Title
ATLAS Measurements of the Higgs Boson Coupling to the Top Quark in the Higgs to Diphoton Decay Channel

Author
Dickinson, Jennet Elizabeth

Publication Date
2020

Peer reviewed|Thesis/dissertation

ATLAS Measurements of the Higgs Boson Coupling to the Top Quark in the Higgs to Diphoton Decay Channel

by

Jennet Elizabeth Dickinson

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 Marjorie Shapiro, Chair
Professor Karl Van Bibber
Professor Barbara Jacak

Spring 2020

ATLAS Measurements of the Higgs Boson Coupling to the Top Quark in the Higgs to Diphoton Decay Channel

Abstract

ATLAS Measurements of the Higgs Boson Coupling to the Top Quark in the Higgs to Diphoton Decay Channel

by

Jennet Elizabeth Dickinson

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Marjorie Shapiro, Chair

During Run 2 of the Large Hadron Collider, the ATLAS experiment recorded proton–proton collision events with a center of mass energy of 13 TeV, the highest energy ever achieved in a collider. Analysis of this dataset, which corresponds to an integrated luminosity of 139 fb^{-1}, has provided new opportunities for precision measurements of the Higgs boson. The interaction between the Higgs boson and the top quark, the heaviest known particle, is a fundamental probe of the Higgs mechanism, which describes how Standard Model particles obtain mass.

The coupling of the Higgs boson to the top quark can be directly probed through the production of a Higgs boson in association with a top-antitop quark pair ($t\bar{t}H$) at the LHC. A statistical combination of ATLAS searches in multiple Higgs boson decay channels using 79.8 fb^{-1} of the Run 2 dataset yielded the first observation of $t\bar{t}H$ production at the level of 6.3σ.

The Higgs to diphoton ($\gamma\gamma$) decay channel is among the most sensitive for $t\bar{t}H$ measurements due to the excellent diphoton mass resolution of the ATLAS detector and the clean signature of this decay. A powerful multivariate analysis was developed to select $t\bar{t}H(\gamma\gamma)$ events based on the momenta of jets, photons, and leptons in the final state. The $t\bar{t}H$ cross section times $H \to \gamma\gamma$ branching fraction was measured to be $\sigma_{t\bar{t}H} \times B_{\gamma\gamma} = 1.59\ ^{+0.43}_{-0.39}$ fb using the full Run 2 ATLAS dataset. Top quarks were reconstructed with high probability in selected $t\bar{t}H(\gamma\gamma)$ events.

Following the observation of the $t\bar{t}H$ process, a direct measurement of the CP properties of the Higgs-top interaction was carried out in the $H \to \gamma\gamma$ decay channel. A CP-sensitive multivariate categorization was developed using reconstructed top quark variables, yielding an observed $t\bar{t}H$ significance of 5.2σ and an upper limit on the cross section of single-top-plus-Higgs (tH) production of 11.6 times the SM expectation. The observed data excludes a fully CP odd Higgs-top coupling at the level of 3.9σ, and the CP mixing angle is constrained to $|\alpha| > 43^\circ$ at 95% confidence level.

Contents

Contents **i**

List of Figures **v**

List of Tables **xiv**

1 Introduction **1**

2 Phenomenology **4**
2.1 Theory of strong interactions 4
2.1.1 Factorization of proton–proton collisions 7
2.2 Theory of electroweak interactions 9
2.2.1 The Higgs mechanism 10
2.2.2 Fermion masses and interactions 11
2.3 Higgs boson physics 12
2.3.1 Production mechanisms 12
2.3.2 Decay channels 17
2.3.3 Coupling measurement framework 19
2.4 Beyond the Standard Model Higgs 21
2.4.1 CP violation in the Higgs sector 21

3 The ATLAS experiment **25**
3.1 The Large Hadron Collider 25
3.2 The ATLAS detector 27
3.2.1 Inner Detector 28
3.2.2 Calorimeters 31
3.2.3 Muon Spectrometer 34
3.2.4 Trigger 36

4 Object Reconstruction **39**
4.1 Tracks and vertices 40
4.2 Photons 40
4.2.1 Primary vertex selection in diphoton events 42

4.3 Electrons . . . 43
4.4 Muons . . . 44
4.5 Hadronic Jets . . . 45
4.5.1 Flavor Tagging . . . 45
4.6 Missing transverse energy . . . 46

5 Data and Monte Carlo samples **48**
5.1 Data: LHC Run 2 . . . 48
5.1.1 Data control samples . . . 50
5.2 Simulated samples . . . 53
5.2.1 SM Higgs boson samples . . . 53
5.2.2 BSM Higgs samples . . . 55
5.2.3 Background samples . . . 57

6 Statistical model **58**
6.1 Motivation for categorization . . . 58
6.2 Signal and background shapes . . . 59
6.2.1 Signal model . . . 60
6.2.2 Background model . . . 61
6.3 Likelihood model . . . 63
6.3.1 Treatment of systematics . . . 64
6.3.2 Generation of Asimov datasets . . . 65
6.4 Sources of systematic uncertainty . . . 66
6.4.1 Theoretical uncertainties . . . 67
6.4.2 Experimental uncertainties . . . 69
6.5 Weighting data for presentation . . . 70

7 Machine Learning techniques **72**
7.1 Multivariate analysis . . . 72
7.1.1 Classification . . . 73
7.2 Boosted Decision Trees . . . 75
7.2.1 Boosting . . . 77
7.2.2 Multiclass BDT . . . 79
7.3 Hyper-parameter optimization . . . 79

8 Reconstruction of top decays **81**
8.1 Reconstruction strategy . . . 83
8.1.1 Hadronic Channel . . . 83
8.1.2 Leptonic Channel . . . 84
8.2 Training of the Top Reco BDT . . . 85
8.2.1 Constructing signal and background samples . . . 85
8.2.2 Training variables . . . 87

8.3 Performance . . . 91
8.3.1 Hadronic channel . . . 92
8.3.2 Leptonic channel . . . 95

9 Selection of $t\bar{t}H(\gamma\gamma)$ events **97**
9.1 Poisson number-counting significance . . . 97
9.2 Categorization of events . . . 99
9.2.1 Hadronic channel . . . 99
9.2.2 Leptonic channel . . . 106
9.3 Decomposition of continuum background . . . 113
9.3.1 Hadronic Channel . . . 113
9.3.2 Leptonic Channel . . . 116

10 First observation of $t\bar{t}H$ Production **118**
10.1 Combination of Higgs decay channels . . . 118
10.1.1 Input analyses . . . 118
10.1.2 Systematic uncertainties . . . 120
10.1.3 Results of combined analysis . . . 120
10.2 $t\bar{t}H$ in the diphoton decay channel . . . 124
10.2.1 Signal and background shapes . . . 124
10.2.2 Systematic uncertainties . . . 124
10.2.3 Results . . . 126

11 CP-sensitive categorization of $t\bar{t}H(\gamma\gamma)$ events **131**
11.1 Number-counting limit on CP mixing . . . 131
11.2 Sensitive observables . . . 132
11.3 Multivariate categorization strategy . . . 135
11.3.1 Hadronic channel . . . 138
11.3.2 Leptonic channel . . . 143
11.4 Determination of category boundaries . . . 148

12 Measurement of CP properties in $t\bar{t}H$ and tH **153**
12.1 Treatment of loop processes . . . 153
12.1.1 Kinematics of gluon fusion . . . 154
12.1.2 Effective loop couplings . . . 155
12.2 Signal and background model . . . 156
12.2.1 Signal and background shapes . . . 156
12.2.2 Signal yield dependence on mixing angle . . . 157
12.3 Systematic uncertainties . . . 162
12.4 Results . . . 165
12.4.1 Sensitivity to SM $t\bar{t}H$ and tH . . . 165
12.4.2 Constraints on CP mixing . . . 169

13 Conclusions **175**

A Search for TeV-scale gravity **177**

B Multijet modeling with MG5_aMC@NLO **185**

C Radiation damage in the IBL **190**
C.1 Single Event Upset . . . 191
C.2 Measurement of leakage current . . . 193

D Toy BDT Model **199**
D.1 Tuning single hyper-parameters . . . 199
D.2 Dataset Size . . . 209
D.3 Training Packages . . . 211
D.4 Multiclassification . . . 213

E $t\bar{t}H$ in the diphoton decay channel (79.8 fb^{-1}) **220**
E.1 Signal and Background Model . . . 220
E.2 Systematic Uncertainties . . . 221
E.3 Results (79.8 fb^{-1}) . . . 222
E.4 Top reconstruction in the 79.8 fb^{-1} analysis . . . 225
E.4.1 Decomposition of continuum background . . . 226
E.5 Event displays of $t\bar{t}H(\gamma\gamma)$ candidates . . . 227

F Higgs couplings in the diphoton decay channel **232**
F.1 Event Categorization . . . 232
F.1.1 Leptonic VH . . . 237
F.1.2 Vector boson fusion and hadronic VH . . . 237
F.1.3 Gluon fusion . . . 238
F.2 Signal and Background . . . 238
F.3 Systematic uncertainties . . . 241
F.4 Results (79.8 fb^{-1}) . . . 241
F.4.1 Higgs cross section by production mode . . . 243
F.4.2 Measurement of STXS . . . 248

G Multiclassification of $t\bar{t}H$, tH, and background **250**
G.1 Multiclass BDT training . . . 250
G.1.1 Hadronic channel . . . 251
G.1.2 Leptonic channel . . . 252
G.2 Performance . . . 262
G.2.1 Sensitivity estimate at 139 fb^{-1} . . . 265

Bibliography **267**

List of Figures

1.1 Constituents of the Standard Model of particle physics 1

2.1 Measurements of the running of the strong coupling 6
2.2 Feynman diagram of an example proton–proton event containing a hard scatter interaction 7
2.3 Parton distribution functions from the MMH14 PDF set 8
2.4 Functional form of the potential of the scalar field ϕ 10
2.5 Standard Model cross section as a function of proton–proton center of mass energy for different Higgs boson production modes 13
2.6 Lowest order Feynman diagrams of the three Higgs boson production modes with highest cross section in 13 TeV proton–proton collisions 14
2.7 Lowest order Feynman diagrams of the $gg \to ZH$ process 15
2.8 Lowest order Feynman diagrams of the $t\bar{t}H$ process 15
2.9 Lowest order Feynman diagrams of the tHj process in the five-flavor PDF scheme 15
2.10 Lowest order Feynman diagrams of the $tHjb$ process in the four-flavor PDF scheme 16
2.11 Lowest order Feynman diagrams of the tWH process 16
2.12 NLO Feynman diagrams of tW production and $t\bar{t}$ production that produce interference 16
2.13 Tree-level diagrams of Higgs boson decays to fermions and vector bosons . . . 17
2.14 Total width and branching fractions of the Standard Model Higgs boson as a function of Higgs boson mass 18
2.15 Lowest order Feynman diagrams contributing to the $H \to \gamma\gamma$ decay 18
2.16 Predicted and measured values of Higgs boson couplings to Standard Model particles and their relationship to particle mass 21
2.17 NLO $t\bar{t}H$ and tH cross sections as a function of the CP mixing angle 24

3.1 Diagram of the CERN collider complex 26
3.2 Diagram of the ATLAS detector 28
3.3 Diagram of the ATLAS Inner Detector layout. 29
3.4 Diagram of the layout of the ATLAS calorimeter systems. 32
3.5 Diagram of the ATLAS electromagnetic calorimeter 32

3.6 Diagram of ATLAS Tile Calorimeter modules . . . 33
3.7 Diagram of the ATLAS muon spectrometer . . . 34
3.8 Diagram of a cross section of the ATLAS muon spectrometer. . . . 35
3.9 Rate of collisions passing the Level 1 Trigger during a 13 TeV run in 2015 . . . 37
3.10 Rate of collisions passing the High Level Trigger during a 13 TeV run in 2015 . 38

4.1 Example signatures left in the ATLAS detector by different types of particles . 39
4.2 Efficiency of b-jet identification vs. rejection of light jets for ATLAS multivariate b-tagging discriminants in $t\bar{t}$ Monte Carlo . . . 46

5.1 Summary of ATLAS integrated luminosity and pileup conditions during Run 2 data taking . . . 48
5.2 The diphoton trigger rate for each year of Run 2 data taking . . . 49
5.3 Efficiency of the diphoton trigger as a function of the E_T and η of the sub-leading photon . . . 50

6.1 Diphoton invariant mass spectrum in Higgs boson Monte Carlo events with the fitted signal function overlaid, shown for two categories targeting $t\bar{t}H(\gamma\gamma)$ events 61
6.2 Example signal-plus-background Asimov datasets in one $t\bar{t}H(\gamma\gamma)$ analysis category . . . 66
6.3 Measured Higgs boson production cross section in bins of b-jet multiplicity in the $H \to ZZ \to 4\ell$ decay channel with 36.1 fb^{-1} of 13 TeV data . . . 68
6.4 Example diphoton invariant mass distribution in the unweighted and $\ln(1+S/B)$ weighted presentations . . . 71

7.1 Example ROC curves . . . 75
7.2 Diagram of an example tree of depth 3 . . . 76

8.1 Lowest order Feynman diagram of the $t\bar{t}H$ process with an all-hadronic decay of the $t\bar{t}$ system . . . 82
8.2 Training variables for the Top Reco BDT: p_T, η, ϕ, and mass of the W candidate 88
8.3 Training variables for the Top Reco BDT: p_T, η, ϕ, and energy of the b candidate 89
8.4 Training variables for the Top Reco BDT: pseudo-continuous b-tag scores of each jet in the triplet . . . 90
8.5 Training variables for the Top Reco BDT: triplet mass, ΔR between the W jets, and $\Delta R(W, b)$ . . . 90
8.6 Top Reco BDT score . . . 91
8.7 Break down of hadronic $t\bar{t}H(\gamma\gamma)$ Monte Carlo events containing no top candidate, a wrongly identified top candidate, and a correctly identified top candidate in bins of jet multiplicity . . . 93
8.8 Reconstructed top mass and Top Reco BDT score in events passing hadronic preselection . . . 95

8.9 Reconstructed top mass and Top Reco BDT score in events passing leptonic preselection . . . 96

9.1 Hadronic BDT training variables: $p_T/m_{\gamma\gamma}$, η, and ϕ of the leading and subleading photons . . . 101
9.2 Hadronic BDT training variables: missing E_T and ϕ direction of missing E_T . 101
9.3 Hadronic BDT training variables: p_T, η, and ϕ of the three jets with highest p_T 102
9.4 Hadronic BDT training variables: p_T, η, and ϕ of the jets ranked fourth, fifth, and sixth in p_T . . . 103
9.5 Distribution of the hadronic BDT score for $t\bar{t}H$ Monte Carlo signal, non-$t\bar{t}H$ Higgs processes, and data sidebands with 79.8 fb^{-1} of ATLAS data . . . 104
9.6 Predicted number-counting significance of four hadronic $t\bar{t}H(\gamma\gamma)$ categories, with different input variables used in BDT training . . . 106
9.7 Leptonic BDT training variables: $p_T/m_{\gamma\gamma}$, η, and ϕ of the leading and subleading photons . . . 108
9.8 Leptonic BDT training variables: p_T, η, and ϕ of the leading lepton, and missing E_T and ϕ direction of missing E_T . . . 109
9.9 Leptonic BDT training variables: p_T, η, and ϕ of the two jets with highest p_T 110
9.10 Leptonic BDT training variables: p_T, η, and ϕ of the jets ranked third and fourth in p_T . . . 111
9.11 Distribution of the leptonic BDT score for $t\bar{t}H$ Monte Carlo signal, non-$t\bar{t}H$ Higgs processes, and data sidebands with 79.8 fb^{-1} of ATLAS data . . . 112
9.12 Number-counting significance of three leptonic categories, with different input variables used in BDT training . . . 113
9.13 Templates of top mass distribution in $t\bar{t}H$, $t\bar{t}\gamma\gamma$ and $\gamma\gamma$ + jets Monte Carlo . . 114
9.14 Template fits to data in the hadronic $t\bar{t}H(\gamma\gamma)$ categories . . . 115

10.1 Summary of all $t\bar{t}H$ analysis categories, binned in the logarithm of the signal-to-background ratio of the category . . . 121
10.2 Measured $t\bar{t}H$ cross section broken down by ATLAS analysis and Higgs boson decay mode, normalized to the Standard Model prediction . . . 122
10.3 Expected and observed likelihood scans of the signal strength $\mu_{t\bar{t}H}$ for the Run 1, Run 2, and combined datasets . . . 123
10.4 Measured $t\bar{t}H$ cross section as a function of proton–proton center of mass energy 123
10.5 Diphoton invariant mass spectrum in the four hadronic $t\bar{t}H(\gamma\gamma)$ categories with 139 fb^{-1} of ATLAS data . . . 127
10.6 Diphoton invariant mass spectrum in the three leptonic $t\bar{t}H(\gamma\gamma)$ categories with 139 fb^{-1} of ATLAS data . . . 128
10.7 Fitted number of $t\bar{t}H(\gamma\gamma)$ and background events in each analysis category with 139 fb^{-1} of ATLAS data . . . 129
10.8 Weighted diphoton invariant mass spectrum of all seven $t\bar{t}H(\gamma\gamma)$ categories with 139 fb^{-1} of ATLAS data . . . 130

11.1 Particle-level distributions of Higgs boson p_T and η in $t\bar{t}H$, $tHjb$, and tWH Monte Carlo for different values of the CP mixing angle 133
11.2 Particle-level distributions of the angular separation between top and antitop quarks in $t\bar{t}H$ Monte Carlo for different values of the CP mixing angle 134
11.3 Particle-level distributions of top quark p_T and η in $t\bar{t}H$, $tHjb$, and tWH Monte Carlo for different values of the CP mixing angle 134
11.4 Particle-level distributions of m_{tH} in $t\bar{t}H$, $tHjb$, and tWH Monte Carlo for different values of the CP mixing angle 135
11.5 Diagram of the two-dimensional BDT plane and $t\bar{t}H(\gamma\gamma)$ CP categories in the hadronic and leptonic channels 136
11.6 Hadronic background rejection BDT score for $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle 136
11.7 Leptonic background rejection BDT score for $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle 137
11.8 Background rejection BDT score for the sum of $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle in the hadronic and leptonic channels . . 137
11.9 Hadronic CP BDT training variables: $p_T^{\gamma\gamma}$ and $\eta^{\gamma\gamma}$ 139
11.10 Hadronic CP BDT training variables: p_T, η, ϕ with respect to the Higgs boson candidate, and Top Reco BDT score of the first reconstructed top quark . . . 139
11.11 Hadronic CP BDT training variables: p_T, η, ϕ with respect to the Higgs boson candidate, and Top Reco BDT score of the second reconstructed top quark . . 140
11.12 Hadronic CP BDT training variables: $\Delta\eta(t_1, t_2)$ and $\Delta\phi(t_1, t_2)$ 140
11.13 Hadronic CP BDT training variables: m_{t_1H} and $m_{t_1t_2}$ 141
11.14 Hadronic CP BDT training variables: jet multiplicity, b-jet multiplicity, H_T, and missing E_T significance 141
11.15 Hadronic CP BDT training variables: $\Delta R_{\text{min}}^{\gamma j}$ and $\Delta R_{\text{min}}^{\gamma j2}$ 142
11.16 Hadronic CP BDT score for $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle 142
11.17 Two-dimensional BDT distribution in data passing hadronic preselection with $t\bar{t}H + tH$ signal contours overlaid 143
11.18 Leptonic CP BDT training variables: $p_T^{\gamma\gamma}$ and $\eta^{\gamma\gamma}$ 144
11.19 Leptonic CP BDT training variables: p_T, η, ϕ with respect to the Higgs boson candidate, and Top Reco BDT score of the first reconstructed top quark . . . 144
11.20 Leptonic CP BDT training variables: p_T, η, ϕ with respect to the Higgs boson candidate, and Top Reco BDT score of the second reconstructed top quark . . 145
11.21 Leptonic CP BDT training variables: $\Delta\eta(t_1, t_2)$ and $\Delta\phi(t_1, t_2)$ 145
11.22 Leptonic CP BDT training variables: m_{t_1H} and m_{t_1t2} 146
11.23 Leptonic CP BDT training variables: jet multiplicity, b-jet multiplicity, H_T, and missing E_T significance 146
11.24 Leptonic CP BDT training variables: $\Delta R_{\text{min}}^{\gamma j}$ and $\Delta R_{\text{min}}^{\gamma j2}$ 147
11.25 Leptonic CP BDT score for $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle 147

11.26 Two-dimensional BDT distribution in data passing leptonic preselection with $t\bar{t}H + tH$ signal contours overlaid . . . 148
11.27 Summary of the CP odd exclusion $Z_{CP}(90)$ and SM $t\bar{t}H + tH$ significance obtained by each set of boundaries in the brute force scan . . . 149
11.28 Number-counting exclusion as a function of CP mixing angle . . . 152

12.1 Hadronic background rejection BDT score and CP BDT score for ggF Monte Carlo samples . . . 154
12.2 Two-dimensional likelihood scan of κ_g and κ_γ from the ATLAS Higgs coupling combination . . . 156
12.3 Higgs event yields and purity of Higgs signal in the $t\bar{t}H(\gamma\gamma)$ CP categories at 139 fb^{-1} for different values of the CP mixing angle . . . 158
12.4 Parameterization as function of the CP mixing angle of the $t\bar{t}H$ signal yield in the $t\bar{t}H(\gamma\gamma)$ CP categories . . . 160
12.5 Parameterization as function of the CP mixing angle of the $tHjb$ signal yield in the $t\bar{t}H(\gamma\gamma)$ CP categories . . . 161
12.6 Parameterization as function of the CP mixing angle of the tWH signal yield in the $t\bar{t}H(\gamma\gamma)$ CP categories . . . 161
12.7 Inclusive parametrization of all processes dependent on the CP mixing angle . 162
12.8 Diphoton invariant mass spectrum in the $t\bar{t}H(\gamma\gamma)$ CP categories with high CP BDT score . . . 166
12.9 Diphoton invariant mass spectrum in the $t\bar{t}H(\gamma\gamma)$ CP categories with low CP BDT score . . . 167
12.10 Diphoton invariant mass spectrum in the $t\bar{t}H(\gamma\gamma)$ CP categories with intermediate CP BDT score . . . 168
12.11 Distribution of reconstructed top quark mass and reconstructed Higgs boson mass in the $t\bar{t}H(\gamma\gamma)$ CP analysis categories . . . 169
12.12 Fitted signal and background yields in the $t\bar{t}H(\gamma\gamma)$ CP analysis categories . . 170
12.13 Extracted $t\bar{t}H + tH$ signal yields in the $t\bar{t}H(\gamma\gamma)$ CP categories for $\alpha = 0°$, $\alpha = 90°$, and in data . . . 170
12.14 Likelihood scans over CP parameters α and $(\kappa_t \cos\alpha,\ \kappa_t \sin\alpha)$, with κ_g and κ_γ constrained by ATLAS combined Higgs boson data . . . 172
12.15 Likelihood scans over CP parameters α and $(\kappa_t \cos\alpha,\ \kappa_t \sin\alpha)$, with κ_g and κ_γ parameterized in terms of κ_t and α . . . 173
12.16 The $H \to \gamma\gamma$ branching fraction and SM $t\bar{t}H$ and tH cross sections as a function of κ_t . . . 173

A.1 Distribution of H_T and n_{jet} for a rotating black hole signal in 6 extra dimensions 177
A.2 Diagram of the Control Region, Validation Region, and Signal Region definitions 178
A.3 H_T distribution in data with 3.0 fb^{-1} in the $n_{\text{jet}} \geq 3$ bin, with qualifying fit functions overlaid . . . 180

A.4 Comparison of data and multijet Monte Carlo at each step of the bootstrap analysis in the $n_{\text{jet}} \geq 3$ bin . . . 182
A.5 Limit on the production of rotating black holes in 6 extra dimensions . . . 183
A.6 Limit on the production of string balls . . . 183

B.1 Leading jet p_T and H_T for the MG5_aMC@NLO multijet sample sliced in parton-level H_T . . . 186
B.2 Leading jet p_T and H_T for the MG5_aMC@NLO multijet sample sliced in parton-level H_T and unsliced . . . 187
B.3 Leading jet p_T and H_T using two different forms of the scale in MG5_aMC@NLO multijet Monte Carlo . . . 188
B.4 Leading jet p_T and inclusive jet p_T for MG5_aMC@NLO multijet samples using scale S_1 . . . 188
B.5 Variables sensitive to the third jet for MadGraph multijet Monte Carlo samples 189

C.1 Position of IBL staves in the $\hat{x} \times \hat{y}$ plane . . . 191
C.2 Fraction of pixels in which the second most significant TDAC bit flips in LHC fill 6371 . . . 192
C.3 Probability of bit flip due to SEU in the three DICE latches most important for IBL operations . . . 193
C.4 Sample IV curve of a silicon pixel sensor . . . 193
C.5 Measured leakage current in IBL modules as a function of integrated luminosity 196
C.6 Measured leakage current in IBL modules as a function of integrated luminosity, normalized to the measured leakage current in the 3D modules . . . 197
C.7 Measured leakage current in IBL modules as a function of integrated luminosity projected through Run 3 . . . 198

D.1 Distribution of a sample binary dataset without random noise and with random noise . . . 200
D.2 Distribution of BDT score obtained by a single decision tree for different values of MinChildWeight . . . 201
D.3 Distribution of BDT score obtained by a single decision tree for different values of MaxDepth . . . 202
D.4 Distribution of BDT score obtained by a single decision tree for different values of MinLossReduction . . . 203
D.5 Distribution of BDT score obtained after different numbers of boosting rounds 204
D.6 Distribution of BDT score obtained after 10 boosting rounds for different values of the LearningRate . . . 205
D.7 Distribution of BDT score obtained after 10 boosting rounds for different values of the Subsample parameter . . . 206
D.8 Distribution of BDT score obtained after 10 boosting rounds for different values of the L1 Regularization parameter α . . . 207

D.9 Distribution of BDT score obtained after 10 boosting rounds for different values of the L2 Regularization parameter λ . . . 208
D.10 Distribution of BDT score obtained from different numbers of randomly generated training points . . . 209
D.11 Distribution of BDT score obtained for the 10,000 event training set, with different relative normalization of signal and background . . . 210
D.12 Distribution of BDT score obtained from XGBoost and TMVA after different numbers of boosting rounds . . . 212
D.13 Distribution of a sample three-class dataset without random noise and with random noise . . . 213
D.14 Distribution of assigned class label obtained by a single decision tree for different values of MinChildWeight . . . 214
D.15 Distribution of "softprob" BDT scores obtained by a single decision tree for different values of MinChildWeight . . . 215
D.16 Distribution of assigned class label obtained by a single decision tree for different values of MaxDepth . . . 216
D.17 Distribution of "softprob" BDT scores obtained by a single decision tree for different values of MaxDepth . . . 217
D.18 Distribution of assigned class label obtained after different numbers of boosting rounds . . . 218
D.19 Distribution of "softprob" BDT scores obtained after different numbers of boosting rounds . . . 219

E.1 Diphoton invariant mass spectrum in the four hadronic $t\bar{t}H(\gamma\gamma)$ categories with 79.8 fb^{-1} of ATLAS data . . . 222
E.2 Diphoton invariant mass spectrum in the three leptonic $t\bar{t}H(\gamma\gamma)$ categories with 79.8 fb^{-1} of ATLAS data . . . 223
E.3 Fitted number of $t\bar{t}H(\gamma\gamma)$ and background events in each analysis category with 79.8 fb^{-1} of ATLAS data . . . 224
E.4 Weighted diphoton invariant mass spectrum of all seven $t\bar{t}H(\gamma\gamma)$ categories with 79.8 fb^{-1} of ATLAS data . . . 224
E.5 Comparison of the Early Top Reco BDT and the Top Reco BDT described in Chapter 8 in hadronic NTI data events. . . . 226
E.6 Template fit to data in the two hadronic $t\bar{t}H(\gamma\gamma)$ categories with highest S/B at 79.8 fb^{-1} . . . 227
E.7 Visualization of a data event selected in category $t\bar{t}H(\gamma\gamma)$ had 1 . . . 228
E.8 Visualization of a data event selected in category $t\bar{t}H(\gamma\gamma)$ had 1 . . . 229
E.9 Visualization of a data event selected in category $t\bar{t}H(\gamma\gamma)$ lep 1 . . . 230
E.10 Visualization of a data event selected in category $t\bar{t}H(\gamma\gamma)$ lep 1 . . . 231

F.1 Correspondence between Simplified Template Cross Section regions and $H \to \gamma\gamma$ analysis categories . . . 234

F.2 Breakdown of expected Higgs signal by STXS process in each analysis category 235
F.3 Breakdown of expected Higgs signal by production mode in each analysis category 236
F.4 Likelihood scan of the inclusive Higgs signal strength measured in $H \to \gamma\gamma$ with 79.8 fb^{-1} . . . 242
F.5 Weighted diphoton invariant mass spectrum of data from all 29 $H \to \gamma\gamma$ analysis categories with 79.8 fb^{-1} . . . 242
F.6 Weighted diphoton invariant mass spectrum of data from all 29 $H \to \gamma\gamma$ analysis categories with 79.8 fb^{-1}, with different Higgs boson processes treated as signal 244
F.7 Measured cross section by production mode in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}, normalized to the Standard Model prediction . . . 245
F.8 Likelihood scan of Higgs production cross sections in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1} . . . 245
F.9 Measured correlations between production mode cross sections in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1} . . . 246
F.10 Measured cross section ratio to ggF in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1} . . 247
F.11 Likelihood scan of Higgs production cross section ratio to ggF in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1} . . . 247
F.12 Measured Simplified Template Cross Sections in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}, normalized to the Standard Model prediction . . . 248
F.13 Measured correlations between STXS bins in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1} 249

G.1 Hadronic multiclass BDT training variables: $p_T^{\gamma\gamma}$ and $\eta^{\gamma\gamma}$ . . . 252
G.2 Hadronic multiclass BDT training variables: p_T, η, ϕ with respect to the Higgs candidate, and Top Reco BDT score of the first reconstructed top quark . . . 253
G.3 Hadronic multiclass BDT training variables: p_T, η, ϕ with respect to the Higgs candidate, and Top Reco BDT score of the second reconstructed top quark . . 254
G.4 Hadronic multiclass BDT training variables: $\Delta\eta(t_1, t_2)$ and $\Delta\phi(t_1, t_2)$ . . . 255
G.5 Hadronic multiclass BDT training variables: $m_{t_1 H}$ and $m_{t_1 t_2}$ . . . 255
G.6 Hadronic multiclass BDT training variables: jet multiplicity, b-jet multiplicity, H_T, and missing E_T significance . . . 256
G.7 Leptonic multiclass BDT training variables: $p_T^{\gamma\gamma}$ and $\eta^{\gamma\gamma}$ . . . 257
G.8 Leptonic multiclass BDT training variables: p_T, η, ϕ with respect to the Higgs candidate, and Top Reco BDT score of the first reconstructed top quark . . . 258
G.9 Leptonic multiclass BDT training variables: p_T, η, ϕ with respect to the Higgs candidate, and Top Reco BDT score of the second reconstructed top quark . . 259
G.10 Leptonic multiclass BDT training variables: $\Delta\eta(t_1, t_2)$ and $\Delta\phi(t_1, t_2)$ . . . 260
G.11 Leptonic multiclass BDT training variables: $m_{t_1 H}$ and $m_{t_1 t_2}$ . . . 260
G.12 Leptonic multiclass BDT training variables: jet multiplicity, b-jet multiplicity, H_T, and missing E_T significance . . . 261
G.13 Multiclass BDT output in the hadronic channel . . . 262
G.14 Multiclass BDT output in the leptonic channel . . . 263

G.15 Efficiency of each multiclass BDT Class on $t\bar{t}H$, tH, and background samples in the hadronic and leptonic channels 263
G.16 Efficiency of each hadronic multiclass BDT Class on $t\bar{t}H$, tH, and background samples, compared to the efficiency in the hadronic $t\bar{t}H(\gamma\gamma)$ binary BDT categories 264
G.17 Efficiency of each leptonic multiclass BDT Class on $t\bar{t}H$, tH, and background samples, compared to the efficiency in the leptonic $t\bar{t}H(\gamma\gamma)$ binary BDT categories 264

List of Tables

1.1 Summary of properties of the fundamental particles in the Standard Model . . 3

2.1 Most accurate available cross sections for Standard Model Higgs boson production, calculated at $m_H = 125.09$ GeV . . . 13
2.2 Branching fractions of the Standard Model Higgs boson . . . 18
2.3 Production modes of the Standard Model Higgs boson in 13 TeV proton–proton collisions and their parameterization in the κ framework . . . 20
2.4 Decay modes of the Standard Model Higgs boson and their parameterization in the κ framework . . . 20
2.5 NLO inclusive cross sections for the $t\bar{t}H$, $tHjb$, and tWH processes in different CP scenarios . . . 23

3.1 Extent of the sensitive region of each ATLAS Inner Detector subsystem . . . 29
3.2 Extent of each ATLAS calorimeter system . . . 31

5.1 Scale factors for estimating continuum background under the Higgs boson mass peak from data control regions . . . 51
5.2 Continuum background estimated by scaling the NTI and TI sidebands, compared to the fitted background in this region . . . 52
5.3 Summary of Monte Carlo generators used for matrix element calculation and parton showering . . . 53
5.4 Summary of Standard Model Higgs boson Monte Carlo samples . . . 54
5.5 Parameters used in the HC model to produce samples with CP mixing in the Higgs-top coupling . . . 56
5.6 NLO cross section times branching ratio for the $t\bar{t}H$, $tHjb$, tWH, and ggF processes for different CP scenarios, normalized with the K-factor . . . 56

6.1 QCD scale and PDF uncertainties on inclusive Higgs boson production cross sections at $\sqrt{s} = 13$ TeV with $m_H = 125$ GeV . . . 67
6.2 Summary of the theory uncertainties incorporated into the $H \to \gamma\gamma$ likelihood model . . . 69
6.3 Summary of the experimental uncertainties incorporated into the $H \to \gamma\gamma$ likelihood model . . . 70

8.1 Branching fraction of W boson decay modes . . . 81
8.2 Fraction of $t\bar{t}$ decays resulting in various numbers of leptons . . . 82
8.3 Summary of the reconstruction status of the primary and second tops in hadronic events with different jet multiplicity . . . 83
8.4 Summary of the reconstruction status of the primary and second tops in leptonic events with different lepton and jet multiplicities . . . 85
8.5 Summary of signal and background triplets in an example hadronic $t\bar{t}H$ event 86
8.6 Breakdown of the performance of the truth matching algorithm when acceptance cuts are applied to truth jets . . . 87
8.7 Percentage of events passing hadronic preselection that contain a correctly reconstructed object (top, W, or b) . . . 92
8.8 Percentage of events passing hadronic preselection in which the Top Reco BDT correctly reconstructs one or two tops . . . 93
8.9 Percentage of hadronic $t\bar{t}H(\gamma\gamma)$ Monte Carlo events containing no correct top candidate, wrongly identified top candidate, and correctly identified top candidate in bins of jet multiplicity . . . 94
8.10 Percentage of events passing hadronic preselection that contain a correctly reconstructed top . . . 94
8.11 Percentage of events passing leptonic preselection that contain a correctly reconstructed top . . . 95
8.12 Percentage of $t\bar{t}H(\gamma\gamma)$ Monte Carlo events in which top candidates are accurately reconstructed, comparing the Top Reco BDT and KLFitter . . . 96

9.1 Ranges in hadronic BDT score defining the four hadronic $t\bar{t}H(\gamma\gamma)$ categories . 104
9.2 Predicted event yield in the region 123 GeV $< m_{\gamma\gamma} <$ 127 GeV at 79.8 fb^{-1} in the hadronic BDT categories . . . 104
9.3 Predicted number-counting significance at 79.8 fb^{-1} in the hadronic BDT categories . . . 105
9.4 Ranges in leptonic BDT score defining the three leptonic $t\bar{t}H(\gamma\gamma)$ categories . 108
9.5 Predicted event yield in the region 123 GeV $< m_{\gamma\gamma} <$ 127 GeV at 79.8 fb^{-1} in the leptonic BDT categories . . . 109
9.6 Predicted number-counting significance at 79.8 fb^{-1} in the leptonic BDT categories 110
9.7 The top background fraction obtained in the hadronic $t\bar{t}H(\gamma\gamma)$ categories . . . 115
9.8 Data/Monte Carlo scale factors determined from the top background fraction for the hadronic $t\bar{t}H(\gamma\gamma)$ categories . . . 115
9.9 The dilepton and inclusive ≥ 1 lepton event yields in the leptonic $t\bar{t}H(\gamma\gamma)$ categories at 139 fb^{-1}, counted in the $m_{\gamma\gamma}$ sidebands . . . 116
9.10 Data/MC scale factors derived for each continuum background process in the leptonic channel . . . 116
9.11 The top background fraction obtained in the leptonic $t\bar{t}H(\gamma\gamma)$ categories . . . 117

10.1 Breakdown of the systematic uncertainties for the combined $t\bar{t}H$ cross section measurement at $\sqrt{s} = 13$ TeV . . . 121
10.2 Best fit parameter values for the signal $m_{\gamma\gamma}$ shape in each of the seven $t\bar{t}H(\gamma\gamma)$ categories at 139 fb^{-1} . . . 124
10.3 Results of the spurious signal test for the $t\bar{t}H(\gamma\gamma)$ categories at 139 fb^{-1} . . . 125
10.4 Underlying Event/Parton Showering uncertainty on $t\bar{t}H$ signal yields in each analysis category. . . . 125
10.5 Contribution of groups of systematic uncertainties to the total error on the measured $t\bar{t}H$ cross section times $H \to \gamma\gamma$ branching ratio with 139 fb^{-1} . . . 126
10.6 Event yields in the seven $t\bar{t}H(\gamma\gamma)$ categories with 139 fb^{-1} . . . 129

11.1 Optimized $t\bar{t}H(\gamma\gamma)$ CP category boundaries . . . 150
11.2 Predicted values of the figures of merit for the twenty $t\bar{t}H(\gamma\gamma)$ CP categories . 150
11.3 Background-only statistical uncertainty in each $t\bar{t}H(\gamma\gamma)$ CP category compared with the heavy flavor uncertainty on ggF and the UEPS uncertainties on $t\bar{t}H$, $tHjb$, and tWH . . . 151

12.1 Best fit parameter values for the signal $m_{\gamma\gamma}$ shape in each $t\bar{t}H(\gamma\gamma)$ CP category 157
12.2 Results of the spurious signal test for the $t\bar{t}H(\gamma\gamma)$ CP analysis categories . . . 159
12.3 Theory systematics impacting the overall rate of signal processes ($t\bar{t}H$, $tHjb$, and tWH) . . . 163
12.4 Theory systematics impacting the migration of $t\bar{t}H$, $tHjb$, tWH, and ggF events across categories . . . 164
12.5 Observed and expected $t\bar{t}H$ and $tH = tHjb + tWH$ yields per $t\bar{t}H(\gamma\gamma)$ CP category . . . 171

A.1 Functional forms considered for the H_T distribution of the multijet background 179
A.2 Observed and expected event yields at each luminosity step in the bootstrap analysis, shown in bins of inclusive jet multiplicity . . . 181
A.3 Expected and observed limits on the cross section of new physics in bins of inclusive jet multiplicity . . . 184

B.1 H_T slice boundaries used for the MG5_aMC@NLO multijet Monte Carlo sample 186

C.1 Summary of the IBL module groups and their positions in ATLAS . . . 195
C.2 Predicted Φ/L factors and fitted scale factors for each group of IBL modules . 195

D.1 Summary of hyper-parameter names in different BDT training packages . . . 211

E.1 Best fit parameter values for the signal $m_{\gamma\gamma}$ shape in each of the seven $t\bar{t}H(\gamma\gamma)$ categories at 79.8 fb^{-1} . . . 221
E.2 Results of the spurious signal test for the $t\bar{t}H(\gamma\gamma)$ categories at 79.8 fb^{-1} . . . 221
E.3 The top background fraction obtained in the hadronic $t\bar{t}H(\gamma\gamma)$ categories with highest S/B at 79.8 fb^{-1} . . . 226

F.1 Summary of the 29 $H \to \gamma\gamma$ analysis categories 233
F.2 Best fit parameter values for the signal $m_{\gamma\gamma}$ shape in each $H \to \gamma\gamma$ category, excluding the $t\bar{t}H$ categories 239
F.3 Results of the spurious signal test for the $H \to \gamma\gamma$ categories at 79.8 fb^{-1}, excluding the $t\bar{t}H$ categories 240
F.4 Expected and observed significance of each Higgs production mode in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1} 246

G.1 Sensitivity to $t\bar{t}H$ and tH obtained from the multiclass BDT categories and the binary $t\bar{t}H(\gamma\gamma)$ categories at 139 fb^{-1}, where the continuum background yield estimated from scaled NTI data 265
G.2 Sensitivity to $t\bar{t}H$ and tH obtained from the multiclass BDT categories and the binary $t\bar{t}H(\gamma\gamma)$ categories at 139 fb^{-1}, where the continuum background yield estimated by scaling the TI data sidebands 266

Acknowledgments

The work included in this Thesis could not have been accomplished without substantial support from colleagues, friends, and family, including many who have assumed more than one of these roles.

My educational journey has been a long one, and I am fortunate to have found tremendous mentors at every step of the way. Rest assured that I will do my best to pay it forward. I am grateful to you all: to the educators at St. Catherine's for guiding me towards a scientific career; to those at Columbia, especially Georgia, for giving me my first taste of research work; and to my ATLAS collaborators around the globe, especially the group at Lawrence Berkeley National Laboratory. Thank you especially to Haichen, not only for the scientific guidance, but also for pushing me to do more than I knew I could. And to Marjorie, for your constant counsel and support, and for setting an example of the type of scientist I aspire to become.

I would like to thank the friends who helped me to stay (mostly) sane throughout my doctoral work. Sam, Jeffrey, Kimber, Megan, and countless more: I have learned that there is no one place, be it Richmond, New York, Berkeley, or Geneva, where you can be close to everyone. I will miss you when I'm far away and celebrate when we are together again.

And finally, and most importantly, my family has been a constant source of support and encouragement throughout my life. Thank you for instilling in me the deep love of learning without which I would not be here today. I love you, Mom and Dad.

Chapter 1

Introduction

The Standard Model of Particle Physics (SM) has enjoyed widespread experimental success over the last fifty years: its predictions have been confirmed by many experiments studying a large range of energy scales, with few exceptions. The list of fundamental particles included in the Standard Model is complete to the best of our knowledge. These particles are shown in Figure 1.1, and an overview of their basic properties is presented in Table 1.1.

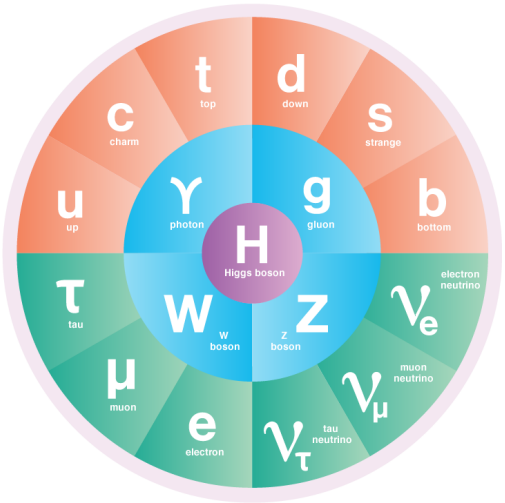


Figure 1.1: Constituents of the Standard Model of particle physics [1].

In addition to a list of fundamental particles, the Standard Model also provides a set of rules governing particle interactions. The Standard Model includes Strong, Electromagnetic, and Weak interactions, which are breifly described in Chapter 2. The well-known Gravitational interaction is not part the Standard Model, and its contribution is typically neglected in high energy physics experiments: the effects of gravity are more than 10^{30} times weaker than SM interactions. Many Beyond the Standard Model (BSM) theories seek to include gravity, and some theories predict signatures that are accessible to current collider experiments [2] [3]. Appendix A includes a search for such signatures.

The final Standard Model particle to be experimentally observed was the Higgs boson, which was discovered in 2012 by the ATLAS and CMS experiments [4] [5] during the first run of the Large Hadron Collider (LHC), which delivered a large dataset of proton–proton collisions with a center of mass energy $\sqrt{s}$ of 7-8 TeV. In the second run of the LHC (2015-2018), an even larger dataset and higher center of mass energy ($\sqrt{s} = 13$ TeV) have facilitated the observation of new Higgs boson production and decay modes, as well as improved measurements of Higgs couplings to fermions and bosons.

The coupling of the Higgs boson to the top quark, the heaviest particle in the Standard Model, is a fundamental probe of the nature of the Higgs boson. The measurement of this coupling is a sensitive test of the Higgs Mechanism, which describes how Standard Model particles obtain mass through their interaction with the Higgs field. Prior to LHC Run 2, the ATLAS and CMS collaborations performed indirect measurements of this coupling [6]. However, these measurements were predicated on assumptions that no BSM physics processes contribute to Higgs boson production. LHC Run 2 presents an opportunity to probe the Higgs-top coupling directly through Higgs boson production in association with a top-antitop quark pair ($t\bar{t}H$). This process is expected to be rare: it makes up only 1% of the Standard Model Higgs boson cross section at $\sqrt{s} = 13$ TeV.

This Thesis presents a series of measurements of $t\bar{t}H$ production using the ATLAS detector at the LHC. The first observation of $t\bar{t}H$ production is reported, which establishes a direct measurement of the interaction between the Higgs boson and quarks. Measurements of combined and single channel $t\bar{t}H$ cross sections are described, and a study of the charge-parity (CP) symmetry of the Higgs-top interaction is presented.

The measurements presented in this work focus on $t\bar{t}H$ production in conjunction with the Higgs boson decay to a pair of photons ($H \to \gamma\gamma$), which has a low branching fraction but clean signature in the ATLAS detector. $H \to \gamma\gamma$ measurements are performed by fitting to the diphoton invariant mass spectrum in data

$$m_{\gamma\gamma} = \sqrt{2E_{\gamma 1}E_{\gamma 2}(1 - \cos\theta_{\gamma\gamma})} \tag{1.1}$$

where $E_{\gamma 1}$ and $E_{\gamma 2}$ represent the energies of two photons separated by an angle $\theta_{\gamma\gamma}$. Due to the high diphoton mass resolution in ATLAS, the Higgs boson signal appears as a narrow peak in this distribution. By using a fit function composed of a smoothly falling background distribution plus a signal peak, the background for $H \to \gamma\gamma$ processes is estimated in a fully data-driven way.

This Thesis is organized as follows: Chapter 2 reviews the phenomenology of collider physics, including a detailed look at Higgs production in proton–proton collisions. Chapter 3 describes the experimental setup, including the Large Hadron Collider and the ATLAS detector. Chapter 4 describes the reconstruction of particles from their signatures in the ATLAS detector. The data and simulated samples used to study $t\bar{t}H$ production in conjunction with the diphoton decay ($t\bar{t}H(\gamma\gamma)$) are detailed in Chapter 5, and the statistical tools required to perform the fit to $m_{\gamma\gamma}$ are presented in Chapter 6. Chapter 7 gives a brief introduction to Machine Learning and the multivariate techniques used in this work. In

Chapter 8, a novel method for reconstruction of top quark decays using Machine Learning is presented. This technique is applied extensively in later Chapters to calculate variables dependent on top quark kinematics. The development of a multivariate analysis selection sensitive to $t\bar{t}H$ production in the $H \to \gamma\gamma$ decay channel is outlined in Chapter 9. Chapter 10 presents the first observation of $t\bar{t}H$ production in ATLAS, including measurements of combined and single channel cross sections. Finally, Chapters 11-12 present a search for violation of charge-parity (CP) symmetry in the interaction between the Higgs boson and the top quark, which could contribute to the observed asymmetry between matter and anti-matter in the universe [7]. Chapter 11 describes the development of multivariate analysis categories sensitive to the CP properties of $t\bar{t}H$ and tH production. Chapter 12 presents the results of this study, including stringent limits on CP mixing. Conclusions and final thoughts are included in Chapter 13. Appendices A-G contain short summaries of additional ATLAS work by the author of this Thesis.

Name	Symbol	Mass	Electric charge	Spin	Antiparticle
Bosons					
Higgs Boson	H	125.09 GeV	0	0	H
W Boson	W^+	80.379 GeV	+1	1	W^-
Z Boson	Z	91.188 GeV	0	1	Z
Photon	γ	0	0	1	γ
Gluon	g	0	0	1	g
Fermions					
Quarks					
Up	u	2.16 MeV	+2/3	1/2	$\bar{u}$
Down	d	4.67 MeV	-1/3	1/2	$\bar{d}$
Strange	s	93 MeV	-1/3	1/2	$\bar{s}$
Charm	c	1.27 GeV	+2/3	1/2	$\bar{c}$
Bottom	b	4.18 GeV	-1/3	1/2	$\bar{b}$
Top	t	172.9 GeV	+2/3	1/2	$\bar{t}$
Leptons					
Electron	e or e^-	511 keV	-1	1/2	$\bar{e}$ or e^+
Muon	μ or μ^-	105 MeV	-1	1/2	$\bar{\mu}$ or μ^+
Tau	τ or τ^-	1.78 GeV	-1	1/2	$\bar{\tau}$ or τ^+
Electron Neutrino	ν_e	< 2 eV	0	1/2	$\bar{\nu}_e$ $(? = \nu_e)$
Muon Neutrino	ν_μ	< 2 eV	0	1/2	$\bar{\nu}_\mu$ $(? = \nu_\mu)$
Tau Neutrino	ν_τ	< 2 eV	0	1/2	$\bar{\nu}_\tau$ $(? = \nu_\tau)$

Table 1.1: Summary of properties of the fundamental particles in the Standard Model, from the Particle Data Group [8].

Chapter 2

Phenomenology

The primary goal of high energy collider experiments is to test the predictions of the Standard Model. Direct searches for Beyond the Standard Model (BSM) physics exploit the high center of mass energy reached by colliders to produce new particles and processes. Indirect measurements rely on the high statistics of collider data to make precision measurements of Standard Model interactions in order to identify deviations that could arise from new physics via loop processes.

This Chapter presents the relevant theory for calculating Standard Model and Beyond the Standard Model predictions. Section 2.1 focuses on the strong interaction and the framework used to perform difficult calculations. The theoretical necessity of the Higgs boson is its role in electroweak symmetry breaking is discussed in the Section 2.2. Production and decay modes of the Higgs boson are described in detail in Section 2.3. Finally, Beyond the Standard Model Higgs scenarios, including the introduction of CP violation in the top Yukawa coupling, are addressed in Section 2.4.

2.1 Theory of strong interactions

In order to study physics at the highest energies and smallest distance scales, collider experiments focus on **hard scatter** interactions, in which there is a large momentum transfer (Q^2) between two incident particles. During the course of a single proton–proton event, many interactions occur with widely varying Q^2. This presents a major challenge for calculating properties of proton–proton events due to the changing behavior of the strong interaction at different energy scales.

Strong interactions between quarks and gluons are described by the non-Abelian $SU(3)_{\text{color}}$ gauge theory known as **quantum chromodynamics (QCD)**. The rate, or **cross section** σ, of any strong interaction can be expanded as a power series in the strong force coupling

α_S:

$$\sigma = \sum_{i}^{\infty} c_i \alpha_S^i \tag{2.1}$$

For small enough α_S, this expansion is **perturbative**: the contribution of each subsequent term is small compared to the previous terms. Summing over a finite number of terms therefore gives a good approximation of σ, and high order terms can be safely neglected. If the highest order term included in a calculation of σ is linear in α_S, the calculation is called "leading order" (LO). If the sum is completed up to the quadratic, cubic, or quartic term in α_S, the calculation is called "next-to-leading order" (NLO), "next-to-next-to-leading order" (NNLO), or "next-to-next-to-next-to-leading order" (N^3LO).

The primary computational challenges in collider physics arise from a unique property of QCD: interactions at low Q^2 (long distance) are stronger than interactions at high Q^2 (short distance). The running of the strong coupling (that is, its dependence on a **renormalization scale** μ_R) is shown in Figure 2.1 for $\mu_R^2 = Q^2$. As the renormalization scale is varied, each term in the perturbative expansion in Equation 2.1 changes, though the infinite sum remains unchanged. Typical values of μ_R^2 used in calculations are $\mathcal{O}(Q^2)$, which ensure that higher order terms remain small.

Due to the self-interaction of gluons, the strong coupling α_S grows large at small Q^2: this results in the **confinement** of quarks and gluons into $SU(3)_{\text{color}}$-singlet states (hadrons). However, at high Q^2 the strong coupling is very small and quarks are not confined to bound states: this is known as **asymptotic freedom**.

This behavior of the strong interaction contrasts strikingly with quantum electrodynamics, for example, where interaction rates fall nearly to zero for particles that are separated by very long distances. This difference in the behavior of these forces is a result of different running of the couplings: the electroweak coupling decreases with Q^2, and its magnitude varies by less than 10% over the range 1 GeV $< Q^2 < 10^4$ GeV [9].

In order to perform perturbative calculations, the smallness of the coupling is imperative. For this reason, calculations of collider phenomena are factorized into two components [10] [11] [12]:

1. Physics at short distance (high Q^2, perturbative in QCD)

2. Physics at long distance (low Q^2, non-perturbative in QCD)

The **factorization scale** μ_F separates these two regimes. Above the factorization scale, α_S is sufficiently small that cross sections can be calculated at fixed order in perturbation theory. Below the factorization scale, predictions rely heavily on data and phenomenological models of low-energy processes. Like the renormalization scale, μ_F^2 is typically taken to be $\mathcal{O}(Q^2)$.

Cross sections and kinematic distributions calculated at high enough order in perturbation theory are independent of the renormalization and factorization scales. However, small

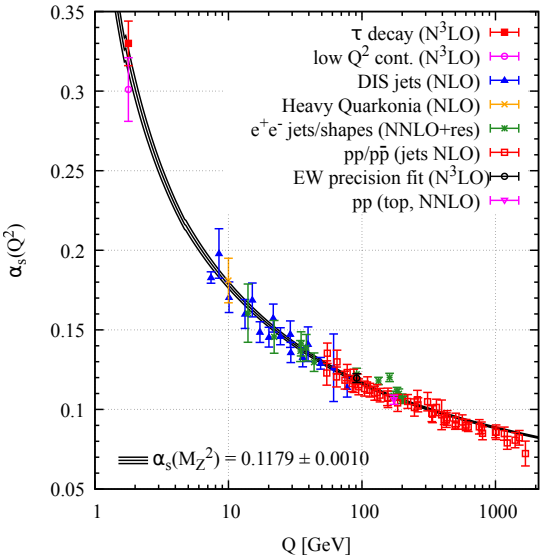


Figure 2.1: Measurements of the running of the strong coupling collected by the Particle Data Group [8].

residual dependence on μ_R and μ_F are expected from finite perturbative expansions: these effects are considered as systematic uncertainties.

2.1.1 Factorization of proton–proton collisions

In proton–proton collisions, one constituent parton (quark, antiquark, gluon) of each incoming proton participates in the hard scatter interaction. The remaining partons contribute low-energy hadronic activity to the final state of the event: this is referred to as the **underlying event (UE)**. Figure 2.2 shows a diagram of a proton–proton event containing a hard scatter interaction of cross section $\hat{\sigma}$.

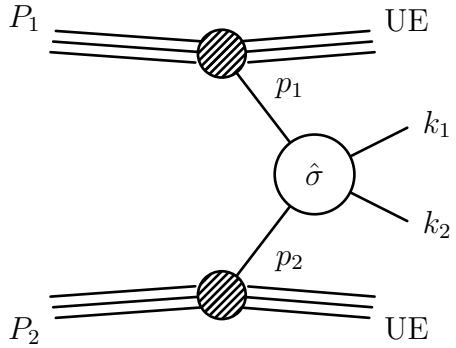


Figure 2.2: Feynman diagram of an example proton–proton event containing a hard scatter interaction of cross section $\hat{\sigma}$.

The incoming proton constituents and their momenta are described by **parton distribution functions (PDFs)**, which are convolved with parton–parton hard scatter cross sections $\hat{\sigma}$ to obtain the full proton–proton cross section σ:

$$\sigma(P_1, P_2) = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij}(p_1, p_2, \alpha_S(\mu_F^2), Q^2/\mu_F^2) \tag{2.2}$$

where i and j represent the different types of partons. The cross section of the hard scatter interaction $\hat{\sigma}$ is calculated to fixed order in perturbation theory for all combinations of incoming partons i, j.

The PDFs f_i depend on:

- the **Björken** x such that $\vec{p} = x\vec{P}$ where p is the parton momentum and P the proton momentum
- the factorization scale μ_F

Measurements of parton distribution functions are primarily based on data from Deep Inelastic Scattering, with additional input data from neutrino scattering and hadron collider measurements [12] [13]. An example of parton distribution functions at two values of $\mu_F^2 = Q^2$ are shown in Figure 2.3. At high values of Q^2, the gluon is by far the dominant proton constituent.

Heavy bottom quarks in the PDF can be treated in two ways: in the five-flavor (5F) PDF scheme, a massless b quark is included among the constituents of the proton; in the

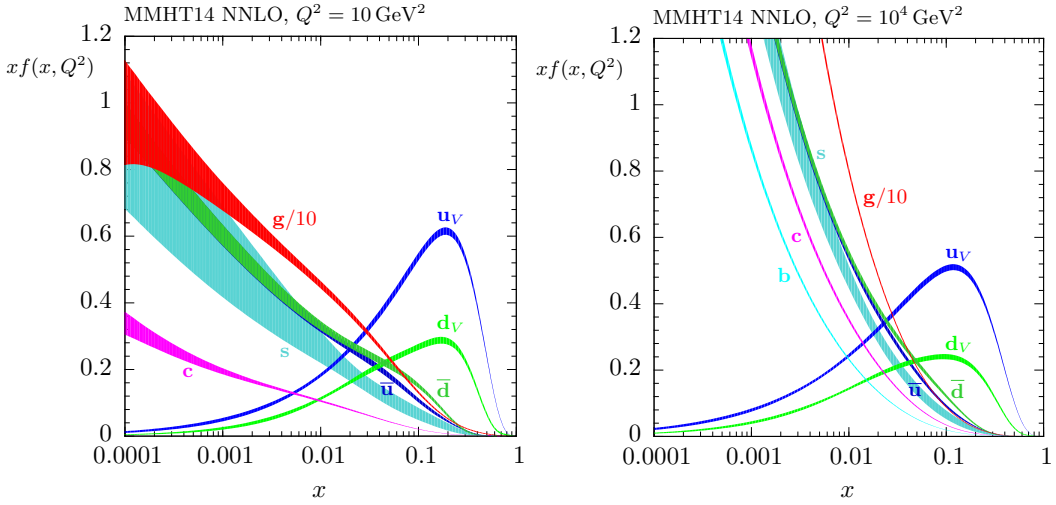


Figure 2.3: Parton distribution functions from the MMH14 PDF set [14], at two values of Q^2. Note the dominance of the gluon PDF at low x.

four-flavor (4F) scheme, the presence of b quarks in the proton is possible only through initial state gluon splitting.

After the calculation of the hard scatter cross section, additional QCD radiation below the factorization scale is modeled by **parton showering** algorithms. Parton showering algorithms simulate chains of low-energy and small-angle QCD radiation using a variety of models, including angular ordered [15] and dipole [16] showering. In order to combine a parton showering algorithm with a hard scatter matrix element calculation, matching and merging algorithms are used to ensure the calculations are consistent and no diagrams are over-counted [17].

During parton showering, quarks and gluons lose energy and pass from the regime of high Q^2 to that of confinement. During this transition, known as **hadronization** [18], partons form $SU(3)_{\mathrm{color}}$-singlet states, resulting in a columnated **jet** of hadrons. Multiple models are available for simulating hadronization and color flow in hadronic final states: the most widely used are the string [19] and cluster [20] models.

Because the underlying event, parton shower, and hadronization (UEPS) are low Q^2 processes that are non-perturbative in QCD, parameters of these models are typically tuned to data rather than predicted *a priori*. These **tunes** are applied extensively in predictions and simulations [21]. Differences in the predictions from multiple UEPS models are taken as a source of systematic uncertainty.

For a discussion and comparison of different models of strong interactions in Monte Carlo, see Appendix B.

2.2 Theory of electroweak interactions

The Glashow-Weinberg-Salam theory unifies the weak and electromagnetic forces under the gauge group $SU(2)_L \times U(1)_Y$ [22] [23] [24]. $SU(2)_L$ is the weak isospin group, and its generators in the spinor representation are the Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};\ \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix};\ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.3}$$

Weak isospin I_3 is the eigenvalue of the operator σ_3. The generator of $U(1)_Y$, the Abelian hypercharge group, is simply denoted 1, and its eigenvalue hypercharge by Y.

A scalar field ϕ is introduced that transforms as a doublet under $SU(2)_L$ and is symmetric under gauge transformations of the form

$$\phi \rightarrow e^{i\alpha^a \sigma^a/2} e^{i\beta/2} \phi \tag{2.4}$$

for all values of $\alpha^1, \alpha^2, \alpha^3$ and β. The electroweak Lagrangian is constructed from a kinetic term and a potential term for ϕ:

$$\mathcal{L} = |\mathcal{D}_\mu \phi|^2 + V(\phi) \tag{2.5}$$

where $\mathcal{D}_\mu$ is the electroweak covariant derivative. To ensure invariance of the kinetic term under the gauge transformation in Equation 2.4, the covariant derivative must take the form

$$\mathcal{D}_\mu \phi = (\partial_\mu - i\frac{g'}{2} B_\mu - i\frac{g}{2} W_\mu^j \sigma_j)\phi \tag{2.6}$$

where $W^{1,2,3}$ and B are the field strengths of the gauge groups $SU(2)_L$ and $U(1)_Y$, respectively. Expanding Equation 2.5 in terms of field tensors and making use of Pauli matrix identities, the Lagrangian becomes

$$\mathcal{L} = |\partial_\mu \phi|^2 + \frac{g'^2}{2} B_\mu B^\mu |\phi|^2 + \frac{g^2}{2} W_\mu^j W_j^\mu |\phi|^2 + \frac{g'g}{4} B_\mu W^{k\mu} (\phi^\dagger \sigma_k \phi + (\sigma_k \phi)^\dagger \phi) + V(\phi) \tag{2.7}$$

The following substitutions simplify the interpretation:

$$\begin{aligned} W_\mu^\pm &= \frac{W_\mu^1 \pm i W_\mu^2}{\sqrt{2}} \\ Z_\mu &= \frac{g' B_\mu - g W_\mu^3}{\sqrt{g'^2 + g^2}} \\ A_\mu &= \frac{g' B_\mu + g W_\mu^3}{\sqrt{g'^2 + g^2}} \end{aligned} \tag{2.8}$$

The transformed field strengths W_μ^+, W_μ^- and Z_μ correspond to the W and Z bosons, and A_μ corresponds to the electromagnetic field and has eigenvalue $Q = I_3 + Y$.

The simplified Lagrangian becomes

$$\mathcal{L} = |\partial_\mu \phi|^2 + \frac{g'^2 + g^2}{2} Z_\mu Z^\mu |\phi|^2 + \frac{g^2}{2} W_\mu^+ W^{-\mu} |\phi|^2 + V(\phi) \tag{2.9}$$

2.2.1 The Higgs mechanism

The theory of weak interactions thus far contains only massless gauge bosons. However, the masses of the W and Z bosons have been measured to be substantial (80.379 GeV and 91.188 GeV, respectively [8]). The introduction of mass for the W and Z bosons while maintaining the gauge invariance of $\mathcal{L}$ is accomplished through **electroweak symmetry breaking**.

The above gauge symmetry is spontaneously broken by means of the Higgs Mechanism [25] [26]. The potential for the doublet ϕ is taken to be

$$V(\phi) = -\mu^2\phi^\dagger\phi + \frac{\lambda}{2}(\phi^\dagger\phi)^2 \tag{2.10}$$

For $\mu^2 > 0$, this potential has a minimum, or **vacuum expectation value**, at $\phi = (0, v)$ where $v = \sqrt{\mu^2/\lambda}$. The shape of this potential is shown in Figure 2.4.

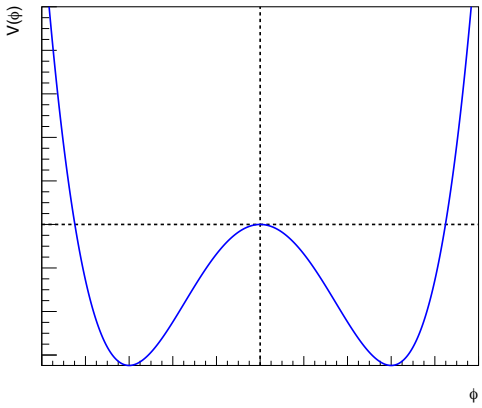


Figure 2.4: Functional form of the potential of the scalar field ϕ.

The field ϕ can be expressed as oscillations about this constant minimum:

$$\phi = (0, v + H) \tag{2.11}$$

where H represents the Higgs boson. Performing this substitution, the Lagrangian becomes

$$\mathcal{L} = (\partial_\mu H)^2 + \frac{g'^2 + g^2}{2} Z_\mu Z^\mu (v + H)^2 + g^2 W_\mu^+ W^{-\mu} (v + H)^2 + V(v + H) \tag{2.12}$$

Thus mass terms for the W and Z bosons have been obtained ($m_W = \frac{1}{2}vg$ and $m_Z = \frac{1}{2}v\sqrt{g'^2 + g^2}$) in addition to interactions between these particles and the Higgs field H. One gauge boson, the photon, remains massless and has no tree-level interactions with the Higgs boson.

The final term in Equation 2.12 gives a mass term for the Higgs boson ($m_H = \sqrt{2\lambda}v$), as well as interactions between the Higgs boson and itself. This Higgs boson self-coupling is a parameter of great interest, and experimental searches are underway [27] [28]. However, unless this coupling is enhanced by Beyond the Standard Model physics, there is no sensitivity to the Standard Model Higgs boson self-interaction at the LHC during Run 2 or even Run 3 [29] [30].

2.2.2 Fermion masses and interactions

The Lagrangian for a fermion field ψ can be written

$$\mathcal{L}_\psi = \bar{\psi} i \not{\mathcal{D}} \psi = \bar{\psi} i \gamma^\mu \mathcal{D}_\mu \psi \tag{2.13}$$

where the co-variant derivative $\mathcal{D}_\mu$ is that defined in Equation 2.6. This can be broken into separate terms for the left- and right-handed components of the field:

$$\bar{\psi} i \not{\mathcal{D}} \psi = \bar{\psi}_L i \not{\mathcal{D}} \psi_L + \bar{\psi}_R i \not{\mathcal{D}} \psi_R \tag{2.14}$$

Since the W boson couples only to the left-handed components of a fermion field, left-handed fermions are represented as doublets in the $SU(2)_L$ gauge group, while right-handed fermions are represented as singlets. Thus, right-handed fermions have weak isospin $I_3 = 0$ and hypercharge $Y = Q$, and left-handed fermions have $I_3 = \frac{1}{2}$ and $Y = Q \pm \frac{1}{2}$.

The Standard Model left-handed lepton doublets are

$$E_L^{(1)} = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} \; ; \; E_L^{(2)} = \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix} \; ; \; E_L^{(3)} = \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \tag{2.15}$$

and the left-handed quark doublets are

$$Q_L^{(1)} = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \; ; \; Q_L^{(2)} = \begin{pmatrix} c_L \\ s_L \end{pmatrix} \; ; \; Q_L^{(3)} = \begin{pmatrix} t_L \\ b_L \end{pmatrix} \tag{2.16}$$

The Lagrangian for a single generation of fermions (for simplicity) can be written

$$\mathcal{L} = \bar{E}_L i \not{D} E_L + \bar{e}_R i \not{D} e_R + \bar{Q}_L i \not{D} Q_L + \bar{u}_R i \not{D} u_R + \bar{d}_R i \not{D} d_R \tag{2.17}$$

Expansion of the covariant derivative gives kinetic terms for each fermion, as well as interactions between the fermions and gauge bosons.

While the mass terms for the W and Z bosons result directly from the Higgs mechanism, no fermion mass terms are explicitly required by electroweak symmetry breaking. However, mass terms can be naturally added with the form

$$\mathcal{L}_e = \lambda_e \bar{E}_{Li} \phi e_{Rj} \tag{2.18}$$

where ϕ is the Higgs doublet and λ_e is the (real-valued) electron Yukawa coupling. Continuing with the first generation leptons as an example, this becomes:

$$\mathcal{L}_e = \frac{-\lambda_e}{\sqrt{2}} \left[\left(\bar{\nu}_{eL}, \bar{e}_L\right) \begin{pmatrix} 0 \\ v+H \end{pmatrix} e_R + \bar{e}_R \left(0, v+H\right) \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} \right] \tag{2.19}$$

This yields both a mass term for the electron and a Higgs boson coupling to the electron:

$$\mathcal{L}_e = \frac{-v\lambda_e}{\sqrt{2}} \bar{e}_L e_R + \frac{-\lambda_e}{\sqrt{2}} H \bar{e}_L e_R \tag{2.20}$$

The electron mass obtained is $m_e = \lambda_e v/\sqrt{2}$, and the strength of the Higgs boson coupling is proportional to m_e. The same procedure can be used to generate mass terms and Higgs boson coupling terms for other charged leptons and down-type quarks. The gauge-invariant transformation $\phi \to -i\sigma_2\phi^*$ gives Yukawa terms of a slightly different form, which generate mass terms for up-type quarks (and Dirac neutrinos):

$$\mathcal{L}_u = -i\lambda_u \bar{Q}_L \sigma_2 \phi^* u_R \tag{2.21}$$

In the case of the top quark, which is the primary interest of this work, the mass and coupling terms are completely analogous to the case of the electron:

$$\mathcal{L}_t = \frac{-v\lambda_t}{\sqrt{2}} \bar{t}_L t_R + \frac{-\lambda_t}{\sqrt{2}} H \bar{t}_L t_R \tag{2.22}$$

with the top mass proportional to the Yukawa coupling via $m_t = \lambda_t v/\sqrt{2}$. Thus for the top quark (and for all fermions), the mass is not predicted directly by the Standard Model. Instead, the Standard Model provides an explicit prediction of the relationship between the mass of fermion f and its coupling to the Yukawa coupling λ_f.

2.3 Higgs boson physics

It was demonstrated in the previous Section that particles with higher mass have stronger couplings to the Higgs boson. This has been confirmed by all measurements of Higgs boson couplings to date [31] [32]. Higgs boson production is therefore dominated by interactions between the Higgs boson and the heaviest Standard Model particles: the top quark, W boson, and Z boson.

2.3.1 Production mechanisms

A Higgs boson can be produced in many ways in proton–proton collisions. Cross sections of the dominant Higgs boson production processes are shown in Figure 2.5 as a function of proton–proton center of mass energy. The best available calculation of the cross sections at

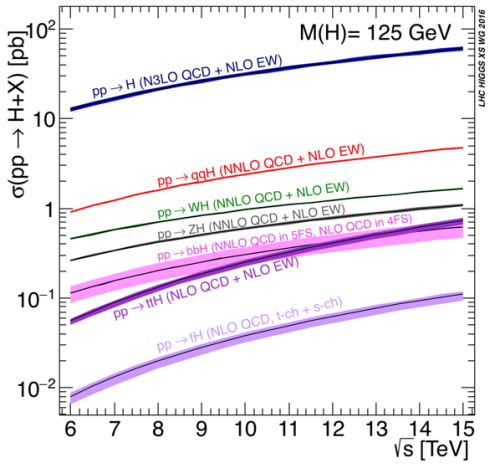


Figure 2.5: Standard Model cross section as a function of proton–proton center of mass energy for different Higgs boson production modes [34].

Process	$\sigma(\sqrt{s} = 13$ TeV) [pb]	Order of calculation
ggF	48.52	N^{3}LO(QCD) + NLO(EW)
VBF	3.779	NNLO(QCD) + NLO(EW)
WH	1.369	NNLO(QCD) + NLO(EW)
$qq/qg \to ZH$	0.7612	NNLO(QCD) + NLO(EW)
$gg \to ZH$	0.1227	NLO + NLL(QCD)
$t\bar{t}H$	0.5065	NLO(QCD) + NLO(EW)
$b\bar{b}H$	0.4863	5FS NNLO or 4FS NLO(QCD)
$tHjb$	0.07426	4FS NLO(QCD)
tWH	0.01517	5FS NLO(QCD)

Table 2.1: Most accurate available cross sections for Standard Model Higgs boson production, calculated at $m_H = 125.09$ GeV [34].

$\sqrt{s} = 13$ TeV is reported in Table 2.1. See [33] for Higgs boson production cross sections at lower center of mass energies.

In proton–proton collisions, the Higgs boson production cross section is dominated by gluon fusion through a quark loop (ggF, or ggH). In the Standard Model, this loop is dominated by the contribution from the top quark due to its high mass. In BSM models, massive particles that are charged under $SU(3)_{\text{color}}$ can also contribute to the rate of ggF production.

Sizable contributions to the total Higgs boson cross section also arise from vector boson fusion (VBF) and vector boson associated production, where a Higgs is radiated off of a W or Z boson ($qq/qg \to VH$ or simply VH, also known as Higgs-strahlung). Feynman diagrams of these three processes are shown in Figure 2.6. An additional small contribution to the ZH final state comes from gluon-initiated production ($gg \to ZH$). Feynman diagrams for this process are shown in Figure 2.7.

The primary focus of this work is a rare process in which a Higgs boson is produced in association with a top-antitop quark pair ($t\bar{t}H$). The lowest order Feynman diagrams for $t\bar{t}H$ production are shown in Figure 2.8. The $t\bar{t}H$ cross section comprises only 1% of the Standard Model Higgs boson cross section at $\sqrt{s} = 13$ TeV.

A Higgs boson is produced in association with a bottom-antibottom pair through diagrams analogous to those shown in Figure 2.8 for $t\bar{t}H$. At $\sqrt{s} = 13$ TeV, the $b\bar{b}H$ cross section is similar to that of $t\bar{t}H$.

Higgs boson production in association with a single top quark (tH) is also predicted by the Standard Model, but with much lower cross section than $t\bar{t}H$ due to destructive interference between Higgs-top and Higgs-W diagrams. The sum of all tH processes is displayed in light purple in Figure 2.5. Many models of BSM physics, including those discussed in Section 2.4.1, predict a large enhancement of tH cross sections: any observation of this process during LHC Run 2 would be a clear sign of new physics.

Figure 2.9 shows the lowest order diagrams for the tHj process as calculated five-flavor (5F) PDF scheme, which includes massless b quarks among the constituents of the proton. Figure 2.10 shows the leading diagrams for the tHj process in the four-flavor (4F) PDF scheme: these are similar to the 5F diagrams, but the incoming b quark is produced by gluon splitting in the 4F scheme.

The primary difference between the 4F and 5F calculations is in which diagrams are considered "higher-order." The two predictions converge if the calculations include terms to high enough order in α_S [35]. Because the shapes of kinematic variables such as Higgs boson p_T and bottom quark p_T are described with smaller uncertainties in the 4F scheme, the tHj process is calculated in the 4F scheme and referred to as $tHjb$ for the remainder of this work. The small contribution from quark-initiated s-channel tHj production is neglected.

The leading diagrams for tWH production are shown in Figure 2.11. At leading order, this process only occurs in the 5F PDF scheme. At next-to-leading order (NLO), there is interference between the tWH and $t\bar{t}H$ processes [36], as demonstrated in Figure 2.12. The treatment of this interference in simulation is discussed in Chapter 5.

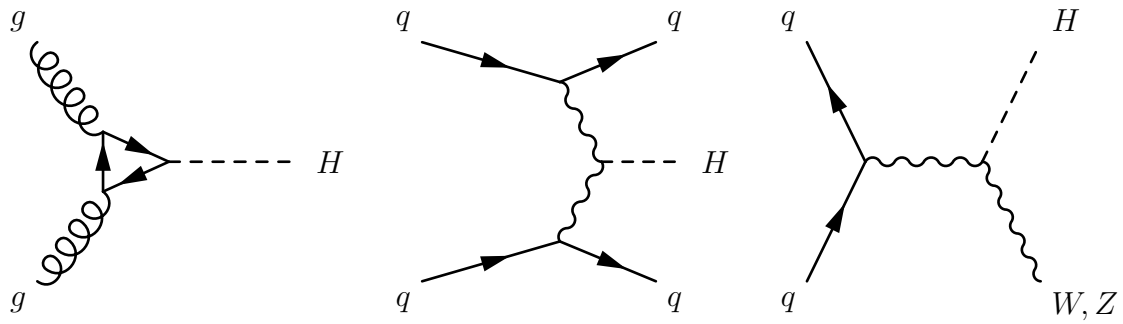


Figure 2.6: Lowest order Feynman diagrams of the three Higgs boson production modes with highest cross section in 13 TeV proton–proton collisions. From left to right: gluon–gluon fusion, vector boson fusion, vector boson associated production.

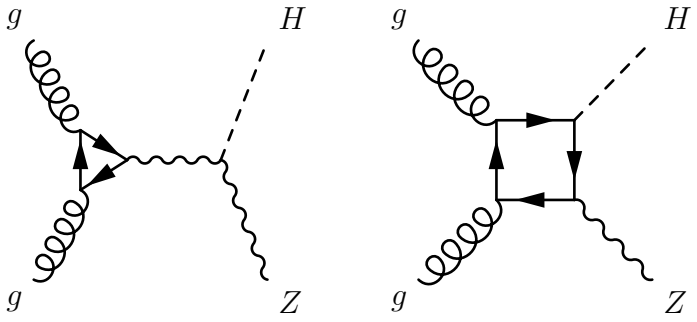


Figure 2.7: Lowest order Feynman diagrams of the $gg \to ZH$ process.

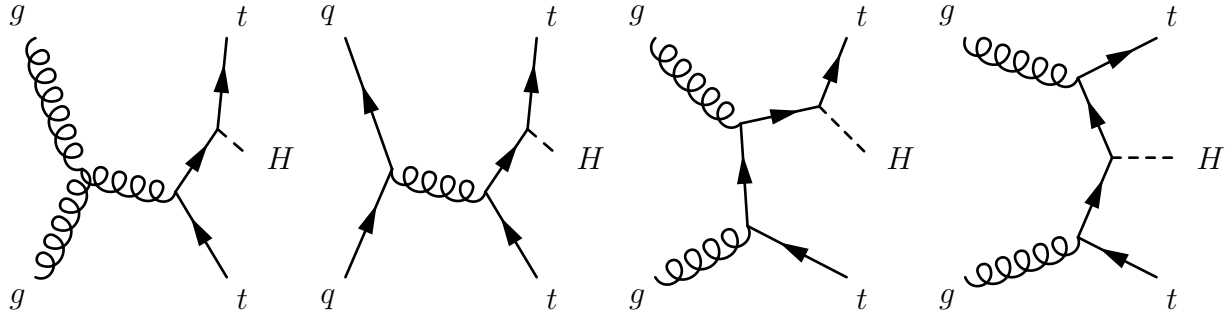


Figure 2.8: Lowest order Feynman diagrams of the $t\bar{t}H$ process. Additional diagrams can be obtained by allowing a Higgs boson to radiate from the antitop, or by crossing the lines of initial gluons in t-channel diagrams.

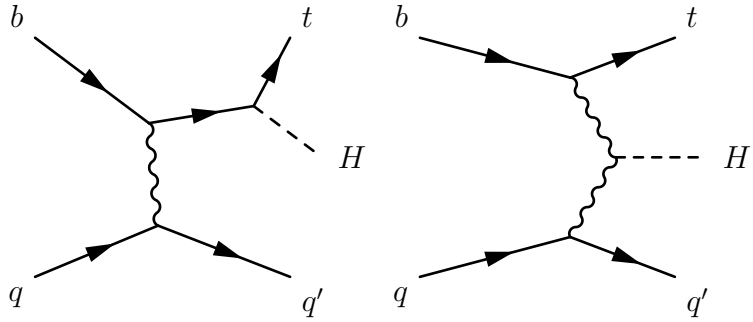


Figure 2.9: Lowest order Feynman diagrams of the tHj process in the five-flavor PDF scheme. Additional diagrams can be obtained by reversing the fermion lines, or by crossing the lines of initial particles.

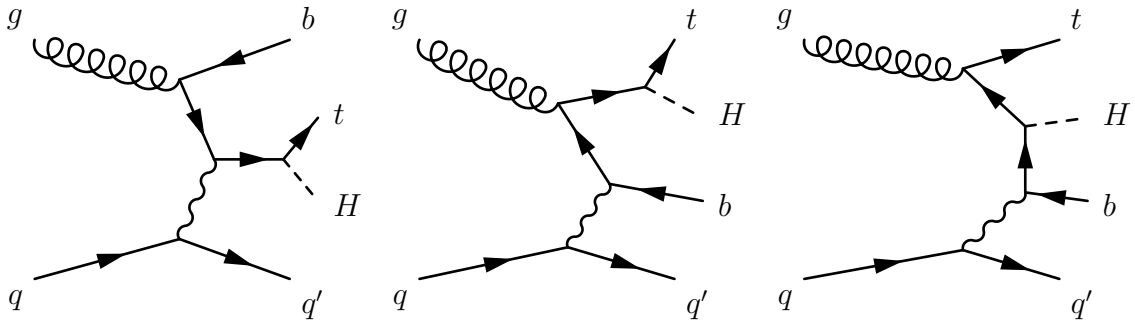


Figure 2.10: Lowest order Feynman diagrams of the $tHjb$ process in a four-flavor PDF scheme. Additional diagrams can be obtained by allowing a Higgs boson to radiate from the W boson, by reversing the fermion lines, or by crossing the lines of initial particles.

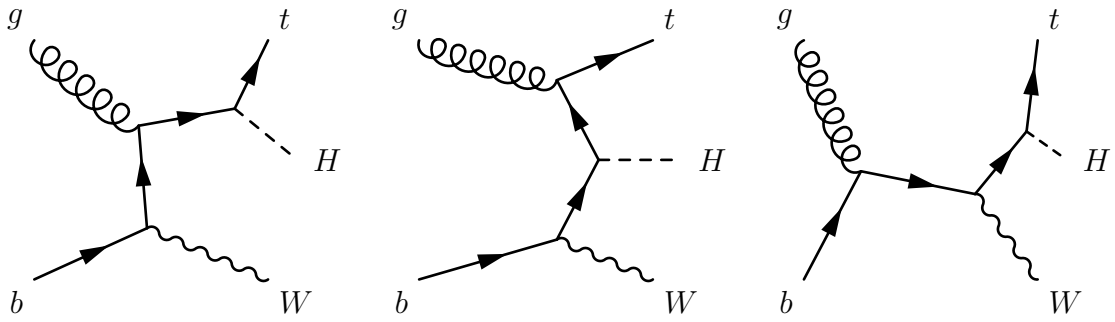


Figure 2.11: Lowest order Feynman diagrams of the tWH process. These diagrams can only occur in a five-flavor PDF scheme. Additional diagrams can be obtained by allowing a Higgs boson to radiate from the W boson, by reversing the fermion lines, or by crossing the lines of initial particles in t-channel diagrams.

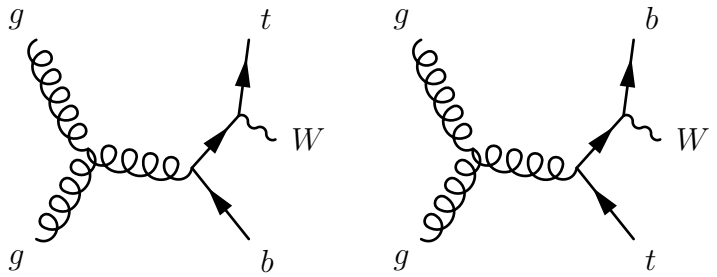


Figure 2.12: NLO Feynman diagrams of tW production (left) and $t\bar{t}$ production (right) that produce interference. These are extended to the tWH and $t\bar{t}H$ case by adding Higgs-strahlung from the t or W lines.

2.3.2 Decay channels

Higgs decays to massive gauge bosons and fermions occur at tree level, as shown in Figure 2.13. Higgs decays to massless particles, including a pair of gluons or photons, occur through loop diagrams and are suppressed accordingly.

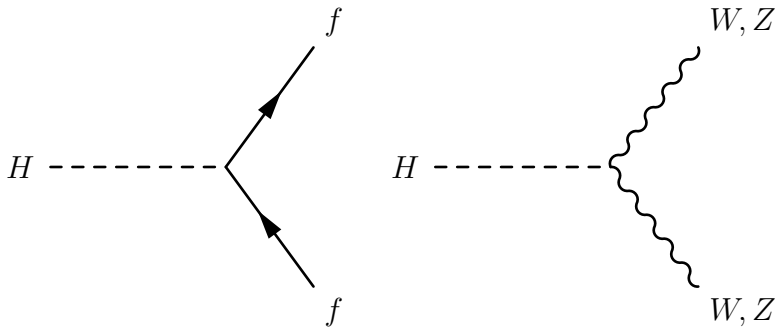


Figure 2.13: Tree-level diagrams Higgs boson decays to fermions (left) and vector bosons (right).

The total width of the SM Higgs boson Γ_H is shown in Figure 2.14 as a function of the Higgs boson mass m_H. Figure 2.14 also shows the branching fractions B of the SM Higgs boson as a function of the Higgs boson mass, where

$$B(H \to X) = \frac{\Gamma(H \to X)}{\Gamma_H} \tag{2.23}$$

The best measurement of the Higgs mass to date comes from the ATLAS/CMS combined Run 1 measurement [37]:

$$m_H = 125.09 \ \pm 0.21 \ \text{(stat)} \ \pm 0.11 \ \text{(syst)} \ \text{GeV} \tag{2.24}$$

Table 2.2 lists the branching fractions of the SM Higgs boson at this value of m_H. Note that the decay of a Higgs boson to top quarks is energetically forbidden.

The number of Higgs boson events produced is larger in decay channels with higher branching fractions. However, when seeking to maximize experimental sensitivity to the Higgs boson, there are additional factors that must be considered. These include the rate of background processes, and the detector's ability to identify and reconstruct the Higgs boson decay products. Despite the low branching fraction, the $H \to \gamma\gamma$ decay channel is among the most powerful for Higgs boson measurements due to the excellent diphoton mass resolution of the ATLAS detector. The bulk of this work will focus on the $H \to \gamma\gamma$ decay, and its advantages are further discussed in later Chapters.

The lowest order diagrams contributing to the $H \to \gamma\gamma$ decay are shown in Figure 2.15. In the Standard Model, this loop is dominated by the Higgs coupling to the W boson, and the decay through a quark loop provides some small sensitivity to the top Yukawa coupling. Destructive interference between the Higgs-top and Higgs-W diagrams suppresses the rate. In BSM models, massive particles that are electrically charged can also contribute to the rate of the $H \to \gamma\gamma$ decay.

Decay mode	Branching fraction [%]
$H \to bb$	57.5 ± 1.9
$H \to WW$	21.6 ± 0.9
$H \to gg$	8.56 ± 0.86
$H \to \tau\tau$	6.30 ± 0.36
$H \to cc$	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 2.2: Branching fractions of the Standard Model Higgs boson [34].

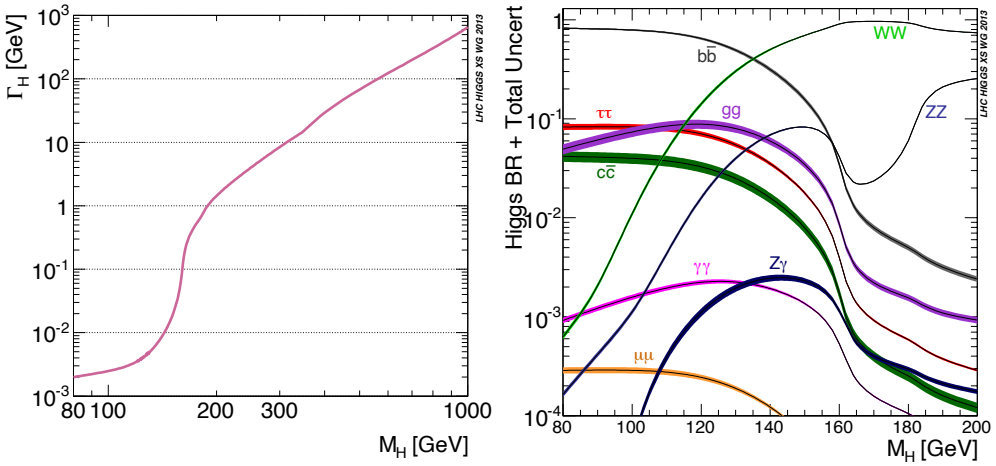


Figure 2.14: Total width (left) and branching fractions (right) of the Standard Model Higgs boson as a function of Higgs boson mass [34].

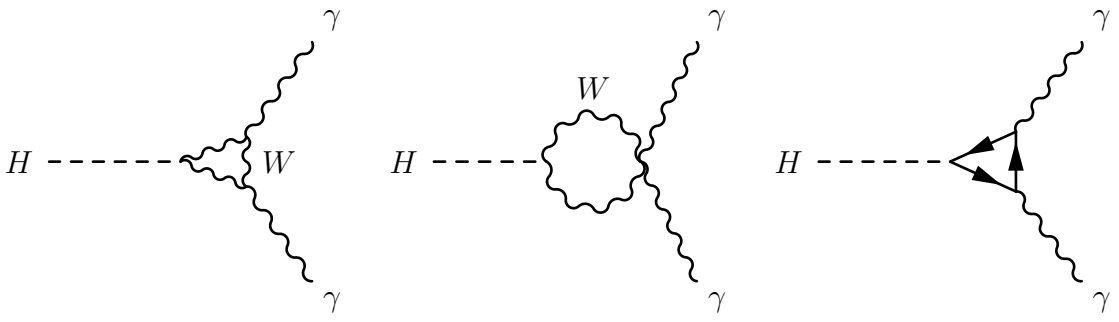


Figure 2.15: Lowest order Feynman diagrams contributing to the $H \to \gamma\gamma$ decay. The fermion loop in the right-most diagram is dominated by the contribution from the top quark.

2.3.3 Coupling measurement framework

When measuring Higgs boson couplings, it is convenient to refer to the ratio to the Standard Model prediction rather than the absolute coupling:

$$\kappa_i = \frac{\lambda_i^{\text{meas}}}{\lambda_i^{SM}} \tag{2.25}$$

for particle type i. This κ parameter is called the **coupling strength**, and measurements of κ that differ from one indicate deviations from the Standard Model prediction. This convention is known as the κ framework [38].

Table 2.3 shows the SM Higgs boson production cross sections calculated in terms of the coupling strengths κ at $\sqrt{s} = 13$ TeV (see [6] for comparable parameterizations at lower center of mass energies). The cross section of $t\bar{t}H$ production is dependent only on the top coupling strength κ_t, and therefore provides a direct probe of the top Yukawa coupling. The cross sections of $tHjb$ and tWH production depend on both the Higgs-top (κ_t) and Higgs-W (κ_W) couplings. The cross term ($\sim \kappa_t\kappa_W$) represents the interference, which provides sensitivity to the relative sign of κ_t and κ_W. Indirect constraints on κ_t also come from top quark contributions to the gluon fusion loop.

Table 2.2 shows the dependence of each partial decay width on the coupling strengths κ at $m_H = 125.09$ GeV. Because the decay of a Higgs boson to top quarks is energetically forbidden, none of the SM Higgs boson decay modes provides a direct probe of κ_t. However, Higgs boson decays through loop processes, including $H \to \gamma\gamma$, have some indirect sensitivity to κ_t.

The resolved coupling modifiers in Tables 2.3-2.2 are calculated assuming the Standard Model. However, the presence of new particles in loop processes would introduce additional terms to these modifiers. In order to account for potential new physics in loop processes, **effective coupling strengths** are often preferred.

For loop-induced production processes such as gluon fusion, the effective coupling strength is defined by taking the ratio of the measured to Standard Model cross section:

$$\kappa_g^2 = \frac{\sigma_{ggF}^{\text{meas}}}{\sigma_{ggF}^{SM}} \tag{2.26}$$

The effective coupling strengths for loop decays are defined in an analogous way from the decay width:

$$\kappa_\gamma^2 = \frac{\Gamma_{\gamma\gamma}^{\text{meas}}}{\Gamma_{\gamma\gamma}^{SM}} \tag{2.27}$$

Because these effective couplings are not explicitly dependent on each particle contributing to the loop, the effective parameterization is more model independent than the resolved parameterization. Depending on the assumptions made in a given statistical analysis, resolved and effective loop couplings can both be useful. See Chapter 12 for an example of an analysis that uses both.

Production mode	Effective modifier	Resolved modifier
$\sigma(ggF)$	κ_g^2	$1.04\kappa_t^2 + 0.002\kappa_b^2 - 0.04\kappa_t\kappa_b$
$\sigma(VBF)$	-	$0.73\kappa_W^2 + 0.27\kappa_Z^2$
$\sigma(qq/qg \to WH)$	-	κ_W^2
$\sigma(qq/qg \to ZH)$	-	κ_Z^2
$\sigma(gg \to ZH)$	-	$2.46\kappa_Z^2 + 0.46\kappa_t^2 - 1.90\kappa_Z\kappa_t$
$\sigma(t\bar{t}H)$	-	κ_t^2
$\sigma(b\bar{b}H)$	-	κ_b^2
$\sigma(tHjb)$	-	$2.63\kappa_t^2 + 3.58\kappa_W^2 - 5.21\kappa_t\kappa_W$
$\sigma(tWH)$	-	$2.91\kappa_t^2 + 2.31\kappa_W^2 - 4.22\kappa_t\kappa_W$

Table 2.3: Production modes of the Standard Model Higgs boson in 13 TeV proton–proton collisions and their parameterization in the κ framework [34].

Decay mode	Effective modifier	Resolved modifier
$H \to bb$	-	κ_b^2
$H \to WW$	-	κ_W^2
$H \to gg$	κ_g^2	$1.11\kappa_t^2 + 0.01\kappa_b^2 - 0.12\kappa_b\kappa_t$
$H \to \tau\tau$	-	κ_τ^2
$H \to cc$	-	κ_c^2
$H \to ZZ$	-	κ_Z^2
$H \to \gamma\gamma$	κ_γ^2	$1.59\kappa_W^2 + 0.07\kappa_t^2 - 0.67\kappa_W\kappa_t$
$H \to Z\gamma$	$\kappa_{Z\gamma}^2$	$1.12\kappa_W^2 - 0.12\kappa_W\kappa_t$
$H \to \mu\mu$	-	κ_μ^2

Table 2.4: Decay modes of the Standard Model Higgs boson and their parameterization in the κ framework [34].

The κ framework is widely used for measurements of Higgs boson couplings. Figure 2.16 shows the measured value of each κ parameter as a function of particle mass. Note that the vertical axis of this Figure differs for bosons and fermions. No significant deviation from the Standard Model expectation (represented as a dashed blue line) has yet been observed.

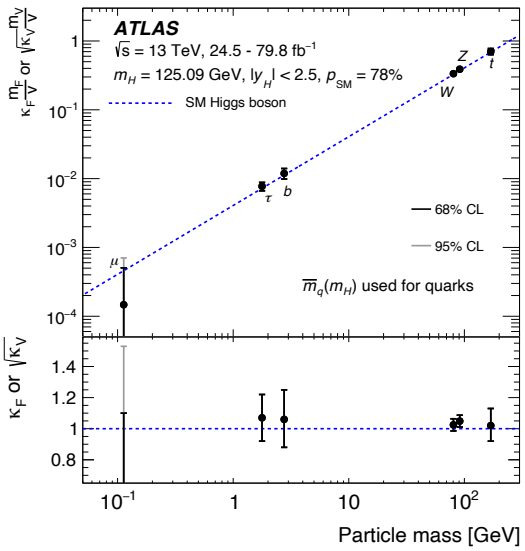


Figure 2.16: Predicted and measured values of Higgs boson couplings to Standard Model particles and their relationship to particle mass [31]. Note that the vertical axis differs for bosons and fermions.

2.4 Beyond the Standard Model Higgs

The simple theory of electroweak symmetry breaking outlined in Section 2.2 can be extended to include additional Higgs bosons and new interactions between the Higgs bosons and SM particles. Such theories include two-Higgs-doublet models (2HDM) [39] [40] and models containing Higgs triplets [41].

These scenarios are being probed by dedicated searches for additional Higgs bosons [42] [43] [44] and global fits of effective field theory (EFT) parameters to collider data [45]. So far, there is no evidence indicating that the Higgs sector is more complex than the single SU(2) doublet model.

2.4.1 CP violation in the Higgs sector

As Higgs boson couplings are measured with greater precision, it becomes possible to explore the charge-parity (CP) symmetry of Higgs interactions. The Standard Model predicts that all Higgs boson interactions are CP-conserving (CP even): that is, if particles and antiparticles are interchanged and spatial coordinates are inverted, all Higgs boson interactions are unchanged. The measurement of even a small CP-violating component in any Higgs boson coupling would constitute a discovery of physics beyond the Standard Model and would potentially address one of the key open questions in particle physics today: the understanding of the matter-antimatter imbalance in the Universe.

Measurements of Higgs couplings to vector bosons have constrained any CP violating component to be small [46] [47] [48]. However, CP-violating contributions to these bosonic couplings can only enter via higher order operators that are suppressed by powers of $1/\Lambda^2$, where Λ is the energy scale of new physics in the EFT.

The CP properties of Higgs-fermion interactions remain largely unexplored, and CP violating terms can enter at the same order as the CP even terms of the SM. These measurements are therefore expected to give higher sensitivity than measurements of Higgs couplings to vector bosons.

As the largest Yukawa coupling, the Higgs-top coupling is a natural choice for testing CP properties. Measurements of electric dipole moments have provided indirect and constraints on the CP properties of the Higgs-top interaction [49], though these rely on assumptions of SM Yukawa couplings for light fermions such as the electron. In addition, hadron collider measurements of ggF production and $H \to \gamma\gamma$ decay at $\sqrt{s} = 8$ TeV have provided some indirect constraints [50], but again the results rely on an assumption of SM contributions to these loops. With LHC data at $\sqrt{s} = 13$ TeV, the CP properties of the Higgs-top interaction can be probed directly using the $t\bar{t}H$ and tH processes.

The Higgs Characterization (HC) model [51] provides a framework for describing different couplings, including scalar, pseudo-scalar, or some combination thereof. Within this model, the effective Lagrangian that describes the top Yukawa coupling can be expressed as:

$$\mathcal{L} = -\frac{m_t}{v}(\kappa_t \cos(\alpha)\bar{t}t + i\kappa_t \sin(\alpha)\bar{t}\gamma_5 t)H \tag{2.28}$$

where α is the CP mixing angle. The Standard Model coupling corresponds to $\alpha = 0°$ and $\kappa_t = 1$, while a fully CP odd coupling is realized when $\alpha = 90°$. In this parameterization, κ_t is restricted to be ≥ 0 and α is in the range $(-180°, 180°)$.

The dependence of the $t\bar{t}H$ and tH cross sections on the model parameters κ_t and α can be calculated from the Lagrangian in Equation 2.28. The $t\bar{t}H$ cross section takes the form

$$\sigma_{t\bar{t}H} = A\kappa_t^2 c_\alpha^2 + B\kappa_t^2 s_\alpha^2 + E\kappa_t^2 s_\alpha c_\alpha \tag{2.29}$$

where s_α and c_α represent the sine and cosine of α, respectively. The coefficient A corresponds to the CP even contribution to the cross section, and B corresponds to the CP odd contribution. The interference between CP even and CP odd is represented by coefficient E, which is expected to be small.

For $tHjb$ and tWH, the interference between the top quark and W boson couplings must also be taken into account. The cross section of both $tHjb$ and tWH take the same form (though with different values of the coefficients):

$$\sigma_{tH} = A\kappa_t^2 c_\alpha^2 + B\kappa_t^2 s_\alpha^2 + C\kappa_t c_\alpha + D\kappa_t s_\alpha + E\kappa_t^2 s_\alpha c_\alpha + F \tag{2.30}$$

The coefficients A, B, and E have the same meanings as in the $t\bar{t}H$ case. The Higgs coupling to the W boson is assumed to obey the Standard Model ($\kappa_W = 1$). The coefficient C corresponds to the interference between the CP even top-Higgs and W-Higgs couplings,

D to the interference between the CP odd top-Higgs and W-Higgs couplings, and F to the contribution of the pure W-Higgs coupling.

Using the HC model and its implementation in the MG5_aMC@NLO Monte Carlo generator [52] [53], the cross sections for $t\bar{t}H$, $tHjb$, and tWH are calculated at multiple values of κ_t and α. The NLO inclusive cross sections calculated by MG5_aMC@NLO are shown in Table 2.5. These cross sections (normalized to the Standard Model prediction) are shown in Figure 2.17 as a function of α with κ_t fixed to unity. The parameterizations of the loop processes ggH (κ_g) and $H \to \gamma\gamma$ (κ_γ) in terms of α are taken from [50].

Parameters		$\sigma(\sqrt{s} = 13$ TeV) [pb]		
κ_t	α	$t\bar{t}H$	$tHjb$	tWH
1	0° (SM)	0.458	0.0606	0.0167
1	15°	0.443	0.0636	0.0184
1	30°	0.396	0.0743	0.0231
1	45°	0.329	0.0958	0.0308
1	60°	0.265	0.1374	0.0422
1	75°	0.217	0.1970	0.0563
1	90°	0.199	0.2707	0.0726
-1	0°	0.458	0.6971	0.1490
0.5	0°	0.1145	0.0939	0.0148
2	0°	1.832	0.2406	0.0924
2	45°	–	0.2055	0.1095

Table 2.5: NLO inclusive cross sections from MG5_aMC@NLO for the $t\bar{t}H$, $tHjb$, and tWH processes in different CP scenarios. For $\kappa_t = 2$ and $\alpha = 45°$, no $t\bar{t}H$ cross section is calculated.

The cross section of $t\bar{t}H$ decreases as the CP odd component becomes larger and is minimized at $\alpha = 90°$. However, the Standard Model point represents a global minimum of the cross section for both tH processes: the rate can only increase with changes in α. Note that at $\alpha = 180°$, the sign of the interference term between the top-Higgs and W-Higgs couplings is flipped with respect to the SM. This switch from destructive to constructive interference accounts for the large change in tH rate at $\alpha = 180°$.

In addition to its impact on overall cross sections, the introduction of a CP odd component to the Higgs-top coupling also alters the shapes of sensitive kinematic observables in $t\bar{t}H$ and tH. Such observables are discussed at length in Chapter 11, which presents a multivariate analysis strategy that exploits the shape dependence of kinematic distributions on the CP mixing angle α.

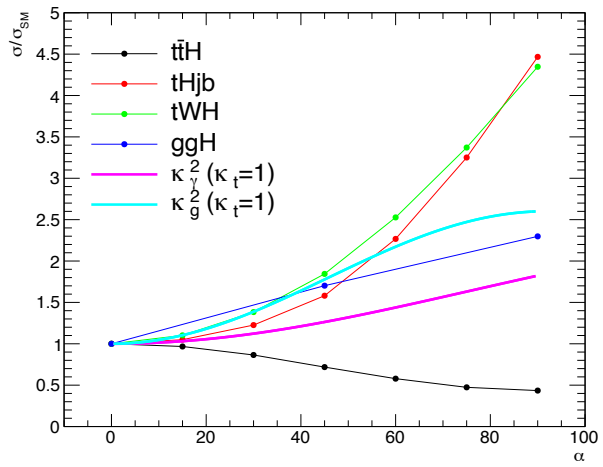


Figure 2.17: NLO $t\bar{t}H$ and tH cross sections as a function of the CP mixing angle α (for $\kappa_t = 1$), normalized to the Standard Model prediction. The dependence of κ_g and κ_γ on α is taken from [50].

Chapter 3

The ATLAS experiment

Experimental high energy physics in the twenty-first century requires collaboration on an unprecedented scale. The design and construction of the world's highest-energy collider, a complex detector the size of the White House, and novel technologies that withstand immense radiation are ambitious projects that necessitate the formation of large international collaborations.

The ATLAS collaboration is made up of over 3000 scientists from 38 nations, all of whom have contributed to the experimental setup described in this Chapter. The first Section describes the accelerator itself, the Large Hadron Collider (LHC). The second describes the layout and function of the general-purpose ATLAS detector and its subsystems.

3.1 The Large Hadron Collider

The **Large Hadron Collider (LHC)** [54] [55] is a high energy particle accelerator located at the *Organisation Européenne pour la Recherche Nucléaire* (CERN) outside of Geneva, Switzerland. The LHC occupies a circular tunnel with a circumference of 27 km, spanning the border between Switzerland and neighboring France. The LHC began collecting proton–proton collision data in the fall of 2009. During the first run of the LHC (2009-2012), protons were collided with center of mass energy $\sqrt{s} = 7$ and 8 TeV. During the second run of the LHC (2015-2018), the center of mass energy was increased to $\sqrt{s} = 13$ TeV, the highest energy ever achieved by a collider.

LHC protons originate from a bottle of hydrogen gas. Once the electrons have been stripped away using an electric field, Linear Accelerator 2 (LINAC2) [56] accelerates the protons using radiofrequency (RF) cavities, which use oscillating electromagnetic fields to accelerate protons to an energy of 50 MeV. The timing of a proton's arrival at a given point in the RF cavity determines how much it will be accelerated (or decelerated): this results in a regular pattern of distinct **proton bunches** [57].

From LINAC2, the protons are injected into three successive synchrotron systems: the Synchrotron Booster (BOOSTER) [58], which accelerates the protons to 1.4 GeV; the Proton

Synchrotron (PS) [59], which accelerates the protons to 25 GeV; and the Super Proton Synchrotron (SPS) [60], which accelerates the protons to 450 GeV. Finally, the protons are injected into the LHC. The layout of the CERN collider complex, including each of these injectors, is shown in Figure 3.1.

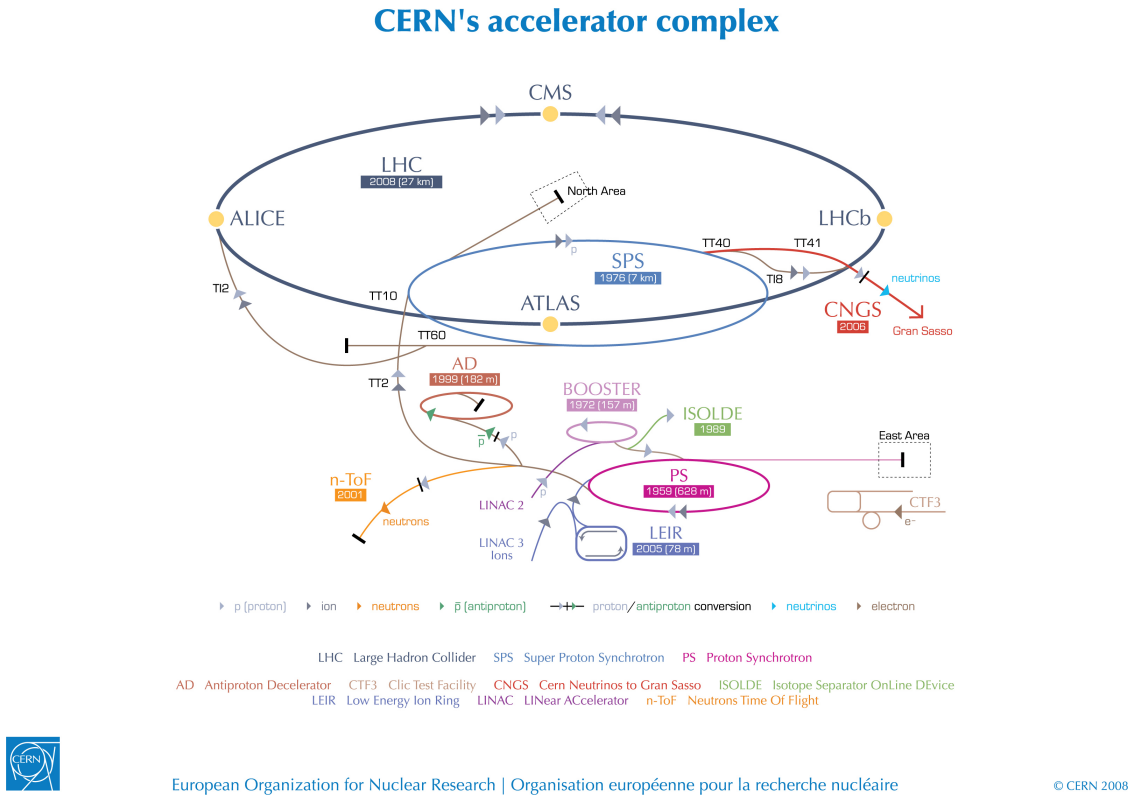


Figure 3.1: Diagram of the CERN collider complex [61].

The LHC itself is composed of eight straight segments, each about half a kilometer long, where RF cavities accelerate protons to 6.5 TeV in bunches separated by 25 ns (50 ns in Run 1). The straight segments are connected by eight rounded arcs, which each contain three superconducting dipole magnets (for bending the particle trajectories) and one superconducting quadrupole magnet (for focusing) [54]. Two beams of protons circulate through the LHC in opposite directions, passing through twin bores in the dipole magnets. The beams cross at four points: the locations of the ATLAS [62], CMS [63], LHCb [64], and ALICE [65] experiments.

Luminosity

For a given process P, the expected rate of interactions at the LHC is given by

$$\frac{dN_P}{dt} = \sigma_P \times \mathcal{L} \tag{3.1}$$

where σ_P is the cross section for process P, and $\mathcal{L}$ is the **instantaneous luminosity** delivered by the machine. For two beams of identical Gaussian profile colliding head-on, the instantaneous luminosity is

$$\mathcal{L} = \frac{n_p^2 n_b f}{4\pi\sigma_x\sigma_y} \tag{3.2}$$

where n_p is the number of protons per bunch, n_b is the number of bunches, f is the revolution frequency, and σ_x and σ_y are the beam resolutions in the plane transverse to the beam. Additional corrections are required to account for beams crossing at slight angles, correlations between σ_x and σ_y, and other effects [66]. The ATLAS detector was designed for a maximum instantaneous luminosity of 10^{34} cm^{-2}s^{-1}.

The total number of interactions of process P produced at the LHC during time T is

$$N_P = \sigma_P \times \int_0^T \mathcal{L}dt' = \sigma_P \times \mathcal{L}_{\text{int}} \tag{3.3}$$

The **integrated luminosity** $\mathcal{L}_{\text{int}}$ is typically measured in **inverse femtobarns (fb^{-1})**. The integrated luminosity delivered to ATLAS during LHC Run 2 was measured by the LUCID-2 detector, which is located around the LHC beam pipe on either side of the ATLAS detector [67].

3.2 The ATLAS detector

The ATLAS detector is a general-purpose particle detector located 100 m underground on the LHC ring. The cylindrical detector geometry uses the LHC beam pipe as its axis, and the forward-backward symmetry gives nearly 4π coverage in solid angle. Figure 3.2 shows a diagram of the ATLAS detector, which extends 44 m long and 25 m high. It is estimated that the detector (including both sensitive area and structural components) weighs 7000 tons.

The detector is described using a right-handed coordinate system with origin corresponding to the proton–proton collision point at the center of the detector. The direction of the incoming proton beams defines the z-axis, the x-axis points towards the center of the LHC ring, and the y-axis points upwards.

Due to the azimuthal symmetry of the detector, cylindrical coordinates r and ϕ (azimuthal angle around the z-axis) provide a convenient description. Each particle's momentum is projected onto the transverse plane to obtain the **transverse momentum**:

$$p_T \equiv \sqrt{p_x^2 + p_y^2} \tag{3.4}$$

The **pseudorapidity** is defined in terms of the polar angle θ:

$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{3.5}$$

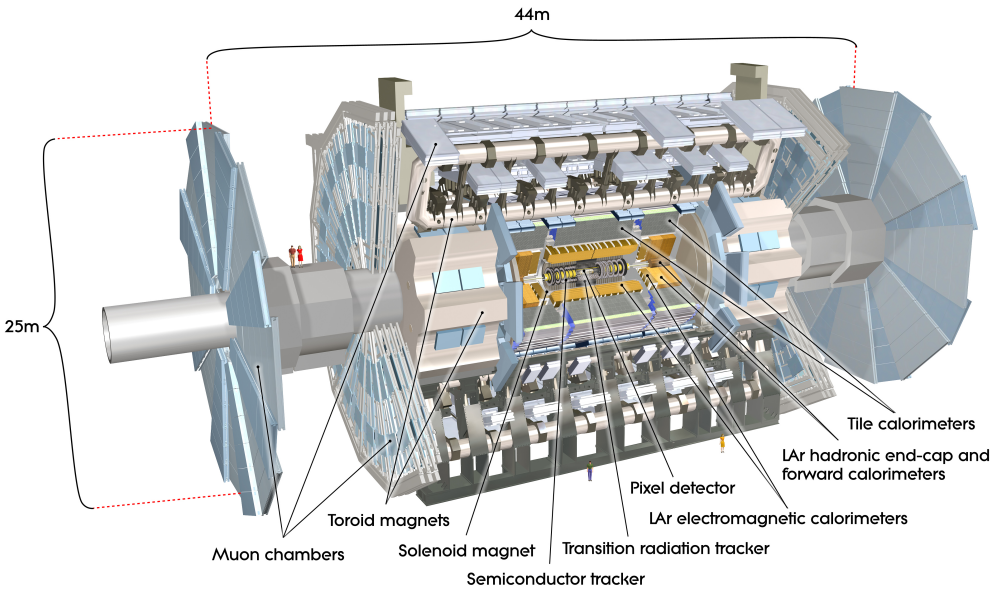


Figure 3.2: Diagram of the ATLAS detector [68].

For a particle with mass much smaller than its momentum ($m \ll |\vec{p}|$), the pseudorapidity approaches the rapidity y. In this limit, pseudorapidity differences are invariant under Lorentz boosts in the $\hat{z}$ direction. Angular distance is measured in terms of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

The following sections detail the design and purpose of each detector system, from inner- to outermost: the Inner Detector, the Calorimeters, and the Muon Spectrometer. Each system is composed of a **barrel** region, with layers arranged on concentric cylinders around the z-axis, and a disk-shaped **endcap**, with layers oriented perpendicular to the beam axis at high $|z|$.

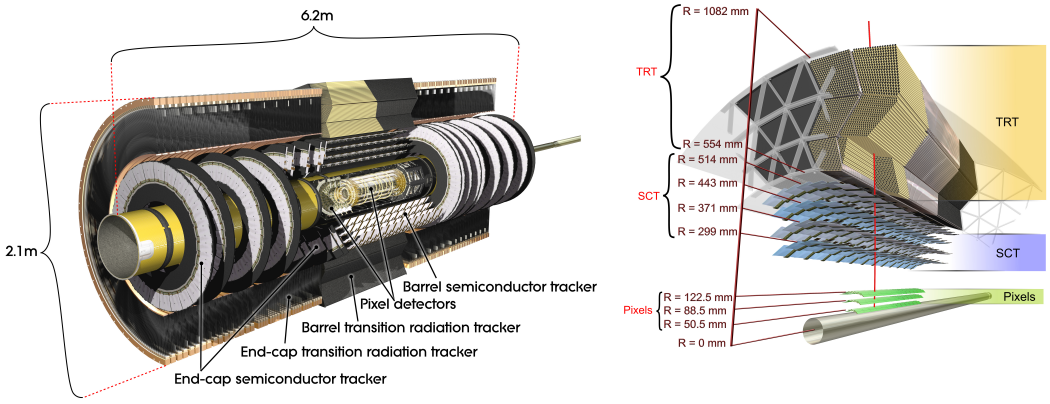
3.2.1 Inner Detector

The ATLAS Inner Detector (ID) [69] is composed of the subsystems closest to the collision point, which provide precise measurements of particle position and momentum needed for tracking. The subsystems include (from inner- to outermost) the Insertable B-Layer (IBL), the Pixel Detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT).

A superconducting solenoid magnet encloses the full ID and immerses all subsystems in a 2 T magnetic field oriented in the $\hat{z}$ direction. The resulting curvature of charged particle trajectories in the transverse plane provides measurements of particle momenta.

Figure 3.3a shows the full layout of the inner detector (including end-caps), and Figure 3.3b shows the radial position of each barrel layer. The IBL is missing from this figure, but

is positioned at $r = 33.25$ mm. The extent of each ID subsystem is summarized in Table 3.1: all subsystems have coverage in the range $|\eta| < 2.5$.



(a) Inner Detector (Full) (b) Inner Detector (Barrel)

Figure 3.3: Diagram of the ATLAS Inner Detector layout [69]. (Left) the full Inner Detector, including end-caps. (Right) cross section of the barrel region of the ID. The Pixel Detector is shown in green, the SCT in blue, and the TRT in orange. The IBL is missing from this figure, but is positioned at $r = 33.25$ mm.

System	Barrel			End-cap		
	Layers	r [mm]	$\|z\|$ [mm]	Disk layers	r [mm]	$\|z\|$ [mm]
IBL	1	33.25	0-330.15	–	–	–
Pixel	3	50.5-122.5	0-400.5	2×5	88.8-149.6	495-650
SCT	4	299-514	0-749	2×9	275-560	810-2797
TRT	73	563-1066	0-712	160	644-1004	848-2710

Table 3.1: Extent of the sensitive region of each ATLAS Inner Detector subsystem [62].

Pixel Detector & Insertable B-Layer

The Pixel detector and Insertable *B*-Layer (IBL) are highly granular detectors designed to provide high-precision measurements of particle trajectories very close to the interaction point. The basic element of these detectors is a silicon sensor with pixelated readout, or **pixel**, which makes high-resolution measurements of ionization energy deposited by charged particles traversing the sensor.

At a radius of only $r = 33.25$ mm, the IBL [70] [71] is the closest ATLAS subsystem to the collision point. The IBL was inserted into ATLAS during LHC Long Shutdown 1 (between LHC Runs 1 and 2), and it now functions as the innermost layer of the Pixel

Detector. The IBL derives its names from the critical role it plays in b-tagging, which relies heavily on identification secondary vertices from the decay of B-hadrons (see Section 4.5.1). Due to its proximity to the interaction point, the IBL is among the ATLAS subsystems most susceptible to radiation damage. See Appendix C for studies of radiation effects in the IBL.

The Pixel Detector [72] is composed of three cylindrical barrel layers (not including the IBL) and ten disk layers (five on each side) of silicon pixels. The barrel layers are named the B-layer, Layer 1, and Layer 2, with the B-layer being the innermost (like the IBL, named for its important role in b-tagging). Each barrel layer is tiled with pixels for two-dimensional segmentation in the $\hat{\phi} \times \hat{z}$ plane. Each pixel module is composed of 61,440 individual pixels and 16 readout chips. The resolution of a pixel is 10 μm in $r\phi$ and 115 μm in z in the barrel layers, and 10 μm in $r\phi$ and 115 μm in r in the endcap disks.

Semiconductor Tracker

Like the IBL and Pixel detector, the Semiconductor Tracker (SCT) [73] relies on silicon semiconductor technology. Each silicon microstrip detector is composed of 768 readout strips, each with a width of 80 μm. In the barrel region, SCT modules are formed from four single-sided silicon microstrip detectors: two are wire-bonded end to end, forming strips that are 12 cm long. Two such pairs are then positioned back-to-back, with one side oriented parallel to the beam direction and the other offset from the z axis by an angle of 40 mrad. This small angle offset allows for measurement of the z coordinate. The SCT barrel is composed of four layers of these modules.

In the endcap, SCT modules are similarly constructed from strip detectors positioned back-to-back with an angular offset of 40 mrad. However, the strips in the endcap are oriented radially and taper to smaller pitch at the origin. Each endcap consists of nine disk layers.

The SCT has $r\phi$ resolution of 17 μm in the barrel and endcap regions. The z coordinate is calculated from the $r\phi$ and stereo measurements (details are available in [74]). The resolution in z is 580 μm in the barrel, and the resolution on r is 580 μm in the endcap.

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) [75] is composed of straw drift tubes filled with xenon-based gas and a central cathode wire. In the barrel region, the TRT straws are positioned parallel to the beam axis and extend to $|\eta| < 0.7$. In the end-cap wheels, TRT straws are positioned radially from the z axis. Each straw is 4 mm in diameter and provides 130 μm resolution in $r\phi$ (but none in z).

The TRT straws allow for continuous tracking, typically providing about 36 space points per charged particle track. The large number of space points compensates for the low resolution compared to the silicon detectors, and TRT inputs make a significant contribution to measurements of charged particle curvature (and hence momentum). Precise measurements in this outer region of the ID are also useful for identifying photon conversions.

The TRT derives its name from its ability to measure transition-radiation photons. The amount of transition radiation measured in the TRT provides discrimination between electrons and charged hadrons (pions).

3.2.2 Calorimeters

The ATLAS calorimeters [76] [77] [78] [79] serve to measure both electromagnetic showers and hadronic jets. The electromagnetic showers spawned by photons and electrons are typically contained in the innermost calorimeter, the liquid argon electromagnetic calorimeter (ECAL). Because hadrons travel through more material than electrons or photons, the hadronic calorimeters make up a larger portion of ATLAS' volume than the ECAL: these are the Tile Calorimeter, Hadronic Endcap Calorimeters (HECs), and Forward Calorimeters (FCAL).

Different materials are used in each calorimeter in order to remain robust against radiation damage in all η regions. The layout of the ATLAS calorimeter systems is shown in Figure 3.4. The number of layers and η coverage of each subsystem are presented Table 3.2.

System	Barrel		End-cap	
	Layers	$\lvert\eta\rvert$ coverage	Layers	$\lvert\eta\rvert$ coverage
Presampler	1	< 1.52	1	$1.5 - 1.8$
ECAL	3 (2)	< 1.35 $(1.35 - 1.475)$	2×2 (3)	$1.375 - 3.2$ $(1.5 - 2.5)$
Tile	3	< 1.0	–	–
Tile Extended	2×3	$0.8 - 1.7$	–	–
HEC	–	–	2×4	$1.5 - 3.2$
FCAL	–	–	2×3	$3.1 - 4.9$

Table 3.2: Extent of each ATLAS calorimeter system [62].

Electromagnetic Calorimeter

The liquid argon Electromagnetic Calorimeter (ECAL) [76] [77] is a sampling calorimeter composed of alternating layers of lead (absorber) and liquid argon (active material). The lead and liquid argon are layered in an accordion structure in order to provide full coverage in ϕ. A pre-sampler detector is located just upstream of the ECAL in order to provide corrections for losses due to material in the ID.

In the barrel region, the ECAL extends from 1.5 m $< r <$ 1.97 m, which corresponds to 24 radiation lengths (X_0) at $\eta = 0$. The ECAL barrel is divided into three layers at different radii. In the central region, the layers are segmented into projective $\eta \times \phi$ regions of sizes 0.003×0.1, 0.025×0.025, and 0.5×0.025 in the first, second, and third ECAL layer respectively. The segmentation of the ECAL barrel becomes coarser for $|\eta| > 1.4$. Due to its fine granularity in η, the first ECAL layer is also referred to as the ECAL strips layer. It provides critical input for the determination of photon vertices and the identification of π^0

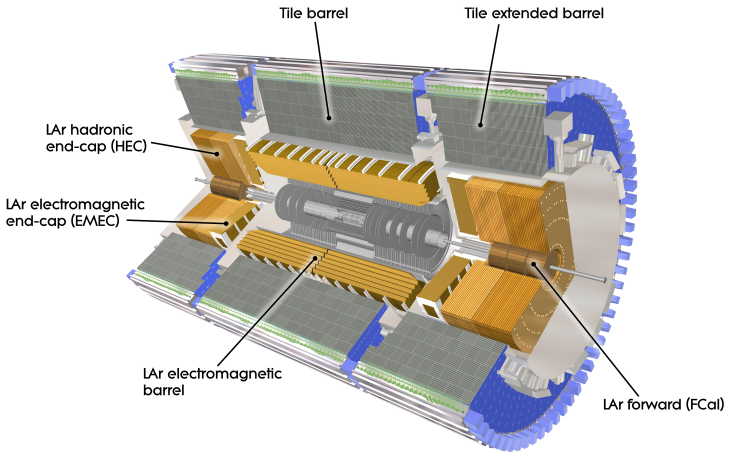


Figure 3.4: Diagram of the layout of the ATLAS calorimeter systems [69].

hadrons (see Section 4.2). Figure 3.5 shows a diagram of the ECAL layout in the barrel and the segmentation of each layer.

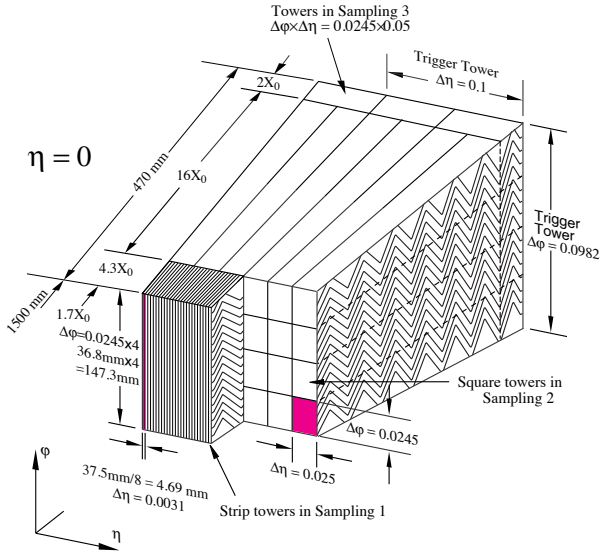


Figure 3.5: Diagram of the ATLAS electromagnetic calorimeter [69].

Each ECAL endcap is composed of two concentric wheels centered at and perpendicular to the z-axis. The total endcap extends radially from 302 mm $< r <$ 2077 mm and longitudinally from 3.7 m $< |z| <$ 4.2 m. The height and angle of the accordion folds in the endcap vary with radius in order to ensure uniform coverage in ϕ.

Though the ECAL provides full coverage in the transition between barrel and endcap ($1.37 < |\eta| < 1.52$, or the "crack"), the large density of material upstream from the ECAL in this region leads to degraded performance. Photon and electron candidates in this region are vetoed.

Tile Calorimeter

The hadronic Tile Calorimeter [78] [79] is a collection of sampling calorimeters using alternating layers of 3 mm-thick plastic scintillator tiles (active material) and 14 mm-thick iron plates (absorber). The calorimeters are the Tile Barrel ($|z| < 2.82$ m), which is divided into two halves, and the Tile Extended Barrels (3.2 m $< |z| <$ 6.11 m). Each half of the Tile Barrel and each Tile Extended Barrel is composed of 64 modules positioned symmetrically in ϕ. Each module is composed of readout cells with the divisions in η giving a projective geometry centered at the detector origin (see Figure 3.6). The readout cells are connected by wavelength-shifting fibers to photomultiplier tubes (PMTs) located at outer radial edge of the module, where the scintillation light is collected.

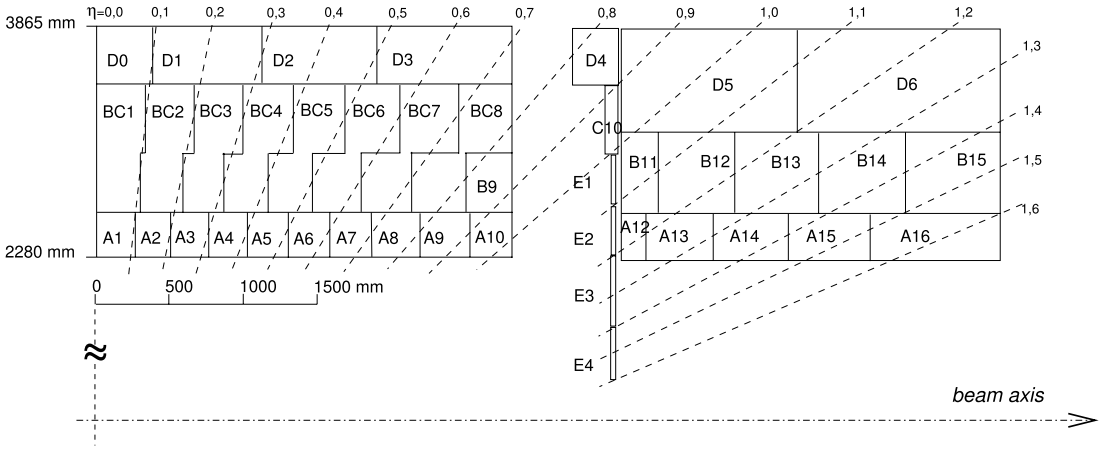


Figure 3.6: Diagram of the cell geometry of a Tile Barrel module (left) and a Tile Extended Barrel module (right) [79]. Dashed lines show constant η, and the projective geometry in η is evident.

For good resolution of high energy jets, the Tile Calorimeter must capture the full radial extent of each hadronic shower to avoid overflow, or punch through, of jet activity into the muon system. The Tile Calorimeter extends from $r = 2.28$ m to $r = 4.25$ m, which corresponds to 9.7 nuclear interaction lengths (λ) at $\eta = 0$.

Calibration of the Tile Calorimeter is performed using laser-based methods, using ^{137}Cs source currents, or by injecting a pulse of known magnitude into the readout electronics. Details on these calibration techniques are available in [79].

Hadronic Endcap Calorimeters

The Hadronic Endcap Calorimeters (HECs) [76] [77] are composed of copper plates (absorber) positioned perpendicular to the beam axis. Each HEC extends from 4.26 m $< |z| <$ 6.12 m and radially from 475 mm $< r <$ 2.03 m. The gaps between copper plates are filled with liquid argon (active material), and three electrodes divide each endcap into four sampling layers at different $|z|$. In the two layers closest to the detector origin, the copper plates are 25 mm thick, compared to 50 mm thick in the outer two layers.

Forward Calorimeters

The Forward Calorimeters (FCAL) [76] [77] extend from 4.7 m $< |z| <$ 5.15 m and radially from 72 mm $< r <$ 455 mm, covering the range $3.1 < |\eta| < 4.9$. In this region, particle fluxes are high, and the FCAL is designed to be robust against radiation damage.

The FCAL consists of three layers on each side of the detector origin. The inner-most layer on each side uses copper as the metal absorber, and the outer two use tungsten. The metal is used to create a matrix of grounded tubes with axes oriented parallel to the beam direction. Each tube contains a concentric cathode rod, and the gaps between each rod and the inner tube edge are filled with liquid argon (active material).

3.2.3 Muon Spectrometer

The largest and outermost subsystem of ATLAS is the Muon Spectrometer (MS) [80] [81], shown in Figure 3.7.

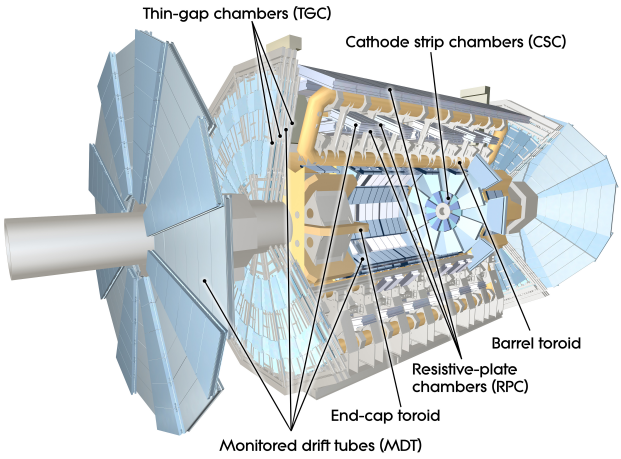


Figure 3.7: Diagram of the ATLAS muon spectrometer [69]. The toroid magnets are shown in orange.

Three superconducting air-core toroid magnets, one barrel toroid and two endcap toroids, deflect charged particles towards the $\pm\hat{z}$ direction. Each toroid consists of 8 coils positioned

at equal distance in ϕ. The magnetic field generated by the barrel toroid is dominant in the region $|\eta| < 1.4$, and the magnetic field generated by the endcap toroid is dominant in the region $1.6 < |\eta| < 2.7$. The region $1.4 < |\eta| < 1.6$, where both barrel and endcap toroids contribute to the magnetic field, is known as the transition region.

Monitored Drift Tubes

In most η regions, Monitored Drift Tubes (MDTs) dominate the precision measurement of muon tracks and momenta. A single MDT consists of a 15 mm-radius aluminum tube and a central cathode wire, with the region in between filled with argon gas.

In the barrel region, MDTs are positioned in three layers ("stations") and divided into sixteen azimuthal regions. The cylindrical barrel stations are centered around the z-axis at $r = 5$ m, 7.5 m, and 10 m. The innermost barrel station is composed of two stacks of four MDT monolayers each, separated by a support structure. The outer two barrel stations are structured similarly, but with only three MDT monolayers on either side of the support. The axis of each drift tube is oriented in the $\hat{\phi}$ direction, providing a projection of particle tracks in the bending plane of the magnetic field ($\hat{r} \times \hat{z}$). The resolution of a single MDT is 80 μm, but higher resolution is achieved by stacking monolayers of MDTs.

In each endcap, the MDTs are arranged in four disk-shaped stations centered at and perpendicular to the beam axis (see Figure 3.8). The innermost station is composed of two stacks of four MDT monolayers, separated by a support structure. The other stations use only three MDT monolayers on either side of the support. Like the barrel, the endcap is divided into sixteen azimuthal regions and the axis of each drift tube is oriented in the $\hat{\phi}$ direction.

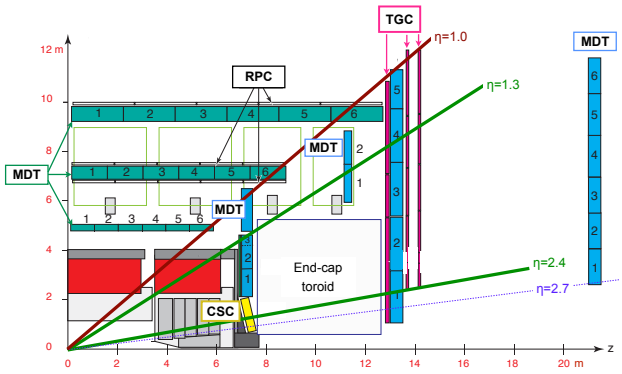


Figure 3.8: Diagram of a cross section of the ATLAS muon spectrometer [82].

Cathode Strip Chambers

While the MDTs perform well in more central regions, their comparatively large radius and the high pressure requirements of argon gas make them ill-suited for regions of very high particle flux. Instead, Cathode Strip Chambers (CSCs) make precision measurements of muon momenta in the very forward region ($2 < |\eta| < 2.7$).

The CSCs are multi-wire proportional chambers, which measure particle trajectories by detecting charge induced on a segmented cathode readout strip by the avalanche formed on the anode wire. The wires are oriented in the $\hat{\phi}$ direction, providing a projection of particle tracks in the bending plane of the magnetic field ($\hat{r} \times \hat{z}$). The resolution of these measurements is about 50 μm.

Each CSC chamber is composed of two layers of four wire planes, separated by a support structure. The chambers are arranged in two pairs of rings at $z = \pm 7$ m, each divided into eight azimuthal segments. Each chamber is angled so that it is perpendicular to a straight particle trajectory starting at the detector origin. The CSCs are shown in yellow in Figure 3.8.

Trigger Chambers

Additional subsystems of the muon spectrometer with lower resolution and faster readout are used for fast muon identification in the trigger system (Section 3.2.4): these include Resistive Plate Chambers (RPCs) in the barrel and Thin Gap Chambers (TGCs) in the endcaps. The location of these chambers is indicated in Figure 3.8. As the studies in this Thesis do not rely on muon-based triggers, the description of these subsystems is left to [80].

3.2.4 Trigger

Due to the 25 ns spacing of LHC proton bunches, collisions in ATLAS occur at a rate of 40 MHz. An average event has a size of $\mathcal{O}(1$ MB) [83], and the resulting data rate is a whopping 1 TB/s. This is far too much data to store and process economically, and the signals from most ATLAS systems require longer than 25 ns to be fully read out. The ATLAS trigger system is responsible for making quick decisions about which collisions to read out and save for analysis, reducing the rate to < 1 kHz.

The trigger system is provided with a list of conditions, known as the **trigger menu**, that specify what combinations of physics objects are interesting enough to be saved. The trigger system first constructs a quick summary of the physics content of each event, then selects those events satisfying one or more of the criteria on the trigger menu. The first step of trigger selection, Level 1, is implemented directly in the detector hardware. The following steps, known as the High Level Trigger, are implemented in software.

The trigger used for analyses in this work requires two photons, with the leading (subleading) photon $p_T > 35$ (25) GeV. The performance of this trigger is reviewed in Chapter 5.

Level 1

The first step in the trigger, Level 1 (L1), is implemented at the hardware level and considers inputs from the muon and calorimeter systems. At L1, only regions of the detector where there is activity, called **Regions of Interest (ROI)**, are considered. This allows for processing of only a fraction of ATLAS' readout channels at L1.

To speed up calculations, L1 triggers do not use the full granularity of the calorimeters. Instead, energy deposits are summed over cells in projective regions called **calorimeter towers**, which have an angular size of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the barrel region. Muon candidates are identified by coincidences in the dedicated Trigger Chambers, which have low resolution but fast readout compared to the MDTs and CSCs. A topological trigger, which calculates geometric or kinematic relationships between L1 trigger objects, was also added for LHC Run 2 [84].

The rate of collisions passing L1 triggers is shown in Figure 3.9 during a 13 TeV run in 2015. The rate is reduced to < 100 kHz after L1 selection. The diphoton trigger used in later Chapters falls under the "Multi EM" group (dark green).

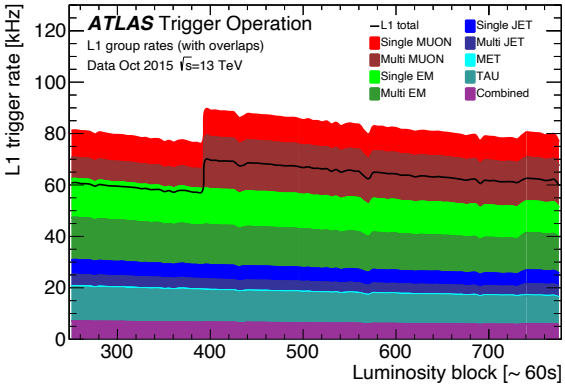


Figure 3.9: Rate of collisions passing the Level 1 Trigger during a 13 TeV run in 2015 [84]. The total (black) is less than the stacked total due to overlap between different triggers.

High Level Trigger

The High Level Trigger (HLT) processes only those events passing L1 triggers. This system is software-based and uses information from all sub-detectors. Tracking information from the ID is added to each event, and information in the MDTs near any L1 muon candidates is also included. Energy deposits in the calorimeters are read out with the full detector granularity at HLT, and these deposits are clustered together using more complex algorithms [84].

The rate of collisions passing HLT triggers is shown in Figure 3.10 during a 13 TeV run in 2015. The rate is reduced to < 1 kHz after HLT selection. The diphoton trigger used in later Chapters falls under the "Photon" group (yellow).

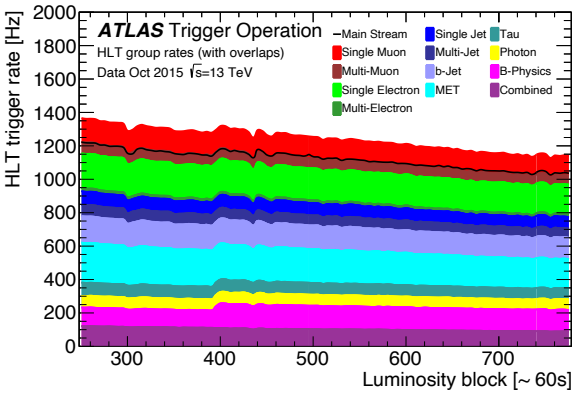


Figure 3.10: Rate of collisions passing the High Level Trigger during a 13 TeV run in 2015 [84]. The total (black) is less than the stacked total due to overlap between different triggers.

Chapter 4

Object Reconstruction

Particles traversing the ATLAS detector are identified and reconstructed based on the combination of signatures they leave in each detector subsystem. Some example particle signatures are shown in Figure 4.1.

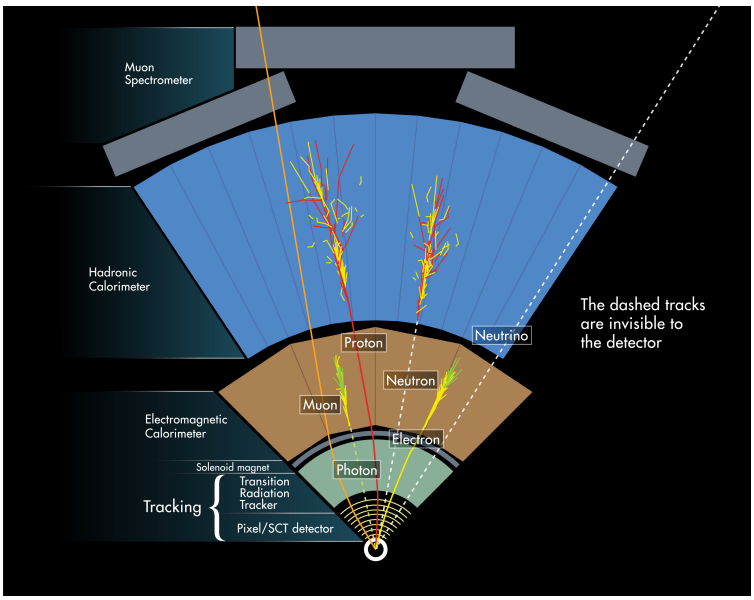


Figure 4.1: Example signatures left in the ATLAS detector by different types of particles, shown in a cross sectional slice of the barrel [85].

The complex final state of $t\bar{t}H(\gamma\gamma)$ events can contain many types of physics objects: the decay of the Higgs boson produces a pair of photons, and the decay of the $t\bar{t}$ system can produce jets (including b-jets) and/or leptons.

The following Sections review the methods for constructing particle tracks and interaction vertices, as well as the reconstruction algorithms and requirements placed on photons, electrons, muons, hadronic jets, and missing transverse energy in ATLAS. Reconstructed

tau leptons are not used in the analyses presented in this Thesis, and the description of tau reconstruction is left to [86].

The reconstruction of each type of physics object is subject to experimental uncertainties due to the detector geometry and resolution, as well as the object reconstruction algorithms. These uncertainties are treated as systematics, and they are fully accounted for in the statistical analyses performed in later Chapters. For a summary of experimental uncertainties, see Table 6.3.

4.1 Tracks and vertices

Charged particle trajectories, or **tracks**, are reconstructed from the pattern of charge deposits ("hits") in the Inner Detector. The high granularity and resolution of the ID systems allows for accurate reconstruction of the position of each hit. The curvature of each track provides a measurement of the particle's charge and momentum with resolution $\Delta p_T/p_T$ that decreases with the p_T of the track [87]. The multi-stage algorithms used to construct tracks from large sets of ID hits are described in [88]. Tracks are required to have $p_T > 500$ MeV and $|\eta| < 2.5$.

Interaction vertices are are determined by an iterative fitting procedure [89]. A fit is performed over the selected tracks and a seed vertex to extract the best fit position of the vertex. Tracks that are incompatible with the calculated vertex are removed, and the fit is repeated. Tracks that are removed can be used to fit additional vertices. Vertex candidates with < 2 compatible tracks are discarded.

Each event can contain multiple vertices, and the number of vertices per event increases with pileup. It is common in ATLAS analyses to select the vertex with the maximum sum of track p_T^2 as the primary interaction vertex. However, $H \to \gamma\gamma$ events can have low charged particle multiplicity in the final state (e.g. gluon fusion), and in these cases it is insufficient to consider tracks alone. For $H \to \gamma\gamma$ measurements, the inputs to primary vertex selection must include not only track information from the ID, but also photon information from the ECAL. The primary vertex selection procedure in diphoton events is described in the following Section.

4.2 Photons

To reconstruct an electromagnetic shower, energy deposits in the calorimeters are collected into topological clusters, or topo-clusters [90]. These topo-clusters are constructed by grouping together calorimeter cells containing energy deposits that exceed the threshold for electronic noise and pileup. Topo-clusters that are associated with a secondary vertex in the ID are designated as converted photon candidates (which have undergone $\gamma \to e^+e^-$ in the ID volume). Topo-clusters that cannot be associated with any tracks or vertices are designated as unconverted photon candidates. Topo-clusters that can be associated with a

single charged particle track are not designated as photon candidates: these are vetoed in order to resolve potential ambiguity between photons and electrons.

In order to separate **prompt** photons, which originate from the primary proton–proton interaction, from background photons, additional identification and isolation criteria are applied. Background photons can be secondary photons produced inside of jets, or hadrons with photon-like signatures (such as π^0 hadrons, which decay to a photon pair). In addition, for the analyses discussed in this Thesis, all photon candidates must have $p_T > 25$ GeV and $|\eta| < 2.37$, with the region $1.37 < |\eta| < 1.52$ (the "crack" containing dead material upstream of the ECAL) vetoed.

High efficiency identification and accurate reconstruction of photons in ATLAS are essential for measurements in the $H \to \gamma\gamma$ decay channel. The photon energy resolution is calculated using multivariate regression trained on simulated samples. The absolute energy scale is calibrated using $Z \to ee$ decays and validated on radiative Z decays [91].

Photon identification

In order to reject background photons, cuts are placed on discriminating variables that are sensitive to differences between prompt and background photons [92]. These variables include:

- Hadronic leakage: energy deposited in the hadronic calorimeter near an ECAL cluster
- Shower shape variables based on the second ECAL layer
- Shower shape variables based on the ECAL strips layer

Because the ECAL strips layer has fine granularity in η, shower shape variables can be constructed that are sensitive to the two-pronged structure characteristic of π^0 hadron decay.

Cuts on the discriminating variables are optimized in bins of $|\eta|$ to create multiple photon identification working points. The loose photon identification criteria give higher photon acceptance but lower background rejection compared to the medium and tight working points. For the tight photon identification working point, the cut values are optimized in bins of E_T as well as $|\eta|$, and performed separately for converted and unconverted photon candidates. The analyses described in following Chapters will make use of photon candidates passing both loose and tight identification requirements.

Photon isolation

Background photons are usually accompanied by nearby hadronic activity, whereas prompt photons are well isolated. The isolation variable is defined as the transverse energy in a cone of angular size $R = 0.2$ around the photon axis. The transverse energy can be calculated from calorimeter clusters (subtracting off the contributions from pileup and the photon candidate

itself):

$$E_T^{\text{cone20}} = \sum_{c \in \text{ cone}} E_T^c - E_T^{\gamma \text{ clusters}} - E_T^{\text{pileup}} \tag{4.1}$$

The transverse energy can also be calculated from tracks with $p_T > 1$ GeV that originate at z within 3 mm of the primary vertex (subtracting off the contribution of any tracks associated with a converted photon candidate):

$$p_T^{\text{cone20}} = \sum_{t \in \text{ cone}} p_T^t - p_T^{\gamma \text{ tracks}} \tag{4.2}$$

Photon candidates are considered isolated if the isolation variables are small compared to the photon transverse energy. For the remainder of this work, photon candidates are considered isolated if they satisfy

$$E_T^{\text{cone20}} < 0.065 \times E_T^\gamma \text{ and } p_T^{\text{cone20}} < 0.05 \times E_T^\gamma \tag{4.3}$$

This corresponds to the loose photon isolation working point defined in [92].

4.2.1 Primary vertex selection in diphoton events

For studies of the $H \to \gamma\gamma$ decay, selection of the correct diphoton vertex is critical to achieve an accurate measurement of the diphoton invariant mass, which depends not only on the energy of each photon candidate, but also on the angle between them (Equation 1.1).

For $H \to \gamma\gamma$ measurements, the inputs to primary vertex selection include both track information from the ID and photon information from the ECAL. The ECAL was designed with fine granularity in η for the purpose of **photon pointing**: the vertex of the diphoton system can be identified with high precision by the ECAL alone (15 mm resolution in z at η=0).

A multivariate algorithm (neural network) is trained to select the primary vertex most compatible with the diphoton vertex from all vertices reconstructed by the ID [93]. For each vertex, the following variables are used to train the neural network:

- the z-position where the photon trajectories intersect the beam axis
- the scalar sum over all tracks associated with the vertex of p_T and p_T^2
- $\Delta\phi$ between the vector sum of the track momenta and the momentum of the diphoton system
- the position of photon conversion vertices, if one or both photons convert to e^+e^- in the ID

The training sample is composed of simulated ggF events. The neural network selects a primary vertex that is within 0.3 mm of the Higgs production vertex in 79% of simulated ggF events at $\sqrt{s} = 13$ TeV. For Higgs production modes with higher multiplicity final states (such as $t\bar{t}H$), this increases to $84 - 97\%$ of events [94]. For $t\bar{t}H(\gamma\gamma)$, the vertex selected by the neural network and the vertex with the maximum sum of track p_T^2 are identical in $> 96\%$ of simulated events.

4.3 Electrons

Like photon candidates, electron candidates are constructed from a topo-cluster of energy deposits in the ECAL. However, the ID signature of an electron differs from that of a photon: electron candidates are associated with a single charged particle track. As for photons, the electron energy resolution is calculated using multivariate regression trained on simulated samples. The absolute energy scale is calibrated using $Z \to ee$ decays and validated on radiative Z decays [91] [92].

For the analyses following in later Chapters, electron candidates must have $p_T > 15$ GeV and $|\eta| < 2.47$ with the region $1.37 < |\eta| < 1.52$ vetoed. In addition, electron candidates must satisfy the identification and isolation criteria detailed in this Section.

Electron identification

In order to separate prompt electrons from secondary or fake electrons, a likelihood discriminant is constructed from sensitive variables [95]. These include

- Hadronic leakage: energy deposited in the hadronic calorimeter near an ECAL cluster
- Shower shape variables based on all three ECAL layers
- Agreement between the ID track and ECAL cluster in measured angle and energy of the electron candidate
- Parameters of the track in the ID, including the amount of transition radiation measured in the TRT

Cuts on the likelihood discriminant are optimized in bins of E_T and $|\eta|$ to give about 88% electron efficiency for typical electroweak processes. Electron candidates in the remainder of this Thesis are required to satisfy these cuts, which correspond to the medium electron identification working point defined in [95].

Electron isolation

As in the case of photons, background electrons are usually accompanied by nearby hadronic activity. The isolation variables used for electrons are analogous to those used for

photons (Equation 4.1 and 4.2). However, because electrons from the decay of very high energy particles can be highly columnated, the track isolation variable $p_T^{\text{varcone20}}$ is defined in a cone of variable size: $R = \min(10 \text{ GeV}/p_T^e, 0.2)$.

Electron candidates are considered isolated if the isolation variables are small compared to the electron transverse energy. For the remainder of this work, electron candidates are considered isolated if they satisfy

$$E_T^{\text{cone20}} < 0.20 \times p_T^e \text{ and } p_T^{\text{varcone20}} < 0.15 \times p_T^e \tag{4.4}$$

This corresponds to the loose electron isolation working point defined in [92].

4.4 Muons

Muon tracks are initially reconstructed separately in the ID (see Section 4.1) and the MS. In the MS, hit patterns in each layer are grouped into straight segments, which are further combined into tracks [96]. Muon tracks with $0.1 < |\eta| < 2.5$ are required to have at least three hits in at least two MS MDT layers. In the very central region of $|\eta| < 0.1$, hits are required in at least one MDT layer and holes larger than one MDT layer are not allowed. Forward muons with $2.5 < |\eta| < 2.7$ are required to leave hits in at least three MS MDT/CSC layers. A χ^2 fit is performed to the MS hits making up each track, and the track is accepted if the χ^2 passes the selection criteria. Muon tracks in the MS and ID are combined by performing a global fit to the muon track using hits in both the ID and MS.

The muon momentum scale and resolution are calculated from $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ decays [96].

For the analyses following in later Chapters, muon candidates must have $p_T > 15$ GeV and $|\eta| < 2.7$ in addition to satisfying the identification and isolation criteria detailed in this Section.

Muon identification

For muon candidates in the central region $|\eta| < 2.5$, the compatibility of the charge-to-momentum ratio measured in the ID and MS is calculated:

$$\frac{|(\frac{q}{p})_{\text{ID}} - (\frac{q}{p})_{\text{MS}}|}{\sqrt{\sigma^2((\frac{q}{p})_{\text{ID}}) + \sigma^2((\frac{q}{p})_{\text{MS}})}} \tag{4.5}$$

where σ indicates the uncertainty on the measurement. Muon candidates are required to have ID/MS compatibility < 7. Because the region $2.5 < |\eta| < 2.7$ is outside of ID acceptance, muon candidates reconstructed in the MS only are accepted in this range.

Muon candidates in the remainder of this Thesis are required to satisfy these cuts, which correspond to the medium muon identification working point defined in [96].

Muon isolation

Prompt muons and muons from the decay of a W, Z, or Higgs boson are well isolated, whereas background muons (e.g. from the decay of hadrons) are typically associated with nearby hadronic activity. The isolation variables used for muons are analogous to those used for electrons. For muon candidates, the variable-sized cone used to calculate the track isolation is allowed to reach a maximum of $R = 0.3$.

Muon candidates are considered isolated if the isolation variables are small compared to the muon transverse energy. For the remainder of this work, muon candidates must satisfy

$$E_T^{\text{cone20}} < 0.30 \times p_T^{\mu} \text{ and } p_T^{\text{varcone30}} < 0.15 \times p_T^{\mu} \tag{4.6}$$

This corresponds to the fixed-cut loose muon isolation working point defined in [96].

4.5 Hadronic Jets

As for photons and electrons, the reconstruction of hadronic jets begins with the formation of topo-clusters from energy deposits in the calorimeters [90] [97]. The topo-clusters are used as input to the **anti-k_T** jet-finding algorithm [98], which groups topo-clusters together based on their p_T and angular separation. Though many jet-finding algorithms have been proposed, the anti-k_T algorithm has the advantage of producing jets of a roughly conical shape, while also satisfying the criteria for infrared and collinear (IRC) safety:

1. Infrared safe: the jet finding is not impacted by the addition/removal of very low p_T constituents

2. Collinear safe: the jet finding is not impacted by splitting one high p_T constituent into two collinear constituents

For the remainder of this work, jets are constructed using the anti-k_T algorithm with $R = 0.4$ and are required to have $p_T >$25 GeV and $|y| < 4.4$. Jets within $\Delta R < 0.4$ of a selected photon or electron are rejected. For jets with $p_T < 50$ GeV and $|\eta| < 2.4$, compatibility with the primary vertex is evaluated using the Jet Vertex Tagger (JVT) likelihood discriminant [99]. In order to suppress contribution from pileup, incompatible jets are rejected.

4.5.1 Flavor Tagging

Jets initiated by bottom quarks contain B-hadrons. Due to their comparatively long lifetimes ($c\tau \sim 450$ μm), these B-hadrons travel a characteristically long distance in the detector before decaying. Thanks to high granularity and good resolution in the pixel detector, such secondary vertices and other track-based observables are measured with high precision. Using multivariate techniques, the track and vertex information is compiled into a single discriminant that separates b-initiated jets from jets initiated by light quarks [100] [101].

Because these discriminants rely on information collected by the Inner Detector, b-tagging is only possible for jets with $|\eta| < 2.4$.

Figure 4.2 shows the performance of the MV2 algorithm for identifying b-jets, with the suffix c20 (c10) indicating that 15% (7%) of the light jet sample is composed of charm jets [102]. The horizontal axis shows the percentage of b-jets that are correctly identified in a simulated sample of $t\bar{t}$ events, and the vertical axis shows the rejection of light jets.

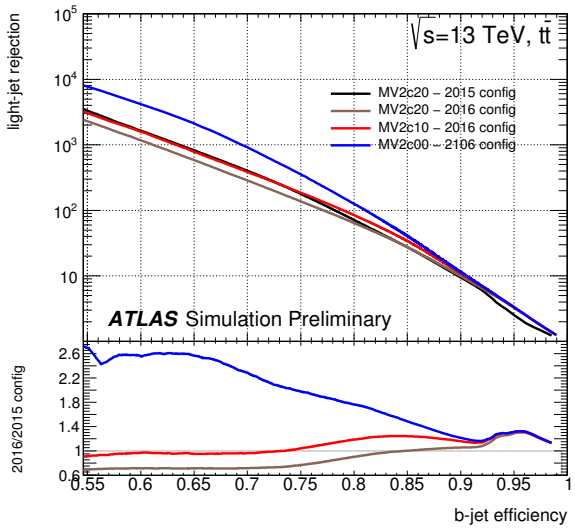


Figure 4.2: Efficiency of b-jet identification vs. rejection of light jets for ATLAS multivariate b-tagging discriminants in $t\bar{t}$ Monte Carlo [102]. The name MV2c20 (MV2c10) indicates that 15% (7%) of the light-jet sample is composed of charm jets.

Multiple working points with different b-jet identification efficiencies are commonly used in ATLAS: these include 65%, 70%, 77%, and 85%. For the remainder of this work, a **b-tagged jet** refers to a jet passing the MV2c10 77% working point: that is, 77% of b-jets in a simulation of $t\bar{t}$ events are correctly identified, which corresponds to $\sim$1 in 100 light jets being misclassified as a b-jet (see Figure 4.2).

It is important to note that the additional tracking space point provided by the IBL led to substantial improvement in ATLAS b-tagging performance. With no changes in the b-tagging algorithm, rejection of light jets increased by a factor of up to four for a fixed b-jet efficiency, with the greatest improvement at working points between 70-80% [103].

4.6 Missing transverse energy

The incoming protons in ATLAS have approximately zero momentum in the plane transverse to the beam. High p_T neutrinos, or BSM particles with very low interaction cross

sections (such as dark matter candidates), pass through the detector without interacting, leaving a large momentum imbalance in the transverse plane: this is referred to as **missing transverse energy (missing E_T)**.

Missing E_T is computed as the negative sum of the transverse momentum of all reconstructed photons, electrons, muons, and jets, plus a soft term reconstructed from tracks that are not associated with these objects [104] [105]. The z component of the missing energy cannot be calculated, since the parton–parton center of mass frame has unknown (and possibly substantial) longitudinal momentum. The missing E_T resolution $\Delta E_T^{\text{miss}} (\sim \sqrt{E_T^{\text{miss}}})$ ranges from about $10 - 25\%$ for $t\bar{t}$ events in ATLAS .

The **missing E_T significance** can be defined in numerous ways. In the following Chapters, an event's missing E_T significance defined as

$$\text{Missing } E_T \text{ significance} \equiv E_T^{\text{miss}} / \sqrt{H_T} \tag{4.7}$$

where H_T is the scalar sum of the p_T of all jets in the event.

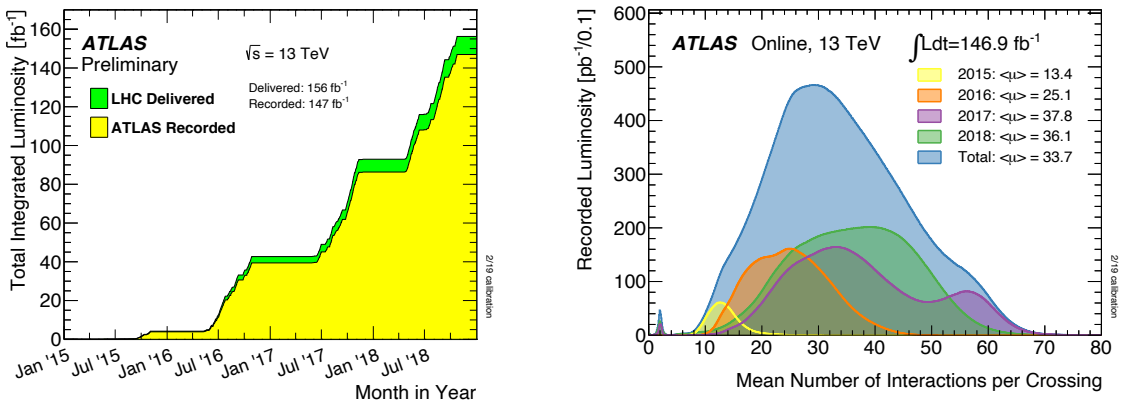
Chapter 5

Data and Monte Carlo samples

5.1 Data: LHC Run 2

The results reported in this Thesis are based on Run 2 proton–proton collision data with center of mass energy $\sqrt{s} = 13$ TeV recorded by the ATLAS detector between 2015 and 2018. The full dataset amounts to an integrated luminosity of 139.0 ± 2.4 fb^{-1} once data quality requirements are imposed (in order to ensure all detector components are operational) [106]. Some results are also reported using partial datasets recorded during 2015-2016 (36.1 ± 1.2 fb^{-1}) and 2015-2017 (79.8 ± 1.6 fb^{-1}).

The integrated luminosity recorded by ATLAS is shown as a function of time in Figure 5.1a. The mean number of interactions per bunch crossing ($< \mu >$) is shown in Figure 5.1b for the full and partial datasets. For the 2015-2016 data taking period, $< \mu >$ was 23, while it increased to 37 for the 2017-2018 data taking period.



(a) Luminosity

(b) Pileup

Figure 5.1: Summary of ATLAS integrated luminosity (a) and pileup conditions (b) during Run 2 data taking.

Data events used in Run 2 $H \to \gamma\gamma$ analyses are initially selected by a trigger requiring two photon candidates with transverse momentum of at least 25 and 35 GeV for the sub-leading and leading photon, respectively. During 2015-2016, photon candidates were required to pass a loose online identification requirement at the trigger level. Due to the increase instantaneous luminosity in 2017-2018, the trigger requirement was tightened to a medium selection to keep the total trigger rate below 20 Hz [107] (see Figure 5.2).

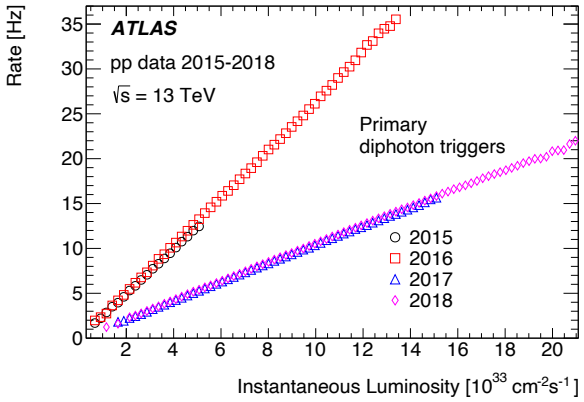
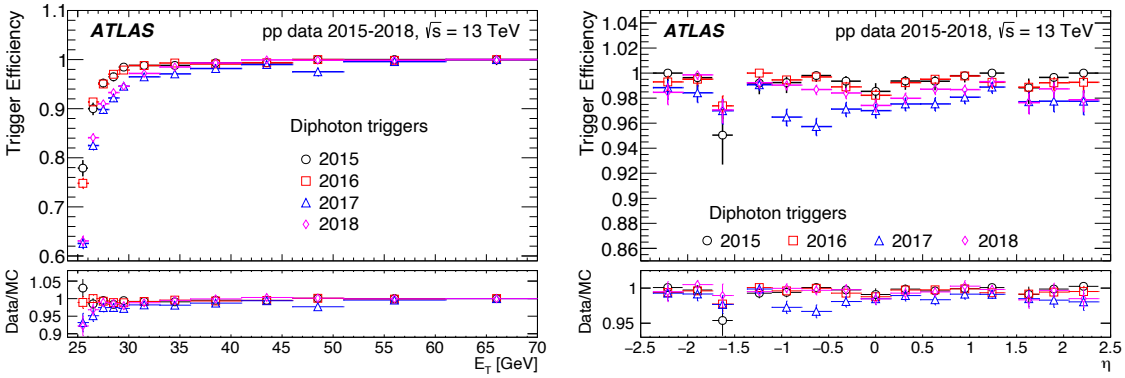


Figure 5.2: The diphoton trigger rate is shown separately for each year of Run 2 data taking [107]. The rate increases linearly with instantaneous luminosity. The difference in slope between the 2015-2016 and 2017-2018 periods is due to the change in photon identification requirement.

The diphoton trigger efficiency is calculated from radiative Z decays. This efficiency is shown as a function of the E_T and η of the lower-p_T photon in Figure 5.3. Once the full diphoton event selection is applied (described in later Chapters), the average trigger efficiency is found to be $> 99\%$ and $> 98\%$ for the 2015-2016 and 2017-2018 data taking periods, respectively.



(a) Sub-leading photon E_T

(b) Sub-leading photon η

Figure 5.3: Efficiency of the diphoton trigger as a function of the E_T and η of the sub-leading (lower p_T) photon. The efficiency is calculated from radiative Z decays. [107]

5.1.1 Data control samples

Events used in $H \to \gamma\gamma$ analysis are required to contain two photons passing the tight identification and isolation requirements described in Section 4.2. Data passing these requirements are referred to as **tight/isolated (TI) data**.

Events passing trigger requirements but failing at least one of the identification or isolation requirements are referred to as **non-tight/isolated (NTI) data**. The size of the NTI dataset (normalized to luminosity) is larger in 2015-2016 than in 2017-2018 due to the switch from looser to tighter photon identification requirement at trigger level.

It is assumed that data events with diphoton invariant mass far from the Higgs signal peak ($|m_{\gamma\gamma} - 125| > 2$) are comprised only of background processes. This region is referred to as the **data sidebands**.

The NTI data and NTI/TI data sidebands are used for multiple purposes in later Chapters, including

- to model the continuum background $m_{\gamma\gamma}$ distribution for the selection of a fit function (Section 6.2.2)
- to model the continuum background in the training of multivariate algorithms (Chapter 9)
- to estimate the continuum background yield under the Higgs signal peak ($m_{\gamma\gamma} = 125 \pm 2$ GeV), for calculation of number-counting significance (Section 9.1)
- to extract the relative contribution of different processes to the continuum background (Section 9.3)

The continuum background yield in the Higgs signal window (n_{TI}^{win}) is estimated by applying a scale factor to the event yield in one of the data sideband regions (either NTI or TI):

$$n_{TI}^{\text{win}} = n_{NTI}^{\text{SB}} \times f_1 \times f_2 \tag{5.1}$$

$$n_{TI}^{\text{win}} = n_{TI}^{\text{SB}} \times f_1 \tag{5.2}$$

The scale factors f_1 and f_2 are calculated separately in the hadronic and leptonic channels according to the following formulae:

$$f_1 = \frac{n_{NTI}^{\text{win}}}{n_{NTI}^{\text{SB}}} \; ; \; f_2 = \frac{n_{TI}^{\text{SB}}}{n_{NTI}^{\text{SB}}} \tag{5.3}$$

The scale factor f_1 scales the number of NTI events in the data sidebands to the number of events in the 125 ± 2 GeV mass window. The second term f_2 scales the number of NTI to the number of TI events by using the ratio calculated in the sidebands. Only f_1 is applied to scale the number of TI sideband events to the number of TI events in the mass window. The scale factors evaluated on the full 139 fb^{-1} of Run 2 data are listed in Table 5.1.

	Preselection	f_1	$f_1 \times f_2$
Hadronic	2γ, 0 e/μ, ≥ 3 jets, ≥ 1 b-jet	0.103	0.013
Leptonic	2γ, ≥ 1 e/μ, ≥ 1 b-jet	0.094	0.016

Table 5.1: Scale factors for estimating continuum background under the Higgs boson mass peak ($m_{\gamma\gamma} = 125 \pm 2$ GeV) from data control regions.

A comparison of yields calculated from the NTI data and the TI data sidebands with the fitted background yield is shown in Table 5.2, using the twenty categories of the $t\bar{t}H(\gamma\gamma)$ CP analysis (Chapter 11) as an example. Because the unscaled NTI sample is much larger than the TI sample, the statistical uncertainty on the scaled NTI prediction is smaller than the uncertainty on the TI. In most categories, the continuum background yield predicted by the scaled NTI data sidebands is smaller than that predicted by TI, and the TI prediction tends to agree better (within uncertainties) with the fitted background. The performance of these background estimates is further discussed in later Chapters.

Category	Scaled NTI SB	Scaled TI SB	Fitted	
$t\bar{t}H(\gamma\gamma)$ CP 1	0.86 ± 0.11	1.9 ± 0.44	1.3	$+0.41/-0.50$
$t\bar{t}H(\gamma\gamma)$ CP 2	1.3 ± 0.13	3.9 ± 0.63	3.0	$+0.55/-0.56$
$t\bar{t}H(\gamma\gamma)$ CP 3	0.90 ± 0.11	1.3 ± 0.37	0.87	$+0.34/-0.32$
$t\bar{t}H(\gamma\gamma)$ CP 4	0.90 ± 0.11	1.5 ± 0.40	1.2	$+0.38/-0.42$
$t\bar{t}H(\gamma\gamma)$ CP 5	1.7 ± 0.15	4.4 ± 0.68	3.9	$+0.63/-0.72$
$t\bar{t}H(\gamma\gamma)$ CP 6	2.0 ± 0.16	3.4 ± 0.59	2.8	$+0.51/-0.57$
$t\bar{t}H(\gamma\gamma)$ CP 7	1.5 ± 0.14	2.5 ± 0.50	2.2	$+0.57/-0.59$
$t\bar{t}H(\gamma\gamma)$ CP 8	6.2 ± 0.28	13.2 ± 1.2	10.4	$+1.1/-1.4$
$t\bar{t}H(\gamma\gamma)$ CP 9	6.3 ± 0.29	6.4 ± 0.81	5.4	$+0.59/-0.77$
$t\bar{t}H(\gamma\gamma)$ CP 10	0.88 ± 0.11	1.8 ± 0.42	1.4	$+0.85/-0.91$
$t\bar{t}H(\gamma\gamma)$ CP 11	12.0 ± 0.39	21.4 ± 1.5	17.6	$+2.6/-2.7$
$t\bar{t}H(\gamma\gamma)$ CP 12	58.5 ± 0.87	69.5 ± 2.7	57.9	$+2.9/-2.9$
$t\bar{t}H(\gamma\gamma)$ CP 13	0.91 ± 0.12	0.85 ± 0.28	0.63	$+0.28/-0.32$
$t\bar{t}H(\gamma\gamma)$ CP 14	1.7 ± 0.16	2.3 ± 0.46	1.9	$+0.41/-0.43$
$t\bar{t}H(\gamma\gamma)$ CP 15	0.83 ± 0.12	1.2 ± 0.34	0.96	$+0.34/-0.35$
$t\bar{t}H(\gamma\gamma)$ CP 16	2.1 ± 0.18	3.5 ± 0.57	2.95	$+0.52/-0.60$
$t\bar{t}H(\gamma\gamma)$ CP 17	2.3 ± 0.19	2.2 ± 0.45	1.75	$+0.45/-0.44$
$t\bar{t}H(\gamma\gamma)$ CP 18	0.85 ± 0.12	1.4 ± 0.36	1.25	$+0.54/-0.55$
$t\bar{t}H(\gamma\gamma)$ CP 19	2.4 ± 0.20	2.5 ± 0.49	2.23	$+0.45/-0.55$
$t\bar{t}H(\gamma\gamma)$ CP 20	3.6 ± 0.24	3.1 ± 0.54	2.92	$+0.42/-0.53$

Table 5.2: Continuum background in the $m_{\gamma\gamma} = 125 \pm 2$ GeV window estimated by scaling the NTI and TI sidebands, compared to the fitted background in this region. Calculated as an example in the twenty $t\bar{t}H(\gamma\gamma)$ CP categories defined in Chapter 11.

5.2 Simulated samples

This Section provides an overview of the generation of the simulated Monte Carlo (MC) samples used in this work. The software used for matrix element calculation and parton showering is summarized in Table 5.3. In samples using Pythia8 or Herwig7 for parton showering, the EvtGen package is used to simulate the decays of heavy hadrons.

Generator	Purpose	Version	References
Pythia8	ME+shower, shower only	8.2.12	[108] [109]
Herwig7	ME+shower, shower only	7.03	[110] [111]
Powheg	ME only	2	[112] [113] [114]
MG5_aMC@NLO	ME only	2.6.2	[52] [53]
Sherpa	ME+shower	1.1	[115] [116] [117] [118]
EvtGen	Decay of heavy hadrons	1.2.0	[119]

Table 5.3: Summary of Monte Carlo generators used for matrix element calculation and parton showering.

Once generated, all simulated Higgs boson events are passed through a Geant4 [120] simulation of the ATLAS detector and reconstructed with the same analysis software used for data [121]. Additional inelastic proton–proton interactions (pileup) are included in the simulation for all generated events such that the average number of interactions per bunch crossing reproduces that observed in data (Figure 5.1b). The pileup collisions were produced using Pythia (8.1.86) with the A2 parameter tune [122] and the MSTW2008lo PDF set [123].

Simulated events are accompanied by a Monte Carlo weight: this weight indicates the relative probability of events in different regions of phase space, while allowing all regions of phase space to be populated somewhat uniformly. Statistical uncertainties on simulated distributions are given by the quadratic sum of the Monte Carlo weights. This corresponds to a variance σ^2 of

$$\sigma^2 = \sum_i w_i^2 \tag{5.4}$$

In order to keep these statistical uncertainties small, simulated samples typically contain $\geq 10\times$ the number of events expected in data.

5.2.1 SM Higgs boson samples

Monte Carlo samples of Higgs boson events with the $H \to \gamma\gamma$ decay were generated for all production modes summarized in Table 5.4. For all samples, the simulation uses Higgs boson mass $m_H = 125$ GeV and width $\Gamma_H = 4.07$ MeV. A **K-factor** is applied to all SM samples that scales the prediction from the generator to the most accurate available theoretical calculation at $m_H = 125.09$ GeV (see Table 2.1). The rate is also scaled by the branching fraction of $H \to \gamma\gamma$ (0.277%).

Process	ME Generator	PDF set	Shower	Tune
ggF	Powheg	PDF4LHC15	Pythia8 (Herwig7)	AZNLO (H7UE)
VBF	Powheg	PDF4LHC15	Pythia8	AZNLO
WH	Powheg	PDF4LHC15	Pythia8	AZNLO
$q\bar{q} \to ZH$	Powheg	PDF4LHC15	Pythia8	AZNLO
$gg \to ZH$	Powheg	PDF4LHC15	Pythia8	AZNLO
$b\bar{b}H$	Powheg	PDF4LHC15	Pythia8	AZNLO
$t\bar{t}H$	Powheg	PDF4LHC15	Pythia8 (Herwig7)	A14 (H7UE)
$t\bar{t}H$	MG5_aMC@NLO	NNPDF30	Pythia8 (Herwig7)	A14 (H7UE)
$tHjb$	MG5_aMC@NLO	NNPDF30	Pythia8 (Herwig7)	A14 (H7UE)
tWH	MG5_aMC@NLO	NNPDF30	Pythia8 (Herwig7)	A14 (H7UE)

Table 5.4: Summary of Standard Model Higgs boson Monte Carlo samples.

Gluon fusion (ggF) is simulated at next-to-next-to-leading-order (NNLO) accuracy in QCD using the Powheg NNLOPS program [124], with the PDF4LHC15 PDF set [125]. The parton-level events are passed to Pythia for computation of parton showering, hadronization and underlying event effects using the AZNLO set of parameters that are tuned to data [126]. The sample is normalized such that it reproduces the total cross section predicted by the N^{3}LO QCD calculation with NLO electroweak corrections applied (Table 2.1). An alternative parton shower sample for ggF uses the same Powheg NNLOPS configuration for the matrix element calculation, but is interfaced instead to Herwig7 with the H7UE parameter tune [111].

Vector boson fusion (VBF), vector boson associated Higgs production (including $q\bar{q} \to WH$, $q\bar{q} \to ZH$, and $gg \to ZH$), and $b\bar{b}H$ production are all generated at next-to-leading-order (NLO) accuracy in QCD using Powheg with the PDF4LHC15 PDF set. The parton shower is generated by Pythia8 using the AZNLO parameter set.

Standard Model $t\bar{t}H$ production is simulated using multiple matrix element generators. All $t\bar{t}H$ samples are normalized to the cross section calculated at NLO in QCD and EW (see Table 2.1). The nominal $t\bar{t}H$ sample is generated at NLO with Powheg and PDF set PDF4LHC15 and interfaced to Pythia8 for parton showering using the A14 parameter tune [122]. An alternative parton shower sample for $t\bar{t}H$ uses the same Powheg configuration for the matrix element calculation, but is interfaced to Herwig7 for parton showering with the H7UE parameter tune.

A Standard Model $t\bar{t}H$ sample is also generated with MG5_aMC@NLO for use in conjunction with the BSM samples described in Section 5.2.2. The matrix element is calculated at NLO accuracy with the NNPDF30 PDF set [127], and the parton shower is generated using Pythia8 and the A14 parameter tune. The dynamical renormalization (μ_R) and factorization (μ_F) scales are defined as one half the scalar sum of the transverse masses of all final state particles ($\frac{1}{2}\sum m_T$). An alternative parton shower sample uses the same MG5_aMC@NLO configuration for matrix the element calculation, but is interfaced to Herwig7 for parton

showering with the H7UE parameter tune.

Higgs production in association with a single top quark, a b quark and a light quark ($tHjb$) is produced at NLO accuracy in QCD (no electroweak correction) using MG5_aMC@NLO with the four-flavor scheme in the NNPDF30 PDF set. The parton-level events are showered using Pythia8 and the A14 parameter tune. The renormalization and factorization scales are defined as one half the scalar sum of the transverse masses of all final state particles ($\frac{1}{2}\sum m_T$). An alternative parton shower sample for $tHjb$ uses the same MG5_aMC@NLO configuration for the matrix element calculation, but is interfaced to Herwig7 for parton showering with the H7UE parameter tune.

Associated production of a Higgs boson with a single top quark and a W boson (tWH) is produced at NLO accuracy in QCD (no electroweak correction) using MG5_aMC@NLO with the NNPDF30 PDF set, and the parton showering is performed using Pythia8 and the A14 parameter tune. The renormalization and factorization scales are defined as one half the scalar sum of the transverse masses of all final-state particles ($\frac{1}{2}\sum m_T$). At NLO, there is interference between the tWH and $t\bar{t}H$ processes [36], as demonstrated in Figure 2.12. The contribution to tWH from doubly resonant diagrams ($t\bar{t}H$ overlap) is removed, and the interference between the diagrams is maintained as a component of the tWH signal. An alternative parton shower sample for tWH uses the same MG5_aMC@NLO configuration for the matrix element calculation, but is interfaced to Herwig7 for parton showering with the H7UE parameter tune.

5.2.2 BSM Higgs samples

The Higgs Characterization (HC) model [51] is implemented in the MG5_aMC@NLO generator, enabling the generation of Monte Carlo samples with different values of the CP mixing parameters defined in the effective Lagrangian in Equation 2.28. The simulated samples and their parameter settings are listed in Table 5.5.

As in the SM case, the MG5_aMC@NLO samples are interfaced with Pythia8 for parton showering and EvtGen, and the NNPDF30 PDF set is used (five-flavor scheme for $t\bar{t}H$ and tWH, four-flavor scheme for $tHjb$). The renormalization and factorization scales are again defined as $\frac{1}{2}\sum m_T$. As a cross check, samples with SM settings ($\alpha = 0°$, $\kappa_t = 1$) are produced in the HC model and compared to the nominal SM samples: agreement is found to be good for all processes. Since the cross section of ggF production also varies with κ_t and α, MG5_aMC@NLO is also used to generate ggF+2 jets samples at three different values of the CP mixing angle α.

The same K-factors derived for the SM case are applied to all BSM samples, since it was found that the K-factors change little with α. For each sample, the cross section (with K-factor applied) times $H \to \gamma\gamma$ branching fraction is shown in Table 5.6.

κ_t	α	$\cos\alpha$	$\kappa_{SM} = 1/\cos\alpha$
1	0° (SM, CP even)	1	1
1	15°	0.965926	1.035276
1	30°	0.866025	1.154701
1	45°	0.707107	1.414214
1	60°	0.5	2
1	75°	0.258819	3.863703
1	90° (CP odd)	0.000001	10^6
-1	0	1	1
0.5	0	1	1
2	0	1	1
2	45	0.707107	1.414214

Table 5.5: Parameters used in the HC model to produce samples with CP mixing in the Higgs-top coupling. For $\alpha = 90°$: due to numerical precision, $\cos\alpha$ strictly equal to zero cannot be generated, and a value approaching it (10^{-6}) and a defined value for κ_{SM} (10^6) are used for the pure CP odd Monte Carlo samples.

Parameters		Normalized $\sigma \times B_{\gamma\gamma}$ [fb]			
κ_t	α	$t\bar{t}H$	$tHjb$	tWH	ggF
1	0° (SM)	1.150	0.169	0.034	33.1
1	15°	1.113	0.177	0.038	–
1	30°	0.995	0.207	0.048	–
1	45°	0.827	0.266	0.064	53.7
1	60°	0.666	0.382	0.087	–
1	75°	0.545	0.548	0.116	–
1	90°	0.500	0.753	0.150	74.4
-1	0°	–	1.980	0.307	–
0.5	0°	–	0.264	0.030	–
2	0°	–	0.666	0.190	–
2	45°	–	0.570	0.226	–

Table 5.6: NLO cross section times branching ratio for the $t\bar{t}H$, $tHjb$, tWH, and ggF processes for different CP scenarios, normalized with the K-factor. Unfilled entries indicate that no sample with these parameters was generated.

5.2.3 Background samples

Backgrounds for $H \to \gamma\gamma$ include non-Higgs processes with ISR/FSR photons and events containing one or more objects that are misidentified as photons. Such processes result in a background spectrum that is continuous and decreasing in the diphoton mass variable $m_{\gamma\gamma}$, and it is therefore referred to as the **continuum background**.

The exact rate and composition of the continuum background is difficult to model, and the background estimation in $H \to \gamma\gamma$ analyses is done using a data-driven method (see Section 6.2.2). However, simulated samples are needed to validate analysis methods and study the composition of the background in data. Large Monte Carlo samples are particularly critical for the determination of the functional form (not the parameters) of the continuum background $m_{\gamma\gamma}$ spectrum in specific regions of phase space, and for the estimation of the uncertainty on the continuum background shape (see description of the spurious signal test in Section 6.2.2).

In regions targeting $t\bar{t}H(\gamma\gamma)$, the dominant contributions to the continuum background arise from $t\bar{t}\gamma\gamma$ and $\gamma\gamma$+jets processes. The $t\bar{t}\gamma\gamma$ background is generated at leading order in QCD using MG5_aMC@NLO and the NNPDF PDF set, and interfaced to Pythia8 for parton showering (A14 tune). The renormalization and factorization scales are defined as $\frac{1}{2}\sum m_T$. Background events from continuum $\gamma\gamma$+jets production are simulated using the Sherpa event generator with the CT10 PDF set. The matrix element is calculated at leading order in QCD with the real emission of up to three additional partons, and interfaced with the Sherpa parton shower using the default parameter tune for the underlying-event activity.

Since the sample size required for good modeling of $\gamma\gamma$+jets and $t\bar{t}\gamma\gamma$ backgrounds is very large, a fast parametric simulation of the ATLAS detector response is used [121] [128].

Chapter 6

Statistical model

The detection of the Higgs to diphoton decay is possible due to the clean signature and excellent photon energy resolution of the ATLAS detector, as well as the high signal-to-background ratio of this channel compared to other Higgs boson decay modes. These factors make $H \to \gamma\gamma$ a powerful channel for Higgs boson measurements, despite the small branching fraction (0.277%). The $H \to \gamma\gamma$ channel is characterized by a peak in the diphoton invariant mass ($m_{\gamma\gamma}$) spectrum, which rises above a smoothly falling background. The size of the Higgs boson signal is extracted from a signal-plus-background maximum likelihood fit to the $m_{\gamma\gamma}$ spectrum in data.

The statistical tools required for a $H \to \gamma\gamma$ analysis are detailed in this Chapter. The first Section explains the statistical advantages of defining multiple analysis categories with different signal-to-background ratios, an approach that is used for the analyses detailed in later Chapters. Section 6.2 describes the parameterization of the shape of the signal and background $m_{\gamma\gamma}$ distributions. The likelihood model is introduced in Section 6.3, and sources of systematic uncertainty are described in Section 6.4. The final Section describes a common statistical procedure used to reweight data for presentation purposes.

6.1 Motivation for categorization

In the analyses included in this Thesis, events passing a simple preselection are divided into multiple categories, and a maximum likelihood fit is performed simultaneously to the $m_{\gamma\gamma}$ spectra in all categories. The partition of selected events into categories helps to maximize the statistical significance Z: as long as the categories contain different ratios of signal to background events (S/B), the statistical significance of signal process S (e.g. $t\bar{t}H$) is higher in the multi-category case than in the single-category case.

The short proof below demonstrates that this is true for categories containing large numbers of data events, where the statistical significance is well approximated by $Z = S/\sqrt{B}$. The conclusion holds for more complex formulations of the likelihood (including that outlined in Section 6.3), assuming that B is well estimated.

A short proof

Any region can be partitioned into two categories such that Category 1 contains S_1 signal and B_1 background events, and Category 2 contains S_2 signal and B_2 background events. The un-partitioned region contains $S_1 + S_2$ signal and $B_1 + B_2$ background events, and the corresponding signal significance is

$$Z = \frac{S_1 + S_2}{\sqrt{B_1 + B_2}} \tag{6.1}$$

Following the partition into two categories, the signal significance is the sum in quadrature of the single-category significances:

$$Z' = \sqrt{\frac{S_1^2}{B_1} + \frac{S_2^2}{B_2}} \tag{6.2}$$

The following demonstrates that $Z' > Z$ when $S_1/B_1 \neq S_2/B_2$. All event yields S_1, S_2, B_1, and B_2 must be ≥ 0.

$$Z'^2 - Z^2 = \frac{S_1^2}{B_1} + \frac{S_2^2}{B_2} - \frac{(S_1 + S_2)^2}{B_1 + B_2} \tag{6.3}$$

Applying a common denominator gives

$$= \frac{(S_1^2 B_2 + S_2^2 B_1)(B_1 + B_2) - B_1 B_2 (S_1 + S_2)^2}{B_1 B_2 (B_1 + B_2)} \tag{6.4}$$

Expanding and canceling like terms gives

$$= \frac{B_1^2 S_2^2 + B_2^2 S_1^2 - 2 B_1 B_2 S_1 S_2}{B_1 B_2 (B_1 + B_2)} \tag{6.5}$$

which simplifies to

$$= \frac{B_1 B_2}{B_1 + B_2} (\frac{S_1}{B_1} - \frac{S_2}{B_2})^2 > 0 \tag{6.6}$$

QED.

6.2 Signal and background shapes

The $H \to \gamma\gamma$ signal peak and the continuum diphoton background are modeled by analytic functions of $m_{\gamma\gamma}$. This Section outlines the criteria for selecting these functions, which are later fitted to data to extract cross section and signal strength results.

6.2.1 Signal model

The **Double-Sided Crystal Ball function (DSCB)** is described by a Gaussian core and two asymmetric exponential tails:

$$f_{\mathrm{DSCB}}(m_{\gamma\gamma}) = N \times \begin{cases} e^{-t^2/2} & \text{if } -\alpha_{\mathrm{low}} \leq t \leq \alpha_{\mathrm{high}} \\ e^{-\frac{1}{2}\alpha_{\mathrm{low}}^2} \left[\frac{1}{R_{\mathrm{low}}}\left(R_{\mathrm{low}} - \alpha_{\mathrm{low}} - t\right)\right]^{-n_{\mathrm{low}}} & \text{if } t < -\alpha_{\mathrm{low}} \\ e^{-\frac{1}{2}\alpha_{\mathrm{high}}^2} \left[\frac{1}{R_{\mathrm{high}}}\left(R_{\mathrm{high}} - \alpha_{\mathrm{high}} + t\right)\right]^{-n_{\mathrm{high}}} & \text{if } t > \alpha_{\mathrm{high}} \end{cases} \tag{6.7}$$

where $t = (m_{\gamma\gamma} - \mu_{\mathrm{CB}})/\sigma_{\mathrm{CB}}$ and $R = \frac{n}{\alpha}$. The parameters μ_{CB} and σ_{CB} describe mean and width of the Gaussian core, and α_{low}, α_{high}, n_{low}, and n_{high} describe the tails.

In each analysis category, the signal parameters are fit to the inclusive SM Higgs boson Monte Carlo sample (the sum of all SM Higgs boson production modes), which are generated with $m_H = 125$ GeV. The signal shape parameters in each category are taken from these fits. The Higgs boson mass has been measured by ATLAS and CMS [37] to be slightly higher than 125 GeV:

$$m_H = 125.09 \pm 0.21 \text{ (stat)} \ \pm 0.11 \text{ (syst) GeV} \tag{6.8}$$

The mean of the DSCB is therefore shifted by $m_H - 125 = 0.09$ GeV for consistency with the measured Higgs boson mass. This rigid translation is viable because the energy scale systematics are large compared to the size of this shift.

The Monte Carlo $m_{\gamma\gamma}$ distribution and fitted signal shape are shown in Figure 6.1 for two example analysis categories targeting the $t\bar{t}H(\gamma\gamma)$ process (details in Chapter 9). In both categories, the DSCB captures the shape of the Monte Carlo signal. The fitted signal resolution (σ_{CB}) varies between categories depending on photon kinematics, and will be reported separately for each analysis in later Chapters.

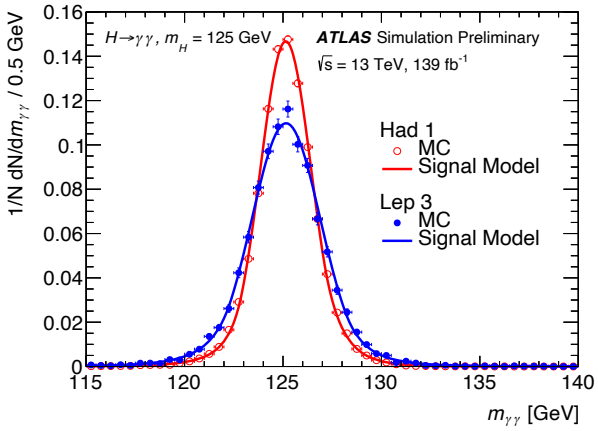


Figure 6.1: Diphoton invariant mass spectrum in Higgs boson Monte Carlo events with the fitted DSCB function overlaid, shown for two categories targeting $t\bar{t}H(\gamma\gamma)$ events with high (low) signal-to-background ratio in red (blue). The $m_{\gamma\gamma}$ resolution (σ_{CB}) varies between the categories due to differences in photon kinematics.

6.2.2 Background model

The continuum background for $H \to \gamma\gamma$ decays is evaluated using a fully data-driven method. Simplified methods (including the estimation of continuum background yield from the data control regions described in Section 5.1.1) are used only for the optimization of analysis cuts and strategies.

The continuum background $m_{\gamma\gamma}$ distribution is modeled by a smoothly falling analytic function. The functional form is chosen using the **spurious signal test** to maximize sensitivity to signal and minimize the systematic uncertainty on the background shape. Full details of the spurious signal test can be found in [129].

The spurious signal test is used only to determine the form of the background function for the final signal-plus-background fit, not the values of the free parameters. The overall background normalization and the values of the parameters are determined in the final signal-plus-background fit.

To perform the spurious signal test, a template $m_{\gamma\gamma}$ distribution is constructed to model the continuum background in each analysis category. The templates are constructed from representative Monte Carlo or data control samples. Each template distribution is checked against the data sidebands to ensure that the chosen sample is a reasonable model of the background distribution in data. A χ^2 fit is used to determine this compatibility and is reported alongside the spurious signal results in later Chapters.

The background template is fitted with a signal-plus-background model in the range 105 GeV $< m_{\gamma\gamma} <$ 160 GeV, where the background shape is the function under evaluation. The signal shape used in this fit is a DSCB function with parameters determined according to the procedure outlined in Section 6.2.1. The number of fitted signal events is computed for Higgs

boson masses varying in intervals of 1 GeV from 121 GeV to 129 GeV. Negative fitted signal is allowed: this indicates that the fit function over-predicts the continuum background. The spurious signal N_{sp} is taken to be the number of fitted signal events with largest magnitude in this 8 GeV window.

The spurious signal test is performed for a variety of candidate background functions:

- Exponential function: $f(m_{\gamma\gamma}) = e^{cm_{\gamma\gamma}}$
- Exponential function of a second order polynomial: $f(m_{\gamma\gamma}) = e^{c_1 m_{\gamma\gamma}^2 + c_2 m_{\gamma\gamma}}$
- Bernstein polynomial of order N=3,4,5: $B_N(m_{\gamma\gamma}) = \sum_{i=0}^{N} c_i \binom{N}{i} m_{\gamma\gamma}^i (1 - m_{\gamma\gamma})^{N-i}$
- Power law function: $f(m_{\gamma\gamma}) = m_{\gamma\gamma}^c$
- Dijet function: $f(m_{\gamma\gamma}) = m_{\gamma\gamma}^{c_1}(1 - m_{\gamma\gamma})^{c_2}$

The selected background function must satisfy at least one of the following criteria:

- $N_{sp} < 0.10 \times N_{s,\text{exp}}$, where $N_{s,\text{exp}}$ is the expected SM number of signal events in the category.
- $N_{sp} < 0.20 \times \sigma_{\text{lumi}}$, where σ_{lumi} is the statistical uncertainty on the fitted number of signal events when fitting the signal-plus-background model to the template. This ensures that the spurious signal is negligible compared with statistical uncertainty.

If more than one candidate function passes the test, the function with the fewest free parameters is selected. In cases where multiple functions with the same number of free parameters pass, the function with the smaller value of N_{sp} is chosen.

In the case of low statistics categories such as $t\bar{t}H$, a single-parameter function is greatly preferred in order to avoid fitting statistical fluctuations in the background templates. The spurious signal criteria are relaxed to accommodate 2σ local statistical fluctuations in the background template. This is done by defining a new variable:

$$\zeta_{sp} = \begin{cases} N_{sp} + 2\Delta_{MC}, & N_{sp} + 2\Delta_{MC} < 0 \\ N_{sp} - 2\Delta_{MC}, & N_{sp} - 2\Delta_{MC} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.9}$$

where Δ_{MC} is the statistical uncertainty on the background template. The new variable ζ_{sp} should then pass the criteria as N_{sp} before. Though the ζ variable is used to choose the background functional form, N_{sp} (not ζ) is used as the spurious signal uncertainty.

The number of spurious signal events in each category is treated as a systematic uncertainty on the signal yield. The selected functions and spurious signal systematics will be reported separately for each analysis in later Chapters.

6.3 Likelihood model

All statistical tests in this work use the profile likelihood ratio as the test statistic:

$$\Lambda(x) = \frac{\mathcal{L}(x, \hat{\hat{\theta}}(x))}{\mathcal{L}(\hat{x}, \hat{\theta})} \tag{6.10}$$

where $\mathcal{L}$ is the likelihood function. In this equation, x represents the parameters of interest (POIs), and θ is the vector of all nuisance parameters (NPs), including systematic uncertainties and parameters describing the shape and normalization of non-Higgs background. The variables $\hat{x}$ and $\hat{\theta}$ denote the values of x and θ that maximize the unconditional likelihood estimate (all NPs are free), while $\hat{\hat{\theta}}(x)$ denotes the values of θ that maximize the conditional likelihood estimate at a fixed x.

The test statistic $-2\ln\Lambda$ is assumed to follow a χ^2 distribution with one degree of freedom [130]. Agreement of the measured x with a test hypothesis is quantified (in the asymptotic approximation) by a p-value from the observed value of $-2\ln\Lambda(x)$.

In the case of a single-category analysis, the likelihood function $\mathcal{L}$ is constructed from a signal-plus-background model. The parameterization of the signal model s and the choice of background functional form b for each category are described in detail in the previous Section. The resulting probability density functions are:

$$s(m_{\gamma\gamma}; \mu_{\mathrm{CB}}, \sigma_{\mathrm{CB}}) \tag{6.11}$$

$$b(m_{\gamma\gamma}; \vec{\xi}) \tag{6.12}$$

The DSCB parameters in the signal function are fixed to the values found by following the procedure in Section 6.2.1. The parameters α^{low}, n^{low}, α^{high}, n^{high} are omitted in the argument of s for the sake of brevity. The background shape parameters $\vec{\xi}$ remain free in the fit.

The most common choice of POI is the **signal strength** μ, which gives the ratio of measured to SM cross sections:

$$\mu_i = \frac{\sigma_i^{\mathrm{meas}}}{\sigma_i^{SM}} \tag{6.13}$$

In order to measure different POI, such as the κ parameters described in Section 2.3.3, the signal strengths can be re-paramtereized (see Table 2.3). Apart from the re-parameterization of the signal strength, the likelihood and test statistic remain unchanged. For the measurement of the CP properties of $t\bar{t}H$, the signal strengths are re-parameterized in terms of POI κ_t and mixing angle α (see Chapter 12).

The likelihood for a single category without systematics takes the form

$$\mathcal{L}(\mu, N_B, \vec{\xi}) = e^{-N_S - N_B} \prod_k \left[N_B b(m_{\gamma\gamma}^k; \vec{\xi}) + N_S s(m_{\gamma\gamma}^k; \mu_{\mathrm{CB}}, \sigma_{\mathrm{CB}}) \right] \tag{6.14}$$

where $m^k_{\gamma\gamma}$ are the measured data points, N_B the number of continuum background events, and N_S the number of measured Higgs boson events, which can be written

$$N_S = \sum_p \mu_p \times N_p \tag{6.15}$$

where N_p is the predicted yield of process p. For a multi-category analysis, the total likelihood function without systematics is the product of the likelihood in each category. All categories use the same parametrization for the likelihood (Equation 6.14), and the POI are correlated across all categories. The background normalization and shape parameters (N^c_B and $\vec{\xi^c}$) are fitted independently in each category c. Measurements of the POI are extracted from an unbinned maximum likelihood fit over the total multi-category likelihood.

6.3.1 Treatment of systematics

Systematic uncertainties are incorporated into the likelihood to account for known sources of imprecision in experimental measurements and theoretical calculations. Each systematic uncertainty has a value (sign and magnitude) σ and an associated **nuisance parameter (NP)** θ that is left free in the maximum likelihood fit. Systematic uncertainties in $H \to \gamma\gamma$ analyses can impact the measured Higgs signal yield from production process p (N_p), the mean of the Higgs signal shape (μ_{CB}), and the width of the Higgs signal shape (σ_{CB}).

For each systematic uncertainty on the Higgs signal yield from production process p, the parameter N_p in the likelihood is modified by a **response term** $(1 + \sigma_i\theta_i)$:

$$N_p \to N_p \prod_i (1 + \sigma_i\theta_i) \tag{6.16}$$

For each systematic uncertainty on the signal mean and resolution, response terms are added to μ_{CB} and σ_{CB} respectively:

$$\mu_{\text{CB}} \to \mu_{\text{CB}} \prod_i (1 + \sigma_i\theta_i) \tag{6.17}$$

$$\sigma_{\text{CB}} \to \sigma_{\text{CB}} \prod_i (1 + \sigma_i\theta_i) \tag{6.18}$$

A **constraint term** associated with each systematic uncertainty modifies the overall likelihood:

$$\mathcal{L}(\mu, \theta) \to \mathcal{L}(\mu, \theta) \prod_i F_i(r_i, \theta_i) \tag{6.19}$$

The means of the constraint terms r_i are **global observables**: these represent the expected values of the NPs θ_i. In the unconditional maximum likelihood fit, all global observables are zero. The constraint term $F(r, \theta)$ takes one of the following forms:

- Gaussian:

$$F(r,\theta) = \exp\left[-\frac{1}{2}\left(\frac{\theta - r}{\sigma}\right)^2\right] \tag{6.20}$$

- Log Normal:

$$F(r,\theta) = \frac{1}{(\theta - r)}\exp\left[-\frac{1}{2}\left(\frac{\ln(\theta - r)}{\sigma}\right)^2\right] \tag{6.21}$$

- Asymmetric: an asymmetric implementation of the log normal constraint with a smooth polynomial interpolation in between [131]

The forms of the constraints for each systematic uncertainty are summarized in Tables 6.2 - 6.3.

For the remainder of this work, the likelihood used in the maximum likelihood fit used includes the systematic uncertainties summarized in Section 6.4.

6.3.2 Generation of Asimov datasets

For each Higgs boson measurement, expected results are evaluated on an **Asimov dataset** [130], which is used to represent the data that would be obtained for a specified value of the POI. Expected results (that is, the results that would be obtained if the data followed the Standard Model) can be derived from a fit to the signal-plus-background Asimov dataset. There are several possibilities for the evaluation of the signal-plus-background Asimov dataset.

Pre-fit Asimov does not use data in the region near the Higgs mass peak, but only data in the sideband region ($|m_{\gamma\gamma} - 125| > 5$ GeV). This Asimov prescription is particularly useful before an analysis is unblinded. A fit is performed to data in the sidebands only, and each global observable is set to the best fit value of the corresponding NP. The SM ($\mu = 1$) signal is super-imposed on the background obtained in this way.

Similarly, **post-fit Asimov** using the full diphoton mass range (105-160 GeV) can be used to compute the expected significance. In this case, a fit is performed to data in the full $m_{\gamma\gamma}$ range, and each global observable fixed to the best fit value of the corresponding NP. This fit can be performed in two ways:

1. Fix the signal strength μ fixed to 1, then perform the fit to data. Set each global observable to the best fit value of the corresponding NP. Generate the $m_{\gamma\gamma}$ distribution.

2. Perform the fit to data with μ profiled. Set each global observable to the best fit value of the corresponding NP. Set μ to 1 and generate the $m_{\gamma\gamma}$ distribution.

The signal-plus-background Asimov datasets (and therefore the expected signal significances) obtained by these different methods are not necessarily the same. Figure 6.2 shows the pre-fit Asimov dataset (green) and the post-fit Asimov datasets obtained by methods 1 (red) and 2 (blue) in an example analysis category targeting the $t\bar{t}H(\gamma\gamma)$ process (details in Chapter 9). Differences are visible in the shape and normalization of the continuum background, as well as in the shape and normalization of the Higgs boson mass peak, due to the different treatment of the global observables in these three prescriptions.

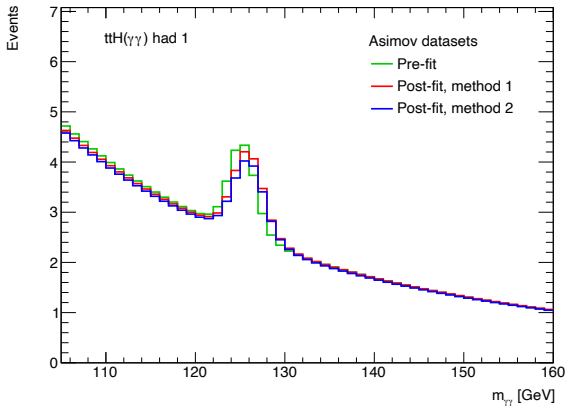


Figure 6.2: Example signal-plus-background Asimov datasets in one $t\bar{t}H(\gamma\gamma)$ analysis category. The pre-fit Asimov dataset is shown in green, and the post-fit Asimov datasets generated with methods 1 and 2 are shown in red and blue respectively.

In the analysis category shown in Figure 6.2, an excess of events in the $m_{\gamma\gamma}$ region 125-130 GeV causes a shift in the position and width of the signal peak in the post-fit Asimov datasets relative to the pre-fit Asimov dataset (which is insensitive to data in the region 120-130 GeV). Because the fit used to generate the Asimov dataset in method 1 assumes a rigid $\mu = 1$ normalization of the signal peak, the observed excess in data results in a higher background normalization relative to method 2. In addition, profiling μ in method 2 can accommodate the excess in observed data without substantial pulls on signal systematic uncertainties. In method 1, these systematics are pulled to fit the excess, resulting in a higher signal normalization than method 2.

For the remainder of this work, expected results are derived on a post-fit Asimov dataset generated by profiling the POI (method 2, corresponding to the blue curve in Figure 6.2).

6.4 Sources of systematic uncertainty

The systematic uncertainties on the $t\bar{t}H(\gamma\gamma)$ measurement can be divided into two primary categories: theoretical uncertainties and experimental uncertainties. The source, form, and impact of these uncertainties is summarized below.

6.4.1 Theoretical uncertainties

Theory systematics are applied to account for missing higher order QCD corrections, imprecise knowledge of the strong coupling constant α_S, and uncertainties on the parton distribution functions (PDFs). These uncertainties can impact the overall Higgs boson production rates as well as the kinematics of Higgs boson events. The latter uncertainties (known as **migration uncertainties**) account for the possibility of event migration between categories, or in/out of selection entirely.

Uncertainties on the predicted Standard Model cross section of process p are not included on the measurement of cross section σ_p. However, these uncertainties are included on measurements of the signal strength μ_p, since this result parameterizes the cross section in terms of the SM prediction. In addition, if a process $q \neq p$ is not measured simultaneously, but is fixed to the SM expectation, then theory uncertainties on the inclusive SM σ_q are included on the measurement of both σ_p and μ_p. Migration uncertainties are included on measurements of both σ_p and μ_p.

The sign and magnitude of each theory uncertainty is calculated separately in each category and for each signal process: this includes the $t\bar{t}H$ process for the cross section measurement described in Chapter 10, and $t\bar{t}H$, $tHjb$, and tWH for the measurement of CP mixing angle described in Chapter 12. Theory uncertainties are evaluated by varying the QCD renormalization and factorization scales (μ_R, μ_F) up and down by a factor of two, and by varying the PDF parameters according to prescriptions laid out in [34] or [13].

For processes with rates much lower than the target process(es), inclusive uncertainties [34] can be applied in every category. The values of these can be found in Table 6.1. These are used for example for $tHjb$, tWH and bbH in Chapter 10.

Process	Order (QCD)	QCD+ [%]	QCD− [%]	PDF [%]
ggF	N^3LO	4.6	6.7	3.2
VBF	NNLO	0.4	0.3	2.1
WH	NNLO	0.5	0.7	1.9
ZH	NNLO	3.8	3.0	1.6
$t\bar{t}H$	NLO	5.8	9.2	3.6
$b\bar{b}H$	5FS NNLO or 4FS NLO	20.1	23.9	–
$tHjb$	LO	6.5	14.7	3.7
tWH	NLO	4.9	6.7	6.3

Table 6.1: QCD scale and PDF uncertainties on inclusive Higgs boson production cross sections at $\sqrt{s} = 13$ TeV with $m_H = 125.09$ GeV [34]. The QCD uncertainties listed for the $b\bar{b}H$ process are combined QCD/PDF uncertainties.

For Higgs boson production modes with large rates (ggF, VBF, and VH), inclusive uncertainties are available. However, existing measurements of V + b-jets [132] [133] and H + b-jets in the $H \to ZZ \to 4\ell$ channel [134] indicate that there is substantial mismodeling in regions where at least one b-jet is required. Figure 6.3 shows a discrepancy of

$\mathcal{O}(100\%)$ between the predicted and measured rates of Higgs boson production with ≥ 1 b-jet, which is far too large to be covered by the inclusive uncertainties in Table 6.1. Since the preselection for the $t\bar{t}H(\gamma\gamma)$ analyses includes a b-jet requirement, a conservative 100% **heavy flavor uncertainty** is assigned to the ggF, VBF, and VH processes to account for this discrepancy.

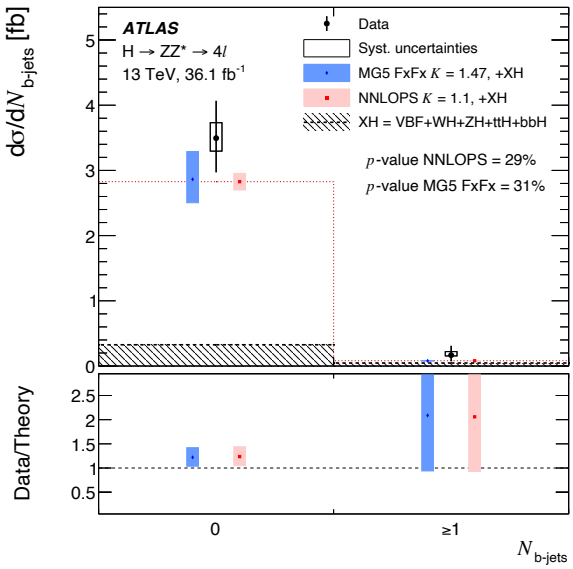


Figure 6.3: Measured Higgs boson production cross section in bins of b-jet multiplicity in the $H \to ZZ \to 4\ell$ decay channel with 36.1 fb^{-1} of 13 TeV data [134]. The discrepancy between data and simulation in the $\geq$1 b-jets bin is $\mathcal{O}(100\%)$.

Migration uncertainties due to Monte Carlo modeling can be calculated by comparing the predictions of different Monte Carlo generators. The **Underlying Event / Parton Showering (UEPS)** systematic accounts for uncertainty in the modeling of the underlying event, parton shower, and hadronization. This uncertainty is evaluated by comparing the same Monte Carlo generator interfaced with different parton showering algorithms (typically Pythia and Herwig). The relative UEPS uncertainty on the yield n_p of process p is

$$\frac{n_p^{\text{Herwig}} - n_p^{\text{Pythia}}}{n_p^{\text{Pythia}}} \tag{6.22}$$

The theory uncertainties and associated nuisance parameters (NPs) are summarized in Table 6.2. The correlation of NPs across categories and the form of the constraint terms are also shown. Details on the different types of constraints are included in Section 6.3.1.

Source	Number of NPs	Constraint type	Correlated
QCD scale	1 per process	asymmetric	yes
LHAPDF: PDF + α_S	1 per process	log normal	yes
PDF4LHC: PDF + α_S	31	log normal	yes
UEPS	1 per process	log normal	yes
Heavy Flavor	1 per process	log normal	yes
$H \to \gamma\gamma$ BR	1	asymmetric	yes

Table 6.2: Summary of the theory uncertainties incorporated into the $H \to \gamma\gamma$ likelihood model. The PDF uncertainty is evaluated either inclusively using the LHAPDF prescription [13] or using the 31 eigenvector PDF4LHC scheme [34].

6.4.2 Experimental uncertainties

Many sources of experimental uncertainty arise from the reconstruction and calibration of physics objects from low-level detector responses. The impact of these uncertainties is obtained by calculating systematic variations provided by ATLAS Combined Performance groups. Experimental uncertainties include

- **Luminosity**, which contributes an uncertainty of 1.7% to the total event yield [106]
- **Trigger**, which contributes an uncertainty of 0.4% to the total event yield [84] [107]
- **Pileup reweighting**
- **Jet uncertainties**, including uncertainties on tracks, vertex identification, jet energy scale and resolution, and flavor composition/tagging [87] [89] [97] [100] [101]
- **Photon uncertainties**, including energy scale, resolution, identification, and isolation [91] [92]
- **Electron uncertainties**, including energy scale, resolution, identification, and isolation [91] [92] [95]
- **Muon uncertainties**, including energy scale, resolution, identification, and isolation [96]
- **Missing** E_T [104] [105]

In a multi-category analysis, the sign and magnitude of these experimental uncertainties are calculated separately in each category.

The choice of an analytic function to model the continuum background also introduces a potential bias to the signal plus background fit. The spurious signal test (Section 6.2.2) quantifies the impact of a given background model on the fitted number of signal events. The spurious signal is therefore included as an experimental systematic uncertainty on the signal yield in each category.

Finally, the uncertainty of the measured Higgs boson mass from the ATLAS and CMS combination [37] is also applied to the mean of the signal DSCB shape (μ_{CB}) by following Equation 6.17.

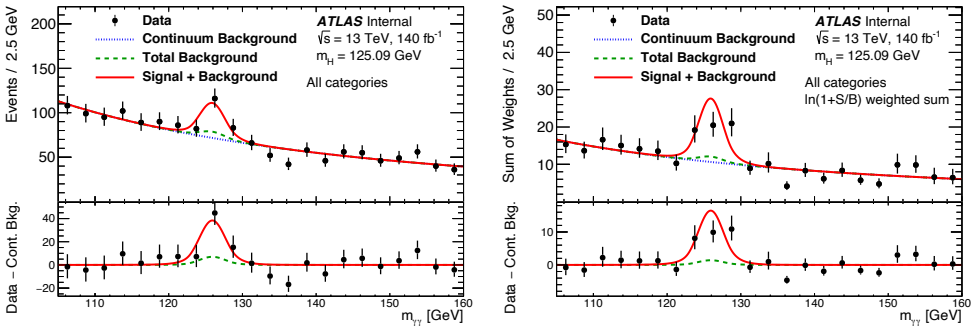
The experimental uncertainties and associated nuisance parameters (NPs) are summarized in Table 6.3. The correlation of NPs across categories and the form of the constraint terms are also shown. Details on the different types of constraints are included in Section 6.3.1.

Source	Number of NPs	Constraint type	Correlated
Yield			
Luminosity	1	log normal	yes
Trigger efficiency	1	log normal	yes
Photon efficiency	3 (ID, isolation, trigger)	asymmetric	yes
Yield and Migration			
Flavor tagging	12	asymmetric	yes
Jets	25	asymmetric	yes
Jet flavor	14	asymmetric	yes
Electrons	2	asymmetric	yes
Muons	7	asymmetric	yes
Missing E_T	3	asymmetric	yes
Pileup	1	asymmetric	yes
Shape			
Photon energy scale	69 (40 merged)	Gaussian	yes
Photon energy resolution	9 (5 merged)	log normal	yes
Spurious signal	1 per category	Gaussian	no
Measured m_H	1	Gaussian	yes

Table 6.3: Summary of the experimental uncertainties incorporated into the $H \to \gamma\gamma$ likelihood model.

6.5 Weighting data for presentation

Every data event carries equal weight in the statistical analysis. However, it can be useful to look at weighted data in order to visualize the power of a multi-category analysis. A weight of $\ln(1 + S/B)$ can be applied to each data event, where S (B) is the expected signal (background) in the analysis category containing the event. This weighting procedure serves to magnify the contribution of the most sensitive categories, which have the highest S/B. Figure 6.4 shows an example of this weighting procedure applied to diphoton mass distribution in the $t\bar{t}H(\gamma\gamma)$ categories described in Chapter 9.



(a) Unweighted (b) Weighted

Figure 6.4: Example diphoton invariant mass distribution in the unweighted (left) and ln(1+ S/B) weighted (right) presentations. The discussion of the particulars of this data is deferred until Section 10.2.

The central value of a single bin in the weighted data distribution is given by the weighted sum over categories:

$$\lambda = \sum_c w_c n_c \tag{6.23}$$

where n_c is the unweighted event yield and $w_c = \ln(1 + S_c/B_c)$ is the calculated weight in category c. The statistical uncertainty on each bin of the weighted data distribution is calculated by following the procedure outlined in [135]. A Poisson distribution $P(n_c/f, \lambda/f)$ is constructed for the scaled variable n_c/f with mean λ/f, where

$$f = \frac{\sum_c w_c^2 n_c}{\sum_c w_c n_c} \tag{6.24}$$

Statistical uncertainties are calculated on this distribution, and the scaling is undone by multiplying by f. The variance thus obtained is

$$f^2\sigma^2 = f\lambda = \sum_i w_i^2 n_i \tag{6.25}$$

The statistical error bars on Figure 6.4b and similar Figures are derived according to this prescription.

Chapter 7

Machine Learning techniques

The goal of Machine Learning (ML) is to make predictions about unknown data based on a large set of example data. As Machine Learning tools have developed and matured, they have been applied with great success to the proton–proton dataset provided by the LHC and the detailed Monte Carlo simulations produced by theorists.

This Chapter provides an overview of the use of the Machine Learning techniques used for the physics analyses included in this Thesis. Section 7.1 provides an overview of multivariate analysis (MVA) strategies, with a focus on classification problems. Section 7.2 describes a class of ML models known as **Boosted Decision Trees (BDTs)**, which are applied extensively throughout the following Chapters. The methods used to optimize ML models through hyper-parameter tuning are described in Section 7.3. For a detailed study of BDT hyper-parameter tuning with a toy dataset, see Appendix D.

7.1 Multivariate analysis

A physics analysis that makes use of Machine Learning techniques is called a **multivariate analysis (MVA)**. Given properties, or "features," $\vec{x}$ of a proton–proton event, an MVA aims predict another property y of the same event. Common features used in MVAs include the energy, flavor, and directional information of the physics objects (or combinations of objects) described in Chapter 4.

The target property y is predicted by finding a function F that approximates F_{true} such that

$$F_{\text{true}}(\vec{x}) = y \tag{7.1}$$

The allowed functional forms of the function F depend on the type of Machine Learning model used (Boosted Decision Tree, Neural Network, Support Vector Machine, etc.). The **complexity** of the model is related to the number of free parameters in the function F: a very complex model has many free parameters, while a simple model has few. The choice of **hyper-parameters** $\vec{h}$ determines the complexity of a model (see Section 7.3).

To begin MVA development, events with known $(\vec{x}, y)$ (e.g. from simulation) are divided into three statistically independent subsets:

1. A **training set**: example points to which the model's free parameters are fit
2. A **validation set**: used select the model's hyper-parameters (see Section 7.3)
3. A **testing set**: used to estimate the final performance of the model

These datasets need not be equal in size. In order to ensure that no bias is introduced, it is important that the model parameters and the hyper-parameters be optimized using independent datasets. The testing set is not used in optimization of the model, but kept as a statistically independent dataset to use in later analysis stages.

Using the training set, the free parameters of the model are chosen to minimize the **loss function**, which quantifies the accuracy of the estimated function F. A common choice of loss function is the mean squared error (MSE), or Euclidean distance between the predictions $F(\vec{x}_i)$ and observation y_i:

$$\text{MSE Loss } = L(F) = \sum_{i=1}^{N} (F(\vec{x}_i) - y_i)^2 \tag{7.2}$$

Another common choice for the loss function is the log loss, which takes the form

$$\text{Log Loss } = L(F) = \sum_{i=1}^{N} -\left[y_i \ln(F(\vec{x}_i)) + (1 - y_i) \ln(1 - F(\vec{x}_i))\right] \tag{7.3}$$

A model that is too complex will result in $F(\vec{x})$ that gives a low value of the loss function when evaluated on the training set, but not when evaluated on the validation set: this is known as **over-training**. A model that is not complex enough will result in $F(\vec{x})$ that performs poorly (high value of the loss function) on both the training and validation sets.

7.1.1 Classification

The MVAs used in the following Chapters address questions of **classification**, where the MVA aims to predict a label indicating what physics processes the event originated from.

A simple model will give a good approximation of F_{true} if the features $\vec{x}$ have very different shapes in the processes under consideration. For interactions that produce complex multi-particle final states, it is often the case that many features have modest shape differences, but no single feature offers powerful discrimination. In such cases, a classifier is a useful tool for combining several features into a single powerful discriminant.

Binary Classification

Binary classification is the simplest and most common type of classification algorithm. In binary classification, there are only two labels to choose from:

$$F_{\text{true}}(\vec{x}) = \begin{cases} 1 \text{ if signal (e.g. } t\bar{t}H) \\ 0 \text{ otherwise} \end{cases} \tag{7.4}$$

The output of a binary classifier evaluated on an event with features $\vec{x}$ is $F(\vec{x}) \in [0, 1]$, with values near 0 indicating more background-like and values near 1 indicating more signal-like. A threshold cut c_{th} is placed on the classifier output such that events with $F(\vec{x}) > c_{\text{th}}$ are labelled signal, and events with $F(\vec{x}) < c_{\text{th}}$ are labelled background.

The overall performance of a binary classifier model depends on:

- **Signal acceptance**, or true positive rate (a): the fraction of signal events that are labelled correctly as signal
- **Background rejection**, or true negative rate (r): the fraction of background events that are labelled correctly as background

An ideal classifier has 100% signal acceptance and 100% background rejection.

The **Receiver Operating Characteristic (ROC)** is commonly used to quantify the performance of a binary classifier. This curve is inscribed by scanning the threshold cut c_{th} and plotting

$$(r(c_{\text{th}}), a(c_{\text{th}})) \tag{7.5}$$

Some example ROC curves are shown in Figure 7.1. The blue curve shows an ideal classifier, and the red curve shows the equivalent of random guessing. Two more realistic models, named Classifier 1 and Classifier 2, are shown in green and black respectively. The performance of Classifier 2 is better than Classifier 1, since the signal acceptance is higher for each value of the background rejection.

The area under the curve (AUC) can also be computed for each ROC curve to measure the relative performance of different classifiers. A perfect classifier has ROC AUC = 1, and a random classifier has ROC AUC = 0.5. ROC curves and ROC AUC will be used to optimize the classifiers developed for the MVAs in later Chapters.

Multiclassification

Multiclassification problems require the function F to divide data in to $N_c > 2$ classes. In this case, the truth label $F_{\text{true}}(\vec{x})$ of an event in class c can be expressed a vector $\{\delta_{c1}, ..., \delta_{cN_c}\}$, where δ_{ij} is the Dirac delta function.

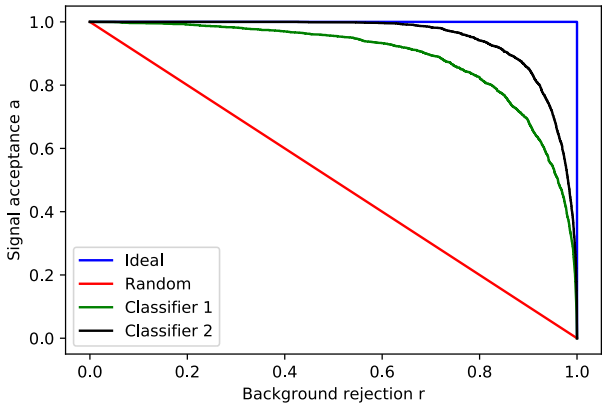


Figure 7.1: Example ROC curves. The blue and red lines show an ideal and a random classifier, respectively. The green and black curves show examples of more realistic classifiers.

The output of a multiclassier takes the form $\{F_1(\vec{x}), ..., F_{N_c}(\vec{x})\}$ where $F_c(\vec{x}) \in [0, 1]$ for all classes c. This can be normalized to give a vector of the probabilities of the event falling into each class:

$$P_c(\vec{x}) = \frac{F_c(\vec{x})}{\sum_{i=1}^{N_c} F_i(\vec{x})} \tag{7.6}$$

The predicted class label assigned to each event corresponds to the class with highest probability in the output vector $\{P_1(\vec{x}), ..., P_{N_c}(\vec{x})\}$.

7.2 Boosted Decision Trees

Boosted Decision Tree (BDT) models have historically been a favorite MVA tool for high energy physicists. The basic building block of the BDT is the decision tree, which divides the feature space into rectangular regions (or "leaves") by a sequence of binary splits [136].

Before training a BDT, each event in the training set is assigned an event weight. For background events, the event weight is simply the Monte Carlo event weight (see Section 5.2). For signal events, the event weight is the product of the Monte Carlo event weight and a scale factor **RelativeNorm**. The RelativeNorm parameter specifies the expected relative normalization of the signal and background processes: if signal is expected to be rare compared to background, RelativeNorm is much less than 1.

Given training data $(\vec{x}_1, y_1), ..., (\vec{x}_N, y_N)$ with event weights $(w_1, ..., w_N)$, the BDT prediction $f(\vec{x}_i)$ is a constant assigned to the leaf R_j containing $\vec{x}_i$:

$$f(\vec{x}_i) = a_j \text{ for } R_j \ni \vec{x}_i \tag{7.7}$$

Once leaves $\{R\}$ are defined, the constants $\vec{a}$ are fully determined by minimizing the loss function. The mean squared error loss function in Equation 7.2 is minimized by

$$a_j = \frac{\sum_{y_i \in R_j} w_i y_i}{\sum_{y_i \in R_j} w_i} \tag{7.8}$$

This is equivalent to the fraction of signal observations ($y_i = 1$) contained in the leaf R_j. For binary BDTs in the following Chapters, the inverse of the ROC AUC is used as the loss function that is minimized in the BDT training.

The leaf definitions are obtained by a sequence of binary partitions of the feature space. A greedy algorithm is used to determine the splitting variable x_n and cut value c such that the leaves $R_1 = \{\vec{x} : x_n < c\}$ and $R_2 = \{\vec{x} : x_n > c\}$ minimize the loss function. The same greedy algorithm is then applied separately to the leaves R_1 and R_2, then to their daughter leaves, etc. until some stopping criterion is reached. The constants $\vec{a}$, the splitting variables $\vec{x}_n$, and the cut values $\vec{c}$ are all free parameters of the BDT model.

Figure 7.2 shows a diagram of an example tree obtained in this way. In some events, a feature may be incalculable: for example, the leading lepton p_T is "missing" in an event containing no leptons. Features with missing values are deemed to pass all threshold cuts, as shown in Figure 7.2.

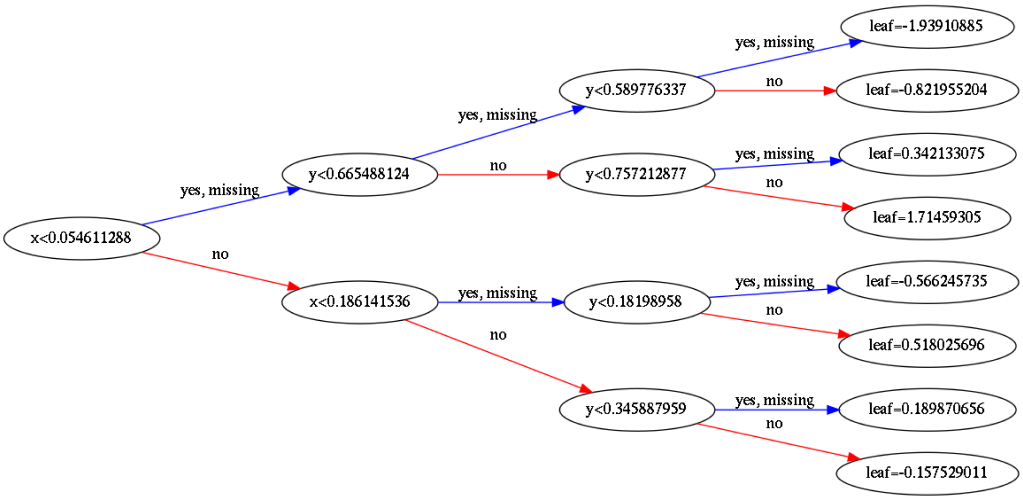


Figure 7.2: Diagram of an example tree of depth 3, which partitions the feature space into eight rectangular regions. The sequence of cuts applied to features x and y is outlined, and the constant output for each leaf is listed.

If the splitting is allowed to proceed indefinitely, the final tree will have a single leaf for each event in the training set. Several hyper-parameters can be adjusted to define stopping conditions and prevent such egregious over-training:

- **MaxDepth**: the depth of the tree is limited to this maximum value. The number of leaves in the tree is at most 2^{MaxDepth}.
- **MinChildWeight**: the number of events in each leaf is required to be above this threshold.

Additional conditions introducing penalties for large BDT weights (values of $\vec{a}$) and small improvements can be expressed as terms in a generalized loss function, called the **objective function**:

$$\text{Objective function } = C(f, \vec{a}) = L(f) + \Omega(\vec{a}) \tag{7.9}$$

where the "complexity function" Ω contains the following penalty terms:

- **L1 regularization**: controls the magnitude of the penalty terms linear in a_i.
- **L2 regularization**: controls the magnitude of the penalty terms quadratic in a_i.
- **MinLossReduction**: each splitting is required to improve the objective by at least this amount.

For a tree with weights $\vec{a} = \{a_1, ..., a_M\}$:

$$\Omega(\vec{a}) = \gamma M + \alpha \sum_{i=1}^{M} |a_i| + \lambda \sum_{i=1}^{M} a_i^2 \tag{7.10}$$

where α and λ are the L1 and L2 regularization hyper-parameters, and γ represents the minimum loss reduction. When these parameters are non-zero, the leaf definitions are chosen to minimize the full objective function.

7.2.1 Boosting

Decision trees are considered weak learners: that is, their predictions tend to have a large variance when trained on different subsets of the same data. In particular, when the first splitting of each tree places a cut on a different feature, the predicted outcome can be very different. **Additive Ensembles**, which sum the predictions of many trees, have much higher accuracy than a single tree [136].

The predicted outcome of an ensemble of T trees on features $\vec{x}$ is

$$F(\vec{x}) = \sum_{t=1}^{T} \beta_t f_t(\vec{x}) \tag{7.11}$$

The sequence of functions f_t and coefficients β_t are determined by a boosting algorithm. Popular boosting algorithms include Adaptive Boosting [137], Logit Boosting [138], and

Gradient Boosting [139]. The Boosted Decision Trees described throughout this Thesis use **Gradient Boosting**, as implemented in the eXtreme Gradient Boosting (XGBoost) software package [140]. The algorithm for Gradient Boosting is outlined below.

An initial decision tree is trained according to the procedure already described, and β_0 is taken to be 1:

$$F^{(0)}(\vec{x}) = f_0(\vec{x}) = a_j^{(0)} \text{ for } R_j^{(0)} \ni \vec{x} \tag{7.12}$$

Then, while $m < T$:

- For each event i, compute the pseudo-residuals for round m:

$$r_i^{(m)} = -\frac{\partial L(f(\vec{x}_i))}{\partial f(\vec{x}_i)}\bigg|_{f=f_{m-1}} \tag{7.13}$$

- Re-weight the training data by taking $w_i' = w_i r_i^{(m)}$. This serves to enhance the importance of training instances that are poorly predicted by preceding trees. Each subsequent round therefore focuses on those events that are hardest to classify.

- Train a single tree f_m on the re-weighted training data.

- Compute α_m so that the loss function is minimized:

$$L(F^{(m-1)} + \alpha_m f_m) \tag{7.14}$$

- Set $\beta_m = \nu\alpha_m$, where ν is the **LearningRate** hyper-parameter. The introduction of the LearningRate parameter ($0 \leq \nu \leq 1$) slows the convergence of the model, ensuring that narrow minima are not overlooked.

- Update the model

$$F^{(m)}(\vec{x}) = F^{(m-1)}(\vec{x}) + \beta_m f_m(\vec{x}) = \sum_{t=1}^{m} \beta_t f_t(\vec{x}) \tag{7.15}$$

This procedure concludes after T rounds, or when early stopping conditions are met. The number of rounds T (**Rounds**) can be treated as a hyper-parameter that requires tuning. Alternatively, an early stopping condition can be implemented which terminates the boosting after k rounds pass without considerable improvement in the loss function.

In order to minimize over-training, a different random subsample of the training data can be used for each boosting round. This procedure is often referred to as bagging. The **Subsample** hyper-parameter can be tuned to specify what fraction of the total dataset is randomly selected at each boosting round.

The impact of tuning individual BDT hyper-parameters is explored in detail in Appendix D. Strategies for performing optimization of a multi-dimensional hyper-parameter space are discussed in Section 7.3.

7.2.2 Multiclass BDT

The binary BDT described above can be generalized to a multiclass BDT, which divides data into $N_c > 2$ classes. N_c binary BDTs are trained, where the signal of BDT i is taken to be events in class i, and the background is taken to be events from all other classes. The output of the multiclass BDT is the vector of the outputs of each binary BDT $\{F_1(\vec{x}), ..., F_{N_c}(\vec{x})\}$.

The multiclass loss function is minimized at each boosting step. To compute the multiclass loss function, the log loss function (Equation 7.3) is computed for each binary tree and summed over all classes:

$$L_c(F_c) = \sum_{i=1}^{N} - \left[\delta_{ic} \ln(F_c(\vec{x}_i)) + (1 - \delta_{ic}) \ln(1 - F_c(\vec{x}_i))\right] \tag{7.16}$$

$$\text{Multiclass log loss} = \sum_{c=1}^{N_c} L_c(F_c) \tag{7.17}$$

7.3 Hyper-parameter optimization

The complexity of an ML model is determined by the choice of its hyper-parameters. Selection of appropriate values for all hyper-parameters is essential to avoid models that are highly sub-optimal (under-trained) or biased (over-trained). The optimal values of a model's hyper-parameters depend on the dataset, and are impossible to predict *a priori*.

In order to optimize hyper-parameters, the loss function $L(F_{\vec{h}})$ is re-parameterized as a function of n hyper-parameters $\vec{h} = (h_1, ..., h_n)$. The re-parametrized loss function $L(\vec{h})$ is then evaluated on the validation set, and the values of $\vec{h}$ that minimize L are chosen as the optimal hyper-parameters. The BDT hyper-parameters that are tuned in the following Chapters have been introduced above: MaxDepth, MinChildWeight, MinLossReduction, L1 Regularization, L2 Regularization, Rounds, LearningRate, Subsample, and RelativeNorm.

Grid scan

The most straightforward approach to hyper-parameter optimization is to perform a grid scan over an n-dimensional rectangular phase space, and select from all points tested the $\vec{h}$ that gives minimal $L(\vec{h})$. This method is easily parallelized, but computationally expensive: scanning n hyper-parameters over s steps each requires training s^n BDTs. The interval of the scan and the size of each step must both be determined before running the scan, and poor choices of these values result in sub-optimal tuning.

The grid scan method of hyper-parameter optimization is applied in the training of the Top Reco BDT described in Chapter 8.

Gaussian processes minimization

A more complex but more efficient approach to hyper-parameter optimization is Bayesian minimization of $L(\vec{h})$ using Gaussian Processes [141], as implemented in the package scikit-optimize [142].

In this method, it is assumed that that $L(\vec{h})$ follows a multivariate Gaussian. Given the allowed range of each hyper-parameter, N_s starting points in the multi-dimensional hyper-parameter space are randomly selected. For each starting point $\vec{h}_i$, the BDT is trained using the corresponding hyper-parameter settings and $L(\vec{h}_i)$ is evaluated. The following procedure is then applied:

- Using the known points $(\vec{h}, L(\vec{h}))$, obtain posterior function L_m.
- Find the $\vec{h}_m$ that minimizes L_m.
- Train the BDT using hyper-parameters $\vec{h}_m$ to get $(\vec{h}_m, L(\vec{h}_m))$.
- Save the minimum of the known points $(\vec{h}, L(\vec{h}))$.
- Increment $m = m + 1$ and repeat until the function L_{N_p} has been obtained.

The hyper-parameters corresponding to the minimum L of the known points are taken to be optimal, and this training model is selected. Throughout this work, $N_s = 10$ and $N_p = 100$.

The Gaussian processes method of hyper-parameter optimization is applied in the training of the BDTs described in Chapters 9 and 11.

Chapter 8

Reconstruction of top decays

Three heavy particles are produced in $t\bar{t}H$ events (one Higgs boson and two top quarks), each of which decays to lighter particles immediately after they are produced. As a result, $t\bar{t}H$ events have final states with particularly high object multiplicity. In the case of $t\bar{t}H(\gamma\gamma)$, the Higgs boson is easy to identify and reconstruct with good mass resolution. However, the $t\bar{t}$-system is more complicated.

The decay of a top quark via $t \to bW$ is dominant, and all other modes are neglected. However, the subsequent decay of the W boson allows for many possibilities, which are enumerated in Table 8.1.

Decay mode	Branching fraction [%]
$W \to e\nu_e$	10.75 ± 0.13
$W \to \mu\nu_\mu$	10.57 ± 0.15
$W \to \tau\nu_\tau$	11.25 ± 0.20
$W \to qq'$	67.60 ± 0.27

Table 8.1: Branching fraction of W boson decay modes from the Particle Data Group [8].

Events are grouped into different channels based on the number of leptons produced in the decay of the $t\bar{t}$ system. Because only electrons and muons are classified as leptons in the following Chapters, top decays to τ are only considered leptonic if the τ decay produces a prompt e or μ. The breakdown shown in Table 8.2 is obtained by folding the W decay rates in Table 8.1 with the branching fraction of $\tau \to e\bar{\nu}_e\nu_\tau$ or $\mu\bar{\nu}_\mu\nu_\tau$.

The all-hadronic (0ℓ) $t\bar{t}$ decay makes up more than half of $t\bar{t}H$ events. A Feynman diagram of this final state is shown in Figure 8.1. Assuming that each final state parton corresponds approximately to one jet in $t\bar{t}H(\gamma\gamma)$ events (due to small rates of highly boosted objects), the hadronic decay of the top quark can be reconstructed by grouping together the correct set of three jets. For example, the jets corresponding to b_1, q_1, and q_1' in Figure 8.1 can be combined into a reconstructed top candidate.

	0 τ_{had}	1 τ_{had}	2 τ_{had}	Total
0 ℓ	45.7%	9.9%	0.5%	56.1%
1 ℓ	34.1%	3.7%	0%	37.8%
2 ℓ	6.4%	0%	0%	6.4%

Table 8.2: Fraction of $t\bar{t}$ decays resulting in various numbers of leptons ($\ell = e$ or μ).

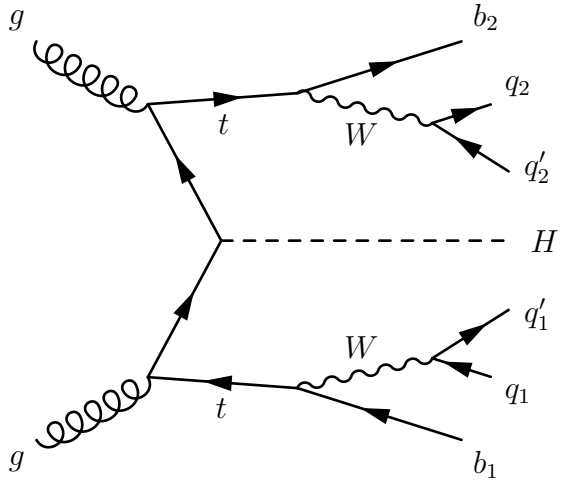


Figure 8.1: Lowest order Feynman diagram of the $t\bar{t}H$ process with an all-hadronic decay of the $t\bar{t}$ system. Additional quarks and gluons, the result of pileup or QCD radiation, can further complicate this final state.

With many jets in each final state, finding the correct groupings of three jets (or **triplets**) is a tricky combinatorial problem. Additional complications include the presence of extra jets resulting from pileup or QCD radiation, or the failure of some jets from top quark decay to meet fiducial requirements.

This Chapter describes a novel method for reconstructing the decay of top quarks using a **Top Reconstruction BDT (Top Reco BDT)**. This algorithm is applied in later Chapters to distinguish between $t\bar{t}H$, tH, and backgrounds, as well as between SM and BSM $t\bar{t}H$ and tH scenarios. Section 8.1 gives an overview of the strategy for reconstructing the full $t\bar{t}$ system given the output of the Top Reco BDT, which is trained to identify the triplet most likely to be the result of a hadronic top quark decay. The training of the Top Reco BDT, including the construction of signal and background samples and the training variables tested, is described in Section 8.2.2. Section 8.3 reviews the performance of the Top Reconstruction BDT in events with different n_ℓ.

8.1 Reconstruction strategy

8.1.1 Hadronic Channel

The hadronic channel preselection targets the all-hadronic decay of the $t\bar{t}$ system by requiring that each event contain 2 photons passing tight identification/isolation criteria, 0 leptons, ≥ 1 b-jet and ≥ 3 total jets. The same hadronic preselection is used throughout later Chapters. Since all preselected hadronic events contain at least three jets, there are always enough jets to construct at least one triplet. In events with more than three jets, all possible triplets are constructed. The Top Reco BDT score is evaluated for each triplet, and the triplet with highest score is selected as the primary top candidate.

In events with six or more jets, a second hadronic top can be fully reconstructed by following the same procedure and excluding jets already assigned to the first top. In events with four or five jets, the second top can only be partially reconstructed: it is taken to be the sum of the jets remaining after the primary top is reconstructed. In events containing exactly three jets, no second top is reconstructed.

Table 8.3 contains a summary of the reconstruction status of the primary and second tops in hadronic events with different jet multiplicity.

n_{jets}	Top 1 (had)	Top 2 (had)
3	full	none
4-5	full	partial
≥ 6	full	full

Table 8.3: Summary of the reconstruction status of the primary and second tops in hadronic events with different jet multiplicity.

Because the average number of jets in $t\bar{t}H(\gamma\gamma)$ events passing hadronic selection is ~ 5.5, many events only contain a single reconstructible top. For this reason, the Top Reco BDT strategy laid out above prioritizes reconstruction of the primary top.

Alternative approaches are available that instead focus on the $t\bar{t}$ system. Rather than sequentially selecting the two top candidates with highest Top Reco BDT score, the average of the two Top Reco BDT scores can be maximized. Tests of this alternative method resulted in very similar performance to the sequential method. However, the computation time needed to calculate the average Top Reco BDT score of each triplet pair is too large for this method to be practically useful. Another alternative for reconstructing the $t\bar{t}$ system would be to train a BDT to identify the correct six jets corresponding to the all-hadronic top decays, as in [143].

8.1.2 Leptonic Channel

The Top Reco BDT is trained to identify hadronic top decays, which are comprised of a b candidate and a hadronic W candidate (see the following Section). However, the same BDT can be applied to semi-leptonic top decays, which are instead comprised of a b candidate and a leptonic W candidate.

In events passing single-lepton preselection (2 photons passing tight identification/isolation criteria, 1 e or μ, and $\geq$ 1 b-jet), a leptonic W candidate is constructed from the lepton and missing transverse energy, which is attributed to a neutrino. The W candidate four-momentum is derived using a dedicated algorithm [144].

The lepton has well-measured four-momentum $(\vec{q}, E_\ell)$ and the neutrino is assigned four-momentum $(\vec{p}, E_\nu)$. The lepton-plus-neutrino system is then assigned the invariant mass of a W boson:

$$m_W^2 = (E_\ell + E_\nu)^2 - (\vec{q} + \vec{p})^2 \tag{8.1}$$

The neutrino is assumed to be massless, and the transverse components of the neutrino momentum are taken to be identical to the missing transverse energy ($p_x = E_x^{\text{miss}}$ and $p_y = E_y^{\text{miss}}$). This gives the following constraint on the z-component of the neutrino momentum, p_z:

$$m_W^2 = \left(E_\ell + \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2 + p_z^2}\right)^2 - \left(\vec{q} + (E_x^{\text{miss}}, E_y^{\text{miss}}, p_z)\right)^2 \tag{8.2}$$

If the transverse mass $m_T < m_W$, then Equation 8.2 has two real solutions for p_z. The solution with smaller $|p_z|$ is chosen, since this gives a momentum vector $\vec{p}$ that is closer in ΔR to the true neutrino momentum in more than half of events.

If $m_T > m_W$, the two solutions to Equation 8.1 have a nonzero imaginary component. In this case, the difference between the transverse energy and neutrino transverse momentum (which is assumed to be zero in Equation 8.1) is adjusted so that the imaginary component vanishes. This is equivalent to requiring $m_W = m_T$ in Equation 8.2, or

$$(\vec{q}_T + \vec{p}_T)^2 = 0 \tag{8.3}$$

The resulting quadratic relationship between p_x and p_y gives two solutions for expressing p_y in terms of p_x: $p_y^{(1)}(p_x)$ and $p_y^{(2)}(p_x)$. For each solution, the distance between the neutrino transverse momentum and the missing transverse energy is calculated:

$$\delta^{(1)}(p_x) = \sqrt{(p_x - E_x^{\text{miss}})^2 + (p_y^{(1)}(p_x) - E_y^{\text{miss}})^2} \tag{8.4}$$

$$\delta^{(2)}(p_x) = \sqrt{(p_x - E_x^{\text{miss}})^2 + (p_y^{(2)}(p_x) - E_y^{\text{miss}})^2} \tag{8.5}$$

The distances $\delta^{(1)}$ and $\delta^{(2)}$ are minimized by $p_x^{(1)}$ and $p_x^{(2)}$, respectively. The corrected missing transverse energy is chosen to be

$$(p_x^{(1)}, p_y^{(1)}(p_x^{(1)})) \text{ or } (p_x^{(2)}, p_y^{(2)}(p_x^{(2)})) \tag{8.6}$$

The selected solution corresponds to the smaller of the minimized $\delta^{(1)}$ and $\delta^{(2)}$. Substituting the corrected missing transverse energy into Equation 8.2 is guaranteed to yield exactly one real solution for p_z.

Once a W candidate has been constructed, the Top Reco BDT score is evaluated for every (leptonic W, jet) combination. The primary top is reconstructed from the leptonic W and the jet giving the highest Top Reco BDT score.

Since the single-lepton preselection requires one lepton and at least one b-jet in every event, there are always enough objects to reconstruct a semi-leptonic top. No top candidate can be reconstructed in dilepton events with this method, as this would require the missing E_T to be split between two semi-leptonic top decays.

In events with at least three jets remaining after semi-leptonic top reconstruction, a hadronic top can also be reconstructed from the jet triplet with highest BDT score. In events with two or three jets, the second top can only be partially reconstructed: it is taken to be the sum of the jets remaining after the primary top is reconstructed. In events containing only a single jet, no second top is reconstructed.

Table 8.3 contains a summary of the reconstruction status of the primary and second tops in leptonic events with different lepton and jet multiplicities.

n_ℓ	n_{jets}	Top 1 (semi-lep)	Top 2 (had)
1	1	full	none
1	2-3	full	partial
1	≥ 4	full	full
2	any	none	none

Table 8.4: Summary of the reconstruction status of the primary and second tops in leptonic events with different lepton and jet multiplicities.

8.2 Training of the Top Reco BDT

8.2.1 Constructing signal and background samples

The goal of the Top Reco BDT is to identify jet triplets that correspond to the decay products of the same top quark: such triplets are designated as signal. The set of background triplets includes all possible wrong combinations. See Table 8.5 for a summary of the signal and background triplets in an event similar to that shown in Figure 8.1. A triplet is correctly matched if its constituent W candidate (jet pair) and b candidate (single jet) are correctly matched to the corresponding truth particles, and these truth particles are daughters of the same top quark.

Signal and background samples are constructed from simulated $t\bar{t}H(\gamma\gamma)$ events using the following procedure:

	b candidate	W candidate
Signal	b_1	$q_1 + q_1'$
Signal	b_2	$q_2 + q_2'$
Background	b_1	$q_2 + q_2'$
Background	b_1	$q_1 + q_2'$
Background	b_1	$q_2 + q_1'$
Background	b_1	b_2 + anything
Background	b_2	$q_1 + q_1'$
Background	b_2	$q_1 + q_2'$
Background	b_2	$q_2 + q_1'$
Background	b_2	b_1 + anything
Background	q_1, q_1', q_2 or q_2'	anything
Background	QCD/pileup	anything
Background	anything	QCD/pileup + anything

Table 8.5: Summary of signal and background triplets in an example hadronic $t\bar{t}H$ event. b and q represent the jets that are truth matched to the corresponding particles in Figure 8.1.

1. Find the W and b particle daughters of a truth top.
2. Follow the simulated decay chain of the W and b to identify the set of final state hadrons descending from each particle. Identify the two (one) truth jets containing the most hadrons descended from the W (b) (weighted by p_T).
3. Match each truth jet to the closest reconstructed jet in ΔR.
4. If this results in a set of three distinct jets (i.e. no jet is used twice in the triplet), this triplet is truth matched and marked as a signal triplet.
5. All other sets of three distinct jets are considered background triplets.

The training is performed on Powheg+Pythia8 $t\bar{t}H(\gamma\gamma)$ Monte Carlo events passing hadronic preselection. Details of this Monte Carlo sample are included in Chapter 5.

If all three truth jets selected at Step 2 are required to be within acceptance ($p_T > 25$ GeV, $|\eta| < 4.0$), only 59% of events passing hadronic preselection contain at least one truth matched triplet. In events with no truth matched triplet, either the top quark is matched to truth jets that fall out of acceptance, or multiple top decay products are merged into a single jet. Table 8.6 shows the possible failure modes of the truth matching algorithm and how frequently they occur. Truth jets falling below the p_T acceptance threshold are the main cause of failures in the truth matching.

If no cuts are placed on the truth jets matched to a top quark, 79% of events passing hadronic preselection contain at least one truth matched triplet. Relaxing the cuts on truth

	Truth jet has $p_T < 25$ GeV	Truth jet has $\|\eta\| > 4.0$	Truth jets are merged
Leading top	41%	$< 0.1\%$	5%
Sub-leading top	47%	$< 0.1\%$	3%

Table 8.6: Breakdown of the performance of the truth matching algorithm when acceptance cuts are applied to truth jets. Jets falling below the p_T acceptance threshold are the main cause of failures in the truth matching. The leading top is defined as the one with higher p_T.

jets allows triplets to be matched through truth jets that re-enter acceptance after reconstruction. The remaining 21% of events with no truth matched triplet are dominated by events where two top decay products correspond to a single jet.

The total number of possible jet triplets per event depends on the jet multiplicity:

$$n_{\text{triplet}} = \frac{1}{2} n_{\text{jet}}(n_{\text{jet}} - 1) \tag{8.7}$$

The factor of $\frac{1}{2}$ is due to the indistinguishability of the two jets comprising a W candidate. With an average of ~ 5.5 jets per event, the number of background triplets is far larger than the number of signal triplets (≤ 2 per event). Only 10% of background triplets are used in the training, and this subset still contains ten times more triplets than the signal training sample, which includes all truth matched triplets in 50% of Monte Carlo events. The other 50% of Monte Carlo events are reserved for validation and testing.

8.2.2 Training variables

The discriminating variables used to separate the signal (truth matched) triplets from background triplets are:

- The four-vector information of the W boson (dijet) candidate
- The four-vector information of the b quark candidate
- The pseudo-continuous b-tag score of each of the three jets (that is, an integer indicating whether the jet is b-tagged at 85, 77, 70, or 65% working point)
- ΔR between the W and b candidates
- ΔR between the two jets comprising the W candidate
- The trijet mass, or mass of the top candidate

The signal and background distributions for each of these variables are shown in Figures 8.2-8.5. Compared to background triplets, truth matched triplets are composed of W and b candidates with high p_T and central η, and a peak is visible at the W mass in signal triplet distribution (Figures 8.2-8.3). The b-tag score is high for truth matched b candidates and low for truth matched W candidates (Figure 8.4). Figure 8.5 shows that the jets composing the W candidate are highly columnated in signal triplets, as are the W and b candidates. There is a clear peak in the trijet mass distribution of signal triplets at the top quark mass (Figure 8.5).

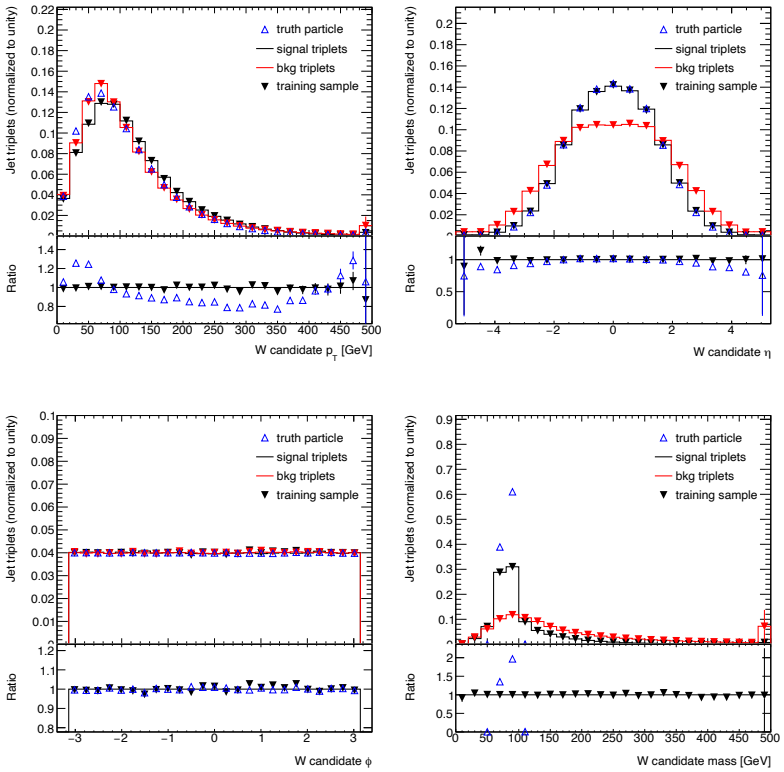


Figure 8.2: Training variables for the Top Reco BDT: p_T, η, ϕ, and mass of the W candidate. Truth matched jet triplets (signal) and unmatched triplets (background) are shown in black and red, respectively. Due to high statistics, agreement is very good between the training (filled triangles) and testing (lines) samples. The open blue triangles show the distribution of truth particles.

The Top Reco BDT is trained using the XGBoost package [140] with these variables as input features. The Top Reco BDT discriminant (Figure 8.6) shows that signal triplets tend to have a very high BDT score, while the random background is spread evenly over the full BDT range. The area under the ROC curve is calculated to be 90.24%. The most important features for BDT training are calculated using the SHAP package [145]: the b-tag score of

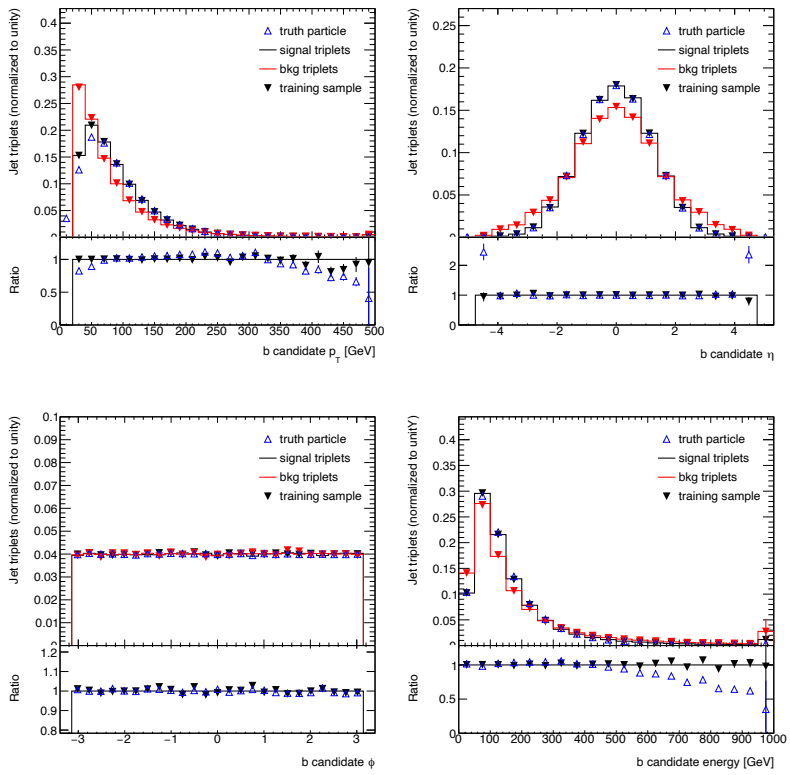


Figure 8.3: Training variables for the Top Reco BDT: p_T, η, ϕ, and energy of the b candidate. Truth matched jet triplets (signal) and unmatched triplets (background) are shown in black and red, respectively. Due to high statistics, agreement is very good between the training (filled triangles) and testing (lines) samples. The open blue triangles show the distribution of truth particles, where applicable.

the b candidate is the dominant variable, followed by the triplet mass and the mass of the W candidate.

Additional training variables were tested, including

- The p_T, η, and ϕ of the top candidate (trijet system)
- The four-vector information of the W boson candidate *in the rest frame of the triplet*
- The four-vector information of the b candidate *in the rest frame of the triplet*
- ΔR between the W and b candidates *in the rest frame of the triplet*
- ΔR between the two jets comprising the W candidate *in the rest frame of the triplet*
- The number of jets in the event

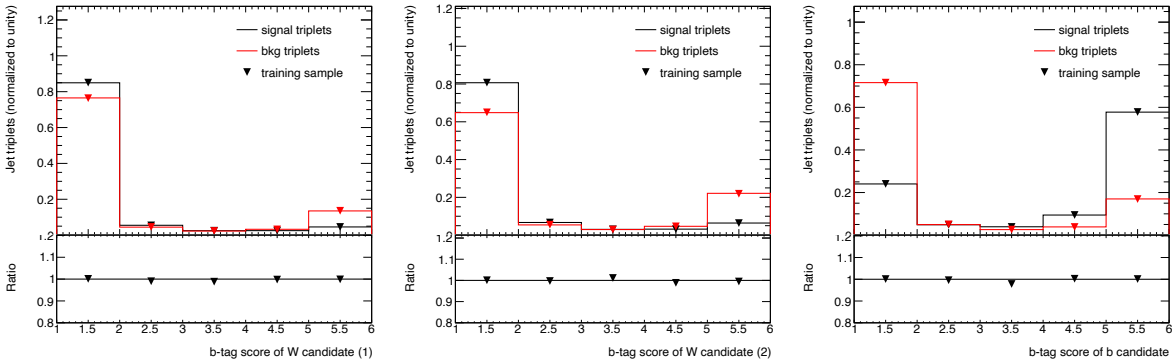


Figure 8.4: Training variables for the Top Reco BDT: pseudo-continuous b-tag scores of each jet in the triplet. High values indicate that the b candidate is b-tagged at a tight working point. Truth matched jet triplets (signal) and unmatched triplets (background) are shown in black and red, respectively. Due to high statistics, agreement is very good between the training (filled triangles) and testing (lines) samples.

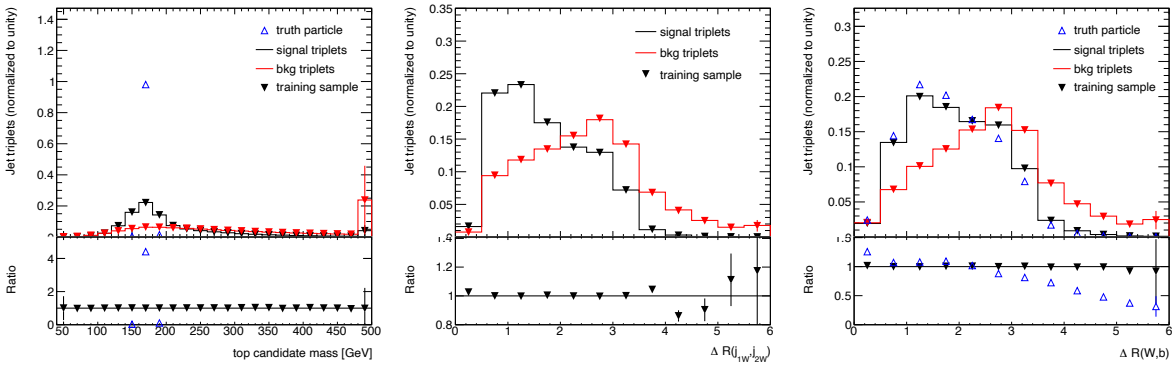


Figure 8.5: Training variables for the Top Reco BDT: triplet mass, ΔR between the W jets, and $\Delta R(W, b)$. Truth matched jet triplets (signal) and unmatched triplets (background) are shown in black and red, respectively. Due to high statistics, agreement is very good between the training (filled triangles) and testing (lines) samples. The open blue triangles show the distribution of truth particles, where applicable.

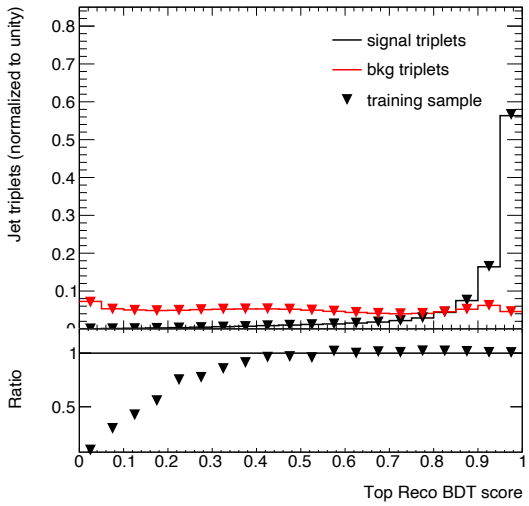


Figure 8.6: Top Reco BDT score. Signal peaks near 1, while the background is evenly spread across the full BDT score range.

- The p_T, η, ϕ, and mass of the system of jets recoiling from the triplet

However, the inclusion of these variables resulted in poorer performance of the Top Reco BDT, as demonstrated in the next Section.

8.3 Performance

The performance of the Top Reco BDT is quantified by counting the number of events in which the algorithm identifies a top candidate (jet triplet) that is truth matched to a top quark. Similarly, the accuracy of W and b identification is quantified by counting number of events in which the selected W and b candidates are truth matched to the corresponding particle. Accuracies are calculated in Powheg+Pythia8 $t\bar{t}H(\gamma\gamma)$ Monte Carlo unless otherwise indicated. The subset of Monte Carlo events used to evaluate performance is statistically independent of that used in training.

The performance of several training setups is compared in this Section. The trainings are named as follows:

1. Nominal: the training variables listed at the beginning of Section 8.2.2
2. Nominal w/o m_t: same training variables used in Nominal, but without m_t
3. Nominal w/o m_W, m_t: same training variables used in Nominal, but without m_W and m_t

4. Boosted: same training variables used in Nominal, but with all kinematic variables calculated *in the rest frame of the triplet*

5. Boosted w/o m_t: same training variables used in Boosted, but without m_t

6. Boosted w/o m_W, m_t: same training variables used in Boosted, but without m_W and m_t

7. Boosted + p_T^t w/o m_W, m_t: same training variables used in Boosted w/o m_W, m_t, plus the p_T of the triplet

Some comparisons are also made to the simple sum-p_T method, which takes the jet triplet with highest p_T as the top candidate.

The performance of the nominal Top Reco BDT is also compared on multiple Higgs Monte Carlo samples.

8.3.1 Hadronic channel

The accuracy of each Top Reco BDT training setup is evaluated for all $t\bar{t}H(\gamma\gamma)$ events passing hadronic preselection. The results are shown in Table 8.7, where the accuracy is shown separately for top, W, and b identification. Note that in some events, the W and b are both correctly identified, but the top is not: in these cases, the W and b identified are not daughters of the same top quark. The Nominal training setup is selected for use in later Chapters based on its high accuracy in this channel.

Training	t-candidate	W-candidate	b-candidate
Nominal	**37%**	**46%**	**57%**
Nominal w/o m_t	35%	45%	57%
Nominal w/o m_W, m_t	35%	43%	57%
Boosted	25%	33%	53%
Boosted w/o m_t	17%	26%	50%
Boosted w/o m_W, m_t	17%	25%	51%
Boosted + p_T^t w/o m_W, m_t	19%	28%	51%

Table 8.7: Percentage of events passing hadronic preselection that contain a correctly reconstructed object (top, W, or b) in $t\bar{t}H(\gamma\gamma)$ Monte Carlo. Shown for different BDT reconstruction algorithms.

The accuracy of top reconstruction for up to two tops is shown in different $t\bar{t}H(\gamma\gamma)$ Monte Carlo samples in Table 8.8. No strong dependence on generator or showering algorithm is observed.

The accuracy of the Top Reco BDT depends strongly on jet multiplicity. The fraction of events that contain no truth matched tops (irreducible background) is higher in events with

	Powheg+Pythia8	MG5_aMC@NLO+Pythia8	MG5_aMC@NLO+Herwig++
1 top correct	37%	36%	33%
2 tops correct	5%	4%	4%

Table 8.8: Percentage of events passing hadronic preselection in which the Top Reco BDT correctly reconstructs one or two tops, compared across several $t\bar{t}H(\gamma\gamma)$ Monte Carlo samples.

fewer jets. At very high jet multiplicity, the total number of jet triplets increases rapidly (see Equation 8.7), and the number of background triplets (combinatorial background) dominates the signal. These effects are demonstrated in Figure 8.7 and Table 8.9, which show the dependence of the Top Reco BDT accuracy on the number of jets in the event. The Top Reco BDT performs best in events with five to seven jets.

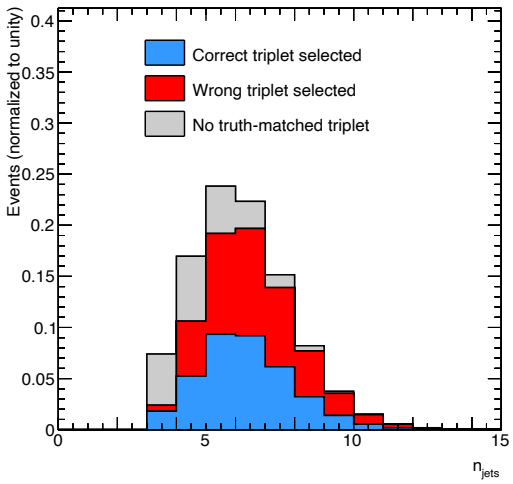


Figure 8.7: Breakdown of hadronic $t\bar{t}H(\gamma\gamma)$ Monte Carlo events containing no top candidate (gray), a wrongly identified top candidate (red), and a correctly identified top candidate (blue). Shown in exclusive bins of jet multiplicity.

The accuracy of each training setup in multiple $H \to \gamma\gamma$ Monte Carlo samples is shown in Table 8.10. The Nominal training has the strongest performance in all three samples.

The Top Reco BDT is also applied in two $H \to \gamma\gamma$ samples where no top quark is present at truth level: ggF and VBF. The kinematics of the selected top candidate in each sample are shown in Figure 8.8. In samples containing true top quarks in the final state ($t\bar{t}H$ and tWH), a top candidate is reconstructed with high p_T and central η. The tops reconstructed in $t\bar{t}H$ and tWH have high Top Reco BDT score and mass peak near 173 GeV. In the samples that do not contain true top quarks (ggF and VBF), only weak peaks are visible in these distributions.

n_{jets}	No truth-matched top	Incorrectly identified top	Correctly identified top
3	58%	18%	24%
4	37%	11%	30%
5	19%	44%	37%
6	12%	50%	38%
7	8%	54%	38%
8	6%	58%	36%
9	5%	61%	34%
10	4%	64%	32%
11	5%	64%	31%
12	4%	67%	29%
13	4%	69%	27%
14	1%	72%	27%

Table 8.9: Percentage of hadronic $t\bar{t}H(\gamma\gamma)$ Monte Carlo events containing no correct top candidate, wrongly identified top candidate, and correctly identified top candidate in bins of jet multiplicity.

	$t\bar{t}H$	$tHjb$	tWH
Maximum in MC	79%	51%	52%
Nominal	**37%**	**31%**	**21%**
Nominal w/o m_t	35%	30%	20%
Nominal w/o m_W, m_t	35%	29%	20%
Boosted	25%	25%	15%
Boosted w/o m_t	17%	19%	10%
Boosted w/o m_W, m_t	17%	19%	10%
Boosted + p_T^t w/o m_W, m_t	19%	19%	11%
Sum p_T	4%	4%	3%

Table 8.10: Percentage of events passing hadronic preselection that contain a correctly reconstructed top in $t\bar{t}H(\gamma\gamma)$ and $tH(\gamma\gamma)$ Monte Carlo. Shown for different BDT reconstruction algorithms.

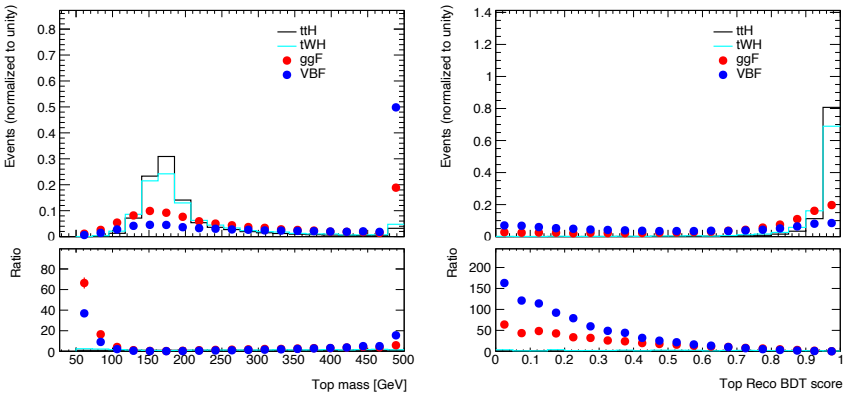


Figure 8.8: Reconstructed top mass and Top Reco BDT score in events passing hadronic preselection. There are clear shape differences between samples with truth tops ($t\bar{t}H$, tWH) and without (ggF, VBF).

8.3.2 Leptonic channel

The accuracy of each Top Reco BDT training setup is evaluated for all $t\bar{t}H(\gamma\gamma)$ and $tWH(\gamma\gamma)$ events passing single-lepton preselection: the results are shown in Table 8.11.

	$t\bar{t}H$	tWH
Nominal	**76%**	**47%**
Nominal w/o m_t	77%	48%
Nominal w/o m_W, m_t	76%	48%
Boosted	71%	44%
Boosted w/o m_t	70%	43%
Boosted w/o m_W, m_t	71%	44%
Boosted + p_T^t w/o m_W, m_t	74%	46%
Sun p_T	19%	9%

Table 8.11: Percentage of events passing leptonic preselection that contain a correctly reconstructed top in $t\bar{t}H(\gamma\gamma)$ and $tWH(\gamma\gamma)$ Monte Carlo. Shown for different BDT reconstruction algorithms.

In the leptonic channel, the Top Reco BDT is also applied to a $t\bar{t}\gamma\gamma$ Monte Carlo sample and a sample where no top quark is present at truth level ($ZH(\gamma\gamma)$). The kinematics of the reconstructed semi-leptonic top candidate in each sample are shown in Figure 8.9. In samples containing true top quarks in the final state ($t\bar{t}H$, tWH, $t\bar{t}\gamma\gamma$), a top candidate is reconstructed with high p_T and central η. The tops reconstructed in $t\bar{t}H$, tWH and $t\bar{t}\gamma\gamma$

have high Top Reco BDT score and mass peak near 173 GeV. In the ZH sample, weaker peaks are visible in these distributions.

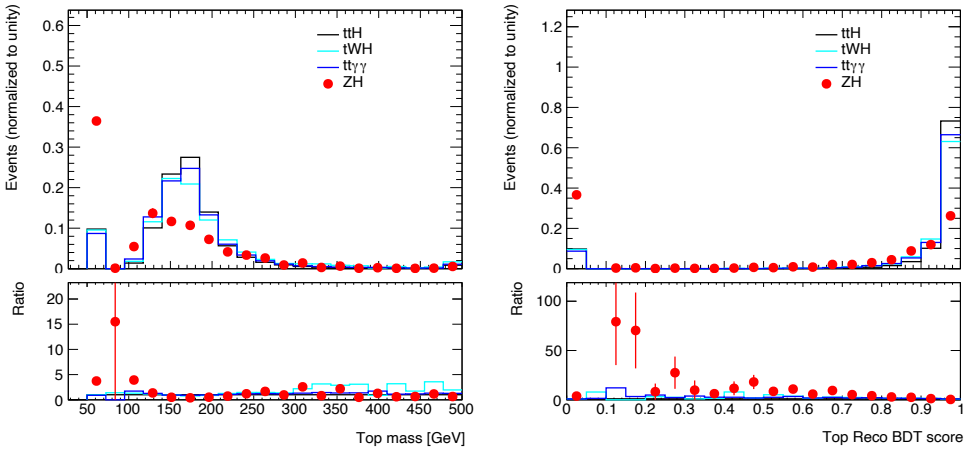


Figure 8.9: Reconstructed top mass and Top Reco BDT score in events passing leptonic preselection. There are clear shape differences between samples with truth tops ($t\bar{t}H$, tWH, $t\bar{t}+\gamma\gamma$) and without (ZH). The underflow bin represents dilepton events, where no top candidate is reconstructed.

In single lepton $t\bar{t}$ events, a log likelihood method is often used to assign final state objects to different parent top quarks. A comparison of one such algorithm (KLFitter [146]) is included in Table 8.12. Here the kinematic range is restricted to the range in which the KLFitter transfer functions are valid (that is, require all objects to have $|\eta| < 2.5$). Performance of the two algorithms is similar in this restricted η range. However, the Top Reco BDT has the advantage of being more widely applicable, both to the full η range and the case of all-hadronic $t\bar{t}$ decays.

Accuracy	KLFitter	Top Reco BDT
Semi-leptonic top correct	25%	23 %
Hadronic top correct	56%	59%
Both tops correct	19 %	20%

Table 8.12: Percentage of $t\bar{t}H(\gamma\gamma)$ Monte Carlo events in which top candidates are accurately reconstructed. The accuracy of KLFitter and the Top Reco BDT is compared.

Chapter 9

Selection of $t\bar{t}H(\gamma\gamma)$ events

In the measurement of $t\bar{t}H(\gamma\gamma)$, events passing a simple preselection are divided into multiple categories, and a maximum likelihood fit is performed simultaneously to the $m_{\gamma\gamma}$ spectra in all categories. The preselection defines two channels:

1. Hadronic channel: 2 photons, 0 e/μ, ≥ 3 jets, ≥ 1 b-tagged jet (77% working point)
2. Leptonic channel: 2 photons, ≥ 1 e/μ, ≥ 1 b-tagged jet (77% working point)

The hadronic preselection targets the all-hadronic decay fo the $t\bar{t}$ system, and the leptonic preselection targets events containing at least one semi-leptonic top decay.

In each channel, an XGBoost [140] Boosted Decision Tree is trained to separate simulated $t\bar{t}H(\gamma\gamma)$ events from continuum background, which is modeled by the NTI data control region (see Section 5.1.1). The figure of merit used to optimize these BDT trainings is described in Section 9.1. Section 9.2 describes the training and optimization of the BDT categories in both channels. Finally, Section 9.3 presents a study composition of the background in the $t\bar{t}H(\gamma\gamma)$ BDT categories.

9.1 Poisson number-counting significance

Since the full likelihood model is quite complex, the multivariate $t\bar{t}H(\gamma\gamma)$ analysis categories are designed to maximize the **number-counting significance**. This number-counting significance is based on the Poisson distribution, which gives the probability of observing k events given an expectation of λ:

$$P(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{9.1}$$

Assuming the background yield B is precisely known, the Poisson distribution can be used to approximate the likelihood. For an observation of $\mu S + B$ events, agreement of the

observed μ with the null hypothesis ($\mu = 0$) is given by the p-value

$$-2\ln\frac{P(\mu S+B,B)}{P(\mu S+B,\mu S+B)} = 2(\mu S+B)\ln(1+\frac{\mu S}{B}) - 2\mu S \tag{9.2}$$

In the case of $t\bar{t}H$, the background yield B includes an estimate of all non-$t\bar{t}H$ Higgs processes as well as the continuum diphoton background. Rather than exercising the full background estimation machinery, the number-counting significance relies on the estimation of continuum background yield from a data control region described in Section 5.1.1.

The number-counting significance gives a rough estimate of the absolute significance of a given set of categories. However, the *relative* performance of two sets of categories is well quantified by a comparison of their number-counting significances. This makes the number-counting significance a powerful figure of merit for the optimization of the $t\bar{t}H(\gamma\gamma)$ BDT categories.

In a single category, the expected significance of a Standard Model observation ($\mu = 1$) (compared to the null hypothesis) is:

$$Z = \sqrt{2(S+B)\ln(1+\frac{S}{B}) - 2S} \tag{9.3}$$

If multiple categories are defined, the total expected significance Z_{tot} is the quadratic sum of the single-category significances.

The statistical uncertainty on Z_{tot} (denoted Y_{tot}) is dominated by the statistical uncertainty on the estimate of continuum background. All Higgs Monte Carlo samples have sufficiently high statistics that their contribution to this uncertainty is negligible. The total significance can be expressed as a function of integers n_i, which represent the number of raw events in the data control region in category i (i.e. the number of events in the NTI control region or TI sidebands). Standard propagation of uncertainties from n_i to Y_{tot} gives the following formula for the statistical uncertainty on Z_{tot}:

$$Y_{\text{tot}}^2 = \sum_i \left(\frac{\partial Z_{\text{tot}}}{\partial n_i}\right)^2 n_i \tag{9.4}$$

In cases where there is a large systematic uncertainty on one or more components of the background, it can also be useful to include this uncertainty in the number-counting significance. For example, it is possible to incorporate the heavy flavor uncertainty on gluon fusion (see Section 6.4) into the number-counting significance. The statistical and systematic uncertainties on the signal yield can be written

$$\sigma_{\text{stat}} = S/Z \tag{9.5}$$

$$\sigma_{\text{syst}} = \lambda n_{ggF} \tag{9.6}$$

where λ is the magnitude of the systematic uncertainty on ggF. The total significance then becomes

$$Z_{\text{total}} = \frac{S}{\sqrt{\sigma^2_{\text{stat}} + \sigma^2_{\text{syst}}}} \tag{9.7}$$

Such systematic uncertainties are included in calculations where indicated.

9.2 Categorization of events

Candidate $t\bar{t}H(\gamma\gamma)$ events must contain two photons with $p_T/m_{\gamma\gamma} > 0.35$ (0.25) for the leading (sub-leading) photon. The diphoton invariant mass must fall in the range 105-160 GeV, and both photons must pass the tight identification and isolation requirements described in Section 4.2.

Events are divided into two channels targeting different decays of the $t\bar{t}$ system using the hadronic and leptonic preselection above. In each channel, a Boosted Decision Tree is trained to separate simulated $t\bar{t}H(\gamma\gamma)$ events from continuum background. In both channels, the continuum background sample used in training corresponds to the NTI data control sample, where at least one photon fails the identification/isolation requirement (see Section 5.1.1). This sample has the high statistics necessary for BDT training, and it is statistically independent of the TI dataset on which the final maximum likelihood fit is performed.

The discriminating variables used in training are low-level inputs such as jet and lepton four-vectors. In most training variables, the true continuum background (the TI data sidebands) is well modeled by the NTI data control sample (see Figures 9.1-9.4 and 9.7-9.10). Discrepancies between NTI and TI are small compared to the difference between $t\bar{t}H(\gamma\gamma)$ simulation and the NTI data control sample.

Once the BDTs are trained, four hadronic categories and three leptonic categories with different signal-to-background ratios are defined by placing cuts on the BDT discriminants. Events with low BDT score are rejected entirely. The remainder of this Section the BDT training and category definition procedure in detail.

The seven categories defined in this Section were first published by ATLAS in $t\bar{t}H(\gamma\gamma)$ measurements at 79.8 fb^{-1} ([147], Section 10.1, Appendix E), and the same categories are used in ATLAS $t\bar{t}H(\gamma\gamma)$ measurements at 139 fb^{-1} (Section 10.2). In addition, the same hadronic and leptonic BDT discriminants are used to define categories for the study of CP violation in $t\bar{t}H(\gamma\gamma)$ included in Chapters 11-12.

9.2.1 Hadronic channel

The hadronic channel targets the all-hadronic decay of the $t\bar{t}$ system. In addition to the photon requirements detailed above, hadronic channel events must pass a preselection requiring at least three jets, with at least one b-tag at the 77% working point. Events with electrons or muons are vetoed.

A hadronic BDT is trained to separate $t\bar{t}H$ Monte Carlo (signal) from the continuum background, which is modeled by the NTI data control region. Both samples are divided into three subsets: 60% of each sample is used for training, another 20% for validation, and the remaining 20% for estimating the final sensitivity. In order to increase statistics in the training and validation sets, the cuts on photon $p_T/m_{\gamma\gamma}$ are replaced with flat cuts of $p_T >$ 35 (25) GeV on the leading (sub-leading) photon.

The variables used to train the hadronic BDT are:

- $p_T/m_{\gamma\gamma}$, η, ϕ of the two photons. Photon p_T is scaled by $m_{\gamma\gamma}$ to reduce sculpting of the diphoton mass spectrum.
- p_T, η, ϕ, and E of the six jets with highest p_T. If there is no jet n, dummy values are given to the BDT in the place of jet n kinematics.
- Boolean b-tag flag (77% working point) for each of the six jets with highest p_T. If there is no jet n, a dummy value is given to the BDT in the place of a b-tag flag for jet n.
- Missing E_T and ϕ direction of missing E_T. In addition to all-hadronic $t\bar{t}$ decays, the hadronic preselection also captures top decays to $b\tau\nu_\tau$ where the τ decays hadronically (see Table 8.2). These events are the source of true missing E_T in the hadronic channel.

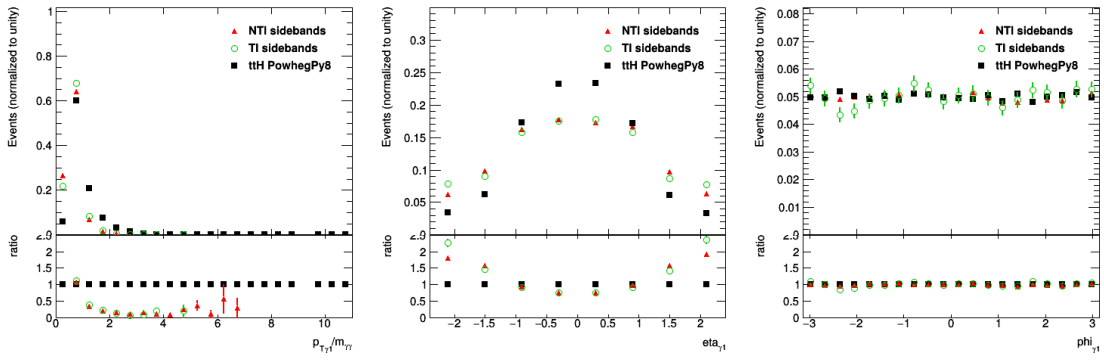
Figures 9.1-9.4 show the signal and background distributions in each of these training variables. The resulting hadronic BDT discriminant is shown in Figure 9.5.

Four cuts are placed on the hadronic BDT discriminant to create five orthogonal regions: events in the region with lowest BDT score are rejected, and the remaining regions make up the four hadronic analysis categories. The cuts forming the category boundaries are treated as hyper-parameters of the analysis, and both the boundary values and BDT hyper-parameters are chosen to maximize the expected Standard Model $t\bar{t}H$ number-counting significance evaluated on the validation set. This significance is given by Equation 9.3 in each category, and the total significance is given by the sum in quadrature over all categories.

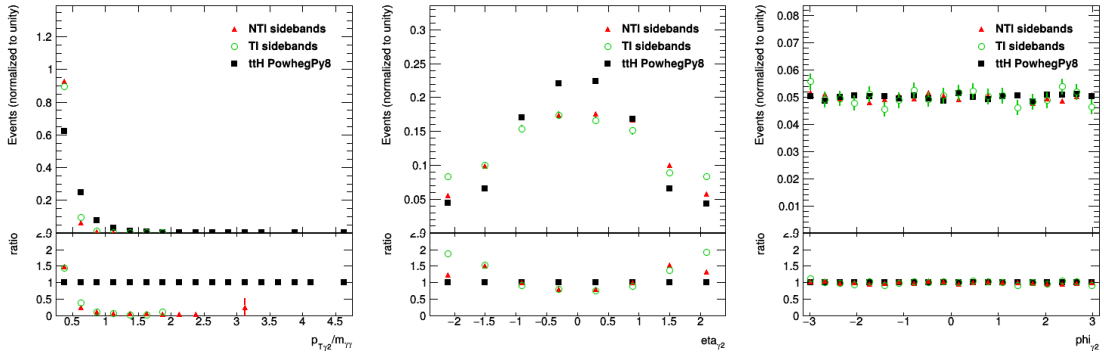
The category boundaries are determined from a brute force scan of all possible boundary combinations. In order to have sufficient data in each category to perform the maximum likelihood fit described in Chapter 6, a minimum of 0.8 continuum background events (predicted from NTI data) is required in each category in the range $123 < m_{\gamma\gamma} < 127$ GeV. This corresponds to ~ 10 events in the full fit range of 105-160 GeV.

The optimal boundaries are shown in Table 9.1. 85% of the $t\bar{t}H$ signal events passing hadronic preselection are selected into these four categories, and 89% of the preselected continuum background events are rejected.

The predicted Higgs boson yield in each category is shown in Table 9.2, along with the estimated continuum background derived by scaling the NTI sidebands. These yields are calculated in the region $123 < m_{\gamma\gamma} < 127$ GeV. The expected number-counting significances of the total Higgs boson signal (Z_H), the $t\bar{t}H$ signal ($Z_{t\bar{t}H}$), and the single top plus Higgs signal (Z_{tH}) are shown in Table 9.3.

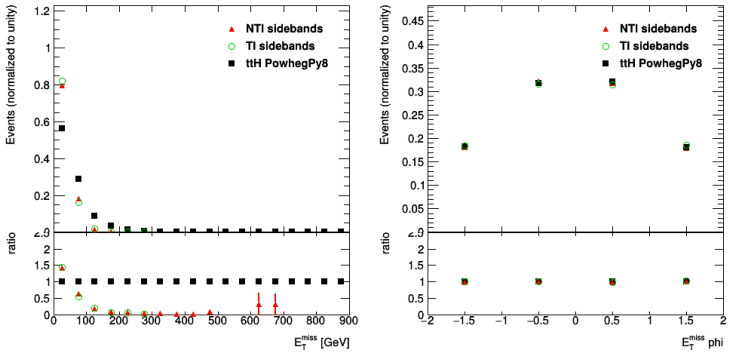


(a) Leading photon $p_T/m_{\gamma\gamma}$ (b) Leading photon η (c) Leading photon ϕ



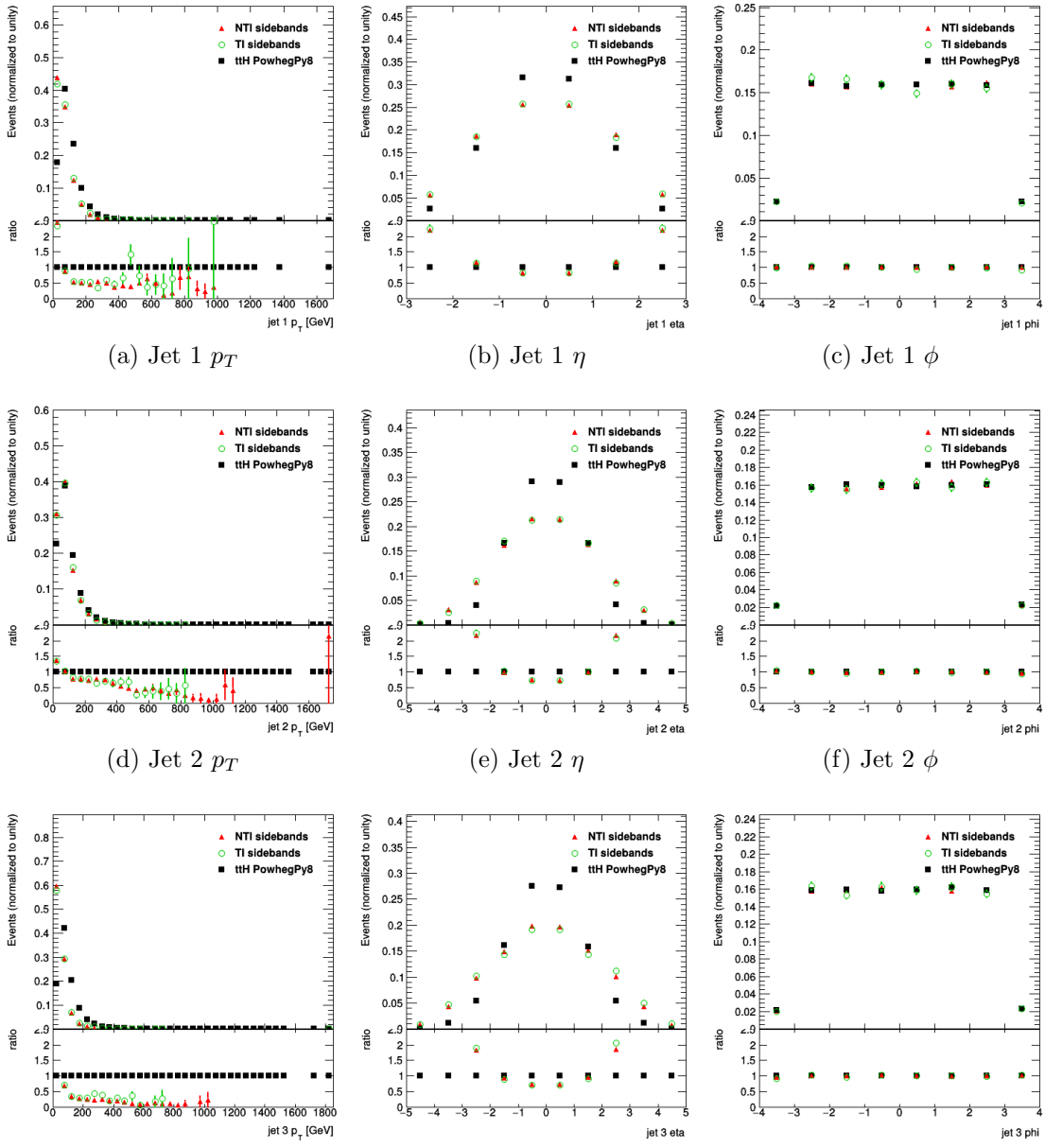
(d) Sub-leading photon $p_T/m_{\gamma\gamma}$ (e) Sub-leading photon η (f) Sub-leading photon ϕ

Figure 9.1: Hadronic BDT training variables: $p_T/m_{\gamma\gamma}$, η, and ϕ of the leading and sub-leading photons. Black squares show $t\bar{t}H$ Monte Carlo signal, and red and green points represent the NTI and TI data sidebands, respectively.



(a) Missing E_T (b) ϕ of missing E_T

Figure 9.2: Hadronic BDT training variables: missing E_T and ϕ direction of missing E_T. Black squares show $t\bar{t}H$ Monte Carlo signal, and red and green points represent the NTI and TI data sidebands, respectively.

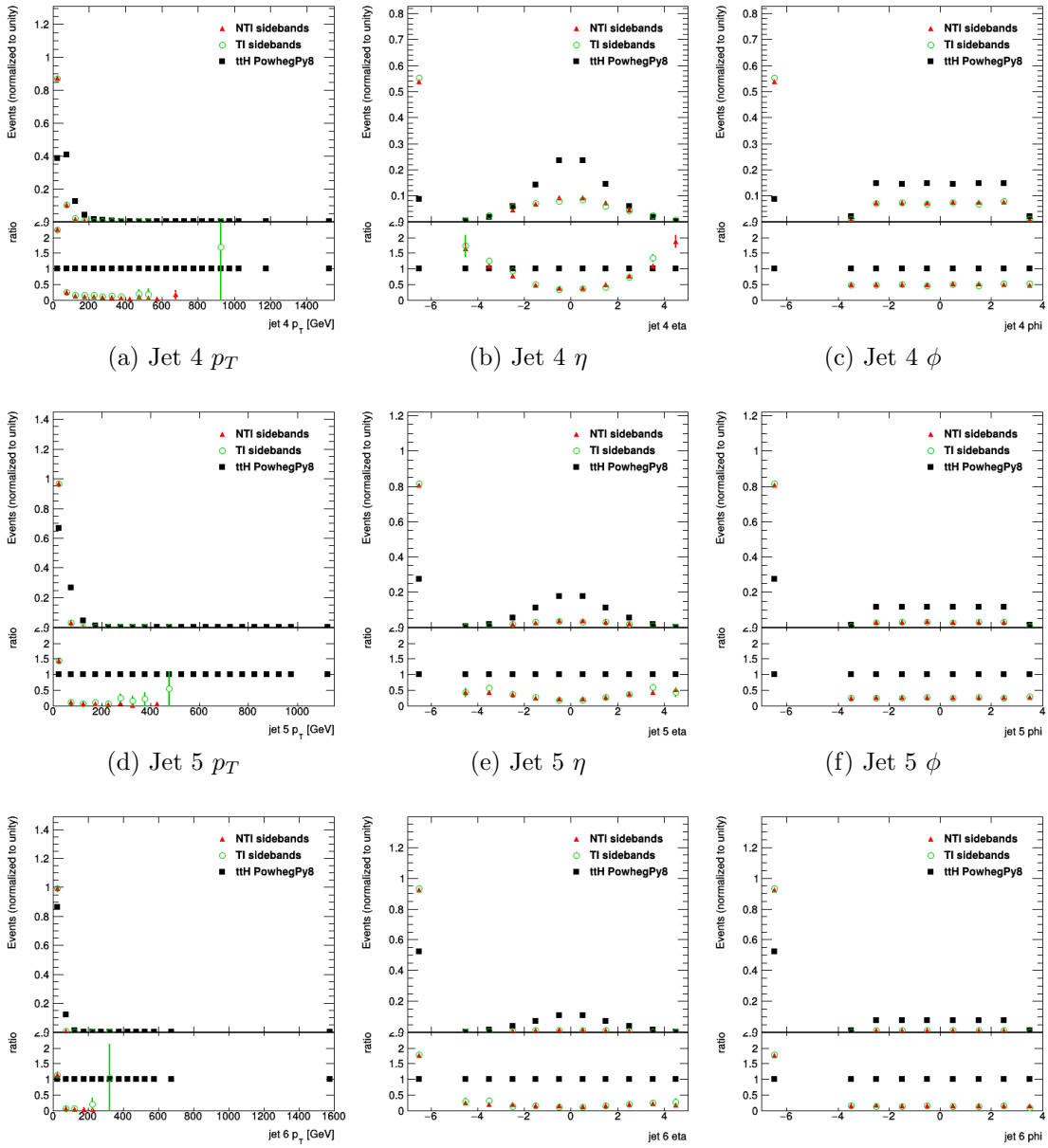


(a) Jet 1 p_T (b) Jet 1 η (c) Jet 1 ϕ

(d) Jet 2 p_T (e) Jet 2 η (f) Jet 2 ϕ

(g) Jet 3 p_T (h) Jet 3 η (i) Jet 3 ϕ

Figure 9.3: Hadronic BDT training variables: p_T, η, and ϕ of the three jets with highest p_T. Black squares show $t\bar{t}H$ Monte Carlo signal, and red and green points represent the NTI and TI data sidebands, respectively.



(a) Jet 4 p_T (b) Jet 4 η (c) Jet 4 ϕ

(d) Jet 5 p_T (e) Jet 5 η (f) Jet 5 ϕ

(g) Jet 6 p_T (h) Jet 6 η (i) Jet 6 ϕ

Figure 9.4: Hadronic BDT training variables: p_T, η, and ϕ of the jets ranked fourth, fifth, and sixth in p_T. Black squares show $t\bar{t}H$ Monte Carlo signal, and red and green points represent the NTI and TI data sidebands, respectively. Events with no jet n are added to the underflow (left-most) bin for the kinematics of jet n.

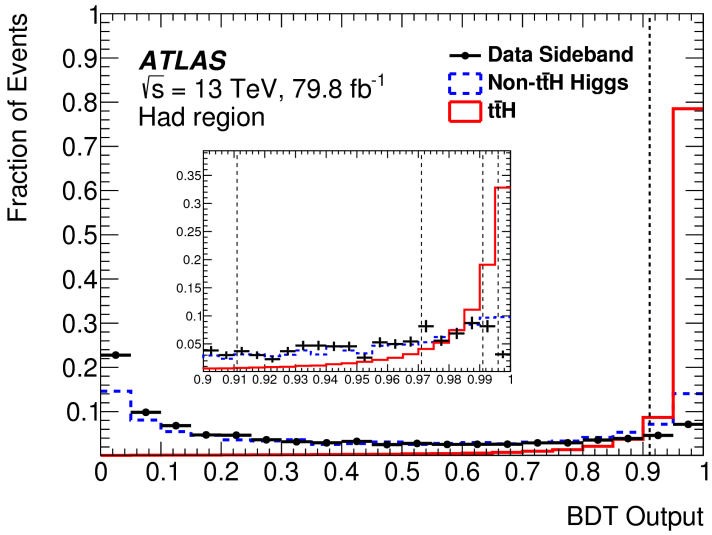


Figure 9.5: Distribution of hadronic BDT score for $t\bar{t}H$ Monte Carlo signal (red), non-$t\bar{t}H$ Higgs processes (blue), and data sidebands (black) with 79.8 fb^{-1} of ATLAS data. Signal (red) and background (blue, black) processes peak at opposite ends of the BDT spectrum. Dashed vertical lines indicate the position of category boundaries, which are most visible in the inlaid plot that zooms in to the signal-like end of the spectrum.

Category	BDT range
$t\bar{t}H$ had 4	[0.911-0.971]
$t\bar{t}H$ had 3	[0.971-0.991]
$t\bar{t}H$ had 2	[0.991-0.996]
$t\bar{t}H$ had 1	[0.996-1]

Table 9.1: Ranges in hadronic BDT score defining the four hadronic $t\bar{t}H(\gamma\gamma)$ categories.

	$t\bar{t}H$	$tHjb$	tWH	ggF	VBF	WH	ZH	cont. bkg.
$t\bar{t}H(\gamma\gamma)$ had 4	2.4	0.154	0.0657	0.657	0.19	0.122	0.338	26.5
$t\bar{t}H(\gamma\gamma)$ had 3	3.92	0.11	0.102	0.471	0.0911	0.0618	0.207	11.4
$t\bar{t}H(\gamma\gamma)$ had 2	2.95	0.0469	0.0835	0.146	0.0245	0.0178	0.0669	2.44
$t\bar{t}H(\gamma\gamma)$ had 1	3.93	0.0299	0.131	0.0897	0.0158	0.0159	0.0518	0.759

Table 9.2: Predicted event yield in the region 123 GeV $< m_{\gamma\gamma} <$ 127 GeV at 79.8 fb^{-1} in the hadronic BDT categories. The continuum background estimate is derived by scaling the data yield in the NTI data sidebands.

	Z_H	$Z_{t\bar{t}H}$	Z_{tH}
$t\bar{t}H(\gamma\gamma)$ had 4	0.746	0.448	0.0399
$t\bar{t}H(\gamma\gamma)$ had 3	1.38	1.06	0.0527
$t\bar{t}H(\gamma\gamma)$ had 2	1.81	1.53	0.0547
$t\bar{t}H(\gamma\gamma)$ had 1	3.23	2.73	0.0728
Combined	4.02	3.33	0.113

Table 9.3: Predicted number-counting significance at 79.8 fb^{-1} in the hadronic BDT categories. The continuum background estimate is derived by scaling the data yield in the NTI data sidebands.

Many combinations of hadronic BDT training variables were tested before settling on the list above, including:

- High-level jet variables (HL): n_{jets}, $n_{\text{cen. jets}}$, $n_{b\text{-jets}}$, missing E_T, H_T
- Photons: $p_T/m_{\gamma\gamma}$ and η for the two photons
- Jets: p_T, η and pseudo-continuous b-tag score for up to six jets
- Tops: Top Reco BDT score (see Chapter 8), p_T, η, and m for the reconstructed top (t_1), p_T, η, and m of the system of jets recoiling against the top (t_2), $\Delta R(t, \gamma\gamma)$, $\Delta R(t_1, t_2)$, $\Delta R(t_2, \gamma\gamma)$

Figure 9.6 compares the performance of several combinations of these variables in terms of number-counting significance. The set of high-level variables includes no photon-related variables, and there is near-perfect agreement between significance calculating using NTI sidebands and TI sidebands. The introduction of the photon variables increases the sensitivity significantly (compare, for example, "Tops, HL, photons" and "Tops, HL"). It is also interesting to note that replacing the low-level jet kinematics with reconstructed top variables gives very similar performance (compare "Tops, HL, photons" and "Jets, HL, photons").

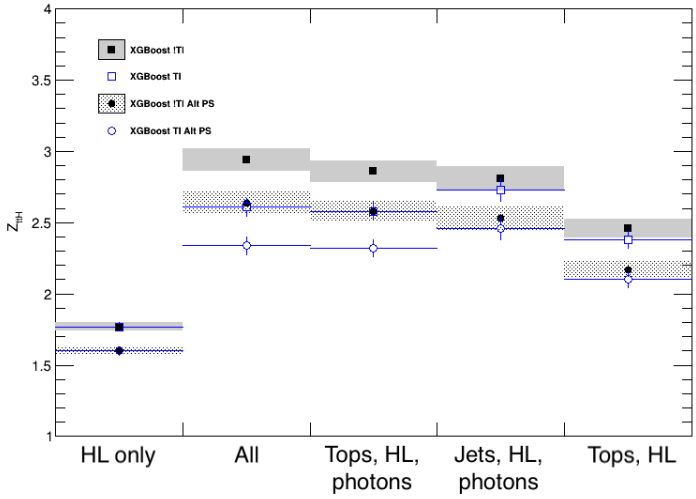


Figure 9.6: Predicted number-counting significance of four hadronic categories, with different input variables used in BDT training. Black (blue) points represent significance calculations with continuum background estimated from the NTI (TI) data sidebands. Square points show significances that use the $t\bar{t}H$ yield from the nominal Monte Carlo sample, and circles show significances that use the yield from a sample using an alternative parton showering algorithm. The errors shown are statistical.

9.2.2 Leptonic channel

The leptonic channel targets final states with at least one top decaying semi-leptonically to an electron or muon. In addition to the photon requirements detailed above, leptonic channel events must pass a preselection requiring at least one b-jet and at least one electron or muon.

A leptonic BDT is trained to separate $t\bar{t}H$ Monte Carlo (signal) from the continuum background, which is modeled by the NTI data control region. The $t\bar{t}H$ signal is required to contain at least one jet that is b-tagged at the 77% working point: 60% of this sample is used for training, another 20% for validation, and the remaining 20% for estimating the final sensitivity. The NTI data sample with ≥ 1 lepton contains limited statistics, and four statistically independent datasets are defined: one for training, one for BDT hyper-parameter optimization (validation 1), one for category boundary optimization (validation 2), and one for testing.

In the training and validation 1 samples, events are required to contain zero b-tagged jets and at least one un-tagged jet in the central η region, effectively modeling the background from non-top QCD processes. In addition, the cuts on photon $p_T/m_{\gamma\gamma}$ are replaced with flat cuts of $p_T > 35$ (25) GeV on the leading (sub-leading) photon, and the range of diphoton invariant masses is extended to $m_{\gamma\gamma} > 80$ GeV. 75% of this sample is used for training, and the other 25% for optimization of the BDT hyper-parameters.

In the validation 2 and testing samples, the same selection is applied that is used for the signal sample (≥ 1 b-jet, ≥ 1 lepton): 50% of this sample is used to optimize the BDT category boundaries, and the other 50% is reserved for estimating the final sensitivity.

The variables used to train the leptonic BDT are:

- $p_T/m_{\gamma\gamma}$, η, and ϕ of the two photons. Photon p_T is scaled by $m_{\gamma\gamma}$ to reduce sculpting of the diphoton mass spectrum.
- p_T, η, and ϕ of up to two leptons. If there is no second lepton, dummy values are given to the BDT in the place of second lepton kinematics.
- p_T, η, ϕ, and E of the four jets with highest p_T. If there is no jet n, dummy values are given to the BDT in the place of jet n kinematics.
- Boolean b-tag flag (77% working point) for each of the four jets with highest p_T. If there is no jet n, a dummy value is given to the BDT in the place of a b-tag flag for jet n.
- Missing E_T and ϕ direction of missing E_T.

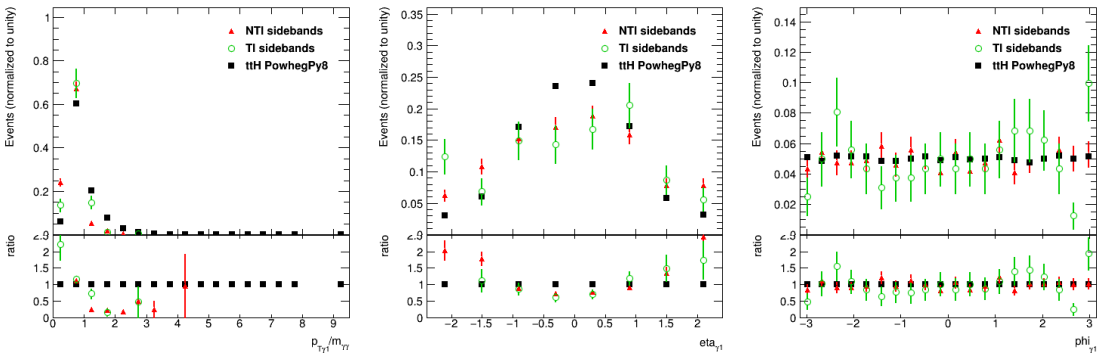
Figures 9.7-9.10 show the signal and background distributions in each of these training variables. The resulting leptonic BDT discriminant is shown in Figure 9.11. The peak in the NTI background at high BDT score is composed primarily of $t\bar{t}\gamma\gamma$ events, which are not well rejected by the leptonic BDT (see Section 9.3.2).

Three cuts are placed on the leptonic BDT discriminant to create four orthogonal regions: events in the region with lowest BDT score are rejected, and the remaining regions make up the three leptonic analysis categories. The cuts forming the category boundaries are chosen to maximize the expected Standard Model $t\bar{t}H$ number-counting significance evaluated on the validation 2 set. This significance is given by Equation 9.3 in each category, and the total significance is given by the sum in quadrature over all categories.

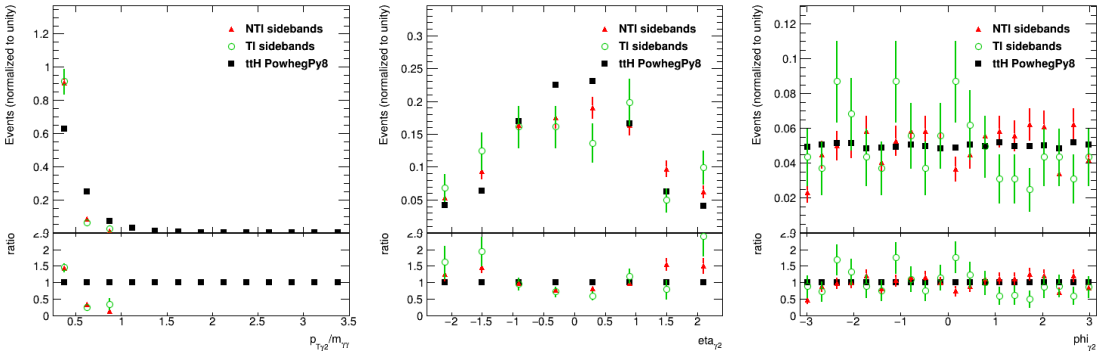
The category boundaries are determined from a brute force scan of all possible boundary combinations. In order to have sufficient data in each category to perform the maximum likelihood fit, a minimum of 0.8 continuum background events (predicted from NTI data) is required in each category in the range $123 < m_{\gamma\gamma} < 127$ GeV. This corresponds to ~ 10 events in the full fit range of 105-160 GeV.

The optimal boundaries are shown in Table 9.4. 97% of the $t\bar{t}H$ signal events passing leptonic preselection (≥ 1 b-jet, ≥ 1 lepton) are selected into these four categories, and 43% of the preselected continuum background events are rejected.

The predicted Higgs boson yield in each category is shown in Table 9.5, along with the estimated continuum background derived by scaling the NTI sidebands. These yields are calculated in the region $123 < m_{\gamma\gamma} < 127$ GeV. The expected number-counting significances of the total Higgs boson signal (Z_H), the $t\bar{t}H$ signal ($Z_{t\bar{t}H}$), and the single top plus Higgs signal (Z_{tH}) are shown in Table 9.6.



(a) Leading photon $p_T/m_{\gamma\gamma}$ (b) Leading photon η (c) Leading photon ϕ

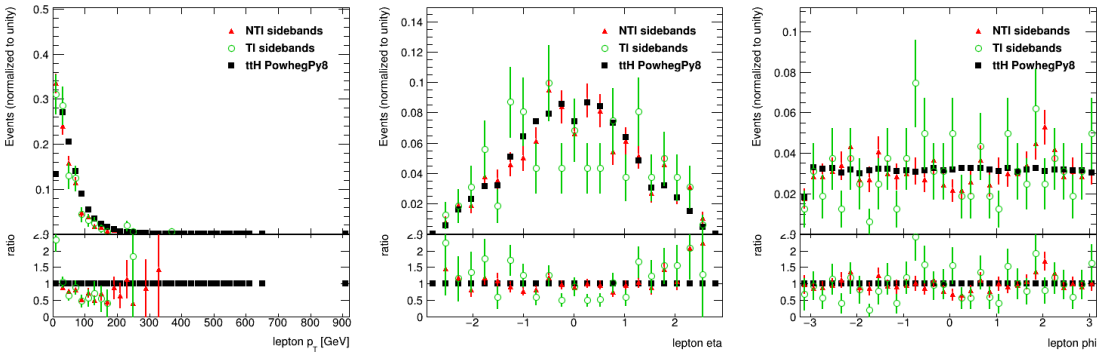


(d) Sub-leading photon $p_T/m_{\gamma\gamma}$ (e) Sub-leading photon η (f) Sub-leading photon ϕ

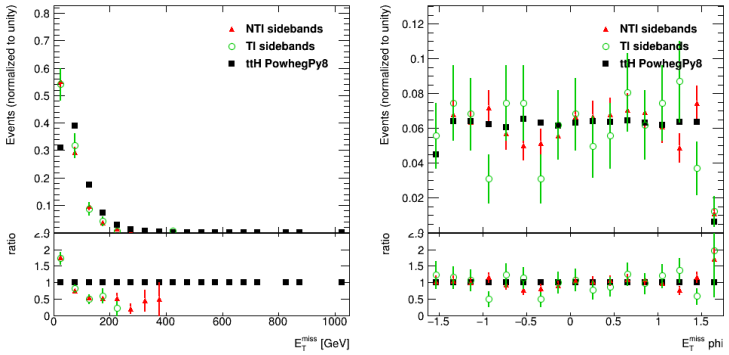
Figure 9.7: Leptonic BDT training variables: $p_T/m_{\gamma\gamma}$, η, and ϕ of the leading and sub-leading photons. Black squares show $t\bar{t}H$ Monte Carlo signal, and red and green points represent the NTI and TI data sidebands, respectively.

Category	BDT range
$t\bar{t}H$ lep 3	[0.705-0.942]
$t\bar{t}H$ lep 2	[0.942-0.987]
$t\bar{t}H$ lep 1	[0.987-1]

Table 9.4: Ranges in leptonic BDT score defining the three leptonic $t\bar{t}H(\gamma\gamma)$ categories.



(a) Lepton p_T (b) Lepton η (c) Lepton ϕ

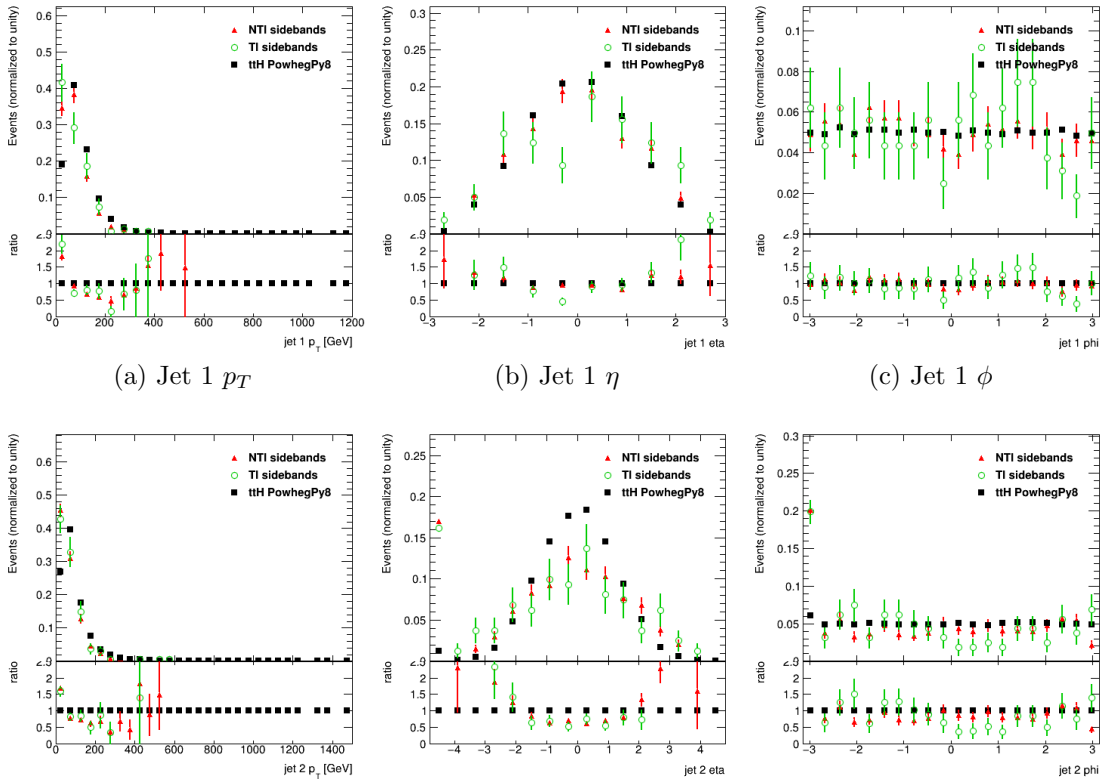


(d) Missing E_T (e) ϕ of missing E_T

Figure 9.8: Leptonic BDT training variables: (top) p_T, η, and ϕ of the leading lepton, and (bottom) missing E_T and ϕ direction of missing E_T. Black squares show $t\bar{t}H$ Monte Carlo signal, and red and green points represent the NTI and TI data sidebands, respectively.

	$t\bar{t}H$	$tHjb$	tWH	ggF	VBF	WH	ZH	cont. bkg.
$t\bar{t}H(\gamma\gamma)$ lep 3	0.66	0.0634	0.022	0.028	0.00102	0.0265	0.0189	2.84
$t\bar{t}H(\gamma\gamma)$ lep 2	1.85	0.0556	0.052	0	0	0.0226	0.013	2.35
$t\bar{t}H(\gamma\gamma)$ lep 1	3.98	0.0307	0.123	0.00587	0	0.00956	0.00994	1.67

Table 9.5: Predicted event yield in the region 123 GeV $< m_{\gamma\gamma} <$ 127 GeV at 79.8 fb^{-1} in the leptonic BDT categories. The continuum background estimate is derived by scaling the data yield in the NTI data sidebands.



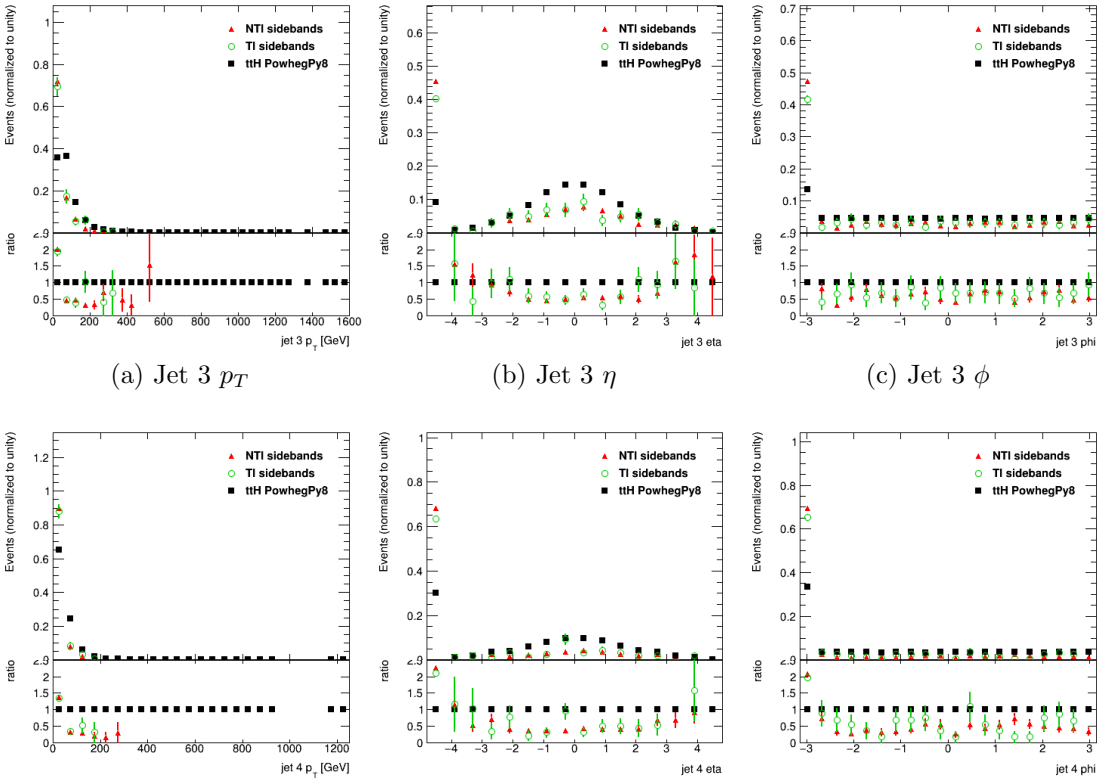
(a) Jet 1 p_T (b) Jet 1 η (c) Jet 1 ϕ

(d) Jet 2 p_T (e) Jet 2 η (f) Jet 2 ϕ

Figure 9.9: Leptonic BDT training variables: p_T, η, and ϕ of the two jets with highest p_T. Black squares show $t\bar{t}H$ Monte Carlo signal, and red and green points represent the NTI and TI data sidebands, respectively. Events with no jet n are added to the underflow (left-most) bin for the kinematics of jet n.

	Z_H	$Z_{t\bar{t}H}$	Z_{tH}
$t\bar{t}H(\gamma\gamma)$ lep 3	0.465	0.368	0.045
$t\bar{t}H(\gamma\gamma)$ lep 2	1.16	1.06	0.0521
$t\bar{t}H(\gamma\gamma)$ lep 1	2.5	2.33	0.0643
Combined	2.8	2.58	0.0942

Table 9.6: Predicted number-counting significance at 79.8 fb^{-1} in the leptonic BDT categories. The continuum background estimate is derived by scaling the data yield in the NTI data sidebands.



(a) Jet 3 p_T (b) Jet 3 η (c) Jet 3 ϕ

(d) Jet 4 p_T (e) Jet 4 η (f) Jet 4 ϕ

Figure 9.10: Leptonic BDT training variables: p_T, η, and ϕ of the jets ranked third and fourth in p_T. Black squares show $t\bar{t}H$ Monte Carlo signal, and red and green points represent the NTI and TI data sidebands, respectively. Events with no jet n are added to the underflow (left-most) bin for the kinematics of jet n.

Many combinations of leptonic BDT training variables were tested before settling on the list above, including:

- High-level jet variables (HL): $n_{\text{cen. jets}}$, missing E_T, H_T, transverse mass m_T, p_T(lepton + missing E_T)
- Photons: $p_T/m_{\gamma\gamma}$ and η for the two photons
- Jets: p_T, η, ϕ and b-tag flag for up to four jets
- Leps: p_T, η and ϕ of up to two leptons
- Tops: Top Reco BDT score, p_T, η, and m for the reconstructed top (t_1), p_T, η, and m of the system of jets recoiling against the top (t_2)

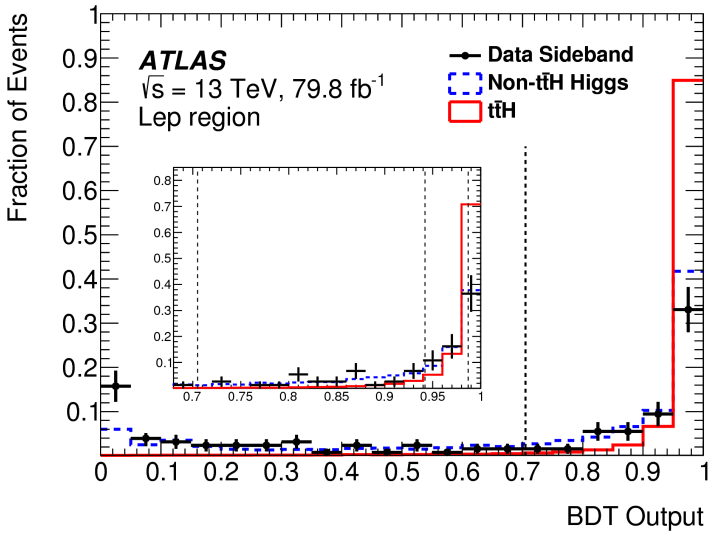


Figure 9.11: Distribution of leptonic BDT score for $t\bar{t}H$ Monte Carlo signal (red), non-$t\bar{t}H$ Higgs processes (blue), and data sidebands (black) with 79.8 fb^{-1} of ATLAS data. Signal (red) and background (blue, black) processes peak at opposite ends of the BDT spectrum. Dashed vertical lines indicate the position of category boundaries, which are most visible in the inlaid plot that zooms in to the signal-like end of the spectrum.

Figure 9.12 compares the performance of three combinations of these variables in terms of number-counting significance. Once again, replacing the low-level jet and lepton kinematics with reconstructed top variables gives very similar performance (compare "Tops, photons" and "Jets, leps, MET, photons").

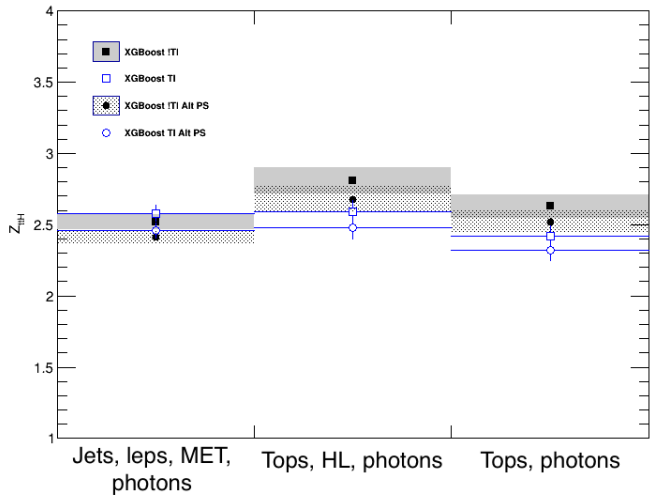


Figure 9.12: Number-counting significance of three leptonic categories, with different input variables used in BDT training. Black (blue) points represent significance calculations with continuum background estimated from the NTI (TI) data sidebands. Square points show significances that use the $t\bar{t}H$ yield from the nominal Monte Carlo sample, and circles show significances that use the yield from a sample using an alternative parton showering algorithm. The errors shown are statistical.

9.3 Decomposition of continuum background

The continuum diphoton background included in the $t\bar{t}H$ analysis categories consists mainly of two processes: $\gamma\gamma$ + jets and $t\bar{t}\gamma\gamma$, with photons that can be either real or fake. The total continuum background is estimated using the data-driven method already described, but the relative contribution from each of these sources in the $t\bar{t}H$ categories has so far remained unexplored. This Section introduces several methods for decomposing the continuum background. Monte Carlo samples containing real photons are used to model the shapes of both real-photon and fake-photon background processes.

9.3.1 Hadronic Channel

In the hadronic categories, the relative contribution of $t\bar{t}\gamma\gamma$ and $\gamma\gamma$ + jets is determined using a template fit method. A probability distribution function (PDF) f is constructed for each background process as a function of an event variable x. The PDFs are summed with a relative weight factor r, giving a combined PDF of

$$r \times f_{t\bar{t}\gamma\gamma}(x) + (1-r) \times f_{\gamma\gamma}(x) + \frac{n_{t\bar{t}H}^{SM}}{n_{\text{data}}} \times f_{t\bar{t}H}(x) \tag{9.8}$$

The final term in this PDF represents the expected signal: $n_{t\bar{t}H}^{SM}$ is the number of SM signal events, n_{data} is the observed number of data events, and $f_{t\bar{t}H}(x)$ is the signal PDF. Equation 9.8 is fitted to data, and the best fit value of r gives the fraction of $t\bar{t}\gamma\gamma$ making up the total continuum background: this is referred to as the **top background fraction**.

The template fit method is most successful when the background PDFs $f_{t\bar{t}\gamma\gamma}$ and $f_{\gamma\gamma}$ are very different in shape. The variable x is chosen to be the reconstructed top candidate mass in order to exploit the shape difference between samples with and without true top quark (see Chapter 8 for details on the top reconstruction). Figure 9.13 shows the templates $f_{t\bar{t}\gamma\gamma}$, $f_{\gamma\gamma}$, and $f_{t\bar{t}H}$, which use a bin size of 20 GeV. The statistical uncertainty on the Monte Carlo templates is eclipsed by the size of the marker points.

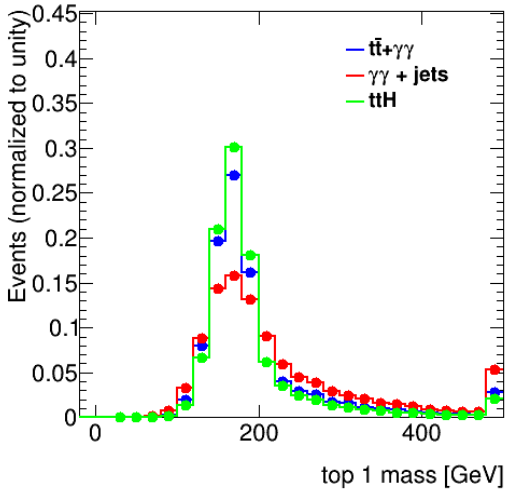
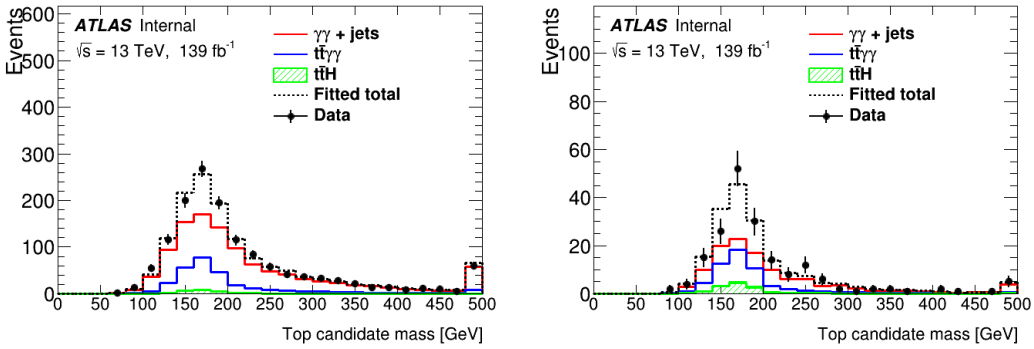


Figure 9.13: The templates, normalized to unit area, of the top mass distribution in the $t\bar{t}H$, $t\bar{t}\gamma\gamma$ and $\gamma\gamma$ + jets Monte Carlo samples.

The template fits are shown in Figure 9.14 in two regions: one region corresponding to all four hadronic $t\bar{t}H(\gamma\gamma)$ categories, and another region corresponding to the two hadronic categories with highest signal-to-background ratio (S/B).

The top background fractions obtained in each region are summarized in Table 9.7. The uncertainty on the top background fraction is divided into two components: a contribution from the statistics of the data, and a contribution from the statistics of the templates. The latter is determined by running 1000 toys per template fit and randomly varying the templates bin-by-bin. The variations are subject to a Gaussian constraint with a width corresponding to the 1σ statistical error on the bin. The standard deviation of the top background fractions from toys is taken as the error due to template statistics. The statistical uncertainty on the templates is small compared to the statistical uncertainty in data.

Using the top background fraction extracted from data, the number of $t\bar{t}\gamma\gamma$ and $\gamma\gamma$+jets



(a) $t\bar{t}H$ had 1-4 (b) $t\bar{t}H$ had 1-2

Figure 9.14: Template fits to data in the hadronic $t\bar{t}H(\gamma\gamma)$ categories. The contributions from $t\bar{t}\gamma\gamma$ (blue) and $\gamma\gamma$+ jets (red) are normalized to their fitted values of r and $(1-r)$. The $t\bar{t}H$ contribution (green) is fixed to the SM expectation. The fitted combined PDF is shown in the dashed black line. The left plot shows the fit for all four hadronic BDT categories, and the right shows only the two categories with highest S/B.

	$n^{SM}_{t\bar{t}H}/n_{\text{data}}$	r	data	temp
$t\bar{t}H$ had 1-4	2%	21%	±6%	±2%
$t\bar{t}H$ had 1-2	7%	31%	±16%	±5%

Table 9.7: The top background fraction obtained in the hadronic $t\bar{t}H(\gamma\gamma)$ categories. The uncertainty is split into components coming from data statistics (data) and from the template statistics (temp).

events in each region can be extracted:

$$n_{t\bar{t}\gamma\gamma} = r n_{\text{data}} \tag{9.9}$$
$$n_{\gamma\gamma} = (1-r) n_{\text{data}} \tag{9.10}$$
$$\tag{9.11}$$

The ratio of this number of events to the prediction from Monte Carlo gives a data/Monte Carlo scale factor for each process. These scale factors are shown in Table 9.8.

	$k_{t\bar{t}\gamma\gamma}$	data	temp	$k_{\gamma\gamma}$	data	temp
$t\bar{t}H(\gamma\gamma)$ had 1-4	4.87	±1.50	±0.44	1.00	±0.08	±0.02
$t\bar{t}H(\gamma\gamma)$ had 1-2	3.31	±1.67	±0.58	1.21	±0.31	±0.11

Table 9.8: Data/Monte Carlo scale factors determined from the top background fraction for the hadronic $t\bar{t}H(\gamma\gamma)$ categories. The uncertainty is split into components coming from data statistics (data) and from the template statistics (temp).

The same procedure can be applied in the data sideband region only, with $n_{t\bar{t}H}^{SM} = 0$. The top background fractions and data/Monte Carlo scale factors obtained from the sidebands agree with those reported in Tables 9.7-9.8 within the statistical uncertainty.

9.3.2 Leptonic Channel

In the leptonic channel, the dilepton events provide a simple handle for background decomposition. Since it is highly unlikely that $\gamma\gamma$ + jets events contain more than one lepton, all dilepton events are taken to be the result of $t\bar{t}$ decays. In the data sidebands ($|m_{\gamma\gamma} - 125$ GeV$| > 5$ GeV), no contamination from $t\bar{t}H(\gamma\gamma)$ is expected, and all dilepton events are therefore attributed to $t\bar{t}\gamma\gamma$.

The number of dilepton events predicted by $t\bar{t}\gamma\gamma$ Monte Carlo is scaled to the dilepton event yield in the data sidebands (eight events at 139 fb^{-1}). This corresponds to a data/MC scale factor of 2.91 ± 1.3, with the uncertainty coming from the statistical error on the number dilepton events in data. This same scale factor is applied to the predicted yield of $t\bar{t}\gamma\gamma$ events containing ≥ 1 lepton, and the scaled prediction is subtracted from the number of ≥ 1 lepton events in data. The remaining events are attributed to $\gamma\gamma$ + jets. The data/MC scale factor for $\gamma\gamma$ + jets is extracted by solving:

$$n_{\text{data}}^{SB} = k_{t\bar{t}\gamma\gamma} n_{t\bar{t}\gamma\gamma}^{SB} + k_{\gamma\gamma} n_{\gamma\gamma}^{SB} \tag{9.12}$$

Tables 9.9 and 9.10 show the values of each parameter in this formula.

	Data yield	Predicted $t\bar{t}\gamma\gamma$ yield	Predicted $\gamma\gamma$ + jets yield
Dilepton	8	2.75	0
≥ 1 leptons	118	29.42	3.92

Table 9.9: The dilepton and inclusive ≥ 1 lepton event yields in the three leptonic $t\bar{t}H(\gamma\gamma)$ categories at 139 fb^{-1}, counted in the $m_{\gamma\gamma}$ sidebands. Shown in data, $t\bar{t}\gamma\gamma$ Monte Carlo, and $\gamma\gamma$ + jets Monte Carlo.

	$k_{t\bar{t}\gamma\gamma}$	$k_{\gamma\gamma}$
Dilepton	2.91 ± 1.3	n/a
≥ 1 leptons	2.91 ± 1.3	8.25 ± 7.22

Table 9.10: Data/MC scale factors derived for each continuum background process in the leptonic channel. The reported statistical error is due to the limited number of dilepton events.

Applying the $t\bar{t}\gamma\gamma$ Monte Carlo scale factor to data in the full $m_{\gamma\gamma}$ range and including the Standard Model prediction for $t\bar{t}H$ gives the fractions listed in Table 9.11. As expected, the $t\bar{t}\gamma\gamma$ process is the dominant background in these categories, since $\gamma\gamma$+ jets is better rejected by a leptonic BDT that is trained on 0 b-jet background (refer to Section 9.2.2).

	$n^{SM}_{t\bar{t}H}/n_{\text{data}}$	r
$t\bar{t}H(\gamma\gamma)$ lep 1-3	9%	$66\% \pm 23\%$

Table 9.11: The top background fraction obtained in the leptonic $t\bar{t}H(\gamma\gamma)$ categories.

Chapter 10

First observation of $t\bar{t}H$ Production

The observation of $t\bar{t}H$ production at the LHC represents the first direct evidence of Higgs boson coupling to quarks. The first observation of $t\bar{t}H$ above the 5σ threshold for discovery combined searches in multiple Higgs boson decay channels, including $H \to \gamma\gamma$ [147] [148]. The 2018 ATLAS $t\bar{t}H$ combination is discussed in Section 10.1. Section 10.2 describes the $t\bar{t}H(\gamma\gamma)$ results from ATLAS using the full Run 2 dataset, which gives a single-channel observed $t\bar{t}H$ significance of 4.9σ (4.5σ expected) [149].

10.1 Combination of Higgs decay channels

The full ATLAS $t\bar{t}H$ program includes searches in multiple Higgs boson decay channels. A statistical combination of all available $t\bar{t}H$ analyses was performed 2018 using a partial Run 2 dataset. The analyses included in this combination are summarized in Section 10.1.1. Section 10.1.2 includes a discussion of the treatment of systematic uncertainties, and the combined results are included in Section 10.1.3.

10.1.1 Input analyses

Run 2 $t\bar{t}H(b\bar{b})$ with 36.1 fb^{-1}

The $t\bar{t}H(b\bar{b})$ analysis [150] targets the $H \to b\bar{b}$ decay in conjunction with at least one semi-leptonic top quark decay. The analysis included in the combination uses 36.1 fb^{-1} of ATLAS data collected in 2015-2016. Events are divided according to lepton multiplicity and b-jet multiplicity at different working points. Within each region, categories are defined using a variety of multivariate techniques. Categories containing high fractions of $t\bar{t}H$ are designated as signal regions. Control regions are used to constrain the rate of $t\bar{t}$ background processes, including $t\bar{t}+b\bar{b}$, $t\bar{t}+b$, $t\bar{t}+$ charm, and $t\bar{t}+$ light jets.

In the dilepton channel, four control regions and three signal regions are defined. In the single-lepton channel, six control regions and six signal regions are defined, including one signal region that targets boosted (high p_T) Higgs bosons and top quarks using large-radius

($R = 1.0$) jets. The $t\bar{t}H$ signal strength is determined by fitting all 19 categories. The rate of each background process is left free in the fit, with constraining power coming from the control regions.

Run 2 $t\bar{t}H$(multi-lepton) with 36.1 fb^{-1}

The $t\bar{t}H$(multi-lepton) analysis [143] targets $H \to WW$, $H \to \tau\tau$, and non-resonant $H \to ZZ$ decays. The analysis included in the combination uses 36.1 fb^{-1} of ATLAS data collected in 2015-2016. Eight signal regions are defined based on lepton flavor and multiplicity, with two categories corresponding to the 4ℓ final state. In addition, four control regions are defined to target the $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$, and diboson backgrounds.

The $t\bar{t}H$ signal strength is determined by fitting all twelve categories. The rate of each background process is left free in the fit, with constraining power coming from the control regions.

Run 2 $t\bar{t}H(\gamma\gamma)$ with 79.8 fb^{-1}

The $t\bar{t}H(\gamma\gamma)$ analysis targets the $H \to \gamma\gamma$ decay using the seven analysis categories described in Chapter 9. The analysis included in this $t\bar{t}H$ combination uses 79.8 fb^{-1} of 13 TeV ATLAS data collected between 2015 and 2017 (results included in Appendix E). The same analysis procedure is later applied to the full Run 2 dataset (139 fb^{-1}), and these results are reported in Section 10.2.

Run 2 $t\bar{t}H(ZZ \to 4\ell)$ with 79.8 fb^{-1}

The $t\bar{t}H(ZZ \to 4\ell)$ analysis [147] targets resonant $H \to ZZ$ decay with both Z bosons decaying leptonically to e or μ. This analysis uses 79.8 fb^{-1} of 13 TeV ATLAS data collected between 2015 and 2017. Events are selected that contain ≥ 1 b-jet and ≥ 4 leptons (two same-flavor/opposite-sign pairs) with 115 GeV $< m_{4\ell} <$ 130 GeV. Events are then sorted into two groups: had-enriched (≥ 4 jets and exactly 4 leptons), and lep-enriched (≥ 2 jets and ≥ 5 leptons). The had-enriched sample is further divided into one "Tight" category and one "Loose" category based on a BDT discriminant. This BDT is trained to separate $t\bar{t}H(ZZ \to 4\ell)$ Monte Carlo events from simulated background using the following training variables: p_T^{jj}, m^{jj},$\Delta\eta^{jj}$, $\Delta R_{\min}(j, Z)$, $\Delta R(j_1, 4\ell)$, η^{ZZ}_{Zepp}, E_T^{miss}, n_{jet}, $n_{b\text{-jet}}$, H_T, and $H \to 4\ell$ matrix element.

Run 1 $t\bar{t}H$

The Run 1 ATLAS search for $t\bar{t}H$ production [151] combined three Higgs boson decay channels: $H(b\bar{b})$ [152], H(multi-lepton) [153], and $H(\gamma\gamma)$ [151]. With 4.5 fb^{-1} at $\sqrt{s} = 7$ TeV and 20.3 fb^{-1} at $\sqrt{s} = 8$ TeV, the Run 1 observed $t\bar{t}H$ significance is 2.5σ, compared to 1.5σ expected. This corresponds to an observed signal strength of $\mu_{t\bar{t}H} = 1.82^{+0.83}_{-0.75}$.

10.1.2 Systematic uncertainties

The Run 2 $t\bar{t}H(b\bar{b})$ and $t\bar{t}H$(multi-lepton) analyses use data from 2015-2016. The experimental uncertainties on physics objects (jets, leptons, photons, etc.) are mostly correlated between these two analyses. The modeling uncertainty on $t\bar{t}V$, tV, 4-top, and diboson backgrounds are treated as fully correlated. The modeling uncertainty on $t\bar{t}$ in the $t\bar{t}H(b\bar{b})$ analysis is not correlated with $t\bar{t}H$(multi-lepton) because the two analyses probe different regions of phase-space. See [143] for full discussion of correlations.

The Run 2 $t\bar{t}H(ZZ \to 4\ell)$ and $t\bar{t}H(\gamma\gamma)$ analyses use data from 2015-2017 and a later version of ATLAS analysis software than the 2015-2016 analyses. The experimental uncertainties on physics objects are treated as fully correlated between these two analyses, but mostly uncorrelated with $t\bar{t}H(b\bar{b})$ and $t\bar{t}H$(multi-lepton).

The main experimental systematics that are correlated in the combination of these four analyses are uncertainties on the electron/photon energy scale and resolution, and on the electron efficiencies. The uncertainty on overall luminosity is treated as partially correlated between the 2015-2016 and 2015-2017 datasets: a 1.51% common uncertainty is assigned to both periods, and an additional 1.61% (1.27%) is applied for the 2015-2016 (2015-2017) period.

Theory uncertainties, including uncertainties on the $t\bar{t}H$ signal prediction, are fully correlated across all four 13 TeV analyses. The theory systematics are dominated by the Monte Carlo modeling of $t\bar{t}$ background processes and uncertainties on the $t\bar{t}H$ signal (including QCD scale and UEPS uncertainties).

The dominant sources of systematic uncertainty on the 13 TeV cross section measurement are summarized in Table 10.1. To determine the contribution from each group of systematics, the corresponding nuisance parameters are fixed to their best fit values from the unconditional maximum likelihood fit. The resulting error is then subtracted (in quadrature) from the total error to obtain the error due only to the systematics in question.

In the combination of Run 1 and Run 2 $t\bar{t}H$ results, almost all systematics are treated as uncorrelated. The only exceptions are the QCD scale uncertainty on $t\bar{t}H$, uncertainties on the Higgs boson branching fractions, and uncertainties on the electron/photon energy scale and resolution.

10.1.3 Results of combined analysis

Run 2

A summary of all $\sqrt{s} = 13$ TeV analysis categories (including control regions) is shown in Figure 10.1. The $t\bar{t}H$ signal is clearly visible above the background in the categories with high signal-to-background ratio (on the right-hand side of the plot). The combined $t\bar{t}H$ significance is observed to be 5.8σ in the 13 TeV analyses, corresponding to a signal strength of $\mu_{t\bar{t}H} = 1.32\ ^{+0.32}_{-0.28}$. The expected significance is 4.9σ.

Uncertainty source	$\Delta\sigma/\sigma$ [%]
Theory uncertainties	11.9
$t\bar{t}$ + heavy flavor	9.9
$t\bar{t}H$	6.0
Non-$t\bar{t}H$ Higgs production	1.5
Other backgrounds	2.2
Experimental uncertainties	9.3
Fake leptons	5.2
Jets and missing E_T	4.9
Electrons and photons	3.2
Luminosity	3.0
Tau leptons	2.5
Flavor tagging	1.8
Monte Carlo statistics	4.4

Table 10.1: Breakdown of the systematic uncertainties for the combined $t\bar{t}H$ cross section measurement at $\sqrt{s} = 13$ TeV. This is shown as the uncertainty due to each group of systematics ($\Delta\sigma$), as a fraction of the total cross section (σ). The numbers are obtained from quadratic subtraction of the total uncertainty and the uncertainty of the fit with the corresponding group fixed.

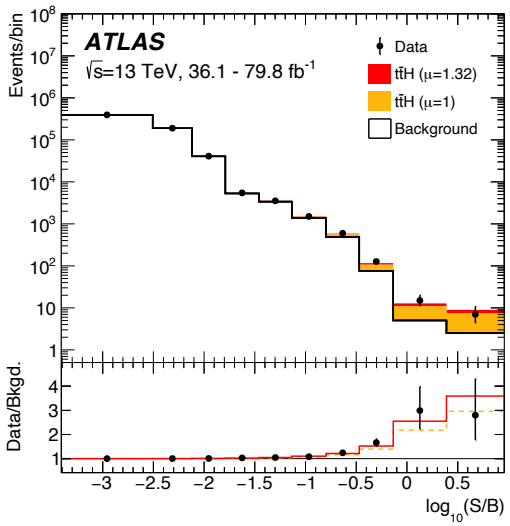
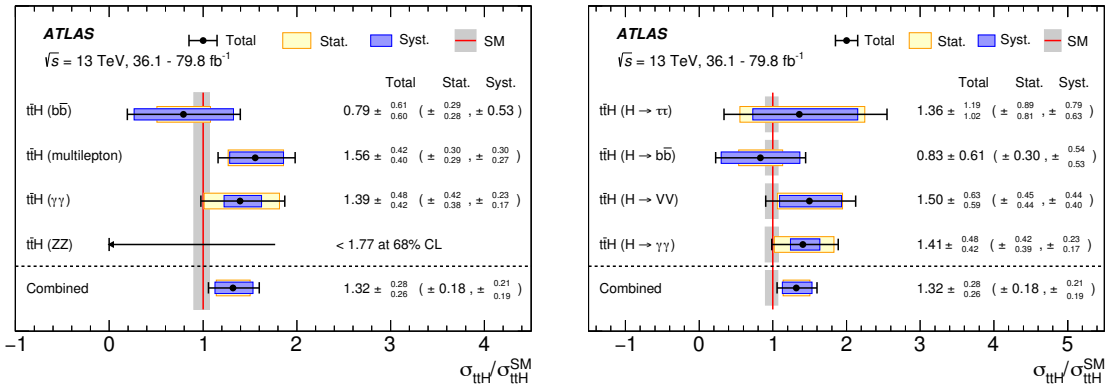


Figure 10.1: Summary of all $t\bar{t}H$ analysis categories, binned in the logarithm of the signal-to-background ratio (S/B) of the category. The $t\bar{t}H$ signal is clearly visible above the background in the most powerful categories (those with highest S/B).

The combined $t\bar{t}H$ cross section at $\sqrt{s} = 13$ TeV is measured to be

$$\sigma_{t\bar{t}H} = 670 \pm 90 \text{ (stat.) } {}^{+110}_{-100} \text{ (syst.) fb} \tag{10.1}$$

This is in agreement with the Standard Model prediction of $507\ {}^{+35}_{-50}$ fb. The measured cross section from each constituent analysis is shown in Figure 10.2a. No events are observed in the $t\bar{t}H \to ZZ \to 4\ell$ categories, so an upper limit is placed on the cross section in this channel. The measured $t\bar{t}H$ cross section in each Higgs boson decay channel is shown in Figure 10.2b. The $H \to b\bar{b}$ and $H \to \gamma\gamma$ decay channel cross sections differ slightly from the analysis-level cross sections due to some small contamination of other Higgs boson decays in these categories.



(a) Per analysis

(b) Per decay mode

Figure 10.2: Measured $t\bar{t}H$ cross section broken down by (a) ATLAS analysis and (b) Higgs boson decay mode, normalized to the Standard Model prediction. The yellow (blue) bars represent the statistical (systematic) component of the uncertainty on the measured cross section. The Standard Model prediction is shown in red, with theory errors in gray.

Run 1 and Run 2

Combining the four Run 2 $t\bar{t}H$ analyses with the ATLAS Run 1 analysis gives an observed $t\bar{t}H$ significance of 6.3σ, compared to 5.1σ expected. The combined signal strength is $\mu_{t\bar{t}H} = 1.38\ {}^{+0.31}_{-0.27}$. Figure 10.3 shows the likelihood scans of the Run 1, Run 2, and combined datasets.

The $t\bar{t}H$ cross section is shown as a function of center of mass energy in Figure 10.4. The data correspond to Run 1 ($\sqrt{s} = 8$ TeV only) and Run 2 (Equation 10.1) cross section measurements. The pink line represents the theory calculation at NLO in QCD + NLO in EW, with the uncertainty shown as a purple band.

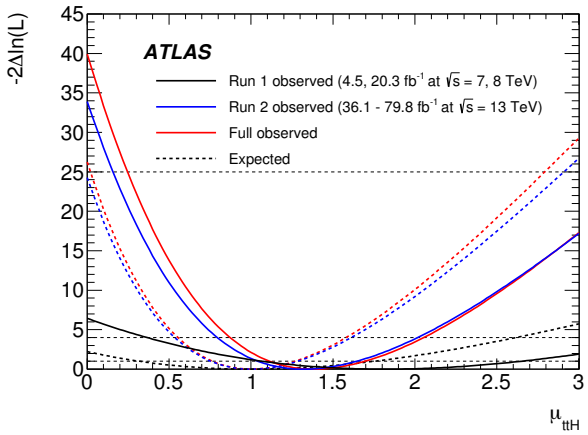


Figure 10.3: Expected (dashed) and observed (solid) likelihood scans of the signal strength $\mu_{t\bar{t}H}$ for the Run 1 (black), Run 2 (blue), and combined datasets (red).

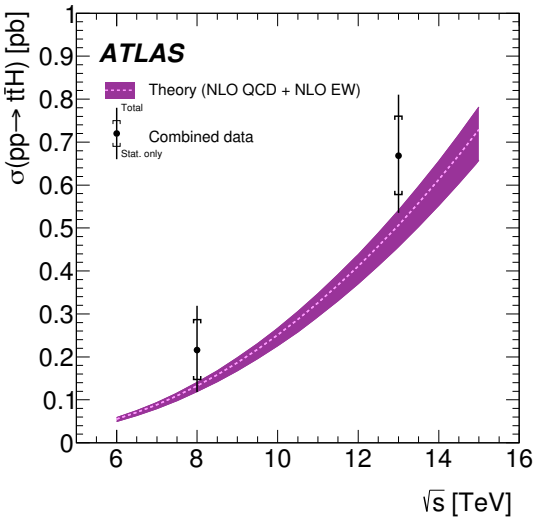


Figure 10.4: Measured $t\bar{t}H$ cross section as a function of proton–proton center of mass energy $\sqrt{s}$. Data is shown in black with total and statistical only uncertainties denoted. The pink line represents the theory calculation at NLO in QCD + NLO in EW, with the uncertainty shown as a purple band.

10.2 $t\bar{t}H$ in the diphoton decay channel

The $t\bar{t}H(\gamma\gamma)$ categorization described in Chapter 9 was developed and optimized on the partial LHC Run 2 dataset of 79.8 fb^{-1}, collected during 2015-2017. This same categorization is applied to the full Run 2 dataset of 139 fb^{-1}, and no re-optimization of the categories is performed. This Section reports the full Run 2 results, which are included in [149]. Earlier results at 79.8 fb^{-1} are described in Appendix E.

10.2.1 Signal and background shapes

In each analysis category, the signal shape is determined by fitting a Double-Sided Crystal Ball (DSCB) function to a sample of inclusive Higgs boson Monte Carlo (see Section 6.2.1). The best fit DSCB parameters for each $t\bar{t}H(\gamma\gamma)$ category are shown in Table 10.2, along with S_{90}, the smallest $m_{\gamma\gamma}$ range containing 90% of Higgs signal.

Category	μ_{CB} [GeV]	σ_{CB} [GeV]	α_{low}	α_{high}	n_{low}	n_{high}	S_{90} [GeV]
$t\bar{t}H(\gamma\gamma)$ had 4	125.15	1.55	1.7	1.5	5.0	13.3	3.00
$t\bar{t}H(\gamma\gamma)$ had 3	125.16	1.52	1.7	1.5	5.0	12.8	2.96
$t\bar{t}H(\gamma\gamma)$ had 2	125.14	1.46	1.7	1.4	4.6	16.6	2.84
$t\bar{t}H(\gamma\gamma)$ had 1	125.14	1.29	1.7	1.5	4.7	16.6	2.48
$t\bar{t}H(\gamma\gamma)$ lep 3	125.16	1.74	1.6	1.7	6.1	5.8	3.30
$t\bar{t}H(\gamma\gamma)$ lep 2	125.13	1.62	1.7	1.4	4.5	21.6	3.13
$t\bar{t}H(\gamma\gamma)$ lep 1	125.12	1.44	1.7	1.4	4.9	13.5	2.80

Table 10.2: Best fit parameter values for the signal $m_{\gamma\gamma}$ shape in each of the seven $t\bar{t}H(\gamma\gamma)$ categories at 139 fb^{-1}. Note that these are the fitted values for the MC samples at $m_H = 125$ GeV.

For the hadronic $t\bar{t}H(\gamma\gamma)$ categories, the functional form of the continuum background is extracted from templates created from NTI data. To enhance statistics in these templates, events with no b-tagged jets are also included, and the b-tag status of all central jets are set to "true" when computing BDT score. For the leptonic $t\bar{t}H(\gamma\gamma)$ categories, the background templates are constructed from $t\bar{t}\gamma\gamma$ Monte Carlo, since this process dominates the continuum background once the BDT selection has been applied. In order to enhance statistics, no photon identification or isolation cuts are applied to the leptonic background templates.

The functional form of the continuum background and the spurious signal are listed in Table 10.3.

10.2.2 Systematic uncertainties

The experimental uncertainties included in the $t\bar{t}H(\gamma\gamma)$ measurement are those summarized in Section 6.4.2. A conservative 100% heavy flavor uncertainty is applied to the

Category	Function	N_{sp}	χ^2/ndof	Prob(χ^2) [%]
$t\bar{t}H(\gamma\gamma)$ had 4	Power Law	+1.07	0.82	82.2
$t\bar{t}H(\gamma\gamma)$ had 3	Power Law	+0.658	1.14	22.2
$t\bar{t}H(\gamma\gamma)$ had 2	Power Law	+0.756	1.31	6.47
$t\bar{t}H(\gamma\gamma)$ had 1	Exponential	+0.316	1.33	5.45
$t\bar{t}H(\gamma\gamma)$ lep 3	Exponential	+0.165	0.87	73.8
$t\bar{t}H(\gamma\gamma)$ lep 2	Power Law	+0.430	1.33	5.45
$t\bar{t}H(\gamma\gamma)$ lep 1	Exponential	+0.241	1.51	0.981

Table 10.3: Results of the spurious signal test for the $t\bar{t}H(\gamma\gamma)$ categories at 139 fb^{-1}. The χ^2/ndof indicates the performance of a background only fit to the template in the $m_{\gamma\gamma}$ range 105-160 GeV, but is in general not a part of the spurious signal criteria.

predicted ggF, VBF, and VH yields in each category, and inclusive uncertainties from [34] are applied to the $b\bar{b}H$, $tHjb$, and tWH yields.

For the $t\bar{t}H$ signal yields, QCD scale variations and PDF uncertainties (31 eigenvector scheme) are evaluated separately in each category. In addition, an Underlying Event/Parton Showering (UEPS) uncertainty is calculated on the $t\bar{t}H$ yield in each category by comparing predictions from Monte Carlo samples using different showering algorithms: MG5_aMC@NLO+Pythia8 and MG5_aMC@NLO+Herwig++. The UEPS uncertainties are listed in Table 10.4.

Category	UEPS [%]
$t\bar{t}H$ had 4	+2.2
$t\bar{t}H$ had 3	−3.2
$t\bar{t}H$ had 2	−6.7
$t\bar{t}H$ had 1	−11.4
$t\bar{t}H$ lep 3	+5.1
$t\bar{t}H$ lep 2	−2.9
$t\bar{t}H$ lep 1	−9.0

Table 10.4: Underlying Event/Parton Showering (UEPS) uncertainty on $t\bar{t}H$ signal yields in each analysis category. The uncertainty is obtained by comparing event yields from $t\bar{t}H$ Monte Carlo samples with two different showering algorithms: MG5_aMC@NLO+Pythia8 and MG5_aMC@NLO+Herwig++.

The main sources of systematic uncertainty on the measurement of $t\bar{t}H$ cross section times $H \to \gamma\gamma$ branching fraction are summarized in Table 10.5. To determine the contribution from each group of systematics, the corresponding nuisance parameters are fixed to their best fit values from the unconditional maximum likelihood fit. The resulting error is then

subtracted (in quadrature) from the total error to obtain the error due only to the systematics in question.

Uncertainty source	$\Delta\sigma_{\text{low}}/\sigma$ [%]	$\Delta\sigma_{\text{high}}/\sigma$ [%]
Theory uncertainties	6.6	9.7
Underlying Event/Parton Shower	5.0	7.2
Heavy flavor	4.0	3.4
QCD scale	3.3	4.7
PDF+α_S	0.3	0.5
Experimental uncertainties	7.8	9.1
Photon energy resolution	5.5	6.2
Photon energy scale	2.8	2.7
Jet/missing E_T	2.3	2.7
Spurious signal	2.1	2.0
Photon efficiency	1.9	2.7
Luminosity and trigger	1.6	2.3
Higgs mass	1.6	1.5
Pileup	1.0	1.5
Flavor tagging	0.9	1.1
Leptons	0.4	0.6

Table 10.5: Contribution of groups of systematic uncertainties to the total error on the measured $t\bar{t}H$ cross section times $H \to \gamma\gamma$ branching ratio with 139 fb^{-1}. This is shown as the uncertainty due to each group of systematics ($\Delta\sigma$), as a fraction of the total cross section (σ). The numbers are obtained from quadratic subtraction of the total uncertainty and the uncertainty of the fit with the corresponding group fixed.

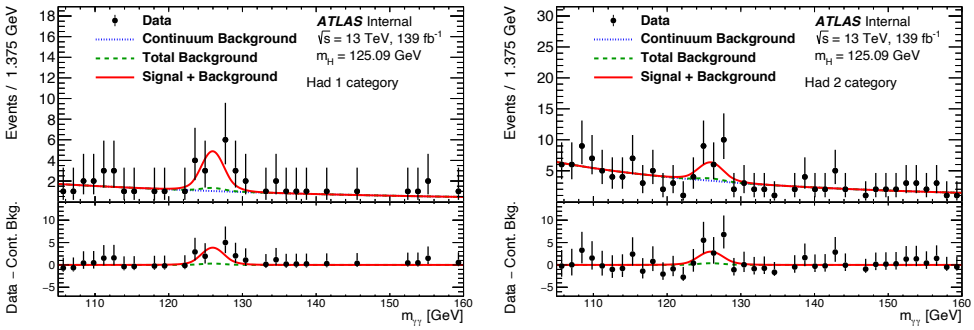
The dominant theory uncertainty is the UEPS uncertainty: since up to six jets in a single event are used in the BDT training (see Chapter 9), dependence on the parton showering algorithm is expected. The photon energy scale and resolution are the dominant experimental systematics, and their role will be discussed in the next Section.

10.2.3 Results

The diphoton mass spectrum in each hadronic (leptonic) category is shown in Figure 10.5 (Figure 10.6), with the signal-plus-background fits overlaid. The observed significance of $t\bar{t}H(\gamma\gamma)$ is 4.9σ, which corresponds to a signal strength of

$$\mu_{t\bar{t}H} = 1.38\ ^{+0.41}_{-0.36} = 1.38\ ^{+0.33}_{-0.31}\ (\text{stat.})\ ^{+0.13}_{-0.11}\ (\text{exp.})\ ^{+0.22}_{-0.14}\ (\text{th.}) \tag{10.2}$$

The expected $t\bar{t}H(\gamma\gamma)$ significance on the full Run 2 dataset (assuming the Standard Model) is 4.5σ (pre-fit) and 4.2σ (post-fit).



(a) $t\bar{t}H(\gamma\gamma)$ had 1

(b) $t\bar{t}H(\gamma\gamma)$ had 2

(c) $t\bar{t}H(\gamma\gamma)$ had 3

(d) $t\bar{t}H(\gamma\gamma)$ had 4

Figure 10.5: Diphoton invariant mass spectrum in the four hadronic $t\bar{t}H(\gamma\gamma)$ categories with 139 fb^{-1} of ATLAS data. $t\bar{t}H(\gamma\gamma)$ had 1 (a) corresponds to the tightest cut on BDT score, and $t\bar{t}H(\gamma\gamma)$ had 4 (d) to the loosest. The data is shown in black, with the fitted continuum background distribution in blue, the total background (including non-$t\bar{t}H$ Higgs processes) in green, and the $t\bar{t}H$ signal plus background in red.

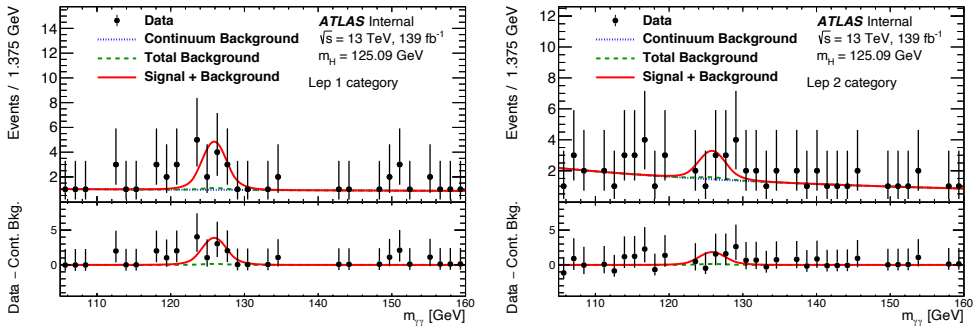
Figure 10.7 shows the fitted signal and background yields in each category, calculated in the $m_{\gamma\gamma}$ window containing 90% of the Higgs boson signal (S_{90} in Table 10.2). These yields are also listed in Table 10.6.

In order to visually enhance the contribution of the most powerful analysis categories, the data and fit in each category are weighted by $\ln(1+S/B)$ where S (B) is the expected signal (background) in the smallest $m_{\gamma\gamma}$ window containing 90% of the Higgs boson signal. This weighting procedure serves to magnify the contribution to the $m_{\gamma\gamma}$ spectrum of categories with highest S/B. The weighted diphoton mass spectrum (summed over all categories) is shown in Figure 10.8. A description of the calculation of errors on this plot is included in Section 6.5.

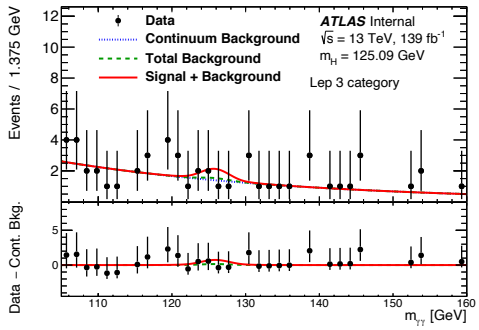
The $t\bar{t}H$ cross section times the $H \to \gamma\gamma$ branching fraction is measured to be

$$\sigma_{t\bar{t}H} \times B_{\gamma\gamma} = 1.59\ ^{+0.43}_{-0.39}\ \text{fb} = 1.59\ ^{+0.38}_{-0.36}\ (\text{stat.})\ ^{+0.15}_{-0.12}\ (\text{exp.})\ ^{+0.15}_{-0.11}\ (\text{th.})\ \text{fb} \tag{10.3}$$

This is compared to a Standard Model prediction of $1.15\ ^{+0.09}_{-0.12}$ fb.



(a) $t\bar{t}H(\gamma\gamma)$ lep 1 (b) $t\bar{t}H(\gamma\gamma)$ lep 2



(c) $t\bar{t}H(\gamma\gamma)$ lep 3

Figure 10.6: Diphoton invariant mass spectrum in the three leptonic $t\bar{t}H(\gamma\gamma)$ categories with 139 fb^{-1} of ATLAS data. $t\bar{t}H(\gamma\gamma)$ lep 1 (a) corresponds to the tightest cut on BDT score, and $t\bar{t}H(\gamma\gamma)$ lep 3 (c) to the loosest. The data is shown in black, with the fitted continuum background distribution in blue, the total background (including non-$t\bar{t}H$ Higgs processes) in green, and the $t\bar{t}H$ signal plus background in red.

Characterizing the Higgs peak

In the seven $t\bar{t}H$ categories, the shape of the Higgs mass peak in data is broader and centered at higher m_H than expected. As a result, the photon energy scale and resolution systematics are pulled to fit the peak in data, making these the dominant experimental uncertainties. A number of tests are performed to quantify the difference between the observed and expected Higgs signal peak, and no significant tension with the expected Higgs boson signal shape is observed.

If the Higgs boson mass is allowed to float along with the $t\bar{t}H$ signal strength, the best fit occurs at $m_H = 126.45\ ^{+0.65}_{-0.63}$ GeV and $\mu_{t\bar{t}H} = 1.40\ ^{+0.42}_{-0.36}$. The tension between this measurement and the expectation ($m_H = 125.09$ GeV) is at the level of 2.15σ.

In order to fit the width of the Higgs mass peak, a scale factor is assigned to σ_{CB}, the resolution of the DSCB function. A scale factor of 1.0 corresponds to the expected width, while a scale factor larger than 1.0 corresponds to a broadened peak. This scale factor is

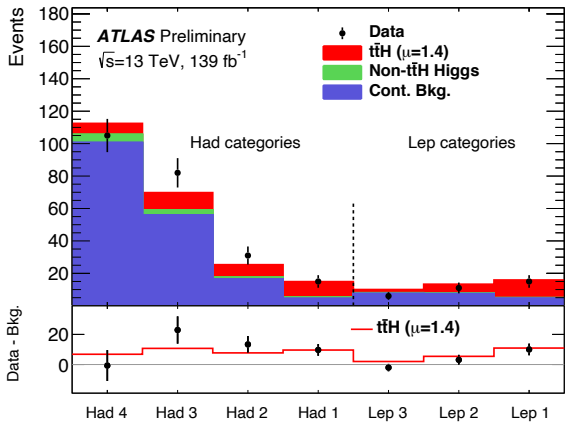


Figure 10.7: Fitted number of $t\bar{t}H(\gamma\gamma)$ and background events in each analysis category with 139 fb^{-1} of ATLAS data. The fitted continuum background yield is shown in blue, the non-$t\bar{t}H$ Higgs background in green, and the $t\bar{t}H$ signal (normalized to the fitted $\mu = 1.38$) in red. The total number of data events per category is overlaid in black.

Category	SM $t\bar{t}H$	SM non-$t\bar{t}H$ Higgs	Cont. bkg.	SM Total	Data
$t\bar{t}H(\gamma\gamma)$ had 1	6.94	0.81	4.50	12.25	15
$t\bar{t}H(\gamma\gamma)$ had 2	5.58	1.11	16.53	23.22	31
$t\bar{t}H(\gamma\gamma)$ had 3	7.73	3.07	56.02	66.83	82
$t\bar{t}H(\gamma\gamma)$ had 4	4.92	5.03	100.65	110.59	105
$t\bar{t}H(\gamma\gamma)$ lep 1	7.89	0.42	4.57	12.88	15
$t\bar{t}H(\gamma\gamma)$ lep 2	3.92	0.43	7.46	11.82	11
$t\bar{t}H(\gamma\gamma)$ lep 3	1.45	0.49	7.54	9.48	6

Table 10.6: Event yields in the seven $t\bar{t}H(\gamma\gamma)$ categories with 139 fb^{-1}. The observed yield (Data) is compared with the sum of expected $t\bar{t}H$ signal, background from non-$t\bar{t}H$ Higgs production, and continuum background. The continuum background yield is extracted from the unconditional maximum likelihood fit. Event yields are counted in the smallest $m_{\gamma\gamma}$ window containing 90% of the expected signal.

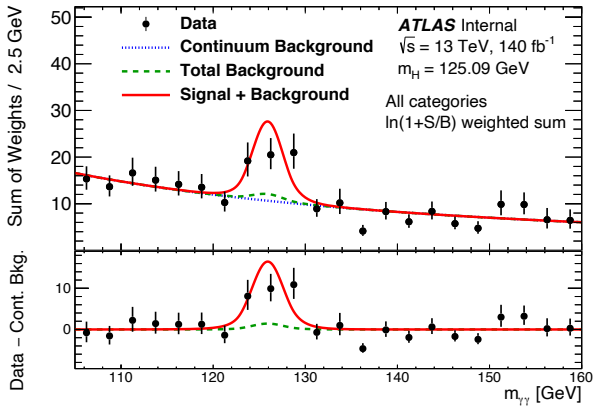


Figure 10.8: Weighted diphoton invariant mass spectrum of all seven $t\bar{t}H(\gamma\gamma)$ categories with 139 fb^{-1} of ATLAS data. The data is shown in black, with the fitted continuum background distribution in blue, the total background (including non-$t\bar{t}H$ Higgs processes) in green, and the $t\bar{t}H$ signal plus background in red. The data and fit in each category are weighted by $\ln(1 + S/B)$, where S (B) is the expected signal (background) in the smallest $m_{\gamma\gamma}$ window containing 90% of the Higgs signal.

allowed to vary along with $\mu_{t\bar{t}H}$ (for fixed m_H). The best fit is at a scale factor of 1.45 $^{+0.45}_{-0.32}$ and $\mu_{t\bar{t}H} = 1.52\ ^{+0.46}_{-0.39}$. The tension between this measurement and the expected width (scale factor = 1.0) is 1.53σ.

Allowing all three parameters ($\mu_{t\bar{t}H}$, m_H, and width scale factor) to vary simultaneously, the best fit value lies at $m_H = 126.33\ ^{+0.64}_{-0.63}$ GeV, width scale factor of 1.35 $^{+0.38}_{-0.29}$, and $\mu_{t\bar{t}H} = 1.53\ ^{+0.45}_{-0.38}$,. The tension between this measurement and the expectation (scale factor = 1.0 and $m_H = 125.09$ GeV) is 1.68σ.

A rigid signal shape is obtained by fixing the photon energy scale and resolution systematics to their expected values. The observed significance in this model is reduced to 4.5σ, corresponding to a signal strength of $\mu_{t\bar{t}H} = 1.14\ ^{+0.38}_{-0.32}$. Within uncertainties, these results with fixed resolution are consistent with results where the mean and width of the signal peak are allowed to vary.

Chapter 11

CP-sensitive categorization of $t\bar{t}H(\gamma\gamma)$ events

With the observation of the $t\bar{t}H$ process, the coupling between the Higgs boson and top quark has been directly established. The remainder of this work focuses on the direct measurement of the CP structure of the top Yukawa coupling using $t\bar{t}H$ and tH production in the $H \to \gamma\gamma$ decay channel.

As discussed in Chapter 2, the CP nature of the coupling (scalar, pseudoscalar or mixed) impacts the overall rates of $t\bar{t}H$ and tH production (Table 5.6) as well as sensitive kinematic distributions [50] [154] [155]. Because of the large enhancement in the tH cross section in BSM CP scenarios, the $tHjb$ and tWH processes are considered as signal in addition to $t\bar{t}H$. This Chapter presents the development of multivariate $t\bar{t}H(\gamma\gamma)$ CP categories that are sensitive to changes in $t\bar{t}H$ and tH kinematics with different mixing angles α. Results derived from this categorization using the full $H \to \gamma\gamma$ statistical machinery are presented in Chapter 12.

Section 11.1 reviews the figures of merit used to optimize the $t\bar{t}H(\gamma\gamma)$ CP category definitions. The most sensitive kinematic variables are reviewed in Section 11.2. Section 11.3 describes the training of CP BDTs to separate CP even and CP odd hypotheses, and Section 11.4 describes the procedure used to optimize the cut values defining category boundaries.

11.1 Number-counting limit on CP mixing

Following a procedure analogous to that presented in Section 9.1, the Poisson distribution $P(k, \lambda)$ (Equation 9.1) can be used to approximate the likelihood and a p-value can be calculated for various CP scenarios. Assuming the CP even hypothesis (SM, $\alpha = 0^\circ$ and $\kappa_t = 1$), the expected single-category exclusion of a CP hypothesis with mixing angle α and coupling strength κ_t is given by the p-value

$$-2\ln\frac{P(S_{0,1}+B, S_{\alpha,\kappa_t}+B)}{P(S_{0,1}+B, S_{0,1}+B)} = \tag{11.1}$$

$$2(S_{0,1}+B)\ln(\frac{S_{0,1}+B}{S_{\alpha,\kappa_t}+B}) - 2(S_{0,1}+B) + 2(S_{\alpha,\kappa_t}+B) \tag{11.2}$$

where S_{α,κ_t} is the number of signal events predicted at mixing angle α and coupling strength κ_t. In the measurement of CP properties of the Higgs-top coupling, the signal yield S_{α,κ_t} includes $t\bar{t}H$, $tHjb$, and tWH processes. The background yield B includes an estimate of other Higgs boson processes as well as the continuum diphoton background. Rather than exercising the full background estimation machinery, the number-counting significance relies on the estimation of continuum background yield from a data control region described in Section 5.1.1.

One figure of merit used to optimize the $t\bar{t}H(\gamma\gamma)$ CP categories is the expected exclusion of the CP odd hypothesis ($\alpha = 90^\circ$ and $\kappa_t = 1$):

$$Z_{CP}(90) = \sqrt{2(S_{0,1}+B)\ln(\frac{S_{0,1}+B}{S_{90,1}+B}) - 2(S_{0,1}+B) + 2(S_{90,1}+B)} \tag{11.3}$$

The other figure of merit is the expected significance of SM $t\bar{t}H + tH$ (Equation 9.3, with S = SM $t\bar{t}H + tHjb + tWH$): this is referred to as $Z_{t\bar{t}H+tH}$. In order to obtain the total $Z_{t\bar{t}H+tH}$ and $Z_{CP}(90)$, the single-category significances are added in quadrature.

The number-counting significance can be used to estimate the impact of systematic uncertainties on the signal and background. Since the measurement of CP in $t\bar{t}H(\gamma\gamma)$ is dominated by the statistical uncertainty on the number of background events, it is enough to compare the magnitude of systematic uncertainties to the background-only statistical uncertainty:

$$\delta = \frac{S_{0,1}}{Z_{t\bar{t}H+tH}} \tag{11.4}$$

For $\sigma_{\text{syst}} << \delta$, the impact of systematics is expected to be nearly negligible compared to the statistical uncertainty on the measurement. It is verified in Section 11.4 that the main theory uncertainties are small compared to δ in the optimized BDT categories.

Assuming $\kappa_t = 1$, a number-counting estimate of the 95% confidence limit on the mixing angle α can also be derived from the exclusion $Z_{CP}(\alpha)$. A curve can be calculated by interpolating $Z_{CP}(\alpha)$ calculated at all available Monte Carlo points (see Section 5.2.2). The value of α where this curve intersects $Z = 1.96$ corresponds to the 95% confidence limit.

11.2 Sensitive observables

Using $t\bar{t}H$ and tH Monte Carlo samples generated with $\kappa_t = 1$ and different values of the CP mixing angle α (details in Section 5.2.2), the following distributions are obtained at particle level (i.e. before simulating the parton shower, hadronization, and detector response).

The most powerful shape discrimination between CP even ($\alpha = 0°$, $\kappa_t = 1$) and CP odd ($\alpha = 90°$, $\kappa_t = 1$)comes from the p_T of the Higgs boson. In $t\bar{t}H$, $tHjb$, and tWH, the Higgs boson tends to have higher p_T and more central in η in CP hypotheses with higher values of the mixing angle α. Figure 11.1 shows the Higgs boson p_T and η for different values of α.

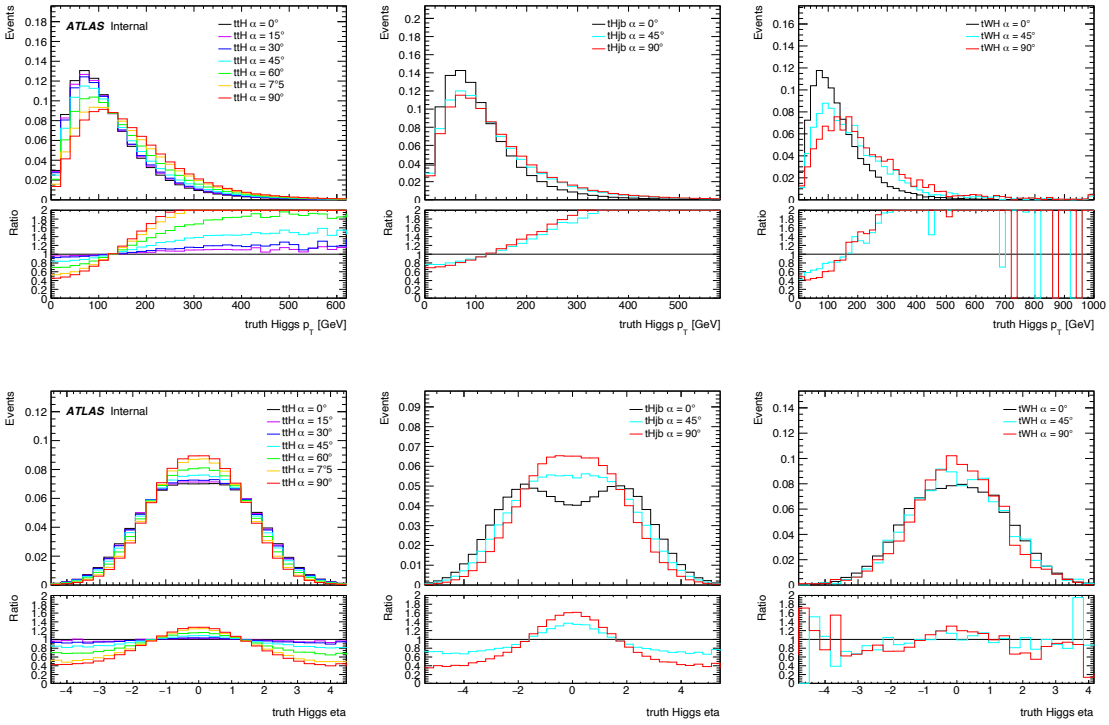


Figure 11.1: Particle-level distributions of Higgs boson p_T and η in $t\bar{t}H$ (left), $tHjb$ (center), and tWH (right) Monte Carlo for different values of the CP mixing angle α. The coupling strength κ_t is fixed to 1.

The angular separation between top quarks also provides discrimination between CP hypotheses in $t\bar{t}H$. Figure 11.2 shows the angular distances $\Delta\eta$ and $\Delta\phi$ between the top and antitop, which are larger for higher values of α.

The p_T and η of the top quarks are shown in Figure 11.3. These variables give some separation between CP even and CP odd in $tHjb$, but their discrimination power is limited in $t\bar{t}H$ and tWH. Another powerful variable is the invariant mass of the top quark plus Higgs boson system, shown in Figure 11.4. In the case of $t\bar{t}H$ and tWH, the m_{tH} is generally larger in CP odd than in CP even, while the opposite trend is visible in $tHjb$.

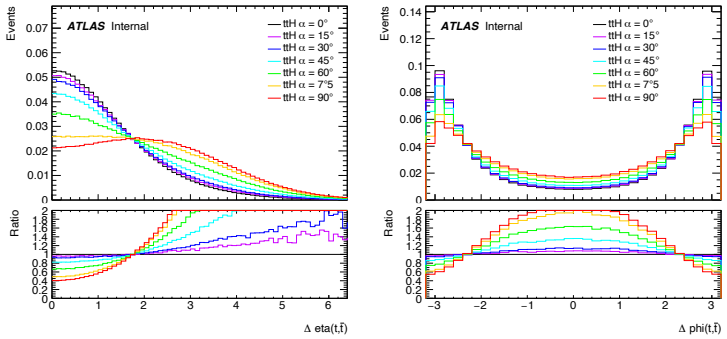


Figure 11.2: Particle-level distributions of the angular separation between top and antitop quarks in $t\bar{t}H$ Monte Carlo for different values of the CP mixing angle α. The coupling strength κ_t is fixed to 1.

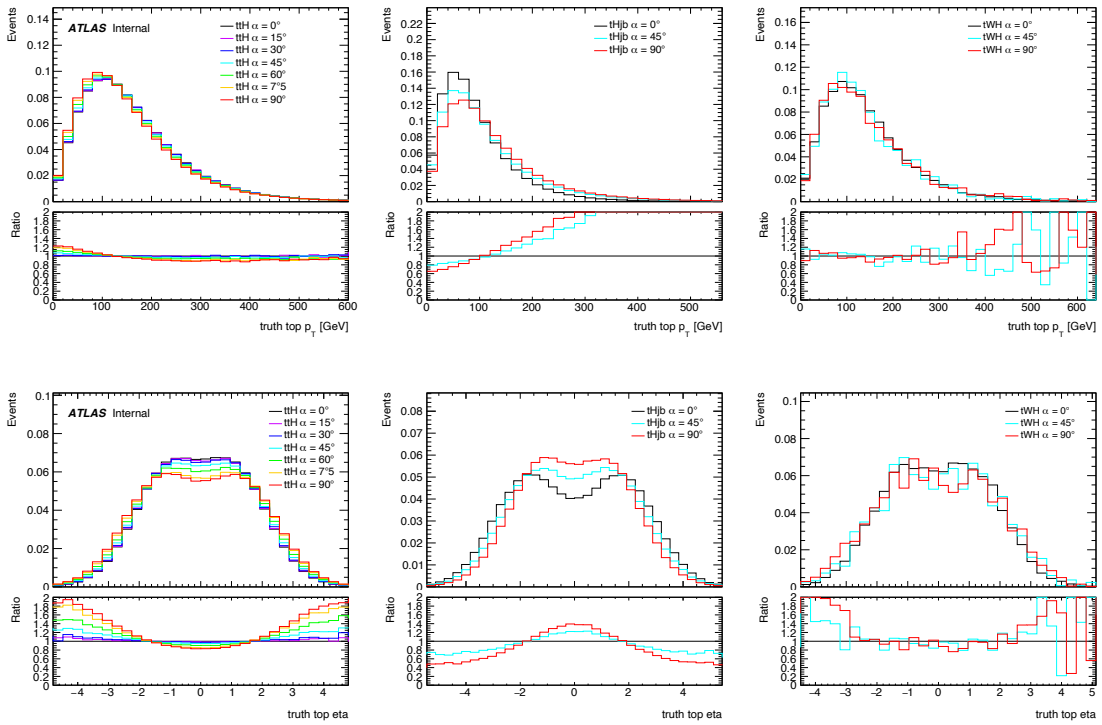


Figure 11.3: Particle-level distributions of top quark p_T and η in $t\bar{t}H$ (left), $tHjb$ (center), and tWH (right) Monte Carlo for different values of the CP mixing angle α. The coupling strength κ_t is fixed to 1.

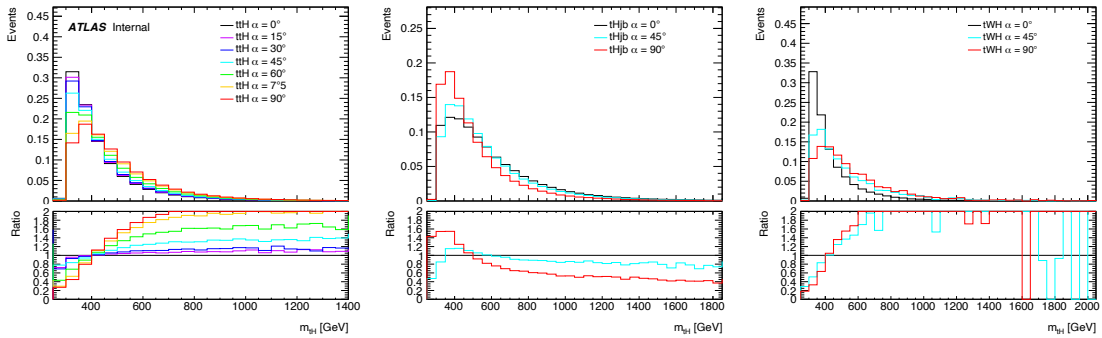


Figure 11.4: Particle-level distributions of m_{tH} in $t\bar{t}H$ (left), $tHjb$ (center), and tWH (right) Monte Carlo for different values of the CP mixing angle α. The coupling strength κ_t is fixed to 1.

11.3 Multivariate categorization strategy

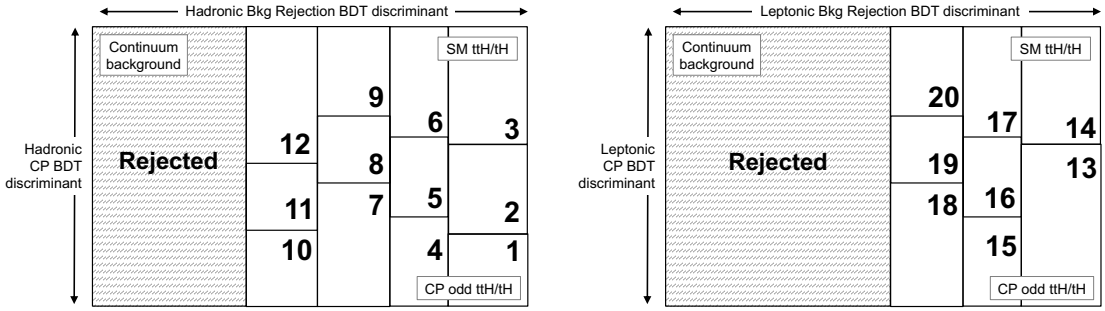
As in the Standard Model $t\bar{t}H(\gamma\gamma)$ analysis (Chapter 9), events are preselected into two channels targeting different decays of the $t\bar{t}$ system: the hadronic channel targets the all-hadronic $t\bar{t}$ decay, and the leptonic channel targets final states with at least one semi-leptonic top decay. The preselection for both channels is unchanged with respect to the SM $t\bar{t}H(\gamma\gamma)$ analysis.

The $t\bar{t}H(\gamma\gamma)$ CP analysis categories are based on a partition of a two-dimensional BDT space in each channel (see the schematic in Figure 11.5). The horizontal axis of the two-dimensional BDT plane is formed by the background rejection BDT discriminant, which separates $t\bar{t}H(\gamma\gamma)$ signal from background processes. The hadronic and leptonic background rejection BDTs are the same discriminants used in the SM $t\bar{t}H(\gamma\gamma)$ analysis and described in detail in Chapter 9.

The background rejection BDTs reject continuum background well, and acceptance at the signal-like end of the spectrum is high for both $t\bar{t}H$ and tH. Figures 11.6 and 11.7 show that there is weak dependence on CP mixing angle, and that this dependence is stronger in tH than in $t\bar{t}H$. The tight/isolated (TI) sideband data ($|m_{\gamma\gamma} - 125| > 5$) is overlaid, giving the approximate shape of the continuum background.

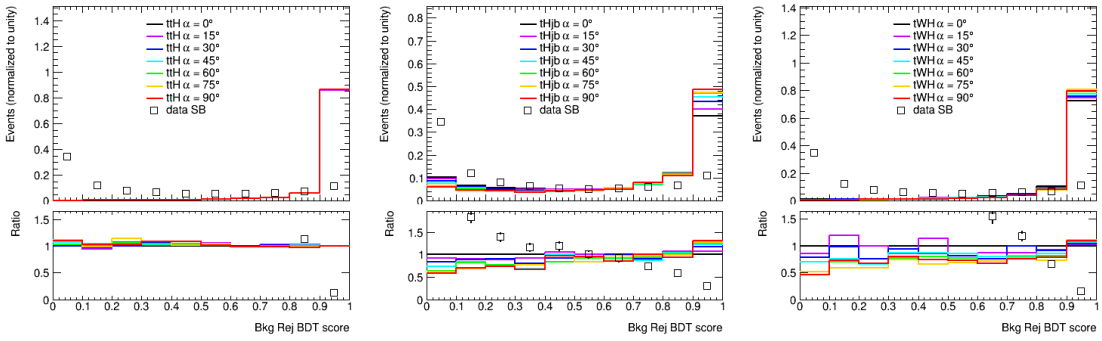
Figure 11.8 shows the background rejection BDT discriminants for the sum of $t\bar{t}H$, $tHjb$ and tWH, with relative weights according to the predicted cross sections in Table 5.6. Here, the dependence on mixing angle is due primarily to the changing fraction of the tH processes.

The vertical axis of the two-dimensional BDT plane (Figure 11.5) is formed by the CP BDT, which separates simulated CP even ($\alpha = 0°$, $\kappa_t = 1$) and CP odd ($\alpha = 90°$, $\kappa_t = 1$) $t\bar{t}H + tH$ signals. The following sections describe the training of the hadronic and leptonic CP BDTs.



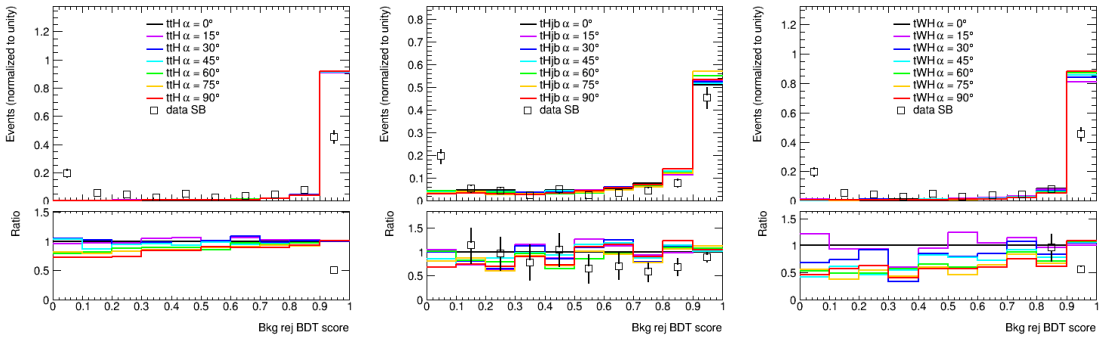
(a) Hadronic categories (b) Leptonic categories

Figure 11.5: Diagram of the two-dimensional BDT plane and $t\bar{t}H(\gamma\gamma)$ CP categories in (left) the hadronic channel and (right) the leptonic channel. The horizontal axis represents the background rejection BDT score, and the vertical axis represents the CP BDT score. Vertical and horizontal lines denote the category boundaries, and the shaded regions contain rejected events.



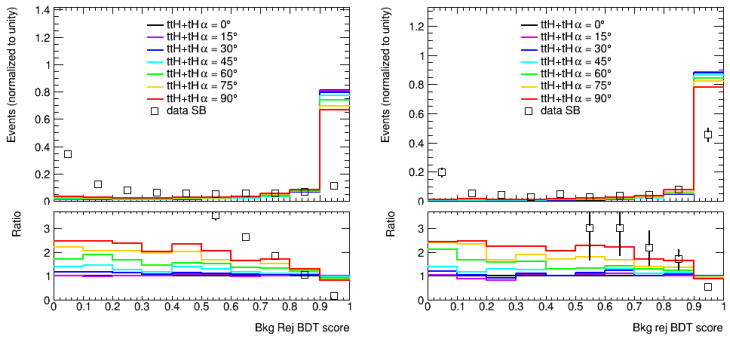
(a) $t\bar{t}H$ had (b) $tHjb$ had (c) tWH had

Figure 11.6: Hadronic background rejection BDT score for $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle. The coupling strength κ_t is fixed to 1. The open squares show TI data in the sideband region.



(a) $t\bar{t}H$ lep (b) $tHjb$ lep (c) tWH lep

Figure 11.7: Leptonic background rejection BDT score for $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle. The coupling strength κ_t is fixed to 1. The open squares show TI data in the sideband region.



(a) Hadronic $t\bar{t}H + tH$ (b) Leptonic $t\bar{t}H + tH$

Figure 11.8: Background rejection BDT score for the sum of $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle in (a) the hadronic channel and (b) the leptonic channel. The coupling strength κ_t is fixed to 1. The open squares show TI data in the sideband region.

11.3.1 Hadronic channel

Following the preselection in Chapter 9, hadronic channel events must contain zero leptons (e or μ) and at least three jets, one or more of which is b-tagged at the 77% working point.

The hadronic CP BDT is trained to separate CP even and CP odd signal events passing hadronic preselection using XGBoost [140]. The training samples are constructed from MG5_aMC@NLO + Pythia8 Monte Carlo samples of $t\bar{t}H$, $tHjb$, and tWH. The composition of $t\bar{t}H$, $tHjb$, and tWH in each sample is determined by the predicted cross section of each process given in Table 5.6. The training is performed on 50% of each sample, while 25% is reserved for validation (including early stopping of the BDT and determination of category boundaries), and the final 25% is reserved for sensitivity evaluation.

The discriminating variables used in the hadronic CP BDT training are:

- p_T and η of the diphoton system (Higgs boson candidate)
- p_T, η, ϕ (with respect to the Higgs boson candidate), and Top Reco BDT score of the reconstructed top candidate and the second top candidate (see Chapter 8). In the case where no second top is reconstructed (either partially or fully), missing values are passed to XGBoost for the kinematics of the second top.
- Angles $\Delta\eta$ and $\Delta\phi$ between the top candidates. In the case where no second top is reconstructed (either partially or fully), missing values are passed to XGBoost.
- Two-object invariant masses m_{t_1H}, and $m_{t_1t_2}$. In the case where no second top is reconstructed (either partially or fully), the value of m_{t_1} is passed to XGBoost for the $m_{t_1t_2}$ variable.
- $H_T = \sum_{\text{jet } j} p_T^j$
- The minimum ΔR between a photon and a jet in the event
- The second-smallest ΔR between a photon and a jet in the event (out of all photon-jet combinations)
- Jet multiplicity and b-jet multiplicity (77% working point)
- Missing E_T significance (Equation 4.7)

Figures 11.9 -11.15 compare the Monte Carlo distribution of each training variable for different values of the mixing angle α, with the coupling strength κ_t fixed to 1.

The BDT hyper-parameters are chosen to maximize the area under the ROC curve evaluated on the validation set. The ROC AUC of the hadronic CP BDT is 78.39%. The distributions of the output BDT score for $t\bar{t}H$, $tHjb$ and tWH are shown separately for each process in Figure 11.16.

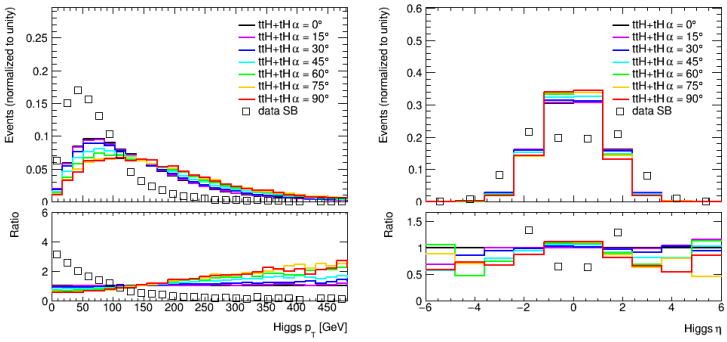


Figure 11.9: Hadronic CP BDT training variables: $p_T^{\gamma\gamma}$ and $\eta^{\gamma\gamma}$. The open squares show TI data in the sideband region.

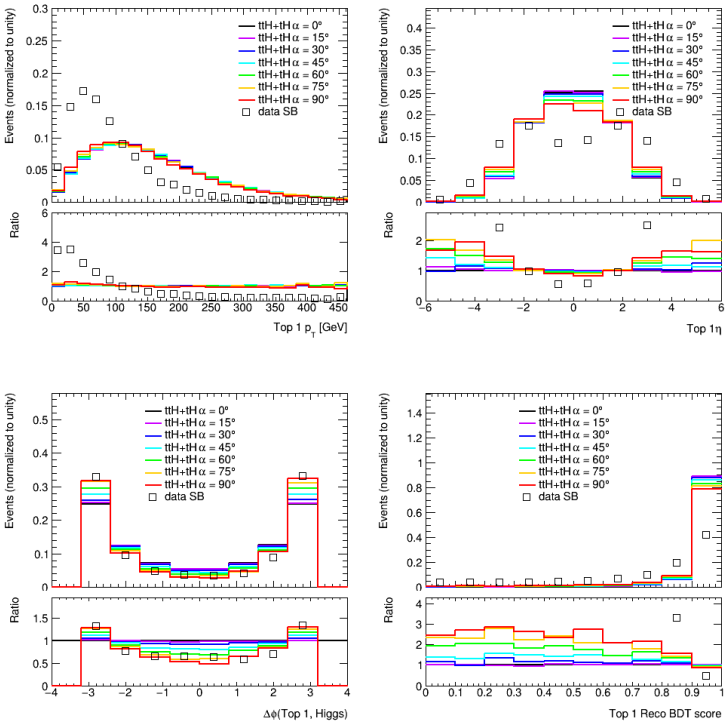


Figure 11.10: Hadronic CP BDT training variables: p_T, η, ϕ with respect to the Higgs boson candidate, and Top Reco BDT score of the first reconstructed top quark. The open squares show data in the sideband region.

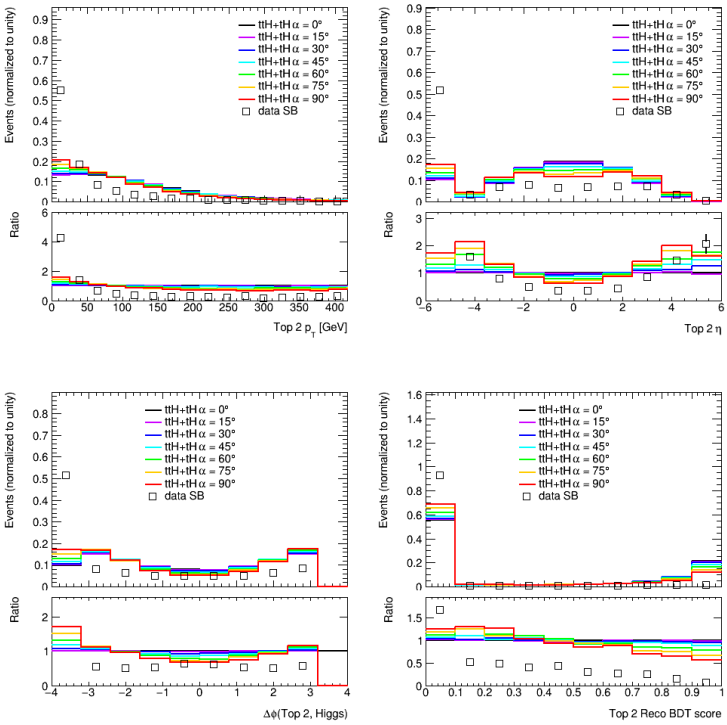


Figure 11.11: Hadronic CP BDT training variables: p_T, η, ϕ with respect to the Higgs boson candidate, and Top Reco BDT score of the second reconstructed top quark. The underflow bins represent events where no second top is reconstructed (either partially or fully). The open squares show data in the sideband region.

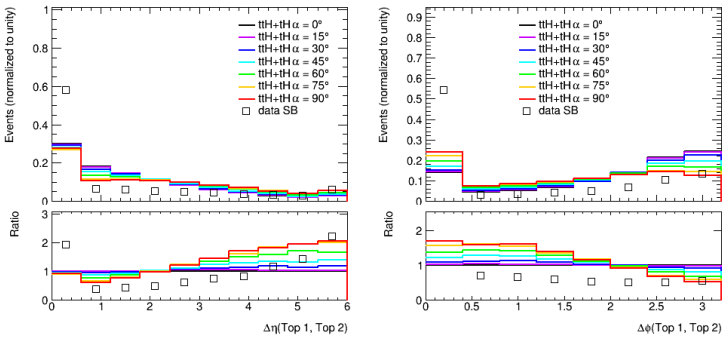


Figure 11.12: Hadronic CP BDT training variables: $\Delta\eta(t_1, t_2)$ and $\Delta\phi(t_1, t_2)$. The underflow bins represent events where no second top is reconstructed (either partially or fully). The open squares show TI data in the sideband region.

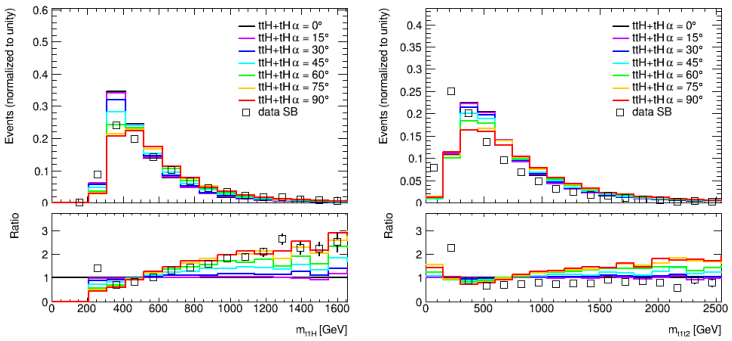


Figure 11.13: Hadronic CP BDT training variables: m_{t_1H} and $m_{t_1t_2}$. The open squares show TI data in the sideband region.

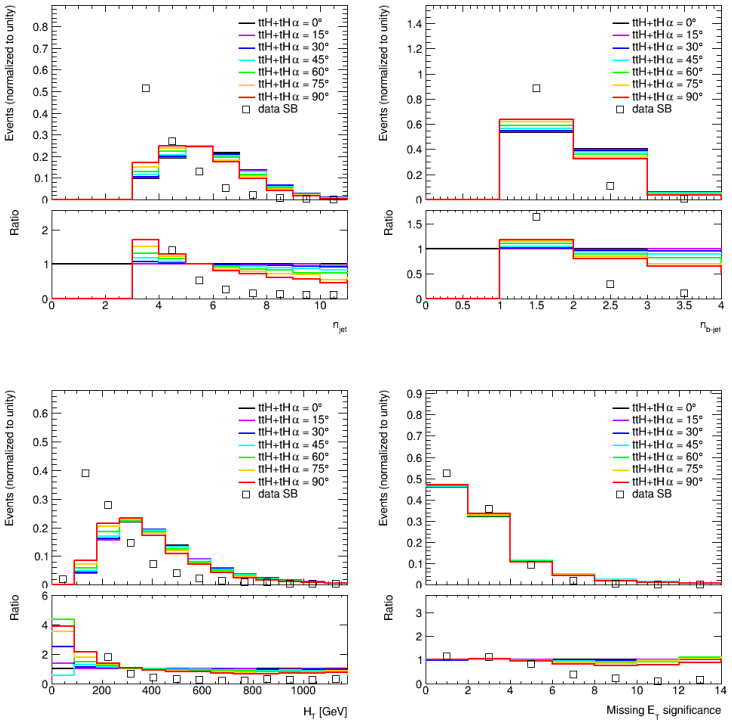


Figure 11.14: Hadronic CP BDT training variables: jet multiplicity, b-jet multiplicity, H_T, and missing E_T significance. The open squares show TI data in the sideband region.

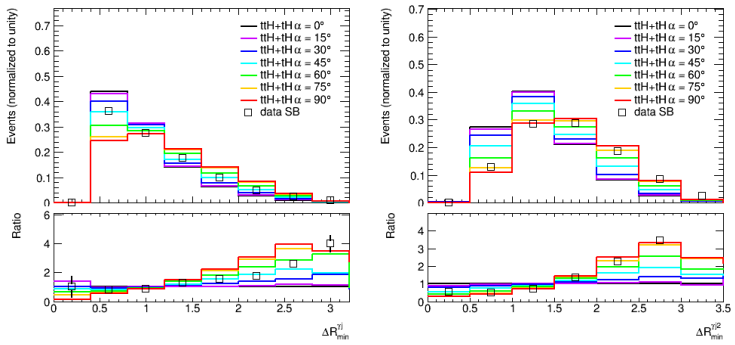
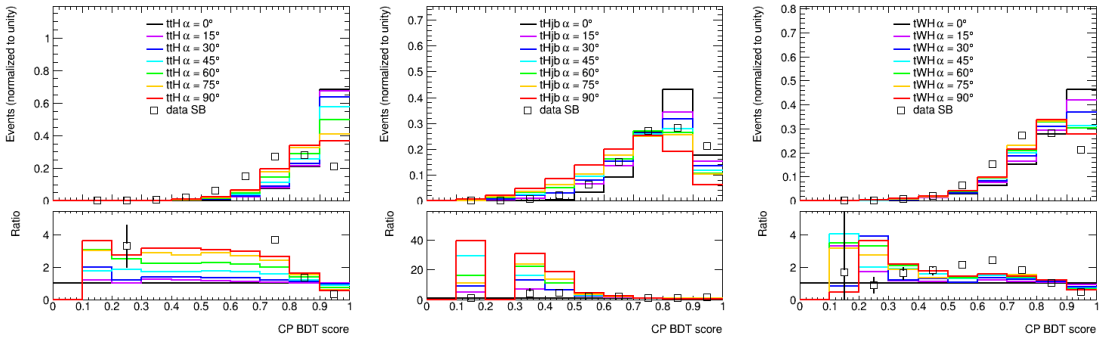


Figure 11.15: Hadronic CP BDT training variables: $\Delta R^{\gamma j}_{\min}$ and $\Delta R^{\gamma j 2}_{\min}$. The open squares show TI data in the sideband region.



(a) $t\bar{t}H$ had (b) $tHjb$ had (c) tWH had

Figure 11.16: Hadronic CP BDT score for $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle. The coupling strength κ_t is fixed to 1. The open squares show TI data in the sideband region.

Figure 11.17 shows the two-dimensional hadronic BDT plane in data passing hadronic preselection. For ease of visualization, the axes here represent transformations of the raw BDT scores. The inner (outer) contours capture 25% (50%) of the $t\bar{t}H + tH$ signal for the CP even and CP odd hypotheses. The projection onto the background rejection BDT and CP BDT axes are shown in the right-hand panels. Data and signal samples are normalized to unit area for a comparison of BDT shapes. It is clear that the background (the dominant component in the data) is concentrated at low background rejection BDT score, while the signal is concentrated at high background rejection BDT score. In addition, the CP even and CP odd signal hypotheses are separated by the CP BDT score. The SM expectation (CP even) is concentrated in the upper right corner of the 2D plane.

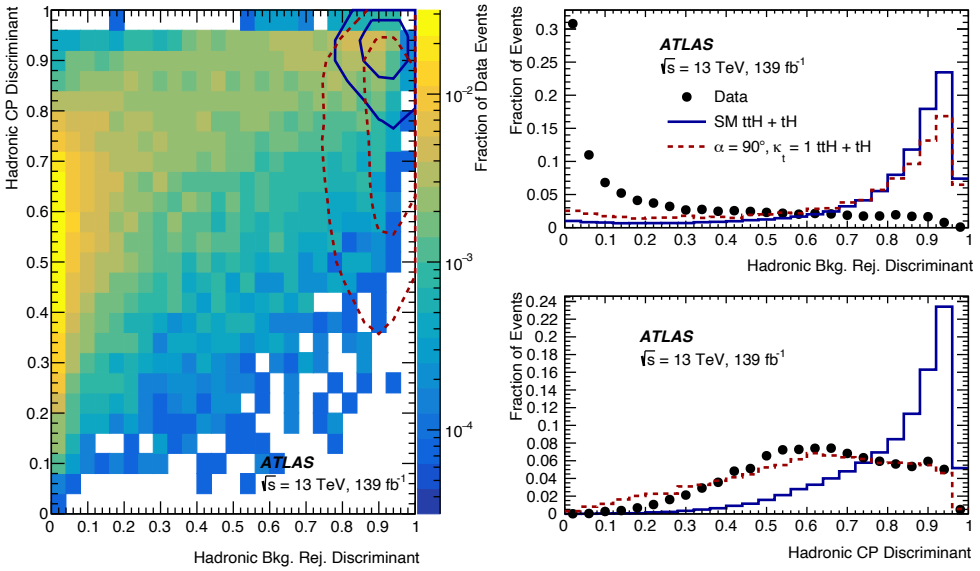


Figure 11.17: Two-dimensional BDT distribution in data passing hadronic preselection. The horizontal axis shows the background Rejection BDT discriminant and the vertical axis shows the CP BDT discriminant. The inner (outer) contours capture 25% (50%) of the $t\bar{t}H + tH$ events for the CP even (blue) and CP odd (red) hypotheses. The right panel shows projections onto the BDT discriminant axes. Contributions from CP even (blue), CP odd (red) $t\bar{t}H + tH$ processes and the data (black) are shown and normalized to unit area.

11.3.2 Leptonic channel

Again following the preselection in Chapter 9, leptonic channel events must contain at least one lepton (e or μ) and at least one b-jet.

The leptonic CP BDT is trained to separate CP even and CP odd events passing leptonic preselection using XGBoost [140]. The construction of training samples and the training variables used are analogous to those used in training of the hadronic CP BDT

The training samples are constructed from MG5_aMC@NLO + Pythia8 Monte Carlo samples of $t\bar{t}H$, $tHjb$, and tWH, with the composition of $t\bar{t}H$, $tHjb$, and tWH determined by the predicted cross sections in Table 5.6. The training is performed on 50% of each sample, while 25% is reserved for validation (including early stopping of the BDT and determination of category boundaries), and the final 25% is reserved for sensitivity evaluation.

The discriminating variables used in the leptonic CP BDT training are the same as those used in the hadronic training (though top quarks are reconstructed differently, as described in Chapter 8). In dilepton events (where no top quarks are reconstructed), missing values are passed to XGBoost for all top-related variables. Figures 11.18 -11.24 compare the Monte Carlo distribution of each training variable for different values of the mixing angle α, with the coupling strength κ_t fixed to 1.

The BDT hyper-parameters are chosen to maximize the area under the ROC curve. The

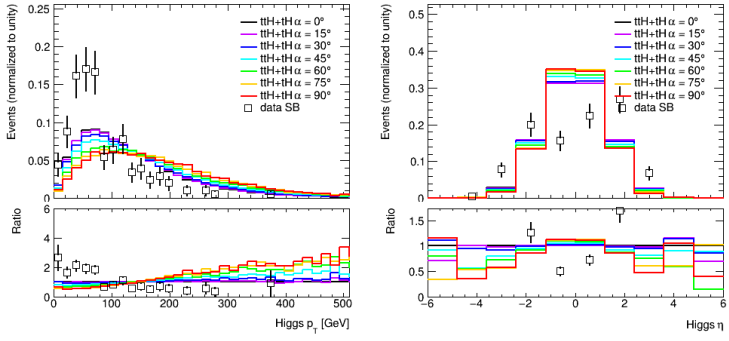


Figure 11.18: Leptonic CP BDT training variables: $p_T^{\gamma\gamma}$ and $\eta^{\gamma\gamma}$. The open squares show TI data in the sideband region.

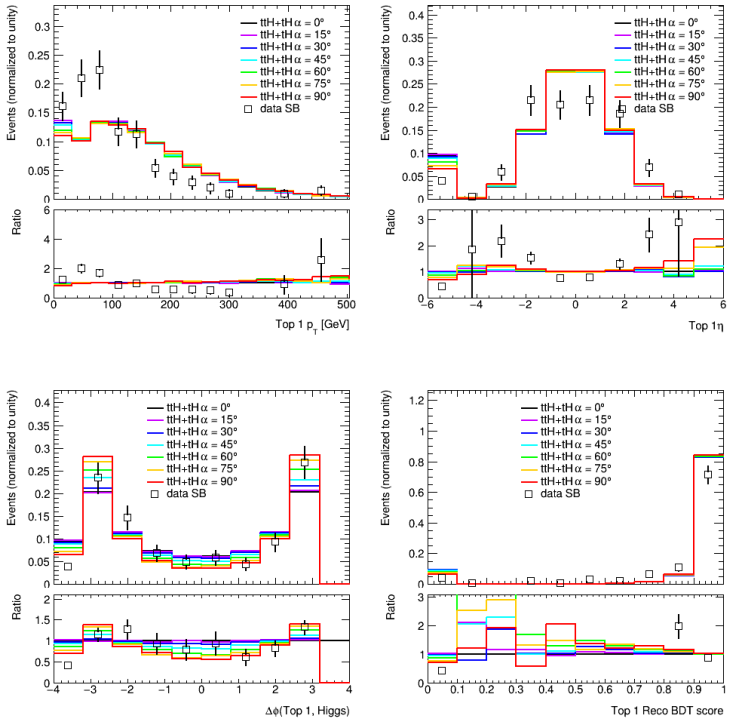


Figure 11.19: Leptonic CP BDT training variables: p_T, η, ϕ with respect to the Higgs boson candidate, and Top Reco BDT score of the first reconstructed top quark. The underflow bin corresponds to events containing ≥ 2 leptons, where no top is reconstructed. The open squares show TI data in the sideband region.

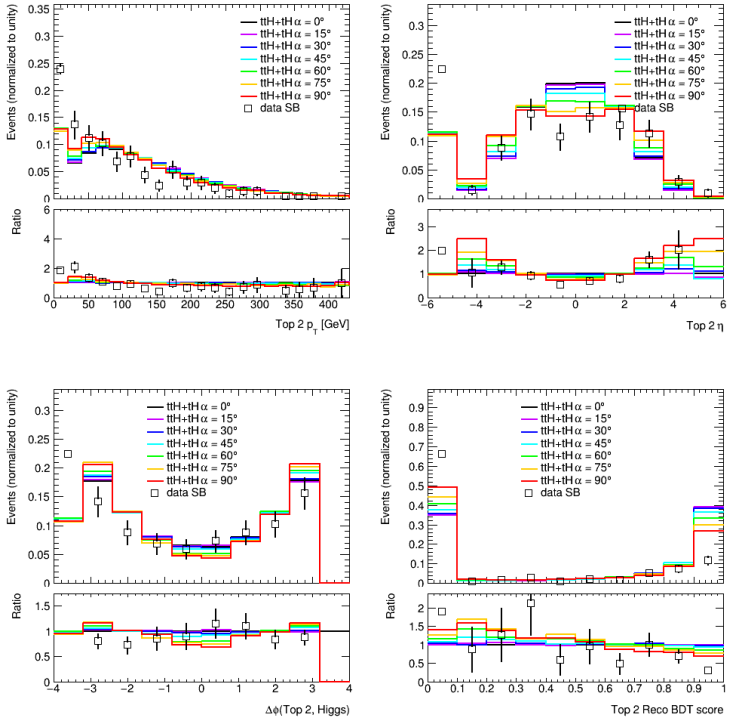


Figure 11.20: Leptonic CP BDT training variables: p_T, η, ϕ with respect to the Higgs boson candidate, and Top Reco BDT score of the second reconstructed top quark. The underflow bins in p_T, η, and ϕ represent events containing ≥ 2 leptons or where no second top is reconstructed (either partially or fully). The underflow bin in Top 2 Reco BDT score represents events where the second top is not fully reconstructed. The open squares show TI data in the sideband region.

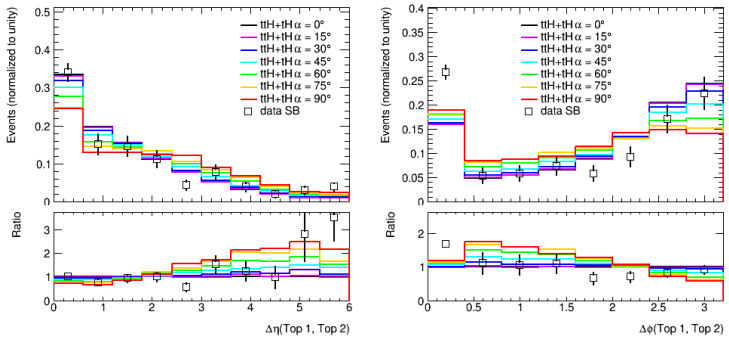


Figure 11.21: Leptonic CP BDT training variables: $\Delta\eta(t_1, t_2)$ and $\Delta\phi(t_1, t_2)$. The underflow bins represent events containing ≥ 2 leptons or where no second top is reconstructed (either partially or fully). The open squares show TI data in the sideband region.

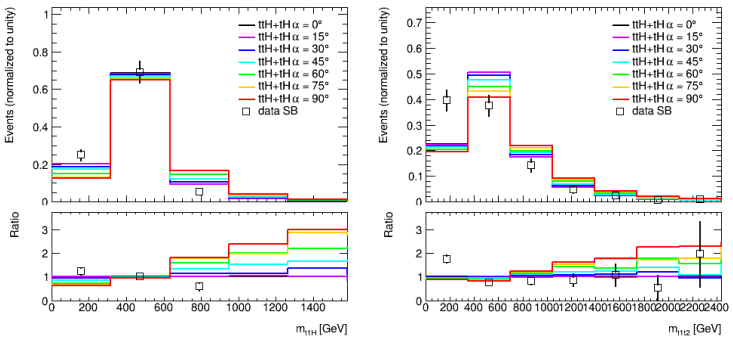


Figure 11.22: Leptonic CP BDT training variables: $m_{t_1 H}$ and $m_{t_1 t_2}$. The underflow bins contain events with ≥ 2 leptons, where no top is reconstructed. The open squares show TI data in the sideband region.

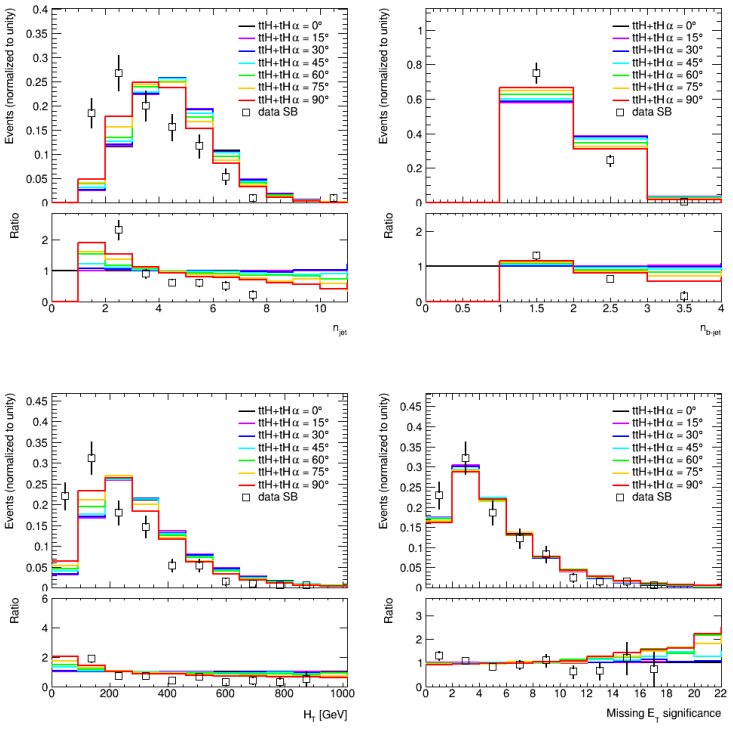


Figure 11.23: Leptonic CP BDT training variables: jet multiplicity, b-jet multiplicity, H_T, and missing E_T significance. The open squares show TI data in the sideband region.

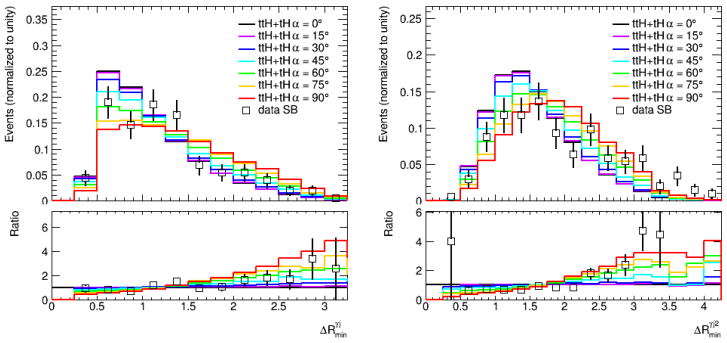
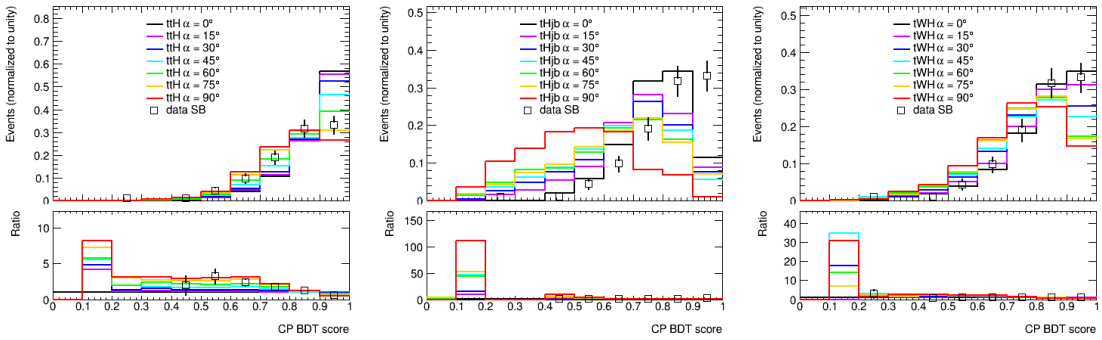


Figure 11.24: Leptonic CP BDT training variables: $\Delta R^{\gamma j}_{\min}$ and $\Delta R^{\gamma j 2}_{\min}$. The open squares show TI data in the sideband region.

ROC AUC for the leptonic CP BDT is 76.69%. The distributions of the output BDT score for $t\bar{t}H$, $tHjb$ and tWH are shown separately for each process in Figure 11.25.



(a) $t\bar{t}H$ lep (b) $tHjb$ lep (c) tWH lep

Figure 11.25: Leptonic CP BDT score for $t\bar{t}H$, $tHjb$, and tWH for different values of the CP mixing angle. The open squares show TI data in the sideband region.

Figure 11.26 shows the two-dimensional leptonic BDT plane in data passing leptonic preselection. For ease of visualization, the axes here represent transformations of the raw BDT scores. The inner (outer) contours capture 25% (50%) of the $t\bar{t}H + tH$ signal for the CP even and CP odd hypotheses. The projection onto the background rejection BDT and CP BDT axes are shown in the right-hand panels. Data and signal samples are normalized to unit area for a comparison of BDT shapes. Due to low statistics in data compared to the hadronic channel, the binning in the 2D BDT plane is coarser here. It is clear that the background (the dominant component in the data) is concentrated at low background rejection BDT score, while the signal is concentrated at high background rejection BDT score. In addition, the CP even and CP odd signal hypotheses are separated by the CP

BDT score. The SM expectation (CP even) is concentrated in the upper right corner of the 2D plane.

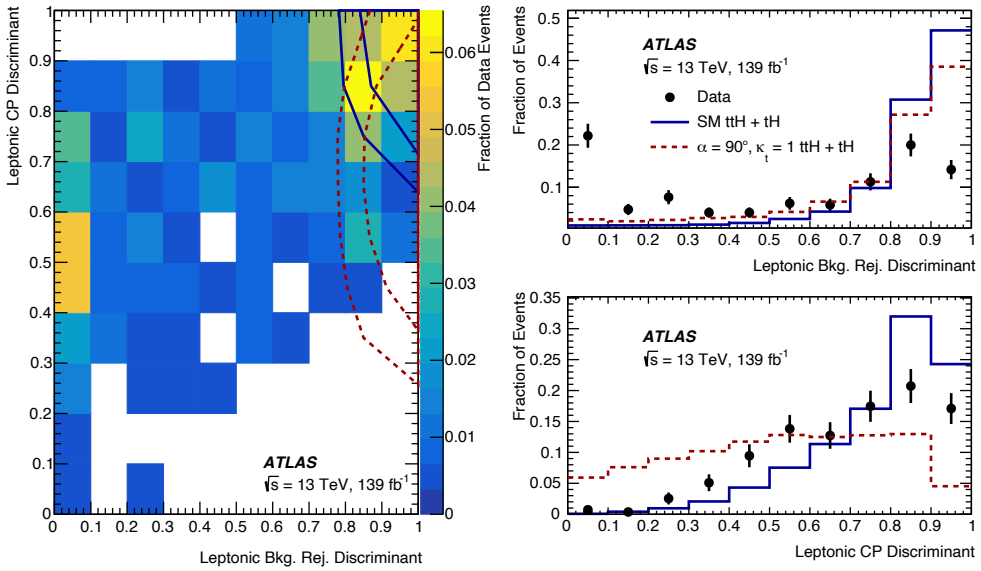


Figure 11.26: Two-dimensional BDT distribution in data passing leptonic preselection. The horizontal axis shows the background rejection BDT discriminant and the vertical axis shows the CP BDT discriminant. The inner (outer) contours capture 25% (50%) of the $t\bar{t}H + tH$ signal events for the CP even (blue) and CP odd (red) hypotheses. The right panel shows projections onto the BDT discriminant axes. Contributions from CP even (blue), CP odd (red) $t\bar{t}H + tH$ processes and the data (black) are shown and normalized to unit area.

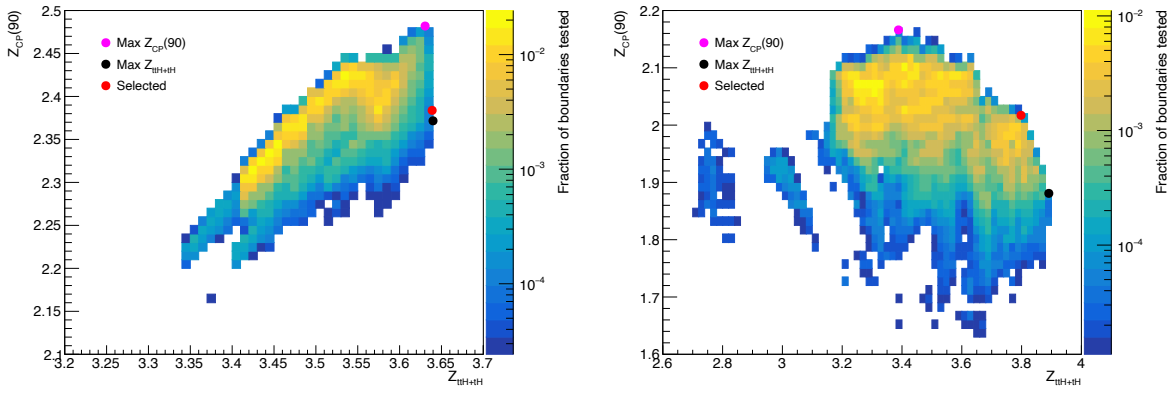
11.4 Determination of category boundaries

The optimal choice of category boundaries maximizes both the expected $t\bar{t}H + tH$ significance ($Z_{t\bar{t}H+tH}$) and the expected number-counting exclusion of the CP odd hypothesis ($Z_{CP}(90)$ from Equation 11.3). A brute force scan is performed over all possible boundary combinations, and the boundary values are chosen to maximize $Z_{t\bar{t}H+tH}$ and $Z_{CP}(90)$. For boundary optimization, both figures of merit are evaluated on the validation set, since the cuts forming the category boundaries are considered hyper-parameters of the analysis.

Up to four (three) categories are created by placing cuts on the hadronic (leptonic) background rejection BDT discriminant. Events falling below the loosest cut are rejected. Within each background rejection category, up to three CP categories are defined. Merging of categories is allowed by overlapping of the category boundaries. The maximum allowed number of categories is twelve (nine) for hadronic (leptonic). The selected categorization contains twelve hadronic and eight leptonic categories: two rather than three CP BDT categories are created in the tightest leptonic background rejection BDT region.

In order to have sufficient data in each category to perform the maximum likelihood fit described in Chapter 6, a minimum of 0.8 continuum background events (predicted from NTI data) is required in each category in the range $123 < m_{\gamma\gamma} < 127$ GeV. This corresponds to ~ 10 events in the full fit range of 105-160 GeV.

The results of the brute force scan are shown in Figure 11.27. It is clear that the same set of boundaries does not necessarily maximize both $Z_{t\bar{t}H+tH}$ and $Z_{CP}(90)$. The final category boundaries are selected by allowing $Z_{CP}(90)$ to be lower than the maximal value by up to 0.15σ: this corresponds to a small drop of 1° in the number-counting 95% confidence limit on the mixing angle α.



(a) Hadronic channel

(b) Leptonic channel

Figure 11.27: Summary of the CP odd exclusion $Z_{CP}(90)$ and SM $t\bar{t}H + tH$ significance $Z_{t\bar{t}H+tH}$ obtained by each set of boundaries in the brute force scan. The maximum exclusion of the CP odd hypothesis is shown in magenta, the maximum SM $t\bar{t}H + tH$ significance in black, and the selected category boundaries in red.

The chosen boundaries are listed in Table 11.1. The categories are labeled with numbers 1-20 according to Figure 11.5. Categories 1-12 are in the hadronic channel, and categories 13-20 are in the leptonic channel. The most CP even-like categories are those with high CP BDT score (3, 6, 9, 12, 14, 17, 20), and the most CP odd-like categories are those with low CP BDT score (1, 4, 7, 10, 13, 16, 19).

The expected significance of $t\bar{t}H$ and tH signal in the CP even and CP odd scenarios is shown in Table 11.2. This Table also includes the expected exclusion of the CP odd and maximal CP mixing ($\alpha = 45°$, $\kappa_t = 1$) hypotheses, assuming SM observation.

The background-only statistical uncertainty δ (Equation 11.4) is shown for each category in Table 11.3 and compared to the major theory uncertainties that will impact the result: these include the 100% heavy flavor uncertainty on ggF, as well as on the Underlying Event/Parton Showering uncertainties on $t\bar{t}H$, $tHjb$, and tWH signal. All of these systematics

Category	Bkg Rej BDT score	CP BDT score
$t\bar{t}H(\gamma\gamma)$ CP 1	$[0.995 - 1]$	$[0 - 0.90]$
$t\bar{t}H(\gamma\gamma)$ CP 2	$[0.995 - 1]$	$[0.90 - 0.97]$
$t\bar{t}H(\gamma\gamma)$ CP 3	$[0.995 - 1]$	$[0.97 - 1]$
$t\bar{t}H(\gamma\gamma)$ CP 4	$[0.991 - 0.995]$	$[0 - 0.88]$
$t\bar{t}H(\gamma\gamma)$ CP 5	$[0.991 - 0.995]$	$[0.88 - 0.96]$
$t\bar{t}H(\gamma\gamma)$ CP 6	$[0.991 - 0.995]$	$[0.96 - 1]$
$t\bar{t}H(\gamma\gamma)$ CP 7	$[0.981 - 0.991]$	$[0 - 0.84]$
$t\bar{t}H(\gamma\gamma)$ CP 8	$[0.981 - 0.991]$	$[0.84 - 0.96]$
$t\bar{t}H(\gamma\gamma)$ CP 9	$[0.981 - 0.991]$	$[0.96 - 1]$
$t\bar{t}H(\gamma\gamma)$ CP 10	$[0.909 - 0.981]$	$[0 - 0.61]$
$t\bar{t}H(\gamma\gamma)$ CP 11	$[0.909 - 0.981]$	$[0.61 - 0.86]$
$t\bar{t}H(\gamma\gamma)$ CP 12	$[0.909 - 0.981]$	$[0.86 - 1]$
$t\bar{t}H(\gamma\gamma)$ CP 13	$[0.988 - 1]$	$[0 - 0.91]$
$t\bar{t}H(\gamma\gamma)$ CP 14	$[0.988 - 1]$	$[0.91 - 1]$
$t\bar{t}H(\gamma\gamma)$ CP 15	$[0.915 - 0.988]$	$[0 - 0.82]$
$t\bar{t}H(\gamma\gamma)$ CP 16	$[0.915 - 0.988]$	$[0.82 - 0.93]$
$t\bar{t}H(\gamma\gamma)$ CP 17	$[0.915 - 0.988]$	$[0.93 - 1]$
$t\bar{t}H(\gamma\gamma)$ CP 18	$[0.252 - 0.915]$	$[0 - 0.72]$
$t\bar{t}H(\gamma\gamma)$ CP 19	$[0.252 - 0.915]$	$[0.72 - 0.86]$
$t\bar{t}H(\gamma\gamma)$ CP 20	$[0.252 - 0.915]$	$[0.86 - 1]$

Table 11.1: Optimized $t\bar{t}H(\gamma\gamma)$ CP category boundaries.

$t\bar{t}H$ significance (if even)	5.00σ
$t\bar{t}H$ significance (if odd)	2.75σ
tH significance (if even)	0.321σ
tH significance (if odd)	2.23σ
$Z_{CP}(90^\circ)$	3.02σ
$Z_{CP}(45^\circ)$	1.25σ

Table 11.2: Predicted values of the figures of merit for the twenty $t\bar{t}H(\gamma\gamma)$ CP categories, calculated from event yields in the $m_{\gamma\gamma}$ region 125 ± 2 GeV. Continuum background is calculated by scaling NTI data sidebands according to Equation 5.2.

are found to be much less than δ, and are therefore expected to have negligible impact on the final results.

Category	δ	n_{ggF}	UEPS $= n^{\text{Herwig}} - n^{\text{Pythia}}$		
			$t\bar{t}H$	$tHjb$	tWH
$t\bar{t}H(\gamma\gamma)$ CP 1	1.462	0.247	-0.157	0.008	0.010
$t\bar{t}H(\gamma\gamma)$ CP 2	1.581	0.044	-0.049	-0.007	-0.005
$t\bar{t}H(\gamma\gamma)$ CP 3	1.136	0.002	0.098	0.000	0.000
$t\bar{t}H(\gamma\gamma)$ CP 4	1.275	0.208	-0.039	-0.043	-0.038
$t\bar{t}H(\gamma\gamma)$ CP 5	1.534	0.054	-0.008	-0.020	-0.006
$t\bar{t}H(\gamma\gamma)$ CP 6	1.554	0.009	0.080	-0.005	-0.007
$t\bar{t}H(\gamma\gamma)$ CP 7	1.597	0.448	-0.053	-0.020	-0.014
$t\bar{t}H(\gamma\gamma)$ CP 8	2.713	0.186	0.015	-0.010	-0.002
$t\bar{t}H(\gamma\gamma)$ CP 9	2.623	0.021	0.094	0.007	-0.001
$t\bar{t}H(\gamma\gamma)$ CP 10	1.363	0.557	-0.003	0.019	-0.012
$t\bar{t}H(\gamma\gamma)$ CP 11	3.858	1.250	-0.053	-0.038	0.026
$t\bar{t}H(\gamma\gamma)$ CP 12	7.860	0.600	0.007	-0.025	0.006
$t\bar{t}H(\gamma\gamma)$ CP 13	1.398	0.001	-0.149	-0.024	-0.006
$t\bar{t}H(\gamma\gamma)$ CP 14	1.640	0.001	0.156	- 0.008	0.001
$t\bar{t}H(\gamma\gamma)$ CP 15	1.168	0.004	-0.089	-0.024	-0.011
$t\bar{t}H(\gamma\gamma)$ CP 16	1.643	0.000	-0.015	-0.012	-0.014
$t\bar{t}H(\gamma\gamma)$ CP 17	1.697	0.001	-0.036	0.001	0.000
$t\bar{t}H(\gamma\gamma)$ CP 18	1.082	0.008	-0.016	0.032	-0.027
$t\bar{t}H(\gamma\gamma)$ CP 19	1.627	0.008	-0.024	0.022	-0.010
$t\bar{t}H(\gamma\gamma)$ CP 20	1.973	0.003	-0.014	-0.013	0.013

Table 11.3: Background-only statistical uncertainty (δ) in each $t\bar{t}H(\gamma\gamma)$ CP category compared with the 100% heavy flavor uncertainty (n_{ggF}) and the UEPS uncertainties on $t\bar{t}H$, $tHjb$, and tWH. All columns are in units of events at 139 fb^{-1} and are calculated in the range $m_{\gamma\gamma} = 125 \pm 2$ GeV.

The number-counting rejection $Z_{CP}(\alpha)$ as a function of mixing angle is shown in Figure 11.28. The 95% confidence limit is shown as a horizontal dashed line. The limit is calculated to be $\alpha = 59.0°$ when background is estimated from the NTI sidebands, and $\alpha = 64.4°$ when background is estimated from the TI sidebands.

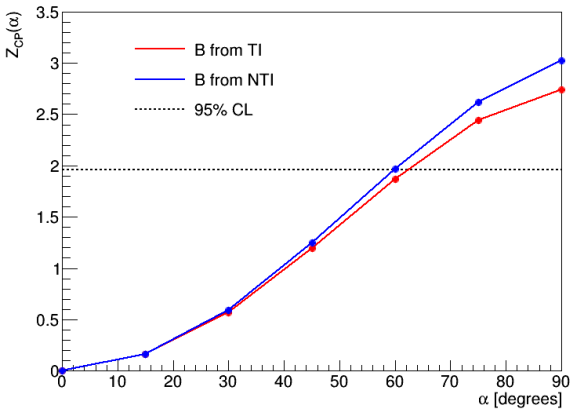


Figure 11.28: Number-counting exclusion as a function of the CP mixing angle α. Continuum background yield is estimated from NTI sidebands (blue) and TI sidebands (red).

Chapter 12

Measurement of CP properties in $t\bar{t}H$ and tH

A direct measurement of the CP properties of the Higgs-top interaction is performed with the full Run 2 ATLAS dataset of 139 fb^{-1} using the twenty $t\bar{t}H(\gamma\gamma)$ CP categories defined in the previous Chapter. This measurement focuses the $t\bar{t}H$ and tH production modes in the $H \to \gamma\gamma$ decay channel, while remaining as insensitive as possible to the gluon fusion and $H \to \gamma\gamma$ loop processes. Section 12.1 describes the impact of CP violation in Higgs-top interactions on ggF and $H \to \gamma\gamma$ and two possibilities for their treatment in the statistical analysis. The second Section describes the $m_{\gamma\gamma}$ signal and background shapes in the $t\bar{t}H(\gamma\gamma)$ CP categories, as well as the derivation of the $t\bar{t}H$ and tH signal yield dependence on the mixing parameters α and κ_t. Section 12.3 discusses the major systematics, and the results are presented in Section 12.4.

12.1 Treatment of loop processes

The contribution of the top quark to the gluon fusion and $H \to \gamma\gamma$ loops gives these processes sensitivity to CP violation in the Higgs-top interaction. However, this sensitivity is only indirect, as many other models of BSM physics predict new loop contributions that change the rates and kinematics of these processes. In order to measure the effects of CP violation in Higgs-top interactions directly, the $t\bar{t}H(\gamma\gamma)$ CP analysis aims to be as insensitive to these loop processes as possible.

The $H \to \gamma\gamma$ decay kinematics are not expected to change with CP mixing angle, and it is demonstrated in Section 12.1.1 that the shape of the ggF yield in the $t\bar{t}H(\gamma\gamma)$ CP categories is independent of the mixing angle α. As a result, the introduction of CP mixing affects only the overall rate of ggF and $H \to \gamma\gamma$ in the analysis categories. The treatment of these rates with effective loop couplings is discussed in Section 12.1.2.

12.1.1 Kinematics of gluon fusion

The ggF contribution to the hadronic $t\bar{t}H(\gamma\gamma)$ CP categories is studied in detail. The efficiency of the hadronic preselection on SM ggF is 0.5%, which is reduced to $< 0.1\%$ with the cut on the hadronic background rejection BDT discriminant at 0.909 (see Table 11.1). In the leptonic channel, ggF contamination is neglected altogether, since the efficiency of the leptonic preselection on SM ggF is $< 0.01\%$.

Figure 12.1 shows the hadronic background rejection BDT and CP BDT discriminants for multiple models of SM ggF (Powheg+Pythia8, Powheg+Herwig7, and MG5_aMC@NLO+Pythia8), as well as two CP-violating scenarios (MG5_aMC@NLO+Pythia8). Though CP mixing in the Higgs-top coupling alters some kinematic variables in ggF (such as $\Delta\phi(j,j)$) [51], the BDTs in Figure 12.1 (which are optimized on $t\bar{t}H$ and tH) are mostly insensitive to this. The background rejection BDT shape does not depend appreciably on the CP mixing angle α, and the CP BDT shape shows only a weak dependence.

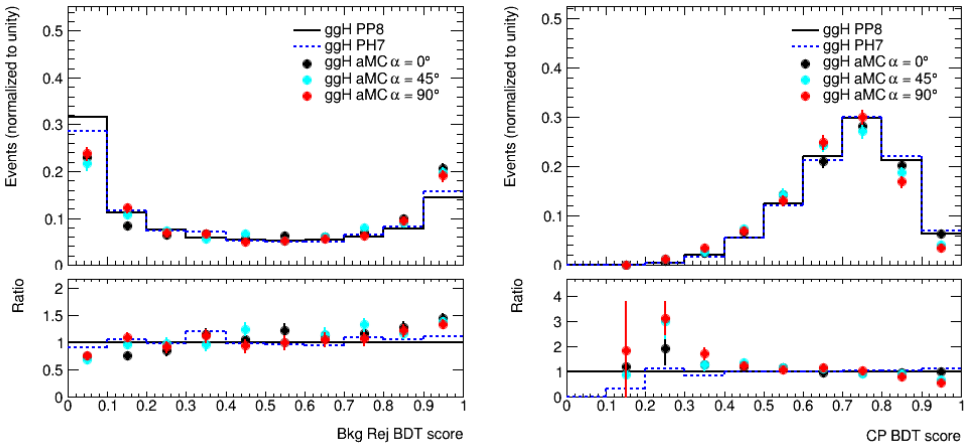


Figure 12.1: Hadronic background rejection BDT score (a) and CP BDT score (b) for ggF Monte Carlo samples.

In the statistical analysis of the $t\bar{t}H(\gamma\gamma)$ CP categories, several theoretical uncertainties are applied to allow migration of ggF events across categories. These uncertainties are derived from a comparison of SM Monte Carlo samples using different Monte Carlo generators and different parton showering algorithms (Table 12.4). With these conservative systematics applied, the shape of ggF yield in the CP BDT categories is assumed to be invariant under the mixing angle α.

12.1.2 Effective loop couplings

The overall rates of the ggF and $H \to \gamma\gamma$ processes are scaled by the square of the effective loop couplings κ_g and κ_γ, respectively:

$$\mu_{ggF} = \kappa_g^2 \tag{12.1}$$
$$\mu_{\gamma\gamma} = \kappa_\gamma^2 \tag{12.2}$$

Depending on the treatment of these loop couplings, the data are interpreted in the following two scenarios:

1. κ_g and κ_γ are constrained using data from the ATLAS combined Higgs coupling measurement [31].

2. κ_g and κ_γ are parameterized as function of κ_t and α and are varied coherently for $t\bar{t}H$, tH, ggF, and $H \to \gamma\gamma$.

Ideally, no model-dependent assumptions should be made about the ggF and $H \to \gamma\gamma$ processes: the first interpretation is therefore treated as the baseline. Details of each are included below.

Constraints from data

In order to constrain the ggF and $H \to \gamma\gamma$ processes, the dataset selected by the $t\bar{t}H(\gamma\gamma)$ CP analysis alone is insufficient. Since only one Higgs boson decay channel is selected, it is impossible to separate $H \to \gamma\gamma$ from the production mode cross section. And since the analysis rejects ggF as a background by design, the $t\bar{t}H(\gamma\gamma)$ CP categories are left with little sensitivity to this process. Therefore knowledge from other Higgs boson decay channels must be applied as well.

The ATLAS Higgs coupling combination [31] provides the measurement of κ_g and κ_γ shown in Figure 12.2. The measured best fit couplings, errors, and correlation are re-derived with all $t\bar{t}H$ categories excluded from the combination. This ensures that the dataset used to constrain the ggF and $H \to \gamma\gamma$ rates is orthogonal to the $t\bar{t}H(\gamma\gamma)$ CP categories and independent of the measured $t\bar{t}H$ cross section. In addition, the systematic uncertainties between the re-derived combination and the $t\bar{t}H(\gamma\gamma)$ CP categories can be neglected.

The re-derived constraints on κ_g and κ_γ are:

$$\kappa_g = 1.034 \pm 0.067;\ \kappa_\gamma = 0.984 \pm 0.064;\ \rho(\kappa_g, \kappa_\gamma) = -0.47 \tag{12.3}$$

Since the ggF and $H \to \gamma\gamma$ measurements are dominated by high-statistics categories, the impact of removing the $t\bar{t}H$ categories is small.

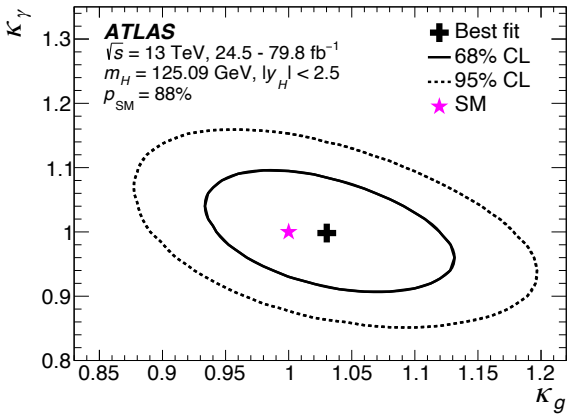


Figure 12.2: Two-dimensional likelihood scan of κ_g and κ_γ from the ATLAS Higgs coupling combination [31]. The best fit value of $(\kappa_g, \kappa_\gamma)$ is shown with 1 and 2σ contours.

Parameterize coherently with $t\bar{t}H$ and tH

Assuming that the only new physics impacting the ggF and $H \to \gamma\gamma$ loops is the model of CP violation described by Equation 2.28, the effective loop couplings κ_g and κ_γ can be directly parameterized as a function of κ_t and α. The ggF and $H \to \gamma\gamma$ rates can then be varied coherently with $t\bar{t}H$ and tH production modes.

The parameterization of κ_g and κ_γ in terms of κ_t and α is derived in [50] and can be re-parameterized in terms of κ_t and α:

$$\kappa_g^2 = \kappa_t^2 c_\alpha^2 + 2.6\kappa_t^2 s_\alpha^2 + 0.11\kappa_t c_\alpha(\kappa_t c_\alpha - 1) \tag{12.4}$$

$$\kappa_\gamma^2 = (1.28 - 0.28\kappa_t c_\alpha)^2 + (0.43\kappa_t s_\alpha)^2 \tag{12.5}$$

where s_α and c_α represent the sine and cosine of the mixing angle α, respectively.

12.2 Signal and background model

12.2.1 Signal and background shapes

In each $t\bar{t}H(\gamma\gamma)$ CP category, the signal shape is determined by fitting a Double-Sided Crystal Ball (DSCB) function to a sample of inclusive Higgs boson Monte Carlo (see Section 6.2.1). The best fit DSCB parameters for all twenty categories are shown in Table 12.1, along with S_{90}, the smallest $m_{\gamma\gamma}$ range containing 90% of Higgs signal. The signal shape parameters are derived from the inclusive SM Higgs boson Monte Carlo sample. Using $t\bar{t}H + tH$ Monte Carlo with CP mixing angle $\alpha \neq 0°$ has negligible impact on the signal shape parameterization.

The functional form of the continuum background is extracted from templates constructed from $t\bar{t}\gamma\gamma$ Monte Carlo, since this process dominates the continuum background

Category	μ_{CB} [GeV]	σ_{CB} [GeV]	α_{low}	α_{high}	n_{low}	n_{high}	S_{90} [GeV]
$t\bar{t}H(\gamma\gamma)$ CP 1	125.10	1.16	1.6	1.5	6.2	15.5	4.3
$t\bar{t}H(\gamma\gamma)$ CP 2	125.12	1.35	1.8	1.5	4.1	15.5	5.2
$t\bar{t}H(\gamma\gamma)$ CP 3	125.27	1.64	1.7	1.7	4.5	11.9	6.0
$t\bar{t}H(\gamma\gamma)$ CP 4	125.13	1.21	1.6	1.5	4.8	5.7	4.7
$t\bar{t}H(\gamma\gamma)$ CP 5	125.17	1.50	1.7	1.5	6.9	15.0	5.6
$t\bar{t}H(\gamma\gamma)$ CP 6	125.17	1.64	1.6	1.4	8.0	35.9	6.3
$t\bar{t}H(\gamma\gamma)$ CP 7	125.09	1.21	1.6	1.5	5.6	11.6	4.7
$t\bar{t}H(\gamma\gamma)$ CP 8	125.18	1.58	1.6	1.7	5.8	7.6	5.9
$t\bar{t}H(\gamma\gamma)$ CP 9	125.21	1.72	1.7	1.7	4.4	12.2	6.3
$t\bar{t}H(\gamma\gamma)$ CP 10	125.10	1.29	1.8	1.6	13.6	7.9	4.5
$t\bar{t}H(\gamma\gamma)$ CP 11	125.15	1.55	1.7	1.6	4.7	7.6	5.4
$t\bar{t}H(\gamma\gamma)$ CP 12	125.21	1.70	1.5	1.5	8.5	13.3	6.5
$t\bar{t}H(\gamma\gamma)$ CP 13	125.11	1.29	1.6	1.5	6.8	11.1	5.0
$t\bar{t}H(\gamma\gamma)$ CP 14	125.17	1.61	1.7	1.5	5.2	15.0	6.1
$t\bar{t}H(\gamma\gamma)$ CP 15	125.15	1.45	1.6	1.6	5.7	6.0	5.4
$t\bar{t}H(\gamma\gamma)$ CP 16	125.19	1.70	1.7	1.7	6.7	14.4	6.2
$t\bar{t}H(\gamma\gamma)$ CP 17	125.20	1.71	1.5	1.5	7.9	14.4	6.7
$t\bar{t}H(\gamma\gamma)$ CP 18	125.03	1.68	1.8	1.6	18.7	12.7	5.7
$t\bar{t}H(\gamma\gamma)$ CP 19	125.17	1.76	2.2	1.5	1.4	10.2	6.6
$t\bar{t}H(\gamma\gamma)$ CP 20	125.14	1.78	1.6	1.6	10.8	14.7	6.6

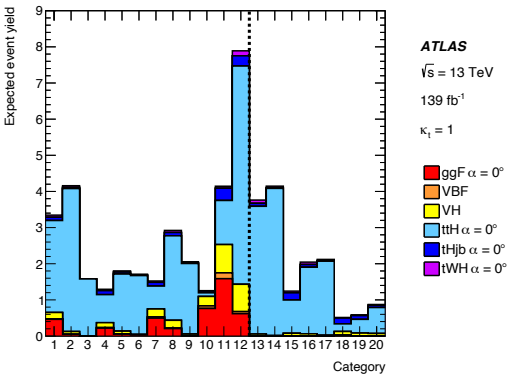
Table 12.1: Best fit parameter values for the signal $m_{\gamma\gamma}$ shape in each $t\bar{t}H(\gamma\gamma)$ CP category. Note that these are the fitted values for the MC samples at $m_H = 125$ GeV.

once the BDT selection has been applied. Because the $t\bar{t}\gamma\gamma$ background template in category 10 has exceptionally high statistical uncertainty ($\Delta_{MC} = 3.06$), the background shape and spurious signal are derived instead from a template built from NTI data ($\Delta_{MC} = 1.91$). In all other categories, the $t\bar{t}\gamma\gamma$ Monte Carlo template has a smaller statistical uncertainty than the NTI background template.

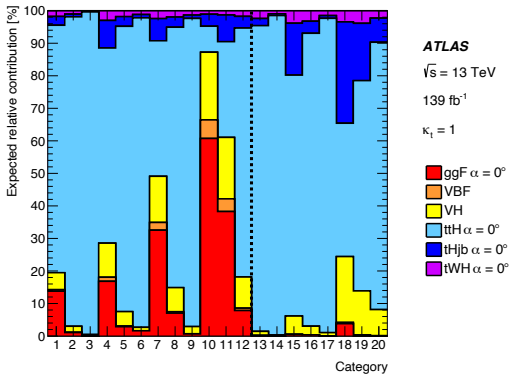
The functional form of the continuum background and the spurious signal are listed in Table 12.2.

12.2.2 Signal yield dependence on mixing angle

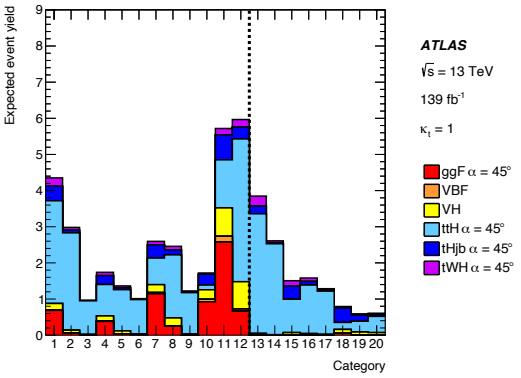
Event yields and signal purities in the twenty CP BDT categories are shown in Figure 12.3 for $\alpha = 0°$ (top row), $\alpha = 45°$ (middle row), and $\alpha = 90°$ (bottom row). The ggF event yields shown are from the MG5_aMC@NLO+Pythia8 Monte Carlo samples in which the ggF rate changes coherently with the CP mixing angle α. The changing composition of the $t\bar{t}H(\gamma\gamma)$ CP categories is clearly visible, particularly the increase in the tH rate with α.



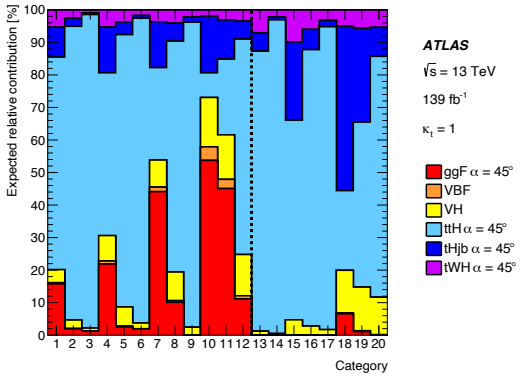
(a) $\alpha = 0°$



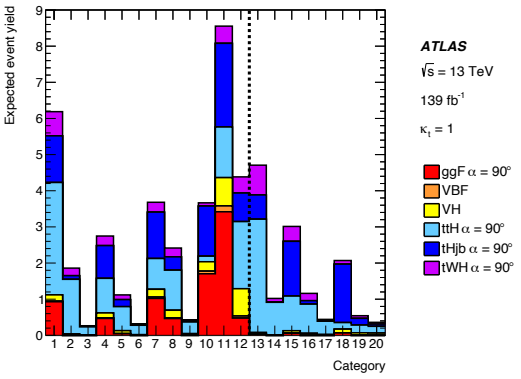
(b) $\alpha = 0°$



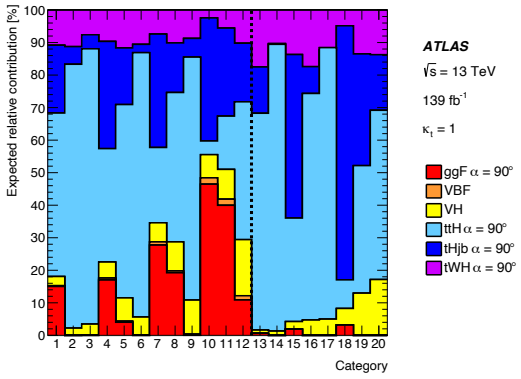
(c) $\alpha = 45°$



(d) $\alpha = 45°$



(e) $\alpha = 90°$



(f) $\alpha = 90°$

Figure 12.3: (Left) Higgs event yields in the $t\bar{t}H(\gamma\gamma)$ CP categories at 139 fb^{-1} at different values of the CP mixing angle α. (Right) Purity of the Higgs yield in each category for different values of α. Yields are calculated in the smallest $m_{\gamma\gamma}$ window containing 90% of the fitted signal.

Category	Function	N_{sp}	χ^2/ndof	Prob(χ^2) [%]
$t\bar{t}H(\gamma\gamma)$ CP 1	Exponential	+0.426	0.621	90
$t\bar{t}H(\gamma\gamma)$ CP 2	Power Law	−0.297	0.71	82
$t\bar{t}H(\gamma\gamma)$ CP 3	Exponential	+0.260	1.69	3
$t\bar{t}H(\gamma\gamma)$ CP 4	Exponential	+0.311	0.672	74
$t\bar{t}H(\gamma\gamma)$ CP 5	Power Law	−0.566	1.15	29
$t\bar{t}H(\gamma\gamma)$ CP 6	Exponential	+0.335	0.573	94
$t\bar{t}H(\gamma\gamma)$ CP 7	Exponential	+0.644	1.91	1
$t\bar{t}H(\gamma\gamma)$ CP 8	Exponential	−1.23	0.677	85
$t\bar{t}H(\gamma\gamma)$ CP 9	Power Law	−0.264	0.742	79
$t\bar{t}H(\gamma\gamma)$ CP 10	Power Law	−1.31	0.795	62
$t\bar{t}H(\gamma\gamma)$ CP 11	Exponential	+4.50	1.18	26
$t\bar{t}H(\gamma\gamma)$ CP 12	Exponential	+3.05	0.981	48
$t\bar{t}H(\gamma\gamma)$ CP 13	Exponential	+0.199	0.981	48
$t\bar{t}H(\gamma\gamma)$ CP 14	Exponential	+0.129	0.996	46
$t\bar{t}H(\gamma\gamma)$ CP 15	Power Law	−0.205	1.06	39
$t\bar{t}H(\gamma\gamma)$ CP 16	Exponential	+0.398	0.841	66
$t\bar{t}H(\gamma\gamma)$ CP 17	Power Law	−0.393	1.38	12
$t\bar{t}H(\gamma\gamma)$ CP 18	Exponential	+0.761	0.91	57
$t\bar{t}H(\gamma\gamma)$ CP 19	Power Law	−0.498	0.718	81
$t\bar{t}H(\gamma\gamma)$ CP 20	Exponential	−0.135	0.618	90

Table 12.2: Results of the spurious signal test for the $t\bar{t}H(\gamma\gamma)$ CP analysis categories. The χ^2/ndof indicates the performance of a background only fit to the template in the $m_{\gamma\gamma}$ range 105-160 GeV, but is in general not a part of the spurious signal criteria.

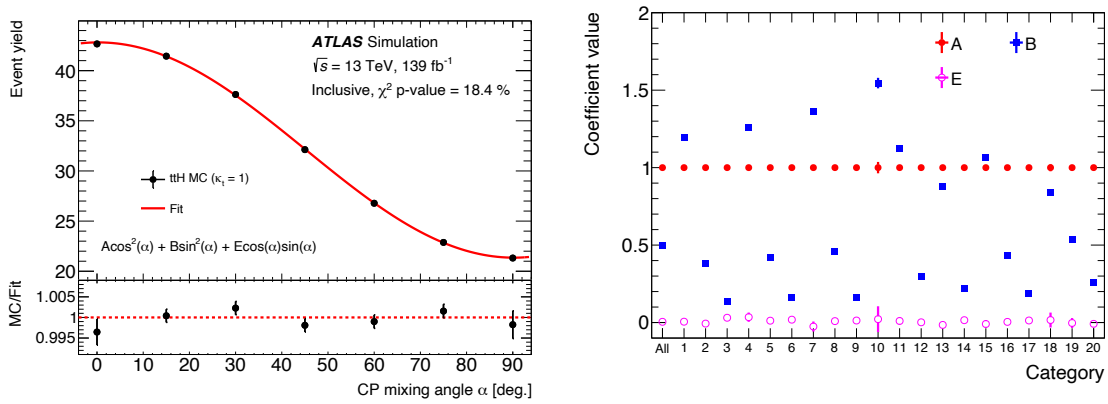
To measure the mixing angle α and coupling strength κ_t, the signal yield must be parameterized as function of these two parameters. The signal yield in each category takes the same form as the $t\bar{t}H$ and tH cross sections in Chapter 2, with the parameters fitted independently in each category.

The $t\bar{t}H$ yield in category i is parameterized using the form of the cross section in Equation 2.29:

$$y_i = A_i\kappa_t^2 c_\alpha^2 + B_i\kappa_t^2 s_\alpha^2 + E_i\kappa_t^2 s_\alpha c_\alpha \tag{12.6}$$

where s_α and c_α represent the sine and cosine of α, respectively. For a discussion of the diagrams corresponding to each term, refer to Section 2.4.1. The coefficients A, B, and E are determined in each category by fitting to the signal yield predicted by Monte Carlo sample at different mixing angles.

The inclusive parameterization of the $t\bar{t}H$ signal yield (sum of all twenty $t\bar{t}H(\gamma\gamma)$ CP categories) is shown in Figure 12.4a, while Figure 12.4b shows the fitted values of the coefficients A, B, and E in each category. As expected, interference between CP even and CP odd (coefficient E) is small.



(a) Inclusive $t\bar{t}H$ fit

(b) $t\bar{t}H$ coefficients

Figure 12.4: (Left) Parameterization as function of the CP mixing angle of the inclusive $t\bar{t}H$ signal yield in the $t\bar{t}H(\gamma\gamma)$ CP categories. The yields are normalized to 139 fb^{-1}. (Right) Fitted parameterization coefficients for $t\bar{t}H$. The coefficient values have been normalized to SM.

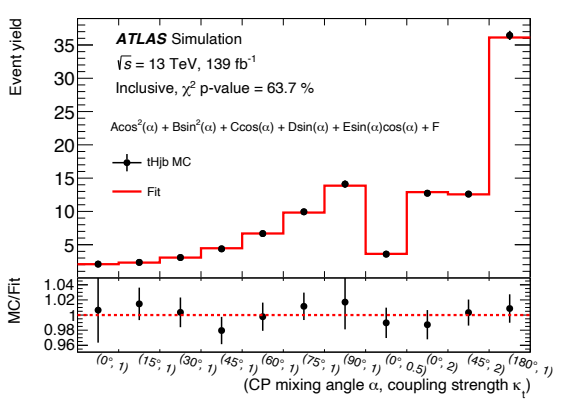
For $tHjb$ and tWH, the interference between Higgs couplings to the top quark and the W boson must also be taken into account. The $tHjb$ and tWH yield in category i is parameterized using the form of the cross section in Equation 2.30 (with different values of the coefficients for $tHjb$ and tWH):

$$y_i = A_i\kappa_t^2 c_\alpha^2 + B_i\kappa_t^2 s_\alpha^2 + C_i\kappa_t c_\alpha + D_i\kappa_t s_\alpha + E_i\kappa_t^2 s_\alpha c_\alpha + F_i \tag{12.7}$$

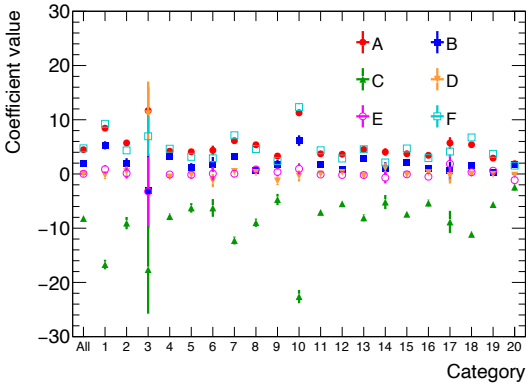
For a discussion of the diagrams corresponding to each term, refer to Section 2.4.1. The coefficients A, B, C, D, E and F are determined in each category by fitting to the signal yield predicted by Monte Carlo. In the case of tH, the fit is performed over both the mixing angle α and the coupling strength κ_t in order to break the degeneracy among the parameters A, B, and F. Several $tHjb$ and tWH Monte Carlo samples with $\kappa_t \neq 1$ are generated for this purpose (see Section 5.2.2 for details).

The inclusive parameterizations of the $tHjb$ and tWH signal yield are shown in Figures 12.5a and 12.6a, respectively. Figures 12.5b and 12.6b show the fitted values of the coefficients A, B, C, D, E, and F in each category. As expected, the interference between the CP even Higgs-top and Higgs-W couplings (coefficient C) is large and negative.

The signal parametrization is summarized in Figure 12.7 for all processes whose rates depend on the CP mixing angle α. The parameterizations of ggF and $H \to \gamma\gamma$ from Equation 12.4 are shown as dashed lines.

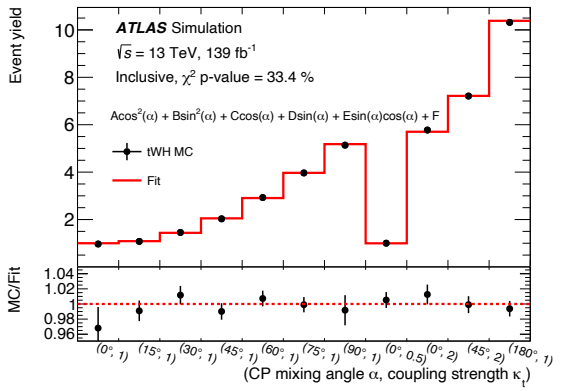


(a) Inclusive $tHjb$ fit

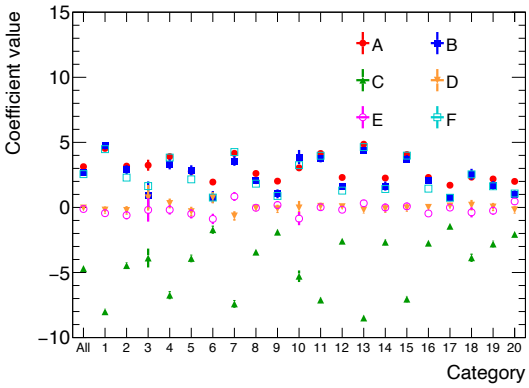


(b) $tHjb$ coefficients

Figure 12.5: (Left) Parameterization as function of the CP mixing angle of the inclusive $tHjb$ signal yield in the $t\bar{t}H(\gamma\gamma)$ CP categories. The yields are normalized to 139 fb^{-1}. (Right) Fitted parameterization coefficients for $tHjb$. The coefficient values have been normalized to SM.



(a) Inclusive tWH fit



(b) tWH cofficients

Figure 12.6: (Left) Parameterization as function of the CP mixing angle of the inclusive tWH signal yield in the $t\bar{t}H(\gamma\gamma)$ CP categories. The yields are normalized to 139 fb^{-1}. (Right) Fitted parameterization coefficients for tWH. The coefficient values have been normalized to SM.

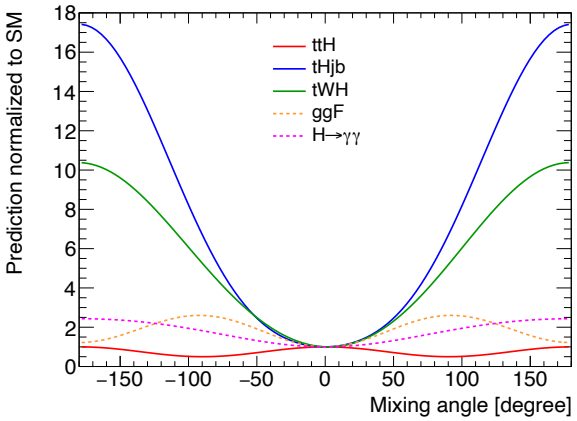


Figure 12.7: Inclusive parametrization of all processes dependent on the CP mixing angle α, with $\kappa_t = 1$. The yields are normalized to the Standard Model prediction.

12.3 Systematic uncertainties

The experimental systematics calculated for the $t\bar{t}H(\gamma\gamma)$ CP analysis are those summarized in Table 6.3. The theory systematics in Table 6.2 are calculated per process and per category for the $t\bar{t}H(\gamma\gamma)$ CP analysis, with magnitudes summarized in this Section.

The QCD and PDF uncertainties are calculated in each category for each signal process ($t\bar{t}H$, $tHjb$, and tWH). The magnitudes of these uncertainties are shown in Table 12.3. The PDF uncertainties on $t\bar{t}H$, $tHjb$, and tWH are calculated according to the LHAPDF [13] prescription and treated as a single correlated nuisance parameter. In the case of ggF, VBF, and VH, a 100% heavy flavor uncertainty is applied to the normalization of each process to account for the uncertainty on the total cross section.

The Underlying Event/Parton Showering (UEPS) systematic is calculated by comparing the yield between SM samples interfaced with two different showering algorithms: Pythia8 and Herwig7. The UEPS uncertainty is calculated separately for each category and for each process ($t\bar{t}H$, $tHjb$, tWH, and ggF): the matrix element generator used for this comparison is MG5_aMC@NLO for $t\bar{t}H$ and tH, and Powheg for ggF.

For $t\bar{t}H$ and ggF, and additional Monte Carlo generator uncertainty is applied, which accounts for differences in the predicted event yield per category from different matrix element calculations. For $t\bar{t}H$, this comparison is performed between the nominal MG5_aMC@NLO+Pythia8 and an alternative Powheg+Pythia8 sample. For ggF, the nominal sample is generated with Powheg+Pythia8 and the alternative with MG5_aMC@NLO+Pythia8.

	$t\bar{t}H$ [%]			$tHjb$ [%]			tWH [%]		
Category	QCD+	QCD−	PDF	QCD+	QCD−	PDF	QCD+	QCD−	PDF
$t\bar{t}H(\gamma\gamma)$ CP 1	6.0	9.4	2.0	7.0	7.7	2.3	2.6	4.8	2.8
$t\bar{t}H(\gamma\gamma)$ CP 2	5.5	8.9	1.6	3.6	4.4	2.6	9.2	15.5	2.9
$t\bar{t}H(\gamma\gamma)$ CP 3	2.8	7.4	1.5	13.1	10.8	1.3	6.4	39.7	2.6
$t\bar{t}H(\gamma\gamma)$ CP 4	7.7	10.3	2.2	11.6	10.5	1.3	1.7	2.5	2.8
$t\bar{t}H(\gamma\gamma)$ CP 5	6.1	9.2	1.7	8.3	8.1	3.9	5.2	9.1	2.3
$t\bar{t}H(\gamma\gamma)$ CP 6	5.8	8.9	1.4	1.2	2.6	2.4	12.5	19.7	2.0
$t\bar{t}H(\gamma\gamma)$ CP 7	9.1	10.9	2.0	6.8	7.5	2.5	2.1	2.7	3.0
$t\bar{t}H(\gamma\gamma)$ CP 8	6.2	9.2	1.6	4.4	5.7	7.0	4.8	7.4	2.5
$t\bar{t}H(\gamma\gamma)$ CP 9	6.2	9.1	1.5	3.5	4.2	5.0	9.0	14.0	2.4
$t\bar{t}H(\gamma\gamma)$ CP 10	9.5	10.9	2.4	7.0	7.8	5.5	4.7	4.6	3.3
$t\bar{t}H(\gamma\gamma)$ CP 11	7.3	9.9	1.9	7.1	7.7	1.3	2.1	2.6	2.9
$t\bar{t}H(\gamma\gamma)$ CP 12	6.0	9.0	1.5	2.8	4.7	1.0	3.8	4.5	2.2
$t\bar{t}H(\gamma\gamma)$ CP 13	6.8	9.7	1.7	2.8	5.1	2.2	3.6	4.6	2.7
$t\bar{t}H(\gamma\gamma)$ CP 14	6.7	9.4	1.5	0.6	2.1	1.5	13.9	21.9	2.1
$t\bar{t}H(\gamma\gamma)$ CP 15	6.6	9.5	1.7	9.7	9.4	2.1	1.8	2.6	2.8
$t\bar{t}H(\gamma\gamma)$ CP 16	4.5	8.2	1.4	5.2	6.2	1.8	4.5	3.8	2.0
$t\bar{t}H(\gamma\gamma)$ CP 17	4.8	7.9	1.3	5.4	5.7	1.5	2.8	3.7	2.1
$t\bar{t}H(\gamma\gamma)$ CP 18	6.1	8.9	1.4	7.2	7.9	1.6	3.7	3.7	2.9
$t\bar{t}H(\gamma\gamma)$ CP 19	4.7	8.1	1.3	7.7	8.0	1.3	1.2	1.0	2.2
$t\bar{t}H(\gamma\gamma)$ CP 20	4.0	7.1	1.4	2.3	3.7	3.5	4.8	7.0	2.1

Table 12.3: Theory systematics impacting the overall rate of signal processes ($t\bar{t}H$, $tHjb$, and tWH): the relative QCD scale uncertainties and PDF uncertainties (shown in %).

	$t\bar{t}H$ [%]		$tHjb$ [%]	tWH [%]	ggF [%]	
Category	MC gen	UEPS	UEPS	UEPS	MC gen	UEPS
$t\bar{t}H(\gamma\gamma)$ CP 1	+1.4	−6.3	+9.1	+10.6	−87.0	−6.7
$t\bar{t}H(\gamma\gamma)$ CP 2	−4.3	−1.4	−19.5	−13.3	−51.5	+53.7
$t\bar{t}H(\gamma\gamma)$ CP 3	−9.7	+7.4	0.0	−7.4	0.0	−50.2
$t\bar{t}H(\gamma\gamma)$ CP 4	+2.5	−5.4	−40.7	−36.1	−16.0	+30.7
$t\bar{t}H(\gamma\gamma)$ CP 5	−5.6	−0.6	−45.2	−15.0	+28.1	+27.3
$t\bar{t}H(\gamma\gamma)$ CP 6	−7.2	+6.1	−24.1	−35.8	−77.2	−96.7
$t\bar{t}H(\gamma\gamma)$ CP 7	+5.3	−8.8	−21.5	−14.3	−12.4	−1.5
$t\bar{t}H(\gamma\gamma)$ CP 8	−5.7	+0.8	−12.3	−2.8	+0.4	+103.5
$t\bar{t}H(\gamma\gamma)$ CP 9	−7.8	+6.0	+47.9	−5.4	+7.0	+20.7
$t\bar{t}H(\gamma\gamma)$ CP 10	−1.8	−2.7	+44.1	−26.7	−38.2	+29.2
$t\bar{t}H(\gamma\gamma)$ CP 11	+2.2	−5.0	−12.0	+8.1	−18.8	+29.2
$t\bar{t}H(\gamma\gamma)$ CP 12	−3.1	+0.1	−11.1	+2.6	+19.0	+58.8
$t\bar{t}H(\gamma\gamma)$ CP 13	+3.5	−4.6	−35.0	−8.2	−96.9	+9.6
$t\bar{t}H(\gamma\gamma)$ CP 14	−2.6	+4.6	−37.3	+2.4	0.0	+58.6
$t\bar{t}H(\gamma\gamma)$ CP 15	+8.8	−10.9	−14.1	−6.5	−35.6	+7.1
$t\bar{t}H(\gamma\gamma)$ CP 16	+0.5	−1.0	−19.2	−23.9	0.0	0.0
$t\bar{t}H(\gamma\gamma)$ CP 17	+0.9	−2.2	+35.4	+15.7	0.0	−100.0
$t\bar{t}H(\gamma\gamma)$ CP 18	+7.5	−9.1	+21.5	−18.4	−203.2	+111.7
$t\bar{t}H(\gamma\gamma)$ CP 19	+5.6	−7.9	+27.7	−13.1	+94.3	−73.6
$t\bar{t}H(\gamma\gamma)$ CP 20	+3.4	−2.5	−24.0	+25.3	−86.6	−12.5

Table 12.4: Theory systematics impacting the migration of $t\bar{t}H$, $tHjb$, tWH, and ggF events across categories: the relative UEPS uncertainties and Monte Carlo generator uncertainties (shown in %).

12.4 Results

The diphoton mass spectra of the data in each of the twenty $t\bar{t}H(\gamma\gamma)$ CP categories are shown in Figures 12.8-12.10. Figure 12.8 shows the most CP even-like categories (those with high CP BDT score: 3, 6, 9, 12, 14, 17, 20), Figure 12.9 shows the most CP odd-like categories (those with low CP BDT score: 1, 4, 7, 10, 13, 16, 19), and Figure 12.10 shows the categories with intermediate CP BDT score (2, 5, 8, 11, 15, 18).

The reconstructed Higgs boson and top quark masses in data are shown in Figure 12.11. The sum of all categories is shown with data in each category weighted by $\ln(1+S/B)$, where S (B) is the fitted signal (background) in the smallest $m_{\gamma\gamma}$ window containing 90% of the signal. The fitted curves in the projection are similarly weighted. This weighting procedure is used to visually emphasize the contribution of the most powerful analysis categories (that is, the categories with highest signal-to-background ratio). A description of the calculation of errors on this plot is included in Section 6.5. A clear peak is visible at the Higgs boson mass and top quark mass, indicating the presence of both particles in the events in the most sensitive categories.

12.4.1 Sensitivity to SM $t\bar{t}H$ and tH

The sensitivity to $t\bar{t}H$ is evaluated by setting $\alpha = 0^\circ$ and fixing the yield of all non-$t\bar{t}H$ production processes (including $tHjb$ and tWH) to the SM expectation. Theory and experimental uncertainties are retained on all event yields. The data is compared to the null hypothesis where $\mu_{t\bar{t}H} = 0$. The observed $t\bar{t}H$ significance is 5.2σ (compared to 4.4σ expected), and the measured signal strength is

$$\mu_{t\bar{t}H} = 1.43\ ^{+0.39}_{-0.34} = 1.43\ ^{+0.33}_{-0.31}\ (\text{stat.})\ ^{+0.13}_{-0.11}\ (\text{exp.})\ ^{+0.17}_{-0.09}\ (\text{th.}) \tag{12.8}$$

An upper limit on the cross section of tH production ($tHjb+tWH$) is derived by setting $\alpha = 0^\circ$ and fixing the cross sections of all Higgs boson processes except $t\bar{t}H$, $tHjb$, and tWH to their SM expectations. A fit is performed where the tH and $t\bar{t}H$ signal strengths are left free.

The limit on the tH cross section is calculated using the CL_s procedure [156]. The observed upper limit on the tH cross section is 11.6 times the SM expectation (compared to 11.7 expected). At the time of writing, this limit is the strongest available on the tH process [157].

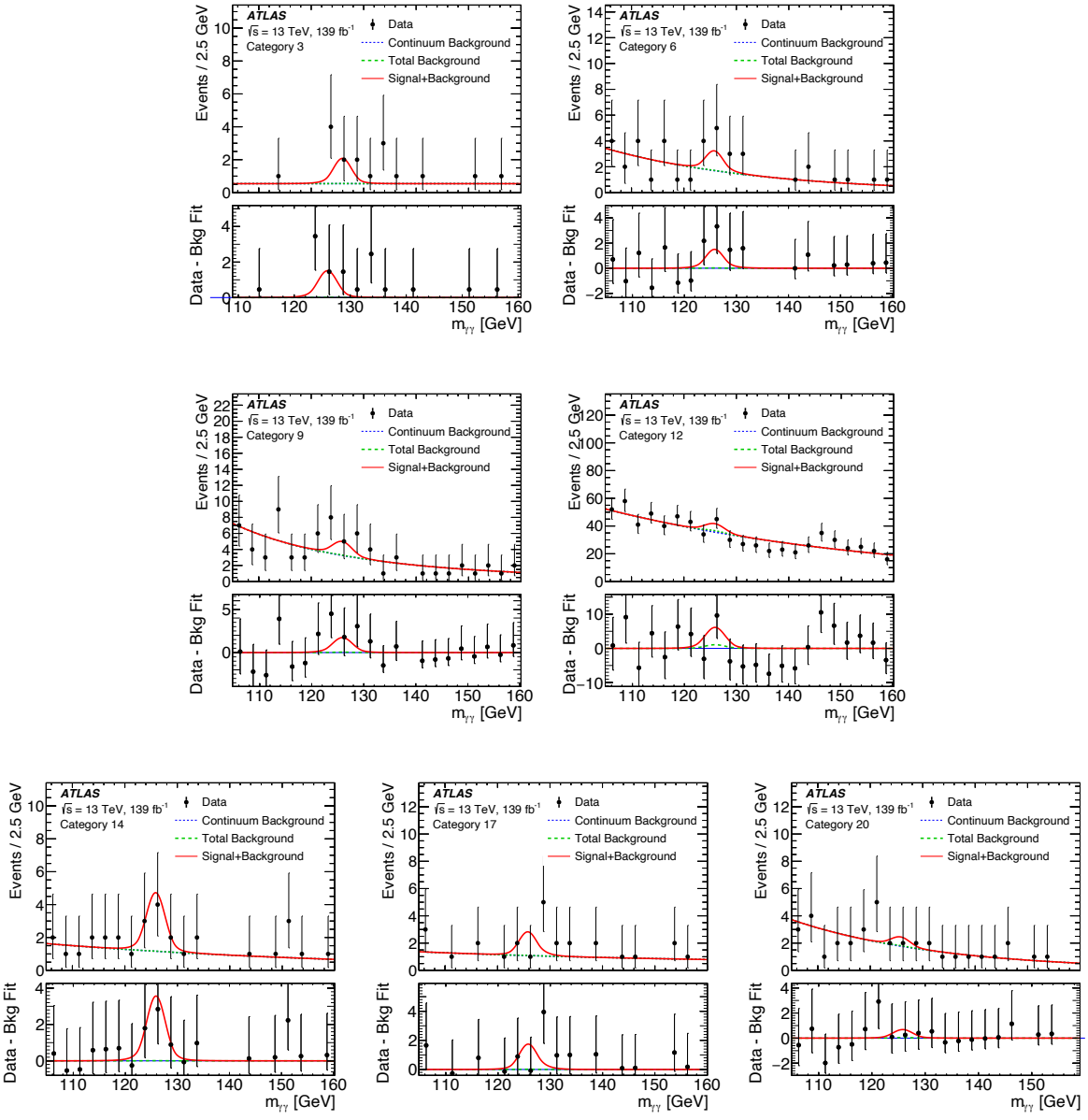


Figure 12.8: Diphoton invariant mass spectrum in the $t\bar{t}H(\gamma\gamma)$ CP categories with high CP BDT score (CP even-like). The signal-plus-background fit is overlaid in red, the total background in dashed green, and the continuum background in dashed blue.

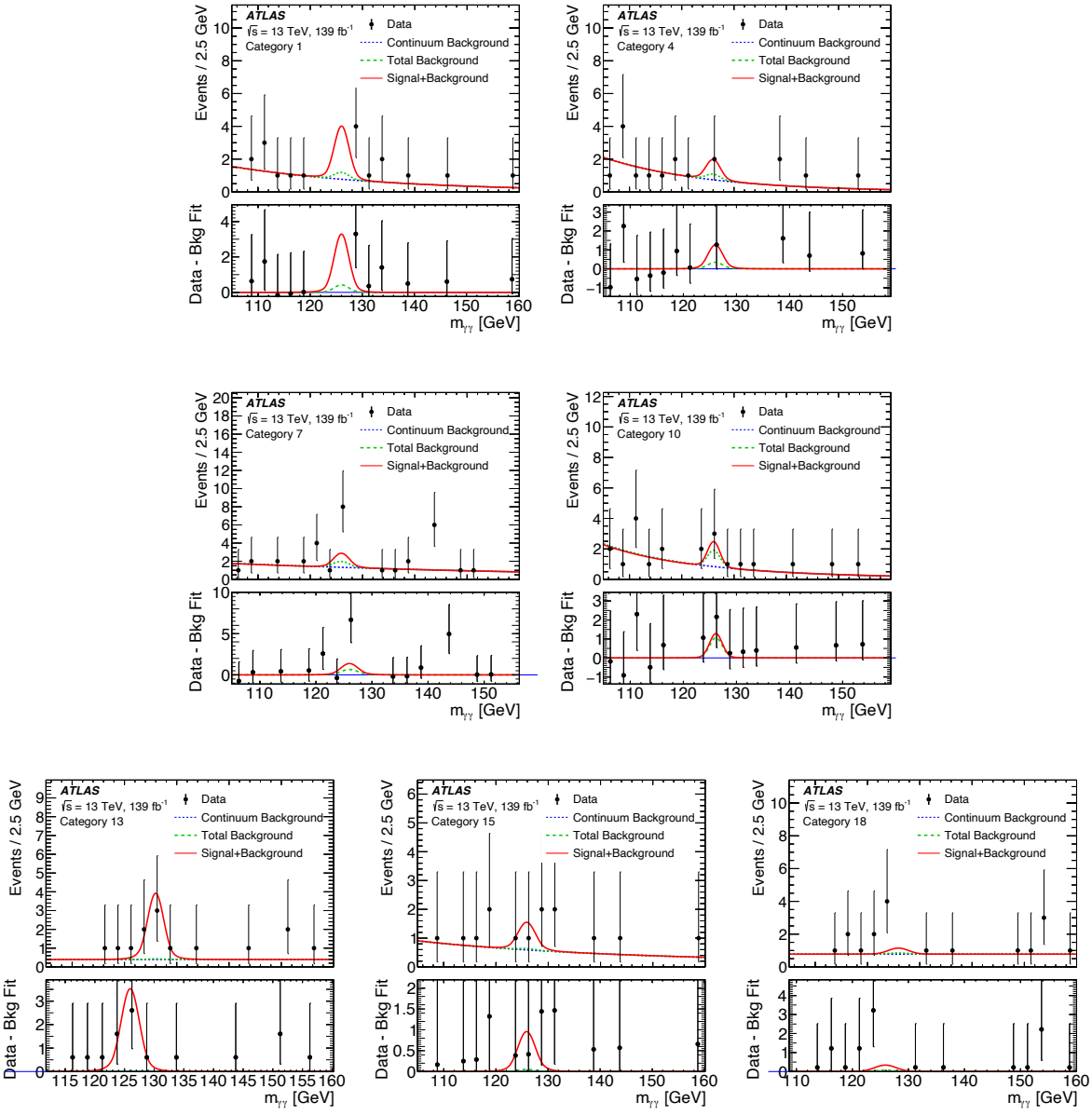


Figure 12.9: Diphoton invariant mass spectrum in the $t\bar{t}H(\gamma\gamma)$ CP categories with low CP BDT score (CP odd-like). The signal-plus-background fit is overlaid in red, the total background in dashed green, and the continuum background in dashed blue.

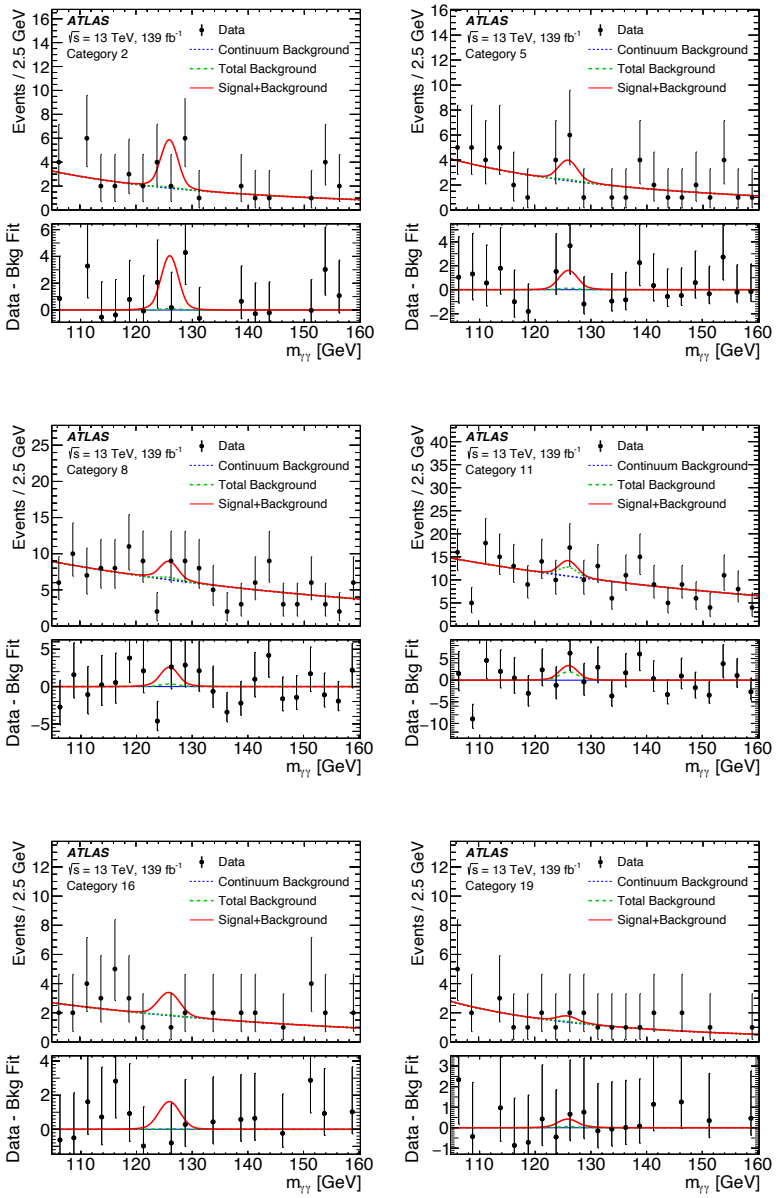


Figure 12.10: Diphoton invariant mass spectrum in the $t\bar{t}H(\gamma\gamma)$ CP categories with intermediate CP BDT score. The signal-plus-background fit is overlaid in red, the total background in dashed green, and the continuum background in dashed blue.

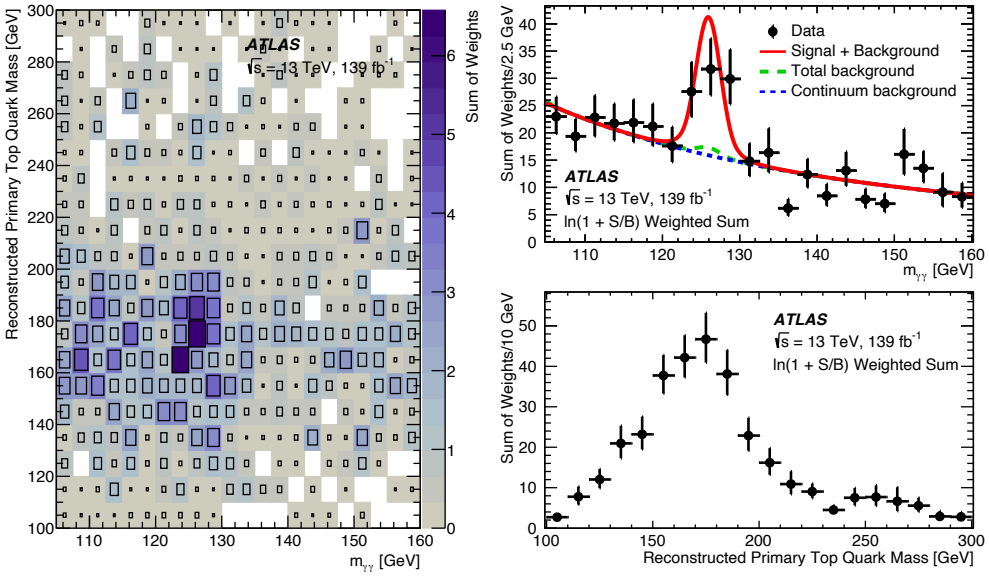


Figure 12.11: Distribution of reconstructed top quark mass and reconstructed Higgs boson mass in the $t\bar{t}H(\gamma\gamma)$ CP analysis categories. The right panels show the projections onto the Higgs boson mass and top quark mass axes. In the upper panel, the fitted continuum background (blue), the total fitted background including non-$t\bar{t}H/tH$ Higgs production (green), and the signal-plus-background fit (red) are shown. All categories are shown, with data in each category weighted by $\ln(1 + S/B)$ of the category. The fitted curves in the projection are similarly weighted.

12.4.2 Constraints on CP mixing

A fit is performed to the data in all twenty $t\bar{t}H(\gamma\gamma)$ CP categories where the CP mixing angle α and the coupling strength κ_t are left free, and the effective loop couplings κ_g and κ_γ are constrained by combined ATLAS Higgs data (Equation 12.3). The fitted signal and background yields are shown in Figure 12.12, calculated in the smallest $m_{\gamma\gamma}$ window containing 90% of the $t\bar{t}H + tH$ signal. The bottom panel compares the fitted signal and data minus the fitted background. Using the method described in [158], the compatibility between the observed data and the fit in Figure 12.12 is found to have a p-value of 35%.

Table 12.5 shows the expected signal yield for three CP mixing hypotheses: Standard Model ($\alpha = 0°$, $\kappa_t = 1$), maximal CP mixing ($\alpha = 45°$, $\kappa_t = 1$), and fully CP odd ($\alpha = 90°$, $\kappa_t = 1$). The expected background yields and observed data are also shown.

A comparison is performed between the data-minus-background and the extracted signal from different CP mixing angles. These fits are performed with α fixed and κ_t free. In order to obtain high enough statistics for a visual interpretation, categories are merged in to three bins: CP even-like (categories 3, 6, 9, 12, 14, 17, 20), CP odd-like (1, 4, 7, 10, 13, 16, 19), intermediate (2, 5, 8, 11, 15, 18). The results are shown in Figure 12.13. The data clearly favor the $\alpha = 0°$ hypothesis over $\alpha = 90°$.

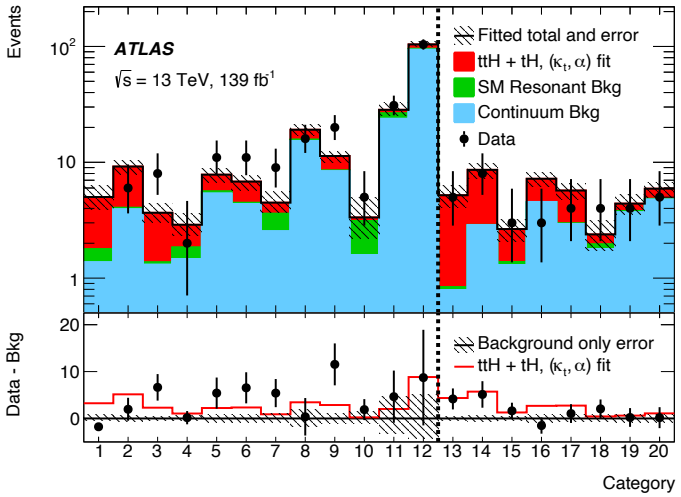


Figure 12.12: Fitted signal and background yields in the $t\bar{t}H(\gamma\gamma)$ CP analysis categories, calculated in the smallest $m_{\gamma\gamma}$ window containing 90% of the $t\bar{t}H + tH$ signal. Note that the uncertainties in the hashed bands are correlated with the statistical error on the data. The bottom panel compares the fitted signal and data minus the fitted background.

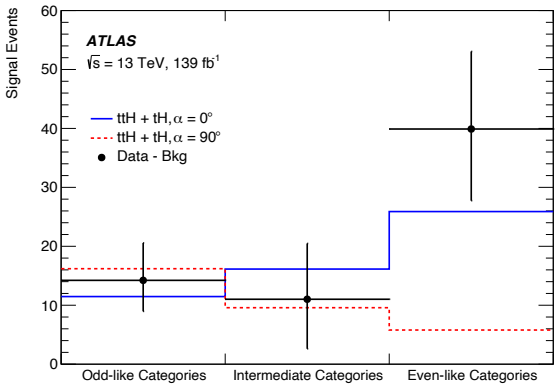


Figure 12.13: Extracted $t\bar{t}H+tH$ signal yields in the $t\bar{t}H(\gamma\gamma)$ CP categories for $\alpha = 0°$ (red), $\alpha = 90°$ (blue), and in data. Yields are calculated in the smallest $m_{\gamma\gamma}$ interval containing 90% of the signal. The twenty categories are merged into three bins: CP odd-like, Intermediate, and CP even-like. $t\bar{t}H + tH$ signal yields are extracted fits performed with α fixed and κ_t free. The data is shown with fitted background subtracted.

Category	Expected $\kappa_t = 1$, $\alpha = 0°$		Expected $\kappa_t = 1$, $\alpha = 45°$		Expected $\kappa_t = 1$, $\alpha = 90°$		Resonant Background	Continuum Background	Data
	$t\bar{t}H$	tH	$t\bar{t}H$	tH	$t\bar{t}H$	tH			
$t\bar{t}H(\gamma\gamma)$ CP 1	2.1	0.14	2.3	0.55	2.5	1.67	0.40	1.4	0
$t\bar{t}H(\gamma\gamma)$ CP 2	3.5	0.067	2.4	0.12	1.3	0.377	0.12	3.9	6
$t\bar{t}H(\gamma\gamma)$ CP 3	1.6	0.007	0.93	0.015	0.21	0.040	0.060	1.3	8
$t\bar{t}H(\gamma\gamma)$ CP 4	0.63	0.12	0.72	0.28	0.79	0.85	0.37	1.5	2
$t\bar{t}H(\gamma\gamma)$ CP 5	1.5	0.063	1.1	0.091	0.63	0.28	0.23	5.4	11
$t\bar{t}H(\gamma\gamma)$ CP 6	1.6	0.026	0.95	0.023	0.26	0.071	0.10	4.4	11
$t\bar{t}H(\gamma\gamma)$ CP 7	0.51	0.12	0.60	0.39	0.70	1.2	1.1	2.5	9
$t\bar{t}H(\gamma\gamma)$ CP 8	2.3	0.12	1.7	0.22	1.04	0.59	0.42	15.2	16
$t\bar{t}H(\gamma\gamma)$ CP 9	1.9	0.046	1.14	0.045	0.31	0.11	0.11	8.3	20
$t\bar{t}H(\gamma\gamma)$ CP 10	0.094	0.064	0.12	0.30	0.15	1.0	1.5	1.6	5
$t\bar{t}H(\gamma\gamma)$ CP 11	1.1	0.38	1.2	0.80	1.22	2.3	2.6	23.7	31
$t\bar{t}H(\gamma\gamma)$ CP 12	5.8	0.39	3.7	0.52	1.72	1.4	1.8	93.5	104
$t\bar{t}H(\gamma\gamma)$ CP 13	2.9	0.15	2.7	0.42	2.55	1.2	0.042	0.79	5
$t\bar{t}H(\gamma\gamma)$ CP 14	3.9	0.052	2.4	0.073	0.86	0.17	0.013	2.9	8
$t\bar{t}H(\gamma\gamma)$ CP 15	0.73	0.19	0.75	0.46	0.78	1.3	0.075	1.3	3
$t\bar{t}H(\gamma\gamma)$ CP 16	1.7	0.13	1.3	0.17	0.76	0.50	-0.030	4.5	3
$t\bar{t}H(\gamma\gamma)$ CP 17	1.9	0.048	1.1	0.057	0.35	0.11	0.024	2.9	4
$t\bar{t}H(\gamma\gamma)$ CP 18	0.18	0.15	0.17	0.36	0.15	1.2	0.17	1.8	4
$t\bar{t}H(\gamma\gamma)$ CP 19	0.32	0.12	0.25	0.18	0.17	0.45	0.092	3.7	4
$t\bar{t}H(\gamma\gamma)$ CP 20	0.68	0.079	0.43	0.079	0.18	0.21	0.076	4.8	5

Table 12.5: Observed and expected $t\bar{t}H$ and $tH = tHjb + tWH$ yields per $t\bar{t}H(\gamma\gamma)$ CP category, calculated in the smallest $m_{\gamma\gamma}$ window containing 90% of the fitted signal. The expected yields assume $\kappa_t = 1$ for various α.

The two-dimensional likelihood scan of $\kappa_t \cos\alpha$ and $\kappa_t \sin\alpha$ is shown in Figure 12.14a. The one-dimensional likelihood scan over CP mixing angle α (with the coupling strength κ_t profiled) is shown in Figure 12.14b. Systematic uncertainties have negligible impact on the limit. Assuming that the SM (CP even) hypothesis is true, the expected 95% confidence limit is $|\alpha| > 63°$, and the observed data exclude $|\alpha| > 43°$ at 95% confidence level. The probability to observe a limit on α that is stronger than the expectation is 11% (evaluated on 500 toys).

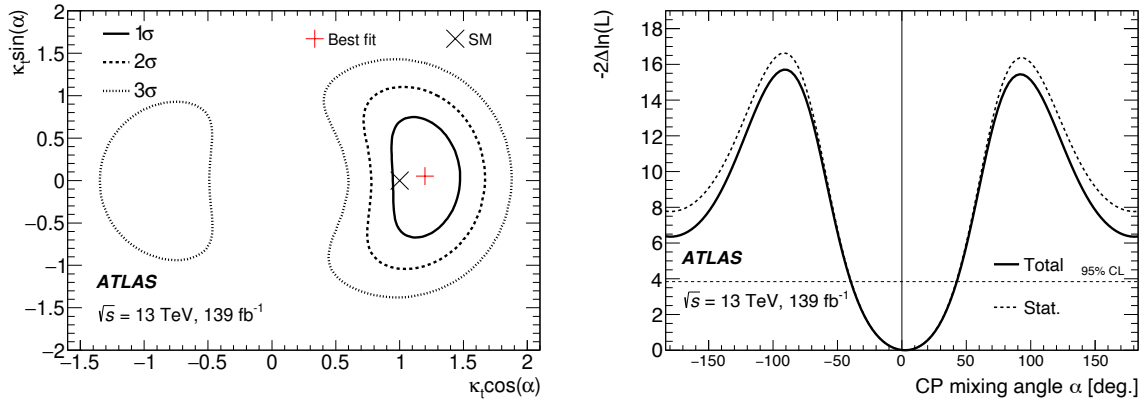


Figure 12.14: (Left) The 1, 2, and 3σ exclusion contours in the ($\kappa_t \cos\alpha$, $\kappa_t \sin\alpha$) plane, calculated with κ_g and κ_γ constrained by ATLAS combined Higgs boson data. (Right) One-dimensional likelihood scan over the CP mixing angle α. The coupling strength κ_t is profiled.

The CP odd hypothesis ($\alpha = 90°$) is rejected at the 3.9σ level, and $\alpha = 180°$ is rejected at the 2.5σ level. The measured mixing angle (in degrees) is

$$\alpha = 0.0\ ^{+25.1}_{-21.2}\ (\text{stat.})\ ^{+2.8}_{-1.7}\ (\text{syst.}) \tag{12.9}$$

Parametrize loop processes in terms of κ_t and α

Assuming there is no new physics other than CP mixing in the Higgs-top interaction, the effective loop coupling strengths κ_g and κ_γ are directly parameterized as functions of κ_t and α using Equation 12.4. This result is more model-dependent than the primary interpretation above. Here, the constraints from combined ATLAS Higgs data are not used, but only data in the twenty $t\bar{t}H(\gamma\gamma)$ CP analysis categories.

The two-dimensional likelihood scan of $\kappa_t \cos\alpha$ and $\kappa_t \sin\alpha$ is shown in Figure 12.15a. The entire negative branch of $\kappa_t \cos\alpha$ is rejected at above 3σ. However, high values of κ_t, up to $\kappa_t \sim 6$, are allowed since no knowledge of the constraints from ggF and $H \to \gamma\gamma$ are applied in this interpretation. The region of allowed values contains a hole at $\kappa_t = 4.7$: near this value of κ_t, the $H \to \gamma\gamma$ branching fraction drops to zero (see Figure 12.16), a

scenario excluded by the observation of $\mu_{t\bar{t}H} > 0$ in the $t\bar{t}H(\gamma\gamma)$ CP analysis categories. The one-dimensional likelihood scan of α with κ_t profiled is shown in Figure 12.15b. Systematic uncertainties have negligible impact on the limit.

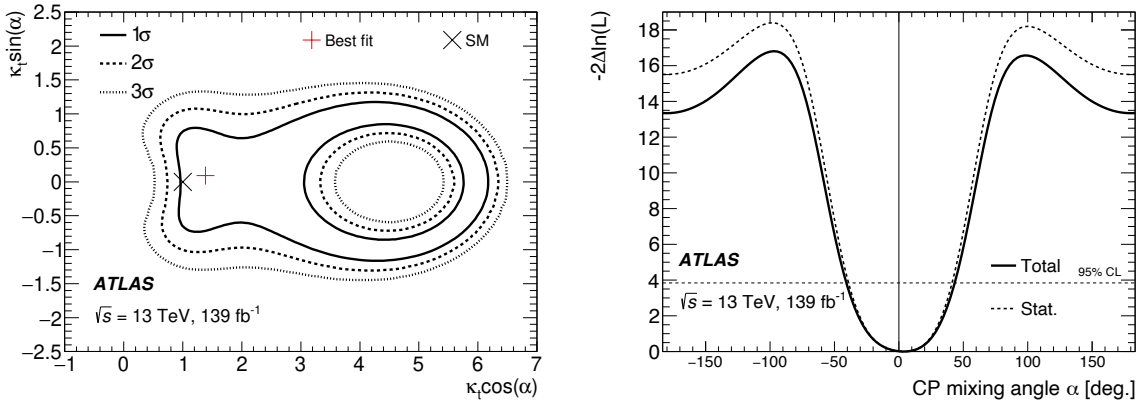


Figure 12.15: (Left) The 1, 2, and 3σ exclusion contours in the $(\kappa_t \cos\alpha,\ \kappa_t \sin\alpha)$ plane, with κ_g and κ_γ parameterized in terms of κ_t and α according to Equation 12.4. (Right) One-dimensional likelihood scan over the CP mixing angle α. The coupling strength κ_t is profiled.

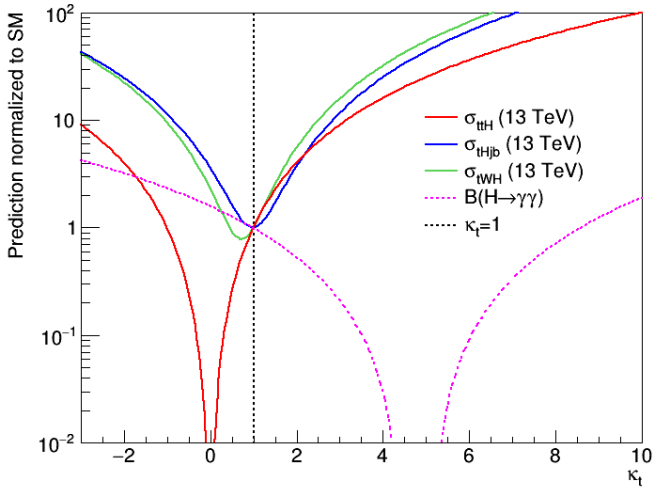


Figure 12.16: The $H \to \gamma\gamma$ branching fraction and SM $t\bar{t}H$ and tH cross sections as a function of κ_t. The drop in the $H \to \gamma\gamma$ branching fraction corresponds to the inner excluded region in Figure 12.15.

Assuming that the SM (CP even) hypothesis is true, the expected 95% confidence limit is $|\alpha| > 56°$. The observed data exclude $|\alpha| > 43°$ at 95% confidence level, and the CP odd

hypothesis is rejected at the 4.0σ level. The measured mixing angle (in degrees) is

$$\alpha = 0.1\ ^{+26.6}_{-23.2}\ (\text{stat.})\ ^{+7.4}_{-5.2}\ (\text{syst.}) \tag{12.10}$$

The region $\kappa_t \cos\alpha > 1.5$ has been definitively excluded ($> 95\%$ CL) by ATLAS combined Higgs data [31], and an interpretation combining this full dataset with the parameterization in Equation 12.4 would yield a stronger limit than either presented here. In particular, it would eliminate the region $\kappa_t \cos\alpha > 1.5$ from the second interpretation, which is not excluded by data in the $t\bar{t}H(\gamma\gamma)$ CP categories alone. However, this limit would be dominated by the statistical power of ggF (and $H \to \gamma\gamma$, to a lesser extent) rather than direct Higgs-top production. This analysis is therefore left to a future combination of ATLAS Higgs data.

Chapter 13

Conclusions

During Run 2 of the Large Hadron Collider, protons were collided at $\sqrt{s} = 13$ TeV, the highest center of mass energy achieved by a collider to date. During this time, the ATLAS experiment collected a dataset of unprecedented size, corresponding to a luminosity of 139 fb^{-1}. Analysis of this dataset has provided fundamental insight into the nature of the Higgs boson and its interaction with the top quark, the heaviest particle in the Standard Model.

By performing a statistical combination of ATLAS $t\bar{t}H$ searches in multiple Higgs decay modes, Higgs boson production in association with a top-antitop quark pair was observed for the first time in the partial Run 2 dataset at the level of 6.3σ (5.1σ expected). This observation directly established the interaction of the Higgs boson with quarks and serves as powerful validation of the theory of Higgs-quark couplings in the SM.

Due to the excellent diphoton mass resolution of the ATLAS detector and the clean signature of the $H \to \gamma\gamma$ decay, this channel has proven to be one of the most powerful for $t\bar{t}H$ measurements. Using machine learning techniques, a sophisticated multivariate analysis was developed based on the momenta of jets, photons, and leptons. The resulting categories contain a high ratio of $t\bar{t}H$ signal to background. Top quarks were reconstructed with high probability in selected $t\bar{t}H(\gamma\gamma)$ events, verifying the presence of both a top quark and a Higgs boson in the final state. The measurement of the $t\bar{t}H$ cross section in the $H \to \gamma\gamma$ channel is among the most precise single channel measurements to date: a cross section times branching fraction of $\sigma_{t\bar{t}H} \times B_{\gamma\gamma} = 1.59\ ^{+0.43}_{-0.39}$ fb is measured using the full Run 2 ATLAS dataset.

With the $t\bar{t}H$ process directly established, a direct measurement of the CP properties of the Higgs-top interaction was then carried out in the $H \to \gamma\gamma$ decay channel. A CP-sensitive multivariate categorization was developed using reconstructed top quark variables, yielding a single-channel observed $t\bar{t}H$ significance of 5.2σ (4.4σ expected) and an observed upper limit on the tH cross section is 11.6 times the SM expectation (11.7 expected). The observed data excludes a fully CP odd Higgs-top coupling at the level of 3.9σ. The CP mixing angle is constrained to $|\alpha| > 43°$ at 95% confidence level ($|\alpha| > 63°$ expected).

With Run 2 brought to a successful conclusion, the LHC experiments (ATLAS included) are undergoing large-scale upgrades in preparation for the High-Luminosity LHC (HL-LHC),

which will expand the existing proton–proton dataset by a factor of ten or more over the next few decades. The exploration of the Higgs boson and its couplings will continue at the HL-LHC through high precision and differential Higgs measurements, in addition to searches for new production and decay modes.

Appendix A

Search for TeV-scale gravity

The increase in proton–proton center of mass energy from $\sqrt{s} = 8$ TeV to 13 TeV in LHC Run 2 opened up a large phase space for new physics that had not been explored by previous experiments. This Appendix describes an early Run 2 search for signatures of TeV-scale gravity, such as microscopic black holes or string balls [2] [3], in final states with many high p_T jets [159].

The production of microscopic black holes and subsequent decay to jets are generated using the CHARYBDIS2 event generator [160]. Figure A.1 shows the distribution of jet multiplicity (n_{jet}) and scalar sum of jet p_T (H_T) for an example signal point: a model of rotating black holes in 6 extra dimensions, where the Planck scale in $4+6$ dimensions (M_D) is 2 TeV and the black hole threshold mass (M_{th}) is 7 TeV. This signal point is just beyond the limit on this process set by ATLAS using 8 TeV data [161]. The signal is concentrated around $H_T \sim 5.5$ TeV and $n_{\text{jet}} \sim 8$.

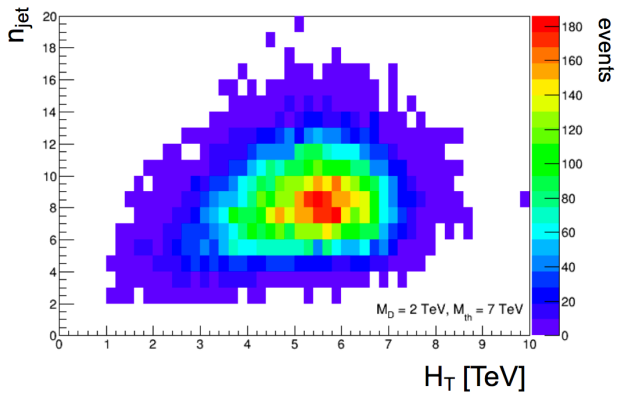


Figure A.1: Distribution of H_T and n_{jet} for a rotating black hole signal in 6 extra dimensions, where the Planck scale in $4+6$ dimensions (M_D) is 2 TeV and the black hole threshold mass (M_{th}) is 7 TeV.

This search targets signals with high jet multiplicity and large H_T, such as that shown

in Figure A.1. Jets are required to have $p_T > 50$ GeV and $|\eta| < 2.8$. In order to ensure that the trigger is fully efficient in the selected region, each event must have $H_T > 1$ TeV and leading jet $p_T > 200$ GeV. The analysis is performed in six inclusive jet multiplicity bins: $n_{\text{jet}} \geq 3 - 8$.

The dominant background passing this selection is QCD multijet production, which is estimated using a data-driven method. In each n_{jet} bin, the H_T range is divided into three regions as shown in Figure A.2: a Control Region (CR) at low H_T, a Validation Region (VR) at intermediate H_T, and a Signal Region (SR) at high H_T. The region boundaries C, V, and S are optimized using a Pythia8 multijet Monte Carlo sample that is normalized to data at low H_T:

1. C is chosen so that the background extrapolation uncertainty is minimized.
2. S is chosen so that the background extrapolation uncertainty in the SR is 0.5 events.
3. V is chosen so that 20 background events are predicted in the range $V < H_T < S$.

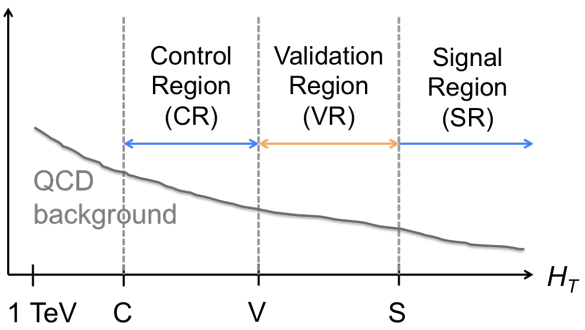


Figure A.2: Diagram of the Control Region, Validation Region, and Signal Region definitions.

The multijet background is estimated separately in each n_{jet} bin by fitting to data in the Control Region. The functional forms considered for the shape of the multijet background are listed in Table A.1. To evaluate the performance of each fit function, pseudo-experiments (PEs) are generated from the Pythia8 multijet Monte Carlo. A candidate function qualifies if the fit converges and the function decreases monotonically in at least 95% of PEs. Qualifying functions are ranked based on their extrapolation uncertainty (determined using PEs). The function with smallest extrapolation uncertainty is selected as the baseline: extrapolation of this function to the VR and SR gives the central values of the background prediction.

The uncertainty on the background prediction has two components:

1. PE: extrapolation uncertainty calculated from the pseudo-experiments
2. DD: data-driven systematic uncertainty

	Functional form	p_1	p_2
1	$f_1(x) = \frac{p_0(1-x)^{p_1}}{x^{p_2}}$	(0,+∞)	(0,+∞)
2	$f_2(x) = p_0(1-x)^{p_1} e^{p_2 x^2}$	(0,+∞)	(-∞ ,+∞)
3	$f_3(x) = p_0(1-x)^{p_1} x^{p_2 x}$	(0,+∞)	(-∞ ,+∞)
4	$f_4(x) = p_0(1-x)^{p_1} x^{p_2 \ln x}$	(0,+∞)	(-∞ ,+∞)
5	$f_5(x) = p_0(1-x)^{p_1}(1+x)^{p_2 x}$	(0,+∞)	(0,+∞)
6	$f_6(x) = p_0(1-x)^{p_1}(1+x)^{p_2 \ln x}$	(0,+∞)	(0,+∞)
7	$f_7(x) = \frac{p_0}{x}(1-x)^{[p_1 - p_2 \ln x]}$	(0,+∞)	(0,+∞)
8	$f_8(x) = \frac{p_0}{x^2}(1-x)^{[p_1 - p_2 \ln x]}$	(0,+∞)	(0,+∞)
9	$f_9(x) = \frac{p_0(1-x^{1/3})^{p_1}}{x^{p_2}}$	(0,+∞)	(0,+∞)
10	$f_{10}(x) = p_0(1-x^{1/3})^{p_1} x^{p_2 \ln x}$	(0,+∞)	(-∞ ,+∞)

Table A.1: Functional forms considered for the H_T distribution of the multijet background.

The systematic uncertainty from pseudo-experiments is the difference in the number of events in the SR between the Monte Carlo prediction and the fit extrapolation. The data-driven systematic uncertainty is derived by extrapolating all qualifying fit functions to the VR and SR. This gives a projected background yield that is different from the baseline. Any qualifying function that does not agree with data in the VR within 95% confidence level is rejected. The data-driven systematic uncertainty is taken to be the maximum difference in the SR background projection between the baseline and the remaining fit functions.

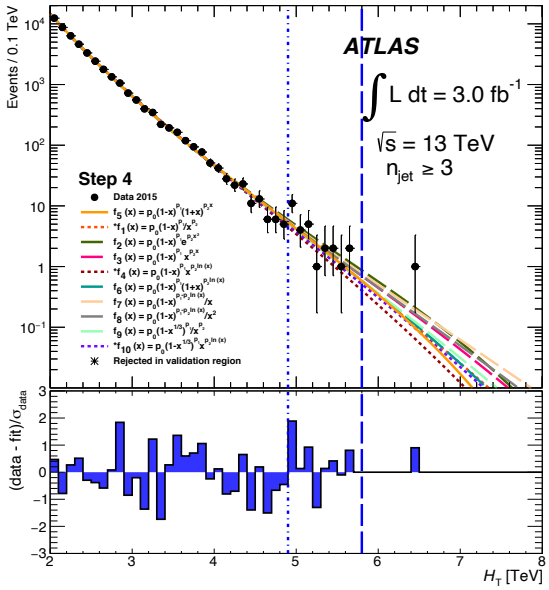


Figure A.3: H_T distribution in data with 3.0 fb^{-1} in the $n_{\text{jet}} \geq 3$ bin, with qualifying fit functions overlaid. Functions rejected by data in the Validation Region are marked with an asterisk. The region boundary V (S) is shown as a blue dotted (dashed) vertical line.

Bootstrap analysis method

The analysis strategy outlined above relies on the assumption that there is no significant signal contamination in the Control Region. However, if the analysis is performed directly with a few inverse femtobarns of 13 TeV data, then there are allowed signal points that could impact the H_T spectrum in the CR and bias the background fit.

In order to ensure that the analysis is not biased by possible signal contamination, a **boostrap analysis method** is applied: the analysis is repeated and at several steps in integrated luminosity. If no new physics is observed at Step i, then Step $i+1$ (at about $10\times$ higher luminosity) is safe from signal contamination in the CR. The Run 1 ATLAS exclusion limit is taken as a starting point, and the 13 TeV analysis is performed at four luminosity steps: 6.5 pb^{-1}, 74 pb^{-1}, 0.44 fb^{-1}, and 3.0 fb^{-1}. The total integrated luminosity used in the analysis is 3.6 fb^{-1}.

Figure A.4 shows the observed H_T distribution in data at each luminosity step in the $n_{\text{jet}} \geq 3$ bin. The region boundaries C, V, and S (overlaid in blue) are calculated independently for each n_{jet} bin and for each luminosity step. A signal point near the exclusion limit set by each step is shown in red.

The observed and predicted number of events in each region are shown in Table A.2 for each luminosity step and each jet multiplicity bin. No significant excess is observed. At Step 1, only the $n_{\text{jet}} \geq 3$ bin has enough statistics to derive a result. Statistics are too low to derive a result in the $n_{\text{jet}} \geq 8$ bin until Step 3.

$n_{\text{jet}} \geq$	VR (obs)	VR (exp)	SR (obs)	SR (exp)
3	19	20.4 ± 4.4 (PE) ± 2.6 (DD)	0	0.65 ± 0.46 (PE) ± 0.64 (DD)

(a) Step 1: 6.5 pb^{-1}

$n_{\text{jet}} \geq$	VR (obs)	VR (exp)	SR (obs)	SR (exp)
3	23	27.1 ± 3.7 (PE) ± 9.6 (DD)	1	1.42 ± 0.41 (PE) $^{+4.3}_{-1.42}$ (DD)
4	27	25.4 ± 3.2 (PE) ± 15.5 (DD)	0	1.62 ± 0.46 (PE) $^{+9.2}_{-1.62}$ (DD)
5	21	18.9 ± 2.9 (PE) ± 9.9 (DD)	0	1.32 ± 0.48 (PE) $^{+5.1}_{-1.32}$ (DD)
6	18	20.7 ± 3.3 (PE) ± 10.4 (DD)	0	1.19 ± 0.48 (PE) $^{+13.3}_{-1.19}$ (DD)
7	29	22.2 ± 3.7 (PE) ± 7.0 (DD)	0	0.81 ± 0.36 (PE) ± 0.60 (DD)

(b) Step 2: 74 pb^{-1}

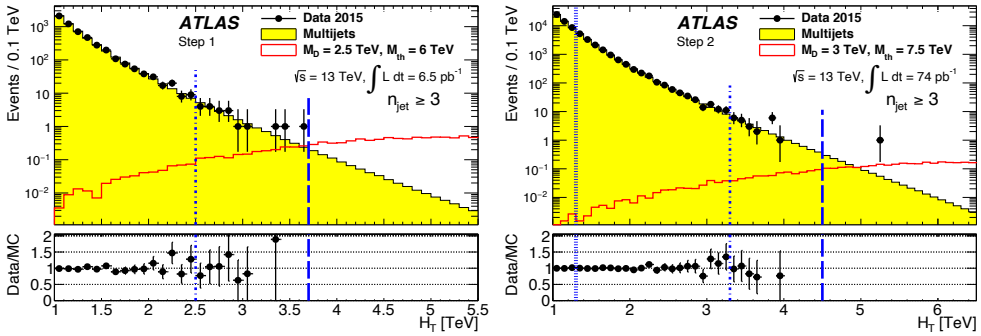
$n_{\text{jet}} \geq$	VR (obs)	VR (exp)	SR (obs)	SR (exp)
3	21	20.4 ± 2.7 (PE) ± 10.5 (DD)	2	1.46 ± 0.42 (PE) $^{+4.37}_{-1.46}$ (DD)
4	23	29.9 ± 3.9 (PE) ± 8.1 (DD)	2	1.95 ± 0.46 (PE) $^{+4.06}_{-1.95}$ (DD)
5	17	21.4 ± 3.4 (PE) ± 7.1 (DD)	1	1.56 ± 0.51 (PE) $^{+3.47}_{-1.56}$ (DD)
6	19	28.3 ± 4.3 (PE) ± 6.3 (DD)	0	1.44 ± 0.40 (PE) $^{+2.13}_{-1.44}$ (DD)
7	28	24.7 ± 3.8 (PE) ± 4.5 (DD)	0	0.96 ± 0.39 (PE) $^{+1.74}_{-0.96}$ (DD)
8	25	31.8 ± 4.7 (PE) ± 1.4 (DD)	2	2.86 ± 0.40 (PE) ± 0.70 (DD)

(c) Step 3: 0.44 fb^{-1}

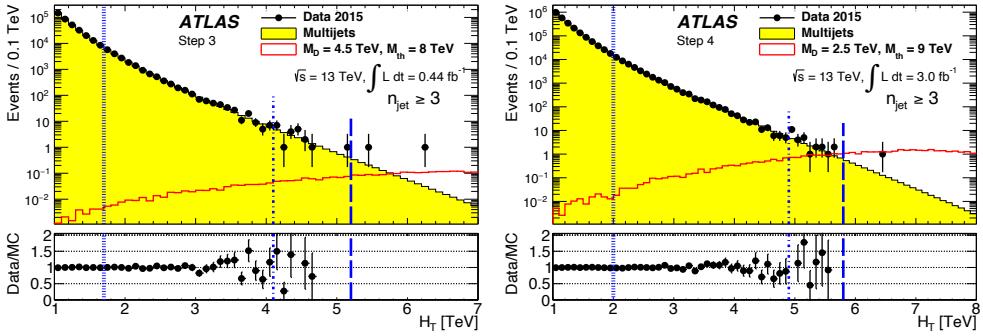
$n_{\text{jet}} \geq$	VR (obs)	VR (exp)	SR (obs)	SR (exp)
3	28	19.5 ± 3.6 (PE) ± 4.1 (DD)	1	2.10 ± 0.51 (PE) ± 1.78 (DD)
4	27	20.8 ± 2.3 (PE) ± 6.4 (DD)	2	2.36 ± 0.52 (PE) ± 2.12 (DD)
5	26	22.3 ± 2.6 (PE) ± 6.8 (DD)	2	1.95 ± 0.45 (PE) $^{+2.10}_{-1.95}$ (DD)
6	20	20.3 ± 2.9 (PE) ± 5.4 (DD)	3	1.82 ± 0.49 (PE) $^{+1.91}_{-1.82}$ (DD)
7	14	20.7 ± 4.1 (PE) ± 1.7 (DD)	0	0.53 ± 0.36 (PE) ± 0.22 (DD)
8	19	18.2 ± 4.9 (PE) ± 3.5 (DD)	0	0.43 ± 0.36 (PE) ± 0.26 (DD)

(d) Step 4: 3.0 fb^{-1}

Table A.2: Observed and expected event yields at each luminosity step in the bootstrap analysis, shown in bins of inclusive jet multiplicity. At Step 1, only the $n_{\text{jet}} \geq 3$ bin has enough statistics to derive a result. Statistics are too low to derive a result in the $n_{\text{jet}} \geq 8$ bin until Step 3.



(a) Step 1: 6.5 pb^{-1} (b) Step 2: 74 pb^{-1}



(c) Step 3: 0.44 fb^{-1} (d) Step 4: 3.0 fb^{-1}

Figure A.4: Comparison of data and multijet Monte Carlo at each step of the bootstrap analysis in the $n_{\text{jet}} \geq 3$ bin. The Monte Carlo prediction from Pythia8 (yellow) is normalized to data at low H_T. The region boundaries C, V, and S are overlaid in blue, and a signal point near the exclusion limit is shown in red for each luminosity step.

Limits on new physics

Limits are derived on multiple signal models simulated using the CHARYBDIS2 event generator [160]. Figure A.5 shows the exclusion of rotating black holes in 6 extra dimensions in terms of the model parameters M_D and M_{th}. Figure A.6 shows the exclusion of string balls in terms of the string coupling (g_S), the string scale (M_S), and the string ball threshold mass (M_{th}).

A model-independent limit on the cross section of new physics at high H_T can be derived from the limit on the number of signal events N_S:

$$\sigma = \frac{N_S}{L}\frac{\epsilon^{\text{reco}}}{\epsilon^{\text{true}}} \tag{A.1}$$

Here, L is the integrated luminosity and ϵ^{true} (ϵ^{reco}) represents the analysis selection efficiency on true (reconstructed) signal events. The ratio of efficiencies $\epsilon^{\text{reco}}/\epsilon^{\text{true}}$ is model dependent, so a conservative estimate is used to obtain the cross section limit. The expected and

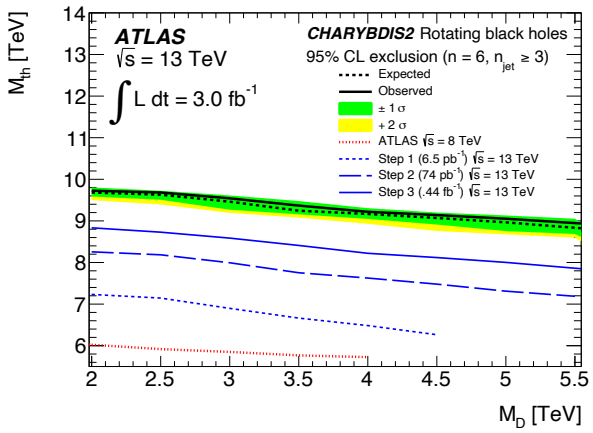
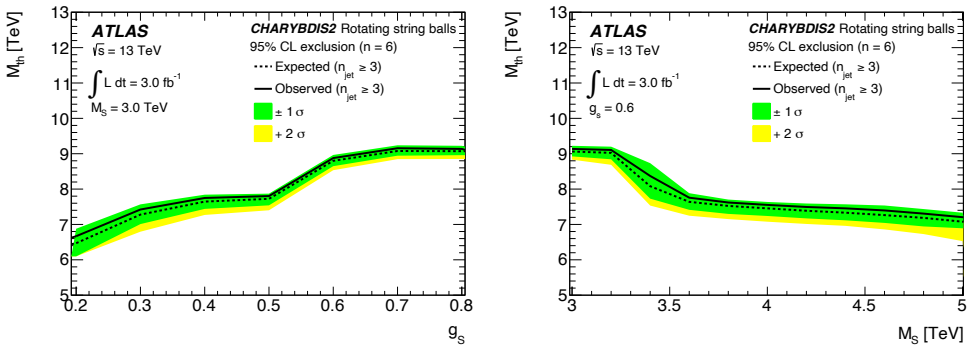


Figure A.5: Limit on the production of rotating black holes in 6 extra dimensions generated using CHARYBDIS2. The Run 1 ATLAS limit is shown in red, and the limit obtained at each step in the bootstrap in the analysis is shown in blue.



(a) $M_S = 3.0$ TeV (b) $g_S = 0.6$

Figure A.6: Limit on the production of string balls generated using CHARYBDIS2.

observed cross section limits are reported in Table A.3 for each inclusive jet multiplicity bin. The H_T range of the Signal Region is also shown.

$n_{\text{jet}} \geq$	$H_{\text{T}} > H_{\text{T}}^{\text{min}}$ (TeV)	Expected limit (fb)	Observed limit (fb)
3	5.8	$1.63^{+0.70}_{-0.57}$	1.33
4	5.6	$1.77^{+0.70}_{-0.57}$	1.77
5	5.5	$1.56^{+0.73}_{-0.50}$	1.75
6	5.3	$1.52^{+0.69}_{-0.50}$	2.15
7	5.4	$1.02^{+0.36}_{-0.0}$	1.02
8	5.1	$1.01^{+0.29}_{-0.0}$	1.01

Table A.3: Expected and observed limits on the cross section of new physics in bins of inclusive jet multiplicity.

Appendix B

Multijet modeling with MG5_aMC@NLO

Multijet final states constitute a major background in many ATLAS analyses, especially searches for Beyond the Standard Model physics (see for example Appendix A). Because pure QCD processes are challenging to model in Monte Carlo, background predictions of multijet processes are often derived using data-driven methods. In order for multijet Monte Carlo samples to be useful for analysis, a thorough understanding of their inputs and features is essential.

This appendix presents the development of the first leading order $2 \to 4$ multijet sample generated for use in ATLAS at $\sqrt{s} = 13$ TeV. The MG5_aMC@NLO generator is selected for this sample because of its ability to perform leading order multileg and NLO calculations, and because of its interface to multiple showering packages. For references associated with each Monte Carlo generator software, see Table 5.3.

Throughout this Appendix, the MG5_aMC@NLO sample is compared to a leading order $2 \to 2$ Pythia sample generated with version 8.1.86. This sample uses PDF set NNPDF23 [162] and the A14 parameter tune, and EvtGen for bottom and charm hadron decays. This Pythia sample is found to agree well with data in important jet distributions, including leading and inclusive jet p_T [163].

The MG5_aMC@NLO multijet sample is generated using MG5_aMC@NLO version 2.3.3. The contribution from diagrams with up to four partons in the final state is calculated at leading order in α_S. Additional jets are generated by the parton shower, which is performed by Pythia 8.2.12. The matching/merging is done according to the CKKW-L algorithm [164] with a merging scale of 30 GeV. The EvtGen program is used for bottom and charm hadron decays.

The NNPDF30 PDF set [127] is selected for the baseline sample with the A14 parameter tune [122]. The CT10NLO PDF [165] set is chosen as an alternative PDF set, and a small test sample is generated in order to examine the impact of the choice of PDF on jet variables. The LHAPDF PDF reweighting procedure [13] is applied to this test sample, and the reweighting is found to close within 15%. The total effect of PDF choice on the MG5_aMC@NLO sample

is taken to be the envelope of these reweighted distributions. This gives an uncertainty on the total jet cross section of about 10-20% for jet $p_T < 4.5$ TeV. The impact of PDF choice grows large in the high p_T tail.

Slicing

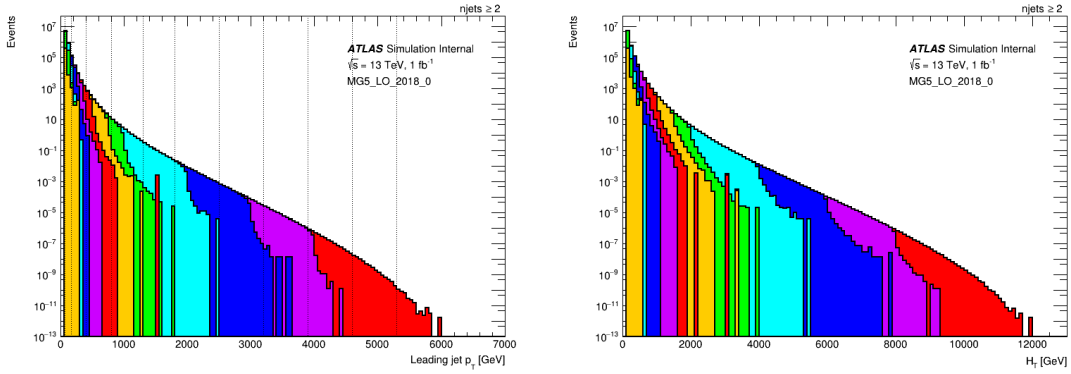
The MG5_aMC@NLO sample is composed of twelve regions (or "slices") in parton-level H_T (scalar sum p_T). Slicing at the parton level rather than particle level removes the need for filters with low efficiency and therefore long generation times. The contribution from each H_T slice to the full p_T and H_T distributions can be seen in Figure B.1. The boundaries of the H_T slices are listed in Table B.1.

Slice name	HT1	HT2	HT3	HT4	HT5	HT6
Min H_T [GeV]	0	100	200	300	500	700
Max H_T [GeV]	100	200	300	500	700	1000

Slice name	HT7	HT8	HT9	HT10	HT11	HT12
Min H_T [GeV]	1000	1500	2000	4000	6000	8000
Max H_T [GeV]	1500	2000	4000	6000	8000	13000

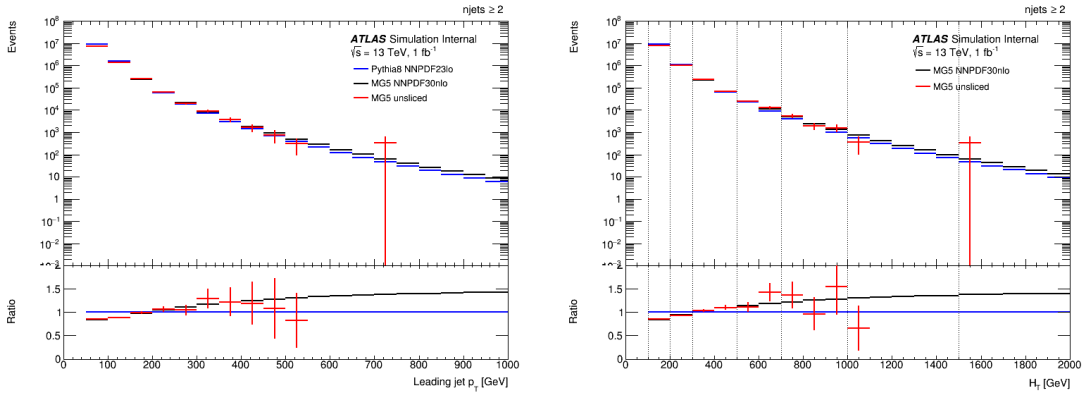
Table B.1: H_T slice boundaries used for the MG5_aMC@NLO multijet Monte Carlo sample.

The H_T sliced sample agrees well with an H_T inclusive test sample, and it has been verified that the distribution is smooth at the slice boundaries (Figure B.2).



(a) Leading jet p_T (b) H_T

Figure B.1: Leading jet p_T and H_T for the MG5_aMC@NLO multijet sample sliced in parton-level H_T. Each slice is represented in a different color. The dashed black lines mark the boundaries of the H_T slices.



(a) Leading jet p_T (b) H_T

Figure B.2: Leading jet p_T and H_T for the MG5_aMC@NLO multijet sample sliced in parton-level H_T (black) and unsliced (red). The ratio in the bottom panel is to the Pythia8 sample. Agreement is good in the region where the unsliced sample has sufficient statistics, including at H_T slice boundaries (vertical dashed lines).

Scale variations

Two forms are tested for the renormalization and factorization scales μ_R and μ_F:

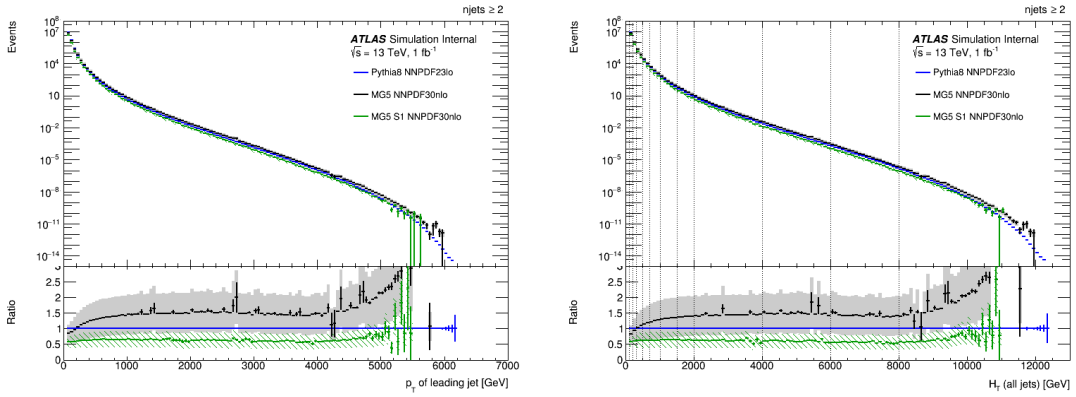
- S_0: the transverse mass of the $2 \to 2$ system resulting from k_T clustering (the default for LO generation in MG5_aMC@NLO)
- S_1: the total transverse energy in the event:

$$S_1 = \sum_{i=1}^{N} \frac{E_i p_{T,i}}{\sqrt{p_{x,i}^2 + p_{y,i}^2 + p_{z,i}^2}} \tag{B.1}$$

The renormalization and factorization scales (μ_R and μ_F) are varied independently up and down by a factor of two to provide an uncertainty band on each distribution. The total scale uncertainty on the MadGraph sample is taken to be the envelope of these eight variations. The impact of scale variations on the jet cross section is dominated by μ_R rather than μ_F. These scale variations give an uncertainty on the total jet cross section of about 40%.

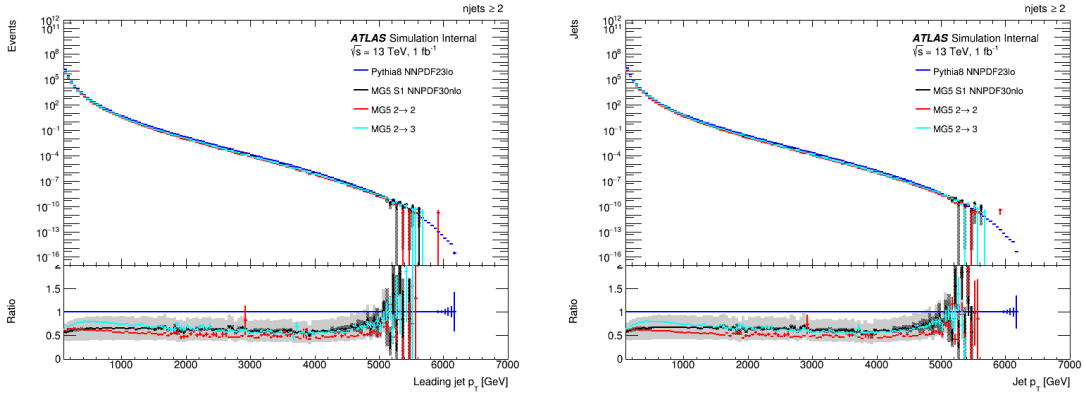
The leading jet p_T and H_T distributions are shown in Figure B.3 using scales S_0 and S_1. The sample using scale S_1 agrees better in shape with the Pythia8 sample in both variables. In addition, the uncertainty due to scale variations is smaller on the sample with scale S_1. This form of the scale is therefore chosen for the final MG5_aMC@NLO sample.

Test samples with $2 \to 2$, $2 \to 3$, and $2 \to 4$ partons in the matrix element were generated using scale S_1. Figure B.4 shows that the scale variations on the $2 \to 4$ multileg sample (light gray band) cover the differences between these calculations.



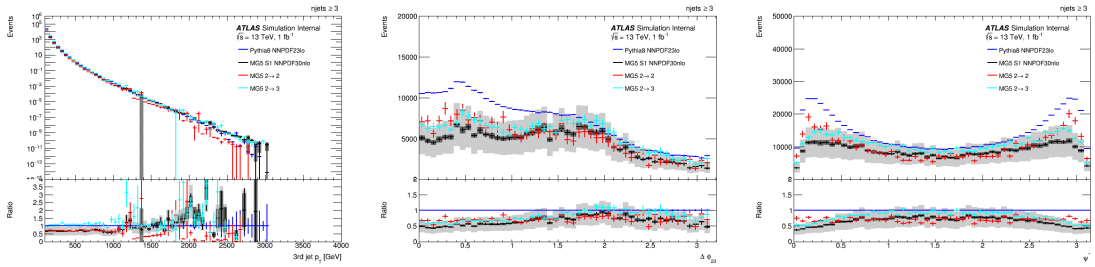
(a) Leading jet p_T (b) H_T

Figure B.3: Leading jet p_T and H_T using two different forms of the scale in MG5_aMC@NLO multijet Monte Carlo. The default scale S_0 is shown in black and S_1 in green. The uncertainty bands show the envelope of distributions with μ_R and μ_F varied up and down by a factor of two. The ratio is shown to the Pythia8 sample. All jets have $p_T > 50$ GeV and $|\eta| < 5.0$.



(a) Leading jet p_T (b) Inclusive jet p_T

Figure B.4: Leading jet p_T and inclusive jet p_T for MG5_aMC@NLO multijet samples using scale S_1. The light gray band shows scale variations on the $2 \to 4$ sample, and the dark gray band shows the PDF uncertainty on this sample. The ratio is shown to the Pythia8 sample.



(a) Third jet p_T (b) $\Delta\phi(j_2, j_3)$ (c) ψ^*

Figure B.5: Variables sensitive to the third jet for MadGraph multijet Monte Carlo samples. The light gray band shows scale variations on the $2 \to 4$ MadGraph sample, and the dark gray band shows the impact of PDF choice on this sample. The ratio is shown to the Pythia8 sample.

Variables sensitive to the third jet (Figure B.5) can be used to probe the differences between the $2 \to 2$ and $2 \to 3, 4$ samples. The $2 \to 3+$ samples have a harder p_T spectrum for the third jet than $2 \to 2$, as well as a higher average jet multiplicity. All MG5_aMC@NLO samples show good agreement in the angular separation between the second and third jets ($\Delta\phi(j_2, j_3)$), but there is a large discrepancy with Pythia. The angle between the plane formed by the leading jet momentum/beam and the plane formed by the second/third leading jet momentum (ψ^*) [166] exhibits shape differences between the $2 \to 2$ and $2 \to 3+$ samples, and the MG5_aMC@NLO $2 \to 2$ sample is most similar in shape to Pythia.

In an optimal multijet Monte Carlo setup, the data provide a guide. Additional measurements at $\sqrt{s} = 13$ TeV, particularly measurements of variables sensitive to the third jet, will shed further light on the relative performance of the available multijet samples. Such comparisons are underway, and this MG5_aMC@NLO sample is included in a comprehensive comparison of ATLAS multijet Monte Carlo samples [163].

Appendix C

Radiation damage in the IBL

This Appendix presents two offline studies of the effects of radiation damage on the ATLAS Insertable B-Layer (IBL). Before describing these studies in detail, a description of the IBL layout and electronics are provided that is beyond the scope of Chapter 3. Section C.1 reports measurements of the rate of Single Event Upset (SEU) in the IBL obtained by reading back pixel configuration latches. Section C.2 describes a measurement of leakage current in the IBL and includes projections for Run 3.

The IBL: a closer look

The IBL is composed of silicon pixel sensors with two different designs:

1. **Planar sensors** are used in the central region ($|z| < 24$ cm). These sensors are of the same design used in other layers of the Pixel detector. Each planar sensor consists of 26,880 pixels, each with surface area of 250×50 μm^2 and thickness of 200 μm.

2. **3D sensors** are used in the forward region. This new sensor design is used for the first time in ATLAS [71]. Each 3D sensor has a surface area of 250×50 μm^2 and thickness of 230 μm.

The readout chip used in the IBL is the FE-I4B, which is built in 130 nm CMOS technology [167] [168]. Each planar IBL module is composed of two FE-I4B chips and one planar sensor, and each 3D module is composed of a single FE-I4B chip and one 3D sensor. Each pixel in the sensor is electrically bonded to one channel of an FE-I4B chip to form a module.

Fourteen carbon fiber staves support the IBL modules. Each stave is 64.3 cm long and 3 cm wide, and is oriented along the beam direction. Figure C.1 shows the orientation of the IBL staves and the numbering scheme used to label them. Each stave holds twelve planar and eight 3D modules. Four 3D modules are located on the A side ($|z| > 0$ cm) and four on the C side ($|z| < 0$ cm). In total, there are 32 FE-I4B readout chips on each stave.

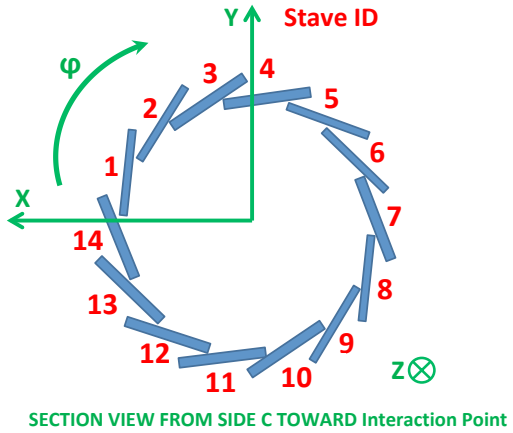


Figure C.1: Position of IBL staves in the $\hat{x} \times \hat{y}$ plane [169].

C.1 Single Event Upset

Single Event Upset (SEU) is an effect of ionizing radiation where charge injected into a circuit's memory cell(s) leads to a flip in one or more logic states. At the LHC, SEUs are caused by recoil nuclei and hadronic showers resulting from the interaction of particles in the detector material near the memory cells affected [170]. Such interactions are difficult to model, but an understanding of SEU rates is essential for detector operations. This Section presents an in-situ measurement of the probability of bit flips in the radiation hard memory cells known as Dual Interlocked CElls, or DICE latches [171] [172], which store the configuration of each pixel in the IBL.

Each DICE latch contains thirteen bits, which correspond to the following configuration settings:

- 0: Output enable
- 1-5: Threshold tuning DAC (TDAC) value. 1 = most significant bit (MSB)
- 6-7: Selection of charge injection capacitor
- 8: HitBus (input to logical OR of all pixel discriminator outputs in the matrix)
- 9-12: Time-over-threshold (ToT) tuning DAC (FDAC) value. ToT represents the length of time that single pixel discriminator is over threshold in counts of an externally supplied clock (nominally 40 MHz). 12 = most significant bit (MSB)

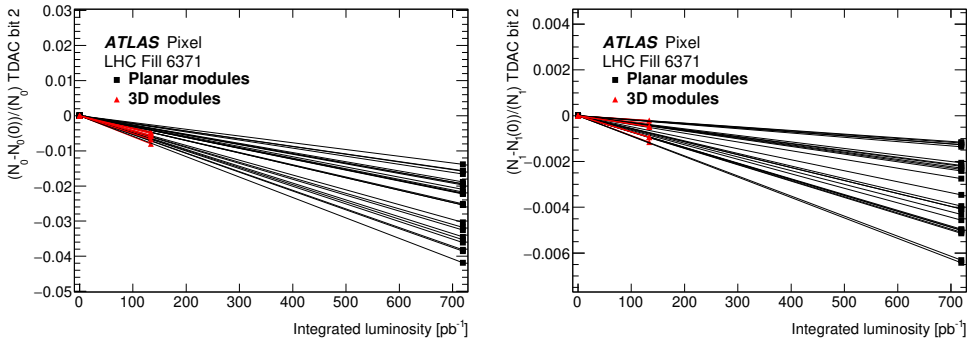
During data taking, more than 99% of pixels are enabled: that is, they have a value of one stored in bit 0. Both capacitor selection bits and the HitBus bit also store values of one. The 5-bit TDAC stores an average value of fifteen, and the 4-bit FDAC stores an average value of seven.

The SEU rate is measured by copying the values of the pixel latches (which are radiation hard) into a register which is not radiation hard. After allowing a certain fluence of particles to traverse the device, the latch values are re-copied, read back, and compared to their original values. The fraction of pixels in which the bit state flips after taking data of integrated luminosity L depends on the probability of $0 \to 1$ ($\sigma_{0\to1}$) and $1 \to 0$ ($\sigma_{1\to0}$) transitions due to SEU:

$$\frac{N_1(0) - N_1}{N_1} = \sigma_{1\to0} \times L \tag{C.1}$$

$$\frac{N_0(0) - N_0}{N_0} = \sigma_{0\to1} \times L \tag{C.2}$$

Figure C.2 shows the change in the fraction of ones (a) and zeros (b) for a sample bit (the second most significant TDAC bit) during LHC fill 6371 in 2017. The first point is zero by construction. Because the register cannot be read back while the FE-I4B chip is receiving trigger signals, there are only two points for the planar modules (black): one before the start of collisions, and the other after beam dump. For the 3D modules (red), the register is read back after about 150 pb^{-1} of data taking.

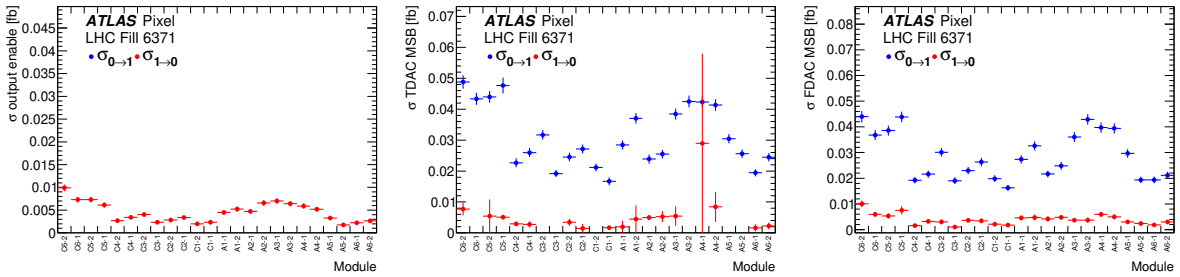


(a) $0 \to 1$ (b) $1 \to 0$

Figure C.2: Fraction of pixels in which TDAC-bit2 flips, measured at the beginning of LHC fill 6371 and after about 150 pb^{-1} of integrated luminosity for 3D modules (red) and after the end of LHC fill for planar modules (black).

From read back measurements like those shown in Figure C.2, the SEU probabilities $\sigma_{0\to1}$ and $\sigma_{1\to0}$ can be extracted for each of the thirteen pixel latches. The resulting probabilities are shown in Figure C.3 for the three bits that are most important for operations: the output enable bit, the TDAC most significant bit, and the FDAC most significant bit. The probability $\sigma_{0\to1}$ is not calculated for the output enable bit because most pixels are enabled initially.

For the TDAC and FDAC most significant bits, the $0 \to 1$ transition is about five times as frequent as the $1 \to 0$ transition. For additional studies of this asymmetry, refer to [169].



(a) 0: output enable (b) 5: TDAC MSB (c) 12: FDAC MSB

Figure C.3: Probability of bit flip due to SEU in the three DICE latches in the three DICE latches most important for IBL operations: (a) the output enable bit, (b) the most significant TDAC bit, and (c) the most significant FDAC bit. The probability of the $0 \to 1$ ($1 \to 0$) transition is shown in blue (red).

C.2 Measurement of leakage current

The **leakage current** is a small current that is present in a reverse-biased semiconductor sensor, even when no particles are traversing the sensor (see Figure C.4). The leakage current increases linearly with non-ionizing energy loss (NIEL), or **fluence**, which measures the radiation that damages the silicon pixel sensors. Fluence is measured in units of MeV neutron-equivalent radiation per cm^2 (abbreviated MeV n_{eq}/cm^2) and is proportional to integrated luminosity, which is well measured in ATLAS.

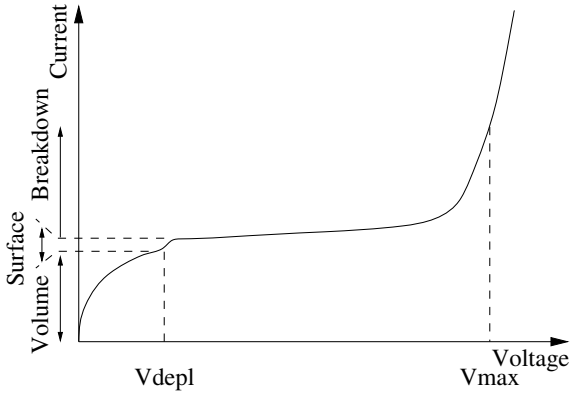


Figure C.4: Sample IV curve of a silicon pixel sensor [173]. The leakage current is the small current measured at voltages between Vdepl and Vmax.

The Hamburg model [174] provides a phenomenological model for predicting the leakage

current after n time steps:

$$I_{\text{leak}} = \frac{\Phi V}{L_{\text{tot}}} \times \sum_{i=1}^{n} L_i \left[\alpha_I \exp\left(-\sum_{j=1}^{n} \frac{t_j}{\tau(T_j)} \right) + \alpha_0^* - \beta \log\left(\sum_{j=1}^{n} \frac{\Theta(T_j) \times t_j}{t_0} \right) \right] \tag{C.3}$$

Here, L_i is the integrated luminosity, t_i is the duration, and T_i is the temperature during the i^{th} time step. L_{tot} is the total integrated luminosity, Φ is the fluence, and V is the sensor volume. The measured values of α_I, α_0^*, and β are

$$\alpha_I = (1.23 \pm 0.06) \times 10^{-17} \text{ A/cm} \tag{C.4}$$

$$\alpha_0^* = 7.07 \times 10^{-17} \text{ A/cm} \tag{C.5}$$

$$\beta = (3.29 \pm 0.18) \times 10^{-18} \text{ A/cm} \tag{C.6}$$

The factors τ and Θ are

$$\tau^{-1}(T_j) = 1.2^{+5.3}_{-1.0} \times 10^{13} \text{ s}^{-1} \times \exp\left[(-1.1 \pm 0.05) \frac{\text{eV}}{k_B T_j} \right] \tag{C.7}$$

$$\Theta(T_j) = \exp\left[-\frac{(1.30 \pm 0.14) \text{ eV}}{k_B} (\frac{1}{T} - \frac{1}{T_R}) \right] \tag{C.8}$$

The reference temperature T_R is taken to be 0°C for this study.

Given a prediction for the fluence Φ and measurements of temperature, duration, and integrated luminosity at each time step, the Hamburg model can be used to calculate the leakage current as a function of integrated luminosity. The Hamburg model predictions are overlaid on all measurements below.

To measure the leakage current, readings of high voltage, temperature, and uncorrected leakage current are compiled for each IBL module group listed in Table C.1. Each module group in this table corresponds to four FE-I4B chips in the IBL. The high voltage, temperature, and leakage current readings are averaged over time steps of one **luminosity block (LB)**. The duration of a luminosity block is usually around one minute, and the experimental conditions are assumed to be constant during this period.

The high voltage of the planar sensors (M1-M3) began at 80 V in 2015, increased to 150 V during 2016, to 300 V at the start of 2017, and finally to 400 V at the start of 2018. The high voltage of the 3D sensors began at 20 V in 2015, and increased to 40 V for the remainder of the run. The increase in the high voltage settings is necessary to ensure that the sensors are fully depleted after experiencing more and more radiation damage. Measurements of the leakage current recorded during an LB when the high voltage differs from the setting by $\geq$ 1V are vetoed: most vetoed points correspond to times when the high voltage is off because ATLAS is not taking data.

The temperature of the IBL sensors varies with time and data taking conditions. However, the leakage current can be corrected to a constant reference temperature of $T_R = 0$°C by applying a correction factor:

$$I_{\text{leak}}(T_R) = I_{\text{leak}}(T) \exp\left(-\frac{E_g}{2k_B} \left[\frac{1}{T_R} - \frac{1}{T} \right] \right) \tag{C.9}$$

Module group	Module type	z [cm]
LI_Sx_A4	3D	> 24
LI_Sx_A3	Planar	[16, 24]
LI_Sx_A2	Planar	[8, 16]
LI_Sx_A1	Planar	[0, 8]
LI_Sx_C1	Planar	[-8, 0]
LI_Sx_C2	Planar	[-16, -8]
LI_Sx_C3	Planar	[-24, -16]
LI_Sx_C4	3D	$<$ -24

Table C.1: Summary of the IBL module groups and their positions in ATLAS. The label x corresponds to the stave number of the module group.

The silicon band gap energy E_g is taken to be 1.21 eV.

The measured leakage current as a function of Run 2 luminosity is shown in Figure C.5. Measurements are averaged over all staves at the same z position and over the A and C sides (which exhibit very similar behavior). The Hamburg model prediction is overlaid for comparison, with a z-dependent scale factor applied to normalize the predicted fluence/luminosity (Φ/L) factor to data. The predicted Φ/L from FLUKA and the fitted scale factors are reported in Table C.2. The simulation is found to over-predict Φ/L at high $|z|$.

The sudden drops in the measured leakage current correspond to technical stops and year-end shutdowns. The IBL is warm (room temperature) during this time, and substantial annealing takes place. The level of annealing predicted by the Hamburg model is slightly less than what is observed in data.

Modules	Predicted Φ/L	Scale factor
A4	58.45 $\times 10^{11}$	0.75
A3	59.90 $\times 10^{11}$	0.93
A2	62.40 $\times 10^{11}$	0.96
A1	65.45 $\times 10^{11}$	1.04
C1	65.45 $\times 10^{11}$	1.04
C2	62.40 $\times 10^{11}$	0.94
C3	59.90 $\times 10^{11}$	0.91
C4	58.45 $\times 10^{11}$	0.76

Table C.2: Predicted Φ/L factors from FLUKA [(1 MeV n_{eq}/cm^2) $\times$ fb] and fitted scale factors for each group of IBL modules.

The ratio of two leakage current measurements is predicted to be the ratio of the fluences multiplied by the depleted volume. The measured ratio to the 3D modules and corresponding Hamburg model predictions are shown in Figure C.6. The discontinuity near 35 fb^{-1} corresponds to the increase in high voltage in 2016: before the voltage change, the IBL

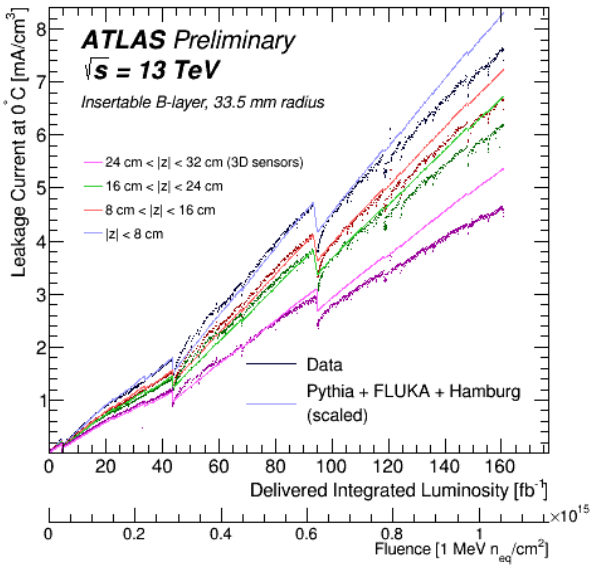


Figure C.5: Measured leakage current in IBL modules as a function of integrated luminosity. The leakage current is averaged over modules at similar $|z|$ and normalized by the sensor volume. The temperature is corrected to 0°C. The data (points) are compared to a Hamburg model simulation (lines), which is normalized to data independently for each module group.

sensors were running under-depleted. Following this change, the ratio is approximately flat for the remainder of Run 2.

Using the scale factors in Table C.2, the Hamburg model predictions for the IBL leakage current can be extended to the end of LHC Run 3. Multiple temperature and luminosity scenarios are under consideration for Run 3, but the maximum leakage current reached is similar in all scenarios.

Figure C.7 shows the Run 2 measured and Run 2-3 predicted leakage current in each module group. This projection corresponds to a total Run 3 luminosity of 320 fb^{-1}, with shutdowns (and therefore warm time for the IBL) after each 80 fb^{-1} period. Since Figure C.5 has demonstrated that the Hamburg model tends to under-predict the amount of annealing, this projection is considered conservative.

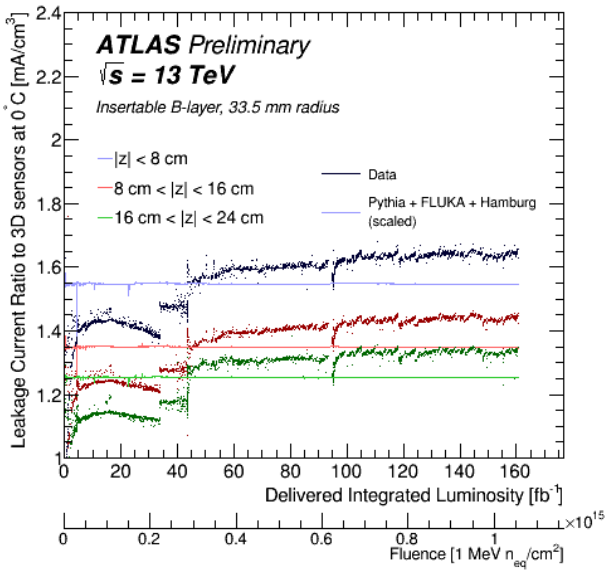


Figure C.6: Measured leakage current in IBL modules as a function of integrated luminosity, normalized to the measured leakage current in the 3D modules. The leakage current is averaged over modules at similar $|z|$ and normalized by the sensor volume. The temperature is corrected to 0°C. The data (points) are compared to a Hamburg model simulation (lines), which is normalized to data independently for each module group.

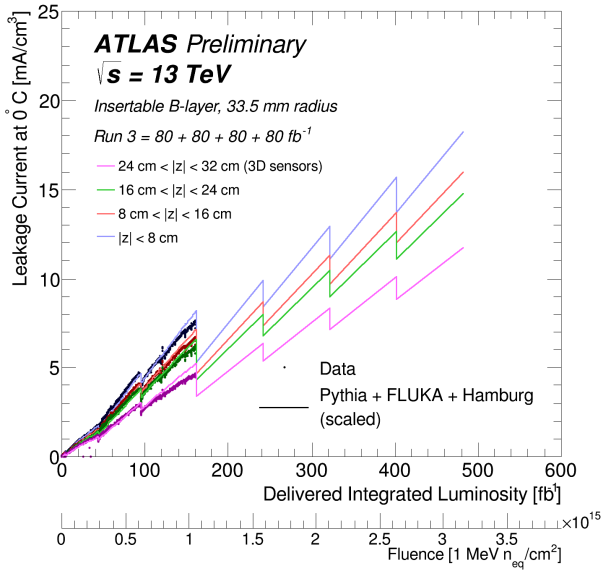


Figure C.7: Measured leakage current in IBL modules as a function of integrated luminosity. The leakage current is averaged over modules at similar $|z|$ and normalized by the sensor volume. The temperature is corrected to 0°C. The data (points) are shown through the end of Run 2, and the Hamburg model simulation (lines) is projected through the end of Run 3.

Appendix D

Toy BDT Model

This Appendix contains a toy example of the BDT method applied to a simple two-dimensional dataset. It is intended to illustrate the structure of a BDT discriminant and the impact of hyper-parameter choices. The code used to perform these tests is available at `https://github.com/jennetd/toy-bdt`.

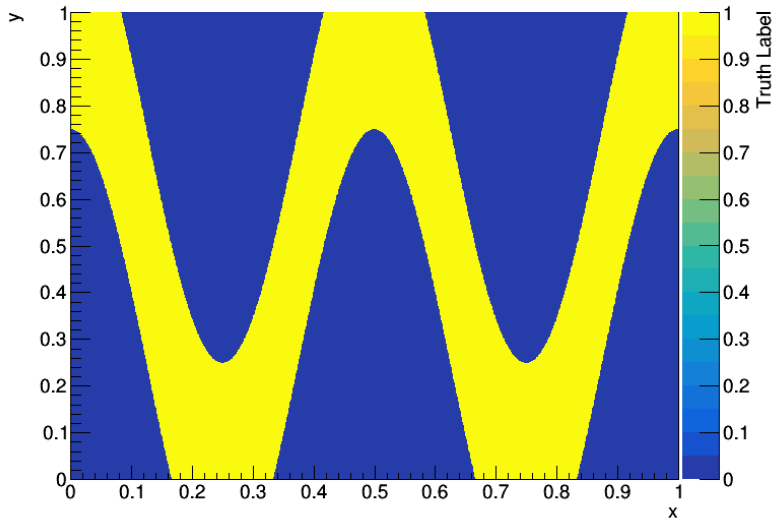
The dataset contains only two features, x and y, for ease of visualization. The training data is composed of 1,000,000 random points in (x, y), with signal (Label = 1) and background (Label = 0) defined by the following:

$$\text{Label} = f(x,y) = \begin{cases} 1 \text{ if } y > A\cos(n\pi x) + y_0 \text{ and } y < A\cos(n\pi x) + y_0 + 1 \\ 0 \text{ otherwise} \end{cases} \tag{D.1}$$

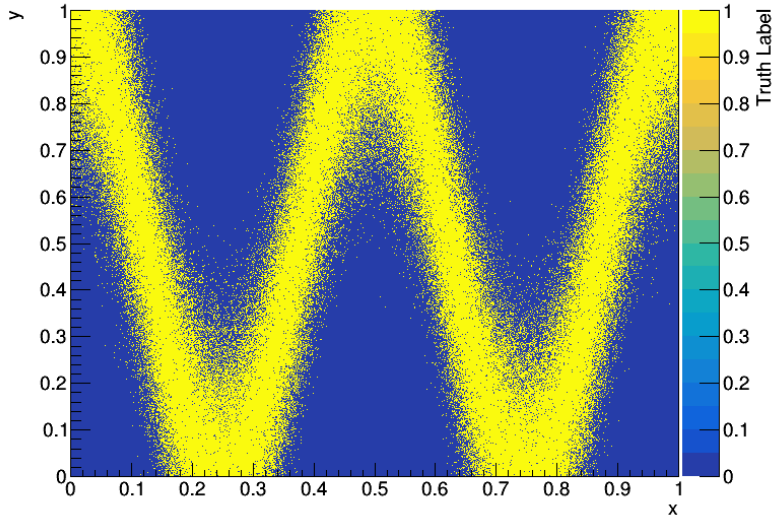
The distribution of signal and background in the (x, y) plane is shown in Figure D.1a for $A = 0.5$, $n = 4$ and $y_0 = 0.5$. This choice of parameters gives a deterministic relationship between the data label (0 or 1) and features (x, y), which is not the case in most systems. A more realistic model introduces some small random noise: Label $= f(x, y + \epsilon)$, where ϵ is a random number generated according to a gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.1$. A distribution of signal and background with the addition of this random noise is shown in Figure D.1b.

D.1 Tuning single hyper-parameters

A single tree (Rounds = 1) is trained in XGBoost on the example dataset. The BDT performance is tested for range of values for each tree hyper-parameter. For a discussion of the mathematical meaning of each hyper-parameter, see Chapter 7.



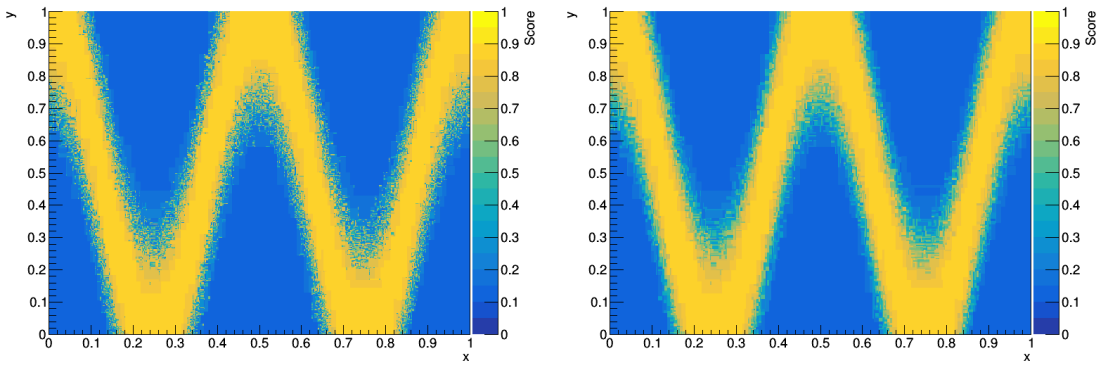
(a)



(b)

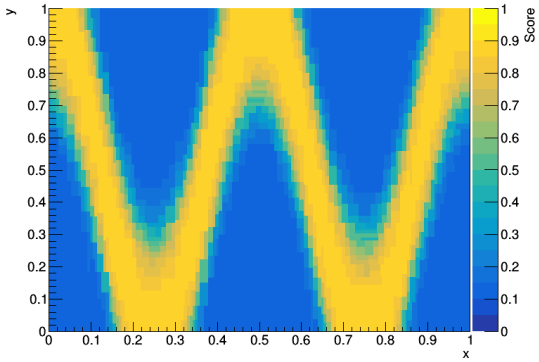
Figure D.1: Distribution in (x, y) of a sample binary dataset without random noise (a) and with random noise (b). Yellow represents the signal (Label = 1) and blue represents the background (Label = 0), according to Equation D.1 with $A = 0.5$, $n = 4$ and $y_0 = 0.5$.

Figure D.2 shows the BDT score as a function of training variables (x, y) for different values of MinChildWeight. MaxDepth is fixed to infinity and MinLossReduction to zero, so that MinChildWeight alone controls the size of the tree. Figure D.2a shows an over-trained example: MinChildWeight = 1, and each leaf contains a single training event. Figure D.2d shows an under-trained example: MinChildWeight = 1000, and the tree does not grow large enough to resolve the signal well.

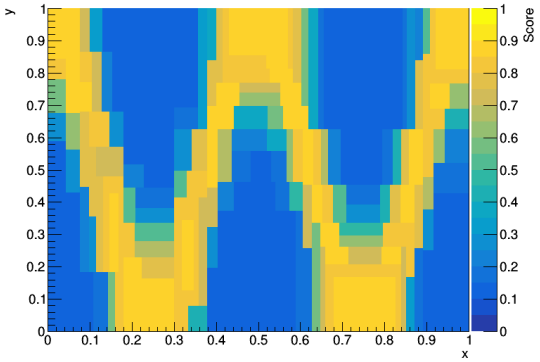


(a) MinChildWeight = 1

(b) MinChildWeight = 10



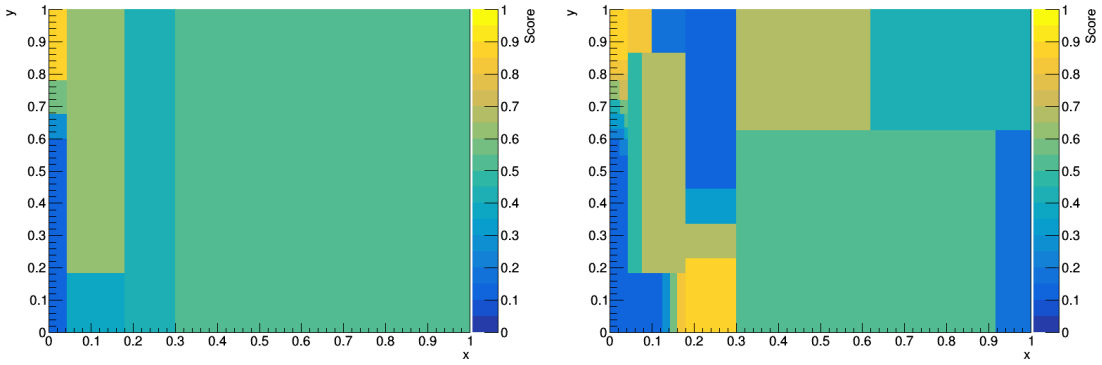
(c) MinChildWeight = 100



(d) MinChildWeight = 1000

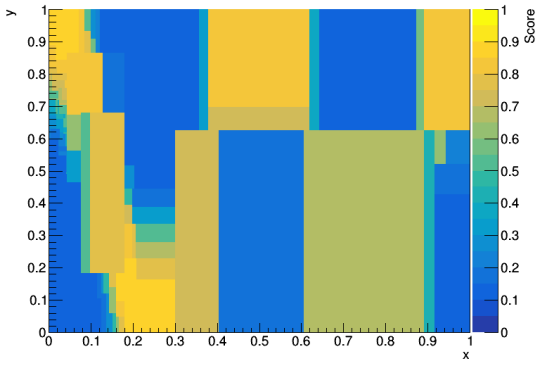
Figure D.2: Distribution of BDT score in (x, y) obtained by a single decision tree for different values of MinChildWeight. MaxDepth is fixed to infinity and MinLossReduction to zero, so that MinChildWeight alone controls the size of the tree.

Figure D.3 shows the BDT score as a function of training variables (x, y) for different values of MaxDepth. MinChildWeight is fixed to one and MinLossReduction to zero, so that MaxDepth alone controls the size of the tree. Figure D.2a shows a highly under-trained example: MaxDepth $= 3$, meaning that the tree has at most $2^3 = 8$ leaves. Higher values of MaxDepth allow for more leaves, and better resolution of the signal (see Figure D.2d).

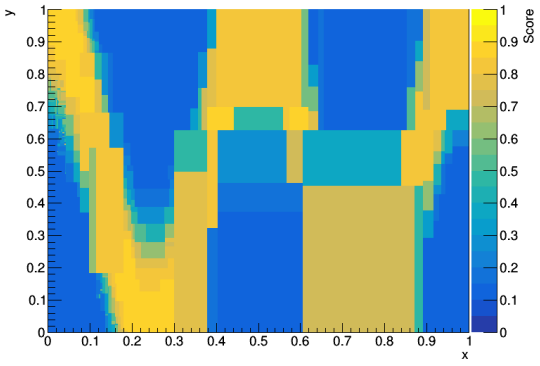


(a) MaxDepth = 3

(b) MaxDepth = 5



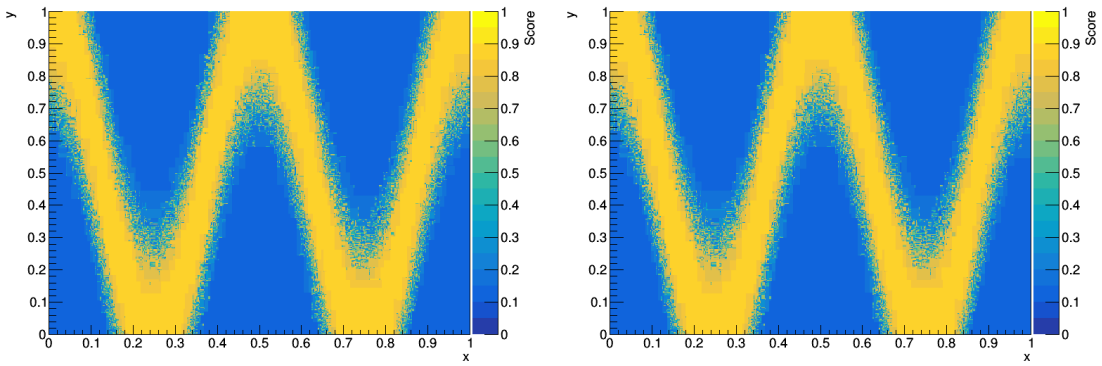
(c) MaxDepth = 7



(d) MaxDepth = 9

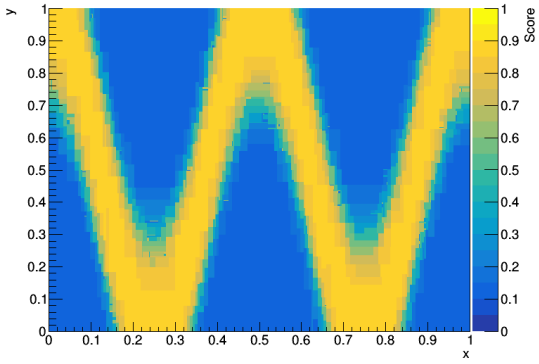
Figure D.3: Distribution of BDT score in (x, y) obtained by a single decision tree for different values of MaxDepth. MinChildWeight is fixed to one and MinLossReduction to zero, so that MaxDepth alone controls the size of the tree.

Figure D.4 shows the BDT score as a function of training variables (x, y) for different values of MinLossReduction. MinChildWeight is fixed to one and MaxDepth to infinity, so that MinLossReduction alone controls the size of the tree. Figure D.4a shows an over-trained example: MinLossReduction = 0, and each leaf contains a single training event. Figure D.4d shows an under-trained example: MinLossReduction = 100, and the tree does not grow large enough to resolve the signal well.

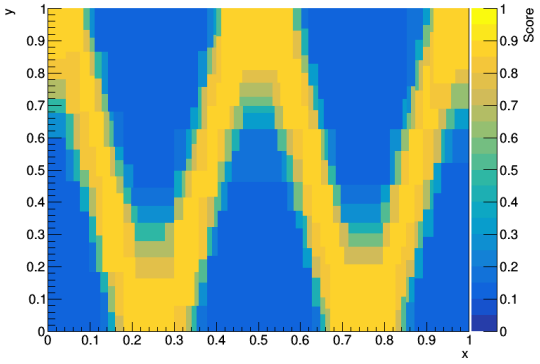


(a) MinLossReduction = 0

(b) MinLossReduction = 1



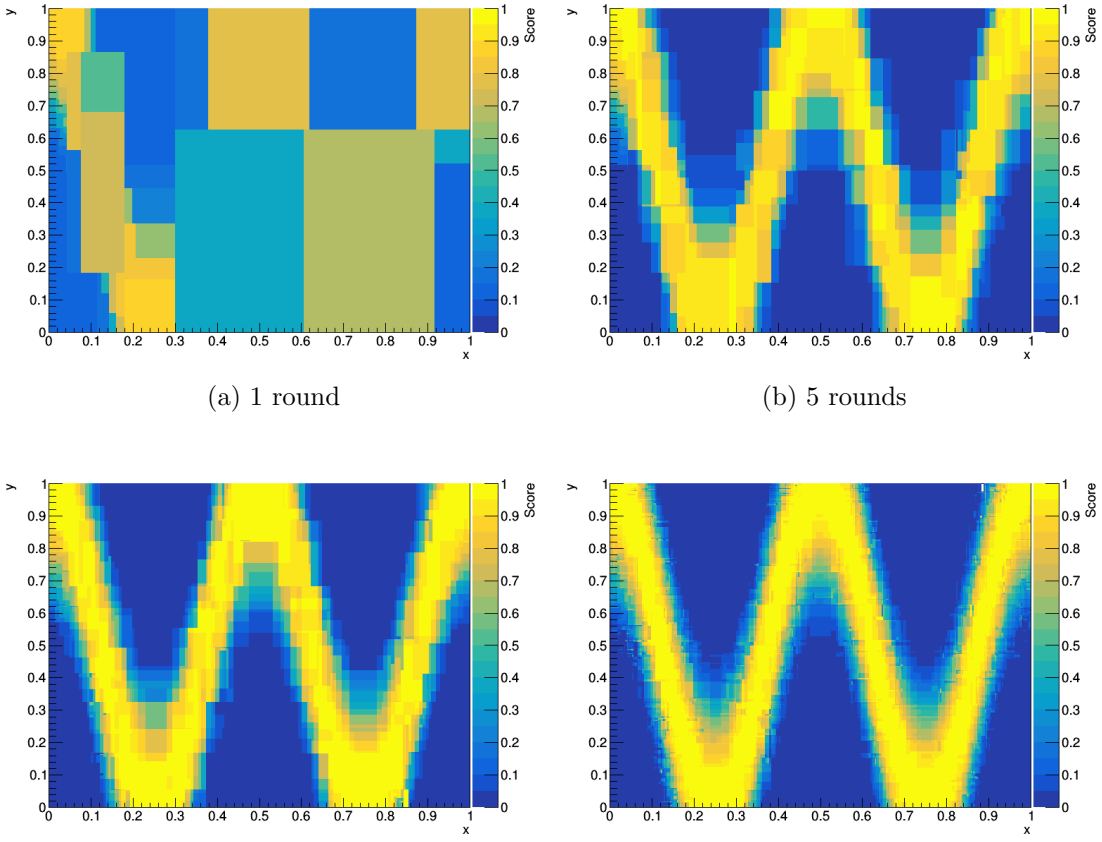
(c) MinLossReduction = 10



(d) MinLossReduction = 100

Figure D.4: Distribution of BDT score in (x, y) obtained by a single decision tree for different values of MinLossReduction. MinChildWeight is fixed to one and MaxDepth to infinity, so that MinLossReduction alone controls the size of the tree.

The power of gradient boosting is illustrated in Figure D.5, which shows the BDT score as a function of training variables (x, y) at different steps in the boosting process. MaxDepth is fixed to six, MinChildWeight to one, and MinLossReduction to zero so that the depth controls the size of the tree. After only 5 rounds, the signal can be resolved with high probability.



(a) 1 round

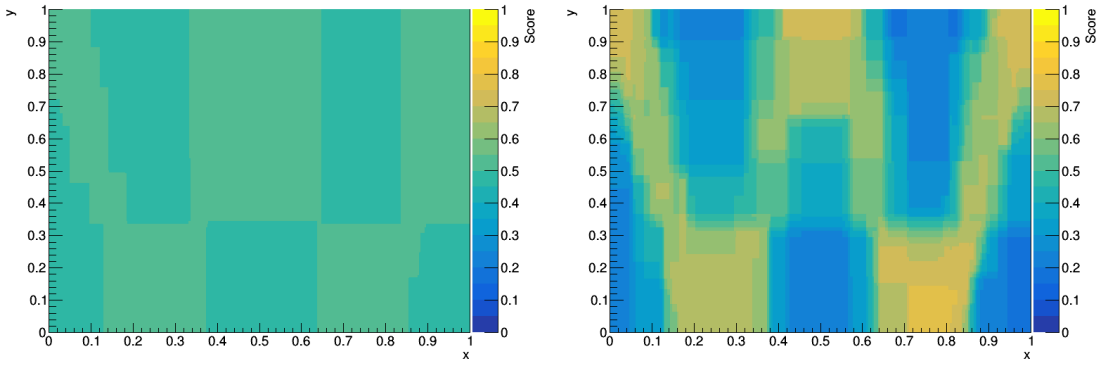
(b) 5 rounds

(c) 10 rounds

(d) 50 rounds

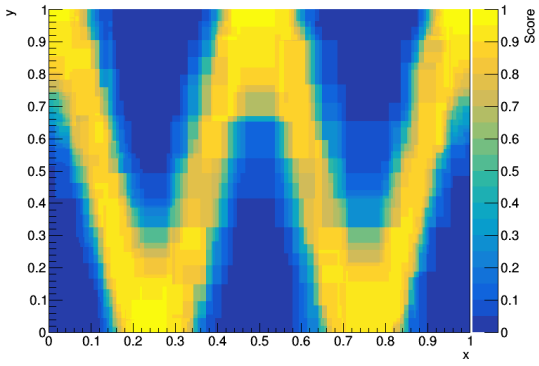
Figure D.5: Distribution of BDT score in (x, y) obtained after different numbers of boosting rounds. MaxDepth is fixed to six, MinChildWeight to one and MinLossRed to zero, so that the depth controls the size of the trees.

Figure D.6 shows the BDT score as a function of training variables (x, y) for different values of the LearningRate. MaxDepth is fixed to six, MinChildWeight to one, and MinLossReduction to zero so that the depth controls the size of the tree. With larger values of the LearningRate, the training converges faster and the separation between signal and background becomes starker after the same number of boosting rounds (10).

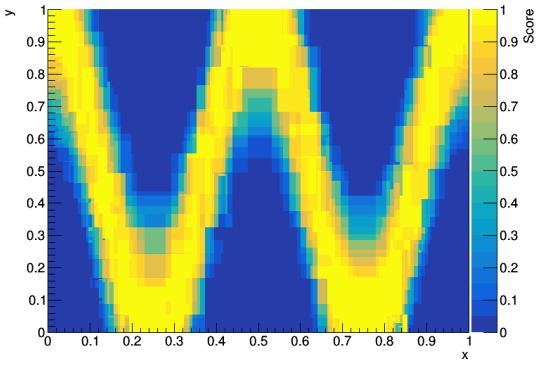


(a) LearningRate = 0.01

(b) LearningRate = 0.1



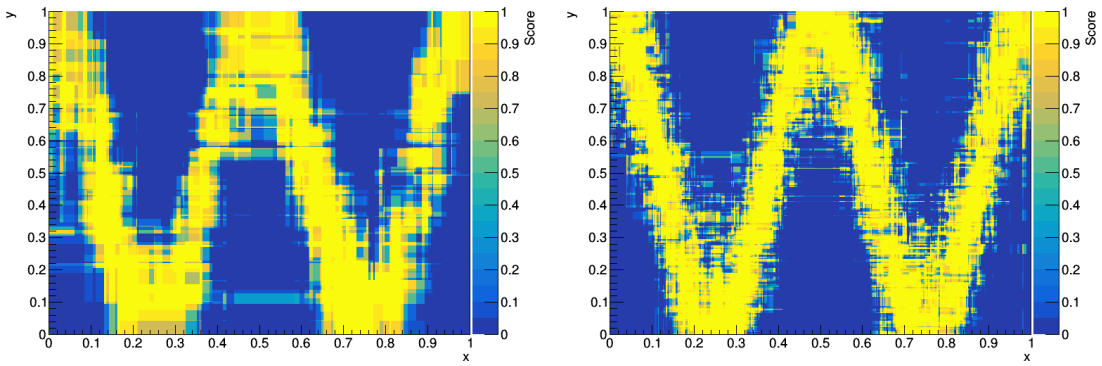
(c) LearningRate = 0.5



(d) LearningRate = 1.0

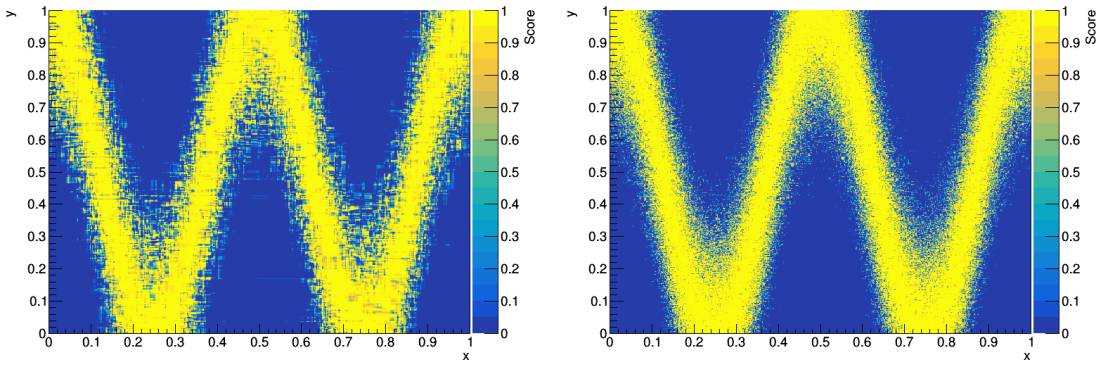
Figure D.6: Distribution of BDT score in (x, y) obtained after 10 boosting rounds for different values of the LearningRate. MaxDepth is fixed to six, MinChildWeight to one, and MinLossRed to zero, so that the depth controls the size of the trees.

Figure D.7 shows the BDT score as a function of training variables (x, y) for different values of the Subsample parameter. MaxDepth is fixed to infinity, MinChildWeight to one, and MinLossReduction to zero, a scenario that results in severe overtraining in a single tree (see Figure D.2a). The Subsample parameter can mitigate the effects of overtraining by using a different subset of the training data in each boosting round. Smaller values of Subsample correspond to less overlap in the training samples, and therefore less overtraining.



(a) Subsample = 0.001

(b) Subsample = 0.01

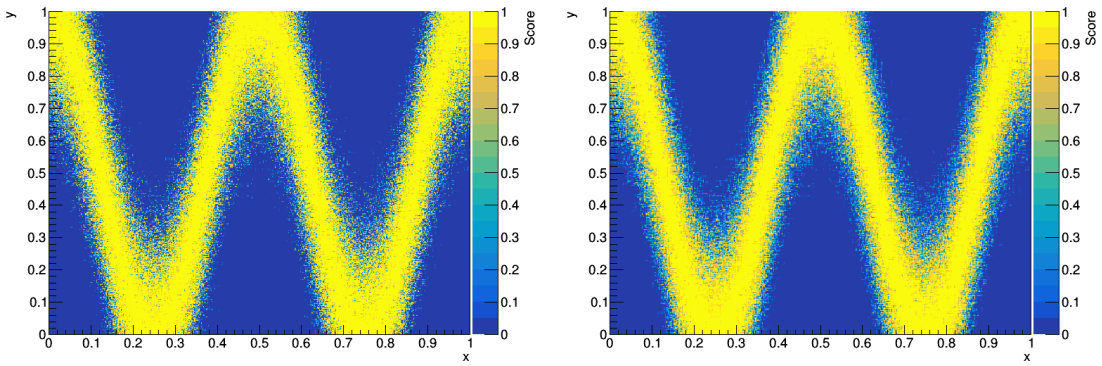


(c) Subsample = 0.1

(d) Subsample = 1

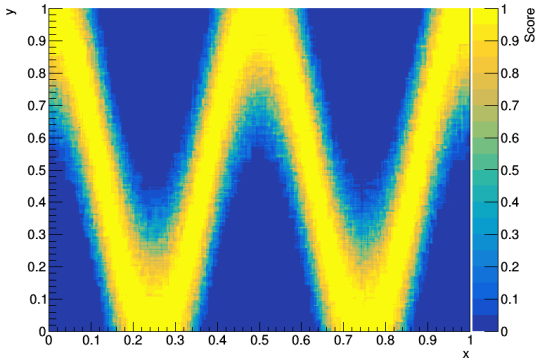
Figure D.7: Distribution of BDT score in (x, y) obtained after 10 boosting rounds for different values of the Subsample parameter. MaxDepth is fixed to infinity, MinChildWeight to one, and MinLossRed to zero, so that the depth controls the size of the trees. Subsample alone suppresses overtraining.

Figure D.8 shows the BDT score as a function of training variables (x, y) for different values of the L1 Regularization parameter α. MaxDepth is fixed to infinity, MinChildWeight to one, and MinLossReduction to zero, a scenario that results in severe overtraining in a single tree. The Subsample parameter is fixed to one. The α parameter can mitigate the effects of overtraining by restricting the magnitude of the leaf weights. Larger values of α correspond to tighter restrictions on the weights, and therefore less overtraining.

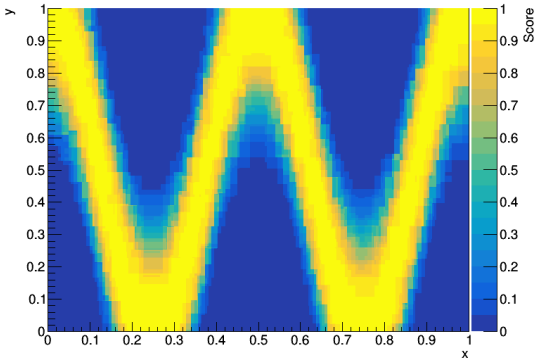


(a) L1 Reg $\alpha = 0$

(b) L1 Reg $\alpha = 1$



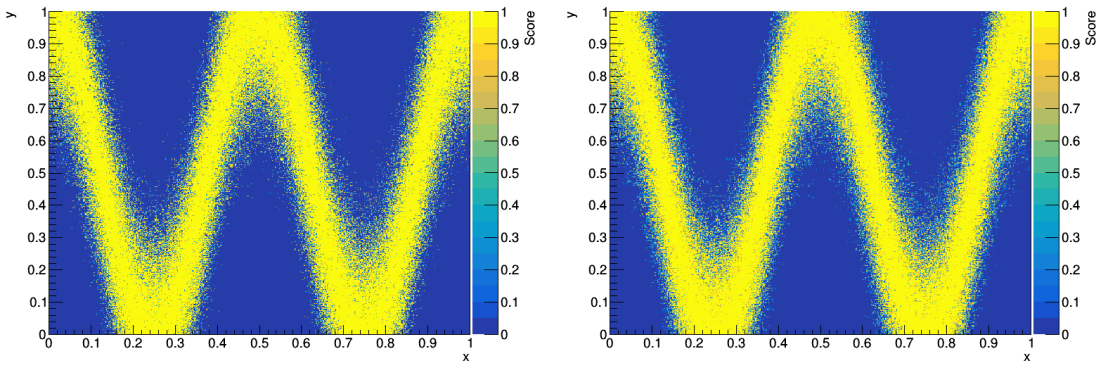
(c) L1 Reg $\alpha = 10$



(d) L1 Reg $\alpha = 100$

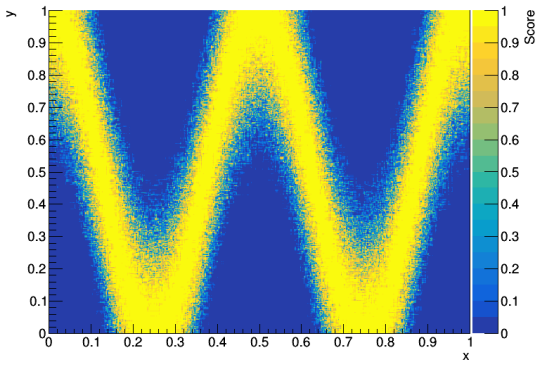
Figure D.8: Distribution of BDT score in (x, y) obtained after 10 boosting rounds for different values of the L1 Regularization parameter α. MaxDepth is fixed to infinity, MinChildWeight to one, and MinLossRed to zero, so that the depth controls the size of the trees. Subsample is set to one, and α alone suppresses overtraining.

Figure D.9 shows the BDT score as a function of training variables (x, y) for different values of the L2 Regularization parameter λ. Again, MaxDepth is fixed to infinity, MinChildWeight to one, and MinLossReduction to zero, a scenario that results in severe overtraining in a single tree. The Subsample parameter is fixed to one. The λ parameter can mitigate the effects of overtraining by restricting the magnitude of the leaf weights. Larger values of λ correspond to tighter restrictions on the weights, and therefore less overtraining.

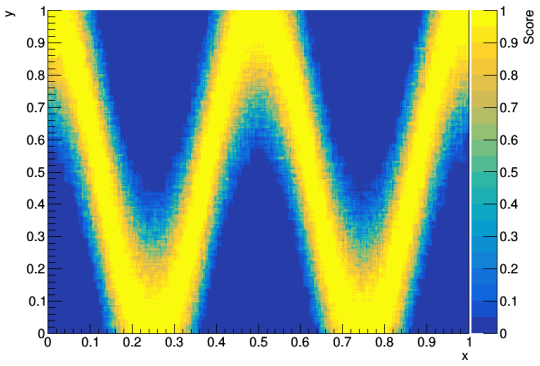


(a) L2 Reg $\lambda = 0$

(b) L2 Reg $\lambda = 1$



(c) L2 Reg $\lambda = 10$

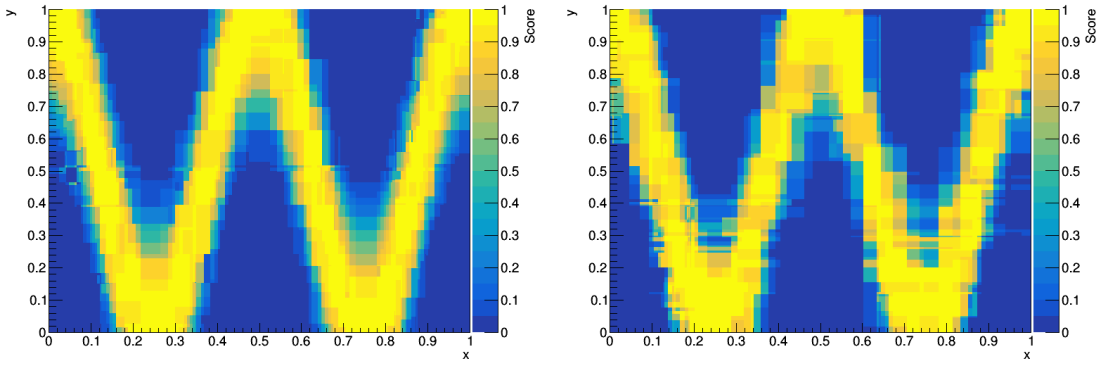


(d) L2 Reg $\lambda = 100$

Figure D.9: Distribution of BDT score in (x, y) obtained after 10 boosting rounds for different values of the L2 Regularization parameter λ. MaxDepth is fixed to infinity, MinChildWeight to one, and MinLossRed to zero, so that the depth controls the size of the trees. Subsample is set to one and α to zero, and λ alone suppresses overtraining.

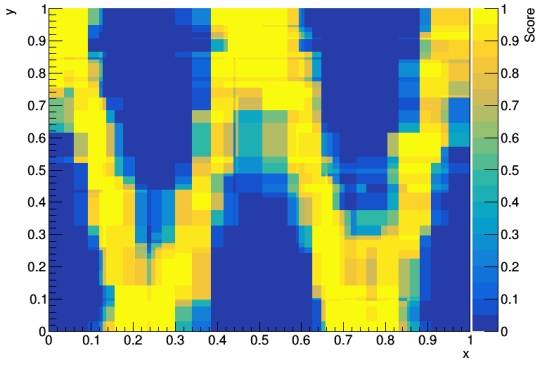
D.2 Dataset Size

The performance of a BDT decreases as training statistics become very low. Figure D.10 demonstrates the worsening performance of a BDT trained on datasets of decreasing size, all obeying Equation D.1. The hyper-parameters are identical for each training dataset, and tuned so that overtraining is not an issue. More training data gives a better approximation of Equation D.1 (Figure D.10a), and with too small a training set Equation D.1 is not learned at all (Figure D.10d).

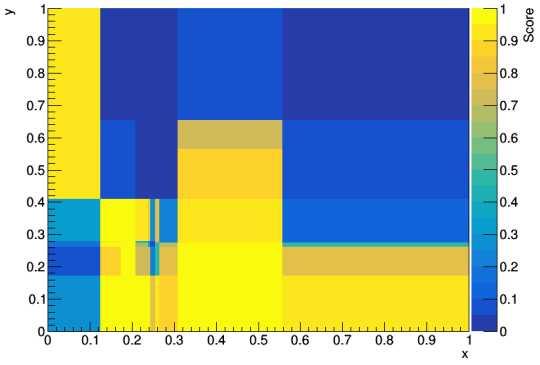


(a) 1,000,000 points

(b) 10,000 points



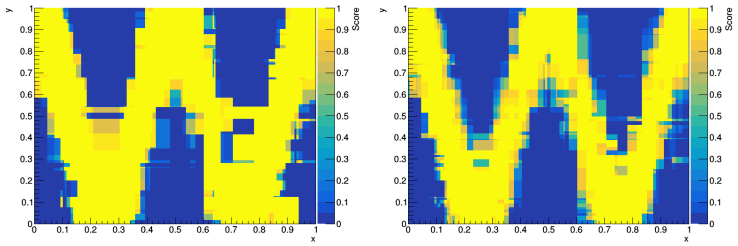
(c) 2,500 points



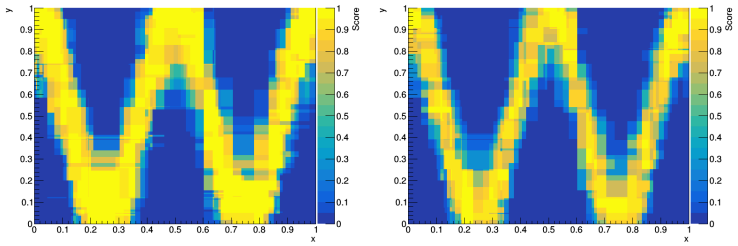
(d) 100 points

Figure D.10: Distribution of BDT score in (x, y) obtained from different numbers of randomly generated training points. More training data gives a better approximation of the desired function shown in Figure D.1a. The hyper-parameters are identical for each training dataset, and tuned so that overtraining is not an issue.

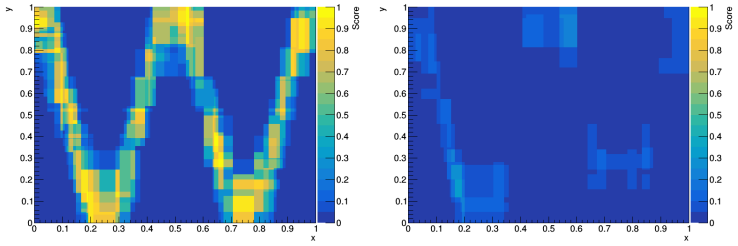
A realistic training dataset can often be very asymmetric: the ratio of signal events to background is very small in measurements of rare physics process such as $ttH(\gamma\gamma)$. Figure D.11 shows the change in BDT output as a function of the RelativeNorm parameter, which gives the relative normalization of signal and background (S/B). In cases where S/B is very large (Figure D.11a, b), the BDT score is more often signal-like. In the case where S/B is very small (Figure D.11e, f), the BDT score is more often background-like.



(a) $S/B = 100$ (b) $S/B = 10$



(c) $S/B = 2$ (d) $S/B = 0.5$



(e) $S/B = 0.1$ (f) $S/B = 0.01$

Figure D.11: Distribution of BDT score in (x, y) obtained for the 10,000 event training set, with different relative normalization of signal (S) and background (B). In the case where S/B is very large (a), the BDT score is more often signal-like (yellow). In the case where S/B is very small (f), the BDT score is more often background-like (blue).

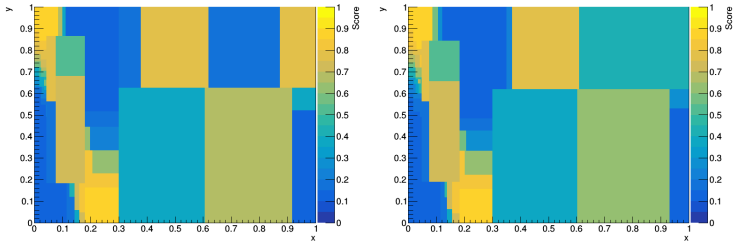
D.3 Training Packages

The first application of machine learning to $t\bar{t}H(\gamma\gamma)$ categorization [94] used the TMVA package [175] to train a simple five-variable BDT with gradient boosting. Subsequent more sophisticated analyses (see Chapters 9 and 11) have relied primarily on the XGBoost package [140].

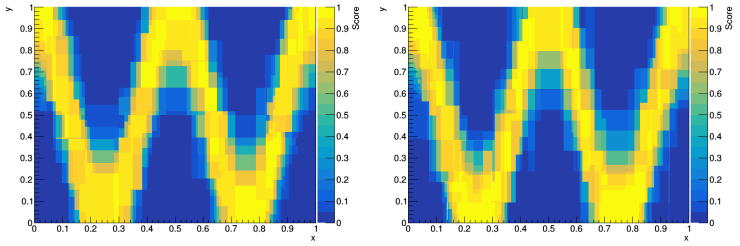
	XGBoost	TMVA
Rounds	`num_rounds`	`NTrees`
MaxDepth	`max_depth`	`MaxDepth`
MinChildWeight	`min_child_weight`	`MinNodeSize`
MinLossReduction	`gamma`	–
Regularization	`lambda`, `alpha`	–
LearningRate	`eta`	`Shrinkage`
Subsample	`subsample`	`BaggedSampleFraction`

Table D.1: Summary of hyper-parameter names in different BDT training packages.

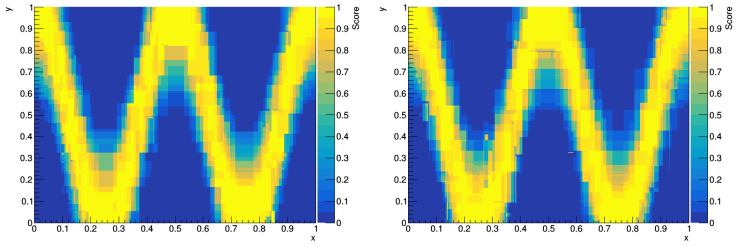
A comparison of the two algorithms is shown in Figure D.12. The same binary training data as above (1,000,000 points) is used in all training comparisons. Hyper-parameters in each package are tuned to identical values (as far as is possible) according to Table D.1. MaxDepth is fixed to six, MinChildWeight to one, and MinLossRed to zero, so that the depth controls the size of the tree. Subsample is fixed to one, and LearningRate is fixed to one. The performance appears to be very similar between the two training packages.



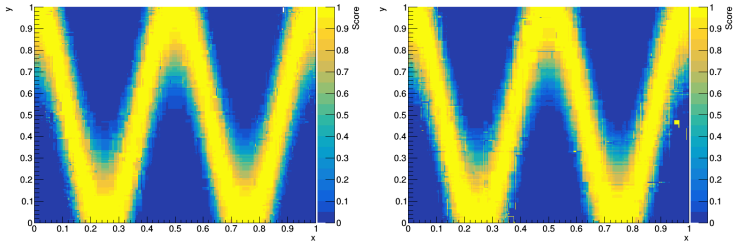
(a) XGBoost - 1 round (b) TMVA - 1 round



(c) XGBoost - 5 rounds (d) TMVA - 5 rounds



(e) XGBoost - 10 rounds (f) TMVA - 10 rounds



(g) XGBoost - 50 rounds (h) TMVA - 50 rounds

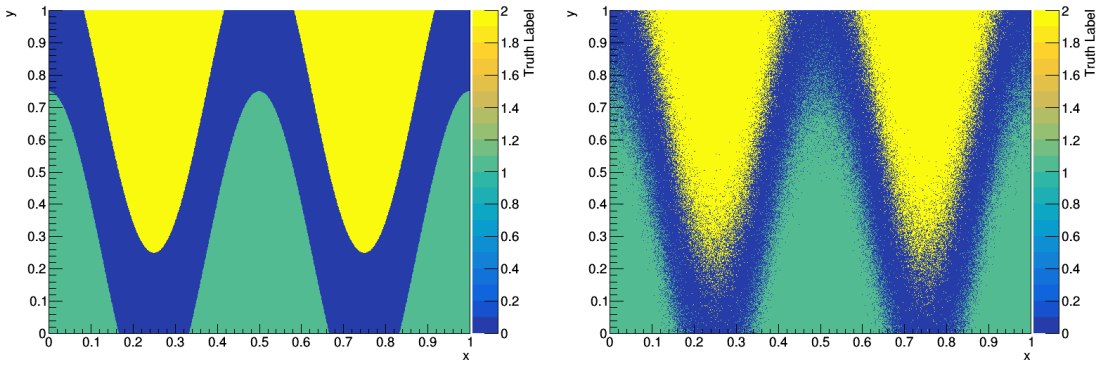
Figure D.12: Distribution of BDT score in (x, y) obtained from XGBoost (left) and TMVA (right) after different numbers of boosting rounds. The hyper-parameters are tuned to identical values according to Table D.1. MaxDepth is fixed to six, MinChildWeight to one, and MinLossRed to zero. Subsample is fixed to one, and LearningRate is fixed to one.

D.4 Multiclassification

The multiclass toy dataset again contains only two features, x and y, for ease of visualization. The training data is composed of 1,000,000 random points in (x, y), with three regions (Label $\in (0, 1, 2)$) defined by the following:

$$\text{Label} = g(x, y) = \begin{cases} 2 \text{ if } y > A\cos(n\pi x) + y_0 \\ 1 \text{ if } y < A\cos(n\pi x) + y_0 + 1 \\ 0 \text{ otherwise} \end{cases} \tag{D.2}$$

The distribution of signal and background in the (x, y) plane is shown in Figure D.13a for $A = 0.5$, $n = 4$ and $y_0 = 0.5$. As in the binary dataset, some small random noise is introduced: Label $= g(x, y + \epsilon)$, where ϵ is a random number generated according to a gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 0.1$. A distribution of the three classes with the addition of this random noise is shown in Figure D.13b.

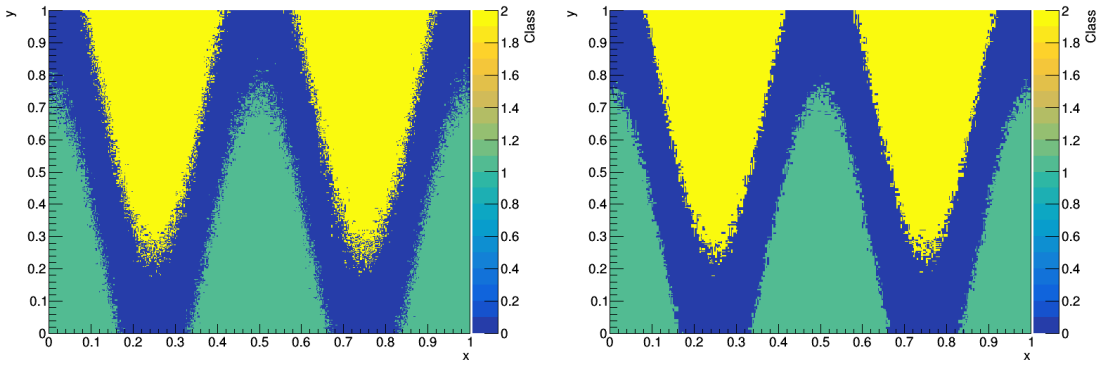


(a) (b)

Figure D.13: Distribution in (x, y) of a sample three-class dataset without random noise (a) and with random noise (b). Yellow represents Label = 2, green represents Label = 1, and blue represents Label = 0, according to Equation D.2 with $A = 0.5$, $n = 4$ and $y_0 = 0.5$.

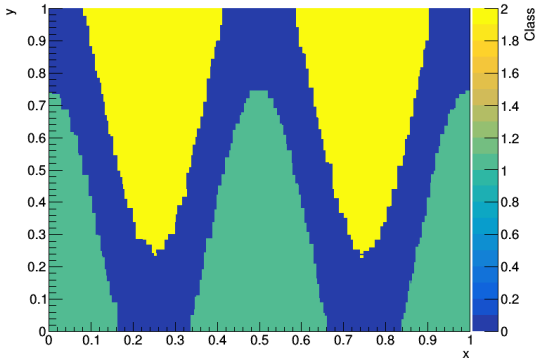
A single tree (Rounds = 1) is trained in XGBoost on the example dataset. The BDT performance is tested for range of values for each tree hyper-parameter.

Figure D.15 shows the BDT "softprob" score as a function of training variables (x, y) for different values of MinChildWeight. MaxDepth is fixed to infinity and MinLossReduction to zero, so that MinChildWeight alone controls the size of the tree. Figure D.15a shows an over-trained example: MinChildWeight = 1, and each leaf contains a single training event. Figure D.15d shows an under-trained example: MinChildWeight = 1,000, and the tree does not grow large enough to resolve the signal well. The assigned class label (corresponding to $\max\{P_0, P_1, P_2\}$) for each (x, y) pair is shown in Figure D.14.

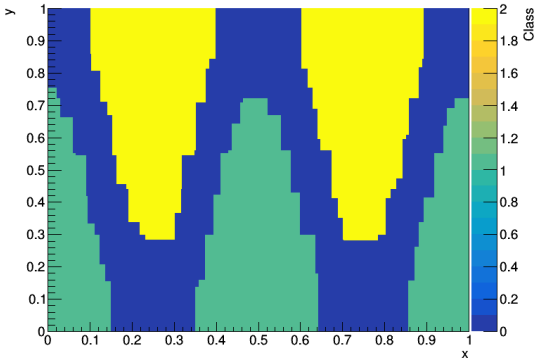


(a) MinChildWeight = 1

(b) MinChildWeight = 10

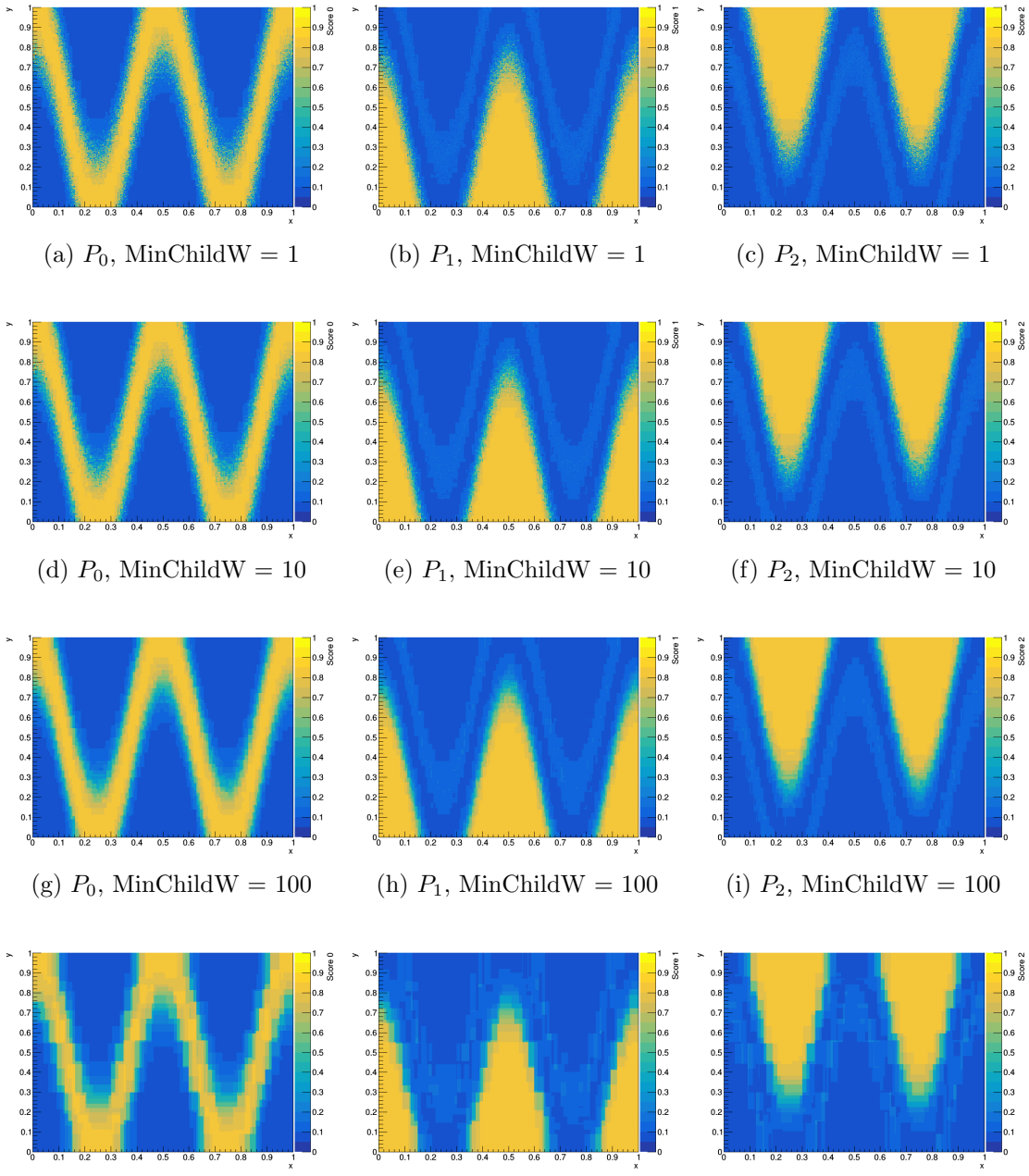


(c) MinChildWeight = 100



(d) MinChildWeight = 1,000

Figure D.14: Distribution of assigned class label in (x, y) obtained by a single decision tree for different values of MinChildWeight. MaxDepth is fixed to infinity and MinLossReduction to zero, so that MinChildWeight alone controls the size of the tree.



(a) P_0, MinChildW = 1 (b) P_1, MinChildW = 1 (c) P_2, MinChildW = 1

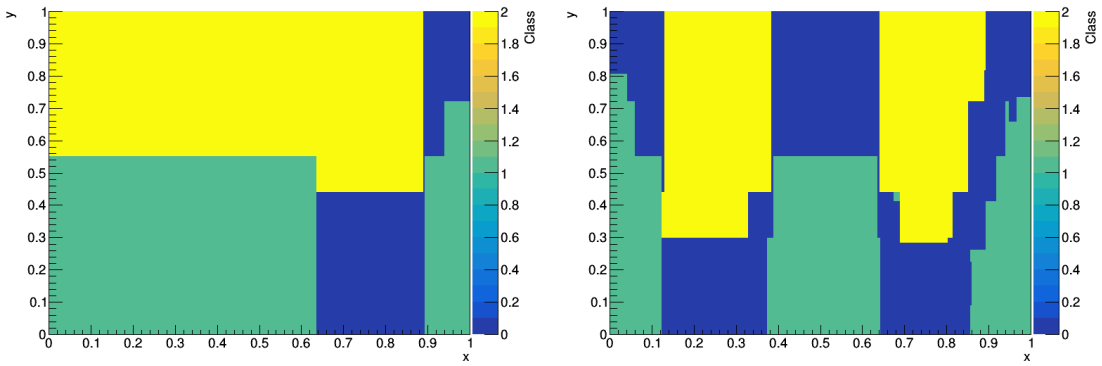
(d) P_0, MinChildW = 10 (e) P_1, MinChildW = 10 (f) P_2, MinChildW = 10

(g) P_0, MinChildW = 100 (h) P_1, MinChildW = 100 (i) P_2, MinChildW = 100

(j) P_0, MinChildW = 1,000 (k) P_1, MinChildW = 1,000 (l) P_2, MinChildW = 1,000

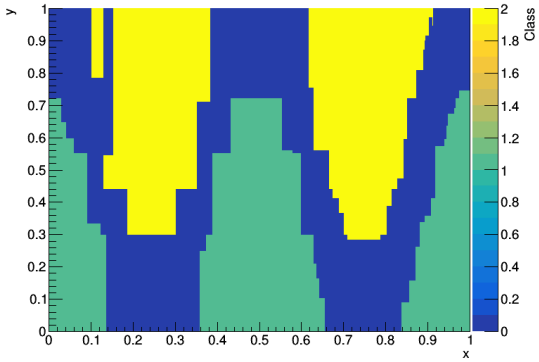
Figure D.15: Distribution of "softprob" BDT scores in (x, y) obtained by a single decision tree for different values of MinChildWeight. MaxDepth is fixed to infinity and MinLossReduction to zero, so that MinChildWeight alone controls the size of the tree.

Figure D.17 shows the BDT "softprob" score for each class as a function of training variables (x, y) for different values of MaxDepth. MinChildWeight is fixed to one and MinLossReduction to zero, so that MaxDepth alone controls the size of the tree. Figure D.15a shows a highly under-trained example: MaxDepth = 3, meaning that the tree has at most $2^3 = 8$ leaves. The assigned class label (corresponding to $\max\{P_0, P_1, P_2\}$) for each (x, y) pair is shown in Figure D.16.

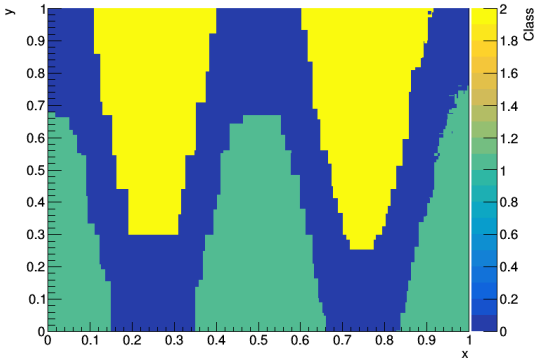


(a) MaxDepth = 3

(b) MaxDepth = 5

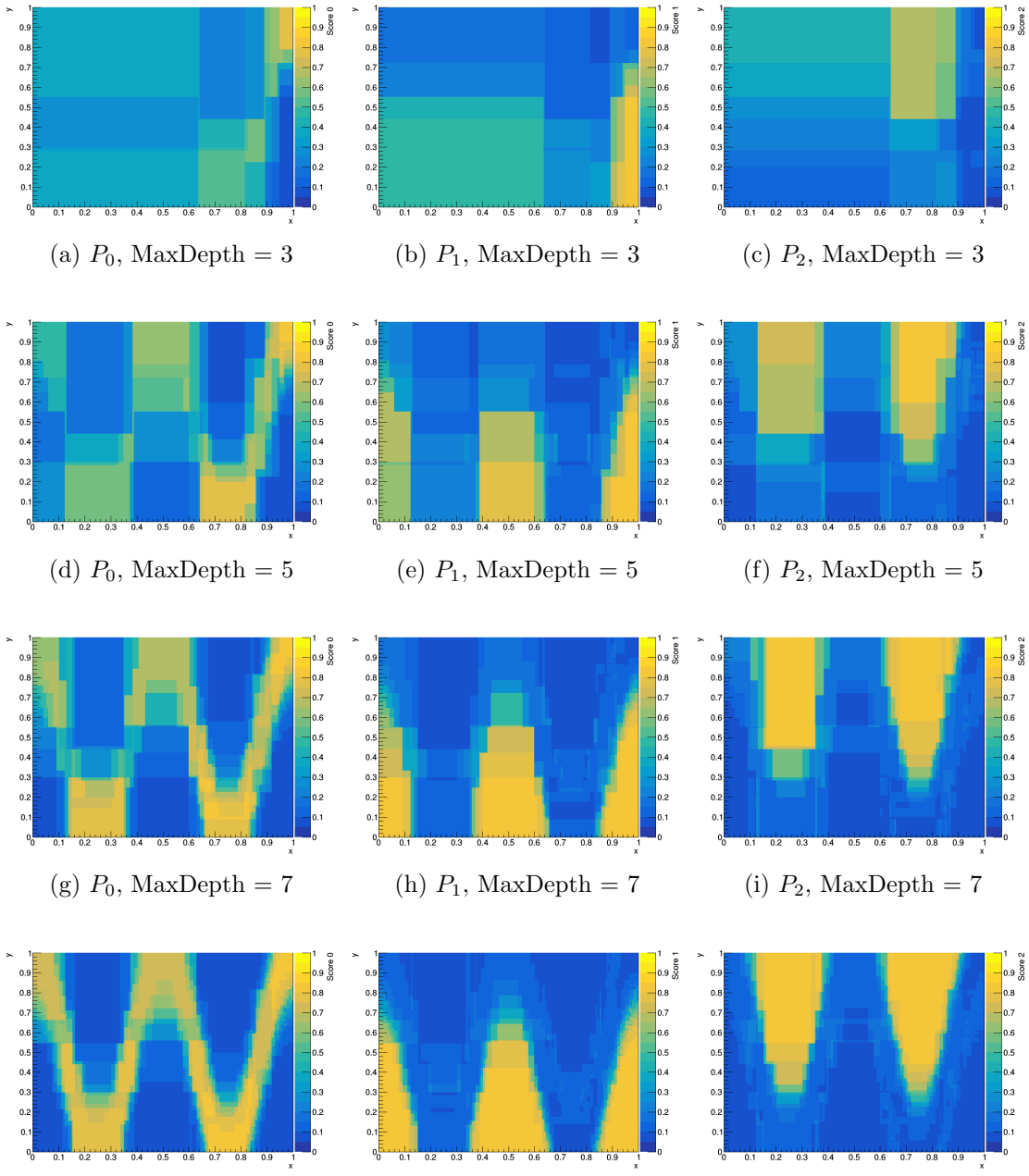


(c) MaxDepth = 7



(d) MaxDepth = 9

Figure D.16: Distribution of assigned class label in (x, y) obtained by a single decision tree for different values of MaxDepth. MinChildWeight is fixed to one and MinLossReduction to zero, so that MaxDepth alone controls the size of the tree.



(a) P_0, MaxDepth = 3 (b) P_1, MaxDepth = 3 (c) P_2, MaxDepth = 3

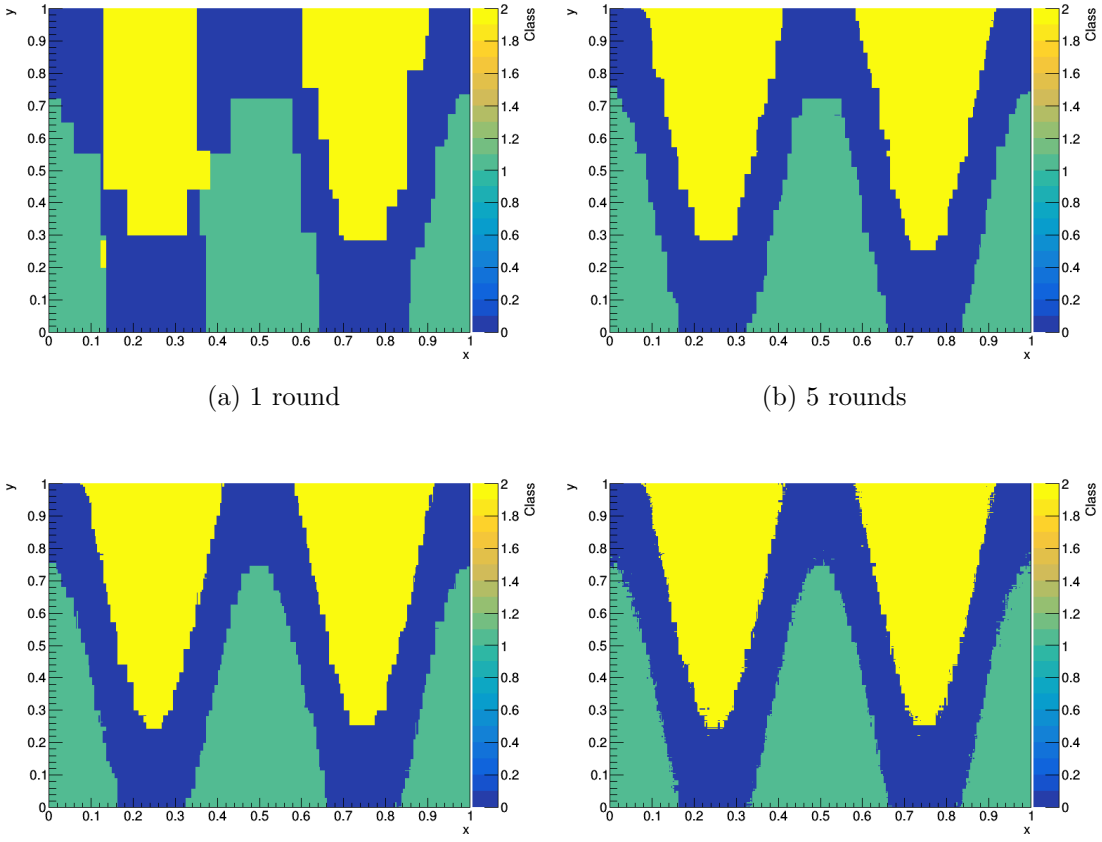
(d) P_0, MaxDepth = 5 (e) P_1, MaxDepth = 5 (f) P_2, MaxDepth = 5

(g) P_0, MaxDepth = 7 (h) P_1, MaxDepth = 7 (i) P_2, MaxDepth = 7

(j) P_0, MaxDepth = 9 (k) P_1, MaxDepth = 9 (l) P_2, MaxDepth = 9

Figure D.17: Distribution of "softprob" BDT scores in (x, y) obtained by a single decision tree for different values of MaxDepth. MinChildWeight is fixed to one and MinLossReduction to zero, so that MaxDepth alone controls the size of the tree.

The power of gradient boosting is illustrated in Figure D.19, which shows the BDT "softprob" score as a function of training variables (x, y) at different steps in the boosting process. After only 5 rounds, the three classes can be resolved with high probability. The assigned class label (corresponding to $\max\{P_0, P_1, P_2\}$) for each (x, y) pair is shown in Figure D.18 as a function of number of boosting rounds.



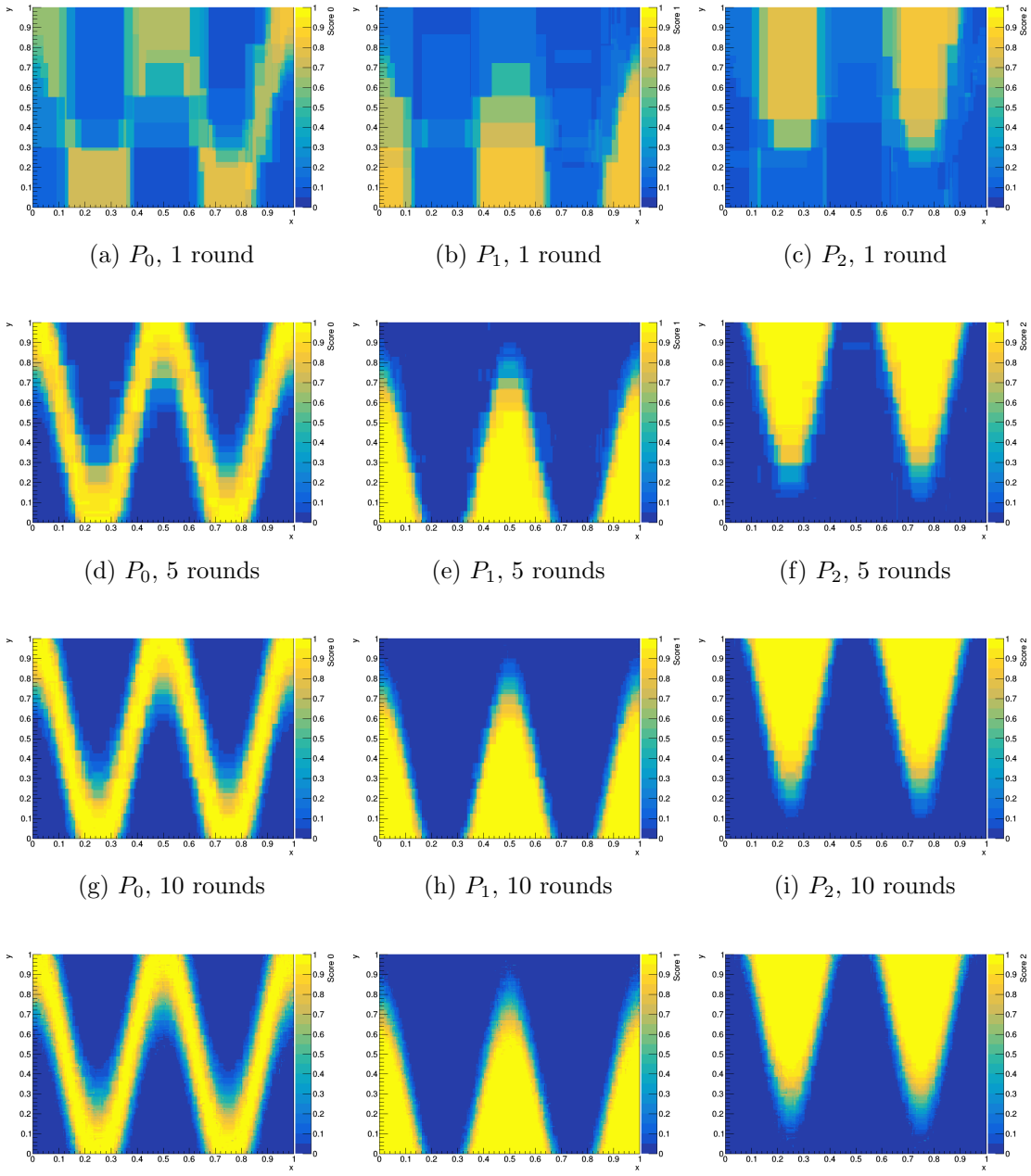
(a) 1 round

(b) 5 rounds

(c) 10 rounds

(d) 50 rounds

Figure D.18: Distribution of assigned class label in (x, y) obtained after different numbers of boosting rounds. MaxDepth is fixed to six, MinChildWeight to one, and MinLossRed to zero, so that the depth controls the size of the tree.



(a) P_0, 1 round (b) P_1, 1 round (c) P_2, 1 round

(d) P_0, 5 rounds (e) P_1, 5 rounds (f) P_2, 5 rounds

(g) P_0, 10 rounds (h) P_1, 10 rounds (i) P_2, 10 rounds

(j) P_0, 50 rounds (k) P_1, 50 rounds (l) P_2, 50 rounds

Figure D.19: Distribution of "softprob" BDT scores in (x, y) obtained after different numbers of boosting rounds. MaxDepth is fixed to six, MinChildWeight to one, and MinLossRed to zero, so that the depth controls the size of the trees.

Appendix E

$t\bar{t}H$ in the diphoton decay channel (79.8 fb^{-1})

The $t\bar{t}H(\gamma\gamma)$ categorization described in Section 9.2 was developed and optimized on the partial LHC Run 2 dataset of 79.8 fb^{-1}, collected during 2015-2017. This Appendix describes the results on this partial dataset, which are superseded by the full Run 2 results in Section 10.2.

E.1 Signal and Background Model

In each analysis category, the signal shape is determined by fitting a Double-Sided Crystal Ball (DSCB) function to a sample of inclusive Higgs boson Monte Carlo (see Section 6.2.1). The best fit DSCB parameters for each category are shown in Table E.1. Differences in the signal parameters with respect to the full Run 2 analysis (Table 10.2), particularly the signal resolution σ_{CB}, result from the addition of a high-pileup Monte Carlo sample representative of the data taking conditions in 2018.

For the hadronic $t\bar{t}H(\gamma\gamma)$ categories, the functional form of the continuum background is extracted from templates created from NTI data. To enhance statistics in these templates, events with no b-tagged jets are also included, and the b-tag status of all central jets are set to "true" when computing the BDT score. For the leptonic $t\bar{t}H(\gamma\gamma)$ categories, the background templates are constructed from $t\bar{t}+\gamma\gamma$ Monte Carlo, since this process dominates the continuum background once the BDT selection has been applied. In order to enhance statistics, no photon ID or isolation cuts are applied.

The functional form of the continuum background and the spurious signal are listed in Table E.2.

Category	μ_{CB} [GeV]	σ_{CB} [GeV]	α_{high}	α_{low}	n_{high}	n_{low}
$t\bar{t}H(\gamma\gamma)$ had 4	125.27	1.63	1.9	1.9	3.8	3.4
$t\bar{t}H(\gamma\gamma)$ had 3	125.27	1.59	2.0	1.8	3.8	4.3
$t\bar{t}H(\gamma\gamma)$ had 2	125.27	1.46	1.6	1.6	5.5	4.9
$t\bar{t}H(\gamma\gamma)$ had 1	125.24	1.32	1.7	1.7	6.5	4.5
$t\bar{t}H(\gamma\gamma)$ lep 3	125.26	1.73	1.9	1.7	3.6	4.9
$t\bar{t}H(\gamma\gamma)$ lep 2	125.24	1.68	1.9	1.8	4.2	4.4
$t\bar{t}H(\gamma\gamma)$ lep 1	125.22	1.45	1.7	1.6	5.3	4.9

Table E.1: Best fit parameter values for the signal $m_{\gamma\gamma}$ shape in each of the seven $t\bar{t}H(\gamma\gamma)$ categories at 79.8 fb^{-1}. Note that these are the fitted values for the MC samples at $m_H = 125$ GeV.

Category	Function	N_{sp}	χ^2/ndof
$t\bar{t}H(\gamma\gamma)$ had 4	Power Law	+0.96	0.98
$t\bar{t}H(\gamma\gamma)$ had 3	Power Law	+0.63	0.88
$t\bar{t}H(\gamma\gamma)$ had 2	Exponential	+0.36	0.72
$t\bar{t}H(\gamma\gamma)$ had 1	Power Law	+0.11	0.86
$t\bar{t}H(\gamma\gamma)$ lep 3	Exponential	+0.24	1.26
$t\bar{t}H(\gamma\gamma)$ lep 2	Power Law	+0.33	0.79
$t\bar{t}H(\gamma\gamma)$ lep 1	Power Law	+0.12	0.97

Table E.2: Results of the spurious signal test for the $t\bar{t}H(\gamma\gamma)$ categories at 79.8 fb^{-1}. The χ^2/ndof indicates the performance of a background only fit to the template in the $m_{\gamma\gamma}$ range 105-160 GeV, but is in general not a part of the spurious signal criteria.

E.2 Systematic Uncertainties

The experimental uncertainties included in the $t\bar{t}H(\gamma\gamma)$ measurement are those summarized in Section 6.4.2.

A conservative 100% heavy flavor uncertainty is applied to the predicted ggF, VBF, and VH yields in each category, and inclusive uncertainties from [34] are applied to the $b\bar{b}H$, $tHjb$, and tWH yields. For the $t\bar{t}H$ signal yields, QCD scale variations and PDF uncertainties (31 eigenvector scheme) are evaluated separately in each category. In addition, a UEPS uncertainty is calculated by comparing event yields predicted by Monte Carlo samples using different showering algorithms: $t\bar{t}H$ MG5_aMC@NLO+Pythia8 and MG5_aMC@NLO+Herwig++. The magnitudes of the UEPS uncertainty are summarized in Table 10.4.

The dominant uncertainty is the UEPS: since up to six jets in a single event are used in the BDT training (see Chapter 9), dependence on the parton showering algorithm is to be expected. The QCD scale uncertainty on the $t\bar{t}H$ yield is the second most dominant uncer-

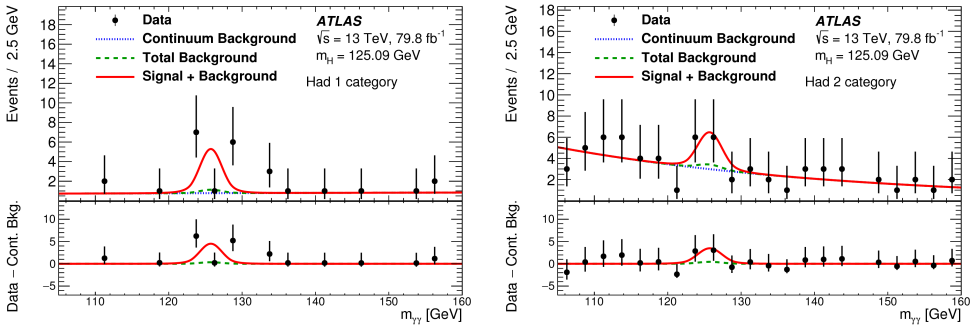
tainty. The photon energy scale and resolution are the dominant experimental systematics.

E.3 Results (79.8 fb^{-1})

The diphoton mass spectrum in each hadronic (leptonic) category is shown in Figure E.1 (Figure E.2) with the signal-plus-background fits overlaid. The observed $t\bar{t}H(\gamma\gamma)$ significance is 4.1σ, which corresponds to a signal strength of

$$\mu_{t\bar{t}H} = 1.41\ ^{+0.53}_{-0.44} \tag{E.1}$$

The expected $t\bar{t}H(\gamma\gamma)$ significance on this dataset (assuming the Standard Model) is 3.7σ (from pre-fit Asimov).



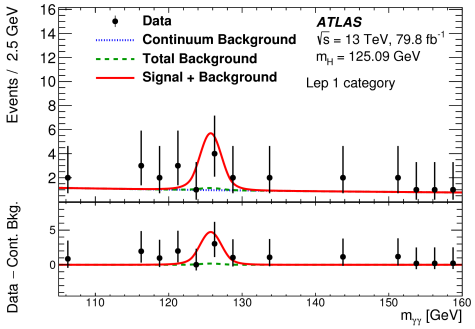
(a) $t\bar{t}H(\gamma\gamma)$ had 1

(b) $t\bar{t}H(\gamma\gamma)$ had 2

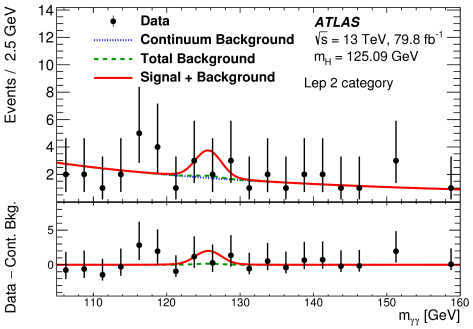
(c) $t\bar{t}H(\gamma\gamma)$ had 3

(d) $t\bar{t}H(\gamma\gamma)$ had 4

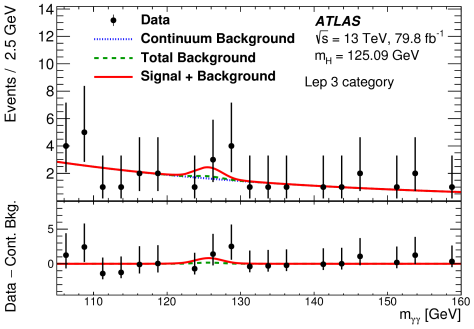
Figure E.1: Diphoton invariant mass spectrum in the four hadronic $t\bar{t}H(\gamma\gamma)$ categories with 79.8 fb^{-1} of ATLAS data. $t\bar{t}H(\gamma\gamma)$ had 1 (a) corresponds to the tightest cut on BDT score, and $t\bar{t}H(\gamma\gamma)$ had 4 (d) to the loosest. The data is shown in black, with the fitted continuum background distribution in blue, the total background (including non-$t\bar{t}H$ Higgs processes) in green, and the $t\bar{t}H$ signal plus background in red.



(a) $t\bar{t}H(\gamma\gamma)$ lep 1



(b) $t\bar{t}H(\gamma\gamma)$ lep 2



(c) $t\bar{t}H(\gamma\gamma)$ lep 3

Figure E.2: Diphoton invariant mass spectrum in the three leptonic $t\bar{t}H(\gamma\gamma)$ categories with 79.8 fb^{-1} of ATLAS data. $t\bar{t}H(\gamma\gamma)$ lep 1 (a) corresponds to the tightest cut on BDT score, and $t\bar{t}H(\gamma\gamma)$ lep 3 (c) to the loosest. The data is shown in black, with the fitted continuum background distribution in blue, the total background (including non-$t\bar{t}H$ Higgs processes) in green, and the $t\bar{t}H$ signal plus background in red.

Figure E.3 shows the fitted signal and background yields in each category, calculated in the $m_{\gamma\gamma}$ window containing 90% of the Higgs signal (S_{90}).

In order to visually enhance the contribution of the most powerful analysis categories, the data and fit in each category are weighted by $\ln(1 + S/B)$ where S (B) is the expected signal (background) in the smallest $m_{\gamma\gamma}$ window containing 90% of the Higgs signal. This weighting procedure serves to magnify the contribution to the $m_{\gamma\gamma}$ spectrum of the categories with highest S/B. The weighted diphoton mass spectrum (summed over all categories) is shown in Figure 10.8.

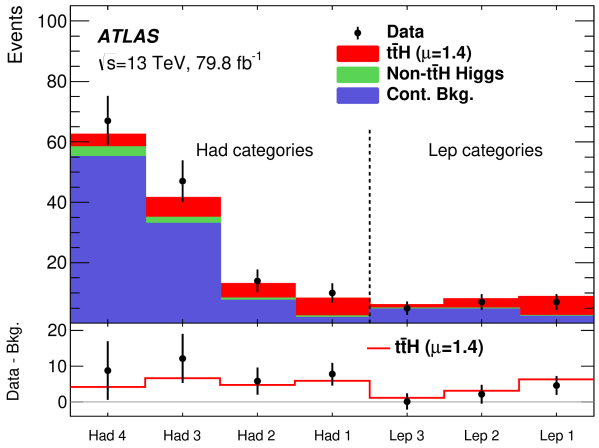


Figure E.3: Fitted number of $t\bar{t}H(\gamma\gamma)$ and background events in each analysis category with 79.8 fb^{-1} of ATLAS data. The fitted continuum background yield is shown in blue, the non-$t\bar{t}H$ Higgs background in green, and the $t\bar{t}H$ signal (normalized to the fitted $\mu = 1.41$) in red. The total number of data events per category is overlaid in black.

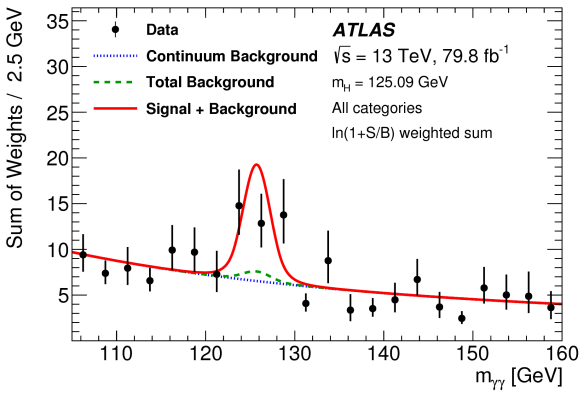


Figure E.4: Weighted diphoton invariant mass spectrum of all seven $t\bar{t}H(\gamma\gamma)$ categories with 79.8 fb^{-1} of ATLAS data. The data is shown in black, with the fitted continuum background distribution in blue, the total background (including non-$t\bar{t}H$ Higgs processes) in green, and the $t\bar{t}H$ signal plus background in red. The data and fit in each category are weighted by $\ln(1 + S/B)$, where S (B) is the expected signal (background) in the smallest $m_{\gamma\gamma}$ window containing 90% of the Higgs signal.

E.4 Top reconstruction in the 79.8 fb^{-1} analysis

At the time of the 79.8 fb^{-1} analysis, the BDT algorithm used to reconstruct top quarks was in an early stage of development. This Early Top Reco BDT differs from the finalized Top Reco BDT described in Chapter 8 in three major ways:

1. **Background sample**: The Early Top Reco BDT was designed for discrimination between $t\bar{t}H$ and non-top backgrounds, so the set of background triplets was constructed from all combinations of three jets in the NTI data control sample. The final Top Reco BDT was designed to correctly identify the jet triplet corresponding to the top in as many $t\bar{t}H$ events as possible, so the set of background triplets was constructed from all non truth-matched triplets in $t\bar{t}H$ Monte Carlo.

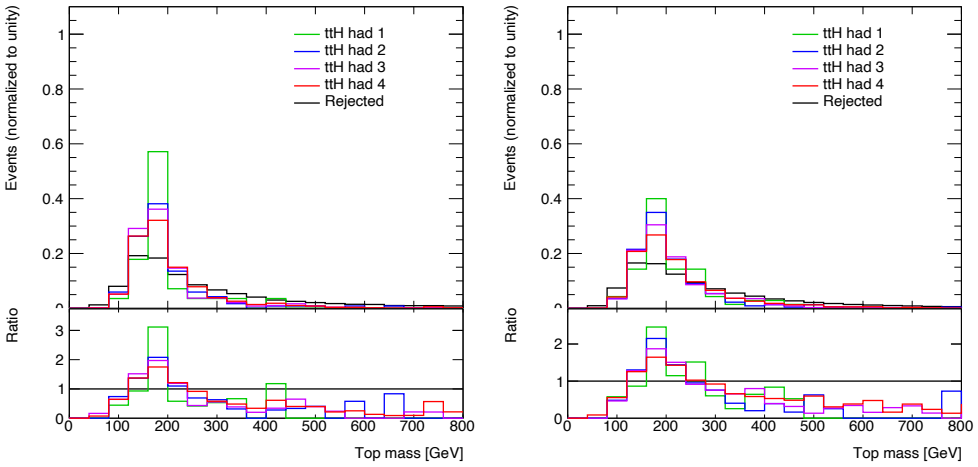
2. **Signal sample**: The Early Top Reco BDT required that the jet truth-matched to the bottom quark be b-tagged for all signal triplets. For the final Top Reco BDT, this requirement on the signal is lifted, and instead the b-tag scores of all three jets were used as training variables.

3. **Tri-jet mass**: The Early Top Reco BDT does not use the triplet mass as a training variable. In the final Top Reco BDT, the addition of this variable improves the accuracy by 2%.

The introduction of these changes increased the accuracy of the top reconstruction from only 19% in $t\bar{t}H$ events with the Early Top Reco BDT to 37% with the final Top Reco BDT described in Chapter 8.

Figure E.5 shows the mass of the reconstructed top quark in the NTI data control region as reconstructed by the Early (left) and final (right) Top Reco BDTs. The distribution is shown in each hadronic $t\bar{t}H(\gamma\gamma)$ category (recall from Chapter 9 that had 1 corresponds to the highest S/B), and the distribution of events passing hadronic preselection but not entering the BDT categories is shown in black.

The two Top Reco BDTs have similar qualitative behavior: as the S/B increases, the reconstructed top mass peaks more strongly at $m_t = 173$ GeV. However, because the triplet mass is used as a training variable in the final Top Reco, this BDT sculpts the top candidate mass into a sharper peak compared to the Early Top Reco.



(a) Final Top Reco (b) Early Top Reco

Figure E.5: Comparison of the Early Top Reco BDT and the Top Reco BDT described in Chapter 8 in hadronic NTI data events.

E.4.1 Decomposition of continuum background

In order to study the continuum background in the hadronic $t\bar{t}H$ categories at 79.8 fb^{-1}, the template fit procedure described in Chapter 9.3 was used to determine the relative contribution of $t\bar{t}\gamma\gamma$ and $\gamma\gamma$ + jets processes. The top quark mass reconstructed by the Early Top Reco BDT is chosen as the PDF distribution.

The Early Top Reco BDT does not suffer from sculpting of the top candidate mass distribution to the extent that the final Top Reco BDT does. As a result, templates built from $t\bar{t}\gamma\gamma$ and $\gamma\gamma$ + jets Monte Carlo samples have a larger shape difference in the Early Top Reco candidate mass than in the final Top Reco candidate mass. For the decomposition of the continuum background, it is therefore advantageous to use the Early Top Reco BDT, despite its lower top identification accuracy.

The template fit to Equation 9.8 are shown in Figure E.6. The fit is performed to data in the two hadronic categories with highest signal-to-background ratio (S/B). The top background fraction obtained is shown in Table 9.7, along with the uncertainty due to the statistics of the data. The statistical uncertainty on the Monte Carlo templates is small compared to the statistical uncertainty in data.

	$n^{SM}_{t\bar{t}H}/n_{\text{data}}$	Top bkg fraction
$t\bar{t}H(\gamma\gamma)$ had 1-2	10%	58% ± 19%

Table E.3: The top background fraction obtained in the two hadronic $t\bar{t}H(\gamma\gamma)$ categories with highest S/B at 79.8 fb^{-1}.

The top background fraction measured in 79.8 fb^{-1} of data using the Early Top Reco

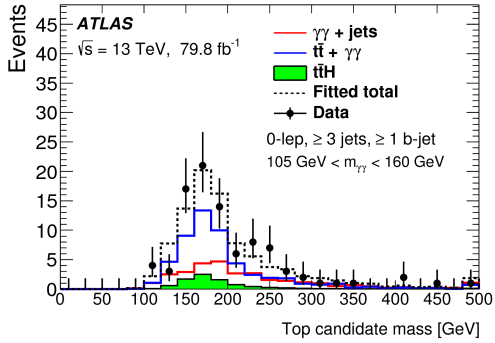


Figure E.6: Template fit to data in the two hadronic $t\bar{t}H(\gamma\gamma)$ categories with highest S/B at 79.8 fb^{-1}. The contributions from $t\bar{t}\gamma\gamma$ (blue) and $\gamma\gamma$+ jets (red) are normalized to their fitted values. The $t\bar{t}H$ contribution (green) is fixed to the SM expectation. The fitted combined PDF is shown in the dashed black line.

BDT is 58% ± 19%. As reported in Section 9.3, the top background fraction measured in 139 fb^{-1} of data using the final Top Reco BDT is 31% ± 16% (neglecting the uncertainty due to template statistics, for a proper comparison). The measurements agree within statistical errors. Because of the disadvantageous sculpting of the top mass templates constructed from the final Top Reco BDT, the precision on the top background fraction measurement does not improve much with the luminosity increase from 79.8 to 139 fb^{-1}.

E.5 Event displays of $t\bar{t}H(\gamma\gamma)$ candidates

Figures E.7-E.10 present visualizations of data events selected into the $t\bar{t}H(\gamma\gamma)$ categories with highest signal-to-background ratio.

In these images, jets reconstructed using the anti-k_t algorithm with $R = 0.4$ are represented by cones, with blue cones indicating that the jet is b-tagged at the 77% working point. Muon candidates are represented as a red track passing through the detector, and electron candidates as a green line leading to green deposits in the electromagnetic calorimeter. Photon candidates correspond to green towers in the electromagnetic calorimeter with no associated track.

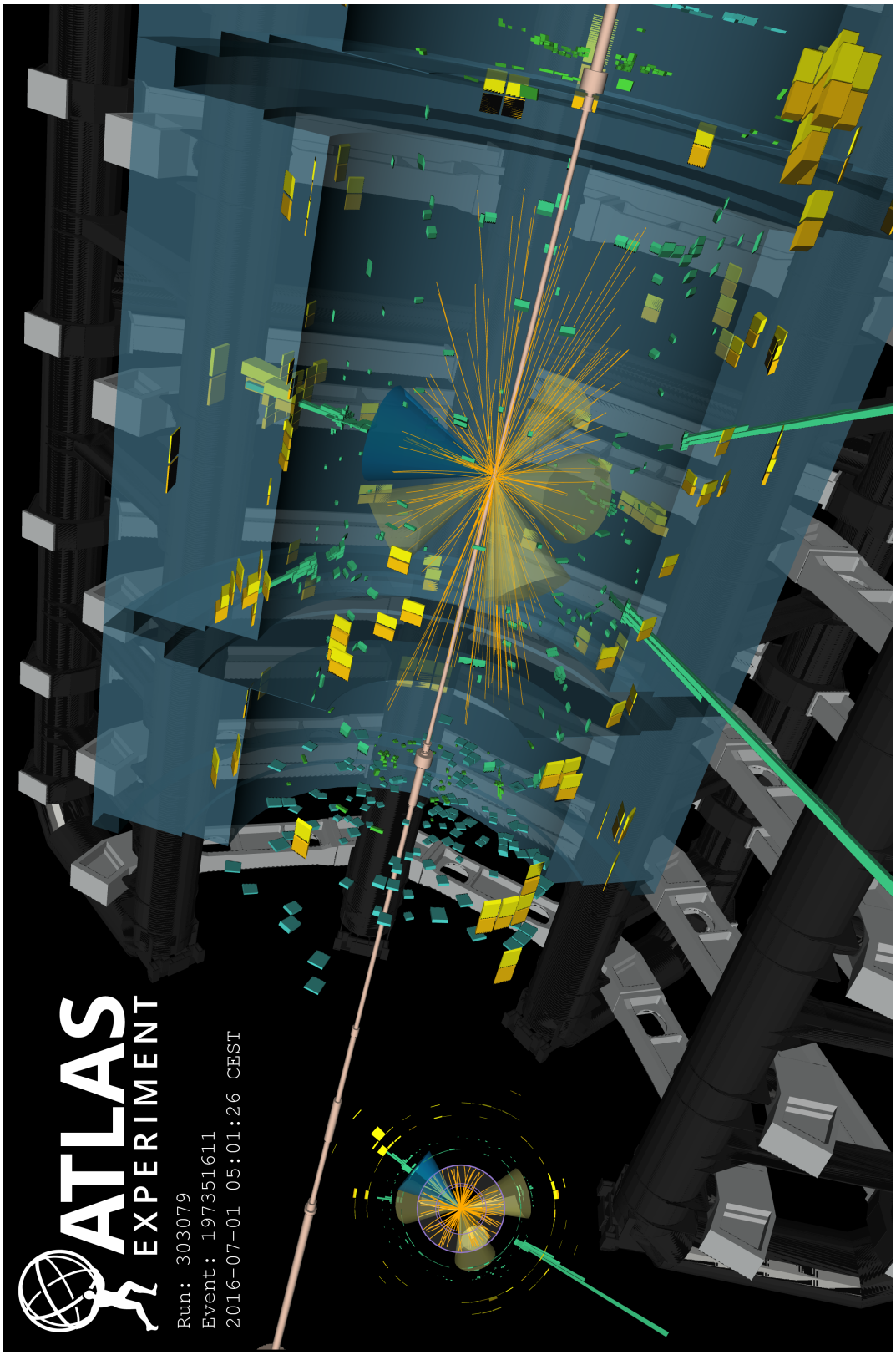


Figure E.7: Visualization of a data event selected in category $t\bar{t}H(\gamma\gamma)$ had 1. This event contains six jets (one of which is b-tagged), and two photon candidates with a diphoton mass of 125.4 GeV.

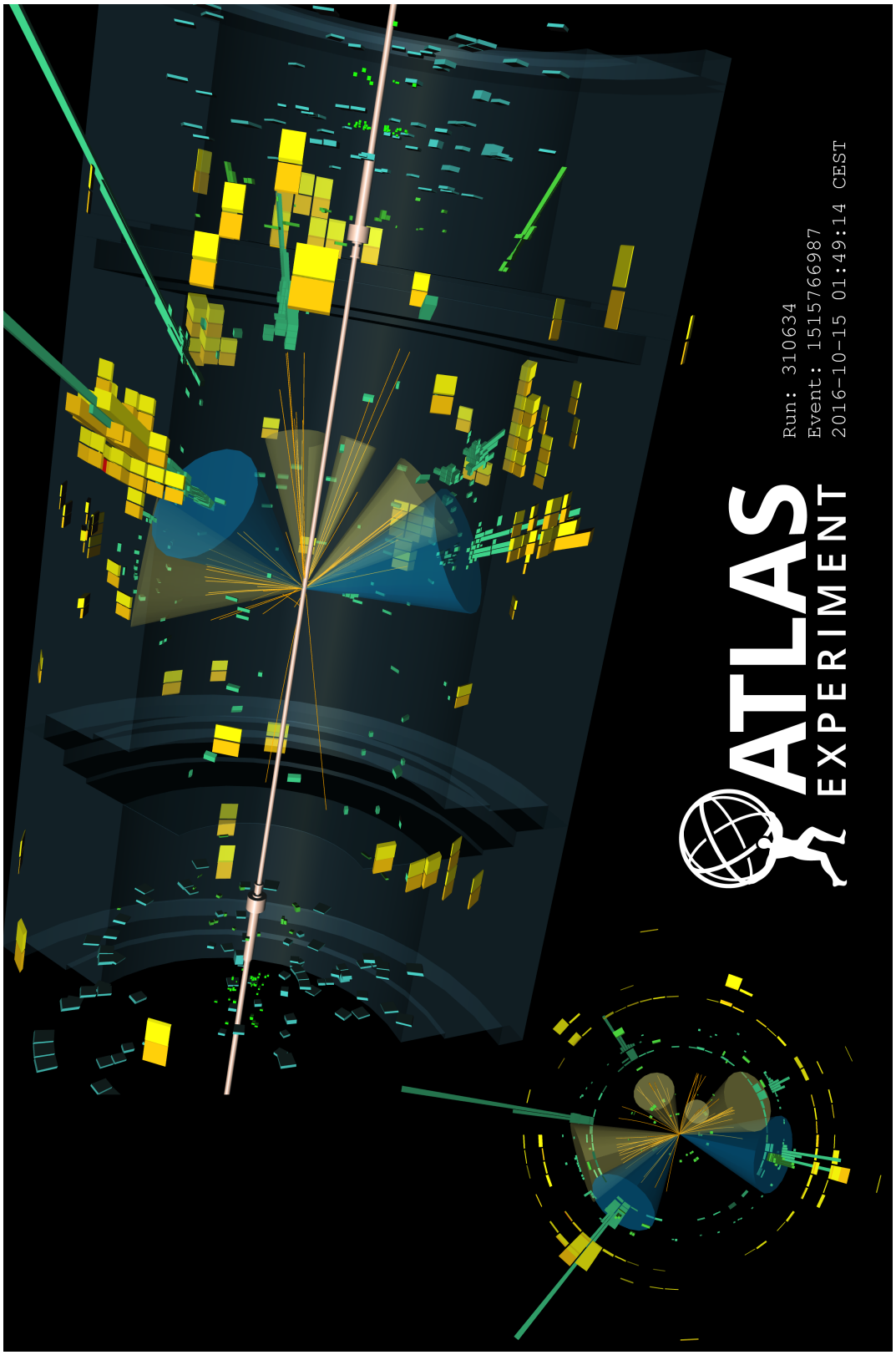


Figure E.8: Visualization of a data event selected in category $t\bar{t}H(\gamma\gamma)$ had 1. This event contains six jets (two of which are b-tagged), and two photon candidates with a diphoton mass of 124.3 GeV.

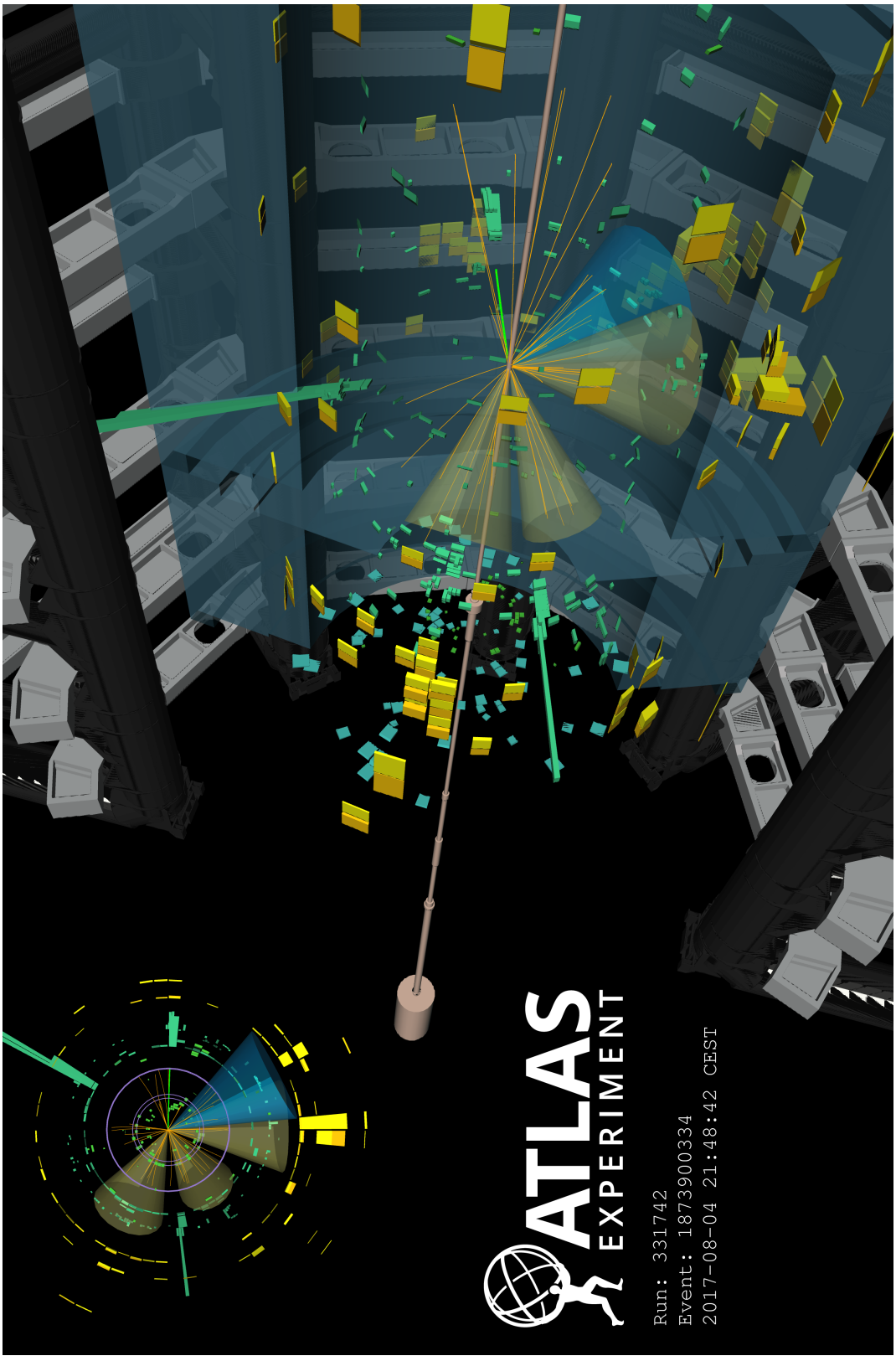


Figure E.9: Visualization of a data event selected in category $t\bar{t}H(\gamma\gamma)$ lep 1. This event contains four jets (one of which is b-tagged), one electron candidate with $p_T = 33.9$ GeV, and two photon candidates with a diphoton mass of 125.2 GeV.

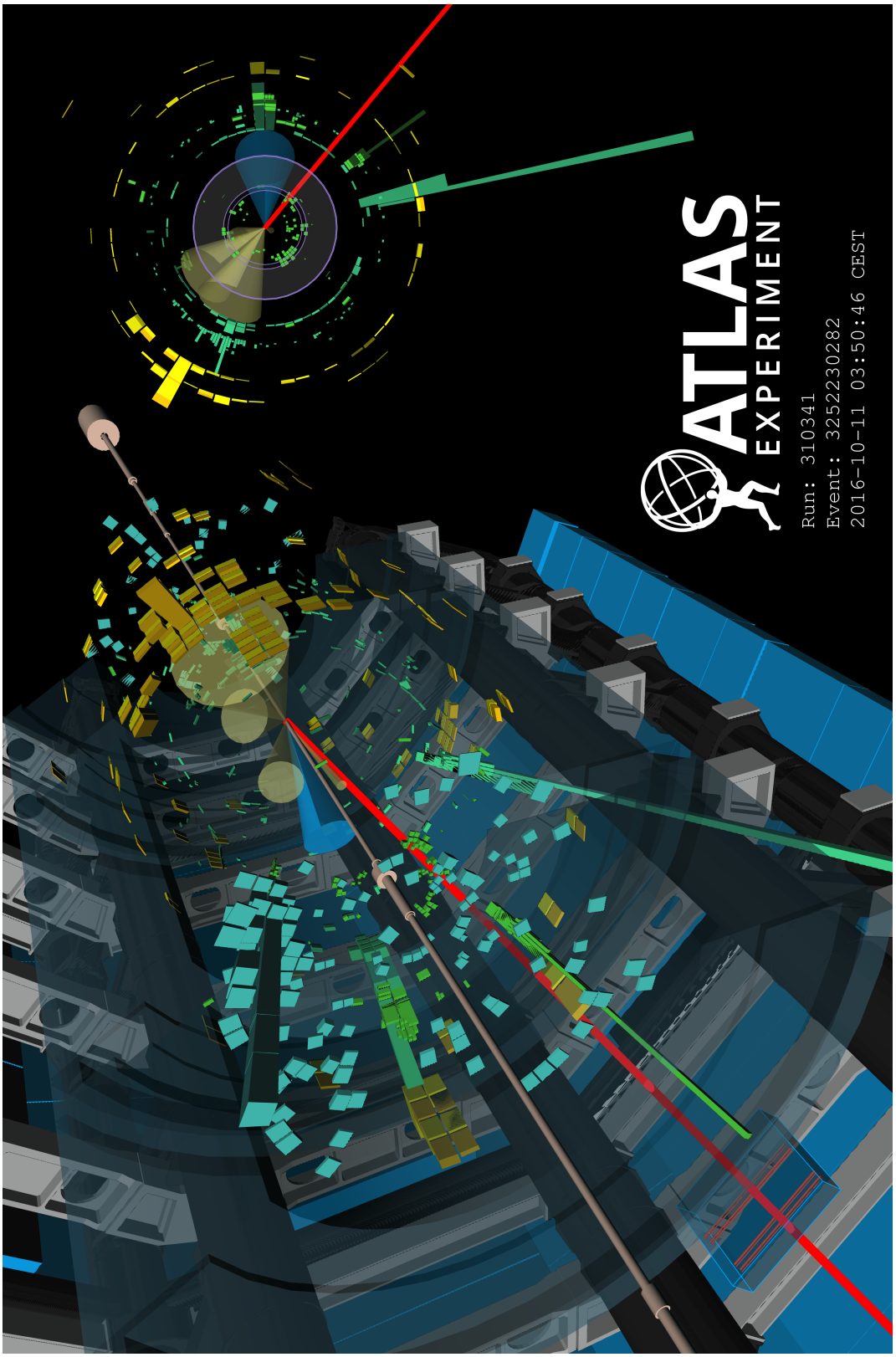


Figure E.10: Visualization of a data event selected in category $t\bar{t}H(\gamma\gamma)$ lep 1. This event contains five jets (one of which is b-tagged), one muon candidate with $p_T = 65.0$ GeV, and two photon candidates with a diphoton mass of 126.3 GeV.

Appendix F

Higgs couplings in the diphoton decay channel

This Appendix describes the full measurement of Higgs couplings in the $H \to \gamma\gamma$ decay channel with 79.8 fb^{-1} of 13 TeV ATLAS data. Section F.1 describes the selection of diphoton events into categories targeting all Higgs boson production modes. Section F.2 details the parameterization of the signal and background $m_{\gamma\gamma}$ shapes, and Section F.3 reviews the dominant systematic uncertainties. Results are presented in Section F.4.

The inclusive Higgs boson cross section is measured, as well as the individual cross sections for four Higgs boson production modes: ggF, VBF, VH, and $t\bar{t}H + tH$. In addition, the rate of $H \to \gamma\gamma$ events is measured in the **Simplified Template Cross Section (STXS)** scheme [34] [176], which gives a coarse differential binning in kinematic variables like jet multiplicity, Higgs candidate p_T, and jet p_T. The STXS regions targeted are shown on the left-hand side of Figure F.1. The boxes denote STXS region measured, including merging of categories due to low statistics.

The measurements reported in this Appendix were supplied as input to the ATLAS Higgs combination [31], which included results from analyses targeting all Higgs boson production and decay modes.

F.1 Event Categorization

All events included in the $H \to \gamma\gamma$ categories are required to have two photons passing the identification and isolation criteria outlined in Section 4.2. The categories are designed to target different Higgs boson production modes, and the categories targeting the rarest processes are filled first. Events failing selection for one production mode are considered for the next rarest mode. In total, the $H \to \gamma\gamma$ analysis includes:

1. Seven categories targeting $t\bar{t}H$
2. Five categories targeting leptonic VH

3. Seven categories targeting VBF and hadronic VH

4. Ten categories targeting gluon fusion

The seven $t\bar{t}H$ categories are those described in Chapter 9. The remaining categories are described in the following sub-sections.

Table F.1 summarizes all 29 categories and their definitions. The STXS regions corresponding to each category are shown in Figure F.1, and the contribution from each STXS bin to the Higgs signal in each category is shown in Figure F.2. Figure F.3 shows the contribution from each production mode to the total Higgs signal in each category.

Category label	Selection
ttH lep BDT1	$N_{\text{lep}} \geq 1,\ N_{b-\text{jet}} \geq 1,\ \text{BDT}_{\text{ttHlep}} > 0.987$
ttH lep BDT2	$N_{\text{lep}} \geq 1,\ N_{b-\text{jet}} \geq 1,\ 0.942 < \text{BDT}_{\text{ttHlep}} < 0.987$
ttH lep BDT3	$N_{\text{lep}} \geq 1,\ N_{b-\text{jet}} \geq 1,\ 0.705 < \text{BDT}_{\text{ttHlep}} < 0.942$
ttH had BDT1	$N_{\text{lep}} = 0,\ N_{\text{jets}} \geq 3,\ N_{b-\text{jet}} \geq 1,\ \text{BDT}_{\text{ttHhad}} > 0.996$
ttH had BDT2	$N_{\text{lep}} = 0,\ N_{\text{jets}} \geq 3,\ N_{b-\text{jet}} \geq 1,\ 0.991 < \text{BDT}_{\text{ttHhad}} < 0.996$
ttH had BDT3	$N_{\text{lep}} = 0,\ N_{\text{jets}} \geq 3,\ N_{b-\text{jet}} \geq 1,\ 0.971 < \text{BDT}_{\text{ttHhad}} < 0.991$
ttH had BDT4	$N_{\text{lep}} = 0,\ N_{\text{jets}} \geq 3,\ N_{b-\text{jet}} \geq 1,\ 0.911 < \text{BDT}_{\text{ttHhad}} < 0.971$
VH dilep	$N_{\text{lep}} \geq 2,\ 70\,\text{GeV} \leq m_{\ell\ell} \leq 110\,\text{GeV}$
VH lep High	$N_{\text{lep}} = 1,\ \lvert m_{e\gamma} - 89\,\text{GeV}\rvert > 5\,\text{GeV},\ p_{\text{T}}^{\ell+E_{\text{T}}^{\text{miss}}} > 150\,\text{GeV}$
VH lep Low	$N_{\text{lep}} = 1,\ \lvert m_{e\gamma} - 89\,\text{GeV}\rvert > 5\,\text{GeV},\ p_{\text{T}}^{\ell+E_{\text{T}}^{\text{miss}}} < 150\,\text{GeV},\ E_{\text{T}}^{\text{miss}}\ \text{significance} > 1$
VH MET High	$150\,\text{GeV} < E_{\text{T}}^{\text{miss}} < 250\,\text{GeV},\ \ E_{\text{T}}^{\text{miss}}\ \text{significance} > 9\ \ \text{or}\ \ E_{\text{T}}^{\text{miss}} > 250\,\text{GeV}$
VH MET Low	$80\,\text{GeV} < E_{\text{T}}^{\text{miss}} < 150\,\text{GeV},\ \ E_{\text{T}}^{\text{miss}}\ \text{significance} > 8$
qqH BSM	$N_{\text{jets}} \geq 2,\ p_{\text{T,j1}} > 200\,\text{GeV}$
VH had BDT tight	$60\,\text{GeV} < m_{\text{jj}} < 120\,\text{GeV},\ \ \text{BDT}_{\text{VH}} > 0.78$
VH had BDT loose	$60\,\text{GeV} < m_{\text{jj}} < 120\,\text{GeV},\ 0.35 < \text{BDT}_{\text{VH}} < 0.78$
VBF high-p_{T}^{Hjj} BDT tight	$\lvert\Delta\eta_{jj}\rvert > 2,\ \lvert\eta_{\gamma\gamma} - 0.5(\eta_{\text{j1}} + \eta_{\text{j2}})\rvert < 5,\ p_{\text{T}}^{Hjj} > 25\,\text{GeV},\ \ \text{BDT}_{\text{VBF}}^{\text{high}} > 0.47$
VBF high-p_{T}^{Hjj} BDT loose	$\lvert\Delta\eta_{jj}\rvert > 2,\ \lvert\eta_{\gamma\gamma} - 0.5(\eta_{\text{j1}} + \eta_{\text{j2}})\rvert < 5,\ p_{\text{T}}^{Hjj} > 25\,\text{GeV},\ -0.32 < \text{BDT}_{\text{VBF}}^{\text{high}} < 0.47$
VBF low-p_{T}^{Hjj} BDT tight	$\lvert\Delta\eta_{jj}\rvert > 2,\ \lvert\eta_{\gamma\gamma} - 0.5(\eta_{\text{j1}} + \eta_{\text{j2}})\rvert < 5,\ p_{\text{T}}^{Hjj} < 25\,\text{GeV},\ \ \text{BDT}_{\text{VBF}}^{\text{low}} > 0.87$
VBF low-p_{T}^{Hjj} BDT loose	$\lvert\Delta\eta_{jj}\rvert > 2,\ \lvert\eta_{\gamma\gamma} - 0.5(\eta_{\text{j1}} + \eta_{\text{j2}})\rvert < 5,\ p_{\text{T}}^{Hjj} < 25\,\text{GeV},\ 0.26 < \text{BDT}_{\text{VBF}}^{\text{low}} < 0.87$
ggF 2J BSM	$N_{\text{jets}} \geq 2,\ p_{\text{T}}^{\gamma\gamma} \geq 200\,\text{GeV}$
ggF 2J High	$N_{\text{jets}} \geq 2,\ p_{\text{T}}^{\gamma\gamma} \in [120, 200]\ \text{GeV}$
ggF 2J Med	$N_{\text{jets}} \geq 2,\ p_{\text{T}}^{\gamma\gamma} \in [60, 120]\ \text{GeV}$
ggF 2J Low	$N_{\text{jets}} \geq 2,\ p_{\text{T}}^{\gamma\gamma} \in [0, 60]\ \text{GeV}$
ggF 1J BSM	$N_{\text{jets}} = 1,\ p_{\text{T}}^{\gamma\gamma} \geq 200\,\text{GeV}$
ggF 1J High	$N_{\text{jets}} = 1,\ p_{\text{T}}^{\gamma\gamma} \in [120, 200]\ \text{GeV}$
ggF 1J Med	$N_{\text{jets}} = 1,\ p_{\text{T}}^{\gamma\gamma} \in [60, 120]\ \text{GeV}$
ggF 1J Low	$N_{\text{jets}} = 1,\ p_{\text{T}}^{\gamma\gamma} \in [0, 60]\ \text{GeV}$
ggF 0J Fwd	$N_{\text{jets}} = 0,$ one photon with $\lvert\eta\rvert > 0.95$
ggF 0J Cen	$N_{\text{jets}} = 0,$ two photons with $\lvert\eta\rvert \leq 0.95$

Table F.1: Summary of the 29 $H \to \gamma\gamma$ analysis categories.

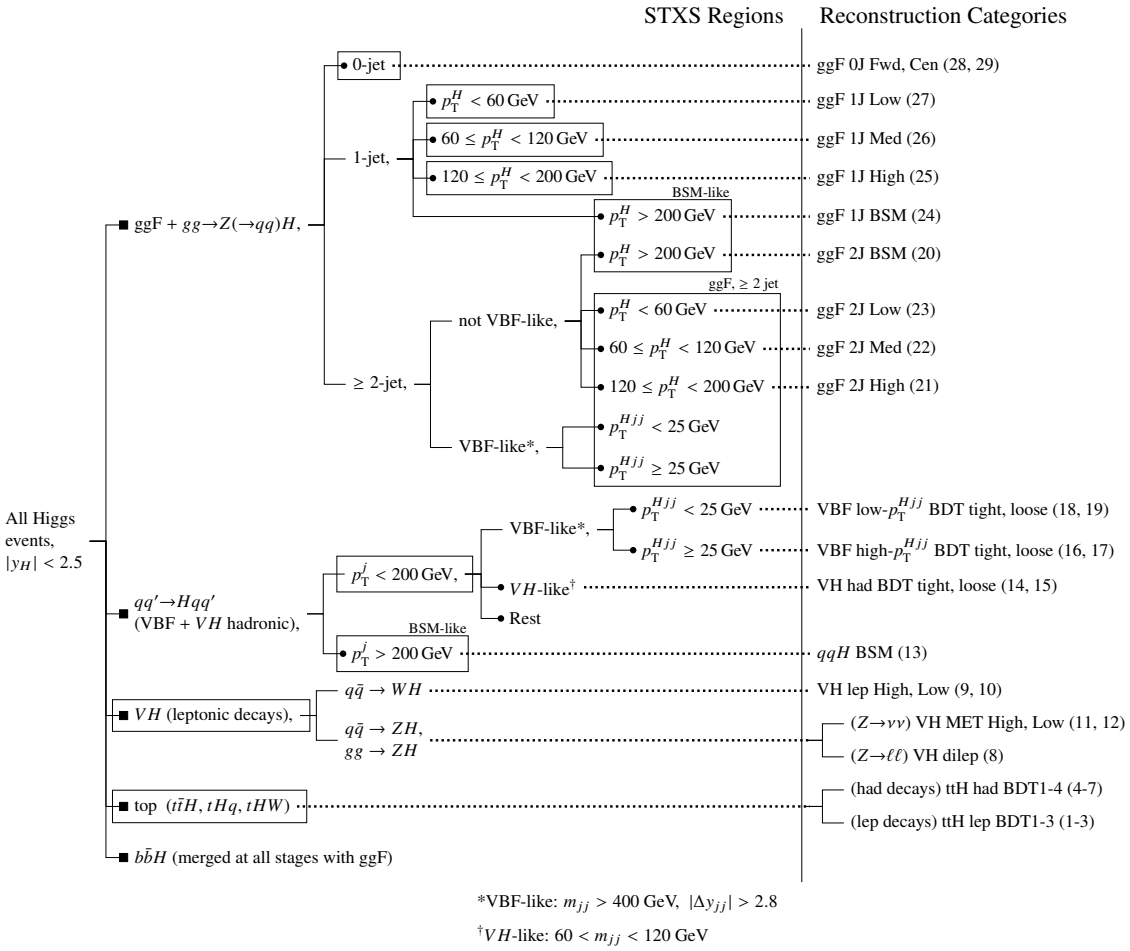


Figure F.1: Correspondence between Simplified Template Cross Section regions (left) and $H \to \gamma\gamma$ analysis categories (right).

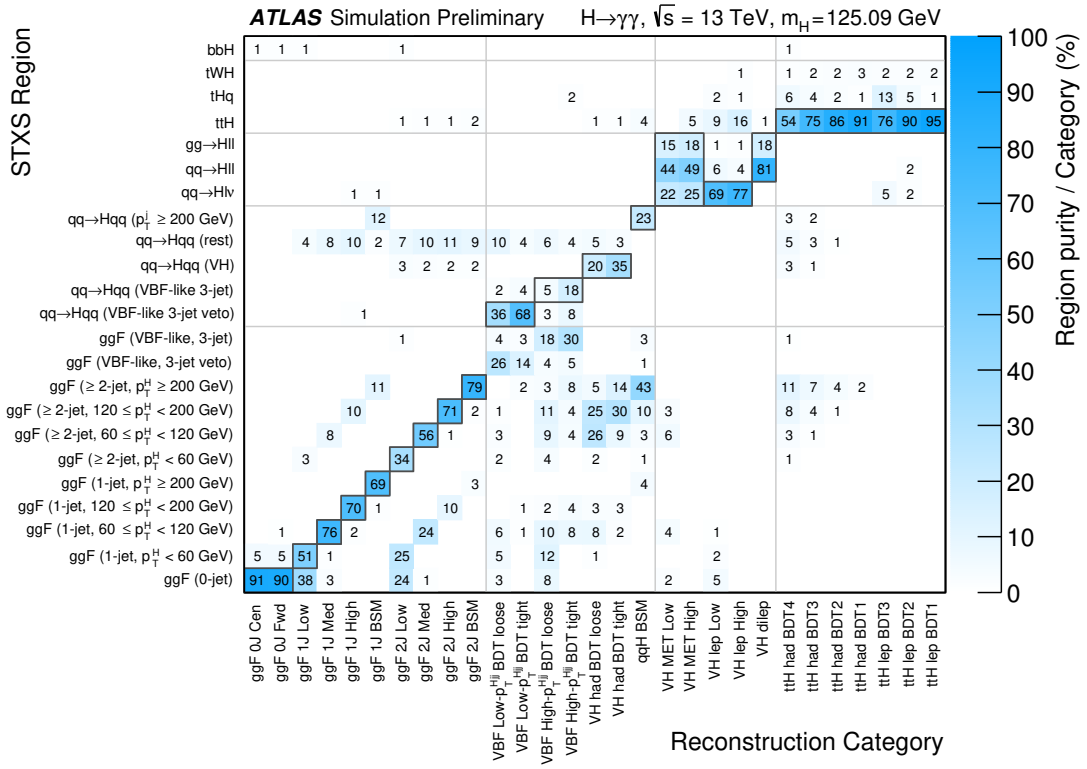


Figure F.2: Breakdown of expected Higgs signal by STXS process (y-axis) in each analysis category (x-axis). The regions bounded in light gray contain categories corresponding to the same Higgs production mode. The entries outlined in dark gray indicate the target process(es) corresponding to each category. The categories are generally very pure in the target process, though there is some contamination (e.g. between ggF and VBF categories).

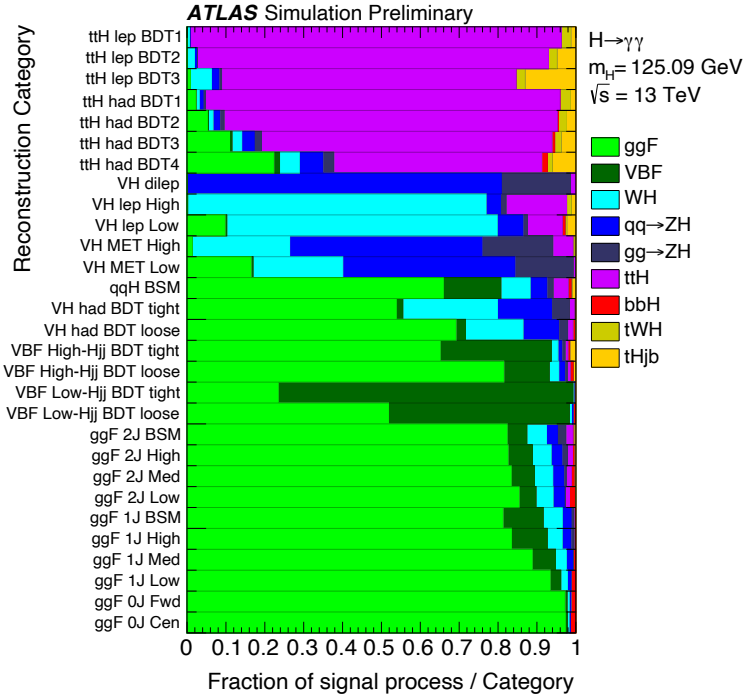


Figure F.3: Breakdown of expected Higgs signal by production mode in each analysis category. The categories are generally very pure in the target process, though there is some contamination (e.g. between ggF and VBF categories).

F.1.1 Leptonic VH

The VH dilepton category targets $qq/qg \to ZH$ production where the Z boson decays to charged leptons: events in this category must contain two same-flavor opposite-sign leptons with dilepton invariant mass $m_{\ell\ell}$ between 70 GeV and 110 GeV.

The VH single lepton categories target $qq/qg \to WH$ production where the W boson decays to a charged lepton (e or μ) and a neutrino. Events in these categories must contain exactly one electron or muon. Events where the electron and either of the two photons have $m_{e\gamma} \in [84, 94]$ GeV are vetoed. Two categories are defined by dividing events by the missing E_T significance and the transverse momentum of the lepton plus missing E_T system: low ($p_T^\ell + E_T^{\text{miss}} < 150$ GeV and E_T^{miss} significance > 1.0) and high ($p_T^\ell + E_T^{\text{miss}} > 150$ GeV).

Events with zero leptons (e or μ) fall into the VH MET categories if they contain high missing E_T. These categories target $qq/qg \to ZH$ production where the Z boson decays to neutrinos. Two such categories are defined: low MET (80 GeV $< E_T^{\text{miss}} < 150$ GeV and E_T^{miss} significance > 8.0), or high MET ($E_T^{\text{miss}} > 150$ GeV and E_T^{miss} significance > 9.0, or $E_T^{\text{miss}} > 250$ GeV).

F.1.2 Vector boson fusion and hadronic VH

The qqH BSM category contains events with at least two jets and very high leading jet p_T (> 200 GeV).

Hadronic VH events target WH and ZH production where the vector boson decays hadronically. Events in these categories must contain two jets with dijet invariant mass 60 GeV $< m_{jj} <$ 120 GeV. A BDT is trained to separate simulated VH events from a background sample constructed from simulated diphoton background and data-driven estimates of photon+jet and dijet backgrounds from NTI data. The BDT uses training variables $\Delta\phi$ and Δy between the diphoton and dijet systems, $\cos\theta^*_{\gamma\gamma,jj}$, dijet invariant mass, diphoton p_T orthogonal to the diphoton thrust axis in the transverse plane, p_T of the diphoton system, $H_T + E_T^{\gamma 1} + E_T^{\gamma 2}$, minimum ΔR between a photon and a jet, the dijet transverse momentum, and Zeppenfeld η ($|\eta_{\gamma\gamma} - \frac{1}{2}(\eta_{j1} + \eta_{j2})|$). Using the BDT output as the discriminating variable, events are classified into two categories: tight (BDT score > 0.78) and loose (BDT score < 0.78).

In the four VBF categories, events are required to contain at least two jets. The two jets with highest p_T must be separated by a pseudorapidity difference $|\Delta\eta(j1, j2)| > 2.0$. In addition, events in the VBF categories must satisfy $|\eta_H - 0.5(\eta_{j1} + \eta_{j2})| < 5.0$, where η_H is the pseudorapidity of the Higgs candidate (diphoton system).

Another BDT is trained to separate simulated VBF events from a background sample constructed from simulated diphoton background and data-driven estimates of photon+jet and dijet backgrounds from NTI data. The BDT uses six training variables: dijet invariant mass, $\Delta\eta$ between the leading jets, $\Delta\phi$ between the diphoton and dijet systems, diphoton p_T orthogonal to the diphoton thrust axis in the transverse plane, minimum ΔR between a photon and a jet, and Zeppenfeld η. Four exclusive regions are defined based on the BDT

discriminant and the transverse momentum of system composed of the Higgs candidate and two leading jets (p_T^{Hj1j2}). The cut values are listed in Table F.1.

F.1.3 Gluon fusion

Events failing the $t\bar{t}H$, VH, and VBF selection are divided among the ten ggF categories. The events are divided into bins of jet multiplicity (0,1,$\geq$ 2). Events with 1 or $\geq$ 2 jets are further divided into four bins in Higgs candidate p_T: low (0-60 GeV), med (60-120 GeV), high (120-200 GeV), and BSM ($>$ 200 GeV). The ggF events with 0 jets are split into two bins in based on the pseudorapidity of the photons: forward events (fwd) contain at least one photon with $|\eta| > 0.95$, and central events (cen) do not.

F.2 Signal and Background

In each analysis category, the signal shape is determined by fitting a Double-Sided Crystal Ball (DSCB) function to a sample of inclusive Higgs boson Monte Carlo (see Section 6.2.1). The best fit DSCB parameters in each category are summarized in Table F.2. The $t\bar{t}H$ categories are included in Table E.1.

The shape of the continuum background in the ggF and VBF categories is derived from templates constructed from diphoton Monte Carlo (see Section 5.2.3). The templates are constructed by adding together $m_{\gamma\gamma}$ distributions representing true diphoton ($\gamma\gamma$) events, single photon+jet (γj), and dijet (jj) events. The shape of the γj distribution is modeled using the data sidebands where exactly one photon fails the identification and/or isolation criteria. The shape of the jj distribution is modeled using the data sidebands where both photons fail the identification and/or isolation criteria. The relative fraction of $\gamma\gamma$, γj, and jj is calculated by a template fit of these three distributions to the TI data sidebands.

In the VH categories, background templates are constructed from $V + \gamma\gamma$ Monte Carlo generated with Sherpa following the setup used for the diphoton background sample described in Section 5.2.3. The Monte Carlo templates are normalized so that the number of sideband events is equal to that in the tight/isolated data. Statistics in the Monte Carlo sample are low in some VH categories, resulting in large spurious signal.

The functional form of the continuum background and the spurious signal are listed in Table F.3 for the ggF, VBF, and VH categories. The background functional form and spurious signal for the $t\bar{t}H$ categories at 79.8 fb^{-1} are included in Appendix E. The χ^2/ndof indicates the performance of a background only fit to the template in the $m_{\gamma\gamma}$ range 105-160 GeV, but is in general not a part of the spurious signal criteria.

Category	μ_{CB} [GeV]	σ_{CB} [GeV]	α_{low}	α_{high}	n_{low}	n_{high}
ggH 0J cen	125.22	1.60	1.6	1.9	808.4	4.1
ggH 0J fwd	125.17	2.00	1.5	1.7	865.2	5.9
ggH 1J low	125.22	1.81	1.5	1.7	687.7	5.5
ggH 1J med	125.19	1.72	1.4	1.7	371.7	5.1
ggH 1J high	215.19	1.54	1.4	1.7	45.0	5.5
ggH 1J BSM	125.15	1.30	1.5	1.6	17.5	5.5
ggH 2J low	125.28	1.84	1.4	1.7	426.4	5.0
ggH 2J med	125.23	1.73	1.5	1.7	34.8	5.0
ggH 2J high	125.18	1.55	1.5	1.8	30.7	4.9
ggH 2J BSM	125.17	1.37	1.5	1.7	20.7	5.8
VBF low p_T^{Hjj}, loose	125.19	1.74	1.5	1.7	35.2	4.8
VBF low p_T^{Hjj}, tight	125.17	1.54	1.5	1.6	33.7	6.4
VBF high p_T^{Hjj}, loose	125.26	1.72	1.4	1.7	26.3	4.7
VBF high p_T^{Hjj}, tight	125.24	1.57	1.5	1.7	26.2	5.4
VH had loose	125.20	1.64	1.5	1.7	25.1	4.8
VH had tight	125.19	1.46	1.5	1.7	22.3	5.2
qqH BSM	125.15	1.33	1.5	1.6	17.6	5.9
VH MET low	125.19	1.70	1.5	1.5	6.1	5.4
VH MET high	125.19	1.48	1.6	1.7	11.0	6.1
VH lep low	125.19	1.75	1.5	1.7	23.1	4.7
VH lep high	125.14	1.38	1.4	1.6	13.8	5.0
VH dilep	125.20	1.59	1.8	1.7	6.0	4.2

Table F.2: Best fit parameter values for the signal $m_{\gamma\gamma}$ shape in each $H \to \gamma\gamma$ category, excluding the $t\bar{t}H$ categories (which are shown in Table E.1.)

Category	Function	N_{sp}	χ^2/ndof	Prob(χ^2) [%]
ggH 0J cen	ExpPoly2	−48.3	1.12	25.1
ggH 0J fwd	ExpPoly2	−228	1.1	29.6
ggH 1J low	ExpPoly2	+45.9	1.18	17.5
ggH 1J med	ExpPoly2	+40.7	1.59	0.445
ggH 1J high	PowerLaw	−9.79	1.03	41.6
ggH 1J BSM	Exponential	−2.63	1.23	12.1
ggH 2J low	ExpPoly2	−38.1	0.957	56.3
ggH 2J med	ExpPoly2	+21.7	1	47.1
ggH 2J high	PowerLaw	+7.43	1.28	8.27
ggH 2J BSM	Exponential	+0.426	1.11	26.8
VBF low p_T^{Hjj}, loose	PowerLaw	+10.3	1.05	38.2
VBF low p_T^{Hjj}, tight	Exponential	+5.16	0.922	63.7
VBF high p_T^{Hjj}, loose	PowerLaw	−12.2	0.895	69
VBF high p_T^{Hjj}, tight	PowerLaw	−17.4	1.37	3.75
VH had loose	PowerLaw	+6.45	1	47.3
VH had tight	Exponential	−1.81	1.07	33.1
qqH BSM	Exponential	−11.7	1.02	43.8
VH MET low	Exponential	+2.45	2.61	0.001
VH MET high	PowerLaw	+1.94	0.989	47.9
VH lep low	Exponential	−4.67	1	46.4
VH lep high	Exponential	−0.598	1.56	3.52
VH dilep	PowerLaw	+0.608	1.51	4.53

Table F.3: Results of the spurious signal test for the $H \to \gamma\gamma$ categories at 79.8 fb^{-1}, excluding the $t\bar{t}H$ categories (which are shown in Table E.2). The χ^2/ndof indicates the performance of a background only fit to the template in the $m_{\gamma\gamma}$ range 105-160 GeV, but is in general not a part of the spurious signal criteria.

F.3 Systematic uncertainties

The experimental uncertainties included in the $H \to \gamma\gamma$ couplings measurement are those summarized in Section 6.4.2.

For each Higgs signal process, QCD scale variations and PDF uncertainties (31 eigenvector scheme) are evaluated separately in each category. The heavy flavor uncertainty in the $t\bar{t}H$ categories is retained, and UEPS uncertainties are evaluated on each Higgs production mode.

For the measurement of the ggF cross section, the UEPS uncertainty on ggF and the spurious signal in the 0-jet ggF categories dominate the systematic uncertainty, along with the photon identification efficiency. For the VBF cross section, the UEPS uncertainty on VBF and the experimental jet uncertainties are the dominant systematics. The dominant uncertainties for the measurement of the VH cross section are the spurious signal in the leptonic VH categories.

For the measurement of the $t\bar{t}H + tH$ cross section, the UEPS uncertainty on $t\bar{t}H$ and the heavy flavor uncertainties on ggF and VH are dominant. Pulls on the photon energy scale and resolution are not as strong as seen in the observed data shown in Appendix E. This is due to the constraining power of the higher statistics categories dedicated to other production modes.

F.4 Results (79.8 fb^{-1})

The inclusive cross section is extracted from the fit with a single POI μ. For the inclusive fit, the number of Higgs boson signal events can be expressed as

$$N_S^c = \mu \times \sum_p \left[N_p^{SM,c}(1 + \sigma_L \theta_L) \prod_i (1 + \sigma_{\epsilon,p}^{i,c} \theta_\epsilon^i) \prod_j (1 + \sigma_{T,p}^{j,c} \theta_T^j) \right] \tag{F.1}$$

where the systematic uncertainty magnitudes σ and nuisance parameters θ have the same definitions given in Chapter 6.

The measured inclusive signal strength is

$$\mu = 1.06\ ^{+0.13}_{-0.12} = 1.06\ \pm 0.08\ (\text{stat})\ ^{+0.10}_{-0.09}\ (\text{syst}) \tag{F.2}$$

This is in good agreement with the SM prediction of $1.00^{+0.13}_{-0.12}$. The likelihood scan for the inclusive cross section is shown in Figure F.4.

Figure F.5 shows the weighted diphoton invariant mass spectrum summed over all 29 $H \to \gamma\gamma$ analysis categories. In order to visually enhance the contribution of the categories targeting each production mode, the data and fit in each category are weighted by $\ln(1+S/B)$ where S (B) is the expected signal (background) in the smallest $m_{\gamma\gamma}$ window containing 90% of the Higgs signal. The signal here is taken to be all Higgs boson events, and the calculation of error bars in Figure F.5 follows the prescription outlined in Section 6.5.

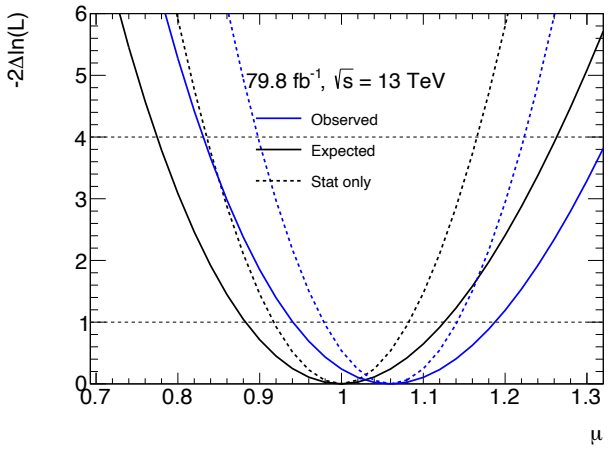


Figure F.4: Likelihood scan of the inclusive Higgs signal strength measured in $H \to \gamma\gamma$ with 79.8 fb^{-1}.

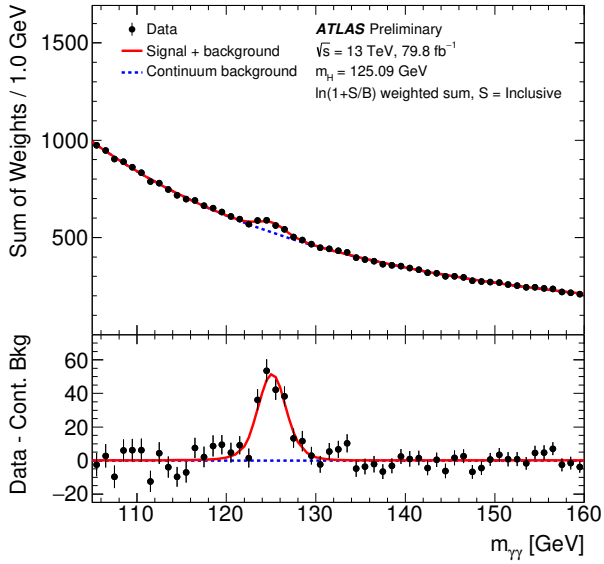


Figure F.5: Weighted diphoton invariant mass spectrum of data from all 29 $H \to \gamma\gamma$ analysis categories with 79.8 fb^{-1}. Each data event is weighted by $\ln(1 + S_c/B_c)$, where c is the category containing the event, and S_c and B_c are the fitted Higgs signal and continuum background in category c, respectively.

F.4.1 Higgs cross section by production mode

The cross sections of individual Higgs boson production modes (p = ggF+$b\bar{b}H$, VBF, VH, $t\bar{t}H + tH$) are fitted simultaneously with POIs μ_p corresponding to each process:

$$N_S^c = \sum_p \left[\mu_p \times N_p^{SM,c}(1 + \sigma_L \theta_L) \prod_i (1 + \sigma_{\epsilon,p}^{i,c} \theta_\epsilon^i) \prod_j (1 + \sigma_{T,p}^{j,c} \theta_T^j) \right] \tag{F.3}$$

where the systematic uncertainty magnitudes σ and nuisance parameters θ have the same definitions given in Chapter 6.

Figure F.6 shows the weighted diphoton invariant mass spectrum summed over all 29 $H \to \gamma\gamma$ analysis categories. In order to visually enhance the contribution of the categories targeting each production mode, the data and fit in each category are weighted by $\ln(1+S/B)$ where S (B) is the expected signal (background) in the smallest $m_{\gamma\gamma}$ window containing 90% of the Higgs signal. In each panel of this figure, a different production mode is treated as signal in the calculation of the category weight. The calculation of error bars in Figure F.6 follows the prescription outlined in Section 6.5.

The measured cross section time $H \to \gamma\gamma$ branching fraction for each production mode is:

$$\sigma_{ggF} \times B_{\gamma\gamma} = 98.28\ ^{+14.71}_{-14.08}\ \text{fb} = 98.28\ \pm 11.18\ (\text{stat})\ ^{+9.15}_{-8.13}\ (\text{syst})\ \text{fb} \tag{F.4}$$

$$\sigma_{VBF} \times B_{\gamma\gamma} = 11.15\ ^{+3.38}_{-2.92}\ \text{fb} = 11.15\ ^{+2.56}_{-2.40}\ (\text{stat})\ ^{+2.24}_{-1.60}\ (\text{syst})\ \text{fb} \tag{F.5}$$

$$\sigma_{VH} \times B_{\gamma\gamma} = 4.84\ ^{+2.66}_{-2.47}\ \text{fb} = 4.84\ ^{+2.40}_{-2.26}\ (\text{stat})\ ^{+1.13}_{-1.00}\ (\text{syst})\ \text{fb} \tag{F.6}$$

$$\sigma_{t\bar{t}H+tH} \times B_{\gamma\gamma} = 1.50\ ^{+0.56}_{-0.48}\ \text{fb} = 1.50\ ^{+0.49}_{-0.44}\ (\text{stat})\ ^{+0.27}_{-0.19}\ (\text{syst})\ \text{fb} \tag{F.7}$$

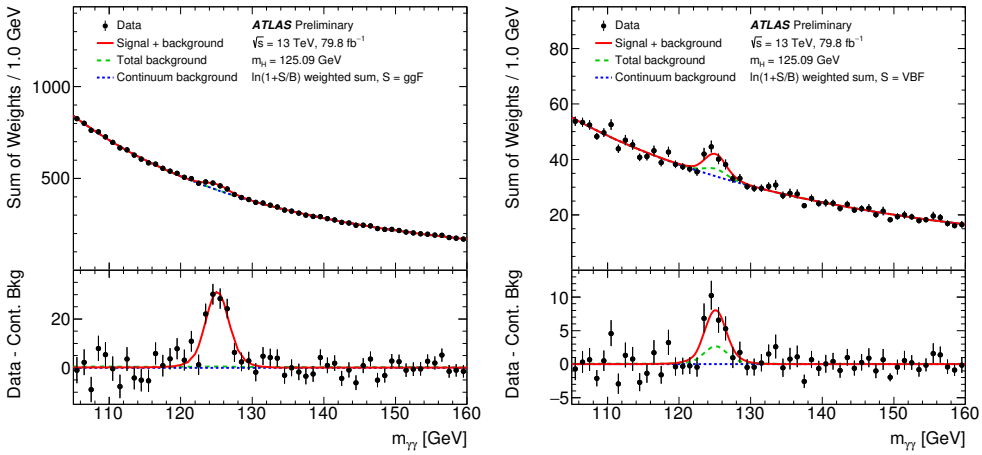
The uncertainty on these measurements is dominated by statistical error at 79.8 fb^{-1}, though the statistical and systematic errors are similar in magnitude for the gluon fusion cross section.

The measured production mode cross sections are shown in Figure F.7, normalized to the SM prediction. The red line represents the Standard Model prediction, with theory error included in the gray uncertainty band. Measurements are shown as black points, with the inclusive uncertainty shown as black error bars. The uncertainty is also shown in separate statistical (yellow) and systematic (blue) components. The likelihood scan for each cross section measurement is shown in Figure F.8. No significant deviation from the Standard Model is observed.

The measured correlations between the production mode cross sections are displayed in Figure F.9. The correlations are overall small, with the correlation between the ggF and VBF modes being the largest.

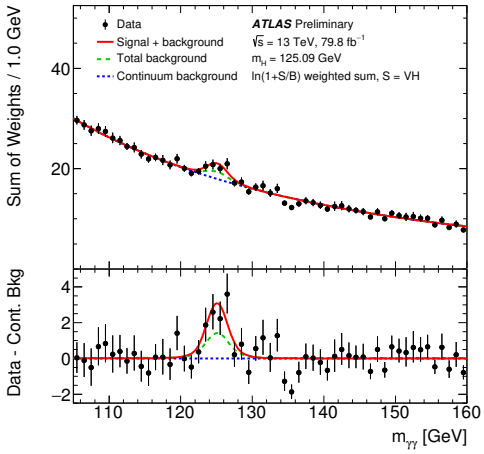
The expected and observed significance of each production mode is reported in Table F.4.

In addition to direct measurements of the production mode cross sections, ratios of the production mode cross sections to the ggF cross section are calculated. This allows for

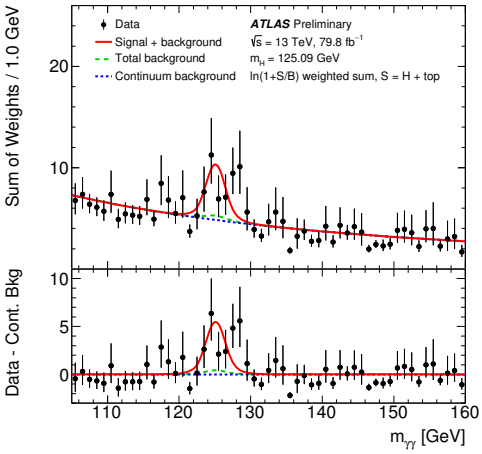


(a) $S = \text{ggF}$

(b) $S = \text{VBF}$



(c) $S = VH$



(d) $S = t\bar{t}H + tH$

Figure F.6: Weighted diphoton invariant mass spectrum of data from all 29 $H \to \gamma\gamma$ analysis categories with 79.8 fb^{-1}, with different Higgs boson processes treated as signal. Each data event is weighted by $\ln(1 + S_c/B_c)$, where c is the category containing the event. In subfigures (a), (b), (c), and (d), S_c corresponds to the fitted ggF, VBF, VH, and $t\bar{t}H + tH$ signal, respectively. The background B_c in each case is corresponds to the fitted continuum background plus the fitted Higgs background from non-signal processes.

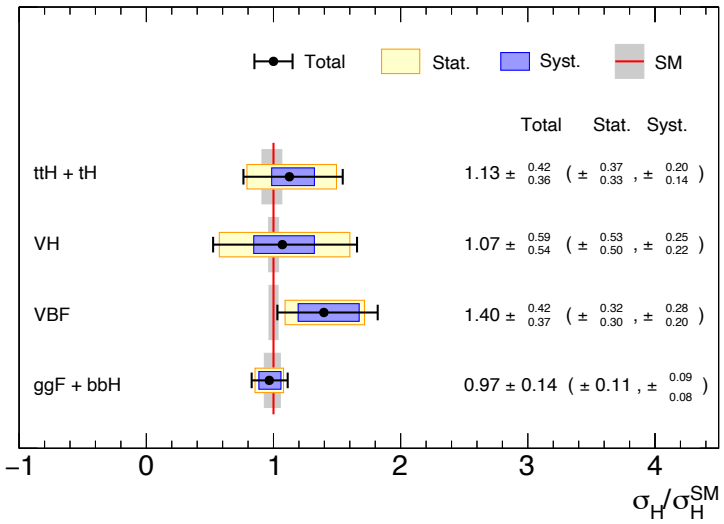


Figure F.7: Measured cross section by production mode in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}, normalized to the Standard Model prediction. The red line represents the Standard Model prediction, with theory error included in the gray uncertainty band. Measurements are shown as black points, with the inclusive uncertainty shown as black error bars. The uncertainty is also shown in separate statistical (yellow) and systematic (blue) components. No significant deviation from the Standard Model prediction is observed.

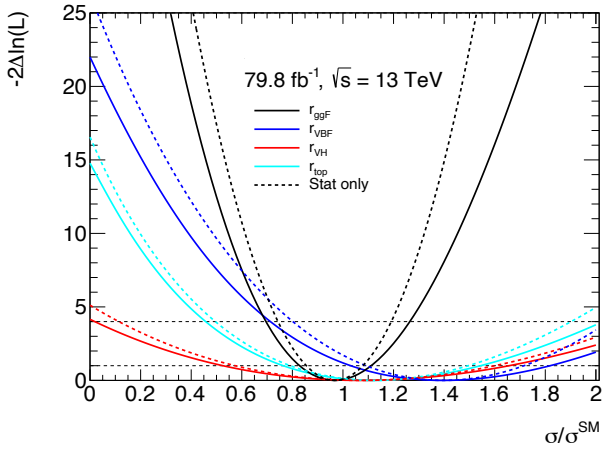


Figure F.8: Likelihood scan of Higgs production cross sections in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}. The ggF, VBF, VH, and $t\bar{t}H + tH$ production modes are shown in black, blue, cyan, and red respectively. The statistical only errors are indicated by dashed lines. No significant deviation from the Standard Model is observed.

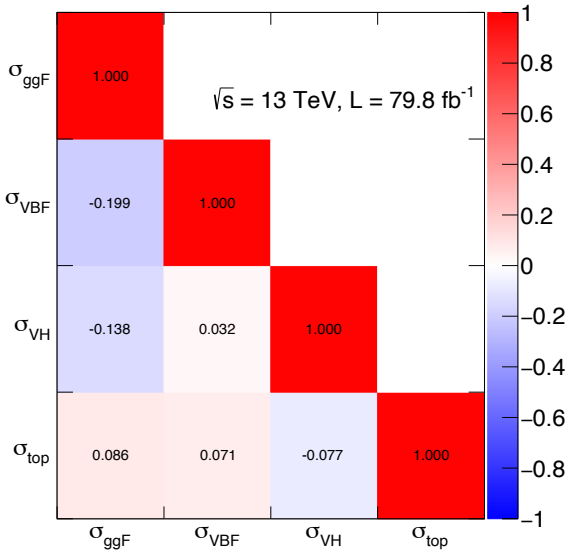


Figure F.9: Measured correlations between production mode cross sections in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}.

Production mode	Expected significance	Observed significance
ggF	7.9σ	7.6σ
VBF	3.6σ	4.7σ
VH	1.9σ	2.0σ
$t\bar{t}H + tH$	3.6σ	3.8σ

Table F.4: Expected and observed significance of each Higgs production mode in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}.

the cancellation of some common systematic uncertainties. The measured cross section ratios are shown in Figure F.10, normalized to the SM expectation. The red line represents the Standard Model prediction, with theory error included in the gray uncertainty band. Measurements are shown as black points, with the inclusive uncertainty shown as black error bars. The uncertainty is also shown in separate statistical (yellow) and systematic (blue) components. The likelihood scan for each cross section ratio is shown in Figure F.11. No significant deviation from the Standard Model is observed.

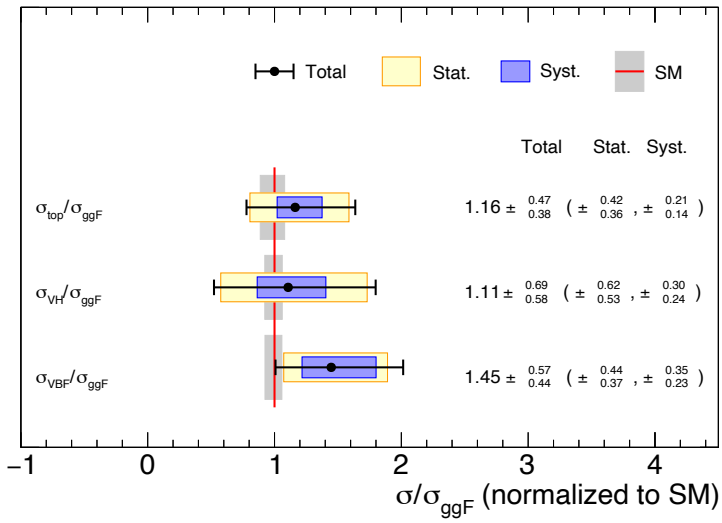


Figure F.10: Measured ratio of production cross section to ggF in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}. The ratio is normalized to the Standard Model prediction. The red line represents the Standard Model prediction, with theory error included in the gray uncertainty band. Measurements are shown as black points, with the inclusive uncertainty shown as black error bars. The uncertainty is also shown in separate statistical (yellow) and systematic (blue) components. No significant deviation from the Standard Model is observed.

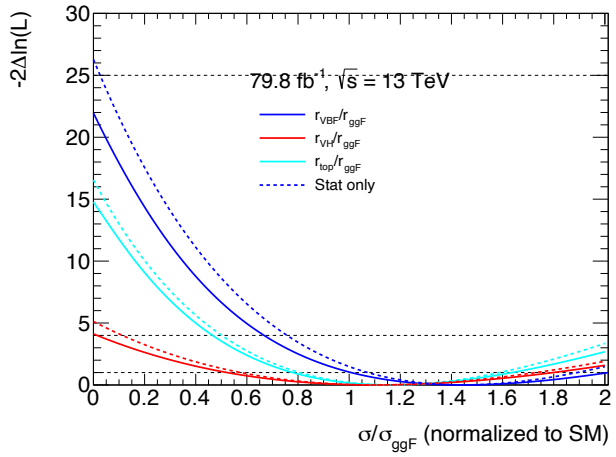


Figure F.11: Likelihood scan of Higgs production cross section ratio to ggF in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}. The ratios of VBF, VH, and $t\bar{t}H + tH$ to ggF are shown in blue, cyan, and red respectively. The statistical only errors are indicated by dashed lines. No significant deviation from the Standard Model is observed.

F.4.2 Measurement of STXS

For the measurement of STXS rates, a POI μ_b is associated with each STXS bin b. The exception is the ggF and VBF BSM bins, where the sum is fitted with POI μ_{BSM} and the difference is profiled. All categories are fit simultaneously. The measured STXS rates are shown in Figure F.12, normalized to the SM expectation. The red line represents the Standard Model prediction, with theory error included in the gray uncertainty band. Measurements are shown as black points, with the inclusive uncertainty shown as black error bars. The uncertainty is also shown in separate statistical (yellow) and systematic (blue) components. No significant deviation from the Standard Model is observed.

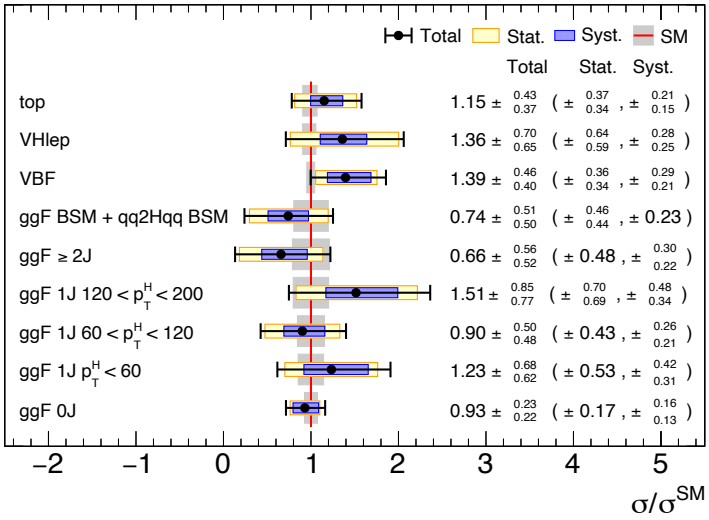


Figure F.12: Measured Simplified Template Cross Sections in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}, normalized to the Standard Model prediction. The red line represents the Standard Model prediction, with theory error included in the gray uncertainty band. Measurements are shown as black points, with the inclusive uncertainty shown as black error bars. The uncertainty is also shown in separate statistical (yellow) and systematic (blue) components. No significant deviation from the Standard Model prediction is observed.

The measured correlations between the STXS measurements are displayed in Figure F.9. The correlations are overall small, with the correlation between ggF bins of different jet multiplicity being the largest.

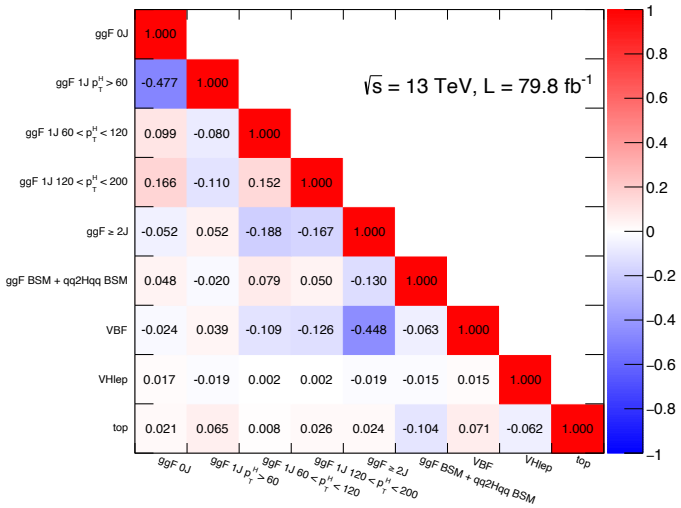


Figure F.13: Measured correlations between STXS bins in the $H \to \gamma\gamma$ channel with 79.8 fb^{-1}.

Appendix G

Multiclassification of $t\bar{t}H$, tH, and background

Higgs boson production in association with a single top quark (tH) is predicted by the Standard Model, but with much lower cross section than $t\bar{t}H$ due to destructive interference between Higgs-top and Higgs-W diagrams. Many models of BSM physics, including those discussed in Section 2.4.1, predict a large enhancement of tH cross sections: any observation of this process during LHC Run 2 would be a clear sign of new physics. However, the analysis of $H \to \gamma\gamma$ couplings described in Appendix F includes no categories designed specifically to target tH production.

This Appendix presents a study of multiclass BDT performance for separation of $t\bar{t}H$, tH, and continuum background. This can be viewed an extension of the binary $t\bar{t}H(\gamma\gamma)$ strategy presented in Chapter 9, with the addition of a third class dedicated to tH. The goal of this multiclass approach is to enhance sensitivity to tH without sacrificing $t\bar{t}H$ significance.

G.1 Multiclass BDT training

Separate multiclass BDTs are trained in the hadronic and leptonic channels using XGBoost [140]. Each multiclass BDT is trained to separate the following three samples:

- Class 0: not tight/isolated (NTI) data, the control sample representing the continuum background
- Class 1: MG5_aMC@NLO+Pythia8 $tHjb(\gamma\gamma)$ Monte Carlo
- Class 2: Powheg+Pythia8 $t\bar{t}H(\gamma\gamma)$ Monte Carlo

Only 50% of each sample is used for training, with 25% of each sample reserved for validation (including hyper-parameter optimization) and 25% reserved for testing.

No tWH Monte Carlo included in the training. The separation between $t\bar{t}H$ and tH is found to be worse when the tH training sample includes $tHjb + tWH$ than when it includes

$tHjb$ alone. This is due to the low cross section of tWH (even compared to $tHjb$) and the kinematic similarities between tWH and $t\bar{t}H$.

The same set of training variables are used for the hadronic and leptonic multiclass BDTs. These variables are:

- p_T and η of the diphoton system (Higgs candidate)
- p_T, η, ϕ (with respect to the Higgs candidate), and Top Reco BDT score of the reconstructed top candidate and the second top candidate (see Chapter 8). In the case where no second top is reconstructed (either partially or fully), a missing value is passed to XGBoost for Top 2 Reco BDT score.
- Angles $\Delta\eta$ and $\Delta\phi$ between the top candidates. In the case where no second top is reconstructed (either partially or fully), a missing value is passed to XGBoost.
- Two-object invariant masses m_{t_1H}, and $m_{t_1t_2}$. In the case where no second top is reconstructed (either partially or fully), a missing value is passed to XGBoost.
- $H_T = \sum_{\text{jet } j} p_T^j$
- Jet multiplicity and b-jet multiplicity (77% working point)
- Missing E_T significance (Equation 4.7)

Additional training variables were tested, but found to yield similar or worse performance. The alternative variable sets tested were:

- The above variables plus the momenta of up to 4 (2) jets
- Instead of the above variables, use only the momenta of photons, 0 (up to 2) leptons, and up to 6 (4) jets in the hadronic (leptonic) channel

Multiple possibilities were investigated for the relative normalization of the training samples. Normalizing each sample ($t\bar{t}H$, tH, and NTI data) to the predicted SM yield resulted in poor separation: the overwhelming majority of events were predicted to belong to the Class with highest cross section (Class 0, continuum background). Instead, each sample is normalized to equal yield. An additional test was performed where the relative normalizations of the Classes are optimized as hyper-parameters: the performance is found to be similar to the case where the Classes are normalized to equal yield.

G.1.1 Hadronic channel

Following the preselection in Chapter 9, hadronic channel events must contain zero leptons (e or μ) and at least three jets, one or more of which is b-tagged. Figures G.1 -G.6 compare the distribution of each training variable in $t\bar{t}H$, $tHjb$, and tWH Monte Carlo, as

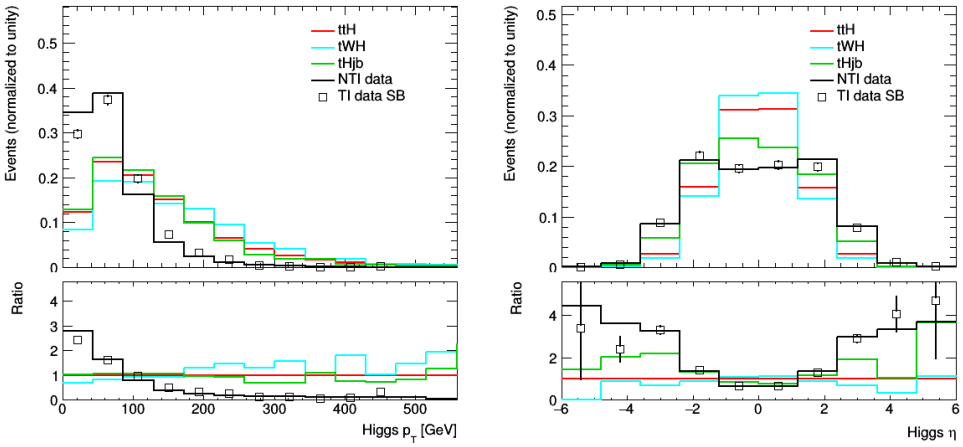


Figure G.1: Hadronic multiclass BDT training variables: $p_T^{\gamma\gamma}$ and $\eta^{\gamma\gamma}$. The open squares show TI data in the sideband region.

well as in NTI data. The TI data sidebands are overlaid for comparison to the NTI data control sample used as background in the training.

The BDT hyper-parameters are chosen to maximize the error evaluated on the validation set. The error is defined as the ratio of the number of misclassified events to the total number of events. Similar performance is obtained when the parameters are chosen to maximize the multiclass log-loss.

G.1.2 Leptonic channel

Again following the preselection in Chapter 9, leptonic channel events must contain at least one lepton (e or μ) and at least one b-jet. Figures G.7 -G.12 compare the distribution of each training variable in $t\bar{t}H$, $tHjb$, and tWH Monte Carlo, as well as in NTI data. The TI data sidebands are not overlaid due to low statistics in the leptonic channel.

The BDT hyper-parameters are chosen to maximize the error evaluated on the validation set. The error is defined as the ratio of the number of misclassified events to the total number of events. Similar performance is obtained when the parameters are chosen to maximize the multiclass log-loss.

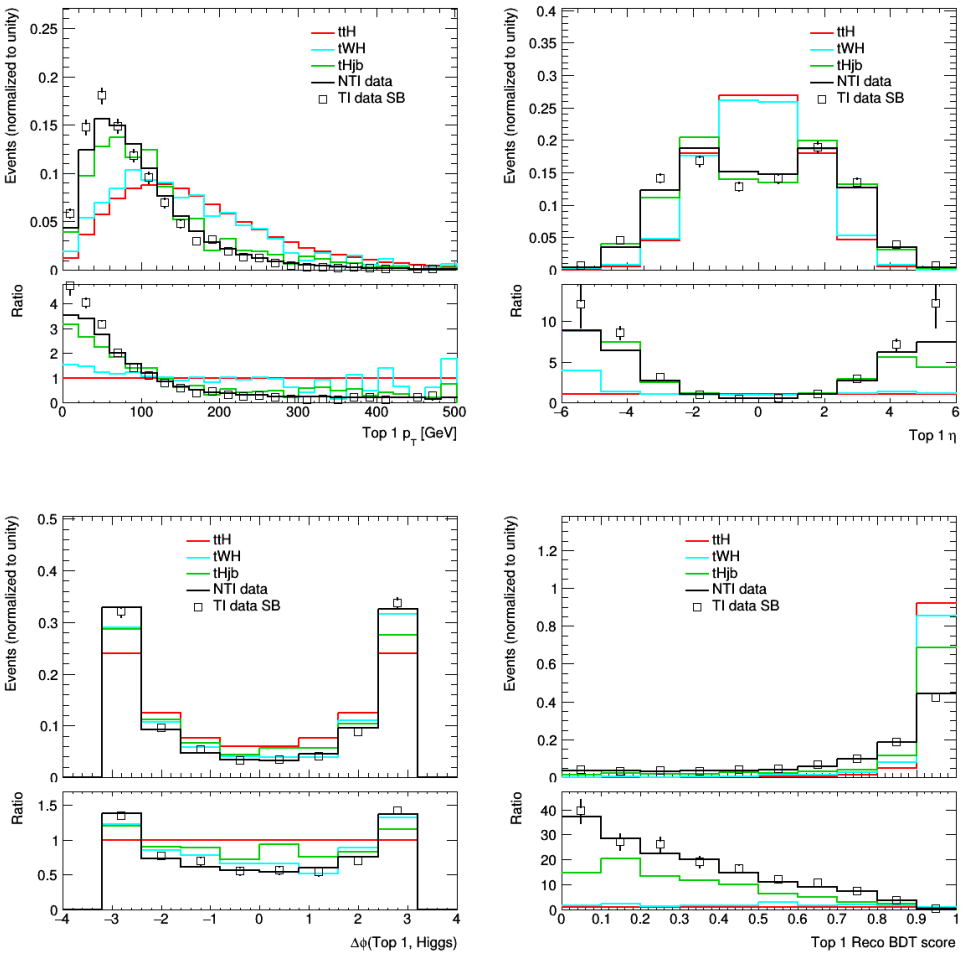


Figure G.2: Hadronic multiclass BDT ttraining variables: p_T, η, ϕ with respect to the Higgs candidate, and Top Reco BDT score of the first reconstructed top quark. The open squares show TI data in the sideband region.

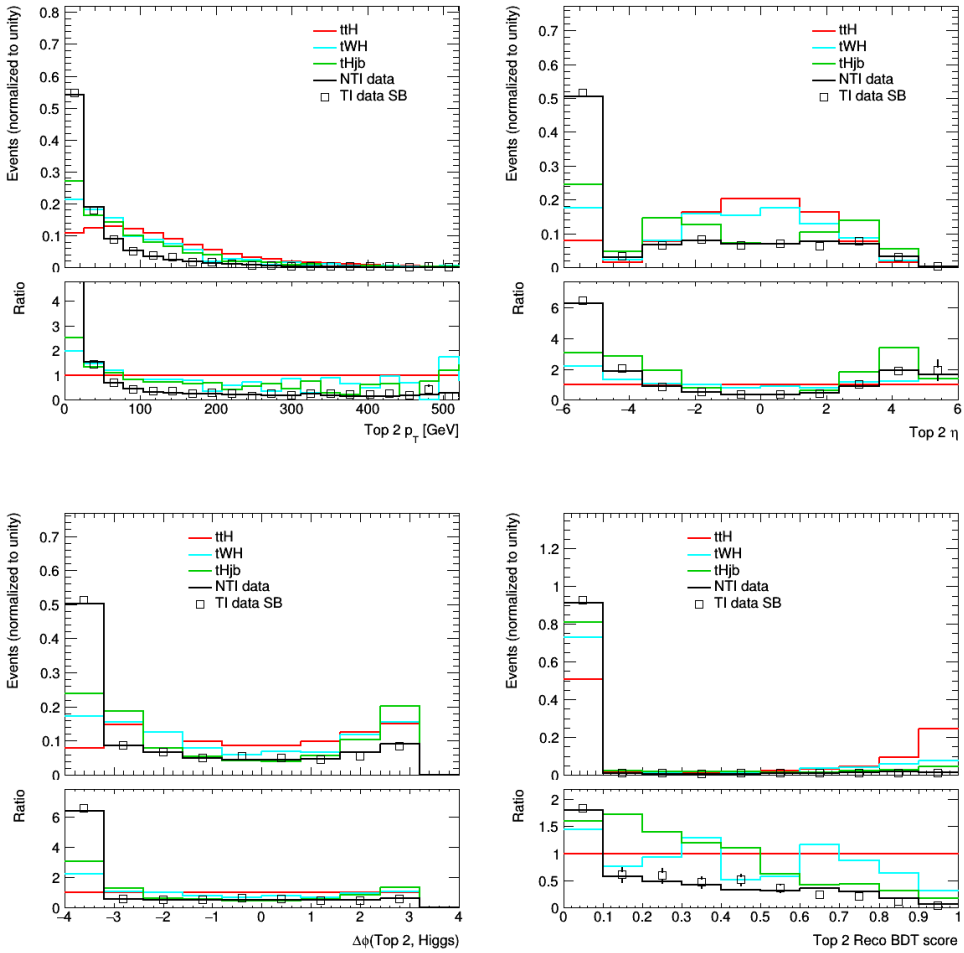


Figure G.3: Hadronic multiclass BDT training variables: p_T, η, ϕ with respect to the Higgs candidate, and Top Reco BDT score of the second reconstructed top quark. The underflow bins in p_T, η, and ϕ represent events where no second top is reconstructed (either partially or fully). The underflow bin in Top 2 Reco BDT score represents events where a second to cannot be fully reconstructed. The open squares show TI data in the sideband region.

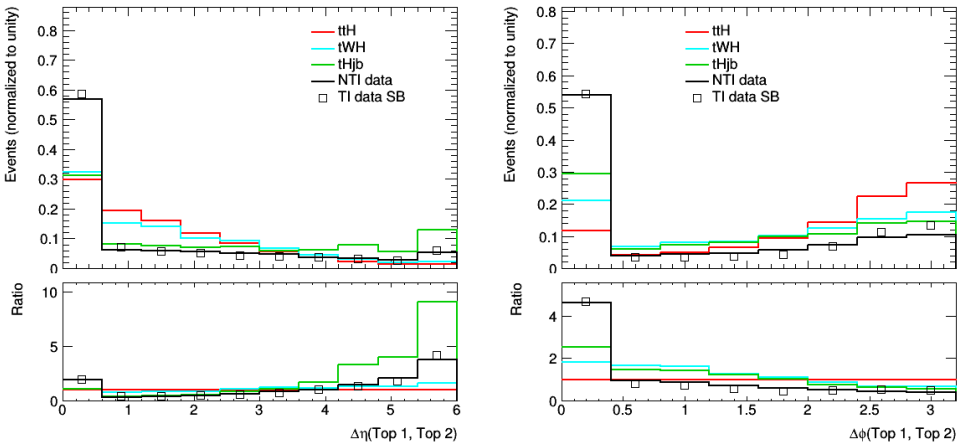


Figure G.4: Hadronic multiclass BDT training variables: $\Delta\eta(t_1, t_2)$ and $\Delta\phi(t_1, t_2)$. The underflow bins represent events where no second top is reconstructed (either partially or fully). The open squares show TI data in the sideband region.

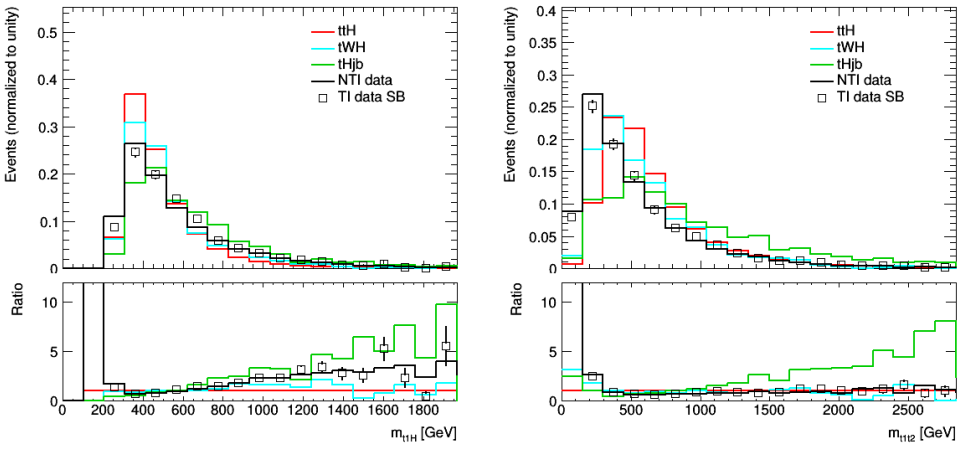


Figure G.5: Hadronic multiclass BDT training variables: $m_{t_1 H}$ and $m_{t_1 t_2}$. The underflow bin in $m_{t_1 t_2}$ represent events where no second top is reconstructed (either partially or fully). The open squares show TI data in the sideband region.

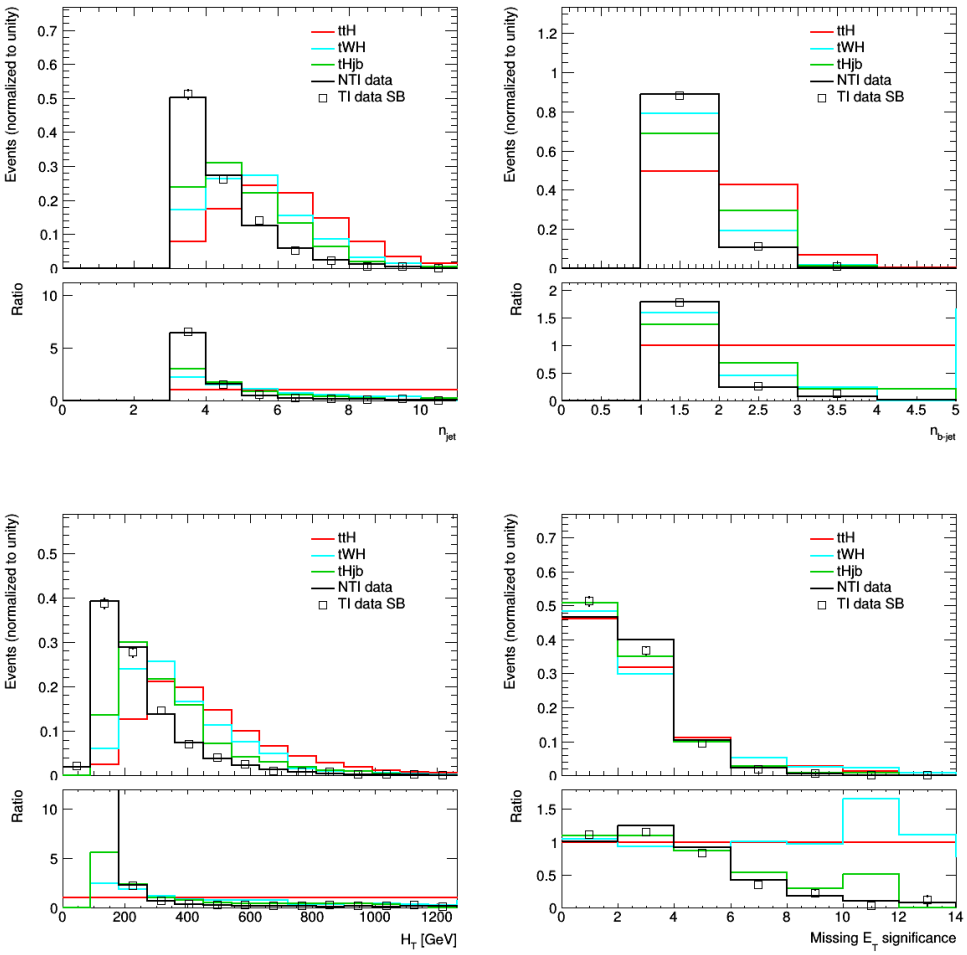


Figure G.6: Hadronic multiclass BDT training variables: jet multiplicity, b-jet multiplicity, H_T, and missing E_T significance. The open squares show TI data in the sideband region.

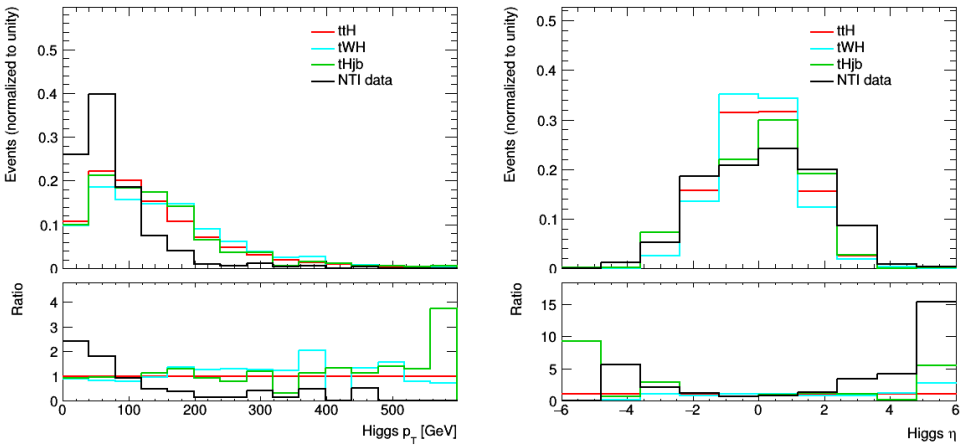


Figure G.7: Leptonic multiclass BDT training variables: $p_T^{\gamma\gamma}$ and $\eta^{\gamma\gamma}$.

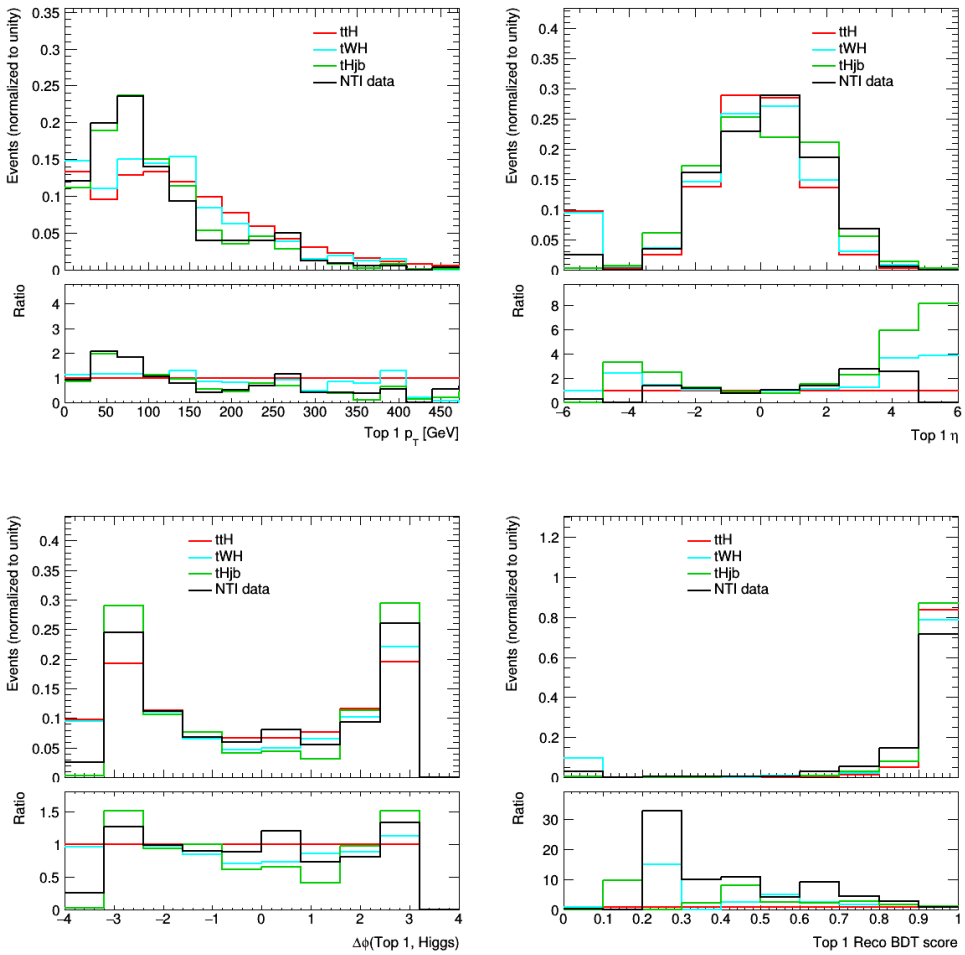


Figure G.8: Leptonic multiclass BDT training variables: p_T, η, ϕ with respect to the Higgs candidate, and Top Reco BDT score of the first reconstructed top quark. The underflow bin corresponds to events containing ≥ 2 leptons, where no top is reconstructed.

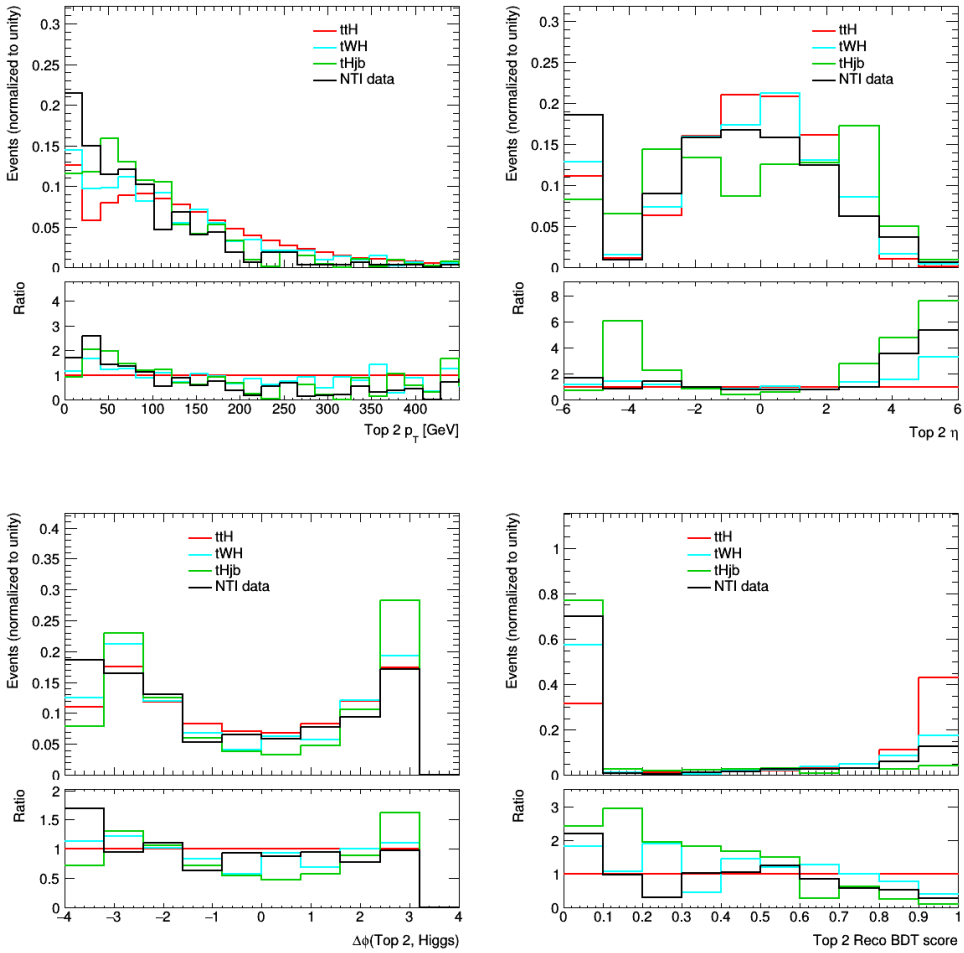


Figure G.9: Leptonic multiclass BDT training variables: p_T, η, ϕ with respect to the Higgs candidate, and Top Reco BDT score of the second reconstructed top quark. The underflow bins in p_T, η, and ϕ represent events containing ≥ 2 leptons or where no second top is reconstructed (either partially or fully). The underflow bin in Top 2 Reco BDT score represents events where the second top is not fully reconstructed.

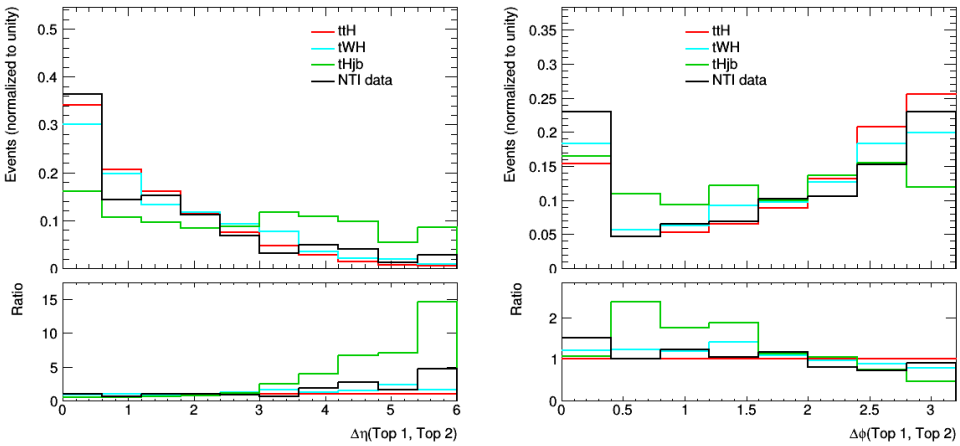


Figure G.10: Leptonic multiclass BDT training variables: $\Delta\eta(t_1, t_2)$ and $\Delta\phi(t_1, t_2)$. The underflow bins represent events containing ≥ 2 leptons or where no second top is reconstructed (either partially or fully).

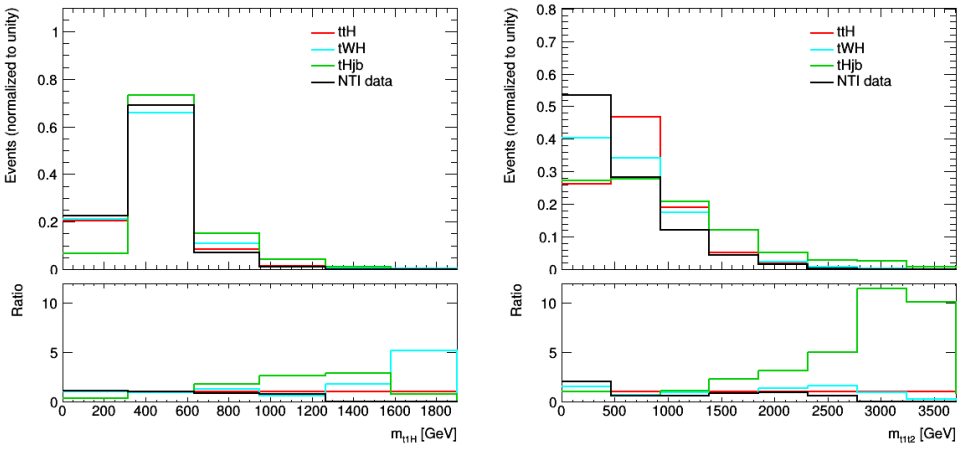


Figure G.11: Leptonic multiclass BDT training variables: $m_{t_1 H}$ and $m_{t_1 t_2}$. The underflow bin in $m_{t_1 t_2}$ represent events containing ≥ 2 leptons or where no second top is reconstructed (either partially or fully).

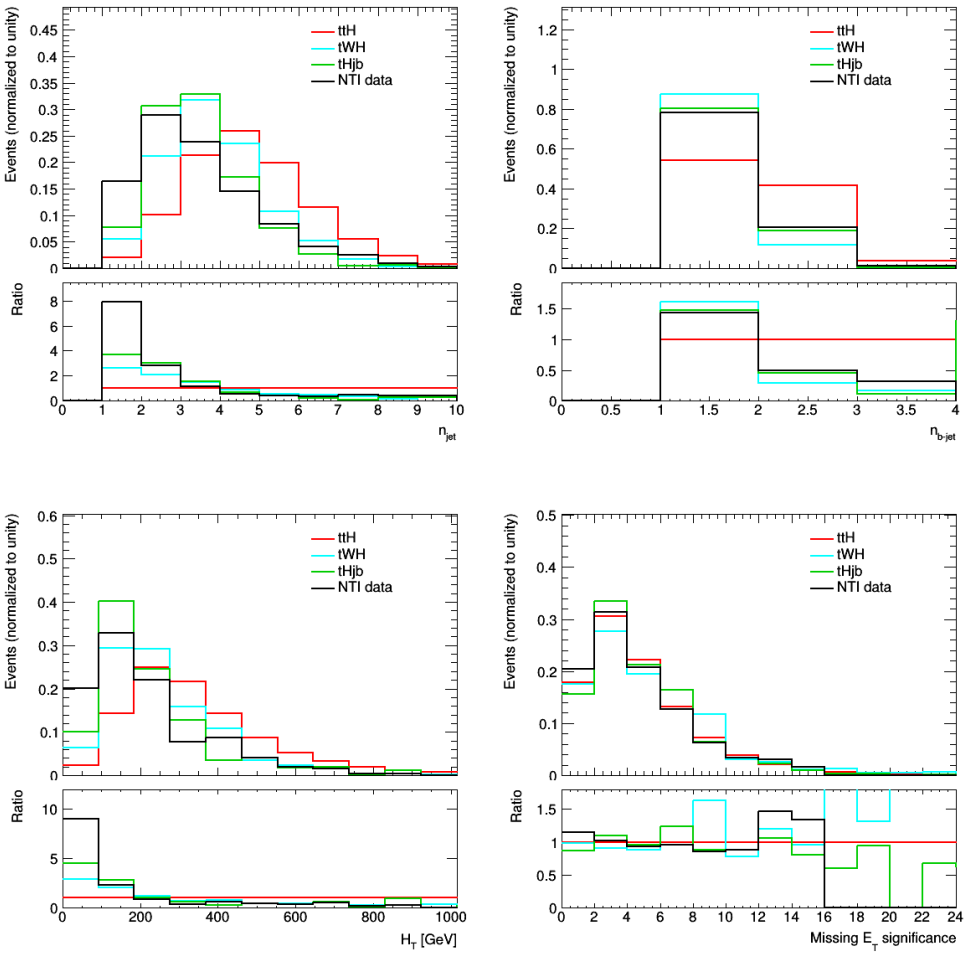
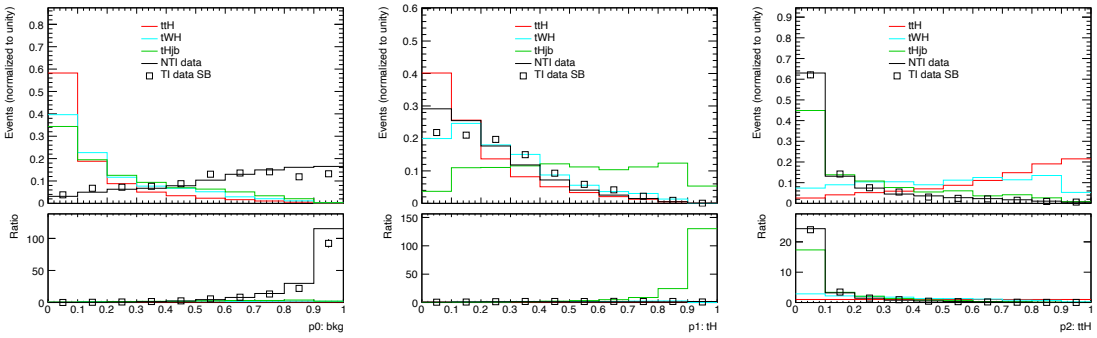


Figure G.12: Leptonic multiclass BDT training variables: jet multiplicity, b-jet multiplicity, H_T, and missing E_T significance.

G.2 Performance

The output of the multiclass BDT is a vector (P_0, P_1, P_2), where P_i is the probability that the event belongs to Class i.

Figure G.13 shows the scores P_0, P_1, and P_2 for the hadronic multiclass BDT for $t\bar{t}H$ Monte Carlo, $tHjb$ Monte Carlo, and NTI data. Each sample behaves as expected: for sample j, P_j should peak near 1 (sample j should be very j-like) and P_k should peak near 0 for $k \neq j$ (sample j should not be very k-like). The tWH Monte Carlo is similar in shape to $t\bar{t}H$ Monte Carlo in all three scores, reflecting the similarity between tWH and $t\bar{t}H$ kinematics. There is good agreement between NTI data and TI data sidebands.

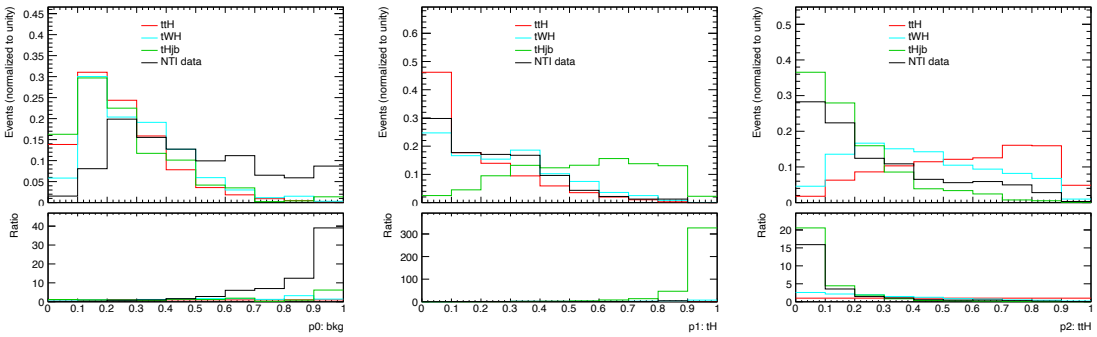


(a) Hadronic P_0 (bkg) (b) Hadronic P_1 (tH) (c) Hadronic P_2 ($t\bar{t}H$)

Figure G.13: Multiclass BDT output in the hadronic channel. $t\bar{t}H$ Monte Carlo is shown in red, $tHjb$ Monte Carlo in green, tWH Monte Carlo in cyan, and the data control regions in black. P_i is the probability that the event belongs to Class i, where $i = 0$ corresponds to background, $i = 1$ corresponds to tH, and $i = 2$ corresponds to $t\bar{t}H$.

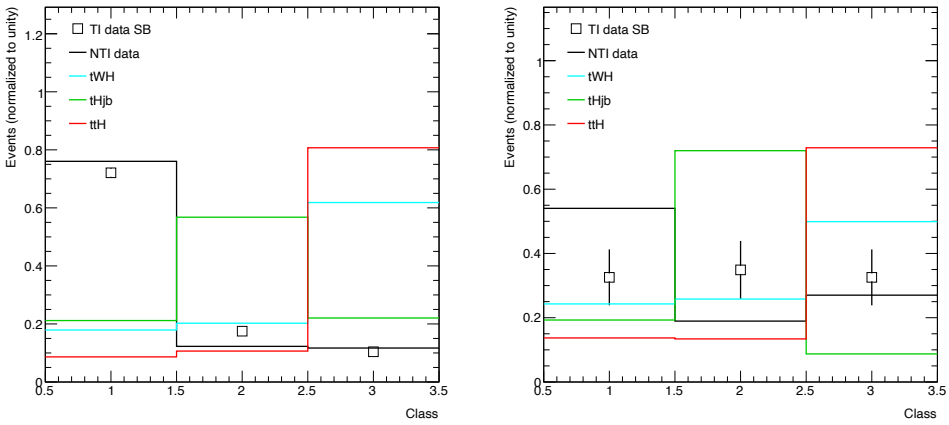
Figure G.14 shows the scores P_0, P_1, and P_2 for the leptonic multiclass BDT for $t\bar{t}H$ Monte Carlo, $tHjb$ Monte Carlo, and NTI data. Each sample again behaves as expected, and the tWH Monte Carlo is again similar in shape to $t\bar{t}H$ Monte Carlo in all three scores. The TI data sideband distributions are not shown due to very low statistics in the leptonic channel.

Each event is assigned to a class based on which probability is highest. If $P_0 > P_1,\ P_2$, then the event falls into Class 0; if $P_1 > P_0,\ P_2$, then the event falls into Class 1; and if $P_2 > P_0,\ P_1$, then the event falls into Class 2. The efficiency of these Classes is shown in Figure G.15 for the hadronic and leptonic channels. As expected, the background (NTI data) is concentrated in Class 0, $tHjb$ Monte Carlo is concentrated in Class 1, and $t\bar{t}H$ Monte Carlo is concentrated in Class 2. The TI sideband data is overlaid for comparison. Though agreement between NTI data and the TI data sidebands is good in the hadronic channel, the agreement in the leptonic channel is rather poor and the statistical errors are large.



(a) Leptonic P_0 (bkg) (b) Leptonic P_1 (tH) (c) Leptonic P_2 ($t\bar{t}H$)

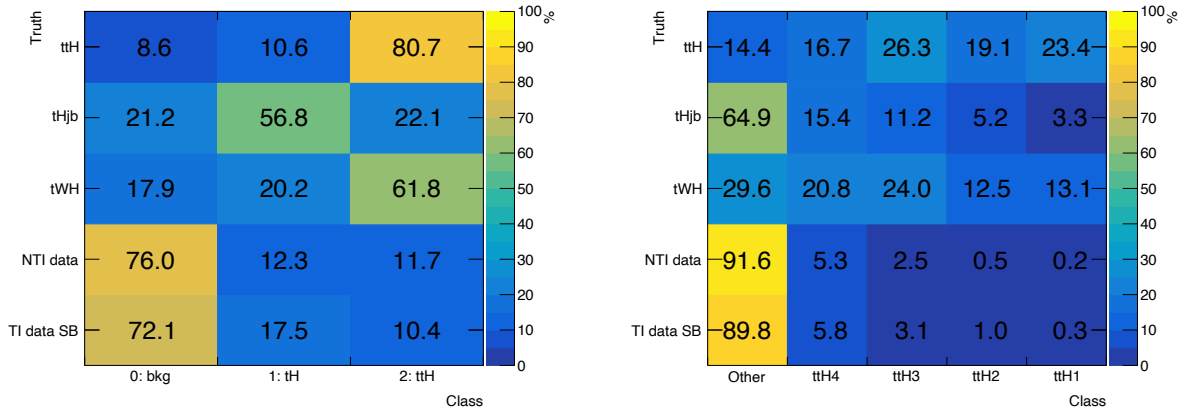
Figure G.14: Multiclass BDT output in the leptonic channel. $t\bar{t}H$ Monte Carlo is shown in red, $tHjb$ Monte Carlo in green, tWH Monte Carlo in cyan, and the NTI data control region in black. P_i is the probability that the event belongs to Class i, where $i = 0$ corresponds to background, $i = 1$ corresponds to tH, and $i = 2$ corresponds to $t\bar{t}H$.



(a) Hadronic channel (b) Leptonic channel

Figure G.15: Efficiency of each multiclass BDT Class on $t\bar{t}H$, tH, and background samples in the hadronic (left) and leptonic (right) channels.

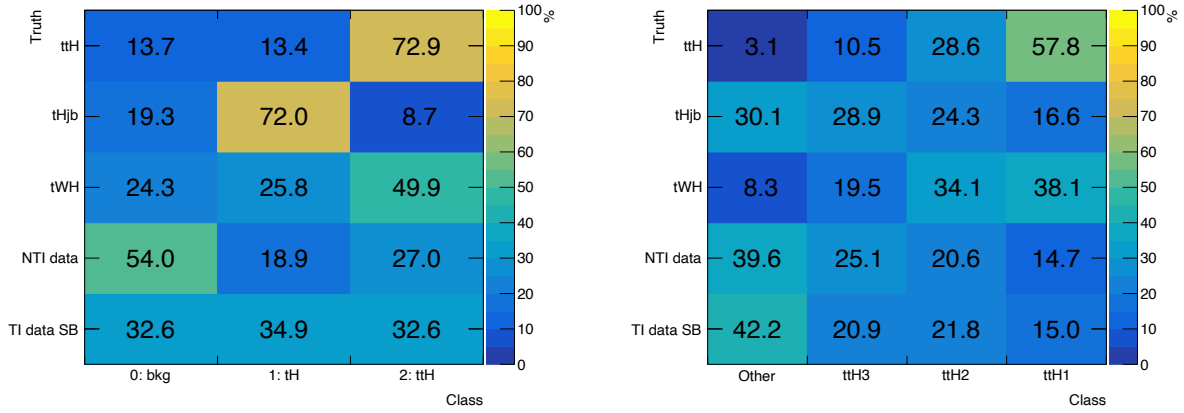
The efficiency of the multiclass BDT Classes can be compared to the efficiency of the $t\bar{t}H(\gamma\gamma)$ categories defined in Chapter 9 using binary BDTs. Figures G.16 and G.17 show these comparisons for the hadronic and leptonic channel, respectively.



(a) Hadronic multiclassification

(b) Hadronic binary categories

Figure G.16: Efficiency of each hadronic multiclass BDT Class on $t\bar{t}H$, tH, and background samples (left), compared to the efficiency in the hadronic $t\bar{t}H(\gamma\gamma)$ binary BDT categories (right).



(a) Leptonic multiclassification

(b) Leptonic binary categories

Figure G.17: Efficiency of each leptonic multiclass BDT Class on $t\bar{t}H$, tH, and background samples (left), compared to the efficiency in the leptonic $t\bar{t}H(\gamma\gamma)$ binary BDT categories (right).

G.2.1 Sensitivity estimate at 139 fb^{-1}

To evaluate the $t\bar{t}H$ significance $Z_{t\bar{t}H}$ using the multiclass BDT strategy, fourteen categories are considered:

- Class 2 is divided into four (three) categories based on the hadronic (leptonic) binary $t\bar{t}H(\gamma\gamma)$ BDT score
- Class 1 is divided into four (three) categories based on the hadronic (leptonic) binary $t\bar{t}H(\gamma\gamma)$ BDT score

Because the binary BDT has not been optimized independently in Class 1 and Class 2, the sensitivity calculated this way is not optimal, but gives an estimate of the performance.

To evaluate the tH significance Z_{tH} using the multiclass BDT strategy, eighteen categories are considered:

- The same fourteen categories used to calculate $Z_{t\bar{t}H}$
- Hadronic (leptonic) events in Class 2 that are rejected by the hadronic (leptonic) binary $t\bar{t}H(\gamma\gamma)$ BDT
- Hadronic (leptonic) events in Class 1 that are rejected by the hadronic (leptonic) binary $t\bar{t}H(\gamma\gamma)$ BDT

Most tH events are rejected by the binary BDTs (as shown in Figures G.16-G.17) and therefore fall into these additional four categories. Again, the sensitivity calculated this way is not optimal, but gives an estimate of the performance.

The single-category significances are evaluated according to Equation 9.3, with $S = t\bar{t}H$ for $Z_{t\bar{t}H}$ and $S = tHjb + tWH$ for Z_{tH}. Table G.1 (G.2) shows the $t\bar{t}H$ and tH significances where the continuum background is estimated from the NTI data (TI data sidebands). The hadronic, leptonic, and total significances are obtained by adding the relevant single-category significances in quadrature.

Significance	Multiclass	Binary	% difference
$Z_{t\bar{t}H}$ had	3.66	3.58	2.1
Z_{tH} had	0.190	0.168	11.5
$Z_{t\bar{t}H}$ lep	3.26	3.42	-4.9
Z_{tH} lep	0.242	0.184	24.0
$Z_{t\bar{t}H}$ total	4.91	4.95	-0.8
Z_{tH} total	0.308	0.249	19.2

Table G.1: Sensitivity to $t\bar{t}H$ and tH obtained from the multiclass BDT categories and the binary $t\bar{t}H$ categories at 139 fb^{-1}, where the continuum background yield estimated from scaled NTI data.

Significance	Multiclass	Binary	% difference
$Z_{t\bar{t}H}$ had	3.84	3.12	18.8
Z_{tH} had	0.160	0.161	-0.6
$Z_{t\bar{t}H}$ lep	3.68	3.43	6.8
Z_{tH} lep	0.214	0.186	13.1
$Z_{t\bar{t}H}$ total	5.32	4.64	12.8
Z_{tH} total	0.268	0.246	8.2

Table G.2: Sensitivity to $t\bar{t}H$ and tH obtained from the multiclass BDT categories and the binary $t\bar{t}H$ categories at 139 fb^{-1}, where the continuum background yield estimated by scaling the TI data sidebands.

The sensitivity estimate depends somewhat on the choice of control sample used to estimate the continuum background, as seen in previous Chapters. This is particularly true in the leptonic channel. However, in both background estimation schemes the multiclass BDT approach achieves similar $t\bar{t}H$ sensitivity to the binary BDT categories, and $\mathcal{O}(10\%)$ higher tH significance. Optimization of the categories defined within each Class (e.g. using dedicated binary BDTs) promises to bring additional sensitivity.

Bibliography

[1] Kurt Riesselmannl. "The Standard Model of particle physics". In: *Symmetry* (2015).

[2] Nima Arkani-Hamed, Savas Dimopoulos, and G.R. Dvali. "The Hierarchy problem and new dimensions at a millimeter". In: *Phys. Lett. B* 429 (1998), pp. 263–272. DOI: 10.1016/S0370-2693(98)00466-3. arXiv: hep-ph/9803315.

[3] Lisa Randall and Raman Sundrum. "A Large mass hierarchy from a small extra dimension". In: *Phys. Rev. Lett.* 83 (1999), pp. 3370–3373. DOI: 10.1103/PhysRevLett.83.3370. arXiv: hep-ph/9905221.

[4] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[5] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[6] Georges Aad et al. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex].

[7] A. D. Sakharov. "Violation of CP in variance, C asymmetry, and baryon asymmetry of the universe". In: *Phys. Usp.* 34.5 (1991), pp. 392–393. DOI: 10.1070/PU1991v034n05ABEH002497. URL: https://ufn.ru/en/articles/1991/5/h/.

[8] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[9] Salvatore Mele. "Measurements of the running of the electromagnetic coupling at LEP". In: *Proceedings, 26th International Symposium on Physics in Collision (PIC 2006): Buzios, Brazil, July 6-9, 2006*. 2006. arXiv: hep-ex/0610037 [hep-ex]. URL: http://www.slac.stanford.edu/econf/C060706/pdf/0610037.pdf.

[10] John C. Collins, Davison E. Soper, and George F. Sterman. "Factorization of Hard Processes in QCD". In: vol. 5. 1989, pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.

[11] Andy Buckley et al. "General-purpose event generators for LHC physics". In: *Phys. Rept.* 504 (2011), pp. 145–233. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].

[12] R. Keith Ellis, W. James Stirling, and B. R. Webber. "QCD and collider physics". In: *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* 8 (1996), pp. 1–435.

[13] Andy Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *Eur. Phys. J.* C75 (2015), p. 132. DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph].

[14] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J.* C75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[15] G. Corcella et al. "HERWIG 6: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)". In: *JHEP* 01.hep-ph/0011363. CAVENDISH-HEP-99-03. CERN-TH-2000-284. RAL-TR-2000-048 (2000), 010. 91 p. DOI: 10.1088/1126-6708/2001/01/010. URL: https://cds.cern.ch/record/478828.

[16] Stefano Catani et al. "The Dipole formalism for next-to-leading order QCD calculations with massive partons". In: *Nucl. Phys. B* 627 (2002), pp. 189–265. DOI: 10.1016/S0550-3213(02)00098-6. arXiv: hep-ph/0201036.

[17] Johan Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *Eur. Phys. J. C* 53 (2008), pp. 473–500. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].

[18] Bryan R Webber. "Hadronization". In: hep-ph/9411384. CAVENDISH-HEP-94-17 (1994). URL: https://cds.cern.ch/record/272660.

[19] Bo Andersson et al. "Parton Fragmentation and String Dynamics". In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[20] Richard D. Field and Stephen Wolfram. "A QCD Model for e+ e- Annihilation". In: *Nucl. Phys. B* 213 (1983), pp. 65–84. DOI: 10.1016/0550-3213(83)90175-X.

[21] *ATLAS Pythia 8 tunes to 7 TeV datas*. Tech. rep. ATL-PHYS-PUB-2014-021. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1966419.

[22] S.L. Glashow. "Towards a Unified Theory: Threads in a Tapestry". In: *Rev. Mod. Phys.* 52 (1980), pp. 539–543. DOI: 10.1103/RevModPhys.52.539.

[23] Steven Weinberg. "Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions". In: *Rev. Mod. Phys.* 52 (1980), pp. 515–523. DOI: 10.1103/RevModPhys.52.515.

[24] Abdus Salam. "Gauge Unification of Fundamental Forces". In: *Rev. Mod. Phys.* 52 (1980), pp. 525–538. DOI: 10.1103/RevModPhys.52.525.

[25] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[26] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[27] Georges Aad et al. "Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector". In: *Phys. Lett. B* 800 (2020), p. 135103. DOI: 10.1016/j.physletb.2019.135103. arXiv: 1906.02025 [hep-ex].

[28] Albert M Sirunyan et al. "Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s}$ = 13 TeV". In: *Phys. Rev. Lett.* 122.12 (2019), p. 121803. DOI: 10.1103/PhysRevLett.122.121803. arXiv: 1811.09689 [hep-ex].

[29] *Measurement prospects of the pair production and self-coupling of the Higgs boson with the ATLAS experiment at the HL-LHC*. Tech. rep. ATL-PHYS-PUB-2018-053. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2652727.

[30] *Sensitivity projections for Higgs boson properties measurements at the HL-LHC*. Tech. rep. CMS-PAS-FTR-18-011. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2647699.

[31] Georges Aad et al. "Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s}$ = 13 TeV collected with the ATLAS experiment". In: *Phys. Rev. D* 101.1 (2020), p. 012002. DOI: 10.1103/PhysRevD.101.012002. arXiv: 1909.02845 [hep-ex].

[32] *Combined Higgs boson production and decay measurements with up to 137 fb-1 of proton-proton collision data at sqrts = 13 TeV*. Tech. rep. CMS-PAS-HIG-19-005. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2706103.

[33] J R Andersen et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties". In: (July 2013). Ed. by S Heinemeyer et al. DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ph].

[34] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. Tech. rep. FERMILAB/CERN, 2017. URL: https://cds.cern.ch/record/2227475.

[35] Federico Demartin et al. "Higgs production in association with a single top quark at the LHC". In: *Eur. Phys. J.* C75.6 (2015), p. 267. DOI: 10.1140/epjc/s10052-015-3475-9. arXiv: 1504.00611 [hep-ph].

[36] Federico Demartin et al. "tWH associated production at the LHC". In: *Eur. Phys. J.* C77.1 (2017), p. 34. DOI: 10.1140/epjc/s10052-017-4601-7. arXiv: 1607.05862 [hep-ph].

[37] Georges Aad et al. "Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex].

[38] A. David et al. "LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle". In: (Sept. 2012). arXiv: 1209.0040 [hep-ph].

[39] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Phys. Rept.* 516 (2012), pp. 1–102. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph].

[40] Pierre Fayet. "Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions". In: *Phys. Lett. B* 69 (1977), p. 489. DOI: 10.1016/0370-2693(77)90852-8.

[41] A. Arhrib et al. "The Higgs Potential in the Type II Seesaw Model". In: *Phys. Rev.* D84 (2011), p. 095005. DOI: 10.1103/PhysRevD.84.095005. arXiv: 1105.1925 [hep-ph].

[42] Georges Aad et al. "Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using *pp* collisions at $\sqrt{s} = 13$ TeV". In: (2020). arXiv: 2002.12223 [hep-ex].

[43] Morad Aaboud et al. "Search for doubly charged scalar bosons decaying into same-sign W boson pairs with the ATLAS detector". In: *Eur. Phys. J.* C79.1 (2019), p. 58. DOI: 10.1140/epjc/s10052-018-6500-y. arXiv: 1808.01899 [hep-ex].

[44] Morad Aaboud et al. "Search for charged Higgs bosons decaying into top and bottom quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 11 (2018), p. 085. DOI: 10.1007/JHEP11(2018)085. arXiv: 1808.03599 [hep-ex].

[45] *Constraints on an effective Lagrangian from the combined $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels using 36.1 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data collected with the ATLAS detector*. Tech. rep. ATL-PHYS-PUB-2017-018. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2293084.

[46] Georges Aad et al. "Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector". In: *Eur. Phys. J.* C76.12 (2016), p. 658. DOI: 10.1140/epjc/s10052-016-4499-5. arXiv: 1602.04516 [hep-ex].

[47] V. Khachatryan et al. "Combined search for anomalous pseudoscalar HVV couplings in VH(H $\to b\bar{b}$) production and H $\to$ VV decay". In: *Phys. Lett.* B759 (2016), pp. 672–696. DOI: 10.1016/j.physletb.2016.06.004. arXiv: 1602.04305 [hep-ex].

[48] CMS Collaboration. "Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state". In: (2019). arXiv: 1901.00174 [hep-ex].

[49] Joachim Brod, Ulrich Haisch, and Jure Zupan. "Constraints on CP-violating Higgs couplings to the third generation". In: *JHEP* 11 (2013), p. 180. DOI: 10.1007/JHEP11(2013)180. arXiv: 1310.1385 [hep-ph].

[50] John Ellis et al. "Disentangling Higgs-Top Couplings in Associated Production". In: *JHEP* 04 (2014), p. 004. DOI: 10.1007/JHEP04(2014)004. arXiv: 1312.5736 [hep-ph].

[51] Federico Demartin et al. "Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction". In: *Eur. Phys. J.* C74.9 (2014), p. 3065. DOI: 10.1140/epjc/s10052-014-3065-2. arXiv: 1407.5089 [hep-ph].

[52] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[53] Stefano Frixione and Bryan R. Webber. "Matching NLO QCD computations and parton shower simulations". In: *JHEP* 06 (2002), p. 029. arXiv: hep-ph/0204244.

[54] Thomas Sven Pettersson and P Lefèvre. *The Large Hadron Collider: conceptual design*. Tech. rep. CERN-AC-95-05-LHC. 1995. URL: https://cds.cern.ch/record/291782.

[55] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[56] "Linear accelerator 2". In: (2012). URL: https://cds.cern.ch/record/1997427.

[57] "Radiofrequency cavities". In: (2012). URL: https://cds.cern.ch/record/1997424.

[58] "The Proton Synchrotron Booster". In: (2012). URL: https://cds.cern.ch/record/1997372.

[59] "The Proton Synchrotron". In: (2012). URL: https://cds.cern.ch/record/1997189.

[60] "The Super Proton Synchrotron". In: (2012). URL: https://cds.cern.ch/record/1997188.

[61] Christiane Lefèvre. "The CERN accelerator complex. Complexe des accélérateurs du CERN". 2008. URL: https://cds.cern.ch/record/1260465.

[62] G. Aad et al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[63] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[64] Jr. Alves A.Augusto et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[65] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[66] Werner Herr and B Muratori. "Concept of luminosity". In: (2006). DOI: 10.5170/CERN-2006-002.361. URL: https://cds.cern.ch/record/941318.

[67] G. Avoni et al. "The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS". In: *JINST* 13.07 (2018), P07017. DOI: 10.1088/1748-0221/13/07/P07017.

[68] Joao Pequenao. "Computer generated image of the whole ATLAS detector". 2008. URL: https://cds.cern.ch/record/1095924.

[69] G. Aad et al. "The ATLAS Inner Detector commissioning and calibration". In: *Eur. Phys. J. C* 70 (2010), pp. 787–821. DOI: 10.1140/epjc/s10052-010-1366-7. arXiv: 1004.5293 [physics.ins-det].

[70] *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. 2010. URL: https://cds.cern.ch/record/1291633.

[71] B. Abbott et al. "Production and Integration of the ATLAS Insertable B-Layer". In: *JINST* 13.05 (2018), T05008. DOI: 10.1088/1748-0221/13/05/T05008. arXiv: 1803.00844 [physics.ins-det].

[72] G. Aad et al. "ATLAS pixel detector electronics and sensors". In: *JINST* 3 (2008), P07007. DOI: 10.1088/1748-0221/3/07/P07007.

[73] A. Ahmad et al. "The Silicon microstrip sensors of the ATLAS semiconductor tracker". In: *Nucl. Instrum. Meth. A* 578 (2007), pp. 98–118. DOI: 10.1016/j.nima.2007.04.157.

[74] *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331063.

[75] T. Akesson et al. "Status of design and construction of the Transition Radiation Tracker (TRT) for the ATLAS experiment at the LHC". In: *Nucl. Instrum. Meth. A* 522 (2004). Ed. by C. Favuzzi et al., pp. 131–145. DOI: 10.1016/j.nima.2004.01.033.

[76] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331061.

[77] G. Aad et al. "Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions". In: *Eur. Phys. J. C* 70 (2010), pp. 723–753. DOI: 10.1140/epjc/s10052-010-1354-y. arXiv: 0912.2642 [physics.ins-det].

[78] *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331062.

[79] G. Aad et al. "Readiness of the ATLAS Tile Calorimeter for LHC collisions". In: *Eur. Phys. J. C* 70 (2010), pp. 1193–1236. DOI: 10.1140/epjc/s10052-010-1508-y. arXiv: 1007.5423 [physics.ins-det].

[80] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331068.

[81] G. Aad et al. "Commissioning of the ATLAS Muon Spectrometer with Cosmic Rays". In: *Eur. Phys. J. C* 70 (2010), pp. 875–916. DOI: 10.1140/epjc/s10052-010-1415-2. arXiv: 1006.4384 [physics.ins-det].

[82] Georges Aad et al. "Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s} = 8$ TeV". In: *Eur. Phys. J. C* 75 (2015), p. 120. DOI: 10.1140/epjc/s10052-015-3325-9. arXiv: 1408.3179 [hep-ex].

[83] Georges Aad et al. "Performance of the ATLAS Trigger System in 2010". In: *Eur. Phys. J. C* 72 (2012), p. 1849. DOI: 10.1140/epjc/s10052-011-1849-1. arXiv: 1110.1530 [hep-ex].

[84] Morad Aaboud et al. "Performance of the ATLAS Trigger System in 2015". In: *Eur. Phys. J. C* 77.5 (2017), p. 317. DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661 [hep-ex].

[85] Joao Pequenao and Paul Schaffner. "How ATLAS detects particles: diagram of particle paths in the detector". 2013. URL: https://cds.cern.ch/record/1505342.

[86] *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s} = 13$ *TeV*. Tech. rep. ATLAS-CONF-2017-029. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2261772.

[87] *Early Inner Detector Tracking Performance in the 2015 data at* $\sqrt{s} = 13$ *TeV*. Tech. rep. ATL-PHYS-PUB-2015-051. Geneva: CERN, 2015. URL: http://cds.cern.ch/record/2110140.

[88] T Cornelissen et al. *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*. Tech. rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1020106.

[89] Morad Aaboud et al. "Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC". In: *Eur. Phys. J. C* 77.5 (2017), p. 332. DOI: 10.1140/epjc/s10052-017-4887-5. arXiv: 1611.10235 [physics.ins-det].

[90] Georges Aad et al. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *Eur. Phys. J. C* 77 (2017), p. 490. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex].

[91] Morad Aaboud et al. "Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data". In: *JINST* 14.03 (2019), P03017. DOI: 10.1088/1748-0221/14/03/P03017. arXiv: 1812.03848 [hep-ex].

[92] Georges Aad et al. "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data". In: *JINST* 14.12 (2019), P12006. DOI: 10.1088/1748-0221/14/12/P12006. arXiv: 1908.00005 [hep-ex].

[93] Georges Aad et al. "Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector". In: *Phys. Rev. D* 90.11 (2014), p. 112015. DOI: 10.1103/PhysRevD.90.112015. arXiv: 1408.7084 [hep-ex].

[94] The ATLAS Collaboration. "Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector. Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 98.arXiv:1802.04146. 5 (2018), 052005. 136 p. DOI: 10.1103/PhysRevD.98.052005. URL: https://cds.cern.ch/record/2302492.

[95] Morad Aaboud et al. "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 79.8 (2019), p. 639. DOI: 10.1140/epjc/s10052-019-7140-6. arXiv: 1902.04655 [physics.ins-det].

[96] Georges Aad et al. "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV". In: *Eur. Phys. J. C* 76.5 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].

[97] M. Aaboud et al. "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 96.7 (2017), p. 072002. DOI: 10.1103/PhysRevD.96.072002. arXiv: 1703.09665 [hep-ex].

[98] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[99] Georges Aad et al. "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector". In: *Eur. Phys. J. C* 76.11 (2016), p. 581. DOI: 10.1140/epjc/s10052-016-4395-z. arXiv: 1510.03823 [hep-ex].

[100] Georges Aad et al. "Performance of b-Jet Identification in the ATLAS Experiment". In: *JINST* 11.04 (2016), P04008. DOI: 10.1088/1748-0221/11/04/P04008. arXiv: 1512.01094 [hep-ex].

[101] Georges Aad et al. "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 79.11 (2019), p. 970. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex].

[102] *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*. Tech. rep. ATL-PHYS-PUB-2016-012. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2160731.

[103] *Plots of b-tagging performance before and after the installation of the Insertable B-Layer*. 2017. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2017-005/.

[104] Morad Aaboud et al. "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 78.arXiv:1802.08168. 11 (2018), 903. 66 p. DOI: 10.1140/epjc/s10052-018-6288-9. URL: https://cds.cern.ch/record/2305380.

[105] *E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC p-p collisions*. Tech. rep. ATLAS-CONF-2018-023. Geneva: CERN, 2018. URL: http://cds.cern.ch/record/2625233.

[106] *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. Tech. rep. ATLAS-CONF-2019-021. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2677054.

[107] Georges Aad et al. "Performance of electron and photon triggers in ATLAS during LHC Run 2". In: *Eur. Phys. J.* C80.1 (2020), p. 47. DOI: 10.1140/epjc/s10052-019-7500-2. arXiv: 1909.00761 [hep-ex].

[108] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. "A brief introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), p. 852. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[109] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[110] M. Bähr et al. "Herwig++ physics and manual". In: *Eur. Phys. J. C* 58 (2008), p. 639. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[111] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[112] Paolo Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[113] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[114] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[115] T. Gleisberg et al. "Event generation with SHERPA 1.1". In: *JHEP* 02 (2009), p. 007. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph].

[116] Stefan Höche et al. "QCD matrix elements and truncated showers". In: *JHEP* 05 (2009), p. 053. DOI: 10.1088/1126-6708/2009/05/053. arXiv: 0903.1219 [hep-ph].

[117] Tanju Gleisberg and Stefan Höche. "Comix, a new matrix element generator". In: *JHEP* 12 (2008), p. 039. DOI: 10.1088/1126-6708/2008/12/039. arXiv: 0808.3674 [hep-ph].

[118] Steffen Schumann and Frank Krauss. "A Parton shower algorithm based on Catani-Seymour dipole factorisation". In: *JHEP* 03 (2008), p. 038. DOI: 10.1088/1126-6708/2008/03/038. arXiv: 0709.1027 [hep-ph].

[119] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152. DOI: 10.1016/S0168-9002(01)00089-4.

[120] S. Agostinelli et al. "GEANT4—a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[121] G. Aad et al. "The ATLAS Simulation Infrastructure". In: *Eur. Phys. J. C* 70 (2010), pp. 823–874. DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568 [physics.ins-det].

[122] *Summary of ATLAS Pythia 8 tunes*. Tech. rep. ATL-PHYS-PUB-2012-003. Geneva: CERN, 2012. URL: https://cds.cern.ch/record/1474107.

[123] A. D. Martin et al. "Parton distributions for the LHC". In: *Eur. Phys. J.* C63 (2009), pp. 189–285. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].

[124] Keith Hamilton et al. "NNLOPS simulation of Higgs boson production". In: *JHEP* 10 (2013), p. 222. DOI: 10.1007/JHEP10(2013)222. arXiv: 1309.0017 [hep-ph].

[125] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys. G* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[126] Georges Aad et al. "Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *JHEP* 09 (2014), p. 145. DOI: 10.1007/JHEP09(2014)145. arXiv: 1406.3660 [hep-ex].

[127] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[128] *The new Fast Calorimeter Simulation in ATLAS*. Tech. rep. ATL-SOFT-PUB-2018-002. Geneva: CERN, 2018. URL: http://cds.cern.ch/record/2630434.

[129] Ahmed Hasib et al. "Supporting document : Probing Higgs production modes and couplings with the $H \to \gamma\gamma$ channel with the Run 2 of LHC in the ATLAS experiment". In: ATL-COM-PHYS-2016-1784 (2016). URL: https://cds.cern.ch/record/2238687.

[130] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[131] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. CERN-OPEN-2012-016. New York: New York U., 2012. URL: https://cds.cern.ch/record/1456844.

[132] Georges Aad et al. "Measurement of differential production cross-sections for a Z boson in association with b-jets in 7 TeV proton-proton collisions with the ATLAS detector". In: *JHEP* 10 (2014), p. 141. DOI: 10.1007/JHEP10(2014)141. arXiv: 1407.3643 [hep-ex].

[133] Georges Aad et al. "Measurement of the cross-section for W boson production in association with b-jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *JHEP* 06 (2013), p. 084. DOI: 10.1007/JHEP06(2013)084. arXiv: 1302.2929 [hep-ex].

[134] Morad Aaboud et al. "Measurement of inclusive and differential cross sections in the $H \to ZZ^* \to 4\ell$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 10 (2017), p. 132. DOI: 10.1007/JHEP10(2017)132. arXiv: 1708.02810 [hep-ex].

[135] G. Bohm and G. Zech. *Introduction to statistics and data analysis for physicists*. DESY, 2010. ISBN: 9783935702416. URL: http://www-library.desy.de/preparch/books/vstatmp_engl.pdf.

[136] Hastie, Trevor and Tibshirani, Robert and Friedman, Jerome. *The Elements of Statistical Learning*. 2nd ed. Springer, 2008.

[137] Yoav Freund and Robert E. Schapire. "A Short Introduction to Boosting". In: *In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*. Morgan Kaufmann, 1999, pp. 1401–1406.

[138] J. Friedman, T. Hastie, and R. Tibshirani. "Additive Logistic Regression: a Statistical View of Boosting". In: *The Annals of Statistics* 38.2 (2000).

[139] Jerome H. Friedman. "Greedy Function Approximation: A Gradient Boosting Machine". In: *Annals of Statistics* 29 (2000), pp. 1189–1232.

[140] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. URL: http://doi.acm.org/10.1145/2939672.2939785.

[141] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN: 026218253X.

[142] Tim Head et al. *scikit-optimize/scikit-optimize: v0.5.2*. Version v0.5.2. Mar. 2018. DOI: 10.5281/zenodo.1207017. URL: https://doi.org/10.5281/zenodo.1207017.

[143] Morad Aaboud et al. "Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector". In: *Phys. Rev.* D97.7 (2018), p. 072003. DOI: 10.1103/PhysRevD.97.072003. arXiv: 1712.08891 [hep-ex].

[144] Julia Bauer. "Prospects for the Observation of Electroweak Top Quark Production with the CMS Experiment. Perspektiven zur Beobachtung der elektroschwachen Produktion einzelner Top-Quarks mit dem CMS Experiment". Presented on 7 May 2010. 2010. URL: https://cds.cern.ch/record/1308713.

[145] Scott Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *CoRR* abs/1705.07874 (2017). arXiv: 1705.07874. URL: http://arxiv.org/abs/1705.07874.

[146] Johannes Erdmann et al. "A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework". In: *Nucl. Instrum. Meth.* A748 (2014), pp. 18–25. DOI: 10.1016/j.nima.2014.02.029. arXiv: 1312.5595 [hep-ex].

[147] The ATLAS Collaboration. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". In: *Phys. Lett. B* 784.arXiv:1806.00425 (2018), 173–191. 19 p. DOI: 10.1016/j.physletb.2018.07.035. URL: https://cds.cern.ch/record/2621167.

[148] Albert M Sirunyan et al. "Observation of $t\bar{t}H$ production". In: *Phys. Rev. Lett.* 120.23 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610 [hep-ex].

[149] *Measurement of Higgs boson production in association with a $t\bar{t}$ pair in the diphoton decay channel using 139 fb^{-1} of LHC data collected at $\sqrt{s} = 13$ TeV by the ATLAS experiment*. Tech. rep. ATLAS-CONF-2019-004. Geneva: CERN, 2019. URL: http://cds.cern.ch/record/2668103.

[150] Morad Aaboud et al. "Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev.* D97.7 (2018), p. 072016. DOI: 10.1103/PhysRevD.97.072016. arXiv: 1712.08895 [hep-ex].

[151] Georges Aad et al. "Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment". In: *Eur. Phys. J.* C76.1 (2016), p. 6. DOI: 10.1140/epjc/s10052-015-3769-y. arXiv: 1507.04548 [hep-ex].

[152] Georges Aad et al. "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 75.7 (2015), p. 349. DOI: 10.1140/epjc/s10052-015-3543-1. arXiv: 1503.05066 [hep-ex].

[153] Georges Aad et al. "Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector". In: *Phys. Lett. B* 749 (2015), pp. 519–541. DOI: 10.1016/j.physletb.2015.07.079. arXiv: 1506.05988 [hep-ex].

[154] Fawzi Boudjema et al. "Lab-frame observables for probing the top-Higgs interaction". In: *Phys. Rev.* D92.1 (2015), p. 015019. DOI: 10.1103/PhysRevD.92.015019. arXiv: 1501.03157 [hep-ph].

[155] S. Amor Dos Santos et al. "Probing the CP nature of the Higgs coupling in $t\bar{t}h$ events at the LHC". In: *Phys. Rev.* D96.1 (2017), p. 013004. DOI: 10.1103/PhysRevD.96.013004. arXiv: 1704.03565 [hep-ph].

[156] A L Read. "Presentation of search results: theCLstechnique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313. URL: https://doi.org/10.1088.

[157] Albert M Sirunyan et al. "Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at $\sqrt{s} =$ TeV". In: *Phys. Rev. D* 99.9 (2019), p. 092005. DOI: 10.1103/PhysRevD.99.092005. arXiv: 1811.09696 [hep-ex].

[158] R. D. Cousins. *Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms*. Tech. rep. UCLA, 2013. URL: http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf4.

[159] Georges Aad et al. "Search for strong gravity in multijet final states produced in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC". In: *JHEP* 03 (2016), p. 026. DOI: 10.1007/JHEP03(2016)026. arXiv: 1512.02586 [hep-ex].

[160] James A. Frost et al. "Phenomenology of Production and Decay of Spinning Extra-Dimensional Black Holes at Hadron Colliders". In: *JHEP* 10 (2009), p. 014. DOI: 10.1088/1126-6708/2009/10/014. arXiv: 0904.0979 [hep-ph].

[161] Georges Aad et al. "Search for low-scale gravity signatures in multi-jet final states with the ATLAS detector at $\sqrt{s} = 8$ TeV". In: *JHEP* 07 (2015), p. 032. DOI: 10.1007/JHEP07(2015)032. arXiv: 1503.08988 [hep-ex].

[162] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys.* B867 (2013), pp. 244–289. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[163] *Multijet simulation for 13 TeV ATLAS Analyses*. Tech. rep. ATL-PHYS-PUB-2019-017. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2672252.

[164] R. Kuhn S. Catania F. Krauss and B.R. Webber. "QCD Matrix Elements + Parton Showers". In: *JHEP* 11 (2001), p. 63. DOI: 10.1088/1126-6708/2001/11/063.

[165] J. Pumplin et al. "New generation of parton distributions with uncertainties from global QCD analysis". In: *JHEP* 07 (2002), p. 012. arXiv: hep-ph/0201195.

[166] Robert Matthew Carey. "Angular Distributions of Three Jet Events in Proton-Antiproton Collisions at the Fermilab Tevatron". In: (1989).

[167] M. Garcia-Sciveres et al. "The FE-I4 pixel readout integrated circuit". In: *Nuclear Instruments and Methods in Physics Research* A 636 (2011), S155–S159.

[168] FE-I4 Collaboration. "The FE-I4B integrated circuit guide". In: (2012).

[169] G. Balbi et al. "Measurements of Single Event Upset in ATLAS IBL". In: (Apr. 2020). arXiv: 2004.14116 [physics.ins-det].

[170] M. Garcia-Sciveres and N. Wermes. "A review of advances in pixel detectors for experiments with high rate and radiation". In: *Reports on Progress in Physics* 81 (2018), nbr. 6.

[171] T. Calin, M. Nicolaidis and R. Velazco. "Upset hardened memory by design for submicron CMOS technology SEU". In: *IEEE Trans. Nucl. Sci.* 43 (1996), p. 2874.

[172] M. Menouni et al. "SEU tolerant memory design for the ATLAS pixel readout chip". In: *JINST* 8 (2013), p. C02026.

[173] N. Wermes et al. *Pixel Detectors, From Fundamentals to Applications*. Springer-Verlag, 2006. ISBN: 3-540-28332-3. DOI: 10.1007/3-540-28333-1.

[174] Michael Moll. "Radiation damage in silicon particle detectors: Microscopic defects and macroscopic properties". PhD thesis. Hamburg U., 1999.

[175] Andreas Hoecker et al. "TMVA: Toolkit for Multivariate Data Analysis". In: *PoS* ACAT (2007), p. 040. arXiv: physics/0703039.

[176] Marco Delmastro et al. *Simplified Template Cross Sections – Stage 1.1*. Tech. rep. arXiv:1906.02754. 14 pages, 3 figures. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2669925.