	±0.012	+0.029 / -0.024	±0.027
single-top s-, t-channel	+0.009 / -0.008	±0.009	+0.007 / -0.006	±0.006
$t\bar{t}$ and single-top Wt	+0.016 / -0.014	±0.015	+0.032 / -0.027	±0.030
W+jets	+0.044 / -0.039	±0.042	+0.052 / -0.045	±0.048
Z+jets	+0.045 / -0.036	±0.041	+0.069 / -0.059	±0.064
Diboson	+0.109 / -0.089	±0.099	+0.164 / -0.116	±0.140
Experimental	+0.037 / -0.030	±0.034	+0.091 / -0.075	±0.083
Leptons	+0.002 / -0.002	±0.002	+0.003 / -0.002	±0.003
E_T^{miss}	+0.013 / -0.011	±0.012	+0.008 / -0.009	±0.009
Jets	+0.029 / -0.024	±0.026	+0.048 / -0.043	±0.045
Flavor tagging (b-jets)	+0.010 / -0.009	±0.009	+0.009 / -0.008	±0.009
Flavor tagging (c-jets)	+0.009 / -0.005	±0.007	+0.071 / -0.056	±0.064
Flavor tagging (l-jets)	+0.009 / -0.005	±0.007	+0.039 / -0.026	±0.033
Flavor tagging (τ-jets)	+0.002 / -0.001	±0.002	+0.004 / -0.004	±0.004
Pile-up	+0.002 / -0.002	±0.002	+0.009 / -0.005	±0.007
Luminosity	+0.006 / -0.005	±0.005	+0.009 / -0.007	±0.008

Table 5.10: Detailed breakdown of NP contributions to errors on VZ signal strengths adapted from [72].

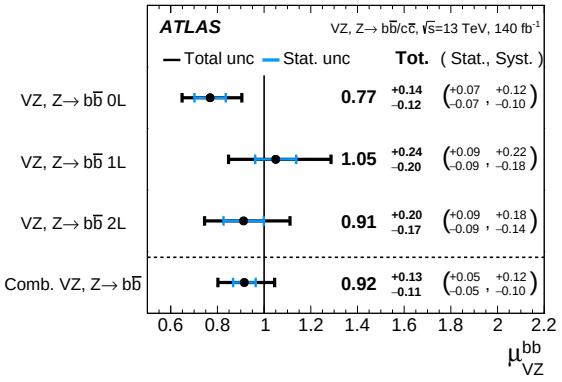


(a)

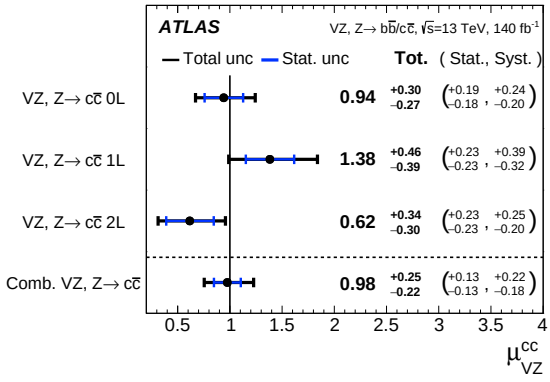


(b)

Figure 5.17: Figures showing observed values and uncertainties of $\mu_{VZb\bar{b}}$ (a) and $\mu_{VZc\bar{c}}$ (b) decorrelated in the WZ and ZZ processes from [104].



(a)



(b)

Figure 5.18: Figures showing observed values and uncertainties of $\mu_{VZb\bar{b}}$ (a) and $\mu_{VZc\bar{c}}$ (b) decorrelated in the zero, one and two lepton channels from [104].

or very nearly within one standard deviation of the Standard Model expectation. The only exception is observed for the zero lepton $VZ(\to b\bar{b})$ signal strength, which falls within two standard deviations of theory predictions. None of the observed differences are large enough to warrant evidence for theory disagreement.

5.4.2 $VH(\to b\bar{b}/c\bar{c})$ Measurement Results

The primary combined analysis measurement of the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ processes resulted in signal strengths of

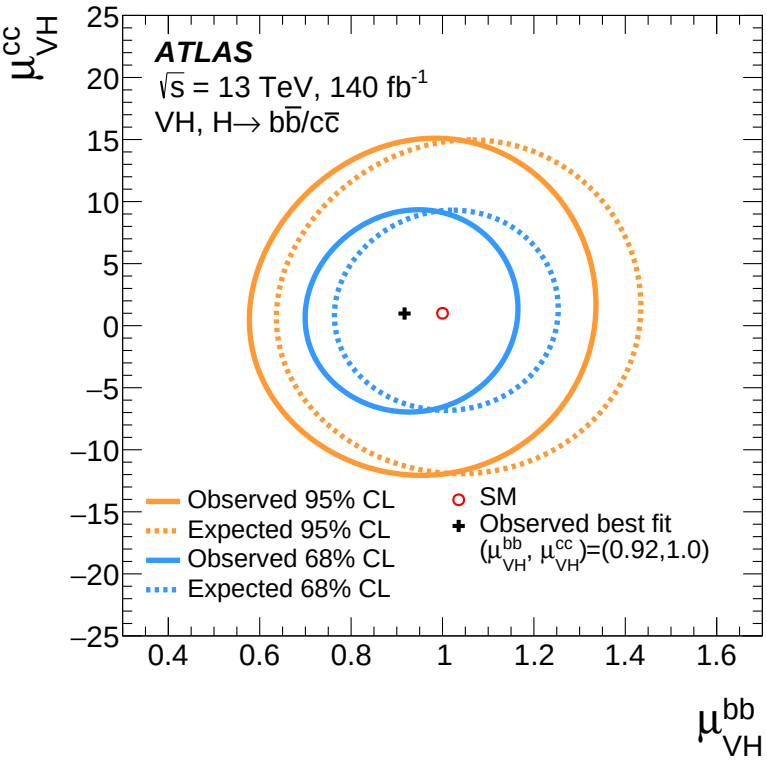


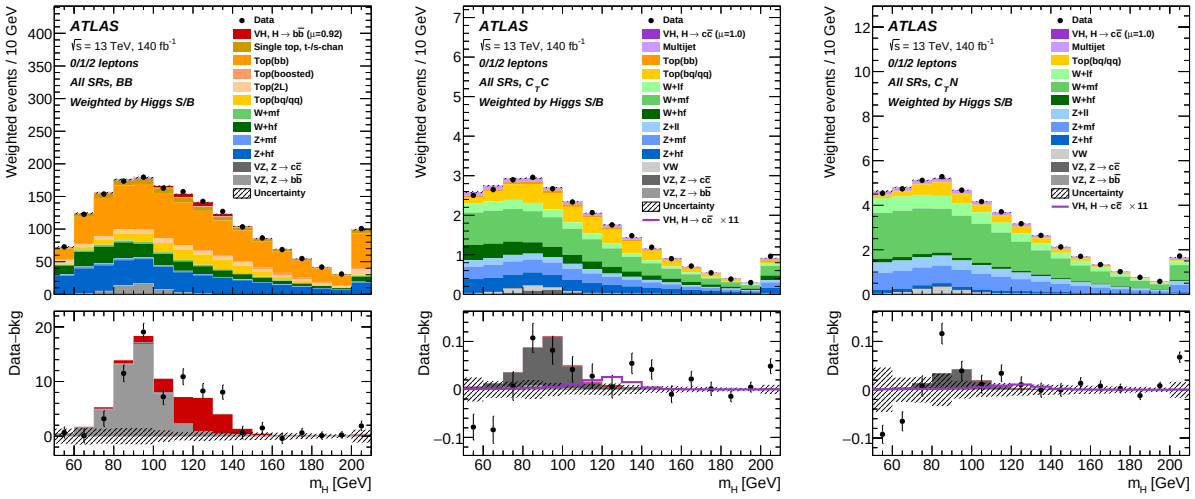
Figure 5.19: Contours showing observed and expected confidence intervals for $\mu_{VHb\bar{b}}$ and $\mu_{VHc\bar{c}}$ from [104].

$$\mu_{\mathrm{VHb\bar{b}}} = 0.92^{+0.16}_{-0.15} = 0.92 \pm 0.10(\mathrm{stat.})^{+0.13}_{-0.11}(\mathrm{syst.}) \tag{5.13}$$

$$\mu_{\mathrm{VHc\bar{c}}} = 1.0^{+5.4}_{-5.2} = 1.0^{+4.0}_{-3.9}(\mathrm{stat.})^{+3.7}_{-3.5}(\mathrm{syst.}) \tag{5.14}$$

In addition to one dimensional intervals, both parameters are probed simultaneously in a two dimensional fit resulting in contours for the two correlated signal strength parameters as shown in Fig. 5.19, with correspondingly larger uncertainties on each. The $VH(\to b\bar{b})$ process was measured to an observed significance of 7.4σ compared to the expected significance of 8.0σ. The 95% confidence level upper limit for the $VH(\to c\bar{c})$ process was expected as $10.6 \times$ SM and observed as $11.5 \times$ SM. While this result does not support or exclude the possibility of physics beyond the Standard Model, the observed signal strength for $VH(\to c\bar{c})$ is consistent with expectation within the constraints of the large associated uncertainties.

In the κ-framework [190] discussed in section 2.3.3, the measured signal strengths are reparameterized in terms of the coupling modifiers κ_b and κ_c as



(a) (b) (c)

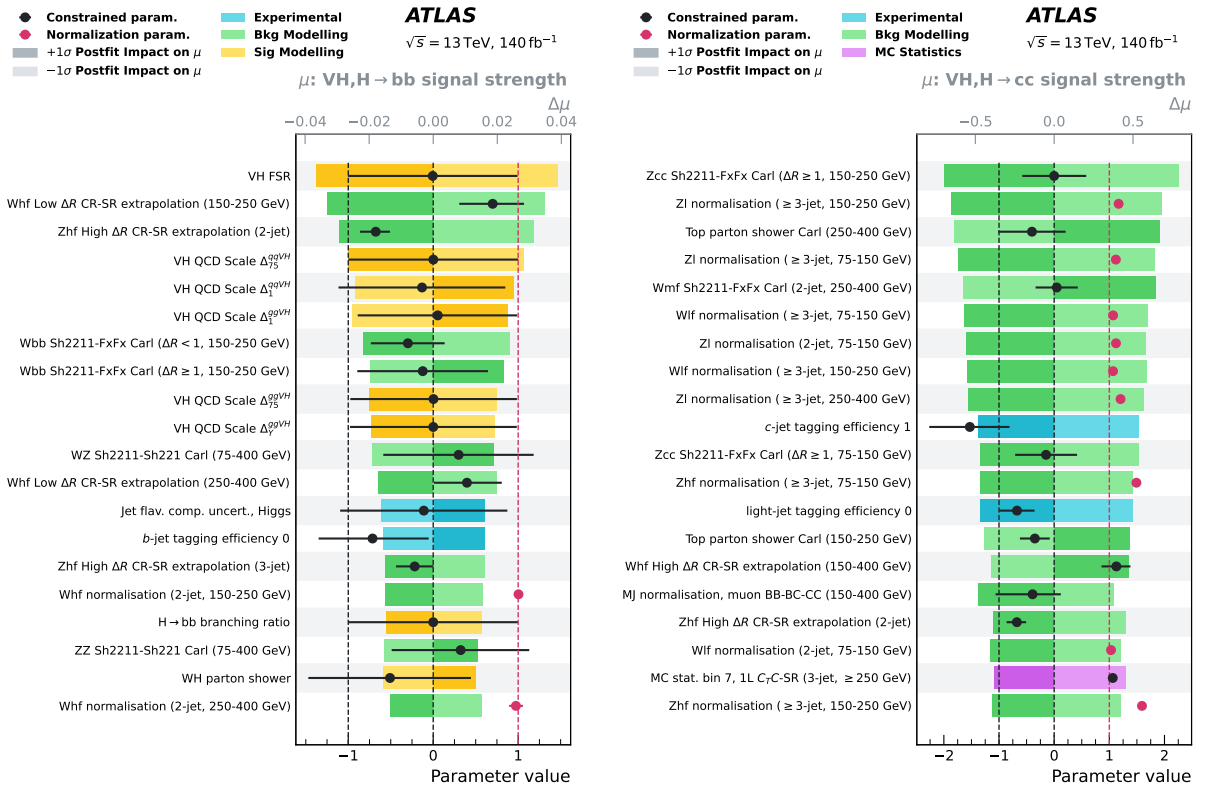
Figure 5.20: Figures showing Higgs S/B weighted event distributions across BB (a), C_TC (b) and C_TN (c) analysis regions.

$$\mu_{VHbb} = \frac{\kappa_b^2}{1 + BR_{Hbb}^{SM}(\kappa_b^2 - 1) + BR_{Hcc}^{SM}(\kappa_c^2 - 1)} \tag{5.15}$$

$$\mu_{VHcc} = \frac{\kappa_c^2}{1 + BR_{Hbb}^{SM}(\kappa_b^2 - 1) + BR_{Hcc}^{SM}(\kappa_c^2 - 1)} \tag{5.16}$$

In each case, the Higgs width is assumed to be consistent with Standard Model predictions and κ constants are taken to impact Higgs boson decays only. Results from fitting κ_b provide observed (expected) 95% CL intervals of $0.65 < |\kappa_b| < 1.37$ ($0.72 < |\kappa_b| < 1.62$). For κ_c, the equivalent observed (expected) limits are $|\kappa_c| < 4.2$ ($|\kappa_c| < 4.1$). When considering the ratio $|\kappa_c/\kappa_b|$ the same procedure decouples the result from assumptions on the Higgs width, leading to observed (expected) 95% CL upper limits of $|\kappa_c/\kappa_b| < 3.6$ ($|\kappa_c/\kappa_b| < 3.5$). These ratios are lower than that of bottom and charm quark masses, suggesting the Higgs coupling to the latter is weaker overall as expected.

To better visualize the signal components of the VH measurement, Fig. 5.20 shows data and MC distributions across all analysis SRs weighted by the total signal to background (S/B) ratio of each particular region based on MC. This greatly emphasizes overall signal contributions due to overrepresentation of signal rich regions. The bottom part of each plot shows the VH (in red) and VZ (in purple) components with all other backgrounds subtracted from the data distributions to further emphasize expected signal contributions in data. Yields across all bins generally show good agreement between data and MC and are a useful visual showcase for the statistical power of this measurement.



(a) (b)

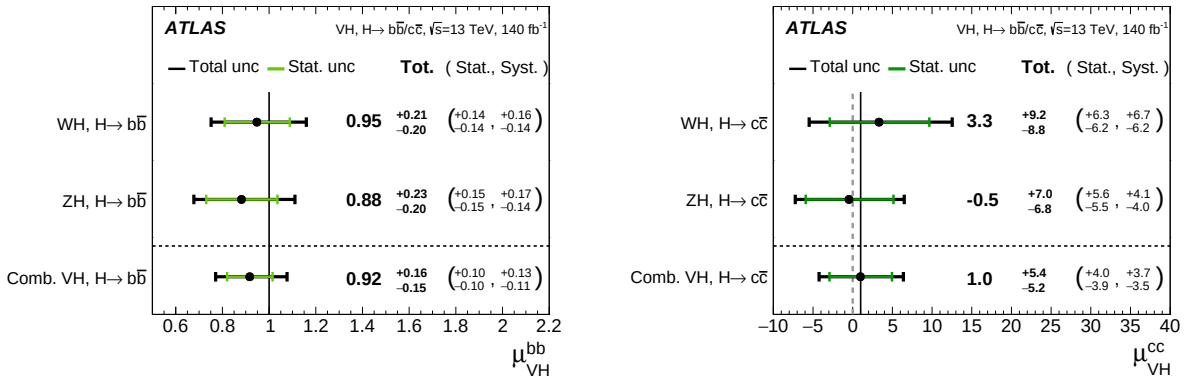
Figure 5.21: Figures showing ranking of systematic uncertainties based on $\Delta\mu$ impact on measured signal strengths for $VH(\to b\bar{b})$ (a) and $VH(\to c\bar{c})$ (b) alongside pulls and constraints for selected NPs from [104].

Contributions from systematics to signal strength uncertainties in the 1+1 POI fit are summarized in the NP "ranking" shown in Fig. 5.21 and broken down in detail in table 5.11. In the $VH(\to b\bar{b})$ measurement, the most impactful uncertainties stem from signal and V+jets modeling. Data statistics are also a large contributor, but have a smaller impact than systematic uncertainties overall. For $VH(\to c\bar{c})$, statistical uncertainties are dominant with additional large contributions from V+jets and $t\bar{t}$ modeling. Signal modeling is less impactful in this regime due to the process's small expected cross section. Experimental systematics make up the smallest uncertainty component across both measurements in the VH regime, but are not negligible. The largest contributors in this category are flavor tagging and general jet systematics.

As in the VZ analysis space, a number of additional fits with decorrelated POIs are performed. Signal strength results from process decorrelated 2+2 POI fits splitting WH and ZH are shown in Fig. 5.22 while those decorrelating the POIs across lepton channels

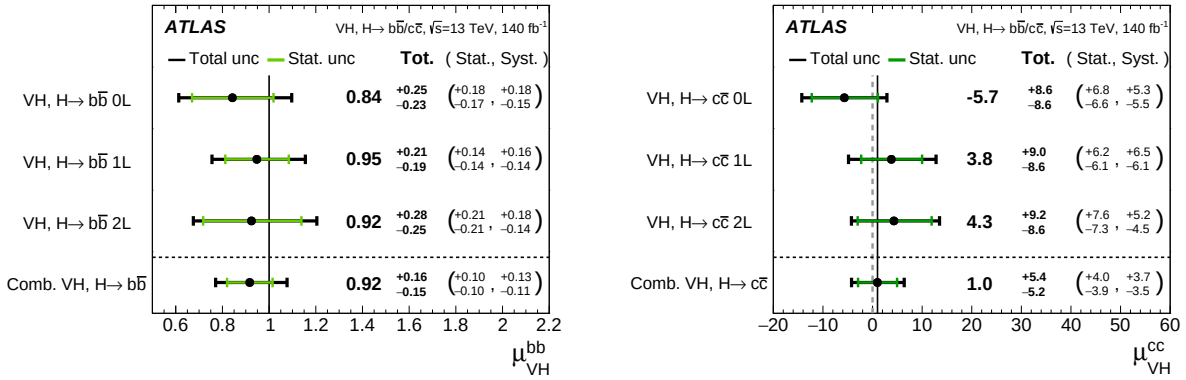
Nuisance parameters	Impact on $\mu_{VHb\bar{b}}$ error		Impact on $\mu_{VHc\bar{c}}$ error	
Total	+0.160 / -0.146	±0.153	+5.407 / -5.215	±5.311
Data statistical	+0.098 / -0.097	±0.097	+3.970 / -3.903	±3.936
Data statistical only	+0.090 / -0.089	±0.090	+3.701 / -3.647	±3.674
Floating normalizations	+0.035 / -0.032	±0.034	+1.301 / -1.170	±1.235
$VH(\to c\bar{c}/b\bar{b})$ normalization	+0.008 / -0.006	±0.007	+0.585 / -0.070	±0.328
Top($e\mu$) CR statistical	+0.010 / -0.009	±0.009	+0.078 / -0.091	±0.084
Systematic	+0.126 / -0.110	±0.118	+3.670 / -3.460	±3.565
MC statistical	+0.024 / -0.022	±0.023	+1.642 / -1.589	±1.616
VH modeling	+0.087 / -0.065	±0.076	+0.937 / -0.510	±0.723
Background modeling	+0.082 / -0.077	±0.080	+3.044 / -2.918	±2.981
multi-jet	+0.006 / -0.005	±0.005	+0.574 / -0.532	±0.553
single-top s-, t-channel	+0.010 / -0.010	±0.010	+0.167 / -0.160	±0.164
$t\bar{t}$ and single-top Wt	+0.019 / -0.017	±0.018	+1.056 / -0.988	±1.022
W+jets	+0.056 / -0.052	±0.054	+1.448 / -1.387	±1.417
Z+jets	+0.043 / -0.040	±0.042	+1.791 / -1.738	±1.765
Diboson	+0.034 / -0.031	±0.033	+0.528 / -0.513	±0.521
Experimental	+0.043 / -0.037	±0.040	+1.536 / -1.368	±1.452
Leptons	+0.004 / -0.003	±0.003	+0.256 / -0.237	±0.246
E_T^{miss}	+0.010 / -0.009	±0.010	+0.226 / -0.231	±0.229
Jets	+0.030 / -0.024	±0.027	+1.031 / -0.999	±1.015
Flavor tagging (b-jets)	+0.021 / -0.019	±0.020	+0.272 / -0.314	±0.293
Flavor tagging (c-jets)	+0.014 / -0.012	±0.013	+0.853 / -0.597	±0.725
Flavor tagging (l-jets)	+0.006 / -0.004	±0.005	+0.746 / -0.570	±0.658
Flavor tagging (τ-jets)	+0.001 / -0.004	±0.002	+0.077 / -0.051	±0.064
Pile-up	+0.008 / -0.008	±0.008	+0.246 / -0.222	±0.234
Luminosity	+0.008 / -0.005	±0.006	+0.083 / -0.070	±0.077

Table 5.11: Detailed breakdown of NP contributions to errors on VH signal strengths adapted from [72].



(a) (b)

Figure 5.22: Figures showing observed values and uncertainties of $\mu_{VHb\bar{b}}$ (a) and $\mu_{VHc\bar{c}}$ (b) decorrelated in the WZ and ZZ processes from [104].



(a) (b)

Figure 5.23: Figures showing observed values and uncertainties of $\mu_{VHb\bar{b}}$ (a) and $\mu_{VHc\bar{c}}$ (b) decorrelated in the zero, one and two lepton channels from [104].

are shown in Fig. 5.23. Expected and observed 95% confidence level limits on the $VH(\to c\bar{c})$ signal strength are also presented in Fig. 5.24 for the latter. Decorrelated fit results broadly agree with the main 1+1 POI fit as all fitted signal strength values lie within 1σ of theory expectation. WH and ZH show similar sensitivity for $VH(\to b\bar{b})$ and ZH is more sensitive in the $VH(\to c\bar{c})$ regime. The one lepton channel provides the smallest uncertainties on $\mu_{VHb\bar{b}}$ while the zero lepton channel does so for $\mu_{VHc\bar{c}}$ by a relatively smaller margin.

The final set of VH fit results come from the STXS (Stage 1.2) [16] split of the $VH(\to b\bar{b})$ signal strength POI into 10 and 13 bins respectively. The results of these 10+1 and 13+1 POI fits are shown in Fig. 5.25. These plots illustrate the measurement sensitivity across

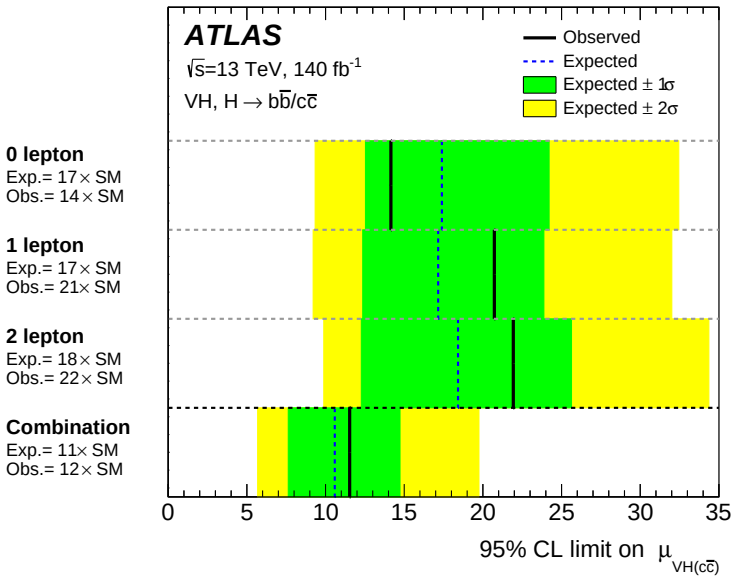
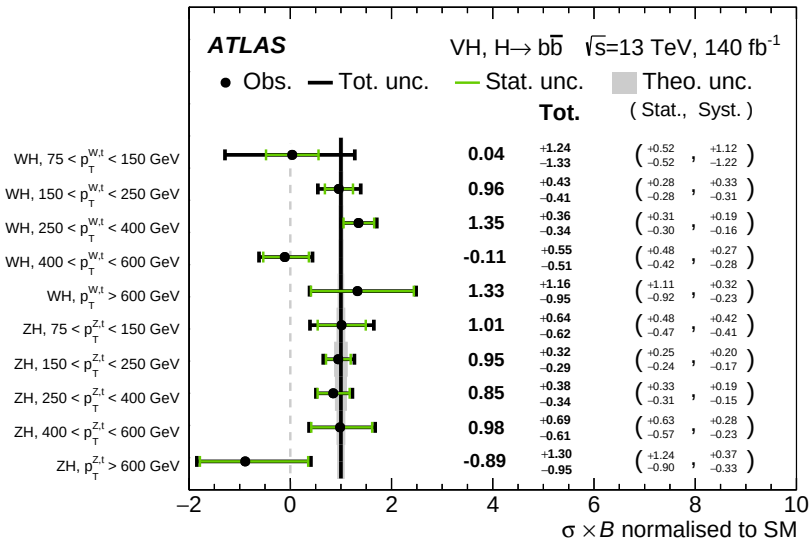


Figure 5.24: Visualization of expected and observed limits for $\mu_{VHc\bar{c}}$ in fit decorrelated by lepton regions.

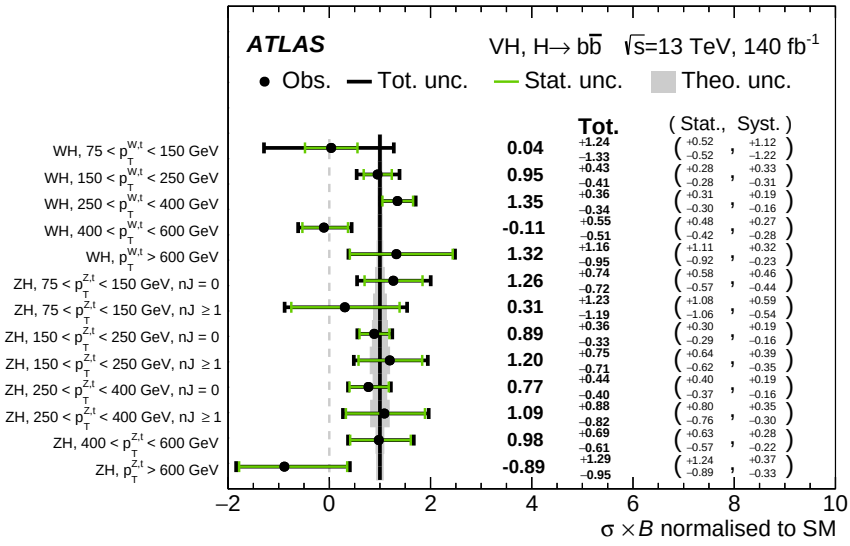
different p_T^V and N_{jet} bins. Events with moderate p_T^V between 150 GeV and 400 GeV with no additional jets provide the most significant signal strength contributions. Analysis sensitivity decreases rapidly in the low and high transverse momentum regimes. All STXS decorrelated POIs show good agreement with Standard Model theory with no deviations larger than 2σ.

5.4.3 Final Fit Model Validation

The final fit model is nearly identical for VH and VZ fits and almost all best fit NP values are within 1σ of each other between the two, with no concerning deviations (i.e. large unexpected pulls or shifts in floating normalizations) observed. Detailed results are discussed for the VH fit only. All post-fit NP values are relatively consistent with their pre-fit expectations [72]. Most floating normalization factors are within 30% of unity with an overall decrease in top background normalization and corresponding increase in V+jets backgrounds. A selection of V+hf and V+mf normalization factors have post-fit values up to twice as large as nominally expected. Outside of floating normalizations, deviations of best fit NP values of more than 1σ from pre-fit expectations are present but not frequent and do not show concerning patterns. Acceptance and shape uncertainties for larger backgrounds are often constrained by the fit model in a consistent fashion between Asimov and data fits. On the experimental side, the analysis has some sensitivity to constrain both leading flavor tagging as well as small-R jet systematics (particularly JER), which have conservative pre-fit estimates. The impact of the



(a)



(b)

Figure 5.25: Figures showing observed values and uncertainties of $\mu_{VHb\bar{b}}$ in the 10 (a) and 13 (b) $VH(\to b\bar{b})$ POI STXS split from [104].

final fit model configuration on post-fit yields can be seen across all signal regions for the lowest jet multiplicity regions in Fig. 5.26 for resolved $VH(\to b\bar{b})$, Figs. 5.27 and 5.28 for $VH(\to c\bar{c})$ and Fig. 5.29 for boosted $VH(\to b\bar{b})$.

5.5 Conclusions

The fully combined run 2 $V(\to ll)H(\to b\bar{b}/c\bar{c})$ measurement sets the most stringent limit on $\mu_{VHb\bar{b}}$ and $\mu_{VHc\bar{c}}$ to date with significantly better sensitivity compared to prior ATLAS measurements using the same data. The observed upper limit for $VH(c\bar{c})$ of 11.5 $\times$ SM compares favorably to both the most recent ATLAS result of 26 $\times$ SM [81] as well as the more recent CMS result of 14 $\times$ SM [119]. The main sources of improvement stem from the application of sophisticated machine learning techniques. These consist of a BDT-based analysis MVA and the DNN-based jet flavor identification algorithm DL1r. Both areas hold potential for further improvements through the use of improved algorithms that better take low level information into account in comparison to BDTs and DNNs. A series of such innovations are presented in the space of jet flavor identification in chapter 6 and their application as well as an upgraded MVA strategy is shown for a new $VH(\to c\bar{c})$ measurement in chapter 7.

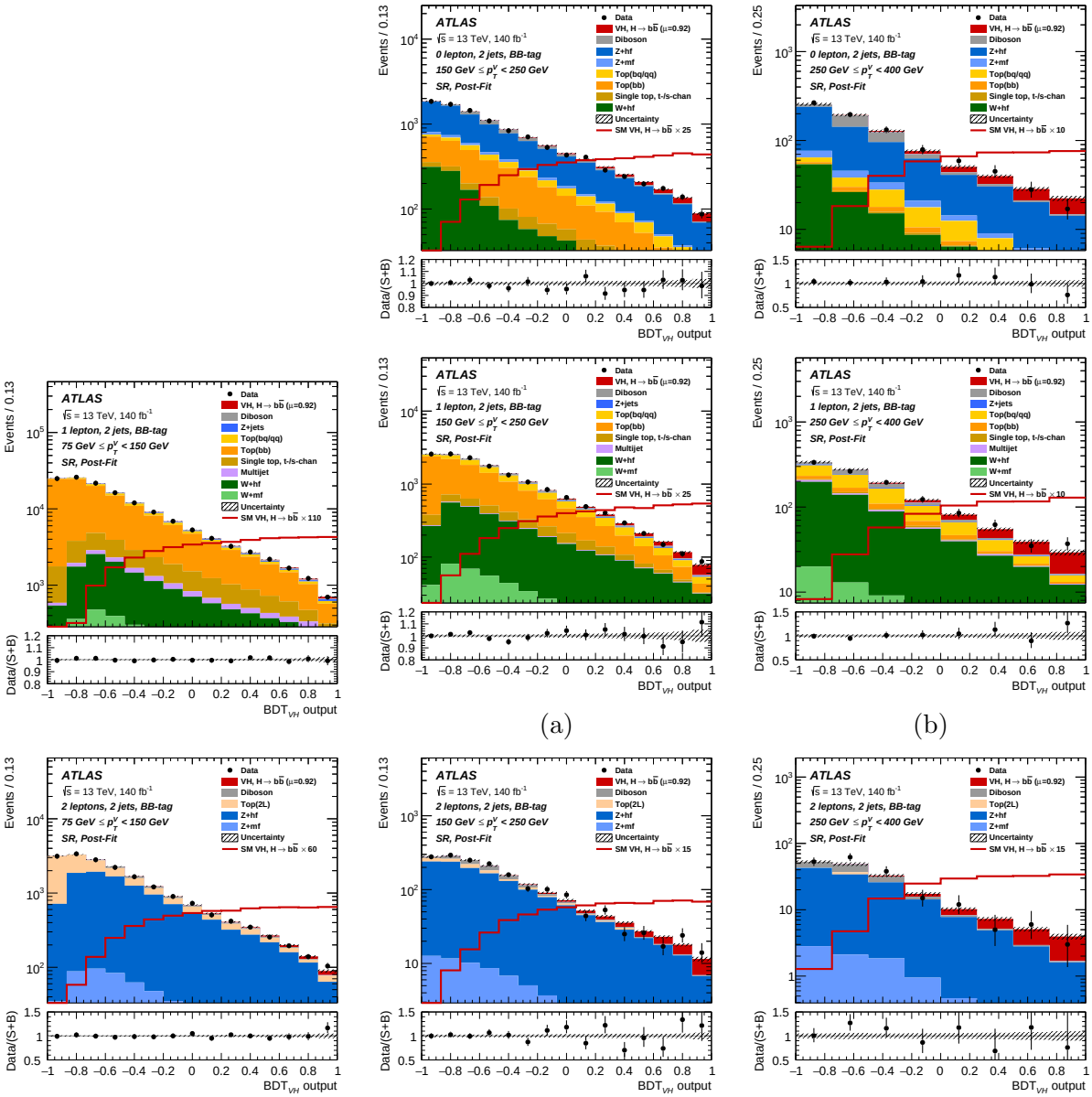


Figure 5.26: Figures showing post-fit distributions for all $N_{\text{jet}} = 2$ signal regions in the resolved $VH(\to b\bar{b})$ regime arranged horizontally by p_T^V category and vertically by lepton channel from [104].

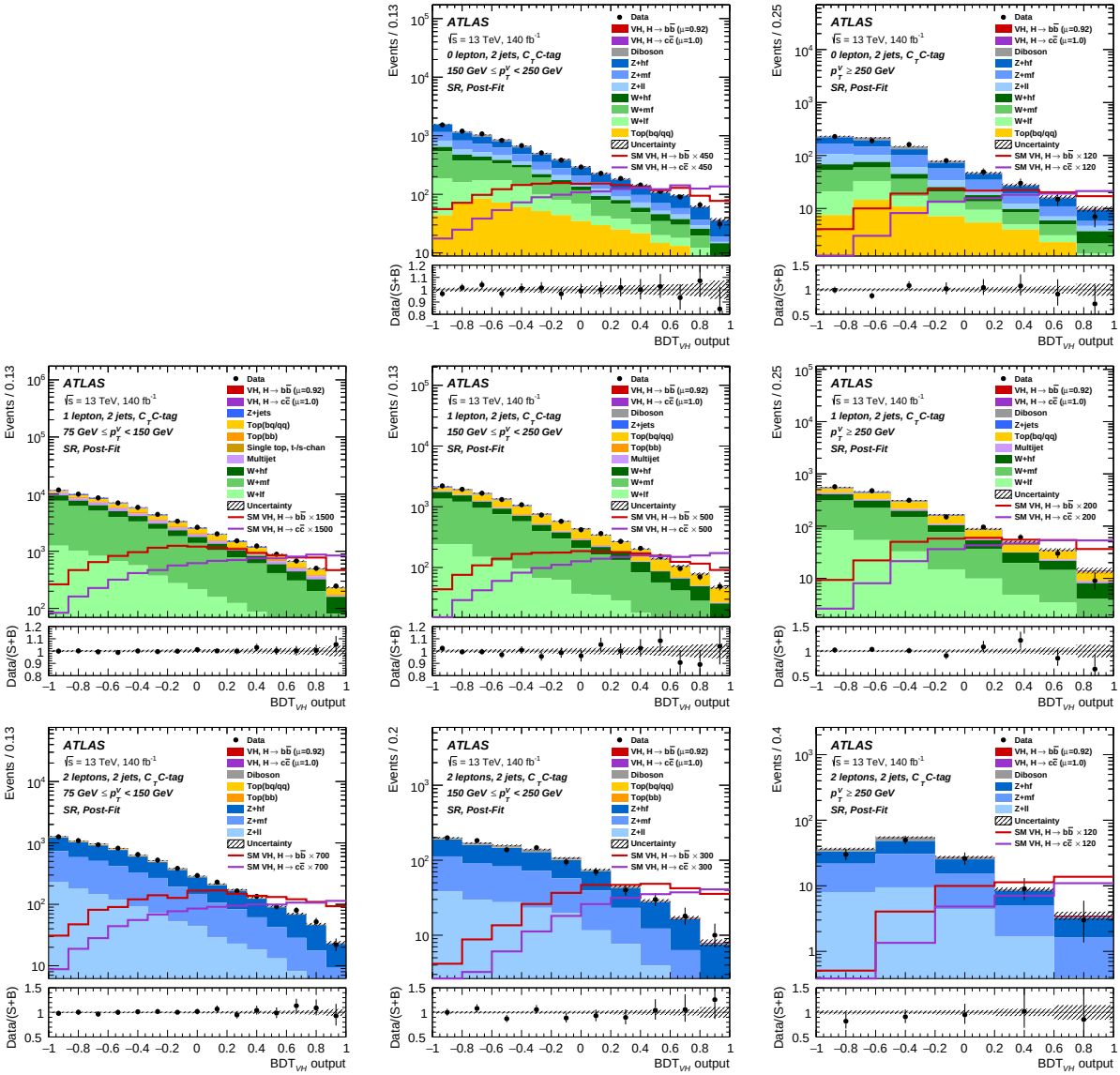


Figure 5.27: Figures showing post-fit distributions for all $N_{\text{jet}} = 2$ $C_T C$ signal regions in the $VH(\to b\bar{b})$ regime arranged horizontally by p_T^V category and vertically by lepton channel from [104].

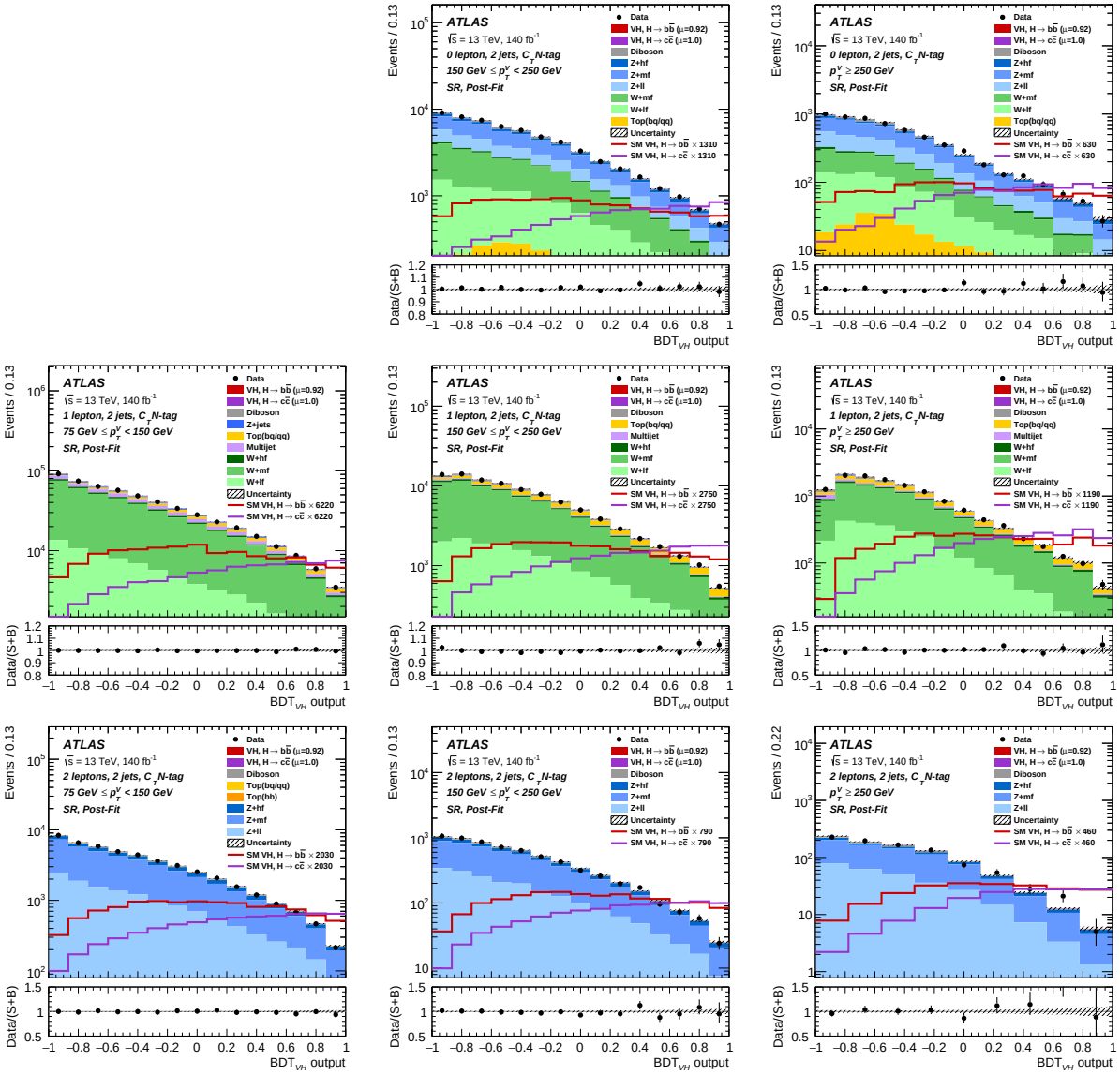


Figure 5.28: Figures showing post-fit distributions for all $N_{\text{jet}} = 2$ C_TN signal regions in the $VH(\to c\bar{c})$ regime arranged horizontally by p_T^V category and vertically by lepton channel from [104].

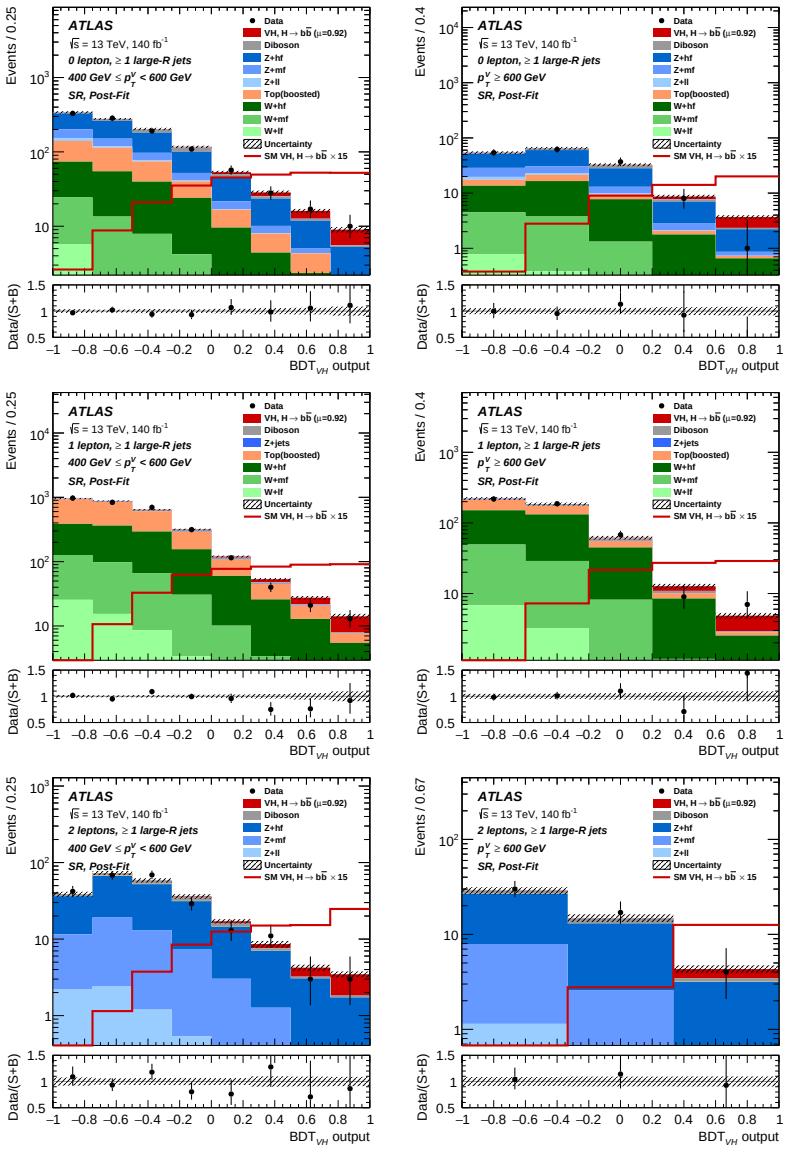


Figure 5.29: Figures showing post-fit distributions for all signal regions in the boosted $VH(\to b\bar{b})$ regime arranged horizontally by p_T^V category and vertically by lepton channel from [104].

Chapter 6

Advances in Flavor Tagging using Machine Learning

Improving the sensitivity of a $H \to c\bar{c}$ measurement depends primarily on the identification of charm-jets through jet flavor tagging and separating them from bottom- and light-jet populations. The ATLAS run 2 approach to this problem is characterized by the use of physically motivated algorithms that exploit individual track features such as the longitudinal and transverse IPs and perform explicit secondary vertex (SV) reconstructions combined via a lightweight DNN as described in section 4.2. The introduction of this DNN (called DL1) presented the first widely adopted use of neural networks for flavor tagging in the ATLAS experiment and brought significant performance improvements with it [90]. In the context of $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$, this approach enabled the results presented in chapter 5, leading to the most sensitive measurement of these process up to that point. This methodology does not generally allow for the consideration of correlations between features across different tracks, which are relevant due to the comparatively high track multiplicity in heavy flavor hadron decays. Modern machine learning techniques such as Graph Neural Networks (GNNs) [191, 164] and Transformers [215] allow for a fundamental shift in strategy that enables encoding much more complex relationships between jet constituents than the previous approach, providing enormous potential benefits for measurements of the charm Yukawa coupling.

This chapter details a set of new machine learning based jet flavor identification algorithms developed for analyses of run 3 ATLAS data. Advances made in the IP-based flavor identification algorithm `RNNIP` [50] are expanded to explicit SV reconstruction using GNNs (section 6.1). These techniques are then generalized via the introduction of new Transformer-based flavor tagging algorithms in the low and high p_T regimes on small-R and large-R jets via `GN2` (section 6.2) and `GN2X` (section 6.4) respectively. An evolution of the former called `GN3` which incorporates jet charge identification and allows for separation of c and $\bar{c}$ in $H \to c\bar{c}$ decays is also presented (section 6.3). Finally, an updated version of `GN2X` called `GN3X` is shown (section 6.4), which provides the baseline for the discussions of chapter 7.

6.1 SVGNN: Vertex Reconstruction with Graph Attention Networks

The existence and characteristics of secondary decay vertices provide a powerful discriminant between bottom, charm and light jets, as is evident in algorithms such as SV1 [54] and JetFitter [61]. In IP-based algorithms, RNNIP [50] shows that taking into account correlations between different tracks can result in significant performance benefits compared to finetuned fitting algorithms such as IP2D and IP3D [47]. This approach can be extended to SV reconstruction by considering sophisticated machine learning techniques. RNNIP uses an LSTM [152] architecture, which is inherently useful for extracting information from ordered sequence data. Jets, which are fundamentally collections of unordered particle tracks thus do not entirely fit this description. Orderings on tracks can be imposed by features such as p_T or IP significance, but a more natural approach requires enforcing permutation invariance in the algorithm itself. Graph Neural Networks (GNNs) [191, 164] are designed to encode information from permutation invariant sets of objects expressed as graphs. A variant of this architecture is used to define a new machine learning based secondary vertex reconstruction algorithm called SVGNN based on previous work using a "Deep Sets" [221] architecture [195]. This section explores the SVGNN architecture and its performance compared to other algorithms such as SV1 and JetFitter.

6.1.1 Description of the SVGNN Architecture

The problem of SV reconstruction in jets can be framed as an edge classification problem on a fully-connected graph. Here, the graph represents a single small-R jet and each node is a track with an associated set of features. Tracks are matched to jets via delta-R association as described in section 4.1.1. The edges connecting all node pairs represent the respective relationships between them, which can be classified as connecting two tracks that originate from a common SV or not. SVGNN is a GNN based algorithm built with pyTorch and DGL [218] that takes this approach. It consists of 3 different neural network components visualized in Fig. 6.1, which turn individual sets of track features into edge classification scores: a node initialization network, a graph (GAT) network and an edge classification network.

Each track used as input to the network has an associated feature vector of 34 different features consisting of kinematics and IPs with corresponding uncertainties as well as lower level information carried over from track reconstruction. The full list of input features is provided in table 6.1. The set of track feature vectors $\mathbf{x}_i$ is simultaneously used as input for the node initialization network as well as the graph network. The former applies a series of affine transformations defined by $\mathbf{h}_i^{node} = \mathbf{w}_{node}\mathbf{x}_i + \mathbf{b}_{node}$ to the feature set of each node i. Here $\mathbf{h}_i^{node}$ represents an encoding of track feature information that is independent for each track and $\mathbf{w}_{node}$ and $\mathbf{b}_{node}$ are two learnable vectors in each layer. The graph network component takes into account all neighboring track feature vectors via a series of Graph Attention (GATv2Conv) [19] transformations defined as

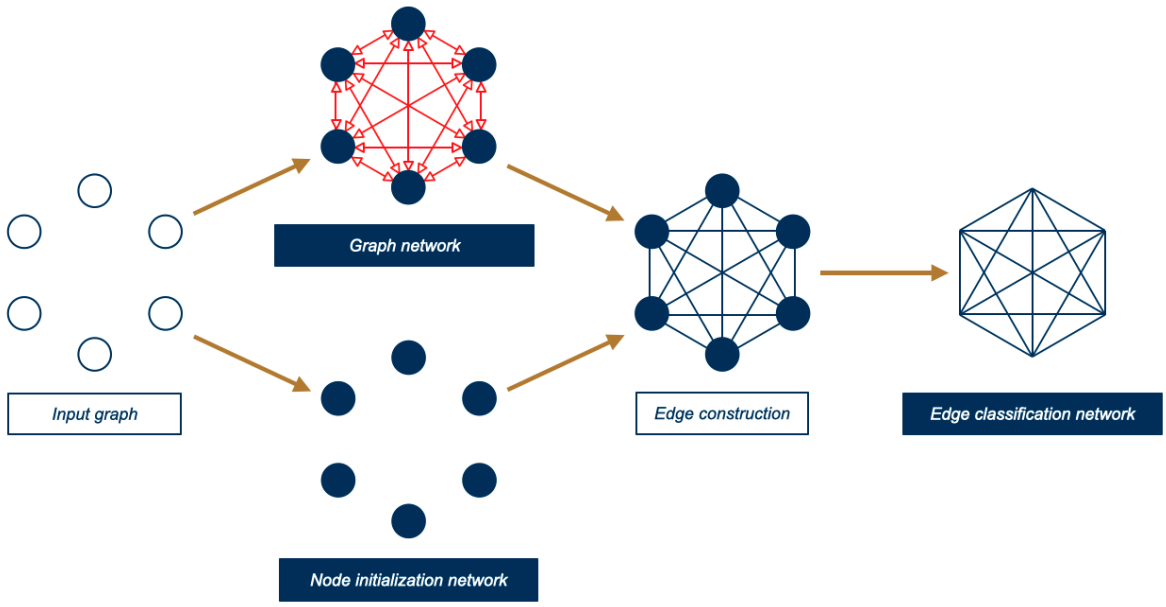


Figure 6.1: Visualization of the architecture of the SVGNN algorithm. Here, nodes represent individual tracks and edges represent relationships between them. Information originally associated with each node is transformed into edge feature sets, which are classified to identify SV connections.

$$\mathbf{h}_i^{\text{graph}} = \sum_{j=1}^{n_{\text{tracks}}} \alpha_{i,j} w_{graph} \mathbf{x}_j \tag{6.1}$$

This results in set of conditional representations of the individual track feature matrices, each of which contains information from all tracks across the jet. In practice, this is how correlations between tracks are taken into account. The makeup of this information is determined by the $\alpha_{i,j}$, which are attention weights that determine the relative contribution of a track's neighbor to the conditional representation of its feature vector. They are defined as

$$\alpha_{i,j} = \frac{\exp(\mathbf{a}^T \text{LeakyReLU}(w_{graph}(\mathbf{x}_i + \mathbf{x}_j)))}{\sum_{k=1}^{n_{\text{tracks}}} \exp(\mathbf{a}^T \text{LeakyReLU}(w_{graph}(\mathbf{x}_i + \mathbf{x}_k)))} \tag{6.2}$$

The GNN layers have learnable parameters $\mathbf{w}_{graph}$ and $\mathbf{a}$ and "LeakyReLU" represents a non-linear "Leaky Rectified Linear Unit" function defined by $\text{LeakyReLU}(x) = \max(0.01x, x)$ [176]. Each GATv2Conv layer can consist of multiple heads, which apply independently optimized $\mathbf{a}$ vectors, such that the total dimensionality of the output feature set is equally split between them.

Outputs from the node initialization and the graph network are concatenated for each track, resulting in an encoded feature set that contains both independent and conditional features. An edge is considered to be defined by the concatenation of a pair of encoded track feature sets. The edge classification network applies a second series of affine transformations to each pair, deriving a final edge classification score between 0 and 1. An edge score close to 1 indicates a high network assigned likelihood of the two tracks being a part of the same secondary vertex.

Table 6.1: Summary of input variables for each track in `SVGNN`.

Variable	**Description**
p_T^J	Associated jet p_T
η^J	Associated jet η
ϕ^J	Associated jet ϕ
q/p_T, $\sigma_{q/p}$	Track q/p_T and uncertainty of q/p
η, σ_η	Track η and uncertainty
ϕ, σ_ϕ	Track ϕ and uncertainty
d_0, σ_{d0}	Track d_0 and uncertainty
z_0, σ_{z0}	Track z_0 and uncertainty
$cov(x, y)$ where $x, y \in \{q/p, \eta, \phi, d_0, z_0\}$	Pairwise covariances for track q/p, η, ϕ, d_0 and z_0
n_{PixHits}, n_{SCTHits}, n_{BLHits}	Number of hits registered in the Pixel detector, SCT and IBL
n_{PixHoles}, n_{SCTHoles}	Number of missing hits along track path (holes) in the Pixel detector and SCT
$n_{\text{PixShared}}$, $n_{\text{SCTShared}}$, n_{BLShared}	Number of hits shared with other tracks in the Pixel detector, SCT and IBL
n_{PixSplit}, n_{BLSplit}	Number of hits split between clusters in the Pixel detector and IBL
w_{PV}	PV association weight for track provided by JVT

To ensure the network is a universal approximator [154], all affine and GATv2Conv layers are connected through intermediate non-linearities called activation functions, mostly "Rectified Linear Units" (ReLU) defined as $\text{ReLU}(x) = \max(0, x)$. For the final output layer, a Sigmoid function defined by $\sigma(x) = \frac{1}{1+e^{-x}}$ is used instead to restrict possible outputs to probabilities between 0 and 1. Batch normalization [157] is also applied to each set of inputs before the first layers. A detailed description of the `SVGNN` architecture is provided in table 6.2.

6.1.2 Training of `SVGNN`

To optimize the learnable parameters, the algorithm is trained on a set of MC simulated bottom, charm and light jets. The optimization is steered by backpropagation gradient

Table 6.2: Detailed summary of SVGNN architecture with layer types and sizes.

Component	**Layers**
Node initialization	BatchNorm Affine (34 in, 128 out) + ReLU Affine (128 in, 128 out) + ReLU
Graph network	BatchNorm GATv2Conv (34 in, 384 out, 2 heads) + ReLU GATv2Conv (384 in, 1152 out, 2 heads) + ReLU
Edge classification	Affine (1152×2 + 128×2 in, 512 out) + ReLU Affine (512 in, 256 out) + ReLU Affine (256 in, 64 out) + ReLU Affine (64 in, 1 out) + Sigmoid

descent [189] through minimization of a binary cross-entropy[1] loss function defined by

$$\mathcal{L} = -\frac{1}{N_{edges}} \sum_{i=1}^{N_{edges}} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right] \tag{6.3}$$

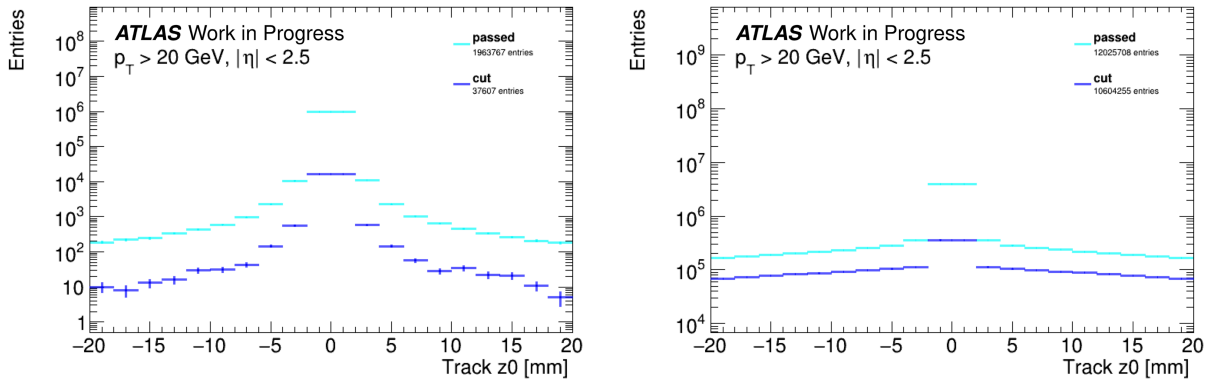
Here, y_i is the true label for each edge and $\hat{y}_i$ is the SVGNN prediction. The labels themselves are based on the MC truth record, with edges between tracks originating from the decay of the same heavy flavor hadron labeled as 1 and all other edges as 0. In bottom-jets, tracks from the decay of the bottom hadron are treated as being from a separate vertex than those from the subsequent charm hadron decay. Since negative labels far outnumber positive ones, a multiplicative factor $(1 - f_{pos})/{f_{pos}}^2$ is applied to all positive label contributions to the loss function, thereby ensuring equal contributions from the two classes.

Before training, a series of basic kinematic cuts $p_T^J > 20$ GeV and $|\eta^J| < 2.5$ are performed on jets. In remaining jets, tracks that do not meet the criteria $p_T > 650$ MeV, $|\eta| < 2.5$ and $|z_0| < 20$ mm are pruned. Overall, Fig. 6.2 shows the impact of the track based cuts in the training dataset. A total of 47% of tracks not originating in decays of heavy flavor particles are cut while only fewer than 2% of tracks from bottom or charm hadron decays are affected. The training dataset consists of around 1.7 million jets[3] taken from $t\bar{t}$ and $ZH \to \ell\ell c\bar{c}$ samples detailed in table 6.3. The average graph consists of 8 nodes after cuts. Each "step" of loss function minimization via backpropagation gradient descent occurs simultaneously on a "batch" of 1000 jets via the ADAM optimizer [163]. Gradient information from each batch

[1]Cross-entropy generally refers to a measure of difference between two probability distributions. As a loss function it is used to minimize the difference between the network's predicted probability distribution for a binary variable and the one given by its underlying truth labels.

[2]Here f_{pos} is the total fraction of positive labels.

[3]Split as approximately 1 million light-jets, 400,000 charm-jets and 300,000 bottom-jets.



(a) (b)

Figure 6.2: Plots showing distribution of z_0 of tracks passing and failing cuts for `SVGNN` training originating from (a) heavy flavor particle decays and (b) other sources

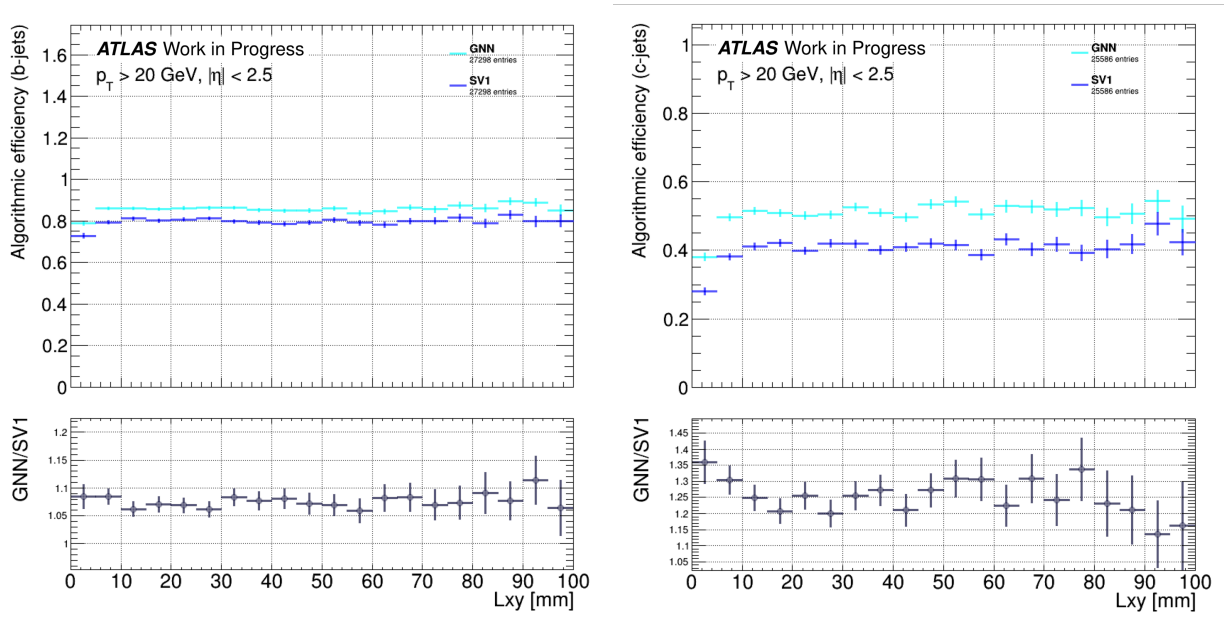
is used to update model parameters throughout the training period for a total of 40 passes through the training dataset, or "epochs". At the end of each epoch, model performance is evaluated on a separate "validation" dataset, which measures how well model optimization on training data generalizes to a statistically independent sample. After some period, the training and validation performance generally diverge due to overoptimization on the training data (i.e. "overtraining"), leading to suboptimal generalizability. To combat this, the final set of model parameters is taken from the epoch that shows the best performance (i.e. minimum loss value) on the validation dataset.

Process	**Matrix element generator**	**Matrix element PDF**	**PS and hadronization**	**UE tune**	**HF quark decay**
$t\bar{t}$	`Powheg Box` v2 [3]	NNPDF3.0NLO [12]	`Pythia` 8.230 [196]	A14 [205]	`EvtGen` 1.6.0 [169]
$VH(\to c\bar{c})$	`Pythia` 8.306	NNPDF2.3LO	`Pythia` 8.306	A14	`EvtGen` 1.7.0

Table 6.3: Monte Carlo samples used for `SVGNN` training.

6.1.3 `SVGNN` Performance

Performance of the `SVGNN` model is evaluated by comparing sets of reconstructed secondary vertices against predictions from `SV1` and MC truth information. The output edge scores of the model are first converted to vertex objects with associated tracks via an iterative algorithm. For each graph, the highest scoring edge is compared against a threshold score. If the



(a) (b)

Figure 6.3: Plots showing vertex reconstruction efficiency in SVGNN and SV1 as a function of PV-SV separation (L_{xy}) for (a) bottom-jets and (b) charm-jets with SV's

edge passes, its associated nodes are denoted as a reconstructed secondary vertex. Additional edges are added in descending order of their classification score until no more edges can be added without the average edge score going below the set threshold. A single secondary vertex is reconstructed for each jet to facilitate comparisons with the SV1 algorithm, but no inherent restriction exists on SV multiplicity in SVGNN in principle.

Vertex reconstruction efficiency comparisons between SV1 and SVGNN with an edge score threshold of 0.6 are shown in Fig. 6.3 for bottom- and charm-jets. The efficiency denotes the total fraction of jets with secondary vertices that were identified correctly after track pruning and is shown to be improved by around 5-10% for bottom-jets and 20-25% for charm-jets in SVGNN compared to SV1. The fraction of jets without SVs where a vertex was incorrectly identified, denoted by the fake rate is shown in Fig. 6.4. The fake rate is generally consistent between the two algorithms, with SVGNN showing improvements in low p_T jets.

Vertex reconstruction quality[4] was found to be similar between the two algorithms as well, except for a minor improvement in the number of falsely associated tracks in SVGNN. A complete summary of SVGNN vs. SV1 performance in terms of efficiency and fake rate for multiple threshold edge scores ranging from 0.1 to 0.9 is given in Fig. 6.5. This allows for additional flexibility in the balancing between the two performance metrics for the SVGNN algorithm compared to its predecessors. Overall, the results indicate significant increases in

[4]Quality refers to the correct and incorrect association of individual tracks to a given vertex

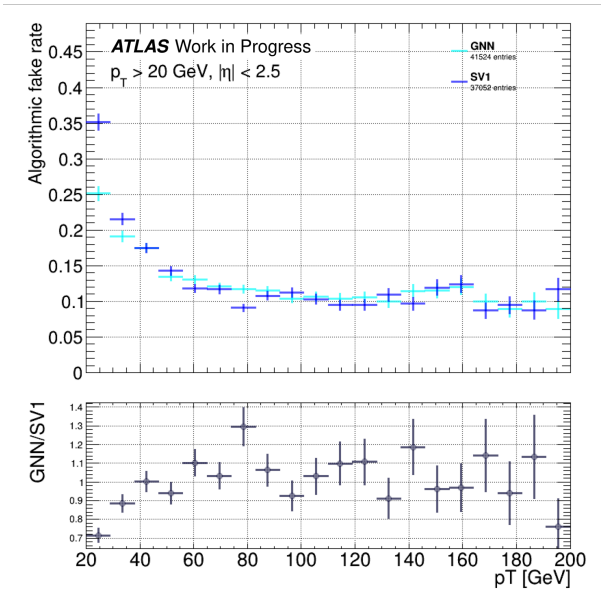


Figure 6.4: Plot showing vertex reconstruction fake rate in `SVGNN` and `SV1` as a function of jet p_T in jets without SV's.

the efficiency of SV reconstruction by applying machine learning techniques that allow for the exploitation of correlations between the features of different tracks.

6.2 `GN2`: Transformer-based Jet Flavor Identification

The general approach to small-R jet flavor identification has relied on a set of low level SV reconstruction and IP-based algorithms whose outputs are combined into a single discriminant. `SVGNN` proves that SV reconstruction performance can benefit from access to track-level correlations compared to traditional fitting algorithms. This idea can be expanded to jet flavor identification as a whole by removing the need for intermediate algorithms entirely and learning directly from low level track features to predict the jet class probabilities. While GNNs provide a naturally permutation invariant approach to this problem, Transformers [215] are a more computationally efficient solution [160] to extracting information from sets of related objects that retain the same advantages[5]. `GN2` [110] is a Transformer-based algorithm that performs jet flavor classification through an approach based on auxiliary training objectives, which guide the algorithm during the optimization process by emphasizing the learning of physically meaningful information. This section details the `GN2` algorithm and its performance as well as studies related to the integration of relational features.

[5]The two architectures are mathematically equivalent the case of fully-connected graphs.

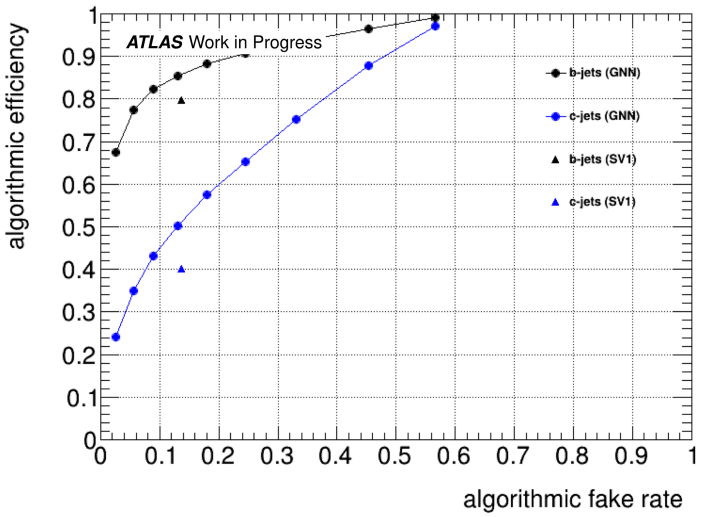


Figure 6.5: Plots showing vertex reconstruction efficiency vs. fake rate for `SVGNN` and `SV1`.

6.2.1 The Architecture of `GN2` and Auxiliary Tasks

The approach taken to the design of `GN2` is similar to the one outlined for `SVGNN` in section 6.1.1, viewing individual jets as graphs with nodes representing tracks and edges their relationships. An overview of the algorithm is shown in Fig. 6.6, which details each of its components. `GN2` uses ReLU activation functions throughout unless noted otherwise.

A set of normalized[6] track and jet level features summarized in table 6.4 is associated with each node and encoded using a track initialization network consisting of sequential affine transformations with learnable parameters. Up to 40 tracks[7] matched to each jet via delta-R association are considered with selection criteria detailed in table 6.5. The resulting track representations of dimension $(n_{\text{tracks}}, n_{\text{features}})$ are then duplicated and encoded separately through one additional affine layer each to produce a set of representations denoted $\mathbf{k_i}$, $\mathbf{q_i}$ and $\mathbf{v_i}$. The number of separate encodings i is determined by the number of so-called attention heads. These are used as inputs for a series of "Transformer Encoder" layers following the "Pre-LN Transformer" implementation [220] shown in Fig. 6.7. This architecture applies a layer normalization operation [7], which normalizes all input elements across the feature space for training stability purposes. The core self-attention mechanism is applied after layer normalization and is defined as

[6]Meaning scaled to have a standard deviation of 1 and shifted to a mean of 0 across the training dataset

[7]The number 40 was chosen because it represents a realistic upper bound for the number of tracks in most small-R jets while minimizing computational complexity, which grows with the square of the number of tracks.

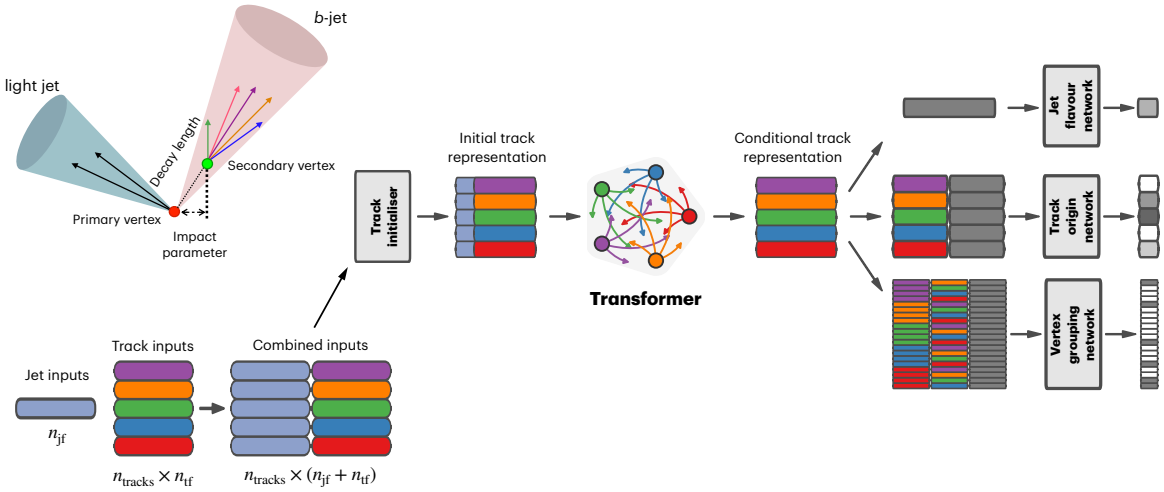


Figure 6.6: Diagram showing the different components of the GN2 algorithm from [110].

$$h = \text{LayerNorm}\left(\mathbf{x} + \text{Concat}(a_1, ..., a_n)\mathbf{w}\right) \tag{6.4}$$

As usual, $\mathbf{w}$ represents a learnable multiplication matrix and n is the total number of attention heads. Each attention head a_i is defined by the respective $\mathbf{k_i}$, $\mathbf{q_i}$ and $\mathbf{v_i}$ encodings as

$$\mathbf{a_i} = \text{Softmax}\left(\frac{\mathbf{q_i}\mathbf{k_i}^T}{c}\mathbf{v_i}\right) \tag{6.5}$$

Here, c is an arbitrary constant set in the model. This mechanism allows for the exchange of feature information between different track encodings with learnable parameters via the internal dot product $\mathbf{q_i}\mathbf{k_i}^T$. Each head has a different set of parameters, allowing for increased flexibility in expressing relationships. The resulting encoding h contains mixed information from across each track in the jet. After application of the Transformer layers, a pooling operation [141] is performed to build a global jet representation $\mathbf{h}^{\text{jet}}$ out of all individual track encodings $\mathbf{h}_i^{\text{track}}$ as

$$\mathbf{h}^{\text{jet}} = \sum_{k=1}^{n_{\text{tracks}}} \text{Softmax}\left(\text{Affine}\left(\mathbf{h}_i^{\text{track}}\right)\left(\mathbf{h}_i^{\text{track}}\right)^T\right) \tag{6.6}$$

From here, the model is split into three separate output components, which have different training objectives. They consist of a main graph classification task to identify jet flavor and two auxiliary tasks: a node classification task to identify the origin of individual tracks and an edge classification task to find SV connections. Each task network consists of multiple affine layers applied to combinations of the track and global jet representations from the

Jet Input	**Description**
p_T	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	**Description**
q/p	Track charge divided by momentum
$d\eta$	Pseudorapidity of the track, relative to the jet η
$d\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Closest distance from the track to the PV in the transverse plane
$z_0 \sin\theta$	Closest distance from the track to the PV in the longitudinal plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0)$	Lifetime signed longitudinal IP significance
n_{PixHits}	Number of pixel hits
n_{SCTHits}	Number of SCT hits
n_{IBLHits}	Number of hits in innermost pixel layer
n_{BLHits}	Number of hits in next to innermost pixel layer
$n_{\text{IBLShared}}$	Number of shared hits in innermost pixel layer
n_{IBLSplit}	Number of split hits in innermost pixel layer
$n_{\text{PixShared}}$	Number of shared pixel hits
n_{PixSplit}	Number of split pixel hits
$n_{\text{SCTShared}}$	Number of shared SCT hits
n_{PixHoles}	Number of pixel holes

Table 6.4: Summary of input variables for each track in GN2 adapted from [82].

Parameter	**Selection**
p_T	> 500 MeV
$\lvert\eta\rvert$	< 2.5
$\lvert d_0\rvert$	< 3.5 mm
$\lvert z_0 \sin\theta\rvert$	< 5 mm
n_{SCTHits}	≥ 8
$n_{\text{SCTShared}}$	≤ 1
n_{SCTHoles}	≤ 2
n_{PixHoles}	≤ 1

Table 6.5: Track selection used for inputs to GN2 [110].

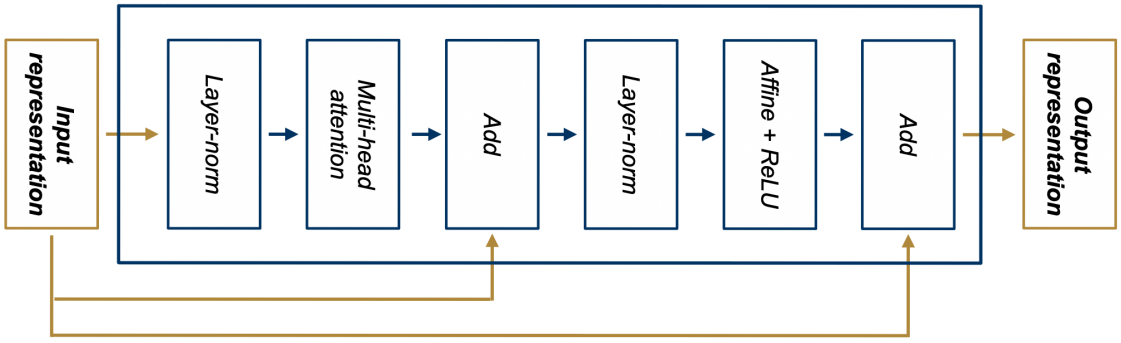


Figure 6.7: Diagram showing "Pre-LN" Transformer architecture as implemented in GN2 based on [220].

preceding Transformer and pooling layers. The main jet classification network is applied to the global representation only to classify each jet by flavor as originating from bottom, charm or light quarks or hadronic τ decays via the corresponding output probabilities p_b, p_c, p_u and p_τ. The latter represents a new addition to GN2 relative to the DL1 series of algorithms. All truth labels are derived via the delta-R association method described in section 4.2.1. A "Softmax" function defined by

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N_{\text{classes}}}(e^{x_j})} \tag{6.7}$$

is applied to the output to convert raw scores into probability distributions by ensuring all class probabilities sum to one. The node and edge classification networks are applied to the track representations directly, which are all extended by the concatenation of the full global jet representation. The node task classifies each track individually by its origin as summarized in table 6.6. The edge task classifies pairs of concatenated track representations to identify edges as either connecting two tracks that originate from a common vertex or not. This considers all vertices in the jet based on the spatial separation of track origin points including the primary event vertex, material interactions and long lived particle decays that do not originate from bottom or charm hadrons.

Optimization of the model occurs through backpropagation gradient descent [189] on a combined loss function given by $\mathcal{L} = \mathcal{L}_{\text{jet}} + \alpha \sum^{N_{\text{tracks}}} \mathcal{L}_{\text{track}} + \beta \sum^{N_{\text{edges}}} \mathcal{L}_{\text{edge}}$ with $\alpha = 0.5$ and $\beta = 1.5$. The latter two parameters indicate the relative importance of each loss component in the overall optimization and were tuned to maximize performance on the jet classification task. The jet and node losses are categorical cross-entropy $\mathcal{L}_{\text{jet/track}} = -\sum_{i=1}^{N_{\text{classes}}} y_i \log(\hat{y}_i)$ while the edge loss is binary cross-entropy simplified as $\mathcal{L}_{\text{edge}} = -\log(\hat{y}_{\text{vertex}})$. In these terms $\hat{y}_i$ class probability and y_i the binary truth label for class i. This truth label is "one-hot encoded", which means it is represented as a vector of length N_{classes} containing all zeroes except for a one in the spot that represents the true class. To balance class contributions in

Truth Origin	**Decscription**
Pileup	From a pp collision other than the primary interaction
Fake	Created from the hits of multiple particles
Primary	Does not originate from any secondary decay
fromB	From the decay of a b-hadron
fromBC	From a c-hadron decay, which itself is from the decay of a b-hadron
fromC	From the decay a c-hadron
OtherSecondary	From other secondary interactions and decays

Table 6.6: Truth origin categories for track classification auxiliary task adapted from [82].

the categorical loss functions, each term is weighted by n_{max}/n_i where n_{max} is the number of instances of the most represented class, to ensure classes are equally represented in the minimization if there is a statistical imbalance. The combined loss allows the two auxiliary tasks to encode relevant physics information related to the presence of SVs and the origins of each track into the common representation seen by all tasks, allowing for increased interpretability and model performance as presented in section 6.2.2 below. A detailed breakdown of the full model architecture is presented in table 6.7. Overall `GN2` consists of over 2.3 million trainable parameters.

The `GN2` algorithm is trained on jets simulated under both run 2 (i.e. $\sqrt{s} = 13$ TeV) and run 3 (i.e. $\sqrt{s} = 13.6$ TeV) conditions using `Powheg` [3], `Pythia` [196] and `EvtGen` [169], as described in table 6.8. Two physical processes are considered: $t\bar{t}$ with low momentum jets of $20\text{GeV} < p_T < 250\text{GeV}$ and Z' with high momentum jets covering $250\text{GeV} < p_T < 6000\text{GeV}$, where the latter is a hypothetical heavy resonance otherwise similar to a Z boson [168]. Nominal samples are used for training and evaluation while alternative samples are used for cross-checks. A total of 260.75 million jets are sampled for training: 186.25 million jets from $t\bar{t}$ made up of 45 million bottom- and charm-jets, 90 million light-jets and 6.25 million τ-jets as well as 74.5 million jets from Z' made up of 18 million bottom- and charm-jets, 36 million light-jets and 4.5 million τ-jets. A resampling procedure is applied to ensure p_T and η distributions are similar across all jet flavors.

`GN2` is trained with a batch size of 4000 jets using the ADAM optimizer [163] with a cosine annealing [173] based learning rate schedule over the course of 40 epochs, which is found to be enough for training and validation performance to start diverging. The best model parameters are chosen based on performance on a separate validation dataset. To combat overtraining, dropout [199], where a randomly selected 10% of all model parameters are set to zero per update step and weight decay [174], which reduces the magnitude of each parameter by a small percentage every time it is updated are also applied. All model training is done with the "salt" framework [13].

Component	Layers
Track Initializer (per track)	Affine (22 in, 256 out) + ReLU Affine (256 in, 256 out) + ReLU
Conditional Encoder	4×Transformer (256×n_{tracks} in, 256×n_{tracks} out, 8 heads) + ReLU Affine (256×n_{tracks} in, 128×n_{tracks} out) + ReLU
Graph Pooling	Pooling (128×n_{tracks} in, 128×1 out)
Jet Classification Network	Affine (128 in, 128 out) + ReLU Affine (128 in, 64 out) + ReLU Affine (64 in, 32 out) + ReLU Affine (32 in, 4 out) + Softmax
Track Origin Network (per track)	Affine (128 + 128 in, 128 out) + ReLU Affine (128 in, 64 out) + ReLU Affine (64 in, 32 out) + ReLU Affine (32 in, 8 out) + Softmax
Vertexing Network (per track pair)	Affine (128×2 + 128 in, 128 out) + ReLU Affine (128 in, 64 out) + ReLU Affine (64 in, 32 out) + ReLU Affine (32 in, 1 out) + Softmax

Table 6.7: Detailed summary of GN2 architecture with layer types and sizes.

6.2.2 GN2 Performance

Performance of the GN2 algorithm is quantified as described in section 4.2.2 via the bottom and charm discriminants D_b and D_c defined as

$$D_{b/c} = \log\left(\frac{p_{b/c}}{f_{c/b}p_{c/b} + f_\tau p_\tau + (1 - f_{c/b} - f_\tau)p_u}\right) \tag{6.8}$$

The f_x parameters are optimized as $\{f_c = 0.2, f_\tau = 0.05\}$ for D_b and $\{f_b = 0.3, f_\tau = 0.01\}$ for D_c to balance the individual background rejections. Different tunings of these parameters allow for preferencing the rejection of certain backgrounds over others. In this case, charm- (bottom-) and light-jet rejections are preferenced for bottom (charm) identification respective to τ-jet rejection.

GN2 model performance is compared in simulation against the DL1d algorithm, which is a successor to DL1r described in section 4.2.2. This version of DL1 includes inputs from DIPS [70], a "Deep Sets" [221] neural network that also uses constituent based inputs but does not have an attention mechanism. Similar to DL1r, it also does not include an explicit τ-jet output category, thereby limiting its ability to identify hadronic τ decays. Performance is studied under simulated run 3 conditions at $\sqrt{s} = 13.6$ TeV for jets containing no prompt

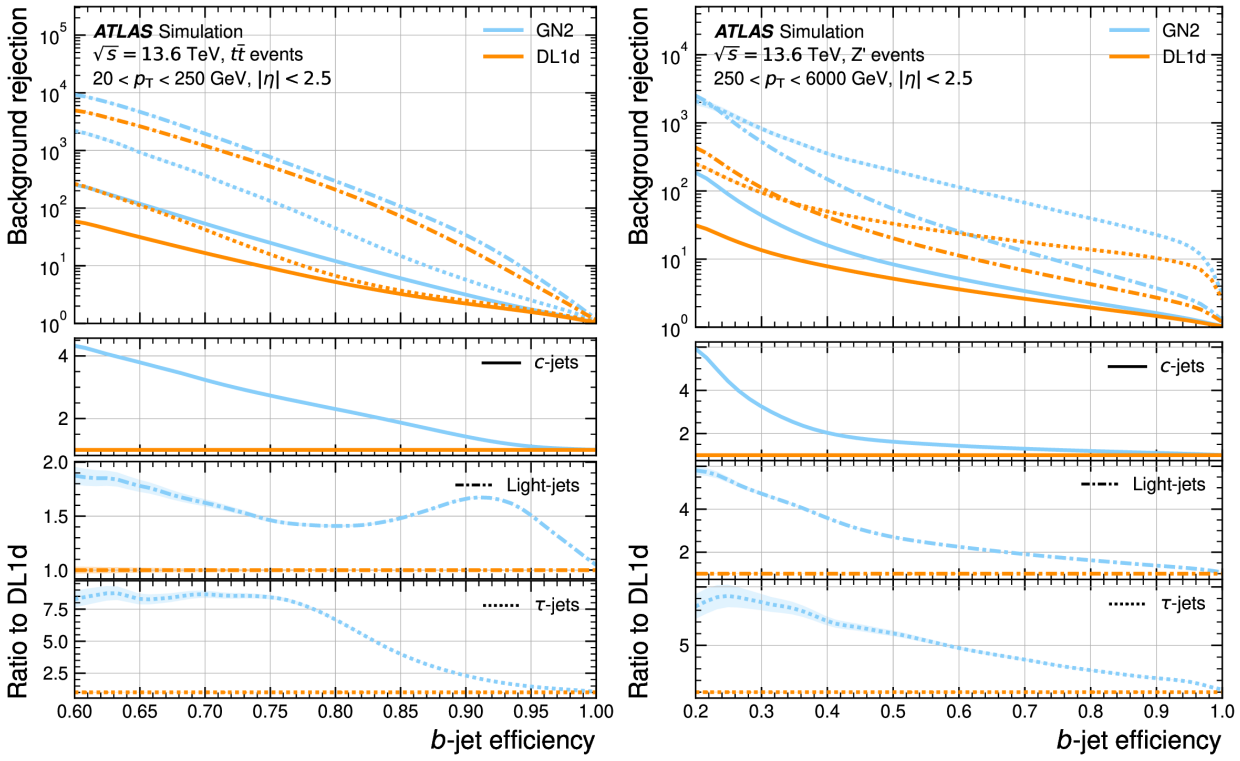
Process	Matrix element generator	Matrix element PDF	PS and hadronization	UE tune	HF quark decay
Nominal					
$t\bar{t}$ ($\sqrt{s} = 13$ TeV)	Powheg Box v2	NNPDF3.0NLO	Pythia 8.230	A14	EvtGen 1.6.0
$t\bar{t}$ ($\sqrt{s} = 13.6$ TeV)	Powheg Box v2	NNPDF3.0NLO	Pythia 8.308	A14	EvtGen 2.1.1
Z' ($\sqrt{s} = 13$ TeV)	Pythia 8.243	NNPDF2.3LO	Pythia 8.243	A14	EvtGen 1.7.0
Z' ($\sqrt{s} = 13.6$ TeV)	Pythia 8.308	NNPDF2.3LO	Pythia 8.308	A14	EvtGen 2.1.1
Alternate					
$t\bar{t}$ ($\sqrt{s} = 13$ TeV)	Powheg Box v2	NNPDF3.0NLO	Herwig 7.1.3	Default	EvtGen 1.6.0
$t\bar{t}$ ($\sqrt{s} = 13$ TeV)	Powheg Box v2	NNPDF3.0NLO	Herwig 7.2.1	Default	EvtGen 1.6.0
$t\bar{t}$ ($\sqrt{s} = 13$ TeV)	Sherpa 2.2.12	NNPDF3.0NNLO	Sherpa 2.2.12	Default	EvtGen 1.6.0

Table 6.8: Monte Carlo samples used for GN2 training (nominal only) and performance evaluation (nominal and alternative) [110].

leptons unless otherwise specified. ROC curves showing background rejections across the bottom-jet identification efficiency spectrum for the $t\bar{t}$ and Z' samples are presented in Fig. 6.8. At a 70% bottom-jet efficiency in the low momentum regime, GN2 achieves a charm-jet rejection of ~ 50, a light-jet rejection of ~ 2000 and a τ-jet rejection of ~ 350. This represents a factor of roughly 3.5, 1.5 and 8 improvement respectively compared to DL1d and about twice that compared to the bottom- and charm-rejections of DL1r. At high momentum, performance is generally worse with a charm-jet rejection of ~ 3, a light-jet rejection of ~ 10 and a τ-jet rejection of ~ 70 at the same 70% bottom-jet identification efficiency. Consistent improvements respective to DL1d and DL1r are also observed in this regime.

For charm-jet identification, similar results are shown in Fig. 6.9. At a 30% charm-jet efficiency in the low momentum $t\bar{t}$ sample, the GN2 bottom-jet rejection is ~ 20, the light-jet rejection ~ 300 and the τ-jet rejection ~ 60 which represents a factor of 1.5, 2 and 15 improvement over DL1d respectively. Comparing directly to DL1r for bottom- and light-jet rejection the overall improvements are around a factor of 2 and 6. As expected, performance is worse across the high momentum regime in the Z' sample with a bottom- and light-jet rejection of ~ 4 and ~ 20 at the same 30% charm-jet efficiency point. The exception is an improvement in the τ-jet rejection at ~ 200. In general, at low momentum light-jets are most easily distinguished from heavy flavor jets while at high momentum hadronic τ-jets are most characteristically different.

To emphasize the impact of the two auxiliary tasks on GN2 performance, a separate model was trained with a single jet flavor identification output component only. This GN2NoAux model is otherwise identical to GN2 and its bottom-jet identification performance is shown in Fig. 6.10. A background rejection degradation of between 10%-20% is observed compared to the nominal version of GN2 across the bottom-jet efficiency space at both high and low momentum. This indicates the power of the defined auxiliary tasks in encoding relevant physical information related to the presence of secondary decay vertices into the network



(a) (b)

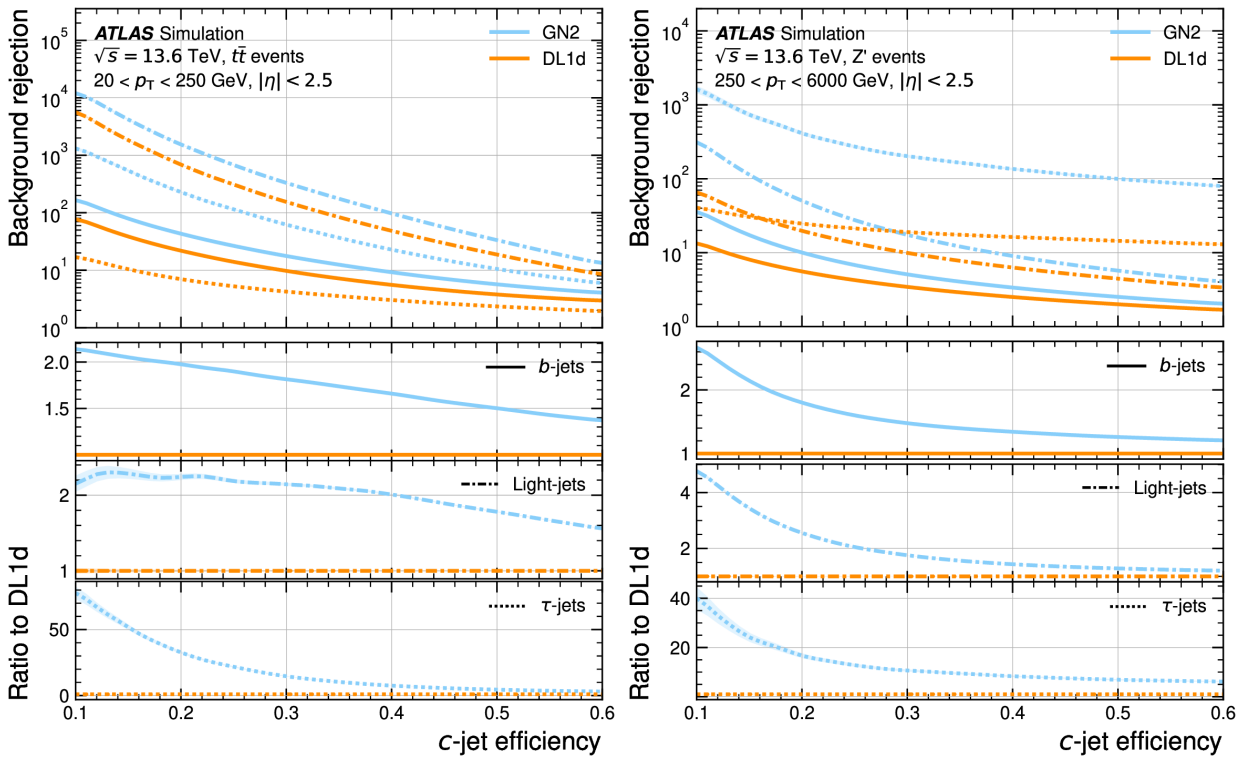
Figure 6.8: Plots showing background rejections vs. bottom-jet efficiency for `DL1d` and `GN2` evaluated on (a) $t\bar{t}$ and (b) Z' from [110].

during training. A third model called `GN2Lep` is also showcased in the same Fig. 6.10. This model is identical to `GN2` with both auxiliary tasks but includes an additional input feature indicating whether a particular track was used in the reconstruction of a lepton[8]. `GN2Lep` represents an initial attempt to introduce lepton features into the network, which is further discussed in section 6.3. The addition of this single feature improves background rejections in bottom-jet identification by up to 25% at low and up to 50% at high momentum.

`GN2` performance is also evaluated on alternate MC samples as listed in table 6.8 to ensure consistency across different generators. Four configurations simulated under run 2 conditions at $\sqrt{s} = 13$ TeV are tested, which cover modifications in parton showering algorithms as well as matrix element calculations. Fig. 6.11 shows the differences in background mistag rates[9] and signal efficiency as a function of p_T at the 77% bottom-jet efficiency working point (WP). Differences in bottom-jet efficiency per p_T bin deviate by less than 3% from their nominal values in each variation, indicating good agreement. For the charm- and light-jet mistag

[8]The "LeptonID" variable marks tracks that likely come from electrons or muons, which can be used to infer the presence of a heavy flavor hadron decay.

[9]Mistag rates are the inverse of background rejections.



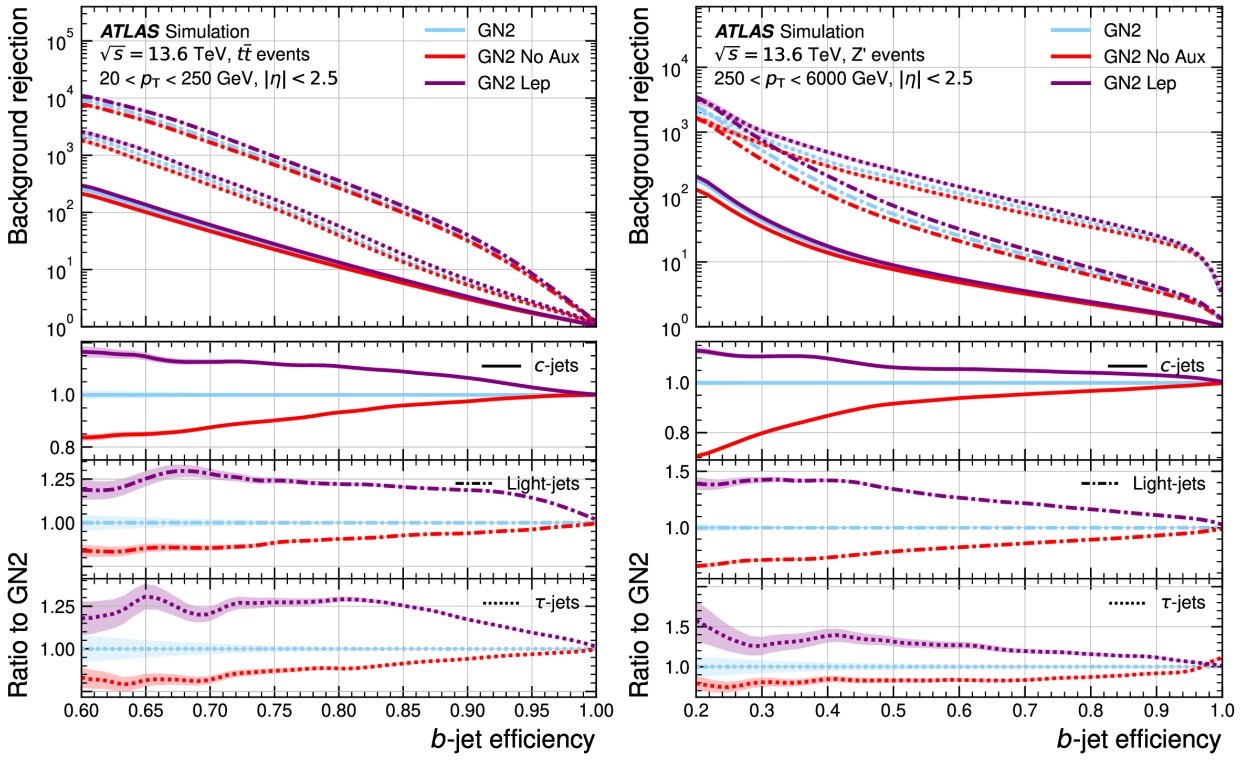
(a) (b)

Figure 6.9: Plots showing background rejections vs. charm-jet efficiency for `DL1d` and `GN2` evaluated on (a) $t\bar{t}$ and (b) Z' from [110].

rates, deviations can reach up to 20% on average and are highest at low jet momentum. Due to miniscule mistag rate values these deviations are not expected to have a large impact. Good agreement between `GN2` evaluation on different MC generators indicates a robust model that does not use information particular to any specific generator in making decisions about jet flavor.

Similar conclusions can be drawn by evaluating `GN2` performance on run 2 data relative to MC simulations via a set of calibration procedures as described in section 4.2.3. Results given in Fig. 6.12 indicate good agreement between `GN2` performance on MC and data within the bounds of calibration-based uncertainties, which is consistent with results from the previous flavor identification algorithm `DL1r`. The largest differences are observed for light-jet rejection values where deviations between MC and data performance in `GN2` exceed the 1σ calibration uncertainty band. Given the large light-jet rejection of `GN2` and corresponding small relative fraction of light-jets at most bottom-jet efficiencies, this is not expected to have a large impact overall. All deviations observed are corrected via "Scale Factors" (SFs) in practice and corresponding uncertainties are included in analyses.

Aside from the main jet flavor identification objective, `GN2` performance can also be



(a) (b)

Figure 6.10: Plots showing background rejections vs. bottom-jet efficiency for `GN2`, `GN2NoAux` and `GN2Lep` evaluated on (a) $t\bar{t}$ and (b) Z' from [110].

evaluated for the auxiliary training objectives. The track origin classification and vertex reconstruction tasks can be used in conjunction to reconstruct secondary decay vertices from bottom and charm hadrons explicitly. Edge scores from the vertexing task are first symmetrized between incoming and outgoing edges between two nodes by averaging the scores since these are not guaranteed to be equal but represent the same connection. A unique index is then assigned to each track which is updated iteratively using the following procedure until no more changes take place: all edges connected to a given node with a score of higher than 0.5[10] are considered and the index of the node is updated based on the minimum index found across its remaining connections. Once this procedure is complete, nodes with a common index are considered to originate from the same vertex. To filter out secondary vertices not originating from heavy flavor decays, scores from the track origin classification task are used. Each track is assigned a track origin category based on its highest probability output class. First, the single vertex containing the most tracks predicted to

[10]The score 0.5 is an arbitrary choice and represents a 50% probability assigned by the model that two tracks originate from the same vertex.

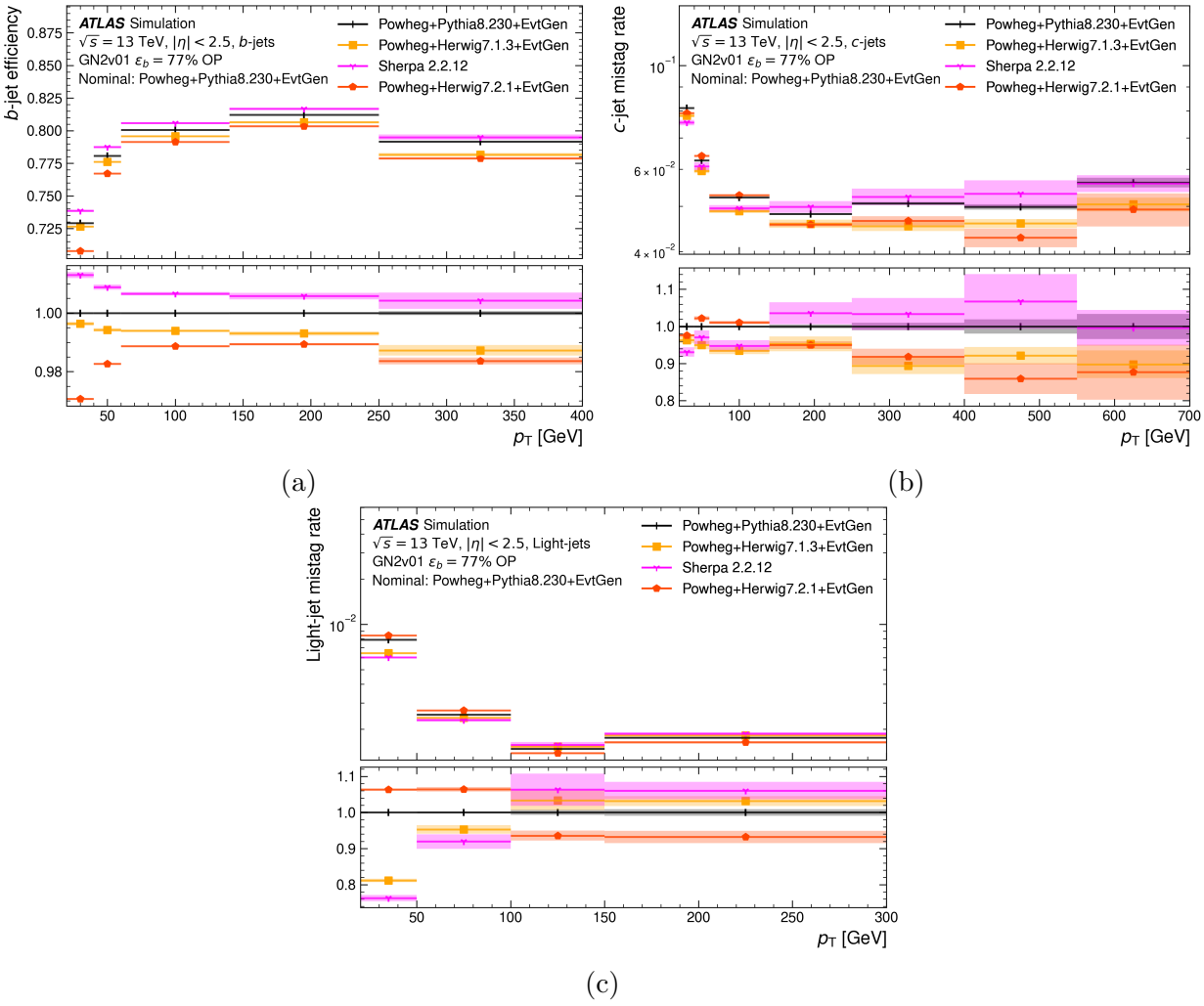


Figure 6.11: Plots showing (a) bottom-jet efficiency, (b) charm-jet rejection and (c) light-jet rejection vs. p_T at 77% bottom-jet efficiency including MC statistical uncertainty bands for GN2 evaluated on different alternate $t\bar{t}$ samples from [110].

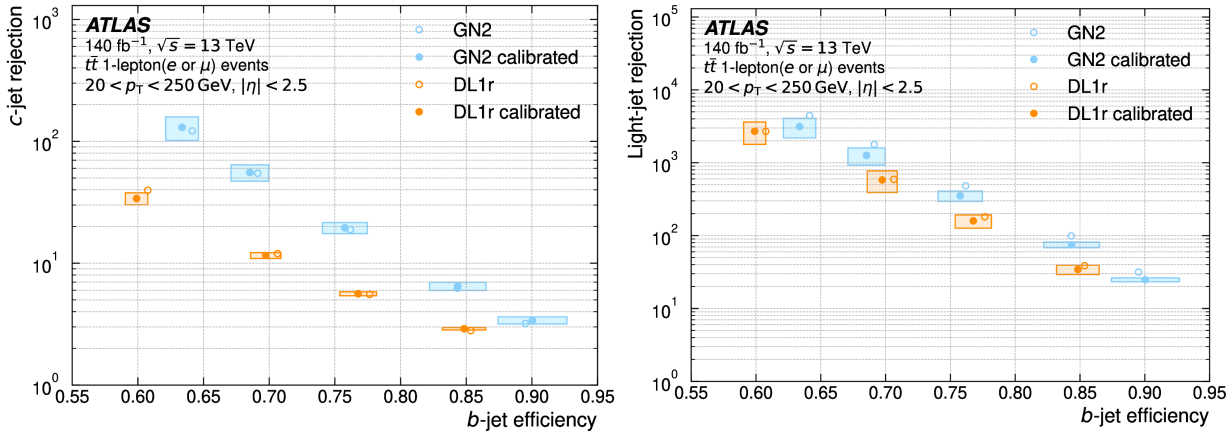
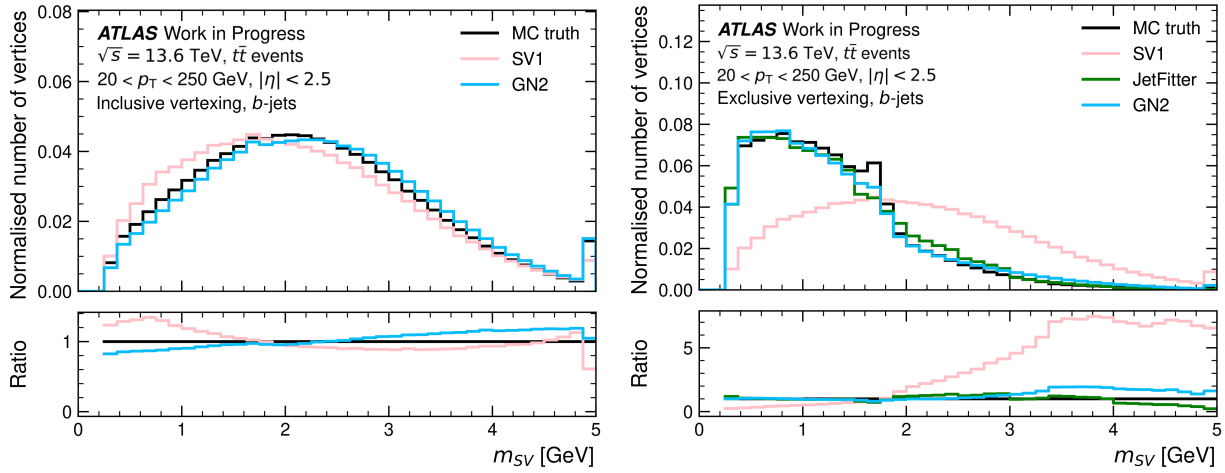


Figure 6.12: Plots showing (a) charm-jet and (b) light-jet rejection vs. bottom-jet efficiency in MC and calibrated in data (with corresponding uncertainties) for `GN2` and `DL1r` in semi-leptonic $t\bar{t}$ events from [110].

originate from the event primary vertex is removed before all vertices without tracks from bottom or charm hadrons (i.e. the "fromB", "fromC" and "fromBC" categories) are also taken out. In "inclusive vertexing" the remaining vertices are merged within each jet while "exclusive vertexing" allows for the existence of multiple distinct vertices.

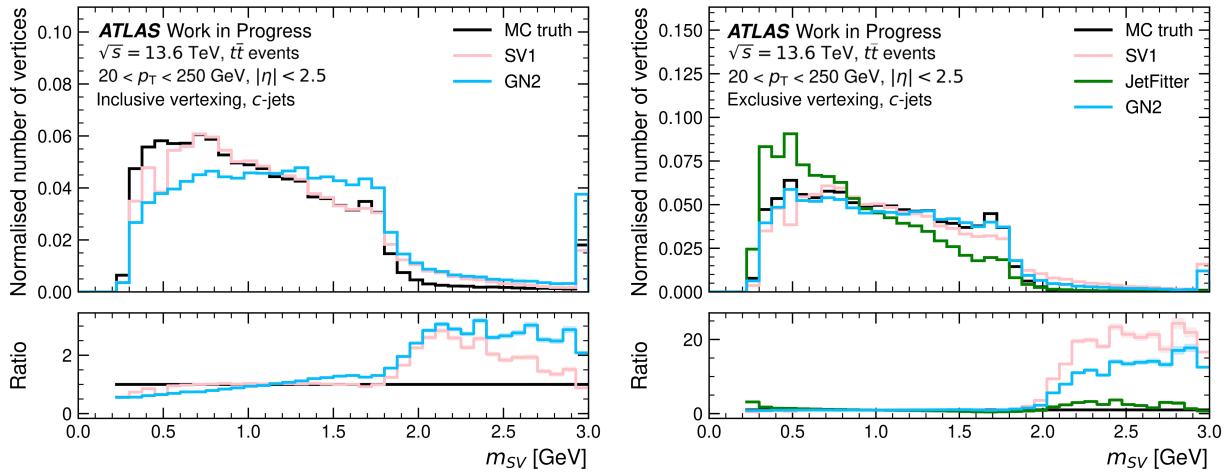
For each true and reconstructed vertex, a mass estimate can be calculated by assuming each associated track has the mass of a π^+, which are generally most prevalent. Comparisons between these mass values are shown for reconstructed bottom and charm vertices in the inclusive and exclusive vertex reconstruction cases in Figs. 6.13 and 6.14. `GN2` is compared against previously used algorithms `SV1` and `JetFitter` detailed in section 4.2.2. The inclusive vertexing strategy is most closely equivalent to `SV1`, given its single vertex limitation while the exclusive strategy enables fairer comparisons to `JetFitter`, which allows for reconstruction of multiple vertices per jet. In these comparisons, `GN2` shows improved agreement with MC truth for vertices in bottom-jets compared to `SV1` with the inclusive and `JetFitter` with the exclusive reconstruction strategy, with deviations from truth across almost all mass bins of less than 15%. In charm-jets the trends are similar for exclusive vertexing, but in the inclusive strategy `GN2` more frequently overestimates the true vertex mass compared to `SV1` leading to discrepancies of around 50% in the most affected bins at $m_{SV} < 2$ GeV. Nevertheless, mass estimation via `GN2` vertices shows improved performance across almost all regimes compared to both `SV1` and `JetFitter` with a much more flexible choice of reconstruction strategy.

To further study vertexing performance, a matching procedure between truth and reconstructed SVs is utilized. This procedure matches all reconstructed vertices to the corresponding truth vertex with the largest number of common tracks. If this criterion results in a tie, track association efficiency (or recall) and purity are used as tiebreakers. The track



(a) (b)

Figure 6.13: Plots showing m_{SV} in bottom-jets with the (a) inclusive and (b) exclusive vertexing strategy for MC truth, `GN2`, `SV1` and `JetFitter` (exclusive only) in semi-leptonic $t\bar{t}$ events.



(a) (b)

Figure 6.14: Plots showing m_{SV} in charm-jets with the (a) inclusive and (b) exclusive vertexing strategy for MC truth, `GN2`, `SV1` and `JetFitter` (exclusive only) in semi-leptonic $t\bar{t}$ events.

association efficiency is defined as the fraction of tracks present in the truth vertex that are also covered by the matched reconstructed vertex while the latter is the fraction of tracks in the reconstructed vertex that are also contained in the truth vertex[11].

Generally, the number of true SVs is expected to be one for bottom- and charm-jets[12] and zero for light-jets. Fig. 6.15 shows the total fraction of bottom- and charm-jets with reconstructed SVs after enforcing quality requirements of a 65% track association efficiency and a 50% track association purity from matching. With the inclusive vertexing strategy this results in around 85% of bottom-jets and 50% of charm-jets having `GN2` reconstructed SVs, compared to 75% and 40% for `SV1`. `SV1` vertex reconstruction also shows a greater dependence on jet p_T with particularly large differences between the algorithms observed at low p_T. The larger fraction of reconstructed vertices in heavy flavor jets with `GN2` does not result in a significantly higher number of falsely reconstructed vertices in light-jets as seen in Fig. 6.16. Both algorithms falsely identify SVs in light-jets roughly at rates between 2% and 9% depending on jet p_T. `GN2` performs worse in this respect at jet $p_T < 70$ GeV, which accounts for at least some of the observed differences between the two algorithms in heavy flavor vertex identification in this part of the phase space. At larger values of $p_T > 70$ GeV the trend inverts and `GN2` outperforms `SV1` across all jet flavors.

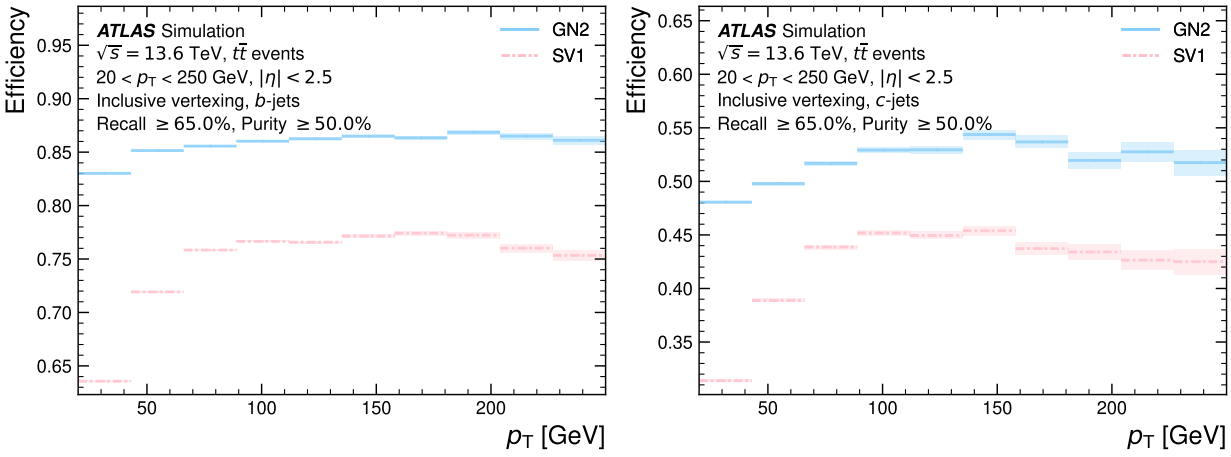
To examine overall vertex quality, Figs. 6.17 and 6.18 show track assignment efficiency and purity for all vertices reconstructed using the inclusive strategy in bottom- and charm-jets matched to truth vertices. Average assignment efficiency lies between 95-97% for `GN2` and 89-91% for `SV1` in both jet species across the p_T spectrum with identical trends as in Fig. 6.15. Assignment purity is higher for `SV1` compared to `GN2` at low p_T and vice versa at high p_T, with a lower inversion point for charm-jets than bottom-jets. The purity range is high overall at 92-94% for bottom-jets and 84-92% for charm-jets in `GN2` and 91-96% and 85-93% respectively in `SV1`.

6.2.3 Edge Features in `GN2`

The performance improvements observed in `GN2` respective to `DL1d` come from the ability of the Transformer-based algorithm to better exploit relational information between constituent tracks via its attention mechanism [215], which is the primary difference between the two architectures. This is amplified by the presence of auxiliary training objectives which indirectly encode relevant physics into the model's jet representation, leading to increased performance as demonstrated in Fig. 6.10. In the case of `GN2`, information between tracks is derived from kinematic features associated with each individual track by the algorithm itself, but no explicit relational information is provided. An alternate approach suggests feeding

[11] Thus a vertex with high efficiency and low purity contains many falsely associated tracks not present in the truth vertex while one with high purity and low efficiency is missing a large number of tracks that are present in the truth vertex.

[12] This is assuming the inclusive definition of a secondary vertex where bottom and subsequent charm vertices are merged given the prevalence of tertiary vertices in bottom-jets via bottom to charm hadron decays.



(a) (b)

Figure 6.15: Plots showing total fraction of (a) bottom- and (b) charm-jets with (correctly) identified secondary vertices vs. p_T after enforcing vertex quality criteria using inclusive vertexing in semi-leptonic $t\bar{t}$ events.

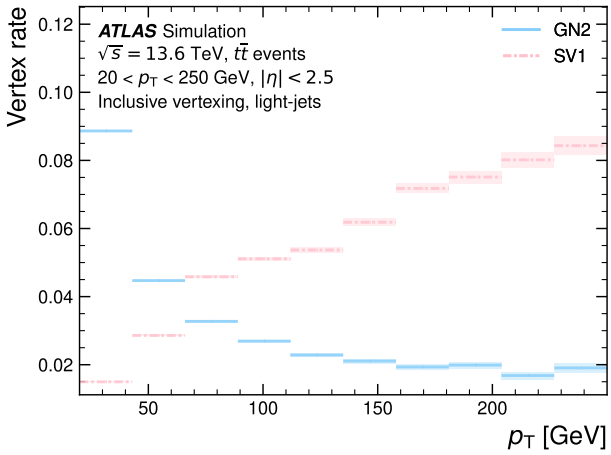
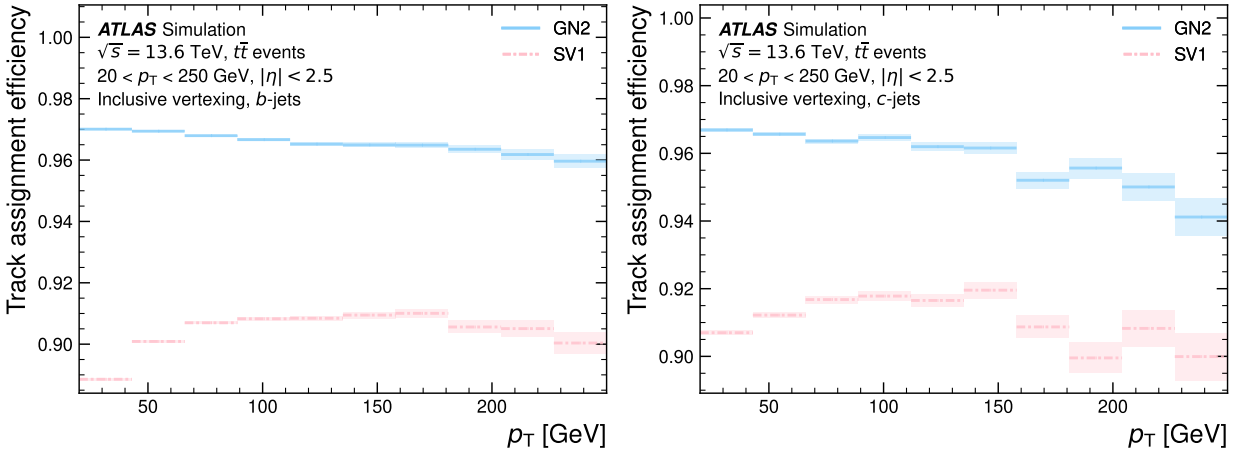
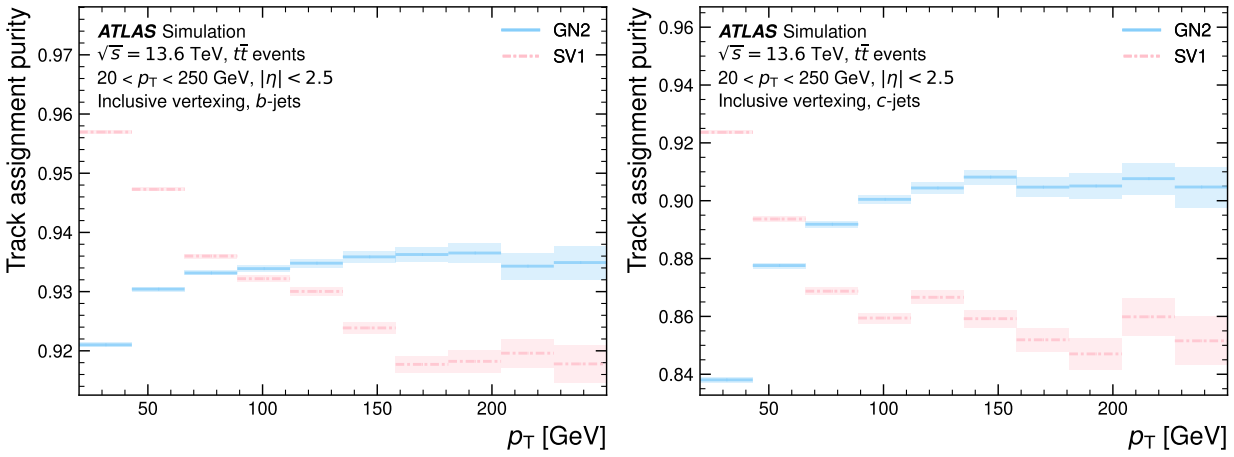


Figure 6.16: Plot showing total fraction of light-jets with (falsely) reconstructed secondary vertices vs. p_T using inclusive vertexing in semi-leptonic $t\bar{t}$ events.



(a) (b)

Figure 6.17: Plots showing track association efficiency for SVs in (a) bottom- and (b) charm-jets vs. p_T using inclusive vertexing in semi-leptonic $t\bar{t}$ events.



(a) (b)

Figure 6.18: Plots showing track association purity for SVs in (a) bottom- and (b) charm-jets vs. p_T using inclusive vertexing in semi-leptonic $t\bar{t}$ events.

some relational information known to be relevant a priori into the network as edge features directly. The auxiliary training objectives of GN2 have demonstrated that an emphasis on encoding physics knowledge enables better training convergence, even when no explicitly new information is provided to the algorithm. This philosophy is extended to the use of edge features, which could be computed from track kinematic information by the network itself. Features under consideration between pairs of tracks i and j are

$$\Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \tag{6.9}$$

$$\Delta k_{Tij} = \min(p_{Ti}, p_{Tj}) \Delta R_{ij} \tag{6.10}$$

$$\Delta z_{ij} = \frac{\min(p_{Ti}, p_{Tj})}{p_{Ti} + p_{Tj}} \tag{6.11}$$

as well as a simple binary flag that marks self-connecting edges. The standard "Pre-LN Transformer" implementation upon which GN2 is based does not allow for the introduction of explicit edge features natively. To facilitate this, some modifications are made to the existing attention mechanism described in section 6.2.1 for the GN2Edge model inspired by the "Edge-augmented Graph Transformer" [155]. Architectural differences between the two are summarized in table 6.9. First, all edge feature sets are encoded into a higher dimensional space via an edge initialization network. This consists of a series of affine transformations applied to all track pairs[13] and is identical in structure to the track initialization network. These encoded edge representations of shape $(n_{\text{tracks}} \times n_{\text{tracks}}, n_{\text{features}})$ are then transformed into $\mathbf{e_i}$ and $\mathbf{g_i}$ vectors shaped as $(n_{\text{tracks}} \times n_{\text{tracks}}, 1)$ via an additional affine layer for each attention head. Here, e_i is a bias term that is added to the attention score matrix and g_i is a scaling term that is restricted by a Sigmoid function. These terms derived from explicit edge features allow the model to emphasize relationships between tracks beyond the learned attention weights. The final implementation modifies each attention head output as

$$\mathbf{a_i} = \text{Softmax}\left(\frac{\mathbf{q_i}\mathbf{k_i}^T}{c}\mathbf{v_i} + \mathbf{e_i}\right) \text{Sigmoid}\left(\mathbf{g_i}\right) \tag{6.12}$$

All other model components are consistent between the GN2 and GN2Edge models. Edge feature information is updated in each Transformer layer via application of an additional affine layer to the attention scores. This provides the model with additional flexibility in representing relational information.

Due to storage constraints, edge features are computed on the fly using track feature information when a batch is loaded. Because of this, training time per epoch is around a factor of 4 larger for the base GN2Edge model than GN2. When viewed as a function of the number of training steps, optimization is observed to be faster for GN2Edge across the board,

[13] Track pairs are ordered and thus include forwards and backwards connections in addition to self-connections for a total of $n_{\text{track}} \times n_{\text{track}}$ edges.

Component	Layers
Edge Initializer (per track pair)	Affine (4 in, 32 out) + ReLU Affine (32 in, 32 out) + ReLU
Conditional Encoder	4×Transformer ($256\times n_{\text{tracks}} + 32 \times n_{\text{tracks}}^2$ in, $256\times n_{\text{tracks}}$ tracks out, 8 heads) + ReLU Affine ($256\times n_{\text{tracks}}$ tracks in, $128\times n_{\text{tracks}}$ tracks out) + ReLU

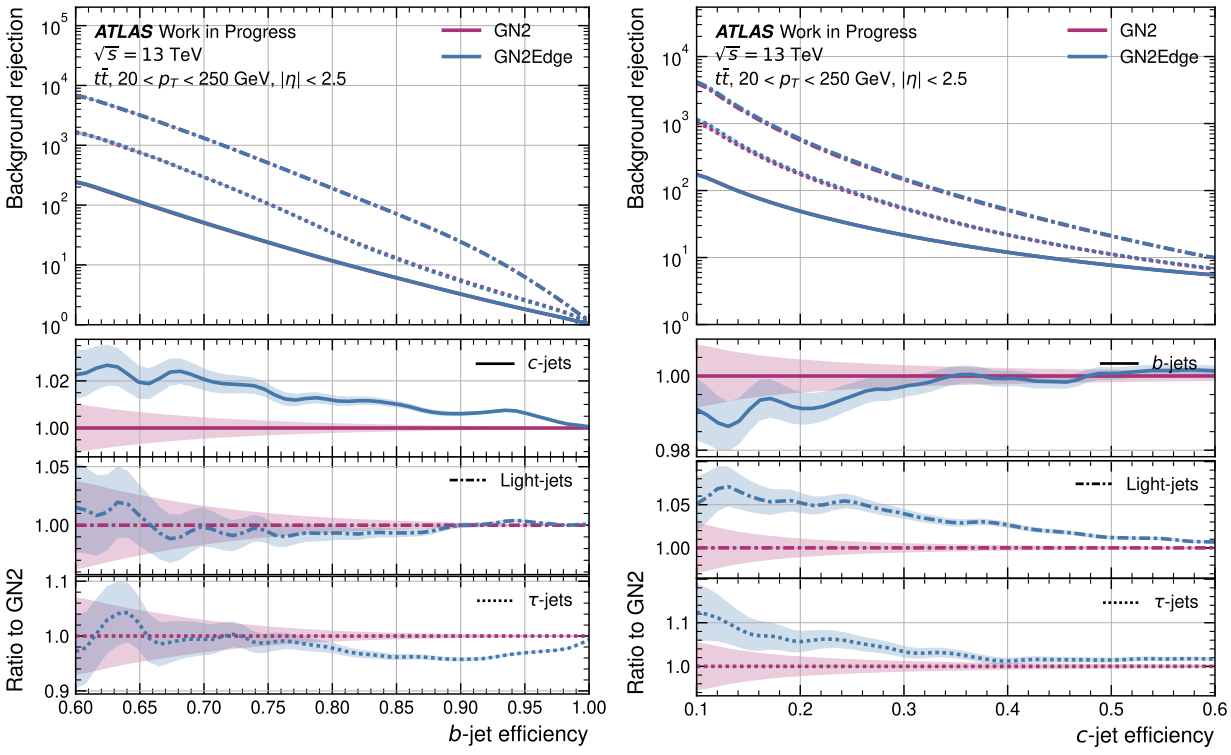
Table 6.9: Overview of `GN2Edge` architecture modifications relative to `GN2` as shown in table 6.7.

indicating the applied edge features contain information that facilitate convergence. The number of trainable parameters of `GN2Edge` does not differ meaningfully from `GN2` despite the presence of an additional initialization network, so differences in size are not expected to play a role.

To quantify performance differences between `GN2` and `GN2Edge`, characteristic bottom- and charm-jet identification ROC curves are shown in Fig. 6.19 for the low p_T $t\bar{t}$ sample. No significant differences are observed between the performance of the two models, indicating the nominal version of `GN2` does not benefit from the introduction of explicit relational features to any notable degree besides faster convergence. Given the relevant physical information contained in the newly introduced edge features, this suggests the model might learn these features implicitly given an adequate amount of training data. To test this hypothesis, two additional versions of `GN2` and `GN2Edge` were trained on smaller proportions of the original dataset. ROC curves shown for models trained on 10% of available data in Fig. 6.20 and on 1% in Fig. 6.21 show the benefit of introducing explicit relational features in scenarios where training data is scarce. In the most extreme case where the models are trained on only 2.6075 million jets, `GN2Edge` outperforms its counterpart by 50% in all background rejections at 70% bottom-jet efficiency and by 30-50% at 30% charm-jet efficiency. While availability of training data is not an issue in jet flavor identification, these results show the applicability of edge features in scenarios where training data is more difficult to generate.

6.3 `GN3`: Jet Charge Identification

`GN2` presents a powerful paradigm shift for jet flavor identification in the ATLAS experiment and is expected to greatly improve the sensitivity of run 3 measurements with hadronic final states in ATLAS. Since deployment of the algorithm, further developments have culminated in an expanded successor called `GN3`, which follows a similar design philosophy to its predecessor with a core Transformer component and multiple training objectives. While `GN2` only takes information from the inner detector into account in the form of particle tracks, `GN3` [107, 208] additionally incorporates the calorimeters and muon spectrometer via "Particle



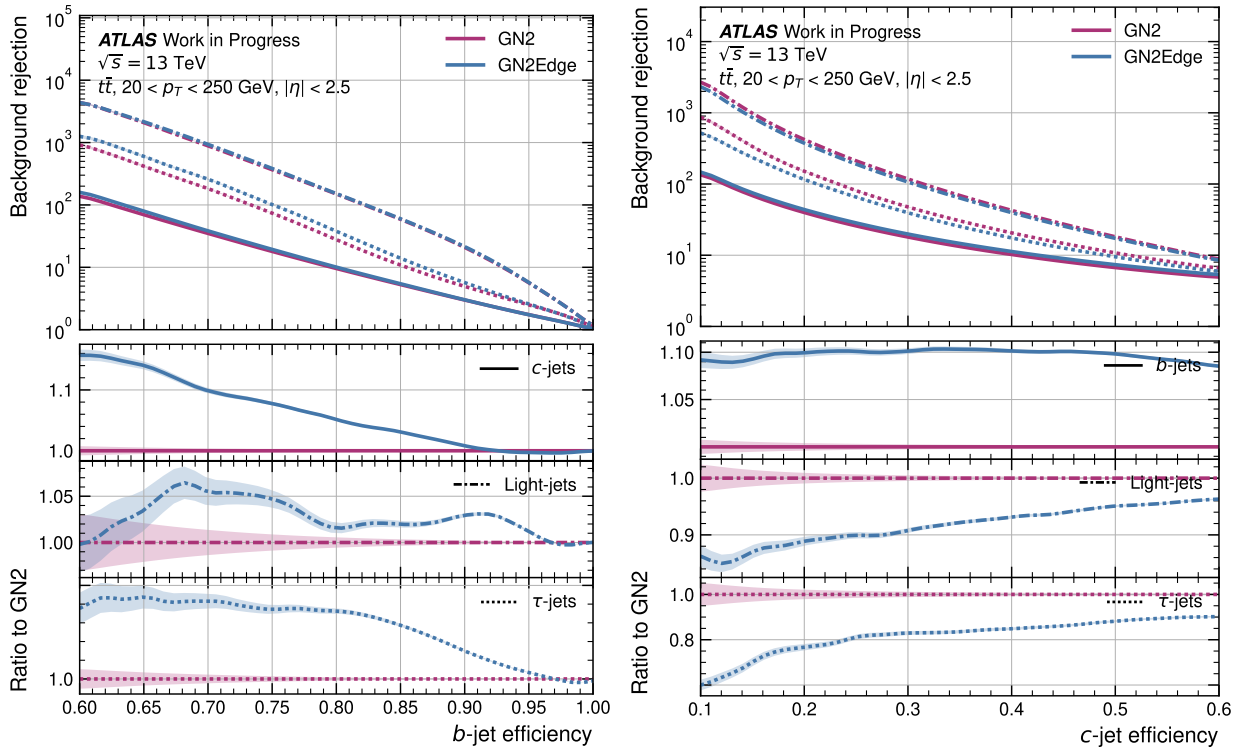
(a) (b)

Figure 6.19: Plots showing background rejections vs. (a) bottom-jet and (b) charm-jet efficiency for GN2 and GN2Edge evaluated on $t\bar{t}$ jets.

Flow" (PFlow) objects as well as reconstructed muons and electrons which are detailed in section 4.1. It is also significantly larger with five times as many trainable parameters at 11.9 million and a 60% larger training dataset of 417,750,000 jets. Furthermore, three additional auxiliary tasks are defined: jet p_T regression, track type classification and jet charge classification. The latter allows for the discrimination of b and $\bar{b}$ as well as c and $\bar{c}$ for the first time, which provides powerful new tools towards measurements of hadronic final states with well defined charge characteristics such as $H \to b\bar{b}$ and $H \to c\bar{c}$. This section introduces the architecture of GN3 relative to its predecessor and details its performance with an emphasis on the new charge identification capabilities.

6.3.1 The Architecture of GN3 and Jet Charge Identification

The core architecture of GN3 remains similar to its predecessor GN2. It follows the same overall structure consisting of an initialization network for each of its input constituents, a Transformer component that allows for the implicit extraction of relational information between constituents and an output network for each training task. Unless noted otherwise,

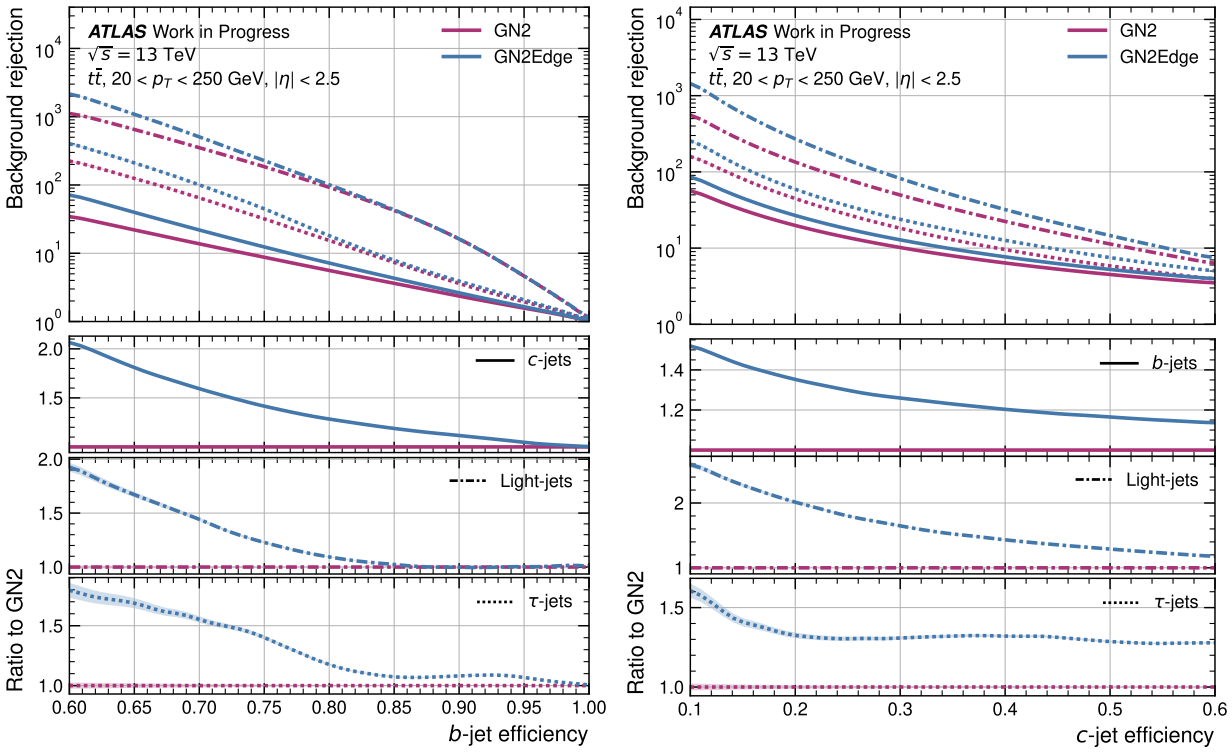


(a) (b)

Figure 6.20: Plots showing background rejections vs. (a) bottom-jet and (b) charm-jet efficiency for GN2 and GN2Edge trained on 10% of the nominal GN2 training data evaluated on $t\bar{t}$ jets.

the mathematical and conceptual details mirror discussions in section 6.2.1 and 6.1.1. GN3 uses "Sigmoid-Weighted Linear Unit" (SiLU) [158] activation functions defined by $\text{SiLU}(x) = x\sigma(x)$ (where $\sigma(x)$ is the Sigmoid function) throughout, which are smoother compared to the ReLU functions used in GN2. A visualization of the different components of GN3 is shown in Fig. 6.22.

The input constituents for GN3 consist of the same track objects used by GN2 with a slightly modified feature set and selection detailed in tables 6.10 and 6.11 as well as PFlow objects and reconstructed electrons. All track features from GN2 are carried over except n_{PixHoles} which is found to have a negligible impact on performance. New track features are exclusively related to leptons and include "lepton ID", which marks whether a track was used in the reconstruction of a "very loose" electron or a "loose" muon as well as four additional variables for tracks associated with reconstructed muons. Rather than keeping separate muon constituents, these features are decorated to tracks directly and filled with placeholder values when applicable. The total number of input tracks is increased to 50 for GN3 with corresponding looser track selection requirements compared to GN2 as summarized in table



(a) (b)

Figure 6.21: Plots showing background rejections vs. (a) bottom-jet and (b) charm-jet efficiency for `GN2` and `GN2Edge` trained on 1% of the nominal `GN2` training data evaluated on $t\bar{t}$ jets.

6.11 to increase the amount of information the algorithm sees. Track to jet matching is done via the ghost association method rather than delta-R to increase the number of matched tracks near the boundary of each jet.

Jet Input	**Description**
p_T	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	**Description**
q/p	Track charge divided by momentum
$d\eta$	Pseudorapidity of the track, relative to the jet η
$d\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Closest distance from the track to the PV in the transverse plane
$z_0 \sin\theta$	Closest distance from the track to the PV in the longitudinal plane
$\sigma(q/p)$	Uncertainty on q/p

$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0)$	Lifetime signed longitudinal IP significance
n_{PixHits}	Number of pixel hits
n_{SCTHits}	Number of SCT hits
n_{IBLHits}	Number of hits in innermost pixel layer
n_{BLHits}	Number of hits in next to innermost pixel layer
$n_{\text{IBLShared}}$	Number of shared hits in innermost pixel layer
n_{IBLSplit}	Number of split hits in innermost pixel layer
$n_{\text{PixShared}}$	Number of shared pixel hits
n_{PixSplit}	Number of split pixel hits
$n_{\text{SCTShared}}$	Number of shared SCT hits
leptonID	Electron and muon reconstruction label associated to the track
muonQuality	Muon reconstruction quality flag (Loose, Medium or Tight)
$\mathcal{M} = \frac{p_{ID} - \Delta E_{Cal} - P_{MS}}{\sigma_{\Delta E_{Cal}}}$	Imbalance of the muon three-momentum between the MS and ID
$\mathcal{R} = \frac{(q/p)_{ID}}{(q/p)_{MS}}$	Ratio of q/p measured independently in the MS and ID
$\mathcal{S} = q\sum_i^{n_{\text{hits}}-1} \frac{\Delta\phi_{i,i+1}}{\sigma_{\Delta\phi_{i,i+1}}}$	Azimuthal angular deviation between scattering centers belonging to a muon track

PFlow Input	**Description**
p_T	PFlow transverse momentum
E	PFlow energy
$d\eta$	Pseudorapidity of the PFlow object, relative to the jet η
$d\phi$	Azimuthal angle of the PFlow object, relative to the jet ϕ
isCharged	Boolean describing whether the PFlow object is charged or not

Electron Input	**Description**
p_T	Electron transverse momentum
E_T	Electron transverse energy
η	Pseudorapidity of the electron
ϕ	Azimuthal angle of the electron
d_0	Closest distance from the electron to the PV in the transverse plane
$\lvert s(d_0)\rvert$	Absolute transverse IP significance
n_{PixHits}	Number of pixel hits
n_{SCTHits}	Number of SCT hits
n_{IBLHits}	Number of hits in innermost pixel layer
p_T^{rel}	Electron transverse momentum projected on the plane perpendicular to the jet direction
p_T^{frac}	Fraction of jet transverse momentum represented by the electron
$\Delta p/p$	Ratio between momentum difference from the perigee to the last measurement point and the momentum at perigee
ΔR	Angular distance between the electron and the jet

R_{had_1}	Ratio of transverse energy in the first layer of the HCAL to that of the EM cluster (for $\|\eta\| < 0.8$ and $\|\eta\| > 1.37$)
R_{had}	Ratio of the transverse energy in the HCAL to that of the EM cluster (for $0.8 < \|\eta\| < 1.37$)
f_1	Ratio of the energy in the first layer to the total energy in the ECAL
f_3	Ratio of the energy in the third layer to the total energy in the ECAL
R_η	Ratio of the sum of energies in a 3×7 rectangle in $\eta\times\phi$ centered around the most energetic cell to that in a 7×7 rectangle
$w_{\eta 2}$	Lateral shower width $\sqrt{(\sum E_i\eta_i^2)/(\sum E_i) - ((\sum E_i\eta_i)/(\sum E_i))^2}$, where sums are over cells in a window of 3×5 cells
R_ϕ	Ratio of the sum of energies in a 3×3 rectangle in $\eta\times\phi$ centered around the most energetic cell to that in a 3×7 rectangle
w_s^{tot}	Total shower width $\sqrt{(\sum E_i(i-i_{\max})^2)/(\sum E_i)}$, where sums are over cells in a window of $\Delta\eta\sim0.0625$ and $i_{\max}$ is the index of the highest energy cell
ΔE_s	Energy difference between the second highest energy cell and the lowest energy cell between the first and second highest energy cells
E_{ratio}	Ratio of the energy difference and sum between the first and second highest energy cells
eProbabilityHT	Likelihood probability based on transition radiation in the TRT
$\Delta\eta_1$	Pseudorapidity difference between the cluster position in the first layer of the ECAL and the extrapolated track
$\Delta\phi_{\text{res}}$	Azimuthal angle difference between the cluster position in the second layer of the ECAL and the extrapolated track multiplied by the charge
E/p	Ratio of the cluster energy to the track momentum
$p_T^{\Delta R_{\leq 0.3}}$	Scalar sum of the transverse momentum of tracks around the electron within a radius of $\Delta R = \min(10/\text{p}_\text{T}, 0.3)$

Table 6.10: Summary of input variables for each constituent in GN3 adapted from [107, 208].

Parameter	**Selection**
p_T	> 500 MeV
$\|\eta\|$	< 2.5
$\|d_0\|$	< 5 mm
$\|z_0\sin\theta\|$	< 5 mm
n_{SCTHits}	≥ 8
n_{SCTHoles}	≤ 1
n_{PixHoles}	$= 0$

Table 6.11: Track selection used for inputs to GN3 adapted from [107].

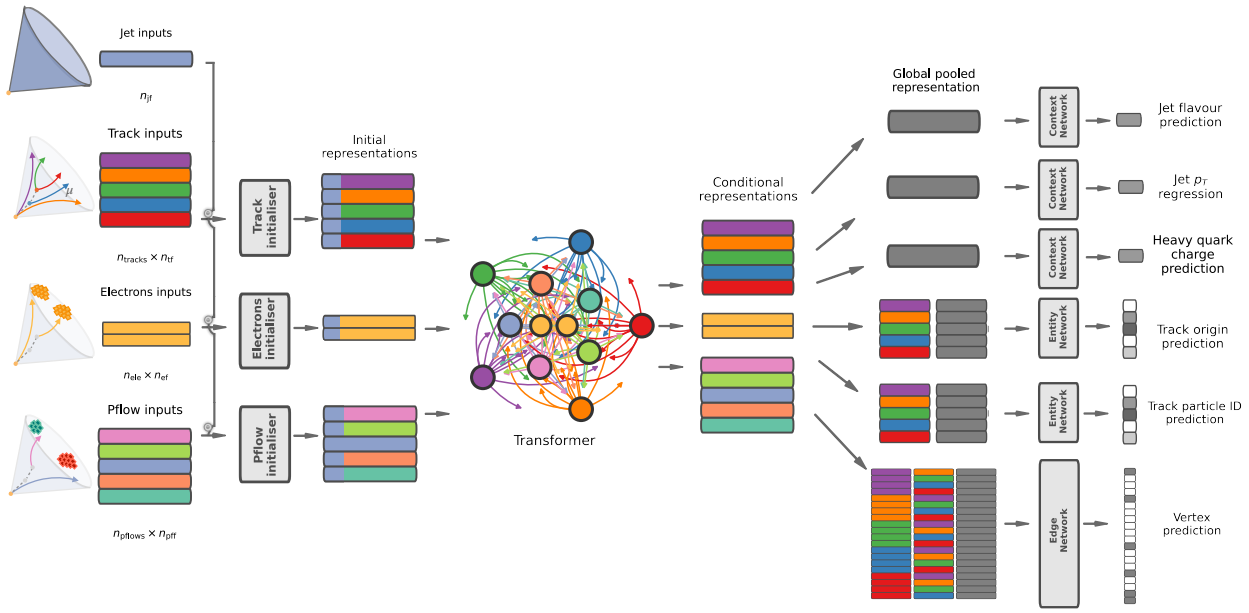


Figure 6.22: Diagram showing the different components of the GN3 algorithm from [208].

In addition to tracks, up to 50 PFlow objects and 10 reconstructed electrons are considered. All object types have separate initialization networks that allow for flexible encoding of their disparate associated feature sets. PFlow objects, which capture calorimeter information, are reconstructed using tracks and calorimeter clusters and consist of charged and neutral components. A total of 5 features are included for each PFlow object, mostly related to kinematic and charge information as detailed in table 6.10. Due to the nature of their reconstruction, some information overlap between tracks and charged PFlow objects is expected. No additional overlap removal processes or selection criteria beyond those detailed in section 4.1.1 are applied for mitigation given the ability of Transformer models to capture this overlap. PFlow objects that are matched to an electron or muon tracks are vetoed entirely due to the more elaborate object reconstruction processes for these objects. Reconstructed electrons meeting a $p_T > 1$ GeV requirement are the final constituent type. They have 28 associated features related to their kinematics and low level reconstruction, which are listed in table 6.10. Jet features are decorated to each constituent separately. Of all new input features, "soft lepton" variables[14] are expected to be the most powerful in identifying heavy flavor jets due to the frequent presence of leptons in heavy flavor decays, with about 40% of bottom-jets containing at least one [107].

[14] "Soft leptons" are distinguished from "prompt leptons" which originate directly from the primary vertex through vector boson or Higgs decays.

After application of the three initialization networks to the three constituent types, encoded representations of all constituents are concatenated resulting in a total of 110 objects passed as inputs to the Transformer Encoder, which is modified respective to GN2 and denoted "TransformerV2". The self-attention mechanism described in section 6.2.1 is replaced by an equivalent but more efficient flash attention mechanism [130] and the affine transformation is replaced by a "Gated Linear Unit" (GLU) [194] defined as

$$h = \text{Affine}\left(\text{SiLU}(\text{Affine}(\mathbf{x}))\text{Affine}(\mathbf{x})\right) \tag{6.13}$$

In addition, 8 learnable "register tokens" [131] are provided to the Transformer, which function as additional constituents with blank feature sets that can be used by the network to aggregate global context information during training without attaching it to pre-existing constituents. The global jet representation is built via pooling in the same fashion as described in section 6.2.1 and is used in the output task heads alongside the track representations built by the Transformer Encoder.

The output task structure is functionally identical between the two algorithms, but with six different task heads in GN3 instead of three. The two main auxiliary tasks (i.e. track origin classification and vertex reconstruction) are carried over unmodified from GN2. The main jet flavor classification task is extended to include six classes via splitting of the light-jet category into up/down-jet (p_{ud}), strange-jet (p_s) and gluon-jet (p_g) categories based on the MC truth record. In general, the light jet probability $p_u = p_{ud} + p_s + p_g$ is used for performance evaluation. The ghost association method is used to determine jet flavor for GN3 rather than delta-R association.

Three new auxiliary tasks are added to GN3: jet p_T regression, track type classification and jet charge identification. The former predicts the ratio of true jet p_T to reconstructed jet p_T to provide an estimate for jet momentum misreconstruction. It uses the global jet representation and is optimized via minimization of an L1 loss function defined for each jet as $\mathcal{L}_{p_T} = |y - \hat{y}|$, where y is the true ratio and $\hat{y}$ the model prediction. Track type classification functions the same as track origin classification but predicts a tracks associated truth particle. It uses six categories to group tracks as originating from pions, kaons, electrons/positrons, muons/antimuons, other charged hadrons or unmatched particles. The final task is jet charge classification, which mirrors the main jet classification task in its implementation but instead aims to separate bottom-jets, antibottom-jets, charm-jets, anticharm-jets and "other" jets[15] from each other. Corresponding output class probabilities are denoted p_b^{charge}, $p_{\bar{b}}^{\text{charge}}$, p_c^{charge}, $p_{\bar{c}}^{\text{charge}}$ and $p_{\text{other}}^{\text{charge}}$. Truth labels for this task are derived from the highest momentum bottom or charm hadron present in a bottom- or charm-jet respectively. The charge task classifies the heavy flavor quark content of these particles at the point of their production. Ambiguous cases with $b\bar{b}$ or $c\bar{c}$ quark content, such as J/Ψ or η_b are infrequent but included in the "other" category. A full summary of the GN3 architecture is provided in table 6.12.

All jet and track classification tasks use categorical cross-entropy loss functions as in GN2. The combined loss is no longer built as a weighted sum but instead as a geometric mean

[15]The "other" category includes light-jets and τ-jets as well as some ambiguous edge cases.

Component	**Layers**
Track Initializer (per track)	Affine (24 in, 512 out) + SiLU Affine (512 in, 512 out) + SiLU
PFlow Initializer (per PFlow object)	Affine (5 in, 512 out) + SiLU Affine (512 in, 512 out) + SiLU
Electron Initializer (per electron)	Affine (28 in, 512 out) + SiLU Affine (512 in, 512 out) + SiLU
Conditional Encoder	4×TransformerV2 (512×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{electrons}} + n_{\text{registers}}$) in, 512×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{electrons}} + n_{\text{registers}}$) out, 8 heads) + SiLU Affine (512×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{electrons}} + n_{\text{registers}}$) in, 256×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{electrons}} + n_{\text{registers}}$) out) + SiLU
Graph Pooling	Pooling (256×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{electrons}} + n_{\text{registers}}$) in, 256×1 out)
Jet Flavor Network	Affine (256 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 6 out) + Softmax
Jet Charge Network	Affine (256 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 5 out) + Softmax
Jet p_T Network	Affine (256 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 1 out) + SiLU
Track Origin Network (per track)	Affine (256 + 256 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 8 out) + Softmax
Track Type Network (per track)	Affine (256 + 256 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 6 out) + Softmax
Vertexing Network (per track pair)	Affine (256×2 + 256 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 1 out) + Softmax

Table 6.12: Detailed summary of `GN3` architecture with layer types and sizes.

of task losses via the Geometric-Loss Strategy (GLS) [32] as $\mathcal{L} = \left(\prod_{i=1}^{N_{\text{tasks}}} \mathcal{L}_i\right)^{1/N_{\text{tasks}}}$. This strategy balances task losses according to their gradients without the need for manual weight parameter optimization. The minimization algorithm used in GN3 is the Lion optimizer [31] with a 1-cycle learning rate schedule [197]. Like its predecessor, GN3 utilizes dropout and weight decay to optimize convergence. It is trained with a lower batch size of 1000 jets due to a higher memory footprint from the increased number of total constituents.

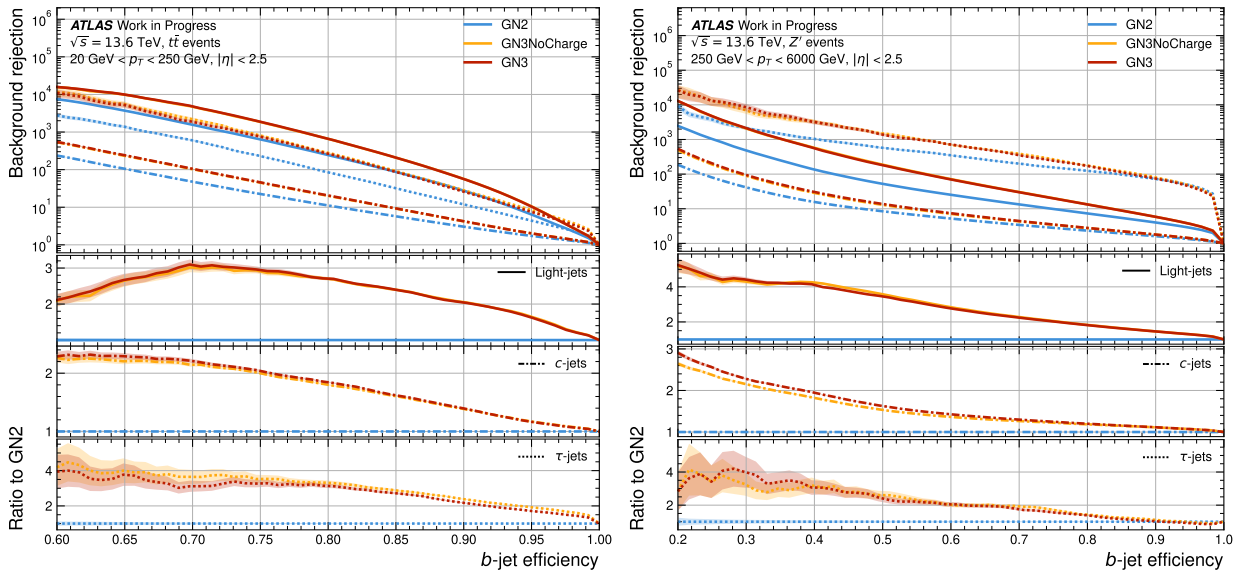
The training dataset of the GN3 algorithm includes jets simulated under run 2 (i.e. $\sqrt{s} =$ 13 TeV) and run 3 (i.e. $\sqrt{s} = 13.6$ TeV) conditions for both $t\bar{t}$ ($20\text{GeV} < p_T < 250\text{GeV}$) and Z' ($250\text{GeV} < p_T < 6000\text{GeV}$) processes from the same underlying nominal samples as GN2 described in table 6.8. A total of 417.75 million jets are sampled for training: 278.5 million jets from $t\bar{t}$ made up of 100 million bottom-jets, 27.5 million charm-jets, 140 million light-jets (split as 55 million up/down-jets, 20 million strange-jets and 65 million gluon-jets) and 11 million τ-jets as well as half as many Z' jets from each category. Like its predecessor, a resampling procedure is applied to the GN3 training dataset to ensure all jet flavors have similar p_T and η distributions.

6.3.2 GN3 Performance

GN3 performance is quantified separately for both flavor and charge classification. The former follows a similar procedure as described in section 6.2.2 with discriminants given by the same equation 6.8. Only the sum of light-jet probabilities $p_u = p_{ud} + p_s + p_g$ is considered in these calculations and the f_x parameters are re-optimized as $\{f_c = 0.2, f_\tau = 0.1\}$ for D_b and $\{f_b = 0.2, f_\tau = 0.1\}$ for D_c. Unless otherwise specified, all results are evaluated using MC simulated jets under run 3 conditions at $\sqrt{s} = 13.6$ TeV. For flavor identification performance, three models are considered: the main GN3 model, GN3NoCharge which is a version identical to GN3 without the charge identification task and GN2 as described in section 6.2. Since GN2 and GN3 use different truth label definitions based on delta-R and ghost association respectively, comparisons are only made between jets where the labels derived from both schemes match.

Efficiencies and background rejections for bottom-jet identification are shown for low p_T $t\bar{t}$ and high p_T Z' jets in Fig. 6.23. GN3 provides a charm-jet rejection of ~ 100, a light-jet rejection of ~ 5000 and a τ-jet rejection of ~ 2000 at 70% bottom-jet identification efficiency. This presents an improvement of factors of 3, 2 and 4 respective to GN2. In the high momentum regime, background rejections are worse overall at the same bottom-jet efficiency. GN3 shows a charm-jet rejection of ~ 4, a light-jet rejection of ~ 30 and a τ-jet rejection of ~ 350, which correspond to improvements of factors of 1.5, 2 and 2. The two GN3 based models show no significant difference in flavor identification performance, which is expected given the conceptual orthogonality of the two objectives. The lack of differences further suggests an absence of optimization issues from the increased model complexity due to the additional output task head.

Similar results are presented for charm-jet identification in $t\bar{t}$ and Z' jets in Fig. 6.24. At



(a) (b)

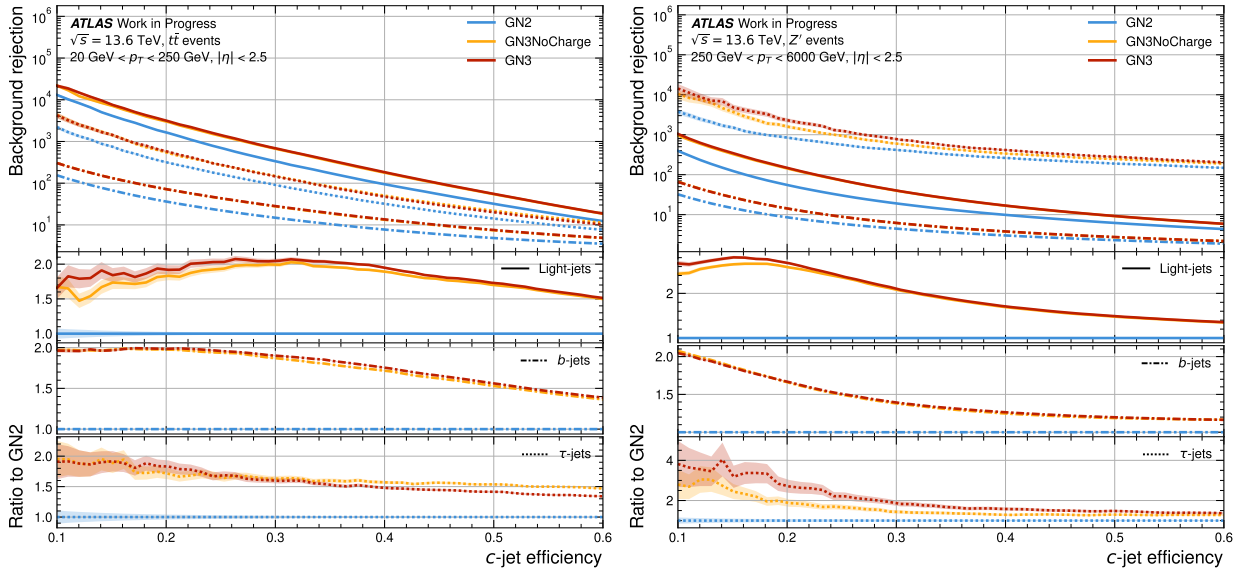
Figure 6.23: Plots showing background rejections vs. bottom-jet efficiency for `GN2`, `GN3NoCharge` and `GN3` evaluated on (a) $t\bar{t}$ and (b) Z'.

the 30% charm-jet efficiency point at low momentum, the bottom-jet rejection on the test sample is observed to be ~ 30, the light-jet rejection ~ 700 and the τ-jet rejection ~ 150. `GN3` thereby improves background rejections by factors of between 1.5 and 2 for each respective to its predecessor. In the Z' sample, performance is degraded overall with a corresponding bottom-jet rejection of ~ 6, a light-jet rejection of ~ 30 and a τ-jet rejection of ~ 700 at the same charm-jet efficiency. As in bottom-jet identification, no notable differences between the `GN3` and `GN3NoCharge` models are observed.

Charge identification with `GN3` requires the definition of a separate set of discriminants using the charge classification output probabilities. In particular, one discriminant is defined to separate jets from b and those from $\bar{b}$ quarks and one to separate jets from c and those from $\bar{c}$ quarks as

$$\mathcal{D}_{b/c}^{\text{charge}} = \log\left(\frac{p_{b/c}^{\text{charge}}}{p_{\bar{b}/\bar{c}}^{\text{charge}}}\right) \tag{6.14}$$

The numerator is chosen to represent q rather than $\bar{q}$, which is an arbitrary choice. Resulting charge discriminant distributions for bottom- and charm-jets are presented in Figs. 6.25 and 6.26. These histograms are normalized respective to the total number of bottom- and charm-jets across both q and $\bar{q}$ categories. Imbalances between q and $\bar{q}$ populations are not observed at a significant level. In general, charge identification is notably more difficult in the Z' sample, where a greater overlap in $\mathcal{D}_{b/c}^{\text{charge}}$ between q and $\bar{q}$ induced jet populations

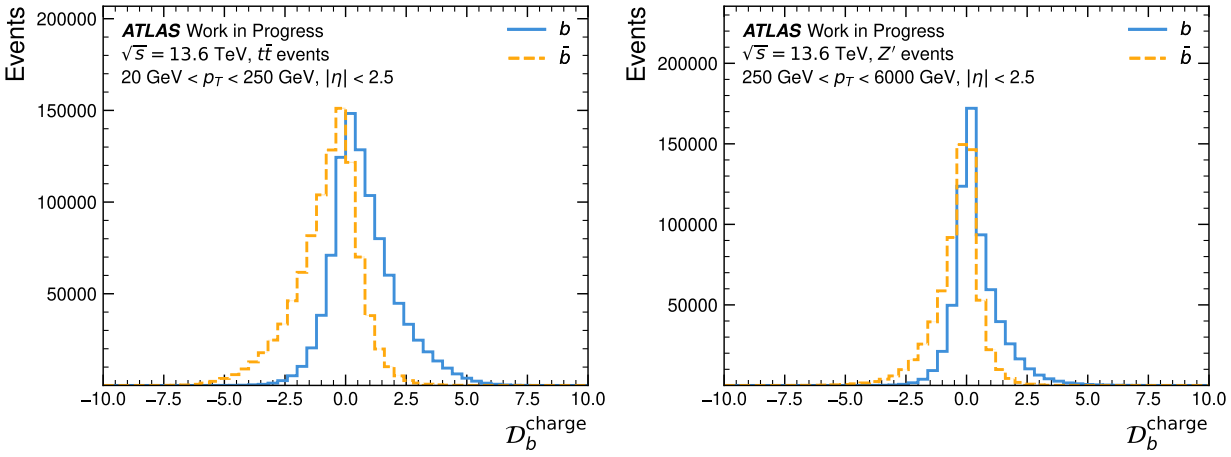


(a) (b)

Figure 6.24: Plots showing background rejections vs. charm-jet efficiency for `GN2`, `GN3NoCharge` and `GN3` evaluated on (a) $t\bar{t}$ and (b) Z'.

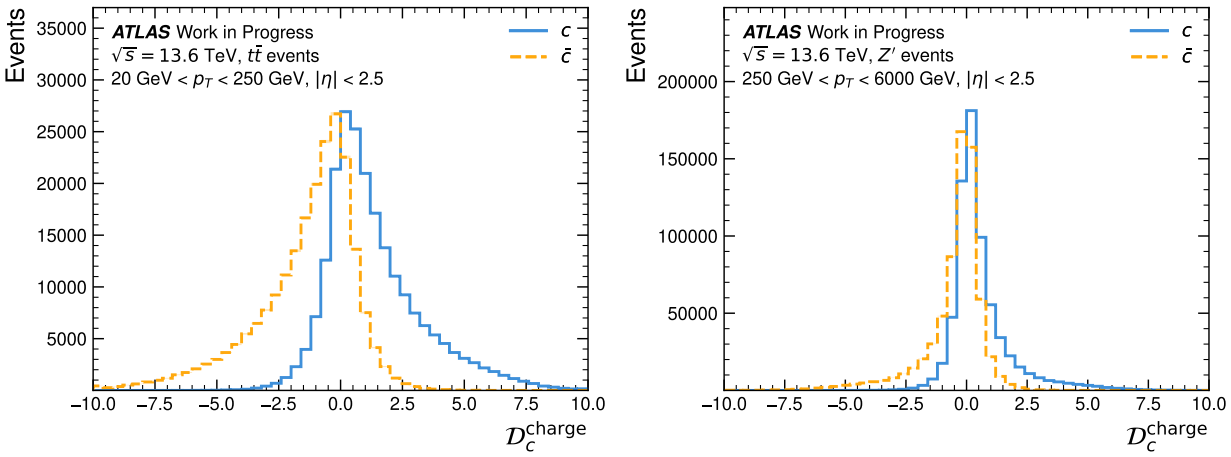
can be seen for both bottom and charm. In the $t\bar{t}$ sample, c and $\bar{c}$ show a greater degree of separation than b and $\bar{b}$, which runs contrary to results from flavor identification where charm-jets are generally more difficult to identify than bottom-jets. A large reason for this is the higher complexity of bottom hadron decays due to the presence of tertiary vertices, which makes them more easily identifiable. This can present a disadvantage when it comes to charge identification due to the higher number of charged particle tracks originating from each decay compared to prompt charm-jets.

The charge discriminants are built to optimally separate heavy flavor jets originating from the same quark type by their charge without providing any discrimination power against jets from other categories. This means, the application of a $\mathcal{D}_b^{\text{charge}}$ or $\mathcal{D}_c^{\text{charge}}$ cut is most powerful when preceded by a selection based on the corresponding flavor identification discriminant D_b or D_c. To indicate the impact of contamination from other jet flavors in an analysis scenario, a series of $\mathcal{D}_b^{\text{charge}}$ and $\mathcal{D}_c^{\text{charge}}$ distributions after the application of successively looser cuts on D_b or D_c corresponding to certain bottom- and charm-jet identification efficiencies are shown in Figs. 6.27 and 6.28. These distributions are normalized to the total number of jets passing a particular cut across all categories and utilize MC event weights for each jet. In bottom-jets, the distribution shapes don't change significantly between the 70% and 85% WPs, indicating the challenge of reconstructing bottom-jet charge even in cases where a jet's flavor is easily identifiable. In charm-jets, differences in distribution shapes are notable between all applied flavor tagging selections. Jets passing tighter WP cuts show more



(a) (b)

Figure 6.25: Histograms showing $\mathcal{D}_b^{\text{charge}}$ for GN3 evaluated on true bottom-jets in (a) $t\bar{t}$ and (b) Z'.



(a) (b)

Figure 6.26: Histograms showing $\mathcal{D}_c^{\text{charge}}$ for GN3 evaluated on true charm-jets in (a) $t\bar{t}$ and (b) Z'.

separated c and $\bar{c}$ populations, suggesting relatively straightforward charge identification for obvious charm-jets.

Charge identification results can be summarized in ROC curves as presented in Fig. 6.29, which show $\bar{q}$ inefficiency[16] as a function of q efficiency. As noted previously, the difference in charge identification power between bottom- and charm-jets is evident at low momentum. Selecting a sample with 50% of all b or c induced jets rejects around 85% of $\bar{b}$ induced jets and 95% of $\bar{c}$ induced jets respectively at low momentum and 80% of both at high momentum.

To isolate the momentum dependence of charge identification further, Figs. 6.30 and 6.31 show q efficiency variations and $\bar{q}$ rejection at a fixed 50% charge identification efficiency (calculated separately for $t\bar{t}$ and Z') as a function of jet p_T. While overall charge identification efficiency remains fixed, bin-by-bin variations show more stable efficiency across the p_T spectrum for bottom-jets compared to charm-jets, which have slightly higher variations at the extreme ends. In the $t\bar{t}$ sample, $\bar{q}$ rejection is universally higher by a factor of around two for charm-jets than for bottom-jets, regardless of p_T. This indicates previously observed performance differences do not stem from kinematic differences between bottom- and charm-jets. In Z', the differences are less pronounced but $\bar{q}$ rejection remains larger for charm-jets up to 2 TeV where the trend reverses.

Further insight into charge identification performance can be gained by separating results based on the species of heavy flavor hadron used to define the jet flavor and jet charge labels[17]. The three most important hadron species and their corresponding excited states are considered for both bottom- and charm-jets as detailed in Fig. 6.32. Only mesons are investigated further due to their higher occurrence frequencies and longer lifetimes compared to baryons as detailed in section 4.2.1. The jets corresponding to these particles are classified based on their heavy flavor quark composition at production, which is $u\bar{b}$ for B^+, $d\bar{b}$ for B^0, $s\bar{b}$ for B_s^0, $c\bar{u}$ for D^0, $c\bar{d}$ for D^+ and $c\bar{s}$ for D_s^+ (as well as corresponding excited states). Of these, the bottom mesons B^0 and B_s^0 have the tendency to oscillate between their particle and anti-particle states with oscillation frequencies defined by $\Delta m_d = 0.516 \pm 0.009$ for B^0 [111] and $\Delta m_s = 17.7683 \pm 0.0060$ for B_s^0 [123]. Because of this, reconstructing the jet charge as defined relative to the initial hadron state is expected to be less feasible for neutral bottom mesons, particularly for the rapidly oscillating B_s^0.

ROC curves showing relationships between meson species dependent $\bar{q}$ inefficiency and q efficiency are presented for the low momentum $t\bar{t}$ sample in Fig. 6.33. Across both bottom- and charm-jets, charge identification is most performant for jets containing charged heavy flavor mesons as compared to neutral ones. The rapidly oscillating B_s^0 shows the worst separation overall with a nearly random charge selection as expected. The best separation is achieved between D^{*+} and D^{*-}, which is also the only species where differences between base and excited states are observed. This difference stems from the presence of $D^{*+} \to \pi^+ D^0$ and $D^{*-} \to \pi^- \bar{D}^0$ decay modes where the jet charge can be characterized via the pion directly

[16]The inefficiency is defined as 1 - efficiency for a particular background or signal.

[17]As noted previously, in cases where multiple heavy flavor hadrons are present in a single jet, the one with the highest momentum is chosen.

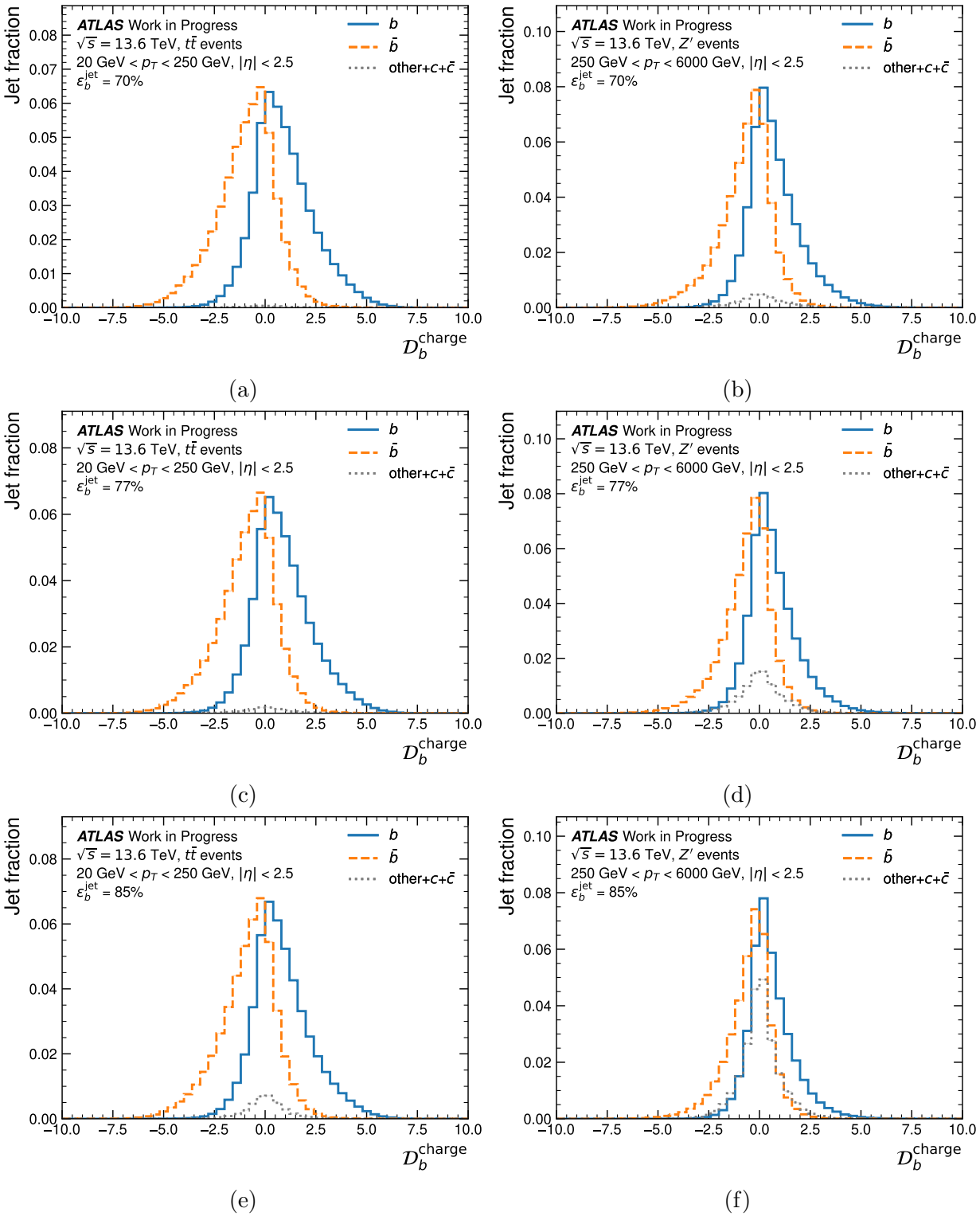


Figure 6.27: Histograms showing $\mathcal{D}_b^{\text{charge}}$ for GN3 evaluated on bottom-jets passing (a,b) 70%, (c,d) 77% and (e,f) 85% ϵ_b^{jet} WPs for (a,c,e) $t\bar{t}$ and (b,d,f) Z'. All WPs are defined respective to the $t\bar{t}$ sample and MC event weights are applied.

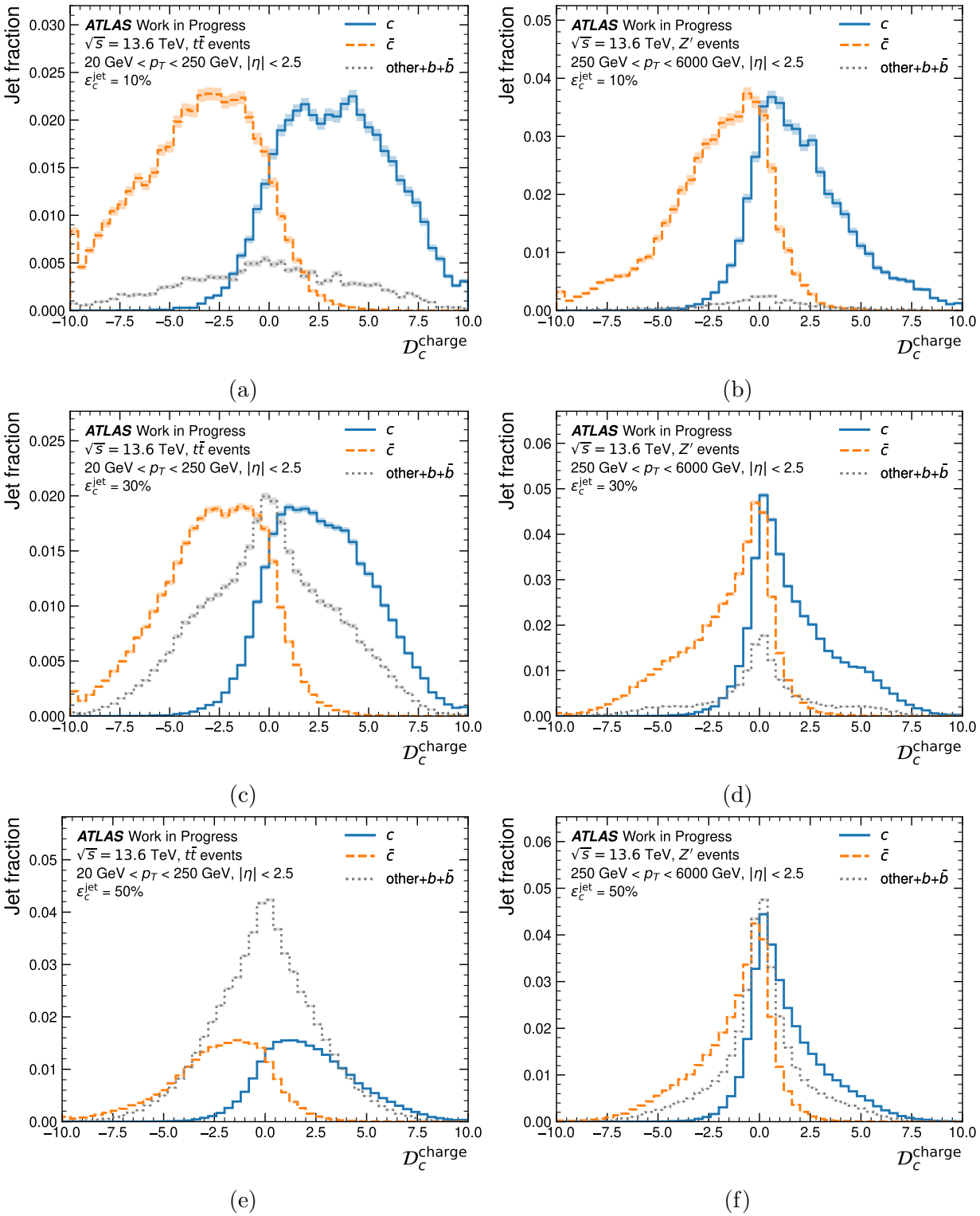
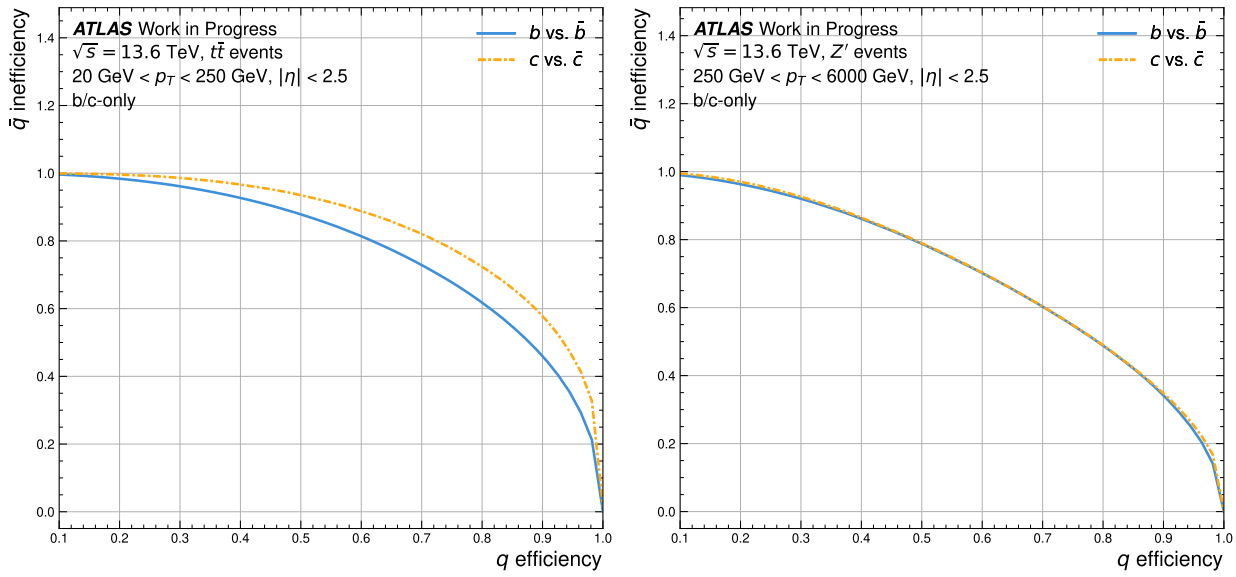
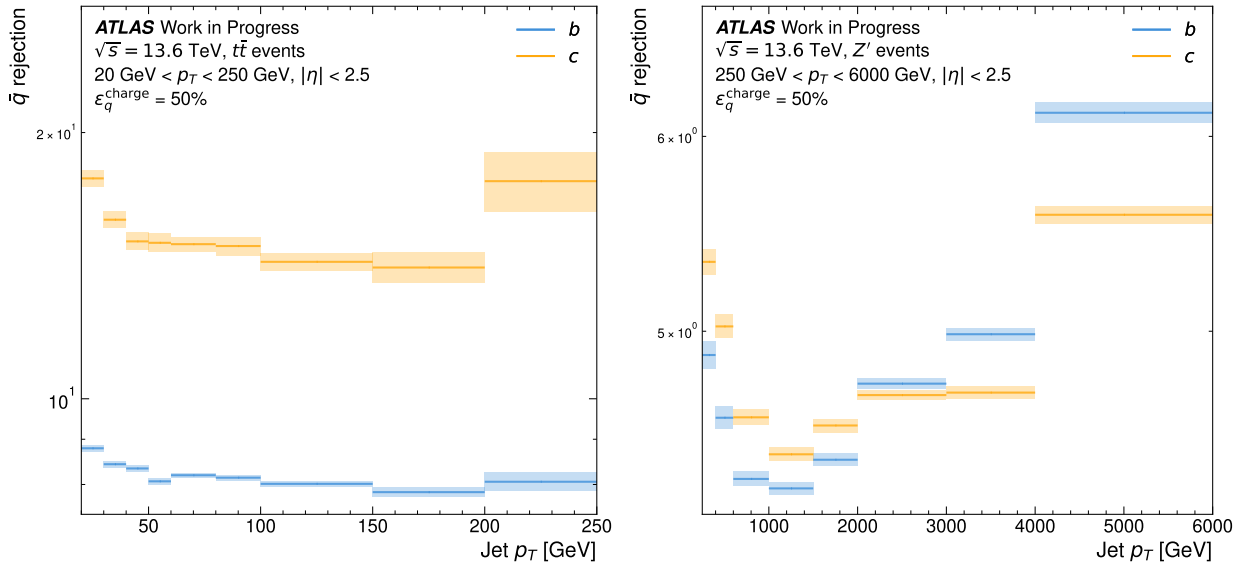


Figure 6.28: Histograms showing $\mathcal{D}_c^{\text{charge}}$ for GN3 evaluated on charm-jets passing (a,b) 10%, (c,d) 30% and (e,f) 50% ϵ_c^{jet} WPs for (a,c,e) $t\bar{t}$ and (b,d,f) Z'. All WPs are defined respective to the $t\bar{t}$ sample and MC event weights are applied.



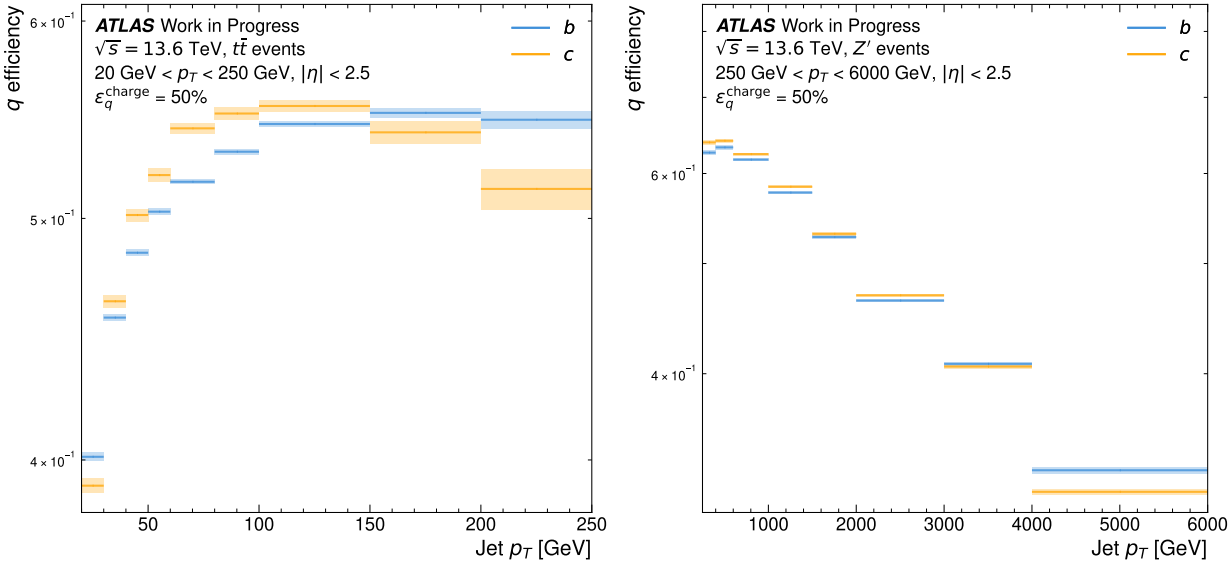
(a) (b)

Figure 6.29: Plots showing $\bar{q}$ inefficiency vs. q efficiency evaluated on (a) $t\bar{t}$ and (b) Z'.



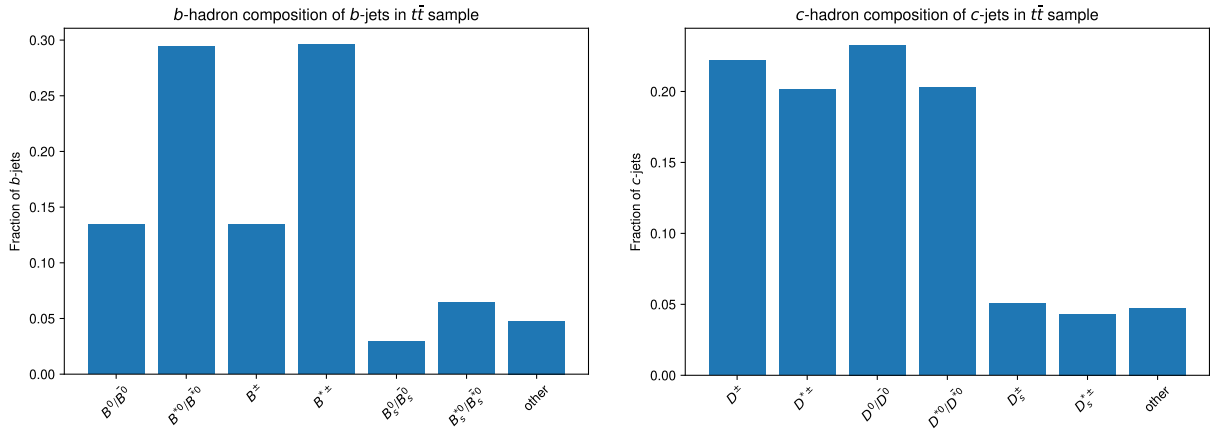
(a) (b)

Figure 6.30: Plots showing q efficiency variations vs. jet p_T at a fixed global $\epsilon_q^{\text{charge}} = 50\%$ with binomial uncerta on (a) $t\bar{t}$ and (b) Z'.



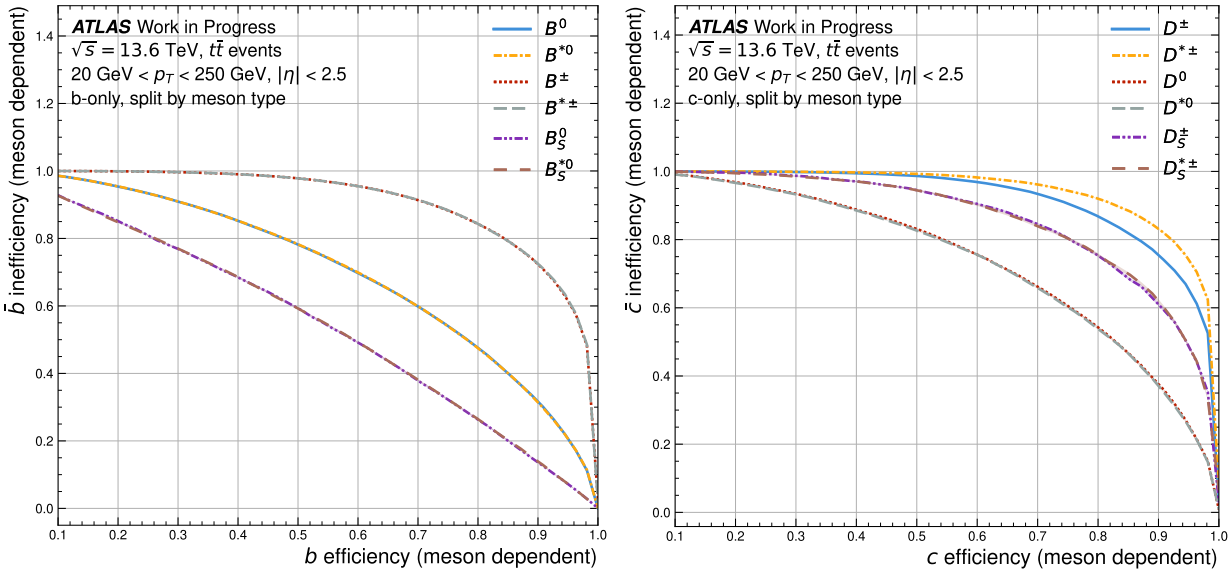
(a) (b)

Figure 6.31: Plots showing $\bar{q}$ efficiency variations vs. jet p_T at a fixed global $\epsilon_q^{\text{charge}} = 50\%$ with binomial uncertainties on (a) $t\bar{t}$ and (b) Z'.



(a) (b)

Figure 6.32: Occurrence frequency of different hadron species in (a) bottom-jets and (b) charm-jets across the $t\bar{t}$ sample.



(a) (b)

Figure 6.33: Plots showing $\bar{q}$ inefficiency vs. q efficiency split by meson species for (a) bottom-jets and (b) charm-jets. All efficiency and inefficiency values are calculated respective to jets containing the meson species in question only.

rather than through the subsequent D^0 decay. The impact of these decays on the overall $D^{*\pm}$ separation can be seen in the $\mathcal{D}_{b/c}^{\text{charge}}$ distributions for different decay modes in Fig. 6.34. From these results it is apparent that GN3 relies information from charge correlated pions in these decays. No other decay modes of excited states provide any such information that could be used to infer jet charge beyond the subsequent base state decays.

GN3 provides large improvements in jet flavor identification over its predecessor GN2 on MC simulated jets and introduces jet charge classification to the model. In conjunction, output information from GN3 presents a promising view of the future of measurements such as $VH(\rightarrow c\bar{c})$. Jet charge classification can be used in this context to further suppress cc and $\bar{c}\bar{c}$ backgrounds and potentially increase measurement sensitivity.

6.4 GN2X and GN3X: Extension to the Boosted Regime

All prior discussions on jet flavor and charge relate to the classification of individual small-R jets only, which are detailed in section 4.1.4. In the context of identifying $H(\rightarrow b\bar{b})$ and $H(\rightarrow c\bar{c})$ decays, this involves the search for two individual bottom- or charm-jets whose combined structure resembles the desired decay. The same flavor identification methodology can be extended to the high momentum (boosted) regime, where heavy resonance decays are collimated enough that they can be reconstructed within a single large-R jet. GN2X and its

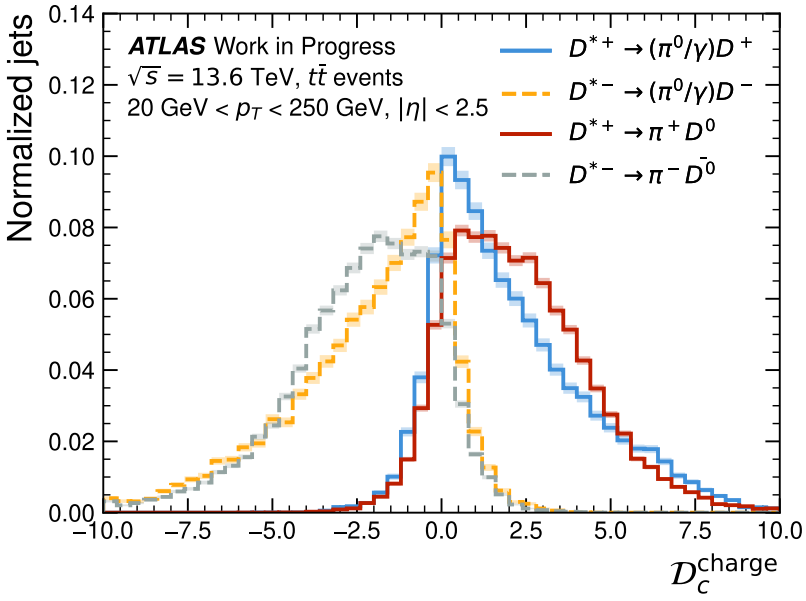


Figure 6.34: Histogram showing $\mathcal{D}_c^{\text{charge}}$ for true charm-jets containing $D^{*\pm}$ mesons split by decay mode. Each distribution is normalized individually to emphasize shape differences.

successor GN3X use the technology of GN2 and GN3 to directly identify $H(\to b\bar{b})$ and $H(\to c\bar{c})$ decays in large-R jets with the same constituent and auxiliary task based approach. This section describes the setup and performance of both algorithms.

6.4.1 GN2X Architecture and Performance

Similar to its DNN-based predecessor detailed in section 4.2.2, GN2X [95] aims to classify large-R jets reconstructed from "Unified Flow Objects" (UFOs) based on their underlying physical process. Jets containing $H(\to b\bar{b})$ or $H(\to c\bar{c})$ decays are separated from those originating from top quarks with subsequent hadronic W decays[18] and multi-jet events (i.e. high-energy QCD radiation). The architecture of GN2X follows GN2 closely and consists of the same structure made up of an initialization network, a series of Transformer Encoder layers and three output task heads for jet flavor classification, track origin classification and secondary vertex reconstruction. GN2X uses up to 100 tracks ghost-associated to the jet as inputs with an identical feature set to its small-R jet counterpart detailed in table 6.4 except for the removal of n_{PixHoles} and the addition of the jet mass m_J. All tracks have to fulfill the same selection criteria as GN2 given in table 6.5.

Architecturally, GN2X is also similar to GN2 with details shown in table 6.13. The main difference between the two lies in the application of an additional Affine layer after each

[18] Large-R jets from top decays can contain either the full structure or be limited to the hadronic W decay only.

Transformer Encoder layer in GN2X. The main jet classification task of the boosted algorithm uses different output classes described by probabilities p_{Hbb}, p_{Hcc}, p_{QCD} and p_{Top}. The loss function is the same weighted sum described in section 6.2.1 and class imbalances are handled with the same loss reweighting strategy that ensures equal overall contributions between classes.

Component	**Layers**
Track Initializer (per track)	Affine (22 in, 256 out) + SiLU Affine (256 in, 192 out) + SiLU
Conditional Encoder	6×(Transformer+Affine) Transformer (192×n_{tracks} in, 256×n_{tracks} out, 4 heads) + SiLU Affine (256×n_{tracks} in, 192×n_{tracks} out) + SiLU Affine (192×n_{tracks} in, 128×n_{tracks} out) + SiLU
Graph Pooling	Pooling (128×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{registers}}$) in, 128×1 out)
Jet Classification Network	Affine (128 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 4 out) + Softmax
Track Origin Network (per track)	Affine (128 + 128 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 8 out) + Softmax
Vertexing Network (per track pair)	Affine (128×2 + 128 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 1 out) + Softmax

Table 6.13: Detailed summary of GN2X architecture with layer types and sizes.

The algorithm is trained on jets simulated under run 2 conditions (i.e. $\sqrt{s} = 13$ TeV) with 200 GeV $< p_T <$ 1.5 TeV, $|\eta| < 2.0$ and 50 GeV $< m_J <$ 200 GeV. A summary of training and evaluation samples is presented in table 6.14. The training dataset consists of a total of 62 million large-R jets, with 15 million each from the $H(\to b\bar{b})$ and $H(\to c\bar{c})$ categories, 10 million from top and 22 million from QCD. Resampling ensures the p_T, $|\eta|$ and mass distributions of all categories are matched to that of $H(\to b\bar{b})$. The latter is a new addition to GN2X, which is intended to make the algorithm agnostic to the jet mass to ensure good background rejection near the Higgs peak. It is trained using a batch size of 1000 with the ADAM optimizer [163] and a one-cycle learning rate schedule [197]. A dropout [199] rate of 10% is also utilized.

Process	Matrix element generator	Matrix element PDF	PS and hadronization	UE tune	HF quark decay
Training					
$q\bar{q} \to ZH$, $Z \to \mu^+\mu^-$	Pythia 8.306	NNPDF2.3LO	Pythia 8.306	A14	EvtGen 1.6.0
$Z' \to t\bar{t}$	Pythia 8.235	NNPDF2.3LO	Pythia 8.235	A14	EvtGen 1.6.0
Multi-jet	Pythia 8.235	NNPDF2.3LO	Pythia 8.235	A14	EvtGen 1.6.0
Evaluation					
$q\bar{q}/gg \to ZH$, $Z \to \ell\bar{\ell}/\nu\bar{\nu}/q\bar{q}$	Powheg Box v2	NNPDF3.0NLO	Pythia 8.212	AZNLO	EvtGen 1.2.0
$t\bar{t}$	Powheg Box v2	NNPDF2.3LO	Pythia 8.230	A14	EvtGen 1.6.0
Multi-jet	Pythia 8.235	NNPDF2.3LO	Pythia 8.235	A14	EvtGen 1.6.0

Table 6.14: Monte Carlo samples used for GN2X training and performance evaluation.

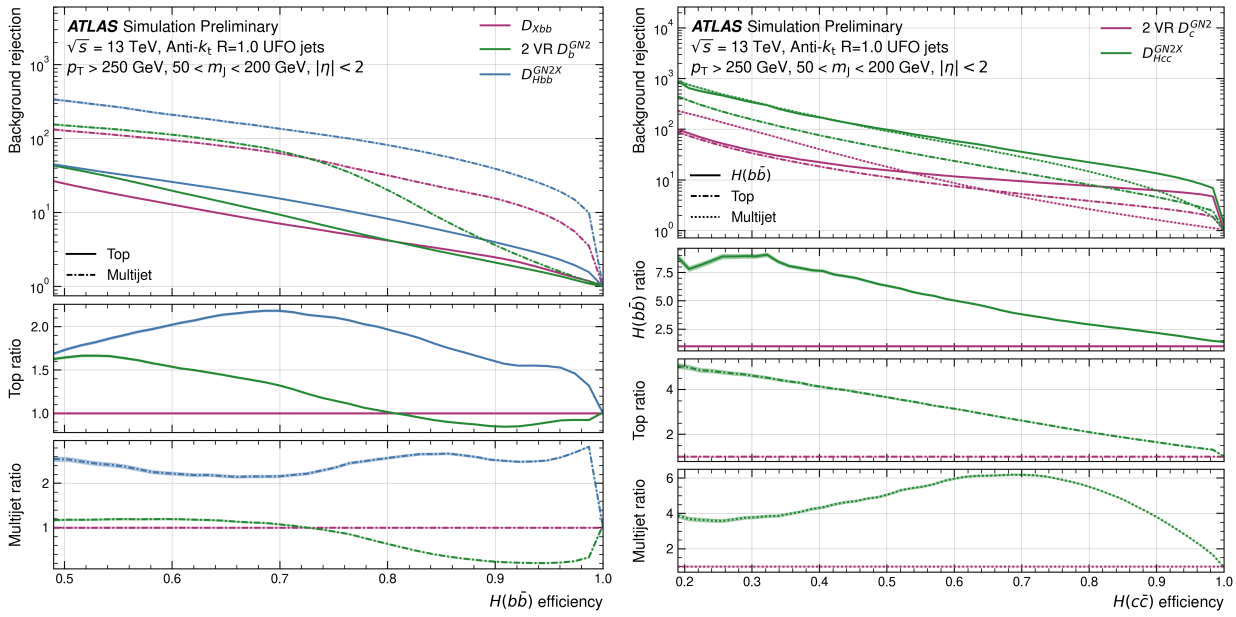
Performance evaluation with GN2X in the large-R jet regime involves the definition of a separate set of discriminants with parameters $\{f_{Hcc} = 0.02, f_{Top} = 0.25\}$ for D_{Hbb} and $\{f_{Hbb} = 0.3, f_{Top} = 0.25\}$ for D_{Hcc} given by

$$D_{Hbb/Hcc} = \log\left(\frac{p_{Hbb/Hcc}}{f_{Hcc/Hbb}p_{Hcc/Hbb} + f_{Top}p_{Top} + (1 - f_{Hcc/Hbb} - f_{Top})p_{QCD}}\right) \tag{6.15}$$

GN2X is compared against the previous DNN-based Xbb algorithm described in section 4.2.2 and an approach where GN2 is used to individually identify the flavor of the leading two VR track-jets[19] contained within the large-R jet. In the latter approach, the minimum D_b or D_c value between the track-jets is used to define the signal efficiency and no correlations between jets are taken into account. ROC curves for both $H(\to b\bar{b})$ and $H(\to c\bar{c})$ identification are presented in Fig. 6.35. GN2X provides a multi-jet rejection of ~ 200 and a top rejection of ~ 30 at an $H(\to b\bar{b})$ efficiency of 70%, which constitutes a $\times 2$ improvement compared to the previous algorithm. The $H(\to c\bar{c})$ identification capabilities of GN2X are a new addition and show a $H(\to b\bar{b})$ and multi-jet rejection of ~ 400 and a top rejection of ~ 200 at an efficiency of 30%.

Given the more physics oriented approach of GN2X, extra precautions are taken to ensure the final algorithm is not significantly dependent on jet mass via resampling. This is done to prevent higher background acceptance near the Higgs mass peak providing a falsely elevated signal strength for $H(\to b\bar{b})$ and $H(\to c\bar{c})$ decays, which is also called mass sculpting. While this can be accounted for with careful background modeling after applying flavor identification selections, the training of a robust model is favored. Histograms of jet mass

[19]These are reconstructed with parameters of $\rho = 30$ GeV, $R_{\min} = 0.02$ and $R_{\max} = 0.4$ as described in section 4.1.4.



(a) (b)

Figure 6.35: Plots showing background rejections vs. (a) $H(\to b\bar{b})$ and (b) $H(\to c\bar{c})$ efficiency for GN2X, the DNN-based Xbb algorithm (for the former only) and the GN2-based strategy from [95].

distributions for $H(\to b\bar{b})$ and QCD jets before and after a 70% $H(\to b\bar{b})$ efficiency selection are shown in Fig. 6.36. No significant evidence of mass sculpting is observed around the Higgs mass peak in these results. Deviations observed in the high and low mass regions are expected given the comparatively lower number of jets far from the Higgs mass present in the training sample[20]. Similar conclusions can be drawn from Fig. 6.37, which compares bin-by-bin efficiency deviations at a 70% identification efficiency for $H(\to b\bar{b})$ and $Z(\to b\bar{b})$. The latter is not used for GN2X training but serves as a lower mass proxy for the former. As observed, GN2X is able to identify $Z(\to b\bar{b})$ jets relatively more frequently than their $H(\to b\bar{b})$ counterparts. Identification is most efficient around the Z boson mass peak, indicating a preference for correctly classifying jets capturing complete on-shell decays over those surrounding the Higgs mass peak. Both of these results provide further support for the mass agnosticism of the GN2X algorithm.

[20]This is the case since all categories are sampled to match the m_J distribution of the $H(\to b\bar{b})$ category, which is relatively tightly centered around the Higgs mass.

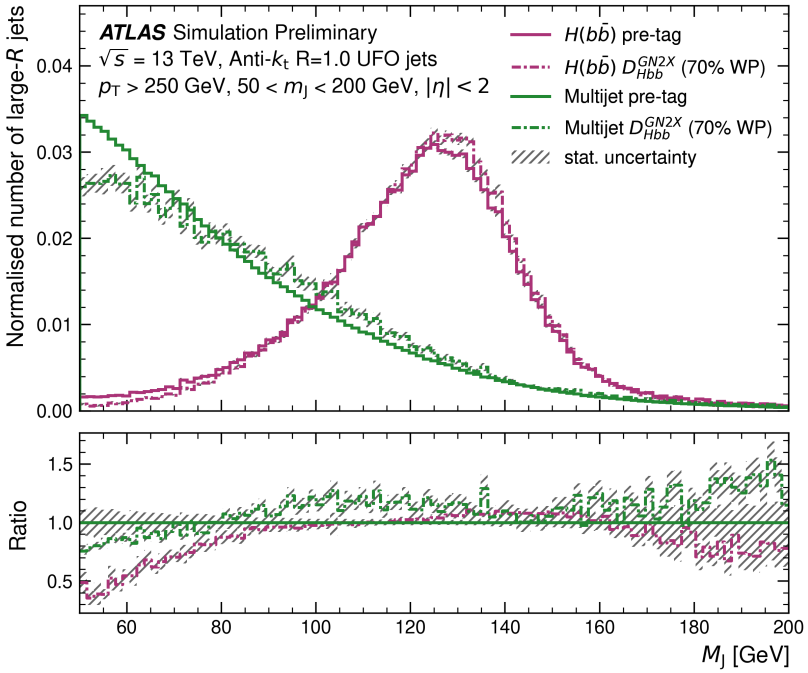
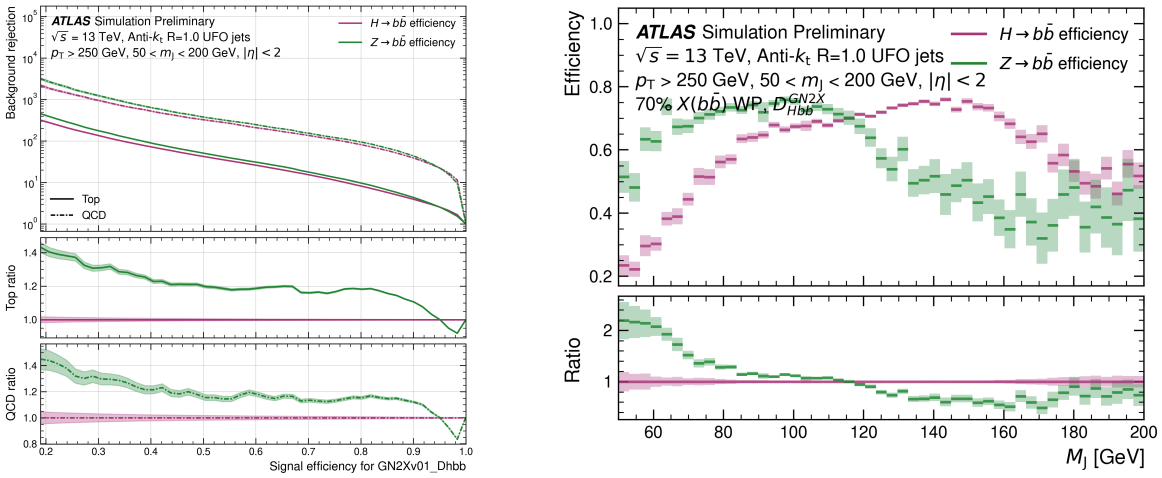


Figure 6.36: Histograms showing m_J distributions for $H(\rightarrow b\bar{b})$ and multi-jet before and after a GN2X based selection from [95].



(a) (b)

Figure 6.37: Plots showing (a) background rejections vs. $H(\rightarrow b\bar{b})$ and $Z(\rightarrow b\bar{b})$ efficiency for GN2X and (b) bin-by-bin variations of $H(\rightarrow b\bar{b})$ and $Z(\rightarrow b\bar{b})$ efficiencies as a function of m_J.

6.4.2 GN3X Architecture and Performance

While GN2X provides a powerful baseline for large-R jet identification in the boosted regime, issues regarding its performance when examining different simulated run periods (i.e. years of data taking) made it infeasible for use in a measurement such as $VH(\to c\bar{c})$. Deviations in the D_{Hbb} distributions of $t\bar{t}$ events were observed to be up to 100% in some bins between events simulated under conditions from 2017 and 2018. This was found to be caused by uneven sampling of MC simulated events from different run periods during the creation of the training dataset and exacerbated by the algorithm's sensitivity to slight differences in detector conditions over time. To remedy this issue and incorporate improvements made with GN3 in the small-R jet regime, a new version called GN3X was developed [108]. GN3X further splits the multi-jet category of GN2X based on the flavor of the dominant two sub-jets as either containing two bottom-jets (bb), one bottom-jet and one other jet (bx), one charm-jet and one light-jet (cx) or two light-jets (ll). It also adds separate categories for prompt hadronic W and $H(\to \tau^+\tau^-)$ decays. GN3X uses the same feature set for tracks as its predecessor as detailed in section 6.4.1, but also includes up to 100 PFlow objects as additional inputs. These objects have the same associated features as shown in 6.10 except for an added binary variable distinguishing charged objects from neutral ones.

On the architecture side, GN3X upgrades the core Transformer Encoder respective to GN2X by shifting to the TransformerV2 implementation with 8 register tokens [131] as described in section 6.3.1. The main task now has 5 additional output categories described by p_{Hbb}, p_{Hcc}, $p_{H\tau\tau}$, p_{Wqq}, p_{QCDbb}, p_{QCDbx}, p_{QCDcx}, p_{QCDll} and p_{Top}. Full architecture details are presented in table 6.15.

GN3X was trained with jets simulated under both run 2 (i.e. $\sqrt{s} = 13$ TeV) and run 3 (i.e. $\sqrt{s} = 13.6$ TeV) conditions with 200 GeV $< p_T <$ 2.1 TeV, $|\eta| < 2.0$ and 50 GeV $< m_J <$ 300 GeV as well as track selection requirements given by table 6.16. Its training dataset consists of 210 million jets made up of 40 million $H(\to b\bar{b})$ and $H(\to c\bar{c})$ jets each, 10 million $H(\to \tau^+\tau^-)$, 35 million top, 80 million QCD[21] and 5 million $W(\to q\bar{q}')$ jets. These jets are taken from the same samples as GN2X, which are detailed in table 6.14, except for the removal of the dedicated $t\bar{t}$ evaluation sample and the addition of a $W' \to WZ \to qqqq$ sample simulated with Pythia 8.306 [196] with an A14 generator tune [205] and NNPDF2.3LO set [12] for additional $W \to qq$ events. The loss is computed via a Geometric-Loss Strategy (GLS) [32] with the Lion optimizer [31] on batches of 8000 jets using the same task losses as in prior iterations.

The additional output categories of GN3X require an updated set of discriminants D_{Hbb} and D_{Hcc} with $\{f_{Hbb} = 0.02, f_{Hcc} = 0.02, f_{H\tau\tau} = 0.02, f_{Top} = 0.25, f_W = 0.1, f_{QCD} = 0.59\}$ given by

[21] Split roughly as 1.5:10:20:20 for ll:cx:bx:bb

Component	**Layers**
Component	**Layers**
Track Initializer (per track)	Affine (22 in, 256 out) + ReLU Affine (256 in, 256 out) + ReLU
PFlow Initializer (per PFlow object)	Affine (4 in, 256 out) + ReLU Affine (256 in, 256 out) + ReLU
Conditional Encoder	8×TransformerV2 (256×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{registers}}$) in, 256×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{registers}}$) out, 8 heads) + ReLU Affine (256×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{registers}}$) in, 128×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{registers}}$) out) + ReLU
Graph Pooling	Pooling (128×($n_{\text{tracks}} + n_{\text{flows}} + n_{\text{registers}}$) in, 128×1 out)
Jet Classification Network	Affine (128 in, 128 out) + ReLU Affine (128 in, 64 out) + ReLU Affine (64 in, 32 out) + ReLU Affine (32 in, 9 out) + Softmax
Track Origin Network (per track)	Affine (128 + 128 in, 128 out) + ReLU Affine (128 in, 64 out) + ReLU Affine (64 in, 32 out) + ReLU Affine (32 in, 8 out) + Softmax
Vertexing Network (per track pair)	Affine (128×2 + 128 in, 128 out) + ReLU Affine (128 in, 64 out) + ReLU Affine (64 in, 32 out) + ReLU Affine (32 in, 1 out) + Softmax

Table 6.15: Detailed summary of `GN3X` architecture with layer types and sizes.

Parameter	**Selection**
p_T	> 400 MeV
$\|\eta\|$	< 2.5
$\|d_0\|$	< 5 mm
$\|z_0 \sin\theta\|$	< 5 mm
n_{SCTHits}	≥ 7
$n_{\text{SCTShared}}$	≤ 1
n_{SCTHoles}	≤ 2
n_{PixHoles}	≤ 1

Table 6.16: Track selection used for inputs to `GN3X` [108].

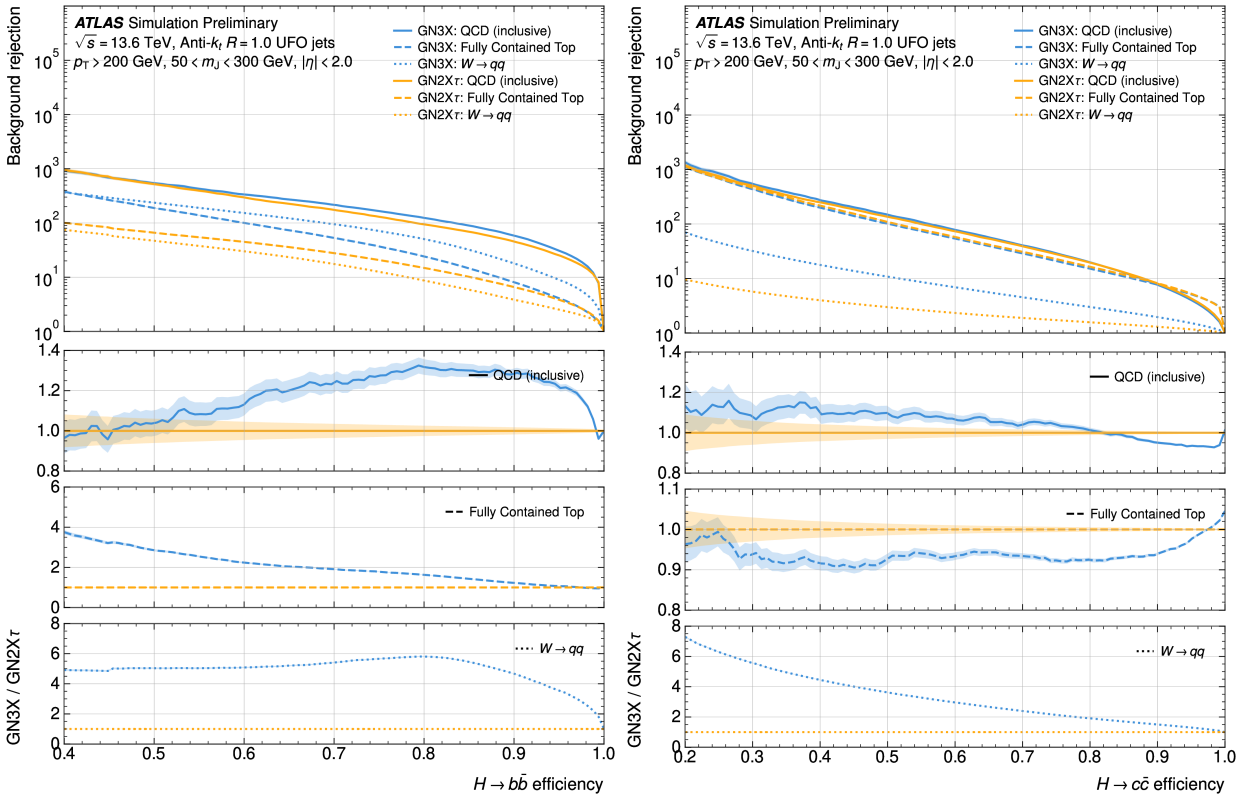
$$D_{Hbb} = \log\left(\frac{p_{Hbb}}{f_{Hcc}p_{Hcc} + f_{H\tau\tau}p_{H\tau\tau} + f_W p_W + f_{Top}p_{Top} + f_{QCD}(p_{QCDbb} + p_{QCDbx} + p_{QCDcx} + p_{QCDll})}\right) \tag{6.16}$$

$$D_{Hcc} = \log\left(\frac{p_{Hcc}}{f_{Hbb}p_{Hbb} + f_{H\tau\tau}p_{H\tau\tau} + f_W p_W + f_{Top}p_{Top} + f_{QCD}(p_{QCDbb} + p_{QCDbx} + p_{QCDcx} + p_{QCDll})}\right) \tag{6.17}$$

Performance of the algorithm is evaluated in comparison to a slightly updated version of GN2X called GN2Xτ, which is structurally identical to the version described in section 6.4.1 except for the inclusion of an $H \to \tau\tau$ output category to facilitate the identification of boosted Higgs to τ decays, the addition of PFlow objects as in GN3X and a slightly larger training dataset size of 64 million jets [102]. ROC curves showcasing background rejections as functions of $H \to b\bar{b}$ and $H \to c\bar{c}$ efficiencies are presented in Fig. 6.38. At 70% $H \to b\bar{b}$ efficiency, GN3X provides a multi-jet rejection of around ~ 250, a top rejection of ~ 50 and a $W \to qq$ rejection of 100. The latter is identified explicitly for the first time with GN3X while the former two are around a factor of $\times 1.25$ and $\times 1.67$ improved compared to baseline GN2X. For $H \to c\bar{c}$ identification, the background rejections are ~ 600 for multi-jet, ~ 400 for top, and ~ 30 for $W \to qq$ at 30% signal efficiency representing $\times 1.5$ and $\times 2$ improvements in multi-jet and top rejections respectively compared to GN2X.

6.5 Conclusions

Over the past decade, modern machine learning has fundamentally transformed jet flavor identification in the ATLAS experiment. Background rejections for bottom-jet identification in small-R jets have improved by more than a factor of 10 at the 70% efficiency WP between the BDT-based MV2c10 and the most recent version of GN3. Charm-jet identification, first introduced in the DL1 series of algorithms, has seen similar improvements of factors between 4 to 6 at the 30% efficiency WP in this timeframe. In addition to performance benefits, the new suite of Transformer-based algorithms are also more expressive through the use of auxiliary training objectives, predicting secondary vertex pairings as well as jet charge alongside the primary objective. In the high momentum regime, GN2X and GN3X are the first algorithms to explicitly identify boosted Higgs to charm decays in ATLAS. These taggers present the core of a new $VH(\to b\bar{b}/c\bar{c})$ measurement for run 3 of data taking as detailed in the following chapter 7.



(a) (b)

Figure 6.38: Plots showing background rejections vs. (a) $H(\to b\bar{b})$ and (b) $H(\to c\bar{c})$ efficiency for GN3X and GN2Xτ from [108].

Chapter 7

$VH(\to c\bar{c})$ in run 3 and beyond

The ATLAS measurements of the semi-leptonic $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ processes on run 2 data presented in chapter 5 lead to significantly improved constraints on the charm Yukawa coupling as detailed in section 5.4.2. This observed increase in measurement sensitivity was driven primarily by the adoption of ML-based jet flavor identification techniques as well as the design of a multivariate analysis (MVA) approach. Flavor tagging advancements discussed in chapter 6 are expected to provide further gains on the sensitivity of $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ measurements due to the significantly improved bottom- and charm-jet identification performance of algorithms such as `GN2` and `GN2X`. Given these considerations as well as the increase in available ATLAS data from run 3, follow up measurements to those presented in 5 on the complete run 2+3 datasets are well motivated and require the development of new analysis strategies.

Of particular interest is the definition of a new analysis regime based on high momentum $VH(\to c\bar{c})$ decays. While events with high di-jet momentum are not excluded in the $VH(\to c\bar{c})$ measurement presented in chapter 5, they follow the same reconstruction strategy as their low momentum counterparts. As a baseline, an Asimov fit without systematic uncertainties restricted to $p_T^V > 400$ GeV using the approach[1] and samples of the run 2 "legacy" analysis resulted in a 95% confidence level upper limit for $\mu_{VHc\bar{c}}$ of $22.56^{+8.84}_{-6.31} \times$ SM compared to $7.00^{+2.74}_{-1.96} \times$ SM without p_T^V restrictions[2]. As Higgs p_T increases, decays become increasingly collimated leading to overlapping jets which are more difficult to characterize by using small-R jet reconstruction and flavor identification techniques. Thus, a similar treatment to events in the boosted $VH(\to b\bar{b})$ regime where the whole Higgs decay is reconstructed as a single large-R jet is expected to improve this result. As part of this strategy, `GN2X` enables direct identification of $H \to c\bar{c}$ decays in large-R jets due to its dedicated output category.

This chapter presents an initial feasibility study on new $VH(\to c\bar{c})$ measurements with `GN2` in the resolved and `GN2X` in the boosted regime using the framework of the previous measurement in section 7.1. This is followed by the exploration of an analysis strategy

[1] In this case, the general approach follows that of the resolved $VH(\to c\bar{c})$ analysis regions.

[2] The value given without p_T^V restrictions is lower than the actual measurement result due to the exclusion of systematic uncertainties in this estimate.

for high momentum leptonic $VH(\to c\bar{c})$ decays using the GN3X algorithm in section 7.2. Preliminary sensitivity estimates utilizing a new MVA strategy based on neural networks are calculated using run 2 MC without consideration of systematic uncertainties. This result is intended to serve as a blueprint for an eventual fully combined resolved and boosted run 2+3 measurement of the same processes analogous to the results from chapter 5.

7.1 Feasibility studies on $VH(\to c\bar{c})$ with GN2 and GN2X

To quantify the potential impact of flavor identification improvements from algorithms such as GN2 and GN2X on measurements of the semi-leptonic $VH(\to c\bar{c})$ process, the full structure and event selection of the run 2 "legacy" analysis is adopted for a run 2 feasibility study based on an Asimov fit with MC simulated samples. This study does not involve the direct application of GN2 or GN2X but rather the scaling of event weights based on estimated tagging efficiencies. All events that meet cuts corresponding to analysis signal regions as defined in section 5.1.3 excluding flavor identification requirements are retained for this study. These events are split into 24 regions based on p_T^V and N_{jet}, where 18 regions with $p_T^V < 400$ GeV are part of the resolved regime and 6 regions with $p_T^V > 400$ GeV make up the new boosted regime as visualized in Fig. 7.1. This corresponds to a simplified structure compared to the full set of "legacy" analysis $VH(\to c\bar{c})$ regions shown in section 5.1.4 with the exception that boosted events are no longer merged into the highest momentum resolved regions. The momentum boundary between the boosted and resolved regime is chosen to be consistent with $VH(\to b\bar{b})$ for the purposes of this study. All regions are binned in the di-jet mass m_{JJ}, which represents an underestimation of the performance compared to an MVA fit.

To simulate the application of GN2 or GN2X, events are weighted based on the signal and background efficiencies of each algorithm as follows. Neither algorithm is applied directly to individual events due to the unavailability of relevant samples. For GN2 in the resolved regime, each event is weighted by a factor $\epsilon_{f1}\epsilon_{f2}$ where $f1$ and $f2$ are the true flavors of the leading two small-R jets and ϵ denotes the corresponding identification efficiency. For this study, a 30% charm-jet efficiency working point (WP) is chosen with a corresponding 5.0% bottom-jet and 0.3% light-jet efficiency derived from the low momentum results in Fig. 6.9. In the boosted regime, each event is weighted based on the category efficiency ϵ_c from the correspondence between samples and the four output categories (i.e. $H \to b\bar{b}$, $H \to c\bar{c}$, Top and QCD). This correspondence is not in general exact, which makes the mapping challenging as a whole. It is unambiguous only for signal, single-top and $t\bar{t}$ samples with explicit GN2X categories. Signal jets in V+jets samples typically originate from gluon radiation and are thus mapped to the QCD output category. Diboson samples are the most challenging and are split by true di-jet flavor: $b\bar{b}$ and $c\bar{c}$ events are added to the respective signal categories given their similar make-up and all others are marked as QCD. While there is no strict relationship between diboson and multi-jet events, the background is small enough that this is not expected to have a significant impact on the estimation and the flavor composition is expected to be similar. For simulated application of GN2 in the boosted regime, a weight

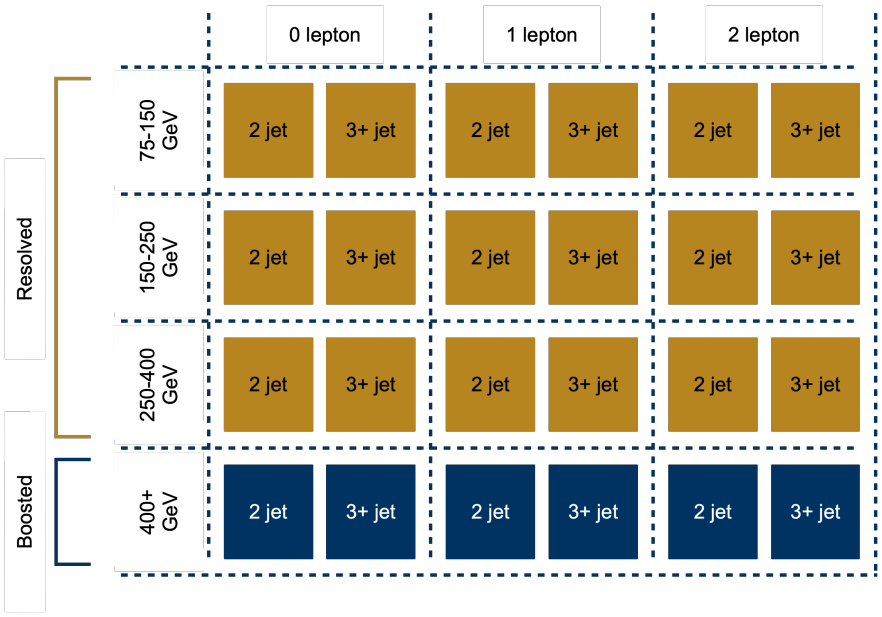


Figure 7.1: Diagram showing individually binned regions considered in the $VH(\to c\bar{c})$ feasibility study.

is derived based on performance of the algorithm on two VR jets contained in each large-R jet (denoted `2VR-GN2` as described in section 6.4.1) while for `GN2X` the large-R jet is used directly. At 30% $H \to c\bar{c}$ efficiency WP, the background efficiencies are 0.3% for $H \to b\bar{b}$, 0.7% for Top and 0.3% for QCD in `GN2X` as well as 2.5% for $H \to b\bar{b}$, 2.5% for Top and 1.0% for QCD in `2VR-GN2` based on results from Fig. 6.35. No veto on bottom-jet or $H \to b\bar{b}$ efficiency is applied for simplicity, but overlap regions are mostly negligible.

A variety of flavor identification strategies are compared via single parameter Asimov fits on $\mu_{VHc\bar{c}}$ using the CL_S method [187] as implemented in `pyhf` [150]. To compare these methods, a 95% CL expected upper limit on the POI is extracted in each case. A baseline sensitivity estimate is extracted by applying a C_TC DL1r-based event selection to mimic results from the "legacy" measurement described in chapter 5[3]. This selection requires one jet to pass the 25% charm-jet efficiency WP of DL1r while the other has to pass the 45% WP and both have to fail the 70% bottom-jet efficiency WP. For consistency, the boosted regions above $p_T^V > 400$ GeV are treated the same as the resolved regions when applying this simplified selection, which means that two small-R jets with C_TC tags are required in the boosted regions and large-R or VR track jets are not considered. All results are given in table 7.1 and are split into resolved only, boosted only and combined categories.

[3]The analysis results are not used directly in this study due to its simplified nature excluding uncertainties and with fewer regions.

Selection	Regions	95% CL upper limit
DL1r (C_TC)	18 resolved	22.68
DL1r (30% ϵ_c)	18 resolved	20.53
GN2 (50% ϵ_c)	18 resolved	15.21
2VR-GN2 (50% $\epsilon_{Hc\bar{c}}$)	6 boosted	29.72
GN2X (50% $\epsilon_{Hc\bar{c}}$)	6 boosted	17.56
DL1r (C_TC)	24 resolved + boosted	21.74
DL1r (30% ϵ_c)	24 resolved + boosted	19.19
GN2+2VR-GN2 (50% ϵ_c, 50% $\epsilon_{Hc\bar{c}}$)	24 resolved + boosted	13.49
GN2+GN2X (50% ϵ_c, 50% $\epsilon_{Hc\bar{c}}$)	24 resolved + boosted	11.39

Table 7.1: Summary of $VH(\to c\bar{c})$ feasibility study results extrapolated from run 2 to run 3 with GN2 and GN2X based jet flavor identification selections.

To ensure consistency between the C_TC selection strategy for DL1r and the reweighting strategy for all other taggers, a second estimate is computed for DL1r following the reweighting strategy. This procedure is based on values shown in Fig. 4.7 with a charm-jet efficiency of 30% and corresponding bottom- and light jet efficiencies of 11.1% and 1.4% applied in both the boosted and resolved regimes. Expected upper limits on the signal strength are slightly improved compared to the C_TC selection with 20.53 in the resolved case and 19.19 in the combined case. Differences between the two strategies are relatively minor compared to those between different strategies and differences between the resolved and combined estimates are consistent.

Both baseline estimated analysis results using a "legacy" C_TC selection show decreased sensitivity compared to the actual measurement result of chapter 5 as expected given the use of the dijet mass as an analysis variable and the simplified structure of analysis regions in this study. Regardless, these results serve as a comparison point for the application of more advanced flavor identification algorithms and are taken to be a conservative estimate in all cases. When considering resolved regions only, simulated application of GN2 improves the Asimov expected limit by almost 50% compared to the estimated "legacy" analysis results, indicating significant expected sensitivity gains from Transformer-based flavor identification algorithms. For boosted regions only, a GN2X selection outperforms a GN2 selection with VR track jets by a factor of more than two, showcasing the power of a dedicated boosted algorithm that can take relations between VR jets within a large-R jet into account. When adding both boosted and resolved regions into a single fit, sensitivity is driven by resolved regions leading to smaller differences between the two large-R jet identification strategies with 2VR-GN2 and GN2X. Both of these compare favorably to the "legacy" selection with 50% improvements in the expected upper limit for $\mu_{VHc\bar{c}}$.

Overall, the estimated sensitivity from just the boosted regions in a Transformer-based analysis is shown to be similar to that of a DL1r-based approach across the whole p_T^V spec-

trum. Given the large estimated benefits for $VH(\to c\bar{c})$ sensitivity from Transformer-based jet flavor identification algorithms as introduced in chapter 6, the application of these algorithms for a new measurement in the $VH(\to c\bar{c})$ space is well motivated.

7.2 Exploratory measurement of $VH(\to c\bar{c})$ in the boosted regime

Measurements of the $VH(\to c\bar{c})$ process in ATLAS have thus far been restricted to the low energy resolved regime using an approach where events with two small-R jets are selected and identified as charm-jets, as was the case in the previous run 2 measurement discussed in chapter 5. Section 6.4 showcases the power of dedicated jet flavor identification strategies for the boosted regime, which are also directly visible in the sensitivity estimates presented in the previous section. These improvements are driven primarily by the ability of algorithms such as `GN2X` to reject jets from gluon splitting much more efficiently than small-R jet based algorithms such as `GN2`, which only identify individual jets by flavor.

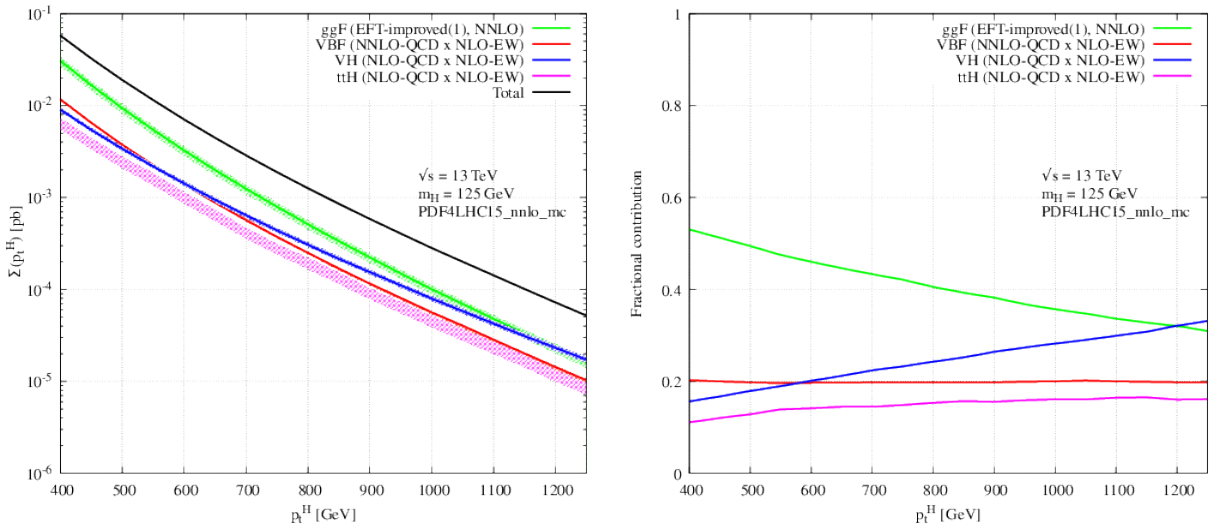
The high momentum regime is particularly interesting for VH analyses since it becomes the dominant Higgs production mode, surpassing VBF at $p_T^H > 600$ GeV and ggF at $p_T^H > 1200$ GeV [14] as detailed in Fig. 7.2. It has also been shown to be a significant driver of sensitivity in the $VH(\to b\bar{b})$ regime, further showcasing its potential as an important part of the phase space for $VH(\to c\bar{c})$ measurements. Finally and most importantly, sensitivity to new physics is expected to be greater at higher energies.

This section presents a proof-of-concept overview of a $VH(\to c\bar{c})$ measurement in the boosted regime. The overall analysis structure is presented first, followed by details on the MVA approach and preliminary sensitivity estimates. A set of future directions is given thereafter, focusing on the integration of this work into a larger combined analysis akin to the results presented in chapter 5.

7.2.1 Analysis Overview

The structure of the proof-of-concept "boosted" $VH(\to c\bar{c})$ analysis is based largely on the previous "legacy" $VH(\to b\bar{b}/c\bar{c})$ analysis presented in chapter 5. It encompasses both W and Z boson associated Higgs production with leptonic vector boson decays as before. Analysis regions are characterized by the number of reconstructed charged leptons as zero lepton, one lepton or two lepton targeting $Z \to \nu_\ell\bar{\nu}_\ell$, $W^- \to \ell^-\bar{\nu}_\ell$ and $Z \to \ell^+\ell^-$ decays respectively, as detailed in section 5.1. Hadronic $Z \to \tau^+\tau^-$ decays are disregarded as they will be included in all-hadronic searches.

The primary difference of this measurement compared to older results lies in the focus on a previously disregarded portion of the $VH(\to c\bar{c})$ phase space via boosted (i.e. high-momentum) Higgs decays. Events are characterized by the momentum of the vector boson decay p_T^V, which is heavily correlated to the Higgs momentum p_T^H. Reconstruction of the Higgs candidate involves a single large-R jet similar to the boosted $VH(\to b\bar{b})$ regions in the



(a) (b)

Figure 7.2: Plots showing (a) total and (b) relative cumulative cross sections for the main Higgs production modes as a function of the applied p_T cut from [14].

previous measurement with a threshold of $p_T^V > 250$ GeV. The $VH(\to b\bar{b})$ process is also considered as before and two POIs are defined: the signal strengths $\mu_{VHb\bar{b}}$ and $\mu_{VHc\bar{c}}$, which are measured across all analysis regions.

The main analysis backgrounds are identical to the previous measurement and are described in detail in section 5.1.1. All background signatures are consistent with the appearance of up to two charged leptons and missing energy from neutrinos as well as two heavy flavor jets. The most important backgrounds are vector bosons with associated QCD radiation denoted V+jets. This background category is further subdivided based on the flavor composition of the two signal candidate jets as "V+hf" (heavy flavor) for two bottom- or charm-jets, "V+lf" (light flavor) for two light-jets and "V+mf" (mixed flavor) for any other combination. Top related backgrounds are the next most important and are dominated by $t\bar{t}$ production in the high momentum regime. Finally, diboson production represents a more minor background category. Nominal analysis samples are summarized in table 7.2, where PS refers to parton shower and UE to underlying event. As before, the Higgs mass is set at 125 GeV with a branching ratio to $b\bar{b}$ of 58% and $c\bar{c}$ of 2.9%.

Event Selection

The analysis event selection is adapted from the previous $VH(\to b\bar{b}/c\bar{c})$ measurement described in section 5.1.3 and consists of two components: the selection of a leptonic V decay candidate and a hadronic Higgs decay candidate. The former is reconstructed with a set

Process	Matrix element generator	Matrix element PDF	PS and hadronization	UE tune	Cross-section order
Signal					
$qq \to VH$	Powheg Box v2 + GoSam + MiNLO	PDF4LHC21	Pythia 8.3	AZNLO	NNLO(QCD) +NLO(EW)
$gg \to ZH$	Powheg Box v2	PDF4LHC21	Pythia 8.3	AZNLO	NLO+NLL
Top quark					
$t\bar{t}$	Powheg Box v2	NNPDF3.0NLO	Pythia 8.3	A14	NNLO+NNLL
single-top (s channel)	Powheg Box v2	NNPDF3.0NLO	Pythia 8.3	A14	NLO
single-top (t channel)	Powheg Box v2	NNPDF3.0NLO	Pythia 8.3	A14	NNLO
Wt	Powheg Box v2	NNPDF3.0NLO	Pythia 8.3	A14	Approx. NNLO+NNLL
V+jets					
V+jets	Sherpa 2.2.14	NNPDF3.0NNLO	Sherpa 2.2.14	Default	NLO
Diboson					
$qq \to VV$	Sherpa 2.2.14	NNPDF3.0NNLO	Sherpa 2.2.14	Default	NLO

Table 7.2: Monte Carlo samples used in the boosted $VH(\to c\bar{c})$ analysis.

of lepton identification criteria while the latter is chosen with jet flavor identification techniques. Reconstruction and calibration techniques for leptons, MET and jets are updated to follow run 3 procedures as detailed in section 4.1.

Lepton selection is carried over entirely from the prior run 2 analysis and is summarized through a set of quality criteria detailed in table 5.2. As before, "VH Loose" cuts are defined to reject low quality lepton candidates in the zero lepton channel while "WH Signal" and "ZH Signal" cuts are kept tighter to identify high quality candidates in the one and two lepton regimes. Events in the zero lepton channel are required to have $E_T^{\text{miss}} > 200$ GeV as well as no "VH Loose" electrons or muons. The one lepton channel requires at least one "WH Signal" and no "VH Loose" lepton as well as $E_T^{\text{miss}} > 50$ GeV. Hadronic τ lepton decays are vetoed to reduce the $t\bar{t}$ background. Finally, in the two lepton channel one "ZH Signal" and one additional "VH Loose" lepton is required, where the two leptons must have opposite charge if they are muons given the higher rate of charge misreconstruction in electrons. The resulting Z candidate must also have $66 < m_{\ell\ell} < 116$ GeV to be consistent with the mass of a Z boson.

Higgs candidates are identified as a single large-R jet in the boosted regime similar to the previous analysis and must have $p_T > 250$ GeV and $|\eta| < 2.0$. Large-R jet reconstruction is adapted from relying only on calorimeter clusters to Unified Flow Objects (UFOs), which

combine information from the tracking detector and calorimeter as described in section 4.1.4. Contrary to the prior result, track jets are not utilized and flavor tagging is instead done directly through explicit identification of the boosted $H \to b\bar{b}$ and $H \to c\bar{c}$ decays via `GN3X` as detailed in section 6.4.2. The bottom and charm tagging discriminants D_{Hbb} and D_{Hcc} are used to determine signal identification WP cuts, which define orthogonal analysis regions for the two processes. In order to decouple the performance of the `GN3X` algorithm from jet kinematics as much as possible, WPs are defined in 260 bins of jet mass ranging from 40 GeV to 300 GeV (with 1 GeV width) and 10 bins of jet p_T between 200 GeV and 3 TeV. These are denoted "flat-mass WPs" and differ from the "fixed-cut" WPs derived for `DL1r`. Their definition is based on a target $H \to b\bar{b}$ or $H \to c\bar{c}$ signal efficiency and a corresponding D_{Hbb} or D_{Hcc} cut that also defines a QCD background efficiency across the training sample. This value for the QCD background efficiency is then enforced separately in each jet mass and p_T bin, providing a different discriminant cut for each. The series of resulting cut values defines the WP whose actual $H \to b\bar{b}$ or $H \to c\bar{c}$ efficiency differs from its original target because of this enforcement. The QCD rejection is used in the definition of each WP instead of the signal efficiency due to the comparatively lower number of available signal events (especially in higher and lower ends of the mass spectrum) which would result in large signal efficiency uncertainties in certain regions of the phase space. For the analysis, an approximate $H \to b\bar{b}$ efficiency WP of 60% and a $H \to c\bar{c}$ efficiency WP of 30% are chosen, corresponding to a bin-by-bin QCD background efficiency of 0.4% and 0.25% respectively. This combination is found to achieve the best estimated sensitivity overall.

Event Categorization

All events passing analysis cuts are sorted into distinct signal regions (SR) and control regions (CR), which are kept orthogonal. All 15 regions (consisting of 9 SRs and 6 CRs) enforce identical jet flavor identification requirements as described above, where events in $VH(\to c\bar{c})$ regions need to pass the 30% $H \to c\bar{c}$, but not the 60% $H \to b\bar{b}$ WP which defines analogous $VH(\to b\bar{b})$ regions. Overlap between the two categories is found to be negligible and preference is given to $H \to b\bar{b}$ for historical reasons only. Signal and control regions are split based on lepton channel and three separate p_T^V bins as 250 GeV $< p_T^V <$ 400 GeV, 400 GeV $< p_T^V <$ 600 GeV and $p_T^V >$ 600 GeV. An overview of all analysis regions is given in Fig. 7.3.

The only type of CRs utilized in the analysis are "Top CRs" intended to constrain the top quark backgrounds in the zero and one lepton channels, which mostly consist of pair production (i.e. $t\bar{t}$). To isolate these events, an additional requirement is placed on small-R jets that do not overlap with the signal candidate large-R jet. If at least one of these small-R jets passes a 60% bottom identification efficiency WP after application of `GN2`, the event is considered a part of a CR. Otherwise, it is added to the corresponding SR. A breakdown of the composition of analysis SRs and CRs is given in Fig. 7.4 across all p_T^V regions and lepton channels. Most CRs are very pure, consisting of $> 90\%$ top quark events across the one lepton channel and between $\sim 45\%$ and $\sim 90\%$ in the zero lepton channel.



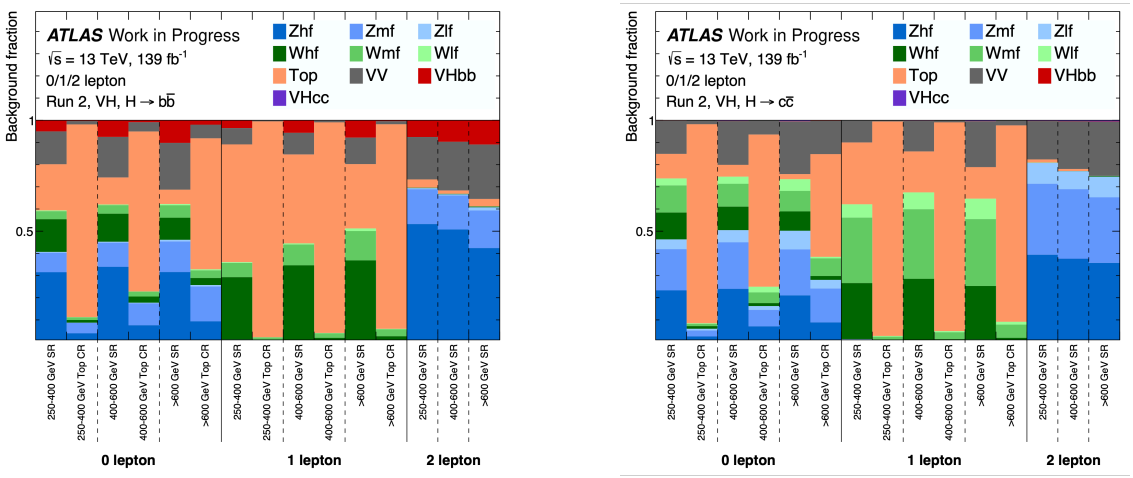
Figure 7.3: Overview of all $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ analysis regions in the partial run 3 boosted measurement.

Events in each region are separated into discrete bins as in the previous measurement. Control regions are binned using the large-R jet mass m_J while signal regions utilize a multivariate analysis (MVA) discriminant optimized to separate signal from background events by considering multiple different variables.

7.2.2 Neural network based MVA Approach

Prior implementations of a separating discriminant between $VH(\to b\bar{b})$ or $VH(\to c\bar{c})$ and relevant backgrounds in the run 2 analysis involved training a set of Boosted Decision Trees (BDTs) individually in each signal region[4] as detailed in section 5.1.4. The simpler structure of the run 3 boosted VH analysis and the power of modern machine learning algorithms, such as Deep Neural Networks (DNNs) [198] and Transformers [215] favor the consideration of a more streamlined and more powerful approach. A number of upgrades to the run 2

[4]In some cases, p_T^V and jet multiplicity split analysis regions were merged for MVA trainings but trainings were always performed separately for each lepton channel.



(a) (b)

Figure 7.4: Prefit composition of SRs and CRs across the $VH(\to b\bar{b})$ (a) and $VH(\to c\bar{c})$ (b) regimes split by lepton channel and p_T^V region.

analysis MVA are examined on $VH(\to c\bar{c})$ using the full set of run 2 nominal MC samples described in section 5.1.2:

- All MVA trainings utilize either Transformers or DNNs rather than BDTs. This decision was based on the more expressive nature and increased architectural flexibility [11] of neural network algorithms as well as the ability to take advantage of existing training infrastructure via the `salt` framework [13]. Performance comparisons are made between the Transformer and DNN based approaches only.

- An approach with an "inclusive" training strategy is considered, where all events passing the vector boson and Higgs candidate selections without enforcement of jet flavor identification cuts are used for training to increase total statistics. In addition, p_T^V cuts are loosened to include all events containing reconstructed large-R jets with the constraint that $p_T^J > 150$ GeV. The motivation for this strategy is the inclusion of slightly out-of-distribution events to allow the neural network to better infer the shapes of background PDFs within the signal regions through their shapes in similar regions of the analysis phase space. This is especially pertinent given limited training statistics in the SRs. Different variations of this approach are compared below.

- The move from BDTs to neural networks provides the opportunity to train multi-class MVA algorithms due to their ability to natively target multiple output values in a single training without requiring significant modifications of the underlying algorithm. Given the differences between background signatures in the analysis, two signal classes and four background classes are defined: "Hcc", "Hbb", "Diboson", "Top", "W+jets",

"Z+jets". A multi-class approach then potentially allows for the definition of entirely new background enriched CRs. It also facilitates a combined $VH(\to b\bar{b}/c\bar{c})$ measurement without requiring separate MVA trainings as long as events from corresponding $VH(\to b\bar{b})$ signal regions are included.

The design of the neural network models is inspired by `DL1` [90] for the DNN and `GN2` [110] for the Transformer with some architectural differences and is described in table 7.3. Model size increases were tested during different stages of development but had a negligible impact on performance. The DNN models are presented a list of event level input features, mirroring the old BDT approach. In the Transformer models, up to 10 small-R jets that do not overlap with the Higgs candidate large-R jet and up to 5 additional large-R jets each have associated feature sets. Event level features are concatenated to these constituent feature sets in the Transformer models. In the DNN approach, all features of the four highest p_T non-overlapping small-R jets are appended to the event level feature list. All input features are summarized in table 7.4. Lepton variables are filled with dummy values when not relevant in a particular channel. Training details are mostly identical to those discussed in chapter 6 with binary cross-entropy loss functions (or linear combinations thereof) used throughout.

A set of different training strategies with variations in training dataset size, number of trained models and minor model architecture differences are pursued. Event selections for the training dataset either follow $VH(\to c\bar{c})$ signal region cuts with 1,463,489 total events[5] or the "inclusive" strategy described above with loosened p_T^V cuts and no flavor identification enforcement resulting in 70,749,481 events[6]. In both cases, MC event weights are used to weight each event during training and output categories are correspondingly scaled to ensure equal contributions to the loss function for each event type (both in the binary and multi-class cases) In addition, k-folding is applied with $k = 2$ such that 45% of events are used for training, 5% for validation and 50% for testing. After training, the model with the lowest loss on the validation dataset is chosen and evaluated. Different strategies are compared primarily using the DNN based approach with a batch size of 2000. They consist of the following:

- **Baseline** follows the philosophy of the prior analysis MVA with separate trainings performed in each lepton channel. The SR only dataset is used and split according to the lepton channel selection criteria. No additional separation is done for the different p_T^V regions.
- **Combined** is functionally similar to "Baseline", but does not divide events based on their lepton channel. Rather it involves training a single MVA model across all analysis SRs.

[5]This SR selection is split roughly as 0.8% Hbb, 65.5% Hcc, 6.8% Diboson, 11.2% W+jets, 13.8% Z+jets and 1.9% Top

[6]The inclusive selection is much more background dominated and split roughly as 2.4% Hbb, 2.7% Hcc, 4.0% Diboson, 35.2% W+jets, 42.7% Z+jets and 13.0% Top

Component	Layers
	DNN
Initializer	Affine (48 in, 256 out) + Mish 2× Affine (256 in, 256 out) + Mish
Output Network	Affine (256 in, 128 out) + Mish Affine (128 out, 64 out) + Mish Affine (64 in, 32 out) + Mish Affine (32 in, 2 (binary) or 6 (multi-class) out) + Softmax
	Transformer
Small-R jet Initializer (per small-R jet)	Affine (30 in, 256 out) + RELU Affine (256 in, 265 out) + ReLU
Large-R jet Initializer (per large-R jet)	Affine (28 in, 256 out) + ReLU Affine (256 in, 256 out) + ReLU
Conditional Encoder	4×Transformer (256×($n_{\text{small-R}} + n_{\text{large-R}}$) in, 256×($n_{\text{small-R}} + n_{\text{large-R}}$ out), 8 heads) + ReLU Affine (256×($n_{\text{small-R}} + n_{\text{large-R}}$) in, 128×($n_{\text{small-R}} + n_{\text{large-R}}$) out) + ReLU
Graph Pooling	Pooling (128×($n_{\text{small-R}} + n_{\text{large-R}}$) in, 128×1 out)
Output Network	Affine (128 in, 128 out) + SiLU Affine (128 in, 64 out) + SiLU Affine (64 in, 32 out) + SiLU Affine (32 in, 6 out) + Softmax

Table 7.3: Detailed summary of DNN and Transformer MVA architectures for the boosted $VH(\to c\bar{c}$ measurement including layer types and sizes.

- **Inclusive** is identical to "Combined" but instead of only training on analysis SRs, this strategy uses the large, background dominated dataset to include events similar to but outside of the analysis phase space. The inclusion of these events greatly increases training statistics at the expense of broadening the scope of event classification for the MVA. Since all events are used during training, two additional variables are added to mark SR events: $\text{bin}^{60\%\text{Hbb}}_{\text{GN3X}}(J)$ and $\text{bin}^{30\%\text{Hcc}}_{\text{GN3X}}(J)$. The validation loss is calculated over all events and is assumed to be similar between SR and non-SR events.

- **LeptonHead** is based on the "Combined" strategy, but uses a modified neural network architecture that has three separate output network components as detailed in table 7.3 running in parallel. Each output component is optimized for classification within a single lepton channel only through a separate loss function that masks contributions

Event Input	**Description**
m_J	Higgs candidate large-R jet mass
p_T^J	Higgs candidate large-R jet transverse momentum
p_T^V	Vector boson transverse momentum
ϕ_V	Vector boson phi
$\lvert\Delta\phi(V,J)\rvert$	Absolute phi difference between Higgs candidate and vector boson
$\Delta y(V,J)$	Rapidity difference between Higgs candidate and vector boson
E_T^{miss}	Missing transverse energy
$n_{\text{small-R}}$	Number of small-R jets not overlapping with Higgs candidate large-R jet
$n_{\text{large-R}}$	Total number of large-R jets
p_T^{j1}, p_T^{j2}	Transverse momenta of leading small-R jets overlapping with Higgs candidate large-R jet
$\Delta R(j1,j2)$, $\Delta R(j1,j3)$, $\Delta R(j2,j3)$	Angular differences between leading three small-R jets overlapping with Higgs candidate large-R jet
$\Delta R(J,j1)$, $\Delta R(J,j2)$, $\Delta R(J,j3)$	Angular differences between leading three small-R jets and Higgs candidate large-R jet.
$p_T^{\ell 1}$, $p_T^{\ell 2}$	Transverse momentum of leptons making up V candidate
$\eta^{\ell 1}$, $\eta^{\ell 2}$	Pseudorapidity of leptons making up V candidate
$\phi^{\ell 1}$, $\phi^{\ell 2}$	Phi of leptons making up V candidate
$\cos\theta_{\ell 1,\ell 2}$	Angular difference between trajectories of leptons making up V candidate
Small-R Jet Input	**Description**
p_T	Non-overlapping small-R jet transverse momentum
η	Non-overlapping small-R jet pseudorapidity
ϕ	Non-overlapping small-R jet phi
m	Non-overlapping small-R jet mass
$\text{bin}_{\text{GN2}}^{70\%\text{b}}(j)$	Binary variable indicating jet passes 70% b-jet WP for GN2
$\text{bin}_{\text{GN2}}^{30\%\text{c}}(j)$	Binary variable indicating jet passes 30% c-jet WP for GN2
Large-R Jet Input	**Description**
p_T	Large-R jet transverse momentum
η	Large-R jet pseudorapidity
ϕ	Large-R jet phi
m	Large-R jet mass

Table 7.4: List of input features used for the analysis MVA studies in the run 3 boosted $VH(\to c\bar{c})$ analysis for the DNN and Transformer approaches.

from all other events. The other components are shared and informed by all three output heads. As such, this strategy provides the benefits of a combined training without sacrificing the potential advantages of a more targeted approach. The loss functions for each lepton channel are added together with weights based on available events in each channel[7] and optimized across all channels simultaneously.

- **TwoHead** is based on the "Inclusive" strategy with architectural modifications similar to the "LeptonHead" algorithm. Rather than splitting output heads by lepton channel, there is one output head which classifies events within the analysis SRs and another for events that do not pass these cuts. The initial shared components are thus informed by out-of-distribution events without the final output score being dominated by them. This is intended to be a more direct way of exploring the benefits of including additional events without diluting the scope of the MVA. The overall loss is a linear combination of the two output losses with the SR component weighted by a factor of 200 to emphasize classification performance in the SRs. The final model is chosen based on the SR loss

[7]This means that the zero lepton component is weighted as 1.0, the one lepton component as 1.1 and the two lepton component as 1.9.

component only.

The "LeptonHead" and "TwoHead" strategies are inspired by the auxiliary task structure present in the jet flavor identification algorithms explored in chapter 6. In each case, events are classified by a single output head focusing on part of the phase space while the intermediate representations within the initial layers of the network are shaped by all events under consideration. This allows the neural network MVA to build information related to all training events into its hidden representation while still allowing individual finetuning per output head.

To measure the performance of each algorithm, a series of ROC curves are shown. These show overall signal efficiency and background rejection within each lepton channel based on MVA discriminants defined as

$$D_{\text{MVA}}^{\text{bin}} = \log(\frac{p_{\text{Hcc}}}{p_{\text{Other}}}) \tag{7.1}$$

$$D_{\text{MVA}}^{\text{mult}} = \log(\frac{p_{\text{Hcc}}}{p_{\text{Hbb}}f_{\text{Hbb}} + p_{\text{Diboson}}f_{\text{Diboson}} + p_{\text{W+jets}}f_{\text{W+jets}} + p_{\text{Z+jets}}f_{\text{Z+jets}} + p_{\text{Top}}f_{\text{Top}}}) \tag{7.2}$$

in the binary and multi-class classification cases respectively. For the multi-class discriminant, the f constants are different in each lepton channel and are chosen to represent the approximate pre-fit fractions of each background category. This is intended to balance out the different background rejections based on their relative importance in each part of the analysis phase space. ROC curves are presented separately for the zero, one and two lepton channels but are not separated by p_T^V region.

Four different binary classification strategies are considered and shown in Fig. 7.5: "Baseline", "Combined", "LeptonHead" and "Inclusive". All approaches present broadly similar performance, except for a notable $\sim 20-30\%$ reduction in background rejection in the "Inclusive" strategy across most of the phase space. These models use the larger, more diluted training dataset which contains events outside of analysis SRs. This decrease in performance indicates the addition of slightly out-of-distribution events does not aid the separation of signal and backgrounds within the SRs. The "Combined" model which is trained on all SRs simultaneously also shows worse performance by up to $\sim 10\%$ compared to the "Baseline" strategy which consists of separate models for each lepton channel. This difference is recovered via the "LeptonHead" approach, which are also trained on events from all lepton channels but use a modified architecture as described above. The "LeptonHead" strategy achieves a separation on par with the "Baseline" strategy, except for a small decrease of $< 5\%$ towards the higher signal efficiency region of the one lepton channel. In the lower end of the signal efficiency distribution this relationship is mostly flipped with the "LeptonHead" models outperforming the "Baseline" models. These differences are small enough that the "LeptonHead" model remains the preferred choice due to its simpler training strategy alone, requiring a factor of 3 fewer total model trainings.

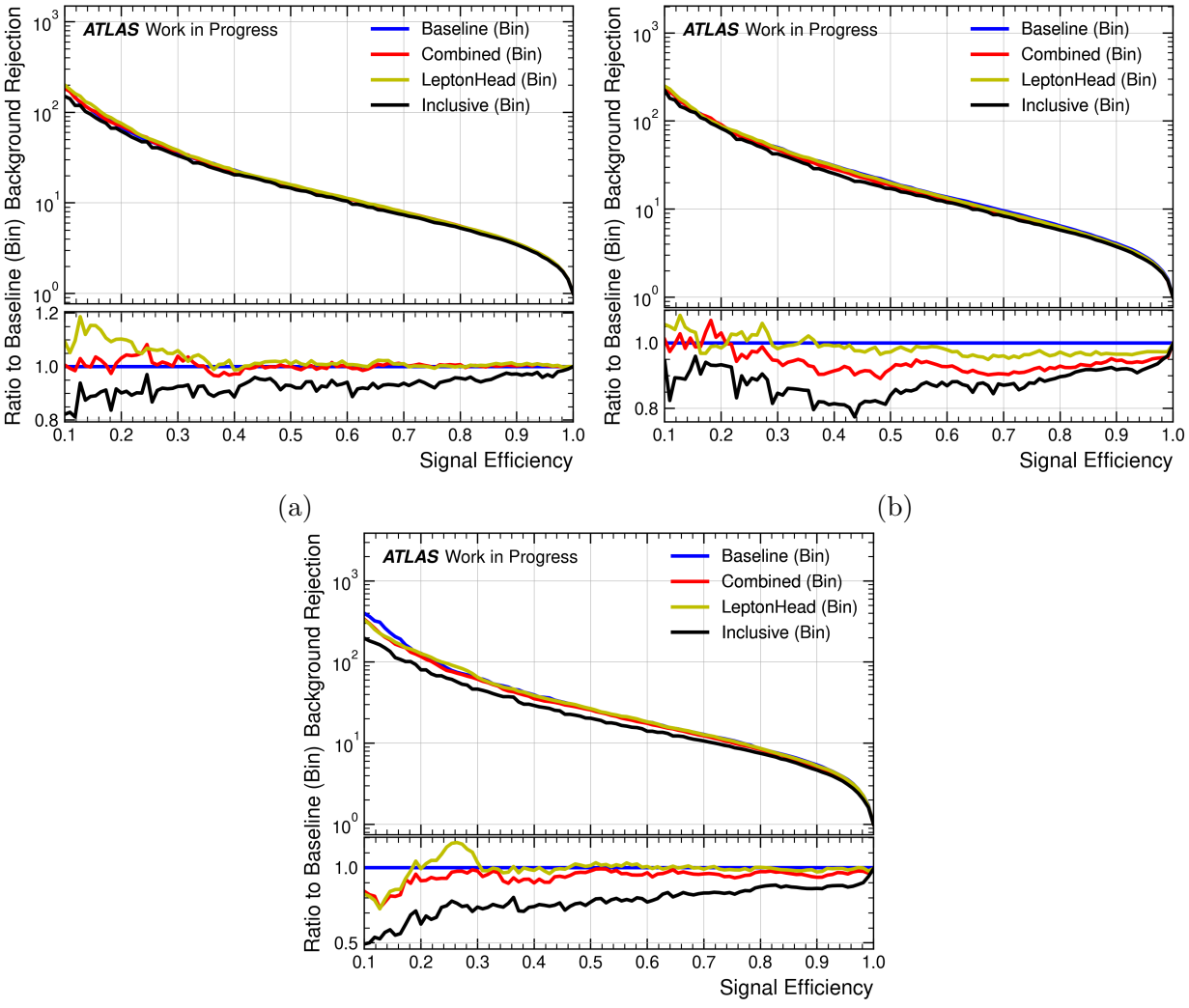


Figure 7.5: ROC curves comparing $H \to c\bar{c}$ signal efficiency vs. background rejection for binary classification MVA models in the (a) zero, (b) one and (c) two lepton channels.

A series of multi-class approaches are also tested as described above. Multi-class versions of the "LeptonHead", "Inclusive" and "TwoHead" strategies are compared with the "Baseline" approach using binary classification in Fig. 7.6. The former, which gives the best performance in the binary classification case, shows significant decreases in background rejection of around $\sim 20\%$ when adapted to the multi-class case. This performance difference stems from the inherently more challenging tasks of separating all backgrounds from each other rather than just identifying $H \to c\bar{c}$. The "Inclusive" strategy recovers this performance difference to provide a set of models that are on par with the simplest binary "Baseline" approach, but in a multi-class setting. The intended benefit of including a large number of slightly out-of-distribution events is thus directly visible in this more complex training objective. Overall, the multi-class strategy provides a series of advantages compared to a more traditional MVA via the ability to produce background enriched control regions using the model's class probability scores as well as the potential to train a single MVA across $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ SRs. The latter further reduces the number of necessary model trainings by a factor of two in a combined analysis scenario. The architecturally more sophisticated "TwoHead" strategy does not provide any additional benefits over the "Inclusive" approach, but is preferenced regardless due to it containing a separate loss component for SR events that can be tracked individually.

A final set of comparisons is made between the "TwoHead" strategy with Transformer models and DNNs in Fig. 7.7. Minor performance differences are observed between the two model types with the Transformer outperforming the DNN by around $\sim 5\%$ across most of the higher end of the signal efficiency space and more significant fluctuations on the lower end. These minor differences likely indicate that the analysis phase space is not complicated enough to benefit from the more advanced constituent-based Transformer approach. In comparison to jet flavor identification where these models are usually applied, the analysis only contains a single signal jet of interest per event alongside generally fewer than 5 additional background jets. These jets frequently also contain relatively little relevant information since many of them come from gluon-splitting.

To provide a first estimate for the analysis sensitivity, the binary "LeptonHead" model is chosen due to its performance advantages and comparative simplicity. Multi-class strategies are not pursued further at this stage, but could provide distinct advantages when considering systematic uncertainties in the future. In particular, the addition of background enriched control regions defined via cuts on background discriminants analogous to equation 7.2 could bring useful constraints on modeling systematics.

7.2.3 Preliminary Fit Results

Preliminary Asimov fit results for a combined boosted $VH(\to b\bar{b}/c\bar{c})$ measurement are produced via a TRExFitter based framework [212] using the aforementioned analysis selections without considering systematic uncertainties. Seven floating normalization factors are defined for the major Top, W+jets and Z+jets backgrounds. The former two are decorrelated in p_T^V and applied in the zero and one lepton channels while the latter is fully correlated

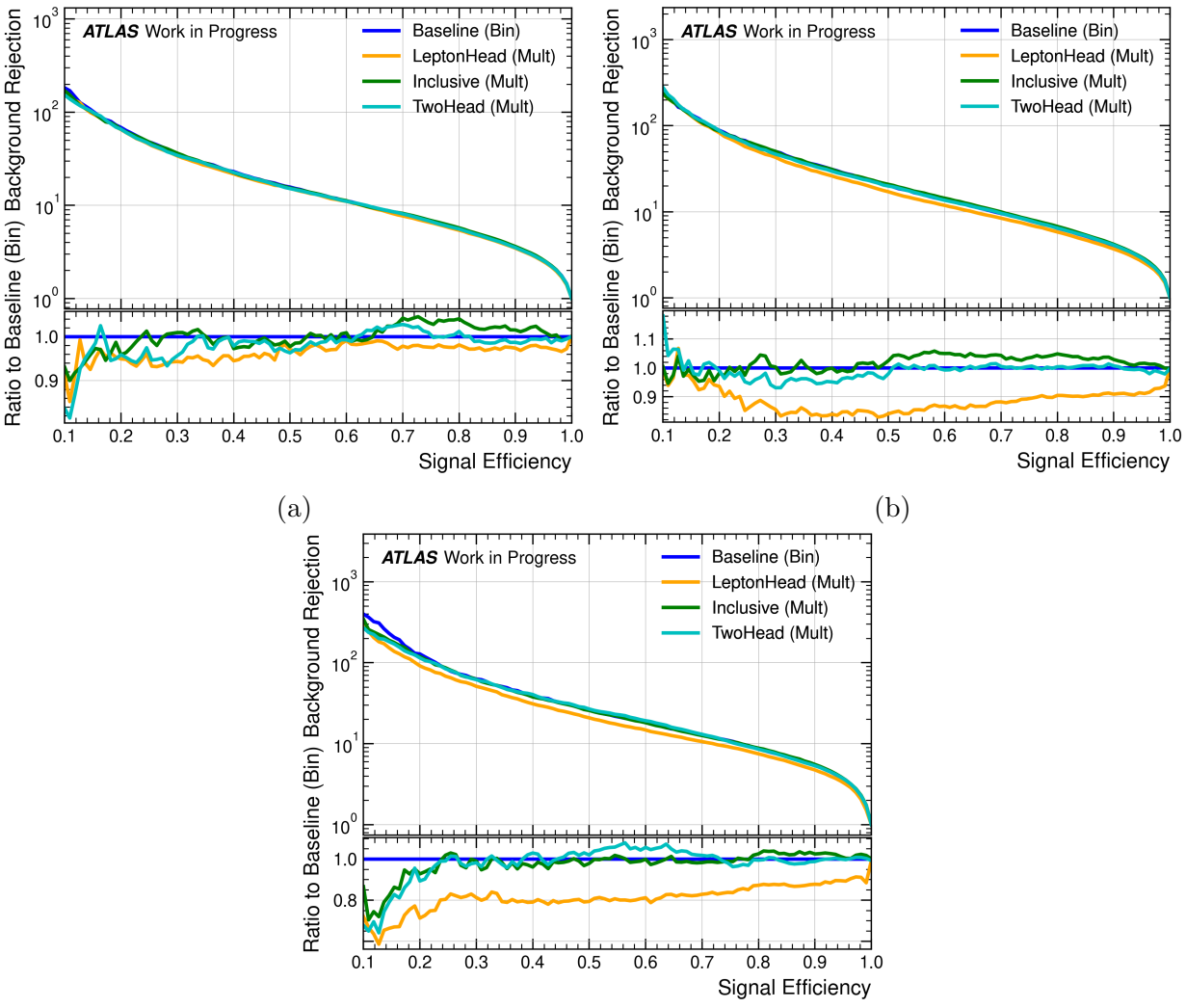


Figure 7.6: ROC curves comparing $H \to c\bar{c}$ signal efficiency vs. background rejection for multi-class classification MVA models in the (a) zero, (b) one and (c) two lepton channels.

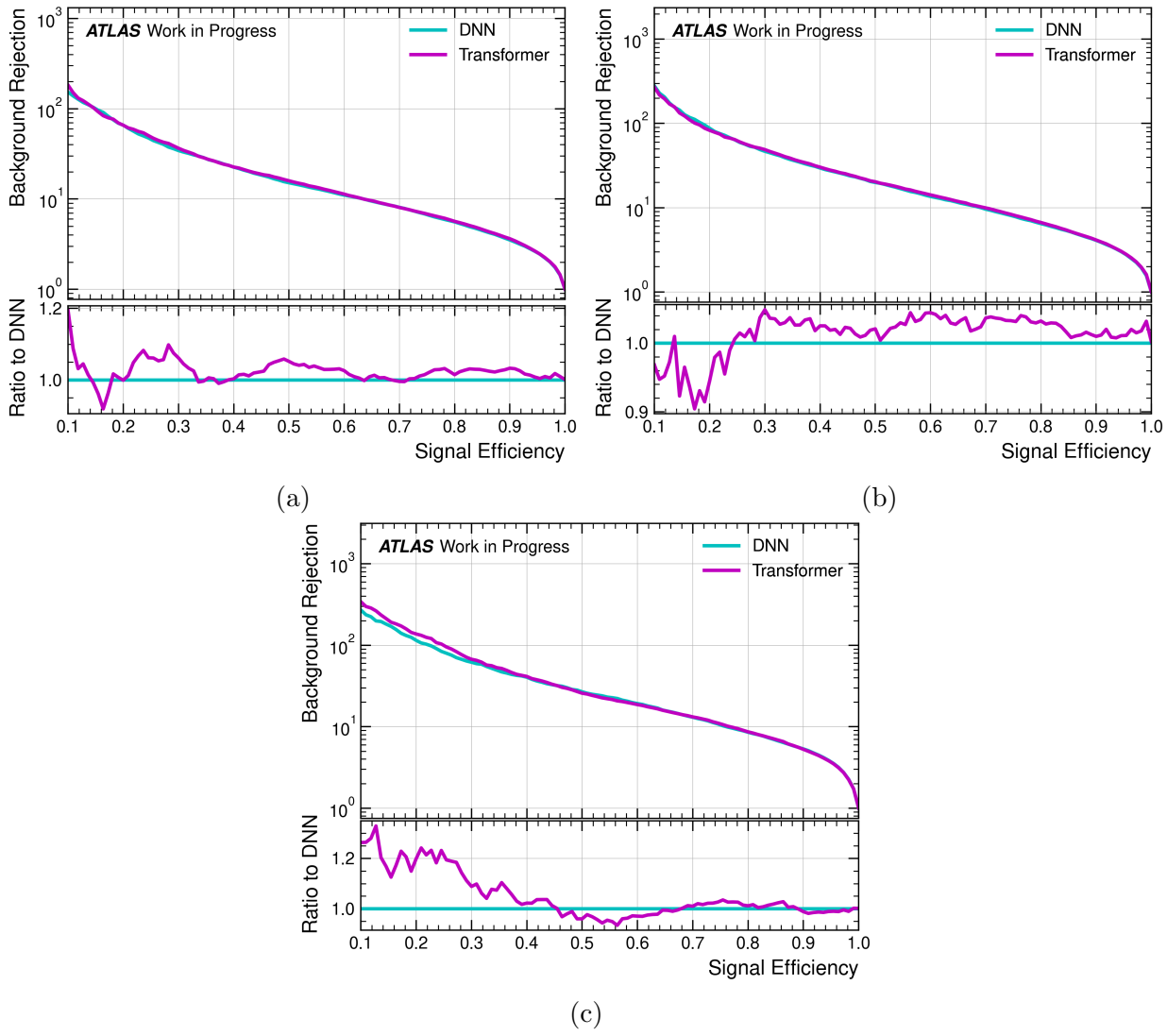
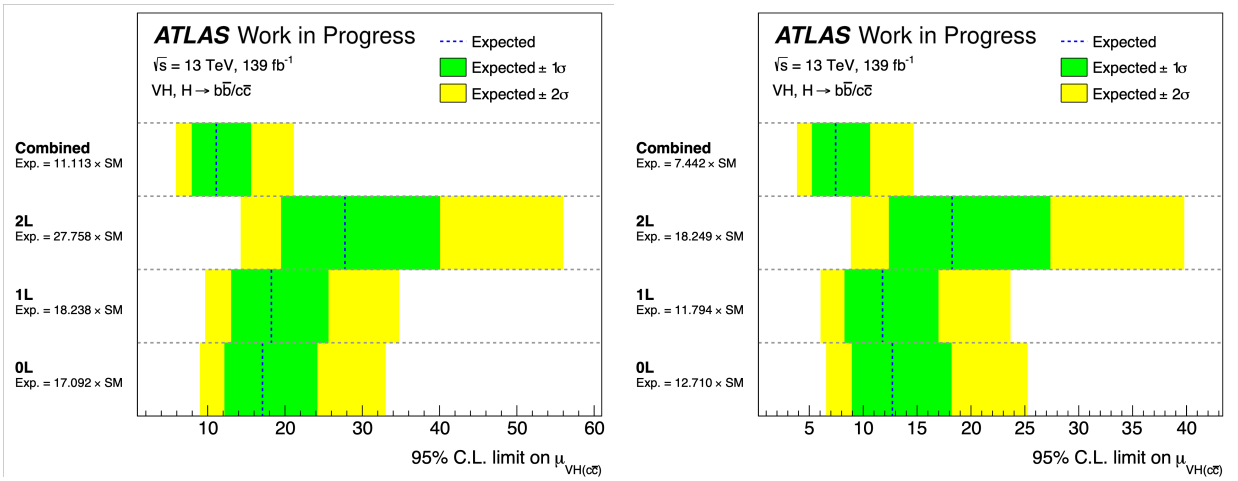


Figure 7.7: ROC curves comparing $H \to c\bar{c}$ signal efficiency vs. background rejection for DNN and Transformer "TwoHead" MVA models in the (a) zero, (b) one and (c) two lepton channels.



(a) (b)

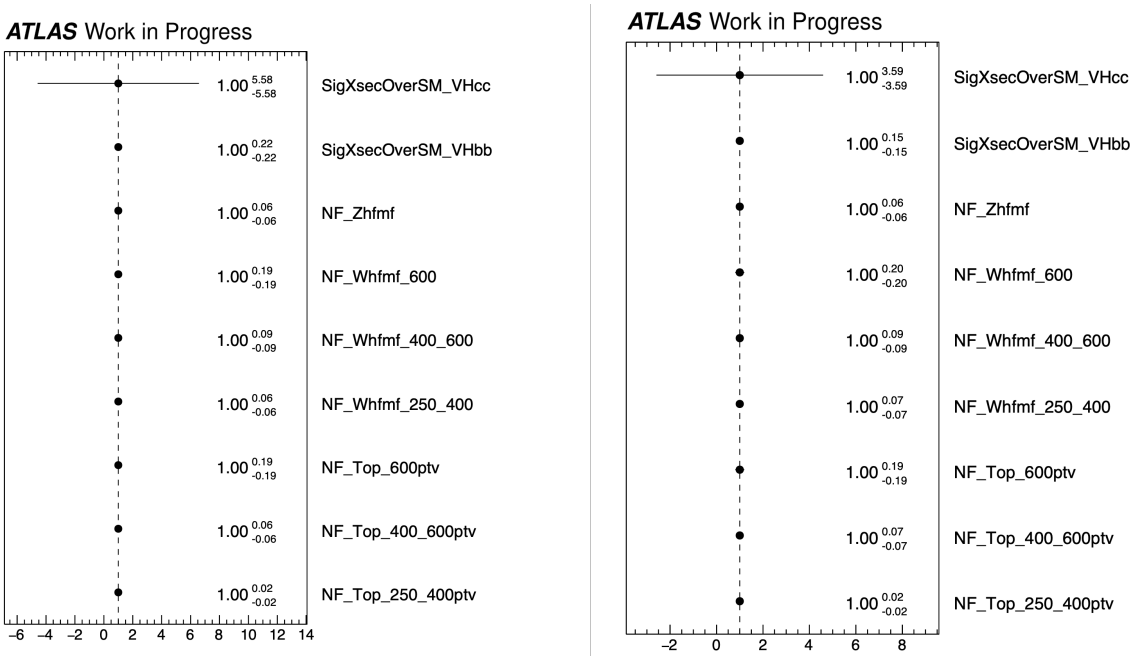
Figure 7.8: Diagrams showing 95% confidence level upper limits for $\mu_{VHc\bar{c}}$ by lepton channel and in an Asimov $VH(\to b\bar{b}/c\bar{c})$ fit using (a) m_J or (b) the MVA discriminant as a discriminating variable.

across all regions and applied in the zero and two lepton channels. For W+jets and Z+jets, heavy and mixed flavor components[8] have a single combined floating normalization while light flavor components have none given their lower occurrence frequency. All floating normalizations are correlated between flavor categories. The parameters of interest are the two signal strengths $\mu_{VHb\bar{b}}$ and $\mu_{VHc\bar{c}}$ which are measured across four fits: one for each lepton channel and one combined between all channels.

Results are compared between fits using the jet mass m_J as a discriminating variable and those using the "LeptonHead" MVA described above. In the latter case, a separate set of models are trained for the $VH(\to b\bar{b})$ SRs but the overall strategy remains identical. A non-linear transformation described in detail in section 5.1.4 is applied to the MVA output scores to maximize analysis sensitivity, producing 8 bins in each region. Figure 7.8 shows results for the expected 95% confidence level upper limits of the POI $\mu_{VHc\bar{c}}$ for each fit.

The two lepton channel provides the smallest expected sensitivity while the zero and one lepton channel are the most sensitive in the m_J and MVA fit respectively due to the larger number of signal events in these regions. The combined upper limit for $\mu_{VHc\bar{c}}$ is 7.442 × SM in the MVA fit and 11.113 × SM in the m_J fit. Both values are improved respective to the measured "legacy" analysis result of 11.5 × SM detailed in chapter 5, with the MVA limit suggesting sensitivity improvements of up to $\sim$ 35%. Uncertainties on all floating normalization factors and both POIs for each fit are shown in Fig. 7.9. While no strong conclusions can be drawn from these numbers given the missing treatment

[8] As before, heavy refers to bb and cc, mixed to bc, bl and cl and light to ll.



(a) (b)

Figure 7.9: Diagrams showing uncertainties on floating normalization factors and POIs in an Asimov $VH(\to b\bar{b}/c\bar{c})$ fit using (a) m_J or (b) the MVA discriminant as a binning variable.

of systematic uncertainties, these expected values demonstrate significant promise for the analysis strategy under investigation. Furthermore, any sensitivity gains from the boosted regime are additive to analogous improvements in the resolved regime stemming from the use of updated flavor identification algorithms in a combined analysis scenario. Both of these facts justify a full investigation of the inclusion of boosted $VH(\to c\bar{c})$ regions in future $VH(\to b\bar{b}/c\bar{c})$ measurements.

While uncertainties on floating normalization factors are similar between the m_J and MVA fit strategies, the MVA fit provides a $\sim 50\%$ improvement in the uncertainties on both POIs respective to the jet mass fit. To further demonstrate its viability, all signal regions binned in the respective MVA discriminants are shown in Fig. 7.10 for $VH(\to b\bar{b})$ and Fig. 7.11 for $VH(\to c\bar{c})$.

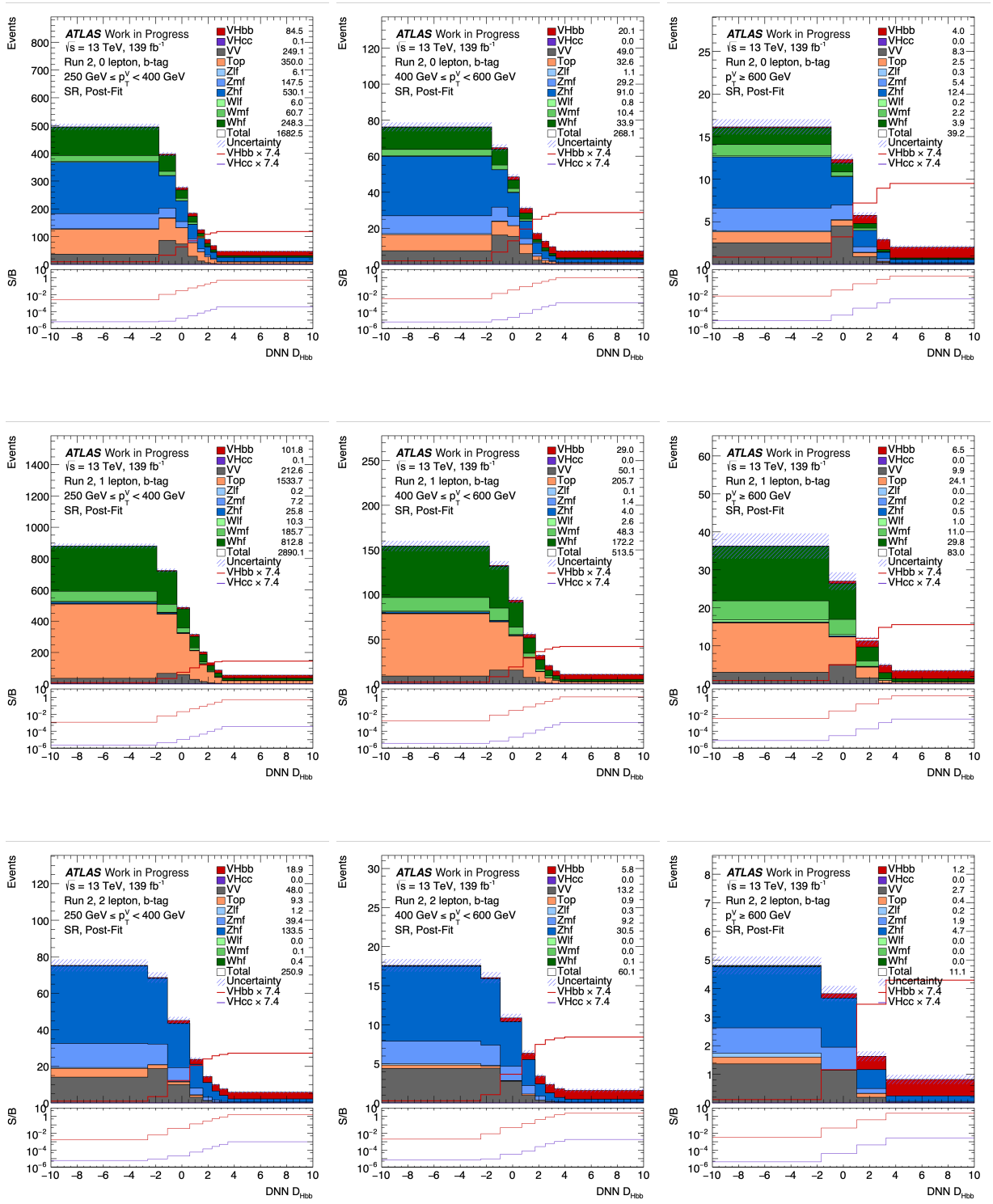


Figure 7.10: Figures showing post-fit Asimov distributions for all signal regions in the boosted $VH(\to b\bar{b})$ regime arranged horizontally by p_T^V category and vertically by lepton channel.

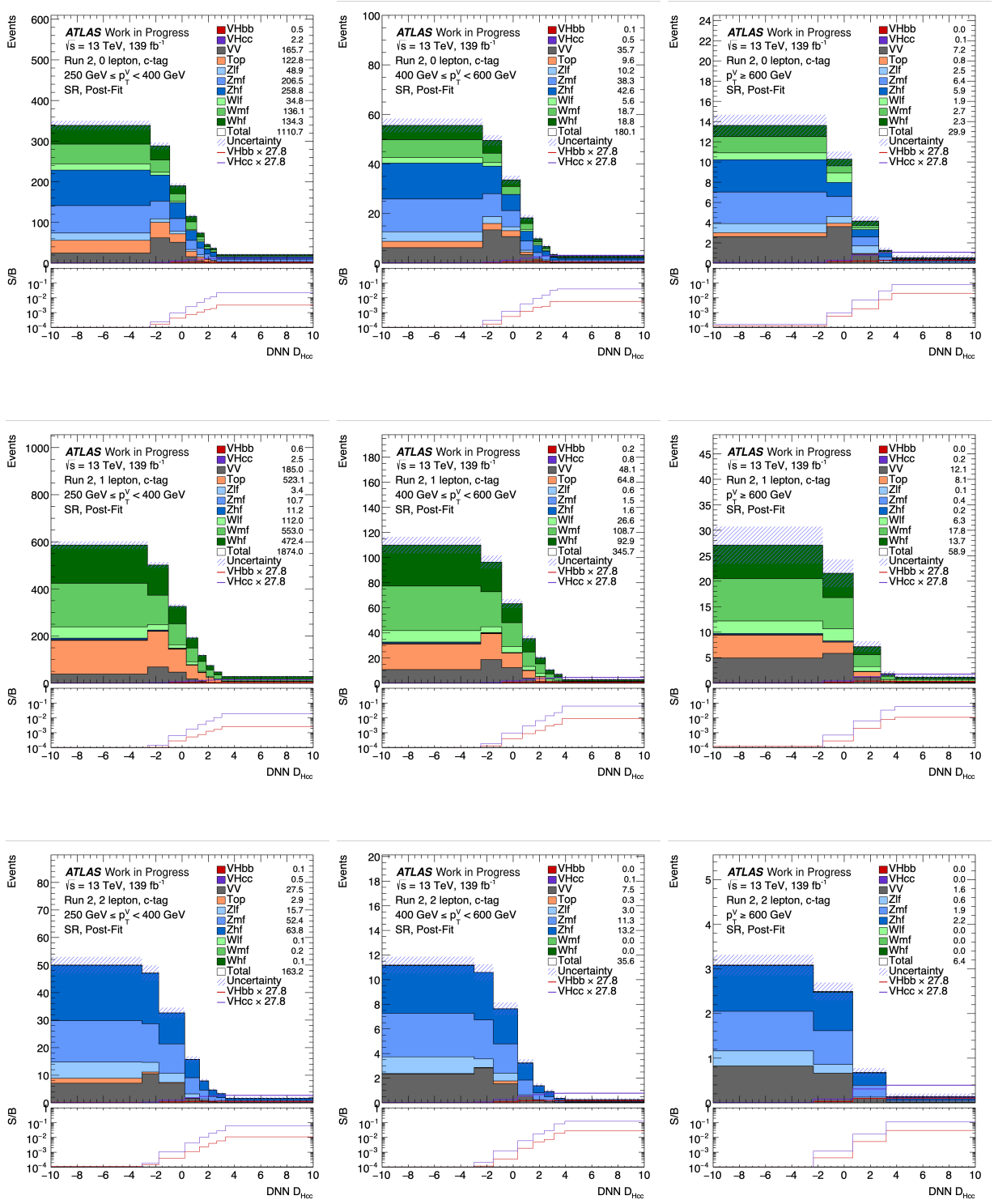


Figure 7.11: Figures showing post-fit Asimov distributions for all signal regions in the boosted $VH(\rightarrow c\bar{c})$ regime arranged horizontally by p_T^V category and vertically by lepton channel.

7.3 Future Work in $VH(\to c\bar{c})$

This chapter provides a blueprint for a future full run 2 plus run 3 combined measurement of the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ processes in the high momentum regime. To facilitate this, techniques from the previous "legacy" analysis summarized in chapter 5 and machine learning based jet flavor identification summarized in chapter 6 are brought together in this previously unexplored portion of the $VH(\to c\bar{c})$ phase space. At an expected sensitivity of $7.442 \times$ SM, this new measurement is expected to be competitive with prior run 2 results despite the stricter p_T^V cuts and modified jet reconstruction. The addition of run 3 data is expected to further improve sensitivity given the factor of two increase in statistics.

Finalization of this analysis requires the development of a full systematics model, including modeling uncertainties as well as experimental systematics updated for run 3 event reconstruction. Once complete, a combination of a `GN2` or `GN3` based resolved and a `GN3X` based boosted measurement of $VH(\to c\bar{c})$ will bring an even greater understanding of the Higgs to charm coupling. While a 5σ observation of $VH(\to c\bar{c})$ requires upgrades to detector and collider technology, the ATLAS experiment has made significant strides over the past few years to broaden our understanding of the charm quark, the Higgs boson and the physical universe as a whole.

Bibliography

[1] J. B. Adams. The CERN Proton Synchrotron. *Nature*, 185:568–572, Feb 1960.

[2] J. B. Adams. The European 300 GeV Accelerator. In *Proceedings of the 8th International Conference on High-Energy Accelerators*, pages 25–31. CERN, 1971.

[3] S. Alioli, P. Nason, C. Oleari, and E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06(IFUM-952-FT, IPPP-10-11, DCPT-10-22):043, 2010.

[4] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), July 2014.

[5] F. T. Avignone, S. R. Elliott, and J. Engel. Double Beta Decay, Majorana Neutrinos, and Neutrino Mass. *Rev. Mod. Phys.*, 80(2):481–516, 2008.

[6] B. Mellado et al. Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. Technical report, CERN, 2012.

[7] J. L. Ba, J. R. Kiros, and G. E. Hinton. Layer Normalization. `https://arxiv.org/abs/1607.06450`, 2016.

[8] C. Badiali, F. A. Di Bello, G. Frattari, E. Gross, V. Ippolito, M. Kado, and J. Shlomi. Parameterization with Neural Networks. `https://arxiv.org/abs/2004.02665`, 2020.

[9] R. Bailey and P. Collier. Standard Filling Schemes for Various LHC Operation Modes. Technical Report LHC-PROJECT-NOTE-323, CERN, 2003.

[10] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. Parton Distributions from LHC, HERA, Tevatron and Fixed Target Data: MSHT20 PDFs. *The European Physical Journal C*, 81(4):341, 2021.

[11] P. Baldi, P. Sadowski, and D. Whiteson. Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*, 5(1), July 2014.

[12] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali. Parton distributions with LHC data. *Nuclear Physics B*, 867(2):244–289, 2013.

[13] J. Barr, D. Biswas, M. Draguet, P. Gadow, E. Haines, O. Karkout, D. Kobylianskii, W. S. Lai, M. Leigh, N. Luongo, I. Oleksiyuk, N. Pond, S. Rettie, A. Vaitkus, S. Van Stroud, and J. Wagner. Salt: Multimodal Multitask Machine Learning for High Energy Physics. *Journal of Open Source Software*, 10(112):7217, 2025.

[14] K. Becker, F. Caola, A. Massironi, B. Mistlberger, P. F. Monni, X. Chen, S. Frixione, T. Gehrmann, N. Glover, K. Hamilton, A. Huss, S. Jones, A. Karlberg, M. Kerner, K. Kudashkin, J. M. Lindert, G. Luisoni, M. Mangano, S. Pozzorini, E. Re, G. P. Salam, E. Vryonidou, and C. Wever. Precise predictions for boosted Higgs production. *SciPost Phys. Core*, 7(CERN-TH-2020-074):001, 2024.

[15] G. Bellini, L. Ludhova, G. Ranucci, and F. L. Villante. Neutrino Oscillations. *Advances in High Energy Physics*, 2014(1):191960, 2014.

[16] N. Berger, C. Bertella, M. Bonanomi, N. Brahimi, T. P. Calvet, M. Calvetti, V. Dao, M. Delmastro, M. Duehrssen-Debling, P. Francavilla, Y. Haddad, S. Heim, J. Jovicevic, O. Kivernyk, M. Moreno Llacer, J. M. Langford, C. Li, G. Marchiori, J. A. McFayden, J. K. L. Michel, P. Milenovic, C. E. Pandini, E. Scott, F. J. Tackmann, K. Tackmann, L. Viliani, M. Xiao, and H." Yang. Simplified Template Cross Sections – Stage 1.1 and 1.2, 2025.

[17] S. N. Bose. Plancks Gesetz und Lichtquantenhypothese. *Zeitschrift für Physik*, 26(1):178–181, 1924.

[18] E. Bothmann, G. Singh Chahal, S. Höche, J. Krause, F. Krauss, S. Kuttimalai, S. Liebschner, D. Napoletano, M. Schönherr, H. Schulz, S. Schumann, and F. Siegert. Event generation with Sherpa 2.2. *SciPost Physics*, 7(3), Sep 2019.

[19] S. Brody, U. Alon, and E. Yahav. How Attentive are Graph Attention Networks? `https://arxiv.org/abs/2105.14491`, 2022.

[20] O. Brüning, M. Klein, S. Myers, and L. Rossi. 70 years at the high-energy frontier with the CERN accelerator complex. *Nature Reviews Physics*, 6(10):628–637, 2024.

[21] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, 2004.

[22] A. Buckley, G. Callea, A. J. Larkoski, and S. Marzani. An optimal observable for color singlet identification. *SciPost Phys.*, 9:026, 2020.

[23] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. De Roeck, J. Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, R. McNulty, A. Morsch, P. Nadolsky, V. Radescu, J. Rojo, and R. Thorne. PDF4LHC recommendations for LHC Run II. *Journal of Physics G: Nuclear and Particle Physics*, 43(2):023001, Jan 2016.

[24] N. Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963.

[25] M. Cacciari, G. P. Salam, and G. Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008.

[26] J. Campbell, T. Neumann, and Z. Sullivan. Single-top-quark production in the t-channel at NNLO. *Journal of High Energy Physics*, 2021(2), Feb 2021.

[27] M. Carena, C. Grojean, M. Kado, and V. Sharma. Status of Higgs boson physics. Technical report, Particle Data Group, October 6 2016.

[28] M. Carena and H. E. Haber. Higgs boson theory and phenomenology. *Prog. Part. Nucl. Phys.*, 50:63–152, 2003.

[29] É. Cartan. Les Groupes Projectifs qui ne laissent invariante aucune multiplicité plane. *Bulletin de la Société Mathématique de France*, 41:53–96, 1913.

[30] F. Cascioli, T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, L. Tancredi, and E. Weihs. zz production at hadron colliders in NNLO QCD. *Physics Letters B*, 735:311–313, July 2014.

[31] X. Chen, C. Liang, D. Huang, E. Real, K. Wang, Y. Liu, H. Pham, X. Dong, T. Luong, C.-J. Hsieh, Y. Lu, and Q. V. Le. Symbolic Discovery of Optimization Algorithms. `https://arxiv.org/abs/2302.06675`, 2023.

[32] S. Chennupati, G. Sistu, S. Yogamani, and S. A Rawashdeh. MultiNet++: Multi-Stream Feature Aggregation and Geometric Loss Strategy for Multi-Task Learning. `https://arxiv.org/abs/1904.08492`, 2019.

[33] L. Clissa, M. Lassnig, and L. Rinaldi. How big is Big Data? A comprehensive survey of data production, storage, and streaming in science and industry. *Frontiers in Big Data*, Volume 6 - 2023, 2023.

[34] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3:S08002, 2008.

[35] The ATLAS Collaboration. ATLAS liquid-argon calorimeter: Technical Design Report. Technical Report CERN-LHCC-96-041, ATLAS-TDR-002, CERN, 1996.

[36] The ATLAS Collaboration. ATLAS tile calorimeter: Technical Design Report. Technical Report CERN-LHCC-96-042, ATLAS-TDR-003, CERN, 1996.

[37] The ATLAS Collaboration. ATLAS muon spectrometer: Technical Design Report. Technical Report CERN-LHCC-97-022, CERN, 1997.

[38] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3:S08003, 2008.

[39] The ATLAS Collaboration. Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung. Technical Report ATLAS-CONF-2012-047, CERN, 2012.

[40] The ATLAS Collaboration. 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(1):1–29, Sep 2012.

[41] The ATLAS Collaboration. A neural network clustering algorithm for the ATLAS silicon pixel detector. *Journal of Instrumentation*, 9(09), Sep 2014.

[42] The ATLAS Collaboration. Operation and performance of the ATLAS semiconductor tracker. *Journal of Instrumentation*, 9(08), Aug 2014.

[43] The ATLAS Collaboration. Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-051, CERN, 2015.

[44] The ATLAS Collaboration. Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *The European Physical Journal C*, 75(1), Jan 2015.

[45] The ATLAS Collaboration. Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC. Technical Report ATL-PHYS-PUB-2015-045, CERN, 2015.

[46] The ATLAS Collaboration. Measurement of $w^{\pm}$ and z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ tev with the ATLAS detector. *Physics Letters B*, 759:601–621, Aug 2016.

[47] The ATLAS Collaboration. Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run. Technical Report ATL-PHYS-PUB-2016-012, CERN, 2016.

[48] The ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in pp collisions $\sqrt{s} = 8$ TeV using the ATLAS detector. *The European Physical Journal C*, 76(11), Oct 2016.

[49] The ATLAS Collaboration. Evidence for the $h \to b\bar{b}$ decay with the ATLAS detector. *Journal of High Energy Physics*, 2017(12), Dec 2017.

[50] The ATLAS Collaboration. Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment. Technical Report ATL-PHYS-PUB-2017-003, CERN, 2017.

[51] The ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS detector. *The European Physical Journal C*, 77(7), Jul 2017.

[52] The ATLAS Collaboration. Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties. *Journal of Instrumentation*, 12(05):P05002–P05002, May 2017.

[53] The ATLAS Collaboration. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton-proton collisions at the LHC. *The European Physical Journal C*, 77(5), May 2017.

[54] The ATLAS Collaboration. Secondary vertex finding for jet flavour identification with the ATLAS detector. Technical Report ATL-PHYS-PUB-2017-011, CERN, 2017.

[55] The ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *The European Physical Journal C*, 77(7), Jul 2017.

[56] The ATLAS Collaboration. ATLAS Event Displays: Higgs boson decaying to two b-quarks. `https://cds.cern.ch/record/2636049`, 2018. Accessed: 2026-04-01.

[57] The ATLAS Collaboration. Evaluation of theoretical uncertainties for simplified template cross section measurements of V-associated production of the Higgs boson. Technical Report ATL-PHYS-PUB-2018-035, CERN, 2018.

[58] The ATLAS Collaboration. Observation of $h \to b\bar{b}$ decays and vh production with the ATLAS detector. *Physics Letters B*, 786:59–86, Nov 2018.

[59] The ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13\ TeV$. *The European Physical Journal C*, 78(11), Nov 2018.

[60] The ATLAS Collaboration. Production and integration of the ATLAS Insertable B-layer. *Journal of Instrumentation*, 13(05), May 2018.

[61] The ATLAS Collaboration. Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC. Technical Report ATL-PHYS-PUB-2018-025, CERN, 2018.

[62] The ATLAS Collaboration. 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(11):970, 2019.

[63] The ATLAS Collaboration. Development of ATLAS Primary Vertex Reconstruction for LHC run 3. Technical Report ATL-PHYS-PUB-2019-015, CERN, 2019.

[64] The ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data. *Journal of Instrumentation*, 14(12), Dec 2019.

[65] The ATLAS Collaboration. Forward jet vertex tagging using the particle flow algorithm. Technical Report ATL-PHYS-PUB-2019-026, CERN, 2019.

[66] The ATLAS Collaboration. In situ calibration of large-radius jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector. *The European Physical Journal C*, 79(2), Feb 2019.

[67] The ATLAS Collaboration. Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016. *The European Physical Journal C*, 79(3), Mar 2019.

[68] The ATLAS Collaboration. Measurement of vh, $h \to b\bar{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector. *Journal of High Energy Physics*, 2019(5), May 2019.

[69] The ATLAS Collaboration. Modelling radiation damage to pixel sensors in the ATLAS detector. *Journal of Instrumentation*, 14(06), Jun 2019.

[70] The ATLAS Collaboration. Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS. Technical Report ATL-PHYS-PUB-2020-014, CERN, 2020.

[71] The ATLAS Collaboration. Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS. Technical Report ATL-PHYS-PUB-2020-019, CERN, 2020.

[72] The ATLAS Collaboration. Measurements of Higgs boson decays to bottom and charm quarks produced in association with a W or Z boson using the ATLAS experiment. Technical Report ATL-COM-PHYS-2020-765, CERN, 2020.

[73] The ATLAS Collaboration. How ATLAS detects particles: visualisation of particle paths in the detector. `https://videos.cern.ch/record/2770812`, May 2021. Accessed: 2025-10-03.

[74] The ATLAS Collaboration. Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *The European Physical Journal C*, 81(8), Aug 2021.

[75] The ATLAS Collaboration. Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physics Letters B*, 816:136204, May 2021.

[76] The ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 81(7), Jul 2021.

[77] The ATLAS Collaboration. Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton-proton collisions. *The European Physical Journal C*, 81(4), Apr 2021.

[78] The ATLAS Collaboration. A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery. *Nature*, 607(7917):52–59, 2022.

[79] The ATLAS Collaboration. Constraints on Higgs boson production with large transverse momentum using $h \to b\bar{b}$ decays in the ATLAS detector. *Physical Review D*, 105(9), May 2022.

[80] The ATLAS Collaboration. Data/MC Modelling for the VH($b\bar{b}/c\bar{c}$) Analysis. Technical Report ATL-COM-PHYS-2022-1140, CERN, 2022.

[81] The ATLAS Collaboration. Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector. *The European Physical Journal C*, 82(8), Aug 2022.

[82] The ATLAS Collaboration. Graph Neural Network Jet Flavour Tagging with the ATLAS Detector. Technical Report ATL-PHYS-PUB-2022-027, CERN, 2022.

[83] The ATLAS Collaboration. Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector. *Eur. Phys. J. C*, 82(95):95, 2022.

[84] The ATLAS Collaboration. Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment. *Journal of High Energy Physics*, 2022(8), Aug 2022.

[85] The ATLAS Collaboration. MVA training and evaluation of the VH(bb/cc) Legacy Analysis. Technical Report ATL-COM-PHYS-2022-1147, CERN, 2022.

[86] The ATLAS Collaboration. Object definition of VH(bb/cc) Legacy Analysis. Technical Report ATL-COM-PHYS-2022-1141, CERN, 2022.

[87] The ATLAS Collaboration. Performance of ATLAS Pixel Detector and Track Reconstruction at the start of Run 3 in LHC Collisions at $\sqrt{s} = 900$ GeV. Technical Report ATL-PHYS-PUB-2022-033, CERN, 2022.

[88] The ATLAS Collaboration. R=0.4 jets input comparison and Monte Carlo calibration with the ATLAS Detector. Technical Report ATL-PHYS-PUB-2022-038, CERN, 2022.

[89] The ATLAS Collaboration. Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data. Technical Report ATL-PHYS-PUB-2022-044, CERN, 2022.

[90] The ATLAS Collaboration. ATLAS flavour-tagging algorithms for the LHC Run 2 *pp* collision dataset. *The European Physical Journal C*, 83(7), Jul 2023.

[91] The ATLAS Collaboration. Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with z + jets events using 139 fb^{-1} of ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 83(8):728, 2023.

[92] The ATLAS Collaboration. Combined Measurement of the Higgs Boson Mass from the $h \to \gamma\gamma$ and $h \to zz^* \to 4\ell$ Decay Channels with the ATLAS Detector Using $\sqrt{s} = 7$, 8, and 13 TeV *pp* Collision Data. *Phys. Rev. Lett.*, 131:251802, Dec 2023.

[93] The ATLAS Collaboration. Measurement of the Higgs boson mass in the $h \to zz^* \to 4\ell$ decay channel using 139 fb^{-1} of $\sqrt{s} = 13$ tev pp collisions recorded by the ATLAS detector at the LHC. *Physics Letters B*, 843, Aug 2023.

[94] The ATLAS Collaboration. Measurement of the Higgs boson mass with $h \to \gamma\gamma$ decays in 140 fb^{-1} of $\sqrt{s} = 13$ TeV *pp* collisions with the ATLAS detector. *Physics Letters B*, 847:138315, 2023.

[95] The ATLAS Collaboration. Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS. Technical Report ATL-PHYS-PUB-2023-021, CERN, 2023.

[96] The ATLAS Collaboration. Electron and photon efficiencies in LHC run 2 with the ATLAS experiment. *Journal of High Energy Physics*, 2024(5), May 2024.

[97] The ATLAS Collaboration. Improving topological cluster reconstruction using calorimeter cell timing in ATLAS. *The European Physical Journal C*, 84(5), May 2024.

[98] The ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3. *Journal of Instrumentation*, 19(05):P05063, 2024.

[99] The ATLAS Collaboration. The ATLAS trigger system for LHC Run 3 and trigger performance in 2022. *Journal of Instrumentation*, 19(06), Jun 2024.

[100] The ATLAS Collaboration. ATLAS Public: Data Quality Results. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/DataQualityResults`, 2025. Accessed: 2025-10-03.

[101] The ATLAS Collaboration. ATLAS Public: Luminosity Public Results Run 3. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3`, 2025. Accessed: 2025-10-03.

[102] The ATLAS Collaboration. Identification of Boosted Higgs Bosons decaying into a pair of collimated τ-leptons decaying hadronically with Transformer Neural Networks in ATLAS. Technical Report ATL-PHYS-PUB-2025-026, CERN, 2025.

[103] The ATLAS Collaboration. Measurement of the associated production of a top-antitop-quark pair and a Higgs boson decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. *The European Physical Journal C*, 85(2), Feb 2025.

[104] The ATLAS Collaboration. Measurements of wh and zh production with Higgs boson decays into bottom quarks and direct constraints on the charm Yukawa coupling in 13 TeV pp collisions with the ATLAS detector. *Journal of High Energy Physics*, 2025(4), Apr 2025.

[105] The ATLAS Collaboration. Muon combined performance with 2022, 2023, and 2024 data. Technical Report ATL-COM-PHYS-2025-020, CERN, 2025.

[106] The ATLAS Collaboration. The performance of missing transverse momentum reconstruction and its significance with the ATLAS detector using 140 fb^{-1} of $\sqrt{s} = 13$ tev pp collisions. *The European Physical Journal C*, 85(6), Jun 2025.

[107] The ATLAS Collaboration. GN3: Multi-task, Multi-modal Transformers for Jet Flavour Tagging in ATLAS. Technical Report ATL-PHYS-PUB-2026-001, CERN, 2026.

[108] The ATLAS Collaboration. GN3X: Improved Transformer-based Tagger for Boosted Higgs Bosons in ATLAS. Technical Report ATL-PHYS-PUB-2026-013, CERN, 2026.

[109] The ATLAS Collaboration. Track and Vertex Reconstruction with the ATLAS Inner Detector. Technical Report CERN-EP-2026-096, CERN, 2026.

[110] The ATLAS Collaboration. Transforming jet flavour tagging at ATLAS. *Nature Communications*, 17(1), Jan 2026.

[111] The Belle II Collaboration. Measurement of the B^0 lifetime and flavor-oscillation frequency using hadronic decays reconstructed in Belle II data. *Phys. Rev. D*, 107:052011, 2023.

[112] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3:S08004, 2008.

[113] The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012.

[114] The CMS Collaboration. Evidence for the Higgs Boson decay to a Bottom Quark-Antiquark pair. *Physics Letters B*, 780:501–532, May 2018.

[115] The CMS Collaboration. Observation of Higgs Boson Decay to Bottom Quarks. *Physical Review Letters*, 121(12), Sept 2018.

[116] The CMS Collaboration. A search for the standard model Higgs boson decaying to charm quarks. *Journal of High Energy Physics*, 2020(3), Mar 2020.

[117] The CMS Collaboration. Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2020(12), Dec 2020.

[118] The CMS Collaboration. W^+W^- boson pair production in proton-proton collisions at $\sqrt{s} = 13$ tev. *Phys. Rev. D*, 102:092001, Nov 2020.

[119] The CMS Collaboration. Search for Higgs Boson Decay to a Charm Quark-Antiquark Pair in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. *Physical Review Letters*, 131(6), Aug 2023.

[120] The CMS Collaboration. Measurement of boosted Higgs bosons produced via vector boson fusion or gluon fusion in the $h \to b\bar{b}$ decay mode using LHC proton-proton collision data at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2024(12), Dec 2024.

[121] The CMS Collaboration. Measurement of the $t\bar{t}h$ and th production rates in the $h \to b\bar{b}$ decay channel using proton-proton collision data at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2025(2), Feb 2025.

[122] The LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation*, 3(LHCb-DP-2008-001):S08005, 2008.

[123] The LHCb Collaboration. Precise determination of the b^0_s-$\bar{b}^0_s$ oscillation frequency. *Nature Physics*, 18(1):1–5, Jan 2022.

[124] The Tevatron Collaboration. Evidence for a Particle Produced in Association with Weak Bosons and Decaying to a Bottom-Antibottom Quark Pair in Higgs Boson Searches at the Tevatron. *Physical Review Letters*, 109(7), Aug 2012.

[125] J. C. Collins, D. E. Soper, and G. F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5(ITP-SB-89-31):1–91, 1989.

[126] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse, and A. Salzburger. The new ATLAS track reconstruction (NEWT). *Journal of Physics: Conference Series*, 119(3):032014, Jul 2008.

[127] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), Feb 2011.

[128] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola, T. Reiter, and F. Tramontano. Automated one-loop calculations with GoSam. *The European Physical Journal C*, 72(3), Mar 2012.

[129] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. Technical report, CERN, 2017.

[130] T. Dao, Daniel Y. Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness. `https://arxiv.org/abs/2205.14135`, 2022.

[131] T. Darcet, M. Oquab, J. Mairal, and P. Bojanowski. Vision Transformers Need Registers. `https://arxiv.org/abs/2309.16588`, 2024.

[132] P. A. M. Dirac. On the Theory of Quantum Mechanics. *Proc. R. Soc. Lond. A*, 112(762):661–677, 1926.

[133] P. A. M. Dirac. The Quantum Theory of the Emission and Absorption of Radiation. *Proceedings of the Royal Society A*, 114:243–265, 1927.

[134] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[135] S. Agostinelli et al. Geant4 — a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[136] R. P. Feynman. Space-Time Approach to Quantum Electrodynamics. *Physical Review*, 76:769–789, 1949.

[137] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. Advantages of the Color Octet Gluon Picture. *Physics Letters B*, 47:365–368, 1973.

[138] 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(2):444–450, 1987.

[139] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nuclear Physics*, 22:579–588, 1961.

[140] L. Gonella. The ATLAS ITk detector system for the Phase-II LHC upgrade. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1045:167597, 2023.

[141] D. Grattarola, D. Zambon, F. M. Bianchi, and C. Alippi. Understanding Pooling in Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 35(2):2708–2718, Feb 2024.

[142] M. Grazzini, S. Kallweit, D. Rathlev, and M. Wiesemann. $w^{\pm}z$ production at hadron colliders in NNLO QCD. *Physics Letters B*, 761:179–183, Oct 2016.

[143] D. J. Gross and F. Wilczek. Ultraviolet Behavior of Non-Abelian Gauge Theories. *Physical Review Letters*, 30:1343–1346, 1973.

[144] LHC Top Physics Working Group. NNLO+NNLL top-quark-pair cross sections. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO`. Accessed: 2026-06-11.

[145] LHC Top Physics Working Group. Predictions at NNLO of Single Top-Quark Production Cross-Sections. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopNNLORef`. Accessed: 2026-06-11.

[146] Particle Data Group. Review of Particle Physics. *Prog. Theor. Exp. Phys.*, 2024(8):083C01, 2024.

[147] J. F. Gunion and H. E. Haber. The CP-conserving two-Higgs-doublet model: The approach to the decoupling limit. *Phys. Rev. D*, 67:075019, 2003.

[148] K. Hamilton, P. Nason, and G Zanderighi. MINLO: multi-scale improved NLO. *Journal of High Energy Physics*, 2012(10), Oct 2012.

[149] P. Hansen. Track Reconstruction. `https://indico.nbi.ku.dk/event/758/attachments/1684/2344/trackingnote.pdf`, September 2015.

[150] L. Heinrich, M. Feickert, G. Stark, and K. Cranmer. pyhf: pure-Python implementation of HistpFactory statistical models. *Journal of Open Source Software*, 6(58):2823, 2021.

[151] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[152] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, Nov 1997.

[153] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A." Zemla. TMVA - Toolkit for Multivariate Data Analysis. `https://arxiv.org/abs/physics/0703039`, 2009.

[154] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

[155] M. S. Hussain, M. J. Zaki, and D. Subramanian. Global Self-Attention as a Replacement for Graph Convolution. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 655–665. ACM, Aug 2022.

[156] J. Illingworth and J. Kittler. A survey of the Hough transform. *Computer Vision, Graphics, and Image Processing*, 44(1):87–116, 1988.

[157] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. `https://arxiv.org/abs/1502.03167`, 2015.

[158] T. Iwasaki, H. Hino, M. Tatsuno, S. Akaho, and N. Murata. Estimation of neural connections from partially observed neural spikes. *Neural Networks*, 108:172–191, 2018.

[159] F. James. MINUIT: Function Minimization and Error Analysis Reference Manual. `"https://cds.cern.ch/record/2296388"`, 1998.

[160] C. K. Joshi. Transformers are Graph Neural Networks. `https://arxiv.org/abs/2506.22084`, 2025.

[161] M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods*. Wiley-VCH, 2008.

[162] N. Kidonakis and N. Yamanaka. Higher-order corrections for tw production at high-energy hadron colliders. *Journal of High Energy Physics*, 2021(5), May 2021.

[163] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Oßptimization. `https://arxiv.org/abs/1412.6980`, 2017.

[164] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*, 2017.

[165] M. Kobayashi and T. Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49(2):652–657, 02 1973.

[166] D. Krohn, J. Thaler, and L.-T. Wang. Jets with variable R. *Journal of High Energy Physics*, 2009(06):059–059, Jun 2009.

[167] D. Krohn, J. Thaler, and L.-T. Wang. Jet trimming. *Journal of High Energy Physics*, 2010(2):84, 2010.

[168] Paul Langacker. The physics of heavy z' gauge bosons. *Rev. Mod. Phys.*, 81:1199–1244, Jul 2009.

[169] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152–155, 2001.

[170] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler. Soft Drop. *Journal of High Energy Physics*, 2014(5):146, 2014.

[171] LHC Higgs Working Group. LHC Higgs Working group Cross Sections and Figures. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGCrossSectionsFigures`, n.d. Accessed: 2025-10-09.

[172] Z. L. Liu and J. Gao. s-channel single top quark production and decay at next-to-next-to-leading-order in QCD. *Physical Review D*, 98(7), Oct 2018.

[173] I. Loshchilov and F. Hutter. SGDR: Stochastic Gradient Descent with Warm Restarts. `https://arxiv.org/abs/1608.03983`, 2017.

[174] I. Loshchilov and F. Hutter. Decoupled Weight Decay Regularization. `https://arxiv.org/abs/1711.05101`, 2019.

[175] M. Lubej. "A pretty diagram of the Standard Model". `https://www-f9.ijs.si/~lubej/SM.pdf`, n.d. Accessed: 2025-09-30.

[176] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, volume 28, pages 1–9, 2013.

[177] M. Magg and C. Wetterich. Neutrino Mass Problem and Gauge Hierarchy. *Physics Letters B*, 94:61–64, 1980.

[178] Z. Maki, M. Nakagawa, and S. Sakata. Remarks on the Unified Model of Elementary Particles. *Progress of Theoretical Physics*, 28(5):870–880, 11 1962.

[179] P. Minkowski. $\mu \to e\gamma$ at a Rate of One out of 10^9 Muon Decays? *Physics Letters B*, 67:421–428, 1977.

[180] I. Neutelings. B-tagging jets. `https://tikz.net/jet_btag/`, November 2024. Accessed: 2025-10-16.

[181] I. Neutelings. CMS coordinate system. `https://tikz.net/axis3d_cms/`, July 2024. Accessed: 2025-10-03.

[182] A. Newborough, M. Buzio, and R. Chritin. Upgrade of the CERN Proton Synchrotron Booster Bending Magnets for 2 GeV Operation. *IEEE Transactions on Applied Superconductivity*, 24(3):1–4, 2014.

[183] O. Aberle et al. *High-Luminosity Large Hadron Collider HL-LHC: Technical design report*. CERN Yellow Reports: Monographs. CERN, 2020.

[184] D. Oliveira Damazio. The ATLAS Trigger System. Technical Report ATL-DAQ-PROC-2024-015, CERN, 2024.

[185] H. Pernegger. The Pixel Detector of the ATLAS experiment for LHC Run-2. *Journal of Instrumentation*, 10(06):C06012, Jun 2015.

[186] H. D. Politzer. Reliable Perturbative Results for Strong Interactions? *Physical Review Letters*, 30:1346–1349, 1973.

[187] A. L. Read. Presentation of search results: the CLs technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693–2704, 2002.

[188] M. Rieger. *The $t\bar{t}H$ Process in the Standard Model of Particle Physics*, pages 5–39. Springer International Publishing, 2021.

[189] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[190] S. Heinemeyer et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. Technical report, CERN, 2013.

[191] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.

[192] J. Schwinger. On Quantum-Electrodynamics and the Magnetic Moment of the Electron. *Physical Review*, 73:416–417, 1948.

[193] M. H. Seymour. Heavy quark pair multiplicity in e^+e^- events. *Nuclear Physics B*, 436(1):163–183, 1995.

[194] Noam Shazeer. Glu variants improve transformer. `https://arxiv.org/abs/2002.05202`, 2020.

[195] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky, H. Maron, and N. Segol. Secondary Vertex Finding in Jets with Neural Networks. *The European Physical Journal C*, 81(6):540, Jun 2021.

[196] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.

[197] L. N. Smith and N. Topin. Super-Convergence: Very Fast Training of Neural Networks using Large Learning Rates. `https://arxiv.org/abs/1708.07120`, 2018.

[198] S. A. Solla, E. Levin, and M. Fleisher. Accelerated Learning in Layered Neural Networks. *Complex Systems*, 2:625–640, 1988.

[199] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[200] I. W. Stewart and F. J. Tackmann. Theory uncertainties for Higgs mass and other searches using jet bins. *Physical Review D*, 85(3), Feb 2012.

[201] G. 't Hooft. Renormalization of Massless Yang-Mills Fields. *Nuclear Physics B*, 33:173–199, 1971.

[202] G. 't Hooft and M. J. G. Veltman. Regularization and Renormalization of Gauge Fields. *Nuclear Physics B*, 44:189–213, 1972.

[203] The ATLAS Collaboration. ATLAS inner detector: Technical Design Report, 1. Technical Report CERN-LHCC-97-016, CERN, Apr 1997.

[204] The ATLAS Collaboration. ATLAS inner detector: Technical Design Report, 2. Technical Report CERN-LHCC-97-017, CERN, Apr 1997.

[205] The ATLAS Collaboration. ATLAS Pythia 8 tunes to 7 TeV data. Technical Report ATL-PHYS-PUB-2014-021, CERN, 2014.

[206] The ATLAS Collaboration. ATLAS EGamma Documentation. `https://atlas-egamma.docs.cern.ch/`, 2026. Accessed: 2026-05-26.

[207] The ATLAS Collaboration. ATLAS Jet/ETmiss Documentation. `https://atlas-jetetmiss.docs.cern.ch/`, 2026. Accessed: 2026-05-01.

[208] The ATLAS Collaboration. Identification of the charge of heavy-flavour jets using transformers with the ATLAS experiment. Technical Report ATL-PHYS-PUB-2026-005, CERN, 2026.

[209] The ATLAS Collaboration. ATLAS Magnet System. `http://atlas-magnet-dfs.web.cern.ch`, n.d. Accessed: 2025-10-03.

[210] The CMS Collaboration. Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production. *Nature Physics*, 18(11):1329–1334, 2022.

[211] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[212] The TRExFitter community. TRExFitter. `https://gitlab.cern.ch/TRExStats/TRExFitter`, 2026.

[213] The umami-hep community. puma: Plotting UMami API. `https://github.com/umami-hep/puma`, 2026.

[214] The umami-hep community. salt. `https://github.com/umami-hep/salt`, 2026.

[215] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention Is All You Need. `https://arxiv.org/abs/1706.03762`, 2017.

[216] W. Verkerke and D. Kirkby. The RooFit toolkit for data modeling. `https://arxiv.org/abs/physics/0306116`, 2003.

[217] M. Vretenar, J. Vollaire, R. Scrivens, C. Rossi, F. Roncarolo, S. Ramberger, U. Raich, B. Puccio, D. Nisbet, R. Mompo, S. Mathot, C. Martin, L. A. Lopez-Hernandez, A. Lombardi, J. Lettry, J. B. Lallement, I. Kozsar, J. Hansen, F. Gerigk, A. Funken, J. F. Fuchs, N. Dos Santos, M. Calviani, M. Buzio, O. Brunner, Y. Body, P. Baudrenghien, J. Bauche, and T. Zickler. *Linac4 design report*, volume 6 of *CERN Yellow Reports: Monographs*. CERN, 2020.

[218] M. Wang, D. Zheng, Z. Ye, Q. Gan, M. Li, X. Song, J. Zhou, C. Ma, L. Yu, Y. Gai, T. Xiao, T. He, G. Karypis, J. Li, and Z. Zhang. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. `https://arxiv.org/abs/1909.01315`, 2020.

[219] S. Weinberg. A Model of Leptons. *Physical Review Letters*, 19:1264–1266, 1967.

[220] R. Xiong, Y. Yang, D. He, K. Zheng, S. Zheng, C. Xing, M. Zhang, Y. Lan, L. Wang, and T.-Y. Liu. On Layer Normalization in the Transformer Architecture. `https://arxiv.org/abs/2002.04745`, 2020.

[221] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov, and A. Smola. Deep Sets. `https://arxiv.org/abs/1703.06114`, 2018.