

**Search for the production of Higgs bosons  
in association with top quarks  
and decaying into bottom quark pairs  
with the ATLAS detector**

by

Alexander Held

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies  
(Physics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2019

© Alexander Held, 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

**Search for the production of Higgs bosons in association with top quarks and decaying into bottom quark pairs with the ATLAS detector**

submitted by Alexander Held in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy  
in Physics

**Examining Committee:**

Oliver Stelzer-Chilton, Physics & Astronomy  
*Supervisor*

Colin Gay, Physics & Astronomy  
*Supervisory Committee Member*

Joanna Karczmarek, Physics & Astronomy  
*University Examiner*

Roman Krens, Chemistry  
*University Examiner*

**Additional Supervisory Committee Members:**

Gary Hinshaw, Physics & Astronomy  
*Supervisory Committee Member*

David Morrissey, TRIUMF  
*Supervisory Committee Member*

# Abstract

The Standard Model of particle physics (SM) describes mass generation of fundamental particles via the Brout-Englert-Higgs mechanism. It predicts Yukawa interactions between the Higgs boson and fermions, with interaction strengths proportional to the fermion masses. The largest Yukawa coupling is that of the top quark, and its value has implications in particle physics and cosmology. As the SM is not a complete theory of nature, detailed measurements of its predictions are a mandatory step towards improving the understanding of nature.

This dissertation presents a search for Higgs boson production in association with a top quark pair, a process directly sensitive to the top quark Yukawa coupling. The search uses  $36.1 \text{ fb}^{-1}$  of data at  $\sqrt{s} = 13 \text{ TeV}$ , collected with the ATLAS detector at the Large Hadron Collider (LHC) in 2015 and 2016. It is designed for Higgs boson decays to bottom quarks, and decays of the top quark pair resulting in final states with one or two electrons or muons. The discrimination between the signal Higgs boson production process and background processes, dominated by the production of top quark pairs, is performed with multivariate analysis techniques. The matrix element method is used and optimized for this search. Possible machine learning extensions of the method are investigated to help overcome its large computational demand. The obtained ratio of the measured cross-section for the signal Higgs boson production process to the prediction of the SM is  $\mu = 0.84^{+0.64}_{-0.61}$ . The expected sensitivity of an extension of the search, using  $139.0 \text{ fb}^{-1}$  of data collected between 2015 and 2018, is  $3.3\sigma$ . Data collected between 2016 and 2018 is also used in a measurement of the ATLAS muon trigger system efficiency.

A statistical combination of searches for Higgs bosons produced in association with top quark pairs is performed, including the search for Higgs boson decays to bottom quarks and additional final states. The combination results in the observation of this Higgs boson production process with an observed significance of  $5.4\sigma$ , compared to an expected sensitivity of  $5.5\sigma$ . It experimentally establishes top quark Yukawa interactions in the SM.

## Lay summary

The Standard Model of particle physics (SM) describes fundamental particles and their interactions. One of these particles is the Higgs boson, which participates in the mechanism that grants masses to fundamental particles. Without this mechanism, the universe would be completely different. The large role played by the Higgs boson in the SM motivates detailed studies of its behavior. Measuring the Higgs boson properties has become possible only in recent years, thanks to a particle collider called the Large Hadron Collider (LHC). This dissertation describes a measurement of the interaction between Higgs bosons and the heaviest known fundamental particles, called top quarks. Producing a Higgs boson at the LHC is rare, and only one out of every hundred Higgs bosons is produced together with top quarks. The measurement of this rare process experimentally establishes a new type of fundamental interaction with Higgs bosons and top quarks, as predicted by the SM.

# Preface

This dissertation is based on data collected with the ATLAS experiment. The operation of this experiment is made possible by an international collaboration of thousands of scientists. Many aspects of data collection and processing are shared among the collaboration, and performed by the members on top of the physics analyses they perform. The simulation of collision events relies on software and efforts from both within and outside the ATLAS collaboration. Analyses performed with the ATLAS experiment use software jointly developed by many members of the collaboration, and rely heavily on the computing resources provided.

All text in this dissertation is written by me, and no text is taken directly from previous public or internal documentation. The ATLAS collaboration releases final results in peer-reviewed journals, and preliminary results that do not undergo an external review process. Both types are reviewed within the collaboration. Figures labeled with "ATLAS" or "ATLAS Simulation" are taken from peer-reviewed publications. The label "ATLAS Preliminary" denotes preliminary public results approved by the collaboration. Figures labeled with "ATLAS work in progress" or "ATLAS Simulation work in progress" are prepared by me and not approved by the collaboration. They use data from the ATLAS detector or official samples of events simulated with ATLAS software. Figures not prepared by me have their reference indicated in the caption.

My primary work is summarized in chapters 6, 7, 9, 10, and 11. Detailed contributions and acknowledgements are listed below. As a member of the ATLAS collaboration, I contributed to two aspects of the common shared efforts enabling the detector operation and analysis of results. Within the muon trigger group, I measured the efficiency of the ATLAS muon trigger system for data collected between 2015 and 2018. The results of this work are summarized in chapter 10. I am also part of the core development and support team of a software package for statistical inference, used for many analyses in the ATLAS collaboration. This package is called TREXFITTER, and is only available to the collaboration. It is used for the results presented in chapters 6 and 9.

Chapters 6 and 7 present a search for the  $t\bar{t}H(b\bar{b})$  process with  $36.1 \text{ fb}^{-1}$  of data. This search was performed by a group of around 70 people, and I was part of a small core team of analyzers for the single-lepton channel. The search has been published as

*ATLAS Collaboration, 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  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector, Phys. Rev. D 97 (2018) 072016, arXiv: 1712.08895 [hep-ex].*

I performed extensive studies of the fit model in the single-lepton channel in close collaboration with other analyzers, particularly T. Calvet. This includes contributions towards the definition of the nominal  $t\bar{t}$  model and the systematic uncertainties assigned to it, as well as the derivation of

systematic uncertainties affecting several small background processes. Independent implementations of the processing of samples and subsequent fit in the single-lepton channel by G. Aad, T. Calvet, and me, were used to validate the fit results. I performed detailed studies of the post-fit modeling of data to validate the fit model. Many of the histograms used to produce the supporting figures accompanying the publication as auxiliary material were produced by me. I developed and optimized the matrix element method (MEM) used in the search. I studied its combination with other multivariate analysis techniques in collaboration with several other analyzers, particularly D. Mori. C. David wrote the software allowing to access variables related to the MEM with a common analysis framework used in the  $t\bar{t}H(b\bar{b})$  group. The work documented in chapter 7 was performed by me, with the exception of the transfer function, which was derived by D. Mori. The description of the MEM in the publication was written by me. I additionally described the method in internal ATLAS documentation, where I also summarized the effect of systematic uncertainties on the modeling of  $t\bar{t}$  production. The MEM implementation was further documented by me in deliverable 1.4 of the AMVA4NewPhysics innovative training network, funded by the European Union's Horizon 2020 research and innovation program. The deliverable was jointly prepared with other members of this network.

Chapter 8 describes the statistical combination of  $t\bar{t}H$  analyses performed with the ATLAS detector. The evidence for the  $t\bar{t}H$  process has been published as

ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. D 97 (2018) 072003, arXiv: 1712.08891 [hep-ex].

For this combination of results, I studied the effect of removing nuisance parameters describing small systematic uncertainties in the  $t\bar{t}H(b\bar{b})$  search. This decreases the fitting time without significantly affecting the results. The subsequent observation of the  $t\bar{t}H$  process has been published in reference [3].

Chapter 9 presents a sensitivity study of a  $t\bar{t}H(b\bar{b})$  search with  $139.0 \text{ fb}^{-1}$  of data. The study was performed within the ATLAS  $t\bar{t}H(b\bar{b})$  group and in collaboration with other analyzers. Within the group, I was responsible for the statistical analysis in the single-lepton channel, and a core analyzer in this channel. In this role, I worked on the definition of the fit model and implemented it in TREXFITTER. I studied the modeling of data, contributed towards defining the event selection and region definitions, and compared two available  $b$ -tagging algorithms. All results of the statistical analysis in the chapter were obtained by me. The analysis of the dataset used in the chapter is ongoing, and will be submitted for publication upon completion of the analysis.

The muon trigger efficiency measurement documented in chapter 10 was performed in collaboration with S. Rettie. The analysis was designed in collaboration with O. Stelzer-Chilton and J. Jovićević, with additional guidance from the ATLAS muon trigger group. I processed the required samples of data and simulated events with software developed within the ATLAS collaboration, and wrote and maintained a framework to process these samples into the histograms needed for the trigger efficiency computation. The tables in section 10.5 were obtained with code written by S. Rettie. The

measurements were described in documentation internal to the ATLAS collaboration, written by S. Rettie and me. A publication summarizing the ATLAS muon trigger performance during Run-2 of the Large Hadron Collider (LHC) is being prepared, and will include the measurements described in the chapter. I am part of the team of editors for this document.

Chapter 11 summarizes two methods of approximating differential cross-sections. The foam approach is based on an idea of T. Carli, and was developed with his guidance. I wrote the framework to test this method, and studied its behavior. The results documented in the chapter for the foam performance are obtained by me. I co-supervised two students together with T. Carli, who studied extensions of the project. T. Sandell optimized the implementation and investigated various methods of improving the performance of the foam. G. Van Goffrier extended the study to more complex final states. The neural network approach was implemented following an idea developed during a stay at the statistics department of EPFL in Lausanne. I developed the framework used for this approach and studied its performance. I supervised two students who contributed to the development of the method. J. Bamber studied optimizations of the neural network architecture and parameters, and contributed to the development of the framework. C. Lewis performed the training for the six-particle final state. All required samples of simulated events were generated by me, and I implemented the differential cross-section calculation for all final states investigated throughout the project. The performance of the neural network approximation for six-particle final states is also described in internal ATLAS documentation I wrote.

# Table of contents

|  |              |
|--|--------------|
| <b>Abstract</b> . . . . .  | <b>iii</b>   |
| <b>Lay summary</b> . . . . .   | <b>iv</b>    |
| <b>Preface</b> . . . . .   | <b>v</b>     |
| <b>Table of contents</b> . . . . .                                     | <b>viii</b>  |
| <b>List of tables</b> . . . . .  | <b>xiv</b>   |
| <b>List of figures</b> . . . . .                                       | <b>xvi</b>   |
| <b>Acronyms</b> . . . . .  | <b>xxvii</b> |
| <b>Acknowledgements</b> . . . . .                                      | <b>xxx</b>   |
| <b>1 Introduction</b> . . . . .  | <b>1</b>     |
| <b>2 The Standard Model of particle physics</b> . . . . .              | <b>3</b>     |
| 2.1 Particles in the Standard Model . . . . .                          | 3            |
| 2.2 Quantum chromodynamics . . . . .                                   | 4            |
| 2.2.1 Running coupling . . . . .                                       | 5            |
| 2.2.2 Parton distribution functions . . . . .                          | 6            |
| 2.2.3 Parton-parton scattering . . . . .                               | 6            |
| 2.3 Electroweak theory . . . . .                                       | 6            |
| 2.3.1 Lagrangian density . . . . .                                     | 7            |
| 2.3.2 Electroweak symmetry breaking . . . . .                          | 10           |
| 2.3.3 Lagrangian density after electroweak symmetry breaking . . . . . | 11           |
| 2.4 Success and limitations of the Standard Model . . . . .            | 13           |
| 2.4.1 Open questions . . . . .   | 14           |
| 2.5 Implications for physics at the Large Hadron Collider . . . . .    | 17           |
| 2.5.1 Top quark . . . . .  | 17           |
| 2.5.2 Higgs boson . . . . .  | 17           |
| 2.5.3 Yukawa couplings and the special role of $t\bar{t}H$ . . . . .   | 20           |
| <b>3 The Large Hadron Collider and the ATLAS experiment</b> . . . . .  | <b>22</b>    |

---

|          |   |           |
|----------|---|-----------|
| 3.1      | The Large Hadron Collider . . . . .                                 | 22        |
| 3.1.1    | Accelerator chain . . . . .   | 22        |
| 3.1.2    | Luminosity and pile-up . . . . .                                    | 23        |
| 3.1.3    | Dataset . . . . .   | 24        |
| 3.2      | The ATLAS detector . . . . .  | 25        |
| 3.2.1    | Coordinate system . . . . .   | 25        |
| 3.2.2    | Detector overview . . . . .   | 26        |
| 3.2.3    | Inner Detector . . . . .  | 26        |
| 3.2.4    | Calorimeters . . . . .  | 28        |
| 3.2.5    | Muon spectrometer . . . . .   | 30        |
| 3.2.6    | Trigger and data acquisition . . . . .                              | 32        |
| 3.2.7    | Data quality requirements and available data for analyses . . . . . | 34        |
| 3.2.8    | Simulation of ATLAS . . . . .                                       | 34        |
| <b>4</b> | <b>Object reconstruction . . . . .</b>                              | <b>36</b> |
| 4.1      | Reconstruction overview . . . . .                                   | 36        |
| 4.2      | Tracks, vertices and energy clusters . . . . .                      | 37        |
| 4.2.1    | Tracks . . . . .  | 37        |
| 4.2.2    | Vertices . . . . .  | 38        |
| 4.2.3    | Energy clusters . . . . .   | 38        |
| 4.3      | Leptons . . . . .   | 39        |
| 4.3.1    | Muons . . . . .   | 39        |
| 4.3.2    | Electrons . . . . .   | 41        |
| 4.3.3    | Tau leptons . . . . .   | 43        |
| 4.4      | Jets and flavor tagging . . . . .                                   | 43        |
| 4.4.1    | Jets . . . . .  | 43        |
| 4.4.2    | Flavor tagging . . . . .  | 44        |
| 4.5      | Missing transverse energy . . . . .                                 | 46        |
| 4.6      | Overlap removal . . . . .   | 46        |
| <b>5</b> | <b>Statistical methods . . . . .</b>                                | <b>48</b> |
| 5.1      | Statistical modeling . . . . .                                      | 48        |
| 5.1.1    | Random variables . . . . .  | 48        |
| 5.1.2    | Common distributions . . . . .                                      | 49        |
| 5.1.3    | Likelihood function . . . . .                                       | 50        |
| 5.2      | Statistical inference . . . . .                                     | 51        |
| 5.2.1    | Parameter estimation . . . . .                                      | 51        |
| 5.2.2    | Hypothesis testing . . . . .  | 51        |
| 5.2.3    | Median significances and the Asimov dataset . . . . .               | 54        |
| 5.3      | Multivariate techniques . . . . .                                   | 55        |

|          |   |           |
|----------|---|-----------|
| 5.3.1    | Boosted decision trees . . . . .  | 55        |
| 5.3.2    | Neural networks . . . . .   | 57        |
| <b>6</b> | <b>Search for Higgs boson production in association with a top quark pair and decaying into a bottom quark pair . . . . .</b> | <b>59</b> |
| 6.1      | Analysis overview . . . . .   | 60        |
| 6.2      | Event selection . . . . .   | 62        |
| 6.2.1    | Dataset . . . . .   | 62        |
| 6.2.2    | Object definitions . . . . .  | 62        |
| 6.2.3    | Definition of the single-lepton and dilepton channels . . . . .   | 63        |
| 6.3      | Modeling . . . . .  | 64        |
| 6.3.1    | $t\bar{t}H$ signal . . . . .  | 64        |
| 6.3.2    | $t\bar{t}$ + jets background . . . . .  | 64        |
| 6.3.3    | Other backgrounds . . . . .   | 66        |
| 6.3.4    | Inclusive modeling of data . . . . .  | 68        |
| 6.4      | Event categorization . . . . .  | 70        |
| 6.4.1    | Region definitions . . . . .  | 70        |
| 6.4.2    | Region composition and signal contributions . . . . .   | 72        |
| 6.5      | Multivariate analysis techniques . . . . .  | 73        |
| 6.5.1    | Reconstruction BDT . . . . .  | 74        |
| 6.5.2    | Likelihood discriminant . . . . .   | 74        |
| 6.5.3    | Matrix element method . . . . .   | 75        |
| 6.5.4    | Classification BDT . . . . .  | 75        |
| 6.6      | Systematic uncertainties . . . . .  | 75        |
| 6.6.1    | Nuisance parameter details . . . . .  | 76        |
| 6.6.2    | Experimental uncertainties . . . . .  | 77        |
| 6.6.3    | Signal and background modeling . . . . .  | 78        |
| 6.6.4    | Summary of systematic uncertainty sources . . . . .   | 81        |
| 6.7      | Statistical analysis and results . . . . .  | 83        |
| 6.7.1    | Fit model details and expected performance . . . . .  | 83        |
| 6.7.2    | Fit to data . . . . .   | 84        |
| 6.7.3    | Dominant nuisance parameters and sources of uncertainty . . . . .   | 86        |
| 6.7.4    | Validation studies . . . . .  | 93        |
| 6.7.5    | Observed significance and upper limits . . . . .  | 94        |
| 6.7.6    | Summary distribution of events . . . . .  | 95        |
| <b>7</b> | <b>The matrix element method for <math>t\bar{t}H(b\bar{b})</math> . . . . .</b>   | <b>97</b> |
| 7.1      | The matrix element method . . . . .   | 97        |
| 7.1.1    | Parton level . . . . .  | 98        |
| 7.1.2    | Reconstructed objects . . . . .   | 99        |

|          |  |            |
|----------|--|------------|
| 7.2      | General approach for $t\bar{t}H(b\bar{b})$ . . . . .   | 99         |
| 7.2.1    | Permutations . . . . .   | 99         |
| 7.2.2    | Transfer function . . . . .  | 101        |
| 7.2.3    | Remaining degrees of freedom . . . . .   | 101        |
| 7.2.4    | Likelihoods and discriminant . . . . .   | 102        |
| 7.2.5    | Systematic uncertainties . . . . .   | 103        |
| 7.3      | Technical implementation . . . . .   | 103        |
| 7.3.1    | Integration . . . . .  | 103        |
| 7.3.2    | Matrix elements . . . . .  | 106        |
| 7.3.3    | Parton distribution functions . . . . .  | 107        |
| 7.3.4    | Transfer function . . . . .  | 107        |
| 7.4      | Results . . . . .  | 110        |
| 7.4.1    | Results for the $SR_1^{\geq 6j}$ signal region . . . . .   | 110        |
| 7.4.2    | Modeling in validation regions . . . . .   | 113        |
| 7.4.3    | Comparison to other methods . . . . .  | 113        |
| 7.5      | System reconstruction with the matrix element method . . . . .   | 116        |
| 7.5.1    | Assignment efficiency . . . . .  | 116        |
| 7.5.2    | Object reconstruction . . . . .  | 118        |
| <b>8</b> | <b>Observation of Yukawa interactions with third generation quarks . . . . .</b>                           | <b>120</b> |
| 8.1      | Evidence for $t\bar{t}H$ . . . . .   | 120        |
| 8.1.1    | Analyses entering the combination . . . . .  | 120        |
| 8.1.2    | Results . . . . .  | 121        |
| 8.2      | Observation of $t\bar{t}H$ . . . . .   | 122        |
| 8.2.1    | Analyses entering the combination . . . . .  | 122        |
| 8.2.2    | Results . . . . .  | 122        |
| 8.2.3    | Top quark Yukawa coupling . . . . .  | 124        |
| 8.3      | Observation of $H \rightarrow b\bar{b}$ . . . . .  | 124        |
| <b>9</b> | <b>Search for <math>t\bar{t}H(b\bar{b})</math> with <math>139 \text{ fb}^{-1}</math> of data . . . . .</b> | <b>126</b> |
| 9.1      | Event selection . . . . .  | 126        |
| 9.1.1    | Dataset . . . . .  | 126        |
| 9.1.2    | Object definitions . . . . .   | 127        |
| 9.1.3    | Definition of the single-lepton channel . . . . .  | 127        |
| 9.2      | Modeling . . . . .   | 127        |
| 9.2.1    | $t\bar{t}H$ signal . . . . .   | 127        |
| 9.2.2    | $t\bar{t}$ + jets background . . . . .   | 127        |
| 9.2.3    | Other backgrounds . . . . .  | 128        |
| 9.2.4    | Inclusive modeling of data . . . . .   | 128        |
| 9.3      | Event categorization . . . . .   | 128        |

---

|           |   |            |
|-----------|---|------------|
| 9.3.1     | Region definitions . . . . .                                      | 129        |
| 9.3.2     | Region composition and signal contributions . . . . .             | 129        |
| 9.4       | Multivariate analysis techniques . . . . .                        | 131        |
| 9.5       | Systematic uncertainties . . . . .                                | 131        |
| 9.5.1     | Experimental uncertainties . . . . .                              | 132        |
| 9.5.2     | Signal and background modeling . . . . .                          | 133        |
| 9.5.3     | Summary of systematic uncertainty sources . . . . .               | 134        |
| 9.6       | Statistical analysis and results . . . . .                        | 134        |
| 9.6.1     | Expected sensitivity . . . . .                                    | 136        |
| 9.6.2     | Dominant nuisance parameters and sources of uncertainty . . . . . | 141        |
| <b>10</b> | <b>Muon trigger efficiency measurement . . . . .</b>              | <b>144</b> |
| 10.1      | Analysis method . . . . .   | 144        |
| 10.2      | Event selection and categorization . . . . .                      | 145        |
| 10.2.1    | Dataset . . . . .   | 145        |
| 10.2.2    | Object definitions . . . . .                                      | 145        |
| 10.2.3    | Definition of the $t\bar{t}$ and $W$ +jets channels . . . . .     | 146        |
| 10.3      | Modeling . . . . .  | 146        |
| 10.3.1    | Comparison with data . . . . .                                    | 146        |
| 10.4      | Systematic uncertainties . . . . .                                | 147        |
| 10.5      | Results . . . . .   | 149        |
| 10.5.1    | Trigger efficiencies . . . . .                                    | 149        |
| 10.5.2    | Scale factors and impact of systematic uncertainties . . . . .    | 152        |
| <b>11</b> | <b>Differential cross-section approximation . . . . .</b>         | <b>155</b> |
| 11.1      | Overview . . . . .  | 155        |
| 11.1.1    | Fully differential cross-sections . . . . .                       | 156        |
| 11.1.2    | Challenges . . . . .  | 156        |
| 11.2      | Final state with two particles . . . . .                          | 157        |
| 11.2.1    | Approximation method . . . . .                                    | 157        |
| 11.2.2    | Implementation . . . . .  | 158        |
| 11.2.3    | Results . . . . .   | 158        |
| 11.3      | Final state with six particles . . . . .                          | 158        |
| 11.3.1    | General considerations . . . . .                                  | 160        |
| 11.3.2    | Implementation . . . . .  | 161        |
| 11.3.3    | Results . . . . .   | 162        |
| 11.4      | Discussion . . . . .  | 165        |
| <b>12</b> | <b>Conclusions and outlook . . . . .</b>                          | <b>167</b> |
|           | <b>Bibliography . . . . .</b>                                     | <b>169</b> |

---

|  |            |
|--|------------|
| <b>Appendices</b> . . . . .  | <b>181</b> |
| A Additional material related to the $t\bar{t}H(b\bar{b})$ analysis with $36.1 \text{ fb}^{-1}$ . . . . .  | 181        |
| A.1 Categorization for the dilepton channel . . . . .  | 181        |
| A.2 Signal region modeling for the dilepton channel . . . . .  | 181        |
| A.3 Correlation between nuisance parameters . . . . .  | 183        |
| B Additional material related to the $t\bar{t}H(b\bar{b})$ analysis with $139.0 \text{ fb}^{-1}$ . . . . . | 186        |
| B.1 Correlation between nuisance parameters . . . . .  | 186        |
| C Additional material related to differential cross-section approximation . . . . .                        | 187        |
| C.1 Differential cross-section for a $2 \rightarrow 2$ process . . . . .                                   | 187        |

# List of tables

|           |  |    |
|-----------|--|----|
| Table 2.1 | Fermion fields with associated charges. $Q$ is the electric charge, $T_L^3$ the weak isospin, and $Y$ the weak hypercharge. . . . .  | 9  |
| Table 2.2 | Branching ratios for the decay of $t\bar{t}$ [10]. . . . .   | 17 |
| Table 2.3 | Branching ratios for the decay of the Higgs boson. The <i>other</i> category contains experimentally challenging final states [34]. . . . .  | 20 |
| Table 4.1 | Operating points of the MV2c10 algorithm, with corresponding $b$ -jet identification efficiencies and rejection factors for $c$ -jets and light jets [75]. . . . .   | 46 |
| Table 5.1 | Probabilities $P$ for Gaussian distributed observable $x$ to fall within $n$ standard deviations of the mean $\mu$ in one experiment, and average amount of experiment repetitions needed for one observation to fall outside of this range. . . . .   | 49 |
| Table 6.1 | Definition of the $t\bar{t}$ + jets components used in the analysis. Additional particle jets are those not originating from a top quark or $W$ boson decay. . . . .   | 65 |
| Table 6.2 | Systematic uncertainty sources affecting the modeling of $t\bar{t}$ + jets. The left column shows the individual sources. Additional details regarding the sources are given in the central column. The column on the right lists on which $t\bar{t}$ components the sources act on, and whether the effect is correlated between the components. Additional details are provided in section 6.6.3 [1]. . . . .  | 80 |
| Table 6.3 | List of the systematic uncertainties affecting the analysis. The type $N$ indicates uncertainties changing normalization of the affected process, uncertainties with type $S + N$ can change both shape and normalization. The amount of different components per source is listed in the third column [1]. . . . .  | 82 |
| Table 6.4 | Expected signal strength measurement in fits to an Asimov dataset. The $\text{SR}_1^{\geq 6j}$ plays an important role in the overall sensitivity of the analysis. . . . .   | 84 |
| Table 6.5 | Contributions to the signal strength uncertainty, grouped by sources. The total statistical uncertainty includes effects from the $k(t\bar{t} + \geq 1b)$ and $k(t\bar{t} + \geq 1c)$ normalization factors, while the intrinsic statistical uncertainty does not. The background model statistical uncertainty includes effects from statistical uncertainties in nominal MC samples and the data-driven fake lepton estimate in the single-lepton channel [1]. . . . . | 94 |

|            |  |     |
|------------|--|-----|
| Table 8.1  | Summary of the signal strength $\mu_{t\bar{t}H}$ and the observed and expected significance measured in the individual analyses used to establish evidence for the $t\bar{t}H$ process, as well as the combination of all analyses. No events are observed in the analysis targeting $H \rightarrow ZZ^* \rightarrow 4l$ , hence the 68% confidence level upper limit is reported for the signal strength [2]. . . . . | 121 |
| Table 8.2  | Observed and expected significance for $H \rightarrow b\bar{b}$ decays. The results are reported separately per production mode, and for the statistical combination of all channels [38]. . . . .   | 125 |
| Table 9.1  | Systematic uncertainty sources affecting the modeling of $t\bar{t}$ + jets. The left column shows the individual sources, while the central column describes how the effect is evaluated. The column on the right lists on which $t\bar{t}$ components the sources act on, and whether the effect is correlated between the components. . . . .  | 134 |
| Table 9.2  | List of the systematic uncertainties affecting the analysis. The type $N$ indicates uncertainties changing normalization of the affected process, uncertainties with type $S + N$ can change both shape and normalization. The amount of different components per source is listed in the third column. . . . .  | 135 |
| Table 9.3  | Contributions to the signal strength uncertainty, grouped by sources. The total statistical uncertainty includes effects from the $k(t\bar{t}+ \geq 1b)$ and $k(t\bar{t}+ \geq 1c)$ normalization factors, while the intrinsic statistical uncertainty does not. . . . .   | 143 |
| Table 10.1 | Summary of trigger efficiencies for data taken between 2016 and 2018, and for simulation. The results are reported separately for the barrel and end-cap regions, and split by channel. Only absolute statistical uncertainties are included. . . . .  | 152 |
| Table 10.2 | Summary of SFs for the years 2016–2018. The results are reported separately for the barrel and end-cap regions, and split by channel. The total absolute uncertainties for the SFs are shown, and their split into statistical and systematic components is also included. . . . .   | 153 |
| Table 10.3 | Average size of relative systematic uncertainty on the SF per source, calculated as the arithmetic mean over the three years 2016–2018 of data-taking. The uncertainties are reported separately per detector region and channel in the measurement. . . . .   | 154 |
| Table 11.1 | Mean relative error obtained when only considering events above a threshold of their normalized differential cross-section. The requirement is listed in the left column, while the central column specifies the fraction of the full test dataset satisfying this requirement. The mean relative error obtained in the dataset with the requirement applied is listed in the right column. . . . .                    | 165 |

# List of figures

|             |   |    |
|-------------|---|----|
| Figure 2.1  | The particle content of the SM, adapted from reference [11] and reference [12], using information from reference [10]. . . . .  | 4  |
| Figure 2.2  | QCD interactions: coupling of quarks and gluons (left), three- and four-point gluon self-interactions (center and right, respectively). The interaction strength is parameterized by the coupling constant $g_s$ . . . . .  | 5  |
| Figure 2.3  | Momentum distributions for partons within a proton for $Q = 100$ GeV, using data from the CT10 PDF set. The first two generations of quarks, including contributions for valence quarks $u_V$ and $d_V$ , as well as gluons are shown. Uncertainties are not drawn. . . . .   | 7  |
| Figure 2.4  | Higgs potential, visualized as a function of the complex field $\phi^0$ . The unbroken case with $\mu^2 > 0$ is shown on the left, with a ground state $\varphi_1 = \varphi_2 = 0$ . The potential for $\mu^2 < 0$ is shown on the right, where the circle of global minima is drawn with a dashed red line. Rotational symmetry is spontaneously broken when a ground state along this circle is chosen. . . . . | 11 |
| Figure 2.5  | Three-point interactions between Higgs boson and $W^+ W^-$ bosons (left), between Higgs boson and two $Z$ bosons (right). . . . .   | 12 |
| Figure 2.6  | Higgs boson coupling to fermions, the coupling strength is proportional to the fermion mass $m_f$ . . . . .   | 13 |
| Figure 2.7  | Summary of ATLAS measurements of total cross-sections of various SM processes, compared to predictions of the SM [25]. . . . .  | 14 |
| Figure 2.8  | Loop corrections to the Higgs boson mass via three-point couplings to fermions $f$ , vector bosons $V$ , self-interactions and a new massive particle $X$ . Contributions from quartic interactions are not shown. . . . .  | 16 |
| Figure 2.9  | Exemplary Feynman diagrams for the gluon–gluon fusion (top left), vector boson fusion (top right), $VH$ (bottom left), and $t\bar{t}H$ (bottom right) processes . . . . .   | 18 |
| Figure 2.10 | Dominant processes for Higgs boson production with associated cross-sections in proton–proton collisions, shown as a function of COM energy. Bands indicate theoretical uncertainties in the cross-section calculation [34]. . . . .  | 18 |
| Figure 3.1  | The CERN accelerator complex relevant for proton–proton collisions in the LHC. Gray arrowheads indicate the proton path. BOOSTER refers to the Proton Synchrotron Booster, PS is the Proton Synchrotron, and SPS is the Super Proton Synchrotron. The figure is adapted from reference [49]. . . . .  | 23 |

|            |  |    |
|------------|--|----|
| Figure 3.2 | Distribution of mean number of interactions per bunch crossing in data recorded by the ATLAS experiment at $\sqrt{s} = 13$ TeV [51]. . . . .   | 24 |
| Figure 3.3 | The ATLAS coordinate system, with ATLAS located at the origin. The $x$ axis points towards the center of the LHC ring (indicated as dotted line), and the $y$ axis up towards the surface. Beams propagate along the $z$ axis. The azimuthal angle $\phi$ and polar angle $\theta$ are also shown for an arbitrary point $P$ . . . . . | 25 |
| Figure 3.4 | The complete ATLAS detector in cutaway view [45]. . . . .  | 26 |
| Figure 3.5 | Cutaway view of the ATLAS ID. The IBL is missing in this visualization [45]. . . . .   | 27 |
| Figure 3.6 | Cutaway view of the ATLAS calorimeter system surrounding the ID and solenoid magnet [45]. The term LAr refers to liquid argon as active material. . . . .  | 29 |
| Figure 3.7 | Cutaway view of the ATLAS MS [45]. . . . .   | 30 |
| Figure 3.8 | Schematic of one quarter of a cross-section through the ATLAS detector [53]. . . . .   | 31 |
| Figure 3.9 | Total integrated luminosity collected by the LHC (green), recorded by the ATLAS detector (yellow) and used for physics analyses (blue), shown as a function of time [51]. . . . .  | 35 |
| Figure 4.1 | Schematic of fundamental particles interacting with the ATLAS detector, adapted from reference [59]. It shows a section of the $x$ - $y$ plane. . . . .  | 36 |
| Figure 5.1 | Relation between significance $Z$ and $p$ -value. . . . .  | 52 |
| Figure 5.2 | Distribution of the test statistic $t_{\text{NP}}$ under hypotheses $H_1$ and $H_0$ , including $p$ -values calculated from an observation $t_{\text{obs}}$ indicated in the shaded areas. . . . .   | 53 |
| Figure 5.3 | Distribution of the test statistic $t_{\mu}$ , the $p$ -value can be obtained via the integral prescription in equation (5.14). . . . .  | 54 |
| Figure 5.4 | Exemplary architecture of a fully connected feedforward neural network with three inputs (drawn as blue circles), two hidden layers (with associated nodes drawn in green), and one output (drawn in purple). Information flows along the lines connecting nodes. . . . .  | 57 |
| Figure 6.1 | Exemplary Feynman diagram for the $t\bar{t}H(b\bar{b})$ topology, with one or two light charged leptons ( $l$ ) in the final state. The different columns listed for the decay products of the $W$ bosons correspond to the alternative topologies considered in the analysis. . . . .   | 60 |
| Figure 6.2 | Exemplary Feynman diagram for the $t\bar{t} + b\bar{b}$ background process. . . . .  | 61 |
| Figure 6.3 | Relative fraction of $t\bar{t} + \geq 1b$ sub-components predicted by the POWHEG+PYTHIA 8 and SHERPA4F samples. The uncertainty for both predictions are also shown, including the sources discussed in section 6.6.3 [1]. . . . .   | 66 |

|             |   |    |
|-------------|---|----|
| Figure 6.4  | Expected distribution of the number of jets per event in the single-lepton channel, compared to data. The uncertainties shown include all sources of systematic uncertainty described in section 6.6, with the exception of the free-floating normalization factors for the $t\bar{t}+ \geq 1b$ and $t\bar{t}+ \geq 1c$ processes. The $t\bar{t}H$ distribution normalized to the total background is overlaid as a dashed red line. . . . .  | 68 |
| Figure 6.5  | Expected distribution of the number of $b$ -tagged jets per event at the four operating points ( <i>very tight</i> , <i>tight</i> , <i>medium</i> , <i>loose</i> ) in the single-lepton channel, compared to data. The uncertainties shown include all sources of systematic uncertainty described in section 6.6, with the exception of the free-floating normalization factors for the $t\bar{t}+ \geq 1b$ and $t\bar{t}+ \geq 1c$ processes. The $t\bar{t}H$ signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram. . . . . | 69 |
| Figure 6.6  | Definition of resolved analysis regions with exactly five jets in the single-lepton channel. The vertical axis shows the $b$ -tagging requirements for the first two jets in each event, while the horizontal axis shows the requirement for the third and fourth jet. Jets are ordered by decreasing tightness of the operating point they satisfy [1]. . . . .  | 71 |
| Figure 6.7  | Definition of resolved analysis regions with six or more jets in the single-lepton channel. The vertical axis shows the $b$ -tagging requirements for the first two jets in each event, while the horizontal axis shows the requirement for the third and fourth jet. Jets are ordered by decreasing tightness of the operating point they satisfy [1]. . . . .   | 71 |
| Figure 6.8  | Composition of background processes in the single-lepton regions. Each pie chart shows the relative contributions per process and region, with the processes defined in section 6.3 [1]. . . . .  | 72 |
| Figure 6.9  | Signal contributions per analysis region in the single-lepton channel, evaluated using the expected amount of $t\bar{t}H$ events ( $S$ ) and background events ( $B$ ) per region. The solid black line, corresponding to the left vertical axis, shows $S/B$ . The dashed red line, corresponding to the right vertical axis, shows $S/\sqrt{B}$ [1]. . . . .  | 73 |
| Figure 6.10 | Measurement of the signal strength $\mu_{t\bar{t}H}$ when fitting the model to data. The two- $\mu$ fit is performed by fitting both dilepton and single-lepton channels, with two separate signal strength parameters affecting them. The nominal fit result, listed in the last row, is obtained by using a single signal strength parameter [1]. .   | 85 |

- Figure 6.11 Overview of the yields in all single-lepton regions pre-fit (top) and post-fit (bottom). The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram. It is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit [1]. . . . . 87
- Figure 6.12 Overview of the yields in all dilepton regions pre-fit (top) and post-fit (bottom). The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram. It is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit [1]. . . . . 88
- Figure 6.13 Comparison between data and the model for the control regions  $CR_{t\bar{t}+ \geq 1c}^{5j}$  (top) and  $CR_{t\bar{t}+ \geq 1c}^{\geq 6j}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. Events with  $H_T^{\text{had}} < 200$  GeV or  $H_T^{\text{had}} > 650$  GeV are included in the leftmost and rightmost bins of the  $CR_{t\bar{t}+ \geq 1c}^{5j}$  distributions, respectively. Similarly, events with  $H_T^{\text{had}} < 200$  GeV or  $H_T^{\text{had}} > 1000$  GeV are also included in the outermost bins of the  $CR_{t\bar{t}+ \geq 1c}^{\geq 6j}$  distributions [1]. . . . . 89
- Figure 6.14 Comparison between data and the model for the signal regions  $SR_1^{5j}$  (top),  $SR_2^{5j}$  (middle) and  $SR^{\text{boosted}}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line [1]. . . . . 90

|             |  |     |
|-------------|--|-----|
| Figure 6.15 | Comparison between data and the model for the signal regions $\text{SR}_1^{\geq 6j}$ (top), $\text{SR}_2^{\geq 6j}$ (middle) and $\text{SR}_3^{\geq 6j}$ (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to $k(t\bar{t}+ \geq 1b)$ and $k(t\bar{t}+ \geq 1c)$ is included pre-fit. The $t\bar{t}H$ signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. The $t\bar{t}H$ distribution normalized to the total background is overlaid as a dashed red line [1]. . . . . | 91  |
| Figure 6.16 | The 20 dominant nuisance parameters in the fit, ranked according to their impact on the signal strength. The empty rectangles correspond to the pre-fit impact, while the filled rectangles show the post-fit impact per nuisance parameter. The upper axis shows the impact $\Delta\mu$ . The pull $\frac{\hat{\theta}-\theta_0}{\Delta\theta}$ of the nuisance parameter is shown as black points, with the vertical black lines visualizing the post-fit nuisance parameter uncertainty $\Delta\hat{\theta}$ [1]. . . . .   | 92  |
| Figure 6.17 | 95% confidence level (CL) upper limits on the signal strength $\mu_{t\bar{t}H}$ , derived in a combined fit to single-lepton and dilepton channels with two independent signal strength parameters (two- $\mu$ fit), as well as a fit with a single signal strength (combined fit) [1]. . . . .  | 95  |
| Figure 6.18 | Post-fit yields of signal ( $S$ ), total background ( $B$ ), and observed data, shown as a function of $\log_{10}(S/B)$ . Contributions of the signal, when scaled to its best-fit signal strength value, are drawn in red, while contributions with the signal strength set to its value excluded at the 95% confidence level are drawn in orange. The lower panel shows the difference of observed data and various fit models to the total background taken from the nominal fit [1]. . . . .   | 96  |
| Figure 7.1  | Topologies for $t\bar{t}H$ production. The three diagrams on the left are initiated by gluons and are considered in the MEM calculation, while the quark–antiquark topology in the diagram on the right is neglected. . . . .  | 107 |
| Figure 7.2  | Transfer function components for $b$ - and light jets, shown as a function of the energy $E^q$ of the quark a jet is associated to. The distributions for light jets, described by a double Gaussian as $W_{\text{light}}$ , are shown as solid lines for different jet energies. The corresponding distributions $W_b$ for $b$ -jets are shown as dashed lines, and are described by crystal ball functions. . . . .  | 109 |
| Figure 7.3  | Distribution of the $t\bar{t}H(b\bar{b})$ signal and $t\bar{t}+b\bar{b}$ background processes in the $\text{SR}_1^{\geq 6j}$ region as a function of the logarithms of the signal and background likelihoods, $L_S$ and $L_B$ . Both processes are normalized to have unit integral. The left- and rightmost bins of the distributions include all events with likelihoods smaller or larger than the edge of these bins, respectively. . . . .  | 111 |

Figure 7.4 Distribution of the  $t\bar{t}H(b\bar{b})$  signal and  $t\bar{t} + b\bar{b}$  background processes in the  $SR_1^{\geq 6j}$  region, both normalized to have unit integral. The left figure shows the distributions as a function of the  $MEM_{D1}$  likelihood ratio, while the right figure shows the transformed version of this variable. . . . . 111

Figure 7.5 Comparison between data and the model for the logarithm of signal likelihood (top), logarithm of background likelihood (middle) and the transformed  $MEM_{D1}$  discriminant (bottom). The figures on the left show the pre-fit model. The post-fit model on the right is obtained from the fit described in section 6.7.2. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. The left- and rightmost bins of the likelihood distributions include all events with likelihoods smaller or larger than the edge of these bins, respectively. . . . . 112

Figure 7.6 Comparison between data and the model for the logarithm of signal likelihood (top), logarithm of background likelihood (middle) and the transformed  $MEM_{D1}$  discriminant (bottom) in the  $SR_2^{\geq 6j}$  region (left) and the  $SR_3^{\geq 6j}$  region (right). The uncertainty bands only include sources related to  $t\bar{t}H$  and  $t\bar{t}$  modeling, with the exception of the  $t\bar{t}+ \geq 1b$  sub-component uncertainty derived from SHERPA4F. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. The left- and rightmost bins of the likelihood distributions include all events with likelihoods smaller or larger than the edge of these bins, respectively. 114

Figure 7.7 Comparison between data and the model for the logarithm of signal likelihood (top left), logarithm of background likelihood (top right) and the transformed  $MEM_{D1}$  discriminant (bottom) in the  $CR_{t\bar{t}+ \geq 1c}^{\geq 6j}$  region. The uncertainty bands only include sources related to  $t\bar{t}H$  and  $t\bar{t}$  modeling, with the exception of the  $t\bar{t}+ \geq 1b$  sub-component uncertainty derived from SHERPA4F. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. The left- and rightmost bins of the likelihood distributions include all events with likelihoods smaller or larger than the edge of these bins, respectively. . . . . 115

|            |   |     |
|------------|---|-----|
| Figure 7.8 | Assignment efficiency of jets to quarks in the permutation with the largest $t\bar{t}H$ likelihood, evaluated with a sample of $t\bar{t}H$ events. The rows correspond to the quark each jet is matched to, while the columns describe the true jet origin. Jets may be truth-matched to multiple quarks. . . . .   | 117 |
| Figure 7.9 | Reconstructed invariant mass of the $b\bar{b}$ system produced in association with the top quark pair. The figure on the left shows the invariant mass of the two jets assigned to the $b$ quarks from the Higgs boson decay in the permutation with the highest $t\bar{t}H$ likelihood, this quantity is interpreted as the reconstructed Higgs boson mass. The figure on the right shows the $b\bar{b}$ system assigned to $b$ quarks that do not originate from top quark decays in the $t\bar{t} + b\bar{b}$ topology, using the permutation with the highest $t\bar{t} + b\bar{b}$ likelihood. Distributions of the $t\bar{t}H$ signal are shown as dashed red lines, the $t\bar{t}$ background is drawn as a solid blue line. All distributions are normalized to unit integral. Only statistical uncertainties are visualized in the figure. . . . . | 119 |
| Figure 8.1 | Results of the $t\bar{t}H$ cross-section measurement, divided by the SM prediction, in the statistical combination. The results per analysis topology are obtained from a fit with four independent cross-section parameters. Statistical and systematic uncertainties are shown in yellow and blue, respectively. The SM prediction is shown in red, with the associated uncertainty indicated as a gray band. No events are observed in the $H \rightarrow ZZ^* \rightarrow 4l$ analysis, and the 68% confidence level upper limit is reported [3]. . . . .   | 123 |
| Figure 9.1 | Expected distribution of the number of jets per event, compared to data. The uncertainties shown include all sources of systematic uncertainty described in section 9.5, with the exception of the free-floating normalization factors for the $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ processes. The $t\bar{t}H$ distribution normalized to the total background is overlaid as a dashed red line. . . . .   | 129 |
| Figure 9.2 | Expected distribution of the number of $b$ -tagged jets per event at the four operating points ( <i>very tight</i> , <i>tight</i> , <i>medium</i> , <i>loose</i> ), compared to data. The uncertainties shown include all sources of systematic uncertainty described in section 9.5, with the exception of the free-floating normalization factors for the $t\bar{t} + \geq 1b$ and $t\bar{t} + \geq 1c$ processes. The $t\bar{t}H$ signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram. Data is not shown in bins where the $t\bar{t}H$ signal is expected to contribute more than 5% to the yield, indicated by a gray hashed area. . . . .   | 130 |
| Figure 9.3 | Composition of background processes per region. Each pie chart shows the relative contributions per process, with the processes defined in section 9.2. . . . .   | 131 |

- Figure 9.4 Signal contributions per region, calculated with the expected amount of  $t\bar{t}H$  events ( $S$ ) and background events ( $B$ ) per region. The histograms show  $S/\sqrt{B}$ , with blue bars for control regions and red bars indicating signal regions.  $S/B$  is also listed for each region. . . . . 132
- Figure 9.5 Overview of the yields in all regions pre-fit (top) and post-fit (bottom). The uncertainty bands include all sources of systematic uncertainty described in section 9.5. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram. It is normalized to the SM prediction. Data is only compared to the pre-fit model, and not shown in bins where the  $t\bar{t}H$  signal is expected to contribute more than 5% to the yield, indicated by a gray hashed area. . . . . 137
- Figure 9.6 Comparison between data and the model for the control regions  $CR^{5j}$  (top) and  $CR^{\geq 6j}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 9.5. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. Events with  $H_T^{\text{had}} < 200$  GeV or  $H_T^{\text{had}} > 800$  GeV are included in the leftmost and rightmost bins of the  $CR^{5j}$  distributions, respectively. Similarly, events with  $H_T^{\text{had}} < 250$  GeV or  $H_T^{\text{had}} > 1150$  GeV are also included in the outermost bins of the  $CR^{\geq 6j}$  distributions. Data is only compared to the pre-fit model. . . . . 138
- Figure 9.7 Comparison between data and the model for the signal regions  $SR_1^{5j}$  (top),  $SR_2^{5j}$  (middle) and  $SR^{\text{boosted}}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 9.5. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. Data is only compared to the pre-fit model, and not shown in bins where the  $t\bar{t}H$  signal is expected to contribute more than 5% to the yield, indicated by a gray hashed area. . . . . 139
- Figure 9.8 Comparison between data and the model for the signal regions  $SR_1^{\geq 6j}$  (top) and  $SR_2^{\geq 6j}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 9.5. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. Data is only compared to the pre-fit model, and not shown in bins where the  $t\bar{t}H$  signal is expected to contribute more than 5% to the yield, indicated by a gray hashed area. . . . . 140

|             |  |     |
|-------------|--|-----|
| Figure 9.9  | The 20 dominant nuisance parameters in the fit, ranked according to their impact on the signal strength. The empty rectangles correspond to the pre-fit impact, while the filled rectangles show the post-fit impact per nuisance parameter. The upper axis shows the impact $\Delta\mu$ . The pull $\frac{\hat{\theta}-\theta_0}{\Delta\theta}$ of the nuisance parameter is shown as black points, with the vertical black lines visualizing the post-fit nuisance parameter uncertainty $\Delta\hat{\theta}$ . MG5 refers to samples generated with MG5_AMC@NLO+PYTHIA 8. . . . .   | 141 |
| Figure 10.1 | Expected distribution of the muon transverse momentum (left) and $E_T^{\text{miss}}$ (right) in the $t\bar{t}$ channel, compared to data. An overall normalization factor is applied to simulation to match data, with an effect smaller than 1%. Only statistical uncertainties are shown for the expected distribution, drawn with dashed lines. The $E_T^{\text{miss}} > 200$ GeV requirement is not applied in the figures showing the $E_T^{\text{miss}}$ distributions. . . . .  | 147 |
| Figure 10.2 | Expected distribution of the muon transverse momentum (left) and $E_T^{\text{miss}}$ (right) in the $W$ +jets channel, compared to data. An overall normalization factor is applied to simulation to match data, with an effect around 10%. Only statistical uncertainties are shown for the expected distribution, drawn with dashed lines. The $E_T^{\text{miss}} > 200$ GeV requirement is not applied in the figures showing the $E_T^{\text{miss}}$ distributions. . . . .  | 148 |
| Figure 10.3 | Muon trigger efficiencies and SFs in the barrel region, measured in the $t\bar{t}$ (left) and $W$ +jets (right) channels, for data recorded in 2016 (top), 2017 (middle), and 2018 (bottom). The upper part of the figures show the trigger efficiencies for data in black, and simulation as a hashed green area. The lower part shows the SF, given by the ratio of efficiency measured in data to simulation. The efficiencies are shown as a function of the reconstructed muon transverse momentum, and the resulting efficiencies and SFs from a fit to muons with $p_T > 100$ GeV are also listed in the figure. Only statistical uncertainties are included. . . . .   | 150 |
| Figure 10.4 | Muon trigger efficiencies and SFs in the end-cap regions, measured in the $t\bar{t}$ (left) and $W$ +jets (right) channels, for data recorded in 2016 (top), 2017 (middle), and 2018 (bottom). The upper part of the figures show the trigger efficiencies for data in black, and simulation as a hashed green area. The lower part shows the SF, given by the ratio of efficiency measured in data to simulation. The efficiencies are shown as a function of the reconstructed muon transverse momentum, and the resulting efficiencies and SFs from a fit to muons with $p_T > 100$ GeV are also listed in the figure. Only statistical uncertainties are included. . . . . | 151 |
| Figure 11.1 | Exemplary Feynman diagram for gluon-initiated $t\bar{t}$ production. . . . .   | 157 |

|             |  |     |
|-------------|--|-----|
| Figure 11.2 | Prediction for the differential cross-section of $t\bar{t}$ production by a foam with 50 000 cells, drawn as dashed blue lines. The distributions are shown as a function of the four variables used to parameterize the fully differential cross-section, and compared to a reference set of events generated with MG5_AMC@NLO. The reference distribution is shown in green, statistical uncertainties are indicated with hashed gray lines. . . . .   | 159 |
| Figure 11.3 | Exemplary Feynman diagram for gluon-initiated $t\bar{t}$ production, with subsequent decays into a final state with six particles. . . . .   | 160 |
| Figure 11.4 | Distribution of one million test events as a function of the logarithm of their normalized differential cross-section, $\log_{10} (1/\sigma_{t\bar{t}} d\sigma_{t\bar{t}})$ , and the corresponding neural network prediction $\log_{10} (1/\sigma_{t\bar{t}} d\hat{\sigma}_{t\bar{t}})$ . The fraction of events per bin is indicated by the color of each bin, drawn with a logarithmic scale. The diagonal gray line indicates where all events would be located for $d\hat{\sigma}_{t\bar{t}} = d\sigma_{t\bar{t}}$ . The test events are distributed like the differential cross-section; fewer events exist in phase space regions with small differential cross-sections. The neural network prediction for a very small fraction of events with small differential cross-sections significantly underestimates their differential cross-sections. The error is largest for such events due to the training set distribution and loss function choice for the neural network. . . . . | 163 |
| Figure 11.5 | Distribution of one million test events. The green histogram shows the distribution of events as a function of the logarithm of their normalized differential cross-section, $\log_{10} (1/\sigma_{t\bar{t}} d\sigma_{t\bar{t}})$ . The dashed blue line shows the distribution as a function of the corresponding network prediction, $\log_{10} (1/\sigma_{t\bar{t}} d\hat{\sigma}_{t\bar{t}})$ . The network prediction significantly underestimates the differential cross-section for a small fraction of events with small differential cross-sections. This is due to the choice of training set distribution and loss function of the neural network. . . . .  | 164 |
| Figure 11.6 | Distribution of the relative error $(d\sigma_{t\bar{t}} - d\hat{\sigma}_{t\bar{t}})/d\sigma_{t\bar{t}}$ for one million test events. The mean relative error is the average absolute value of the relative error across all events. . . . .  | 164 |
| Figure A.1  | Definition of analysis regions with exactly three jets in the dilepton channel. The vertical axis shows the $b$ -tagging requirements for the first two jets in each event, while the horizontal axis shows the requirement for the third jet. Jets are ordered by decreasing tightness of the operating point they satisfy [1]. . . . .   | 181 |
| Figure A.2  | Definition of analysis regions with four or more jets in the dilepton channel. The vertical axis shows the $b$ -tagging requirements for the first two jets in each event, while the horizontal axis shows the requirement for the third and fourth jet. Jets are ordered by decreasing tightness of the operating point they satisfy [1]. . . . .   | 182 |
| Figure A.3  | Composition of background processes in the dilepton regions. Each pie chart shows the relative contributions per process and region, with the processes defined in section 6.3 [1]. . . . .  | 182 |

|            |   |     |
|------------|---|-----|
| Figure A.4 | Signal contributions per analysis region in the single-lepton channel. The solid black line, corresponding to the left vertical axis, shows $S/B$ . The dashed red line, corresponding to the right vertical axis, shows $S/\sqrt{B}$ . $S$ is the number of $t\bar{t}H$ events per region, and $B$ the number of expected background events [1]. . . . .   | 183 |
| Figure A.5 | Comparison between data and the model for the signal regions $SR_1^{\geq 4j}$ (top), $SR_2^{\geq 4j}$ (middle) and $SR_3^{\geq 4j}$ (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to $k(t\bar{t}+ \geq 1b)$ and $k(t\bar{t}+ \geq 1c)$ is included pre-fit. The $t\bar{t}H$ signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. The $t\bar{t}H$ distribution normalized to the total background is overlaid as a dashed red line [1]. . . . . | 184 |
| Figure A.6 | Correlations between the most highly ranked nuisance parameters and the signal strength, determined by the nominal fit to data described in section 6.7.2. All values are in % [1]. . . . .   | 185 |
| Figure B.1 | Correlations between nuisance parameters and signal strength, determined by a fit to an Asimov dataset. Parameters are only included if they have a correlation of at least 25% with at least one other parameter. All values are in %. . . . .   | 186 |

# Acronyms

**3F** three-flavor

**4F** four-flavor

**5F** five-flavor

**BDT** boosted decision tree

**BSM** beyond the Standard Model

**CKM** Cabibbo–Kobayashi–Maskawa

**COM** center-of-mass

**CP** charge conjugation and parity symmetry

**CPU** central processing unit

**CSC** cathod strip chamber

**DGLAP** Dokshitzer–Gribov–Lipatov–Altarelli–Parisi

**DM** dark matter

**EW** electroweak

**EWSB** electroweak symmetry breaking

**FSR** final state radiation

**GPU** graphics processing unit

**HL-LHC** High Luminosity Large Hadron Collider

**HLT** high-level trigger

**IBL** insertable B-layer

**ID** inner detector

- ISR** initial state radiation
- L1** Level-1
- LHC** Large Hadron Collider
- LHD** likelihood discriminant
- LO** leading order
- MC** Monte Carlo
- MDT** monitored drift tube
- MEM** matrix element method
- MPI** multi-parton interaction
- MS** muon spectrometer
- NLO** next-to-leading order
- NNLL** next-to-next-to-leading logarithmic
- NNLO** next-to-next-to-leading order
- PDF** parton distribution function
- PMNS** Pontecorvo–Maki–Nakagawa–Sakata
- PS** parton shower
- QCD** quantum chromodynamics
- QED** quantum electrodynamics
- ReLU** rectified linear unit
- RoI** region of interest
- RPC** resistive plate chamber
- SCT** semiconductor tracker
- SF** scale factor
- SM** Standard Model of particle physics

**TGC** thin gap chamber

**TMVA** Toolkit for Multivariate Data Analysis with ROOT

**TRT** transition radiation tracker

**UE** underlying event

**VEV** vacuum expectation value

# Acknowledgements

This dissertation would not have been possible without many people who I had the pleasure to meet and work with during the last five years. Thank you for making this an amazing experience.

To my supervisor Oliver, thank you for your continuous guidance, for always being available to help, for your advice, and for providing me the freedom to focus on projects I am interested in. I have learned a lot from you, and I am happy to have had you as my supervisor. Thank you to Tancredi for sharing your deep understanding of ATLAS and particle physics. Your experience, insight and advice were invaluable.

Thank you to the members of the AMVA4NewPhysics network, which shaped the second half of my PhD. You taught me lots about machine learning, and encouraged me to follow new ideas. It was a great experience to have been part of this network. Thank you to Tommaso for all your work to make it happen, and the groups at CP3 and EPFL for hosting me.

To the  $t\bar{t}H(b\bar{b})$  group, thank you for providing such a great environment to work in at all times. I am especially grateful to Jelena. Thank you for teaching me most of what I know about the analysis, patiently answering lots of questions, and for the continuous advice and feedback. Thank you to everyone who made the use of the matrix element method possible: Bernd for lots of advice, Dan for the transfer function derivation, and Claire for the interface to the common analysis code. Thanks to everyone else in the group, particularly Georges, Michele and Thomas. It was great to work alongside you on this analysis.

My time working towards a PhD was split between Vancouver and CERN, and it was wonderful to work in both places. Thank you to the UBC and TRIUMF groups for providing such an enjoyable atmosphere. Thanks to everyone in the UBC lab, both previous and current members, especially to Sébastien writing in parallel, and to Tal at CERN. A special thank you to Felix and Wojtek for running Flashy, which was an invaluable resource.

To everyone who read earlier drafts of this document, thank you for all the useful feedback you provided.

Last but not least, to my family and friends: thank you for all your support during the last five years. It was a privilege to have been given the chance to pursue this journey, and I thoroughly enjoyed it.



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 675440 AMVA4NewPhysics.

# 1. Introduction

The Standard Model of particle physics (SM) describes interactions between elementary particles. It is a highly successful theory of nature, and in agreement with decades of experimental tests. One of the central ingredients to the SM is the Brout-Englert-Higgs mechanism, which explains how fermions and the weak gauge bosons acquire mass. This mechanism predicts the existence of a Higgs boson, which was observed in 2012 by both the ATLAS and CMS collaborations at the Large Hadron Collider (LHC). The Higgs boson observation marks the beginning of a new era of experimental particle physics, which studies the details of electroweak symmetry breaking (EWSB).

Despite all its success, the SM has several shortcomings and does not constitute a complete theory of nature. A rich experimental program is dedicated to the search for phenomena predicted by theories that extend the SM. An alternative way forward are precise measurements of the theoretical predictions. Detailed studies of the interactions between the Higgs boson and other elementary particles are necessary to further establish the SM validity. Such studies are possible at the LHC. A possible deviation from the SM predictions can help guide the way towards a more complete theory.

This dissertation presents a measurement of the interaction between Higgs bosons and top quarks. The strength of the Higgs boson interaction with other elementary particles increases with their masses. The top quark, being the heaviest elementary particle in the SM, is the fermion that interacts most strongly with the Higgs boson. This interaction is studied directly in a process called  $t\bar{t}H$ , using proton–proton collisions at the LHC that were recorded by the ATLAS experiment. The  $t\bar{t}H$  process describes the production of a pair of top quarks and a Higgs boson. A specific topology is the focus of this dissertation, where the Higgs boson decays to a pair of bottom quarks. This process is called  $t\bar{t}H(b\bar{b})$ . A measurement of the  $t\bar{t}H$  process can be interpreted in terms of the top quark Yukawa coupling, which has important implications in both particle physics and cosmology.

The dissertation is organized as follows. Chapter 2 introduces the SM, with a focus on the process of EWSB and implications for the experimental analyses performed. The LHC and the ATLAS experiment are described in chapter 3. This is followed by a description of how the ATLAS detector reconstructs physics objects, provided in chapter 4. Chapter 5 introduces the statistical methodology used to interpret results, and includes the description of multivariate analysis techniques employed. A search for the  $t\bar{t}H(b\bar{b})$  process is presented in detail in chapter 6. It uses  $36.1 \text{ fb}^{-1}$  of data collected by the ATLAS experiment in 2015 and 2016. The  $t\bar{t}H(b\bar{b})$  search uses a multivariate analysis technique called the matrix element method (MEM), and its implementation for the search is described in detail in chapter 7. Chapter 8 presents the statistical combination of the  $t\bar{t}H(b\bar{b})$  search with other measurements performed with the ATLAS detector, leading to the observation of both the  $t\bar{t}H$  process and the Higgs boson decay to bottom quark pairs by the ATLAS collaboration in 2018. Between

2015 and 2018, the ATLAS experiment collected a proton–proton collision dataset of  $139.0 \text{ fb}^{-1}$ . The expected sensitivity of a  $t\bar{t}H(b\bar{b})$  analysis with this dataset is studied in chapter 9. Chapter 10 presents a measurement of the efficiency of the ATLAS muon trigger system. This system is essential to the success of many physics analyses studying events where muons are produced, including the  $t\bar{t}H(b\bar{b})$  process. The studies summarized in chapter 11 investigate the feasibility of using machine learning techniques to help reduce the large computational requirements of MEM calculations. A conclusion to the dissertation is provided in chapter 12. The appendices A and B provide additional material related to the  $t\bar{t}H(b\bar{b})$  searches with  $36.1 \text{ fb}^{-1}$  and  $139.0 \text{ fb}^{-1}$  respectively, and appendix C includes additional material related to chapter 11.

This dissertation uses natural units, where the speed of light in vacuum  $c$  and the reduced Planck constant  $\hbar$  are both set to unity, and electric charge is given in units of the electric charge of the positron.

## 2. The Standard Model of particle physics

The SM [4–7] describes interactions between fundamental particles via three out of the four known fundamental forces: electromagnetic, weak, and strong force. The strong force is governed by the theory of quantum chromodynamics (QCD), while the electromagnetic and weak forces are unified in the electroweak (EW) theory. Gravitational interactions are not accounted for by the SM, and described by the theory of general relativity.

The SM is a non-abelian gauge theory, invariant under the gauge group

$$G = SU(3)_C \times SU(2)_L \times U(1)_Y. \quad (2.1)$$

Its Lagrangian density is given by

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}}, \quad (2.2)$$

with the two components described in sections 2.2 and 2.3.

The SM is a Lorentz invariant quantum field theory, reconciling the laws of quantum mechanics and special relativity. This summary is based on reference [8], with some input from reference [9] and experimental measurements summarized in reference [10]. The Einstein summation convention is used in this chapter, repeated indices are summed over.

The chapter starts with an overview of the particle content in the SM in section 2.1, followed by brief descriptions of QCD in section 2.2 and the EW theory in section 2.3. Section 2.4 highlights the success and limitations of the SM. The chapter closes with predictions of the SM for physics at the LHC relevant to this dissertation in section 2.5.

### 2.1 Particles in the Standard Model

The particle content of the SM is shown in figure 2.1. Matter is made up of fermions, which are particles with half integer spin. Particles with integer spin are called bosons.

The fundamental fermions can be split into two classes, quarks and leptons. Quarks carry a fractional electric charge  $Q$  and a color charge. There are six flavors of quarks, split into three different generations. Each generation contains an up-type and a down-type quark, where the types are named after the quarks in the first generation: *up* and *down*. The corresponding quarks of the second and third generations are called *charm* and *strange*, and *top* and *bottom*, respectively. Up-type quarks carry electric charge  $Q = 2/3$ , while down-type quarks have  $Q = -1/3$ . Besides these six quarks, there are six antiquarks with the same masses and opposite charges. Quarks can only be observed in bound states, called hadrons, due to color confinement. Hadrons consisting of three quarks are called baryons, and they are also fermions. Quark–antiquark pairs form mesons, which are bosons. The second

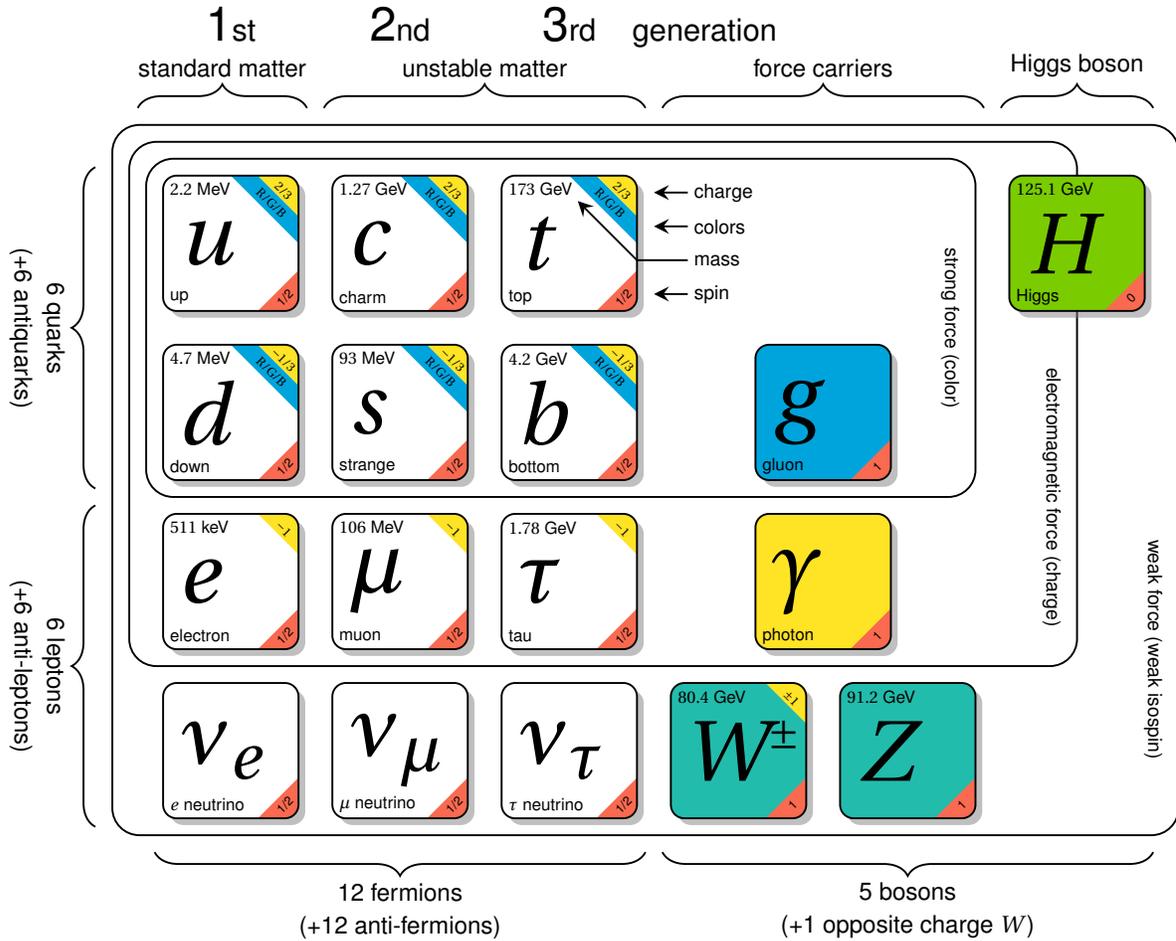


Figure 2.1: The particle content of the SM, adapted from reference [11] and reference [12], using information from reference [10].

class of fundamental fermions are the leptons, again split into three generations: electrons, muons, and tau leptons. Each generation contains a charged lepton with electric charge  $Q = -1$  and the corresponding neutrino of the same flavor.

There are five different types of elementary bosons in the SM. Four types of vector bosons with spin-1 act as gauge bosons, mediating forces between fermions. The scalar spin-0 Higgs boson is a consequence of EWSB, described in section 2.3.2.

Throughout this document, the term electron is generally used to refer to both of the charged first generation leptons. The term positron is only used when a distinction is necessary. Similarly, the distinction between quarks and antiquarks of a specific flavor is only made when explicitly required.

## 2.2 Quantum chromodynamics

Strong interactions are described by the non-abelian  $SU(3)$  gauge theory QCD. Two types of fields are contained in this theory, quark fields  $q_{r\alpha}$ , and gauge fields  $G^i$ , called the gluons. The flavor index  $r$

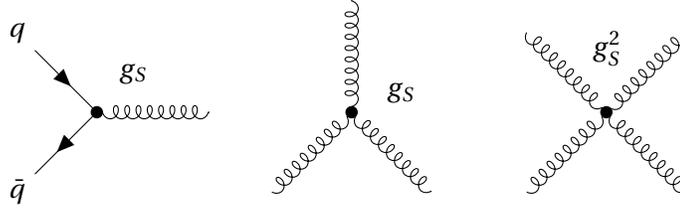


Figure 2.2: QCD interactions: coupling of quarks and gluons (left), three- and four-point gluon self-interactions (center and right, respectively). The interaction strength is parameterized by the coupling constant  $g_s$ .

takes values  $r = u, d, c, s, t, b$ , the color index is  $\alpha = \text{red, green, blue}$ . There are eight Hermitian fields  $G^i = G^{i\dagger}$ ,  $i = 1, \dots, 8$ , corresponding to the eight generators of  $SU(3)$ . The gauge covariant derivative, introduced to keep the quark kinetic terms gauge invariant under local  $SU(3)$  transformations, is given by

$$D_\alpha^{\mu\beta} = \partial^\mu \delta_\alpha^\beta + \frac{ig_s}{\sqrt{2}} G_\alpha^{\mu\beta}. \quad (2.3)$$

It couples quarks to the gluon field, with an interaction strength parameterized by  $g_s$ . The expression  $\delta_\alpha^\beta$  is the Kronecker delta, it is unity for  $\alpha = \beta$  and zero otherwise. The gluon field in matrix notation is

$$G_\alpha^\beta = \sum_{i=1}^8 G^i \frac{\lambda_{\alpha\beta}^i}{\sqrt{2}}, \quad (2.4)$$

where  $\alpha, \beta$  are matrix indices. The Gell-Mann matrices  $\lambda^i$  form a representation of the Lie algebra of  $SU(3)$  and satisfy  $[\lambda^i, \lambda^j] = 2i f_{ijk} \lambda^k$ , defining the antisymmetric structure constants  $f_{ijk}$ . Their normalization is chosen such that  $\text{Tr}[\lambda^i \lambda^j] = 2\delta^{ij}$ . The gluon field strength tensor is

$$G_{\mu\nu}^i = \partial_\mu G_\nu^i - \partial_\nu G_\mu^i - g_s f_{ijk} G_\mu^j G_\nu^k. \quad (2.5)$$

Putting everything together, QCD is described by the Lagrangian density

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G_{\mu\nu}^i G^{i\mu\nu} + \sum_r \bar{q}_r^\alpha i \gamma_\mu D_\alpha^{\mu\beta} q_{r\beta}, \quad (2.6)$$

where  $\gamma_\mu$  are the Dirac matrices. The first term leads to three- and four-point gluon self-interactions, while the quark kinetic term results in quark–gluon coupling. These interactions are depicted in figure 2.2.

Additional quark mass terms  $-\sum_r m_r \bar{q}_r^\alpha q_{r\alpha}$  could be added to the Lagrangian in pure QCD. In the context of the SM, they are instead generated during EWSB, as described in section 2.3. The Lagrangian 2.6 also allows for the addition of a charge conjugation and parity symmetry (CP) violating term, scaled by a free parameter  $\theta_{\text{QCD}}$ . It is experimentally known to be small.

### 2.2.1 Running coupling

Calculations in QCD involve contributions from processes happening at different orders of  $g_s$ . The lowest order of coupling at which a process can take place is called leading order (LO), the next highest

order next-to-leading order (NLO), followed by next-to-next-to-leading order (NNLO). Contributions from higher order processes can be partially absorbed into an effective coupling  $\alpha_s(\mu_R^2) = g_s^2(\mu_R^2)/4\pi$  with renormalization scale  $\mu_R$ . This scale is commonly chosen to be a characteristic energy scale for a given process, such as the momentum transfer  $Q$  in a collision. Quark loop corrections to the gluon propagator decrease the effective coupling  $\alpha_s(\mu_R^2)$  at long distances (equivalent to small energy scale  $\mu_R$ ), while gluon loops increase it. The effect from the gluon loops is dominant. In result, the strong force becomes weaker at high  $\mu_R$ , this phenomenon is called asymptotic freedom. At low energies, QCD is non-perturbative.

The dependence of the effective coupling on the energy scale is expressed by the renormalization group equation. At one loop level, the effective coupling scales like  $\alpha_s(Q^2) \propto \ln(Q^2/\Lambda^2)^{-1}$ , where  $\Lambda$  is a reference energy scale at which the strong coupling becomes large.

### 2.2.2 Parton distribution functions

The parton model describes the constituents of hadrons as point-like particles. Parton distribution functions (PDFs) describe the probability density (as defined in section 5.1) for a specific parton to carry a momentum fraction  $x$  of a proton. These distributions, expressed as  $f(x, Q^2)$ , depend on the momentum transfer  $Q^2$ . The evolution of the PDFs with energy scale  $Q^2$  is given by the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations [13–15].

Figure 2.3 shows an example of the momentum distributions  $xf(x, Q^2)$  of partons in protons. The data is taken from the CT10 PDF set [16] interfaced via LHAPDF [17], with  $Q = 100$  GeV. Protons contain two valence up and one down quark, which carry significant momentum fractions as visible in the figure. The contributions from sea quarks decreases at higher  $x$ .

### 2.2.3 Parton-parton scattering

The cross-sections for processes at hadron colliders can be factorized into two contributions. PDFs describe the colliding partons  $i, j$  within the colliding hadrons  $H_1, H_2$ . The subsequent hard scattering of the partons, described by a cross-section  $\sigma_{ij \rightarrow F}$ , can typically be calculated perturbatively. The total cross-section for producing a final state  $F$ , along with any unobserved  $X$ , from colliding hadrons  $H_1, H_2$  is given by the factorization theorem [18], and can be written as

$$\sigma_{H_1, H_2 \rightarrow F+X} = \sum_{i,j} \int f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \sigma_{ij \rightarrow F} dx_1 dx_2. \quad (2.7)$$

The energy scale  $\mu_F$  is called factorization scale, typically chosen to correspond to a momentum transfer or characteristic momentum of the process. It may also affect the hard scatter cross-section  $\sigma_{ij \rightarrow F}$  through radiation from the initial partons  $i, j$ .

## 2.3 Electroweak theory

The EW theory provides the unified description of electrodynamic and weak interactions. It is a non-abelian gauge theory, based on the  $SU(2)_L \times U(1)_Y$  gauge group. In contrast to QCD, it is chiral

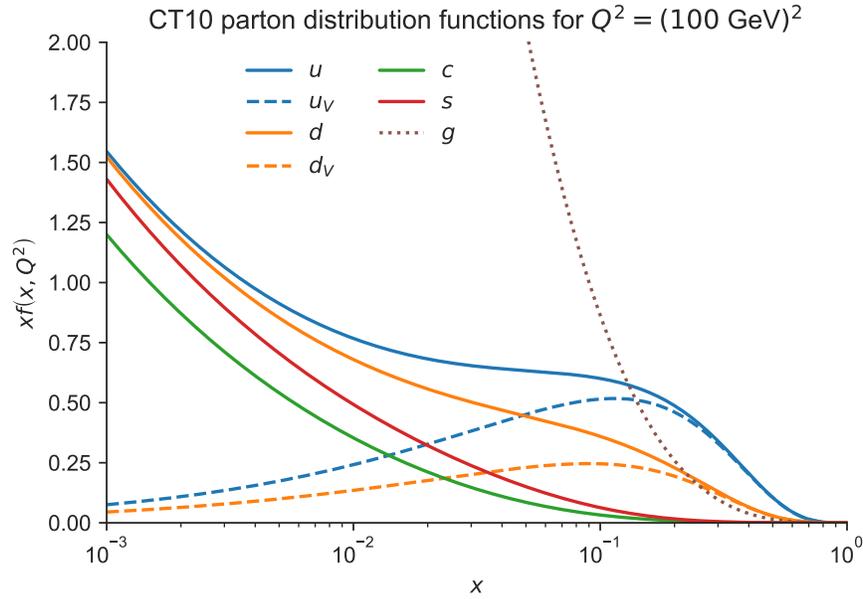


Figure 2.3: Momentum distributions for partons within a proton for  $Q = 100 \text{ GeV}$ , using data from the CT10 PDF set. The first two generations of quarks, including contributions for valence quarks  $u_V$  and  $d_V$ , as well as gluons are shown. Uncertainties are not drawn.

and spontaneously broken via the process of EWSB. An unbroken  $U(1)_Q$  symmetry remains after EWSB, which describes quantum electrodynamics (QED).

The  $SU(2)_L$  part acts on the flavor of left-chiral fermions, indicated by the subscript  $L$ . It introduces three gauge bosons  $W^i$ ,  $i = 1, 2, 3$  with associated coupling  $g$ . The  $U(1)_Y$  group describes a gauge boson  $B$  with coupling  $g'$ , and different interactions with left- and right-chiral fermions. Its subscript  $Y$  refers to the weak hypercharge, a conserved quantity under  $U(1)_Y$  transformations. It is defined as

$$Y = Q - T_L^3, \quad (2.8)$$

where  $Q$  is the electric charge after EWSB, and  $T_L^3$  the weak isospin, which is conserved in weak interactions.

After EWSB, a  $U(1)_Q$  symmetry remains. It describes the photon, a linear combination of  $W^3$  and  $B$ . The massive  $W^\pm$  bosons interact with left-chiral fermions via charged current interactions and can change their flavor. Massive  $Z$  bosons mediate neutral current interactions, and couple to both left- and right-chiral fermions. Flavor changing neutral currents are highly suppressed.

### 2.3.1 Lagrangian density

The EW Lagrangian density is given by

$$\mathcal{L}_{\text{EW}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_f + \mathcal{L}_\phi + \mathcal{L}_{\text{Yuk}}. \quad (2.9)$$

The first term,  $\mathcal{L}_{\text{gauge}}$ , contains the gauge boson kinetic terms and self-interactions. Covariant fermion kinetic terms are described by  $\mathcal{L}_f$ . The Higgs sector behavior is captured by  $\mathcal{L}_\phi$ . Yukawa interactions, which generate fermion masses, are contained in the last term,  $\mathcal{L}_{\text{Yuk}}$ . All terms are described individually in this section.

### Gauge term

The field strength tensors for the  $SU(2)_L$  and  $U(1)_Y$  gauge bosons are given by

$$\begin{aligned} W_{\mu\nu}^i &= \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g\epsilon_{ijk}W_\mu^jW_\nu^k, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu. \end{aligned} \quad (2.10)$$

The definition of  $W_{\mu\nu}^i$  is similar to the  $SU(3)_C$  QCD gluon tensor in equation (2.5), but now there are three tensors  $W_{\mu\nu}^i$ ,  $i = 1, 2, 3$ , corresponding to the three generators of  $SU(2)$ . The Levi-Civita tensor  $\epsilon_{ijk}$  with  $i, j, k = 1, 2, 3$  is totally antisymmetric, with  $\epsilon_{123} = 1$ .

The gauge term in the Lagrangian density is

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}W_{\mu\nu}^iW^{i\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \quad (2.11)$$

The kinetic term for the  $W^i$  bosons introduces three- and four-point gauge boson self interactions, while the  $B$  gauge boson does not self-interact.

### Fermion sector

The fermion sector is split into left-chiral  $SU(2)_L$  doublets and right-chiral singlets. For each fermion field  $f_{mL}^0 = q_{mL}^0, \ell_{mL}^0$  and  $f_{mR}^0 = u_{mR}^0, d_{mR}^0, e_{mR}^0$ , the subscripts  $L$  and  $R$  refer to the left- and right-chiral nature of the field, respectively. The subscript  $m$  refers to the generation. Up-type quarks are written as  $u_m$  with  $m = 1, 2, 3$ ,  $d_m$  are the down-type quarks, and the charged leptons are  $e_m = e, \mu, \tau$ . The superscript 0 indicates that these fields are weak interaction eigenstates. The color index  $\alpha$  carried by the quarks, which transforms under QCD, is not explicitly written. A summary of the fermion fields in the SM, along with their associated charges, is provided in table 2.1.

The gauge covariant derivatives for left-chiral fermion doublets  $f_{mL}^0$  introduce couplings to the three  $SU(2)_L$  gauge bosons, written as a column vector  $\vec{W} = \begin{pmatrix} W^1 & W^2 & W^3 \end{pmatrix}^T$ , with coupling strength  $g$ , and to the  $U(1)_Y$  gauge boson  $B$  with a coupling proportional to weak hypercharge  $Y_f$  and coupling strength  $g'$ . The corresponding weak hypercharge per fermion is listed in table 2.1. The row vector  $\vec{\tau} = \begin{pmatrix} \sigma_1 & \sigma_2 & \sigma_3 \end{pmatrix}$  denotes the Pauli matrices, which satisfy  $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$ . The covariant derivatives for left-chiral fermion doublets  $f_{mL}^0$  are

$$D_\mu f_{mL}^0 = \left( \partial_\mu + \frac{ig}{2}\vec{\tau} \cdot \vec{W}_\mu + ig'Y_f B_\mu \right) f_{mL}^0. \quad (2.12)$$

The gauge covariant derivatives for right-chiral fermions  $f_{mR}^0$  only couple them to the  $B$  gauge boson:

$$D_\mu f_{mR}^0 = (\partial_\mu + ig'Y_f B_\mu) f_{mR}^0. \quad (2.13)$$

Table 2.1: Fermion fields with associated charges.  $Q$  is the electric charge,  $T_L^3$  the weak isospin, and  $Y$  the weak hypercharge.

| fields   | $Q$             | $T_L^3$         | $Y$    |
|--|-----------------|-----------------|--------|
| $q_{mL}^0 = \begin{pmatrix} u_m^0 \\ d_m^0 \end{pmatrix}_L$      | $2/3$<br>$-1/3$ | $1/2$<br>$-1/2$ | $1/6$  |
| $u_{mR}^0$   | $2/3$           | $0$             | $2/3$  |
| $d_{mR}^0$   | $-1/3$          | $0$             | $-1/3$ |
| $\ell_{mL}^0 = \begin{pmatrix} \nu_m^0 \\ e_m^0 \end{pmatrix}_L$ | $0$<br>$-1$     | $1/2$<br>$-1/2$ | $-1/2$ |
| $e_{mR}^0$   | $-1$            | $0$             | $-1$   |

The fermion part of the EW Lagrangian density contains the kinetic terms for all fermion fields for all three generations:

$$\begin{aligned} \mathcal{L}_f = \sum_{m=1}^3 & (\bar{q}_{mL}^0 i\gamma^\mu D_\mu q_{mL}^0 + \bar{u}_{mR}^0 i\gamma^\mu D_\mu u_{mR}^0 + \bar{d}_{mR}^0 i\gamma^\mu D_\mu d_{mR}^0 \\ & + \bar{\ell}_{mL}^0 i\gamma^\mu D_\mu \ell_{mL}^0 + \bar{e}_{mR}^0 i\gamma^\mu D_\mu e_{mR}^0). \end{aligned} \quad (2.14)$$

### Higgs sector

The Higgs sector introduces two complex scalar Higgs fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4 \end{pmatrix}, \quad (2.15)$$

transforming as a doublet under  $SU(2)_L$  with weak hypercharge  $Y_H = 1/2$ . The gauge covariant derivative acting on this doublet is given by

$$D_\mu \phi = \left( \partial_\mu + \frac{ig}{2} \vec{\tau} \cdot \vec{W}_\mu + ig' Y_H B_\mu \right) \phi, \quad (2.16)$$

consistent with the behavior on left-chiral fermion doublets. This covariant derivative introduces three- and four-point interactions between the gauge bosons  $W^i, B$  and the Higgs field in  $\mathcal{L}_\phi$ .

The Lagrangian density of the Higgs sector is given by

$$\mathcal{L}_\phi = (D_\mu \phi)^\dagger D^\mu \phi - V(\phi) \quad (2.17)$$

with a potential term

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (2.18)$$

The first term determines the Higgs boson mass after EWSB, and the second term describes quartic Higgs-field self-interactions. A requirement of  $\lambda > 0$  guarantees that there is a lower bound on the potential.

### Yukawa term

The last term in the EW Lagrangian density describes Yukawa interactions. It has the form

$$\mathcal{L}_{\text{Yuk}} = - \sum_{m,n=1}^3 \left[ \Gamma_{mn}^u \bar{q}_{mL}^0 \tilde{\phi} u_{nR}^0 + \Gamma_{mn}^d \bar{q}_{mL}^0 \phi d_{nR}^0 + \Gamma_{mn}^e \bar{\ell}_{mL}^0 \phi e_{nR}^0 \right] + h.c., \quad (2.19)$$

with a Hermitian conjugate term given by  $h.c.$ . The field  $\tilde{\phi}$  is a conjugate of the Higgs doublet  $\phi$ , defined as  $\tilde{\phi} = i\sigma_2 \phi^\dagger = \begin{pmatrix} \phi^{0\dagger} & -\phi^- \end{pmatrix}^T$ . The  $3 \times 3$  matrices  $\Gamma^u, \Gamma^d, \Gamma^e$  determine the Yukawa couplings and fermion masses after EWSB.

### 2.3.2 Electroweak symmetry breaking

During EWSB, the scalar  $\phi^0$  field acquires a vacuum expectation value and the  $SU(2)_L \times U(1)_Y$  gauge group is broken down to  $U(1)_Q$ . In the SM, this takes place via the Brout-Englert-Higgs mechanism [19–24]. This mechanism generates masses for the gauge bosons  $W^\pm$  and  $Z$ , and fermion masses via Yukawa couplings. Out of the four degrees of freedom in the complex doublet  $\phi$ , three get assigned to the  $W^\pm$  and  $Z$ , and the last degree of freedom is the scalar Higgs boson.

The vacuum expectation value (VEV) of the scalar doublet  $\phi$  is given by its lowest energy state  $\langle 0|\phi|0\rangle$ , which minimizes the potential defined in equation (2.18). For  $\mu^2 > 0$ , this minimum is located at

$$\phi^\dagger \phi = 0, \quad (2.20)$$

where all real scalar fields  $\varphi_i$  have zero VEV, since  $\phi^\dagger \phi = \frac{1}{2} \sum_{i=1}^4 \varphi_i^2$ . This case is visualized in figure 2.4 on the left, projected onto the complex field  $\phi^0$ . There is a single stable global minimum at the origin, preserving the  $SU(2)_L \times U(1)_Y$  symmetry.

When  $\mu^2 < 0$ , a hypersphere of degenerate stable global minima is located at

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} > 0, \quad (2.21)$$

which defines  $v \equiv \sqrt{-\mu^2/\lambda}$ . The choice of a specific ground state breaks the degeneracy, spontaneously breaking rotational symmetry. Figure 2.4 shows this configuration on the right, again projected onto the complex field  $\phi^0$ . The degenerate minima are located along a circle, drawn as a dotted red line. A suitable axis rotation aligns the real component of  $\phi^0$  with the ground state chosen, such that  $\langle 0|\varphi_3|0\rangle = v$  and  $\langle 0|\varphi_1|0\rangle = \langle 0|\varphi_2|0\rangle = \langle 0|\varphi_4|0\rangle = 0$ . The ground state in these coordinates is given by

$$\langle 0|\phi|0\rangle \equiv v = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad (2.22)$$

with  $v^\dagger v = v^2/2$ . Only the neutral field component  $\phi^0$  receives a non-zero VEV, resulting in a  $U(1)_Q$  symmetry after EWSB. Written in unitary gauge, the expansion around this minimum is given by

$$\phi \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \quad (2.23)$$

with a Hermitian scalar field  $H$  representing the Higgs boson.

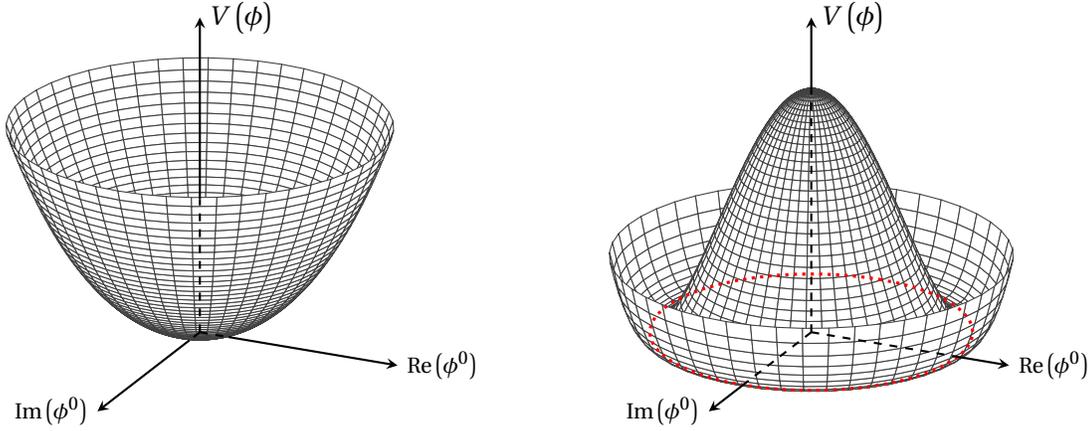


Figure 2.4: Higgs potential, visualized as a function of the complex field  $\phi^0$ . The unbroken case with  $\mu^2 > 0$  is shown on the left, with a ground state  $\varphi_1 = \varphi_2 = 0$ . The potential for  $\mu^2 < 0$  is shown on the right, where the circle of global minima is drawn with a dashed red line. Rotational symmetry is spontaneously broken when a ground state along this circle is chosen.

### 2.3.3 Lagrangian density after electroweak symmetry breaking

The Lagrangian density in equation (2.9) can be rewritten after EWSB in unitary gauge to study the structure of the SM.

#### Gauge and Higgs sectors

It is convenient to define linear combinations of the  $SU(2)_L \times U(1)_Y$  gauge bosons as

$$W_\mu^\pm = \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}}, \quad (2.24)$$

and

$$Z_\mu = -\sin\theta_W B_\mu + \cos\theta_W W_\mu^3, \quad (2.25)$$

where  $\sin\theta_W = g'/g_Z$  and  $\cos\theta_W = g/g_Z$  with  $g_Z = \sqrt{g^2 + g'^2}$ . This defines the weak angle  $\theta_W \equiv \arctan g'/g$ . The remaining linear combination of the  $B_\mu$  and  $W_\mu^3$  gauge bosons is the photon, given by

$$A_\mu = \cos\theta_W B_\mu + \sin\theta_W W_\mu^3. \quad (2.26)$$

The Lagrangian density for the Higgs sector after EWSB is obtained by inserting the expression (2.23) into equation (2.17), using the definitions from equation (2.24) and equation (2.25). It is given by

$$\mathcal{L}_\phi = \frac{g^2 v^2}{4} W^{+\mu} W_\mu^- \left(1 + \frac{H}{v}\right)^2 + \frac{1}{2} \frac{g_Z^2 v^2}{4} Z^\mu Z_\mu \left(1 + \frac{H}{v}\right)^2 + \frac{1}{2} (\partial^\mu H) (\partial_\mu H) - V(H). \quad (2.27)$$

This Lagrangian density describes two charged gauge bosons  $W^\pm$  and a neutral boson  $Z$ , with tree-level masses

$$M_W = \frac{gv}{2}, \quad M_Z = \frac{g_Z v}{2}, \quad (2.28)$$

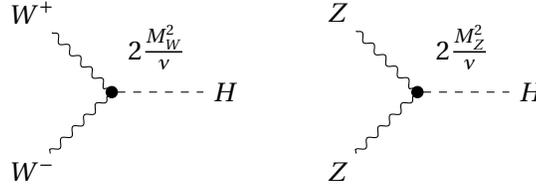


Figure 2.5: Three-point interactions between Higgs boson and  $W^+W^-$  bosons (left), between Higgs boson and two  $Z$  bosons (right).

while the photon remains massless. The masses of the  $Z$  boson and the  $W^\pm$  bosons are related via  $M_W/M_Z = \cos\theta_W$ , alternatively written as  $\sin^2\theta_W = 1 - (M_W/M_Z)^2$ .

The Lagrangian density in equation (2.27) introduces three- and four-point interactions between the  $W^\pm$  and  $Z$  bosons with the Higgs boson  $H$ . The corresponding interaction strengths are proportional to the square of the respective gauge boson masses. The three-point interactions  $HW^+W^-$  and  $HZZ$  are shown in figure 2.5, with interaction strengths proportional to the squares of the  $W^\pm$  and  $Z$  boson masses. Three- and four-point Higgs boson self-interactions are included in the potential term, which is given in unitary gauge by

$$V(H) = -\frac{\mu^4}{4\lambda} - \mu^2 H^2 + \lambda v H^3 + \frac{\lambda}{4} H^4. \quad (2.29)$$

The Higgs boson tree-level mass is appearing in the second term as

$$M_H = \sqrt{-2\mu^2} = \sqrt{2\lambda}v, \quad (2.30)$$

and is related to its quartic coupling.

The gauge interactions from  $\mathcal{L}_{\text{gauge}}$  in equation (2.11) result in three- and four-point interactions of the gauge bosons after EWSB.

### Quantum electrodynamics

The theory of QED is described by the Lagrangian density

$$\mathcal{L}_{\text{QED}} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi - q A_\mu \bar{\psi} \gamma^\mu \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \quad (2.31)$$

for a fermion field  $\psi$  with mass  $m$  and electric charge  $q$ , and field strength tensor  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ . When comparing the fermion Lagrangian density  $\mathcal{L}_f$  after EWSB to  $\mathcal{L}_{\text{QED}}$ , the positron electric charge can be identified as

$$e = g \sin\theta_W. \quad (2.32)$$

It is fully determined by the  $SU(2)_L$  and  $U(1)_Y$  coupling constants.

### Fermions

Fermion interactions with the Higgs field are obtained by inserting the expression (2.23) into equation (2.19). The resulting mass matrices  $M_{mn}^u = \Gamma_{mn}^u v/\sqrt{2}$  and Yukawa couplings to the Higgs field

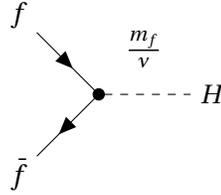


Figure 2.6: Higgs boson coupling to fermions, the coupling strength is proportional to the fermion mass  $m_f$ .

$1/v M_{mn}^u$  need to be diagonalized to obtain the fermion mass eigenstates. The Lagrangian density describing fermion mass eigenstates  $\psi_f$ , where  $f = u_m, d_m, e_m$  ( $m = 1, 2, 3$  is the fermion generation), is

$$\mathcal{L}_\psi = \sum_{f=u,d,e} \bar{\psi}_f \left[ i\gamma^\mu \partial_\mu - m_f \left( 1 + \frac{H}{v} \right) \right] \psi_f, \quad (2.33)$$

with  $\psi_f = \psi_{fL} + \psi_{fR}$ . The kinetic term for the fermions originates from equation (2.14). Fermions couple to the Higgs field with a coupling strength proportional to their masses, visualized in figure 2.6. The Yukawa coupling strength of fermions is defined as

$$y_f = \sqrt{2} \frac{m_f}{v}. \quad (2.34)$$

The relation between the fermion mass eigenstates and weak eigenstates is given by the  $3 \times 3$  unitary Cabibbo–Kobayashi–Maskawa (CKM) matrix  $V_{\text{CKM}}$ ,

$$\begin{pmatrix} d^0 \\ s^0 \\ b^0 \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \quad (2.35)$$

with  $V_{\text{CKM}}^* V_{\text{CKM}} = I_3$ , the  $3 \times 3$  identity matrix. The off-diagonal elements cause flavor changing weak charged current interactions of quarks with transition probabilities proportional to the squared magnitude of the matrix elements,  $|V_{mn}|^2$ .

## 2.4 Success and limitations of the Standard Model

The SM has 19 free parameters:

- nine fermion Yukawa couplings for the three charged leptons and six quarks,
- three coupling constants for gauge interaction strengths:  $g_s$ ,  $g$ ,  $g'$  (alternatively parameterized via the strong coupling  $\alpha_s$ , the QED coupling  $\alpha$ , and the Fermi coupling constant  $G_F$ ),
- two parameters describing the Higgs potential:  $v$ ,  $m_H$ ,
- four mixing angles of the CKM matrix,
- one strong CP phase  $\theta_{CP}$ , usually taken as zero, leading to CP violation in strong interactions.

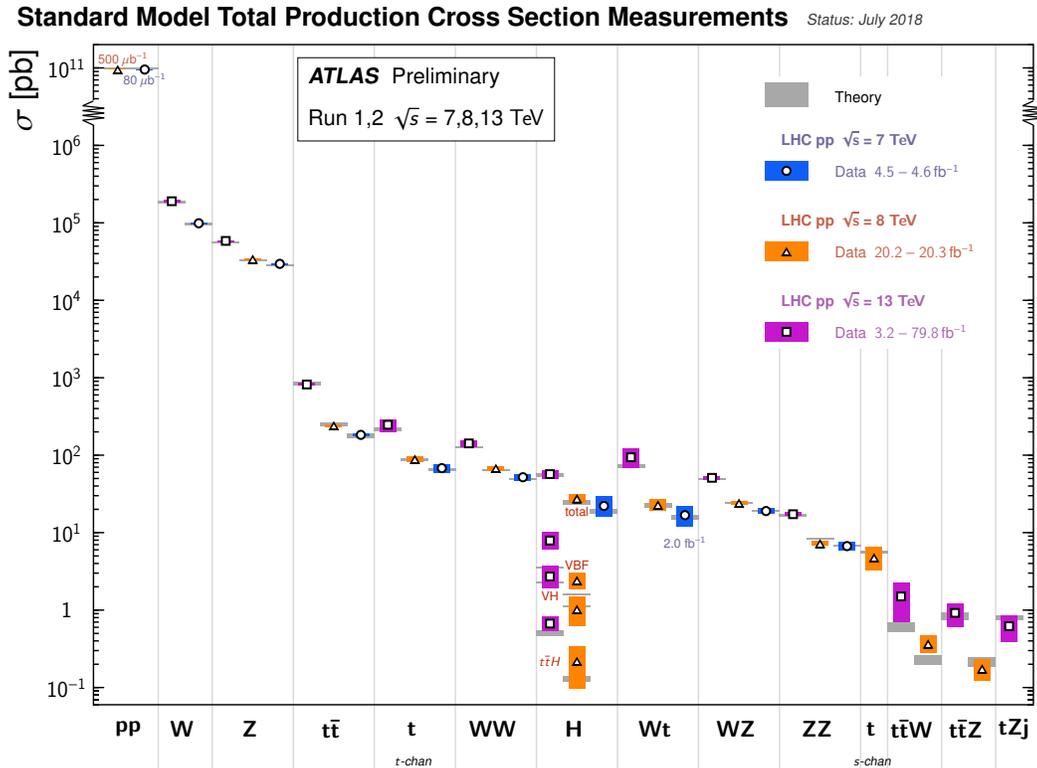


Figure 2.7: Summary of ATLAS measurements of total cross-sections of various SM processes, compared to predictions of the SM [25].

The SM can be extended with seven additional parameters to describe neutrino oscillations and mass generation via couplings to the Higgs field (for Dirac neutrinos). The Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix relates the neutrino mass eigenstates to weak eigenstates, parameterized with four parameters. On top of this, three new Yukawa couplings are needed, bringing the total amount of free parameters in the SM to 26. Alternative mechanisms for neutrino mass generation exist, and the experimental determination of the mechanism responsible for neutrino masses in nature is an active area of research.

Using just these 19 degrees of freedom, the predictions of the SM are in agreement with decades of experimental tests. It is a highly successful theory of nature. Figure 2.7 shows cross-section measurements for a range of processes, measured with the ATLAS detector. They are compared to SM predictions, obtained from calculations performed at least at NLO. The measurements are in great agreement with these predictions, for processes with cross-sections varying over many orders of magnitude.

### 2.4.1 Open questions

Despite all the success, the SM is not the final theory of particle physics. A number of outstanding issues are listed in this section [8, 9].

### Dark matter

When assuming that the mass of galaxies is mostly comprised of luminous stars, the tangential velocity  $v$  of stars with mass  $m$ , located at a radius  $r$  from the center of the galaxy, is given by

$$\frac{mv^2}{r} \approx \frac{Gm}{r^2} M(r), \quad (2.36)$$

with gravitational constant  $G$ . The mass contained within radius  $r$  is given by  $M(r)$ . For galaxies like the Milky Way, most of the luminous mass is in the center. The observed distribution  $M(r)$  decreases slower with radius than expected, hence indicating that a significant amount of non-luminous dark matter (DM) contributes gravitationally.

Gravitational lensing results from the bullet cluster [26], measuring the deflection of light due to the mass in this galaxy cluster, provide additional support for the existence of DM. Furthermore, when fitting the cosmological standard model  $\Lambda$ CDM to the cosmic microwave background power spectrum, the density of baryonic matter makes up only roughly one sixth of the total matter density in the universe [27]. Alternative explanations to these results, which propose modified versions of gravity, are increasingly challenged by the precision of the experimental observations. A detailed historical perspective is provided in reference [28].

Many theories of beyond the Standard Model (BSM) physics predict candidates for DM, and a broad range of experiments is searching for them. Once DM particles are observed and confirmed, the description of them will require an appropriately modified new SM.

### Baryon asymmetry

The baryon asymmetry in the universe describes the observed excess of baryonic matter compared to antimatter. Its origin is unclear, but the SM might require additional sources of CP violation to explain the visible universe today.

### Neutrinos

The discovery of neutrino oscillations between different flavors established their non-zero masses, and current data can be described assuming neutrino mixing between all three flavors [10]. Several open questions remain to be answered about the nature of neutrinos. Neutrinos may be Dirac or Majorana fermions, the latter implying that they are equivalent to their antiparticles and can violate lepton number conservation. A Majorana nature could be confirmed by the observation of neutrinoless double  $\beta$  decays ( $nn \rightarrow ppe^-e^-$ ). The absolute scale of neutrino masses and the mechanism by which they are generated are currently unknown. Neutrino masses may be acquired via Yukawa couplings to the Higgs field, which introduces sterile right-chiral neutrinos that do not couple to the  $W^\pm$  and  $Z$  bosons. A range of other proposed mechanisms exists.

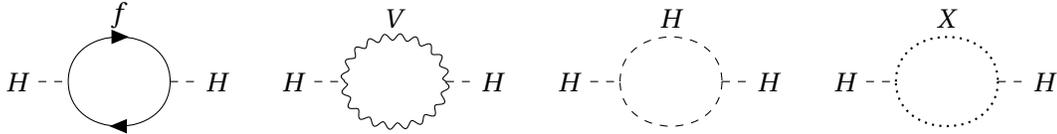


Figure 2.8: Loop corrections to the Higgs boson mass via three-point couplings to fermions  $f$ , vector bosons  $V$ , self-interactions and a new massive particle  $X$ . Contributions from quartic interactions are not shown.

### Vacuum stability and top quark Yukawa coupling

The evolution of the effective Higgs self-coupling  $\lambda$  is given by the renormalization group equation. It predicts that large values of the top quark Yukawa coupling  $y_t$  drive the self-coupling to negative values at high energies. This results in the appearance of a new EW vacuum, with a lower potential than the vacuum after EWSB, which is described in section 2.3.2. For very large values of  $y_t$ , the lifetime of the current vacuum is smaller than the lifetime of the universe. Data indicates that the current EW vacuum in the SM is meta-stable, with a lifetime larger than the lifetime of the universe [29]. A precise measurement of  $y_t$  is an important check of the SM validity in the cosmological context, and can answer the question whether any BSM phenomena are needed for consistency below the Planck scale, which is the energy scale at which gravitational effects are expected to play a significant role [30].

### Theory considerations

The SM is unable to provide explanations for *why* it takes the form described in this chapter. There are three generations of fermions, with the second and third generations playing seemingly no large role in nature. It is unclear why the fermion masses, free parameters for which the SM makes no prediction, vary over many orders of magnitude. The CP-violating strong phase  $\theta_{QCD}$  is known to be very small, with no explanation for why this is the case.

The hierarchy problem arises when embedding the SM into another, more complete, theory of nature. Such a theory is characterized by an energy scale  $\Lambda$  above the EW scale  $v \approx 246$  GeV. If the SM is valid up to this scale, then the Higgs boson mass receives loop-level corrections to the tree-level expression from equation (2.30). The contributions from three-point interactions are visualized in figure 2.8. In the SM, they are dominated by the top quark in the fermion loop. These corrections to the mass are proportional to  $\Lambda$ , the scale at which the integrals over the momenta of the particles contributing to the loops are cut off. When taking this scale to be the Planck scale  $\approx 10^{19}$  GeV, the Higgs boson mass should receive corrections that make it many orders of magnitude larger than its measured value. These corrections could be cancelled by additional couplings of the Higgs field to a new field  $X$ . The coupling strength of such interactions needs to be fine-tuned to precisely cancel out the contributions from SM couplings. An increasingly severe fine-tuning is needed for larger values of  $\Lambda$ . According to the naturalness paradigm, such large cancellations between free parameters of a theory should not occur [31].

Table 2.2: Branching ratios for the decay of  $t\bar{t}$  [10].

| Final state   | Branching ratio |
|---------------|-----------------|
| all-hadronic  | 45.7%           |
| single-lepton | 43.8%           |
| dilepton      | 10.5%           |

## 2.5 Implications for physics at the Large Hadron Collider

This section describes the physics of top quarks, Higgs bosons, and their interplay, in collisions at the LHC.

### 2.5.1 Top quark

With a mass of around 173 GeV, the top quark  $t$  is the heaviest elementary particle of the SM [10]. This allows for decays into  $W$  bosons and  $b$  quarks, which happen before it can form hadrons. Decays into other down-type quarks are in principle possible, but, due to the structure of the CKM matrix with  $|V_{tb}| \approx 1$ , extremely rare and negligible in practice. With its large mass, the top quark has a Yukawa coupling value of approximately unity.

Top quarks are predominantly produced in pairs, as a  $t\bar{t}$  system, at hadron colliders. Table 2.2 shows the branching ratios of the  $t\bar{t}$  system into three different final states. The all-hadronic final state contains decays of both  $W$  bosons into a quark–antiquark pair each. This is the dominant decay mode. A slightly smaller fraction of events decays into the single-lepton final state, where one of the  $W$  bosons decays into a quark–antiquark pair, while the other one decays into a charged lepton and a neutrino. The dilepton final state describes events where both  $W$  bosons decay leptonically. Table 2.2 includes tau leptons in the definition of leptons. In the context of physics analyses, leptonic decays often instead refer to decays to light charged leptons only. Light charged leptons are electrons and muons, including those from decays of tau leptons. Decays of tau leptons to hadronic final states are treated differently experimentally, as described in section 4.3.3.

### 2.5.2 Higgs boson

The detailed investigation of the EWSB process is one of the major physics goals of the LHC. In 2012, both the ATLAS and CMS collaborations published their independent discoveries of a new boson [32, 33], consistent with a SM Higgs boson. Subsequent studies have confirmed the SM nature of the observed particle, and are probing its properties in detail. With the central role played by the Higgs boson in the SM, detailed studies are both an important check of the validity of the SM, and can at the same time reveal possible hints of BSM physics phenomena.

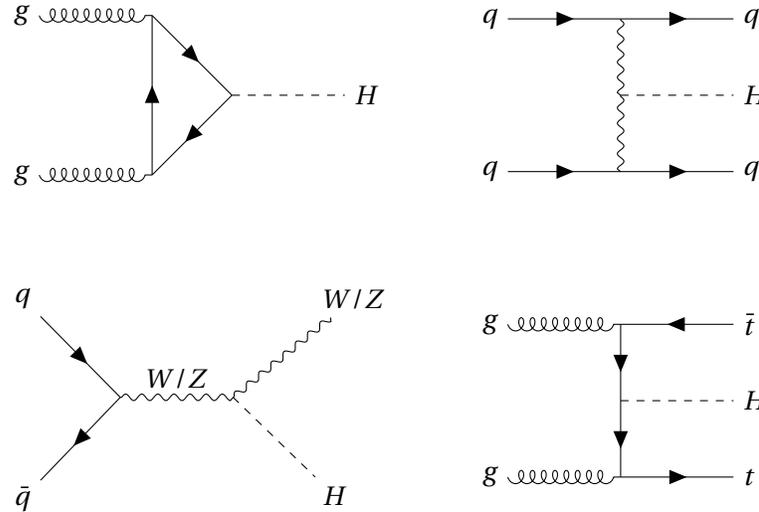


Figure 2.9: Exemplary Feynman diagrams for the gluon–gluon fusion (top left), vector boson fusion (top right),  $VH$  (bottom left), and  $t\bar{t}H$  (bottom right) processes

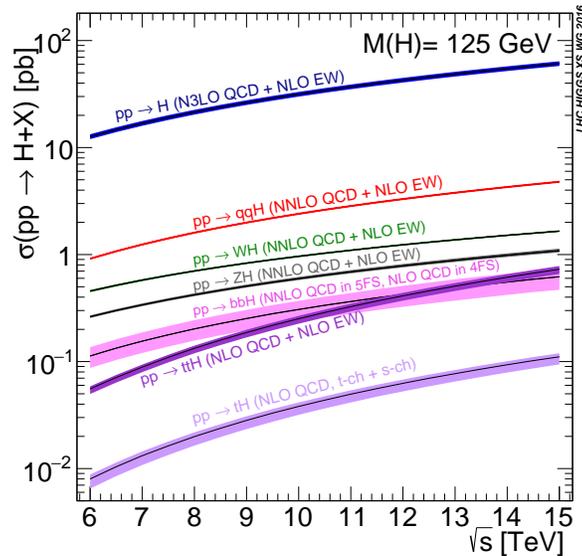


Figure 2.10: Dominant processes for Higgs boson production with associated cross-sections in proton–proton collisions, shown as a function of COM energy. Bands indicate theoretical uncertainties in the cross-section calculation [34].

### Higgs boson production

There are four major Higgs boson production modes accessible in proton–proton collisions at the LHC [34]. Figure 2.9 shows exemplary LO Feynman diagrams for these four modes, while the respective cross-section as a function of the center-of-mass (COM) energy are presented in figure 2.10 for a Higgs boson mass of 125 GeV.

The loop-induced gluon–gluon fusion is the dominant production mode. Due to its large Yukawa

coupling, the largest contributions to the loop in the SM come from virtual top quarks. The contributions from lighter quarks  $q$  are suppressed proportional to the squares of their masses  $m_q$ , by factors  $m_q^2/m_t^2$  for top quark mass  $m_t$ . While this production mode has a large cross-section, it lacks distinctive objects produced in the final state along the Higgs bosons, which would help identify events produced via this production mode experimentally.

The production mode with the second largest cross-section is vector boson fusion, sometimes also more precisely called weak boson fusion. It is initiated by the scattering of two quarks or antiquarks via exchange of a  $W^\pm$  or  $Z$  boson. The weak boson radiates off a Higgs boson. Experimentally, this production mode is characterized by two high-momentum jets (see section 4.4), emitted at small angles from the colliding protons.

Higgs boson production in association with a  $W^\pm$  or  $Z$  vector boson,  $VH$ , is the production mode with the third highest cross-section. It is characterized by the Drell-Yan production of an off-shell weak boson, which radiates off a Higgs boson. The weak boson in the final state can be used experimentally to identify events originating from this process.

The associated production of Higgs bosons with a top quark pair,  $t\bar{t}H$ , is the focus of this dissertation. In this process, the Higgs boson is radiated off a top quark pair. This top quark pair can be produced from either gluon–gluon or quark–antiquark interactions. The relevance of this process is discussed in section 2.5.3. In order to concisely specify a topology where the Higgs boson is produced via  $t\bar{t}H$  and decays to a bottom quark pair, the term  $t\bar{t}H(b\bar{b})$  will be used in this dissertation. Chapter 6 and chapter 9 describe analyses of this topology.

Additional Higgs boson production modes remain experimentally challenging. Higgs boson production in association with a bottom quark pair,  $b\bar{b}H$ , has a comparable cross-section to  $t\bar{t}H$  at a COM energy of  $\sqrt{s} = 13$  TeV. The bottom quark pair signature is not as easily identifiable experimentally as the top quark pair signature of  $t\bar{t}H$ , thus making the isolation of  $b\bar{b}H$  from background processes difficult. Higgs boson production in association with a single top quark,  $tH$ , is sensitive to the sign of the top quark Yukawa coupling and affected by destructive interference between dominant processes contributing to its production. The cross-section for this process is an order of magnitude smaller than  $t\bar{t}H$ , making the process more difficult to observe. Lastly, double Higgs boson production is sensitive to the Higgs boson cubic self-coupling. Due to its very low cross-sections in the SM, the observation of this production requires future runs of the LHC, in order to collect a large amount of integrated luminosity.

### Higgs boson decay

The Higgs boson decays into a wealth of experimentally accessible final states. Its branching ratios in the SM for a mass of 125 GeV are shown in table 2.3.

The  $b\bar{b}$  final state is most common. Decays into gauge vector bosons are suppressed, since one of the bosons must be produced off-shell. The decays into a di-photon final state  $\gamma\gamma$  are loop-induced; this loop is dominated by contributions from virtual top quarks in the SM. Higgs boson decays into  $ZZ^*$  and subsequently four charged leptons, as well as decays into  $\gamma\gamma$ , are comparatively rare. Due to

Table 2.3: Branching ratios for the decay of the Higgs boson. The *other* category contains experimentally challenging final states [34].

| Final state    | Branching ratio |
|----------------|-----------------|
| $b\bar{b}$     | 58.2%           |
| $WW^*$         | 21.4%           |
| $\tau\tau$     | 6.3%            |
| $ZZ^*$         | 2.6%            |
| $\gamma\gamma$ | 0.2%            |
| other          | 11.3%           |

their clean experimental final state signature, these decay modes nevertheless made the dominant contributions to the Higgs boson discovery.

The *other* category includes decays into pairs of gluons (with a branching ratio of 8.2%), charm quarks,  $Z\gamma$ , muons, and other processes with very small cross-sections. These are challenging to measure experimentally. While the branching ratio for the decay to a pair of muons is only 0.02%, a search performed by the ATLAS collaboration with  $79.8 \text{ fb}^{-1}$  of data excludes values larger than roughly twice this prediction at the 95% confidence level, assuming the SM production cross-section [35]. An ATLAS search for the  $Z\gamma$  final state, which has a branching ratio of 0.15%, excludes values larger than roughly seven times the SM prediction using  $36.1 \text{ fb}^{-1}$  of data [36].

### 2.5.3 Yukawa couplings and the special role of $t\bar{t}H$

The  $t\bar{t}H$  process assumes a special role in the SM, as it allows for the direct tree-level measurement of the top quark Yukawa coupling  $y_t$ . The relevance of this parameter is highlighted in section 2.4.1. When assuming that the SM is correct,  $y_t$  can be obtained in various ways. It is related to the top quark mass via equation (2.34). Furthermore, it can be measured from the gluon–gluon fusion Higgs boson production and the loop-induced Higgs boson decay to di-photon final states, since the top quark contribution to the loop dominates. The measurement from these loop-induced processes relies on the assumption that no BSM particles contribute to the loops. In a combined fit to Run-2 data of up to  $79.8 \text{ fb}^{-1}$  recorded by the ATLAS experiment, the uncertainty on  $y_t$  is around 10%, with a central value consistent with the SM prediction [37]. In this fit, the coupling strengths of the Higgs boson to weak gauge bosons, third generation quarks, tau leptons, and muons are all measured simultaneously. This fit assumes no BSM particles coupling to the Higgs bosons.

A direct measurement of  $y_t$  via  $t\bar{t}H$  is an important confirmation of the Brout-Englert-Higgs mechanism and of the SM validity. It tests a fundamental type of interaction in the SM, the Yukawa interaction between Higgs boson and fermions. The ATLAS  $t\bar{t}H(b\bar{b})$  analysis presented in chapter 6 contributed to the observation of both the  $t\bar{t}H$  process [3] and the Higgs boson decay to bottom quark pairs [38]. These processes were independently observed by the CMS collaboration [39, 40], and both CMS and ATLAS observed the Higgs boson coupling to  $\tau$  leptons in 2017 and 2018, respectively [41, 42]. These results establish the observation of Yukawa interactions. The observation was achieved by

analyzing the interactions of Higgs bosons with third generation fermions. The measurement of the interactions with the first and second generation fermions remains a challenge for the future [43].

## 3. The Large Hadron Collider and the ATLAS experiment

The LHC [44] is a hadron accelerator, delivering proton–proton collisions and also supporting collisions of protons with heavy ions or just heavy ions. The decay products of the interactions taking place are recorded by a range of experiments. The four major experiments include two general-purpose experiments, ATLAS [45] and CMS [46]. LHCb [47] specializes in physics with  $b$ -hadrons, and ALICE [48] in heavy ion collisions.

This chapter describes the experimental facilities relevant to the work presented in this dissertation. The LHC is briefly introduced in section 3.1, and relevant details about the ATLAS experiment are given in section 3.2.

### 3.1 The Large Hadron Collider

The LHC is located at the border between Switzerland and France, close to the city of Geneva [44]. It forms a part of the CERN accelerator complex, and serves as the final stage in a chain of accelerators, designed to provide collisions between proton beams at COM energies of  $\sqrt{s} = 14$  TeV with instantaneous luminosities of  $L = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . The LHC is installed in a 26.7 km long tunnel, which lies between 45 m and 170 m below the surface. Beams of charged particles are accelerated in opposite directions within two rings in a vacuum system, using electromagnetic fields in radio frequency cavities operating at 400 MHz. Superconducting magnets deflect the beams via the Lorentz force, keeping them on track within the rings. The beams are focused and collided at four points, with the four experiments ATLAS, CMS, LHCb and ALICE situated at these collision points. Besides proton–proton collisions, the LHC also supports collisions of heavy ions, as well as heavy ions and protons. As those operation modes are not relevant for this dissertation, they are not discussed further.

#### 3.1.1 Accelerator chain

Before proton beams are circulated in the LHC, a range of other accelerators gradually accelerates the beams to increasingly higher energies. The CERN accelerator complex is pictured in figure 3.1.

Hydrogen gas is used as a source of protons, with their electrons stripped off by an electric field. The protons are initially accelerated in the linear accelerator Linac2 to energies of 50 MeV. Subsequently, the Proton Synchrotron Booster accelerates them to 1.4 GeV, followed by further acceleration in the Proton Synchrotron to 25 GeV. The last step before entering the LHC is additional acceleration in the Super Proton Synchrotron to energies of 450 GeV. It takes a minimum of approximately 16

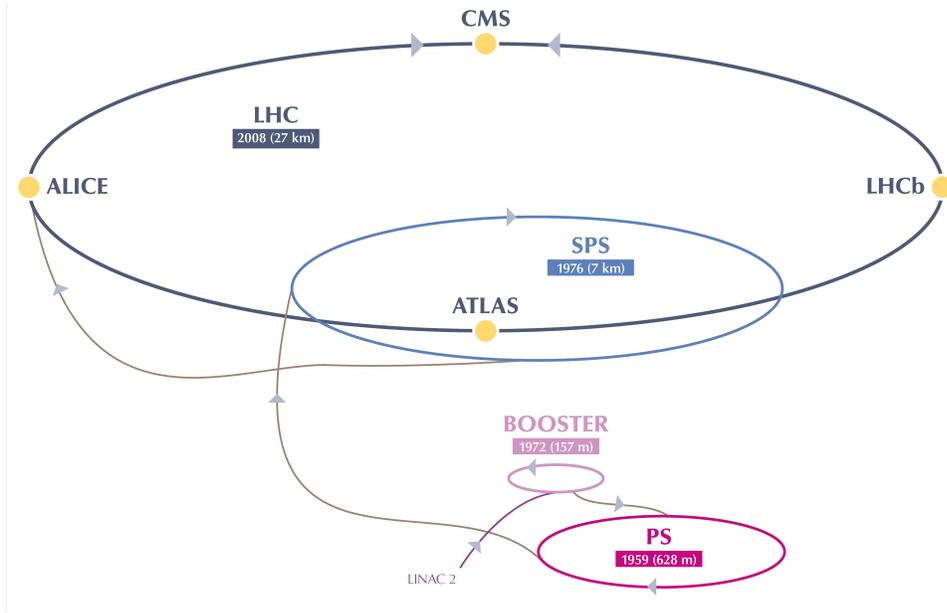


Figure 3.1: The CERN accelerator complex relevant for proton–proton collisions in the LHC. Gray arrowheads indicate the proton path. BOOSTER refers to the Proton Synchrotron Booster, PS is the Proton Synchrotron, and SPS is the Super Proton Synchrotron. The figure is adapted from reference [49].

minutes to fill the LHC with pre-accelerated proton bunches. Within around 20 minutes after the filling, the LHC accelerates the beams to their target collision energies.

### 3.1.2 Luminosity and pile-up

The number of events produced by the LHC is a function of the instantaneous luminosity  $L$  delivered by the machine over time  $t$ ,

$$N_{\text{events}} = \sigma_{\text{event}} \int L dt = \sigma_{\text{event}} \mathcal{L}, \quad (3.1)$$

and is proportional to the relevant cross-section  $\sigma_{\text{event}}$  for producing such events. The time integral over instantaneous luminosity is called the integrated luminosity  $\mathcal{L}$ . The instantaneous luminosity is given by [9]

$$L = f \frac{n_1 n_2}{4\pi \sigma_x \sigma_y} F. \quad (3.2)$$

With the nominal LHC spacing between proton bunches of 25 ns, the collision frequency is  $f = 40$  MHz. The amount of protons per bunch is  $n_1$  and  $n_2$  for the two beams, with up to  $10^{11}$  protons per bunch. Not all bunches are filled with protons in practice. The bunches have root mean square extensions  $\sigma_x$  and  $\sigma_y$  in the two directions perpendicular to the beam propagation direction. Collisions at the LHC are not exactly head-on, and the factor  $F$  contains a description of the geometric effects due to the crossing angle between the beams at the interaction point.

Due to the large amount of protons per bunch at the LHC, each bunch crossing usually results in more than one hard scattering interaction. Interactions besides the interaction of interest are

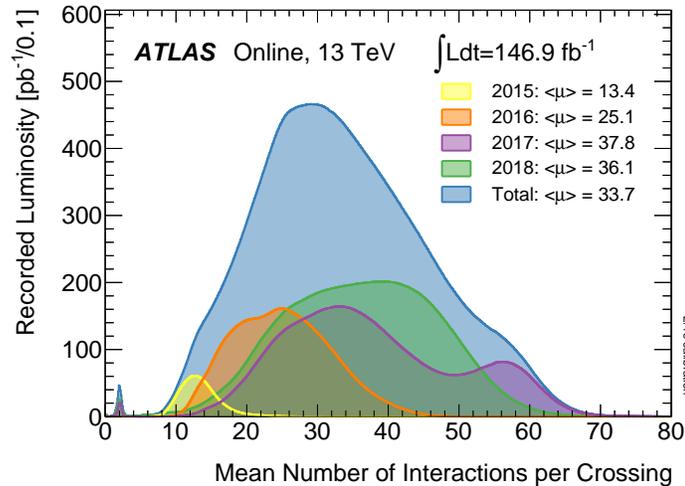


Figure 3.2: Distribution of mean number of interactions per bunch crossing in data recorded by the ATLAS experiment at  $\sqrt{s} = 13$  TeV [51].

called in-time pile-up. Out-of-time pile-up is caused by proton–proton interactions taking place in neighboring bunch crossings around the crossing of interest, which can affect the measurement, since the readout times for detector systems can be longer than the time between two bunches. The number of interactions per bunch crossing is Poisson-distributed, with a mean  $\mu$  proportional to the product of the total inelastic proton–proton cross-section  $\sigma_{\text{inel}}$  and the instantaneous luminosity [50],

$$\mu = \frac{L\sigma_{\text{inel}}}{f}. \quad (3.3)$$

### 3.1.3 Dataset

The data-taking period with LHC collisions of protons at energies of  $\sqrt{s} = 7$  TeV and  $\sqrt{s} = 8$  TeV is called Run-1. Subsequently, Run-2 lasted from 2015–2018, with collision energies of  $\sqrt{s} = 13$  TeV. The dataset collected in Run-2 is analyzed in this dissertation.

The mean number of interactions per bunch crossing is shown in figure 3.2 for the data recorded by the ATLAS detector in Run-2 of the LHC. Different colors show the distribution per data-taking year, with the 2015 contribution in yellow, 2016 in orange, 2017 in purple and 2018 in green. The blue distribution corresponds to the total dataset integrated over all four data-taking years. The average number of interactions per bunch crossing per year is shown in the legend as  $\langle\mu\rangle$ . This quantity, and hence the pile-up, increases over the course of Run-2. The increase is caused by the increased instantaneous luminosity provided by the LHC, which reaches a plateau in 2017 and 2018.

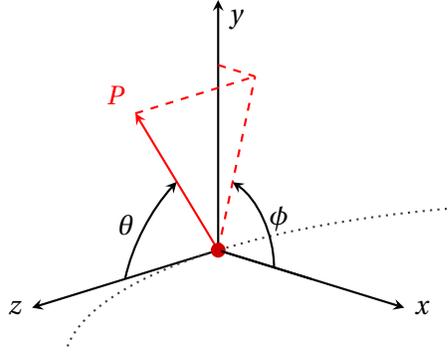


Figure 3.3: The ATLAS coordinate system, with ATLAS located at the origin. The  $x$  axis points towards the center of the LHC ring (indicated as dotted line), and the  $y$  axis up towards the surface. Beams propagate along the  $z$  axis. The azimuthal angle  $\phi$  and polar angle  $\theta$  are also shown for an arbitrary point  $P$ .

## 3.2 The ATLAS detector

The ATLAS detector is a general purpose detector at the LHC [45]. A broad range of high energy particle physics analyses are conducted with ATLAS, including measurements of SM properties and searches for hints of BSM physics. In order to meet the requirements of these analyses, the ATLAS detector is designed with fast electronics, high granularity, good object reconstruction efficiency and resolution.

### 3.2.1 Coordinate system

The right-handed coordinate system used to describe ATLAS is visualized in figure 3.3. Its origin is located at the nominal beam interaction point. The  $z$ -axis is aligned with the beam axis, while the  $x$ - $y$  plane is perpendicular to it. The  $x$ -axis points towards the center of the LHC ring, which is indicated in the figure as a dotted line. The  $y$ -axis is directed up towards the surface. Transverse quantities are defined in the  $x$ - $y$  plane. The transverse momentum  $p_T$  of an object is the momentum component in the  $x$ - $y$  plane,  $p_T = \sqrt{p_x^2 + p_y^2}$ . The azimuthal angle  $\phi$  is measured around the beam axis, while the polar angle  $\theta$  is measured from the beam axis. A common alternative parametrization of the polar angle is given by the pseudorapidity  $\eta = -\ln[\tan \theta/2]$ , which approaches infinity as the polar angle decreases to zero. Differences in pseudorapidity,  $\Delta\eta$ , are Lorentz invariant under boosts along the beam axis. The rapidity  $y = \frac{1}{2} \ln \left[ \frac{E+p_z}{E-p_z} \right]$ , typically used for massive objects, is equivalent to the pseudorapidity in the limit of negligible object mass,  $m \ll E$ . A common measure of distances between object is defined as  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ . This separation can also be defined using rapidity instead of pseudorapidity, as  $\Delta R_y = \sqrt{\Delta y^2 + \Delta\phi^2}$ .

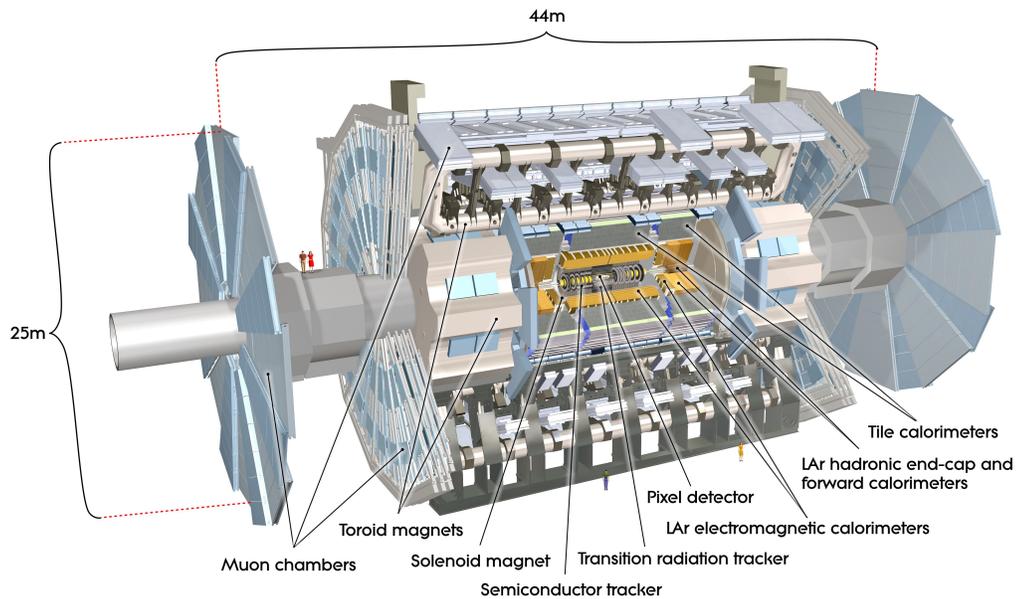


Figure 3.4: The complete ATLAS detector in cutaway view [45].

### 3.2.2 Detector overview

A schematic of the ATLAS detector is shown in figure 3.4. The full detector is 44 m long and 25 m high, roughly rotationally symmetric around the LHC beam pipe, and has a weight of around 7 000 metric tons. It is centered around the interaction point of the colliding LHC beams, and symmetric under reflection across the  $z = 0$  plane. The central barrel region of the detector consists of multiple detector subsystems arranged as concentric cylinders. A disk-shaped end-cap is located at each side of the barrel.

The inner detector (ID) is located at the core of ATLAS, and embedded within a 2 T axial magnetic field provided by a solenoid magnet. It provides tracking of charged particles within the central  $|\eta| < 2.5$  region. Electromagnetic and hadronic calorimeters surround the inner detector. The outermost parts of ATLAS are formed by the muon spectrometer (MS), which is immersed in a magnetic field provided by toroid magnets in barrel and end-caps.

### 3.2.3 Inner Detector

The ATLAS ID system is designed to provide precision tracking information for charged particles up to  $|\eta| < 2.5$ . It is located within a 2 T magnetic field generated by a solenoid magnet with a 2.5 m diameter and a length of 5.3 m in the  $z$  direction. This axial field is parallel to the  $z$ -axis, bending the tracks of charged particles in the  $\phi$  direction. The deflection allows for measurements of momentum and charge of these particles. The precision information provided by the ID is also used to reconstruct the primary vertex, the measured location at which the hard parton scattering of interest in a given event

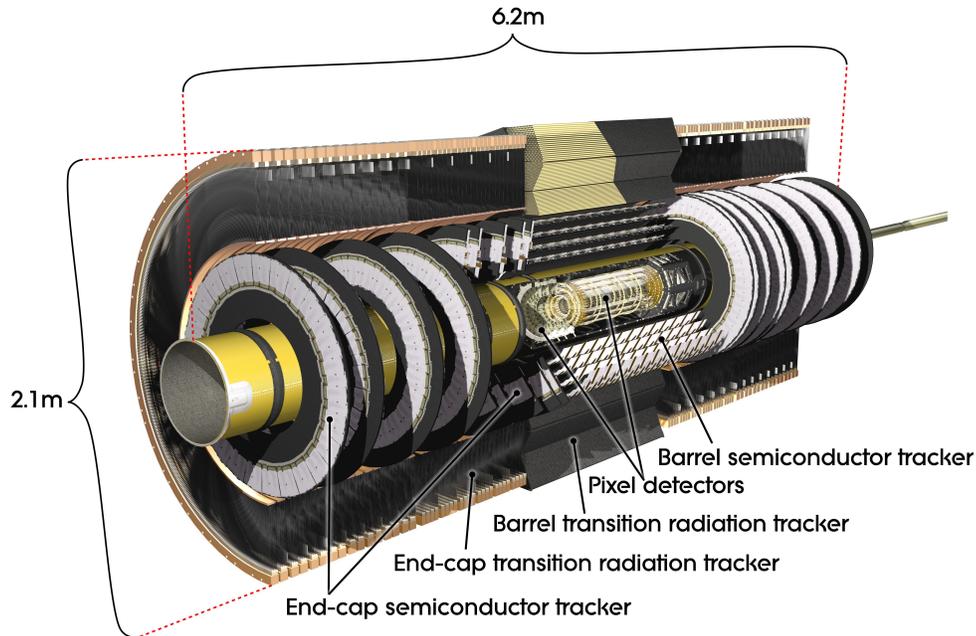


Figure 3.5: Cutaway view of the ATLAS ID. The IBL is missing in this visualization [45].

took place.

Figure 3.5 illustrates the ID layout. Closest to the beam pipe is the pixel detector, followed by the semiconductor tracker (SCT) and the transition radiation tracker (TRT). Precision tracking information within  $|\eta| < 2.5$  is provided by the pixels and the SCT. They are arranged as concentric cylinders centered around the beam pipe in the barrel, and disks in both end-caps.

### Pixel detectors

The innermost subsystem of the ID are silicon pixel detectors. Charged particles passing through them are detected via the electron–hole pairs they create in the semiconductors. There are four layers in the pixel system. The innermost layer, called insertable B-layer (IBL), was installed between Run-1 and Run-2 of the LHC [52]. It has the highest granularity, with a pixel size of  $50\ \mu\text{m}$  in the  $\phi$  direction, and  $250\ \mu\text{m}$  in the  $z$  direction. This layer is located at a radius  $r = 33\ \text{mm}$  from the beam pipe center, and covers the  $|\eta| < 3.0$  region. The pixels in the remaining three layers have sizes of  $50\ \mu\text{m}$  in the  $\phi$  direction, and  $400\ \mu\text{m}$  in the  $z$  direction. In total, the pixel detector system contains around 86 million pixels. The expected hit resolution in the IBL is  $8\ \mu\text{m}$  in the  $\phi$  direction and  $40\ \mu\text{m}$  in the  $z$  direction. The remaining three pixel layers have a decreased resolution of  $10\ \mu\text{m}$  in  $\phi$  and  $115\ \mu\text{m}$  in the  $z$  direction.

### Semiconductor tracker

The SCT consists of silicon strip detectors, with strips of size  $80\ \mu\text{m} \times 12\ \text{cm}$ . It is made up of four layers in the barrel, and two disks in each end-cap. In each barrel layer, one set of strips is parallel to the  $z$  axis, and a second set of strips is rotated by a stereo angle of  $40\ \text{mrad}$  with respect to the first set. The end-cap disks have one set of strips in the radial direction perpendicular to the  $z$  axis, with another set again rotated by  $40\ \text{mrad}$ . The resolution in the SCT is  $17\ \mu\text{m}$  in the  $\phi$  direction, and  $580\ \mu\text{m}$  in the  $z$  direction. The SCT contains around 6 million readout channels.

### Transition radiation tracker

The outermost part of the ID is the TRT. It consists of around 300 000 straw tubes with a diameter of  $4\ \text{mm}$  in the region  $|\eta| < 2.0$ . In the barrel region, the straws are  $144\ \text{cm}$  long and parallel to the  $z$  axis. The straws in the end-caps are  $37\ \text{cm}$  long, and positioned radially, perpendicular to the  $z$  axis. A wire runs through the center of the straws, and they are filled with gas. The wire is held at a potential difference with the tube walls. Charged particles passing through the tubes ionize the gas, and the resulting electrons drift to the anode wire. The resolution in the  $\phi$  direction per straw is  $130\ \mu\text{m}$ . Besides the tracking capabilities provided by the TRT, it is also used for particle identification. The straws are interleaved with polypropylene, and transition radiation is emitted at the material boundaries. Electrons can be distinguished from charged pions due to the larger amount of transition radiation they leave behind.

## 3.2.4 Calorimeters

The solenoid magnet containing the ATLAS ID is surrounded by a calorimeter system. Electromagnetic and hadronic calorimeters provide energy measurements of particles passing through them, covering the range up to  $|\eta| < 4.9$ . The system is designed to absorb the energy of most SM particles originating from the collision, with the exception of muons and neutrinos. Figure 3.6 shows the ATLAS calorimeter system.

### Electromagnetic calorimeter

The electromagnetic calorimeter consists of a barrel and two end-cap components, located in separate cryostats and covering the range  $|\eta| < 3.2$ . It is a sampling calorimeter, consisting of alternating layers of lead absorbing plates and liquid argon as an active medium. Electromagnetic showers develop mostly in the absorbing material, and are measured in the active medium. Lead plates and electrodes are arranged in an accordion shape, with a potential difference applied between them to collect charges from ionization left by particles passing through the calorimeter. The granularity varies across the calorimeter, and it is highest for  $|\eta| < 2.5$ , to match the region where precision tracking information from the ID is available. The highest granularity is  $0.025 \times 0.025$  in  $\eta \times \phi$ . In total, the electromagnetic calorimeter has around 170 000 readout channels.

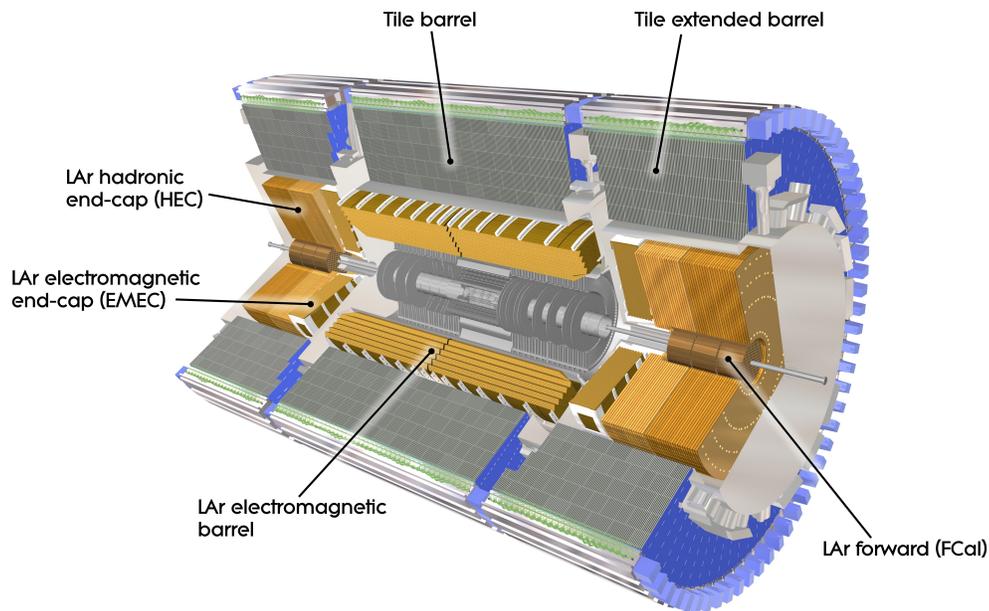


Figure 3.6: Cutaway view of the ATLAS calorimeter system surrounding the ID and solenoid magnet [45]. The term LAr refers to liquid argon as active material.

The electromagnetic calorimeter provides energy measurements of electrons and photons. Electrons at the LHC lose most of their energy via bremsstrahlung and subsequent electron–positron pair production. For photons, electron–positron pair production is the dominant energy loss process. The radiation length  $X_0$ , which is a function of atomic number and mass number of the material through which the particles are passing, characterizes the distance scale over which the energy losses take place. One radiation length is the average distance over which bremsstrahlung reduces the electron energy by a factor of  $1/e$ , and corresponds to roughly  $7/9$  of the mean free path for photon-induced pair production [9]. The electromagnetic calorimeter has a thickness of more than  $22X_0$  in the barrel, and more than  $24X_0$  in the end-caps, to provide good containment of electromagnetic showers originating from bremsstrahlung and pair production.

### Hadronic calorimeters

The hadronic calorimeter system consists of three components, with a total of around 19 000 readout channels. The tile calorimeter covers the central region  $|\eta| < 1.7$ , with a barrel in the region  $|\eta| < 1.0$  and two extended barrels covering  $0.8 < |\eta| < 1.7$ . It is a sampling calorimeter made up of steel as an absorption material and scintillating plastic tiles as active material. The scintillating tiles are read out via photomultiplier tubes. The tile calorimeter has three layers, with the highest granularity of  $0.1 \times 0.1$  in  $\eta \times \phi$  in the first two layers.

The end-cap hadronic calorimeters are located directly outside the electromagnetic calorimeters.

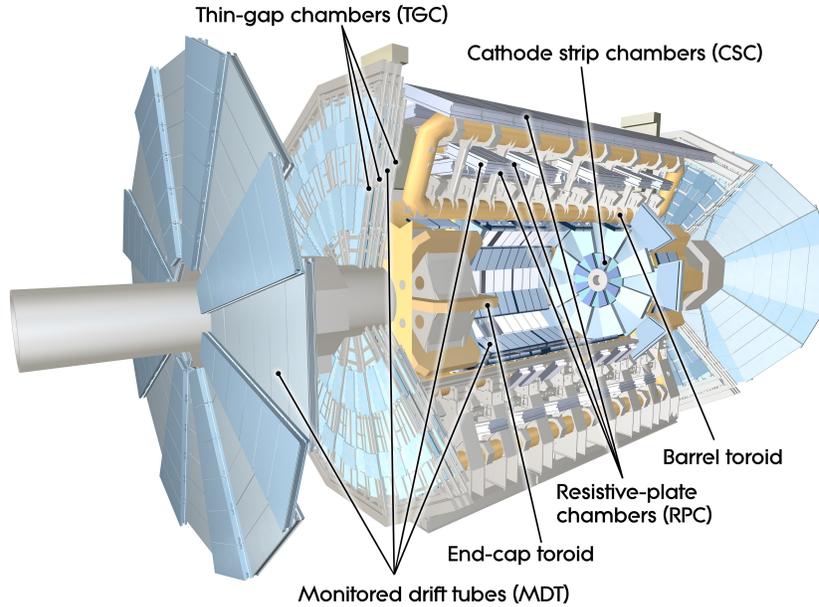


Figure 3.7: Cutaway view of the ATLAS MS [45].

Copper plates are used as an absorbing material, while the active medium is liquid argon. They cover the range  $1.5 < |\eta| < 3.2$ , with the highest granularity of  $0.1 \times 0.1$  in  $\eta \times \phi$  in the region  $|\eta| < 2.5$ .

The forward calorimeter covers the range  $3.2 < |\eta| < 4.9$  and is made up of three layers. The first layer uses copper as absorbing material, optimized for electromagnetic measurements. The remaining layers use tungsten for hadronic measurements. Tubes, aligned parallel to the  $z$  direction, are located within the absorbing material. They contain small gaps of less than 1 mm, filled with liquid argon as active material, and a rod in the center.

Charged hadrons lose energy via ionization of the surrounding material, and both charged and neutral hadrons also lose energy by undergoing strong interactions with nuclei in the surrounding medium. The nuclear interaction length  $\lambda_I$  describes the mean distance between hadronic interactions for relativistic hadrons. The hadronic calorimeter has a thickness of around  $10\lambda_I$ , therefore containing the majority of the energy in hadronic showers within the hadronic calorimeter. These showers also contain electromagnetic components, for example from the decay of neutral pions into two photons.

### 3.2.5 Muon spectrometer

The MS forms the outermost layer of ATLAS. Four different detector systems make up the MS, with more than one million readout channels in total. A cutaway view is shown in figure 3.7. The MS is embedded in a magnetic field produced by a system of three superconducting toroidal magnets, one in the barrel, and one in each end-cap. The magnets provide an average magnetic field of around 0.5 T across the MS, pointing in the  $\phi$  direction and therefore generally perpendicular to the muon

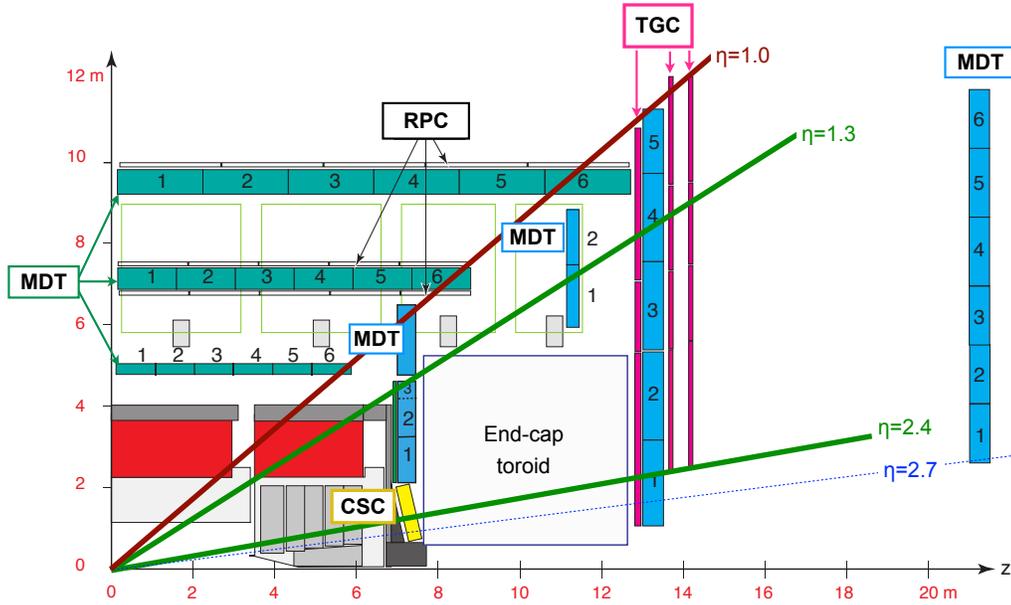


Figure 3.8: Schematic of one quarter of a cross-section through the ATLAS detector [53].

propagation direction. This field deflects muons passing through the MS in the  $\eta$  (or equivalently  $z$ ) direction, allowing for measurement of their momenta.

Figure 3.8 shows a schematic view of a cross-section of a quarter of the detector. The outermost chambers in the barrel region are located at a radius of around 10 m from the beam pipe, and the outermost end-cap disks at  $|z| \approx 21.5$  m. The muon system is not completely symmetric under rotation in the  $\phi$  direction due to gaps needed for detector services and support structure (feet).

### Muon trigger chambers

Two detector systems allowing for fast readout are used to make initial trigger decisions (the trigger system is described in section 3.2.6). Different experimental conditions in the barrel and end-cap regions motivate the use of two detector technologies. A higher granularity is needed in the end-caps to match the momentum resolution in the barrel, and the radiation levels in the end-caps are higher.

Three layers of resistive plate chambers (RPCs) are used in the barrel region  $|\eta| < 1.05$ . They consist of parallel plates with high resistivity held at a potential difference, with a gas mixture in the gap between them. Muons ionize the gas, and the resulting charges are collected on the plates. Besides trigger information, the RPCs provide  $\eta$  and  $\phi$  measurements, with a resolution of around 10 mm in both the  $z$  direction and the plane tangential to the  $\phi$  direction.

The end-caps use thin gap chambers (TGCs) in the region  $1.05 < |\eta| < 2.4$ , which deal with the increased rate requirements due to non-collision background processes. The chambers of these detectors are formed by graphite-coated cathodes, filled with a gas mixture, and contain multiple wires separated by 1.8 mm. The TGCs also provide a measurement of the  $\phi$  coordinate, with a resolution of around 5 mm.

### Precision muon tracking chambers

Two additional detector systems provide high position resolution and precision tracking information. These systems are slower, and only read out after an initial trigger decision has been made.

A system of monitored drift tubes (MDTs) covers the range  $|\eta| < 2.7$ . It is made of aluminum drift tubes with a diameter of 3 cm, filled with a gas mixture. The drift tubes contain a wire in the center, which is held at a potential difference with the tube. Muons passing through the tubes ionize the gas, and the resulting electrons are collected at the central wire. The electron drift time can reach up to 700 ns, and the length of the signal pulse indicates how far from the wire the muon passed through the tube. The drift tubes are aligned tangentially to the  $\phi$  direction to achieve high position accuracy in the  $z$  direction. MDT chambers consist of three to eight layers of drift tubes, achieving an average position resolution of 35  $\mu\text{m}$  per chamber in the  $z$  direction. The measurement in the  $\phi$  direction is obtained from the remaining MS systems.

The forward region  $2.0 < |\eta| < 2.7$  contains cathod strip chambers (CSCs). The CSCs are multi-wire proportional chambers like the TGCs. They are used for their high rate capability and time resolution of 4 ns. The anode wires are arranged in the radial direction, with a spacing of 2.5 mm. On one end, the cathodes form strips perpendicular to the wires, providing the precision measurement in the radial direction. The cathodes on the other end are perpendicular to the wires, and provide a measurement in the  $\phi$  direction. The position resolution is 40  $\mu\text{m}$  in the radial direction and 5 mm in  $\phi$ .

### 3.2.6 Trigger and data acquisition

It is impossible to record all collision events in ATLAS at the 40 MHz collision rate provided by the LHC. The ATLAS trigger and data acquisition system reduces this input down to a rate of around 1 kHz of events, which are recorded and kept for subsequent physics analyses [54, 55]. It consists of two components, the hardware-based Level-1 (L1) trigger and a software-based high-level trigger (HLT). Events accepted at the fast initial L1 step are passed on to the HLT, which runs more precise algorithms and makes a final trigger decision. Pre-scale factors  $N$  can be applied to triggers, causing only every 1 in  $N$  events passing the trigger to be accepted by the trigger. The use of such factors limits the trigger rate, allowing to trigger on objects with less stringent requirements, which are produced more often by the LHC. Triggers used for physics analyses are typically not pre-scaled. A wide range of triggers, summarized in a trigger menu, is used during ATLAS data-taking. They include triggers for events with one or multiple characteristic objects (such as charged leptons with high transverse momenta), combinations of different object types, or triggers for particular event topologies.

#### Level-1

The L1 trigger step searches for events with high transverse momentum objects, including charged leptons, photons, jets, and missing transverse momentum (see chapter 4). It uses data from the calorimeters and MS, and defines coarse regions of interest (RoIs) in which the relevant objects are located, for further processing at the subsequent steps. The L1 trigger reduces the event rate to around

100 kHz, with a 2.5  $\mu$ s maximum latency. The minimum duration between two consecutive events accepted by the L1 trigger is limited, as is the number of events accepted over a given period of time. Events passing this trigger step are buffered for further processing in the HLT.

### High level trigger

After the incoming event rate has been reduced by the L1 trigger system, the software-based HLT system can run more precise and computationally expensive reconstruction algorithms to refine the trigger decision. These reconstruction algorithms generally are run in the RoIs defined at the L1 step. Events passing the HLT criteria are transferred to local storage at the ATLAS site, and then sent to CERN's computing center for storage and subsequent processing. In 2015, the average HLT processing time for events at the highest instantaneous luminosity reached was around 200 ms.

### Muon trigger system

Given its relevance for the trigger efficiency analyses presented in chapter 10 of this dissertation, the muon trigger system is briefly explained in more detail in this section. This system is designed to identify events containing muons over a large spectrum of transverse momentum, with high efficiencies and moderate trigger rates.

The L1 step uses only inputs from the fast muon trigger chambers, the RPCs and TGCs. The system requires a hit coincidence in the trigger chambers, which points back to the beam interaction region. A rough muon candidate track transverse momentum is obtained at this step by comparing it to a track expected from a muon with infinite transverse momentum. The L1 defines RoIs of size  $0.1 \times 0.1$  in  $\eta \times \phi$ , which are passed to the HLT for further processing. Limited mostly due to the effect of detector geometry on the muon trigger chamber distribution, the L1 trigger covers around 80% of the barrel region and 99% of the end-caps.

The trigger decision is refined at the HLT step by incorporating higher resolution and precision tracking information from the MDT and CSC detector systems. A coincidence of generally 2–3 hits in multiple detector layers is required. This step has an efficiency of close to 100% with respect to L1, and reduces the trigger rate by a factor of roughly 100.

The HLT step consists of two stages: an initial fast reconstruction step is followed by precision muon reconstruction. Fast track reconstruction is performed in the RoIs defined at L1, using only information from the MS, with measurements from the MDTs and CSCs. A refined track fit is subsequently performed by combining measurements from the ID and the MS. If candidates pass this stage, they enter the precision HLT step.

High resolution muon reconstruction takes place in the precision step, with inputs from both the ID and MS. Track candidates in the MS are extrapolated back and combined with ID information (called outside-in). A second approach, which starts with ID information and then extrapolates and combines it with the MS (inside-out), recovers inefficiencies at low muon transverse momentum.

Following this step, optional requirements can be applied. The use of muon isolation cuts helps distinguish between prompt and non-prompt muons. Prompt muons originate from the initial

hard scatter process taking place, while non-prompt muons can arise from the decay of charged hadrons. The muon isolation requirement is performed by rejecting events where the scalar sum of the transverse momentum of all tracks in a cone around the muon candidate is large. This sum is expected to be small for prompt muons. Applying muon isolation requirements allows to lower the muon trigger transverse momentum threshold, while maintaining a reasonable trigger rate.

Additional HLT algorithms exist besides the baseline strategy described. There are triggers using only MS information, and do not require information from the ID. In order to circumvent the trigger efficiency loss in the L1 step, the full-scan approach does not rely on L1 RoIs, but searches the entire MS for additional muons. While this is very computationally expensive, it provides a high trigger efficiency. It is suitable for finding additional muons for multi-muon triggers, when one muon has already been found with the baseline approach.

### 3.2.7 Data quality requirements and available data for analyses

Collision events from the LHC happening within ATLAS can only be used in physics analyses if they passed a trigger and were recorded. Additional requirements are made to ensure high quality data. These requirements include that all detector subsystems were operational, and the beams provided by the LHC were stable when the event was recorded. Each event must furthermore contain a vertex (see section 4.2.2) with at least two associated tracks with transverse momenta  $p_T > 400$  MeV.

Figure 3.9 shows the total integrated luminosity during Run-2 as a function of time. The green histogram corresponds to the total amount of data delivered by the LHC, amounting to  $156 \text{ fb}^{-1}$ . Around 94% of this was recorded by the ATLAS detector, a total of  $147 \text{ fb}^{-1}$ , shown in yellow. Additional requirements on the quality of the reconstructed data reduce the integrated luminosity available for physics analyses further, and a total of  $139.0 \text{ fb}^{-1}$  remains to be analyzed in the full Run-2 ATLAS dataset. This contribution is shown in blue.

### 3.2.8 Simulation of ATLAS

The simulation of the ATLAS detector is based on the GEANT4 software toolkit [56] and implemented within the ATLAS simulation framework [57]. It models the interactions of stable particles with the detector, and the resulting signals received by the detector. The analog detector signals are digitized and combined with pile-up effects, simulating both in-time and out-of-time pile-up. Trigger algorithms are simulated, and the events are further reconstructed with the same algorithms used for data, described in section 4. An alternative faster simulation method, called AFII, is also used. It uses a parameterized calorimeter response to electromagnetic and hadronic showers, instead of simulating it in detail [58].

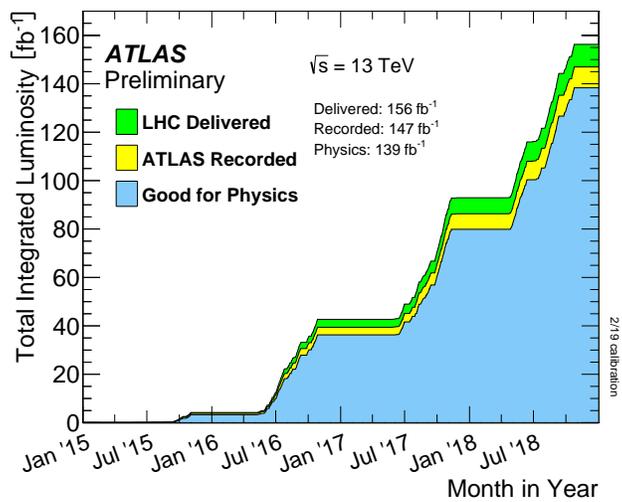


Figure 3.9: Total integrated luminosity collected by the LHC (green), recorded by the ATLAS detector (yellow) and used for physics analyses (blue), shown as a function of time [51].

## 4. Object reconstruction

This chapter provides an overview of the reconstruction of physics objects with the ATLAS described in section 3.2, with a focus on the objects of relevance to this dissertation. Basic components used in the identification of particles, tracks and vertices, are described in section 4.2. Section 4.3 outlines the reconstruction of charged leptons, with a focus on muons and electrons. The approach used to reconstruct jets is described in section 4.4. This section also includes a description of flavor tagging algorithms, which are employed to identify jets containing  $b$ -hadrons. The definition of missing transverse energy is given in section 4.5. The section ends with a description of the procedure used to resolve the overlap between the various physics objects reconstructed. This is described in section 4.6.

### 4.1 Reconstruction overview

A schematic depiction of different fundamental particles interacting with the ATLAS detector is shown in figure 4.1 [59]. Muons pass through the whole detector, and are reconstructed mainly

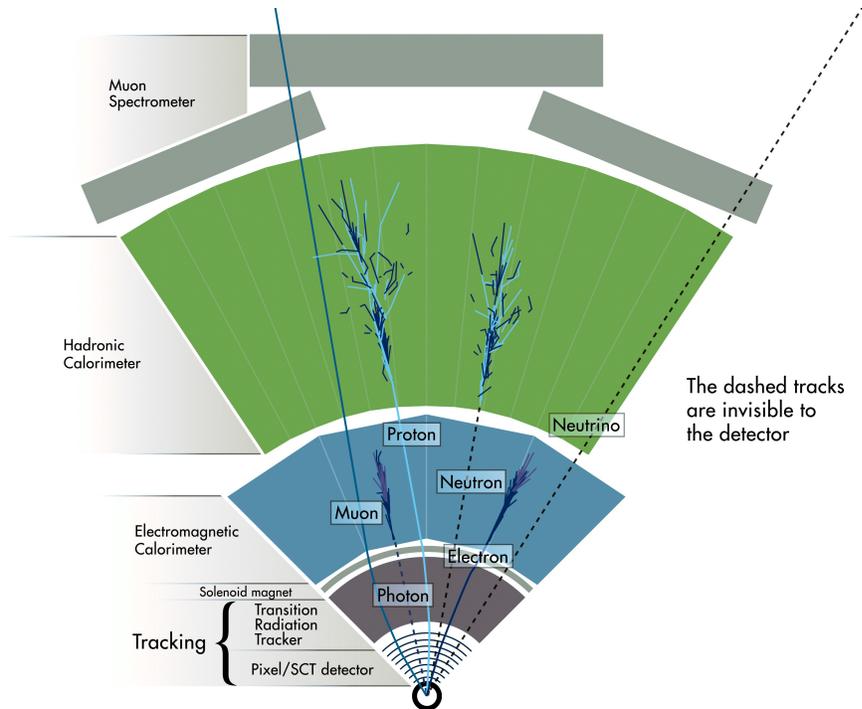


Figure 4.1: Schematic of fundamental particles interacting with the ATLAS detector, adapted from reference [59]. It shows a section of the  $x$ - $y$  plane.

from their interactions with the ATLAS ID and MS system. Since they do not carry electric charge, photons do not interact with the ID, and form a collimated shower in the electromagnetic calorimeter. Electrons cause the same kind of electromagnetic shower, but interact with the ID on top of this. The electrically charged protons also interact with the ID, and then deposit their energy in the hadronic calorimeter, in a shower that is less confined than electromagnetic showers. Neutrons cause the same hadronic shower, but do not interact with the ID due to their lack of electric charge. The ATLAS detector is unable to detect neutrinos directly, however their presence may be inferred by considering momentum conservation. Missing transverse energy, described in section 4.5, can be caused by neutrinos in the SM.

## 4.2 Tracks, vertices and energy clusters

Charged particles passing through the ATLAS ID leave behind tracks. Primary vertices are located at proton–proton interaction points. The energy deposited by particles in the ATLAS calorimeter system is grouped into clusters. Tracks, vertices, and calorimeter energy clusters are inputs to the reconstruction of other physics objects discussed in this chapter.

### 4.2.1 Tracks

The reconstructed trajectories of charged particles passing through the ATLAS detector are called tracks. This section provides an overview of track reconstruction with ATLAS. More detailed descriptions of the algorithms used and their performance are provided in references [60, 61].

Sensor measurements above a threshold in the pixel and SCT detectors are grouped into clusters. The resulting three-dimensional measurements of the cluster positions are called space points. A cluster is typically composed of multiple contributing pixels, and the intersection point of a charged particle with the detector layer is obtained from combining the information of all pixels in the cluster. Clusters can contain charge deposits from multiple particles, and they also may be used in the reconstruction of multiple tracks.

The baseline inside-out tracking algorithm starts by defining sets of three space points, used as seeds for tracks. These seeds are combined with additional space points compatible with the preliminary track trajectory estimated from the three space points, then forming track candidates. This is done using a Kalman filter, which iteratively updates the best track candidate estimate as more space points are added. A score for each track candidate is calculated, with higher scores assigned to candidates more likely representing the trajectories of charged particles. An ambiguity solver then considers track candidates in order of decreasing score. It limits the amount of clusters shared by different tracks, but not identified as being compatible with having originated from multiple tracks. Further quality criteria are applied in this step, and track candidates failing to pass the ambiguity solver are rejected. These criteria include that at least 7 clusters are assigned to a track, and that it contains at most two holes. Holes are defined as the intersection of a track candidate with a detector element

that does not contain a matching cluster, even though it would be expected from the trajectory of a particle following along the track candidate. Tracks are then extended into the TRT.

An additional outside-in algorithm starts with track segments reconstructed in the TRT, and extends them towards the inside of the detector, adding information from the silicon detectors.

### 4.2.2 Vertices

Vertices are locations of particle interactions, and identified via the tracks pointing away from them. Of particular relevance for measurements of particle kinematics are vertices at the points where proton–proton interactions took place. Their reconstruction in ATLAS is briefly described here. Additional details are provided in reference [62].

Vertex reconstruction consists of two steps. In the first step, reconstructed tracks are associated to candidate vertices, this is called vertex finding. During the subsequent vertex fitting step, the vertex position is reconstructed.

The vertex finding starts with a seed position for a candidate vertex. The optimal vertex candidate position is then updated by an iterative fitting procedure, using the tracks found previously. Each track is assigned a weight, which describes its compatibility with the candidate vertex position. Both track weights and vertex candidate position are updated throughout the iterative fit. All tracks that are incompatible with the vertex after the last iteration are not assigned to the vertex, and instead used in the determination of subsequent vertices. New iterations of the vertex finding algorithm are then performed on the tracks not yet assigned to any vertex, this is repeated until no new vertex can be found. The primary vertex of a collision event is defined as the vertex with the largest sum of squared transverse momenta of tracks associated to it. While other definitions exist, this definition is used in the results presented in this dissertation.

### 4.2.3 Energy clusters

Energy deposits in individual cells of the ATLAS calorimeter system are clustered together, forming three-dimensional topological cell clusters, also called topo-clusters. The procedure employed by ATLAS is described in detail in reference [63].

Topo-clusters are built starting from seeds. They are located at cells with high signal significance, which is given by the ratio of the signal to the expected average noise in each specific cell. Neighboring cells are iteratively added to the cluster if they have sufficient signal significance. This process continues until no significant cells are left to be added to the clusters. Cells with insignificant amounts of signal are not included in the clusters, thereby suppressing noise. A topo-cluster does not necessarily contain all of the energy deposited by a particle, especially in hadronic showers, which are less confined than electromagnetic showers. It may contain the full shower or only a fraction of it, or even calorimeter responses to multiple particles.

The ATLAS calorimeters are non-compensating, meaning that the signal of a hadron with the same energy as an electron or photon is smaller. Signals are measured using the electromagnetic energy scale, which correctly measures energies deposited by electrons and photons. Topo-clusters can be

calibrated to correct for the non-compensating nature of the calorimeters and additional effects, such as energy losses in inactive material. They can be interpreted as massless pseudo-particles, characterized by their energy and location in the  $\eta$  and  $\phi$  coordinates.

### 4.3 Leptons

This section describes the reconstruction of charged leptons with the ATLAS detector. The focus lies on electrons and muons, which are most relevant to the work presented in this dissertation. Tau leptons are only briefly described, as the  $t\bar{t}H(b\bar{b})$  analyses in chapter 6 and chapter 9 contain a veto for events with tau lepton with hadronic decay products, to ensure that the events analyzed in this analysis do not overlap with other searches for  $t\bar{t}H$ . Due to the special treatment of tau leptons, and the inability of the ATLAS detector to detect neutrinos directly, the term lepton is used for a reconstructed object in this dissertation when describing a light charged lepton (electron or muon).

#### 4.3.1 Muons

Muons are reconstructed mostly using information from the ATLAS ID and MS detector systems. Details about the relevant algorithms and their performance are given in reference [64].

##### Reconstruction

The first step of muon reconstruction takes place independently in both the ID and the MS. Tracks in the ID are reconstructed as described in section 4.2.1.

In the MS, segments of tracks are formed by combining nearby hits, which follow the trajectory expected from a muon, and are consistent with having originated from the proton–proton interaction point. Segments from different detector layers are combined, first using segments from the middle of the MS as seeds, and then extending the seeds also to the inner and outer layers. At least two segments are required to form a track candidate in general, and segments can be used in multiple track candidates. A  $\chi^2$ -based fit to the hits associated to each track candidate is then performed to determine whether a track candidate is accepted as a track.

Multiple different muon types exist, depending on the detector information used to reconstruct them. *Combined* muons are built from a combined track fit in the ID and MS. Most muons are found with an outside-in approach, which extrapolates the track from the MS to the ID. An inside-out algorithm is also used. *Extrapolated* muons are muons based only on a track in the MS, which is required to be consistent with having originated from the interaction point. These muons extend the muon acceptance in the region  $2.5 < |\eta| < 2.7$ , which is not covered by the ID. *Segment-tagged* muons are built from tracks in the ID and extrapolated to match at least one track segment in the MS. This recovers muons crossing only one MS layer. *Calorimeter-tagged* muons use energy deposits in the calorimeter compatible with having been deposited by a minimum-ionizing particle. These must be matched to an ID track. This muon type recovers inefficiencies in regions of gaps in the MS.

## Identification

During muon identification, further quality requirements are applied to muon candidates to suppress non-muon backgrounds (mostly originating from decays of charged hadrons) and to ensure a good muon momentum measurement. Non-prompt muons from decays of charged hadrons can lead to incompatible momentum measurements in the ID and MS, and the comparison of individual and combined fits helps reject such muons.

The muon identification criteria are collected in so-called operating points. For relevance to this dissertation are the *loose*, *medium*, and *high- $p_T$*  operating points.

The *medium* operating point is optimized to minimize systematic uncertainties related to muon reconstruction and calibration. It uses combined and extrapolated muon tracks. For combined muons, at least three hits in at least two layers of the MDT are required. The requirement is relaxed in the central  $|\eta| < 0.1$  region due to a gap in the MS. Extrapolated muons need to have hits in at least three MDT or CSC layers. A consistency check of the momentum measurement for combined muons in the ID and MS is also applied to reject non-prompt muons. The reconstruction efficiency for muons with transverse momentum above 20 GeV when using this operating point is 96%.

The *loose* operating point includes all muons passing the *medium* operating point requirements, but extends the acceptance further. It is designed for analyses searching for Higgs boson decays to four charged leptons, and increases the reconstruction efficiency to 98% for muons with transverse momentum above 20 GeV. This operating point includes segment- and calorimeter-tagged muons in the  $|\eta| < 0.1$  region.

The *high- $p_T$*  operating point is optimized for momentum resolution of tracks with transverse momenta above 100 GeV. It uses combined muons fulfilling the *medium* operating point requirements. On top of this, tracks need to have at least three hits in at least three MS detector layers. This improves the transverse momentum resolution of muons with very high momenta above 1.5 TeV by around 30%, while reducing reconstruction efficiency by roughly 20%.

## Isolation

Prompt muons are typically produced spatially well separated from other particles. Muon isolation, which refers to detector activity present around a muon candidate, can therefore be used to reject non-prompt muons. The decay of objects at high momenta, resulting in collimated decay products which can include muons, presents one possible exception. Two measures of isolation are used, one track-based and one based on calorimeter information. The track-based isolation is calculated as the scalar sum of transverse momenta of all tracks with  $p_T > 1$  GeV within a cone around the muon, excluding the momentum of the muon track itself. The cone size depends on the muon momentum  $p_T^\mu$  and decreases in size for higher momentum muons:  $\Delta R = \min(10 \text{ GeV}/p_T^\mu, 0.3)$ . The calorimeter-based isolation is obtained by summing the transverse energies of topo-clusters in a cone of size  $\Delta R = 0.2$  around the muon, again excluding the energy deposit from the muon itself, and also correcting for effects from pile-up. Isolation criteria are defined via the ratio of track-based or calorimeter-based isolation to the muon transverse momentum.

The isolation operating points used in this dissertation are *loose*, *gradient*, *FCTight* and *FCTTO*. The *loose*, *gradient*, and *FCTight* (which stands for "fixed cut tight") operating points use both track- and calorimeter-based isolation; the *FCTTO* (which stands for "fixed cut tight track-only") operating point only uses track-based isolation. The *loose* operating point provides a constant 99% efficiency for muons across  $\eta$  and  $p_T$ , while the *gradient* operating point is at least 90% efficient at muon transverse momenta of 25 GeV and 99% efficient at 60 GeV. The *FCTight* and *FCTTO* operating points are designed to be robust to pile-up effects.

### Corrections

In order to match data more accurately, corrections to muon momentum scale and resolution are applied to simulated muons. An additional correction is applied to resolve differences in muon selection efficiencies between data and simulation. These selection efficiencies originate from the association of tracks to vertices, muon identification, and muon isolation. They also include trigger efficiencies when muon triggers are used. All of these corrections are derived from studies making use of the clean decays of  $Z$  bosons and  $J/\psi$  mesons to muon pairs. The trigger efficiency corrections are described in more detail in chapter 10, where they are derived for additional event topologies.

#### 4.3.2 Electrons

Electrons are reconstructed using the tracks they leave in the ID, and their energy deposited in the electromagnetic calorimeter. A detailed overview of reconstruction algorithms and their performance can be found in reference [65].

### Reconstruction

Electrons lose their energy predominantly due to bremsstrahlung and subsequent electron–positron pair production from emitted photons. The energy from an electron is typically deposited within a single cluster in the electromagnetic calorimeter. Interactions of electrons with detector material can already happen before they enter the calorimeter, resulting in radiated photons converting into electron–positron pairs (photon conversions). This can result in multiple tracks being reconstructed in the ID, all originating from the same electron, and all pointing to the same cluster in the calorimeter. Electrons are reconstructed in the  $|\eta| < 2.47$  region of ATLAS, not including the overlap region between the barrel and end-caps.

Electron reconstruction starts by finding energy clusters in the electromagnetic calorimeter with transverse energies above 2.5 GeV. This procedure is more than 99% efficient for electrons with transverse momenta above 15 GeV. The track reconstruction in the ID, described in section 4.2.1, is more than 98% efficient for electron transverse momenta above 10 GeV. Tracks are electron candidates if they geometrically match an energy cluster, have at least four hits in silicon layers of the ID, and are not associated to a vertex that has been identified as originating from a photon conversion [66]. An algorithm selects the track most likely originating from the primary electron in case multiple tracks

match the criteria. The energy cluster size is then extended and calibrated to accurately represent the original electron energy. The electron position in the  $\eta$  and  $\phi$  coordinates is obtained from the track matching the cluster.

### Identification

Likelihood-based identification algorithms are applied to electron candidates in order to ensure high quality electrons for analysis and suppress non-prompt electron backgrounds, which include photon conversions, jets, and non-prompt electrons from heavy-flavor quark decays. The likelihood is calculated from a range of measurements, including the shower shape in the electromagnetic calorimeter, energy deposited in the hadronic calorimeter, details about the track matched to the electron and the matching itself, as well as the transition radiation in the TRT to discriminate against pions. The probability distribution functions of the inputs to the algorithm are derived from simulation, and corrected to accurately model data. Electron identification proceeds by requiring that a likelihood-based discriminant, which increases in value for more electron-like candidates, has at least a specific minimum value, which depends on the operating point used.

The electron identification operating points relevant to the work presented in this dissertation are *tight*, *medium*, and *LooseAndBLayer*. The *tight* operating point is 80% efficient at identifying prompt electrons with transverse energies of 40 GeV, while the *medium* operating point is 88% efficient. The *LooseAndBLayer* operating point is a variation of the 93% efficient *loose* operating point. It includes a requirement of a hit in the innermost pixel layer, which the *tight* operating point does as well. Further tracking requirements are applied for all operating points. These include at least two hits across all pixel layers and at least seven hits in pixel and SCT layers combined.

### Isolations

Like muons, prompt electrons are generally expected to be spatially separated from other particles, and correspondingly detector activity associated to them is expected to be isolated.

Electron isolation is calculated similarly to muon isolation, with track-based and calorimeter-based isolation variables. Track-based isolation is calculated from the scalar sum of transverse momenta of tracks with  $p_T > 1$  GeV around the electron. The track matched to the electron is not included, and neither are nearby tracks likely having originated from photon conversions. The cone has a variable size depending on the electron momentum  $p_T^e$ ,  $\Delta R = \min(10 \text{ GeV}/p_T^e, 0.2)$ . The calorimeter-based isolation is given from the sum over transverse energies of topo-clusters in a cone of size  $\Delta R = 0.2$  around the electron, with the electron energy deposit at the core removed and additional corrections for pile-up applied. Isolation criteria are defined via the ratio of track-based or calorimeter-based isolation to the electron transverse momentum.

The isolation operating points used in this dissertation are *loose* and *gradient*. These operating points use both track- and calorimeter-based isolation. The *loose* operating point is 98% efficient, while the *gradient* operating point is 90% and 99% efficient for electrons with transverse momenta of 25 GeV and 60 GeV, respectively.

## Corrections

A calibration of the electron energy scale and resolution is applied to electrons in data and simulation [67]. It is derived from samples with  $Z$  boson decays to electron pairs, and the corrections are validated in samples with  $J/\psi$  meson decays to electron pairs, as well as  $Z$  boson decays to electron pairs with an additional photon. Electron selection efficiency corrections are derived using the clean decays of  $Z$  bosons and  $J/\psi$  mesons to electron pairs. The selection efficiency in simulation is corrected for effects due to electron reconstruction, identification, and isolation. When electron triggers are used, the acceptance efficiency due to these triggers is also corrected.

### 4.3.3 Tau leptons

Tau leptons decaying into final states with electrons or muons are reconstructed as these lighter leptons. Separate techniques are used for tau leptons decaying into hadronic final states, also called hadronic tau leptons [68]. Their reconstruction starts from jets, which are described in section 4.4.1. Additional criteria on the associated tracks help distinguish lepton candidates from jets. The tau lepton energy is calibrated depending on their transverse momentum. Boosted decision trees (BDTs) are used to identify taus decaying into hadronic final states. Different operating points exist for the tau lepton identification, including the *medium* operating point used for the  $t\bar{t}H(b\bar{b})$  analyses presented in chapter 6 and chapter 9.

## 4.4 Jets and flavor tagging

Jets consist of showers of hadrons originating from partons produced in proton–proton collisions. The showers develop from color-charged particles produced in the hard scattering interaction in collisions at the LHC. Clustering algorithms group together the energy deposits from the shower in the ATLAS calorimeter system into jets. Flavor tagging algorithms are used to distinguish between jets containing hadrons with quarks of different flavors. This section describes jets at ATLAS and the flavor tagging algorithm used for the results in this dissertation to identify jets containing  $b$ -hadrons.

### 4.4.1 Jets

The jet definition relevant for the analyses in this dissertation employs energy deposits in the ATLAS calorimeter system, clustered together to form jets. This section provides an overview of jets in these analyses. Details about reconstruction and calibration of jets are given in reference [69].

#### Formation from energy clusters

Jet reconstruction starts with three-dimensional topological cell clusters described in section 4.2.3, which are clustered together into jets. The clustering is performed with the anti- $k_t$  algorithm [70], implemented in the FASTJET package [71]. In the anti- $k_t$  algorithm, the distance measure between

two objects  $i, j$  (characterized by their transverse momenta, rapidities  $y$ , and azimuthal angles) is defined as

$$d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2}. \quad (4.1)$$

The parameter  $R$  is variable, and set to  $R = 0.4$  for the results presented in this dissertation. After calculating the distances between all objects, they are combined together in order of increasing distance. If  $d_{ij} > p_{T,i}^{-2}$  for any object  $j$ , the object  $i$  is called a jet and not used anymore in the clustering of the remaining objects.

The distance between two objects with low momentum is large compared to the distance between objects with low and high momentum. High momentum objects therefore cluster together with low momentum objects in their vicinity, forming a conical jet with radius  $R$ , as long as no other high momentum objects are nearby. Low momentum objects do not modify the shape of the jet, making the algorithm infrared safe; it is furthermore also collinear safe (and therefore unaffected by collinear gluon emissions).

### Calibration and selection

The topological clusters from which a jet is built are calibrated to the electromagnetic scale, and correctly measure the energy deposits from electromagnetically interacting particles. In the first calibration step, the jet four-momentum is scaled such that the jet points to the primary vertex, while keeping the same energy. After this, contributions to the jet energy from pile-up are removed. The jet energy and direction are then calibrated to match the behavior derived from simulation. It corrects both the jet energy scale and resolves reconstruction biases as a function of  $\eta$ , which occur due to changes in calorimeter granularity. Further corrections are derived by including also tracking and ATLAS MS information, which increase the jet energy resolution. The final step consists of residual corrections to jets in data, accounting for differences between data and simulation. It is derived using well-measured reference objects, such as photons or  $Z$  bosons.

After these calibration steps, quality criteria on jets are applied [72]. Events with jets failing to meet these criteria are not considered for further analysis. The effect of pile-up is mitigated by employing an algorithm called jet vertex tagger [73]. It rejects jets where a significant fraction of the transverse momentum of tracks assigned to the jet is not associated to the primary vertex.

#### 4.4.2 Flavor tagging

The identification of jets containing hadrons of a specific flavor is called flavor tagging. It is used to distinguish between the kind of parton a given jet originated from. This section focuses on the description of a so-called  $b$ -tagging algorithm, which is designed to identify jets containing  $b$ -hadrons. Such jets are called  $b$ -jets, and originate from bottom quarks produced in the initial proton–proton scattering. Their identification is crucial for the identification of processes like  $t\bar{t}H(b\bar{b})$ , where many bottom quarks are expected to be produced, and subsequently many  $b$ -tagged jets are expected in a

reconstructed event. In contrast to  $b$ -jets, the so-called light jets originate from the first generation and  $s$  quarks, while  $c$ -jets originate from charm quarks.

This section provides an overview of the  $b$ -tagging algorithm used for the work presented in this dissertation. More details are provided in reference [74]. The performance of the algorithm version relevant for the  $t\bar{t}H(b\bar{b})$  analysis in chapter 6 is summarized in reference [75]. The  $t\bar{t}H(b\bar{b})$  analysis in chapter 9 and the muon trigger efficiency measurement in chapter 10 use an updated algorithm version, which is described in reference [76].

### Algorithm overview

The most important ingredients to the  $b$ -tagging algorithm are jets, tracks reconstructed in the ID and the primary vertex. Tracks are associated to jets based on their separation  $\Delta R$ , within a cone of varying size depending on the jet transverse momentum.

The identification of  $b$ -hadrons in jets makes use of their long lifetime and high mass. With a lifetime around 1.5 ps, a  $b$ -hadron with a transverse momentum of 50 GeV travels around 4.5 mm in the transverse direction before it decays. Tracks associated to the hadron can be identified by their large impact parameters. The transverse impact parameter  $d_0$  is the closest distance between primary vertex and track in the transverse direction. The longitudinal impact parameter  $z_0$  measures this closest distance in the  $z$  direction. Tracks significantly displaced from the primary vertex can be used to reconstruct a secondary vertex where the  $b$ -hadron decay takes place. A dedicated algorithm aims at reconstructing the  $b$ -hadron decay chain [77].

The BDT based MV2 algorithm combines the information from various other algorithms, using impact parameter information, the reconstruction of a secondary vertex and the  $b$ -hadron decay chain. It is designed to correctly identify  $b$ -jets. The MV2c10 algorithm used for the results in this dissertation is trained with  $b$ -jets as signal, and a mixture of 93% light jets and 7%  $c$ -jets as background. The jets used for the training are taken from a sample of simulated  $t\bar{t}$  events.

### Operating points

The MV2c10 algorithm is used with a range of operating points, achieving different efficiencies and rejection rates for  $b$ -jets and other jets, respectively. The  $b$ -tagging efficiency is the rate at which true  $b$ -jets are correctly identified as such. For a given efficiency, the algorithm performance is quantified by the rejection of other jets, such as  $c$ -jets and light jets. A rejection factor  $r$  means that one in  $r$  jets will mistakenly be tagged as a  $b$ -jet. Larger  $b$ -jet efficiencies result in lower rejection of other jets.

Four different operating points are defined for the MV2c10 algorithm, corresponding to  $b$ -jet efficiencies of 60%, 70%, 77% and 85%. These operating points are referred to as *very tight*, *tight*, *medium* and *loose*, respectively. By applying the algorithm to every jet in an event, each jet can be classified into one of five classes. When a jet satisfies any of the four operating points, it is called a  $b$ -jet. If it fails to satisfy the *loose* operating point, it is instead classified as untagged. The tightest operating point satisfied by a jet is used to refer to it, since it also satisfies all operating points with higher efficiency by design.

Table 4.1: Operating points of the MV2c10 algorithm, with corresponding  $b$ -jet identification efficiencies and rejection factors for  $c$ -jets and light jets [75].

| operating point   | $b$ -jet efficiency | $c$ -jet rejection factor | light jet rejection factor |
|-------------------|---------------------|---------------------------|----------------------------|
| <i>very tight</i> | 60%                 | 34                        | 1538                       |
| <i>tight</i>      | 70%                 | 12                        | 381                        |
| <i>medium</i>     | 77%                 | 6                         | 134                        |
| <i>loose</i>      | 85%                 | 3                         | 33                         |

An overview of the operating points and their performance is shown in table 4.1, taken from [75]. The performance is evaluated using a sample of simulated  $t\bar{t}$  events.

### Calibration

The performance of the MV2c10 algorithm is evaluated using various event topologies enriched in  $b$ -,  $c$ - and light jets. Using these measurements, scale factors for the  $b$ -jet tagging efficiency and  $c$ - and light jet mis-tag rates are derived. These scale factors are applied to simulation to match the performance measured in data, and depend on the true jet flavor.

## 4.5 Missing transverse energy

Conservation of four-momentum implies that the vector sum of all objects produced in a collision at the LHC is equal to the sum over colliding partons. As the LHC collides protons head-on, the transverse momentum of the system containing all objects produced in a collision should vanish. Not all of the objects in this system are always detected by ATLAS; neutrinos leave the detector unseen. The resulting momentum imbalance is restored by adding the so-called missing transverse energy to the system. It can be quantified by an energy, denoted by  $E_T^{\text{miss}}$ , and an associated azimuthal angle. Adding this contribution to the system of all visible particles will balance the total vector sum to have vanishing transverse momentum.

The missing transverse energy in an event is calculated as the negative of the vector sum of the transverse momenta of all reconstructed, calibrated objects [78]. For the results in this dissertation, this sum includes electrons, muons and jets. An additional term is added to the sum to account for energy deposits not associated to any of these reconstructed objects. This term is built from charged particle tracks in the ID, which are assigned to the primary vertex, but not to any reconstructed objects. Overlap between the different physics objects is removed in the calculation of  $E_T^{\text{miss}}$ , in order to avoid double counting of contributions.

## 4.6 Overlap removal

After the reconstruction of the various objects described in this chapter, an overlap removal procedure is used to avoid double counting detector responses in the reconstruction of multiple objects. Such

double counting can happen for example when an electron showers and deposits energy in the electromagnetic calorimeter, and this deposit is also reconstructed as a jet.

The overlap removal employed for the results in this dissertation is briefly described here. If there is a jet candidate within  $\Delta R_\gamma = 0.2$  of an electron candidate, the closest jet to this electron is removed. In case there is another jet left within  $\Delta R_\gamma = 0.4$  of the electron candidate after this step, the electron candidate is removed as well. Muon candidates are required to not be within  $\Delta R_\gamma = 0.4$  of a jet candidate, or removed from the event otherwise. An exception to this treatment is used if the jet candidate has two or less tracks associated to it, in which case the jet candidate is removed and the muon candidate is kept instead. This accounts for muons losing a significant amount of energy in the calorimeter. Candidates for tau leptons decaying into hadronic final states are rejected if they are within  $\Delta R_\gamma = 0.2$  of an electron or muon.

## 5. Statistical methods

This chapter describes several statistical techniques used to interpret measurements at the LHC within a frequentist approach. The notion of probability in this approach refers to the relative frequency of an outcome of a repeatable experiment. In contrast, Bayesian statistics includes prior subjective knowledge to express probability density functions for parameters. The chapter starts with a description of the basic ingredients needed for statistical inference in section 5.1, followed by details regarding inference techniques in section 5.2. A brief introduction to two multivariate analysis techniques, BDTs and neural networks, is included in section 5.3. The first two sections in this chapter follow the overview of statistical techniques relevant to high energy physics provided in reference [10]. A summary of the procedures used for the Higgs boson discovery can be found in reference [79]. More information about a broad range of multivariate analysis techniques can be found in references [80, 81].

### 5.1 Statistical modeling

This section introduces the basic ingredients necessary for statistical inference.

#### 5.1.1 Random variables

A random variable is the outcome of a repeatable experiment. This outcome of an experiment is denoted as an observation  $x$ . Depending on the experiment, observations take on either discrete or continuous values. The continuous case is assumed in the following. Using the probability density function  $f(x; \alpha)$ , the probability for an observation to lie in the range between  $x$  and  $dx$  is given by  $f(x; \alpha) dx$ . The probability density function may depend on one or more additional parameters, denoted by  $\alpha$ . It is normalized to unity, such that the probability for an observation to take on any allowed value  $x$  is exactly one. The probability for an observation to take on a value  $x \leq b$  is given by the cumulative distribution function

$$F(b) = \int_{-\infty}^b f(x; \alpha) dx. \quad (5.1)$$

For any function  $u(x)$  of a random variable  $x$ , its expected value is given by

$$E[u(x)] = \int_{-\infty}^{\infty} u(x) f(x; \alpha) dx. \quad (5.2)$$

The mean of a probability density function is given by  $\mu = E[x]$ , and the variance by  $\sigma^2 = E[x^2] - \mu^2$ . The square root of the variance is the standard deviation  $\sigma$ .

Table 5.1: Probabilities  $P$  for Gaussian distributed observable  $x$  to fall within  $n$  standard deviations of the mean  $\mu$  in one experiment, and average amount of experiment repetitions needed for one observation to fall outside of this range.

| $n$ | $P$                   | $\frac{1}{1-P}$ |
|-----|-----------------------|-----------------|
| 1   | 0.683                 | 3.15            |
| 2   | 0.954                 | 22.0            |
| 3   | 0.997                 | 370             |
| 4   | $1 - 6 \cdot 10^{-5}$ | 15800           |
| 5   | $1 - 6 \cdot 10^{-7}$ | 1 740 000       |

If  $x$  and  $y$  are two random variables, then  $f(x, y; \alpha)$  is called the joint probability distribution function. The marginal probability density function of  $x$  is then given by

$$f_1(x; \alpha) = \int_{-\infty}^{\infty} f(x, y; \alpha) dy, \quad (5.3)$$

obtained via marginalization over  $y$ . The covariance of  $x$  and  $y$  is defined as

$$\text{cov}[x, y] = E[(x - \mu_x)(y - \mu_y)] = E[(xy)] - \mu_x \mu_y, \quad (5.4)$$

with  $\mu_x$  and  $\mu_y$  being the means of  $x$  and  $y$ , respectively. For  $x = y$ ,  $\text{cov}[x, x] = \sigma_x^2$ , with  $\sigma_x$  being the standard deviation for  $x$ . The correlation between two variables  $x$  and  $y$  is given by  $\text{cov}[x, y]/\sigma_x \sigma_y$ .

### 5.1.2 Common distributions

A few commonly used probability density functions are described in this section.

#### Poisson distribution

The Poisson distribution

$$\text{Poisson}(n; \nu) = \frac{\nu^n e^{-\nu}}{n!} \quad (5.5)$$

gives the probability to observe  $n$  events occurring independently in an interval, where the expected rate of events  $\nu$  is  $\nu > 0$  per interval. For this distribution,  $\sigma^2 = \nu$ .

#### Gaussian distribution

The Gaussian or normal distribution  $\mathcal{N}$  is given by

$$\mathcal{N}(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (5.6)$$

with mean  $\mu$  and variance  $\sigma^2$ . The probabilities  $P(x \in [\mu - n\sigma, \mu + n\sigma])$  for observation  $x$  to be within the range  $[\mu - n\sigma, \mu + n\sigma]$  in one experiment are given in table 5.1. The third column describes how many times on average an experiment would have to be repeated for an observable to fall outside of the range  $[\mu - n\sigma, \mu + n\sigma]$ .

### $\chi^2$ distribution

For  $n$  independent Gaussian random variables  $x_1, x_2, \dots, x_n$ , the variable  $z = \sum_{i=1}^n (x_i - \mu_i)^2 / \sigma_i^2$  is distributed like a  $\chi^2$  probability distribution function with  $n$  degrees of freedom. It is written as  $\chi^2(n)$  and given for  $z \geq 0$  as

$$f_{\chi^2}(z; n) = \frac{z^{\frac{n}{2}-1} e^{-\frac{z}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}. \quad (5.7)$$

The gamma function is  $\Gamma(n) = (n-1)!$  for integer  $n > 0$ . For  $n = 1$  degrees of freedom,  $f_{\chi^2}(z; 1) = e^{-z/2} / \sqrt{2\pi z}$ .

#### 5.1.3 Likelihood function

The expression  $L(\alpha) = P(x|\alpha)$  defines the likelihood function for a hypothesis  $\alpha$ , given an observation  $x$ . It specifies the probability to obtain an observation  $x$  under a specific hypothesis. This hypothesis  $\alpha$  is usually specified by a parameter of interest  $\mu$ , as well as nuisance parameters  $\theta$ ,  $\alpha = (\mu, \theta)$ . A typical choice for the parameter of interest is a signal strength  $\mu = \sigma_{\text{obs}} / \sigma_{\text{SM}}$ , given by the ratio of a measured cross-section to the prediction from the SM. Nuisance parameters  $\theta$  encode additional degrees of freedom in the likelihood, representing systematic uncertainties.

For an experiment measuring event counts across  $N$  different bins  $i$ , with the expected counts under hypothesis  $\alpha$  given by  $v_i(\alpha)$ , the likelihood of an observation  $x$  characterized by event counts  $x_i$  per bin is given by a product of Poisson terms,

$$L(\alpha) = P(x|\alpha) = \prod_{i=1}^N \text{Poisson}(x_i; v_i(\alpha)). \quad (5.8)$$

#### Systematic uncertainties

There are typically many different sources of systematic uncertainty which affect the expected counts in a bin  $v_i(\mu, \theta)$ . The parameters  $\theta$  describe these effects and can increase the uncertainties on the parameter of interest  $\mu$ . To reduce the impact of these uncertainties, statistically independent subsidiary measurements with data  $y$  can be used to build a joint model expressing the total likelihood for observations  $x$  and  $y$ , given all parameters  $\mu, \theta$ . This joint model is

$$L(\mu, \theta) = P(x|\mu, \theta) P(y|\theta). \quad (5.9)$$

The subsidiary measurements usually do not depend on the parameter of interest  $\mu$ .

In many practical applications, the subsidiary measurement is approximated by a model. Gaussian distributions  $\mathcal{N}$  are a common choice for this. Consider a subsidiary observation  $y_i$ , used to constrain a nuisance parameter  $\theta_i$ . Given an estimator for this nuisance parameter  $\hat{\theta}_i$  (which can be obtained by finding the parameter value maximizing the likelihood, described in section 5.2.1), and its standard deviation  $\hat{\sigma}_{\theta_i}$ , the subsidiary measurement can be approximated as

$$P(y_i|\theta_i) \rightarrow \mathcal{N}(\hat{\theta}_i; \theta_i, \hat{\sigma}_{\theta_i}). \quad (5.10)$$

Nuisance parameters are often re-defined for convenience, such that  $\hat{\theta} \equiv \theta_0 = 0$  and  $\hat{\sigma}_{\theta} \equiv \Delta\theta = 1$ .

## 5.2 Statistical inference

Depending on the scientific question examined, a range of different inference methods exist to gain insights from measured data. This section provides an overview of techniques relevant to the work in this dissertation.

### 5.2.1 Parameter estimation

An estimate of any parameter  $\alpha_i$  can be obtained via the method of maximum likelihood, by solving

$$\frac{\partial P(x|\alpha)}{\partial \alpha_i} = 0. \quad (5.11)$$

The estimators solving this set of equations are given by  $\hat{\alpha}$  and are called maximum likelihood estimators. They are unbiased: their expected value agrees with the true parameter value  $E[\hat{\alpha}_i] = \alpha_i$ .

An estimate for the covariance matrix  $V_{ij} = \text{cov}[\hat{\alpha}_i, \hat{\alpha}_j]$  is obtained from

$$(\hat{V}^{-1})_{ij} = -\left. \frac{\partial^2 P(x|\alpha)}{\partial \alpha_i \partial \alpha_j} \right|_{\hat{\alpha}}. \quad (5.12)$$

The estimate for the variance of a parameter  $\alpha_i$  is given at  $\hat{V}_{ii}$ .

So-called conditional maximum likelihood estimators are obtained when maximizing the likelihood for a given value of one of the parameters. The parameter values  $\hat{\theta}_\mu$  maximize the likelihood  $L(\alpha)$  with  $\alpha = (\mu, \theta)$  for a given setting of  $\mu$ .

Solutions to equation (5.11) and equation (5.12) are typically calculated numerically; the MINUIT software [82, 83] is used for the applications in this dissertation.

### 5.2.2 Hypothesis testing

In a hypothesis test, two different hypotheses  $H_0, H_1$  are compared with each other to determine whether the null hypothesis  $H_0$  can be rejected in favor of the alternative  $H_1$ . In a typical use case, the hypotheses are distinguished by a signal strength  $\mu = \mu^{\text{obs}}/\mu^{\text{SM}}$ , which is the ratio of a measured cross-section to the prediction from the SM. The null hypothesis specifies a signal strength  $\mu = 0$ , while the alternative hypothesis predicts a signal strength consistent with the SM,  $\mu = 1$ . The rejection of  $H_0$  is required to claim discovery of the signal process affected by  $\mu$ . As stated by the Neyman-Pearson lemma [84], the likelihood ratio

$$\lambda_{\text{NP}}(x) = \frac{f(x|H_1)}{f(x|H_0)} \quad (5.13)$$

maximizes the statistical power to reject  $H_0$  in favor of  $H_1$ . A scalar function of the data, such as the likelihood ratio described in equation (5.13), is called a test statistic  $t(x)$ . While  $\lambda_{\text{NP}}(x)$  is an optimal test statistic, it can only be used if the probability density functions  $f(x|H_i)$  can be evaluated. When this is not possible, common alternatives include the use of BDTs and neural networks (see section 5.3), and the matrix element method described in detail in chapter 7.

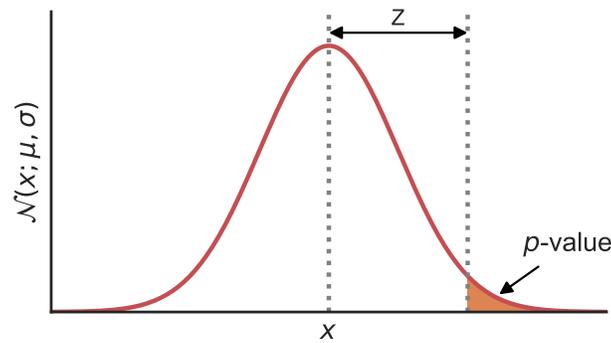


Figure 5.1: Relation between significance  $Z$  and  $p$ -value.

Given the probability density function  $f(t|H_0)$ , and assuming that larger values of  $t$  indicate increased discrepancy with  $H_0$ , the  $p$ -value

$$p = \int_{t_{\text{obs}}}^{\infty} f(t|H_0) dt \quad (5.14)$$

quantifies the level of discrepancy between the observed test statistic  $t_{\text{obs}}$ , calculated from measured data, and the expectation when assuming that  $H_0$  is true. When  $H_0$  is true, the  $p$ -value will be uniformly distributed in the interval  $[0, 1]$ . The  $p$ -value can be converted into the significance  $Z$  via

$$Z = \Phi^{-1}(1 - p), \quad (5.15)$$

where  $\Phi(x) = \int_{-\infty}^x \mathcal{N}(y; 0, 1) dy$  is the Gaussian cumulative distribution function and  $\Phi^{-1}$  is its inverse. The result of a hypothesis test comparing a hypothesis including a new particle and a null hypothesis without this particle present is called *evidence* in high energy physics if  $Z \geq 3$ , and *observation* of this particle for  $Z \geq 5$ . A threshold of  $p = 0.05$  is usually applied when performing a test to reject a hypothesis containing a new signal process in favor of a background-only hypothesis. The relation between significance and  $p$ -value is visualized in 5.1. In the presence of nuisance parameters, the  $p$ -value generally depends on those.

### The CL<sub>S</sub> method

Instead of working with the  $p$ -value directly, a common alternative for the derivation of limits is defined by the CL<sub>S</sub> method [85]. Let  $p_{\mu}$  be the  $p$ -value derived under a hypothesis specifying a signal with strength  $\mu$  that is being tested. It represents the probability to obtain an observed result equally or less compatible with the signal hypothesis than the observed one. Let  $p_0$  be the  $p$ -value describing the probability to obtain a result equally or less compatible with a background-only (no signal,  $\mu = 0$ ) hypothesis. The CL<sub>S</sub> method modifies  $p_{\mu}$  to determine whether the signal hypothesis may be rejected:

$$\text{CL}_S(\mu) = \frac{p_{\mu}}{1 - p_0}. \quad (5.16)$$

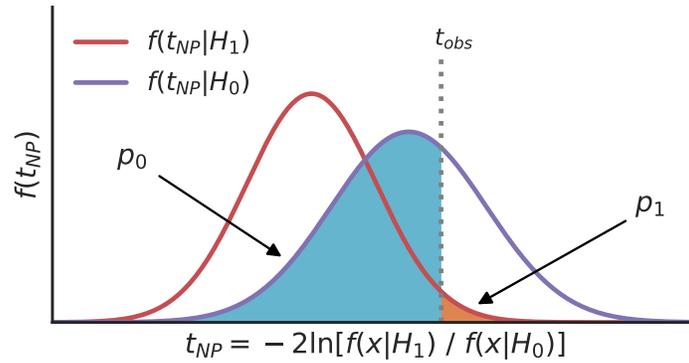


Figure 5.2: Distribution of the test statistic  $t_{NP}$  under hypotheses  $H_1$  and  $H_0$ , including  $p$ -values calculated from an observation  $t_{obs}$  indicated in the shaded areas.

In experiments with little sensitivity, the distributions of test statistics under signal and background-only hypotheses may overlap significantly. If the observed data fluctuates downwards compared to the expectation from the background-only hypothesis, the upper limit derived on  $\mu$  may be very low. For large  $p_0$ , the resulting value for  $CL_S(\mu)$  penalizes  $p_\mu$  to mitigate this effect. Models to which the test is not sensitive are therefore not excluded.

Figure 5.2 shows an example with a test statistic defined as  $t_{NP} = -2 \ln [f(x|H_1)/f(x|H_0)]$ . The test statistic distribution under hypothesis  $H_0$ , shown as  $f(t_{NP}|H_0)$ , is concentrated along higher values of  $t_{NP}$  than the distribution under hypothesis  $H_1$ . The  $p$ -value  $p_1$  calculated to reject hypothesis  $H_1$  in this example is around 2%. When calculating  $CL_S$ , it gets penalized by the large value of  $p_0$  and increases by roughly a factor of 3. It is not possible to reject  $H_1$  at the 95% confidence level in this case, as the sensitivity of the measurement is insufficient.

### Profile likelihood

The profile likelihood ratio is a test statistic defined as

$$\lambda_\mu = \frac{L(\mu, \hat{\theta}_\mu)}{L(\hat{\mu}, \hat{\theta})} \quad (5.17)$$

in order to dissolve the dependence on nuisance parameters  $\theta$ . The parameters  $\hat{\mu}, \hat{\theta}$  are maximum likelihood estimators, while  $\hat{\theta}_\mu$  is the conditional maximum likelihood estimator for a given  $\mu$ . For convenience, the test statistic

$$t_\mu = -2 \ln [\lambda_\mu], \quad (5.18)$$

defined as a function of the profile likelihood ratio, is commonly used. Increasing values of  $t_\mu$  correspond to larger discrepancies of the observed data with the hypothesis parameter setting  $\mu$ . Wilk's theorem [86] states that in the limit of sufficiently large data samples, and observations generated with a signal strength parameter  $\mu'$ ,  $t_\mu$  is distributed like a  $\chi^2$  distribution with as many degrees of

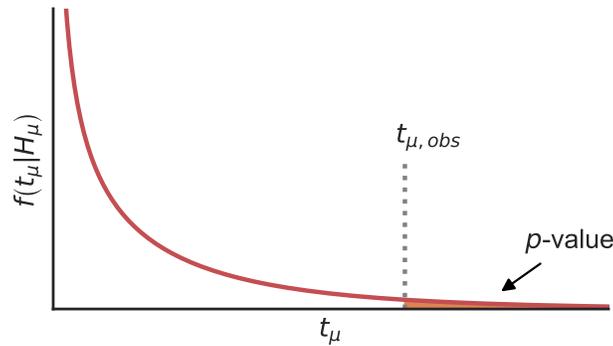


Figure 5.3: Distribution of the test statistic  $t_\mu$ , the  $p$ -value can be obtained via the integral prescription in equation (5.14).

freedom as dimensions in  $\mu$  for  $\mu' = \mu$ . An example with one degree of freedom is shown in 5.3. When testing for other settings  $\mu' \neq \mu$ , the test statistic  $t_\mu$  follows a non-central  $\chi^2$  distribution.

### Discovery test statistic

The test statistic

$$q_0 = \begin{cases} t_0 = -2 \ln [\lambda_0] & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0 \end{cases} \quad (5.19)$$

is used to test for the discovery of a new signal by rejecting the  $\mu = 0$  hypothesis. This is assuming  $\mu \geq 0$ , and a data fluctuation resulting in  $\hat{\mu} < 0$  is not interpreted as evidence for a signal. Instead, only increasing values of  $\hat{\mu} > 0$  result in increasingly large values of  $q_0$  and thereby an increased incompatibility with the null hypothesis  $\mu = 0$ . The discovery significance  $Z_0$  is given by [87]

$$Z_0 = \sqrt{q_0}. \quad (5.20)$$

### Test statistic for upper limits

For testing the upper limit on a signal strength parameter, the test statistic

$$\tilde{q}_\mu = \begin{cases} \tilde{t}_\mu = -2 \ln [\tilde{\lambda}_\mu] & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu \end{cases} \quad (5.21)$$

is defined. The test statistic  $\tilde{\lambda}_\mu$  is equivalent to  $\lambda_\mu$  for  $\hat{\mu} \geq 0$ . For  $\hat{\mu} < 0$ , it is given by  $\tilde{\lambda}_\mu = L(\mu, \hat{\theta}_\mu) / L(0, \hat{\theta}_0)$ . Fluctuations in a measurement resulting in  $\hat{\mu} > \mu$  are not regarded as making the observed data less compatible with the hypothesis signal.

### 5.2.3 Median significances and the Asimov dataset

It can be very expensive computationally to build the probability density functions  $f(t_\mu | H_\mu)$  for test statistics. This procedure relies on repeatedly generating datasets  $x$  distributed according to

each hypothesis  $H_\mu$  that needs to be tested, and evaluating the test statistic for each of them. When evaluating the performance of an experiment, the median discovery significance for a given signal process can be evaluated by constructing the distribution  $f(q_0|\mu=0)$ , and evaluating the median test statistic  $\bar{q}_0$  across many different simulated datasets distributed under the signal hypothesis. The median significance with which a background-only hypothesis is expected to be excluded is obtained by calculating the  $p$ -value of this median  $\bar{q}_0$  via equation (5.14) and converting it to a significance. Similarly, expected median upper limits can be evaluated after constructing  $f(\bar{q}_\mu|\mu=0)$ , by finding the value of  $\mu$  for which the median  $p$ -value is 0.05. The value obtained is the median upper limit for  $\mu$  at the 95% confidence level.

### Asimov dataset

Analytic approximations for the distribution of the test statistics mentioned above exist in the large sample limit [87]. The method to obtain them uses the Asimov dataset, which is defined in such a way that the estimators for all parameters  $\alpha = (\mu, \theta)$  obtained on this dataset correspond to their true values. Let  $\mu'$  be the signal strength parameter used in the generation of the Asimov dataset, and  $\theta$  the nuisance parameters. The profile likelihood ratio evaluated on this dataset is given by

$$\lambda_{\mu,A} = \frac{L_A(\mu, \hat{\theta})}{L_A(\mu', \theta)}, \quad (5.22)$$

where  $L_A$  is the likelihood of the Asimov dataset. With a test statistic defined as  $q_{\mu,A} = -2\ln[\lambda_{\mu,A}]$ , the variance of  $\hat{\mu}$  can be calculated:

$$\sigma_A^2 = \frac{(\mu - \mu')^2}{q_{\mu,A}}. \quad (5.23)$$

In a test for discovery,  $\mu = 0$  to exclude a background-only hypothesis. When calculating the median exclusion significance for hypothesis  $\mu$  under the assumption that no signal exists,  $\mu' = 0$ .

The Asimov dataset generated with signal strength parameter  $\mu'$  can furthermore be used to estimate median significances. Assuming a signal strength  $\mu'$ , the discovery significance is given by  $\text{med}(Z_0|\mu') = \sqrt{q_{0,A}}$ . The median exclusion significance for a signal strength  $\mu$ , assuming true signal strength  $\mu' = 0$ , is  $\text{med}(Z_\mu|\mu' = 0) = \sqrt{\bar{q}_{\mu,A}}$ .

## 5.3 Multivariate techniques

This section provides a brief introduction into two multivariate techniques relevant to this dissertation, BDTs and neural networks. The description is based on references [80, 81].

### 5.3.1 Boosted decision trees

BDTs are a common multivariate technique in high energy physics. They are used in many places, for example in the identification of objects from collision remnants in the detector, and as discriminants to distinguish between events originating from various processes in physics analyses.

### Decision trees

Let  $\vec{X} = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)$  be a set of data, where each data point  $i$  is described by a set of features  $\vec{x}_i = (x_i^1, x_i^2, \dots, x_i^M)$ , as well as a label (or set of labels)  $z_i$ . The feature set  $\vec{x}_i$  corresponds to observable information, such as kinematics of collision remnants reconstructed in a detector. The label  $z_i$  is not observable, but here it is assumed that it is possible to generate a simulated set of events  $(\vec{x}_i, z_i)$ . Discrete labels are used to separate events into different classes. In the context of high energy physics, each event may be assigned for example a label of  $z_i = 0$  if no Higgs boson is produced in the final state of the event, and a label  $z_i = 1$  otherwise. Continuous labels can also be used.

Decision trees provide an approximate model  $\hat{z}_i = \hat{f}(\vec{x}_i)$  to assign a label to a set of features for each data point  $i = 1..N$ . They partition the feature space into hyperrectangles, with a label  $\hat{z}_i$  assigned to each hyperrectangle. Only binary partitioning, which is most commonly used, is considered in the following; decision trees can then be visualized as binary trees. At the root of the tree, the full feature space has not been partitioned yet. The space is then recursively split into hyperrectangles, by applying cuts along hyperplanes, until a stopping criterion has been reached. Each split decision is indicated by a node, the terminal nodes are also called leaves. Typical stopping criteria are a maximum tree depth and a minimum amount of simulated events remaining in a node during construction of the tree.

### Constructing decision trees

The construction of decision trees is done by recursively determining and applying the best possible next cut, according to some metric. Each cut splits a hyperrectangle into two. For classification problems with discrete labels  $k = 1, 2, \dots, K$ , the Gini index is given by

$$G = \sum_{k=1}^K p_k (1 - p_k), \quad (5.24)$$

where  $p_k$  is the fraction of events in class  $k$  at a given node. The index is bounded by  $\max G = 1 - \frac{1}{K}$ , reached if an equal fraction of events from all classes is present at the node, and  $\min G = 0$ , reached if only events from one class are present. Cut decisions are taken such that the sum of Gini indices of the child nodes, weighted by the relative amount of events contained in them, is minimized.

### Boosting

Boosting is a powerful method to extend the performance of decision trees for classification problems. It relies on constructing a weak classifier, which performs just slightly better than pure guessing, and iterative applications of this classifier on modified versions of the data. After each step, events that were misclassified by the latest classifier receive larger weights, while correctly classified events receive smaller weights. The final model is then obtained via a weighted average of the individual classifiers.

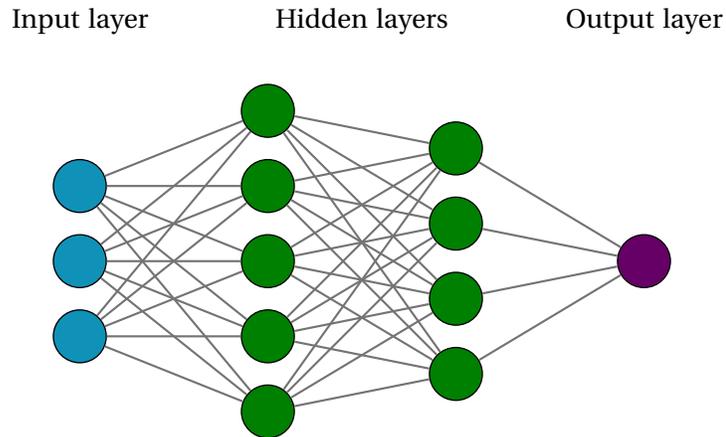


Figure 5.4: Exemplary architecture of a fully connected feedforward neural network with three inputs (drawn as blue circles), two hidden layers (with associated nodes drawn in green), and one output (drawn in purple). Information flows along the lines connecting nodes.

### 5.3.2 Neural networks

The term neural network is used for a wide range of machine learning methods. The focus in this section is on fully connected feedforward artificial neural networks; this architecture is used in chapter 11. Such a network has the goal of approximating a function  $f(\vec{x})$ . The term neural network originates from its use as a model for the human brain, describing neurons connected via synapses.

A neural network consists of layers of nodes. Figure 5.4 visualizes an example. The nodes of the input layer, drawn as blue circles, correspond to features  $\vec{x}$  provided to the neural network, with one feature per node. Nodes in all the remaining layers calculate a derived feature from the combination of all nodes in the respective previous layer. The inputs to each node are visualized via lines in the figure. Each node performs a linear combination of its inputs, weighted by a vector  $\vec{\alpha}$ , and also allows for the addition of a bias term  $\alpha_0$ . This bias term can be thought of as an additional node in the previous layer with a constant value. The output of a node  $y^j$  is given by

$$y^j = \sigma(\vec{\alpha}^j \cdot \vec{r} + \alpha_0^j), \quad (5.25)$$

where  $\sigma(v)$  is called an activation function. The values for  $\vec{\alpha}^j$  and  $\alpha_0^j$  are learned during the training of the neural network. The vector  $\vec{r}$  contains the output of all nodes in the previous layer. When calculating the output of a node in the first hidden layer, it is equivalent to the network inputs,  $\vec{r} = \vec{x}$ .

The term feedforward to describe this architecture refers to the flow of information from the input nodes forward towards the output of the network. It is fully connected, as each node receives inputs from all nodes in the previous layer.

The hidden layers receive their name since their associated node values are not observed, and only act as intermediate steps in the calculation of the network output. The output of the network is given in the final layer, the output layer. It may have more than one node, for example for a network

designed for a classification problem. When using the network in a regression problem, one output node is typically used.

### Loss function and training

The weights  $\theta$  in the network are learned during a training procedure, which minimizes a loss function  $L_\theta$ . These weights contain  $\vec{\alpha}^j, \alpha_0^j$ , and can also contain additional parameters affecting  $L_\theta$ . The loss function measures the performance of the network. The regression task studied in section 11.3 uses the mean absolute error between network output and true function value as loss function, defined as

$$L_\theta = \frac{1}{N} \sum_{i=1}^N |f(\vec{x}_i) - g_\theta(\vec{x}_i)|. \quad (5.26)$$

The network output is given by  $g_\theta(\vec{x}_i)$ , it depends on the set of weights  $\theta$ . This loss function is evaluated by considering a set of  $N$  events, calculating the absolute error in the network output for each event, and averaging the results. The minimization of the loss function  $L_\theta$  can be performed with gradient descent methods. During this iterative process, updates to the network weights  $\theta$  are calculated via the chain rule of calculus, minimizing the value of  $L_\theta$  evaluated with a set of training events.

### Activation functions

For  $\sigma(v) = v$ , the neural network is a linear model of its inputs. The use of other activation functions introduces non-linearity to the neural network, and thus allows the network to describe non-linear functions of its inputs. Two types of activation functions are used in this dissertation, they are called rectified linear unit (ReLU) and softplus. The ReLU function is defined by

$$\sigma(v) = \begin{cases} 0 & v < 0, \\ v & v \geq 0. \end{cases} \quad (5.27)$$

Its output is zero for negative inputs, and equal to its input otherwise. The softplus function is

$$\sigma(v) = \log(e^v + 1), \quad (5.28)$$

with positive output. It approaches the ReLU function for both very small and very large input values.

### Architecture and hyperparameters

The performance of a neural network not only depends on learning the weights  $\theta$ , but also on parameters of the model that are not learned. This includes the architecture of the model, specified by the amount of hidden layers and nodes per layer, and the activation functions used. The set of inputs to the network can be tuned, as can the choice of algorithm performing the loss function minimization, and the hyperparameters associated to this algorithm.

## 6. Search for Higgs boson production in association with a top quark pair and decaying into a bottom quark pair

This chapter summarizes a search for  $t\bar{t}H$  production with Higgs boson decays to bottom quark pairs with the ATLAS detector, performed with  $36.1 \text{ fb}^{-1}$  of data collected in 2015 and 2016 at  $\sqrt{s} = 13 \text{ TeV}$  during Run-2 of the LHC. The result of the search was published in 2018, measuring a signal strength of  $\mu_{t\bar{t}H} = 0.84_{-0.61}^{+0.64}$  [1]. This signal strength is defined as the ratio of the measured cross-section to the cross-section predicted by the SM, as  $\mu_{t\bar{t}H} = \sigma_{t\bar{t}H}^{\text{obs}}/\sigma_{t\bar{t}H}^{\text{SM}}$ .

A measurement of the Higgs boson production process in association with top quark pairs,  $t\bar{t}H$ , provides a direct probe of the Yukawa sector of the SM. This production process is sensitive to the top quark Yukawa coupling  $y_t$ , a parameter with implications exceeding particle physics, as discussed in section 2.4.1. Section 2.5.3 describes the role of the  $t\bar{t}H$  process in determining this coupling in more detail. The determination of  $y_t$  via loop-induced couplings of gluons or photons to the Higgs boson relies on assumptions about BSM particles contributing to these loops in the Higgs boson production and decay. In contrast to this, the tree-level measurement of  $y_t$  via the more rare  $t\bar{t}H$  process does not rely on such assumptions.

The cross-section of  $t\bar{t}H$  is proportional to the square of the top quark Yukawa coupling. A measurement of the  $t\bar{t}H$  cross-section can therefore be interpreted in terms of  $y_t$ .

A range of  $t\bar{t}H$  analyses were conducted prior to the search described in this chapter. During Run-1 of the LHC, both the ATLAS and CMS collaborations performed dedicated searches for  $t\bar{t}H$  with three different Higgs boson decay topologies. Decays to bottom quark pairs, photon pairs, and final states with multiple charged electrons or muons (via Higgs boson decays to weak gauge bosons and tau leptons) were analyzed. An ATLAS analysis of the  $t\bar{t}H(b\bar{b})$  topology using  $20.3 \text{ fb}^{-1}$  of Run-1 data measured a signal strength of  $\mu_{t\bar{t}H} = 1.5 \pm 1.1$  [88]. The combination of various  $t\bar{t}H$  final states analyzed with Run-1 data from ATLAS and CMS resulted in a  $t\bar{t}H$  signal strength measurement of  $\mu_{t\bar{t}H} = 2.3_{-0.6}^{+0.7}$  [89].

This chapter is organized as follows. Section 6.1 provides a brief overview of the analysis approach and challenges. The definitions of the objects used in the analysis and the basic event selection are described in section 6.2. The expected kinematic distributions of events produced via the  $t\bar{t}H$  process and background processes are obtained from simulation and a data-driven technique; details about this are provided in section 6.3. Section 6.4 describes the categorization of events into different regions. The multivariate analysis techniques employed to distinguish between the  $t\bar{t}H$  signal and background processes are presented in section 6.5. Systematic uncertainties affecting the  $t\bar{t}H(b\bar{b})$

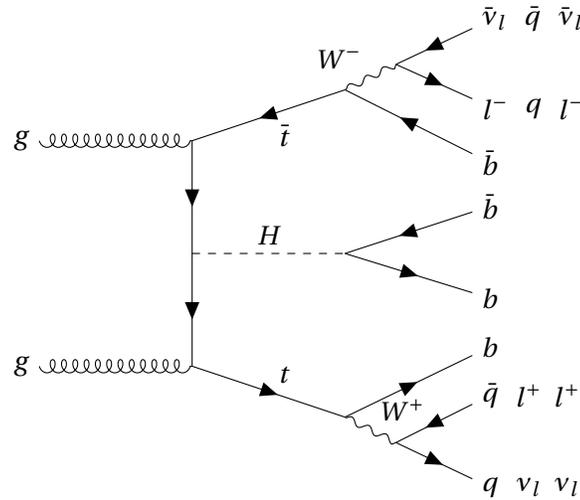


Figure 6.1: Exemplary Feynman diagram for the  $t\bar{t}H(b\bar{b})$  topology, with one or two light charged leptons ( $l$ ) in the final state. The different columns listed for the decay products of the  $W$  bosons correspond to the alternative topologies considered in the analysis.

search are listed in section 6.6. Lastly, section 6.7 presents the results of the statistical analysis.

## 6.1 Analysis overview

The analysis presented in this chapter targets the  $t\bar{t}H(b\bar{b})$  topology, with one or two light charged leptons (electrons or muons) originating from the  $W$  bosons produced in the top quark pair decay. Figure 6.1 shows one of the Feynman diagrams for this topology. Two channels are considered in the analysis, with events assigned to them depending on the amount of reconstructed light charged leptons. Events with one reconstructed light charged lepton are analyzed in the single-lepton channel, while those with two light charged leptons are contained in the dilepton channel. Within both of these channels, several regions are defined depending on the number of reconstructed jets and  $b$ -tagging information. The single-lepton channel also contains a dedicated region targeting the decay of top quarks with high transverse momentum. Given that  $W$  bosons decay hadronically with a branching ratio of roughly  $2/3$  [10], the single-lepton channel contains more events than the dilepton channel.

The signal extraction is performed via a combined profile likelihood fit to 19 non-overlapping regions across both channels. In 10 of these regions, a small fraction of signal events (less than 1.5% of the total amount of events) is expected, and they mostly serve to constrain systematic uncertainties associated to background modeling. The remaining 9 regions have larger contributions of the  $t\bar{t}H(b\bar{b})$  signal, and a variety of multivariate techniques are employed there to discriminate between  $t\bar{t}H(b\bar{b})$  and other processes.

While the analysis is designed for  $H \rightarrow b\bar{b}$  decays, all SM Higgs boson decay modes are considered as signal. The contributions from other decay modes are small, they make up around 1–4% of the total signal in the most sensitive signal regions.

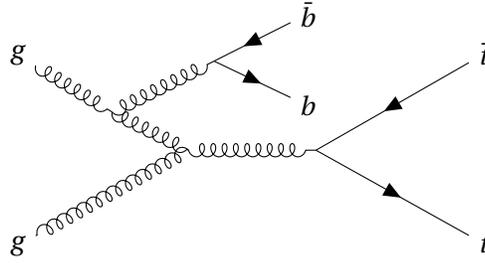


Figure 6.2: Exemplary Feynman diagram for the  $t\bar{t} + b\bar{b}$  background process.

The presence of at least one light charged lepton in the final state allows for an efficient way to select events for the analyses, using electron and muon triggers. Requiring at least one light charged lepton to be present also suppresses background contributions from QCD multi-jet production. A dedicated analysis for  $t\bar{t}H(b\bar{b})$  with a fully hadronic final state was performed by ATLAS in Run-1 of the LHC [90]. The ATLAS analysis of Run-2 data in this final state is ongoing; the analysis design differs significantly from the analysis presented in this chapter in order to deal with the multi-jet background.

### Main background: $t\bar{t} + b\bar{b}$

The largest experimental challenge in the analysis arises from the modeling of top quark pair production with additional  $b$ -jets, called  $t\bar{t} + \geq 1b$ . A subset of  $t\bar{t} + \geq 1b$ , the  $t\bar{t} + b\bar{b}$  background, arises from top quark pair production with an additional emitted gluon splitting to a bottom quark pair. An exemplary Feynman diagram for this signature is shown in figure 6.2. The correct description of this process is difficult, and the large uncertainties associated to the predicted distribution of  $t\bar{t} + b\bar{b}$  events in the analysis limit the overall sensitivity.

### Signal–background discrimination

Further challenges arise from the similarity of the  $t\bar{t}H(b\bar{b})$  signal to other background processes. In particular, the final states of  $t\bar{t} + b\bar{b}$  and  $t\bar{t}H(b\bar{b})$  contain the same partons. A successful discrimination between these two processes relies on small differences in kinematic distributions.

The invariant mass distribution of the  $b\bar{b}$  system from the decay of the Higgs boson is sharply peaked around the Higgs boson mass, while the invariant mass of the system produced from an emitted gluon follows a more broad distribution. Reconstruction of the Higgs boson invariant mass suffers from combinatorial ambiguity, since additional  $b$ -jets from the top quark decays are present in every event. The jets originating from the Higgs boson decay may also not have been  $b$ -tagged, or not have been reconstructed. Additional jets may have been  $b$ -tagged mistakenly. Due to the large amount of objects expected in the final state, the efficiency to reconstruct and identify all correctly is low. Even for the cases where all objects were correctly identified, the finite ATLAS detector resolution

considerably widens the invariant mass peak expected from the  $b$ -jet system from the Higgs boson decay.

Additional small differences between  $t\bar{t} + b\bar{b}$  and  $t\bar{t}H(b\bar{b})$  are expected, and diluted by detector effects. They are also affected by combinatorial ambiguity; in order to take advantage of some of these effects, jets need to be matched to partons. The angular distribution of the  $b\bar{b}$  system originating from a Higgs boson decay in the Higgs boson frame differs from the corresponding distribution of a gluon. This difference is due to the spin-0 nature of the SM Higgs boson, compared to the spin-1 gluon.

The analysis uses a range of multivariate techniques to perform system reconstruction and discrimination of  $t\bar{t}H$  from the background processes present. These are described in section 6.5.

## 6.2 Event selection

This section summarizes the requirements on events to be considered in the analysis. The analyzed dataset is briefly described in section 6.2.1. Section 6.2.2 lists additional details about the object definitions used. If no further details are given, the object definitions follow the description in chapter 4. Lastly, section 6.2.3 specifies how the single-lepton and dilepton channels are defined, and which events they contain.

### 6.2.1 Dataset

Events considered in this analysis are taken from proton–proton collisions at  $\sqrt{s} = 13$  TeV, delivered in 2015 and 2016 by the LHC, and recorded by the ATLAS detector. All events are required to fulfill the quality criteria listed in section 3.2.7. The dataset corresponds to an integrated luminosity of  $36.1 \pm 0.8 \text{ fb}^{-1}$ ; the uncertainty is derived with a method similar to reference [50]. The mean number of interactions per bunch crossing in this dataset is 24, with a distribution ranging from around 8 to 45 interactions. Figure 3.2 shows these distributions, but also includes data recorded by ATLAS that does not fulfill the quality criteria for physics analyses.

### 6.2.2 Object definitions

Electrons are required to have  $p_T > 10$  GeV and be reconstructed within  $|\eta| < 2.47$ . They are removed if they fall into the transition region between the calorimeter barrel and end-cap, located at  $1.37 < |\eta| < 1.52$ . Electrons need to satisfy the *LooseAndBLayer* identification operating point, and the *loose* isolation operating point.

Muons have the same transverse momentum requirement of  $p_T > 10$  GeV, and need to be located within  $|\eta| < 2.5$ . They have to satisfy the *loose* identification operating point, as well as the *loose* isolation operating point.

Jets are required to have  $p_T > 25$  GeV and be located within  $|\eta| < 2.5$  after their calibration. The boosted region within the single-lepton channel uses an additional jet definition, so-called large- $R$  jets. It targets the decay of objects with high momenta, where the decay products are collimated and

not resolved as individual jets. The selected standard jets are re-clustered [91] into large- $R$  jets using the anti- $k_t$  algorithm with  $R = 1.0$ . Only large- $R$  jets with  $p_T > 50$  GeV are considered.

The overlap removal procedure described in section 4.6 is applied on these objects, using the standard  $R = 0.4$  jets, not the large- $R$  jets. After the overlap removal, the lepton requirements are tightened further. Electrons need to pass the *tight* identification operating point, while muons need to fulfill the *medium* identification operating point. Both also need to pass the respective *gradient* isolation operating point requirements.

### 6.2.3 Definition of the single-lepton and dilepton channels

All events for this analysis were recorded with single light lepton (electron and muon) triggers. These triggers are highly efficient above their thresholds, and events are required to pass either a trigger with a lower transverse momentum threshold and a lepton isolation requirement, or a trigger with higher threshold and no isolation requirement. For muons, the transverse momentum thresholds of the triggers used are 20 GeV and 26 GeV for data recorded in 2015 and 2016, respectively, with additional isolation requirements applied. Without the isolation requirements, the threshold is 50 GeV. The lowest electron trigger thresholds are 24 GeV and 26 GeV with isolation requirements. Additional electron triggers with thresholds of 60 GeV and 120 GeV (for data recorded in 2015) and 140 GeV (for 2016), using increasingly relaxed identification criteria, are also used. Events considered in this analysis are required to pass any of these triggers. They also need to contain a reconstructed lepton with  $p_T > 27$  GeV, which matches the lepton reconstructed by the trigger, defined by proximity in  $\Delta R$ .

#### Single-lepton channel

The single-lepton channel selects events with at least five jets and exactly one reconstructed light lepton, no other light leptons with  $p_T > 10$  GeV must be present. Events with more than one hadronic tau lepton are removed. This requirement avoids the selection of events used in other searches for  $t\bar{t}H$  with different Higgs boson final states.

The boosted region within the single-lepton channel targets events with at least one top quark produced at high transverse momentum. Higgs boson candidates are defined as large- $R$  jets with  $p_T > 200$  GeV, which contain at least two jets. At least two of the contained jets need to be  $b$ -tagged at the *loose* operating point. Top quark candidates are formed by large- $R$  jets with  $p_T > 250$  GeV, containing at least two jets, out of which exactly one passes the *loose*  $b$ -tagging operating point. Events in the boosted region contain at least one Higgs boson candidate, at least one top quark candidate, as well as an additional jet  $b$ -tagged at the *loose* operating point. The  $b$ -tagging requirements are less stringent than in the resolved regions to retain a sufficient  $t\bar{t}H$  selection efficiency.

Events which do not fall into the boosted region are instead considered for the remaining regions of the single-lepton channel, called the resolved regions. Among the five required jets, they need to have at least two jets passing the *very tight*  $b$ -tagging operating point, or at least three passing the *medium* operating point.

## Dilepton channel

The dilepton channel requires events to have two reconstructed light leptons with opposite electric charge. For events with two electrons, the lepton with lower transverse momentum needs to satisfy  $p_T > 15$  GeV, while the threshold is  $p_T > 10$  GeV for events with at least one muon. If both leptons have the same flavor, their invariant mass is required to be higher than 15 GeV and not within the 83–99 GeV range. The latter requirement suppresses events originating from  $Z$  boson decays. At least three jets are required in the dilepton channel, and at least two of them need to be  $b$ -tagged at the *medium* operating point. Events with at least one hadronic tau lepton are removed.

## 6.3 Modeling

The expected contributions to the analysis from various processes are modeled mostly with Monte Carlo (MC) simulation; the contribution of fake and non-prompt leptons in the single-lepton channel is estimated from data. The GEANT4-based full simulation of the ATLAS detector is used for the majority of MC samples. Some samples used to build templates for estimating systematic uncertainties are instead simulated with the AFII method. The simulation of ATLAS is described in reference [57]. Pile-up interactions are simulated with PYTHIA 8.186 [92], and simulated events are reweighted to correspond to the pile-up profile in data. EVTGEN 1.2.0 [93] is used to decay  $b$ - and  $c$ -hadrons for all samples except those produced with SHERPA [94]. The top quark mass in all simulations is set to 172.5 GeV.

### 6.3.1 $t\bar{t}H$ signal

Samples for the expected  $t\bar{t}H$  signal distributions are produced with the MADGRAPH5\_AMC@NLO [95] generator in version 2.3.2, performing the matrix element calculation at NLO in QCD. The term MG5\_AMC@NLO will be used in the following to refer to this event generator. The parton distribution functions are provided by the NNPDF3.0NLO [96] set. Both renormalization and factorization scales are set to  $\mu_R = \mu_F = 0.5 H_T$ , with the  $H_T$  variable defined as the sum of the transverse masses  $\sqrt{p_T^2 + m^2}$  of all final state particles per event. Parton showering and hadronization are performed by PYTHIA 8.210 [97], with free model parameters set to the A14 tune [98]. The Higgs boson mass in the simulation is set to 125 GeV, and its branching ratios are calculated with HDECAY [34, 99]. The production cross-section for  $t\bar{t}H$  is  $507_{-50}^{+35}$  fb [34], calculated at NLO accuracy in QCD and including NLO EW corrections.

### 6.3.2 $t\bar{t}$ + jets background

Top quark pair production, the  $t\bar{t}$  process, is the dominant background in this analysis. It is modeled with the POWHEG-BOX v2 event generator [100–103] at NLO using the NNPDF3.0NLO PDF set in the five-flavor (5F) scheme. This event generator will be referred to as POWHEG. The setup is tuned to describe data in a more inclusive phase space than the analysis in this chapter is using [104].

Table 6.1: Definition of the  $t\bar{t}$  + jets components used in the analysis. Additional particle jets are those not originating from a top quark or  $W$  boson decay.

| $t\bar{t}$ + jets component | definition  |
|-----------------------------|---|
| $t\bar{t} + \geq 1b$        | $\geq 1$ additional particle jets matched to $\geq 1$ $b$ -hadrons                                |
| $t\bar{t} + b$              | one additional particle jet matched to one $b$ -hadron  |
| $t\bar{t} + b\bar{b}$       | two additional particle jets matched to one $b$ -hadron each                                      |
| $t\bar{t} + B$              | one additional particle jet matched to two or more $b$ -hadrons                                   |
| $t\bar{t} + \geq 3b$        | remaining $t\bar{t} + \geq 1b$ events, excluding $t\bar{t} + b$ (MPI/FSR)                         |
| $t\bar{t} + b$ (MPI/FSR)    | additional particle jet from MPI and FSR  |
| $t\bar{t} + \geq 1c$        | not $t\bar{t} + \geq 1b$ , and $\geq 1$ additional particle jets matched to $\geq 1$ $c$ -hadrons |
| $t\bar{t} + \text{light}$   | neither $t\bar{t} + \geq 1b$ nor $t\bar{t} + \geq 1c$   |

Renormalization and factorization scales are set to the transverse mass  $\sqrt{p_{\perp}^2 + m^2}$  of the top quark, evaluated in the reference frame where the  $t\bar{t}$  center of mass is at rest. PYTHIA 8.210 with the same A14 tune is used for parton showering and hadronization. The production cross-section for  $t\bar{t}$  is  $832_{-51}^{+46}$  pb. It is evaluated with Top++2.0 [105] at NNLO in QCD, including next-to-next-to-leading logarithmic (NNLL) corrections [106–109].

### $t\bar{t}$ + jets classification into components

The  $t\bar{t}$  + jets background is split into multiple components. In order to perform the split, so-called particle jets are built with the anti- $k_t$  algorithm with  $R = 0.4$ , using stable particles (with mean lifetimes  $\tau > 3 \cdot 10^{-11}$  s) from the MC simulation as input. The number of  $b$ - and  $c$ -hadrons within  $\Delta R < 0.4$  is then counted for each particle jet. An event is classified as  $t\bar{t} + \geq 1b$  if it has at least one particle jet containing at least one  $b$ -hadron, and this jet does not originate from a top quark or  $W$  boson decay. All of the remaining events are classified as  $t\bar{t} + \geq 1c$  if they have at least one particle jet containing at least one  $c$ -hadron, and this jet does not originate from a  $W$  boson decay. The rest of the events are labeled  $t\bar{t} + \text{light}$ ; the name refers to the smaller masses of the  $u$ ,  $d$  and  $s$  quarks compared to the  $c$  and  $b$  quarks.

The  $t\bar{t} + \geq 1b$  component is split further into sub-components. For the following definitions, again only particle jets not originating from a top quark or  $W$  boson decay are considered. Events containing a single particle jet with exactly one  $b$ -hadron matched to it are labeled as  $t\bar{t} + b$ . If they contain exactly two particle jets with exactly one  $b$ -hadron matched to each, they are labeled as  $t\bar{t} + b\bar{b}$ . Events with a single particle jet containing two or more  $b$ -hadrons are called  $t\bar{t} + B$ . The remaining events are categorized as  $t\bar{t} + \geq 3b$ . A special category exists for events containing  $b$ -jets originating from multi-parton interaction (MPI) and final state radiation (FSR), which is gluon radiation from top quark decay products. This affects 10% of the events, and such events are categorized as  $t\bar{t} + b$  (MPI/FSR) instead. Table 6.1 summarizes the different  $t\bar{t}$  + jets components used in the analysis.

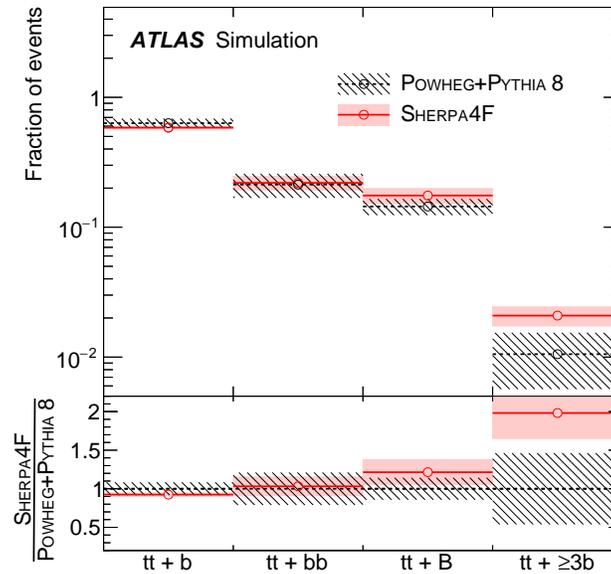


Figure 6.3: Relative fraction of  $t\bar{t} + \geq 1b$  sub-components predicted by the POWHEG+PYTHIA 8 and SHERPA4F samples. The uncertainty for both predictions are also shown, including the sources discussed in section 6.6.3 [1].

### Reweighting

A MC sample for  $t\bar{t} + b\bar{b}$  is produced with SHERPA+OPENLOOPS [94, 110, 111], describing the production of the two additional  $b$ -jets at NLO precision. SHERPA is used in version 2.1.1, with the CT10 PDF set [16, 112] in the four-flavor (4F) scheme. This sample will be referred to as SHERPA4F. As it describes the two additional  $b$ -jets at NLO and takes into account the  $b$  quark mass, it is the most precise theoretical prediction for the  $t\bar{t} + \geq 1b$  process available for the analysis. In the nominal POWHEG+PYTHIA 8  $t\bar{t}$  sample, additional  $b$ -jets come from the parton shower (PS).

The relative contributions from the various  $t\bar{t} + \geq 1b$  sub-components in the POWHEG+PYTHIA 8 sample are reweighted to match the distribution of the SHERPA4F sample. The  $t\bar{t} + b$  (MPI/FSR) component, which is not included in the SHERPA4F prediction, is unaffected by this reweighting. Figure 6.3 shows the fraction of events in the relevant  $t\bar{t} + \geq 1b$  sub-components for both samples. The uncertainty for the POWHEG+PYTHIA 8 prediction is obtained from the modeling uncertainties discussed in section 6.6.3, but not including SHERPA4F related uncertainties. The uncertainty for the SHERPA4F prediction originates from the sources affecting this prediction directly, as discussed in section 6.6.3.

### 6.3.3 Other backgrounds

The remaining background processes have a smaller impact on the analysis. All of the processes described in the following, with the exception of  $t\bar{t}V$ , are collectively referred to as non- $t\bar{t}$ . This group

of non- $t\bar{t}$  processes contributes 4–15% across the regions considered.

### Additional simulated backgrounds

The simulation of weak vector boson production with additional jets,  $V$ +jets, is performed with SHERPA 2.2.1 and the NNPDF3.0NNLO PDF set. It allows for the generation of up to two additional partons at NLO, and four at LO [113–115].  $Z$ +jets events containing  $b$ - or  $c$ -jets are weighted by a factor 1.3 to match the distribution of data within a control region around the  $Z$  boson mass window. SHERPA 2.2.1 is also used to simulate diboson (two weak vector boson) production [116].

The production of top quark pairs with additional weak vector bosons,  $t\bar{t}W$  and  $t\bar{t}Z$ , is modeled at NLO with MG5\_AMC@NLO and the NNPDF3.0NLO PDF set. For these  $t\bar{t}V$  samples, PYTHIA 8.210 with the A14 tune is used for parton showering and hadronization. This setup corresponds to the treatment for  $t\bar{t}H$ .

Five different samples are generated for single top quark production. They describe  $s$ - and  $t$ -channel single top quark production, as well as  $Wt$ ,  $tZ$  and  $tWZ$  topologies. Three of these processes are modeled at NLO with POWHEG-BOX v1 and the CT10 PDF set:  $s$ - and  $t$ -channel production, as well as  $Wt$ . PYTHIA 6.428 [117] with parameters set to the Perugia 2012 tune [118] is used for all three samples. The  $t$ -channel sample is produced in the 4F scheme. The diagram removal scheme [119] is used to treat the overlap of  $t\bar{t}$  and  $Wt$  topologies. The  $tZ$  process is instead generated at LO with MG5\_AMC@NLO and PYTHIA 6. The  $tWZ$  process is also generated with MG5\_AMC@NLO, but at NLO and using PYTHIA 8.

Two additional rare process involving multiple top quarks are the production of  $t\bar{t}t\bar{t}$  and  $t\bar{t}WW$ . Both processes are generated with MG5\_AMC@NLO at LO and use PYTHIA 8 for parton showering and hadronization.

Additional backgrounds originate from Higgs boson production mechanisms other than  $t\bar{t}H$ . The gluon–gluon fusion, vector boson fusion and  $VH$  production mechanisms result in topologies very different from  $t\bar{t}H$ , and are all negligible in the analysis. The rare Higgs boson production with a single top quark is modeled with two samples. MG5\_AMC@NLO with CTEQ6L1 [120] PDF set and HERWIG++ [121] for parton showering and hadronization is used for the production with an additional  $W$  boson in the final state,  $tWH$ . A 4F scheme simulation is used for the  $tHqb$  final state, where  $q$  stands for any quark lighter than the bottom quark. The sample for this process is obtained at LO using MG5\_AMC@NLO with the CT10 PDF set and PYTHIA 8.

### Fake and non-prompt leptons

Photons or jets which are misidentified as light leptons are called fake leptons. Both fake and non-prompt leptons will be referred to as fake leptons in the following. In the dilepton channel of the analysis, this background is estimated from simulation in a control region where two leptons with the same electric charge are required. The dominant contribution to the fake lepton background in this channel arises from single-lepton  $t\bar{t}$  events, where one of the two reconstructed leptons is fake and can therefore have the same charge as the real lepton. The fake lepton estimate is normalized to data.

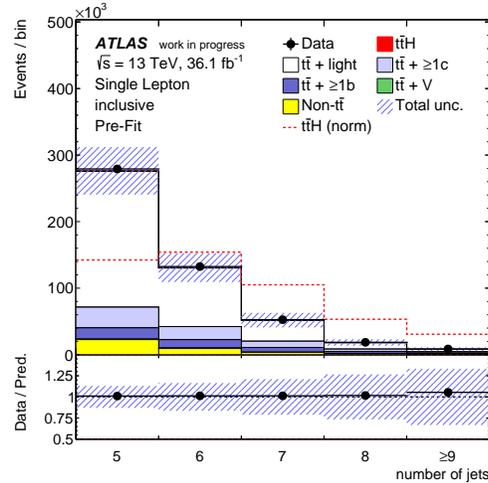


Figure 6.4: Expected distribution of the number of jets per event in the single-lepton channel, compared to data. The uncertainties shown include all sources of systematic uncertainty described in section 6.6, with the exception of the free-floating normalization factors for the  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  processes. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line.

The single-lepton channel makes use of the matrix method [122] to estimate the fake lepton background with a data-driven technique. This method defines a control region with relaxed lepton requirements, composed of events with real and fake leptons. The fraction of events in this control region which also satisfy the nominal analysis requirements is estimated for both real and fake leptons. Events taken from data in the control region are then assigned weights, depending on the lepton kinematics and the measured fractions. These weighted events provide the estimate of fake leptons in the nominal analysis regions. The estimate is statistically consistent with zero events in the three most sensitive single-lepton signal regions ( $SR_1^{\geq 6j}$ ,  $SR_2^{\geq 6j}$ ,  $SR_1^{5j}$ , see section 6.4), and the contribution from fake leptons is neglected there.

### 6.3.4 Inclusive modeling of data

The model described in this section can be compared to data in an inclusive region, containing all events entering the single-lepton channel. Figure 6.4 shows the distribution of the number of jets per event, where the model is in good agreement with data. With this inclusive selection, the background is dominated by  $t\bar{t} + \text{light}$  processes. The uncertainty shown in the figure includes statistical uncertainties and all sources from section 6.6. Uncertainties related to the free-floating normalization factors for the  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  processes are not included.

The number of  $b$ -tagged jets at the four operating points *very tight*, *tight*, *medium*, *loose* are shown in figure 6.5. The model is in agreement with data for all operating points, and the  $t\bar{t} + \geq 1b$  process dominates the events with many  $b$ -tagged jets.

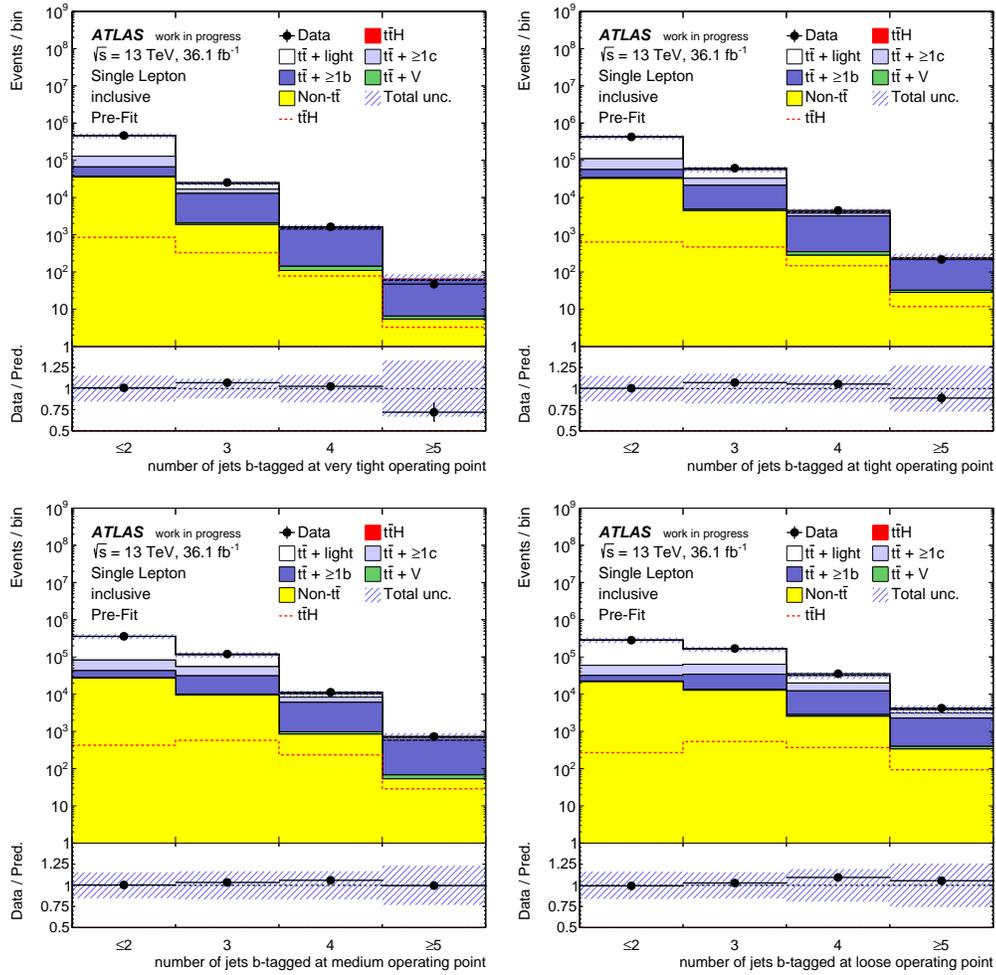


Figure 6.5: Expected distribution of the number of  $b$ -tagged jets per event at the four operating points (*very tight*, *tight*, *medium*, *loose*) in the single-lepton channel, compared to data. The uncertainties shown include all sources of systematic uncertainty described in section 6.6, with the exception of the free-floating normalization factors for the  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  processes. The  $t\bar{t}H$  signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram.

## 6.4 Event categorization

Events in both the single-lepton and dilepton channel are divided into multiple exclusive regions. These regions are defined via the amount of jets present per event, as well as the amount of  $b$ -tagged jets at the four operating points. The expected composition of events varies across the regions defined. Regions with large contributions from the  $t\bar{t}H$  process are called signal regions. The remaining regions are called control regions, and they serve to constrain backgrounds and systematic uncertainties in the analysis. Both signal and control regions enter the fit to measure the  $t\bar{t}H$  signal strength. The simultaneous use of four calibrated  $b$ -tagging operating points is an improvement compared to the previous Run-1  $t\bar{t}H(b\bar{b})$  ATLAS analysis [88].

### 6.4.1 Region definitions

The boosted region in the single-lepton channel is defined as described in section 6.2.3. It will also be referred to as  $\text{SR}^{\text{boosted}}$ . All remaining events in the single-lepton channel are split depending on their jet multiplicity; separate regions are constructed for events with exactly five, and six or more jets. For every event, jets are then considered in decreasing order of tightness of the  $b$ -tagging operating points they satisfy. The operating points are assigned numerical values 1–5, with 1 representing a jet that is not  $b$ -tagged, and 5 for a jet tagged at the *very tight* operating point, as described in section 4.4.2. An event with at least four jets can thus be represented by  $(b_1, b_2, b_3, b_4)$ , with  $b_i \in \{1, 2, 3, 4, 5\}$  and  $b_i \geq b_{i+1}$ .

Exactly four  $b$  quarks are expected from the  $t\bar{t}H(b\bar{b})$  signal, motivating the creation of signal regions requiring four  $b$ -tagged jets. The best signal purity is achieved when using the *very tight* operating point. In the single-lepton channel, the  $\text{SR}_1^{5j}$  and  $\text{SR}_1^{\geq 6j}$  signal regions are therefore defined by requiring the first four jets to be  $b$ -tagged at the *very tight* operating point, and the event to contain exactly five, or at least six jets, respectively. These events can be represented by  $(5, 5, 5, 5)$ .

All remaining regions in the single-lepton channel are defined by grouping together events with similar  $b$ -tagging configurations. The  $\text{SR}_2^{5j}$ ,  $\text{SR}_2^{\geq 6j}$  and  $\text{SR}_3^{\geq 6j}$  signal regions are obtained by merging together configurations enriched in  $t\bar{t} + \geq 2b$  processes. Control regions enriched in  $t\bar{t} + b$ ,  $t\bar{t} + \geq 1c$  and  $t\bar{t} + \text{light}$  collect the rest of the events. Three of these control regions are built with events containing exactly five jets:  $\text{CR}_{t\bar{t}+b}^{5j}$ ,  $\text{CR}_{t\bar{t}+\geq 1c}^{5j}$ , and  $\text{CR}_{t\bar{t}+\text{light}}^{5j}$ . The corresponding control regions for events with at least six jets are  $\text{CR}_{t\bar{t}+b}^{\geq 6j}$ ,  $\text{CR}_{t\bar{t}+\geq 1c}^{\geq 6j}$ , and  $\text{CR}_{t\bar{t}+\text{light}}^{\geq 6j}$ .

The exact definitions of the eleven resolved regions in the single-lepton channel are summarized in figure 6.6 and figure 6.7 for the regions with exactly five, and at least six jets, respectively. These figures show the  $b$ -tagging requirement placed on the first two jets on the vertical axis, and the requirement on third and fourth jet on the horizontal axis. Signal regions have the most stringent  $b$ -tagging requirements. Control regions enriched in  $t\bar{t} + \text{light}$  have the loosest  $b$ -tagging requirements. Regions enriched in  $t\bar{t} + \geq 1c$  are located between regions enriched in  $t\bar{t} + \text{light}$  and signal regions; this is due to the larger mis-tag rate of  $c$ -jets compared to light jets.

The treatment for the dilepton channel is similar, resulting in three signal regions and four control

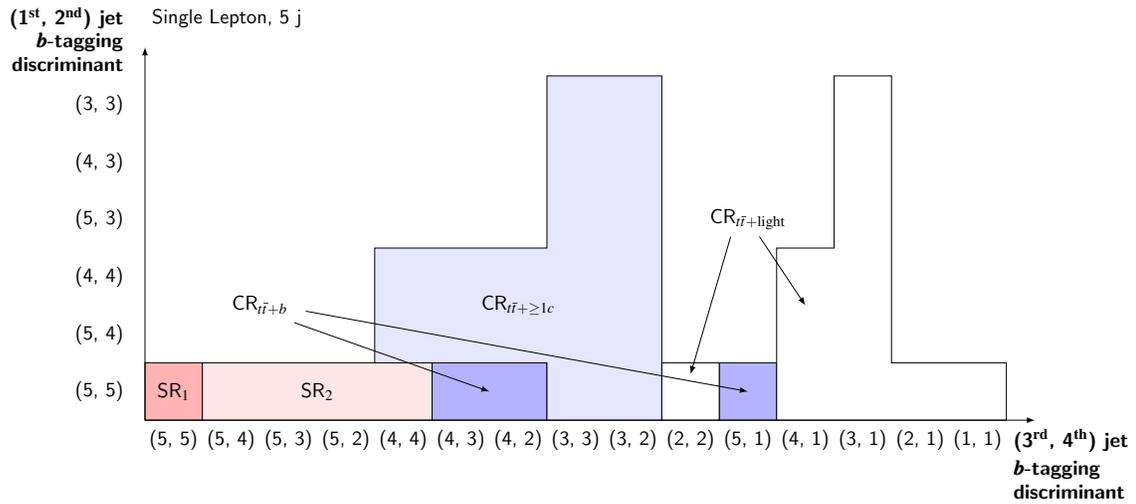


Figure 6.6: Definition of resolved analysis regions with exactly five jets in the single-lepton channel. The vertical axis shows the  $b$ -tagging requirements for the first two jets in each event, while the horizontal axis shows the requirement for the third and fourth jet. Jets are ordered by decreasing tightness of the operating point they satisfy [1].

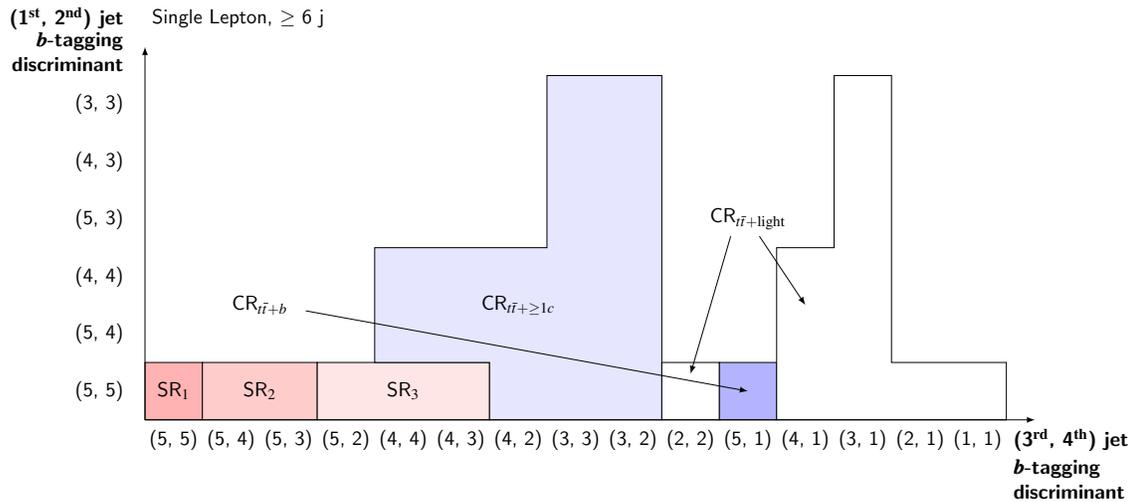


Figure 6.7: Definition of resolved analysis regions with six or more jets in the single-lepton channel. The vertical axis shows the  $b$ -tagging requirements for the first two jets in each event, while the horizontal axis shows the requirement for the third and fourth jet. Jets are ordered by decreasing tightness of the operating point they satisfy [1].

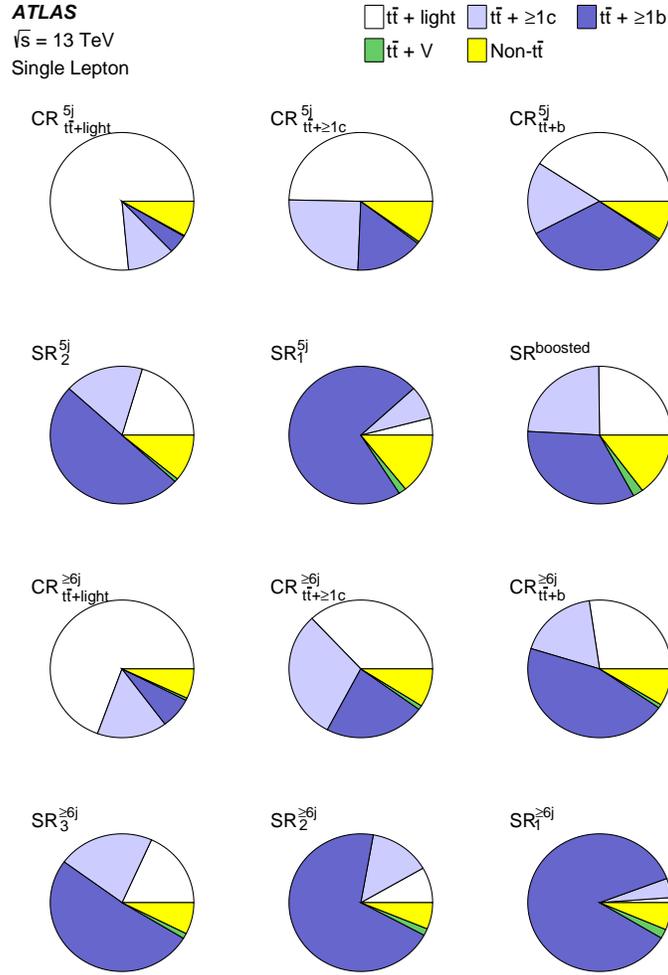


Figure 6.8: Composition of background processes in the single-lepton regions. Each pie chart shows the relative contributions per process and region, with the processes defined in section 6.3 [1].

regions:  $SR_1^{\geq 4j}$ ,  $SR_2^{\geq 4j}$ ,  $SR_3^{\geq 4j}$ , and  $CR_{t\bar{t} \geq 1b}^{3j}$ ,  $CR_{t\bar{t} + \text{light}}^{3j}$ ,  $CR_{t\bar{t} \geq 1c}^{\geq 4j}$ ,  $CR_{t\bar{t} + \text{light}}^{\geq 4j}$ . The corresponding figures are shown in appendix section A.1.

#### 6.4.2 Region composition and signal contributions

The background composition of all single-lepton regions is visualized in figure 6.8. Signal regions are dominated by  $t\bar{t} + \geq 1b$  production. The remaining regions vary in their composition, from regions dominated by  $t\bar{t} + \text{light}$  to regions with substantial  $t\bar{t} + \geq 1c$  or  $t\bar{t} + \geq 1b$  contribution. No region is completely dominated by the  $t\bar{t} + \geq 1c$  background. The relative contribution to the total background from non- $t\bar{t}$  and  $t\bar{t}V$  processes is small compared to the  $t\bar{t}$  background.

Figure 6.9 visualizes the contribution of the signal  $t\bar{t}H$  process to the single-lepton regions. The solid black line, corresponding to the left vertical axis, shows the fraction of expected signal events

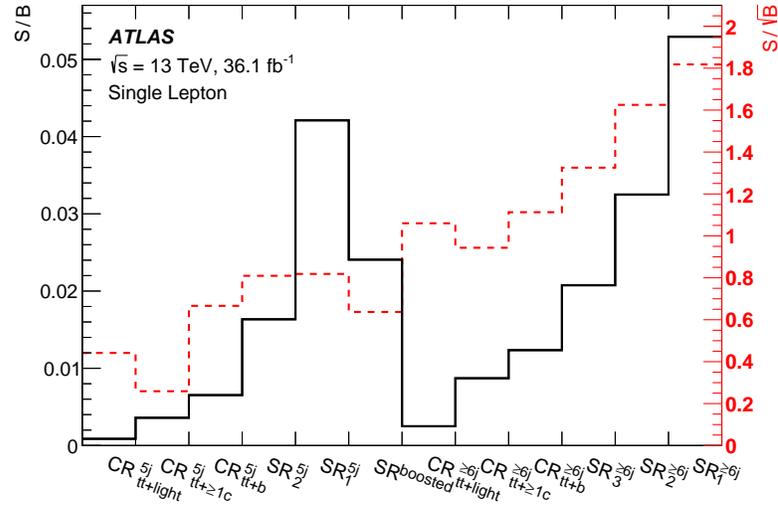


Figure 6.9: Signal contributions per analysis region in the single-lepton channel, evaluated using the expected amount of  $t\bar{t}H$  events ( $S$ ) and background events ( $B$ ) per region. The solid black line, corresponding to the left vertical axis, shows  $S/B$ . The dashed red line, corresponding to the right vertical axis, shows  $S/\sqrt{B}$  [1].

( $S$ ) to the total background ( $B$ ). This fraction is below 1.5% in the control regions, and surpasses 5% only in  $SR_1^{\geq 6j}$ . The dashed red line, corresponding to the right vertical axis, shows  $S/\sqrt{B}$ , which is also highest in  $SR_1^{\geq 6j}$ .

The corresponding figures for the dilepton channel are shown in appendix section A.1.

## 6.5 Multivariate analysis techniques

Multivariate analysis techniques are employed in all signal regions to help isolate the  $t\bar{t}H$  signal process from the backgrounds. The approach for the  $t\bar{t}H(b\bar{b})$  analysis has two stages. In the first stage, various methods of system reconstruction are performed. A reconstruction BDT (see also section 5.3.1) matches jets to partons to obtain candidates for top quarks and the Higgs boson in each event. The likelihood discriminant (LHD) considers the kinematics across possible jet-parton assignments, and calculates a discriminant for each event. The MEM provides another discriminant, built from first principles. All three methods approach the system reconstruction slightly differently, and their combination results in a stronger overall discriminant. The output from these methods is combined with additional information in the second stage, the classification BDT. Both the reconstruction and classification BDT are trained with the Toolkit for Multivariate Data Analysis with ROOT (TMVA) package [123].

### 6.5.1 Reconstruction BDT

The reconstruction BDT is trained to match jets to the partons from the  $t\bar{t}$  and  $t\bar{t}H$  system. It is used in the dilepton and resolved single-lepton regions. Candidates for top quarks,  $W$  bosons and Higgs bosons are built by combining jets and leptons, and multiple permutations (jet–parton assignments) are possible per event. The use of  $b$ -tagging information reduces the total amount of permutations considered. For each permutation, invariant masses of object candidates and combinations of them, as well as angular distances, are calculated. They serve as inputs to the reconstruction BDT. The BDT is trained with simulated  $t\bar{t}H$  events to identify correct permutations. Besides the BDT output itself, reconstructed quantities in the permutation with the highest BDT output are also used as input to the classification BDT. Two different versions of the reconstruction BDT are used; for one version all observables related to the Higgs boson are removed. This reduces the reconstruction efficiency, but improves the discriminating power for variables related to the Higgs boson, such as the invariant mass of the Higgs boson candidate. In  $\text{SR}_1^{\geq 6j}$ , the Higgs boson is correctly reconstructed for 48% and 32% of the simulated  $t\bar{t}H$  events when using or not using information related to the Higgs boson in the reconstruction BDT, respectively.

The large- $R$  jets in the boosted region simplify the combinatorial problem, and no reconstruction BDT is used in that region. The Higgs boson candidate in this region contains the two jets from the Higgs boson decay for 47% of the  $t\bar{t}H$  events.

### 6.5.2 Likelihood discriminant

The LHD is calculated from various one-dimensional probability density functions, which describe signal and background distribution of kinematic variables such as invariant masses and angles. It is used in the resolved single-lepton regions. The probabilities for an event to be consistent with the signal or background hypotheses,  $p^{\text{sig}}$  and  $p^{\text{bkg}}$ , are calculated as a product of the one-dimensional probability density functions, and averaged over jet–parton permutations. The permutations are weighted with  $b$ -tagging information. Two background processes are considered,  $t\bar{t} + b$  and  $t\bar{t} + \geq 2b$ , their likelihoods are added in the calculation of  $p^{\text{bkg}}$ , and weighted by their expected relative contribution in the simulated  $t\bar{t}$  sample. Different distributions are used in regions with exactly five, or six and more jets. An additional hypothesis is included for events where not all of the jets corresponding to the hadronic decay products of the  $W$  boson were reconstructed. The output of the LHD for use in the classification BDT is defined as  $p^{\text{sig}}/p^{\text{sig}}+p^{\text{bkg}}$ . In contrast to the reconstruction BDT, the LHD incorporates information from multiple jet–parton permutations in its output. The LHD however does not account for correlations between the kinematic variables used to build the probability density function templates, while both the reconstruction BDT and the MEM take correlations into account. More information about the LHD method is provided in reference [124].

### 6.5.3 Matrix element method

The MEM provides a strong discriminant between  $t\bar{t}H$  and the  $t\bar{t} + b\bar{b}$  background and is described in detail in chapter 7. In contrast to the LHD, the discriminant is calculated from first principles. Due to its large computational cost, the MEM is only used in the most sensitive signal region  $SR_1^{\geq 6j}$  of the analysis. This choice is made to maximize the separation of  $t\bar{t}H$  from the backgrounds in the region where it is most important for the analysis.

### 6.5.4 Classification BDT

The classification BDT is trained to separate  $t\bar{t}H$  from the  $t\bar{t}$  background. It combines a range of inputs to achieve good discrimination. For every input variable, the expected distribution from simulation is compared to data, and only well modeled variables are used. Different combinations of input variables are used across the nine signal regions, they are listed in reference [1]. The basic inputs are kinematic variables, such as angles between reconstructed objects and invariant masses of combined objects. Information regarding the  $b$ -tagging operating points passed by various jets is also included. An additional ingredient are the three intermediate system reconstruction approaches. Information from the reconstruction BDT is used in the dilepton and the resolved single-lepton regions, such as the output of the reconstruction BDT itself, but also the Higgs boson candidate mass. The LHD is used in the resolved single-lepton regions, and the MEM discriminant enters the classification BDT only in  $SR_1^{\geq 6j}$ . The LHD and MEM discriminant are the most powerful inputs to the classification BDT, followed by the reconstruction BDT output.

## 6.6 Systematic uncertainties

The  $t\bar{t}H(b\bar{b})$  analysis is affected by many sources of systematic uncertainty. All sources can generally affect both the normalization and shape of the distributions on which they act. Exceptions to this are the luminosity uncertainty, as well as cross-section and normalization uncertainties on the various processes considered in the analysis. These sources of uncertainty only affect the normalization of the samples they act on. In the case of normalization uncertainties affecting only a specific process, a variation of the related nuisance parameter can however still result in a shape variation of a distribution containing a sum of processes. A nuisance parameter is introduced for each source of uncertainty.

General notes regarding the treatment of systematic uncertainties in the profile likelihood fit are given in section 6.6.1. This is followed by descriptions of the experimental and modeling uncertainties in section 6.6.2 and section 6.6.3. A summary of all nuisance parameters considered is provided in section 6.6.4.

### 6.6.1 Nuisance parameter details

The nuisance parameters are implemented with a Gaussian constraint as explained in section 5.1.3. The interpolation between the two templates defining the  $\pm 1\sigma$  effect of the systematic variation specified by a nuisance parameter is done with polynomial functions. The extrapolation method beyond this range differs for the normalization and shape components. A linear extrapolation is used for the shape component, while the use of an exponential extrapolation for the normalization component prevents the total yield from samples to become negative. The normalization component thus effectively behaves as if the extrapolation was linear, with a log-normal constraint.

Some nuisance parameters are defined by a variation of the nominal configuration in only a single direction. An important example in this analysis is the comparison between the nominal  $t\bar{t}$  simulation, and a variation where the MC generator is replaced by an alternative setup. In these cases, the variation is defined as the  $+1\sigma$  effect, and the effect of the variation is symmetrized to obtain the corresponding template for the  $-1\sigma$  effect of the nuisance parameter. When both variations of a nuisance parameter are defined, the templates corresponding to the  $\pm 1\sigma$  effects are usually symmetrized. An exception to this are cross-section and normalization uncertainties.

Besides the symmetrization, a smoothing procedure is applied to the templates defining the  $\pm 1\sigma$  effects of nuisance parameters, with the exception of most of the cross-section and normalization uncertainties. This procedure removes the effect of statistical fluctuations in the templates, which lead to artificially enlarged constraints in the profile likelihood fit.

Lastly, the effect of a nuisance parameter acting on a specific sample in a given region is removed from the likelihood function if this effect is negligible. This speeds up the profile likelihood fit, without changing the results. A normalization effect in a given region acting on a sample is removed from the fit model if it is below 1%. The same threshold is used to drop shape effects; a shape size of 1% corresponds to the template in any bin changing the normalization of the nominal sample by more than 1% compared to the average effect.

#### Statistical uncertainties on the model

Statistical uncertainties related to the distributions predicted by the nominal model originate from the finite amount of simulated events in the MC samples, and the finite amount of events in the data-driven fake lepton estimate for the single-lepton channel. These uncertainties reach 12% in two bins considered in the analysis, while the uncertainties in the majority of the remaining bins is significantly below 10%. The nominal model estimate is treated as a subsidiary measurement, with an uncertainty corresponding to the statistical uncertainty in each bin. One nuisance parameter per bin in the analysis is used to describe these statistical uncertainties. A Gaussian constraint controls these statistical variations in each bin, which is a good approximation for Poisson uncertainties for the relatively small statistical uncertainties in this analysis.

### 6.6.2 Experimental uncertainties

The relative uncertainty regarding the integrated luminosity of the dataset used in the analysis is 2.1%, derived with a similar method as in reference [50]. A variation of the pile-up modeling is also included to cover related uncertainties.

#### Leptons

Lepton related systematic uncertainties have a very small impact on the analysis. For electrons, these cover effects related to the trigger, reconstruction, identification, and isolation efficiencies. Two additional nuisance parameter cover the calibration of the electron energy scale and resolution, for a total of six nuisance parameters related to electrons. The treatment of muons is similar. The associated nuisance parameters describe uncertainties related to the muon trigger, the association of tracks to vertices, the muon identification, identification of low-momentum muons, and the muon isolation. These nuisance parameters for muon systematic uncertainties related to efficiency are split into systematic and statistical components of the effects they describe. Muon scale calibration uncertainties are covered by five nuisance parameters, related to the muon momentum scale and resolution, and additional calibrations used. In total, 15 nuisance parameters related to muons in ATLAS are used. Three more nuisance parameters related to tau leptons are considered and found to be negligible.

#### Jets

A total of 23 nuisance parameters are used to describe sources of systematic uncertainty related to jets. The basic jet energy scale calibration is covered by a set of eight nuisance parameters. Additional nuisance parameters are used to describe uncertainties related to the calibration dependence on jet flavor, the jet position in  $\eta$ , pile-up, jets not contained within the calorimeter system and jets with high momentum. This results in 20 nuisance parameters related to the jet energy scale. Two nuisance parameters describe uncertainties related to the jet energy resolution. One more nuisance parameter covers uncertainties related to the jet vertex tagger. Since the events considered in the analysis have many jets, the related uncertainties have a significant impact on the analysis.

#### Flavor tagging

The efficiency to correctly tag  $b$ -jets, and mis-tag rates for  $c$ - and light jets are measured for all operating points used in the analysis, and combined into a global calibration. The related uncertainties take into account correlations between different operating points, and depend on the jet transverse momentum. In the case of the light jet mis-tag rate, the calibration is also dependent on the jet pseudorapidity. Uncertainties regarding the  $b$ -tagging efficiency are split into 30 sources. There are 15 sources describing mis-tag rates for  $c$ -jets, and 80 nuisance parameters are used for light jet mis-tag rate uncertainties. An additional nuisance parameter is used for jets from hadronic decays of tau

leptons. The  $b$ -tagging uncertainty ranges between 2% and 10%, and the  $c$ - and light jet mis-tag rates have uncertainties in the ranges 5–20%, and 10–50%, respectively.

### Missing transverse energy

Uncertainties regarding the energy of objects used to calculate the missing transverse energy are propagated to the measured  $E_T^{\text{miss}}$ . An additional three nuisance parameters describe uncertainties related to energy deposits that enter the  $E_T^{\text{miss}}$  calculation, but are not associated to any reconstructed objects.

### 6.6.3 Signal and background modeling

This section summarizes systematic uncertainties related to the nominal model introduced in section 6.3.

#### $t\bar{t}H$ signal

The  $t\bar{t}H$  cross-section uncertainty is split into two components. The first component contains the QCD scale uncertainties ( $^{+5.8\%}_{-9.2\%}$ ), and the second component the uncertainties related to the PDF and strong coupling ( $\pm 3.6\%$ ) [34]. Three nuisance parameters cover the uncertainties related to the Higgs boson branching ratios into  $b\bar{b}$ ,  $WW^*$ , and the remaining final states. The absolute uncertainty of the branching ratio for the dominant  $b\bar{b}$  is  $^{+1.2\%}_{-1.3\%}$ . All five of these components do not have a significant effect on the  $t\bar{t}H$  shape, so only their normalization effect is considered. An uncertainty regarding the choice of the PS and hadronization model is implemented by comparing the nominal setup, which uses PYTHIA 8, to a sample using HERWIG++.

#### $t\bar{t}$ + jets background

A large amount of systematic uncertainties is associated to the modeling of the  $t\bar{t}$  + jets background. The cross-section uncertainty is 6% [105], implemented as one nuisance parameter and correlated for all  $t\bar{t}$  components. All remaining uncertainties related to  $t\bar{t}$  are implemented separately for the different  $t\bar{t}$  components. Two nuisance parameters control the normalization of the  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  components, with no constraint applied. The normalization of these components is free-floating in the fit, and referred to as  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$ . All samples used to define additional systematic uncertainties are reweighted such that the fractions of  $t\bar{t} + \geq 1b$ ,  $t\bar{t} + \geq 1c$  and  $t\bar{t}$  + light processes they contain corresponds to the nominal  $t\bar{t}$  sample, and such that the  $t\bar{t} + \geq 1b$  sub-components match the SHERPA4F predictions. Dedicated uncertainties describe the  $t\bar{t} + \geq 1b$  sub-component fractions, and the samples used to derive these uncertainties are not reweighted to match the SHERPA4F prediction.

Multiple alternative simulations for the  $t\bar{t}$  background are generated. A  $t\bar{t}$  sample, which will be referred to in the following as SHERPA5F, is generated at NLO with SHERPA 2.2.1 and OPENLOOPS, using the NNPDF3.0NNLO PDF set. This sample is accurate to NLO for up to one additional parton

beyond the  $t\bar{t}$  system, and to LO for up to four more partons. The difference between the nominal POWHEG+PYTHIA 8 and the SHERPA5F samples is used as a systematic uncertainty related to the choice of NLO event generator, and it also varies the PS and hadronization model. Another sample is generated with a configuration similar to the nominal POWHEG+PYTHIA 8 sample, but using HERWIG 7 [125], version 7.0.1, instead for parton showering and hadronization. The difference between the POWHEG+PYTHIA 8 and POWHEG+HERWIG 7 samples is used as nuisance parameters related to the choice of PS and hadronization model. Two alternative POWHEG+PYTHIA 8 samples are compared to the nominal POWHEG+PYTHIA 8 sample to evaluate the uncertainty related to initial state radiation (ISR) and FSR. In these alternative samples, renormalization and factorization scales, a parameter in POWHEG controlling extra radiation, as well as parameters in the A14 tune for the PYTHIA 8 shower are set to different values than in the nominal POWHEG+PYTHIA 8 sample. These three uncertainty sources are implemented with nine nuisance parameters, split between  $t\bar{t}$  components.

Two additional modeling uncertainties for  $t\bar{t}+ \geq 1b$  and  $t\bar{t}+ \geq 1c$  are constructed from additional MC samples. The residual difference between the POWHEG+PYTHIA 8 and SHERPA4F samples is used as an uncertainty for the  $t\bar{t}+ \geq 1b$  sub-components, with the exception of  $t\bar{t}+ b$  (MPI/FSR), which is not included in the SHERPA4F calculation. This uncertainty covers differences between the 5F and NLO 4F scheme calculations of  $t\bar{t}+ \geq 1b$  and  $t\bar{t}+ b\bar{b}$ . A dedicated  $t\bar{t}+ c\bar{c}$  sample is produced in the three-flavor (3F) scheme with MG5\_AMC@NLO at NLO, including the effect of massive  $c$  quarks. Parton showering and hadronization are performed with HERWIG++ [126]. The difference between this sample and the POWHEG+PYTHIA 8 prediction is used as an additional uncertainty for the  $t\bar{t}+ \geq 1c$  process.

The uncertainties related to the fractions of  $t\bar{t}+ \geq 1b$  sub-components predicted by SHERPA4F are evaluated by varying parameters within this simulation. Three of the uncertainties are related to the settings of scales within SHERPA. One more nuisance parameter is used to compare two models for the PS. Two nuisance parameters describe the effect of exchanging the nominal CT10 PDF set by the MSTW2008NLO [127] and NNPDF2.3NLO PDF sets. One nuisance parameter varies the settings of the underlying event (UE) modeling. The UE refers to everything not related to the primary hard scattering process of interest in proton–proton collisions. These seven nuisance parameters are used to build the uncertainty band for the SHERPA4F prediction shown in figure 6.3. An additional 50% normalization uncertainty is added for the  $t\bar{t}+ \geq 3b$  sub-component, which covers the difference between the POWHEG+PYTHIA 8 and SHERPA4F predictions. Lastly, a 50% normalization uncertainty is added for the  $t\bar{t}+ b$  (MPI/FSR) sub-component, which is not described by the SHERPA4F prediction.

A summary of the nuisance parameters for  $t\bar{t}$  modeling is shown in table 6.2. The modeling of  $t\bar{t}+ \text{light}$  processes is described by three nuisance parameters, which compare POWHEG+PYTHIA 8 to SHERPA5F and POWHEG+HERWIG 7, and also include ISR and FSR variations. One additional nuisance parameter is used for  $t\bar{t}+ \geq 1c$ , originating from the comparison to the MG5\_AMC@NLO+HERWIG++ sample, for a total of four nuisance parameters affecting  $t\bar{t}+ \geq 1c$  modeling. The  $t\bar{t}+ \geq 1b$  component is covered by 13 nuisance parameters in total, including the three sources also affecting  $t\bar{t}+ \text{light}$ , the comparison to SHERPA4F, seven variations affecting the SHERPA4F prediction, and finally two

Table 6.2: Systematic uncertainty sources affecting the modeling of  $t\bar{t}$  + jets. The left column shows the individual sources. Additional details regarding the sources are given in the central column. The column on the right lists on which  $t\bar{t}$  components the sources act on, and whether the effect is correlated between the components. Additional details are provided in section 6.6.3 [1].

| Systematic sources                        | Description  | $t\bar{t}$ categories |
|---|--|-----------------------|
| $t\bar{t}$ cross-section                  | Up or down by 6%   | All, correlated       |
| $k(t\bar{t} + \geq 1b)$                   | Free-floating $t\bar{t} + \geq 1b$ normalization                                   | $t\bar{t} + \geq 1b$  |
| $k(t\bar{t} + \geq 1c)$                   | Free-floating $t\bar{t} + \geq 1c$ normalization                                   | $t\bar{t} + \geq 1c$  |
| SHERPA5F vs. nominal                      | NLO event generator choice   | All, uncorrelated     |
| PS & hadronization                        | POWHEG+HERWIG 7<br>vs. POWHEG+PYTHIA 8   | All, uncorrelated     |
| ISR and FSR                               | Variations of $\mu_R$ , $\mu_F$ , and additional<br>POWHEG and PYTHIA 8 parameters | All, uncorrelated     |
| $t\bar{t} + \geq 1b$ SHERPA4F vs. nominal | Comparison of $t\bar{t} + b\bar{b}$ NLO (4F)<br>vs. POWHEG+PYTHIA 8 (5F)           | $t\bar{t} + \geq 1b$  |
| $t\bar{t} + \geq 1c$ 3F vs. 5F scheme     | MG5_AMC@NLO+HERWIG++<br>vs. POWHEG+PYTHIA 8  | $t\bar{t} + \geq 1c$  |
| $t\bar{t} + \geq 1b$ scale variations     | Three components   | $t\bar{t} + \geq 1b$  |
| $t\bar{t} + \geq 1b$ shower recoil scheme | Alternative model scheme   | $t\bar{t} + \geq 1b$  |
| $t\bar{t} + \geq 1b$ PDF (MSTW)           | Compare MSTW vs. CT10  | $t\bar{t} + \geq 1b$  |
| $t\bar{t} + \geq 1b$ PDF (NNPDF)          | Compare NNPDF vs. CT10   | $t\bar{t} + \geq 1b$  |
| $t\bar{t} + \geq 1b$ UE                   | Alternative set of tuned parameters<br>for the underlying event                    | $t\bar{t} + \geq 1b$  |
| $t\bar{t} + \geq 3b$ normalization        | Up or down by 50%  | $t\bar{t} + \geq 1b$  |
| $t\bar{t} + \geq 1b$ MPI                  | Up or down by 50%  | $t\bar{t} + \geq 1b$  |

normalization uncertainties for  $t\bar{t} + \geq 3b$  and  $t\bar{t} + b$  (MPI/FSR).

### Small backgrounds

A 40% normalization uncertainty for  $W$ +jets production is used, derived from parameter variations within the SHERPA simulation. The dominant contribution to this uncertainty comes from variations of the renormalization scale. An additional 30% uncertainty is assigned to events containing  $b$ - or  $c$ -jets, split into two components. One nuisance parameter is used for events with exactly two such jets, another one for events with three or more. This uncertainty covers differences observed when comparing the SHERPA prediction to a sample generated with MG5\_AMC@NLO and PYTHIA 8. An uncertainty of 35% is assigned to the normalization of  $Z$ +jets events. It is split into three components; events are treated separately if they have exactly three jets, at least four jets and are in the dilepton channel, or fall into the single-lepton channel. A normalization uncertainty of 50% is used for diboson

events [116].

The treatment of the  $t\bar{t}V$  cross-section uncertainties is equivalent to the treatment of  $t\bar{t}H$ . They are split into components for the PDF and scale uncertainties, and not correlated between  $t\bar{t}W$  and  $t\bar{t}Z$ . For both  $t\bar{t}W$  and  $t\bar{t}Z$ , the combined impact of the two components is around  $\pm 15\%$  [128]. The comparison between the nominal  $t\bar{t}V$  samples and samples produced with SHERPA is used as an additional modeling uncertainty, not correlated between  $t\bar{t}W$  and  $t\bar{t}Z$ .

One normalization uncertainty each is assigned to the cross-section for the  $s$ - and  $t$ -channel single top quark production, as well as the production of  $Wt$ , and for  $tWZ$ . The cross-section uncertainty for  $tZ$  is split into two components, equivalent to the treatment employed for  $t\bar{t}H$  and  $t\bar{t}V$ . For the  $Wt$  and  $t$ -channel production processes, additional uncertainties related to the choice of PS and hadronization model are derived by comparing the nominal POWHEG+PYTHIA 6 samples to alternative samples generated with POWHEG and HERWIG++. Two more uncertainties are derived from POWHEG+PYTHIA 6 samples, where renormalization and factorization scale, as well as settings in the Perugia shower tune, are all varied; this describes variations in ISR and FSR. An additional uncertainty affecting the  $Wt$  sample is derived by comparing the nominal diagram removal scheme to the alternative diagram subtraction scheme [119]. In total, the single top quark processes are described by six cross-section uncertainties and five nuisance parameters related to modeling of the processes.

Seven more nuisance parameters describe cross-section uncertainties for the remaining background processes with minor contributions. A 50% normalization uncertainty is used for  $t\bar{t}t\bar{t}$ . The uncertainties for  $t\bar{t}WW$ ,  $tWH$  and  $tHqb$  are split into two components each. These components separately describe QCD scale uncertainties and uncertainties related to the PDFs.

### Fake and non-prompt leptons

A 50% uncertainty is used for the normalization of the data-driven fake lepton estimate in the single-lepton channel. It is split into six components for this channel, treating events with electrons and muons separately. Three types of regions are assigned separate nuisance parameters: the boosted region, resolved regions with exactly five jets, and resolved regions with six or more jets. The fake lepton estimate for the dilepton channel is derived from simulation, and a single 25% normalization uncertainty is assigned to it. This results in a total of seven nuisance parameters related to the fake lepton estimate.

#### 6.6.4 Summary of systematic uncertainty sources

Table 6.3 lists all systematic uncertainties affecting the analysis, grouped by their sources. The type of each source indicate whether the nuisance parameter affects only normalization (type  $N$ ) or both shape and normalization (type  $S + N$ ) of the samples on which it acts. Many of the uncertainties are broken down into multiple components, the amount of these components per source is listed in the last column. The cross-sections for the small backgrounds listed at the end of the table affect  $t\bar{t}t\bar{t}$ ,  $tZ$ ,  $tWZ$ ,  $t\bar{t}WW$ ,  $tHjb$  and  $WtH$ .

Table 6.3: List of the systematic uncertainties affecting the analysis. The type  $N$  indicates uncertainties changing normalization of the affected process, uncertainties with type  $S + N$  can change both shape and normalization. The amount of different components per source is listed in the third column [1].

| Systematic uncertainty                             | Type              | Components |
|--|-------------------|------------|
| <i>Experimental uncertainties</i>                  |                   |            |
| Luminosity   | $N$               | 1          |
| Pile-up modeling                                   | $S + N$           | 1          |
| <b>Physics objects</b>                             |                   |            |
| Electron   | $S + N$           | 6          |
| Muon   | $S + N$           | 15         |
| Taus   | $S + N$           | 3          |
| Jet energy scale                                   | $S + N$           | 20         |
| Jet energy resolution                              | $S + N$           | 2          |
| Jet vertex tagger                                  | $S + N$           | 1          |
| $E_T^{\text{miss}}$                                | $S + N$           | 3          |
| <b><math>b</math>-tagging</b>                      |                   |            |
| Efficiency   | $S + N$           | 30         |
| Mis-tag rate ( $c$ )                               | $S + N$           | 15         |
| Mis-tag rate (light)                               | $S + N$           | 80         |
| Mis-tag rate (extrapolation $c \rightarrow \tau$ ) | $S + N$           | 1          |
| <i>Modeling uncertainties</i>                      |                   |            |
| <b>Signal</b>                                      |                   |            |
| $t\bar{t}H$ cross-section                          | $N$               | 2          |
| $H$ branching fractions                            | $N$               | 3          |
| $t\bar{t}H$ modeling                               | $S + N$           | 1          |
| <b><math>t\bar{t}</math> background</b>            |                   |            |
| $t\bar{t}$ cross-section                           | $N$               | 1          |
| $t\bar{t} + \geq 1c$ normalization                 | free-floating $N$ | 1          |
| $t\bar{t} + \geq 1b$ normalization                 | free-floating $N$ | 1          |
| $t\bar{t} + \text{light}$ modeling                 | $S + N$           | 3          |
| $t\bar{t} + \geq 1c$ modeling                      | $S + N$           | 4          |
| $t\bar{t} + \geq 1b$ modeling                      | $S + N$           | 13         |
| <b>Other backgrounds</b>                           |                   |            |
| $W$ +jets normalization                            | $N$               | 3          |
| $Z$ +jets normalization                            | $N$               | 3          |
| Diboson normalization                              | $N$               | 1          |
| $t\bar{t}W$ cross-section                          | $N$               | 2          |
| $t\bar{t}Z$ cross-section                          | $N$               | 2          |
| $t\bar{t}W$ modeling                               | $S + N$           | 1          |
| $t\bar{t}Z$ modeling                               | $S + N$           | 1          |
| Single top cross-section                           | $N$               | 6          |
| Single top modeling                                | $S + N$           | 5          |
| Small background cross-sections                    | $N$               | 7          |
| Fake and non-prompt lepton normalization           | $N$               | 7          |

## 6.7 Statistical analysis and results

This section presents the results of the statistical analysis for the  $t\bar{t}H(b\bar{b})$  search. In the statistical analysis, the profile likelihood ratio introduced in section 5.2.2 is maximized, corresponding to a minimization of  $t_\mu$ . The HISTFACTORY [129] software is used to build the likelihood functions for the model described in section 6.3, including the uncertainties described in section 6.6. HISTFACTORY is based on ROOFIT [130] and used together with ROOSTATS [131] tools. It operates on histograms specifying all relevant distributions needed to build the likelihood functions. The creation of these histograms is performed with the TREXFITTER software. This software is developed for internal use in the ATLAS collaboration. It acts as a steering tool for the statistical analysis, and includes a large range of tools used to study the fit model and fit results in detail. The minimization of  $t_\mu$  is performed with MINUIT [82], implemented in C++ within the ROOT framework [83]. The uncertainties for unconstrained parameters are determined with the MINOS algorithm, which supports asymmetric uncertainties. It varies a parameter in both directions until  $t_\mu$  changes by one unit, thereby obtaining the parameter uncertainties.

The profile likelihood test statistic  $t_\mu$  is constructed as described in section 5.2.2. Systematic uncertainties are implemented according to the prescription from section 5.1.3, with additional details provided in section 6.6.1. All nuisance parameters  $\theta$  encoding prior knowledge are scaled such that  $\hat{\theta} \equiv \theta_0 = 0$  and  $\hat{\sigma}_\theta \equiv \Delta\theta = 1$ . Upper limits are calculated with the CL<sub>s</sub> technique described in section 5.2.2.

### 6.7.1 Fit model details and expected performance

A total of 19 regions enter the simultaneous fit. The four dilepton control regions, as well as  $\text{CR}_{t\bar{t}+b}^{5j}$ ,  $\text{CR}_{t\bar{t}+\text{light}}^{5j}$ ,  $\text{CR}_{t\bar{t}+b}^{\geq 6j}$ , and  $\text{CR}_{t\bar{t}+\text{light}}^{\geq 6j}$ , enter as a single bin each. The distribution of the  $H_T^{\text{had}}$  variable, which is the scalar sum of jet transverse momenta, is used for the  $\text{CR}_{t\bar{t}+\geq 1c}^{5j}$  and  $\text{CR}_{t\bar{t}+\geq 1c}^{\geq 6j}$  regions instead. These distributions, with six and eight bins for  $\text{CR}_{t\bar{t}+\geq 1c}^{5j}$  and  $\text{CR}_{t\bar{t}+\geq 1c}^{\geq 6j}$ , respectively, allow for additional control over the  $t\bar{t}+ \geq 1c$  background. Distributions of the classification BDT are used in all nine signal regions to help isolate the  $t\bar{t}H$  signal. The binning of all distributions is optimized for sensitivity, while keeping the statistical uncertainties related to the model below 20%. This avoids possible bias due to statistical fluctuations in the model. Most distributions entering the fit are shown in section 6.7.2, while several regions of the dilepton channel are included in appendix section A.2.

The expected performance of the analysis can be studied by fitting the model to an Asimov dataset as described in section 5.2.3. This results in an expected signal strength measurement of  $\mu_{t\bar{t}H} = 1.00_{-0.58}^{+0.61}$ , while the free-floating normalization factors for the  $t\bar{t}+ \geq 1b$  and  $t\bar{t}+ \geq 1c$  backgrounds are expected to be measured as  $k(t\bar{t}+ \geq 1b) = 1.00_{-0.08}^{+0.09}$  and  $k(t\bar{t}+ \geq 1c) = 1.00_{-0.20}^{+0.20}$ . The analysis sensitivity is expected to be dominated by the regions in the single-lepton channel, and strongly affected by  $\text{SR}_1^{\geq 6j}$ . When only including the single-lepton channel, the signal strength measurement is expected to be  $\mu_{t\bar{t}H} = 1.00_{-0.65}^{+0.68}$ . When performing the same fit, but with  $\text{SR}_1^{\geq 6j}$  excluded, the uncertainties increase, resulting in  $\mu_{t\bar{t}H} = 1.00_{-0.84}^{+0.85}$ . In contrast to this, a removal of  $\text{SR}_2^{\geq 6j}$  results

Table 6.4: Expected signal strength measurement in fits to an Asimov dataset. The  $\text{SR}_1^{\geq 6j}$  plays an important role in the overall sensitivity of the analysis.

| regions included in fit                                | $\mu_{t\bar{t}H}$      |
|--|------------------------|
| single-lepton and dilepton regions                     | $1.00^{+0.61}_{-0.58}$ |
| single-lepton regions                                  | $1.00^{+0.68}_{-0.65}$ |
| single-lepton regions, without $\text{SR}_2^{\geq 6j}$ | $1.00^{+0.71}_{-0.68}$ |
| single-lepton regions, without $\text{SR}_1^{5j}$      | $1.00^{+0.73}_{-0.70}$ |
| single-lepton regions, without $\text{SR}_1^{\geq 6j}$ | $1.00^{+0.85}_{-0.84}$ |

in  $\mu_{t\bar{t}H} = 1.00^{+0.71}_{-0.68}$ , while removing  $\text{SR}_1^{5j}$  results in  $\mu_{t\bar{t}H} = 1.00^{+0.73}_{-0.70}$ . Table 6.4 summarizes these configurations.

### 6.7.2 Fit to data

The analysis is optimized for sensitivity using the expected signal and background distributions from Asimov datasets. Data is used during the optimization stage only in signal-depleted regions to guide the definition of the background model and its associated uncertainties. This avoids biases in the analysis design due to the knowledge of the distribution of data in regions where significant signal contributions are expected for a SM  $t\bar{t}H$  signal. Fits to data that are sensitive to the signal are performed only after finalizing all decisions regarding the analysis design.

When fitting the model to data, the  $t\bar{t}H$  signal strength is measured as

$$\mu_{t\bar{t}H} = 0.84 \pm 0.29 \text{ (stat.)}^{+0.57}_{-0.54} \text{ (syst.)} = 0.84^{+0.64}_{-0.61}. \quad (6.1)$$

It is compatible with the SM prediction. The statistical uncertainty is evaluated in a second fit. For this fit, all nuisance parameters are set to the values minimizing  $t_\mu$ , the post-fit values  $\hat{\theta}$ . A fit of only the three free-floating parameters  $\mu_{t\bar{t}H}$ ,  $k(t\bar{t}+ \geq 1b)$ ,  $k(t\bar{t}+ \geq 1c)$  to data is performed, and the resulting uncertainty  $\Delta\mu_{t\bar{t}H}$  is interpreted as the statistical uncertainty for the signal strength. The systematic uncertainty component reported in equation (6.1) is obtained by subtracting the statistical component in quadrature from the total uncertainty.

The fit results for the free-floating normalization factors  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  are

$$\begin{aligned} k(t\bar{t}+ \geq 1b) &= 1.24 \pm 0.10, \\ k(t\bar{t}+ \geq 1c) &= 1.63 \pm 0.23. \end{aligned} \quad (6.2)$$

Another fit is performed where two signal strength parameters are used, one scaling  $t\bar{t}H$  in the dilepton channel, and another one scaling the single-lepton channel. The results of this so-called two- $\mu$  fit, as well as the nominal fit result with only a single signal strength parameter, are summarized in figure 6.10. The fit is used to validate the compatibility of the measurement in both channels. The compatibility between the two signal strengths in the two- $\mu$  fit is evaluated with a  $\chi^2$  test with one degree of freedom (corresponding to the additional signal strength parameter in the two- $\mu$  fit,

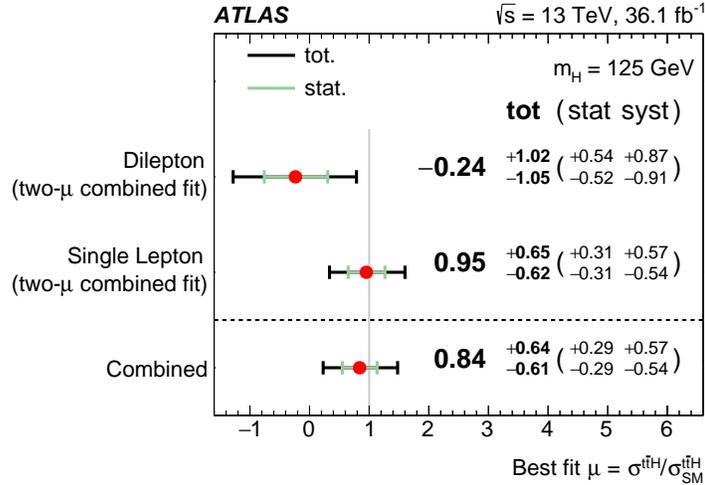


Figure 6.10: Measurement of the signal strength  $\mu_{t\bar{t}H}$  when fitting the model to data. The two- $\mu$  fit is performed by fitting both dilepton and single-lepton channels, with two separate signal strength parameters affecting them. The nominal fit result, listed in the last row, is obtained by using a single signal strength parameter [1].

compared to the nominal fit). It compares the negative logarithm of the likelihood for the nominal and two- $\mu$  fits at the respective best-fit points. The probability to obtain a difference between the two signal strength parameters as large or larger than the one observed is 19%; the results are not incompatible with each other.

Separate fits to the single-lepton and dilepton channels are also performed. When fitting only data in the single-lepton channel, the signal strength is measured as  $\mu_{t\bar{t}H} = 0.67^{+0.71}_{-0.69}$ . The fit to the dilepton channel results in  $\mu_{t\bar{t}H} = 0.11^{+1.36}_{-1.41}$ . Both of these individual measurements are smaller than the signal strength extracted from the combined fit, but compatible with it. This effect is caused by large correlations between systematic uncertainties affecting both channels.

### Distributions before and after the fit

A summary of all single-lepton regions considered in the fit is shown in figure 6.11. Every region is shown as a single bin in this distribution, even though some regions are considered in the fit with multiple bins. The figure includes both the nominal distribution of the model (called pre-fit), as well as the distribution of the model with all parameters set to their best-fit values (called post-fit). For the pre-fit figure, the normalization factors  $k(t\bar{t} \geq 1b)$  and  $k(t\bar{t} \geq 1c)$  are both set to unity, and no uncertainty related to them is used to build the total uncertainty band shown in the figure. The post-fit figure contains the effect of their uncertainties, as determined by the fit. It also takes into account the correlations between all nuisance parameters, and the constraints of nuisance parameters. These constraints arise when the nuisance parameter uncertainty determined in the fit is smaller than the uncertainty originating from the associated constraint. In the figure, the  $t\bar{t}H$  signal is shown

contributing in red to the stacked histogram, and separately drawn as a dashed red line on top of the stacked histogram. The signal is normalized to the SM prediction pre-fit, and to the best-fit signal strength post-fit.

In some regions, the predicted yield from the model is smaller than the observed amount of events in data. This deficit is corrected well in the fit, as visible in the post-fit distribution. The total model uncertainties are also decreased in the post-fit distribution compared to the pre-fit distribution. This is due to correlations between nuisance parameters and their constraints.

Figure 6.12 shows the equivalent distribution of all regions pre- and post-fit for the dilepton channel. The pre-fit deficit visible in some regions is well adjusted by the fit.

The distributions of the control regions in the single-lepton channel that enter the fit with more than one bin are shown in figure 6.13. Figure 6.14 shows the resolved signal regions with exactly five jets, as well as the boosted region. The resolved signal regions with six or more jets are visualized in figure 6.15. The corresponding distributions for the dilepton regions are found in appendix section A.2. In figure 6.14 and figure 6.15, the  $t\bar{t}H$  distribution normalized to the total background is drawn as a dashed red line.

All distributions of the model are in agreement with data within the associated uncertainties both pre- and post-fit. The post-fit uncertainties are reduced due to the correlations between nuisance parameters, and their constraints. The agreement of the model with data is improved post-fit.

The post-fit modeling of data is also studied with distributions that are not directly used in the fit; data is described well by the post-fit model for the wide range of distributions investigated. The post-fit distribution of the MEM discriminant is shown in section 7.4.1. Distributions not directly used in the fit are generally still well described post-fit.

### 6.7.3 Dominant nuisance parameters and sources of uncertainty

The influence of a given nuisance parameter on the fit result can be evaluated by studying the fit with this nuisance parameter fixed to specific values. The impact  $\Delta\mu$  of a nuisance parameter on the signal strength is defined as the shift in the signal strength  $\mu_{t\bar{t}H}$  between the nominal fit and a fit with the nuisance parameter held fixed at  $\hat{\theta} \pm x$ . The pre-fit impact is obtained by considering  $x = \Delta\theta = 1$ . Since nuisance parameters may get constrained during the fit, the post-fit impact of a nuisance parameter, evaluated with  $x = \Delta\hat{\theta} \leq 1$ , can be smaller than its pre-fit impact.

The 20 dominant nuisance parameters in the fit, ranked according to their impact  $\Delta\mu$  on the signal strength, are shown in figure 6.16. The pre-fit impact, where the nuisance parameter is fixed to  $\hat{\theta} \pm \Delta\theta$ , is shown as empty blue and cyan rectangles. The blue rectangle corresponds to fixing the parameter to  $\hat{\theta} + \Delta\theta$ , for the cyan rectangle it is fixed to  $\hat{\theta} - \Delta\theta$ . Similarly, the filled blue and cyan rectangles show the post-fit impact, obtained from fits with the nuisance parameter fixed to  $\hat{\theta} \pm \Delta\hat{\theta}$ . The upper axis on the figure shows the scale of the impact  $\Delta\mu$ . The pull of a nuisance parameter is defined by comparing its best-fit point  $\hat{\theta}$  to its nominal pre-fit value  $\theta_0$ , and dividing the difference by its pre-fit uncertainty. This pull, given by  $\frac{\hat{\theta} - \theta_0}{\Delta\theta}$ , is shown by the black points for the nuisance parameters. The lower axis in the figure shows the corresponding scale; all best-fit points  $\hat{\theta}$  are within

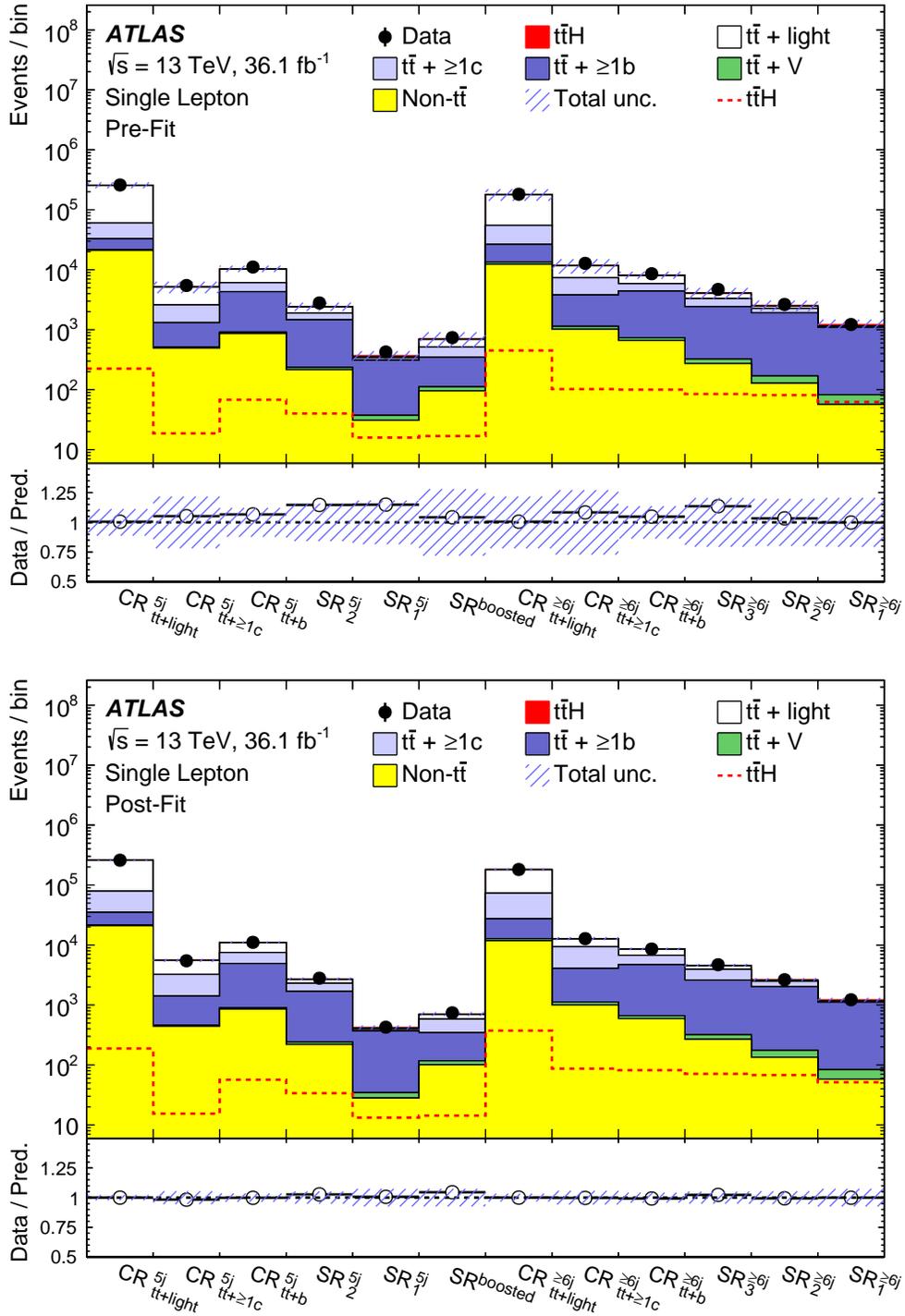


Figure 6.11: Overview of the yields in all single-lepton regions pre-fit (top) and post-fit (bottom). The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram. It is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit [1].

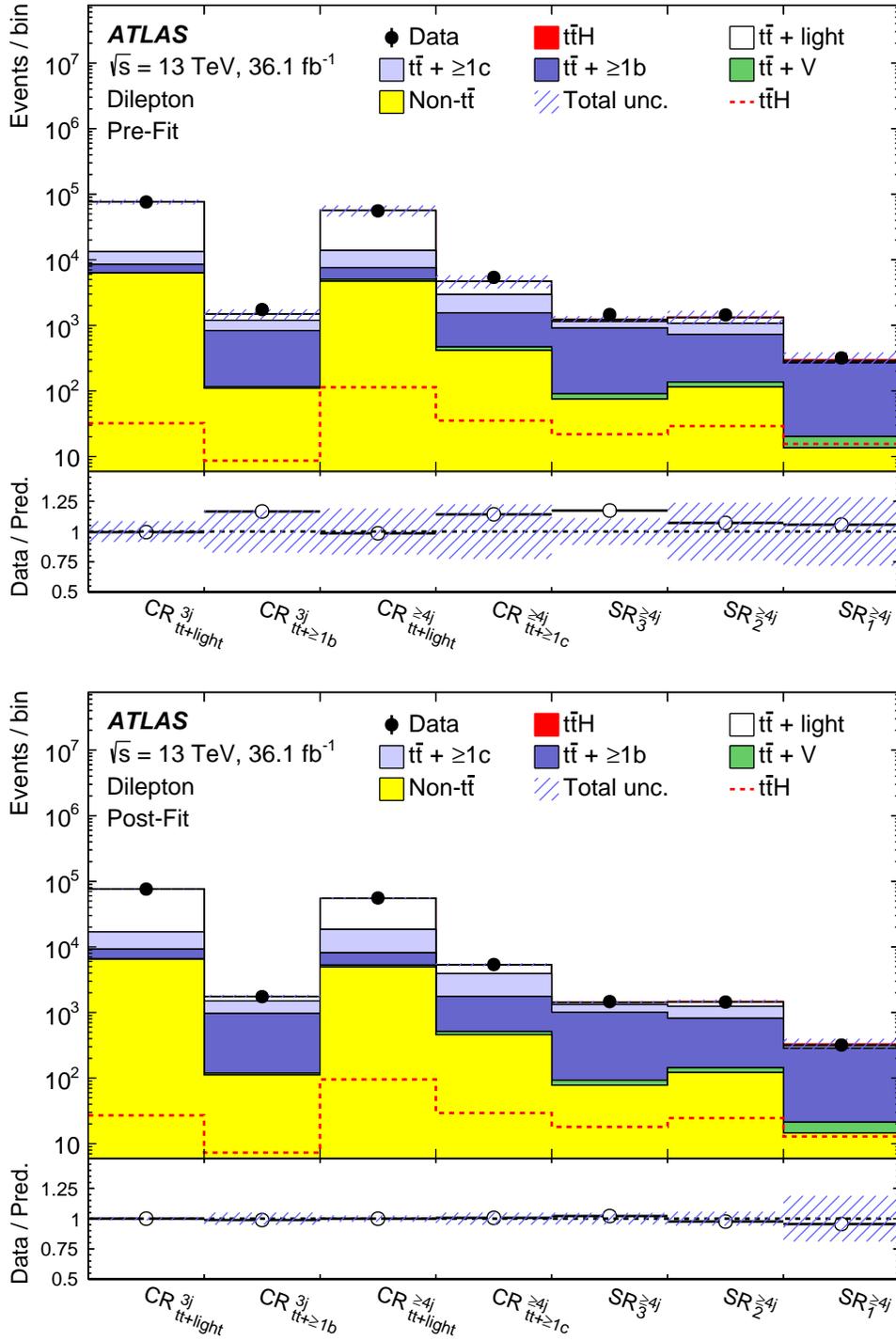


Figure 6.12: Overview of the yields in all dilepton regions pre-fit (top) and post-fit (bottom). The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram. It is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit [1].

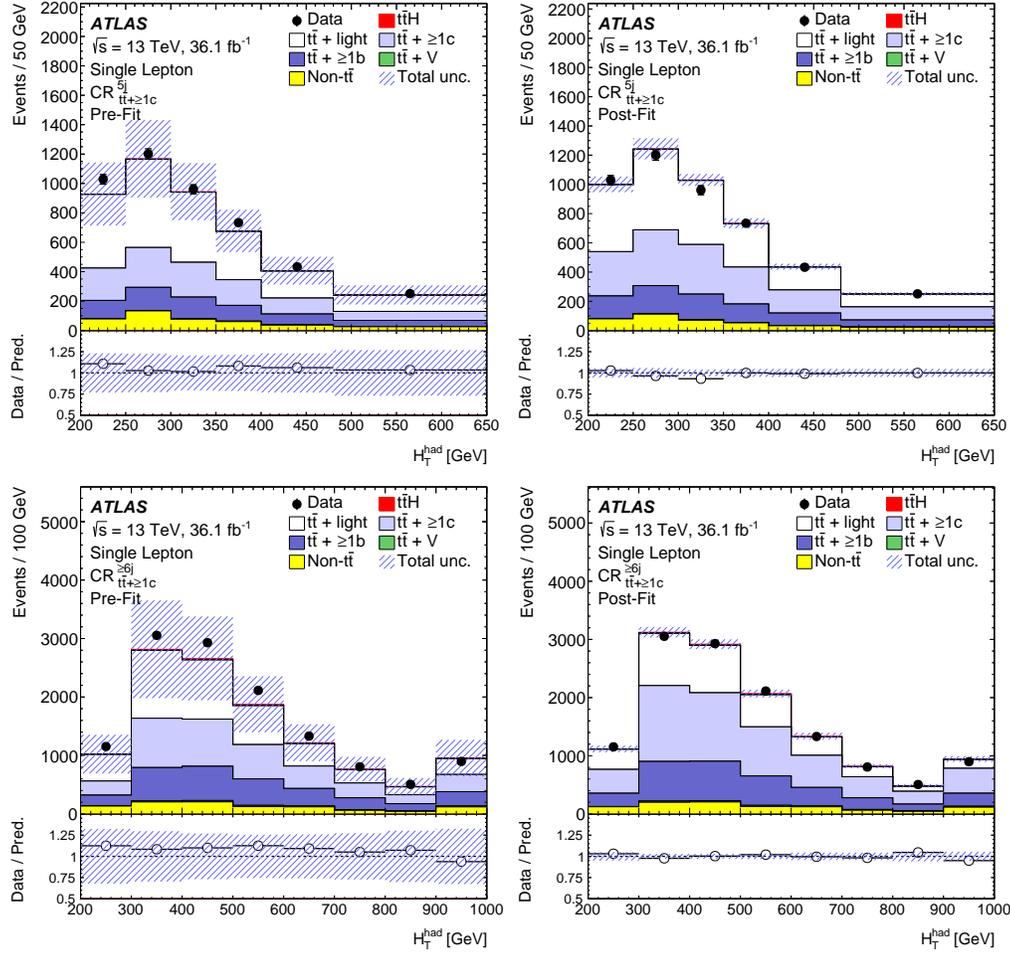


Figure 6.13: Comparison between data and the model for the control regions  $CR_{t\bar{t}+1c}^{5j}$  (top) and  $CR_{t\bar{t}+1c}^{\ge 6j}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. Events with  $H_T^{\text{had}} < 200$  GeV or  $H_T^{\text{had}} > 650$  GeV are included in the leftmost and rightmost bins of the  $CR_{t\bar{t}+1c}^{5j}$  distributions, respectively. Similarly, events with  $H_T^{\text{had}} < 200$  GeV or  $H_T^{\text{had}} > 1000$  GeV are also included in the outermost bins of the  $CR_{t\bar{t}+1c}^{\ge 6j}$  distributions [1].

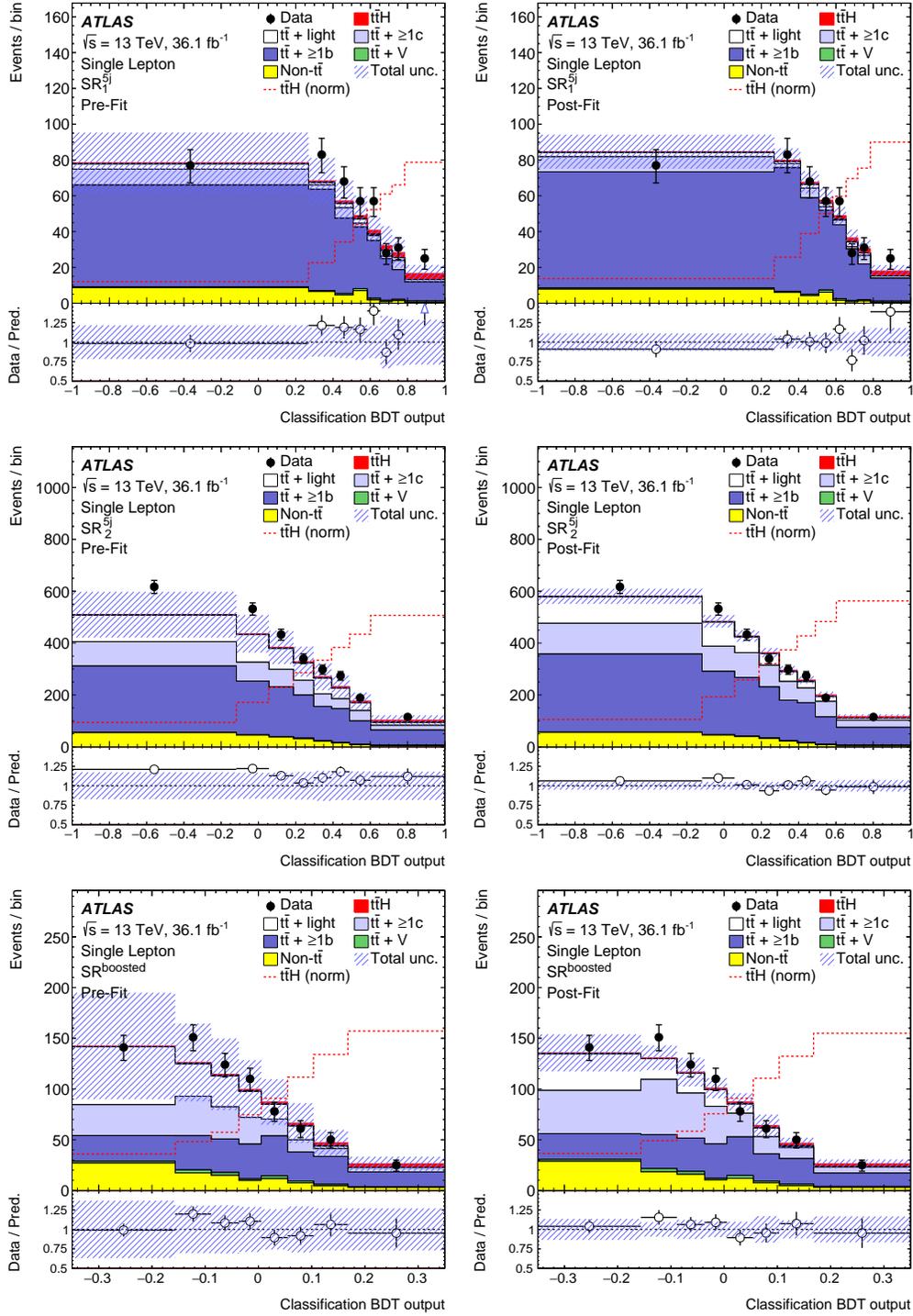


Figure 6.14: Comparison between data and the model for the signal regions  $SR_1^{5j}$  (top),  $SR_2^{5j}$  (middle) and  $SR^{\text{boosted}}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line [1].

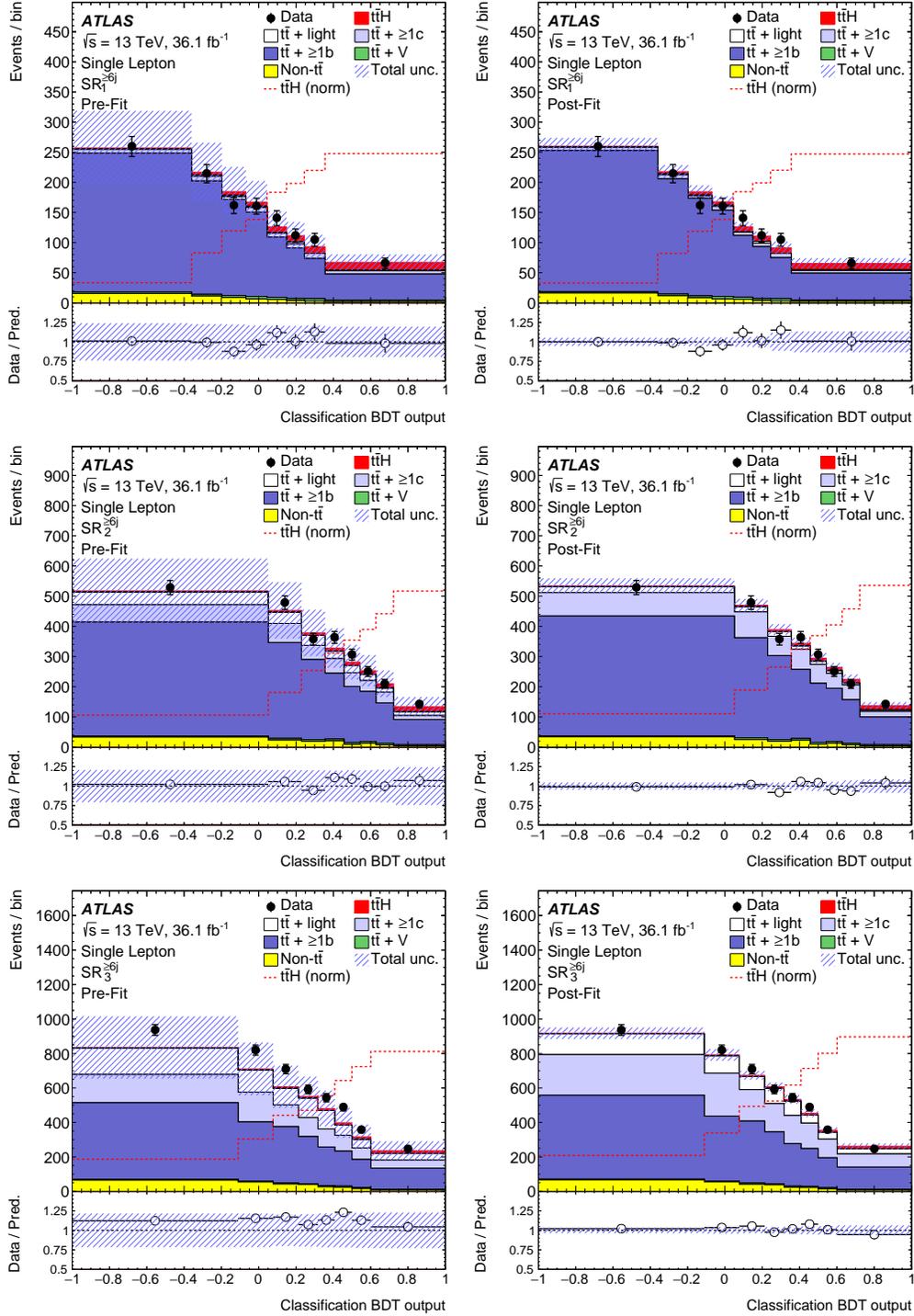


Figure 6.15: Comparison between data and the model for the signal regions  $SR_1^{\geq 6j}$  (top),  $SR_2^{\geq 6j}$  (middle) and  $SR_3^{\geq 6j}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line [1].

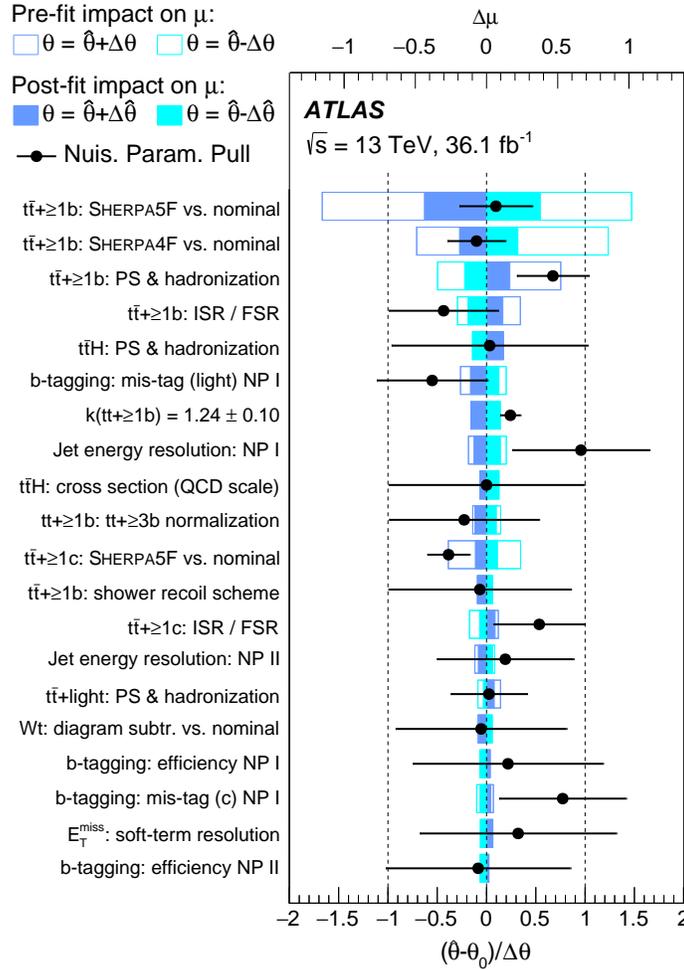


Figure 6.16: The 20 dominant nuisance parameters in the fit, ranked according to their impact on the signal strength. The empty rectangles correspond to the pre-fit impact, while the filled rectangles show the post-fit impact per nuisance parameter. The upper axis shows the impact  $\Delta\mu$ . The pull  $\frac{\hat{\theta} - \theta_0}{\Delta\theta}$  of the nuisance parameter is shown as black points, with the vertical black lines visualizing the post-fit nuisance parameter uncertainty  $\Delta\hat{\theta}$  [1].

the pre-fit uncertainties  $\Delta\theta$ . The vertical black lines show the post-fit nuisance parameter uncertainty  $\Delta\hat{\theta}$ . Neither pre-fit uncertainty  $\Delta\theta$  nor nominal value  $\theta_0$  are defined for the free-floating normalization factor  $k(t\bar{t}+\geq 1b)$ , so they are set to  $\Delta\theta = \theta_0 = 1$  in this figure. The pre-fit impact of  $k(t\bar{t}+\geq 1b)$  is not drawn, since it is not well-defined. Nuisance parameters corresponding to statistical uncertainties in the samples, described in section 6.6.1, are excluded from this figure.

The dominant uncertainty source is related to the modeling of  $t\bar{t}+\geq 1b$ , and described by the comparison of the nominal POWHEG+PYTHIA 8  $t\bar{t}$  sample to the SHERPA5F prediction. Additional uncertainties related to  $t\bar{t}+\geq 1b$  modeling, also derived by comparing the nominal POWHEG+PYTHIA 8  $t\bar{t}$  to alternative samples, follow in the ranking. Multiple systematic uncertainties related to the modeling of  $t\bar{t}H$  show up, and the modeling of  $t\bar{t}+\geq 1c$  and  $t\bar{t}+\text{light}$  also plays a role, albeit much decreased compared to  $t\bar{t}+\geq 1b$ . When considering experimental sources, the dominant contribu-

tions are related to  $b$ -tagging and jet energy resolution. The impact of all uncertainties not included in figure 6.16 is small; when removing them from the fit, the total signal strength uncertainty decreases by 5%. The correlations between the most highly ranked nuisance parameters are shown in appendix section A.3.

### Uncertainties grouped by source

Table 6.5 shows contributions to the measured total signal strength uncertainty  $\Delta\mu$ , grouped by sources of uncertainty. The contributions are obtained by fixing all nuisance parameters within one group to their best-fit values and repeating the fit, resulting in a reduced uncertainty  $\Delta\mu' < \Delta\mu$ . The impact of a group is defined by subtracting this uncertainty in quadrature from the total uncertainty in the nominal fit,  $\sqrt{(\Delta\mu)^2 - (\Delta\mu')^2}$ . Due to correlations between the nuisance parameters in different groups, the quadrature sum of all sources differs from the total uncertainty. The total statistical uncertainty is evaluated by fixing all nuisance parameters except the free-floating normalization factors  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$ , thus including their contribution as well as the contribution from data statistics. In contrast, the intrinsic statistical uncertainty is obtained from a fit where these normalization factors are also fixed to their post-fit values. The statistical uncertainty in background modeling includes both the effects from statistical uncertainties in the nominal MC samples, as well as the data-driven estimation of the fake lepton background in the single-lepton channel.

The modeling of  $t\bar{t}+ \geq 1b$  is the dominant source of uncertainty in the analysis and limits its sensitivity. Another large source of uncertainty is the background model statistical uncertainty due to the finite amount of MC events generated. The experimental sources of uncertainty with the largest impact are related to  $b$ -tagging and jet calibration. The modeling of the  $t\bar{t}H$  signal and  $t\bar{t}+ \geq 1c$  play a smaller role, followed by the remaining sources listed.

#### 6.7.4 Validation studies

In order to test the fit model, a dataset of simulated events is built where the  $t\bar{t}$  sample is replaced by a prediction generated with POWHEG+PYTHIA 6. When fitting the nominal model to this dataset, no significant bias in the signal strength measurement is observed.

Dedicated studies are performed to understand the pulls and constraints observed in the fit. The origin of pulls is investigated by splitting them into multiple components, such that different components act on different regions or samples. Their effect on the signal extraction is also studied by excluding bins enriched in  $t\bar{t}H$  signal. In general, the pulls serve to correct predictions of  $t\bar{t}$ , without biasing the signal strength measurement.

The constraints observed when fitting data are compatible with the constraints expected from fitting the Asimov dataset, and from fitting the dataset built with the POWHEG+PYTHIA 6 prediction for  $t\bar{t}$ .

The modeling of  $t\bar{t}$  has a dominant impact on the analysis. Alternative ways to defined the  $t\bar{t}$  model and associated systematic uncertainties have been studied, and they lead to compatible results [132].

Table 6.5: Contributions to the signal strength uncertainty, grouped by sources. The total statistical uncertainty includes effects from the  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  normalization factors, while the intrinsic statistical uncertainty does not. The background model statistical uncertainty includes effects from statistical uncertainties in nominal MC samples and the data-driven fake lepton estimate in the single-lepton channel [1].

| Uncertainty source   | $\Delta\mu$ |       |
|--|-------------|-------|
| <i>Systematic uncertainties</i>                              |             |       |
| $t\bar{t}H$ modeling   | +0.22       | -0.05 |
| $t\bar{t}+ \geq 1b$ modeling                                 | +0.46       | -0.46 |
| $t\bar{t}+ \geq 1c$ modeling                                 | +0.09       | -0.11 |
| $t\bar{t}+ \text{light}$ modeling                            | +0.06       | -0.03 |
| Other background modeling                                    | +0.08       | -0.08 |
| Background model statistical uncertainty                     | +0.29       | -0.31 |
| $b$ -tagging efficiency and mis-tag rates                    | +0.16       | -0.16 |
| Jet energy scale and resolution                              | +0.14       | -0.14 |
| Jet vertex tagger, pile-up modeling                          | +0.03       | -0.05 |
| Light lepton ( $e, \mu$ ) identification, isolation, trigger | +0.03       | -0.04 |
| Luminosity   | +0.03       | -0.02 |
| Total systematic uncertainty                                 | +0.57       | -0.54 |
| <i>Statistical uncertainties</i>                             |             |       |
| $t\bar{t}+ \geq 1b$ normalization                            | +0.09       | -0.10 |
| $t\bar{t}+ \geq 1c$ normalization                            | +0.02       | -0.03 |
| Intrinsic statistical uncertainty                            | +0.21       | -0.20 |
| Total statistical uncertainty                                | +0.29       | -0.29 |
| Total uncertainty  | +0.64       | -0.61 |

### 6.7.5 Observed significance and upper limits

The observed signal strength represents an excess over the SM background of 1.4 standard deviations. The expected sensitivity, evaluated from a dataset of simulated events, is 1.6 standard deviations. The dataset used for this calculation corresponds to the post-fit model, with all nuisance parameters adjusted to their best-fit values.

A signal strength  $\mu_{t\bar{t}H} > 2.0$  is excluded at the 95% confidence level. Figure 6.17 summarizes 95% confidence level upper limits on the signal strength  $\mu_{t\bar{t}H}$ . The dashed black line shows the expected median limit, corresponding to the median of the distribution of pseudo-experiments generated under a background-only hypothesis with  $\mu_{t\bar{t}H} = 0$ . Green and yellow bands correspond to the ranges containing 68% and 95% of these limits. The expected limit under a hypothesis including the SM  $t\bar{t}H$  signal is drawn as a dashed red line. All expected limits are calculated using the dataset of simulated events corresponding to the post-fit model. The observed limits are shown as solid black lines. The limits for the two- $\mu$  fits are derived by simultaneously fitting the single-lepton and dilepton channels, but using two independent signal strength parameters for them.

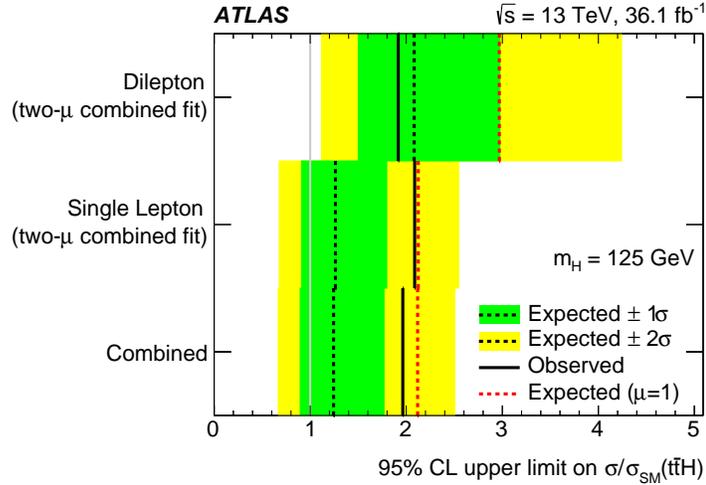


Figure 6.17: 95% confidence level (CL) upper limits on the signal strength  $\mu_{t\bar{t}H}$ , derived in a combined fit to single-lepton and dilepton channels with two independent signal strength parameters (two- $\mu$  fit), as well as a fit with a single signal strength (combined fit) [1].

### 6.7.6 Summary distribution of events

Figure 6.18 shows the post-fit distribution of  $t\bar{t}H$  signal ( $S$ ) and total background ( $B$ ) events as a function of  $\log_{10}(S/B)$ , as well as the distribution of data. It is built by evaluating  $\log_{10}(S/B)$  of every bin in the analysis, and combining the bins with similar  $\log_{10}(S/B)$  to form this distribution. The evaluation is performed with the SM prediction for the signal strength,  $\mu_{t\bar{t}H}$ . The upper panel shows the total background, as well as the contribution from  $t\bar{t}H$  signal with signal strength corresponding to its best-fit value in red, and corresponding to the signal strength excluded at the 95% confidence level in orange. The lower panel of the figure shows the difference between data and post-fit background model taken from the nominal fit, divided by its uncertainty (shown with hashed lines). The result of a background-only fit with a fixed signal strength  $\mu_{t\bar{t}H} = 0$  is shown as a dashed black line. It underestimates the yields observed in data in bins of high  $\log_{10}(S/B)$ . The red line corresponds to the nominal fit, including the signal strength at its best-fit value. The orange dashed line represents the signal strength scaled to its value excluded at the 95% confidence level. Events with  $\log_{10}(S/B) < -2.7$  are included in the first bin.

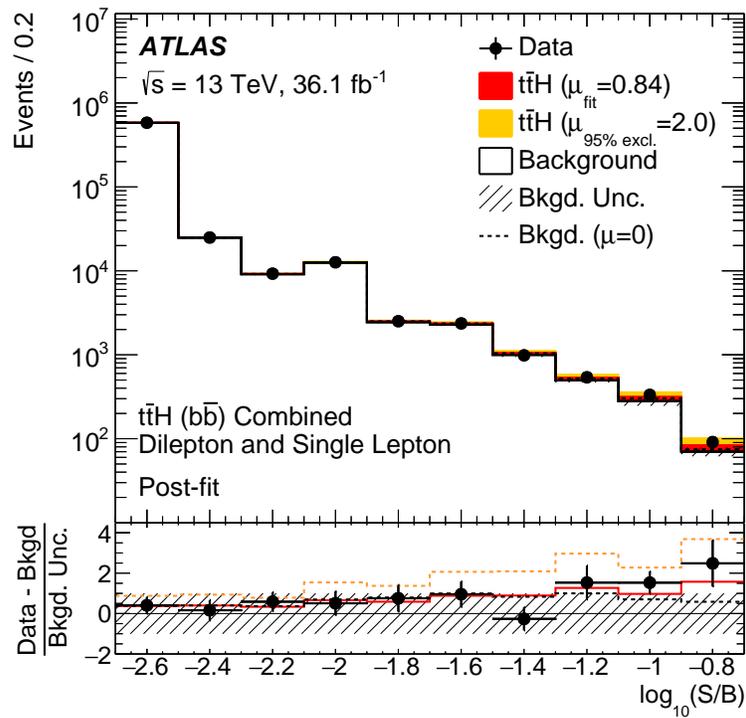


Figure 6.18: Post-fit yields of signal ( $S$ ), total background ( $B$ ), and observed data, shown as a function of  $\log_{10}(S/B)$ . Contributions of the signal, when scaled to its best-fit signal strength value, are drawn in red, while contributions with the signal strength set to its value excluded at the 95% confidence level are drawn in orange. The lower panel shows the difference of observed data and various fit models to the total background taken from the nominal fit [1].

## 7. The matrix element method for $t\bar{t}H(b\bar{b})$

As discussed in section 5.2.2, the optimal test statistic to distinguish between two hypotheses is given by a likelihood ratio. In the  $t\bar{t}H(b\bar{b})$  search, the two hypotheses are distinguished by the signal strength  $\mu_{t\bar{t}H}$ . The signal hypothesis predicts that both  $t\bar{t}H$  and background processes take place and result in a set of reconstructed objects in the ATLAS detector for every event. The background hypothesis with  $\mu_{t\bar{t}H} = 0$  specifies that no observed events are due to the  $t\bar{t}H$  process. Events generated under both of these hypotheses can be generated with MC methods, but the individual likelihoods cannot be evaluated directly. These likelihoods involve the integration over millions of random variables in the description of parton showering, hadronization, and interactions with the detector. When factorizing the contributions to the likelihood into parton level kinematics and all remaining effects, the likelihood ratio can be estimated directly with machine learning methods [133].

The MEM provides an approximation for the individual likelihoods. They are calculated from first principles, with sufficient simplifications to make the calculation computationally feasible. This method was first used for a top quark measurement by the D0 Collaboration [134], following an original proposal in reference [135]. It has been used in many analyses since, including the  $t\bar{t}H(b\bar{b})$  search performed by the ATLAS collaboration in Run-1 of the LHC [88]. An overview of the MEM from an experimental viewpoint is provided in reference [136], and multiple software implementations exist [137, 138].

This chapter describes an implementation of the MEM for the single-lepton channel of the  $t\bar{t}H(b\bar{b})$  search with  $36.1 \text{ fb}^{-1}$  of Run-2 data from the LHC. Details about this search are provided in chapter 6. The implementation defines a discriminant between  $t\bar{t}H$  and the dominant  $t\bar{t} + b\bar{b}$  background, and is used as an input to the classification BDT in the  $\text{SR}_1^{\geq 6j}$  region. As the MEM calculation involves an integral over a function that is computationally expensive, the implementation focuses on providing a good discriminant at a reasonable computational cost.

The chapter starts with a general description of the MEM in section 7.1, followed by more specific details regarding the implementation for  $t\bar{t}H(b\bar{b})$  in section 7.2. Technical aspects for the MEM calculation are discussed in section 7.3. The performance and modeling of data is summarized in section 7.4, and section 7.5 describes additional studies where the method is used to reconstruct the  $t\bar{t}H$  and  $t\bar{t} + b\bar{b}$  system.

### 7.1 The matrix element method

The MEM approximates the probability density  $f(\vec{X}|\alpha)$  for an observed set of reconstructed objects  $\vec{X}$  in an event and hypothesis  $\alpha$ . Two hypotheses are considered here. In the signal hypothesis, all

events are produced via the signal process  $S$ . The background hypothesis specifies that events are only produced via background processes  $B$ . The corresponding likelihoods are  $L_S$  and  $L_B$ , respectively. If the signal process exist, events generally should be produced both via signal and background processes. The aim of the method described in this chapter is however not the test of  $\mu_{t\bar{t}H} > 0$  versus  $\mu_{t\bar{t}H} = 0$ , where both hypotheses include contributions from background processes. Instead, the likelihood ratio  $L_S/L_B$  is used as a discriminant to distinguish whether any observed event is more compatible with having been produced by a  $t\bar{t}H$  process, than with having been produced by a background process.

The full expression for  $f(\vec{X}|\alpha)$  involves the description of parton production, their showering and hadronization, and the subsequent detector interactions and the reconstruction into observed objects. The MEM factorizes the contributions into a parton level process, and groups together all remaining contributions. These two components will be described in the following, resulting in an approximate expression for the likelihoods  $L_S$  and  $L_B$ .

### 7.1.1 Parton level

The probability density is proportional to the differential cross-section  $d\sigma$ , which can be written as [8]

$$d\sigma_\alpha(p_a, p_b, \vec{Y}) = \frac{1}{F} (2\pi)^4 \delta^4\left(p_a + p_b - \sum_{i=1}^N p_i\right) |\mathcal{M}_\alpha(p_a, p_b, \vec{Y})|^2 d\vec{\Phi}_{\vec{Y}} \quad (7.1)$$

at the parton level for hypothesis  $\alpha$ . The squared matrix element,  $|\mathcal{M}_\alpha|^2$ , describes the transition probability from the interacting initial states  $p_a$  and  $p_b$  to the  $N$  final state partons  $p_i$ , all characterized by their four-momenta. The flux factor  $F$  in the laboratory frame is given by  $F = 4\sqrt{(p_a \cdot p_b)^2 - m_a^2 m_b^2}$ . The Dirac delta distribution  $\delta^4$  enforces conservation of four-momentum. In the following,  $p_{\text{net}} = p_a + p_b - \sum_i p_i$  will be used. The final state configuration at parton level is called  $\vec{Y}$ , and described by

$$d\vec{\Phi}_{\vec{Y}} = \prod_{i=1}^N \frac{d^3\vec{p}_i}{(2\pi)^3 2E_i}, \quad (7.2)$$

with three degrees of freedom per on-shell final state parton, and a total of  $N$  final state partons.

The initial states  $p_a$  and  $p_b$  carry unknown momentum fractions  $x_1$  and  $x_2$  in proton–proton collisions. As described in section 2.2.3, the differential cross-section can be factorized into the hard scatter differential cross-section  $d\sigma_\alpha(p_a, p_b, \vec{Y})$ , and contributions from PDFs. The resulting differential cross-section,

$$d\sigma_\alpha(\vec{Y}) = \sum_{j,k} \int_{x_1, x_2} f_j(x_1) f_k(x_2) d\sigma_\alpha(p_a, p_b, \vec{Y}) dx_1 dx_2, \quad (7.3)$$

is marginalized over the colliding partons by summing all flavors  $i, j$  and integrating over all allowed momentum fractions  $x_1, x_2$ . The total cross-section is obtained when also performing the integral over all final state configurations,

$$\sigma_\alpha = \int_{\vec{Y}} d\sigma_\alpha(\vec{Y}). \quad (7.4)$$

The parton-level probability density is given by

$$f(\vec{Y}|\alpha) = \frac{1}{\sigma_\alpha} d\sigma_\alpha(\vec{Y}). \quad (7.5)$$

### 7.1.2 Reconstructed objects

The correspondence between parton level and reconstructed objects  $\vec{X}$  is described by a transfer function and a sum over all possible permutations of objects,  $\sum_{\text{perm.}} T(\vec{X}|\vec{Y})$ . The permutations describe assignments between partons and reconstructed objects. The joint probability density  $f(\vec{X}, \vec{Y}|\alpha)$  for an observed set of reconstructed objects  $\vec{X}$ , parton-level configuration  $\vec{Y}$ , and hypothesis  $\alpha$  is written as

$$f(\vec{X}, \vec{Y}|\alpha) = \left[ \sum_{\text{perm.}} T(\vec{X}|\vec{Y}) \right] f(\vec{Y}|\alpha), \quad (7.6)$$

factorizing into a term describing the parton level, and another term describing all other effects. The corresponding likelihood for the observation  $\vec{X}$  to be consistent with hypothesis  $\alpha$  is obtained when performing the marginalization over the space of parton level configurations,

$$L_\alpha = P(\vec{X}|\alpha) = \frac{1}{\sigma_\alpha} \sum_{\text{perm.}} \int_{\vec{Y}} T(\vec{X}|\vec{Y}) d\sigma_\alpha(\vec{Y}). \quad (7.7)$$

The expression for the likelihood in equation (7.7) can be used to build the likelihood ratio  $L_S/L_B$ , which in turn provides a discriminant between signal and background events.

## 7.2 General approach for $t\bar{t}H(b\bar{b})$

The MEM relies on the calculation of likelihoods of the form given in equation (7.7). They are calculated for both the signal  $t\bar{t}H(b\bar{b})$  process and the dominant background in the  $\text{SR}_1^{\geq 6j}$  region, which is  $t\bar{t} + b\bar{b}$ . Both processes have  $N = 8$  final state partons at LO, described by the likelihood

$$L_\alpha = \frac{1}{\sigma_\alpha} \sum_{\text{perm.}} \sum_{j,k} \int \frac{1}{F} T(\vec{X}|\vec{Y}) f_j(x_1) f_k(x_2) (2\pi)^4 \delta^4(p_{\text{net}}) |\mathcal{M}_\alpha|^2 dx_1 dx_2 \prod_{i=1}^8 \frac{d^3 \vec{p}_i}{(2\pi)^3 2E_i}. \quad (7.8)$$

This likelihood requires a sum over the type of initial partons colliding, and an integration over their two momentum fractions. It also includes an integration over 24 kinematic degrees of freedom of the final state partons, and four constraints for four-momentum conservation. After resolving the delta distribution, an integration over a 22-dimensional phase space remains. The computation of this integral is not feasible when having to consider many different assignments of reconstructed objects to partons, while having to perform this calculation for millions of events. Several approximations simplify the calculation significantly.

### 7.2.1 Permutations

The likelihood calculation relies on the description of the relation between partons and reconstructed objects by the transfer function. A unique identification of the parton from which a given reconstructed object originated is generally not possible, and contributions from all possible assignments need to be considered. They are described by the sum  $\sum_{\text{perm.}}$  in equation (7.8).

Both the number of partons and the number of reconstructed objects can in general vary across events. Only LO processes are considered here, which guarantees exactly eight final state partons in

every event. All events in the  $SR_1^{\geq 6j}$  region have at least six reconstructed jets and exactly one reconstructed light charged lepton. There is no ambiguity in the assignment of the single reconstructed lepton to the parton level lepton. It is possible to uniquely assign one jet to each quark in every permutation, but there may be events where no jet corresponding to a given quark is reconstructed. Such cases can be addressed by integrating over the degrees of freedom describing the jet. This is computationally expensive and not done here. An example is briefly discussed in section 7.5.1.

Events can have more than six reconstructed jets due to higher order corrections to the LO matrix element calculation, pile-up interactions, or imperfect object reconstruction in the detector. For such events, it is in principle possible to sum over contributions from each possible subset of six jets. In each subset, multiple permutations need to be considered. The combinatorial complexity quickly increases for events with more than six jets, and it becomes computationally prohibitive to consider all possible permutations. The computational cost of the likelihood in equation (7.8) scales linearly with the amount of permutations considered. Only a limited amount of permutations is thus considered in practice.

### Jet selection strategy

Four jets need to be assigned to  $b$  quarks in every permutation, and a unique set of four jets is chosen for every event. Jets are ordered by decreasing tightness of the  $b$ -tagging operating point they satisfy. In the  $SR_1^{\geq 6j}$  region, at least four jets are guaranteed to pass the *very tight* operating point. The ambiguity in the ordering between jets satisfying the same operating point is resolved by additionally ordering such jets in decreasing order of their transverse momentum. The top four jets in this ranking are the selected  $b$ -jets, which are assigned to  $b$  quarks in the permutations considered.

Two more jets need to be selected and assigned to the quarks originating from the decay of the  $W$  boson. The invariant mass  $m_{jj}$  of every pair of two jets, excluding jets that were already selected as  $b$ -jets, is calculated. For every pair, the distance  $|m_W - m_{jj}|$ , with  $m_W = 80.4$  GeV, is calculated. The pair that minimizes this distance is selected and assigned to the two quarks from the  $W$  boson decay. These two jets are called the selected light jets.

### Jet permutations

There is no ambiguity in the assignment of the selected light jets to the quarks originating from the  $W$  boson decay. The four selected  $b$ -jets need to be assigned to the four  $b$  quarks. Two are assigned to the decay products of the Higgs boson, and one each is assigned to the  $b$  quarks originating from the decay of both top quarks. There are four possible quarks to assign each selected  $b$ -jet to. The likelihood is unchanged when exchanging the jets assigned to the Higgs boson decay products. This results in  $4!/2 = 12$  possible permutations that need to be considered for every event.

When not relying on  $b$ -tagging information in the jet selection strategy, a selection of six jets could be made, and these jets could be assigned to all six quarks. The likelihood is unchanged when exchanging jets assigned to the Higgs boson decay products, and when exchanging those assigned to the  $W$  boson final states. There are thus  $6!/2 \cdot 2 = 180$  permutations for every event in this case. This

would slow down the likelihood calculation by more than an order of magnitude; the use of  $b$ -tagging information is essential to control the computational cost.

Permutations could also be considered only if they satisfy kinematic constraints. When calculating the  $t\bar{t}H$  signal likelihood, permutations could be rejected if the invariant mass of the two  $b$ -jets assigned to the decay of the Higgs boson is very different from the Higgs boson mass. No such kinematic constraints are used in this implementation.

## 7.2.2 Transfer function

The transfer function is assumed to factorize into components, with one component per reconstructed and selected object,

$$\sum_{\text{perm.}} T(\vec{X}|\vec{Y}) = W^{\text{lep}}(p^l, \text{reco}|p^l) W^\nu(p^\nu, \text{reco}|p^\nu) \left[ \sum_{\text{perm.}} \prod_{i=1}^6 W_i^{\text{jet}}(p_i^{\text{jet}}|p_i^q) \right]. \quad (7.9)$$

It describes multiple assignments of reconstructed objects to partons, with one-to-one correspondences between objects in each permutations. When calculating the likelihood for a given event, specified by a set of reconstructed objects  $\vec{X}$ , contributions from regions of the parton level phase space are generally suppressed by the transfer function if the parton level kinematics  $\vec{Y}$  are very different from the measurement  $\vec{X}$ . The quantities  $p^l$ ,  $p^\nu$ , and  $p^q$  refer to the lepton, neutrino, and quark four-vectors on parton level, while  $p^l, \text{reco}$ ,  $p^\nu, \text{reco}$ , and  $p^{\text{jet}}$  are the corresponding quantities for the associated reconstructed objects.

In the case of charged leptons, the measurement of the ATLAS detector  $p^{l, \text{reco}}$  agrees well with the parton level lepton kinematics  $p^l$ . The corresponding component of the transfer function is approximated by a delta distribution,  $W^{\text{lep}}(p^l, \text{reco}|p^l) = \delta^3(\vec{p}^{l, \text{reco}} - \vec{p}^l)$ . Leptons on both parton level and the reconstructed objects are treated as massless.

There is no direct measurement of the neutrino four-momentum  $p^{\nu, \text{reco}}$ . Its kinematics can be constrained by requiring four-momentum conservation. It can also be related to the  $E_T^{\text{miss}}$  measurement of the ATLAS detector, which is not done in the implementation described here. The transfer function component  $W^\nu(p^{\nu, \text{reco}}|p^\nu)$  is treated as a uniform distribution.

The detector resolution for jet measurements is worse than for leptons. The measured jet direction is approximated to exactly correspond to the parton direction, and described by a delta distribution. The relation between measured jet energy and parton energy is described by a transfer function component. Details about this treatment are provided in section 7.3.4.

## 7.2.3 Remaining degrees of freedom

With these chosen approximations for the transfer function, the dimensionality of the integral in equation (7.8) can be reduced by resolving all delta distributions. The integral is performed over a 22-dimensional phase space. The lepton kinematics between reconstructed lepton and parton level lepton are assumed to match exactly, reducing the dimensionality by three. The directions for all six jets are also assumed to match exactly between parton and jet, resulting in an additional reduction by

two degrees of freedom for each of the six jets. After inserting and resolving these delta distributions into equation (7.8), an integral over a seven-dimensional phase space remains. This integral has to be evaluated for each of the 12 permutations considered in every event.

### Further possible simplifications

The integral may be simplified further when considering details of the kinematics in  $t\bar{t}H$  and  $t\bar{t} + b\bar{b}$  processes. The total width of the Higgs boson is predicted to be  $\Gamma_H = 4$  MeV by the SM [10], significantly below the resolution of the ATLAS detector. The  $t\bar{t}H$  matrix element significantly suppresses contributions from phase space regions where  $m_{bb} - m_H \gg \Gamma_H$ . In this expression,  $m_{bb}$  is the invariant mass of the two  $b$  quarks originating from the Higgs boson decay. Contributions to the likelihood from such regions are negligible. An additional constraint  $\delta(m_{bb} - m_H)$  can be introduced to reduce the degrees of freedom by one when calculating the  $t\bar{t}H$  signal likelihood.

The decay width of  $W$  bosons and top quarks are three orders of magnitude larger compared to the Higgs bosons, but similar arguments can be applied to reduce the degrees of freedom in the integration further. Particularly interesting is the decay of the  $W$  bosons into leptons. The charged lepton is assumed to be well measured, and the transverse components of the neutrino can be constrained by requiring four-momentum conservation in each event. One degree of freedom remains for the neutrino when treating it as massless and on-shell. When forcing the  $W$  boson to decay exactly on-shell, the resulting quadratic function can be solved for the neutrino momentum in the  $z$  direction.

The two approaches were tested, and both found to not improve the performance of the method. Consequently, neither of these additional simplifications are adopted for the  $t\bar{t}H(b\bar{b})$  analysis. The special treatment for the Higgs boson decay is only possible when calculating the  $t\bar{t}H$  likelihood, and thus leads to a different approach than in the calculation of the  $t\bar{t} + b\bar{b}$  likelihood. It results in a slight loss in discrimination power of the likelihood ratio. Requiring an on-shell decay of the  $W$  boson to leptons reduces the degrees of freedom in the integral required for the likelihood calculation, but also introduces two solutions that both need to be considered separately. This slows down the calculation, and slightly decreases discrimination power when implemented for both  $t\bar{t}H$  and  $t\bar{t} + b\bar{b}$  likelihoods. When only using this approximation in one of the likelihoods, but not the other, the discrimination power reduces more significantly.

### 7.2.4 Likelihoods and discriminant

For every event, the signal and background likelihood is calculated according to equation (7.8). The signal likelihood  $L_S$  is calculated using a  $t\bar{t}H$  matrix element, while the background likelihood  $L_B$  is calculated for the  $t\bar{t} + b\bar{b}$  process. In the calculation, several of the constant factors in the expression for the likelihood are not included, and additional constant factors are applied to prevent numerical underflow. The logarithm of the likelihoods  $\log_{10}(L_S)$  and  $\log_{10}(L_B)$ , visualized in section 7.4, is proportional to the logarithm of the likelihood in equation (7.8) up to a constant shift. The resulting

calculated variables can thus not be interpreted as a probability density  $P(\vec{X}|\alpha)$  without suitable normalization. This has no impact on the performance of the method.

The MEM discriminant is defined as the logarithm of the likelihood ratio,

$$\text{MEM}_{D1} = \log_{10}(L_S) - \log_{10}(L_B). \quad (7.10)$$

It is designed to discriminate between events originating from the  $t\bar{t}H$  process, which are characterized by large values of  $\text{MEM}_{D1}$ , and events from the  $t\bar{t} + b\bar{b}$  process. The  $t\bar{t} + b\bar{b}$  process dominates the  $\text{SR}_1^{\geq 6j}$  region, where the discriminant is used, and contributions from other processes are small. While the  $\text{MEM}_{D1}$  discriminant is not designed to discriminate against other processes besides  $t\bar{t} + b\bar{b}$ , it still helps discriminate against them.

### 7.2.5 Systematic uncertainties

No systematic uncertainty inherent to the calculation of  $\text{MEM}_{D1}$  is used in the  $t\bar{t}H(b\bar{b})$  analysis. The discriminant is a function of event properties, like any other multivariate analysis technique. It is calculated in the same way for both data and simulated events. An incorrect implementation of the likelihood calculation does not invalidate the method, but decreases its performance. Uncertainties covering the difference between the approximate event likelihood calculated via the MEM and the true likelihood are therefore not needed. The method relies on the correct modeling of kinematic features present in data. All systematic uncertainties described in section 6.6 are therefore propagated through the MEM calculation, and also through all other multivariate techniques used in the analysis.

## 7.3 Technical implementation

The likelihood calculation is steered by a framework based on reference [139]. It allows for the implementation of the integrand in equation (7.8) in both CUDA [140] and OPENCL [141]. The calculation can be performed both with central processing units (CPUs) and graphics processing units (GPUs). In both architectures, the integral evaluation is parallelized across many cores. Initial studies for the method were mostly performed with GPUs and CUDA, while the final implementation was using OPENCL on CPUs. The choice was motivated by the availability of resources. This implementation calculates both likelihoods on a Intel Xeon E5-2650 v2 CPU in 1.8 s. Around one million events per day can be processed on a cluster with 368 cores, which was used for the majority of the calculations.

This section lists additional details regarding the implementation of the integration, PDFs, the matrix elements, and transfer function.

### 7.3.1 Integration

The phase space integral in equation (7.8) is performed with MC methods, using the VEGAS algorithm [142]. An integral over a volume  $V$  can be approximated by sampling the integrand  $f(\vec{x})$  with  $N$  points

drawn from a probability density  $g(\vec{x})$ .

$$\int_V f(\vec{x}) d\vec{x} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{f(\vec{x}_i)}{g(\vec{x}_i)} \quad (7.11)$$

The standard deviation of the approximation decreases as  $1/\sqrt{N}$ , independent of the dimensionality of the integral. The VEGAS algorithm is adaptive and focuses on sampling regions where the integrand is large. It starts out using a uniform probability density  $g(\vec{x})$ , and the integration volume  $V$  is divided into a uniform grid of hypercubes. The size of the hypercubes is then adjusted, resulting in small hypercubes where the integrand is large, and large hypercubes where it is small. The number of integration points per hypercube remains the same, and the hypercube density is equivalent to the probability density  $g(\vec{x})$ . The algorithm thus focuses the sampling on the regions that contribute most to the total integral.

### Phase space details

As discussed in section 7.2.3, seven degrees of freedom remain to be integrated over in the likelihood evaluation. They are chosen to be the energies of the six quarks, as well as the momentum of the neutrino in the  $z$  direction. The momentum of the neutrino in the transverse directions is chosen such that the total transverse momentum of the LO parton system vanishes. The directions of the quarks are fixed to the directions of the jets they are assigned to in each permutation. The jet four-momenta are built from their measured energies, as well as their position in the  $\eta$  and  $\phi$  coordinates. Their transverse momenta are adjusted such that the invariant mass of selected  $b$ -jets is 4.7 GeV, corresponding to the  $b$  quark mass in the matrix element calculation, and the invariant mass of selected light jets vanishes. The same treatment is employed for the reconstructed lepton, which receives a vanishing invariant mass.

The integration is performed over the six quark energies  $E^q$ , with the integration volume restricted by a requirement depending on the energy of the jet assigned to the quark,  $E^q \in I_t(E^{\text{jet}})$ . The interval  $I_t$  is centered around the measured jet energy, and depends on the jet type  $t$ , with a different definition for  $b$ - and light jets. The definition of  $I_t$  is determined from the transfer function described in section 7.3.4. A distinction is made between  $b$ - and light jets, as a different parametrization is used for the respective transfer function components. The interval is defined as

$$I_t = [E^{\text{jet}} \cdot (1 - r), E^{\text{jet}} \cdot (1 + r)], \quad (7.12)$$

with parameter  $r$  defined separately for  $b$ - and light jets as  $r_b$  and  $r_l$ , respectively. These parameters are

$$\begin{aligned} r_b &= a \cdot \left( \frac{3.73}{E^{\text{jet}}} + 0.0736 \right), \\ r_l &= a \cdot \left( \frac{3.19}{E^{\text{jet}}} + 0.0856 \right). \end{aligned} \quad (7.13)$$

The setting  $a = 5$  is used, which limits the integration volume for faster evaluation of the integral, while not restricting the volume too much to avoid compromising the performance of the MEM<sub>D1</sub>

discriminant. The integration volume is furthermore restricted to regions with positive parton energy. Contributions from outside the integration volume considered are suppressed by the transfer function, and thus only have a small contribution to the likelihoods.

For light jets, the intervals contain more than 99% of the quark energy probability density distribution,  $W(E^q|E^{\text{jet}})$ , for any given jet energy. They contain around 80% for  $b$ -jets with  $E^{\text{jet}} = 25$  GeV, 90% for 50 GeV and around 95% for  $b$ -jets with  $100 \text{ GeV} < E^{\text{jet}} < 500$  GeV.

The integration range for the neutrino momentum in the  $z$  direction is restricted to be in the interval  $p_z^{\nu} \in [-1 \text{ TeV}, 1 \text{ TeV}]$  in the laboratory frame.

### Alternative parameterizations

The choice of the six quark energies as integration variables aligns the peak structure of the integrand, which is induced by the transfer function, with these integration variables. This is desirable for faster convergence of the integration. Additional peaks of the integrand due to the decay of intermediate particles into the final state objects are not aligned with the integration variables. It is possible to perform a change of integration variables to instead align the integration variable with these intermediate resonances. The integral over the energy of both  $b$  quarks originating from the Higgs boson decay can be replaced by an integral over the energy of one decay product and the invariant mass of the Higgs boson. No significant improvement in computation speed or power of the MEM<sub>D1</sub> discriminant is observed in this case. A similar transformation can be done for the two quarks from the decay of the  $W$  boson, again with no improvement in performance observed. Neither of these alternative integration variable choices are adopted in the final implementation of the method.

### Integration process

The integral over the seven degrees of freedom of equation (7.8) is performed with VEGAS, and optimized to provide a good discriminant within a reasonable amount of time. This requirement corresponds to being able to process around a million events per day.

The integration for the signal and background likelihoods are performed separately, and each of the 12 permutations per likelihood are also integrated separately. The integration procedure consists of multiple phases. Each phase can contain several rounds of integration. In one round of integration, up to 1024 phase space points are evaluated. The VEGAS grid of hypercubes is updated after every round to provide an improved sampling of the integrand in subsequent rounds.

The first phase consists of a single round of integration. The result for the integral obtained from this round of integration is not used any further, but this phase provides a first update to the hypercube grid. Subsequent integration rounds therefore yield better results.

The main integration phase follows, where three rounds of integration are performed. An estimate for the integral is obtained by combining the results from the three rounds. If the estimated uncertainty for the integral is below 1%, the integration process stops. This is the case only for a small fraction of events.

After the main integration phase, the contributions to the signal and background likelihoods from each permutation are compared. The likelihoods are  $L_\alpha$  for the two hypotheses, and consist of a sum of contributions  $L_\alpha^i$  per permutation  $i$ , which are all evaluated in separate integrals. If the likelihood of an individual permutation  $i$  is less than 1% of the largest likelihood for any permutation for this hypothesis, no further integration is performed for this permutation  $i$ . The contribution from such a permutation to  $L_\alpha$  is small, and precision for this permutation is thus less important. Generally the individual contributions to the likelihood range over multiple orders of magnitude, and the removal of small contributions speeds up the last integration phase, while not compromising the accuracy for the full likelihood.

The final integration phase consists of up to three more rounds of integration for all permutations that are not pruned in the previous phase. The integration process is stopped early if the estimated uncertainty for the integral is below 1% after any of these rounds.

The total likelihood is then obtained by summing the individual contributions from all 12 permutations for both the signal and background hypotheses.

### 7.3.2 Matrix elements

The matrix elements for the  $t\bar{t}H$  and  $t\bar{t} + b\bar{b}$  processes are provided by MADGRAPH5\_AMC@NLO [95]. Only LO Feynman diagrams are considered; the extension to NLO is challenging and comes at a significant computational cost [143]. The intermediate top quark pair in every diagram is required to decay to  $W$  bosons and bottom quarks. Only single-lepton final states are considered, and the  $W^+$  boson in all diagrams is forced to decay to leptons, while the  $W^-$  boson decays to quarks. The same matrix elements are used regardless of the charge and flavor of the lepton, which is equivalent to the assumption of lepton universality and invariance under charge conjugation.

Two topologies of diagrams can be distinguished, those initiated by interactions of gluons, and diagrams with a quark–antiquark pair in the initial state. The total cross-section for both processes is dominated by contributions from gluon–gluon diagrams. Figure 7.1 visualizes examples of the  $t\bar{t}H$  topologies. The three diagrams on the left are initiated by gluon–gluon interactions, while the diagram on the right is initiated by a quark–antiquark interaction.

In the MEM calculation, diagrams initiated by quark–antiquark interactions are not included, and only gluon–gluon interactions are considered. This is done for both the  $t\bar{t}H$  and  $t\bar{t} + b\bar{b}$  matrix elements. The calculation time is reduced by around 30% with this removal of diagrams, without significantly impacting the discrimination power of the likelihood ratio.

The matrix element code provided by MADGRAPH5\_AMC@NLO is optimized for performance. Non-physical helicity settings for the particles involved in the interaction are removed from the calculation. They do not contribute to the matrix element, and their removal results in a significant decrease in calculation time.

The renormalization scale is set to

$$\mu_R = \sqrt{\left(\sum_{i=1}^8 E_i\right)^2 - \left(\sum_{i=1}^8 p_{z,i}\right)^2}, \quad (7.14)$$

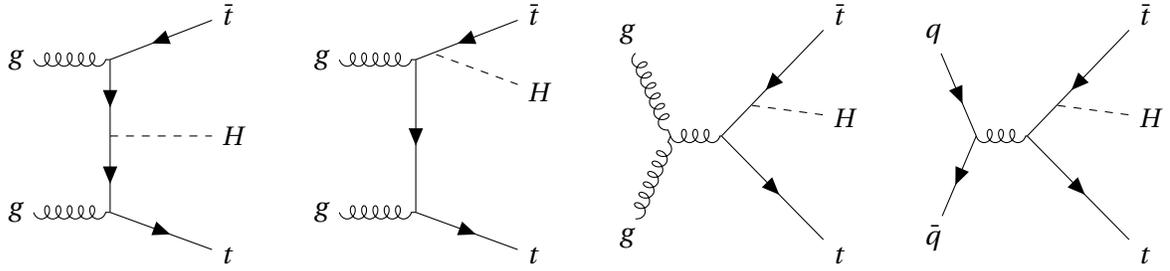


Figure 7.1: Topologies for  $t\bar{t}H$  production. The three diagrams on the left are initiated by gluons and are considered in the MEM calculation, while the quark–antiquark topology in the diagram on the right is neglected.

where the sum runs over the eight final state partons produced. This choice has a negligible impact.

The top quark mass in the calculation is set to 173 GeV, the Higgs boson mass is 125 GeV, the  $W$  boson mass is 80.42 GeV, and the bottom quark mass is 4.7 GeV.

### Signal hypothesis

The matrix element for the  $t\bar{t}H$  signal contains diagrams with an intermediate top quark pair and Higgs boson. The Higgs boson is forced to decay to a pair of bottom quarks. Diagrams initiated by quark–antiquark interactions are manually removed from the calculation.

### Background hypothesis

The matrix element for  $t\bar{t} + b\bar{b}$  describes the production of a pairs of top quarks and bottom quarks. Only diagrams initiated by gluon–gluon interactions are considered.

### 7.3.3 Parton distribution functions

PDFs are obtained from the CT10 PDF set [16], via an interface provided by LHAPDF [17]. The PDF distributions are saved to a grid, parameterized via momentum fraction  $x$  carried by the parton and momentum transfer  $Q^2$ . The integrand in the likelihood calculation uses this PDF grid with a linear interpolation, instead of evaluating the PDF via LHAPDF for every phase space point. This speeds up the calculation at a negligible loss of accuracy. The factorization scale choice is equivalent to the renormalization scale choice, listed in equation (7.14). No significant difference is observed when instead using a factorization scale setting of  $\mu_F = 173$  GeV.

### 7.3.4 Transfer function

The implementation for the transfer function used in the likelihood calculation provides the relation between the energies of reconstructed jets observed in the ATLAS detector, and the energies of the quarks they are associated to, in the form  $W(E^{\text{jet}}|E^q)$ . It suppresses contributions to the likelihood

from parton level configurations that are very different from the kinematics of the associated reconstructed jets. The transfer function component  $W(E^{\text{jet}}|E^q)$  is derived separately for  $b$ - and light jets, denoted as  $W_b(E^{\text{jet}}|E^q)$  and  $W_{\text{light}}(E^{\text{jet}}|E^q)$ , respectively.

These distributions are derived with a sample of  $t\bar{t}$  events generated with POWHEG+PYTHIA 6 [100–103, 117]. Events considered in the derivation need to have at least four jets, out of which at least two are  $b$ -tagged at the *loose* operating point. The distributions are subsequently validated with the nominal POWHEG+PYTHIA 8 sample used in the  $t\bar{t}H(b\bar{b})$  analysis, which is described in section 6.3.2. They describe the POWHEG+PYTHIA 8 sample well. Jets are matched to quarks produced in the decay of the top quark pair system, by requiring their distance to be  $\Delta R < 0.3$ . If multiple quarks are within this distance of the jet, the jet is matched to the quark that is closest in  $\Delta R$ . The components  $W_b(E^{\text{jet}}|E^q)$  and  $W_{\text{light}}(E^{\text{jet}}|E^q)$  are derived in a fit of the  $E^q - E^{\text{jet}}$  distribution, using  $b$  quarks or the lighter first and second generation quarks, respectively. Two different functional forms are used to describe  $b$ - and light jets.

### Light jets

Double Gaussian distributions describe the transfer function component for light jets, parameterized as

$$W_{\text{light}}(E^{\text{jet}}, E^q) = \frac{1}{\sqrt{2\pi}(\sigma_1 + A\sigma_2)} \left[ \exp\left(-\frac{(E^q - E^{\text{jet}} - \mu_1)^2}{2\sigma_1^2}\right) + A \cdot \exp\left(-\frac{(E^q - E^{\text{jet}} - \mu_2)^2}{2\sigma_2^2}\right) \right], \quad (7.15)$$

with the parameters  $\mu_1, \mu_2, \sigma_1, \sigma_2, A$  determined by the fit.

The resulting transfer function component  $W_{\text{light}}(E^{\text{jet}}, E^q)$  is visualized in figure 7.2 with solid lines, shown as a function of quark energy for various light jet energies. The distribution  $W_{\text{light}}$  suppresses the integrand in the MEM likelihood calculation in phase space regions where the quark energy is very different from the measured energy of the jet it is associated to. The width of the distribution increases with jet energy.

### $b$ -jets

The description for  $b$ -jets makes use of a crystal ball function [144, 145], which provides better fit results for such jets than the double Gaussian function. The core of the crystal ball function distribution is Gaussian, and it includes a tail region described by a power law. The transfer function component is described by

$$W_b(E^{\text{jet}}, E^q) = \frac{1}{\sigma(A+B)} \begin{cases} \exp\left(-\frac{(E^q - E^{\text{jet}} - \mu)^2}{2\sigma^2}\right), & \frac{E^q - E^{\text{jet}} - \mu}{\sigma} < \alpha, \\ C\left(D + \frac{E^q - E^{\text{jet}} - \mu}{\sigma}\right)^{-n}, & \frac{E^q - E^{\text{jet}} - \mu}{\sigma} \geq \alpha, \end{cases} \quad (7.16)$$

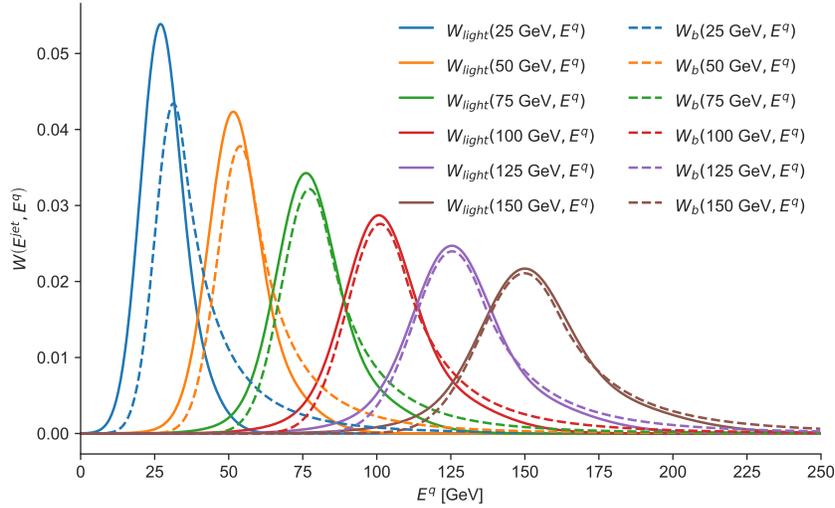


Figure 7.2: Transfer function components for  $b$ - and light jets, shown as a function of the energy  $E^q$  of the quark a jet is associated to. The distributions for light jets, described by a double Gaussian as  $W_{\text{light}}$ , are shown as solid lines for different jet energies. The corresponding distributions  $W_b$  for  $b$ -jets are shown as dashed lines, and are described by crystal ball functions.

with  $A, B, C, D$  defined as

$$\begin{aligned}
 A &= \frac{n}{|\alpha|(n-1)} \exp\left(-\frac{(|\alpha|)^2}{2}\right), \\
 B &= \sqrt{\frac{\pi}{2}} \left(1 + \operatorname{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right), \\
 C &= \left(\frac{n}{|\alpha|}\right)^n \exp\left(-\frac{(|\alpha|)^2}{2}\right), \\
 D &= \frac{n}{|\alpha|} - |\alpha|,
 \end{aligned} \tag{7.17}$$

The error function is defined as  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ . The parameters  $\mu, \sigma, \alpha, n$  are determined by the fit.

The fit result for  $W_b(E^{\text{jet}}, E^q)$  is shown in figure 7.2 with dashed lines for various  $b$ -jet energies. Compared to the double Gaussian used to describe light jets, the power law behavior is visible in the tail for high jet energies. This tail describes jets with a measured energy that underestimates the energy of the associated quark. It is larger for  $b$ -jets than for light jets due to the production of muons and neutrinos in the decay of  $b$ -hadrons.

## 7.4 Results

This section summarizes the results obtained with the MEM implementation described in this chapter. The separation of a classifier, such as the  $\text{MEM}_{\text{D1}}$  likelihood ratio, is defined by [123]

$$\langle S^2 \rangle = \frac{1}{2} \sum_{i=1}^N \frac{(S_i - B_i)^2}{S_i + B_i}, \quad (7.18)$$

evaluated using a distribution with  $N$  bins. The normalized signal and background yield per bin is denoted as  $S_i$  and  $B_i$ , with normalization condition  $\sum_i S_i = \sum_i B_i = 1$ . For identical distributions of signal and background, the separation is  $\langle S^2 \rangle = 0$ . Distributions that do not overlap in any bin have  $\langle S^2 \rangle = 1$ , which is the maximum possible separation. The separation depends on the choice of binning, and the values listed in this section are calculated using distributions with eight bins.

The  $\text{MEM}_{\text{D1}}$  discriminant as defined in equation (7.10) is used in the  $\text{SR}_1^{\geq 6j}$  region of the  $t\bar{t}H(b\bar{b})$  analysis as an input to the classification BDT described in section 6.5.4. A sigmoid function is used for most distributions shown in this section to map the discriminant into the interval  $[0, 1]$ . The function is  $(1 + \exp(-\text{MEM}_{\text{D1}} - 4))^{-1}$ , indicated on the relevant figures.

The expected distributions of signal and background processes for all figures in this section are obtained with the model described in section 6.3, unless specified otherwise.

### 7.4.1 Results for the $\text{SR}_1^{\geq 6j}$ signal region

Figure 7.3 shows the logarithm of the signal and background likelihoods described in section 7.2.4 for the  $\text{SR}_1^{\geq 6j}$  region. The normalized distributions of the  $t\bar{t}H(b\bar{b})$  signal and  $t\bar{t} + b\bar{b}$  background are drawn with dashed red and solid blue lines, respectively. The contributions from the signal have on average a larger signal likelihood  $L_S$  than the  $t\bar{t} + b\bar{b}$  background. The distributions for the background likelihoods  $L_B$  are very similar. The  $\text{MEM}_{\text{D1}}$  discriminant for the two processes is visualized in figure 7.4. The left side shows the discriminant directly, while the right side shows the transformed version used in the following. The likelihood ratio  $L_S/L_B$  is on average larger for the  $t\bar{t}H$  signal than the background, hence the contribution from the signal peaks towards the right of the distribution, while the  $t\bar{t} + b\bar{b}$  background peaks at a lower value of the discriminant. The separation between  $t\bar{t}H(b\bar{b})$  and  $t\bar{t} + b\bar{b}$  is  $\langle S^2 \rangle = 13\%$  and  $\langle S^2 \rangle = 14\%$ , respectively. For a signal efficiency of 50%, the  $t\bar{t} + b\bar{b}$  efficiency is 24%. This corresponds to a rejection factor of four when correctly identifying every second  $t\bar{t}H$  event. When including all background processes, and  $t\bar{t}H$  decays into the remaining final states besides  $H \rightarrow b\bar{b}$ , the separation between  $t\bar{t}H$  and the total background reduces to 12% for both distributions. The separation achieved by the classification BDT, shown in figure 6.15, reaches  $\langle S^2 \rangle = 20\%$ .

A comparison between data and the model is provided in figure 7.5. It shows the logarithms of the signal and background likelihoods, as well as the transformed  $\text{MEM}_{\text{D1}}$  discriminant. The pre-fit model is shown on the left, and the post-fit model, obtained from the combined fit described in section 6.7.2, is on the right. The model is in good agreement with data within its associated uncertainties, both pre- and post-fit.

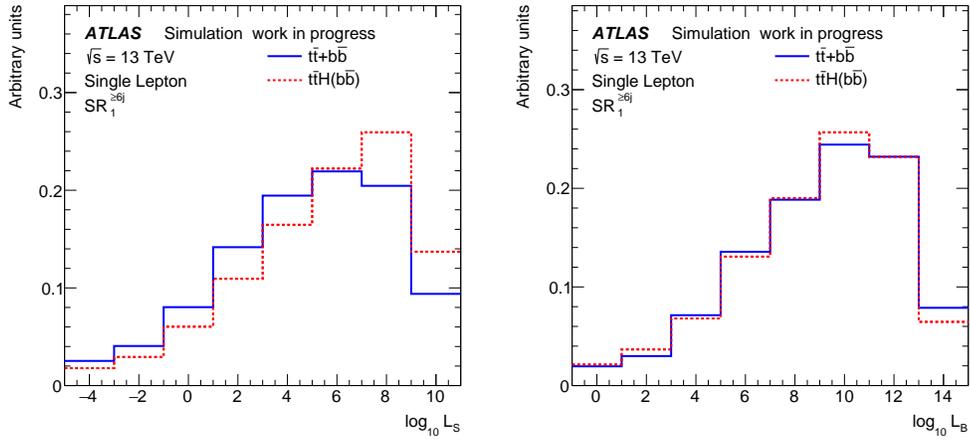


Figure 7.3: Distribution of the  $t\bar{t}H(b\bar{b})$  signal and  $t\bar{t} + b\bar{b}$  background processes in the  $SR_1^{\geq 6j}$  region as a function of the logarithms of the signal and background likelihoods,  $L_S$  and  $L_B$ . Both processes are normalized to have unit integral. The left- and rightmost bins of the distributions include all events with likelihoods smaller or larger than the edge of these bins, respectively.

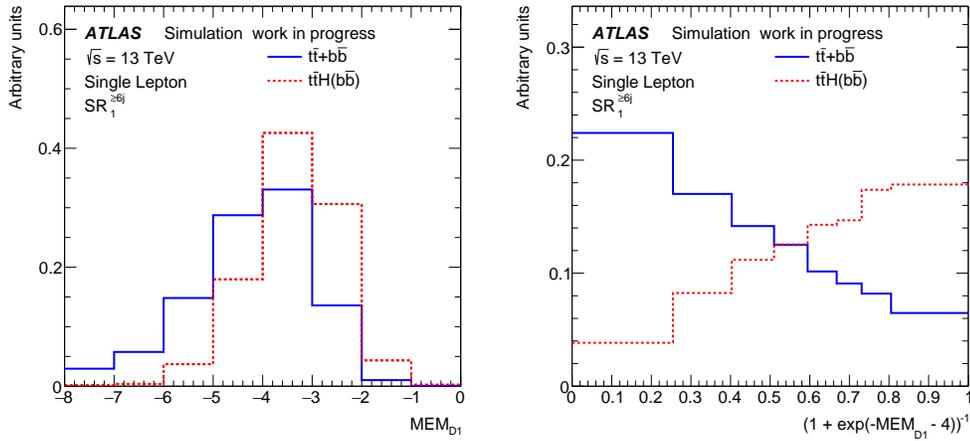


Figure 7.4: Distribution of the  $t\bar{t}H(b\bar{b})$  signal and  $t\bar{t} + b\bar{b}$  background processes in the  $SR_1^{\geq 6j}$  region, both normalized to have unit integral. The left figure shows the distributions as a function of the  $MEM_{D1}$  likelihood ratio, while the right figure shows the transformed version of this variable.

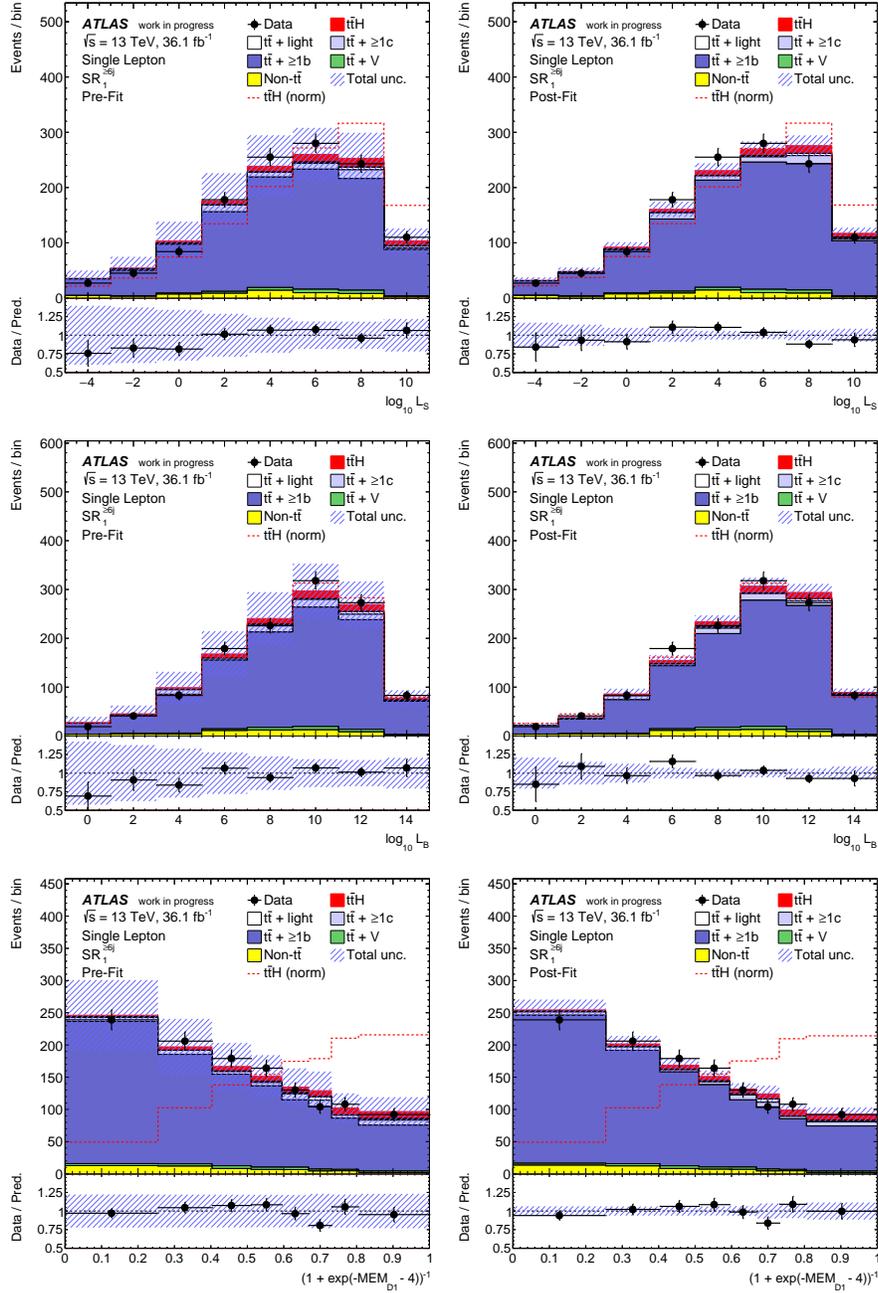


Figure 7.5: Comparison between data and the model for the logarithm of signal likelihood (top), logarithm of background likelihood (middle) and the transformed  $\text{MEM}_{D1}$  discriminant (bottom). The figures on the left show the pre-fit model. The post-fit model on the right is obtained from the fit described in section 6.7.2. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. The left- and rightmost bins of the likelihood distributions include all events with likelihoods smaller or larger than the edge of these bins, respectively.

### 7.4.2 Modeling in validation regions

The MEM likelihoods are calculated in additional regions to validate the modeling of data. These regions are  $SR_2^{\geq 6j}$ ,  $SR_3^{\geq 6j}$  and  $CR_{t\bar{t}+\geq 1c}^{\geq 6j}$ , corresponding to the remaining regions with at least six jets in the  $t\bar{t}H(b\bar{b})$  analysis that are considered in the combined fit with more than one bin each. The  $b$ -jet selection strategy is slightly modified from the description in section 7.2.1. The set of jets satisfying the *loose*  $b$ -tagging operating point is ordered by decreasing transverse momentum, and the top four jets are used as selected  $b$ -jets.

The amount of MC events in these regions is significantly larger than the amount of events in the  $SR_1^{\geq 6j}$  region, resulting in increased computation time needed to process all systematic variations for all events in these regions. For this reason, only the dominant systematic uncertainties are evaluated. They include all modeling uncertainties related to the  $t\bar{t}H$  signal and the  $t\bar{t}$  background. An exception is the  $t\bar{t}+\geq 1b$  sub-component uncertainty derived from the comparison of POWHEG+PYTHIA 8 and SHERPA4F, which is not evaluated. Modeling uncertainties related to small backgrounds, as well as experimental uncertainties, are not considered.

Figure 7.6 shows the logarithms of the signal and background likelihoods, as well as the transformed  $MEM_{D1}$  discriminant, in the two signal regions  $SR_2^{\geq 6j}$  and  $SR_3^{\geq 6j}$ . All distributions are shown pre-fit, as the post-fit model cannot be evaluated without processing all variations defining systematic uncertainties. The model generally describes data well, but underestimates data in the  $SR_3^{\geq 6j}$  region. The shape of data and the model agrees for all distributions, validating the MEM implementation. Figure 7.7 shows the same distributions for the  $CR_{t\bar{t}+\geq 1c}^{\geq 6j}$  region. No issues in the modeling of the MEM distributions are observed. The separation between  $t\bar{t}H$  and the total background is  $\langle S^2 \rangle = 8\%$  in  $SR_2^{\geq 6j}$ , 6% in  $SR_3^{\geq 6j}$ , and 5% in  $CR_{t\bar{t}+\geq 1c}^{\geq 6j}$ .

### 7.4.3 Comparison to other methods

The MEM implementation provides a strong discriminant between  $t\bar{t}H$  and  $t\bar{t}+b\bar{b}$  in the  $SR_1^{\geq 6j}$  signal region. In the  $t\bar{t}H(b\bar{b})$  analysis, it is combined with additional multivariate techniques described in section 6.5. The three methods, reconstruction BDT, LHD and the MEM, are complimentary.

Compared to the LHD and reconstruction BDT approaches, the MEM does not rely on having sufficiently many simulated events available to derive the LHD template distributions and for BDT training. The transfer function in this implementation is derived from a MC sample. It can instead be approximated by considering the detector resolution, which can be approximated without the need for dedicated MC samples for the processes considered in a specific analysis.

The results of the MEM as implemented here depend on MC samples only through the transfer function. The LHD templates and BDT training can bias these methods towards features specific to the samples used to build the templates and for training. When using these methods, it is therefore important to validate that all relevant features are also present in data.

Both the reconstruction BDT and the MEM use correlations between objects, which the LHD does not. The LHD and MEM both combine information from multiple permutations (assignments of jets

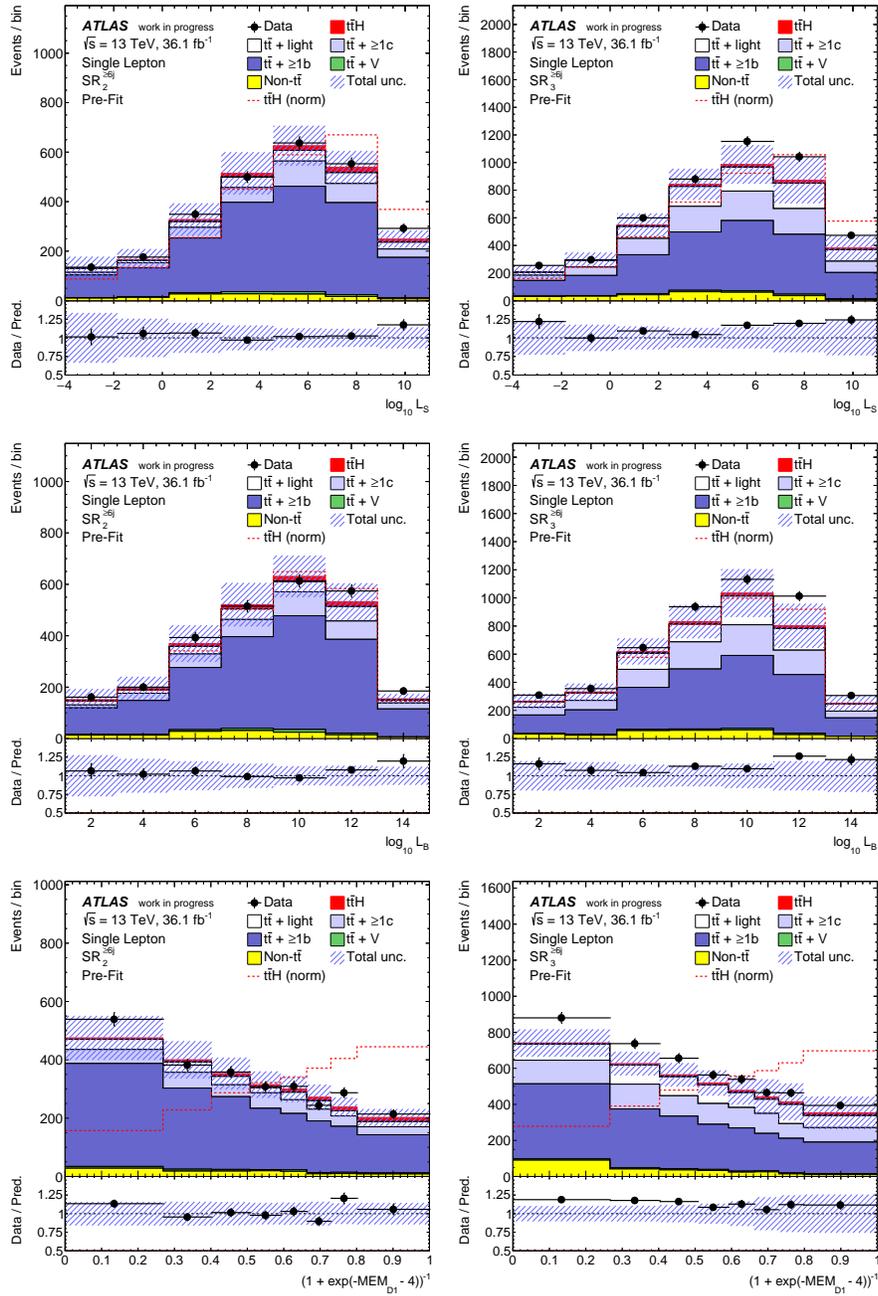


Figure 7.6: Comparison between data and the model for the logarithm of signal likelihood (top), logarithm of background likelihood (middle) and the transformed  $MEM_{D1}$  discriminant (bottom) in the  $SR_2^{\geq 6j}$  region (left) and the  $SR_3^{\geq 6j}$  region (right). The uncertainty bands only include sources related to  $t\bar{t}H$  and  $t\bar{t}$  modeling, with the exception of the  $t\bar{t} + \geq 1b$  sub-component uncertainty derived from SHERPA4F. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. The left- and rightmost bins of the likelihood distributions include all events with likelihoods smaller or larger than the edge of these bins, respectively.

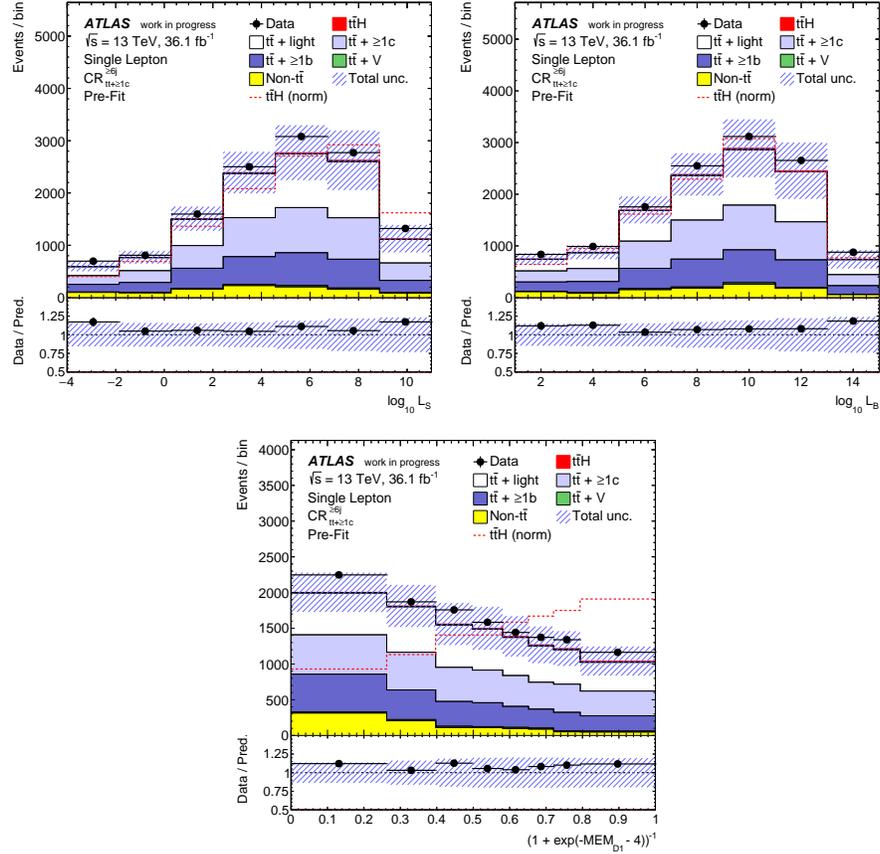


Figure 7.7: Comparison between data and the model for the logarithm of signal likelihood (top left), logarithm of background likelihood (top right) and the transformed  $\text{MEM}_{D1}$  discriminant (bottom) in the  $\text{CR}_{t\bar{t}\ge 1c}^{\ge 6j}$  region. The uncertainty bands only include sources related to  $t\bar{t}H$  and  $t\bar{t}$  modeling, with the exception of the  $t\bar{t} + \ge 1b$  sub-component uncertainty derived from SHERPA4F. No uncertainty related to  $k(t\bar{t} + \ge 1b)$  and  $k(t\bar{t} + \ge 1c)$  is included. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. The left- and rightmost bins of the likelihood distributions include all events with likelihoods smaller or larger than the edge of these bins, respectively.

to partons) into the resulting discriminants. Additional permutations are considered in the case of the LHD, which also includes hypotheses beyond what is considered in this MEM implementation. The addition of additional hypotheses in the MEM is conceptually straightforward, but computationally expensive. Increasing the dimensionality of the integral in the likelihoods to account for objects that were not reconstructed results in significantly increased integration times.

## 7.5 System reconstruction with the matrix element method

The MEM likelihoods are evaluated per jet permutation before being summed together. The assignment of jets to quarks in the permutation with the highest likelihood can be considered to study properties of the reconstructed  $t\bar{t}H$  and  $t\bar{t} + b\bar{b}$  system. The results summarized in this section are obtained in a region defined by requiring at least six jets, out of which at least four are  $b$ -tagged at the *tight* operating point. Events for the  $t\bar{t}H$  process are simulated with MADGRAPH5\_AMC@NLO [95] and HERWIG++ [121]. The  $t\bar{t}$  background is simulated with POWHEG+PYTHIA 6 [100–103, 117], the sample used is enriched with contributions from  $t\bar{t} + \geq 1b$ .

### 7.5.1 Assignment efficiency

The true origin of a jet is defined by matching it to quarks. A jet is matched to any of the six quarks expected from the LO  $t\bar{t}H$  system in the single-lepton channel if the jet is within  $\Delta R < 0.3$  of the quark. This matching is called truth-matching. Each jet may be truth-matched to multiple quarks.

If a jet in the highest likelihood permutation is assigned to a quark that it also truth-matched to, then the assignment by the MEM is declared to be correct. This is done no matter whether the jet is also truth-matched to additional quarks or not. The MEM assignment efficiency can be calculated with these definitions.

The confusion matrix in figure 7.8 summarizes the MEM performance for object assignments. It is evaluated on a sample of  $t\bar{t}H$  events, and using the  $t\bar{t}H$  likelihood. The rows correspond to the six different quarks needed for the LO  $t\bar{t}H$  topology. The columns list the true origin of each jet, determined by truth-matching. The last column contains jets that are not truth-matched to any quark in the LO  $t\bar{t}H$  system. Quarks from the decay of the Higgs boson and those from the decay of the  $W$  boson are distinguished according to their transverse momenta. The following nomenclature is used:

- $b_1$  from  $H$ :  $b$  quark from Higgs boson decay with highest transverse momentum,
- $b_2$  from  $H$ : second  $b$  quark from Higgs boson decay,
- $b$  from  $t_{\text{lep}}$ :  $b$  quark from top quark decay  $t_{\text{lep}} \rightarrow Wb \rightarrow l\nu b$ , with the  $W$  boson decaying to leptons,
- $b$  from  $t_{\text{had}}$ :  $b$  quark from top quark decay  $t_{\text{had}} \rightarrow Wb \rightarrow q\bar{q}b$ , with the  $W$  boson decaying to quarks,
- $q_1$  from  $W$ : quark from  $W$  boson decay with highest transverse momentum,
- $q_2$  from  $W$ : second quark from  $W$  boson decay.

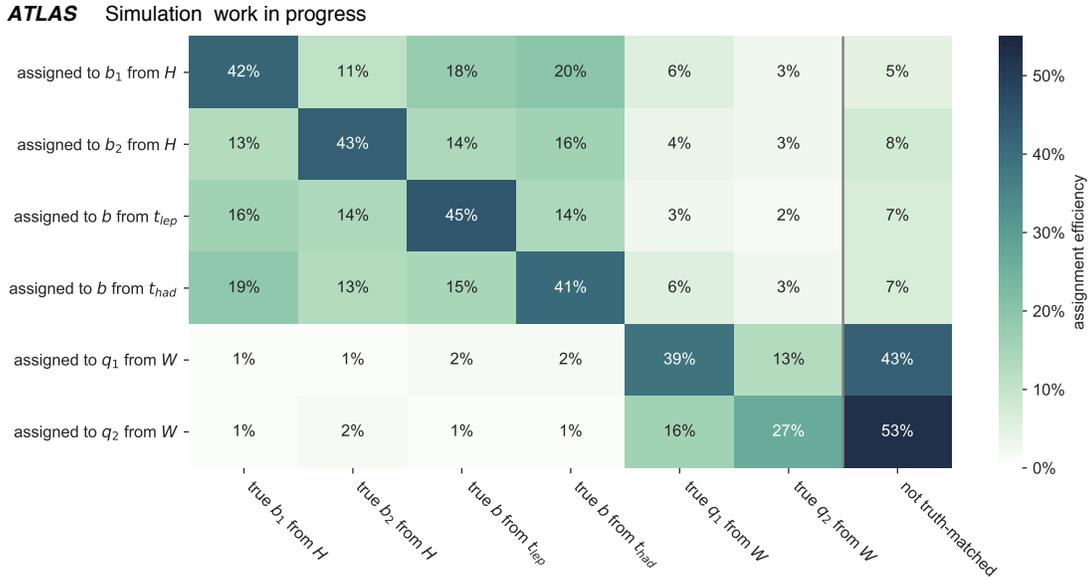


Figure 7.8: Assignment efficiency of jets to quarks in the permutation with the largest  $t\bar{t}H$  likelihood, evaluated with a sample of  $t\bar{t}H$  events. The rows correspond to the quark each jet is matched to, while the columns describe the true jet origin. Jets may be truth-matched to multiple quarks.

The correct assignments for the selected  $b$ -jets to the four  $b$  quarks in the  $t\bar{t}H$  system are found with an efficiency of slightly over 40%. The chance to assign a selected  $b$ -jet to the wrong  $b$  quark is around 10–20% per quark. The  $b$ -jet selection is very efficient, as the four selected  $b$ -jets are truth-matched to the four  $b$  quarks for around 85–90% of all events. This selection efficiency is evaluated with figure 7.8. It is obtained by summing the assignment efficiency of jets truth-matched to  $b$  quarks from the  $t\bar{t}H$  system, and assigned to  $b$  quarks in the permutation with the highest  $t\bar{t}H$  likelihood. Only a small fraction of events contain jets that are truth-matched to  $b$  quarks but assigned to  $W$  boson decay products. The assignment efficiency for the quark  $q_2$  is significantly smaller than the remaining assignment efficiencies. The majority of jets assigned to  $q_2$  are not truth-matched to any of the quarks from the  $t\bar{t}H$  system. The jet originating from  $q_2$  is frequently not reconstructed, as it fails to satisfy the  $p_T > 25$  GeV threshold. Both of the jets truth-matched to the decay products  $q_1$  and  $q_2$  of the  $W$  boson are present in only around 50% of the events. The sum of assignment efficiencies across each row is slightly larger than 100%; this is due to jets that are truth-matched to multiple quarks. Reconstruction efficiencies for the  $b$  quarks from top quark decays and the quarks from  $W$  boson decays are similar when considering the assignment efficiency for  $t\bar{t}$  events and using the  $t\bar{t} + b\bar{b}$  likelihood.

### Discriminant dependence on assignment efficiency

The  $\text{MEM}_{D_1}$  discrimination power is largest when all reconstructed objects associated to the LO  $t\bar{t}H$  and  $t\bar{t} + b\bar{b}$  systems are present in an event and selected for the jet permutations in the MEM

calculation. The separation  $\langle S^2 \rangle$  increases by 20% when evaluating the discriminant only for events with exactly six jets, instead of events with six or more jets. This is due to the simplified jet selection, which results in increased assignment efficiencies. When considering events with exactly six jets, and also requiring that at least one jet in every event is truth-matched to the quark from the  $W$  boson decay with lower transverse momentum, the separation increases by 30%.

The low assignment efficiency for the hadronic decay products of the  $W$  boson adversely affects the performance of the  $\text{MEM}_{\text{D1}}$  discriminant. An additional topology can be considered, where instead of integrating over the energies of the two quarks resulting from the decay of the  $W$  boson, the integral is performed over the  $W$  boson directly. This increases the dimensionality of the integral by one, and the assignment of jets to the  $W$  boson is no longer needed. The separation obtained in a test where the integration is performed over the three momentum components  $p_x, p_y, p_z$  of the  $W$  boson is only  $\langle S^2 \rangle = 1\%$ . The integration for this configuration is significantly slower than for the nominal configuration, and this hypothesis is not used in the final implementation of the MEM for the  $t\bar{t}H(b\bar{b})$  analysis.

### 7.5.2 Object reconstruction

The two jets assigned to the  $b$  quarks from the Higgs boson decay in the permutation with the highest  $t\bar{t}H$  likelihood can be combined and interpreted as a reconstructed Higgs boson. Similarly, the two jets assigned to the two  $b$  quarks produced in association with the top quark pair in the  $t\bar{t} + b\bar{b}$  process can be combined in the permutation with the highest  $t\bar{t} + b\bar{b}$  likelihood. The invariant mass distributions of the resulting reconstructed objects are visualized in figure 7.9. The Higgs boson mass is shown on the left, and the mass of the  $b\bar{b}$  system in the  $t\bar{t} + b\bar{b}$  case is on the right. The figures show distributions for both the  $t\bar{t}H$  signal as a dashed red line, and the  $t\bar{t}$  background as a solid blue line, both normalized to unit integral.

The  $t\bar{t}H$  process shows a peak in the reconstructed Higgs boson invariant mass, located around the 125 GeV Higgs boson mass used in the simulation. The peak for the  $t\bar{t}$  background is less pronounced. This peak appears also for the  $t\bar{t}$  background since the likelihood of a permutation where the invariant mass of the two jets is very different from the Higgs boson mass is suppressed. The suppression can take place via the transfer function in phase space regions where the invariant mass of the quarks from the Higgs boson decay is close to the Higgs boson mass. In regions where their invariant mass is very different from the Higgs boson mass, the transfer function may be larger, but the matrix element is smaller. The permutation chosen to calculate the reconstructed Higgs boson mass is the one with the largest likelihood, which is therefore biased towards selecting a configuration where the invariant mass of the jets assigned to the Higgs boson is close to the Higgs boson mass.

The  $t\bar{t} + b\bar{b}$  likelihood is not biased by the Higgs boson propagator. The distribution of the reconstructed mass of the  $b\bar{b}$  system in the permutation with the highest  $t\bar{t} + b\bar{b}$  likelihood is qualitatively different for the  $t\bar{t}H$  and  $t\bar{t}$  distributions. The  $t\bar{t}$  distribution is smoothly falling after a threshold effect due to the jet transverse momentum requirement of  $p_T > 25$  GeV in the analysis. The distribution of the  $t\bar{t}H$  signal peaks at higher values. If the assignment of jets to  $b$  quarks originating from top quark

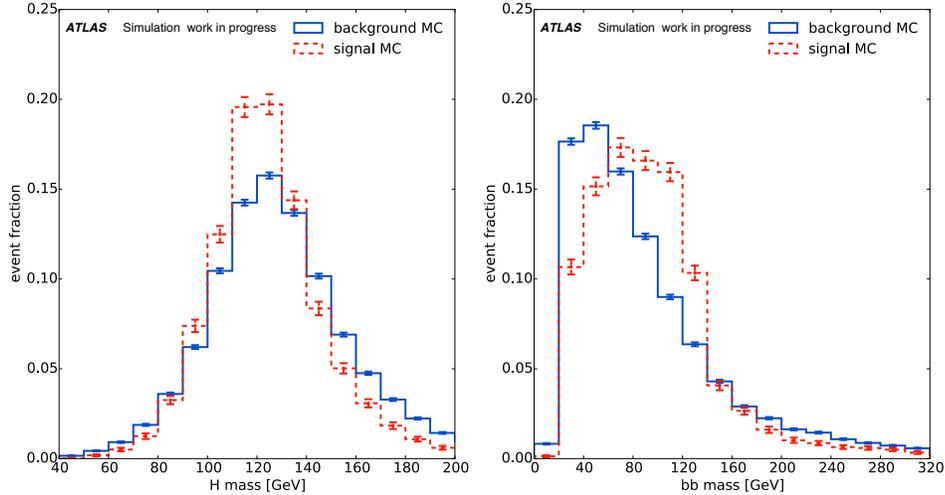


Figure 7.9: Reconstructed invariant mass of the  $b\bar{b}$  system produced in association with the top quark pair. The figure on the left shows the invariant mass of the two jets assigned to the  $b$  quarks from the Higgs boson decay in the permutation with the highest  $t\bar{t}H$  likelihood, this quantity is interpreted as the reconstructed Higgs boson mass. The figure on the right shows the  $b\bar{b}$  system assigned to  $b$  quarks that do not originate from top quark decays in the  $t\bar{t} + b\bar{b}$  topology, using the permutation with the highest  $t\bar{t} + b\bar{b}$  likelihood. Distributions of the  $t\bar{t}H$  signal are shown as dashed red lines, the  $t\bar{t}$  background is drawn as a solid blue line. All distributions are normalized to unit integral. Only statistical uncertainties are visualized in the figure.

decays is performed correctly, the remaining jets assigned to  $b$  quarks will on average have a larger invariant mass for  $t\bar{t}H$  events than for  $t\bar{t}$  events, since they originate from Higgs boson decays (if correctly selected).

The reconstructed Higgs boson and  $b\bar{b}$  masses provide discrimination between  $t\bar{t}H$  and background processes. Information from such reconstructed variables can be combined with the  $\text{MEM}_{\text{D1}}$  discriminant to further help distinguish  $t\bar{t}H$  from other processes. The assignment efficiencies from the reconstruction BDT in the  $t\bar{t}H(b\bar{b})$  analysis are slightly higher than those of the MEM, as the BDT is specifically optimized for system reconstruction. The classification BDT thus includes inputs from the reconstruction BDT which describe the kinematics of reconstructed objects. No information about reconstructed objects from the MEM is used in the analysis.

## 8. Observation of Yukawa interactions with third generation quarks

The  $t\bar{t}H(b\bar{b})$  analysis described in chapter 6 is statistically combined with a range of other analyses, resulting in the observation of Yukawa interactions with third generation quarks by the ATLAS collaboration. This chapter summarizes the evidence [2] and subsequent observation [3] of the  $t\bar{t}H$  process in section 8.1 and section 8.2. The observation of  $H \rightarrow b\bar{b}$  [38] is briefly summarized in section 8.3. These results establish the Yukawa interactions of top and bottom quarks predicted by the SM. The CMS collaboration also independently observed these interactions [39, 40], and both the CMS and ATLAS collaborations observed Yukawa interactions of tau leptons [41, 42].

The measurements reported in this chapter include the  $t\bar{t}H$  signal strength  $\mu_{t\bar{t}H}$ , defined like in chapter 6 as the ratio of the measured cross-section to the SM prediction:  $\mu_{t\bar{t}H} = \sigma_{t\bar{t}H}^{\text{obs}} / \sigma_{t\bar{t}H}^{\text{SM}}$ .

### 8.1 Evidence for $t\bar{t}H$

Evidence for the  $t\bar{t}H$  process is obtained by statistically combining searches with the ATLAS detector for  $t\bar{t}H$  with four different final states, which all use  $36.1 \text{ fb}^{-1}$  of Run-2 LHC data with  $\sqrt{s} = 13 \text{ TeV}$  [2]. This includes the  $t\bar{t}H(b\bar{b})$  search [1], described in detail in chapter 6. Higgs boson decays to  $Z$  bosons and subsequently four light leptons (electrons and muons) [146] are analysed in a targeted search. The remaining final states with multiple leptons are considered separately [2]; these multi-lepton final states originate from Higgs boson decays to  $Z$  and  $W$  bosons, as well as decays to tau leptons. Lastly, Higgs boson decays to two photons are included [147].

#### 8.1.1 Analyses entering the combination

The  $t\bar{t}H$  analysis for the  $H \rightarrow ZZ^* \rightarrow 4l$  topology selects events with four leptons, which form two pairs of leptons with the same flavor and opposite charge each. The invariant mass of the four lepton system is required to be in a window around the Higgs boson mass, and additional jet and  $b$ -tagging requirements are used to select events corresponding to the  $t\bar{t}H$  topology. A total of 0.5 events are expected, out of which 0.4 events are expected to originate from  $t\bar{t}H$  production. No events are observed.

The multi-lepton search analyzes  $t\bar{t}H$  final states with seven different topologies, defined via different combinations of light leptons and tau leptons decaying to hadrons. Events selected for the analysis need to have various combinations of leptons for the different topologies, while jet and  $b$ -tagging requirements are included to select the  $t\bar{t}H$  phase space. Dedicated multivariate

## 8. Observation of Yukawa interactions with third generation quarks

Table 8.1: Summary of the signal strength  $\mu_{t\bar{t}H}$  and the observed and expected significance measured in the individual analyses used to establish evidence for the  $t\bar{t}H$  process, as well as the combination of all analyses. No events are observed in the analysis targeting  $H \rightarrow ZZ^* \rightarrow 4l$ , hence the 68% confidence level upper limit is reported for the signal strength [2].

|                                     | $\mu_{t\bar{t}H}$   | Significance |             |
|-------------------------------------|---------------------|--------------|-------------|
|                                     |                     | Observed     | Expected    |
| $H \rightarrow b\bar{b}$            | $0.8^{+0.6}_{-0.6}$ | $1.4\sigma$  | $1.6\sigma$ |
| $H \rightarrow ZZ^* \rightarrow 4l$ | $< 1.9$             | -            | $0.6\sigma$ |
| Multi-lepton                        | $1.6^{+0.5}_{-0.4}$ | $4.1\sigma$  | $2.8\sigma$ |
| $H \rightarrow \gamma\gamma$        | $0.6^{+0.7}_{-0.6}$ | $0.9\sigma$  | $1.7\sigma$ |
| Combination                         | $1.2^{+0.3}_{-0.3}$ | $4.2\sigma$  | $3.8\sigma$ |

analysis methods are used in most regions enriched in  $t\bar{t}H$  signal. Additional regions are included in a combined fit to control background contributions.

The  $t\bar{t}H$  analysis with loop-induced Higgs boson decays to a system of photons,  $H \rightarrow \gamma\gamma$ , extracts the  $t\bar{t}H$  signal from a fit of the di-photon invariant mass spectrum. The signal contribution results in a peak over a smooth background distribution, this peak is located around the Higgs boson mass. Besides requiring two photons in the event selection, the analysis contains channels for final states of the  $t\bar{t}$  system with zero leptons or at least one lepton. Additional jet and  $b$ -tag requirements complete the event selection. Different categories are defined via a BDT, which is trained to identify  $t\bar{t}H$  signal events. A combined fit to the di-photon invariant mass spectra in all categories is used to measure the  $t\bar{t}H$  signal.

### 8.1.2 Results

Table 8.1 summarizes the  $t\bar{t}H$  signal strength,  $\mu_{t\bar{t}H}$ , measured in the four individual analyses, and in the statistical combination of all analyses. It also includes the observed and expected significance of the measurements, compared to the background-only hypothesis where the  $t\bar{t}H$  signal is absent. Due to the lack of observed events in the  $H \rightarrow ZZ^* \rightarrow 4l$  analysis, the 68% confidence level upper limit on the signal strength is calculated with the  $CL_s$  method described in section 5.2.2.

The statistical combination of the four analyses results in a signal strength measurement of

$$\mu_{t\bar{t}H} = 1.17 \pm 0.19 \text{ (stat.) }^{+0.27}_{-0.23} \text{ (syst.)}, \quad (8.1)$$

dominated by systematic uncertainty. Leading systematic uncertainties in the measurement are the modeling of  $t\bar{t}$  in the  $t\bar{t}H(b\bar{b})$  analysis, the cross-section uncertainty for  $t\bar{t}H$ , and the non-prompt and fake lepton estimate in the multi-lepton analysis.

The background-only hypothesis, where  $\mu_{t\bar{t}H} = 0$ , is excluded at  $4.2\sigma$ , with an expected sensitivity of  $3.8\sigma$  when using the SM prediction for the  $t\bar{t}H$  signal. This result establishes evidence for the

$t\bar{t}H$  process. The sensitivity is dominated by the multi-lepton channel, followed by contributions of similar size from  $H \rightarrow \gamma\gamma$  and  $H \rightarrow b\bar{b}$  final states.

The corresponding cross-section for the best-fit signal strength is measured as  $\sigma_{t\bar{t}H} = 590^{+160}_{-150}$  fb, in agreement with the SM prediction of  $\sigma_{t\bar{t}H} = 507^{+35}_{-50}$  fb.

## 8.2 Observation of $t\bar{t}H$

The observation of the  $t\bar{t}H$  process by the ATLAS collaboration is achieved by combining the searches for multi-lepton and  $H \rightarrow b\bar{b}$  final states, which were also used to establish evidence for  $t\bar{t}H$ , with searches using a larger dataset recorded at  $\sqrt{s} = 13$  TeV for the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4l$  topologies [3]. A combination with  $t\bar{t}H$  searches performed with Run-1 LHC data recorded at  $\sqrt{s} = 7$  TeV and  $\sqrt{s} = 8$  TeV is also performed. The top quark Yukawa coupling is extracted in a combined fit of  $t\bar{t}H$  searches and additional Higgs boson analyses performed at  $\sqrt{s} = 13$  TeV.

### 8.2.1 Analyses entering the combination

The  $H \rightarrow b\bar{b}$  topology is discussed in detail in chapter 6, and the multi-lepton search is briefly described in section 8.1. Both of these searches use a dataset of  $36.1 \text{ fb}^{-1}$  and enter the statistical combination.

The  $t\bar{t}H$  analysis targeting the  $H \rightarrow ZZ^* \rightarrow 4l$  topology is updated to use  $79.8 \text{ fb}^{-1}$  of data [3]. The approach to this search is similar to the approach for the prior search with  $36.1 \text{ fb}^{-1}$ , described in section 8.1.1. A BDT is used in one of the regions enriched in signal to further improve the sensitivity. No events are observed, while 1.1 events are expected. The  $t\bar{t}H$  process is expected to contribute 0.6 of these expected events.

The analysis for  $t\bar{t}H$  with loop-induced Higgs boson decays to di-photon final states is also updated to use  $79.8 \text{ fb}^{-1}$  of data [3]. A similar approach is used as for the search with the  $36.1 \text{ fb}^{-1}$  dataset described in section 8.1.1. The inclusion of four-momentum information for the reconstructed objects results in an improved BDT performance for this search.

Additional searches for  $t\bar{t}H$  enter the combination when also considering data from Run-1 of the LHC. These target the  $H \rightarrow b\bar{b}$  final state [88, 90], multi-lepton final states [148] and di-photon final states [149] of the Higgs boson.

### 8.2.2 Results

Figure 8.1 shows the  $t\bar{t}H$  cross-section measurement, divided by the SM prediction, which is obtained in a combination of analyses using only  $\sqrt{s} = 13$  TeV data from Run-2 of the LHC. The results per analysis topology are obtained from a combined fit with four individual parameters for the  $t\bar{t}H$  cross-section, which independently scale the  $t\bar{t}H$  contributions in each topology. The result of the combined fit with only one cross-section parameter is also shown. The 68% confidence level upper limit is reported for the  $H \rightarrow ZZ^* \rightarrow 4l$  analysis, where no events are observed. Statistical and systematic uncertainties affecting the measurements are drawn as yellow and blue rectangles,

## 8. Observation of Yukawa interactions with third generation quarks

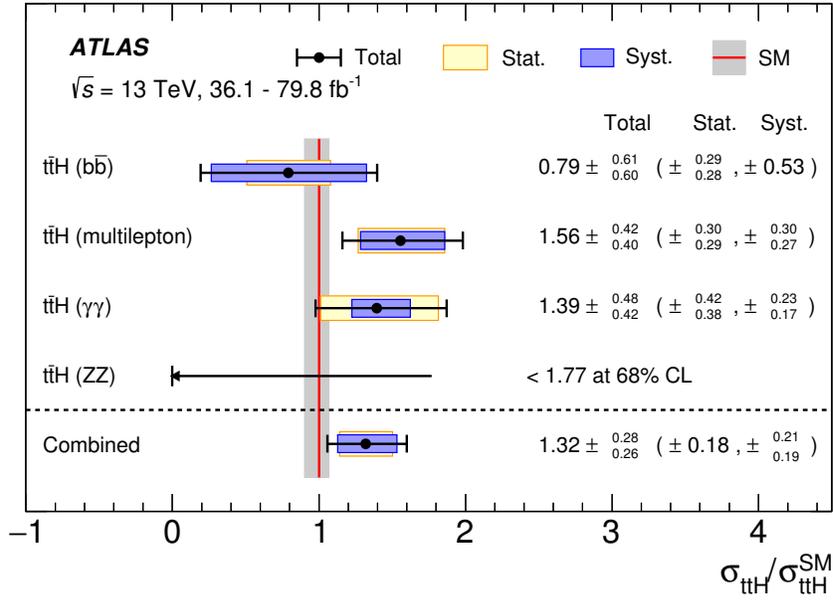


Figure 8.1: Results of the  $t\bar{t}H$  cross-section measurement, divided by the SM prediction, in the statistical combination. The results per analysis topology are obtained from a fit with four independent cross-section parameters. Statistical and systematic uncertainties are shown in yellow and blue, respectively. The SM prediction is shown in red, with the associated uncertainty indicated as a gray band. No events are observed in the  $H \rightarrow ZZ^* \rightarrow 4l$  analysis, and the 68% confidence level upper limit is reported [3].

respectively. The total uncertainty is shown as a black bar. The SM prediction is shown in red, with a gray bar indicating uncertainties related to the  $t\bar{t}H$  cross-section prediction. These uncertainties related to the  $t\bar{t}H$  cross-section are not included in the systematic uncertainty reported for the measurement; this is in contrast to signal strength measurements, where they are included.

The statistical combination of the four searches for  $t\bar{t}H$  results in the observation of the  $t\bar{t}H$  process. The background-only hypothesis is excluded at  $5.8\sigma$ , with an expected sensitivity of  $4.9\sigma$ , when using only Run-2 data with  $\sqrt{s} = 13$  TeV. When also including  $t\bar{t}H$  searches performed with the ATLAS detector at  $\sqrt{s} = 7$  TeV and  $\sqrt{s} = 8$  TeV during Run-1 of the LHC, the exclusion is at  $6.3\sigma$ , with an expected sensitivity of  $5.1\sigma$ .

The larger dataset and analysis improvements in the search for di-photon final states are the dominant contribution to the increase in sensitivity compared to the  $t\bar{t}H$  combination described in section 8.1. The observed significance when only considering this search targeting the  $H \rightarrow \gamma\gamma$  topology is  $4.1\sigma$ , with an expected sensitivity of  $3.7\sigma$ . Contributions to the combination from  $H \rightarrow ZZ^* \rightarrow 4l$  are small, with an expected sensitivity of  $1.2\sigma$ , but no observed events. The  $t\bar{t}H(b\bar{b})$  analysis has a smaller impact in this combination than in the  $t\bar{t}H$  combination establishing evidence for the process.

Statistical and systematic effects have a similar impact on the combined measurement. Contribu-

tions to the systematic uncertainty are dominated by  $t\bar{t}$  modeling in the  $t\bar{t}H(b\bar{b})$  analysis, the  $t\bar{t}H$  cross-section uncertainty, and the non-prompt and fake lepton estimate in the multi-lepton search.

The  $t\bar{t}H$  cross-section is measured at  $\sqrt{s} = 13$  TeV as  $\sigma_{t\bar{t}H} = 670 \pm 90^{+110}_{-100}$  fb. It agrees with the SM prediction of  $\sigma_{t\bar{t}H} = 507^{+35}_{-50}$  fb.

### 8.2.3 Top quark Yukawa coupling

An interpretation of ATLAS Higgs boson measurements in terms of Yukawa couplings can be performed by allowing these couplings to vary. The effective coupling  $\kappa_t$  scales the top quark Yukawa coupling as a multiplicative factor. When combining the  $t\bar{t}H$  analyses performed at  $\sqrt{s} = 13$  TeV with other analyses using up to  $79.8 \text{ fb}^{-1}$  of Run-2 data from the LHC, the effective coupling is measured as  $\kappa_t = 1.02^{+0.11}_{-0.10}$  [37]. The value is consistent with the SM, which predicts  $\kappa_t = 1$ . In this measurement, the effective couplings to weak gauge bosons, bottom quarks, tau leptons, and muons are also extracted simultaneously with  $\kappa_t$ , and are all compatible with their SM predictions. It assumes no BSM particles coupling to the Higgs boson. The ratio of effective couplings to top quarks and gluons is measured to be  $\kappa_t/\kappa_g = 1.10^{+0.15}_{-0.14}$ , also compatible with the SM prediction.

## 8.3 Observation of $H \rightarrow b\bar{b}$

The statistical combination of ATLAS searches for multiple Higgs boson production modes, with Higgs boson decays to  $b\bar{b}$ , results in the observation of  $H \rightarrow b\bar{b}$  [38]. The dominant contribution to this combination comes from the search of  $VH$  production of Higgs bosons, with decays  $H \rightarrow b\bar{b}$ . Both the search with Run-1 data [150] and Run-2 data [38] are included. The  $t\bar{t}H(b\bar{b})$  search [1] described in chapter 6 enters the combination, as does the corresponding Run-1 search [88]. Two searches for vector boson fusion Higgs production with  $H \rightarrow b\bar{b}$  are included as well, using Run-1 data [151] and Run-2 data [152] of the LHC.

The observed significance for  $H \rightarrow b\bar{b}$  decays from this combination is  $5.4\sigma$ , with an expected sensitivity of  $5.5\sigma$ . Table 8.2 summarizes the observed and expected significance per production mode, and for the combination. The result per production mode includes two searches each, using both Run-1 and Run-2 data. The overall sensitivity is dominated by the  $VH$  channel. The  $t\bar{t}H$  channel has a small contribution; the leading contribution to the  $t\bar{t}H$  significance comes from the Run-2  $t\bar{t}H(b\bar{b})$  analysis.

With the cross-sections for Higgs boson production processes fixed to the SM prediction, the  $H \rightarrow b\bar{b}$  signal strength  $\mu_{H \rightarrow b\bar{b}}$  measures the ratio of the observed branching ratio to the SM prediction. In the combination of all channels, the signal strength is measured as

$$\mu_{H \rightarrow b\bar{b}} = 1.01 \pm 0.20 = 1.01 \pm 0.12 \text{ (stat.) }^{+0.16}_{-0.15} \text{ (syst.)}, \quad (8.2)$$

dominated by systematic uncertainty and consistent with the SM prediction.

## 8. Observation of Yukawa interactions with third generation quarks

---

Table 8.2: Observed and expected significance for  $H \rightarrow b\bar{b}$  decays. The results are reported separately per production mode, and for the statistical combination of all channels [38].

| Process             | Significance |             |
|---------------------|--------------|-------------|
|                     | Observed     | Expected    |
| Vector boson fusion | $0.9\sigma$  | $1.5\sigma$ |
| $t\bar{t}H$         | $1.9\sigma$  | $1.9\sigma$ |
| $VH$                | $5.1\sigma$  | $4.9\sigma$ |
| Combination         | $5.5\sigma$  | $5.4\sigma$ |

The effective coupling strength scaling the bottom Yukawa coupling is extracted in a combination of analyses, as described in section 8.2.3, as  $\kappa_b = 1.06^{+0.19}_{-0.18}$  [37]. It is compatible with the SM prediction of  $\kappa_b = 1$ .

## 9. Search for $t\bar{t}H(b\bar{b})$ with $139\text{ fb}^{-1}$ of data

With the Run-2 of the LHC completed, and  $139.0\text{ fb}^{-1}$  of data recorded by the ATLAS experiment available for physics analyses, this chapter provides an outlook on the sensitivity of a  $t\bar{t}H(b\bar{b})$  analysis with this data. Many of the details of this study follow the description of the analysis published with  $36.1\text{ fb}^{-1}$  [1], which is documented in chapter 6. This chapter highlights important aspects and differences to the treatment described previously.

The analysis presented here is designed to provide a robust baseline configuration for the next iteration of the  $t\bar{t}H(b\bar{b})$  analysis. Region definitions and multivariate analysis techniques are simplified, and only the single-lepton channel is considered. MC samples for physics analyses with the ATLAS experiment are centrally produced for use by the whole collaboration. The modeling and treatment of systematic uncertainties in the study documented in this chapter differs in several regards from the  $36.1\text{ fb}^{-1}$  analysis, mostly caused by the availability of MC samples.

The design of the ATLAS search for  $t\bar{t}H(b\bar{b})$  with the full Run-2 dataset is not finalized at the time of writing this chapter. In order to not bias decisions taken regarding the analysis design, data is not shown in bins of any distributions where the  $t\bar{t}H$  signal is expected to contribute more than 5% to the bin yield. Bins affected by this blinding are indicated with a gray hashed area. In addition to this, this chapter does not include any fits to data, and only shows the expected sensitivity obtained from fitting an Asimov dataset.

### 9.1 Event selection

The event selection for the analysis is very similar to the selection described in section 6.2. Relevant changes in object definitions and the definition of the single-lepton channel are listed in this section.

#### 9.1.1 Dataset

The analysis uses proton–proton collision events provided by the LHC between 2015 and 2018 at  $\sqrt{s} = 13\text{ TeV}$ . They are recorded by the ATLAS detector and required to fulfill the quality criteria from section 3.2.7. The integrated luminosity of this dataset is  $139.0 \pm 2.4\text{ fb}^{-1}$ . Figure 3.2 shows the mean amount of interactions in each bunch crossing per year, including data recorded by ATLAS but not used for physics analyses. Its mean value of 34 interactions for this dataset is increased compared to the analysis performed with  $36.1\text{ fb}^{-1}$ .

### 9.1.2 Object definitions

The basic object definitions correspond to those listed in section 6.2.2, with minor changes. Electrons need to satisfy the *medium* identification operating point, while muons need to satisfy the *loose* identification operating point. No isolation requirements are applied for either of these light leptons. Jet definitions are unchanged. After applying the overlap removal, electron and muon identification operating points are tightened to *tight* and *medium*, respectively. Electrons also need to satisfy the *gradient* isolation operating point, while muons need to fulfill the *FCTTO* isolation operating point.

### 9.1.3 Definition of the single-lepton channel

Events are recorded with single electron and muon triggers, with thresholds as described in section 6.2.3. The thresholds from 2016 stay the same for 2017 and 2018. An additional muon trigger is added in the barrel region for 2017–2018, with a transverse momentum threshold of 60 GeV.

Events are required to have at least five jets, exactly one reconstructed light lepton, and not more than one hadronic tau lepton. The boosted region definition is unchanged. All remaining events are considered for the resolved regions, provided they contain at least three jets satisfying the *tight*  $b$ -tagging operating point.

## 9.2 Modeling

The modeling of all processes relevant to this analysis is very similar to the setup described in section 6.3. This section focuses on differences compared to the previous setup. All processes are modeled with MC simulation. Contributions from fake and non-prompt leptons are negligible and therefore not included. The nominal samples used to describe  $t\bar{t}$  production are simulated with the AFII method, described in reference [58]. Additional events are simulated in regions of phase space the analysis is targeting. This decreases the sizeable uncertainty related to background model statistical uncertainties that was observed in the  $t\bar{t}H(b\bar{b})$  analysis with  $36.1\text{ fb}^{-1}$ , as listed in table 6.5.

### 9.2.1 $t\bar{t}H$ signal

The  $t\bar{t}H$  signal is modeled with POWHEG-BOX v2 [100–102, 153], referred to in the following as POWHEG, at NLO in QCD and using the NNPDF3.0NLO PDF set [96]. Parton showering and hadronization are simulated with PYTHIA 8.230 [97], using the A14 set of free parameters [98].

### 9.2.2 $t\bar{t}$ + jets background

The dominant background process in this analysis is top quark pair production. It is modeled with the POWHEG generator [103] and an updated version of PYTHIA, using 8.230 with the A14 tune. The split of  $t\bar{t}$  into components is performed as described in section 6.3.2. No reweighting to a prediction from a 4F scheme sample is performed.

### 9.2.3 Other backgrounds

The modeling of  $V$ +jets and  $t\bar{t}V$  processes is unchanged. Diboson samples use a combination of SHERPA 2.2.1 and 2.2.2 [94, 113–115]. Single top quark production in the  $s$ - and  $t$ -channel, as well as  $Wt$ , is simulated at NLO with POWHEG and the NNPDF3.0NLO PDF set. The  $t$ -channel is simulated in the 4F scheme. The  $tZ$  and  $tWZ$  processes are simulated with MADGRAPH5\_AMC@NLO [95], in the following abbreviated as MG5\_AMC@NLO. Parton showering and hadronization are performed with PYTHIA 8 for all five of these single top quark processes. The production of four top quarks,  $t\bar{t}t\bar{t}$ , is simulated at NLO with MG5\_AMC@NLO and PYTHIA 8; the  $t\bar{t}WW$  process is not considered separately. No additional Higgs production mechanisms are considered, including the associated production with a single top quark.

### 9.2.4 Inclusive modeling of data

Figure 9.1 compares the model described to data, showing the distribution of the number of jets per event. It includes all events satisfying the event selection for the single-lepton channel. The uncertainty band drawn as a hashed blue area combines statistical and systematic uncertainties, described in section 9.5, with the exception of uncertainties from the free-floating normalization factors for the  $t\bar{t}+ \geq 1b$  and  $t\bar{t}+ \geq 1c$  processes. Due to the tightened  $b$ -tagging requirements, the contribution from  $t\bar{t}$ +light processes is smaller than in figure 6.4, which shows the corresponding distribution for the  $36.1\text{ fb}^{-1}$  analysis.

The agreement with data is worse in figure 9.1 than in the corresponding figure 6.4 for the  $36.1\text{ fb}^{-1}$  analysis. While many small differences contribute to this, a large contribution comes from an improved jet calibration. Differences between the predicted distribution and data are covered by systematic uncertainties, and the normalization difference can be corrected with free-floating  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  normalization factors. The background model describes data well in fits to data performed with signal-depleted regions.

Figure 9.2 shows the number of  $b$ -tagged jets at the operating points *very tight*, *tight*, *medium*, and *loose*. The model underestimates data, particularly for events with many  $b$ -tagged jets. No events with only two  $b$ -tagged jets at the *loose* operating point are considered in the analysis, since either three  $b$ -tagged jets are required at the *tight* operating point for the resolved selection, or four  $b$ -tagged jets at the *loose* operating point for the boosted selection.

The modeling improves considerably when applying the best-fit normalization factors for  $t\bar{t}+ \geq 1b$  and  $t\bar{t}+ \geq 1c$  obtained in the previous analysis, which are given in equation (6.2). With this scaling applied, model and data agree well within the uncertainties related to the model, for the distributions of the number of jets and  $b$ -tagged jets.

## 9.3 Event categorization

Events in the single-lepton channel are divided into five signal regions and two control regions. The region design is simplified compared to the previous configuration described in section 6.4.

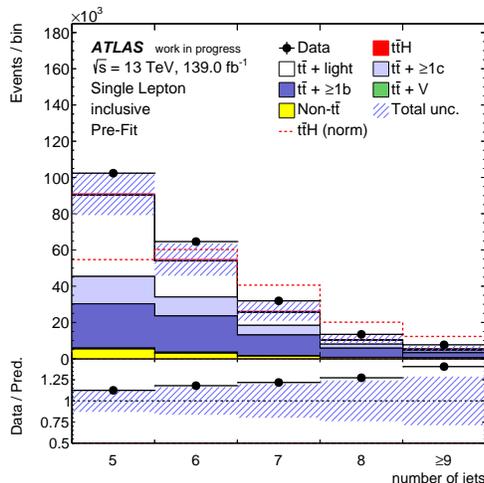


Figure 9.1: Expected distribution of the number of jets per event, compared to data. The uncertainties shown include all sources of systematic uncertainty described in section 9.5, with the exception of the free-floating normalization factors for the  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  processes. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line.

### 9.3.1 Region definitions

The  $\text{SR}^{\text{boosted}}$  signal region remains as defined in section 6.4.1. The resolved regions are divided according to jet multiplicity and  $b$ -tagging requirements. Separate regions are formed for events with exactly five jets, and for events with six or more jets. The  $\text{SR}_1^{5j}$  and  $\text{SR}_1^{\geq 6j}$  signal regions are unchanged, requiring at least four jets that are  $b$ -tagged at the *very tight* operating point. Two control regions are defined by requiring exactly three jets  $b$ -tagged at the *tight* operating point, these regions are  $\text{CR}^{5j}$  and  $\text{CR}^{\geq 6j}$ , for events with exactly five or six and more jets, respectively. All remaining events enter two intermediate signal regions. They require at least four jets that are  $b$ -tagged at the *tight* operating point, but less than four jets  $b$ -tagged at the *very tight* operating point. The two resulting signal regions are called  $\text{SR}_2^{5j}$  and  $\text{SR}_2^{\geq 6j}$ .

### 9.3.2 Region composition and signal contributions

The composition of background processes in the seven regions is summarized in figure 9.3. The  $t\bar{t} + \geq 1b$  process dominates in contribution to signal regions. Control regions have a larger amount of  $t\bar{t} + \text{light}$ , while  $t\bar{t} + \geq 1c$  is not dominant in any region. Contributions from  $t\bar{t}V$  and non- $t\bar{t}$  processes are small, but not negligible.

Signal contributions to the seven regions are summarized in figure 9.4. For each region, the fraction of expected signal events ( $S$ ) to the total background ( $B$ ) is listed. The histograms furthermore visualize  $S/\sqrt{B}$ ; blue bars are used to indicate control regions, while red bars are used in signal regions. The relative contribution of signal is increased in the  $\text{SR}_1^{\geq 6j}$  and  $\text{SR}_1^{5j}$  signal regions compared to the  $36.1 \text{ fb}^{-1}$   $t\bar{t}H(b\bar{b})$  analysis, which uses these regions as well. This is caused by the improved  $b$ -tagging

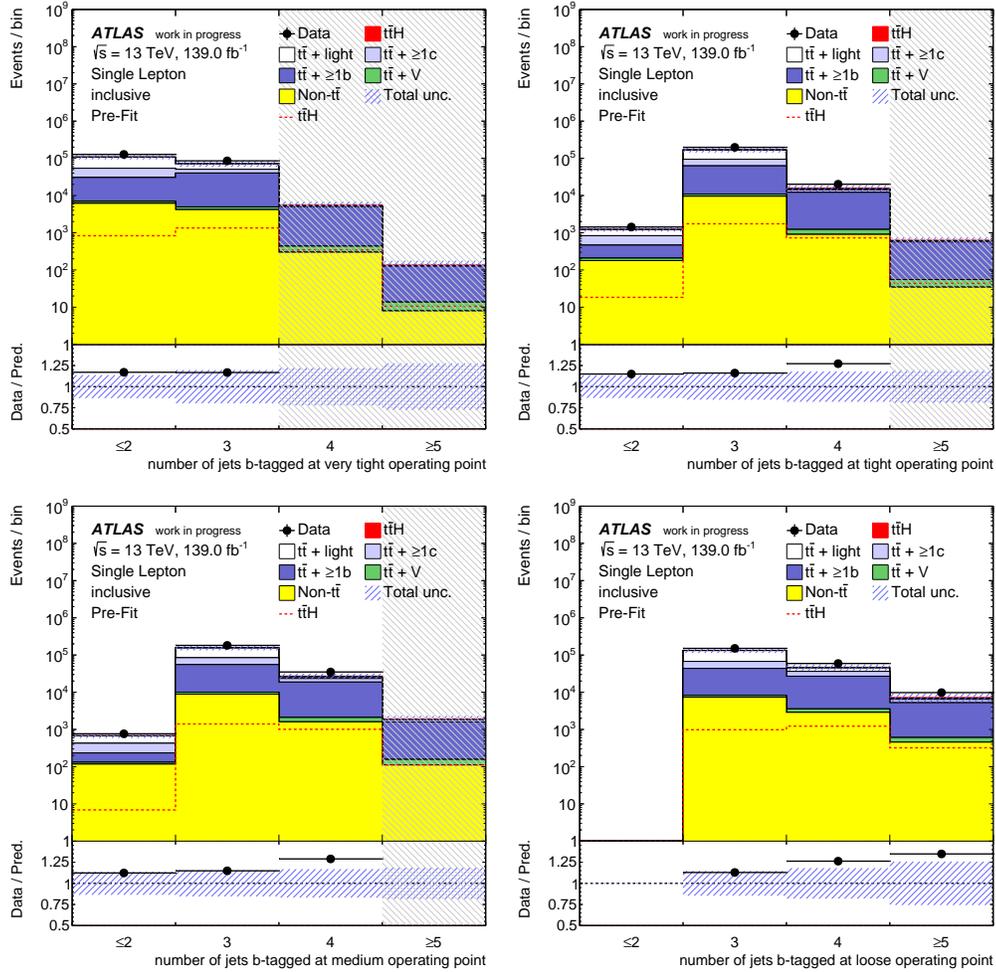


Figure 9.2: Expected distribution of the number of  $b$ -tagged jets per event at the four operating points (*very tight*, *tight*, *medium*, *loose*), compared to data. The uncertainties shown include all sources of systematic uncertainty described in section 9.5, with the exception of the free-floating normalization factors for the  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  processes. The  $t\bar{t}H$  signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram. Data is not shown in bins where the  $t\bar{t}H$  signal is expected to contribute more than 5% to the yield, indicated by a gray hashed area.

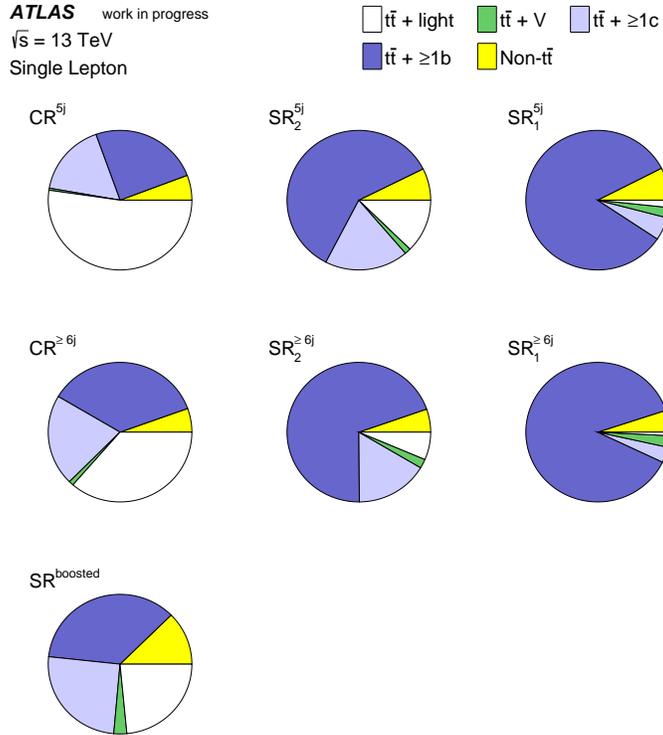


Figure 9.3: Composition of background processes per region. Each pie chart shows the relative contributions per process, with the processes defined in section 9.2.

algorithm. There are also substantial increases in  $S/\sqrt{B}$  due to the large increase in the size of the dataset used.

## 9.4 Multivariate analysis techniques

The multivariate analysis approach employed in the signal regions to discriminate between the  $t\bar{t}H$  signal and background processes is very similar to the approach for the previous analysis, described in section 6.5. It consists of two stages, with a reconstruction BDT and LHD employed in the first stage. The second stage consists of the classification BDT. Information from the reconstruction BDT and LHD is included in all resolved signal regions. Variables related to kinematic quantities and  $b$ -tagging are also used.

## 9.5 Systematic uncertainties

The implementation of systematic uncertainties follows the treatment described in section 6.6.1. This section summarizes changes with respect to the previous analysis.

**ATLAS** work in progress  
 $\sqrt{s} = 13\text{ TeV}, 139.0\text{ fb}^{-1}$   
 Single Lepton

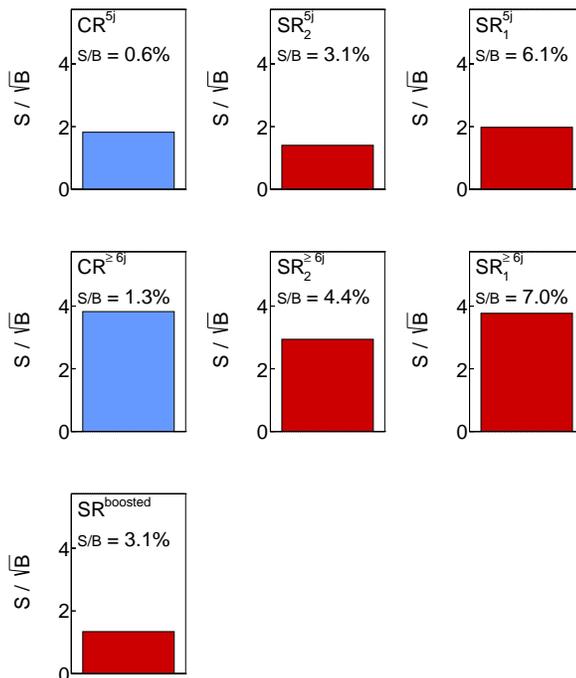


Figure 9.4: Signal contributions per region, calculated with the expected amount of  $t\bar{t}H$  events ( $S$ ) and background events ( $B$ ) per region. The histograms show  $S/\sqrt{B}$ , with blue bars for control regions and red bars indicating signal regions.  $S/B$  is also listed for each region.

### 9.5.1 Experimental uncertainties

A relative uncertainty of 1.7% is assigned to the total size of the dataset used, derived with a method similar to reference [50]. The majority of experimental uncertainties closely corresponds to the description in section 6.6.2.

#### Leptons

Systematic uncertainties related to the electron energy scale calibration are split into two components, while the remaining nuisance parameters associated to leptons are unchanged. This results in a total of seven nuisance parameters related to electrons. The 15 nuisance parameters associated to muons are unchanged. No uncertainties related to tau leptons are considered; they are expected to be negligible given their effect in the  $36.1\text{ fb}^{-1}$  analysis.

#### Jets

Uncertainties related to the jet energy scale calibration are described by a set of 31 nuisance parameters, split into more components than in the previous analysis. The treatment of nuisance parameters

related to jet energy resolution is also updated, resulting in eight associated nuisance parameters. One more nuisance parameter is related to the jet vertex tagger, resulting in a total of 40 nuisance parameters associated to jets.

### Flavor tagging

The  $b$ -tagging calibration used in this analysis describes uncertainties related to  $b$ -tagging efficiency, split into 45 sources. Mis-tag rates for  $c$ - and light jets are split into 20 sources each, for a total of 85 nuisance parameters related to  $b$ -tagging.

### Missing transverse energy

The treatment of uncertainties related to the  $E_T^{\text{miss}}$  calculation is unchanged, with three associated nuisance parameters.

## 9.5.2 Signal and background modeling

The treatment of systematic uncertainties related to the modeling of processes follows section 6.6.3, with differences mostly due to the availability of MC samples.

### $t\bar{t}H$ signal

Uncertainties related to the  $t\bar{t}H$  cross-section and Higgs boson branching ratios are unchanged. A comparison of the nominal POWHEG+PYTHIA 8 sample to a sample produced with MG5\_AMC@NLO and PYTHIA 8 is implemented as a nuisance parameter capturing effects due to the event generator choice. The choice of PS and hadronization model is described by a comparison of the nominal sample to one generated with POWHEG and HERWIG 7 [125]. Uncertainties related to ISR and FSR are described by two nuisance parameters. The ISR nuisance parameter compares the nominal configuration to a variation of renormalization and factorization scale, as well as the tune used in the PS. The nuisance parameter related to FSR varies the renormalization scale.

### $t\bar{t}$ + jets background

The  $t\bar{t}$  cross-section uncertainty treatment remains the same, as do the free-floating normalization factors  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$ . All other  $t\bar{t}$  modeling nuisance parameters are split across the three  $t\bar{t}$  components. Samples used to estimate modeling uncertainties are reweighted to match the  $t\bar{t}+ \geq 1b$ ,  $t\bar{t}+ \geq 1c$  and  $t\bar{t}$  + light fractions predicted by the nominal POWHEG+PYTHIA 8 sample.

A  $t\bar{t}$  sample generated with MG5\_AMC@NLO and PYTHIA 8 is compared to the nominal sample to estimate the uncertainty related to the choice of NLO event generator. The choice of PS and hadronization model is described with an uncertainty derived from the comparison to a sample produced with POWHEG+HERWIG 7. Uncertainties related to ISR and FSR modeling are derived with the same method as used for  $t\bar{t}H$ , and described with two nuisance parameters per  $t\bar{t}$  component.

Table 9.1: Systematic uncertainty sources affecting the modeling of  $t\bar{t} + \text{jets}$ . The left column shows the individual sources, while the central column describes how the effect is evaluated. The column on the right lists on which  $t\bar{t}$  components the sources act on, and whether the effect is correlated between the components.

| Systematic sources       | Description   | $t\bar{t}$ categories |
|--------------------------|---|-----------------------|
| $t\bar{t}$ cross-section | Up or down by 6%                                    | All, correlated       |
| $k(t\bar{t}+ \geq 1b)$   | Free-floating $t\bar{t}+ \geq 1b$ normalization     | $t\bar{t}+ \geq 1b$   |
| $k(t\bar{t}+ \geq 1c)$   | Free-floating $t\bar{t}+ \geq 1c$ normalization     | $t\bar{t}+ \geq 1c$   |
| Event generator          | MG5_AMC@NLO+PYTHIA 8 vs. POWHEG+PYTHIA 8            | All, uncorrelated     |
| PS & hadronization       | POWHEG+HERWIG 7 vs. POWHEG+PYTHIA 8                 | All, uncorrelated     |
| ISR                      | Variations of $\mu_R$ , $\mu_F$ , and PYTHIA 8 tune | All, uncorrelated     |
| FSR                      | Variation of $\mu_R$                                | All, uncorrelated     |

A summary of the nuisance parameters for  $t\bar{t}$  is shown in table 9.1. Four sources of uncertainty affect each of the  $t\bar{t}$  components separately.

### Small backgrounds

The normalization uncertainties assigned to all small background processes remain the same, with the exception of the normalization of  $Z+\text{jets}$  events. These events are assigned a single nuisance parameter describing a 35% uncertainty.

For the  $Wt$ ,  $s$ - and  $t$ -channel, additional samples are generated with POWHEG+HERWIG 7. The comparison of those samples to the nominal POWHEG+PYTHIA 8 samples is used to describe uncertainties related to the choice of PS and hadronization model, with one nuisance parameter for each sample. A comparison between the nominal  $Wt$  sample, and another one generated with MG5\_AMC@NLO and PYTHIA 8, is implemented as a nuisance parameter describing the event generator choice. An additional uncertainty for the  $Wt$  sample is derived from the comparison of the nominal diagram removal scheme to the diagram subtraction scheme [119]. This results in a total of five nuisance parameters related to modeling of single top quark processes.

### 9.5.3 Summary of systematic uncertainty sources

An overview of systematic uncertainties in the analysis is provided in table 9.2. Nuisance parameters affecting only normalization are indicated by type  $N$ , those affecting both shape and normalization are type  $S + N$ . The amount of components per source is listed in the last column.

## 9.6 Statistical analysis and results

This section summarizes the expected sensitivity of the analysis to the  $t\bar{t}H$  signal. The statistical treatment corresponds to the description from section 6.7. Seven regions enter the simultaneous fit.

Table 9.2: List of the systematic uncertainties affecting the analysis. The type  $N$  indicates uncertainties changing normalization of the affected process, uncertainties with type  $S + N$  can change both shape and normalization. The amount of different components per source is listed in the third column.

| Systematic uncertainty                  | Type              | Components |
|---|-------------------|------------|
| <i>Experimental uncertainties</i>       |                   |            |
| Luminosity                              | $N$               | 1          |
| Pile-up modeling                        | $S + N$           | 1          |
| <b>Physics objects</b>                  |                   |            |
| Electron                                | $S + N$           | 7          |
| Muon                                    | $S + N$           | 15         |
| Jet energy scale                        | $S + N$           | 31         |
| Jet energy resolution                   | $S + N$           | 8          |
| Jet vertex tagger                       | $S + N$           | 1          |
| $E_T^{\text{miss}}$                     | $S + N$           | 3          |
| <b><math>b</math>-tagging</b>           |                   |            |
| Efficiency                              | $S + N$           | 45         |
| Mis-tag rate ( $c$ )                    | $S + N$           | 20         |
| Mis-tag rate (light)                    | $S + N$           | 20         |
| <hr/>                                   |                   |            |
| <i>Modeling uncertainties</i>           |                   |            |
| <b>Signal</b>                           |                   |            |
| $t\bar{t}H$ cross-section               | $N$               | 2          |
| $H$ branching fractions                 | $N$               | 3          |
| $t\bar{t}H$ modeling                    | $S + N$           | 4          |
| <b><math>t\bar{t}</math> background</b> |                   |            |
| $t\bar{t}$ cross-section                | $N$               | 1          |
| $t\bar{t} + \geq 1c$ normalization      | free-floating $N$ | 1          |
| $t\bar{t} + \geq 1b$ normalization      | free-floating $N$ | 1          |
| $t\bar{t} + \text{light}$ modeling      | $S + N$           | 4          |
| $t\bar{t} + \geq 1c$ modeling           | $S + N$           | 4          |
| $t\bar{t} + \geq 1b$ modeling           | $S + N$           | 4          |
| <b>Other backgrounds</b>                |                   |            |
| $W$ +jets normalization                 | $N$               | 3          |
| $Z$ +jets normalization                 | $N$               | 1          |
| Diboson normalization                   | $N$               | 1          |
| $t\bar{t}W$ cross-section               | $N$               | 2          |
| $t\bar{t}Z$ cross-section               | $N$               | 2          |
| Single top cross-section                | $N$               | 6          |
| Single top modeling                     | $S + N$           | 5          |
| $t\bar{t}t\bar{t}$ normalization        | $N$               | 1          |

In the two control regions,  $\text{CR}^{5j}$  and  $\text{CR}^{\geq 6j}$ , the distribution of the  $H_{\text{T}}^{\text{had}}$  variable is used. The remaining five signal regions use distributions of the classification BDT. Statistical uncertainties related to the model are below 5% across all bins.

### 9.6.1 Expected sensitivity

The expected analysis sensitivity is evaluated in a fit of the model to an Asimov dataset. The resulting signal strength and free-floating normalization factors are

$$\begin{aligned}\mu_{t\bar{t}H} &= 1.00 \pm 0.18 \text{ (stat.)}_{-0.24}^{+0.29} \text{ (syst.)} = 1.00_{-0.31}^{+0.34}, \\ k(t\bar{t}+ \geq 1b) &= 1.00_{-0.19}^{+0.24}, \\ k(t\bar{t}+ \geq 1c) &= 1.00_{-0.35}^{+0.45}.\end{aligned}\tag{9.1}$$

The sensitivity increases substantially compared to the  $36.1 \text{ fb}^{-1}$  analysis, where the signal strength expected when fitting the single-lepton channel model to an Asimov dataset was  $\mu_{t\bar{t}H} = 1.00_{-0.65}^{+0.68}$ . The increase can only partially be attributed to the larger dataset, given that the  $36.1 \text{ fb}^{-1}$  analysis sensitivity was already limited by systematic uncertainties. Despite the larger dataset, the level of expected constraints on nuisance parameters is overall similar, as the selection requirements in the  $139.0 \text{ fb}^{-1}$  analysis are tightened. Less events enter the fit compared to the  $39.1 \text{ fb}^{-1}$  analysis. The uncertainty source with the second largest impact in the  $36.1 \text{ fb}^{-1}$  analysis, given by the comparison of the SHERPA4F prediction for  $t\bar{t}+ \geq 1b$  to the nominal configuration, does not have an equivalent in this analysis. Additional uncertainties related to  $t\bar{t}+ \geq 1b$  modeling derived from the SHERPA4F sample are also not considered, leading to an increase in sensitivity. The improved  $b$ -tagging calibration results in more powerful signal regions, which improve the performance further.

The expected significance over the SM background prediction is  $3.3\sigma$ , thereby surpassing the threshold for evidence for the  $t\bar{t}H(b\bar{b})$  process. This sensitivity is likely to decrease when considering a comparison between a 4F scheme and 5F scheme prediction for  $t\bar{t}$  as an additional uncertainty. The analysis presented in this chapter is however not optimized for sensitivity, and a dedicated optimization of the region definitions and multivariate analysis techniques can increase the sensitivity further. The inclusion of the dilepton channel, which is not considered in this chapter, will also result in a sensitivity increase.

### Distributions before and after the fit

Figure 9.5 shows a summary of the yield in all regions considered in the fit. The pre-fit model is shown at the top, with the post-fit model below, obtained from the fit to the Asimov dataset. The comparison between the post-fit model and data is not meaningful, hence data is not included in the corresponding figure. No uncertainty related to  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  is included pre-fit. The expected  $t\bar{t}H$  signal is shown in red contributing to the stacked histogram, and also drawn overlaid as a dashed red line. The expected uncertainties decrease post-fit due to correlations between nuisance parameters and their constraints.

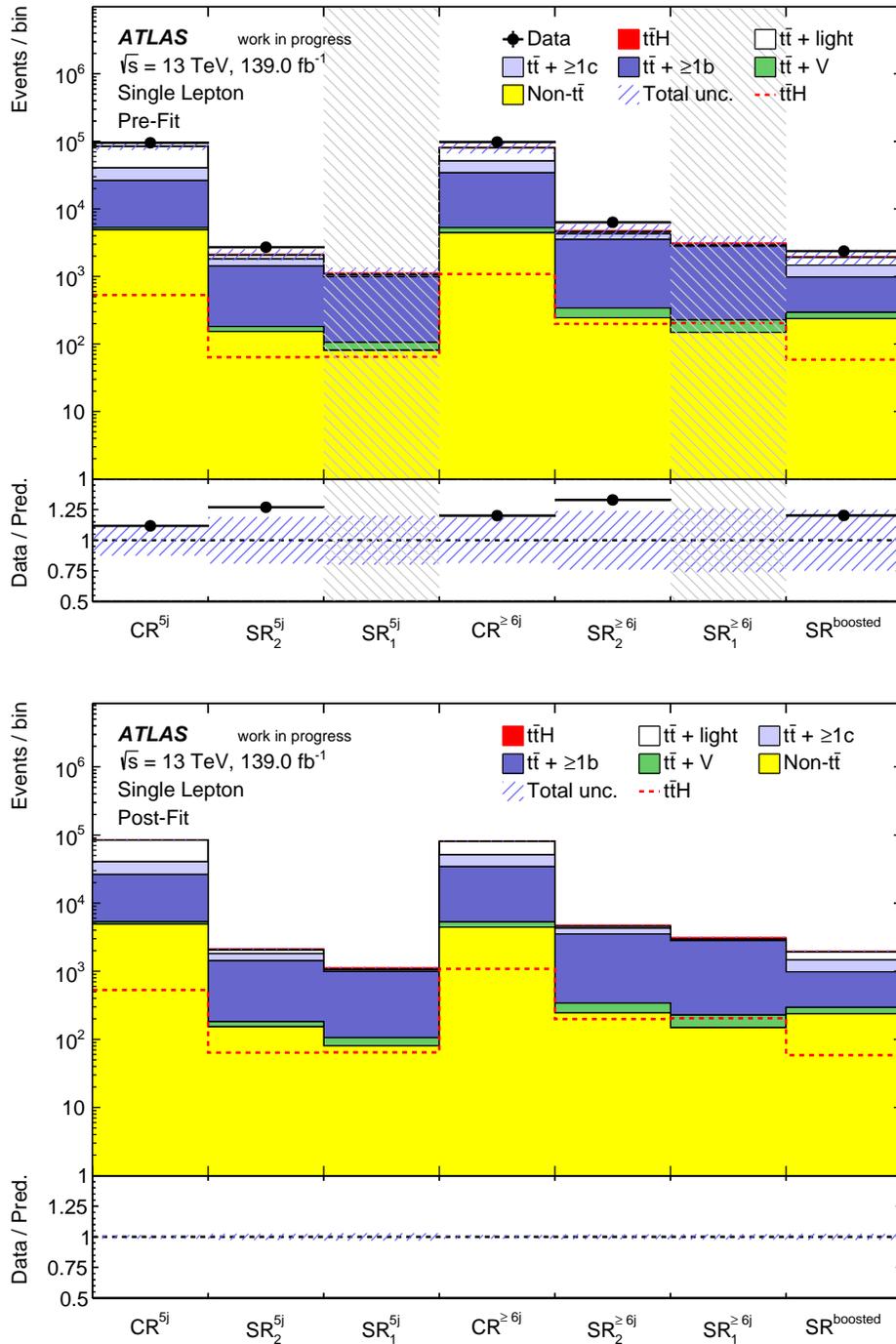


Figure 9.5: Overview of the yields in all regions pre-fit (top) and post-fit (bottom). The uncertainty bands include all sources of systematic uncertainty described in section 9.5. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal is shown both in the stacked histogram, contributing in red, as well as a dashed red line drawn on top of the stacked histogram. It is normalized to the SM prediction. Data is only compared to the pre-fit model, and not shown in bins where the  $t\bar{t}H$  signal is expected to contribute more than 5% to the yield, indicated by a gray hashed area.

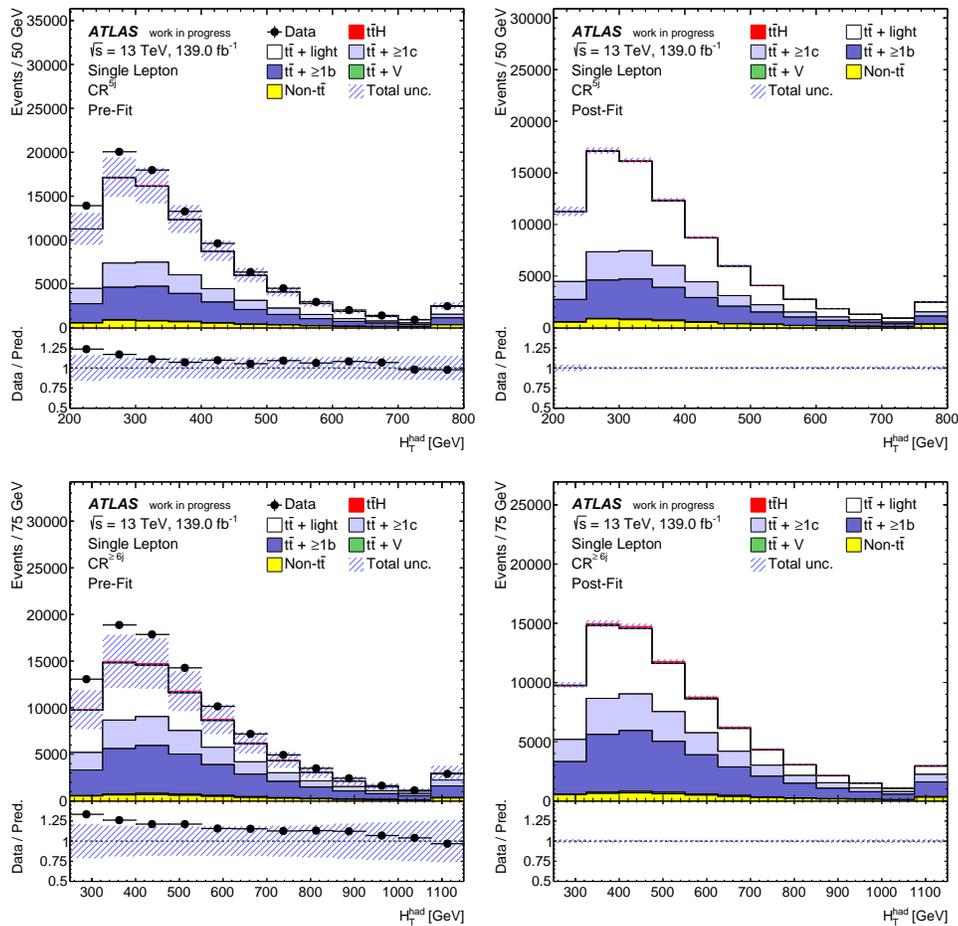


Figure 9.6: Comparison between data and the model for the control regions  $\text{CR}^{5j}$  (top) and  $\text{CR}^{\geq 6j}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 9.5. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. Events with  $H_T^{\text{had}} < 200\text{ GeV}$  or  $H_T^{\text{had}} > 800\text{ GeV}$  are included in the leftmost and rightmost bins of the  $\text{CR}^{5j}$  distributions, respectively. Similarly, events with  $H_T^{\text{had}} < 250\text{ GeV}$  or  $H_T^{\text{had}} > 1150\text{ GeV}$  are also included in the outermost bins of the  $\text{CR}^{\geq 6j}$  distributions. Data is only compared to the pre-fit model.

The  $H_T^{\text{had}}$  distributions in the two control regions which enter the fit,  $\text{CR}^{5j}$  and  $\text{CR}^{\geq 6j}$ , are shown in figure 9.6. Figure 9.7 shows the classification BDT distributions in the  $\text{SR}_1^{5j}$ ,  $\text{SR}_2^{5j}$ , and  $\text{SR}^{\text{boosted}}$  regions, while figure 9.8 shows the corresponding distributions in the  $\text{SR}_1^{\geq 6j}$  and  $\text{SR}_2^{\geq 6j}$  regions. All distributions are shown with the binning used in the fit. The  $t\bar{t}H$  distribution, normalized to the total background prediction, is overlaid as a dashed red line in the signal region distributions. Data is only included in pre-fit distributions.

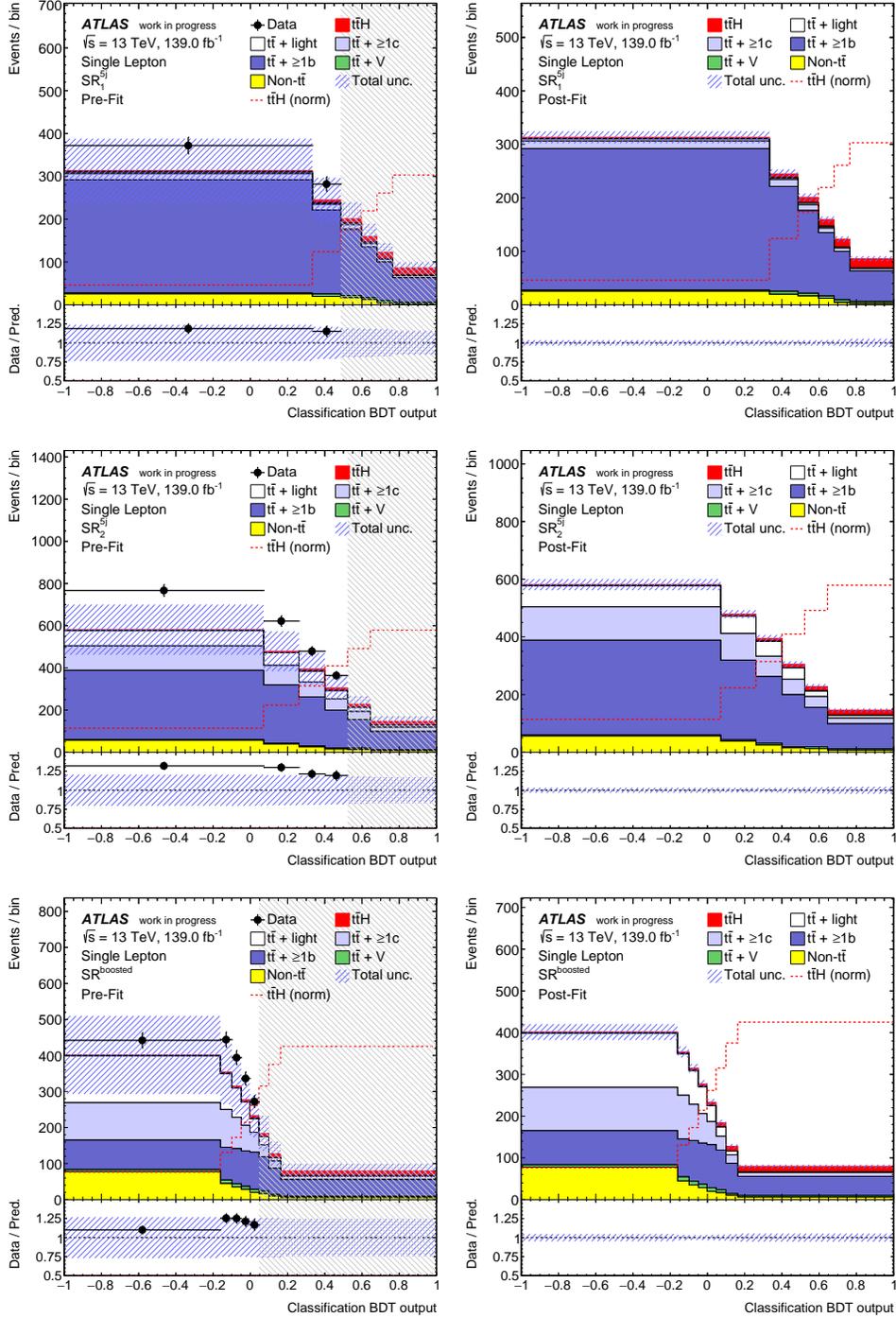


Figure 9.7: Comparison between data and the model for the signal regions  $SR_1^{5j}$  (top),  $SR_2^{5j}$  (middle) and  $SR^{\text{boosted}}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 9.5. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. Data is only compared to the pre-fit model, and not shown in bins where the  $t\bar{t}H$  signal is expected to contribute more than 5% to the yield, indicated by a gray hashed area.

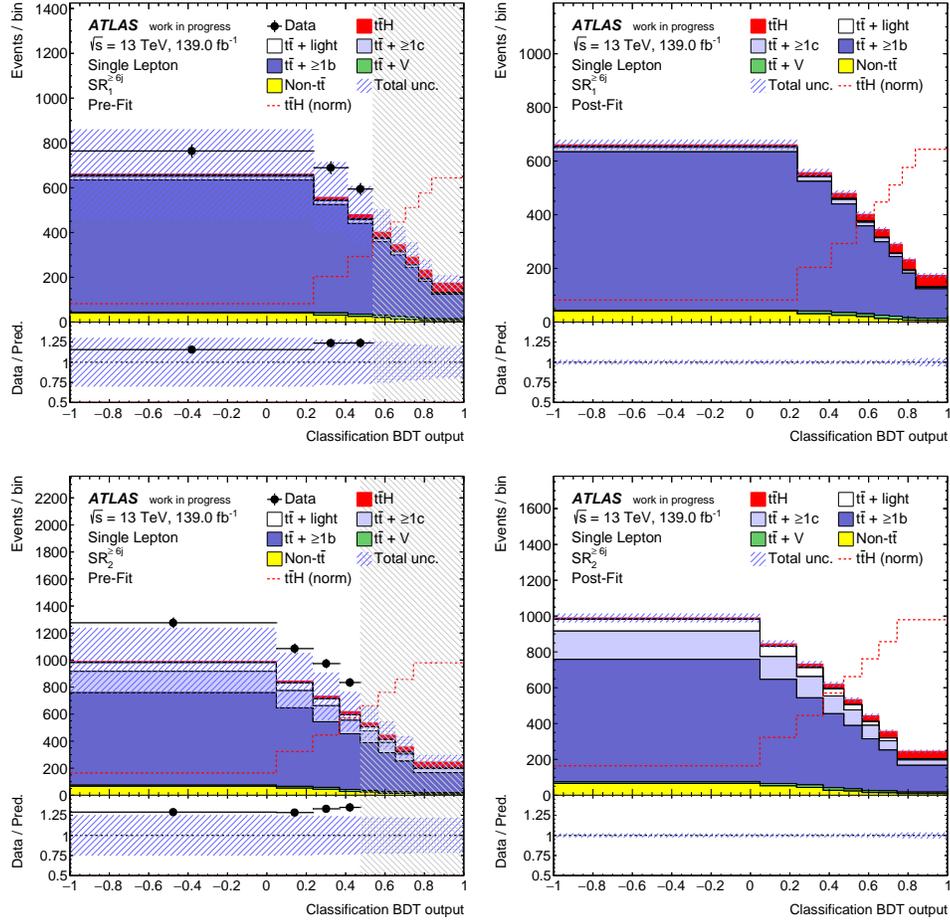


Figure 9.8: Comparison between data and the model for the signal regions  $SR_1^{\geq 6j}$  (top) and  $SR_2^{\geq 6j}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 9.5. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line. Data is only compared to the pre-fit model, and not shown in bins where the  $t\bar{t}H$  signal is expected to contribute more than 5% to the yield, indicated by a gray hashed area.

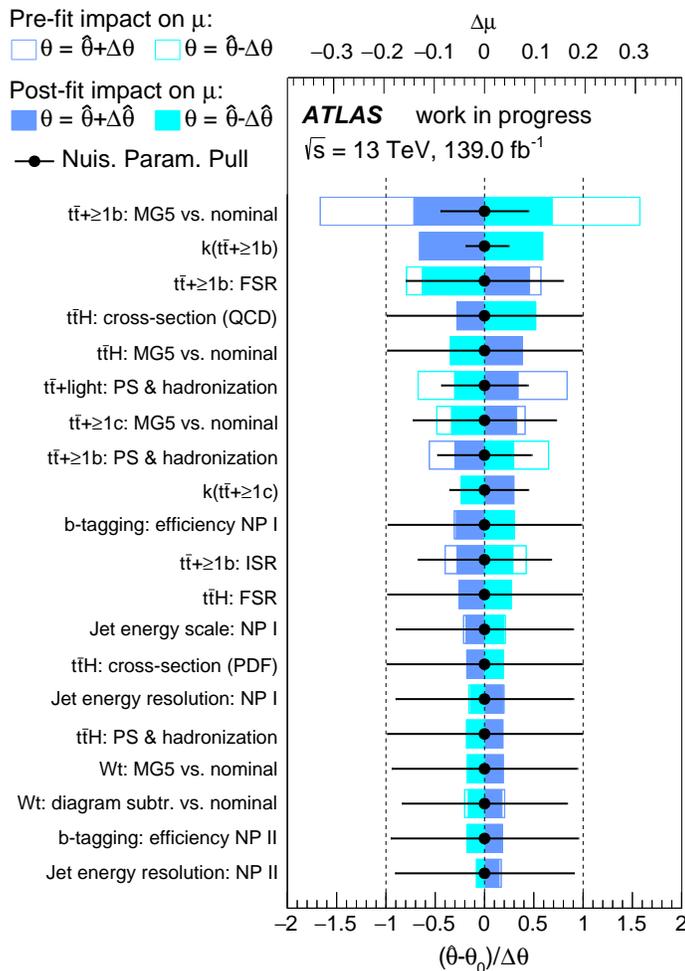


Figure 9.9: The 20 dominant nuisance parameters in the fit, ranked according to their impact on the signal strength. The empty rectangles correspond to the pre-fit impact, while the filled rectangles show the post-fit impact per nuisance parameter. The upper axis shows the impact  $\Delta\mu$ . The pull  $\frac{\hat{\theta} - \theta_0}{\Delta\theta}$  of the nuisance parameter is shown as black points, with the vertical black lines visualizing the post-fit nuisance parameter uncertainty  $\Delta\hat{\theta}$ . MG5 refers to samples generated with MG5\_AMC@NLO+PYTHIA 8.

### 9.6.2 Dominant nuisance parameters and sources of uncertainty

All nuisance parameters are ranked according to their impact, as defined in section 6.7.3. The 20 dominant contributions are summarized in figure 9.9. Pre- and post-fit impact are drawn with empty and filled rectangles, respectively. The upper axis shows the impact  $\Delta\mu$  of each nuisance parameter, while the pull is indicated on the lower axis, and drawn with black points. No pulls are present in this fit to the Asimov dataset by design. Constraints are indicated by the vertical black lines. The central value and pre-fit uncertainty for  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  are not defined, and set to  $\Delta\theta = \theta_0 = 1$  in the figure. MG5 refers to samples generated with MG5\_AMC@NLO+PYTHIA 8.

Systematic uncertainties in the analysis are dominated by the modeling of  $t\bar{t} + \geq 1b$ . The largest individual uncertainty in the analysis is related to the choice of the event generator for  $t\bar{t} + \geq$

$1b$ , defined by comparing the nominal POWHEG+PYTHIA 8 setup to the sample produced with MG5\_AMC@NLO+PYTHIA 8. It predicts large shape variations for  $t\bar{t}+ \geq 1b$  in the most sensitive signal regions, which have an effect specifically in the bins most enriched in  $t\bar{t}H$  signal. This nuisance parameter also predicts shape variations for  $t\bar{t}+ \geq 1b$  in the control regions and is constrained by the fit. The corresponding nuisance parameter is the most dominant in the  $36.1 \text{ fb}^{-1}$  analysis as well. The  $k(t\bar{t}+ \geq 1b)$  normalization factor has the second highest impact, followed by  $t\bar{t}+ \geq 1b$  FSR. Both  $k(t\bar{t}+ \geq 1b)$  and the  $t\bar{t}+ \geq 1b$  FSR nuisance parameter affect the  $t\bar{t}+ \geq 1b$  normalization; the  $t\bar{t}+ \geq 1b$  FSR nuisance parameter defines a variation of the  $t\bar{t}+ \geq 1b$  normalization of around 15% per region, with only a small effect on the shape of the  $t\bar{t}+ \geq 1b$  distributions. The fit determines that it is strongly anti-correlated with  $k(t\bar{t}+ \geq 1b)$ , with a correlation of around 75%. A correlation matrix for this fit is provided in appendix section B.1. The nuisance parameter assigned to  $t\bar{t}+ \geq 1b$  PS and hadronization model similarly affects the  $t\bar{t}+ \geq 1b$  normalization, and predicts variations around 10–25% per region, with slightly larger predicted shape variations. The variations decrease the  $t\bar{t}+ \geq 1b$  yield, and this nuisance parameter is strongly correlated with  $k(t\bar{t}+ \geq 1b)$  as well.

Uncertainties related to the modeling of  $t\bar{t}H$  follow in the ranking, related to the  $t\bar{t}H$  cross-section and the choice of event generator. Several additional nuisance parameters related to  $t\bar{t}H$  also are included within the 20 most dominant contributions. The impact of  $t\bar{t}+ \geq 1c$  and  $t\bar{t}+ \text{light}$  modeling is smaller than the impact of  $t\bar{t}+ \geq 1b$  modeling. The dominant experimental uncertainties are related to  $b$ -tagging and jet calibration.

### Uncertainties grouped by source

Table 9.3 shows contributions to the total predicted uncertainty  $\Delta\mu$  for the  $t\bar{t}H$  signal strength, grouped by sources of uncertainty. The method to obtain these results is equivalent to the description in section 6.7.3. The dominant source of uncertainty is the modeling of  $t\bar{t}+ \geq 1b$ , with smaller contributions from the other components  $t\bar{t}+ \geq 1c$  and  $t\bar{t}+ \text{light}$ . The modeling of  $t\bar{t}H$  is also a large source of uncertainty. The statistical uncertainties related to the background model are significantly reduced in impact compared to the  $36.1 \text{ fb}^{-1}$  analysis due to the use of additional samples to populate the phase space selected by the analysis with more MC events.

Table 9.3: Contributions to the signal strength uncertainty, grouped by sources. The total statistical uncertainty includes effects from the  $k(t\bar{t}+ \geq 1b)$  and  $k(t\bar{t}+ \geq 1c)$  normalization factors, while the intrinsic statistical uncertainty does not.

| Uncertainty source                         | $\Delta\mu$ |       |
|--|-------------|-------|
| <i>Systematic uncertainties</i>            |             |       |
| $t\bar{t}H$ modeling                       | +0.18       | -0.08 |
| $t\bar{t}+ \geq 1b$ modeling               | +0.19       | -0.19 |
| $t\bar{t}+ \geq 1c$ modeling               | +0.07       | -0.07 |
| $t\bar{t}+ \text{light}$ modeling          | +0.08       | -0.07 |
| Other background modeling                  | +0.07       | -0.07 |
| Experimental uncertainties                 | +0.15       | -0.14 |
| Background model statistical uncertainties | +0.04       | -0.05 |
| Total systematic uncertainty               | +0.29       | -0.24 |
| <i>Statistical uncertainties</i>           |             |       |
| $t\bar{t}+ \geq 1b$ normalization          | +0.13       | -0.12 |
| $t\bar{t}+ \geq 1c$ normalization          | +0.05       | -0.05 |
| Intrinsic statistical uncertainty          | +0.12       | -0.11 |
| Total statistical uncertainty              | +0.18       | -0.18 |
| Total uncertainty                          | +0.34       | -0.31 |

# 10. Muon trigger efficiency measurement

The ATLAS muon trigger system identifies events containing one or more muons at various thresholds of transverse momentum. It is described in section 3.2.6. Muons are produced by a wide range of processes, and muon triggers are consequently used in many different physics analyses in the ATLAS collaboration. They are an essential ingredient to the  $t\bar{t}H(b\bar{b})$  analyses presented in chapter 6 and chapter 9.

This chapter describes a measurement of the ATLAS muon trigger efficiency for muons with  $p_T > 100$  GeV. The measurement is performed in two channels, dominated by contributions from  $t\bar{t}$  and  $W$ +jets processes. It is complimentary to measurements performed with decays of  $Z$  bosons to muon pairs (which use a method similar to reference [53]), where the muons typically have smaller transverse momenta. The efficiency is computed with data recorded in 2016–2018, and compared to the expected efficiency from the simulation of the ATLAS detector. The simulation can then be corrected to match the efficiency measured in data. This correction uses the ratio of efficiencies measured in data to the efficiency measured in simulation; the ratio is called scale factor (SF).

Section 10.1 outlines the method used for the trigger efficiency measurement in this chapter. The event selection and definitions of the two channels are provided in section 10.2. Section 10.3 describes the samples of simulated events used in this measurement. Systematic uncertainties affecting the measurement are listed in section 10.4, followed by a presentation of the results in section 10.5.

## 10.1 Analysis method

The efficiency of a trigger is evaluated from a set of events containing objects that the trigger should identify, and then measuring the fraction of events in which a positive trigger decision is made. In order to not bias the measurement, the event selection should not be correlated to the performance of the trigger under investigation. The tag-and-probe method is used to achieve this. Events are selected by a tag trigger, and are required to also contain the object that the probe trigger is designed to identify. The probe trigger efficiency is the fraction of these events in which the probe trigger sends a positive trigger decision.

For the muon trigger efficiency measurements in this chapter, the tag trigger identifies events based on their missing transverse energy,  $E_T^{\text{miss}}$ . The muons used to study the trigger performance mostly originate from  $W$  boson decays, and are produced together with neutrinos. Events with high momentum muons are therefore also generally expected to have a significant amount of  $E_T^{\text{miss}}$ . The event selection requires events to pass the  $E_T^{\text{miss}}$  tag trigger, and to contain exactly one muon.

The probe trigger is a logical *or* between two muon triggers. They identify events with  $p_T > 26$  GeV

muons and an isolation requirement, or muons with  $p_T > 50$  GeV and no isolation requirement, respectively. These two triggers perform very similarly for the muon transverse momentum range  $p_T > 100$  GeV targeted by this measurement, and are therefore combined.

The trigger efficiency is measured separately for each year of data-taking. It varies due to small changes to the trigger algorithms and the trigger chambers in the ATLAS MS system implemented throughout Run-2 of the LHC. No muon trigger SFs are applied in the derivation of the trigger efficiencies.

The trigger efficiency can also be measured in  $t\bar{t}$  decays resulting in one muon and one electron. In this case, an electron tag trigger can be used. The trigger efficiencies obtained with this topology are consistent with those shown in section 10.5.

## 10.2 Event selection and categorization

This section summarizes the selection of events entering the trigger efficiency measurement. Two different channels, one enriched in  $t\bar{t}$  events and another enriched in  $W$ +jets events, are used independently to measure the muon trigger efficiency.

### 10.2.1 Dataset

The measurement uses events from proton–proton collisions at  $\sqrt{s} = 13$  TeV, delivered by the LHC in 2016–2018, and recorded by the ATLAS detector. The events need to fulfill all quality criteria specified in section 3.2.7. The resulting dataset corresponds to an integrated luminosity of  $135.7 \text{ fb}^{-1}$ . Figure 3.2 shows the distribution of the mean number of interactions per bunch crossing for data recorded during the three years considered in this measurement. It also includes additional data recorded by ATLAS that is not used for physics analyses. Data recorded in 2015 is not considered due to the small size of this dataset, which has an integrated luminosity of only  $3.2 \text{ fb}^{-1}$ . Trigger efficiency measurements for the 2015 dataset are performed with  $Z$  boson decays to muon pairs. The statistical uncertainties for trigger efficiencies measured with the  $t\bar{t}$  and  $W$ +jets topologies are very large, making the measurement with these topologies less useful for this small dataset.

### 10.2.2 Object definitions

Muons considered in this analysis must have  $p_T > 25$  GeV, and need to be within  $|\eta| < 2.5$ . They are required to satisfy the *medium* identification operating point, and no requirement is applied regarding their isolation.

Electrons need to have  $p_T > 25$  GeV and be reconstructed within  $|\eta| < 2.47$ , but are removed if they fall into the  $1.37 < |\eta| < 1.52$  transition region between calorimeter barrel and end-cap. They need to satisfy the *tight* identification operating point and the *gradient* isolation operating point.

Jets are required to have  $p_T > 25$  GeV and be located within  $|\eta| < 2.5$ . The overlap removal procedure described in section 4.6 is applied.

### 10.2.3 Definition of the $t\bar{t}$ and $W$ +jets channels

Events in this measurement are recorded with  $E_T^{\text{miss}}$  triggers. They require a threshold of  $E_T^{\text{miss}} > 50\text{--}60$  GeV at L1, and  $E_T^{\text{miss}} > 100\text{--}110$  GeV at the HLT stage across the three years 2016–2018. All selected events are then required to satisfy  $E_T^{\text{miss}} > 200$  GeV. This puts them far above the  $E_T^{\text{miss}}$  trigger thresholds, and outside the range where the  $E_T^{\text{miss}}$  trigger simulation does not model the behavior in data well.

All events need to contain exactly one muon with transverse momentum above 27 GeV, and no electrons with transverse momentum above 25 GeV.

#### $t\bar{t}$ channel

Events in the  $t\bar{t}$  channel need to contain at least four jets. At least one jet has to be  $b$ -tagged at the *medium* operating point. The resulting selection of events is dominated by those originating from  $t\bar{t}$  production.

#### $W$ +jets channel

The  $W$ +jets channel selects mostly events from  $W$ +jets production. It requires events to have at most four jets, and no jet that is  $b$ -tagged at the *medium* operating point. This selection guarantees that there is no overlap between both channels.

## 10.3 Modeling

The muon trigger efficiency obtained in the simulation of ATLAS is evaluated using samples very similar to those described in section 9.2. The event topology targeted in this trigger efficiency measurement is different from the  $t\bar{t}H(b\bar{b})$  analysis, with relaxed  $b$ -tagging and a large amount of  $E_T^{\text{miss}}$  required. Consequently, the samples considered and their treatment are slightly different. The  $t\bar{t}$  process is modeled with POWHEG+PYTHIA 8 [97, 100–103], but is not split into multiple components. Both  $W$ +jets and  $Z$ +jets processes are simulated with SHERPA 2.2.1 [94, 113–115]. Events from the  $Z$ +jets process containing  $b$ - or  $c$ -jets are not scaled by the factor 1.3 used in the  $t\bar{t}H(b\bar{b})$  analysis due to the different phase space region used in this trigger efficiency measurement. The remaining processes have small contributions, and are produced as described in section 9.2.3. These include the production of single top quark,  $t\bar{t}V$ , diboson, and  $t\bar{t}t\bar{t}$  final states. No Higgs boson production processes are considered in this measurement, as their contributions to the  $t\bar{t}$  and  $W$ +jets channels are negligible.

### 10.3.1 Comparison with data

The modeling of the muon transverse momentum and pseudorapidity, compared to data, is shown in figure 10.1 for the  $t\bar{t}$  channel. The corresponding distributions for the  $W$ +jets channel are shown in figure 10.2. These figures only show statistical uncertainties for the prediction. The  $E_T^{\text{miss}}$  distributions

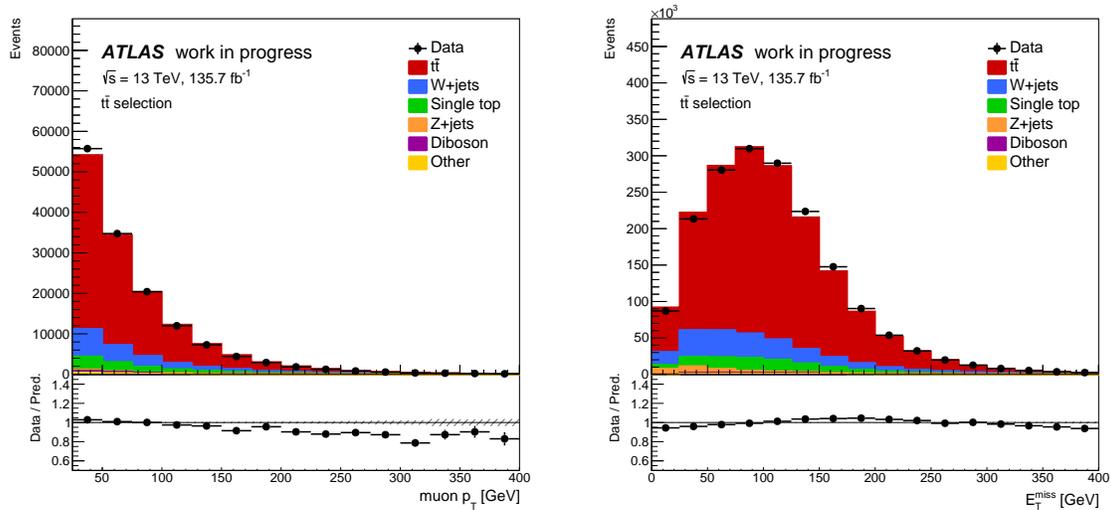


Figure 10.1: Expected distribution of the muon transverse momentum (left) and  $E_T^{\text{miss}}$  (right) in the  $t\bar{t}$  channel, compared to data. An overall normalization factor is applied to simulation to match data, with an effect smaller than 1%. Only statistical uncertainties are shown for the expected distribution, drawn with dashed lines. The  $E_T^{\text{miss}} > 200$  GeV requirement is not applied in the figures showing the  $E_T^{\text{miss}}$  distributions.

are obtained by removing the  $E_T^{\text{miss}} > 200$  GeV requirement. The simulation slightly underestimates data, and an overall normalization factor is applied to the simulation to correct for this effect. In the  $t\bar{t}$  channel, this correction is smaller than 1%. The correction for the  $W$ +jets channel is around 10%. Both of these corrections are within the systematic uncertainties affecting the modeling of the simulation in both channels. An overall correction does not influence the trigger efficiency measurement, as the normalization factor cancels out in the ratio of events used to calculate an efficiency.

In the  $t\bar{t}$  channel, the prediction slightly underestimates data for low muon transverse momenta, and overestimates it at high  $p_T$ . In the  $W$ +jets channel, the prediction underestimates data by a constant factor, independent of the muon  $p_T$ . The  $E_T^{\text{miss}}$  distribution in the  $t\bar{t}$  channel shows a slight discrepancy between data and prediction, and a constant offset in the  $W$ +jets channel above the  $E_T^{\text{miss}} > 200$  GeV threshold. The bulk of events in data are modeled well by the prediction. Systematic uncertainties are introduced to cover effects from mis-modeling that may affect the measurement, they are described in section 10.4.

## 10.4 Systematic uncertainties

This section summarizes the sources of systematic uncertainty considered in the measurement. Trigger efficiencies and scale factors are measured independently for the nominal configuration and the systematic variations. The size of each systematic uncertainty is given by the absolute difference between the measurement obtained with the nominal configuration and with the variation, specific to

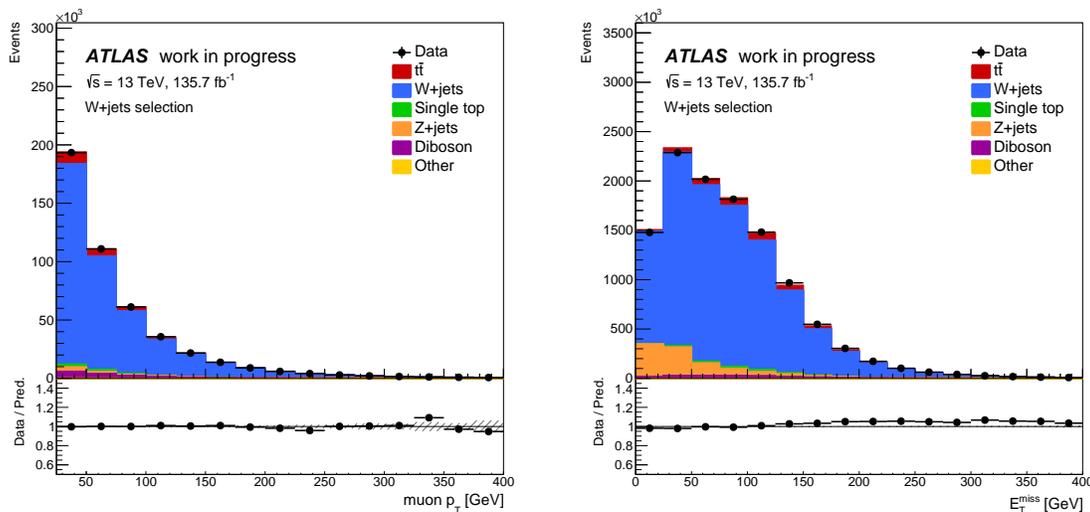


Figure 10.2: Expected distribution of the muon transverse momentum (left) and  $E_T^{\text{miss}}$  (right) in the  $W$ +jets channel, compared to data. An overall normalization factor is applied to simulation to match data, with an effect around 10%. Only statistical uncertainties are shown for the expected distribution, drawn with dashed lines. The  $E_T^{\text{miss}} > 200$  GeV requirement is not applied in the figures showing the  $E_T^{\text{miss}}$  distributions.

each systematic source. All uncertainties arising from the systematic sources considered are assumed to not be correlated, and added in quadrature to obtain the total systematic uncertainty.

The distinction between the  $t\bar{t}$  and  $W$ +jets channels relies on  $b$ -tagging. A variation of the  $b$ -tagging operating point affects the composition of processes contained in both channels, and is considered as a systematic uncertainty. For the  $t\bar{t}$  channel, the nominal *medium*  $b$ -tagging operating point is replaced by the *tight*  $b$ -tagging operating point to define the systematic variation. In the  $W$ +jets channel, the variation from the nominal *medium* operating point to the *loose* operating point is used. The variations for both channels tighten the requirements for events to be considered in the analysis.

The  $E_T^{\text{miss}} > 200$  GeV requirement in the measurement is considered as an additional source of uncertainty. Raising the  $E_T^{\text{miss}}$  threshold to define a systematic uncertainty is not feasible, as any substantial increase in this threshold limits the amount of events remaining for the efficiency measurement. The systematic uncertainty derived from that would be dominated by statistical fluctuations. Instead, a variation of  $E_T^{\text{miss}} > 150$  GeV is considered as a source of uncertainty. By bringing this threshold closer to the trigger threshold, the variation covers possible mis-modeling of the  $E_T^{\text{miss}}$  trigger.

The effect of pile-up on the measurement is considered by raising the jet transverse momentum threshold to  $p_T > 30$  GeV, compared to the nominal requirement of  $p_T > 25$  GeV. This tightened requirement rejects more pile-up events.

The isolation of the reconstructed muon may affect the trigger efficiency, particularly for triggers

that themselves include isolation requirements. In the nominal event selection for the efficiency measurements, no isolation requirements for the reconstructed muons are included. The effect of isolation is evaluated via a systematic uncertainty, calculated by comparing the nominal configuration to two other variations. In one variation, the *FCTight* isolation operating point is applied to the reconstructed muon, while the other variation uses the *FCTTO* operating point. The systematic uncertainty is defined as the larger of the two variations, compared to the nominal configuration.

Lastly, the effect of muon identification requirements is considered. The associated systematic uncertainty is defined as the difference between the nominal measurement, obtained with the *medium* identification operating point, and a measurement using the *high- $p_T$*  operating point.

Additional systematic uncertainties related to muon calibration, similar to those described in section 6.6.2, are negligible.

## 10.5 Results

This section lists the results of the trigger efficiency measurement and the SFs derived. All results are reported separately for each of the years of data-taking considered, split between the MS barrel and end-cap regions, and for both the  $t\bar{t}$  and  $W$ +jets channels. Trigger efficiencies and SFs are listed in section 10.5.1, only including statistical uncertainties. Section 10.5.2 describes the SF results including systematic uncertainties.

### 10.5.1 Trigger efficiencies

Figure 10.3 shows the trigger efficiencies for muons in the barrel region of the MS, derived for the logical *or* between the two triggers described in section 10.1. The corresponding results for the end-cap regions are presented in figure 10.4. The figures show the  $t\bar{t}$  channel on the left, and the  $W$ +jets results on the right. Each row corresponds to one year of data-taking, with 2016 on top, 2017 in the middle, and 2018 on the bottom. The upper part of each figure shows the trigger efficiency as a function of the reconstructed muon transverse momentum. The measurement in data is reported with black points, along with the statistical uncertainty. The predicted trigger efficiency from simulation is shown as a green hashed area, with the size of the area indicating the statistical uncertainty. In the lower part of the figure, the trigger SF is shown. It is calculated as the ratio of the trigger efficiency measured in data to the measurement in simulation, with a statistical uncertainty given by the sum in quadrature of the statistical uncertainties of the efficiencies entering the calculation. All figures also report the measured efficiencies in data and simulation, as well as the measured SF. These are obtained by fitting a constant term to the binned efficiency and SF distributions, only considering the bins with muon  $p_T > 100$  GeV. The efficiencies and SFs show no large dependence on the muon transverse momentum. When including a term proportional to the muon  $p_T$  in the fit, no significant slopes are found. The uncertainty shown in the figure for the trigger efficiency and SF results is the absolute statistical uncertainty, and does not include sources of systematic uncertainty.

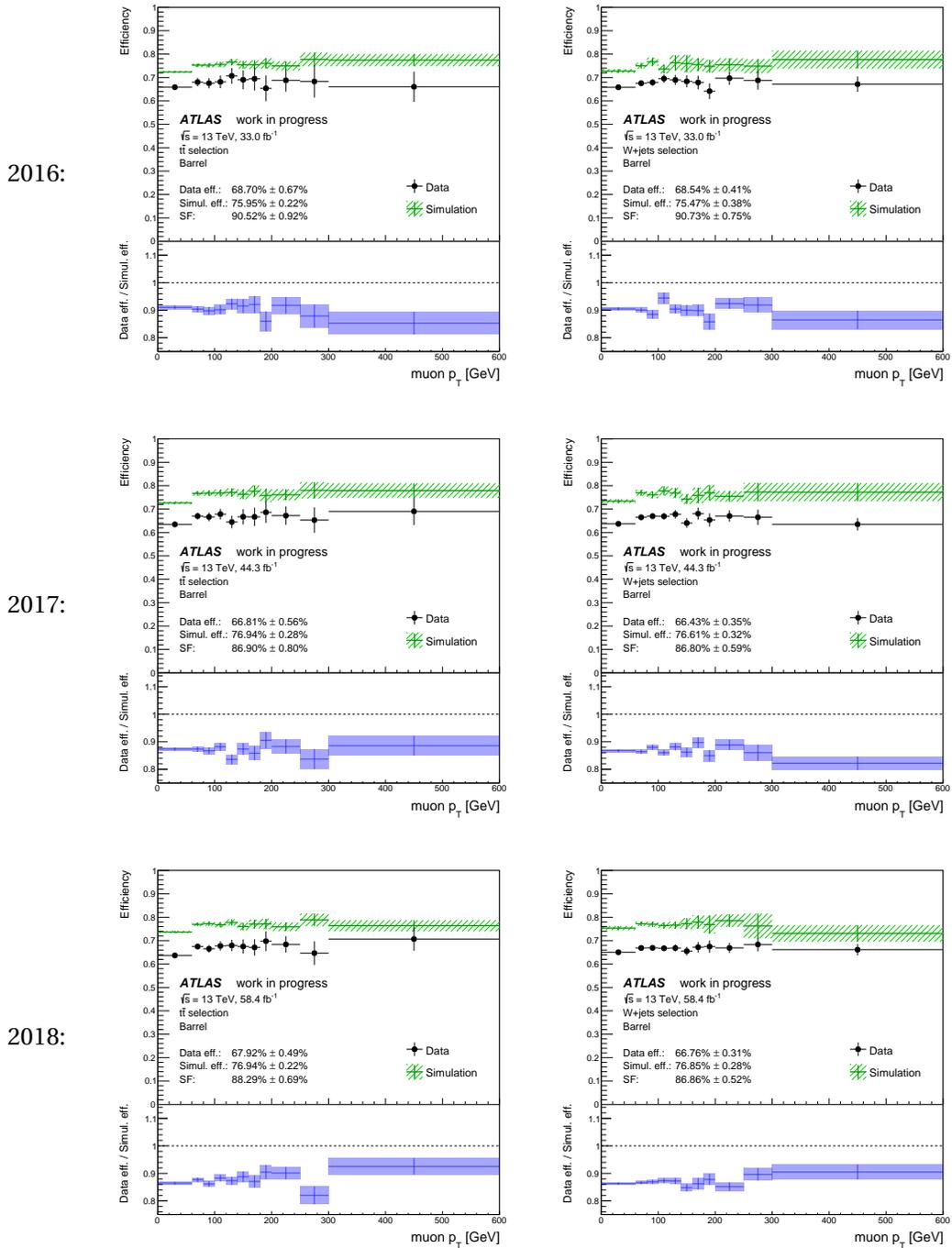


Figure 10.3: Muon trigger efficiencies and SFs in the barrel region, measured in the  $t\bar{t}$  (left) and  $W$ +jets (right) channels, for data recorded in 2016 (top), 2017 (middle), and 2018 (bottom). The upper part of the figures show the trigger efficiencies for data in black, and simulation as a hashed green area. The lower part shows the SF, given by the ratio of efficiency measured in data to simulation. The efficiencies are shown as a function of the reconstructed muon transverse momentum, and the resulting efficiencies and SFs from a fit to muons with  $p_T > 100 \text{ GeV}$  are also listed in the figure. Only statistical uncertainties are included.

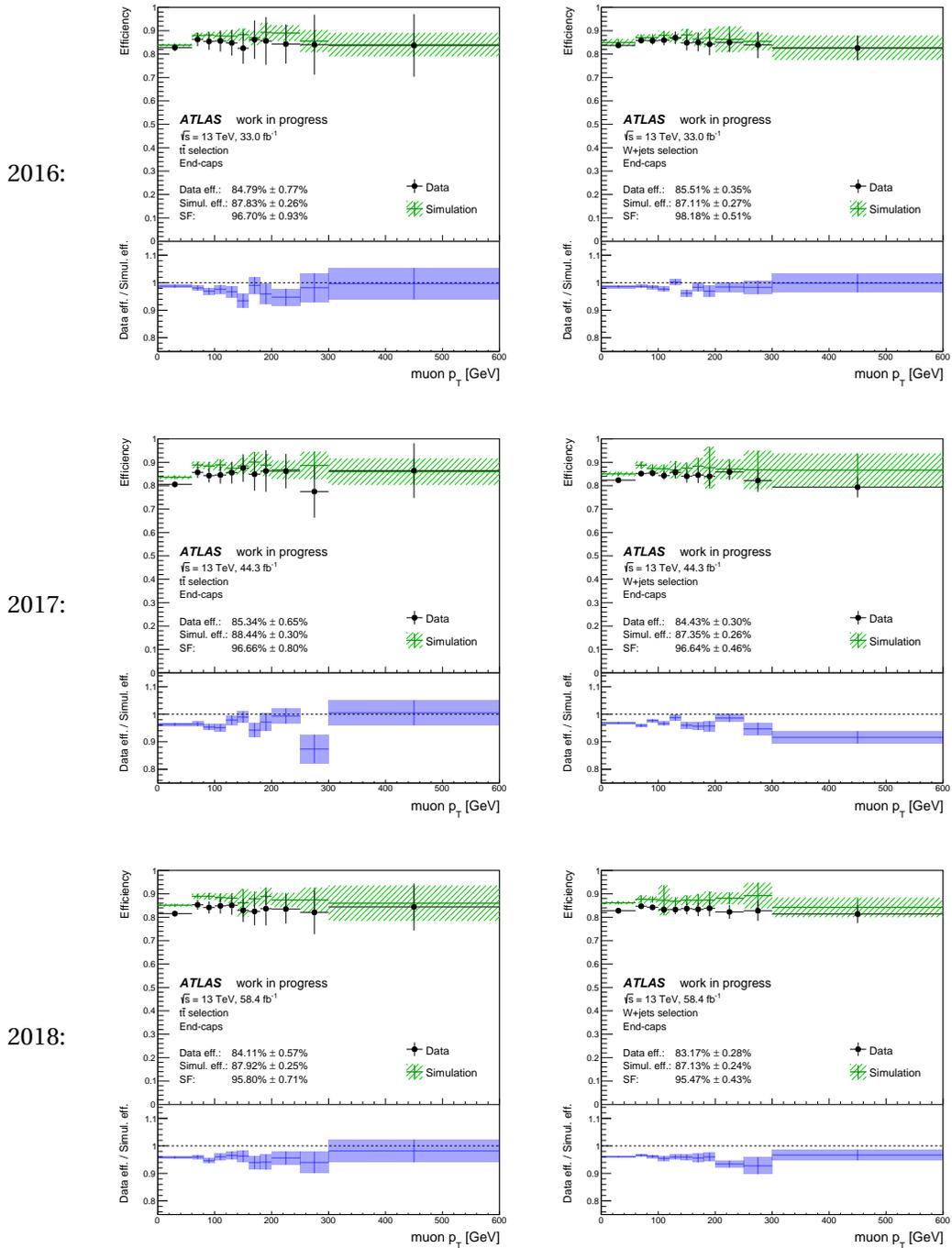


Figure 10.4: Muon trigger efficiencies and SFs in the end-cap regions, measured in the  $t\bar{t}$  (left) and  $W$ +jets (right) channels, for data recorded in 2016 (top), 2017 (middle), and 2018 (bottom). The upper part of the figures show the trigger efficiencies for data in black, and simulation as a hashed green area. The lower part shows the SF, given by the ratio of efficiency measured in data to simulation. The efficiencies are shown as a function of the reconstructed muon transverse momentum, and the resulting efficiencies and SFs from a fit to muons with  $p_T > 100$  GeV are also listed in the figure. Only statistical uncertainties are included.

Table 10.1: Summary of trigger efficiencies for data taken between 2016 and 2018, and for simulation. The results are reported separately for the barrel and end-cap regions, and split by channel. Only absolute statistical uncertainties are included.

| Year | Region   | Channel    | Data efficiency    | Simulation efficiency |
|------|----------|------------|--------------------|-----------------------|
| 2016 | Barrel   | $t\bar{t}$ | $68.7\% \pm 0.7\%$ | $76.0\% \pm 0.2\%$    |
|      |          | $W$ +jets  | $68.5\% \pm 0.4\%$ | $75.5\% \pm 0.4\%$    |
|      | End-caps | $t\bar{t}$ | $84.8\% \pm 0.8\%$ | $87.8\% \pm 0.3\%$    |
|      |          | $W$ +jets  | $85.5\% \pm 0.4\%$ | $87.1\% \pm 0.3\%$    |
| 2017 | Barrel   | $t\bar{t}$ | $66.8\% \pm 0.6\%$ | $76.9\% \pm 0.3\%$    |
|      |          | $W$ +jets  | $66.4\% \pm 0.4\%$ | $76.6\% \pm 0.3\%$    |
|      | End-caps | $t\bar{t}$ | $85.3\% \pm 0.7\%$ | $88.4\% \pm 0.3\%$    |
|      |          | $W$ +jets  | $84.4\% \pm 0.3\%$ | $87.4\% \pm 0.3\%$    |
| 2018 | Barrel   | $t\bar{t}$ | $67.9\% \pm 0.5\%$ | $76.9\% \pm 0.2\%$    |
|      |          | $W$ +jets  | $66.8\% \pm 0.3\%$ | $76.9\% \pm 0.3\%$    |
|      | End-caps | $t\bar{t}$ | $84.1\% \pm 0.6\%$ | $87.9\% \pm 0.3\%$    |
|      |          | $W$ +jets  | $83.2\% \pm 0.3\%$ | $87.1\% \pm 0.2\%$    |

The measured trigger efficiencies generally agree between the  $t\bar{t}$  and  $W$ +jets channels within their statistical uncertainties. They vary across the three years considered due to changes in the ATLAS MS, as well as changes to the trigger algorithms. The trigger efficiencies measured in the barrel region are significantly below the efficiencies in the end-caps. This is mostly caused by the limited coverage of the L1 trigger system in the barrel region, where detector support structure and elevator paths to access the inner parts of ATLAS necessitate gaps in the RPC coverage. The L1 system covers 99% of the end-caps, and around 80% of the barrel.

The simulation overestimates the trigger efficiency in the RPCs, and consequently the trigger efficiency measured in data is significantly below the efficiency obtained from simulation in the barrel region. In the end-caps, the agreement between data and simulation is better, with SFs closer to unity.

A summary of the efficiencies measured for muons with  $p_T > 100$  GeV for the years 2016–2018 and both channels is provided in table 10.1. Only the absolute statistical uncertainties are included in the table.

The statistical uncertainties for the trigger efficiency measured in data are smaller for the  $W$ +jets channel than for the  $t\bar{t}$  channel, both in the barrel and end-cap regions. The efficiencies in simulation have comparable statistical uncertainties between the channels.

### 10.5.2 Scale factors and impact of systematic uncertainties

The measured muon trigger SFs for muons with  $p_T > 100$  GeV recorded in the years 2016–2018, split between barrel and end-cap regions, and split by channel, are summarized in table 10.2. The results show the total absolute uncertainty for the SFs, as well as their split into statistical and systematic components. The total uncertainty is calculated as the sum in quadrature of these two components.

Table 10.2: Summary of SFs for the years 2016–2018. The results are reported separately for the barrel and end-cap regions, and split by channel. The total absolute uncertainties for the SFs are shown, and their split into statistical and systematic components is also included.

| Year | Region   | Channel    | SF                 | (stat.) | (syst.) |
|------|----------|------------|--------------------|---------|---------|
| 2016 | Barrel   | $t\bar{t}$ | $90.5\% \pm 1.2\%$ | (0.9%)  | (0.8%)  |
|      |          | $W$ +jets  | $90.7\% \pm 1.1\%$ | (0.8%)  | (0.8%)  |
|      | End-caps | $t\bar{t}$ | $96.7\% \pm 2.6\%$ | (0.9%)  | (2.5%)  |
|      |          | $W$ +jets  | $98.2\% \pm 0.9\%$ | (0.5%)  | (0.7%)  |
| 2017 | Barrel   | $t\bar{t}$ | $86.9\% \pm 1.0\%$ | (0.8%)  | (0.6%)  |
|      |          | $W$ +jets  | $86.8\% \pm 0.7\%$ | (0.6%)  | (0.4%)  |
|      | End-caps | $t\bar{t}$ | $96.7\% \pm 1.1\%$ | (0.8%)  | (0.7%)  |
|      |          | $W$ +jets  | $96.6\% \pm 1.1\%$ | (0.5%)  | (1.0%)  |
| 2018 | Barrel   | $t\bar{t}$ | $88.3\% \pm 1.7\%$ | (0.7%)  | (1.6%)  |
|      |          | $W$ +jets  | $86.9\% \pm 1.2\%$ | (0.5%)  | (1.1%)  |
|      | End-caps | $t\bar{t}$ | $95.8\% \pm 1.7\%$ | (0.7%)  | (1.5%)  |
|      |          | $W$ +jets  | $95.5\% \pm 0.8\%$ | (0.4%)  | (0.7%)  |

All systematic components are added in quadrature to obtain the total systematic uncertainty.

The SFs vary significantly between the three years, caused by differences in the active trigger chambers in the MS. The mis-modeling of the RPC efficiency in simulation needs to be corrected with SFs that deviate more significantly from unity in the barrel region compared to the end-caps. The measurements in the  $t\bar{t}$  and  $W$ +jets channels per year and detector region are compatible with each other within their uncertainties. Relative statistical uncertainties in the measurement reach up to 1%, while the systematic uncertainties are generally slightly larger. The measured SFs are in good agreement with those derived from decays of  $Z$  bosons to muon pairs, when considering their associated uncertainties.

A detailed look at the systematic uncertainties, split into the components described in section 10.4, is provided in table 10.3. It lists the arithmetic mean of the relative systematic uncertainty on the SF per source, averaged over the three years 2016–2018 of data-taking.

The smallest uncertainty source is the variation of the muon isolation requirement, followed by the variation of the  $b$ -tagging operating point used to define the channels. The variation of the  $E_T^{\text{miss}}$  requirement and the jet  $p_T$  threshold have a comparable impact, and the largest source of uncertainty is the variation of the muon identification operating point. The impact per source of systematic uncertainty is generally larger in the  $t\bar{t}$  channel than in the  $W$ +jets channel. A contribution to this effect arises from the larger statistical uncertainties in the data efficiency measurement in the  $t\bar{t}$  channel compared to the  $W$ +jets channel; a statistical fluctuation in the determination of the SF measured with the systematic variation can increase the impact from a systematic source.

Table 10.3: Average size of relative systematic uncertainty on the SF per source, calculated as the arithmetic mean over the three years 2016–2018 of data-taking. The uncertainties are reported separately per detector region and channel in the measurement.

| Uncertainty source            | Region   | Channel    |           |
|-------------------------------|----------|------------|-----------|
|                               |          | $t\bar{t}$ | $W$ +jets |
| $E_T^{\text{miss}}$ threshold | Barrel   | 0.5%       | 0.3%      |
|                               | End-caps | 0.9%       | 0.3%      |
| $b$ -tagging operating point  | Barrel   | 0.2%       | 0.1%      |
|                               | End-caps | 0.3%       | 0.2%      |
| Jet $p_T$ threshold           | Barrel   | 0.7%       | 0.3%      |
|                               | End-caps | 0.5%       | 0.1%      |
| Muon isolation                | Barrel   | 0.1%       | 0.1%      |
|                               | End-caps | 0.1%       | 0.1%      |
| Muon identification           | Barrel   | 0.5%       | 0.6%      |
|                               | End-caps | 1.0%       | 0.7%      |

# 11. Differential cross-section approximation

The MEM presented in chapter 7 is a powerful analysis technique with large computational demand. The calculation of MEM likelihoods relies on two ingredients: the transfer function and the fully differential cross-section. Equation (7.7) shows the general form of the likelihoods and is repeated here:

$$L_\alpha = \frac{1}{\sigma_\alpha} \sum_{\text{perm.}} \int_{\vec{Y}} T(\vec{X}|\vec{Y}) d\sigma_\alpha(\vec{Y}). \quad (11.1)$$

In this expression,  $T(\vec{X}|\vec{Y})$  is the transfer function,  $\vec{X}, \vec{Y}$  are the kinematics of reconstructed objects and partons, respectively, and  $\alpha$  denotes the process of interest. The transfer function is generally fast to evaluate. The evaluation of the differential cross-section  $d\sigma_\alpha/\sigma_\alpha$  is much slower for processes involving the interaction of many particles.

The calculation of the MEM discriminant for the  $t\bar{t}H(b\bar{b})$  analysis described in chapter 6 requires around  $10^{12}$  evaluations of  $d\sigma_\alpha$ . This includes the calculation for all systematic variations needed for the statistical analysis in the  $\text{SR}_1^{\geq 6j}$  region. When the event selection and kinematics of the selected event in the analysis change, which happens while optimizing it for sensitivity, the calculation has to be repeated. Consequently, the MEM implementation needs to make sufficient approximations to limit the calculation time to a reasonable level.

This chapter investigates the feasibility of approximating  $d\sigma_\alpha$  with machine learning methods. A sufficiently precise approximation  $d\hat{\sigma}_\alpha$  can replace the exact expression  $d\sigma_\alpha$  in the likelihood evaluation without significant adverse effect. If such an approximation is possible, it can be optimized for speed and may be used to significantly reduce the computation time needed for MEM calculations.

Section 11.1 provides a general introduction to the method and discusses its challenges. An example for a scattering process with two particles in the final state is provided in section 11.2. Section 11.3 demonstrates the feasibility of the method for a complex scattering process where six particles are produced. An alternative approach for faster MEM calculations with machine learning techniques is discussed in section 11.4 and compared to the approach studied in this chapter.

## 11.1 Overview

The expression for a fully differential cross-section of a scattering process at the LHC has the form

$$d\sigma_\alpha \propto \delta^4(p_{\text{net}}) dx_1 dx_2 \prod_{i=1}^N \frac{d^3\vec{p}_i}{(2\pi)^3 2E_i}, \quad (11.2)$$

with momentum fractions  $x_1$  and  $x_2$  carried by the colliding partons,  $N$  final state particles, and a four-vector  $p_{\text{net}}$  corresponding to the difference in four-momentum between all incoming and

outgoing particles. While  $d\sigma_\alpha$  is an expression with  $3N + 2$  degrees of freedom, it is non-zero only when four-momentum is conserved. The phase space where four-momentum is conserved has  $3N - 2$  degrees of freedom, with three degrees of freedom per final state particle, two degrees of freedom related to the colliding partons, and the four constraints applied. As it is computationally inexpensive to check whether four-momentum is conserved in any calculation of interest, the approximation of  $d\sigma_\alpha$  is implemented in a phase space where four-momentum conservation is enforced.

### 11.1.1 Fully differential cross-sections

After partial integration over four degrees of freedom, the fully differential cross-section for the scattering of two partons into two final states  $p_i = (E_i, \vec{p}_i)$  is given by

$$d\sigma_{2 \rightarrow 2} = \frac{1}{64 (2\pi)^2 E_B^4} \sum_{j,k} \frac{1}{x_1 x_2 E_1 E_2} f_j(x_1) f_k(x_2) |\mathcal{M}_{2 \rightarrow 2}|^2 p_T |\vec{p}_1| |\vec{p}_2| dp_T d\phi_1 d\eta_1 d\eta_2, \quad (11.3)$$

with additional details provided in appendix section C.1. The energy of the colliding protons is  $E_B$ , and  $p_T$  is the transverse momentum of the final state particles, which is equal for both in this case.

The extension to final states involving additional particles is straightforward; a phase space factor  $d^3 \vec{p}_i / (2\pi)^3 2E_i$  per final state particle extends the expression. The production of six final state particles can be parameterized as

$$d\sigma_{2 \rightarrow 6} = \frac{1}{1024 (2\pi)^{14} E_B^4} \sum_{j,k} \frac{1}{x_1 x_2} f_j(x_1) f_k(x_2) |\mathcal{M}_{2 \rightarrow 6}|^2 \left[ \prod_{i=1}^5 \frac{1}{E_i} p_{T_i} |\vec{p}_i| dp_{T_i} d\phi_i d\eta_i \right] \frac{1}{E_6} dp_{z_6} \quad (11.4)$$

### 11.1.2 Challenges

Obtaining a good approximation for a differential cross-section becomes increasingly challenging for final states with many particles. The differential cross-section can generally not be factorized into components that separately describe the dependence on each kinematic variable; an approximation needs to correctly model correlations between kinematic variables. The expression can vary over many orders of magnitude across phase space, and the matrix element can be highly non-uniform when describing the effect of resonances in the scattering process. A suitable parametrization of the differential cross-section helps mitigate this problem, motivating the use of an expression differential in the invariant masses of internal resonances. An analytical change of variables to achieve this is not always possible [137].

The differential cross-section across the phase space with  $3N - 2$  degrees of freedom vanishes in many regions. These regions describe events where the colliding partons have more energy than what is provided by the beam of protons,  $E_B$ . The PDFs vanish in these regions, and the differential cross-section approaches zero in their vicinity as  $\lim_{x \rightarrow 1} f(x) = 0$ . The distribution of such regions with vanishing cross-section depends on the parametrization chosen.

A suitable parameterization choice for the differential cross-section simplifies the approximation task. No matter the parameterization, the expression generally remains a function with many distinct features in its distribution across the phase space, which all need to be captured by a good approximation.

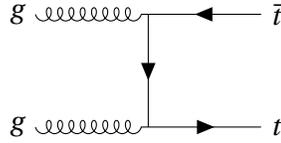


Figure 11.1: Exemplary Feynman diagram for gluon-initiated  $t\bar{t}$  production.

## 11.2 Final state with two particles

This section demonstrates the approximation of the differential cross-section for the production of a two-particle system as a proof of concept. A possible parameterization for such a process is given in equation (11.3). The production of two particles in head-on collisions is invariant under a global rotation in the  $\phi$  direction, motivating the parametrization as a function of  $d\phi_1$  to take advantage of this symmetry.

The process studied in this section is the production of a top quark pair, without its subsequent decays. Only gluon-initiated collisions are considered; no significant impact is expected when including  $t\bar{t}$  production via quark–antiquark annihilation. An exemplary Feynman diagram is shown in figure 11.1.

### 11.2.1 Approximation method

The approximation of  $d\sigma_{t\bar{t}}$  is performed with the Foam algorithm [154]. This algorithm is designed for MC event generation and integration.

The Foam algorithm is defined for an integrand  $\rho$ , defined across a hypercube. The algorithm samples the integrand across this hypercube, and then divides the hypercube into two cells. These cells are hyperrectangles, obtained by a split along a hyperplane. In subsequent steps, the distribution of the integrand across each cell is sampled, and further binary splits of cells are performed until a maximum amount of cells is reached. The resulting grid of cells is called a foam. The foam defines an approximation  $\rho'$  for the integrand  $\rho$ ; this approximation is constant across each cell. An event weight  $w = \rho'/\rho$  is defined for each event generated when sampling cells. The implementation used in this study optimizes the binary splits of cells such that the variance of weights  $w$  is minimized.

When identifying  $\rho = d\sigma_{t\bar{t}}$ , a foam can be built to obtain an approximation  $\rho' = d\hat{\sigma}_{t\bar{t}}$ . This foam spans a four-dimensional hypercube, with edges along  $dp_T$ ,  $d\phi_1$ ,  $d\eta_1$ , and  $d\eta_2$ . As the differential cross-section is constant across  $d\phi_1$ , no cell splits are made in this direction. The foam defines a lookup table for the differential cross-section  $d\sigma_{t\bar{t}}$  with a constant approximation  $d\hat{\sigma}_{t\bar{t}}$  across each cell. Cells are large where  $d\sigma_{t\bar{t}}$  is approximately constant, and small where it varies quickly as a function of the edges of the hypercube. Due to the binary tree structure of the foam,  $d\hat{\sigma}_{t\bar{t}}$  is fast to evaluate. The quality of the approximation  $d\hat{\sigma}_{t\bar{t}}$  is limited by the amount of cells used in the foam.

### 11.2.2 Implementation

A foam with 50 000 cells is constructed to describe  $d\sigma_{t\bar{t}}$ . The implementation of the function uses the form specified in equation (11.3). The LO matrix element is provided by MADGRAPH5\_AMC@NLO [95], denoted as MG5\_AMC@NLO in the following. The CTEQ6L1 [120] PDF set is used, evaluated with the LHAPDF [17] interface. Renormalization and factorization scale are set to the top quark mass in the simulation, which is 172.5 GeV for this study. The beam energy is set to  $E_B = 6.5$  TeV. A sample of one million simulated events is produced with MG5\_AMC@NLO with equivalent settings. This sample provides the reference for the differential cross-section distributions considered to evaluate the foam performance.

The distribution of the foam in one dimension  $u$  is obtained by marginalizing over the three remaining dimensions. This is done by considering all cells that overlap with the interval  $[u_1, u_2]$ . The prediction  $d\hat{\sigma}_{t\bar{t}}$  in all these cells is summed, weighted by the overlap between each cell and the interval. This process can be repeated for many intervals to obtain an expression for the normalized differential cross-section in one variable,  $1/\sigma_{t\bar{t}} d\sigma_{t\bar{t}}/du$ .

### 11.2.3 Results

Figure 11.2 shows the foam approximation of the normalized differential cross-section as a function of each of the four variables describing the full phase space for this process. The distribution of the foam is shown as a dashed blue line, overlaid on top of a reference distribution obtained with the events generated with MG5\_AMC@NLO. The reference distribution is shown in green, with statistical uncertainties indicated as hashed gray lines. No statistical uncertainty related to the approximation by the foam is included.

The foam describes all four variables very well. The transverse momentum distribution is accurately approximated across four orders of magnitude. The distribution of the azimuthal angle  $\phi$  is completely uniform due to the design of the foam, and therefore accurately describes the rotational symmetry. Both pseudorapidity distributions for the top quarks are also accurately described.

This method works well for final states with limited degrees of freedom. Its performance depends on the parameterization chosen for the foam, and the dimensionality of the foam corresponds to the amount of degrees of freedom in the process. The performance significantly reduces for processes involving the production of many particles. Since the construction of foams with hundreds of millions of cells becomes computationally prohibitive, the average amount of cell splits per dimension of the process decreases when adding more final state particles. The following section provides an alternative method for such cases.

## 11.3 Final state with six particles

The approximation of differential cross-sections with neural networks (see section 5.3.2) offers several benefits. A neural network can handle any amount of input parameters, so it is possible to provide inputs beyond the  $3N - 2$  parameters required to fully describe final states with  $N$  particles. This

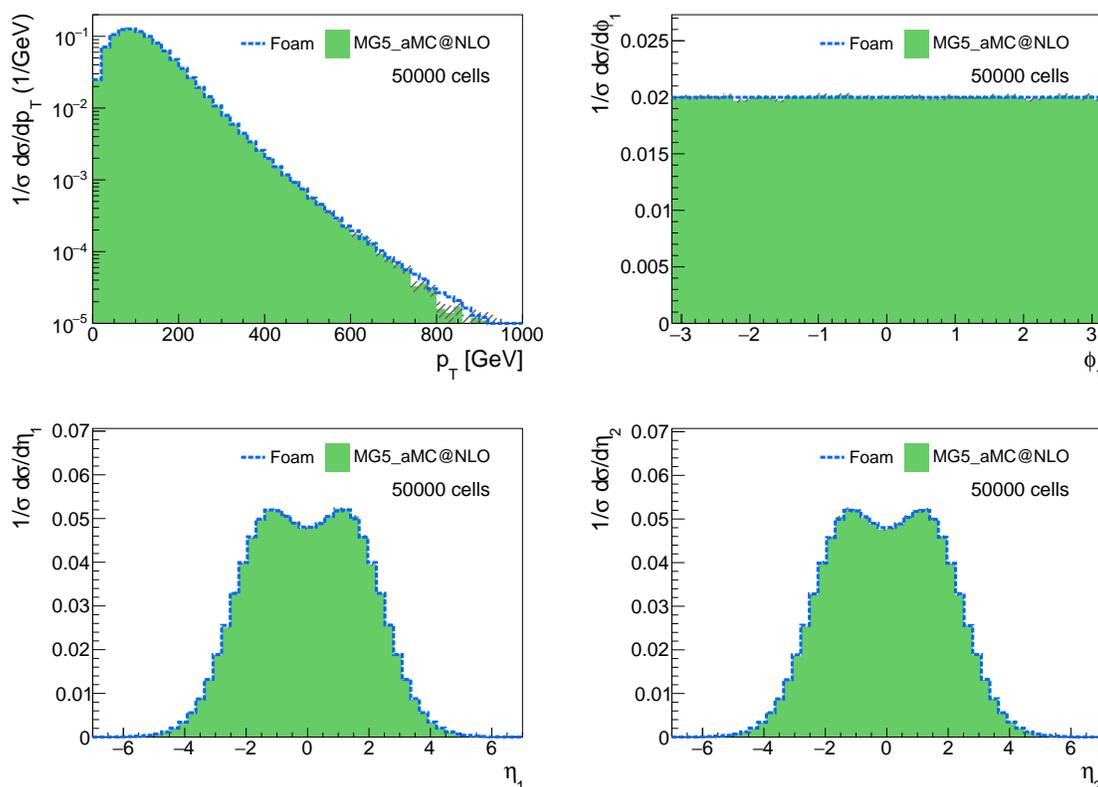


Figure 11.2: Prediction for the differential cross-section of  $t\bar{t}$  production by a foam with 50 000 cells, drawn as dashed blue lines. The distributions are shown as a function of the four variables used to parameterize the fully differential cross-section, and compared to a reference set of events generated with MG5\_AMC@NLO. The reference distribution is shown in green, statistical uncertainties are indicated with hashed gray lines.

allows to use all physically motivated variables, which strongly affect the differential cross-section, such as the momentum fractions  $x_1$ ,  $x_2$  and the invariant masses of internal resonances. A foam built in  $d > 3N - 2$  dimensions approximates a function that is non-zero only in specific regions defined by a delta distribution, and thus is not useful in practice. A neural network with  $d > 3N - 2$  input nodes does not suffer from this issue; it can make use of all input variables for the function approximation task, and does not need to learn the relation between input variables. An implementation with neural networks can also take advantage of highly optimized libraries to increase the computational performance.

This section describes the approximation of the fully differential cross-section for  $t\bar{t}$  production, with subsequent decays of the top quarks into  $W$  bosons and bottom quarks, and also including the decays of the  $W$  bosons. Figure 11.3 visualizes this process.

There are six final state particles, for a total of 16 degrees of freedom. Top quark pair production is a common background process for many analyses at the LHC. It represents an important step towards the  $t\bar{t} + b\bar{b}$  process, for which the MEM likelihoods are calculated for the  $t\bar{t}H(b\bar{b})$  analysis

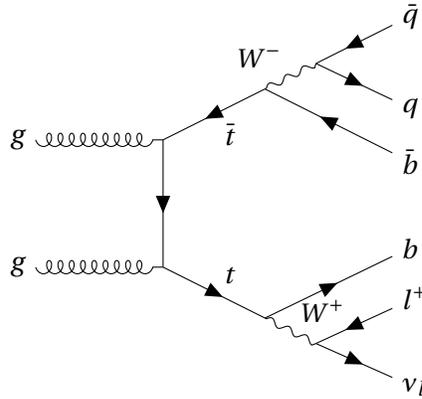


Figure 11.3: Exemplary Feynman diagram for gluon-initiated  $t\bar{t}$  production, with subsequent decays into a final state with six particles.

as described in chapter 7. This process also includes four resonances, in form of the two top quarks and two  $W$  bosons. It provides a good benchmark for the performance of neural networks when approximating its differential cross-section  $d\sigma_{t\bar{t}}$ .

### 11.3.1 General considerations

The differential cross-section for the production of six particles is given in equation (11.4). When dividing this expression by the total cross-section,

$$f(\vec{Y}) = \frac{1}{\sigma_{t\bar{t}}} d\sigma_{t\bar{t}}, \quad (11.5)$$

the resulting distribution describes the probability density for generating a final state with a configuration  $\vec{Y}$ . A set of samples can be generated in the phase space  $\vec{Y}$ , and each sample labeled with its associated probability density. The configuration  $\vec{Y}$  is characterized by a set of kinematic variables, which can be used as input features to a neural network. The network is then trained to approximate  $f(\vec{Y})$ , or equivalently  $d\sigma_{t\bar{t}}$ .

The probability density  $f(\vec{Y})$  varies across many orders of magnitude. For practical applications, it is particularly important to accurately approximate regions of the phase space where the probability density is large. Such regions contribute most to the calculation of a MEM likelihood. The performance of a neural network approximation depends on the distribution of training events across the space  $\vec{Y}$ , and on the loss function chosen for the training. The training events for this study are distributed like  $f(\vec{Y})$ . Most events are concentrated in regions where the probability density is high. Contributions to the loss function are dominated by such regions, hence the performance of the network will generally be best in regions where the probability density is large.

As described in section 11.1.2, large regions of phase space have a vanishing probability density. It is possible to also generate events in these regions, and to add them to the set of events used for network training. No performance improvements are obtained when including such events in this study, so the network training only uses events with non-zero associated probability densities. The

network will therefore generally predict a non-zero differential cross-section even in regions where the differential cross-section vanishes due to PDFs. These regions are characterized by  $x_1 \geq 1$  or  $x_2 \geq 2$ ; the condition is computationally inexpensive to check. The neural network prediction in such a region can be replaced by a vanishing differential cross-section if required.

### 11.3.2 Implementation

This section describes details about the generation of events used to train the neural network and tests its performance. It also includes details about the network architecture.

#### Event generation

The  $t\bar{t}$  process is simulated at LO with MG5\_AMC@NLO, using the CTEQ6L1 PDF set. Only gluon-initiated interactions are considered, and only the decay topology visualized in figure 11.3 is used. The positively charged  $W^+$  boson is forced to decay into leptons, while the negatively charged  $W^-$  decays into quarks. Final states other than bottom quarks are treated as massless. The top quark mass is set to 173 GeV, the  $W$  boson mass to 80.42 GeV, and the bottom quark mass to 4.7 GeV. Both renormalization and factorization scale are set to the  $Z$  boson mass. The beam energy is set to  $E_B = 6.5$  TeV. Neither of these choices are expected to have significant qualitative impact of the neural network approximation quality. The total cross-section for this process is  $\sigma_{t\bar{t}} = 77.2$  pb =  $198 \cdot 10^{-9}$  GeV $^{-2}$ . A set of 50 million events is generated for network training. Tests of the network performance use a separate dataset, generated with the same configuration. The events in all datasets used for training and testing of the neural network are distributed according to the probability density  $f(\vec{Y})$  given in equation (11.5).

#### Differential cross-section calculation

The calculation of the differential cross-section is implemented according to equation (11.4). The LO matrix element is provided by MG5\_AMC@NLO, the CTEQ6L1 PDF set is used with the LHAPDF interface. All settings correspond to those used in the event generation. The normalized differential cross-section is calculated for every generated event for use in network training and testing.

#### Neural network design

The neural network is implemented with Keras [155], using the TensorFlow [156] backend.

A set of 34 input variables is used. These include the 16 variables parameterizing the differential cross-section. Those variables are  $p_T$ ,  $\eta$  and  $\phi$  for each of the four final state quarks and the final state charged lepton, as well as the neutrino momentum in the  $z$  direction. An additional set of 18 variables is calculated for every event and included in the inputs to the neural network. These variables include the invariant mass, as well as  $p_T$ ,  $\eta$  and  $\phi$  for each of the four resonances in the scattering process, the two top quarks and the two  $W$  bosons. The momentum fractions  $x_1$  and  $x_2$  form the last two input variables. The use of this set of 34 input variables significantly improves the quality of the neural network approximation compared to using just 16 variables that fully specify

the differential cross-section. A scaling is applied to the inputs for the neural network to speed up the training procedure. The normalized differential cross-section, which the network is designed to approximate, is also scaled for the training procedure, such that its distribution over the training set has a standard deviation of unity.

The neural network used to predict the differential cross-section as a function of the inputs is a fully connected feedforward neural network. It has nine hidden layers, with 512, 384, 320, 320, 256, 256, 192, 128, 64 nodes. ReLU activation functions are used throughout, with the exception of the last layer. The last layer uses a softplus activation function, and the network thus cannot predict a negative differential cross-section. The network uses the mean absolute error as the loss function for training. This quantity is given by the mean absolute difference between the network prediction for the scaled differential cross-section, and the reference value for it. The minimization of the loss function is performed with the ADADELTA [157] algorithm.

The network training is performed with 80% of the training set. The remaining 20% are used to validate the neural network performance during training. Training over 500 epochs takes around 24 hours on a NVIDIA Tesla K40 GPU when using the set of 50 million training events.

### 11.3.3 Results

Figure 11.4 shows the distribution of one million events from a test dataset as a function of their differential cross-section and the corresponding network prediction. The horizontal axis specifies the logarithm of the normalized differential cross-section,  $\log_{10}(1/\sigma_{i\bar{i}} d\sigma_{i\bar{i}})$ . The network prediction is provided along the vertical axis, as  $\log_{10}(1/\sigma_{i\bar{i}} d\hat{\sigma}_{i\bar{i}})$ . The fraction of events in each bin across this two-dimensional plane is indicated by the bin color, and a logarithmic scale is used. All events are located along the diagonal gray line for a network where  $d\hat{\sigma}_{i\bar{i}} = d\sigma_{i\bar{i}}$ .

Most events are located in bins along the diagonal, over several orders of magnitude of their differential cross-sections. The network provides an accurate estimate for the differential cross-sections of these events. The approximation  $d\hat{\sigma}_{i\bar{i}}$  is less accurate for events where  $\log_{10}(1/\sigma_{i\bar{i}} d\sigma_{i\bar{i}}) < -20$ . This affects a small fraction of events. Only few events in the training sample are located in this region. The absolute error for these events is small, since their associated differential cross-sections are small. Both of these effects combined mean that the penalty in the neural network loss function when not accurately modeling this region is small.

A one-dimensional visualization is provided in figure 11.5. It shows the distribution of one million events as a function of  $\log_{10}(1/\sigma_{i\bar{i}} d\sigma_{i\bar{i}})$  in green, and the distribution as a function of the network prediction  $\log_{10}(1/\sigma_{i\bar{i}} d\hat{\sigma}_{i\bar{i}})$  as a dashed blue line. The neural network predicts very small values of the differential cross-section for a small fraction of events, while their actual differential cross-section can be several orders of magnitude larger.

#### Relative error

The relative error of the network prediction for any event is given by  $(d\sigma_{i\bar{i}} - d\hat{\sigma}_{i\bar{i}})/d\sigma_{i\bar{i}}$ . Figure 11.6 visualizes the distribution of this quantity for one million test events. The distribution is normalized

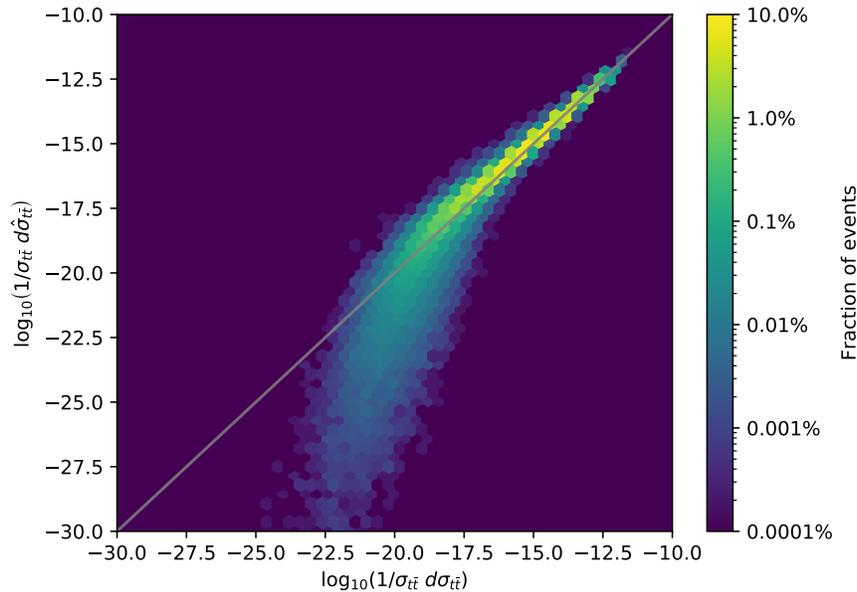


Figure 11.4: Distribution of one million test events as a function of the logarithm of their normalized differential cross-section,  $\log_{10}(1/\sigma_{i\bar{i}} d\sigma_{i\bar{i}})$ , and the corresponding neural network prediction  $\log_{10}(1/\sigma_{i\bar{i}} d\hat{\sigma}_{i\bar{i}})$ . The fraction of events per bin is indicated by the color of each bin, drawn with a logarithmic scale. The diagonal gray line indicates where all events would be located for  $d\hat{\sigma}_{i\bar{i}} = d\sigma_{i\bar{i}}$ . The test events are distributed like the differential cross-section; fewer events exist in phase space regions with small differential cross-sections. The neural network prediction for a very small fraction of events with small differential cross-sections significantly underestimates their differential cross-sections. The error is largest for such events due to the training set distribution and loss function choice for the neural network.

to unit integral.

The largest fraction of events is located in the bin corresponding to  $d\sigma_{i\bar{i}} - d\hat{\sigma}_{i\bar{i}} = 0$ . The distribution is not symmetric around zero due to its definition; the bin at  $(d\sigma_{i\bar{i}} - d\hat{\sigma}_{i\bar{i}})/d\sigma_{i\bar{i}} = 1$  contains all events where the network significantly underestimates the differential cross-section. On average, the predicted differential cross-section is 0.5% below the true value. The fraction of events with  $(d\sigma_{i\bar{i}} - d\hat{\sigma}_{i\bar{i}})/d\sigma_{i\bar{i}} < -1$  is 12%. For these events, the network predicts a differential cross-section at least twice as large as the true value.

The mean relative error is calculated by averaging the absolute value of the relative error across all events. It is 50.5% for the full set of events. The mean relative error depends on the differential cross-section distribution of the events considered; it decreases for events with larger differential cross-sections  $d\sigma_{i\bar{i}}$ . Table 11.1 summarizes this dependence.

The mean relative error reaches 6.7% for events with  $\log_{10}(1/\sigma_{i\bar{i}} d\sigma_{i\bar{i}}) > -13.8$ , which corresponds to 10% of the full dataset. The mean relative error can be compared to the typical ratio between signal and background likelihoods in a MEM calculation. The left part of figure 7.4 shows that

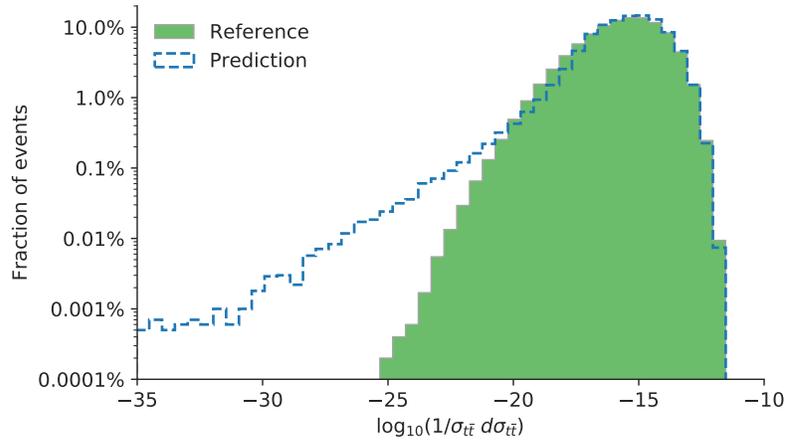


Figure 11.5: Distribution of one million test events. The green histogram shows the distribution of events as a function of the logarithm of their normalized differential cross-section,  $\log_{10}(1/\sigma_{ii} d\sigma_{ii})$ . The dashed blue line shows the distribution as a function of the corresponding network prediction,  $\log_{10}(1/\hat{\sigma}_{ii} d\hat{\sigma}_{ii})$ . The network prediction significantly underestimates the differential cross-section for a small fraction of events with small differential cross-sections. This is due to the choice of training set distribution and loss function of the neural network.

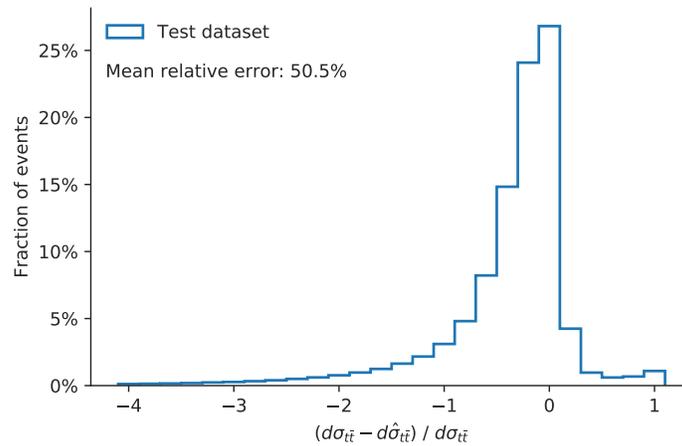


Figure 11.6: Distribution of the relative error  $(d\sigma_{ii} - d\hat{\sigma}_{ii})/d\sigma_{ii}$  for one million test events. The mean relative error is the average absolute value of the relative error across all events.

Table 11.1: Mean relative error obtained when only considering events above a threshold of their normalized differential cross-section. The requirement is listed in the left column, while the central column specifies the fraction of the full test dataset satisfying this requirement. The mean relative error obtained in the dataset with the requirement applied is listed in the right column.

| Requirement   | Dataset fraction | Mean relative error |
|---|------------------|---------------------|
| -   | 100%             | 50.5%               |
| $\log_{10}(1/\sigma_{t\bar{t}} d\sigma_{t\bar{t}}) > -16.6$ | 75.0%            | 26.0%               |
| $\log_{10}(1/\sigma_{t\bar{t}} d\sigma_{t\bar{t}}) > -15.5$ | 50.0%            | 16.3%               |
| $\log_{10}(1/\sigma_{t\bar{t}} d\sigma_{t\bar{t}}) > -14.6$ | 25.0%            | 8.1%                |
| $\log_{10}(1/\sigma_{t\bar{t}} d\sigma_{t\bar{t}}) > -13.8$ | 10.0%            | 6.7%                |

these likelihoods frequently differ by more than an order of magnitude. An error of around 50% in the likelihood calculation may thus still be small enough to obtain good performance of a MEM discriminant.

## 11.4 Discussion

This chapter demonstrates the feasibility of approximating differential cross-sections with machine learning techniques. Section 11.2 shows the use of a foam to describe a four-dimensional probability density. This approach works well for distributions with limited degrees of freedom, but performs significantly worse when describing processes where many final state particles are produced. The neural network shown in section 11.3 approximates the 16-dimensional probability density for the production of  $t\bar{t}$  with subsequent decays to a six-particle final state with a mean relative error of around 50%. The performance is significantly better in phase space regions of high probability density, which contribute most to the evaluation of a likelihood in a MEM implementation. A more accurate modeling in other regions can be obtained when using a different loss function, or a training set with a different distribution.

Another approach is possible to reduce the computation time for MEM calculations. While the implementation in this chapter approximates the differential cross-section, MEM likelihoods can also be approximated directly for every event. A MEM likelihood is a function of a reconstructed set of objects, marginalized over all possible parton level configurations, and weighted by the transfer function. This marginalization smoothens out the effect of resonances in the scattering process, and the resulting distribution can be easier to approximate than the differential cross-section itself. If the likelihoods are approximated directly, no subsequent integration is required anymore, resulting in a fast method to approximate MEM likelihoods.

There are several drawbacks to directly approximating MEM likelihoods. The generation of a training set is computationally expensive. Convergence of the integral over all parton level configurations needs to be guaranteed for every event in the training set to avoid adversely affecting the training of the neural network. The MEM likelihood depends on the transfer function choice, and any change in transfer function requires the generation of a new training set, and a repeated network training.

The training set for the direct approximation of the differential cross-section is significantly faster to generate, and the network is independent of the transfer function. Once trained, it is thus widely useable.

## 12. Conclusions and outlook

The Higgs boson discovery in 2012 was a milestone for the field of particle physics. It marked the beginning of an experimental program measuring the predictions of the Brout-Englert-Higgs mechanism in detail. The coupling of the Higgs boson to fermions, realized in the SM via Yukawa interactions, is one of these predictions. In 2017 and 2018, the ATLAS and CMS collaborations independently achieved the observation of Yukawa interactions involving top quarks, bottom quarks, and tau leptons. These results represent a significant validation of the SM. Measurements of the Yukawa interactions involving first and second generation fermions remain a challenge for the future of particle physics.

This dissertation presented an experimental study of the Higgs boson interaction with top quark pairs, a process directly sensitive to the top quark Yukawa coupling. A search targeted at Higgs boson decays to bottom quarks was performed with  $36.1 \text{ fb}^{-1}$  of data at  $\sqrt{s} = 13 \text{ TeV}$ , delivered by the LHC and recorded by the ATLAS experiment in 2015 and 2016. The MEM is used in this search alongside other multivariate techniques to help distinguish between the  $t\bar{t}H(b\bar{b})$  signal and background processes. The measured signal strength for the  $t\bar{t}H$  process is  $\mu = 0.84^{+0.64}_{-0.61}$ , consistent with the prediction by the SM. This measurement improved the sensitivity compared to the corresponding Run-1 result by the ATLAS collaboration [88]. Limiting factors to the sensitivity were the modeling of the  $t\bar{t} + \geq 1b$  background, and statistical uncertainties due to the finite amount of simulated events. New techniques and progress in the understanding of  $t\bar{t} + \geq 1b$  modeling are required to significantly reduce the associated systematic uncertainties. The statistical uncertainties are straightforward, though computationally expensive, to overcome.

The CMS collaboration measured a signal strength of  $\mu = 1.15^{+0.32}_{-0.29}$  for  $t\bar{t}H$  in the  $H \rightarrow b\bar{b}$  channel when combining data collected in 2016 and 2017, corresponding to a total of  $77.4 \text{ fb}^{-1}$  [158]. This result constitutes evidence for the  $t\bar{t}H(b\bar{b})$  process. Based on studies of the single-lepton channel in the ATLAS  $t\bar{t}H(b\bar{b})$  analysis with the full  $139.0 \text{ fb}^{-1}$  dataset of proton-proton collisions from Run-2 of the LHC, the ATLAS experiment is also expected to provide evidence for this process. The expected sensitivity in the single-lepton channel is  $3.3\sigma$ .

While the Higgs boson branching ratio for the decay to bottom quarks is the largest, rare final states play an increasingly important role in the sensitivity to  $t\bar{t}H$  production. The observation of  $t\bar{t}H$  with the ATLAS detector was obtained by combining the  $t\bar{t}H(b\bar{b})$  search with analyses targeting additional Higgs boson final states. It resulted in an observed significance of  $5.4\sigma$ , with an expected sensitivity of  $5.5\sigma$ . The result establishes Yukawa couplings in the SM. The  $t\bar{t}H$  cross-section in the combination of  $t\bar{t}H$  channels is measured to be  $\sigma_{t\bar{t}H} = 670 \pm 90^{+110}_{-100} \text{ fb}$ , in agreement with the SM prediction of  $\sigma_{t\bar{t}H} = 507^{+35}_{-50} \text{ fb}$ . The sensitivity is dominated by loop-induced Higgs boson decays to

di-photon final states. This topology is still statistically limited with a dataset of  $139.0 \text{ fb}^{-1}$  [159], and will benefit from a larger dataset.

Following the LHC, the High Luminosity Large Hadron Collider (HL-LHC) [160] will begin operation in 2026 and is expected to deliver a proton–proton collision dataset of  $3\,000 \text{ fb}^{-1}$  at  $\sqrt{s} = 14 \text{ TeV}$ . An extrapolation to this dataset shows an expected relative uncertainty of a  $t\bar{t}H(b\bar{b})$  cross-section measurement of 14%, and a 5% cross-section uncertainty in the combination of  $t\bar{t}H$  channels [161]. The top quark Yukawa coupling is expected to be measured with an uncertainty of 3% [161]. Besides the increased precision in inclusive measurements at the HL-LHC, the large dataset will allow differential measurements of  $t\bar{t}H$  with high precision. Such measurements provide a detailed investigation of the SM, and constrain possible contributions from BSM phenomena.

A feasibility study was performed to investigate the performance of machine learning techniques in the approximation of differential cross-sections. The fully differential cross-section for  $t\bar{t}$  production, with subsequent decay into a six-particle final state, can be described by a neural network with a mean relative error of 50%. The error significantly decreases in the most populated regions of phase space, where precision is especially needed. Machine learning techniques are likely going to be used in the future to enhance MEM techniques and help overcome its large computational requirements.

# Bibliography

- [1] ATLAS Collaboration, *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*, Phys. Rev. D 97 (2018) 072016, arXiv: 1712.08895 [hep-ex] (cited on pages v, 59, 66, 71–73, 75, 80, 82, 85, 87–92, 94–96, 120, 124, 126, 181–185).
- [2] ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. D 97 (2018) 072003, arXiv: 1712.08891 [hep-ex] (cited on pages vi, 120, 121).
- [3] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. B 784 (2018) 173, arXiv: 1806.00425 [hep-ex] (cited on pages vi, 20, 120, 122, 123).
- [4] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. 22 (1961) 579 (cited on page 3).
- [5] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. 19 (1967) 1264 (cited on page 3).
- [6] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C680519 (1968) 367 (cited on page 3).
- [7] G. 't Hooft and M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. B 44 (1972) 189 (cited on page 3).
- [8] P. Langacker, *The Standard Model and Beyond; 2nd ed.* High Energy Physics, Cosmology and Gravitation, CRC Press, (2017), ISBN: 978-1420079067 (cited on pages 3, 14, 98, 187).
- [9] M. Thomson, *Modern Particle Physics*, Cambridge University Press, (2013), ISBN: 978-1107034266 (cited on pages 3, 14, 23, 29).
- [10] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D 98 (2018) 030001 (cited on pages 3, 4, 15, 17, 48, 60, 102).
- [11] D. Galbraith and C. Burgard, *Standard Model infographic developed at the CERN webfest*, (2012), URL: <https://cds.cern.ch/record/1473657> (visited on Aug. 28, 2019) (cited on page 4).
- [12] S. Gadatsch, *The Higgs Boson*, CERN-THESIS-2015-113, PhD thesis: University of Amsterdam, (2015) (cited on page 4).
- [13] V. N. Gribov and L. N. Lipatov, *Deep inelastic  $e p$  scattering in perturbation theory*, Sov. J. Nucl. Phys. 15 (1972) 438 (cited on page 6).

- 
- [14] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. B 126 (1977) 298 (cited on page 6).
- [15] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and  $e^+e^-$  Annihilation by Perturbation Theory in Quantum Chromodynamics*. Sov. Phys. JETP 46 (1977) 641 (cited on page 6).
- [16] M. Guzzi et al., *CT10 parton distributions and other developments in the global QCD analysis*, (2011), arXiv: 1101.0561 [hep-ph] (cited on pages 6, 66, 107).
- [17] A. Buckley et al., *LHAPDF6: parton density access in the LHC precision era*, Eur. Phys. J. C 75 (2015) 132, arXiv: 1412.7420 [hep-ph] (cited on pages 6, 107, 158).
- [18] J. C. Collins and D. E. Soper, *The Theorems of Perturbative QCD*, Ann. Rev. Nucl. Part. Sci. 37 (1987) 383 (cited on page 6).
- [19] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. 13 (1964) 321 (cited on page 10).
- [20] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. 12 (1964) 132 (cited on page 10).
- [21] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. 13 (1964) 508 (cited on page 10).
- [22] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. 13 (1964) 585 (cited on page 10).
- [23] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. 145 (1966) 1156 (cited on page 10).
- [24] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. 155 (1967) 1554 (cited on page 10).
- [25] ATLAS Collaboration, *Summary plots from the ATLAS Standard Model physics group*, (2019), URL: <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/> (visited on Aug. 28, 2019) (cited on page 14).
- [26] D. Clowe et al., *A direct empirical proof of the existence of dark matter*, Astrophys. J. 648 (2006) L109, arXiv: astro-ph/0608407 [astro-ph] (cited on page 15).
- [27] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, (2018), arXiv: 1807.06209 [astro-ph.CO] (cited on page 15).
- [28] G. Bertone and D. Hooper, *History of dark matter*, Rev. Mod. Phys. 90 (2018) 045002, arXiv: 1605.04909 [astro-ph.CO] (cited on page 15).
- [29] D. Buttazzo et al., *Investigating the near-criticality of the Higgs boson*, JHEP 12 (2013) 089, arXiv: 1307.3536 [hep-ph] (cited on page 16).

- [30] E. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, J. Exp. Theor. Phys. 120 (2015) 335, arXiv: 1411.1923 [hep-ph] (cited on page 16).
- [31] G. F. Giudice, *Naturally Speaking: The Naturalness Criterion and Physics at the LHC*, (2008), arXiv: 0801.2562 [hep-ph] (cited on page 16).
- [32] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 (2012) 1, arXiv: 1207.7214 [hep-ex] (cited on page 17).
- [33] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 (2012) 30, arXiv: 1207.7235 [hep-ex] (cited on page 17).
- [34] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, (2016), arXiv: 1610.07922 [hep-ph] (cited on pages 18, 20, 64, 78).
- [35] ATLAS Collaboration, *A search for the rare decay of the Standard Model Higgs boson to dimuons in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector*, ATLAS-CONF-2018-026, (2018) (cited on page 20).
- [36] ATLAS Collaboration, *Searches for the  $Z\gamma$  decay mode of the Higgs boson and for new high-mass resonances in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, JHEP 10 (2017) 112, arXiv: 1708.00212 [hep-ex] (cited on page 20).
- [37] ATLAS Collaboration, *Combined measurements of Higgs boson production and decay using up to  $80 \text{ fb}^{-1}$  of proton-proton collision data at  $\sqrt{s} = 13$  TeV collected with the ATLAS experiment*, ATLAS-CONF-2019-005, (2019) (cited on pages 20, 124, 125).
- [38] ATLAS Collaboration, *Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the ATLAS detector*, Phys. Lett. B 786 (2018) 59, arXiv: 1808.08238 [hep-ex] (cited on pages 20, 120, 124, 125).
- [39] CMS Collaboration, *Observation of  $t\bar{t}H$  production*, Phys. Rev. Lett. 120 (2018) 231801, arXiv: 1804.02610 [hep-ex] (cited on pages 20, 120).
- [40] CMS Collaboration, *Observation of Higgs boson decay to bottom quarks*, Phys. Rev. Lett. 121 (2018) 121801, arXiv: 1808.08242 [hep-ex] (cited on pages 20, 120).
- [41] CMS Collaboration, *Observation of the Higgs boson decay to a pair of  $\tau$  leptons with the CMS detector*, Phys. Lett. B 779 (2018) 283, arXiv: 1708.00373 [hep-ex] (cited on pages 20, 120).
- [42] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of  $\tau$ -leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Phys. Rev. D 99 (2019) 072001, arXiv: 1811.08856 [hep-ex] (cited on pages 20, 120).

- [43] G. P. Salam, *Theory vision*, PoS LHCP2018 (2018) 304, arXiv: 1811.11282 [hep-ph] (cited on page 21).
- [44] L. Evans and P. Bryant, *LHC Machine*, JINST 3 (2008) S08001 (cited on page 22).
- [45] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 (2008) S08003 (cited on pages 22, 25–27, 29, 30).
- [46] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST 3 (2008) S08004 (cited on page 22).
- [47] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST 3 (2008) S08005 (cited on page 22).
- [48] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST 3 (2008) S08002 (cited on page 22).
- [49] E. Mobs, *The CERN accelerator complex*, (2016),  
URL: <https://cds.cern.ch/record/2197559> (visited on Aug. 28, 2019) (cited on page 23).
- [50] ATLAS Collaboration,  
*Luminosity determination in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector at the LHC*,  
Eur. Phys. J. C 76 (2016) 653, arXiv: 1608.03953 [hep-ex] (cited on pages 24, 62, 77, 132).
- [51] ATLAS Collaboration, *Luminosity Public Results Run-2*, (2019), URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>  
(visited on Aug. 28, 2019) (cited on pages 24, 35).
- [52] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, ATLAS-TDR-19, (2010) (cited on page 27).
- [53] ATLAS Collaboration, *Performance of the ATLAS muon trigger in pp collisions at  $\sqrt{s} = 8$  TeV*,  
Eur. Phys. J. C 75 (2015) 120, arXiv: 1408.3179 [hep-ex] (cited on pages 31, 144).
- [54] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2010*,  
Eur. Phys. J. C 72 (2012) 1849, arXiv: 1110.1530 [hep-ex] (cited on page 32).
- [55] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*,  
Eur. Phys. J. C 77 (2017) 317, arXiv: 1611.09661 [hep-ex] (cited on page 32).
- [56] GEANT4 Collaboration, *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. A 506 (2003) 250 (cited on page 34).
- [57] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C 70 (2010) 823,  
arXiv: 1005.4568 [physics.ins-det] (cited on pages 34, 64).
- [58] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, (2010) (cited on pages 34, 127).
- [59] J. Pequenao and P. Schaffner,  
*How ATLAS detects particles: diagram of particle paths in the detectors*, (2013),  
URL: <https://cds.cern.ch/record/1505342> (visited on Aug. 28, 2019) (cited on page 36).

- 
- [60] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, Eur. Phys. J. C 77 (2017) 673, arXiv: 1704.07983 [hep-ex] (cited on page 37).
- [61] T. Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*, J. Phys. Conf. Ser. 119 (2008) 032014 (cited on page 37).
- [62] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton-proton collisions at the LHC*, Eur. Phys. J. C 77 (2017) 332, arXiv: 1611.10235 [physics.ins-det] (cited on page 38).
- [63] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C 77 (2017) 490, arXiv: 1603.02934 [hep-ex] (cited on page 38).
- [64] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at  $\sqrt{s}=13$  TeV*, Eur. Phys. J. C 76 (2016) 292, arXiv: 1603.05598 [hep-ex] (cited on page 39).
- [65] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at  $\sqrt{s}=13$  TeV*, Eur. Phys. J. C 79 (2019) 639, arXiv: 1902.04655 [physics.ins-det] (cited on page 41).
- [66] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, Eur. Phys. J. C 79 (2019) 205, arXiv: 1810.05087 [hep-ex] (cited on page 41).
- [67] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015-2016 LHC proton-proton collision data*, JINST 14 (2019) P03017, arXiv: 1812.03848 [hep-ex] (cited on page 43).
- [68] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at  $\sqrt{s}=13$  TeV*, ATLAS-CONF-2017-029, (2017) (cited on page 43).
- [69] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at  $\sqrt{s}=13$  TeV with the ATLAS detector*, Phys. Rev. D 96 (2017) 072002, arXiv: 1703.09665 [hep-ex] (cited on page 43).
- [70] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, JHEP 04 (2008) 063, arXiv: 0802.1189 [hep-ph] (cited on page 43).
- [71] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur. Phys. J. C 72 (2012) 1896, arXiv: 1111.6097 [hep-ph] (cited on page 43).
- [72] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, (2015) (cited on page 44).

- [73] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, Eur. Phys. J. C 76 (2016) 581, arXiv: 1510.03823 [hep-ex] (cited on page 44).
- [74] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST 11 (2016) P04008, arXiv: 1512.01094 [hep-ex] (cited on page 45).
- [75] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, (2016) (cited on pages 45, 46).
- [76] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, (2017) (cited on page 45).
- [77] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, CERN-OPEN-2008-020, (2009), arXiv: 0901.0512 [hep-ex] (cited on page 45).
- [78] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at  $\sqrt{s} = 13$  TeV*, Eur. Phys. J. C 78 (2018) 903, arXiv: 1802.08168 [hep-ex] (cited on page 46).
- [79] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, ATL-PHYS-PUB-2011-11, CMS-NOTE-2011-005, (2011) (cited on page 48).
- [80] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY, (2009), ISBN: 978-0387848570 (cited on pages 48, 55).
- [81] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, (2016), URL: <https://www.deeplearningbook.org> (visited on Aug. 28, 2019) (cited on pages 48, 55).
- [82] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Comput. Phys. Commun. 10 (1975) 343 (cited on pages 51, 83).
- [83] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A 389 (1997) 81 (cited on pages 51, 83).
- [84] J. Neyman, E. S. Pearson, and K. Pearson, *IX. On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 231 (1933) 289 (cited on page 51).
- [85] A. L. Read, *Presentation of search results: The  $CL(s)$  technique*, J. Phys. G 28 (2002) 2693 (cited on page 52).

- [86] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Annals Math. Statist.* 9 (1938) 60 (cited on page 53).
- [87] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* 71 (2011) 1554, arXiv: 1007.1727 [physics.data-an], [Erratum: *Eur. Phys. J. C* 73 (2013) 2501] (cited on pages 54, 55).
- [88] ATLAS Collaboration, *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*, *Eur. Phys. J. C* 75 (2015) 349, arXiv: 1503.05066 [hep-ex] (cited on pages 59, 70, 97, 122, 124, 167).
- [89] ATLAS Collaboration, CMS Collaboration, *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*, *JHEP* 08 (2016) 045, arXiv: 1606.02266 [hep-ex] (cited on page 59).
- [90] ATLAS Collaboration, *Search for the Standard Model Higgs boson decaying into  $b\bar{b}$  produced in association with top quarks decaying hadronically in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, *JHEP* 05 (2016) 160, arXiv: 1604.03812 [hep-ex] (cited on pages 61, 122).
- [91] B. Nachman, P. Nef, A. Schwartzman, M. Swiatlowski, and C. Wanotayaroj, *Jets from Jets: Re-clustering as a tool for large radius jet reconstruction and grooming at the LHC*, *JHEP* 02 (2015) 075, arXiv: 1407.2922 [hep-ph] (cited on page 63).
- [92] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* 178 (2008) 852, arXiv: 0710.3820 [hep-ph] (cited on page 64).
- [93] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* 462 (2001) 152 (cited on page 64).
- [94] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* 02 (2009) 007, arXiv: 0811.4622 [hep-ph] (cited on pages 64, 66, 128, 146).
- [95] 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*, *JHEP* 07 (2014) 079, arXiv: 1405.0301 [hep-ph] (cited on pages 64, 106, 116, 128, 158).
- [96] NNPDF Collaboration, *Parton distributions for the LHC Run II*, *JHEP* 04 (2015) 040, arXiv: 1410.8849 [hep-ph] (cited on pages 64, 127).
- [97] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* 191 (2015) 159, arXiv: 1410.3012 [hep-ph] (cited on pages 64, 127, 146).
- [98] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, ATL-PHYS-PUB-2014-021, (2014) (cited on pages 64, 127).

- [99] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension*, Comput. Phys. Commun. 108 (1998) 56, arXiv: hep-ph/9704448 [hep-ph] (cited on page 64).
- [100] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP 11 (2004) 040, arXiv: hep-ph/0409146 [hep-ph] (cited on pages 64, 108, 116, 127, 146).
- [101] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP 11 (2007) 070, arXiv: 0709.2092 [hep-ph] (cited on pages 64, 108, 116, 127, 146).
- [102] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP 06 (2010) 043, arXiv: 1002.2581 [hep-ph] (cited on pages 64, 108, 116, 127, 146).
- [103] J. M. Campbell, R. K. Ellis, P. Nason, and E. Re, *Top-Pair Production and Decay at NLO Matched with Parton Showers*, JHEP 04 (2015) 114, arXiv: 1412.1828 [hep-ph] (cited on pages 64, 108, 116, 127, 146).
- [104] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, (2016) (cited on page 64).
- [105] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. 185 (2014) 2930, arXiv: 1112.5675 [hep-ph] (cited on pages 65, 78).
- [106] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B 710 (2012) 612, arXiv: 1111.5869 [hep-ph] (cited on page 65).
- [107] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP 12 (2012) 054, arXiv: 1207.0236 [hep-ph] (cited on page 65).
- [108] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP 01 (2013) 080, arXiv: 1210.6832 [hep-ph] (cited on page 65).
- [109] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through  $O(\frac{4}{5})$* , Phys. Rev. Lett. 110 (2013) 252004, arXiv: 1303.6254 [hep-ph] (cited on page 65).
- [110] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. 108 (2012) 111601, arXiv: 1111.5206 [hep-ph] (cited on page 66).
- [111] F. Cascioli, P. Maierhöfer, N. Moretti, S. Pozzorini, and F. Siegert, *NLO matching for  $t\bar{t}b\bar{b}$  production with massive  $b$ -quarks*, Phys. Lett. B 734 (2014) 210, arXiv: 1309.5912 [hep-ph] (cited on page 66).

- [112] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. D 89 (2014) 033009, arXiv: 1302.6246 [hep-ph] (cited on page 66).
- [113] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP 03 (2008) 038, arXiv: 0709.1027 [hep-ph] (cited on pages 67, 128, 146).
- [114] T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, JHEP 12 (2008) 039, arXiv: 0808.3674 [hep-ph] (cited on pages 67, 128, 146).
- [115] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, JHEP 04 (2013) 027, arXiv: 1207.5030 [hep-ph] (cited on pages 67, 128, 146).
- [116] ATLAS Collaboration, *Multi-Boson Simulation for 13 TeV ATLAS Analyses*, ATL-PHYS-PUB-2016-002, (2016) (cited on pages 67, 81).
- [117] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP 05 (2006) 026, arXiv: hep-ph/0603175 [hep-ph] (cited on pages 67, 108, 116).
- [118] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. D 82 (2010) 074018, arXiv: 1005.3457 [hep-ph] (cited on page 67).
- [119] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, *Single-top hadroproduction in association with a W boson*, JHEP 07 (2008) 029, arXiv: 0805.3067 [hep-ph] (cited on pages 67, 81, 134).
- [120] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP 07 (2002) 012, arXiv: hep-ph/0201195 [hep-ph] (cited on pages 67, 158).
- [121] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C 58 (2008) 639, arXiv: 0803.0883 [hep-ph] (cited on pages 67, 116).
- [122] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, ATLAS-CONF-2014-058, (2014) (cited on page 68).
- [123] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide*, CERN-OPEN-2007-007, (2007), arXiv: physics/0703039 [physics.data-an] (cited on pages 73, 110).
- [124] ATLAS Collaboration, *Search for flavour-changing neutral current top quark decays  $t \rightarrow Hq$  in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, JHEP 12 (2015) 061, arXiv: 1509.06047 [hep-ex] (cited on page 74).
- [125] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C 76 (2016) 196, arXiv: 1512.01178 [hep-ph] (cited on pages 79, 133).

- 
- [126] ATLAS Collaboration, *Studies of  $tt+cc$  production with MadGraph5\_aMC@NLO and Herwig++ for the ATLAS experiment*, ATL-PHYS-PUB-2016-011, (2016) (cited on page 79).
- [127] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. C 63 (2009) 189, arXiv: 0901.0002 [hep-ph] (cited on page 79).
- [128] J. M. Campbell and R. K. Ellis,  *$t\bar{t}W^\pm$  production and decay at NLO*, JHEP 07 (2012) 052, arXiv: 1204.5678 [hep-ph] (cited on page 81).
- [129] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016, (2012) (cited on page 83).
- [130] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf C0303241 (2003) MOLT007, arXiv: physics/0306116 [physics] (cited on page 83).
- [131] L. Moneta et al., *The RooStats Project*, PoS ACAT2010 (2010) 057, arXiv: 1009.1003 [physics.data-an] (cited on page 83).
- [132] T. P. Calvet, *Search for the production of a Higgs boson in association with top quarks and decaying into a  $b$ -quark pair and  $b$ -jet identification with the ATLAS experiment at LHC*, CERN-THESIS-2017-258, PhD thesis: Aix-Marseille University, (2017) (cited on page 93).
- [133] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *A Guide to Constraining Effective Field Theories with Machine Learning*, Phys. Rev. D 98 (2018) 052004, arXiv: 1805.00020 [hep-ph] (cited on page 97).
- [134] D0 Collaboration, *A precision measurement of the mass of the top quark*, Nature 429 (2004) 638, arXiv: hep-ex/0406031 [hep-ex] (cited on page 97).
- [135] K. Kondo, *Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models*, J. Phys. Soc. Jap. 57 (1988) 4126 (cited on page 97).
- [136] F. Fiedler, A. Grohsjean, P. Haefner, and P. Schieferdecker, *The Matrix Element Method and its Application in Measurements of the Top Quark Mass*, Nucl. Instrum. Meth. A 624 (2010) 203, arXiv: 1003.1316 [hep-ex] (cited on page 97).
- [137] P. Artoisenet, V. Lemaitre, F. Maltoni, and O. Mattelaer, *Automation of the matrix element reweighting method*, JHEP 12 (2010) 068, arXiv: 1007.3300 [hep-ph] (cited on pages 97, 156).
- [138] S. Brochet et al., *MoMEMta, a modular toolkit for the Matrix Element Method at the LHC*, Eur. Phys. J. C 79 (2019) 126, arXiv: 1805.08555 [hep-ph] (cited on page 97).
- [139] D. Schouten, A. DeAbreu, and B. Stelzer, *Accelerated Matrix Element Method with Parallel Computing*, Comput. Phys. Commun. 192 (2015) 54, arXiv: 1407.7595 [physics.comp-ph] (cited on page 103).

- 
- [140] J. Nickolls, I. Buck, M. Garland, and K. Skadron, *Scalable Parallel Programming with CUDA*, Queue 6 (2008) 40, ISSN: 1542-7730 (cited on page 103).
- [141] J. E. Stone, D. Gohara, and G. Shi, *OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems*, IEEE Des. Test 12 (2010) 66, ISSN: 0740-7475 (cited on page 103).
- [142] G. P. Lepage, *A New Algorithm for Adaptive Multidimensional Integration*, Journal of Computational Physics 27 (1978) 192 (cited on page 103).
- [143] T. Martini and P. Uwer, *Extending the Matrix Element Method beyond the Born approximation: Calculating event weights at next-to-leading order accuracy*, JHEP 09 (2015) 083, arXiv: 1506.08798 [hep-ph] (cited on page 106).
- [144] M. Oreglia, *A Study of the Reactions  $\psi' \rightarrow \gamma\gamma\psi$* , PhD thesis: Stanford University, (1980) (cited on page 108).
- [145] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, PhD thesis: Cracow, INP, (1986) (cited on page 108).
- [146] ATLAS Collaboration, *Measurement of the Higgs boson coupling properties in the  $H \rightarrow ZZ^* \rightarrow 4\ell$  decay channel at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, JHEP 03 (2018) 095, arXiv: 1712.02304 [hep-ex] (cited on page 120).
- [147] ATLAS Collaboration, *Measurements of Higgs boson properties in the diphoton decay channel with  $36 \text{ fb}^{-1}$  of  $pp$  collision data at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Phys. Rev. D 98 (2018) 052005, arXiv: 1802.04146 [hep-ex] (cited on page 120).
- [148] ATLAS Collaboration, *Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector*, Phys. Lett. B 749 (2015) 519, arXiv: 1506.05988 [hep-ex] (cited on page 122).
- [149] ATLAS Collaboration, *Search for  $H \rightarrow \gamma\gamma$  produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector*, Phys. Lett. B 740 (2015) 222, arXiv: 1409.3122 [hep-ex] (cited on page 122).
- [150] ATLAS Collaboration, *Search for the  $b\bar{b}$  decay of the Standard Model Higgs boson in associated  $(W/Z)H$  production with the ATLAS detector*, JHEP 01 (2015) 069, arXiv: 1409.6212 [hep-ex] (cited on page 124).
- [151] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced by vector-boson fusion and decaying to bottom quarks in  $\sqrt{s} = 8$  TeV  $pp$  collisions with the ATLAS detector*, JHEP 11 (2016) 112, arXiv: 1606.02181 [hep-ex] (cited on page 124).
- [152] ATLAS Collaboration, *Search for Higgs bosons produced via vector-boson fusion and decaying into bottom quark pairs in  $\sqrt{s} = 13$  TeV  $pp$  collisions with the ATLAS detector*, Phys. Rev. D 98 (2018) 052003, arXiv: 1807.08639 [hep-ex] (cited on page 124).

- 
- [153] H. B. Hartanto, B. Jager, L. Reina, and D. Wackerroth, *Higgs boson production in association with top quarks in the POWHEG BOX*, Phys. Rev. D 91 (2015) 094003, arXiv: 1501.04498 [hep-ph] (cited on page 127).
- [154] S. Jadach, *Foam: A General purpose cellular Monte Carlo event generator*, Comput. Phys. Commun. 152 (2003) 55, arXiv: physics/0203033 [physics] (cited on page 157).
- [155] F. Chollet et al., *Keras*, (2015), URL: <https://keras.io> (visited on Aug. 28, 2019) (cited on page 161).
- [156] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, (2015), URL: <https://www.tensorflow.org> (visited on Aug. 28, 2019) (cited on page 161).
- [157] M. D. Zeiler, *ADADELTA: An Adaptive Learning Rate Method*, (2012), arXiv: 1212.5701 [cs.LG] (cited on page 162).
- [158] CMS Collaboration, *Measurement of  $t\bar{t}H$  production in the  $H \rightarrow b\bar{b}$  decay channel in 41.5 fb<sup>-1</sup> of proton-proton collision data at  $\sqrt{s} = 13$  TeV*, CMS-PAS-HIG-18-030, (2019) (cited on page 167).
- [159] ATLAS Collaboration, *Measurement of Higgs boson production in association with a  $t\bar{t}$  pair in the diphoton decay channel using 139 fb<sup>-1</sup> of LHC data collected at  $\sqrt{s} = 13$  TeV by the ATLAS experiment*, ATLAS-CONF-2019-004, (2019) (cited on page 168).
- [160] G. Apollinari et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs CERN-2017-007-M, CERN, (2017) (cited on page 168).
- [161] ATLAS Collaboration, *Projections for measurements of Higgs boson cross sections, branching ratios, coupling parameters and mass with the ATLAS detector at the HL-LHC*, ATL-PHYS-PUB-2018-054, (2018) (cited on page 168).

# Appendices

## A Additional material related to the $t\bar{t}H(b\bar{b})$ analysis with $36.1 \text{ fb}^{-1}$

This appendix section contains additional figures related to the  $t\bar{t}H(b\bar{b})$  analysis presented in chapter 6.

### A.1 Categorization for the dilepton channel

The definition of the regions in the dilepton channel follows the method used in the single-lepton channel described in section 6.4. Figure A.1 shows the categorization employed for events containing exactly three jets, while figure A.2 shows the case for events with four or more jets. The background composition of the dilepton regions is shown in figure A.3. Figure A.4 shows the contribution of the  $t\bar{t}H$  process to the dilepton regions.

### A.2 Signal region modeling for the dilepton channel

Figure A.5 shows all regions in the dilepton channel that enter the profile likelihood fit with more than one bin. They are shown both with the pre- and post-fit model. Section 6.7.2 describes further details about how these distributions are obtained.

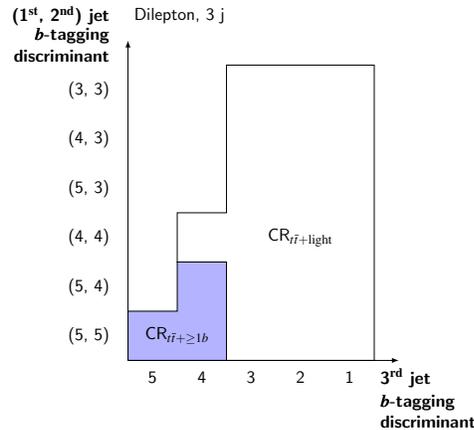


Figure A.1: Definition of analysis regions with exactly three jets in the dilepton channel. The vertical axis shows the  $b$ -tagging requirements for the first two jets in each event, while the horizontal axis shows the requirement for the third jet. Jets are ordered by decreasing tightness of the operating point they satisfy [1].

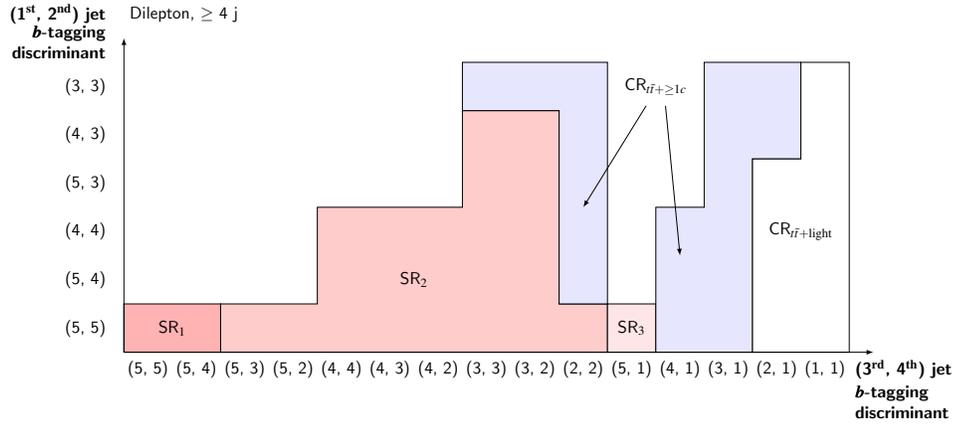


Figure A.2: Definition of analysis regions with four or more jets in the dilepton channel. The vertical axis shows the *b*-tagging requirements for the first two jets in each event, while the horizontal axis shows the requirement for the third and fourth jet. Jets are ordered by decreasing tightness of the operating point they satisfy [1].

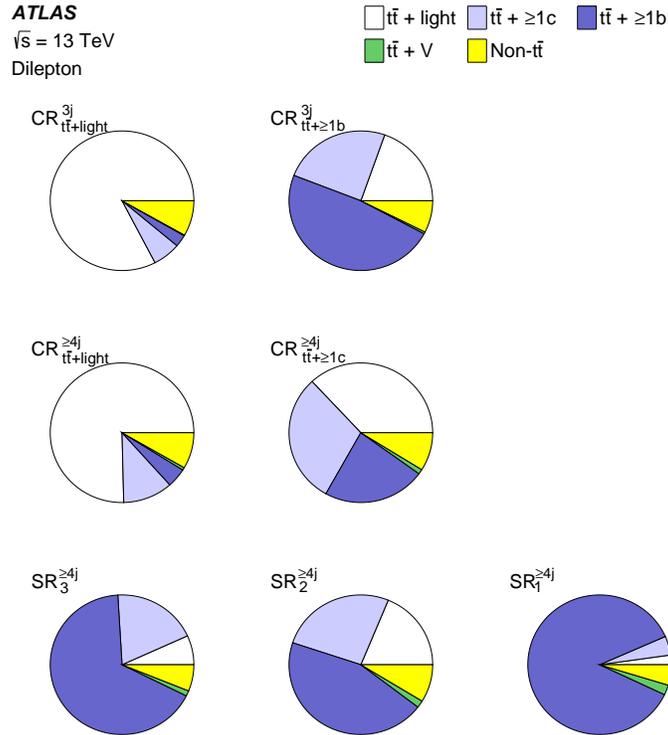


Figure A.3: Composition of background processes in the dilepton regions. Each pie chart shows the relative contributions per process and region, with the processes defined in section 6.3 [1].

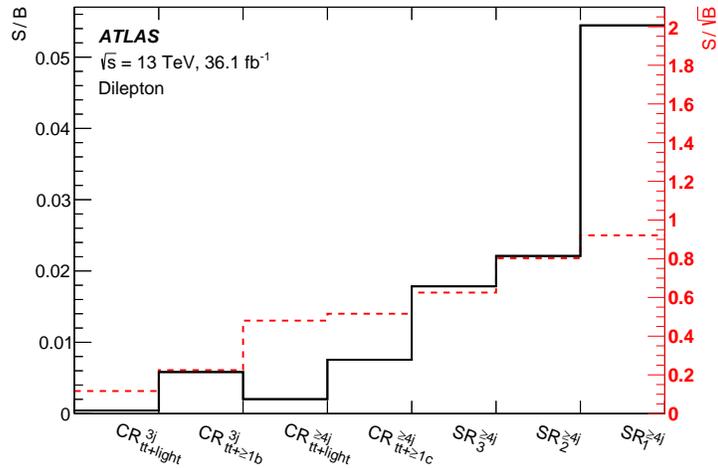


Figure A.4: Signal contributions per analysis region in the single-lepton channel. The solid black line, corresponding to the left vertical axis, shows  $S/B$ . The dashed red line, corresponding to the right vertical axis, shows  $S/\sqrt{B}$ .  $S$  is the number of  $t\bar{t}H$  events per region, and  $B$  the number of expected background events [1].

### A.3 Correlation between nuisance parameters

Figure A.6 shows the correlations between the most highly ranked nuisance parameters, as well as the signal strength, as determined by the nominal fit to data (described in section 6.7.2). All values are given in %.

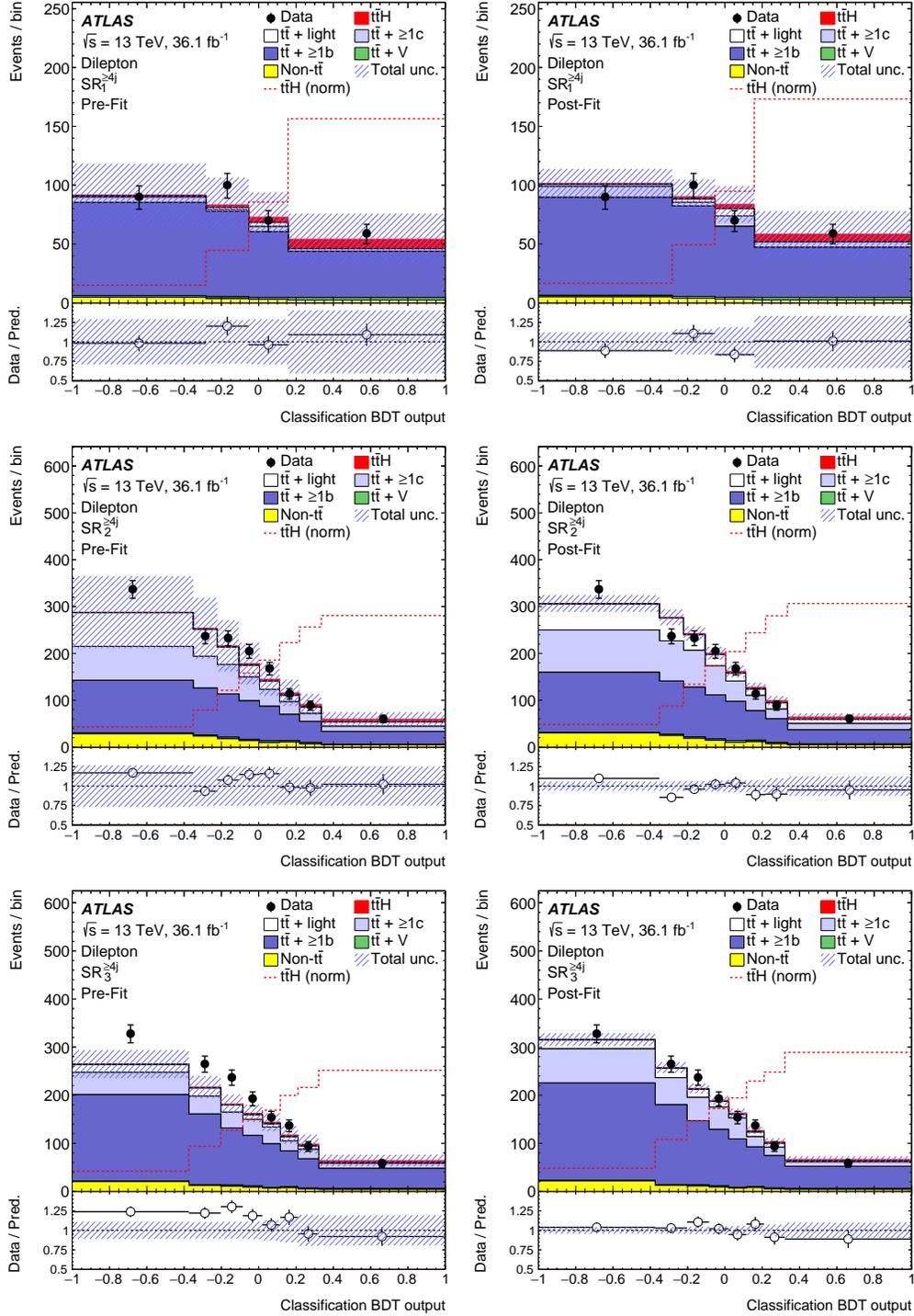


Figure A.5: Comparison between data and the model for the signal regions  $SR_1^{\geq 4j}$  (top),  $SR_2^{\geq 4j}$  (middle) and  $SR_3^{\geq 4j}$  (bottom), with pre-fit on the left and post-fit on the right. The uncertainty bands include all sources of systematic uncertainty described in section 6.6. No uncertainty related to  $k(t\bar{t} + \geq 1b)$  and  $k(t\bar{t} + \geq 1c)$  is included pre-fit. The  $t\bar{t}H$  signal shown in red in the stacked histogram is normalized to the SM prediction pre-fit, and the best-fit signal strength value reported in equation (6.1) post-fit. The  $t\bar{t}H$  distribution normalized to the total background is overlaid as a dashed red line [1].

ATLAS

$\sqrt{s} = 13 \text{ TeV}, 36.1 \text{ fb}^{-1}$

|   |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $\mu_{\text{th}}$   | 100.0 | -14.8 | 0.8   | -66.0 | -32.8 | 26.4  | 19.0  | -13.0 | -15.2 | -11.4 | -14.9 | -8.0  | 7.6   | -9.8  | 4.9   | -8.3  | 5.1   | 6.9   | 3.0   |
| $k(\text{tt} \geq 1\text{b})$   | -14.8 | 100.0 | -38.8 | 24.8  | -22.0 | 37.2  | 7.5   | 3.1   | 5.7   | -17.0 | 8.8   | -9.3  | 15.6  | 9.8   | 27.8  | 7.0   | 6.4   | 38.9  | 9.8   |
| $k(\text{tt} \geq 1\text{c})$   | 0.8   | -38.8 | 100.0 | -2.8  | -6.7  | -19.2 | 2.5   | 33.8  | 4.0   | 16.8  | -17.5 | 16.7  | -13.1 | -22.7 | 18.9  | -1.3  | 3.9   | -10.6 | 9.9   |
| $\text{tt} \geq 1\text{b}$ : SHERPA5F vs. nominal                     | -66.0 | 24.8  | -2.8  | 100.0 | 16.4  | -2.9  | 4.1   | 8.4   | 18.1  | 13.6  | 12.5  | 11.0  | -17.6 | -9.0  | -5.6  | 5.4   | -1.0  | -10.7 | -3.7  |
| $\text{tt} \geq 1\text{b}$ : SHERPA4F vs. nominal                     | -32.8 | -22.0 | -6.7  | 16.4  | 100.0 | -43.2 | -12.6 | -4.0  | 8.5   | -4.9  | 2.4   | -3.2  | -2.3  | -0.3  | -8.8  | 4.2   | 1.6   | -5.7  | -0.7  |
| $\text{tt} \geq 1\text{b}$ : PS & hadronisation                       | 26.4  | 37.2  | -19.2 | -2.9  | -43.2 | 100.0 | 29.9  | -15.2 | -16.3 | 24.2  | -16.3 | 5.1   | 16.9  | -9.3  | -4.9  | -7.2  | -0.4  | 14.2  | 6.8   |
| $\text{tt} \geq 1\text{b}$ : ISR / FSR                                | 19.0  | 7.5   | 2.5   | 4.1   | -12.6 | 29.9  | 100.0 | 4.7   | -19.6 | -6.8  | -13.1 | 8.6   | -15.7 | -16.8 | 6.4   | 2.1   | -5.6  | -24.6 | -7.0  |
| b-tagging: mis-tag (light), NP 0                                      | -13.0 | 3.1   | 33.8  | 8.4   | -4.0  | -15.2 | 4.7   | 100.0 | -5.6  | 3.1   | 23.4  | 4.3   | 0.3   | -14.5 | 4.2   | 1.6   | -2.4  | -9.8  | -5.5  |
| Jet energy resolution: NP 1   | -15.2 | 5.7   | 4.0   | 18.1  | 8.5   | -16.3 | -19.6 | -5.6  | 100.0 | -4.5  | 15.3  | 0.6   | -16.4 | 7.3   | 0.1   | -8.8  | -1.0  | 8.7   | -3.3  |
| $\text{tt} \geq 1\text{b}$ : $\text{tt} \geq 3\text{b}$ normalisation | -11.4 | -17.0 | 16.8  | 13.6  | -4.9  | 24.2  | -6.8  | 3.1   | -4.5  | 100.0 | -5.4  | -25.8 | 1.5   | -7.0  | -3.6  | -4.6  | 4.7   | -2.7  | 8.2   |
| $\text{tt} \geq 1\text{c}$ : SHERPA5F vs. nominal                     | -14.9 | 8.8   | -17.5 | 12.5  | 2.4   | -16.3 | -13.1 | 23.4  | 15.3  | -5.4  | 100.0 | 3.1   | -25.2 | 19.7  | 13.9  | -4.2  | -1.5  | -3.8  | -6.5  |
| $\text{tt} \geq 1\text{b}$ : shower recoil scheme                     | -8.0  | -9.3  | 16.7  | 11.0  | -3.2  | 5.1   | 8.6   | 4.3   | 0.6   | -25.8 | 3.1   | 100.0 | -8.0  | 1.7   | -4.6  | -1.4  | 4.9   | -12.7 | 8.8   |
| $\text{tt} \geq 1\text{c}$ : ISR / FSR                                | 7.6   | 15.6  | -13.1 | -17.6 | -2.3  | 16.9  | -15.7 | 0.3   | -16.4 | 1.5   | -25.2 | -8.0  | 100.0 | -11.3 | 6.9   | 0.5   | -1.3  | 25.4  | 1.5   |
| Jet energy resolution: NP 0   | -9.8  | 9.8   | -22.7 | -9.0  | -0.3  | -9.3  | -16.8 | -14.5 | 7.3   | -7.0  | 19.7  | 1.7   | -11.3 | 100.0 | 24.8  | -8.4  | -3.1  | -10.8 | -9.3  |
| $\text{tt}$ +light: PS & hadronisation                                | 4.9   | 27.8  | 18.9  | -5.6  | -8.8  | -4.9  | 6.4   | 4.2   | 0.1   | -3.6  | 13.9  | -4.6  | 6.9   | 24.8  | 100.0 | -4.6  | 6.6   | 7.5   | 5.3   |
| Wt: diagram subtr. vs. nominal  | -8.3  | 7.0   | -1.3  | 5.4   | 4.2   | -7.2  | 2.1   | 1.6   | -8.8  | -4.6  | -4.2  | -1.4  | 0.5   | -8.4  | -4.6  | 100.0 | -0.1  | 8.0   | -2.1  |
| b-tagging: efficiency, NP 1   | 5.1   | 6.4   | 3.9   | -1.0  | 1.6   | -0.4  | -5.6  | -2.4  | -1.0  | 4.7   | -1.5  | 4.9   | -1.3  | -3.1  | 6.6   | -0.1  | 100.0 | 2.4   | -7.5  |
| b-tagging: mis-tag (c), NP 0  | 6.9   | 38.9  | -10.6 | -10.7 | -5.7  | 14.2  | -24.6 | -9.8  | 8.7   | -2.7  | -3.8  | -12.7 | 25.4  | -10.8 | 7.5   | 8.0   | 2.4   | 100.0 | 5.4   |
| b-tagging: efficiency, NP 0   | 3.0   | 9.8   | 9.9   | -3.7  | -0.7  | 6.8   | -7.0  | -5.5  | -3.3  | 8.2   | -6.5  | 8.8   | 1.5   | -9.3  | 5.3   | -2.1  | -7.5  | 5.4   | 100.0 |

Figure A.6: Correlations between the most highly ranked nuisance parameters and the signal strength, determined by the nominal fit to data described in section 6.7.2. All values are in % [1].

## B Additional material related to the $t\bar{t}H(b\bar{b})$ analysis with $139.0 \text{ fb}^{-1}$

This appendix section contains additional material related to the  $t\bar{t}H(b\bar{b})$  analysis presented in chapter 9.

### B.1 Correlation between nuisance parameters

The correlations between fit parameters are visualized in figure B.1. All values are given in %, and only parameters that have a correlation of at least 25% with at least one other parameter are included. The correlations are determined in a fit to an Asimov dataset, as described in section 9.6.1.

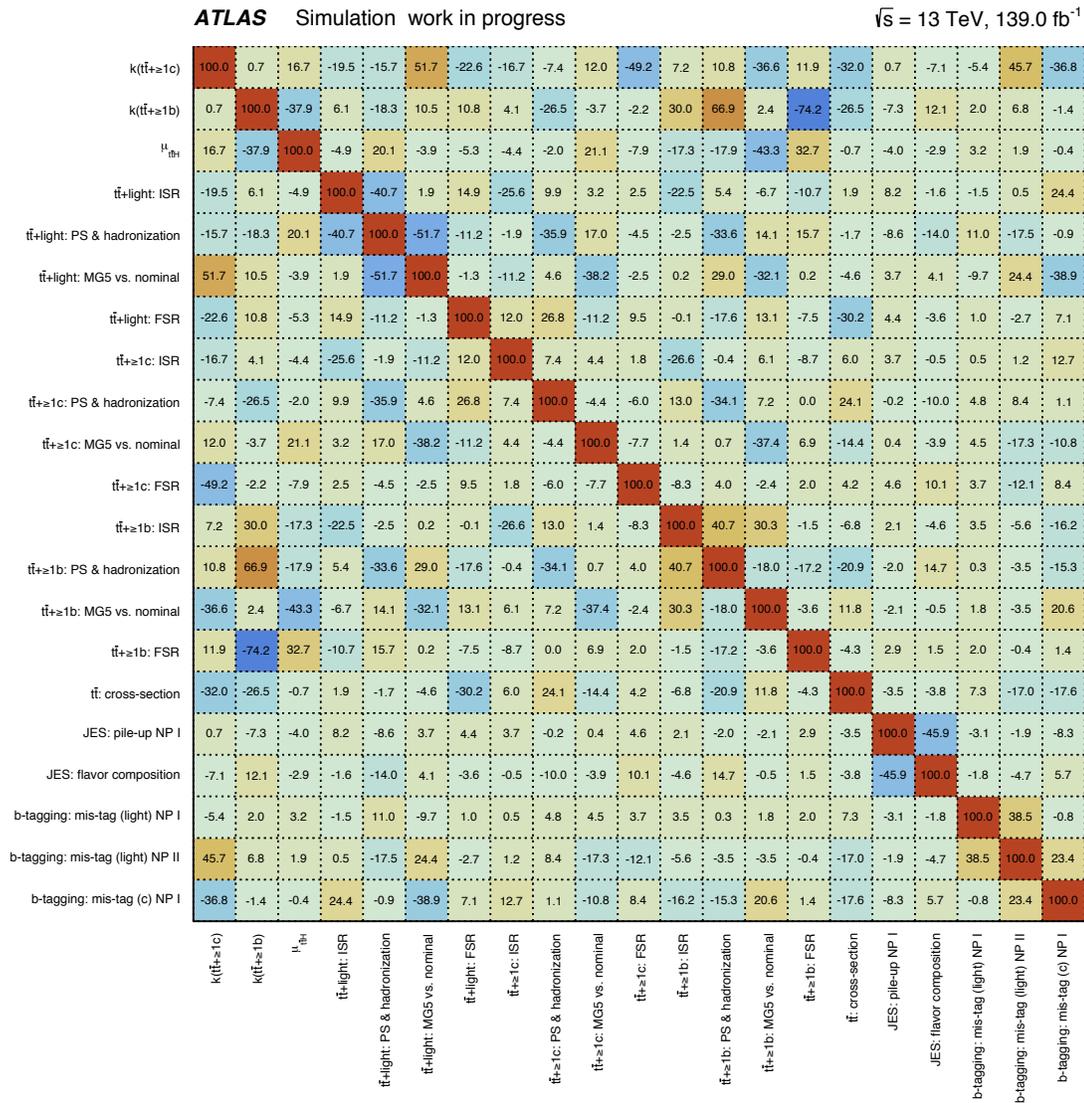


Figure B.1: Correlations between nuisance parameters and signal strength, determined by a fit to an Asimov dataset. Parameters are only included if they have a correlation of at least 25% with at least one other parameter. All values are in %.

## C Additional material related to differential cross-section approximation

This appendix section contains additional material related to chapter 11.

### C.1 Differential cross-section for a $2 \rightarrow 2$ process

This section demonstrates the partial integration to obtain an expression for the differential cross-section with the four-momentum conservation applied. The head-on scattering of two massless particles with four-vectors  $p_a, p_b$  into two final states with four-vectors  $p_1, p_2$  is considered. They are parameterized as

$$p_a = E_B \begin{pmatrix} x_1 \\ 0 \\ 0 \\ x_1 \end{pmatrix}, \quad p_b = E_B \begin{pmatrix} x_2 \\ 0 \\ 0 \\ -x_2 \end{pmatrix}, \quad p_1 = \begin{pmatrix} E_1 \\ p_{x_1} \\ p_{y_1} \\ p_{z_1} \end{pmatrix}, \quad p_2 = \begin{pmatrix} E_2 \\ p_{x_2} \\ p_{y_2} \\ p_{z_2} \end{pmatrix}, \quad (1)$$

with a beam energy  $E_B$  and momentum fractions  $x_1$  and  $x_2$  carried by the colliding particles. The total cross-section can be written as [8]

$$\sigma_{2 \rightarrow 2} = \sum_{j,k} \int \frac{1}{F} f_j(x_1) f_k(x_2) (2\pi)^4 \delta^4(p_a + p_b - p_1 - p_2) |\mathcal{M}_{2 \rightarrow 2}|^2 dx_1 dx_2 \frac{d^3 \vec{p}_1}{(2\pi)^3 2E_1} \frac{d^3 \vec{p}_2}{(2\pi)^3 2E_2}, \quad (2)$$

with PDF contributions  $f_j(x_1) f_k(x_2)$ , and a squared matrix element  $|\mathcal{M}_{2 \rightarrow 2}|^2$  that depends on the process of interest. The flux factor in this case is

$$F = 4\sqrt{(p_a \cdot p_b)^2 - m_a^2 m_b^2} = 8x_1 x_2 E_B^2. \quad (3)$$

Let

$$u \equiv (x_1 + x_2) E_B, \quad v \equiv (x_1 - x_2) E_B, \quad (4)$$

implying that

$$x_1 = \frac{u+v}{2E_B}, \quad x_2 = \frac{u-v}{2E_B}. \quad (5)$$

The Jacobian for a change of integration variables  $x_1, x_2 \rightarrow u, v$  is

$$J = \frac{\partial(x_1, x_2)}{\partial(u, v)} = \begin{vmatrix} \frac{1}{2E_B} & \frac{1}{2E_B} \\ \frac{1}{2E_B} & -\frac{1}{2E_B} \end{vmatrix} = \frac{1}{2E_B^2}. \quad (6)$$

The new variables allow for a partial integration over four dimensions to resolve the four-momentum conservation constraint set by the delta distribution. The integration of equation (2) over the variables  $u, v, p_{x_2}, p_{y_2}$  results in

$$\sigma_{2 \rightarrow 2} = \frac{J}{(2\pi)^2} \sum_{j,k} \int \frac{1}{F} f_j(x_1) f_k(x_2) |\mathcal{M}_{2 \rightarrow 2}|^2 \frac{d^3 \vec{p}_1}{2E_1} \frac{dp_{z_2}}{2E_2}. \quad (7)$$

An additional change of variables can be performed using  $dp_{x_1} dp_{y_1} = p_T dp_T d\phi_1$ , where the transverse momentum  $p_T$  is the same for both final state particles in this case. The integration over the remaining momentum components in  $z$  direction can be substituted using  $\eta_i = \text{arcsinh}(p_{z_i}/p_T)$ , such that  $dp_{z_i} = |\vec{p}_i| d\eta_i$ . The fully differential cross-section with these new variables is given by

$$d\sigma_{2 \rightarrow 2} = \frac{1}{64 (2\pi)^2 E_B^4} \sum_{j,k} \frac{1}{x_1 x_2 E_1 E_2} f_j(x_1) f_k(x_2) |\mathcal{M}_{2 \rightarrow 2}|^2 p_T |\vec{p}_1| |\vec{p}_2| dp_T d\phi_1 d\eta_1 d\eta_2. \quad (8)$$