Search for the $H \to b\bar{b}$ Decay of the Standard Model Higgs Boson in Associated Production with Vector Bosons Using ATLAS Data at $\sqrt{s}$ = 8 TeV

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Stephan Hageböck
aus
Bonn

Bonn, Januar 2018

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

1. Gutachter:	Prof. Dr. Norbert Wermes
2. Gutachter:	Prof. Dr. Ian C. Brock
Tag der Promotion:	03.09.2018
Erscheinungsjahr:	2018

Acknowledgements

I would like to thank Dr. Jan Therhaag and Dr. Götz Gaycken for being the first to propose, and to actively work for using machine-learning algorithms for the $VH \to Vb\bar{b}$ search in ATLAS. Without their collaboration, the implementation of the searches presented in this work would not have been possible. I especially want to thank Götz Gaycken for always being available for giving physics- and computing-related advice, and for implementing a remarkable data analysis framework that facilitated all our research.

Furthermore, I want to thank Dr. Vadim Kostyukhin, Dr. Eckhard von Törne and Prof. Dr. Norbert Wermes for countless excellent ideas, for guidance and for helping to advocate the use of machine learning for this challenging search. I also want to thank them for questioning ideas and strategies, and thereby constantly helping to improve the analysis. I am grateful to Eckhard von Törne for supervising me during my Diploma thesis, which taught me a lot about scientific work and how to learn from my results.

I want to specially thank Prof. Dr. Norbert Wermes for accepting me as a member of his group, for giving me the opportunity to do this work, for giving me access to the ATLAS collaboration and for funding my research. I especially appreciate that he sends his students to CERN to take part in the endeavours of the scientific community from all over the world.

I want to thank Cusanuswerk for supporting me financially, and I am very grateful for all the non-material support I received. As a member of Cusanuswerk, I met a lot of interesting people, some of which I now call my friends.

Many members of the Bonn group and of the people I met at CERN also have become my friends during these years. They helped to overcome many difficulties that one faces when doing such challenging work, and made the time in Bonn and at CERN very enjoyable – be it volleyball, skiing, scuba diving, hiking, the famous movie nights or the “little reunions”.

I also want to thank my long-term friends, who shared all the good and the bad times with me, who always have been a source of advice, encouragement, delight and energy.

Last but not least, I want to thank my family, who always supported me, especially in doing this work, encouraged me in all my decisions, and taught me a lot in all the years.

Contents

I Preamble **1**

II The Role of the Higgs Boson in the Standard Model of Particle Physics **3**
II 1 Introduction . . . 3
II 2 Gauge Invariance: The Central Guiding Principle of the Standard Model . . . 7
II 2.1 $U(1)$ Gauge Symmetry and QED . . . 8
II 2.2 $SU(3)$ Gauge Symmetry and QCD . . . 10
II 2.3 Gauge Symmetries in Weak Interactions . . . 12
II 3 Gauge Symmetries and Boson Masses . . . 15
II 3.1 Spontaneous Symmetry Breaking and the Goldstone Theorem . . . 15
II 3.2 The Higgs Mechanism . . . 16
II 4 Yukawa Coupling and Fermion Masses . . . 20
II 5 The Higgs Boson . . . 22
II 5.1 Discovery . . . 22
II 5.2 Higgs Production . . . 24
II 5.3 Higgs Decays and Coupling Measurements . . . 26

III Particle Physics at Hadron Colliders **29**
III 1 The Large Hadron Collider . . . 29
III 1.1 Luminosity and PDFs . . . 31
III 2 Collider Detectors: The ATLAS Experiment . . . 34
III 2.1 Tracking Detectors: Measuring the Momenta of Charged Particles . . . 35
III 2.2 Calorimetry: Measuring the Energy of Particles and Jets . . . 37
III 2.3 Muon Spectrometer: Measuring the Momenta of Muons . . . 40
III 2.4 Trigger System . . . 40
III 3 Reconstruction of Particle Collisions in ATLAS . . . 42
III 3.1 Reconstruction of Visible Objects . . . 42
III 3.2 Reconstruction of Missing Transverse Energy . . . 46
III 3.3 Overlap Removal . . . 47
III 3.4 b-tagging . . . 48
III 4 Simulation of Particle Collisions . . . 50
III 4.1 Running Couplings and Renormalisation Scale . . . 50
III 4.2 Factorisation Theorem . . . 54
III 4.3 Parton Showers, Multi-Leg Generators and Hadronisation . . . 55
III 4.4 Other Aspects of Simulating LHC Collisions . . . 58

IV Search for $H \to b\bar{b}$ Decays with ATLAS **61**
IV 1 Backgrounds, Signal Selection and Event Categories . . . 62
IV 1.1 Selection of Objects . . . 66
IV 1.2 Event Selection and Categorisation of Events . . . 68
IV 1.3 b-tagging at Multiple Working Points . . . 73
IV 2 Modelling of Signal and Background Processes . . . 75
IV 2.1 Signal Simulations . . . 75
IV 2.2 b-jet Energy Correction . . . 75
IV 2.3 V+jets . . . 77
IV 2.4 Truth-Flavour Tagging at Multiple Working Points . . . 79
IV 2.5 Top Transverse Momentum Distribution . . . 80
IV 2.6 QCD Multijet Estimation . . . 81
IV 3 A Multivariate Analysis for Detecting a $H \to b\bar{b}$ Signal . . . 87
IV 3.1 Introduction to Boosted Decision Trees . . . 87
IV 3.2 Overfitting and Regularisation . . . 91
IV 3.3 BDTs for $H \to b\bar{b}$ Decays . . . 95
IV 4 Excursion: Covariate Shift and Multivariate Monte Carlo Modelling Checks . . . 100
IV 5 Statistical Analysis of the Data . . . 102
IV 5.1 A Likelihood Model for Extracting a Signal Strength . . . 102
IV 5.2 Parametrisation of Systematic Uncertainties with Nuisance Parameters . . . 104
IV 5.3 Statistical Tests for Detecting a $H \to b\bar{b}$ Signal: The Profile-Likelihood Ratio . . . 106
IV 5.4 Systematic Uncertainties and the Profile-Likelihood Ratio . . . 114
IV 6 The Likelihood Model for the $VH \to Vb\bar{b}$ Search . . . 117

V Beyond the ATLAS $H \to b\bar{b}$ Analysis **127**
V 1 A Lorentz-Invariant MVA . . . 127
V 1.1 Lorentz-Invariant Quantities for a Full Description of the VH System . . . 127
V 1.2 Details on the 1-Lepton Channel . . . 133
V 1.3 Details on the 2-Lepton Channel . . . 135
V 1.4 Better BDT Training With Gradient Boosting . . . 137
V 1.5 Comparison of Lorentz-Invariant and ATLAS VH Classifiers . . . 144
V 2 Improvement of the Likelihood Model . . . 152
V 2.1 Smooth Parametrisations of Systematic Uncertainties . . . 152
V 2.2 Analysis of Nuisance-Parameter Pulls . . . 156
V 2.3 The Treatment of V+bl Normalisations . . . 161

VI Measurement of the $VH \to Vb\bar{b}$ Signal Strength **171**
VI 1 Comparison with ATLAS Results . . . 174
VI 2 Measurement of the Diboson Process . . . 177
VI 2.1 Comparison with ATLAS Diboson Results . . . 179

VII Summary **183**
VII 1 Outlook . . . 184

Bibliography **187**

A Notes on the Introduction **201**
A 1 Notes on the Discussion of Gauge Invariance . . . 201
A 2 A Note on the Higgs Mass . . . 202
A 3 Useful Relations for Four-Vectors . . . 205

B Notes on the ATLAS $H \to b\bar{b}$ Analysis **207**
B 1 The Contribution of sg Scattering in $W+cl$ Events . . . 207
B 2 Extrapolating the $1b$-tag QCD Multijet Template to the $2b$-tag Region . . . 208
B 3 A BDT-driven Determination of Preselection Cuts . . . 209
B 4 Multivariate Modelling Checks and Covariate Shift . . . 211

C Notes on the Lorentz-Invariant Analysis **213**
C 1 Plots of the 3-Jet Regions . . . 213
C 1.1 Lorentz Invariants . . . 213
C 1.2 BDT Classifiers . . . 214
C 2 Breakdown of Systematic Uncertainties: Comparing Lorentz-Invariant MVA and ATLAS Approach . . . 215
C 3 Details on the Validation of the Likelihood Model . . . 217
C 3.1 Invariant Mass of b-jets for $t\bar{t}$ Events . . . 217
C 4 Response of MV1c . . . 218
C 4.1 $1b$-tag Region . . . 218
C 4.2 $2b$-tag Region . . . 221
C 4.3 Nuisance-Parameter Pulls of the Improved Likelihood Model . . . 223
C 4.4 Necessity of an Extension of the Likelihood Model for cl Events . . . 227

D Event Yields **229**

List of Figures **233**

List of Tables **236**

CHAPTER I

Preamble

One of the oldest quests of mankind is to understand the world we are living in. A natural question, in this respect, is to ask what the universe is made of. From Aristotle's four roots fire, water, air and earth, knowledge developed to Leucippus' and Demokritus' philosophy of the ατομος, the indivisible, to the discovery of chemical elements in the 17$^{\text{th}}$ and to the theory of atoms in the 19$^{\text{th}}$ century. A shift of paradigms occurred with the discovery of subatomic particles at the end of the 19$^{\text{th}}$ century, which showed that the atoms were, in fact, not indivisible. The development of quantum theory at the beginning of the 20$^{\text{th}}$ century was the true revolution, though. Einstein's saying "I, at any rate, am convinced that *He* does not throw dice" expresses what many could not believe at the time: the subatomic world is reigned by chance and probability instead of determinism.

Nowadays, the fundamental building blocks of matter are understood to be elementary particles such as the electron or the quarks. The development of quantum field theory in the 20$^{\text{th}}$ century provided the tools to describe their properties, and how they interact. Uniting classical field theory, quantum theory and special relativity, relativistic quantum field theories are among the most accurate theories ever conceived.

To this day, the most successful attempt to formulate a quantum field theory is the development of the Standard Model of elementary particles in the 20$^{\text{th}}$ century. It describes all known elementary particles and three of the four known forces of nature. The Standard Model answered many questions about the building blocks of our universe, but it also raised new ones. One long-standing question has presumably been answered recently: how do elementary particles acquire mass?

The answer to this question is connected with the existence of an elementary particle, the Higgs boson, which was predicted in 1964. Particle physicists have been searching for it ever since. The Higgs boson is the cornerstone of the Standard Model because it has the special role of being connected with the mechanism that gives mass to the other particles: the Higgs mechanism. Furthermore, after the discovery of the τ-neutrino in 2000, all particles predicted by the Standard Model *except* for the Higgs boson had been found. Consequently, the discovery of a Higgs-like particle at CERN's Large Hadron Collider (LHC) in 2012 was regarded as the most important breakthrough for particle physics in decades. Measurements seem to indicate that this particle is the Higgs boson, but some of its properties remain under study to this day. This thesis contributes to this research: if the new particle is indeed the Higgs boson, it must decay to b quarks. Proving this is the goal of the searches presented in this work.

$H \to b\bar{b}$ searches, however, are very challenging measurements at the Large Hadron Collider. They have been attempted several times by the ATLAS [2–4] and CMS collaborations [5, 6] and also at the Tevatron, but without finding conclusive evidence. In this thesis, I will therefore discuss how machine-learning methods were introduced to the ATLAS search [1] for $H \to b\bar{b}$ decays to improve the sensitivity – a strategy pioneered here in Bonn [7].

We will proceed as follows: in chapter II, I will review some of the theoretical foundations of the Standard Model, and discuss the importance of the Higgs boson. In chapter III, I will discuss aspects of particle physics at hadron colliders, and introduce the experimental tools: LHC and ATLAS. Chapter IV is dedicated to the first ATLAS $H \to b\bar{b}$ analysis that used machine learning, which received sizeable contributions from the Bonn $H \to b\bar{b}$ group. I will focus on final states with charged leptons, which I worked on.

Chapter V is dedicated to a $H \to b\bar{b}$ analysis that explores concepts that could not be introduced to the analysis published by ATLAS described in chapter IV. This second $H \to b\bar{b}$ analysis is based on the ATLAS analysis, but the machine-learning part has been revised to reduce uncertainties of the ATLAS analysis. This will be relevant for future $H \to b\bar{b}$ searches because with growing datasets, the reduction of systematic uncertainties is the next big challenge in this search. In chapter VI, I will discuss the results of measurements in ATLAS data recorded in 2012. The thesis concludes with a summary and an outlook to future searches in the $H \to b\bar{b}$ channel in chapter VII.

CHAPTER II

The Role of the Higgs Boson in the Standard Model of Particle Physics

II 1 Introduction

The Standard Model of particle physics [8–12] is the theory that summarises most of today's understanding of how elementary particles interact. It describes the effects of three forces of nature: electromagnetism, the strong force and the weak force. The last fundamental force, gravity, is not part of the Standard Model. Figure II.1 shows the known elementary particles and their quantum numbers, which determine how the forces act on the particles. The particles can be divided according to their spins (■[1]), into matter particles with spin $\frac{1}{2}$, the `fermions`, and force carriers with spin 1, the `bosons`.

Among the fermions, one finds three generations, also known as families. The first generation features the lightest particles, which form the ordinary matter. Particles in the other families have identical quantum numbers (and therefore the same properties), but higher masses than their counterparts in the first generation. Due to the high masses, the members of the second and third generation are unstable, and hence decay to particles of the first generation. The masses in figure II.1 are given in the "natural units" commonly used in particle physics. Setting the fundamental constants to one,

$$\hbar = c = k_{\mathrm{B}} = 1,$$

quantities like energy, momentum, mass, temperature, inverse time and inverse length can be measured in a unit of energy, the `electronvolt`:

$$[E] = [p] = [m] = [T] = \left[\frac{1}{t}\right] = \left[\frac{1}{d}\right] = \mathrm{eV} = 1.60218{\times}10^{-19}\ \mathrm{J}\,.$$

This convention is widely used in particle physics since it simplifies many relations as, for example, the relativistic energy-momentum relation:

$$E^2 = p^2\mathrm{c}^2 + m^2\mathrm{c}^4 \rightarrow p^2 + m^2$$

The three fundamental forces are mediated by the bosons, which are therefore often called "force carriers". Each boson acts on a different kind of charge that determines whether fermions are subject to the force: the quarks carry a colour charge (■), and are thus subject to the strong interaction mediated by

[1] Occasionally, colours in figures will be indicated by these coloured boxes in the text.

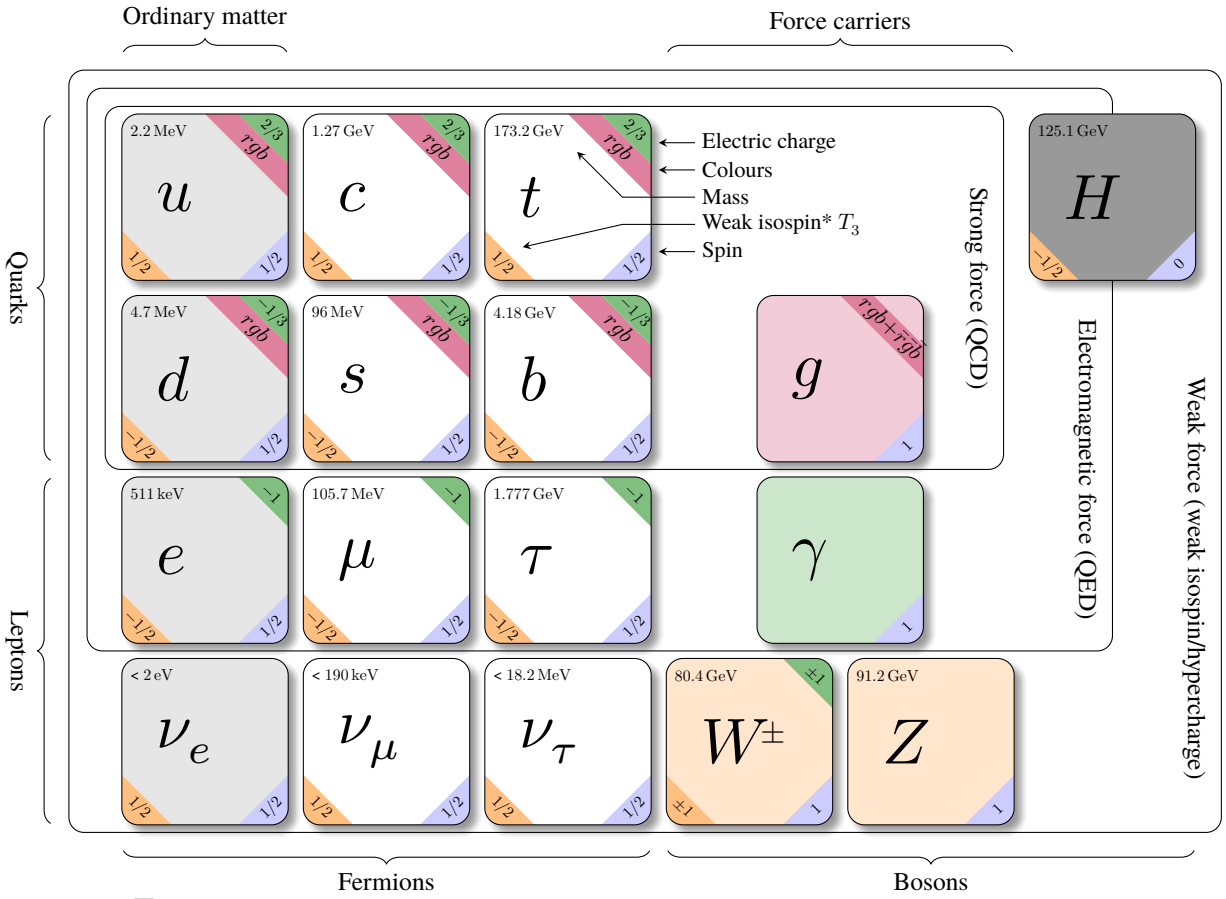


Figure II.1: Particles described by the Standard Model. Quantum numbers of the particles are indicated in coloured boxes. Masses from the PDG [13]. (*) The third component of weak isospin is given for left-handed particles only, since the weak interaction does not act on right-handed particles. For right-handed particles, the weak isospin is always zero. Based on [14], updated and modified.

the gluons g. All electrically charged (■) particles take part in the electromagnetic interaction mediated by the photon γ. All particles, except gluons and photons, also carry weak isospin (■) and weak hypercharge, and therefore take part in the weak interaction mediated by the W and Z bosons. Gravity cannot be described by the Standard Model. At the energy scales accessible for particle physics, gravity is weak, though, and it can therefore be neglected.

The Standard Model is a theory with large predictive power, and has been tested extensively. Its success stems from the fact that by imposing mathematical symmetries on Lagrange densities, it can predict the interactions of bosons and fermions with remarkable accuracy. It has even predicted *the existence* of several particles before any experimental evidence was at hand. Examples for particles predicted by the Standard Model are:

- the gauge bosons W, Z and g,
- the heavier fermions c, b, t,
- the τ neutrino ν_τ
- and the Higgs boson.

The Higgs boson, which is the focus of this thesis, has a special role, because it is indicative of the mechanism that gives mass to the Standard Model particles: the BEH mechanism. The confirmation of its existence in 2012 was a milestone for the experimental tests of the Standard Model; particularly because all the particles predicted by the Standard Model are now confirmed. As an example for how the

Standard Model predicts interactions, it will be shown in section II 2 that photons are *required* to couple to charged fermions if one imposes a gauge symmetry on the equations of motion of a fermion. Utilising this concept will lead to the prediction of the Higgs boson and its interactions.

Despite its success, the Standard Model also has limitations. A non-exhaustive list of these includes: it does not explain the role of gravity, why neutrinos have mass, or why the universe seems to be dominated by dark matter and dark energy. Furthermore, the 19 parameters of the Standard Model cannot be predicted. The Higgs mechanism, for instance, explains *how* particles acquire mass, but cannot predict the masses themselves. The origin of the mass hierarchy shown in figure II.1 is hence unknown. Furthermore, the Standard Model does not have a strong-enough mechanism to favour matter over antimatter (CP-violation), and the energy regime of the Higgs sector is surprisingly low in comparison to the energies where gravity supposedly becomes important: the Planck mass (hierarchy problem). Those many questions that cannot be answered by the Standard Model are good reasons for the existence of an underlying mechanism that includes the Standard Model as an effective theory. Describing such a mechanism would require a new theory of elementary particles, likely including also new particles and interactions. Although the search for new physics is not the aim of this thesis, the measurement of Standard Model observables is nonetheless interesting in this regard, as Standard Model observables might be affected by unknown processes. Consequently, the topic of this thesis is not only to determine whether $H \to b\bar{b}$ decays exist, but also to quantify their probability, that is, to measure the coupling, and compare it to the prediction of the Standard Model.

This is achieved by measuring the probabilities of interactions. Any measurement of such probabilities might provide insights into the validity of the Standard Model, and they are conveniently expressed using the `cross section`. When scattering a beam of particles at a target, one can express the scattering rate as

$$\dot{N} = \mathscr{L} \cdot \sigma, \qquad [\sigma] = \mathrm{m}^2 = 1{\times}10^{28}\,\mathrm{b}.$$

The `Luminosity` $\mathscr{L}$ is the number of particles per unit time and unit area, and σ is the cross section. Cross sections in the quantum regime are usually measured in fractions of the unit barn. For colliding beams of particles, one can relate the cross section to the probability of transitioning from the state $|1,2\rangle$, which is formed by two initial particles 1 and 2, into some final state $|3,\dots,n\rangle$. Such transitions are described by the S matrix: $\langle 3,\dots,n|\, S\, |1,2\rangle$. Separating the S matrix into $S = \mathbb{1} + \mathrm{i}\,T$ and extracting kinematic factors (factors that do not depend on the interaction), one can write down an expression for the cross section in dependence of the invariant `matrix element` $\mathcal{M}$:

$$\begin{aligned}\sigma = &\frac{\mathcal{N}}{2E_1 2E_2\,|v_1 - v_2|}\int \;|\mathcal{M}(p_1,p_2 \to p_3,\dots,p_n)|^2 \\ &\times (2\pi)^4\delta^4(p_1+p_2-p_3-\dots p_n) \times \prod_{f=3}^{n}\frac{1}{2E_f}\frac{\mathrm{d}^3\vec{p}_f}{(2\pi)^3}\end{aligned} \tag{II.1}$$

p_i are the four-momenta of the particles, and E_i their energies. $|v_1 - v_2|$ is the relative velocity of the beams as viewed from the laboratory frame, and $\mathcal{N}$ is a normalisation factor to prevent double counting when identical particles are participating. The δ^4-distribution ensures energy and momentum conservation, $\mathrm{d}^3\vec{p}_f$ is an element of the phase space of possible final states. With a similar rule for particle decays, the reactions of the known elementary particles can be described provided that the amplitudes $\mathcal{M}$ are known.

These amplitudes can be obtained using the "Feynman Rules", which is a set of rules to translate interactions permitted by the Standard Model to scattering amplitudes. To obtain the total probability of an interaction, *all* possible transitions from the initial to the final state under consideration have to be

evaluated. However, not all transitions are equally probable. This allows for the application of perturbation theory, that is, to calculate the approximate total amplitude using a power series in a parameter α, the coupling constant of an interaction, as long as $\alpha \ll 1$. If α is small, amplitudes only need to be evaluated up to a specific order in α because higher-order diagrams only account for very small corrections.

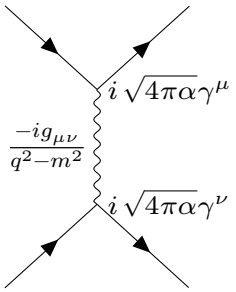


Figure II.2: Feynman diagram showing the scattering of fermions via photon exchange.

Figure II.2 shows a Feynman diagram describing the scattering of two fermions via exchange of a photon. It is leading order (`LO`) in α. For full details on the translation between diagrams and amplitudes, one may refer to textbooks, for example [15, 16]. For this work, the most interesting features of the Feynman rules are:

- Any vertex allowed by a Standard Model interaction can be part of a Feynman diagram.
- At each vertex, conservation of four-momenta has to be ensured.
- Each vertex contributes to the amplitudes with a factor $\propto ig$, where g is a coupling constant specific to the interaction of interest.
- For QED $g = \sqrt{4\pi\alpha_{\text{QED}}} \approx 0.3$, for QCD $g = \sqrt{4\pi\alpha_s} \approx 1.1$[2]. In the present diagram, two vertices contribute with $\sqrt{\alpha_{\text{QED}}}$ each, which is why the diagram II.2 is leading order in α.
- Diagrams with more vertices contribute with lower amplitudes: while this diagram describes an amplitude $\propto \alpha \approx \frac{1}{137}$, the amplitude of a diagram with four vertices would be $\propto \alpha^2 \approx 5{\times}10^{-5}$. This would be `next-to-leading order (NLO)` in α.
- The evolution of particles between vertices is described by propagators $\propto \frac{i}{q^2-m^2}$. These describe how a particle of mass m carries the four-momentum q from one vertex to the next.
- If particles have properties like spin, colour, hypercharge, *etc*., extra factors have to be added to vertices and propagators to ensure the conservation of these quantum numbers.

State-of-the-art cross-section calculations for a large range of processes include perturbative expansions up to `next-to-next-to-leading order (NNLO)`, or sometimes also N^3LO in α_s for QCD.

To study which interactions are allowed by the Standard Model, one can rely on the gauge principle: it predicts interactions of the bosons and the fermions, which defines the set of allowed vertices used in diagrams such as the one shown in figure II.2. Cross sections and decay probabilities will be revisited in section II 5.

[2] The couplings depend on the momentum q that is transferred. The couplings given here are for rather high, LHC-like momentum transfers of $\mathcal{O}(100\,\text{GeV})$.

II 2 Gauge Invariance: The Central Guiding Principle of the Standard Model

> Our basic postulate is that it should be possible to generate strong, weak and electro-magnetic interaction terms (with all their correct symmetry properties and also with clues regarding their relative strengths), by making local gauge transformations on the kinetic-energy terms in the free Lagrangian for all particles.
>
> *Salam, Ward, 1961 [17]*

To explore the connection of gauge invariance and interactions, it is beneficial to use the Lagrangian formalism. "Lagrangians", the density of kinetic and potential energy of a particle, will be denoted with $\mathcal{L}$. To generate the equations of motion, one uses the Euler-Lagrange equations. This will be demonstrated in a moment for QED. For now, one may take as given the Lagrangians for free particles that lead to the Klein-Gordon equation for spin-0 particles, the Dirac equation for spin-$\frac{1}{2}$ particles and the Proca equation for spin-1 particles:

$$\mathcal{L}_{\text{K-G}} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\ \phi^2 \tag{II.2}$$

$$\mathcal{L}_{\text{Dirac}} = \mathrm{i}\,\bar{\psi}\gamma^\mu\partial_\mu\psi - m\ \bar{\psi}\psi \tag{II.3}$$

$$\mathcal{L}_{\text{Proca}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}m_A^2\ A^\mu A_\mu \tag{II.4}$$

$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu$ is the strength tensor of the electromagnetic field. The fields in the Lagrangians assign operators to each point in space, and particles can be seen as excitations of these fields. The Klein-Gordon field ϕ in equation (II.2), for example, is the property of space-time to create a particle at position $\boldsymbol{x}$ as an excitation of the vacuum state $|0\rangle$:

$$\begin{aligned}|\boldsymbol{x}_{KG}\rangle = \phi(\boldsymbol{x})\,|0\rangle &= \int \frac{\mathrm{d}^3p}{(2\pi)^3}\frac{1}{\sqrt{2E_p}}(\mathrm{e}^{\mathrm{i}\,\boldsymbol{p}\cdot\boldsymbol{x}}\,a_p + \mathrm{e}^{-\mathrm{i}\,\boldsymbol{p}\cdot\boldsymbol{x}}\,a_p^\dagger)\,|0\rangle \\ &= \int \frac{\mathrm{d}^3p}{(2\pi)^3}\frac{1}{2E_p}\mathrm{e}^{-\mathrm{i}\,\boldsymbol{p}\cdot\boldsymbol{x}}\,|\boldsymbol{p}\rangle\,.\end{aligned}$$

The particle is a superposition of states $|\boldsymbol{p}\rangle$ with defined momentum $\boldsymbol{p}$ and energy $E_p = \sqrt{|\boldsymbol{p}|^2 + m^2}$, which themselves are the result of creating particles:

$$|\boldsymbol{p}\rangle = \sqrt{2E_p}\,a_p^\dagger\,|0\rangle\,.$$

$a_p^\dagger$ is the operator that creates a particle with momentum $\boldsymbol{p}$ when acting on the vacuum state, and a_p destroys such a particle.

The goal of the following sections is to generate the mass terms on the right-hand side of the Lagrangians in equations (II.2) to (II.4), while simultaneously preserving gauge invariance. One solution to this problem is the Higgs mechanism, which will be discussed in section II 3. The discussion will mostly follow the review of S. Novaes [18], complemented with aspects from P. Skands [19] and A. Djouadi [20].

II 2.1 $U(1)$ Gauge Symmetry and QED

To demonstrate why gauge symmetries are so important for the Standard Model, gauge transformations will now be applied to a Dirac field. One finds that the photon must be the mediator of the electromagnetic force. To do this, one may start with the Lagrangian of a free fermion, equation (II.3), and take the mass term as given. A global gauge (*i.e.* phase) transformation under the one-dimensional unitary group U(1) would affect the fermion field in the following way:

$$\psi \to \psi' = \mathrm{e}^{-\mathrm{i}\,\alpha\,\mathbb{1}_1}\psi \qquad \alpha \in \mathbb{R} \tag{II.5}$$

The one-dimensional identity matrix $\mathbb{1}_1$ was added to the notation in order to emphasise that all elements of Lie groups, which this gauge transformation is, can be written as $E(\theta_1, \dots, \theta_n) = \mathrm{e}^{-i\theta_j G_j}$ with θ_j being the parameters and G_j the generators of the group. A summation over indices appearing twice is implied. Equation (II.3) is invariant under this $U(1)$ transformation, since the complex number $\mathrm{e}^{i\alpha}$ commutes with all elements of the Lagrangian and $\bar{\psi}'\psi' = \bar{\psi}\mathrm{e}^{i\alpha}\mathrm{e}^{-i\alpha}\psi = \bar{\psi}\psi$. This is, however, only a *global* gauge transformation. It becomes *local* if the value of α depends on the space-time coordinate x – a much more restrictive assumption. In this case, the product rule has to be applied to $\mathrm{e}^{-i\,\alpha(x)}\psi$ in the first term of the Lagrangian:

$$\mathcal{L}'_\psi = \mathcal{L}_\psi + \bar{\psi}\gamma^\mu\psi\,\partial_\mu\alpha(x) \tag{II.6}$$

Even though both global and local phase changes should not affect the physics described by this equation, changes in $\alpha(x)$ do affect the Lagrangian. In particular, if the phase α would be allowed to vary with time *and* be observable at any location $\vec{x}$, information could be transmitted faster than the speed of light across the whole universe, violating the principle of causality. The solution to enforce gauge invariance is to add a gauge field to the Lagrangian (II.3):

$$\mathcal{L}_\psi = \mathrm{i}\,\bar{\psi}\gamma^\mu\partial_\mu\psi - m\,\bar{\psi}\psi - q\,\bar{\psi}\gamma^\mu\psi\,A_\mu \tag{II.7}$$

This counter term is meant to cancel the extra term in equation (II.6), and therefore it needs to be a vector field. Usually, this is expressed using the covariant derivative D_μ, since the covariant derivative is invariant under gauge transformations. This leads to the more concise notation:

$$\begin{aligned} \mathcal{L}_\psi &= \mathrm{i}\,\bar{\psi}\gamma^\mu D_\mu\psi - m\,\bar{\psi}\psi, \\ \text{where} \quad D_\mu\psi &\equiv (\partial_\mu + \mathrm{i}\,q\,A_\mu)\psi, \end{aligned}$$

and q is the charge of the fermion. To make this Lagrangian gauge invariant, the new field has to transform in a way that exactly cancels the extra term in the transformed Lagrangian (II.6):

$$A_\mu \to A_\mu{}' = A_\mu + \frac{1}{q}\,\partial_\mu a(x) \tag{II.8}$$

Taking the transformed Lagrangian (II.6) and adding the transformed new field yields the desired gauge invariance:

$$\begin{aligned} \mathcal{L}_\psi \to \mathcal{L}'_\psi &= \mathrm{i}\,\bar{\psi}\gamma^\mu\partial_\mu\psi - m\,\bar{\psi}\psi + \bar{\psi}\gamma^\mu\psi\,\partial_\mu\alpha(x) - q\,\bar{\psi}\gamma^\mu\psi\,A_\mu - q\,\bar{\psi}\gamma^\mu\psi\,\frac{1}{q}\partial_\mu\alpha(x) \\ &= \mathcal{L}_\psi \end{aligned}$$

To inspect the equations of motion described by the Lagrangian (II.7), the Euler-Lagrange equation can

be applied, for example with respect to $\bar{\psi}$:

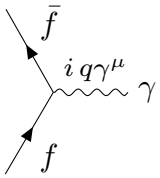
$$\begin{aligned} & \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \bar{\psi})} \right) - \frac{\partial \mathcal{L}}{\partial \bar{\psi}} = 0 \\ \Leftrightarrow \quad & \partial_\mu(0) - \mathrm{i}\,\gamma^\mu \partial_\mu \psi + m\psi + q\gamma^\mu \psi A_\mu = 0 \\ \Leftrightarrow \quad & \left(\mathrm{i}\,\gamma^\mu \partial_\mu - q\gamma^\mu A_\mu - m\right)\psi = 0 \end{aligned}$$

This is the Dirac equation for a fermion in an electromagnetic field A_μ. The postulation of local gauge invariance therefore yields the equations of motion for a fermion in an electromagnetic field! Furthermore, it predicts that the photon couples to charged fermions. The Lagrangian is, however, incomplete if it is meant to describe the interaction of a fermion with a photon. It is missing the terms for the motion of the photon field. The missing terms can be taken from the Proca-Lagrangian equation (II.4) because this equation describes the motion of free vector fields. The first term in the Proca-Lagrangian is invariant under the gauge transformation (II.8), but the mass term is not. Yet if A^μ is massless, which is known to be valid for the photon, gauge invariance is preserved. This yields the Lagrangian of QED:

$$\mathcal{L}_{\text{QED}} = \mathrm{i}\,\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} - q\,\bar{\psi}\gamma^\mu \psi A_\mu \tag{II.9}$$

Several properties of QED can be read off its Lagrangian (II.9):

- The mass of the fermion can be extracted from the $\bar{\psi}\psi$ term and is m, as in the free Dirac Lagrangian equation (II.3). This is, of course, built into the theory, since the Dirac Lagrangian for a massive fermion was taken as given. In section II 4 one will find, though, that the Yukawa mechanism can generate such terms in a gauge-invariant way.
- QED has exactly one interaction vertex connecting two fermion fields and the photon:

  Note that this is the only vertex needed to create the diagram shown in figure II.2. In fact, all Feynman diagrams of QED can be generated now because this is the only possible vertex (the type of fermion can change, though).
- Photons as the mediators of the electromagnetic force have to be massless in order to preserve gauge invariance.
- The coupling of fermions to the photon is proportional to their charge: $q = \sqrt{4\pi\,\alpha}$.

II 2.2 $SU(3)$ Gauge Symmetry and QCD

As for the electromagnetic force, the theory of the strong interaction can be based on a gauge symmetry: $SU(3)_C$. The subscript C indicates that the symmetry is in the $\{r, g, b\}$ space, the colour charge carried by quarks and gluons. The reason for not having a $U(3)_C$ symmetry is group theoretical: a gluon is a superposition of a colour and an anticolour state. For the direct product of such groups, one finds $3 \otimes \bar{3} = 8 \oplus 1$, an octet and a singlet of colour states. All states in the octet carry net colour charge, but the state in the singlet would be colour neutral. It would therefore not be bound by the strong interaction, and act as a free particle similar to the photon. Since such a particle is not known, the symmetry group must be $SU(3)$, which has eight instead of nine states. The $SU(3)$ symmetry manifests in nature as particles with colour charge behaving in the same way, regardless of their colour state. Viewed from outside, the colours of quarks and gluons are completely hidden. The generators of the $SU(3)$ group are the eight 3x3 Gell-Mann λ matrices, such that an element of $SU(3)$ can be written as $E(\theta_1, \dots, \theta_8) = \mathrm{e}^{-\mathrm{i}\,\theta_i \lambda^i}$. This is identical to the definition in equation (II.5), but the symmetry group had only one generator, the one-dimensional unity matrix. The λ matrices do not commute, therefore $SU(3)$ is non-Abelian. This makes the derivation of the gauge-invariant QCD Lagrangian more complicated than in the QED case with only the $U(1)$ symmetry. Details about gauge invariance with non-Abelian groups can be found in [18]. The QCD Lagrangian reads:

$$\begin{aligned}\mathcal{L}_{\mathrm{QCD}} = &-\frac{1}{4}(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu)(\partial_\mu G_\nu^a - \partial_\nu G_\mu^a) + \sum_{\alpha,f} \bar{q}_f^\alpha (i\gamma^\mu \partial_\mu - m_f) q_f^\alpha \\ &- g_s G_a^\mu \sum_f \bar{q}_f^\alpha \gamma_\mu \frac{1}{2} (\lambda^a)_{\alpha\beta} q_f^\beta \\ &+ \frac{g_s}{2} f^{abc} (\partial^\mu G_a^\nu - \partial^\nu G_a^\mu) G_\mu^b G_\nu^c - \frac{g_s^2}{4} f^{abc} f_{ade} G_b^\mu G_c^\nu G_\mu^d G_\nu^e\end{aligned} \tag{II.10}$$

An inspection of the QCD Lagrangian leads to the following insights about QCD:

- The terms in the first line are the kinetic terms for the gluon fields G and for the quarks q, in analogy to the QED Lagrangian.
- The second line contains the analogue of the photon-fermion coupling that appeared in the Dirac Lagrangian in equation (II.7). As for the Dirac Lagrangian, interaction terms appear when the normal derivative is replaced by the covariant derivative in the kinetic terms for the quark fields.
- For QCD, this only works if eight vector fields are added, corresponding to the eight generators of SU(3). Therefore, G_a^μ in the middle line is summed over the entries $\alpha\beta$ of the λ^a matrices, where $a = 1, \dots, 8$. Each non-zero combination defines specific interactions mediated by gluons.
- The eight gluons of QCD are massless as photons are. A mass term of the form $\frac{1}{2} m_g^2 F_{\mu\nu}^a F_a^{\mu\nu}$ with a generalised strength tensor F would break gauge invariance, and can therefore not be in the Lagrangian, unless $m_g = 0$.
- The interaction vertices of QCD are shown in figure II.3.

  Gluons interact with quarks via a QED-like vertex (left). The λ-matrices in the second line of the Lagrangian, however, only permit interactions where a gluon carries both the colour and the anticolour state of the incoming quarks, as illustrated by the subscripts in the leftmost diagram.
- Unlike photons, gluons have three-point and four-point self-interaction vertices caused by the non-Abelian nature of $SU(3)$ (middle and right vertex in figure II.3). These can be attributed to the last

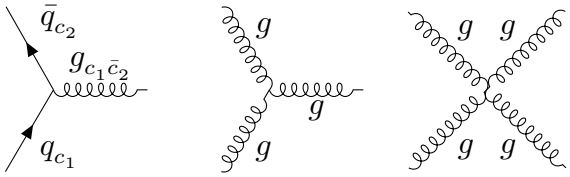


Figure II.3: Interaction vertices of QCD.

two terms in the Lagrangian, which arise because the strength tensor $F_{\mu\nu}$ needs to be generalised in a non-Abelian group:

$$F^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f_{abc} G^b_\mu G^c_\nu$$

f^{abc} are the structure constants of $SU(3)$. These ensure the conservation of the colour charge in the gluon self-interaction vertices.

- The strength of all gluon couplings is given by $g_s = \sqrt{4\pi\alpha_s}$, the coupling constant of the strong interaction.

Again, the requirement of *local* gauge symmetry in the Dirac Lagrangian yields a Lagrangian that naturally includes massless gauge bosons, and predicts all interactions of these gauge bosons with fermions. External knowledge that has to be added, though, is the fact that only quarks carry colour charge, and are thus the only fermions taking part in the strong interaction.

II 2.3 Gauge Symmetries in Weak Interactions

It could be assumed that a strategy similar to the one used in the $SU(3)$ QCD instance would yield the theory of weak interactions with the three gauge bosons W^+, W^-, Z^0. A suitable symmetry group with three generators is $SU(2)$. However, applying this strategy yields massless gauge bosons, which is in clear violation of experimental observations. The great achievement of Brout, Englert and Higgs was to formulate a gauge-invariant, symmetric Lagrangian that would generate mass terms for the heavy gauge bosons. This will be discussed in the next section because the application of gauge invariance in weak interactions is complicated by another fact: weak bosons only interact with left-handed fermions and right-handed antifermions. The chirality ("handedness") of fermions is the eigenvalue of the γ^5 matrix, and can be extracted from an arbitrary state using projection operators:

$$P_L \equiv \frac{1}{2}(1-\gamma^5) \qquad P_R \equiv \frac{1}{2}(1+\gamma^5).$$

The fact that weak interactions only act on left-handed fermions and right-handed antifermions manifests in the Lagrangian in the presence of the P_L projection operator. A weak current for left-handed fermions can be expanded as:

$$\bar{\psi}_L \gamma^\mu \psi_L = (\psi^\dagger P_L \gamma_0)\, \gamma^\mu P_L \psi = (\bar{\psi} P_R)\, \gamma^\mu P_L \psi = \bar{\psi}\gamma^\mu P_L P_L \psi = \bar{\psi}\gamma^\mu \frac{1}{2}(1-\gamma^5)\psi.$$

This gives rise to the term "$V-A$ structure" of weak interactions, since the last term transforms as a vector (γ^μ) minus an axial vector ($\gamma^\mu\gamma^5$).

Fermion mass terms always are a product of left- and right-handed fields:

$$-m\,\bar{\psi}\psi = -m\,\bar{\psi}(\underbrace{P_L + P_R}_{=1})\psi = -m\,\bar{\psi}P_L^2\psi - m\,\bar{\psi}P_R^2\psi = -m\,\bar{\psi}_R\psi_L - m\,\bar{\psi}_L\psi_R \qquad \text{(II.11)}$$

To apply the gauge principle, it must first be understood that quarks and leptons come in two states, as far as the weak interactions are concerned: $SU(2)_L$ doublets (with weak isospin $T = \frac{1}{2}$), on which the three gauge bosons act, and $U(1)$ singlets with $T = 0$ for the right-handed fermions:

$$\begin{pmatrix} q^u \\ q^d \end{pmatrix}_L \quad \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L \equiv L \qquad \text{vs.} \qquad q^u_R, \quad q^d_R, \quad \ell_R \equiv R \qquad \text{(II.12)}$$

Two shorthand notations R and L were defined here. Right-handed neutrino states are not part of the Standard Model. Since right- and left-handed terms need to be mixed for the mass terms as shown in equation (II.11), neutrinos need to be massless in the Standard Model. As far as the Standard Model is concerned, right-handed neutrinos, if they existed, would not interact with other particles, which renders them undetectable.

Given the two symmetry groups $SU(2)_L$ and $U(1)$, one needs to accommodate for the experimental result that the $SU(2)_L$ symmetry is not exact. Neutral weak interactions have a small but non-vanishing right-handed component; this hints at a contribution of the electromagnetic force, since the vector current in the photon interaction term includes both right- and left-handed particles:

$$\bar{\psi}\gamma^\mu\psi = \bar{\psi}_L\gamma^\mu\psi_L + \bar{\psi}_R\gamma^\mu\psi_R$$

A product of the two symmetry groups was proven successful, accommodating the electromagnetic component: $SU(2)_L \otimes U(1)$. Now, local gauge invariance can be imposed, replacing the normal by the

covariant derivative and introducing *four* vector fields, since one group has three, the other one generator:

$$\begin{aligned} SU(2)_L &\rightarrow W_\mu^1,\, W_\mu^2,\, W_\mu^3 \\ U(1)_Y &\rightarrow B_\mu \end{aligned}$$

As for QCD, the strength tensor needs to be generalised, since $SU(2)$ is not Abelian. This gives rise to self-interaction terms as for the gluons. With the shorthand notations R, L from equation (II.12), the Lagrangian for leptons, for example, reads:

$$\begin{aligned} \mathcal{L}_{\text{leptons}} = &\bar{L}\,\mathrm{i}\gamma^\mu\partial_\mu\, L + \bar{R}\,\mathrm{i}\gamma^\mu\partial_\mu\, R \\ &+ \bar{L}\,\mathrm{i}\gamma^\mu\left(\mathrm{i}\frac{g}{2}\tau_i W_\mu^i + \mathrm{i}\frac{g'}{2}YB_\mu\right) L \\ &+ \bar{R}\,\mathrm{i}\gamma^\mu\left(\qquad\qquad \mathrm{i}\frac{g'}{2}YB_\mu\right) R \end{aligned}$$

Here, τ^i are the Pauli matrices, the three generators of $SU(2)$, similar to the λ matrices as generators of $SU(3)$ in equation (II.10). The operator Y is the hypercharge operator, which is the charge corresponding to the $U(1)_Y$ symmetry. It can be seen as the generator of the $U(1)_Y$ group. The constants g and g' are coupling constants, adjusting the relative strength of the two interactions, and they must be determined experimentally. A summation over the three lepton generations is implied. The second and third lines predict interactions between the leptons and bosons, which arise when gauge invariance under the product group is postulated. As experiments have shown ("$V-A$ structure"), there is no interaction between the W^i bosons and right-handed particles. The B boson, however, ensures that right-handed particles also take part in the electroweak interaction.

Rewriting the Lagrangian in the Weinberg, Glashow, Salam model and inserting the Pauli matrices, different properties become visible:

$$\begin{aligned} \mathcal{L}_{\text{leptons}} = &\bar{L}\,\mathrm{i}\gamma^\mu\partial_\mu\, L + \bar{R}\,\mathrm{i}\gamma^\mu\partial_\mu\, R \\ &- \frac{g}{2}\bar{L}\,\gamma^\mu\begin{pmatrix} 0 & W_\mu^1 - \mathrm{i}\,W_\mu^2 \\ W_\mu^1 + \mathrm{i}\,W_\mu^2 & 0 \end{pmatrix} L \\ &- \Big(\tfrac{g'}{2}(\bar{L}\gamma^\mu Y L + \bar{R}\gamma^\mu Y R),\quad \tfrac{g}{2}\bar{L}\gamma^\mu\left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)L\Big) \underbrace{\begin{pmatrix}\cos(\theta_W) & -\sin(\theta_W)\\ \sin(\theta_W) & \cos(\theta_W)\end{pmatrix}\begin{pmatrix}A_\mu\\ Z_\mu\end{pmatrix}}_{\begin{pmatrix}B_\mu\\ W_\mu^3\end{pmatrix}} \end{aligned} \tag{II.13}$$

The off-diagonal elements in the second line of the equation allow for transitions between states in the $SU(2)$ doublets ($\ell \leftrightarrow \nu$). These transitions change the charge of the particles. Therefore, the $W^\pm$ bosons can be identified with the states:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp \mathrm{i}\,W_\mu^2). \tag{II.14}$$

As the third line of the Lagrangian comprises only neutral interactions, it can only be related to Z^0 bosons and to photons. This is why the B_μ and W_μ^3 fields have been written as a mixture of Z_μ and A_μ in equation (II.13). The `Weinberg angle` θ_W determines this mixture. Since electromagnetic interactions do not distinguish between left- and right-handed particles, the Weinberg angle needs to be chosen such that the photon interacts with the left- and right-handed fields with equal strength. Furthermore, photons need to be prevented from interacting with the neutrino component of $\frac{g}{2}\bar{L}\gamma^\mu\left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)L$ in the third line of

equation (II.13), since photons do not interact with neutrinos. Given that $YL = -1L$ and $YR = -2R$, this is achieved when

$$\begin{aligned} & g\sin(\theta_W) = g'\cos(\theta_W), \\ \Leftrightarrow \quad & \sin(\theta_W) = \frac{g'}{\sqrt{g^2+g'^2}}, \quad \text{or} \\ \Leftrightarrow \quad & \cos(\theta_W) = \frac{g}{\sqrt{g^2+g'^2}}. \end{aligned}$$

Writing out the R and L shorthands, and including the Weinberg angle, the neutral part of the Lagrangian (II.13) can be rearranged to[3]:

$$\begin{aligned} \mathcal{L}^{\text{Neutral}}_{\text{Leptons}} = \; & -g\sin(\theta_W)\,\bar{\ell}\gamma^\mu\ell\,A_\mu \\ & -\frac{g}{2\cos(\theta_W)}\sum_{i=\nu_L,\ell_L,\ell_R}\bar{\psi}_i\gamma^\mu(g_V - g_A\gamma^5)\psi_i\,Z_\mu. \end{aligned} \tag{II.15}$$

With the Gell-Mann-Nishijima relation:

$$Q \equiv T_3 + \frac{1}{2}Y, \tag{II.16}$$

and by comparing to the QED Lagrangian (II.9), the coupling constants of the gauge bosons can be read off the Lagrangian:

$$g_V \equiv T_3 - 2Q\sin^2(\theta_W) \tag{II.17}$$

$$g_A \equiv T_3 \tag{II.18}$$

$$e \equiv g\sin(\theta_W) \tag{II.19}$$

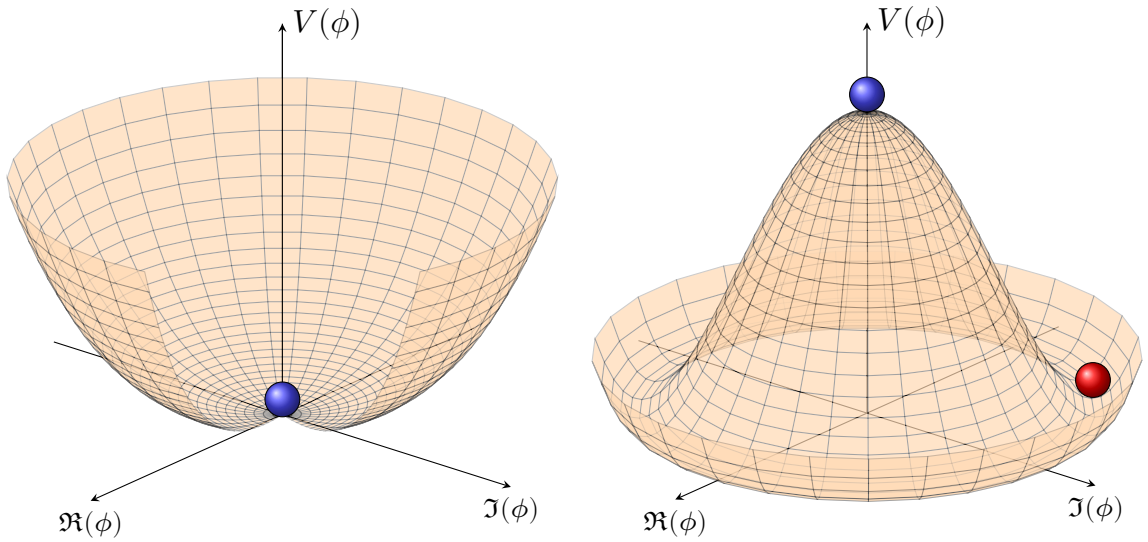
$T_3 = \frac{1}{2}\tau_3$ is the operator for the 3rd component of isospin, which is understood to act only on the $SU(2)$ doublets. For the $U(1)$ singlets, it is zero. The second line of the Lagrangian (II.15), the Z^0 interactions, again shows the $V-A$ structure of weak interactions.

A similar strategy can be employed to obtain the Lagrangian for quarks, but the hypercharges Y have to be adjusted to be compatible with the electric charges of the quarks: $+\frac{2}{3}$ for up- and $-\frac{1}{3}$ for down-type quarks. Furthermore, one needs to include transitions between the quark families, the CKM matrix [21, 22], such that heavy quarks can decay to light quarks mediated by $W^\pm$ bosons. Lastly, the weak boson (self-)interaction terms are missing in the present discussion (they are included pictorially in equation (A.2) on page 202 in the appendix). However, all particles are still massless at this point (except for fermions, where mass terms have been deliberately put into the QED/QCD Lagrangians). Therefore, this discussion will now focus on the Higgs mechanism, which solves the mass problem in a gauge-invariant way.

[3] For details see equation (A.1) on page 201 in appendix A 1

II 3 Gauge Symmetries and Boson Masses

II 3.1 Spontaneous Symmetry Breaking and the Goldstone Theorem



(a) Quartic potential with $a, b > 0$

(b) The "Mexican Hat Potential", $a < 0$

Figure II.4: Illustration of Spontaneous Symmetry Breaking (SSB). Shown are two symmetric quartic potentials of the form $a\,\phi^2 + b\,\phi^4$. **(a)**: a and b are positive. The ground state (blue) does not break the symmetry. **(b)**: a is negative. The state with vanishing field $\phi = 0$ (blue) is symmetric but unstable: a ball in such a potential will roll into a random direction. The transition into one of the states with minimal energy (red) breaks the symmetry. There is an infinite number of degenerate states of minimal energy.

Imposing local gauge invariance on free-particle Lagrangians demonstrated how gauge bosons interact with fermions. The three fundamental forces (except gravity) can all be described using this technique. For giving masses to particles, however, a new concept is necessary: spontaneous symmetry breaking (SSB). This concept can be illustrated with the Lagrangian of a complex scalar field ϕ in a quartic potential:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi^\dagger\partial^\mu\phi - \frac{1}{2}\mu^2\phi^\dagger\phi - \frac{1}{4}\lambda\left(\phi^\dagger\phi\right)^2 \tag{II.20}$$

The coefficients of the quartic potential are called μ, λ following the usual conventions. In order for the potential to be bounded, λ must be greater than zero, but μ^2 need not be. One can differentiate two cases:

1. $\mu^2 \geq 0$: The Lagrangian describes a scalar particle with mass $m = \mu$. Note that the Lagrangian is symmetric with respect to phase changes:

   $$\phi \to \phi' = \mathrm{e}^{i\theta}\phi$$

   The potential of this Lagrangian is shown in figure II.4(a). The ground state of this potential is $\phi_0 = 0$.

2. $\mu^2 < 0$: This potential is shown in figure II.4(b): an interesting feature of this potential is the fact that the ground state is not the state of vanishing field. Since the potential is symmetric with respect to phase changes, there is in fact an infinite number of degenerate minima. If a ball was put into such a potential, it could come to rest at any point of the well as shown for the red ball. Yet the

ground state of this system can be *only one* of the degenerate states. Since the state of vanishing field is unstable, and one of the ground states has to be occupied, the symmetry of the system is said to be "spontaneously broken".

One can assume the ground state (vacuum state) to be at $\Im(\phi) = 0$, without loss of generality:

$$v \equiv \phi_0 = \sqrt{-\frac{\mu^2}{\lambda}}$$

The field can be reparametrised around this state, that is, one can shift the real part of the field by the "vacuum expectation value" v, while the imaginary component remains unchanged:

$$\phi = \phi_1 + \mathrm{i}\,\phi_2 \quad \rightarrow \quad (\phi_1 - v) + \mathrm{i}\,\phi_2$$

Now the Lagrangian becomes:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi^\dagger\partial^\mu\phi - \frac{1}{2}(-2\mu^2)\,\phi_1^2 - \lambda v\,\phi_1(\phi_1^2+\phi_2^2) - \frac{\lambda}{4}(\phi_1^2+\phi_2^2)^2 + c \tag{II.21}$$

This is a Lagrangian of two scalar fields, a massive field ϕ_1 with $m = -2\mu^2$ and a massless field ϕ_2, a so-called "Goldstone boson". The terms proportional to λ are interaction terms of the fields. If a field is excited along the flat direction of the potential, the field is massless. The classical analogy is that one does not need energy to move the red ball along the line of minimal potential. By contrast, the field that corresponds to the direction where the potential rises is massive. Goldstone's theorem [23] states that when an exact symmetry is spontaneously broken, a massless scalar particle appears for each broken generator. For the Lagrangian in equation (II.20), $U(1)$ was spontaneously broken: the potential is $U(1)$-symmetric, but the vacuum state it occupies is not. Therefore, one Goldstone boson, ϕ_2, had to appear. The more interesting result concerning the problem of massless Standard Model particles, though, is the fact that a *massive* scalar particle appeared, *and* that the Lagrangian is still gauge-invariant. Indeed, it is the first mass term *generated* in this discussion. The fermion mass terms seen before in section II 2 were always assumed to be given. However, there is no experimental evidence of Goldstone bosons[4]. One more concept is therefore necessary.

II 3.2 The Higgs Mechanism

Although it was possible to generate a mass term without violating gauge-invariance, the Goldstone mechanism is not the full solution to the mass problem: one needs to combine SSB and local gauge invariance, so that the degrees of freedom appearing as Goldstone bosons transfer to the gauge bosons and make them massive. This is the essence of the Higgs mechanism, also called BEH or Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism, since it was proposed by Englert and Brout [24], Higgs [25–27], Guralnik, Hagen and Kibble [28, 29]. As there are three massive weak gauge bosons and the massless photon, one may start with four fields and break three generators to generate the mass terms. The corresponding symmetry should be the symmetry group of electroweak interactions: $SU(2)_L \otimes U(1)_Y$. This can be achieved by introducing a scalar $SU(2)$ doublet with hypercharge $Y = 1$ and with complex fields, that is, four degrees of freedom:

$$\phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

[4] Pions can be regarded as Goldstone bosons in QCD. They are, however, only pseudo-Goldstone bosons because pions have a mass of 140 MeV.

The superscripts indicate the electric charge of the states, which can be found by applying the charge operator $Q = T_3 + \frac{Y}{2}$. As in section II 2.3, a covariant derivative needs to be introduced to enforce gauge invariance on the Lagrangian, and to generate interactions with the gauge bosons:

$$\mathcal{L} = D_\mu \phi^\dagger D^\mu \phi - \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2 \tag{II.22}$$

$$D_\mu = \partial_\mu + \mathrm{i}\, g \frac{\tau_i}{2} W_\mu^i + \mathrm{i}\, g' \frac{Y}{2} B_\mu \tag{II.23}$$

Given that the potential is symmetric, a particular minimum can be chosen, spontaneously breaking the symmetry:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \qquad \text{with } v = \sqrt{-\frac{\mu^2}{\lambda}} \tag{II.24}$$

Since the generator of $U(1)_{\text{em}}$, $Q = T_3 + \frac{Y}{2}$, annihilates the vacuum state, which is electrically neutral, this minimum preserves the $U(1)_{\text{em}}$ symmetry[5]. The remaining generators, which *are* broken when an arbitrary vacuum state is occupied, will create massive gauge bosons. The number of broken/unbroken generators matches the number of massive/massless gauge bosons described by the Standard Model.

Starting from the vacuum state ϕ_0, adding a field H, and adding the exponentials of the generators of $SU(2)$, the field ϕ can be written in terms of $SU(2)$ transformations. In this notation the Goldstone bosons, which are just degrees of freedom of the $SU(2)$ symmetry, are visible:

$$\begin{aligned} \phi &= \mathrm{e}^{\mathrm{i}\, \frac{\tau^i}{2} \frac{\chi_i}{v}} \begin{pmatrix} 0 \\ \frac{v+H}{\sqrt{2}} \end{pmatrix} \\ &\approx \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{2}(\chi_2 + \mathrm{i}\,\chi_1) \\ v + H - \mathrm{i}\,\frac{1}{2}\chi_3 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathrm{i}\,\sqrt{2}\omega^+ \\ v + H - \mathrm{i}\, z^0 \end{pmatrix} \end{aligned} \tag{II.25}$$

Only terms of linear order in the fields have been kept in the second line. The right-hand part of the second line uses a suggestive notation that hints at the fact that the field can be written as H, the Higgs field, plus Goldstone bosons.

Since the Lagrangian (II.22) is $SU(2)$-symmetric, the parameters χ_i can be chosen such that the Goldstone bosons are eliminated from equation (II.25). However, the Higgs field remains. This is called "unitary gauge". For this, a $SU(2)$ gauge transformation with $\alpha_i = \chi_i / v$ has to be performed, completely cancelling all χ_i contributions. The Lagrangian (II.22) becomes:

$$\mathcal{L} = \left[\left(\partial_\mu + \mathrm{i}\, g \frac{\tau_i}{2} W_\mu^i + \mathrm{i}\, g' \frac{Y}{2} B_\mu \right) \frac{v+H}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right]^2 - \mu^2 \frac{(v+H)^2}{2} - \lambda \frac{(v+H)^4}{4}$$

A mass term similar to that from (II.21) and self-interaction terms proportional to λ appear. Multiplying

[5] It can be shown [18] that all generators that annihilate the vacuum state correspond to unbroken symmetries and a massless boson. Postulating the preservation of the $U(1)$ symmetry while acting on the vacuum state yields:

$$\begin{aligned} & \mathrm{e}^{\mathrm{i}\,\alpha Q} \phi_0 \approx (1 + \mathrm{i}\,\alpha Q)\phi_0 \overset{!}{=} \phi_0 \\ \Leftrightarrow \quad & Q\phi_0 = 0 \end{aligned}$$

out the part $\left[(g\tau_3 W^3 + g'YB)v\binom{0}{1}\right]^2$, one can write:

$$\mathcal{L}^{W^3+B} = \frac{v^2}{8}\begin{pmatrix} B_\mu & W^3_\mu \end{pmatrix}\begin{pmatrix} g'^2 & -gg' \\ -gg' & g^2 \end{pmatrix}\begin{pmatrix} B^\mu \\ W^\mu_3 \end{pmatrix}$$

Being quadratic in the fields, this equation describes the mass terms of the neutral bosons. Since the matrix is symmetric, it can be diagonalised. This yields two eigenvalues:

$$0 = m_A \quad \text{and} \quad \frac{1}{2}\frac{v^2(g^2+g'^2)}{4} = \frac{1}{2}m_z^2$$

This matrix happens to be diagonalisable with a rotation by the Weinberg angle θ_W. The mixing of W^3 and B is such that the photon remains massless, preserving the $U(1)_{\text{em}}$ symmetry. One can thus rotate the Lagrangian into the "mass eigenbasis"[6], and write $W^{1,2}$ as $W^\pm$, see equation (II.14). All boson mass terms can now be identified by inspecting the terms that are quadratic in the fields:

$$\mathcal{L} = \frac{1}{2}\partial_\mu H \partial^\mu H + \frac{1}{2}\frac{g^2}{4}(v+H)^2\left(2\,W^+_\mu W^{-\mu} + \frac{1}{\cos^2(\theta_W)}Z_\mu Z^\mu\right) - \mu^2\frac{(v+H)^2}{2} - \lambda\frac{(v+H)^4}{4}$$

One obtains[7]:

$$\begin{aligned} m_W &= \frac{gv}{2} \\ m_Z &= \frac{gv}{2}\frac{1}{\cos(\theta_W)} \\ m_H &= \sqrt{-2\mu^2} = \sqrt{2\lambda}v \\ m_A &= 0 \end{aligned}$$

In its most expressive form, the Lagrangian therefore reads:

$$\begin{aligned} \mathcal{L} = {} & \frac{1}{4}\lambda v^4 \\ & + \frac{1}{2}\partial_\mu H\,\partial^\mu H - \frac{1}{2}m_H^2 H^2 - \frac{m_H^2}{2v}H^3 - \frac{m_H^2}{8v^2}H^4 \\ & + \frac{1}{2}m_W^2\,(2\,W^+_\mu W^{-\mu})\left(1 + \frac{2}{v}H + \frac{H^2}{v^2}\right) + \frac{1}{2}m_Z^2\,Z_\mu Z^\mu\left(1 + \frac{2}{v}H + \frac{H^2}{v^2}\right) \end{aligned} \tag{II.26}$$

The following points are worth noting:

- The postulate of a quartic potential for a scalar doublet of the form (II.22), in conjunction with the postulate of local gauge invariance and spontaneous symmetry breaking, generates mass terms and interaction terms for the weak gauge bosons.
- Provided the quartic potential exists in nature, the corresponding "Higgs field" would have a non-zero vacuum expectation value, equation (II.24). It was determined using data about the Fermi constant G_F, which can be measured in muon decays [18]:

$$v = \sqrt{-\frac{\mu^2}{\lambda}} = \sqrt{\frac{1}{\sqrt{2}G_F}} \approx 246\,\text{GeV}$$

[6] Compare equation (II.13) in section II 2.3

[7] The W mass term in the Lagrangian has an additional factor of 2. It arises because there are mass terms for two W particles: $2W^+_\mu W^{-\mu} = W_{1\mu}W_1^\mu + W_{2\mu}W_2^\mu$

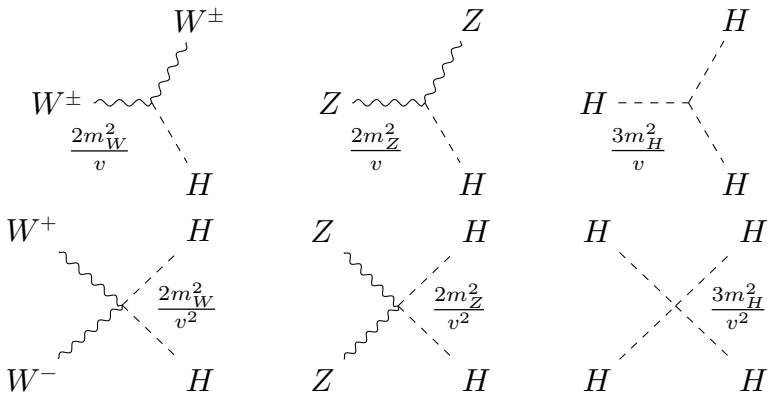


Figure II.5: Interaction vertices between gauge bosons and Higgs. All couplings are proportional to the squared boson mass. The upper left and upper middle diagrams are the most important couplings for this work.

- The vacuum expectation value v has units of mass, and is the only parameter of the Standard Model that is not dimensionless. It also determines the mass scale of all bosons and fermions (see next section). Why v is $\mathcal{O}(100\,\mathrm{GeV})$ is one of the unsolved problems of the Standard Model, called the "hierarchy problem". The name originates from the fact that the only other relevant mass scale is significantly different from v: the Planck mass $m_P = \sqrt{\frac{\hbar c}{G}} \approx 1.2{\times}10^{19}\,\mathrm{GeV}$.
- An excitation of the Higgs field corresponds to creating a particle, the Higgs boson. Peter Higgs was the first to postulate its existence, which lead to the name of the particle.
- As g and λ are free parameters, the masses of W, Z and H cannot be predicted. The prediction of the ratio m_W/m_Z, though, was a great success of the Standard Model. Even though λ is in principle a free parameter, there are intervals for λ that are favoured by the Standard Model. A brief discussion can be found in appendix A 2 starting on page 202. The Higgs mass measured at LHC, $m_H = (125.09 \pm 0.24)\,\mathrm{GeV}$ [30], is right at the boundary of stability, where loop effects could render λ negative, leading to an unstable quartic potential. If there were unknown particles preventing this from happening, they would couple to the Higgs, and thus affect production cross sections or branching fractions. Measuring cross sections and branching fractions – one of the goals of this thesis – might therefore provide clues in the search for new physics.
- The Goldstone bosons predicted by the Goldstone theorem disappeared when a proper gauge transformation was applied. This is in agreement with experimental data that show no signs of Goldstone bosons.
- The degrees of freedom that would be carried by the Goldstone bosons are transferred to the $W^\pm$ and Z bosons: massless vector bosons can only have two spin states (z-component): $S_z = \pm 1$. If they become massive, they acquire also $S_z = 0$, preserving the number of degrees of freedom.
- Imposing SSB and local gauge invariance predicts *all* interactions of the Higgs boson with the gauge bosons: Equation (II.26) shows that the Higgs couples to pairs of W or pairs of Z bosons and to itself via a triple and quartic coupling.
- The couplings to the heavy gauge bosons and also the Higgs self couplings are proportional to the squared mass of the bosons. The Higgs boson therefore couples most strongly to the heaviest particles.
- The quartic Higgs coupling and the coupling $VVHH$ are suppressed with respect to the triple couplings by $\frac{1}{v}$. The so-called "Higgs-Strahlung" diagrams (the upper left and upper middle diagrams in figure II.5) will therefore be the most important couplings for this work.

II 4 Yukawa Coupling and Fermion Masses

Despite the success for vector bosons, the Higgs mechanism cannot explain why fermions are massive. Since directly adding mass terms to the Lagrangian breaks gauge invariance, a new concept is necessary: Yukawa interactions [9, 31, 32]. Adding interactions between the Higgs doublet and fermions yields massive fermions. The subset of the Yukawa Lagrangian for just one lepton generation reads:

$$\mathcal{L}_{\text{Yukawa}} \supset -y\,\bar{\ell}_R \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}^\dagger \begin{pmatrix}\nu\\ \ell\end{pmatrix}_L - y^* \begin{pmatrix}\nu\\ \ell\end{pmatrix}^\dagger_L \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix} \ell_R\,.$$

Note that it contains left- and right-handed terms, a necessary prerequisite for fermion mass terms. The constant y is a complex number, and determines the coupling strength. It can even be chosen real by moving any phase into ℓ_R. In unitary gauge, the Higgs doublet simplifies to $\frac{1}{\sqrt{2}}\left(\begin{smallmatrix}0\\ v+H\end{smallmatrix}\right)$, equation (II.25), and the Lagrangian becomes:

$$\begin{aligned}\mathcal{L}_{\text{Y}} &\supset -y\,\frac{v+H}{\sqrt{2}}\left[\bar{\ell}_R\ell_L + \bar{\ell}_L\ell_R\right]\\ &= -m_\ell\bar{\ell}\ell - \frac{m_\ell}{v}H\bar{\ell}\ell\end{aligned}$$

The fact that fermion mass terms have the structure $RL + LR$ (equation (II.11)) was used to arrive at the second line. Neutrino mass terms are not possible, since a right-handed neutrino singlet is not part of the Standard Model.

For quarks, the procedure is similar. For up-type quarks, one needs the conjugate doublet[8]:

$$\overline{\begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}} \equiv \mathrm{i}\,\tau_2 \begin{pmatrix}\phi^-\\ \phi^{0*}\end{pmatrix} = \begin{pmatrix}\phi^{0*}\\ -\phi^-\end{pmatrix}$$

Hence, the quark part of the Yukawa Lagrangian reads:

$$\mathcal{L}_Y \supset -y_d \begin{pmatrix}\bar{u} & \bar{d}\end{pmatrix}_L \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix} d_R - y_u \begin{pmatrix}\bar{u} & \bar{d}\end{pmatrix}_L \begin{pmatrix}\phi^{0*}\\ -\phi^-\end{pmatrix} u_R + \text{h.c.}$$

In unitary gauge:

$$\mathcal{L}_Y \supset -\frac{y_d v}{\sqrt{2}}\bar{d}d - \frac{y_d}{\sqrt{2}}H\bar{d}d - \frac{y_u v}{\sqrt{2}}\bar{u}u - \frac{y_u}{\sqrt{2}}H\bar{u}u$$

[8] This results from complex-conjugating an infinitesimal gauge transformation, like in equation (II.25), and requiring that the conjugated doublet transforms in the same way as the normal doublet.

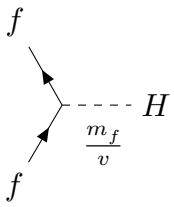


Figure II.6: Interaction vertex between fermions and the Higgs boson.

Points worth noting are:

- Only because the Higgs field has a non-zero vacuum expectation value v, a Yukawa Lagrangian with fermion mass terms can be constructed.
- This mechanism is fundamentally different from the mechanism that gives mass to the gauge bosons. The observation of the Higgs boson is therefore not a proof that fermion masses are caused by the Yukawa mechanism. The measurement carried out in this thesis therefore is a necessary complement to measurements of the well-established Higgs-to-gauge-boson couplings.
- For each Yukawa mass term, a corresponding Higgs-fermion interaction is predicted, figure II.6. The strength of the coupling to fermions is completely determined by only two parameters: the fermion mass and v. Therefore, if the fermion mass is known, one can predict the probability of the interaction shown in figure II.6. This is a very precise prediction, which naturally should be tested in order to decide if Yukawa interactions generate the masses of fermions.

Mass terms for the other fermion generations can be added using the same strategy. The Yukawa couplings y are free parameters, but can be calculated once the mass of a fermion is known. However, the Standard Model cannot explain why the masses of the fermions follow the hierarchy shown at the beginning of this chapter in figure II.1.

The predictions regarding the Higgs couplings have to be tested. The aim of this work is to test three different couplings, focussing on the Yukawa coupling to b quarks in $H \to b\bar{b}$ decays. $H \to b\bar{b}$ is the second-strongest Yukawa coupling, but the strongest that can be probed directly in decays: $H \to t\bar{t}$ decays are kinematically forbidden because the Higgs mass is lower than the top-quark mass. The other two couplings that are accessible in the $H \to b\bar{b}$ channel are the gauge coupling to the W boson in WH production and the gauge coupling to the Z boson in ZH production, see again figure II.5. Other production modes are more challenging because of high b-jet[9] backgrounds at the LHC. The $VH \to Vb\bar{b}$ analysis therefore has to test all three couplings simultaneously. Before the $H \to b\bar{b}$ analysis can be discussed in more detail, though, a closer look at the properties of the Higgs boson is necessary.

[9] A jet is a bundle of collimated particles, which are produced by outgoing quarks or gluons. This is a consequence of the fact that the only free particles permitted by QCD have to be colourless. The formation of jets will be discussed in section III 4.

II 5 The Higgs Boson

II 5.1 Discovery

In July 2012, the ATLAS and CMS collaborations announced the discovery of a new particle that is consistent with the Standard Model's Higgs boson [33, 34]. The main results from the ATLAS publication are shown in figure II.7. Signs of the Higgs boson are visible

- in the invariant mass spectrum of four leptons, the blue contribution in **(a)**,
- in the transverse mass distribution in the search for $H \to WW$ decays, the red contribution in **(b)**,
- and in the invariant mass spectrum of photon pairs, the peak in **(c)**.

Figure II.7(d) shows compelling statistical evidence for a new particle at a mass of $125\,\mathrm{GeV}$. The insets show:

(a) Exclusion limit. The upper limit at $95\,\%$ confidence level on the signal strength μ, see item (c).

(b) p-value. The probability that Standard Model processes without Higgs contributions would yield data that look equally or more "Higgs-like" than the measured data.

(c) Signal strength μ. The processes of interest are referred to as "signal". "Backgrounds" are other known processes with similar signature. In a counting experiment, the number of observed events can therefore be expressed in terms of the signal strength as:

$$N_{\text{total}} = \mu S + B,$$

where S and B are the expected numbers of signal and background events. Background-only Standard Model processes would therefore yield $\mu = 0$, whereas Higgs signal plus background processes yield $\mu = 1$.

All three metrics in figure II.7(d) show that the observed data are incompatible with the background-only hypothesis, and that a particle with a mass of $125\,\mathrm{GeV}$ has been found. Details on these metrics will be discussed in section section IV 5.3, which is dedicated to the statistical analysis of the data.

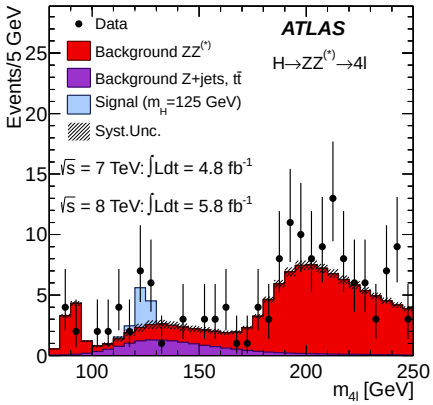
After the discovery, measurements provided evidence for a neutral spin-zero boson with $J^P = 0^+$ [35]. It was proven that the new particle couples to gauge bosons and photons, and evidence for vector-boson-fusion production was also found [36] (see section II 5.2 for details on production modes). Mass measurements [30] revealed

$$m_H = 125.09 \pm 0.21_{\text{stat}} \pm 0.11_{\text{syst}}\ \mathrm{GeV}.$$

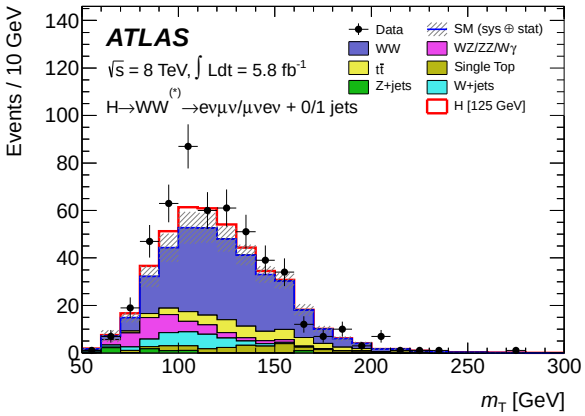
The 2013 Nobel Prize in physics was awarded to Englert and Higgs

> "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider."
>
> *Royal Swedish Academy of Sciences, 2013* [37]

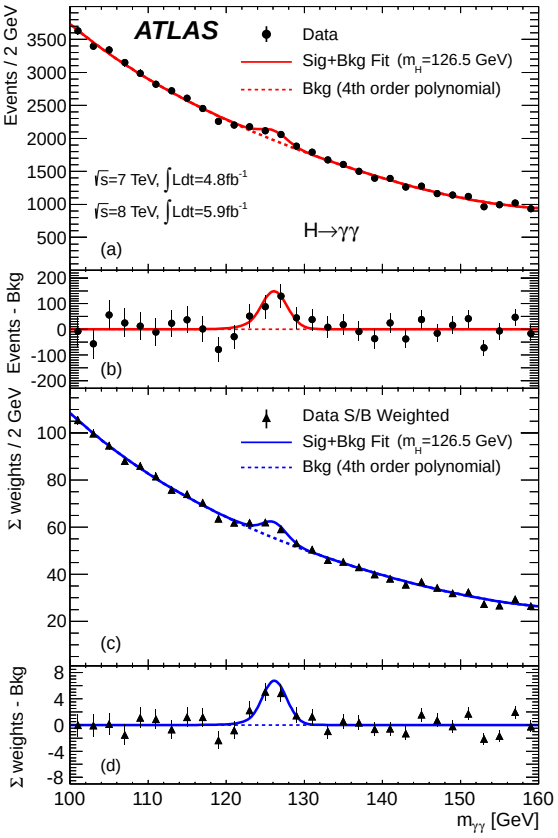
The properties of this newly found particle are in agreement with predictions for the Higgs boson. For this reason, this particle will be called Higgs boson throughout this thesis. It is understood that it remains unclear if *all* properties (couplings, for instance) are exactly as predicted by the Standard Model. Testing this is the main motivation for this work.



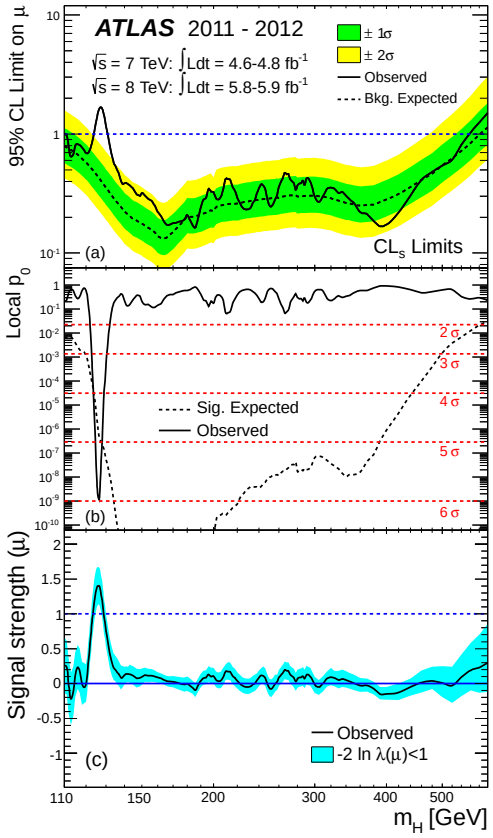
(a) $H \to ZZ$ channel



(b) $H \to WW$ channel



(c) $H \to \gamma\gamma$ channel



(d) Exclusion limit, p-value and signal strength (top to bottom)

Figure II.7: ATLAS results contributing to the Higgs discovery [33]. The figures **(a)** – **(c)** show results of searches for Higgs decays to gauge bosons. **(d)** shows the combination of the search channels. All channels observe an excess of events over the backgrounds predicted by the Standard Model.

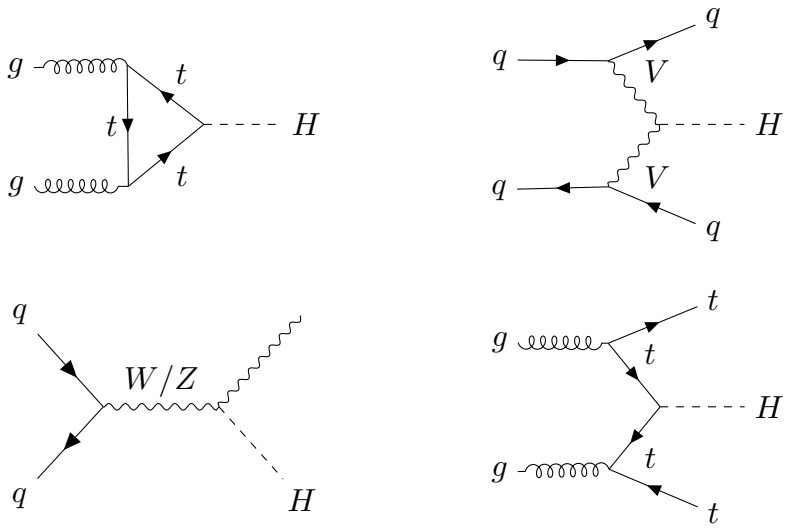


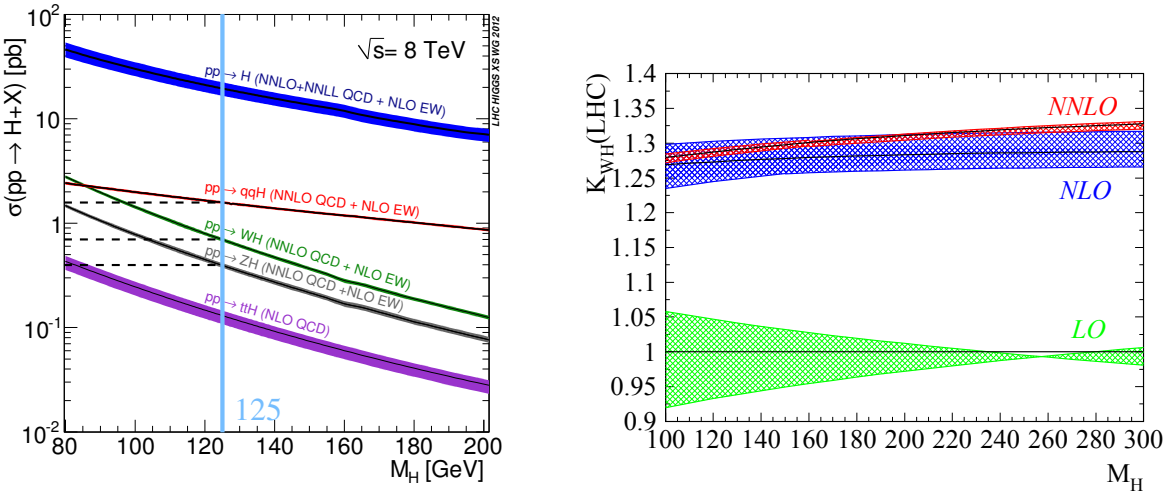
Figure II.8: Dominant Higgs production modes at LHC, ordered by cross section ($\sqrt{s} = 8\,\mathrm{TeV}$). From top left to bottom right: Gluon-gluon fusion (`ggF`) — Vector boson fusion (`VBF`), Associated production with vector bosons (`VH`) — Top fusion (`ttH`).

II 5.2 Higgs Production

With knowledge of the Higgs mass and given the possible couplings to bosons and fermions discussed in section II 3, all production and decay channels can be predicted. The dominant production processes at the Large Hadron Collider are shown in figure II.8. The corresponding cross sections are shown versus the Higgs mass in figure II.9(a). Since the LHC collides protons, which are composite particles, the most likely collisions to produce Higgs bosons are the ones induced by gluons. This is because gluons are the most common partons found in the proton. The details of this will be discussed in section III 1.1. Due to the dominance of gluons, about $88\,\%$ of all Higgs bosons found in the data of 2012 should be the result of gluon-gluon fusion (`ggF`, $21.4\,\mathrm{pb}$, the top left diagram in figure II.8 and most of the blue band in figure II.9(a)). Since the massless gluons do not couple to the Higgs boson, the process is mediated by a quark loop. The loop is mostly populated by top quarks because of their high masses, and therefore strong Higgs-Yukawa coupling. Other quarks are possible to couple as well. With a more than 10 times lower cross section, the fusion of vector bosons can produce a Higgs boson and two jets, which are caused by the scattered quarks shown in figure II.8 (`VBF`, $1.6\,\mathrm{pb}$, red band). With only about $3\,\%$ of the `ggF` cross section, W bosons can be produced and radiate Higgs bosons (`WH/Higgs Strahlung`, $701\,\mathrm{fb}$, green band). Similarly, Z bosons can radiate Higgs bosons (`ZH`, $420\,\mathrm{fb}$, grey band). The cross section is about $40\,\%$ lower than for WH production because the Z coupling to the quarks is suppressed by $\cos(\theta_W)$, compare equation (II.15) on page 14. Again three times less likely is the fusion of top quarks (`ttH`, $133\,\mathrm{fb}$, magenta). A similar process with fusion of b quarks exists ($202\,\mathrm{fb}$), but the experimental signature of the top quarks is identified more easily.

At the end of the LHC's Run 1, `ggF` and `VBF` had been observed, and evidence at the level of 4.4 standard deviations for ttH production was found by combining ATLAS and CMS results [40]. The same publication quotes signs for VH production at the level of 3.5 standard deviations, to which the analysis presented in this thesis, chapter IV, contributed.

Details concerning VH production cross sections can be found in [39, 41]. The calculations include higher-order diagrams, which alter the `LO` cross sections obtained by evaluating the diagrams in figure II.8.



(a) Higgs production modes at LHC ($\sqrt{s} = 8\,\mathrm{TeV}$). **(b)** k-factors for WH production [38].

Figure II.9: Higgs production cross-section predictions in dependence of the Higgs mass [39]. **(a)** Cross sections: An overlay shows the measured Higgs mass of $125.09\,\mathrm{GeV}$. **(b)** k-factors: Impact of the first NNLO QCD calculations ($\sqrt{s} = 14\,\mathrm{TeV}$). The k-factors are correction factors for the LO cross section to obtain cross sections at higher-order accuracy. These are nearly identical for $\sqrt{s} = 8\,\mathrm{TeV}$ and for the ZH process (except for gg-induced diagrams).

NLO corrections in QCD, for instance, increase the VH cross-section predictions by about 25 to 30 %, see figure II.9(b). They are therefore included in form of k-factors. The NLO diagrams comprise:

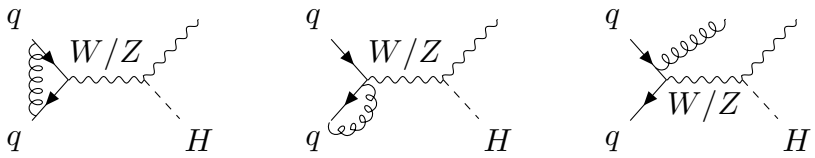


Figure II.10: NLO (QCD) diagrams included in the cross-section calculations for figure II.9(a).

Next-to-next-to-leading order corrections (NNLO, [38]) do not significantly increase the cross section of WH production, see figure II.9(b), but including them reduces the uncertainties of the predictions. The reason for this will be discussed in more detail in the section on Monte Carlo generators, III 4. NNLO diagrams do, however, have a particular impact on ZH production, because they permit a new production mode: gg-induced ZH production. The corresponding diagrams are similar to ggF, but a Z boson is produced after the quark loop, see figure II.11.

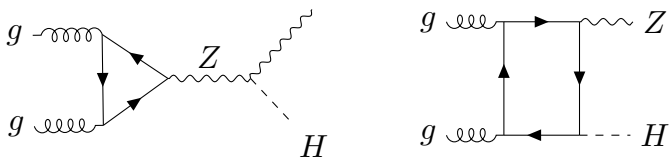


Figure II.11: NNLO (QCD) diagrams included in the cross-section calculations for figure II.9. The gg-induced ZH production is only taken into account when calculating VH at NNLO accuracy.

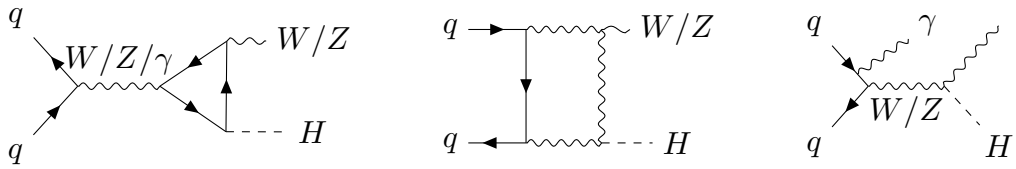


Figure II.12: NLO (EW) diagrams included in the cross-section calculations for figure II.9(a).

Not only QCD processes affect the cross sections: electroweak contributions are included up to NLO [42]. Typical diagrams are shown in figure II.12. The main effect of these diagrams is a reduction of the NNLO (QCD) cross-section predictions. The magnitude of this reduction reaches from 5 to 10 %, and depends on the transverse momentum of the Higgs boson. For the ZH search, this is taken into account by applying corrections to the LO Higgs transverse momentum spectrum that is predicted by Pythia8 [43].

II 5.3 Higgs Decays and Coupling Measurements

The Higgs boson is a heavy particle that couples to all massive particles. It therefore decays almost instantly with a lifetime of 1.6×10^{-22} s to lighter particles[10]. Searches for the Higgs boson can therefore only proceed by recording its decay products. Figure II.13 shows the branching fractions for various decay modes in dependence of the Higgs mass, and table II.1 lists the branching fractions for the observed mass of 125.09 GeV. The branching fractions are uniquely determined by the Higgs couplings, which depend only on the Higgs mass and the masses of the decay particles. Uncertainties of the branching fractions mostly stem from uncertainties of particle masses, and from uncertainties of α_s for loop-mediated decays. Although decays to first-generation fermions should be possible, the Yukawa coupling is so weak that these can be ignored.

The dominant Higgs decay is $H\to b\bar{b}$ (58 %). Despite the high branching fraction, it has not been observed. The analysis presented in this work finds an excess of 2.4 standard deviations – not enough to claim evidence for $H\to b\bar{b}$ decays. It shows, however, how systematic uncertainties in this channel can be reduced. This might benefit future updates of this analysis. Particularly, because the Run-2 successor of the ATLAS $VH\to Vb\bar{b}$ analysis [44] finds evidence for $H\to b\bar{b}$ decays at the level of 3.6 standard deviations, while CMS finds an excess of 3.3 standard deviations [45]. 3.0 and 2.8 standard deviations would have been expected from simulations. To claim an observation of $H\to b\bar{b}$ decays, five standard deviations would have to be reached. If only *statistical* uncertainties were relevant, a dataset with about three times the data statistics used for these two analyses would suffice. Both measurements are, however, limited by *systematic* uncertainties.

The second most abundant decay is $H\to WW$. W bosons couple more strongly to the Higgs than b quarks, but the decay is suppressed because $m_H < 2m_W$: at least one W boson must be virtual for the decay to be kinematically possible. Even though this decay is less likely than $H\to b\bar{b}$, decays to W bosons have been observed. This is because pairs of W bosons provide a more distinctive signature at the LHC [46, 47]. Decays to Z-bosons (2.6 %) are very similar to $H\to WW$ decays, but are even more strongly suppressed since the Z-mass is higher than the W mass. Nevertheless, these decays have been observed [48] since $ZZ\to 4l$ decays provide a very clean signature at hadron colliders. There is evidence for decays to τ leptons from ATLAS [49], and CMS has nearly reached the threshold to claim an observation: $4.9\,\sigma$ [50]. Decays to photons are very rare because they are mediated by a fermion or W loop. Nonetheless, $H\to\gamma\gamma$ was the driving channel for the Higgs discovery, since background levels in the diphoton final state are low at the LHC [51].

[10] From total width of 4 MeV [39], using $\tau=\hbar/\Gamma$.

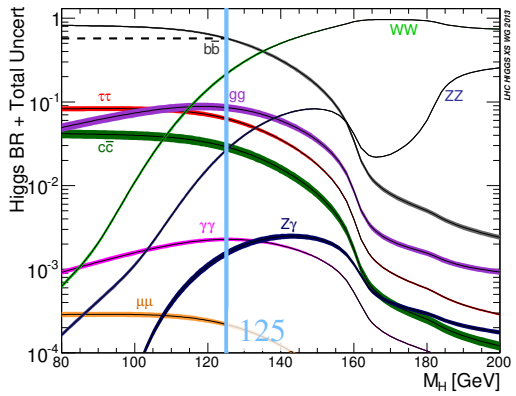


Figure II.13: Higgs decay branching fractions in dependence of m_H [39]

Decay	Coupling	Observed
58.0 % bb	Yukawa	
22.0 % WW	Gauge	✓
8.6 % gg	Yukawa/Gauge loop	
6.3 % $\tau\tau$	Yukawa	(✓)
2.9 % cc	Yukawa	
2.6 % ZZ	Gauge	✓
0.23% $\gamma\gamma$	Yukawa/Gauge loop	✓
0.15% $Z\gamma$	Yukawa/Gauge loop	
0.02% $\mu\mu$	Yukawa	

Table II.1: Higgs decay channels predicted for $m_H = 125.09\,\text{GeV}$ [41]

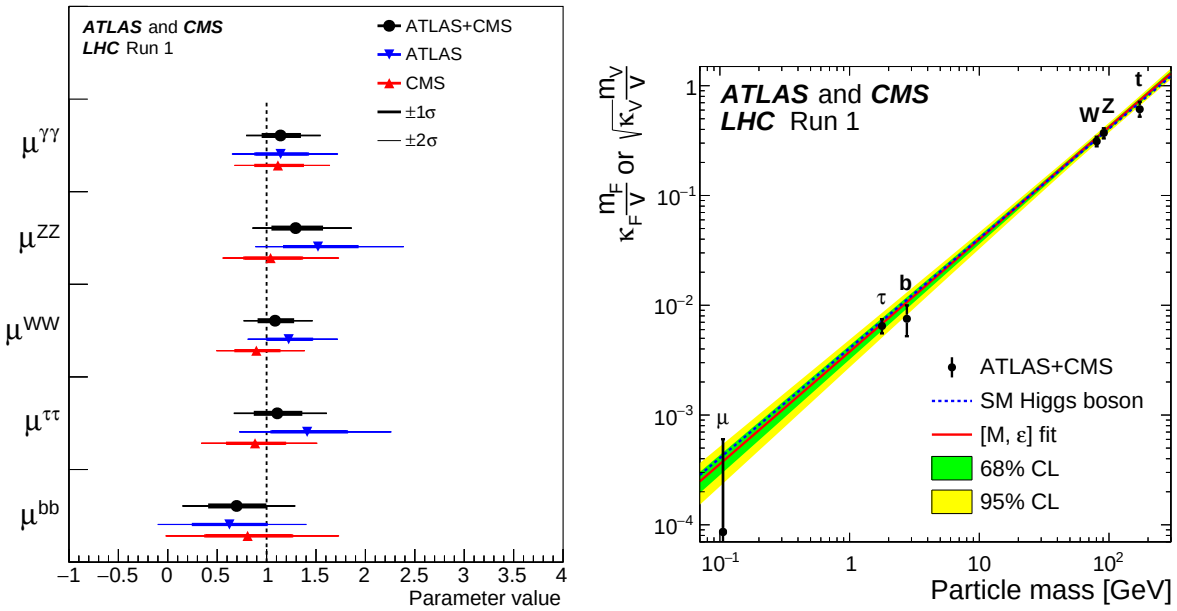
Decays that cannot be observed in the near future are decays to gg, cc, $Z\gamma$ and $\mu\mu$. For gg and cc, background levels at the LHC are too high: the signatures are two light jets for the former and two c-jets for the latter. Both are final states with much higher background levels than for the already challenging $H \to b\bar{b}$ decay, and the decay branching fractions are lower than for $H \to b\bar{b}$. The searches for the $Z\gamma$ [52] and $\mu\mu$ decays [53] are not sensitive because of the very low branching fractions.

ATLAS and CMS results concerning the couplings in the different final states and their corresponding signal strengths are summarised in [40], which also includes results from this thesis: the ATLAS $H \to b\bar{b}$ analysis discussed in chapter IV. Figure II.14(a) illustrates the most important Run-1 results concerning Higgs decays: all signal strength measurements are consistent with Standard Model predictions. Nonetheless, the decay channel that would profit the most from reduced uncertainties is $H \to b\bar{b}$ because the measurements of μ^{bb} are not accurate enough to prove or exclude the existence of these decays.

Figure II.14(b) shows the mass dependence of the measured couplings: the κ-parameters factor out the difference between gauge and Yukawa couplings in order to display the mass dependence in a single plot. Unlike the signal strength μ, which focusses on final states, the κ-parameters measure the couplings across different production channels and different final states. All the couplings align with Standard Model predictions, even though more precision on κ_b, in particular, would be desirable.

$H \to b\bar{b}$ and $H \to \tau\bar{\tau}$ are of marked interest since these are the only decay channels with notable sensitivity to the Yukawa couplings of the Higgs boson. This thesis is contributing to one of the outstanding problems in the Higgs boson research since all accessible couplings to gauge bosons have been proven, while for Yukawa couplings there is evidence at best, for example, in the $H \to \tau\bar{\tau}$ channel. The only other process that is sensitive to Yukawa couplings is `ttH` production, but also here there is only evidence but no proof at the level of five standard deviations.

To conclude the discussion of the importance of the Higgs boson, one should point out that even though the discovery of the Higgs boson was one of the most important milestones of modern particle physics, the measurement of its properties remains challenging. Especially the measurement of the $H \to b\bar{b}$ Yukawa coupling turned out to lack precision in comparison with other channels. Improved sensitivity in $H \to b\bar{b}$ searches will yield certainty about the existence of this coupling with smaller datasets, that is, in the nearer future. In this respect, the results of chapter V are interesting because this thesis shows how systematic uncertainties of the $VH \to Vb\bar{b}$ search can be reduced.



(a) Signal strength in different decay channels

(b) Higgs couplings versus particle mass

Figure II.14: ATLAS and CMS Higgs results at the end of the LHC's Run 1 [40]. **(a)** Measured signal strength in the most sensitive decay channels. **(b)** Higgs couplings versus particle mass. The measured couplings κ are normalised to remove the difference between Yukawa and gauge couplings: $\kappa_f \frac{m_f}{v}$ for fermions, $\sqrt{\kappa_V}\frac{m_V}{v}$ for gauge bosons. A fit to the observed data is superimposed [54], which is consistent with the prediction that the Higgs couplings are proportional to the fermion mass and the squared boson mass. All results are consistent with Standard Model predictions, but improvements in the $H \to b\bar{b}$ channel are needed the most to better test the Standard Model coupling predictions.

CHAPTER III

Particle Physics at Hadron Colliders

In this chapter, important aspects of physics at hadron colliders will be reviewed to understand the experimental constraints for the ATLAS $H \to b\bar{b}$ analysis (chapter IV). The discussion will start with the experimental tools, LHC and ATLAS, review the reconstruction of an LHC collision in ATLAS, and finish with an overview of how such collisions are simulated.

III 1 The Large Hadron Collider

In the search for new particles, particle accelerators proved to be the most successful tools. A multitude of particles, such as antiproton (1955), ν_μ (1962), charm quark (1974), τ (1975), bottom quark (1977), gluon (1979), W/Z bosons (1983), top quark (1995), ν_τ (2000) and Higgs boson (2012), were all detected in collisions produced with particle accelerators. A key parameter of an accelerator is the centre-of-mass energy $E_{\text{cm}} = \sqrt{s}$. It determines which kinds of particles can be produced in a collision. s is one of the Mandelstam variables: in a reaction $1 + 2 \to 3 + 4$, s is the squared energy available in the initial state or – because of energy conservation – in the final state:

$$s = (\boldsymbol{p}_1 + \boldsymbol{p}_2)^2 = (\boldsymbol{p}_3 + \boldsymbol{p}_4)^2 = M^2.$$

The four-vectors $\boldsymbol{p}_i$ denote the four-momenta of the particles in the beams or final state. If the particles 3 and 4 have a common ancestor, its mass is $M = E_{\text{cm}}$. Particles with masses above E_{cm} cannot be produced. Since the centre-of-mass energies reached by accelerators increased over time, more and more particles with higher masses could be discovered.

To date, the Large Hadron Collider (LHC) [56] is the most powerful accelerator. The LHC is a proton collider at CERN, the European laboratory for particle physics in Geneva, Switzerland. It was built in a tunnel with a circumference of $26.7\,\text{km}$, see figure III.1, which previously hosted the Large Electron Positron Collider (LEP), where W and Z bosons were studied. The LHC is the last stage in a chain of accelerators that accelerates proton beams up to energies of $7\,\text{TeV}$ per beam. Protons are extracted from hydrogen gas, and accelerated in a linear accelerator to an energy of $50\,\text{MeV}$, in bunches of about $1.7{\times}10^{11}$ protons. These bunches are subsequently injected into the booster synchrotron ($l = 157\,\text{m}$, $E = 1.4\,\text{GeV}$), into the Proton Synchrotron PS ($l = 628\,\text{m}$, $E = 25\,\text{GeV}$), the Super Proton Synchrotron SPS, where W and Z bosons were discovered ($l = 7\,\text{km}$, $E = 450\,\text{GeV}$), and finally into the LHC. The LHC is currently running at energies of $6.5\,\text{TeV}$ per beam. In 2012, where the data for this thesis were recorded, the energy was $4\,\text{TeV}$ per beam. These beam energies correspond to centre-of-mass energies of $\sqrt{s} = 13$ and $8\,\text{TeV}$.

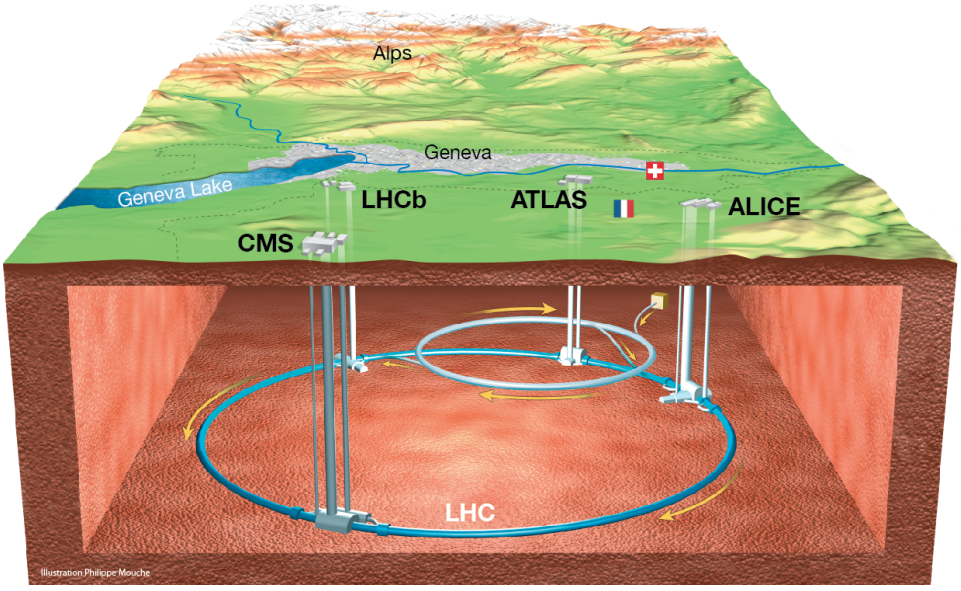


Figure III.1: Overview of the LHC and the SPS [55]. The LHC is a proton-proton collider at CERN, Geneva. It is the last accelerator in a chain of accelerators designed to accelerate protons up to a centre-of-mass energy of $14\,\mathrm{TeV}$. Four large experiments record the particle collisions. The data for this thesis was recorded by ATLAS.

In the LHC, particles circulate in two ultra-high-vacuum tubes with pressures of $1\times10^{-13}\,\mathrm{bar}$ to avoid collisions with gas molecules inside the tube. $1\,232$ superconducting dipole magnets operating at $1.9\,\mathrm{K}$ and currents of up to $13\,\mathrm{kA}$ provide a bending field of up to $8.33\,\mathrm{T}$, which keeps the particles on a closed orbit. The magnets are arranged in eight arcs with a bending radius of $2\,804\,\mathrm{m}$ each. With $q\vec{v}\times\vec{B}=\frac{mv^2}{R}$, the maximum beam energy can be estimated:

$$E_{\mathrm{Beam}}\approx p=qRB=\mathrm{e}\cdot 2\,804\,\mathrm{m}\cdot 8.33\,\mathrm{T}=7\,\mathrm{TeV}. \tag{III.1}$$

The maximum current that can be tolerated by the magnets thus defines the beam energy.

The eight arcs with bending magnets are interleaved with straight sections. In four of these straight sections, the proton beams are focussed and collided. The collision points are surrounded by seven particle detectors: four large and three small experiments. Two of the large experiments use multi-purpose detectors, `ATLAS` [57] and `CMS` [58], which focus on a broad spectrum of proton-proton physics such as the search for the Higgs boson, searches for new physics processes or the measurement of the W mass. The third experiment, `LHCb` [59], is dedicated to b quark physics, and the fourth, `ALICE` [60], to heavy ion physics, which takes most of its data when the LHC collides lead nuclei. Two of the small experiments, `LHCf` and `TOTEM`, focus on physics with protons scattered in forward direction, and are installed along the LHC beam line near ATLAS and CMS. `MoEDAL`, installed near LHCb, searches for magnetic monopoles.

The LHC was commissioned in 2008, but had to be shut down for repairs until 2009. The first sizeable dataset was recorded in 2011 at $\sqrt{s}=7\,\mathrm{TeV}$. In 2012, the centre-of-mass energy was increased to $\sqrt{s}=8\,\mathrm{TeV}$. The runs of these two years are collectively referred to as "Run 1". From 2013 until 2015, upgrades were installed to reach a centre-of-mass energy of $\sqrt{s}=13\,\mathrm{TeV}$. The current "Run 2" is foreseen to end in 2018.

The LHC was designed to prove whether the Higgs boson exists or not because its predecessors, `LEP` and `Tevatron` had neither been able to confirm nor exclude its existence. For LEP, this was due to a lack of energy: being an electron-positron collider, the main production channel is ZH production like in

figure II.8, page 24. LEP reached a centre-of-mass energy of 209 GeV, whereas at least $m_Z + m_H = 216.3\,\mathrm{GeV}$ would have been necessary to produce both Z and H. The Tevatron, a proton-antiproton collider with $\sqrt{s} = 1.96\,\mathrm{TeV}$, had ample energy to produce the Higgs boson, but lacked `luminosity`.

III 1.1 Luminosity and PDFs

The luminosity determines the event rate of a given reaction produced by an accelerator:

$$\begin{aligned} \dot{N} &= \sigma\,\mathscr{L} \\ N &= \sigma \int \mathscr{L}\,\mathrm{d}t \end{aligned} \tag{III.2}$$

N is the number of events, σ the cross section of the reaction, $\mathscr{L}$ is the instantaneous luminosity and $\int \mathscr{L}\,\mathrm{d}t$ the integrated luminosity. The lack of luminosity was the reason that the Tevatron experiments were unable to discover the Higgs boson because statistical uncertainties in counting experiments scale as $\sqrt{N}$. The luminosity of an accelerator can be estimated from the following parameters:

$$\mathscr{L} = \frac{n N_1 N_2 f}{4\pi \sigma_x \sigma_y} \tag{III.3}$$

n is the number of colliding bunches, N_i is the number of particles per bunch, f is the repetition rate, and $\sigma_{x,y}$ is the lateral extension of the bunches assuming the particle density is a Gaussian distribution. $\mathscr{L}$ therefore has the units of inverse area-seconds, and is most conveniently expressed in $\mathrm{fb}^{-1}\,\mathrm{s}^{-1}$ because cross sections in equation (III.2) are often also of the order of fb or pb.

The LHC accelerates particles using eight superconducting radio-frequency cavities operated at 400 MHz, which generate accelerating fields of up to $5\,\mathrm{MV\,m}^{-1}$. The accelerator is designed to hold a bunch of particles in every tenth RF-cycle (the "RF bucket"), leading to a maximal collision rate of 40 MHz, corresponding to 25 ns between collisions. The LHC can hold up to 2 808 colliding bunch pairs with $N_1 = N_2 \approx 1.15{\times}10^{11}$ protons. With revolution frequencies of 11 200 per second and a lateral extension of about 20 µm, this yields an instantaneous luminosity of $\mathscr{L} = 1{\times}10^{34}\,\mathrm{cm}^{-1}\,\mathrm{s}^{-1}$. The lateral extension of the luminous region in ATLAS is actually about 60 µm, but equation (III.3) does not take into account the beam crossing angle and that the beam profile is not Gaussian.

By contrast, the Tevatron reached $N_p \approx 2.8{\times}10^{11}$, $N_{\bar{p}} \approx 0.8{\times}10^{11}$, with only 36 colliding bunch pairs at the end of its lifetime, because antiprotons have to be reproduced constantly [61]. The Tevatron therefore accumulated a dataset of approximately $10\,\mathrm{fb}^{-1}$ during its lifetime from 1986 to 2011. The integrated luminosity delivered to ATLAS during the LHC's Run 1 is $5.5\,\mathrm{fb}^{-1}$ (2011) and $22.7\,\mathrm{fb}^{-1}$ (2012), at $\sqrt{s} = 7$ and 8 TeV [62, 113]. The number of bunches was increased gradually to 1 400 at the end of 2012. Bunch currents were also increased gradually, which is why the dataset accumulated in 2012 is four times larger than the dataset of 2011. Due to data-taking inefficiencies, a slightly smaller dataset of $4.7\,\mathrm{fb}^{-1}$ for 2011 and $20.2\,\mathrm{fb}^{-1}$ for 2012 is usable for physics analyses. It was therefore both the higher energy and the higher luminosity that ensured that the LHC would be able to either confirm or exclude the Higgs boson.

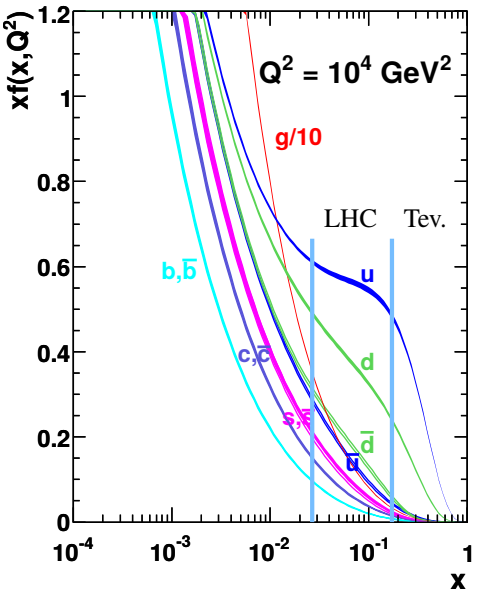


Figure III.3: Parton distribution functions (PDFs) for protons [63]. The minimal x for WH production at rest is marked for the LHC and the Tevatron.

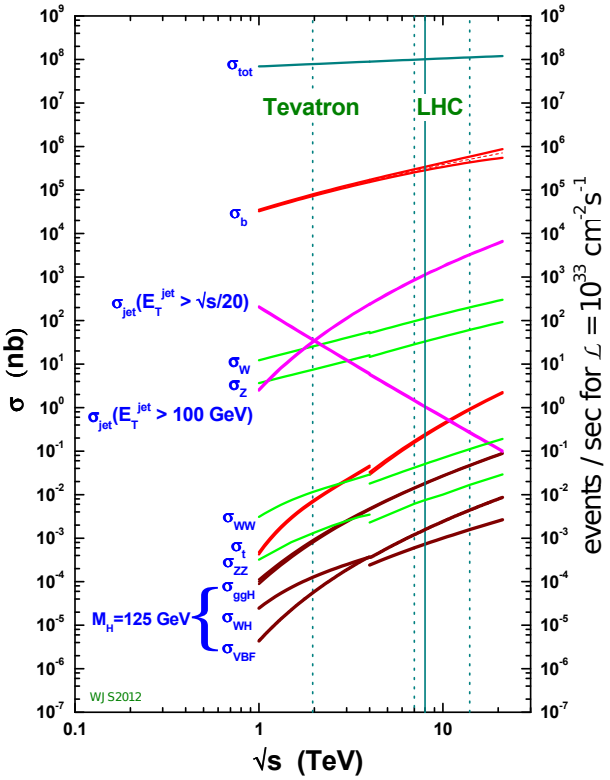


Figure III.4: Cross sections of Higgs and various background processes in dependence of the centre-of-mass energy [64]. The increase of cross sections with higher centre-of-mass energy is caused by the rise of PDFs towards lower x.

Parton Luminosity and Parton Distribution Functions

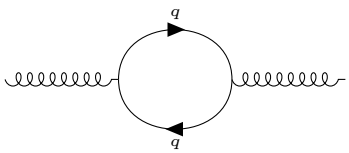


Figure III.2: A gluon temporarily splitting to quarks.

Although the simple discussion of luminosity and cross sections from the previous section is sufficient to describe reaction rates for an e^+e^- collider like LEP, it misses an important aspect of hadron colliders: hadrons are composite particles. When the LHC, for example, collides protons, not only whole protons interact, but also collisions between "partons" occur. Partons are the quarks and gluons found inside hadrons. The proton consists of valence quarks, two up and one down quark, which define the quantum numbers of the proton such as its electric charge. The quarks are bound by exchanging gluons, which also take part in parton collisions. The gluons can temporarily split into quark-antiquark pairs, the "sea quarks", via diagrams like the one shown in figure III.2. Gluons and sea quarks carry a significant fraction of the proton momentum, and thus have to be considered in calculations of reaction rates. The fraction of the proton momentum that is carried by a parton is described by the Bjorken x-variable:

$$p_{\text{parton}} = x \cdot p_{\text{proton}}, \qquad 0 \leq x \leq 1.$$

The probability to find a parton with a certain momentum fraction x inside the proton is described by the parton distribution functions (PDF). They depend on the parton type and a momentum scale Q^2.

Figure III.3 shows the MSTW proton PDFs [63] at a momentum scale of $Q^2 = 1{\times}10^4\,\mathrm{GeV}^2$. For the LHC, this is a moderate momentum transfer of 100 GeV. The probability densities in the figure show that the partons with the highest values of x at this momentum scale are valence quarks. Since the proton has two up (valence) quarks, the u-quark probability density is higher than for the d quark. At $x \lesssim 0.1$, however, gluons completely dominate among all partons: note that the gluon PDF is scaled down by a factor of 10 for better visibility. The rest of the proton momentum is carried by the sea quarks. Because of the dominance of gluons, the LHC is sometimes called a gluon collider. This explains why gluon-initiated Higgs production dominates all other Higgs production channels at the LHC, compare section II 5.2. The LHC provides a very high "parton luminosity" for gluons.

The production channel in the focus of this thesis, Higgs production in association with vector bosons, is dominantly produced in collisions of a quark and an antiquark. In proton collisions, the latter can only be a sea quark. Only at higher orders in α_s ($\geq$NNLO), VH can also be initiated by gluons, see figure II.11 on page 25. Since sea quarks have sizeable probability densities only at low x, the beam energy, and thus the parton luminosities, have a strong influence on production cross sections. This is illustrated by the two markers that have been added to figure III.3. The markers show the minimal x that is required to produce $W^+ + H$ at rest. x for such a collision can be estimated by calculating the minimal s required to produce the particles:

$$\begin{aligned} & s' = (x_1 \cdot E_{\text{beam1}} + x_2 \cdot E_{\text{beam2}})^2 = (m_W + m_H)^2 = (205\,\mathrm{GeV})^2 \\ \Rightarrow \quad & x_{\text{LHC}} = \frac{m_W + m_H}{2E_{\text{beam}}} = 0.026\,, \qquad x_{\text{Tevatron}} = 0.105 \end{aligned}$$

To reach the same value of s', LHC collisions require lower values of x than Tevatron collisions. The LHC therefore operates in a regime of higher parton densities as shown by figure III.3. The parton luminosity for a given reaction strongly increases with the beam energy because the PDFs rise quickly towards low x. When allowing asymmetric collisions with $s' > (205\,\mathrm{GeV})^2$ and $x_1 \neq x_2$, the sea quark PDF at the LHC rises most quickly – much faster than any relevant Tevatron-PDF. It is therefore not only the higher (proton) luminosity, but also the higher parton luminosity in comparison to the Tevatron that lead to the discovery of the Higgs boson at the LHC.

Figure III.4 shows NLO production cross-section calculations after summation over all initial states and integration over all possible x and the MSTW PDFs. The impact of the rising beam energy on the cross sections is clearly visible. The discontinuity at $\sqrt{s} = 4\,\mathrm{GeV}$ is the result of changing from proton-antiproton to proton-proton PDFs. The dark graphs at the bottom show Higgs production cross sections for ggF, VBF and WH production. ZH production is not shown, but it is about on order of magnitude less likely than WH production. Already at $\sqrt{s} = 7\,\mathrm{TeV}$, the LHC outperforms the Tevatron in all production processes, often by an order of magnitude.

The fact that for the LHC the most likely process for WH or ZH production is the collision of a valence quark with a sea quark also has consequences for the kinematics of VH events: the most likely collision is a collision of a valence quark with high x and a sea quark with low x. In most collisions, the VH system will therefore be boosted along the initial direction of the valence quark. At the Tevatron, the boost will be low on average. The high boost at the LHC is visible in the collision shown in figure III.12 on page 49, and will also be relevant for defining observables to select VH events in chapter V.

III 2 Collider Detectors: The ATLAS Experiment

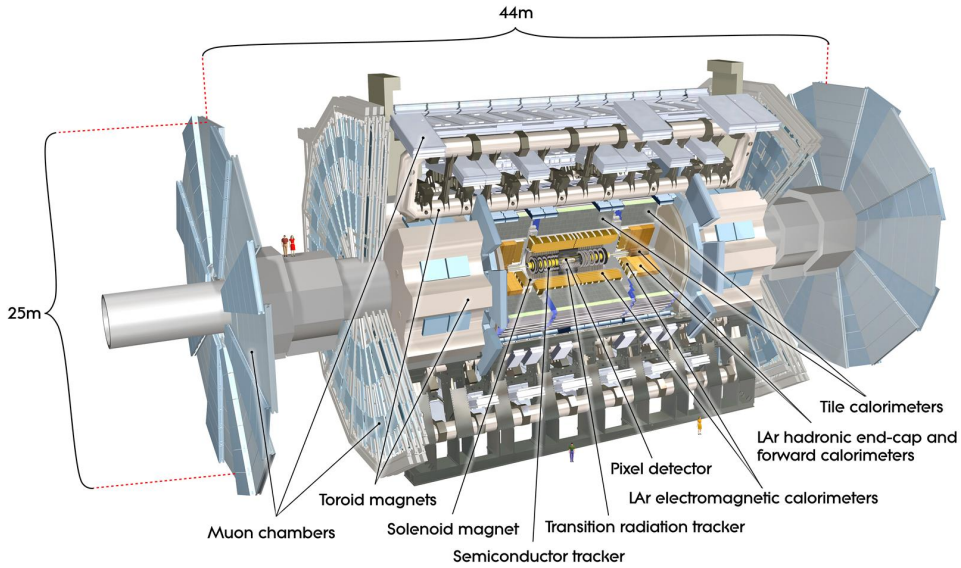


Figure III.5: The ATLAS detector, a multi-purpose detector at the LHC. Collisions occur in the centre of the detector, and are recorded by the following subsystems, from inside to outside: tracking detectors (dark grey), calorimeters (orange and grey), muon system (blue). A solenoid magnet enclosing the tracking detectors and toroid magnets (light grey, muon system) provide magnetic fields for the measurement of charged-particle momenta. Image from [65].

ATLAS [57] in figure III.5 is a barrel-shaped multi-purpose detector at CERN's Large Hadron Collider. ATLAS was designed to detect particles produced in LHC collisions at collision energies of up to $14\,\mathrm{TeV}$ with repetition rates of up to $40\,\mathrm{MHz}$. It covers a solid angle of nearly 4π around the interaction region. The central region is called "barrel" and the regions in forward/backward directions "endcaps".

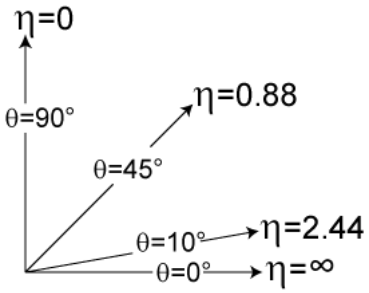


Figure III.6: Pseudorapidity [66]

ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis[1]. The polar angle θ is expressed in terms of the pseudorapidity:

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

$\eta = 0$ is the transverse direction, $\eta = \pm\infty$ is aligned with the beam axis, see sketch in figure III.6. The pseudorapidity is better suited than the polar angle to describe the angular distributions of scattered particles, since most particles are scattered into directions close to the beam axis. With only a few particles scattered into the transverse directions, angular distributions in θ peak close to the beam axis, and are only sparsely populated in transverse direction. The pseudorapidity, by contrast, has an almost flat distribution.

The use of cylindrical coordinates has one more advantage: in a given collision, the momentum fractions of the partons, x, and therefore also the z-momentum of the final state cannot be predicted. However, the sum of the transverse momenta is known to be zero due to the conservation of momenta.

[1] This is the recommended description of the coordinate system for ATLAS papers [67].

The transverse and longitudinal components of the particle momenta are therefore measured separately, which favours cylindrical coordinates. In this way, all transverse quantities remain independent of the longitudinal boost. The Lorentz-invariant analysis in chapter V will even go one step further and make the event selection insensitive also to transverse boosts.

Cylindrical coordinates are also beneficial because ATLAS uses a magnetic field aligned with the beam axis to bend the tracks of charged particles. Measuring the track curvatures yields the transverse momentum components:

$$\vec{p}_{\mathrm{T}} = \begin{pmatrix} p_x \\ p_y \end{pmatrix} = \begin{pmatrix} p_{\mathrm{T}} \cos\phi \\ p_{\mathrm{T}} \sin\phi \end{pmatrix}$$

The longitudinal momentum component is not measured directly because a magnetic field that bends particles in the longitudinal direction would also deflect the beams. The longitudinal component can, however, be calculated using p_{T} and η. From $p_{\mathrm{T}}, \eta, \phi$, one obtains the four-vectors of the particles (neglecting their rest mass):

$$p^{\mu} = \begin{pmatrix} p_{\mathrm{T}} \cosh(\eta) \\ p_{\mathrm{T}} \cos(\phi) \\ p_{\mathrm{T}} \sin(\phi) \\ p_{\mathrm{T}} \sinh(\eta) \end{pmatrix} \tag{III.4}$$

Details on this parametrisation can be found in appendix A 3, page 205.

Angular distances are measured as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}\,.$$

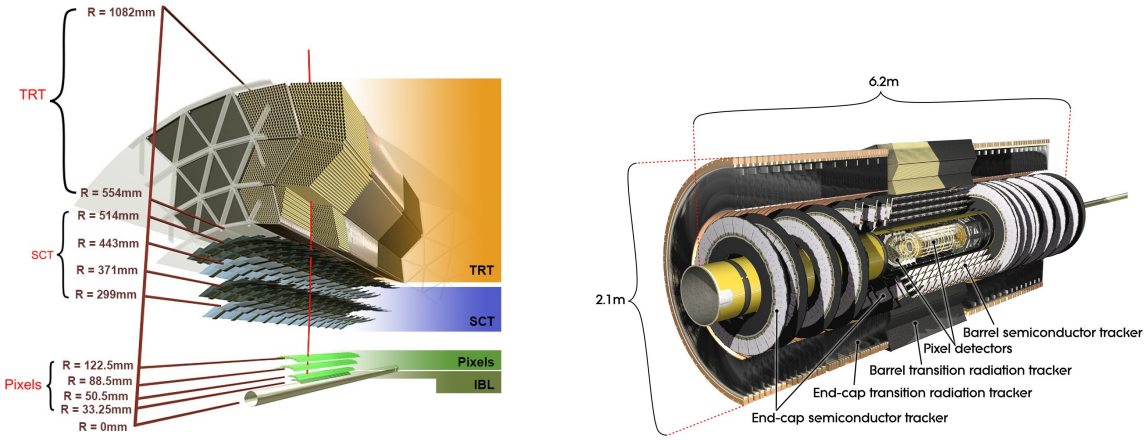
Due to a rotational symmetry, distributions in ϕ are flat, and distributions in η are nearly flat as explained earlier. It is therefore reasonable to measure distances between objects as the magnitude of a vector in the η-ϕ plane. ΔR will be used frequently to measure distances between objects, define the radius of jet cones or to determine if objects are isolated from other activity in an event.

III 2.1 Tracking Detectors: Measuring the Momenta of Charged Particles

The subdetectors of ATLAS will now be discussed from inside to outside: the innermost layers of ATLAS are tracking detectors. These surround the beam pipe, in which the protons are colliding. When charged particles traverse the layers of the tracking detectors, they induce electric charges, which are collected and digitised. This yields "hits", that is, information on which parts of the detectors have been traversed by a charged particle. The hits are connected by the reconstruction software, yielding the trajectories of the particles.

Figure III.7 shows the layout of the tracking detectors. Starting in the centre of the detector, one finds the following subdetectors:

Pixel detector Three layers of pixel sensors and three discs in the forward direction host 80 million silicon pixels with a size of 50 x $400\,\mu\mathrm{m}^2$ distributed over a surface of $1.7\,\mathrm{m}^2$. The pixels yield hit information with a position resolution of $14\,\mu\mathrm{m}$ in transverse and $115\,\mu\mathrm{m}$ in beam direction. High spatial resolution close to the beam pipe is paramount in order to be able to extrapolate the particle trajectories into the interaction region, where no detectors can be installed. Therefore, detectors are as close as $5\,\mathrm{cm}$ to the colliding beams. If all layers are hit, the pixel detector provides three high-resolution points on the particle trajectories.
Figure III.7(a) shows the pixel layers as green surfaces close to the beam pipe. It additionally shows a fourth layer, `IBL`, at a distance of $3.3\,\mathrm{cm}$ from the interaction region. Since IBL was installed after the end of Run 1, it is not relevant for this thesis.



(a) Expanded view of the barrel region [68]

(b) Inner detectors to scale including endcap discs [69]

Figure III.7: Inner detectors (ID) surrounding the interaction region of ATLAS. **(a)** Expanded view of the different detector layers. From inside to outside: the "Insertable B-Layer" IBL (since Run 2), three layers of pixel detectors, four layers of the semiconductor tracker SCT and the transition radiation tracker TRT. The track of a charged particle is indicated in red. **(b)** Discs of pixels and SCT modules to extend the acceptance of the tracking detectors into the forward direction up to $|\eta| = 2.5$.

Semiconductor Tracker SCT consists of four barrel layers and 18 endcap discs hosting $60\,\mathrm{m}^2$ of silicon sensors. The sensors are read out using readout strips spaced at a distance of $80\,\mu\mathrm{m}$, allowing for a position resolution of $17\,\mu\mathrm{m}$ transverse to the strips. The layers are installed at a small crossing angle, providing position information also in direction of the readout strips with a resolution of about $600\,\mu\mathrm{m}$. The layers of double-sided barrel modules provide up to eight points on a particle trajectory. The endcap discs and barrel modules are shown in figure III.7(b).

Transition Radiation Tracker TRT is the outermost tracking detector, consisting of 300 000 straw tubes with a diameter of $4\,\mathrm{mm}$. A tungsten wire in the centre of each straw detects charges induced in the gas-filled tubes. TRT does not provide the z-coordinate of the particles, but can provide $\gg 10$ hits with transverse position resolution of $170\,\mu\mathrm{m}$. When electrons traverse TRT, they induce transition radiation, that is, they radiate photons. This creates much more charge than for heavier particles, and is used to separate electrons from other particles.

The inner detector is surrounded by a superconducting solenoid magnet. It provides a nearly uniform magnetic field of $2\,\mathrm{T}$ aligned with the beams, bending the tracks of charged particles in the transverse plane. The momentum[2] in the transverse plane can be computed from the track curvature using the Lorentz force as for the LHC bending magnets, equation (III.1). The charge sign is determined from the bending direction. The resolution of transverse momentum measurements is [70]:

$$\frac{\sigma_{p_\mathrm{T}}}{p_\mathrm{T}} \approx 5\times10^{-4} \cdot \frac{p_\mathrm{T}}{\mathrm{GeV}} + 0.01 \tag{III.5}$$

The equation shows that the resolution is lower for high-momentum tracks. This is because these are nearly straight, which degrades the measurement of the bending radius. For an electron with $p_\mathrm{T} = 100\,\mathrm{GeV}$, one can expect a transverse momentum resolution of $\sim 6\,\%$, but $\sim 2\,\%$ for $p_\mathrm{T} = 25\,\mathrm{GeV}$.

[2] Strictly speaking, it is $p_\mathrm{T} \cdot q$, which is measured by ATLAS. Since virtually all particles with long-enough lifetimes carry only a single elementary charge, this description is nevertheless accurate.

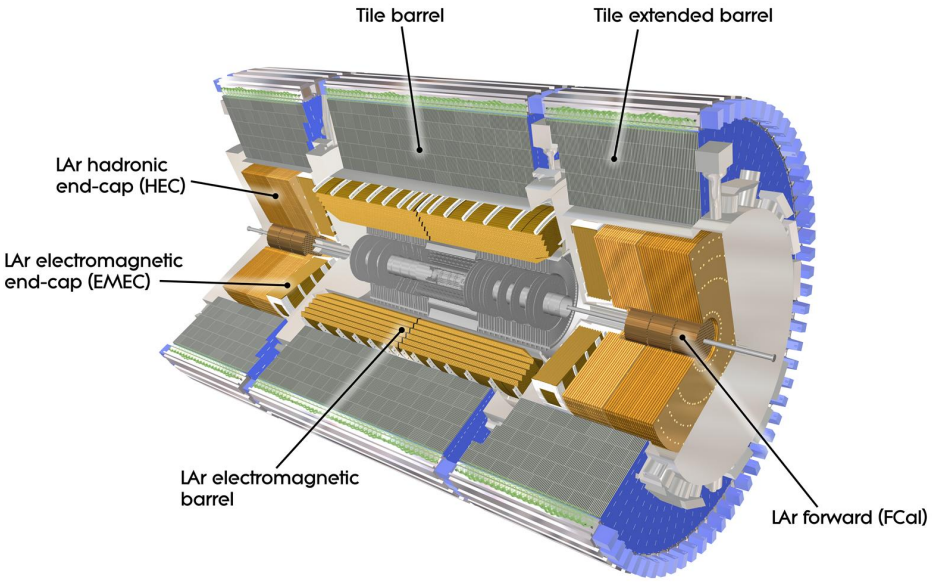


Figure III.8: Calorimeters of the ATLAS detector. The inner orange calorimeter is the electromagnetic liquid argon calorimeter `ECal/LAr`, which detects photons, electrons and positrons. The outer tile calorimeter `HCal/Tile` detects hadrons. Multiple forward calorimeters surround the beam pipe, leading to a total calorimeter coverage of $-5 < \eta < 5$ (angular distance of ~$0.8°$ from the beam axis) [71].

The inner detector covers a pseudorapidity range of $-2.5 < \eta < 2.5$, with full coverage in ϕ. It is one of the most important components for the $H \to b\bar{b}$ analysis: since accurate track information is indispensable to identify jets from b quarks, all events must have jets within the acceptance of the tracking detector in order to be usable for this analysis. Events with "forward jets", that is, jets outside the tracker acceptance, have to be discarded. The identification of b-jets will be discussed in section III 3.4. The tracking detector is also indispensable for the identification of leptons produced in vector boson decays, because such leptons are expected to be isolated from other activity in an event. The tracking detector can detect such additional activity, and thus it helps to identify the VH topology.

III 2.2 Calorimetry: Measuring the Energy of Particles and Jets

Calorimetry is indispensable for detecting outgoing neutral particles because these entirely escape detection in the tracking systems. Therefore, the tracking detectors of ATLAS are enclosed by calorimeters, see figure III.8. Furthermore, when particles with net colour charge are produced in an interaction, they form jets – collimated bundles of hadrons. Single hadrons are most of the time not of interest, contrary to the properties of the particle that sparked the jet. These properties, for example the energy of the initial particle and the amount of hadronic activity, are measured with calorimeters. ATLAS uses four types of calorimeters:

Liquid Argon EM Calorimeter (LAr ECal) The innermost calorimeter is an electromagnetic calorimeter, figure III.9. Electromagnetic calorimeters measure the energy of primarily electromagnetically interacting particles: photons, electrons and positrons[3]. When interacting with the material in

[3] Other particles, of course, also interact electromagnetically, but they are usually not fully absorbed in the EM calorimeter.

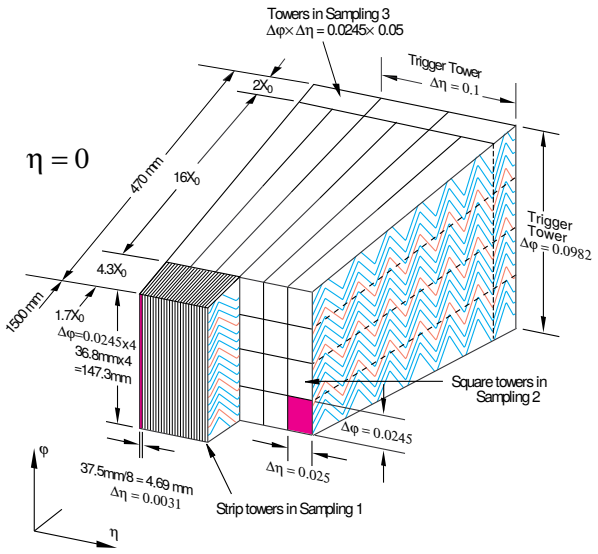


Figure III.9: The ATLAS EM calorimeter. Accordion-shaped liquid argon cells separated by layers of lead measure charge deposited by EM showers. Three layers with different granularities are used to determine shower shapes [72].

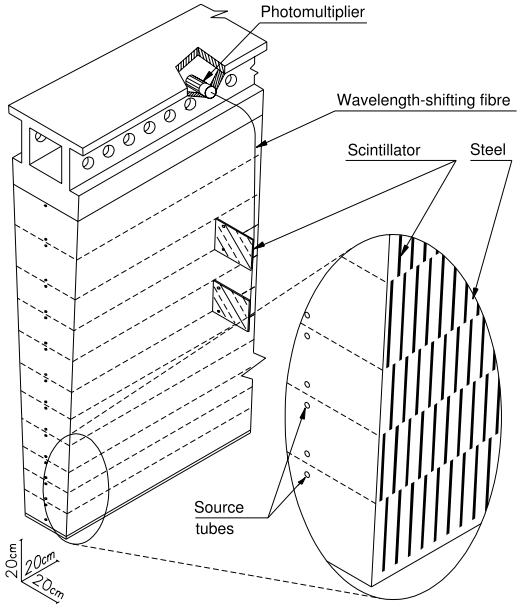


Figure III.10: Wedge of the Tile Calorimeter. Steel as absorber is interleaved with small scintillating tiles. The tiles are connected to photomultiplier tubes for readout [73].

the EM calorimeter, photons convert into electron/positron pairs (pair creation). When electrons and positrons traverse the EM calorimeter, they radiate photons when scattered off hull electrons (Bremsstrahlung). Such photons can again create e^+e^--pairs, which creates showers of particles in the calorimeter. Shower evolution stops when the energy of Bremsstrahlung-photons falls below the pair-creation threshold of $2m_e = 1.022\,\mathrm{MeV}$. The remaining low-energy particles ionise the material of the calorimeter, and the energy of the incoming particles can therefore be measured by collecting the charge deposited in the calorimeter.

The ECal of ATLAS is a sampling calorimeter: it consists of accordion-shaped cells filled with liquid argon, the "active material", where electric charge is collected. The liquid argon cells are interleaved with layers of lead where electromagnetic showers are induced, the "passive" or "absorber material". The energy resolution of the ATLAS EM calorimeter is approximately [13]:

$$\frac{\sigma_E}{E} = \frac{30\,\%}{E} \oplus \frac{10\,\%}{\sqrt{E}} \oplus 0.4\,\%. \tag{III.6}$$

E is the energy of a particle in GeV, $\oplus$ denotes summation in quadrature. The resolution depends on three main factors:

1. **Noise term** $\propto \frac{1}{E}$: Caused by noise (*e.g.* due to electronics), usually negligible unless when measuring low energies.
2. **Stochastic term** $\propto \frac{1}{\sqrt{E}}$: Caused by statistical fluctuations in the shower or in the amount of detected charge. Since the energy deposited in the calorimeter is proportional to the number of particles produced in the shower, the relative resolution scales with $1/\sqrt{E}$. This is because the statistical uncertainty in a counting experiment scales with $\sqrt{N}$ $(\propto \sqrt{E})$.
3. **Constant term**: Caused by non-uniformity of the calorimeter or calibration uncertainties. Dominates at high particle energies.

For an electron with a moderately high energy of $E = 100\,\mathrm{GeV}$, the energy resolution is $\sim 1\,\%$, and it is $\sim 2\,\%$ for $E = 25\,\mathrm{GeV}$. The EM calorimeter therefore outperforms the tracking detector already at $p_\mathrm{T} \gtrapprox 25\,\mathrm{GeV}$, compare with the momentum resolution, equation (III.5), which is $6\,\%$ at $p_\mathrm{T} = 100\,\mathrm{GeV}$ and $2\,\%$ at $25\,\mathrm{GeV}$. For electrons, the calorimetric measurement can therefore be preferred over momentum measurements.

In transverse direction, the calorimeter has three layers (figure III.9): a presampler with very high granularity in η, which detects neutral pions decaying to two photons and particles whose shower already started in the inner detector. The presampler is followed by longer towers with high granularity, which detect the bulk of the EM showers, and allow for measurements of the η- and ϕ-coordinates of the particles. The last layer detects showers that are leaving the calorimeter. This occurs when particles other than electron/positron/photon start showers inside the ECal.

The calorimeter has a barrel segment ($-1.475 < \eta < 1.475$) and two endcap segments covering $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$. The very forward region of the EM calorimeter is not used in the $VH \to Vb\bar{b}$ analysis because electrons cannot be separated from photons in this region. For this, tracking information would be necessary, which is only available up to $|\eta| = 2.5$.

Tile Calorimeter (Tile Cal / HCal) Hadrons such as pions, kaons, protons and neutrons traverse the electromagnetic calorimeter largely unimpeded since they do not emit Bremsstrahlung because of higher masses or missing electric charge. They can, however, interact hadronically with nuclei of the hadron calorimeter. The probability of such interactions scales with the amount of nucleons, which is why the hadron calorimeter consists of steel absorbers surrounding the ECal. These are arranged in 64 wedges with a granularity of $\Delta\phi \cdot \Delta\eta = 0.1 \cdot 0.1$, see figure III.10. To detect showers of hadrons, the steel absorbers are interleaved with scintillating tile layers, which emit light when excited by charged particles in the showers. Neutral pions can also be detected because these decay to photon pairs, which in turn create small EM showers. The scintillator light is guided to photomultipliers by fibres as shown in figure III.10. The amount of light collected is a measure of the energy of the incoming hadrons or jets.

The Tile Cal energy resolution is much worse than the ECal resolution, since a large part of the hadronic showers is absorbed in the steel layers. This leads to a large stochastic term. The energy resolution also varies with the incident particle type. For charged pions, one finds, for example [74]:

$$\frac{\sigma_E}{E} = \frac{160\,\%}{E} \oplus \frac{52\,\%}{\sqrt{E}} \oplus 3\,\% \tag{III.7}$$

For $E_\pi = 100\,\mathrm{GeV}$, the resolution is $\sim 6\,\%$; for $E_\pi = 25\,\mathrm{GeV}$ it is $\sim 13\,\%$. The HCal resolution is the most important uncertainty for the measurement of b-jet energies, which are indispensable for reconstructing Higgs decays in the $H \to b\bar{b}$ analysis. The typical jet energy resolution in ATLAS is about $19, 11$ and $6\,\%$ at $p_\mathrm{T} = 25, 100$ and $500\,\mathrm{GeV}$ [75]. Most jets relevant for the $H \to b\bar{b}$ analysis are found in the range $p_\mathrm{T} = 25$ to $100\,\mathrm{GeV}$, where the resolution is not very high. However, building a hadron calorimeter with better energy resolution would require a much larger and thus much more expensive calorimeter. It therefore has to be accepted as a limiting factor for the $H \to b\bar{b}$ search. The Tile Cal covers a pseudorapidity range of $|\eta| < 1.7$, see figure III.8.

Endcap and Forward Calorimeters The LAr Hadron Endcap HEC is installed behind the LAr forward ECal. Due to the high particle density in the forward regions, the hadron endcap calorimeters also use liquid argon technology: the argon, which is constantly exchanged, does not suffer from radiation damage, contrary to scintillators. The HEC covers $1.5 < |\eta| < 3.2$, and complements the Tile Cal in the forward regions. A forward calorimeter covers $3.1 < |\eta| < 4.9$, uses three LAr layers, and is a combined ECal and HCal.

Table III.1: Subsystems and number of readout channels of the ATLAS detector.

Readout Channels			
(IBL)	12×10^6	LAr	170×10^3
Pixel	80×10^6	HEC&FCal	9×10^3
SCT	6×10^6	Tile	10×10^3
TRT	350×10^3	Muons	1×10^6

The $H\to b\bar{b}$ analysis mostly relies on the central parts of the calorimeters with $|\eta| < 2.5$, since tracking information is necessary for both the identification of b-jets and the identification of leptons. The forward calorimeters are used only to enforce a "jet veto": if two or three jets are present in the central part of the detector, and additional jets are detected in the forward calorimeters, the event is discarded because only two or three jets are expected in $VH\to Vb\bar{b}$ events, and these must be within the acceptance of the tracking detectors to apply b-tagging.

III 2.3 Muon Spectrometer: Measuring the Momenta of Muons

The muon spectrometer is the outermost and largest detector of ATLAS: the light-blue and grey components in figure III.5. It exclusively detects muons since all particles except neutrinos and muons are absorbed inside the calorimeters. Muons, since they are not interacting via the strong force, deposit only small amounts of energy when traversing the calorimeters. The muon spectrometer is implemented as a tracking detector with its own magnetic field to measure track curvatures and thus muon momenta.

The magnetic field in the muon spectrometer is created by eight large toroid coils in the barrel region, providing a field of 0.5 to $4\,\mathrm{T}$, depending on the distance to the coils. Sixteen smaller toroid coils with similar field strength are installed in the endcaps. The muon system uses three layers of tracking detectors, both in the barrel and forward regions. These use a mixture of fast detectors with low position resolution and slow detectors with high position resolution. This allows for both accurate timing and precise position or momentum measurements.

The momentum resolution of the muon system ranges from $1.7\,\%$ at low $|\eta|$ for $p_\mathrm{T}\approx 10\,\mathrm{GeV}$, to $4\,\%$ at large $|\eta|$ for $p_\mathrm{T}\approx 100\,\mathrm{GeV}$ [76]. The momentum measurement from the muon system can be combined with momentum measurements from the inner detector to improve the resolution. Such "combined muons" will mostly be used to identify $VH\to Vb\bar{b}$ events where the vector boson decays into muons. The muon system covers a pseudorapidity range of $|\eta| < 2.7$, a slightly larger acceptance than the inner tracking system.

III 2.4 Trigger System

As discussed in section III 1.1, the LHC is designed for bunch crossing rates of up to $40\,\mathrm{MHz}$, or every $25\,\mathrm{ns}$. The full record of an ATLAS event is about $2\,\mathrm{MB}$ since the data of about 100 million readout channels must be stored, see table III.1. Because of the high collision rate, most events have to be discarded immediately in order to achieve sustainable data rates. A three-level trigger system therefore selects events to be recorded. The level-1 trigger is a hardware system that selects events based on sparse data from the calorimeters and the muon system. It can detect the presence of high-momentum muons, electrons, photons, jets and hadronic decays of τ leptons, as well as missing transverse momentum (see section III 3.2). The hardware systems on level 1 select event candidates with a latency of $2.5\,\mu\mathrm{s}$. Only if an event is accepted by one of the level-1 triggers, the full detector is read out. Since the trigger decision

might be taken as late as $2.5\,\mu s$ after the collision, all subdetectors need to constantly buffer at least 100 events. Since no tracking information is available on level 1 (compare the amount of inner detector channels to the other detector components in table III.1), one cannot implement level-1 b-jet triggers. For the $VH \to Vb\bar{b}$ analysis, this means that trigger decisions must be taken based on decay products of the vector boson and not based on the $H \to b\bar{b}$ decay. Besides reduced background levels, this is one of the main reasons to focus on the VH final state.

Level-1 triggers reduce the event rate to about 75 to 100 kHz. This means that most events have to be discarded without fully reading out the detector. Since most collisions produce objects with low transverse momenta, triggers are usually set to only react to high-transverse-momentum objects. Typical trigger thresholds in 2012 were $p_T \geq 350\,\text{GeV}$ for a single jet, $p_T \geq 18\,\text{GeV}$ for isolated electrons and $p_T \geq 20\,\text{GeV}$ for muons. These trigger thresholds are too high in order to use jet triggers for the $H \to b\bar{b}$ final state.

The level-2 trigger system runs on a computer farm, and has access to the full detector data in the regions of interest (`ROI`) marked by the level-1 triggers. A simple reconstruction of the event is performed, and a trigger decision is taken within $\sim 40\,\text{ms}$. The output rate is 2 to 4 kHz. The final stage ("event filter") performs a more detailed reconstruction, and processes events within a few seconds. These two stages are collectively referred to as "high level trigger" (HLT). HLT was planned for an output rate of 200 Hz, but trigger rates as high as 650 Hz were reached in 2012. To achieve a sufficient reduction of event rates, HLT trigger thresholds need to be significantly higher than the level-1 thresholds. The threshold for isolated electrons or muons, for example, is 24 GeV. For a b-jet, the threshold is 360 GeV, but an additional jet with $p_T \geq 360\,\text{GeV}$ has to be required to keep the trigger rate low enough. Double b-jet triggers with $p_T > 35\,\text{GeV}$ are only available when simultaneously requiring a third and fourth jet because this trigger is seeded by a four-jet-trigger on level 1.

This shows that b-jet triggers are unusable to detect $H \to b\bar{b}$ because the signature of `ggF`, for example, is two b-jets with transverse momenta rarely higher than $\sim 100\,\text{GeV}$. The trigger thresholds are so high because the b-jet production cross section at the LHC is large: Figure III.4 shows that with $\mathscr{L} = 1 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$, the b rate is $\sim 100\,\text{kHz}$. Trigger constraints are slightly better for `VBF` production. Triggers for `VBF` topologies are available, but background levels in the ATLAS $H \to b\bar{b}$ `VBF` analyses [77, 78] are significantly higher than in the already challenging $VH \to Vb\bar{b}$ analysis discussed in this work.

The $VH \to Vb\bar{b}$ channel therefore is the most promising production channel for $H \to b\bar{b}$ searches at the LHC. This is because high-p_T leptons are rare in LHC collisions since protons consist only of strongly interacting particles. The triggers for isolated leptons with $p_T > 24\,\text{GeV}$ are sufficiently low to record most leptonic vector boson decays in $VH \to Vb\bar{b}$ events. Typical lepton momenta are $p_\ell \approx m_W/2 = 40\,\text{GeV}$ and $m_Z/2 = 46\,\text{GeV}$ if the vector bosons are at rest. The momentum in transverse direction is on average lower, but with non-vanishing transverse momenta of the bosons, the leptons easily exceed the trigger thresholds. Hadronic decays of the vector bosons are not targeted in this work because these would again require jet triggers.

III 3 Reconstruction of Particle Collisions in ATLAS

The raw data recorded by the ATLAS experiment cannot be used directly for analysis. Multiple steps of object reconstruction have to be run before collision events can be analysed. The mapping of measured quantities like deposited light or deposited charge to objects like charged particles or jets is carried out by reconstruction algorithms, which use the fact that specific particles only interact with specific subdetectors. The signatures of different particles are summarised in figure III.11.

III 3.1 Reconstruction of Visible Objects

Track and Vertex Reconstruction

The reconstruction of charged particle tracks [79, 80] starts from hits in the inner detector. Pixel and SCT hits are clustered to both improve the accuracy of the position measurement and reduce the number of points (and thus the noise) entering the track-finding stage. A typical track in the barrel region is associated with three pixel clusters, eight clusters in the SCT and ≥ 30 straw hits in the TRT. The reconstruction algorithms also take into account when clusters are missing because tracks may cross detector areas known to be dead. After clustering, an inside-out track-finding algorithm searches for three-point tracks in the silicon detectors. These are extrapolated into the outer regions of the tracking detector by iteratively associating new clusters to the track and updating the track parameters (Kalman filter). This algorithm is designed to find charged particles leaving the collision point. The highest detectable curvatures correspond to momenta as low as a few hundred MeV.

The inside-out algorithm is complemented by a back-tracking algorithm that starts from hits in the TRT, and extrapolates towards the silicon detectors, iteratively associating silicon clusters. This algorithm finds tracks that do not originate from the interaction region, giving access to tracks from particles that are produced in decays outside of the interaction region.

Fully reconstructed tracks are described by six parameters with corresponding uncertainties, which are estimated by fits to all valid hits associated to a track. The space coordinates of a track are described by the transverse distance to the interaction region, d_0, and by z_0, the z-coordinate of the point where the track is closest to the interaction region. The direction of the outgoing particle is described by ϕ and θ or η, and the track momentum or curvature are described by q/p_{T}.

After tracks are reconstructed, vertices are searched [80, 81]: a vertex is a point where a set of tracks cross, and is therefore often the point where the collision occurred or where a particle decayed. Vertices are found by calculating the crossing point of a few tracks and iteratively adding/discarding all tracks that are compatible/incompatible with originating from this crossing point. In this process, the vertex position is updated with each change of the track association. Vertices within the interaction region are called `primary vertex`, vertices outside of the interaction region `secondary vertex`. Secondary vertices may indicate the points where particles with longer lifetimes decayed. This is crucial information for b-tagging, since b-flavoured hadrons have lifetimes such that they can decay several millimetres from the collision point. For a B meson with $p = 60\,\mathrm{GeV}$ and a lifetime of $\mathcal{O}(1\times10^{-12}\,\mathrm{s})$, one can expect a mean decay length of about $4.5\,\mathrm{mm}$, whereas the interaction region extends only about $60\,\mu\mathrm{m}$ in transverse direction.

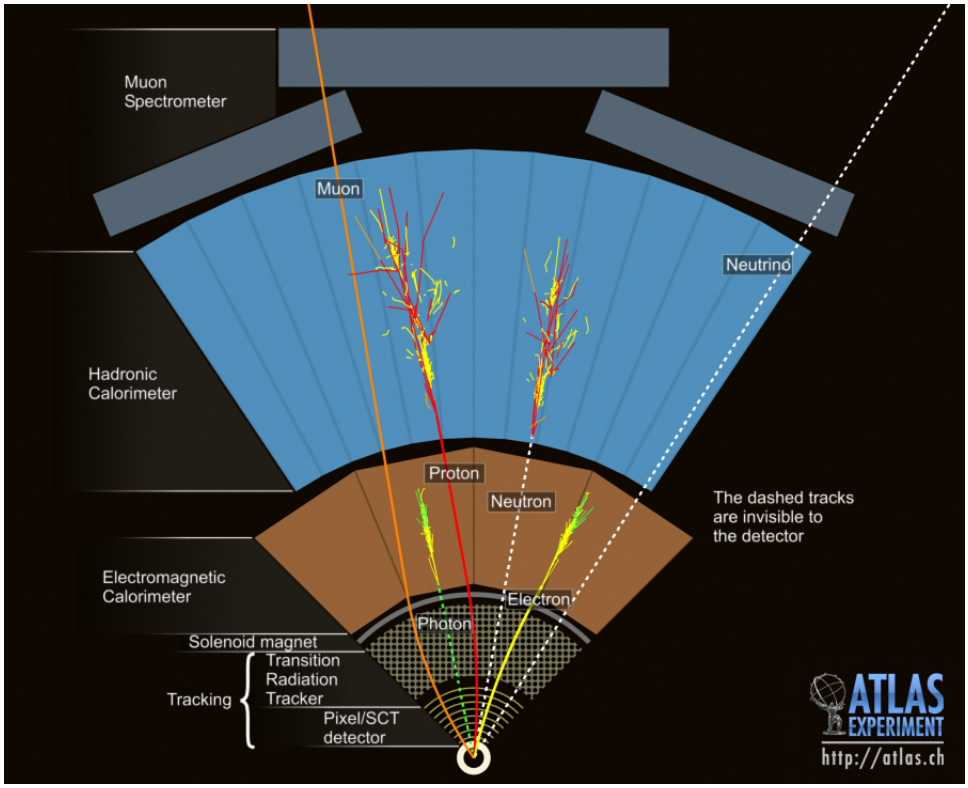


Figure III.11: Interaction of different particles with the subsystems of ATLAS [82]. The signatures of proton and neutron are representative for the signatures of any long-lived charged or neutral hadron. The track curvature can be used to separate particles from antiparticles.

Electron Reconstruction

The reconstruction of electrons[4] [83, 84] starts in the EM calorimeter, where electrons deposit their energy. The calorimeter is divided into a grid of 200x256 (η-ϕ)-towers, the granularity of the middle layer, and a sliding window of 3x5 towers is moved across this grid. Clusters with $E_{\mathrm{T}} > 2.5\,\mathrm{GeV}$ are recorded, and the energy deposited in the other layers of the calorimeter is added to the clusters.

Since electrons are expected to leave a track in the tracking system, tracks within a search window of $\Delta R = 0.3$ from the cluster barycentre are selected as candidate tracks. Without any tracks, the EM cluster would be a photon candidate, compare figure III.11, but photons are not relevant for the $VH \to Vb\bar{b}$ analysis. Two track fits are performed: assuming the candidate tracks are caused by pions and assuming they are caused by electrons. Compared with electrons, pions have a high mass and thus traverse the inner detector without being affected strongly by multiple scattering. Pions therefore create more uniform tracks than electrons. If the pion track fit fails or indicates a low likelihood, the fit using an electron hypothesis is preferred. This fit accounts for bremsstrahlung losses and multiple scattering. If this fit succeeds, a track is accepted in either of the two following cases:

1. If a track with at least four silicon hits matches with the cluster barycentre within a tolerance of $|\Delta\eta| < 0.05$ and $-0.05 < \Delta\phi < 0.2$, the track is accepted. $\Delta\phi$ is defined such that positive distances coincide with the direction the track is bending into. The window is wider towards positive $\Delta\phi$ to prevent failed matches due to track momentum uncertainties. The $\Delta\eta$ requirement is

[4] Throughout this thesis, the name of a particle will often be used to also refer to the corresponding antiparticle.

dropped if the track is a TRT-only track, since the TRT does not provide the longitudinal coordinates.

2. After rescaling the track momentum to the cluster energy and repeating the track-cluster matching, tracks are accepted if they match within a tolerance of $|\Delta\eta| < 0.05$ and $-0.05 < \Delta\phi < 0.1$. This ensures that electrons, which may encounter substantial momentum losses due to the emission of bremsstrahlung, are still accepted.

Track parameters of accepted tracks are estimated using a specialised track fit for electrons, the "Gaussian Sum Filter", which yields the best estimate of track parameters when momentum losses due to scattering and Bremsstrahlung occur.

After assigning tracks to the EM clusters, the cluster energies are determined by enlarging the clusters to 3x7 cells (5x5 in the endcaps), and applying cell-energy calibration factors. If multiple tracks are matched to the same cluster, the best track is selected based on the ΔR distance to the cluster and the number of precision hits in the inner detector.

Despite the track-cluster-matching, the electron reconstruction described here yields a high number of false positives, since many clusters in the LAr calorimeter have associated tracks. Such false positives can be caused by photons or π_0 mesons with nearby tracks or by charged hadrons that deposit a part of their energy in the ECal. This is because the electron reconstruction does not check if particles also deposit energy in the hadron calorimeter. Another source of electron candidates are electrons produced inside jets: these can be real electrons, but they should be attributed to the jet instead of being considered as electrons in their own right. To suppress false positives, a likelihood-based identification method is employed. Electron-like and background-like likelihood functions are combined into the following discriminant:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}, \qquad \mathcal{L}_{S,B} = \prod_{i=1}^{n} L^i_{S,B}(x_i)$$

The terms $L^i(x_i)$ are likelihood functions in different observables like shower shapes, bremsstrahlung losses, track parameters or data from the TRT. Different likelihoods for multiple observables are combined by multiplication. Several working points were defined by selecting sets of L^i with corresponding thresholds for $d_{\mathcal{L}}$. The two working points used in the $VH \to Vb\bar{b}$ analysis are `Very Loose Likelihood` and `Very Tight Likelihood` [83][5]. Typical electron identification efficiencies are $\varepsilon = 77\,\%$ with a background rejection of 1:350 for the very tight working point and $\varepsilon = 98\,\%$ with a background rejection of 1:30 for the very loose working point. In the $VH \to Vb\bar{b}$ analysis, very tight is used in settings where hadrons are frequently misidentified as electrons; very loose is used when misidentified hadrons are rare.

Muon Reconstruction

Since the calorimeters absorb all particles except muons and neutrinos, compare again figure III.11, muon reconstruction is equivalent to finding a track in the muon system [76]. For this, track fragments in the layers of the muon system are connected to full muon-system tracks. If a muon also traverses the inner detector, the tracks of the two detector systems are combined. This leads to different types of muons:

1. **Combined**: Tracks in both inner detector and muon system. Combining these tracks yields a higher-precision measurement of the muon four-momentum. There are two combination strategies: simple statistical combination (`StaCo`) and re-fitting of the tracks, taking into account that muons loose energy when traversing the calorimeters (`MuID` = Muon Identification). The latter strategy is used in the $VH \to Vb\bar{b}$ analysis.

[5] The very loose likelihood working point is not documented in the official ATLAS documentations. It is, however, documented in ATLAS-internal documentation [85].

2. **Stand-Alone**: Muons reconstructed in the muon system but without inner-detector tracks. These muons predominantly occur in the forward regions at $2.5 < |\eta| < 2.7$ since the muon system extends to pseudorapidities of 2.7 whereas the inner detector only extends to 2.5. Stand-alone muon tracks can be extrapolated towards the interaction point, but the track parameters have considerably higher uncertainties because no hits in the inner detector are available.
3. **Segment-Tagged**: Inner detector tracks connected with a track segment in the muon system. This occurs when not all layers of the muon system register the muon because of a strongly curved track, or because regions with low acceptance are traversed.
4. **Calorimeter-Tagged**: The muon system has uninstrumented regions at $\eta \approx 0$, where services for the inner detector and the calorimeters are installed. Muons can nevertheless be identified by connecting inner detector tracks with energy clusters in the calorimeters, provided the muon deposits a part of its energy in the calorimeter. This partly recovers the muon reconstruction inefficiency at $|\eta| < 0.1$. In the $VH \to Vb\bar{b}$ analysis, such muons are only used in this region because the rate of false positives is higher.

Even though different muon quality levels were defined by ATLAS, the strictest level "tight" can be used for all muons in the $VH \to Vb\bar{b}$ analysis with neither notable inefficiency nor notable problems with false positives. This is because virtually all tracks found in the muon system are indeed caused by muons.

Considerable effort was invested in designing ATLAS such that it can reconstruct leptons with high precision. EM calorimeter and muon system in fact yield leptons with so high precision that lepton uncertainties are almost irrelevant for the $VH \to Vb\bar{b}$ analysis – unlike jet (= hadron calorimeter) uncertainties. The only lepton generation with notable inefficiencies and uncertainties are τ leptons. These are in principle equally interesting as electrons and muons, but due to their short lifetime and richer decay spectra they are more difficult to reconstruct. With mean lifetimes of $t_\tau \approx 2.9{\times}10^{-13}\,\mathrm{s}$, τ leptons decay before they can be detected by ATLAS. Most decay to hadrons, which requires new triggers and using the hadron calorimeter for reconstruction. Due to this, the inclusion of $V \to \tau\nu_\tau$ decays into the $VH \to Vb\bar{b}$ analysis would have little benefit: τ-identification efficiencies of 25 to 50 % at background rejection rates of 1:100 to 1:10 [86] could be achieved. This compares unfavourably to $\varepsilon \gtrsim 95\,\%$ with better background suppression for the other lepton generations. Furthermore, including τ leptons would require dedicated τ triggers and τ analysis categories, and entail considerable inefficiencies as well as identification and calibration uncertainties. ATLAS analyses therefore usually focus exclusively on electrons and muons *or* exclusively on τ leptons. This work is part of the former class.

Jet Reconstruction

Although the ATLAS sketch in figure III.11 suggests otherwise, one is rarely concerned with the reconstruction of single hadrons in ATLAS: since quarks and gluons create jets, one usually reconstructs these as a whole. Jet reconstruction starts from pre-clustered calorimeter cells. Since the noise term is dominant at low energies, see equation (III.6), pre-clustering the calorimeter into topological clusters, and only using clusters with high-enough energy suppresses noise (topocluster algorithm [87]). Starting only with cells where a signal-to-noise ratio of $S/N \geq 4$ is found, surrounding cells with $S/N > 2$ are subsequently added to the cluster. Finally, neighbouring cells with lower S/N are added as well. This reduces the amount of input cells for the jet-finding algorithm, and suppresses the low-energy noise in the calorimeter. A discussion of clustering algorithms used in ATLAS can be found in [88].

Jets are subsequently reconstructed [89] using the `FastJet` [90] implementation of the anti-k_T algorithm [91]. Anti-k_T sequentially combines topoclusters into larger objects based on the momentum-weighted distance between clusters. For two objects i and j, this distance is:

$$d_{ij} = \min(p_{\mathrm{T},i}^{-2}\ ,\ p_{\mathrm{T},j}^{-2})\frac{\Delta R_{i,j}^2}{0.4^2}, \qquad \text{with } d_{iB} = p_{\mathrm{T},i}^{-2} \quad \text{and}$$
$$\Delta R_{i,j}^2 = (\phi_i - \phi_j)^2 + (y_i - y_j)^2.$$

The distance d_{ij} is weighted by the inverse, squared transverse momentum of the higher-momentum object. The algorithm combines the two objects with the lowest d_{ij} into a single cluster. This means that objects are likely to be combined if they are either very close or if a high-momentum object is surrounded by low-momentum objects. Therefore, higher-momentum objects in a sense "pull" lower-momentum objects into a cluster. d_{iB} (B for beam) is a stopping criterion: if no further objects with $d_{ij} < d_{iB}$ are found, the jet i is considered final, and all associated clusters are removed from the set of candidate clusters. The denominator in d_{ij} therefore defines the width of the jets: if a jet i is surrounded only by low-momentum objects, the stopping criterion reads:

$$\begin{aligned} & d_{ij} < d_{iB} \\ \Leftrightarrow \quad & p_{\mathrm{T},i}^{-2}\frac{\Delta R_{i,j}^2}{0.4^2} < p_{\mathrm{T},i}^{-2} \\ \Leftrightarrow \quad & \Delta R_{i,j} < 0.4 \end{aligned}$$

Anti-k_{T}[6] therefore produces circle-shaped jets. The standard jet radius in ATLAS is $\Delta R = 0.4$ in the azimuth-rapidity plane.

After the jet clustering, jet energies are calibrated with calibration functions obtained from simulations and measurements in data [89, 92]. The intrinsic resolution of the calorimeters nevertheless limits the achievable jet energy resolution (JER), which is one of the most important uncertainties of $H \to b\bar{b}$ searches. This is because the resolution of the invariant mass of a pair of b-jets scales with the jet energy resolution, and the invariant mass is one of the best observables to separate $H \to b\bar{b}$ decays from background events.

III 3.2 Reconstruction of Missing Transverse Energy

Up to this point, all known particles except for neutrinos are covered by at least one of the reconstruction algorithms. Neutrinos, however, cannot be reconstructed directly because they do not interact with the detector material. Due to momentum conservation, a part of the neutrino momentum can be reconstructed, though: as "missing transverse energy".

The initial state of LHC collisions is well defined in the transverse plane: since the proton beams are aligned with the longitudinal axis, the transverse momentum in the initial state is effectively zero. Transverse momenta that partons acquire inside the proton are negligible compared to the momentum of the proton. While the transverse momentum vanishes, the longitudinal momentum in a collision is $(x_1 - x_2) \cdot E_{\mathrm{Beam}}$. The momentum fractions x_i can neither be predicted nor measured. Conservation of momentum can thus only be imposed in the transverse plane. If neutrinos or unknown particles leave the detector with non-zero transverse momenta, the missing momentum can be reconstructed using the conservation of momentum. This leads to the term "missing transverse momentum":

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = \vec{p}_{\mathrm{T}}^{\,\mathrm{initial}} - \vec{p}_{\mathrm{T}}^{\,\mathrm{final}} = -\vec{p}_{\mathrm{T}}^{\,\mathrm{final}}.$$

The missing transverse momentum is equal to the negative reconstructed or visible momentum in the final state.

[6] The name "anti-k_{T}" originates from the 'k_{T}'-algorithm, which uses the weight $k_{\mathrm{T},i}^2 \equiv p_{\mathrm{T},i}^2$. k_{T}-jets are not circular, though. This poses problems for calibrating the jets, because the jet area is not predictable. $p_{\mathrm{T},i}^{-2}$ was found to yield circle-shaped jets, and dubbed 'anti-k_{T}' because of the negative exponent.

The transverse momenta of muons are measured directly in the muon spectrometer, but for other particles only *energies* are detected in the calorimeters. This leads to the definition of missing transverse *energy*, $E_{\mathrm{T}}^{\mathrm{miss}}$ or MET:

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\mathrm{Calo}} E_{\mathrm{T}}^{i} \begin{pmatrix} \cos\phi_i \\ \sin\phi_i \end{pmatrix} - \sum_{\mathrm{Muons}} \vec{p}_{\mathrm{T}}^{\,i} \tag{III.8}$$

$E_{\mathrm{T}}^{\mathrm{miss}}$ therefore can be seen as an attempt to measure momenta with calorimeters. All energies measured in all calorimeter cells have to be decomposed into a longitudinal and a transverse component, and the transverse components have to be summed, taking into account the ϕ-coordinates of the cells. Since a large number of calorimeter cells has to be summed, especially with contributions from the hadron calorimeter with high uncertainties of up to 20 %[7], $E_{\mathrm{T}}^{\mathrm{miss}}$ has high uncertainties, too. Typical values of $E_{\mathrm{T}}^{\mathrm{miss}}$ range from 10 to 30 GeV, even if the actual missing transverse momentum $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is zero. The accuracy of $E_{\mathrm{T}}^{\mathrm{miss}}$ is usually increased by using information about other reconstructed objects: electrons, muons, jets, τ leptons, photons *etc*., which have been reconstructed and calibrated, are fed back into the calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ in order to make use of the calibration results. The corresponding calorimeter cells are taken out of the sum in equation (III.8). Details on the reconstruction and calibration of $E_{\mathrm{T}}^{\mathrm{miss}}$ can be found in [93, 94].

One fact that should be noted about $E_{\mathrm{T}}^{\mathrm{miss}}$ is that only the sum of *all* missing momentum vectors in the transverse plane can be detected. If multiple invisible particles are produced, $E_{\mathrm{T}}^{\mathrm{miss}}$ can increase or decrease depending on the azimuthal angle between the objects. With $\phi \approx 180°$, objects might be completely invisible if their transverse momenta have similar magnitude. The same holds for particles that escape in beam direction with low transverse momentum. Nevertheless, $E_{\mathrm{T}}^{\mathrm{miss}}$ is an indispensable tool in the VH analysis: leptonic W decays always produce a neutrino, which can be used to detect and reconstruct the transverse momentum component of the W boson. This is one of the best observables to separate $WH \to \ell\nu b\bar{b}$ events from background events. The W reconstruction will be revisited in chapter IV.

III 3.3 Overlap Removal

The reconstruction algorithms described in the previous sections run independently – no information is exchanged between different algorithms during the reconstruction of objects. This can lead to objects being reconstructed multiple times. Since both electrons and jets are reconstructed from clusters in the calorimeters, electrons are also reconstructed as jets, and jets can be reconstructed as electrons. Moreover, muons that radiate photons in or in front of the calorimeter might be detected as electrons, since such muons are visible as a track pointing to a photon cluster. Muons created inside jets might be mistaken as coming directly from the parton collision because the muon reconstruction does not check for the presence of jets. The same can happen for electrons inside jets. One such electron is, for example, visible in the event display in figure III.12 where a green electron track is found among the blue hadron tracks associated to the jet. Despite of two visible electrons, the event is correctly classified as a one-electron event because disambiguities and double counting of objects are resolved by `overlap removal`.

To remove the overlap between different object collections, a hierarchy of objects is defined: higher-priority objects are preferred over lower-priority objects. If two objects are found close to each other in the η-ϕ plane, the lower-priority object is discarded. High-E_{T} electrons, for example, are preferred over jets, combined muons over calorimeter-tagged muons. Muons or electrons produced inside jets are not counted as leptons produced in the collision. The exact definition of the object hierarchy slightly varies depending on which objects are most important for an analysis. The object hierarchy for the $VH \to Vb\bar{b}$ analysis will be discussed in section IV 1.1.

[7] Compare equation (III.7). The energies in most cells will be $\leq \mathcal{O}(\mathrm{GeV})$.

III 3.4 *b*-Tagging

b-tagging is a crucial step of the event reconstruction for the $H \to b\bar{b}$ search. When a jet forms around a b quark, the b quark is bound in a short-lived hadron, which subsequently decays. Such hadrons can be B mesons, for example, with lifetimes of $\tau \approx 1.5{\times}10^{-12}\,\mathrm{s}$. This lifetime is long enough to make the hadron travel several millimetres before it decays, but too short to reach the calorimeters. These are reached only by the lightest mesons like pions and kaons with $\tau \approx 2{\times}10^{-8}\,\mathrm{s}$. This leads to particular signatures for b-jets, which can be identified with precise tracking: when a b hadron decays, several lighter hadrons originate from a slightly displaced secondary vertex. The tracks of such hadrons will usually not cross the interaction region, that is, they will have a non-zero impact parameter. This is the distance between the track and the collision vertex at the point of closest approach. The mean flight length for a B meson with $p = 25\,\mathrm{GeV}$ is, for example:

$$\beta \cdot \mathrm{c} \cdot \tau \cdot \gamma \approx \mathrm{c} \cdot 1.5{\times}10^{-12}\,\mathrm{s} \cdot 5 = 2.25\,\mathrm{mm}.$$

This leads to impact parameters of $\mathcal{O}(0.1\,\mathrm{mm})$, since the outgoing tracks originate from the secondary vertex and not from the interaction region. This can be seen in figure III.12, which shows two b-jets: the bottom right inset shows a jet with two secondary vertices (blue spheres) that are displaced from the primary vertex (green sphere) by several millimetres. Secondary vertices like these two can be caused by a $B \to D \to$ light cascade. Several tracks pass the primary vertex at a distance of a few hundred $\mu\mathrm{m}$. These impact parameters are smaller than the vertex displacements because the b hadrons are boosted along the jet axis. The other jet has no secondary vertex marker, but also has several displaced tracks. Both jets are identified as b-jets by the ATLAS software.

ATLAS uses three classes of b-tagging algorithms [95, 96]:

- Impact-parameter-based: These algorithms detect displaced tracks by comparing the measured impact parameters with typical impact-parameter distributions for b-jets and light jets. If a jet has several displaced tracks, it is likely a b-jet.
- Secondary-vertex-based: If a secondary vertex is found in a jet, it is likely a b-jet. Not all b-jets have reconstructible secondary vertices, though, as is visible for one of the jets in figure III.12.
- Multivariate algorithms: This class of algorithms combines information from multiple sources into a single discriminant. The standard ATLAS b-tagging algorithm for Run-1 data is `MV1`, a multivariate classifier that combines the output of several impact-parameter- and vertex-based b-tagging algorithms, as well as a multivariate algorithm that analyses jet kinematics. It outperforms all other algorithms, because it is the only algorithm that simultaneously uses impact-parameter, vertex and kinematic information.

`MV1c`, which was developed mainly for the $VH \to Vb\bar{b}$ analysis, is a variant of `MV1` with increased charm jet suppression. Instead of training the algorithm to separate b- and light jets, it is trained to separate b-jets from a mixture of charm and light jets. This improves the charm-jet suppression at the cost of a slightly lower light-jet suppression. Yet charm-jet suppression is more important for the $VH \to Vb\bar{b}$ analysis because charm jets are much more likely to be b-tagged.

This is because, firstly, charm jets can have secondary vertices, and secondly, they also have tracks with non-zero impact parameters. The lifetime of charm mesons is shorter than for bottom mesons, but the boost is higher because of the lower mass. With $\tau_D \approx 4{\times}10^{-13}\,\mathrm{s}$, typical displacements of a D meson, for example, are of the order of one millimetre, depending on its boost: $\beta\mathrm{c}\tau\gamma(p = 25\,\mathrm{GeV}) \approx 1.6\,\mathrm{mm}$. The track multiplicity in D decays is lower, though. With lower mass, less energy to create particles is available. One expects about two tracks for D and about five tracks for B mesons. `MV1c` therefore needs

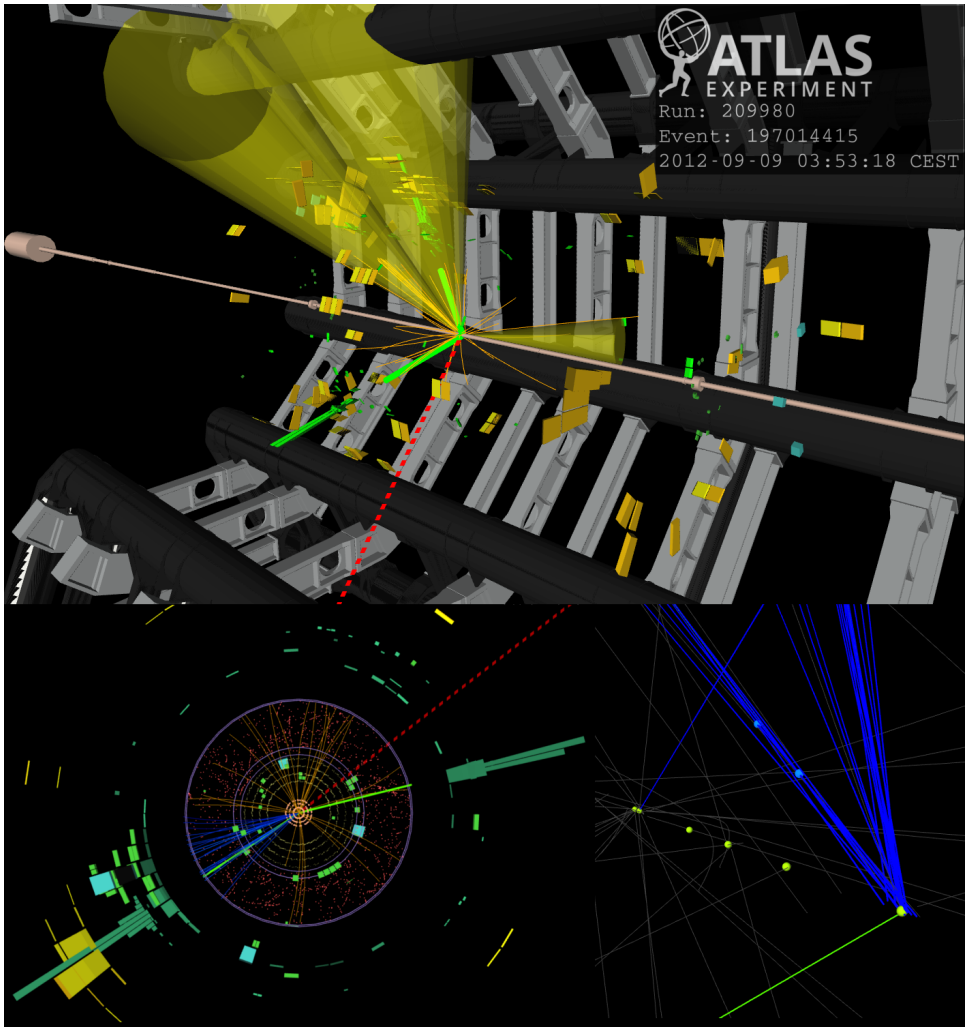


Figure III.12: Event display of a $WH \to e\nu b\bar{b}$ candidate event recorded in 2012. The top image shows two b-tagged jets as yellow cones as well as a pileup jet with low momentum. The green tracks are reconstructed electrons. Energy clusters in the calorimeters are indicated by green (LAr) and orange (HCal) cells. Insets at the bottom show (left) a transverse view of tracking detectors and calorimeters and (right) the interaction region. Tracks associated to the b-tagged jets are shown in blue and tracks from pileup collisions in orange/grey. A green sphere indicates the location of primary vertices. Many tracks are significantly displaced from the primary vertex in the foreground, which is typical for b-jets. The primary vertices in the background are caused by pileup collisions. One jet has two secondary vertices (blue spheres), likely from a $B \to C \to$ light cascade. In the transverse plane, the two jets and the electron+$E_{\mathrm{T}}^{\mathrm{miss}}$ system (green & dashed red line) are balanced because of momentum conservation. This is a typical signature for H and W recoiling against each other.

to discriminate b vertices from c vertices. Such discrimination cannot be achieved if only light jets are considered in the training of the algorithm.

The same holds for the analysis of track impact parameters. As for b-jets, charm jets can have tracks with non-zero impact parameters because particles originate from displaced charm-hadron decays. Impact parameters are usually smaller than for b-jets, though, since charm hadrons have shorter lifetimes. Without dedicated training of `MV1c` against c-jets, the non-zero impact parameters of charm jets would lead to frequent b-tags because the impact parameters in light jets vanish within the tracking uncertainties.

III 4 Simulation of Particle Collisions

The $VH \to Vb\bar{b}$ analysis strongly depends on predictions from "Monte Carlo" simulations: Monte Carlo generators perform cross-section calculations by evaluating the amplitudes of Feynman diagrams, and simulate the resulting particles in order to predict how signal and background processes would appear in a detector like ATLAS. This requires solving integrals in high dimensions, for example, integrals over all possible momenta in both the initial and final state of a reaction. Such integrals usually have to be evaluated numerically. The only method that reliably converges is integration by sampling the function at random points, hence the name "Monte Carlo integration". The sampling is not fully random, though, it is often guided in order to reduce sampling errors: by concentrating the sampling on regions that seem to contribute most strongly to the integral (the `VEGAS` algorithm [97]), accuracy and convergence speed can be improved. A review of Monte Carlo methods can be found in [98, 99]. In the following, selected aspects of QCD and event simulation will be discussed, mostly following the reviews of P. Skands[8] [19] and the particle data group [13].

III 4.1 Running Couplings and Renormalisation Scale

When predicting cross sections, one faces a challenging aspect of QCD: the coupling strength changes with the energy scale probed in the reaction. The strong coupling "constant" α_s varies with the energy scale as shown in figure III.14. Other couplings such as α_em also vary, but for QCD the effect is most noticeable: α_s can become so large that perturbative expansions in α_s become impossible.

The scale dependence of the couplings arises because at high energies, virtual particles may contribute to a coupling, and therefore change the effective coupling strength. The theory of running couplings is described by the `renormalisation group`. Renormalisation is a set of techniques to treat diverging integrals, which occur in quantum field theories. One example of such integrals is the evaluation of higher-order corrections to the `LO` t-channel scattering of fermions. Figure III.13 shows one such `NLO` diagram, an additional fermion loop on the photon line of the `LO` diagram. The corresponding amplitudes have to be integrated over all possible momenta q that can "circulate" in the loop. This integral, however, diverges [15]: each propagator in the loop contributes a factor proportional to $\frac{i}{q^2-m^2}$. For high values of q, the integral over the loop momentum therefore approaches:

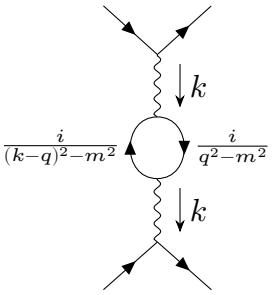


Figure III.13: `NLO` Feynman diagram showing the scattering of fermions via photon exchange with one extra loop.

$$\int_{q_0}^{\infty} \frac{1}{q^4} \cdot q^3 \mathrm{d}q = \ln q \Big|_{q_0}^{\infty} .$$

This integral diverges, which clearly violates unitarity, since this would implicate that the probability for such reactions exceeds unity. The diverging terms can be split off by stopping the integration at an arbitrary cutoff scale, which is a `regularisation` of this integral[9]. To avoid introducing errors by stopping the integration, the integrals have to be split into a finite part and a diverging part. The finite part can be used to evaluate probability amplitudes, but results depend on the cutoff scale. The diverging contributions can be absorbed into the constants of the theory such as masses and couplings, but this

[8] P. Skands contributes to the Monte Carlo generator `Pythia8` [43], which is used for both the simulation of $VH \to Vb\bar{b}$ and for many of the parton showers for backgrounds of the VH analysis.

[9] Other methods of regularisation are usually preferred over a cutoff scale, but this is beyond the scope of this discussion.

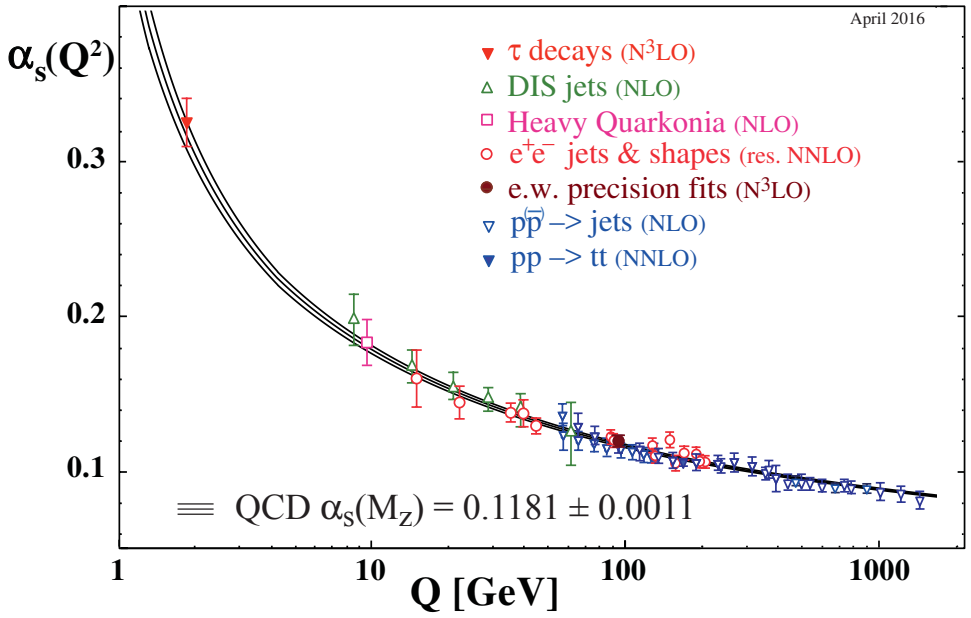


Figure III.14: Scale dependence of the strong coupling constant α_s from [13]. For low momentum transfers, the coupling becomes so strong that perturbation theory cannot be applied. This is the region of "confinement", where interactions that bind the partons in the nucleon take place. At high momentum transfers, partons are in "asymptotic freedom". This is the region of hard collisions, where perturbation theory can be applied. The coloured points show various measurements of α_s, which agree with the predicted evolution of α_s.

leads to scale-dependent constants. This scale dependence, however, cancels the corresponding scale dependence of the finite terms, which renders predictions independent of the choice of scale.

The need for regularisation expresses the fact that a theory such as the Standard Model cannot be expected to describe the interactions of particles up to infinite loop momenta/energies, or equivalently to infinitesimally small length or time scales. It is natural to assume that the Standard Model, at some point, will not yield correct predictions, because processes beyond the framework of the Standard Model have a non-negligible effect. If a theory is renormalisable, which the Standard Model is [11], it can nevertheless be used to make predictions, regardless of a possible breakdown at extreme loop momenta. Renormalisation, in a sense, is a mathematical way to acknowledge that loops with infinite momenta (or similar processes) are beyond the capabilities of the Standard Model.

The evolution of scale-dependent constants can be predicted to several orders in α by solving the renormalisation group equation (RGE), yielding, for example, for α_s [19]:

$$\alpha_s(Q^2) = \alpha_s(\mu_R^2)\frac{1}{1 + b_0\alpha_s(\mu_R^2)\ln\frac{Q^2}{\mu_R^2} + \mathcal{O}(\alpha_s^2)}.$$

The `renormalisation scale` μ_R is a reference point for the evolution of the coupling. The 'constant' α_s can be measured at this scale, and be extrapolated to other scales using the RGE. The coefficient b_0, as well as higher-order terms, can be found in the chapter on QCD in the PDG review [13]. For QCD, which has gluon self-coupling, b_0 is positive. This leads to a decreasing coupling when Q increases. In QED, which has no self-coupling of gauge bosons, b_0 is negative, and the coupling strength therefore increases with energy. Figure III.14 shows the scale dependence of the strong coupling constant and

various measurements at different scales, which agree excellently within uncertainties. Since α_s is small for high Q^2, perturbative expansions in α_s are feasible for hard collisions. These break down, however, when the regime of momentum transfers that bind the nucleons is approached: the size of a proton is $\sim 1\times10^{-15}$ m, which corresponds to energy scales of $Q = \frac{\hbar c}{1\times10^{-15}\,\mathrm{m}} \approx 200\,\mathrm{MeV}$. In this regime, α_s is too large for perturbative expansions. This end of the Q-scale is called "confinement" because the strong coupling binds the partons inside the nucleons to colour-neutral states. The high end of the Q-scale is called "asymptotic freedom" because the partons can be treated as free particles, and interactions between particles can be seen as small perturbations.

It is customary to set the renormalisation scale, at which the evolution of α_s starts, close to the typical momentum transfers in the interactions being simulated. This ensures that the cross-section predictions only weakly depend on the renormalisation scale because the evolution of α_s has to be carried out only over a small energy range. A residual μ_R-dependency always remains, though, because both the Feynman amplitudes and the coupling evolution are expanded in powers of $\alpha_s(\mu_R^2)$, and are not evaluated to all orders. When stopping the expansion at n^{th} order, the residual μ_R-dependence is $\mathcal{O}(\alpha_s^{n+1})$. This can be used to estimate the uncertainty of the matrix element calculation: since the uncertainties of an `N`$^{\text{n}}$`LO` calculation are of the same order as the error due to the μ_R-dependence – both are $\mathcal{O}(\alpha_s^{n+1})$, it is customary to vary the renormalisation scale and take the impact on the cross section as uncertainty of the matrix element calculation. Even though results depend on the choice of renormalisation scheme and the magnitude of the scale variation [100], this yields an estimate of the accuracy of the predictions. The uncertainties of the Higgs production cross sections in figure II.9(b) on page 25, as well as the uncertainties of background cross sections for the $VH \to Vb\bar{b}$ analysis, were obtained in this way. A second source of cross-section uncertainties is the accuracy of the α_s-measurements that are taken as starting point for the RGE-evolution. PDF uncertainties are a third important source of uncertainty. All these are included in the cross-section uncertainties used in the $VH \to Vb\bar{b}$ analysis.

The running couplings, and therefore the validity of perturbative expansions, have an impact on how Monte Carlo simulations can proceed: event generators must treat different stages of the simulations with different strategies. Figure III.15 illustrates this for a proton-proton collision. It is inspired by the sketch of ttH production in the `Sherpa` manual [106], but has been redrawn to show a $ZH \to \ell\bar{\ell}b\bar{b} + g$ event, which is more typical for the $VH \to Vb\bar{b}$ analysis. The most important steps are[10]:

1. ■ Initial state with incoming partons. This state is fully described by specifying the beam energy and incident particle type.
2. ■ Colliding partons. Incoming valence quarks, sea quarks and gluons are described using the PDFs, which are obtained from external measurements. The range of possible Q can be quite large: low-x partons create "soft" collisions, which have to be described with different methods than "hard" collisions produced by high-x partons.
3. ■ Hard interaction. Q is large, and the interactions can be calculated using perturbation theory. Here, the production of Higgs (■) and Z boson (■) is depicted. If the event generator has at least `NLO` or multi-leg accuracy (see section III 4.3), it is also able to describe the emission of additional hard partons such as the gluon (■) shown in the top left part of the figure.
4. ■ Parton shower. All other emissions, as well as gluons splitting to quarks, are carried out by a `parton shower` algorithm. Parton showers will be discussed in section III 4.3.
5. ■ Hadronisation. When the parton shower stops, partons are bound into short-lived hadrons. Since Q is very low at this stage, perturbation theory cannot be applied. This stage is described with phenomenological models.

[10] The numbering of the stages does not imply that the stages have to be simulated in this order. Some stages such as the underlying event can be simulated in parallel.

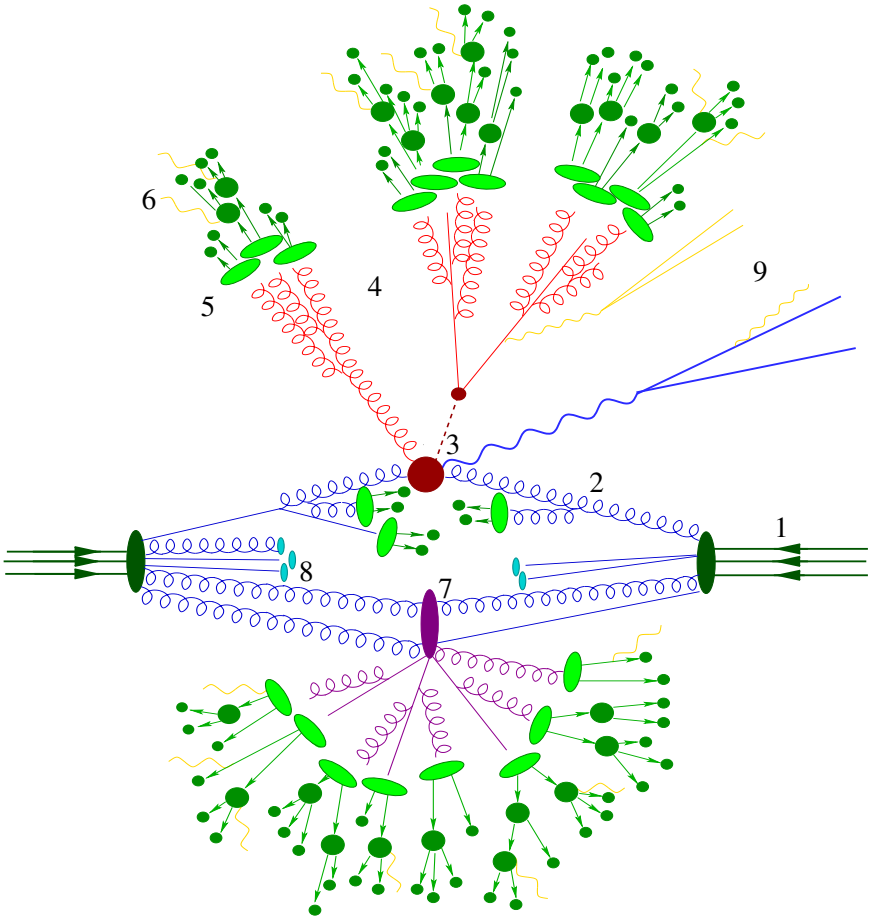


Figure III.15: Pictorial representation of the stages necessary to simulate a proton-proton collision producing a $ZH \to \ell\bar{\ell}b\bar{b}$ event with an additional jet. Inspired by [106], but redrawn for ZH. The numbered stages are explained in the text.

6. Decay of hadrons. Short-lived hadrons decay into longer-lived hadrons such as pions, kaons, *etc*. The decay chains are recorded because these hadrons can reach the detector. All particles in this stage are fed into detector simulations to estimate the response of the detector.
7. Secondary collision. Secondary parton collisions might take place. These are not part of the reaction of interest, but may contribute to the "underlying event". The underlying event needs to be taken into account because $VH \to Vb\bar{b}$ observables (*e.g.* $E_{\mathrm{T}}^{\mathrm{miss}}$) might be altered by additional activity in the event.
8. Proton remnants. The proton remnants are scattered into the very forward direction, and mostly escape detection because they are close to the proton beams. They are of minor importance for the reconstruction of the hard reaction.
9. Photon radiation. Photons may be radiated at any time from any charged object.

As illustrated by the figure, the simulation of collisions is a mixture of processes with high Q and low Q. The low-Q parts require a new approach because perturbative expansions break down when α_{s} is large.

III 4.2 Factorisation Theorem

The low-Q parts of the cross-section calculations can be approached by factorising the cross sections into a perturbative and a non-perturbative part. The factorisation theorem [101] provides a strategy to do this:

$$\mathrm{d}\sigma_{h_1,h_2} = \sum_{i,j} \int_0^1 \int_0^1 \mathrm{d}x_i \mathrm{d}x_j \sum_f \int \mathrm{d}\Phi_f \, \mathrm{PDF}_{i,h_1}(x_i, \mu_F^2) \, \mathrm{PDF}_{j,h_2}(x_j, \mu_F^2) \frac{\mathrm{d}\hat{\sigma}_{ij\to f}}{\mathrm{d}x_i \mathrm{d}x_j \mathrm{d}\Phi_f}$$

This equation implies that for the collision of two hadrons h_1 and h_2, the cross section can be obtained by summing over all initial state partons i, j, by integrating the respective PDFs over all momentum fractions x, by summing over all final states f and by integrating over the possible phase space elements $\mathrm{d}\Phi_f$ (*e.g.* all kinematically possible momenta of the final-state particles). The essence of the factorisation theorem is that the non-perturbative parts that govern the behaviour of the partons inside the protons, that is, the PDFs, and the perturbative part in the matrix element are independent. The total cross section is obtained by calculating the convolution of these factors. Since this holds independently of the hard process, of energy scales, accelerators and detectors, the PDFs can be measured externally, and be used for LHC calculations. The extrapolation of the PDFs between different energy scales is done using the `DGLAP` equations [102], see the following page.

The point where one switches from perturbative calculations to approximation through PDFs is set by the `factorisation scale`, μ_F. It is not a quantity that describes the nature of the interaction but merely a point for stopping matrix element calculations to prevent going into a regime where perturbation theory breaks down.

Skands [19] gives a simple argument why it is possible to separate the two regimes: low-Q interactions that take place inside the hadron must be at the confinement scale, the scale where α_s is large. This is $\Lambda \approx 200\,\mathrm{MeV}$. In natural units, this corresponds to a length scale of $\frac{\hbar \mathrm{c}}{\Lambda} = 1{\times}10^{-15}\,\mathrm{m}$, the size of the proton, and to a time scale of $\frac{\hbar}{\Lambda} = 1{\times}10^{-24}\,\mathrm{s}$. By contrast, hard interactions happen at an energy scale that is considerably higher, for example $\Lambda_\mathrm{hard} = \mathcal{O}(100\,\mathrm{GeV})$. This corresponds to $\frac{\hbar \mathrm{c}}{\Lambda} = 2{\times}10^{-18}\,\mathrm{m}$ and $\frac{\hbar}{\Lambda} = 7{\times}10^{-27}\,\mathrm{s}$. The time scale of the hard interaction is so short in comparison to interactions inside the proton that the colliding partons perceive the inner structure of the proton as frozen. Alternatively, one can argue that the interaction only probes a very small volume inside the proton, and therefore is not sensitive to the physics of confinement.

The factorisation shown above is incomplete, though: the low-Q regime *after* the collision also has to be factorised from the hard regime. This is achieved by including a `fragmentation function`. Similar to the PDFs, the fragmentation function describes the non-perturbative parts after the hard interaction when gluons are radiated, and jets form around the partons. The fragmentation function has to be defined with respect to an observable O:

$$\begin{aligned}\frac{\mathrm{d}\sigma_{h_1,h_2}}{\mathrm{d}O} = &\sum_{i,j} \int_0^1 \int_0^1 \mathrm{d}x_i \mathrm{d}x_j \sum_f \int \mathrm{d}\Phi_f \, \mathrm{PDF}_{i,h_1}(x_i, \mu_F^2) \, \mathrm{PDF}_{j,h_2}(x_j, \mu_F^2) \\ &\cdot \frac{\mathrm{d}\hat{\sigma}_{ij\to f}}{\mathrm{d}x_i \mathrm{d}x_j \mathrm{d}\Phi_f \, \mathrm{d}\hat{O}} \cdot D_f(\hat{O} \to O, \mu_F^2)\end{aligned}$$

O could, for example, be the energy of a jet. In this case, $\hat{O}$ would be the energy of the corresponding parton, and the fragmentation function D_f describes how the parton energy translates to the observed jet energy. To obtain an inclusive cross section for all jet energies, an additional convolution with D has to be performed. The fragmentation function also depends on a factorisation scale that defines at which point the fragmentation function takes over from the hard matrix element calculation. Since no exact calculations are possible, such functions have to be tuned to data using measurements.

III 4.3 Parton Showers, Multi-Leg Generators and Hadronisation

Parton showers are a successful strategy to approximate fragmentation functions. Although the partonic cross sections for some final state F with n extra partons can be predicted if the partons are sufficiently hard and also separated from other objects, this becomes impossible for jets: when soft gluons are emitted from a parton, or when gluons are emitted collinearly with a parton, QCD amplitudes diverge. This can be seen by examining the cross section for the gluon emission shown in figure III.16. Normalised to the cross section of $q\bar{q}$ production, one finds [103]:

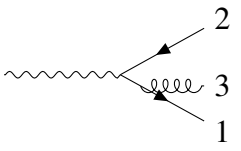


Figure III.16: Gluon radiation in the final state.

$$\frac{\mathrm{d}\sigma(q\bar{q}+g)}{\sigma(q\bar{q})} \approx \frac{\alpha_\mathrm{s}}{2\pi}\frac{\mathrm{d}Q^2}{Q^2}\frac{4}{3}\frac{1+z^2}{1-z}\,\mathrm{d}z. \qquad \text{(III.9)}$$

$Q^2 = m_{1+3}^2$ is the "virtuality" or mass of quark 1 before the gluon emission happens, and z determines how energy is distributed between quark and gluon: $E_q = zE^{\text{ini.}}$, $E_g = (1-z)E^{\text{ini.}}$. Quarks are assumed to be massless, and it is assumed that quark 2 carries most of the energy. If this is not the case, the labels 1 and 2 can be exchanged. The Q^{-2}-dependence of equation (III.9) shows the collinear or mass singularity of QCD: with vanishing Q, which means that the gluon is emitted collinearly with the quark, the amplitude diverges. If the gluon is soft, that is, $1-z \to 0$, the amplitude diverges as well.

Despite the singularities, equation (III.9) is very useful: it can be used to describe gluon emissions for all final states and all energies where the quark mass is small compared with Q. Similar equations exist for the probability of other emissions like gluons splitting to quarks and gluons emitting gluons. The resulting functions are the **A**ltarelli-**P**arisi splitting functions, which are a key ingredient of the DGL**AP** equations [102]:

$$\begin{aligned}
\mathrm{d}\mathcal{P}_{a\to bc} &= \frac{\alpha_\mathrm{s}}{2\pi}\frac{dQ^2}{Q^2}P_{a\to bc}(z)\,\mathrm{d}z \\
P_{q\to gq} &= \frac{4}{3}\frac{1+z^2}{1-z} \\
P_{g\to gg} &= 3\frac{(1-z(1-z))^2}{z(1-z)} \\
P_{g\to q\bar{q}} &= \frac{n_f}{2}(z^2+(1-z)^2) \qquad n_f = \#\text{ of quark flavours}
\end{aligned}$$

Starting from one initial parton, the repetitive application of the splitting functions describes the formation of a jet. This leads to a shower of particles, the parton shower. The singularities in the splitting functions have to be avoided, though, by not adding splits softer or "more collinear" than certain shower-evolution criteria. Different approaches are pursued to keep probabilities finite, see, for example, [99]. Shower evolution is stopped at the `hadronisation scale`, $\mathcal{O}(1\,\mathrm{GeV})$, where α_s starts to rise above levels acceptable for perturbative expansions.

The factorisation theorem creates two regimes: in the rather collinear and soft regime, which defines how jets evolve, parton showers provide a good approximation of all branchings that may occur. In the well-separated and hard regime, where the coupling and thus the branching probability is low, a matrix element calculation can be used.

Often, one needs matrix elements that describe a certain process *plus* the emission of multiple additional partons, such as W production with multiple jets, an important background of the VH analysis. Such matrix elements with multiple well-separated additional partons are called "multi-leg" matrix elements.

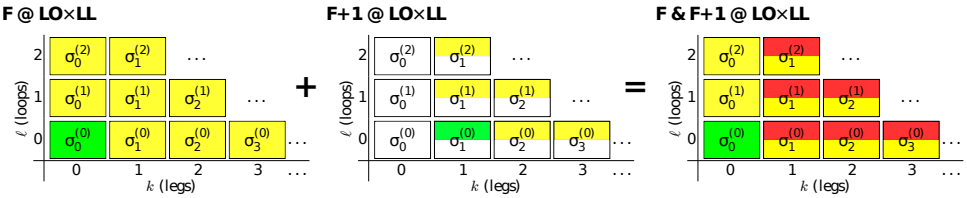


Figure III.17: Double-counting problem for adding a parton (additional leg) to a process F when parton showers are used [19].

The cross section for a process F with k additional legs has the following algebraic structure in QCD:

$$\sigma_{F+k} = \alpha_s^k(\ln^{2k} + \ln^{2k-1} + ... + \ln + \mathcal{F})$$

The notation $\ln^\lambda$ denotes functions of transcendentality λ, for which the logarithms of the given power are representative, while $\mathcal{F}$ is a rational function. For details see Skands [19]. Parton showers can approximate the coefficients of the functions with $\ln^{2k}$ and $\ln^{2k-1}$ for *all* orders in k (`LL`, leading-logarithmic parton shower). Diverging amplitudes due to soft or collinear emissions cancel with virtual corrections from loops, as stated by the KLN theorem [104, 105]: with k additional legs and ℓ virtual loops, probabilities remain finite when performing a full calculation – order by order – for all combinations of k and ℓ with $k + \ell \leq n$ (`resummation`). This is illustrated in the left sketch in figure III.17: the `LO` matrix element (■) predicts the amplitudes for a process without any additional loops or legs, $k = \ell = 0$. Adding a parton shower with appropriate resummation approximates the leading-log (`LL`) coefficients, (partly) covering amplitudes for loops and legs to all orders in n (■). The coverage is only partly, though, because the parton showers only extend to the regions of rather collinear or soft emissions.

To increase the precision of the cross-section prediction, multi-leg generators add matrix elements such as the $F+1$ process (*e.g.* $W+1$ jet). An example of a multi-leg generator is `Sherpa` [106], which is used to simulate all $V+$jets backgrounds of the $VH \to Vb\bar{b}$ analysis. When adding the $F+1$ (hard) matrix element, one has to restrict the amplitudes to the well-separated phase space, though, to avoid divergent terms. This is indicated by half-filled boxes in the middle sketch in figure III.17. Applying the parton shower to the additional parton again yields the `LL` coefficients for amplitudes with more legs or loops. However, since the parton showers for the process F (sketch on the left) also reproduce these coefficients, the sum $F + (F+1)$ at `LOxLL` double counts the leading logarithms (■, right sketch in figure III.17). The same considerations hold for `N`$^{\text{n}}$`LO` generators, but here more legs and loops are covered by the hard matrix element in the first place.

Such double counting is mitigated with `matching`. Different strategies to match parton shower and hard matrix elements are possible. One of them is "slicing": the phase space is sliced into a region exclusively treated by the hard matrix element and a region exclusively treated by the parton shower. This was first introduced in `HERWIG` [107–109], which is used in this work to estimate parton shower uncertainties for the $VH \to Vb\bar{b}$ analysis. Further details can be found in [19].

An important consequence for the $VH \to Vb\bar{b}$ analysis is that cross-section predictions are more challenging than only predicting the amplitudes of the hard matrix element, especially when multiple legs and matching are required. Cross-section uncertainties are therefore larger than indicated by μ_R variations, α_s measurements and PDF uncertainties. In the $VH \to Vb\bar{b}$ analysis, $F+2$ and $F+3$ predictions are required to simulate the $W+$jets and $Z+$jets backgrounds. The choice of the shower algorithm and matching strategy, as well as adjustments of cutoff scales and shower or matching parameters will affect the prediction of the cross sections for such $V+n$ jets processes. Therefore, cross-section

uncertainties beyond the usual scale and α_s uncertainties have to be taken into account. In the $VH \to Vb\bar{b}$ analysis, this is achieved by adding extra cross-section uncertainties for different jet-multiplicity bins. The magnitude of these uncertainties is determined by comparing Monte Carlo generators with different shower implementations and matching procedures. Whenever possible, the normalisations of Monte Carlo predictions will be measured in data to constrain the associated cross-section uncertainties. The same holds for the prediction of the parton flavour in additional legs. For the $VH \to Vb\bar{b}$ analysis, it is not sufficient to only predict/measure the jet multiplicity, but also whether these jets are b-, c- or light-flavour jets. $W+bb$ events are evidently harder to separate from $WH \to \ell\nu b\bar{b}$ events than $W+ll$ events. Jet-flavour uncertainties are therefore also estimated by comparing Monte Carlo generators with varying shower/matching/radiation settings, and – where possible – they are measured or at least constrained using the data. These uncertainties nevertheless remain among the dominant background uncertainties of the $VH \to Vb\bar{b}$ analysis.

Hadronisation

Parton showers cannot run indefinitely, because α_s becomes large at the hadronisation scale $\mu_H \lessapprox 1\,\text{GeV}$. At this scale, parton showers have to be stopped, and partons are bound into colour-neutral hadrons. The generator `Pythia` [43, 110], for example, uses a string model for hadronisation to take into account the attractive potential between partons. This is based on the insight that the potential between two colour charges in the regime of confinement is essentially

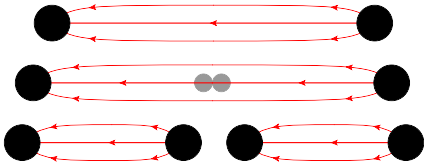


Figure III.18: String model of hadronisation used by `Pythia` [19].

$$V(d) = 0.9\,\text{GeV}\,\text{fm}^{-1} \cdot d\,,$$

where d is the distance between the partons.

The string model is illustrated in figure III.18: if two quarks start to separate, their kinetic energy is converted into potential energy, which is proportional to the length of the strings. If the potential energy is large enough for a new pair of quarks to be produced, the string breaks, and a pair of quarks appears. This process continues until the binding energies and quark flavours match those of known hadrons, that is, bound states of quarks. At this stage, the partons are assigned to colour-neutral hadrons, ensuring that no free coloured particles remain. If the hadrons created in this step are unstable, they decay to lighter hadrons. There is considerable freedom, for example, in selecting the spin states of the hadrons produced, in tuning the number of mesons vs. baryons, in balancing the momentum fractions each string is assigned, *etc*. Hadronisation models therefore have to be tuned to data.

Most Monte Carlo generators used in this thesis use `Pythia8` or `Pythia6` to simulate the parton showers with the string model of hadronisation. Only `Sherpa` [106], the multi-leg generator used for V+jets events, treats matrix element calculation, matching and hadronisation in a single process. `Sherpa`'s hadronisation model, as well as the model of `Herwig` [107], is different from the string model. Comparisons between these three shower generators are therefore used to derive shower and hadronisation uncertainties, which will be used to devise a likelihood model for the statistical analysis of the data in section IV 6.

III 4.4 Other Aspects of Simulating LHC Collisions

Underlying Event The underlying event is part of a proton-proton collision but independent from the hard collision. Secondary collisions within the same proton and the proton remnants (7 and 8 in the list on page 53) lead to additional activity in the detector that is not part of the hard event. The cross section of such Multiple Parton Interactions (MPI) depends on the momentum transfer: the lower the transverse momentum produced in the parton-parton interaction, the higher the cross section. It can, in fact, be so high, that the total inelastic proton-proton cross section is exceeded. For LHC, this happens at $p_{\mathrm{T}} = 4$ to $5\,\mathrm{GeV}$ [19]. Below this energy regime, more than one parton collision is expected per proton collision. However, since high transverse momenta (*i.e.* comparable to typical objects from a $H \to b\bar{b}$ decay) from MPI are rare, these have only minor importance for the $H \to b\bar{b}$ analysis. An ATLAS measurement of underlying event characteristics can be found in [111], details on the simulation of MPI, parton showers and hadronisation in [112].

Pileup Since the LHC collides bunches of $\mathcal{O}(1\times10^{11})$ protons each, multiple proton-proton collisions can occur, or "pile up", in a single bunch crossing. The amount of pileup is determined by the bunch currents and the transverse extension of the bunches. It therefore scales with the luminosity: high luminosities coincide with high pileup, unless higher luminosities are reached by increasing the number of colliding bunches. The typical number of simultaneous collisions in the data of 2012 ranges from 5 to 45, with a mean number of 20.7 [113]. Pileup collisions predominantly produce low-transverse-momentum objects in comparison with the hard collision that triggers the recording of an event. There is usually only one hard collision per event because the hard interaction cross section is low in comparison to the total proton-proton interaction cross section. Objects from pileup collisions can be separated from objects from the hard collision because collision vertices are most of the time separated in the z-direction. Figure III.12 on page 49 shows a part of the interaction region with well-separated collision vertices. The interaction region extends to about $60\,\mu\mathrm{m}$ in xy-direction but to $\sim 10\,\mathrm{cm}$ in z-direction. When jet and lepton tracks are traced back to a primary vertex, the z-distance between the primary vertices can be used to suppress pileup objects. Most of the tracks from pileup collisions in figure III.12 have in fact been removed from the event display for better visibility.

Energy measurements in the calorimeters are also corrected for pileup effects. Contributions from pileup objects to jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ are subtracted from the reconstructed energies, but high levels of pileup lead to higher energy uncertainties, especially for $E_{\mathrm{T}}^{\mathrm{miss}}$. Furthermore, if two collisions are separated by less than $\sim 100\,\mu\mathrm{m}$, their vertices cannot be resolved, any more. When pileup levels are high, objects from different collisions might thus seem to originate from the same vertex. High pileup therefore increases the uncertainties of energy measurements, and slightly worsens the background suppression of object identification algorithms. High pileup also yields higher data statistics, though, which outweighs these disadvantages for many analyses.

Pileup is simulated using a specially tuned version of `Pythia8` [112] with `MSTW2008LO` PDFs [63]. Pileup collisions are simulated independently from hard processes and overlaid onto simulated events.

Detector Simulation The simulation of the detector response is a very important step, which takes considerable computing power. After simulating the parton shower, hadronisation and decay of hadrons, particles are passed to `GEANT4` [114], a software package for simulating the interactions of particles with matter. A complete model of ATLAS, including all detector components such as sensitive and insensitive material (*e.g.* cables, support structures *etc.*) is represented in `GEANT4`.

All interactions such as Bremsstrahlung, calorimeter showers, charge depositions *etc*. are simulated. The response of the detector, as well as trigger decisions are saved, and subsequently processed with the same reconstruction algorithms that are also used for LHC collision data.

Since the detailed simulation of calorimeter showers is time consuming, ATLAS employs a fast simulation method, `ATLFAST II` [115]. Predefined calorimeter showers are selected based on the incident particle type and energy, and are placed into an event instead of calculating the full interaction of the incident particle with the detector. This reduces the simulation time by more than an order of magnitude. Small differences between reconstructed observables obtained with full and fast simulation are observed, though. These are corrected for with additional calibrations. Most of the Monte Carlo samples for the $VH \to Vb\bar{b}$ analysis use the fast detector simulation.

Event Weights Many Monte Carlo generators assign weights to events. These weights are used to re-shape the probability distributions being generated. This is necessary because Monte Carlo generators sample elements of the phase space with pre-computed frequency: phase space elements with higher cross sections are sampled more often. These phase space elements are, however, not infinitely fine-grained, and the cross sections inside such an element may vary. This is corrected by assigning a weight to each Monte Carlo event that adjusts the probability of selecting a phase space element to the probability of actually observing a particular event.

Reweighting Similar to generator weights, event weights can be used when a Monte Carlo generator does not correctly describe the distributions of certain observables. Instead of producing new Monte Carlo samples, the probability distributions can be reshaped by applying correction weights to the events. Such weights can, for example, correct identification efficiencies, trigger response probabilities, the description of kinematic variables, *etc*. Reweighting does not change the values of any observables in an event, since event kinematics are of course unambiguously determined by four-momentum conservation. Reweighting changes the probability of observing an event, thus shaping the distributions observed when sampling many events. The corrections necessary for the $VH \to Vb\bar{b}$ analysis will be discussed in chapter IV.

Reweighting can further be used to estimate systematic uncertainties. For this, multiple Monte Carlo samples are created that differ only with respect to event weights: a "nominal" sample that only has generator and correction weights, $w_{\text{generator}} \cdot \prod w^i_{\text{corr}}$, and samples with altered weights, $w_{\text{generator}} \cdot w_{\text{sys. unc.}} \cdot \prod w^i_{\text{corr}}$. The additional weights parametrise changes in the predicted distributions that would occur if identification efficiencies or selection probabilities were different. In this way, event samples representing various uncertainties can be created without having to generate new Monte Carlo events.

Monte Carlo Truth To be able to analyse cross-section predictions independently of the detector response, particles produced after the matrix element / hadronisation step are saved without undergoing detector simulation and reconstruction. These particles are often referred to as "truth" particles. The truth records can, for example, be used to study the detector response by comparing truth observables to reconstructed quantities. The truth records are, however, generator dependent: since parton showers, hadronisation scale and hadronisation models differ, the only reliable stage for the analysis of truth particles is *after* hadronisation. This is why jet flavour studies or b-tagging calibrations in the $H \to b\bar{b}$ analysis are only based on truth *hadrons*. If truth hadrons with b or c quarks are found inside a jet, the jet is considered to originate from a b or c quark. In other cases, the jet is treated as a light jet. Further applications of the truth records will be discussed in chapter IV.

CHAPTER IV

Search for $H \to b\bar{b}$ Decays with ATLAS

The first multivariate analysis of the $VH \to Vb\bar{b}$ channel in ATLAS

In this chapter, the first of the two analyses discussed in this thesis will be introduced: the ATLAS $VH \to Vb\bar{b}$ analysis [1]. Since the final implementation of this analysis is an effort of a larger group in ATLAS, this chapter will focus on aspects that were studied the most during the work for this thesis. In particular:

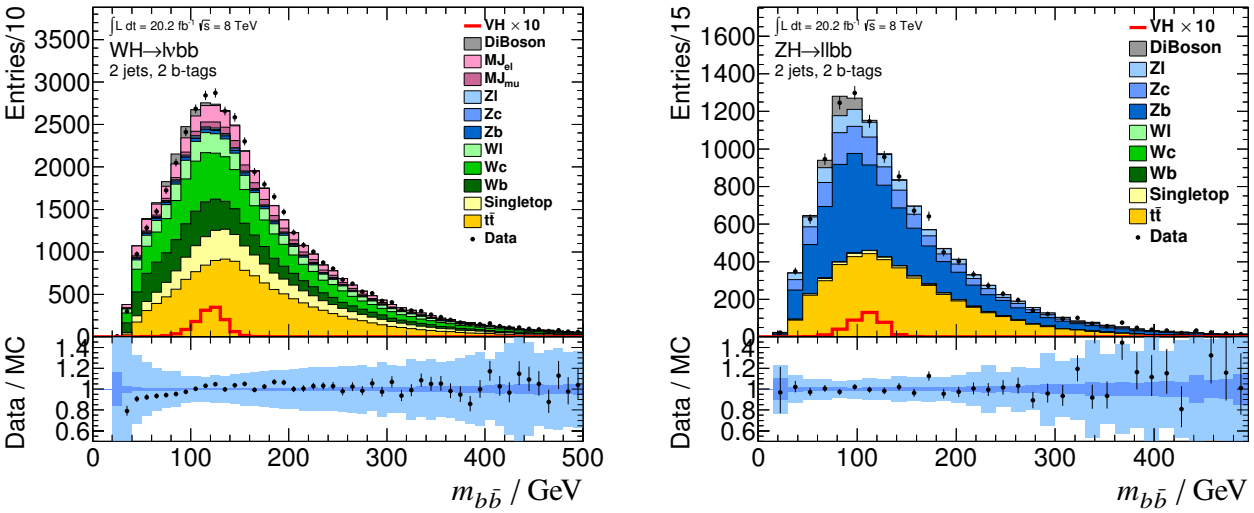
- Usage of multivariate classifiers instead of a cut-based analysis of the dijet mass spectrum
- Optimising the event selection for use with multivariate classifiers
- Introducing b-tagging information into the multivariate analysis
- Truth-flavour tagging with b-tagging at multiple working points
- Resolution correction for b-jets
- Estimation of the QCD multijet background

For other aspects such as the analysis of final states without charged leptons, one may refer to the ATLAS publication [1].

This chapter will start with an introduction to the backgrounds of the $VH \to Vb\bar{b}$ analysis in section IV 1. Afterwards, section IV 2 will focus on how signal and background distributions are estimated, discuss deficits or uncertainties of the Monte Carlo simulations, and review the necessary corrections.

In section IV 3, the use of boosted decision trees (`BDTs`) for the $VH \to Vb\bar{b}$ search in ATLAS will be discussed, which was pioneered at the University of Bonn [7]. Here, it will also be discussed why the event selection outlined in section IV 1 is better for the use with multivariate classifiers than the selection used in previous ATLAS analyses [3, 4].

After a short excursion to using machine-learning methods to check the quality of Monte Carlo simulations in section IV 4, the chapter will finish with the statistical analysis of the data in sections IV 5 and IV 6. It will be shown how BDT classifiers, simulations and measurements are brought together in a likelihood model that can be used to quantify the compatibility of the data with the presence or absence of $VH \to Vb\bar{b}$ decays.



(a) WH: 1 lepton + $E_{\mathrm{T}}^{\mathrm{miss}}$ **(b)** ZH: 2 leptons

Figure IV.1: Backgrounds of the $VH \to Vb\bar{b}$ search. The figures show the invariant mass of two b-tagged jets in events with two jets and one or two leptons. The measured distributions are shown with black markers, and background simulations are shown as coloured histograms. The $H \to b\bar{b}$ signal is superimposed (■), and magnified 10 times. The inset at the bottom shows the ratio of the measured over the simulated distributions. The dark inner band (■) shows the statistical uncertainty of the background prediction, the light blue band (■) shows the quadratic sum of systematic uncertainties. Uncertainties for data points indicate the 68 % confidence interval of a Poisson distribution, and are therefore asymmetric. Background cross sections are scaled to match the data.

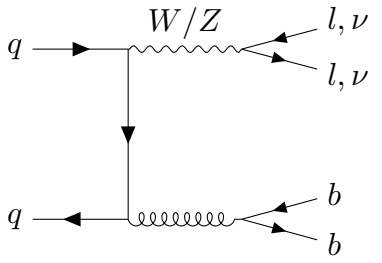
IV 1 Backgrounds, Signal Selection and Event Categories

This thesis is focussed on Higgs production in association with vector bosons because of experimental requirements like the ATLAS triggers or the high levels of b-jet production at the LHC, which were discussed in chapter III. If vector bosons decay into leptons, clear signatures for the ATLAS trigger system are created because in comparison to all-hadronic final states, leptons are rare in LHC collisions.

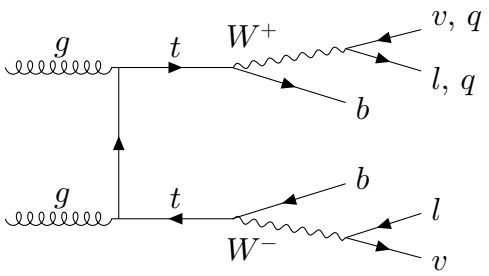
Figure IV.1 shows distributions and relative importance of the backgrounds of the $VH \to Vb\bar{b}$ search. All processes with the signature 'two b-jets and one or two leptons' are relevant for this search. The figures show `ROOT` [120] histograms of the invariant mass of the two b-jets, the most powerful observable to separate background processes from $H \to b\bar{b}$ production. $VH \to Vb\bar{b}$ simulations are shown with a red line, peaking at the Higgs mass of $\sim 125\,\mathrm{GeV}$. The width of the Higgs peak is determined by the jet energy resolution of the calorimeters (section III 2.2), which is one of the dominant uncertainties of the $VH \to Vb\bar{b}$ search. The following background processes need to be taken into account in order to correctly describe the data:

■ **W+jets** A leptonically decaying W boson produced in association with jets, figure IV.2(a). The final state in these reactions is identical to WH production if the jets being produced are b-jets. If the invariant mass of the jets is close to 125 GeV, the W+bb component is therefore irreducible. The jet multiplicities and quark flavours of the W+jets backgrounds cannot be predicted accurately, though, and are therefore measured in data where possible.

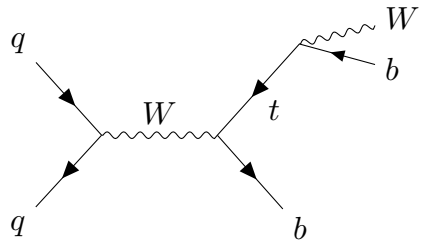
■ **Z+jets** Z+jets production, also figure IV.2(a), is similar to ZH production if two b-jets and two leptons are produced. Since the Z boson has a higher mass than the W boson and because $Z \to \ell\bar{\ell}$ is less likely than $W \to \ell\nu$, it is about one order of magnitude less likely than W+jets production.



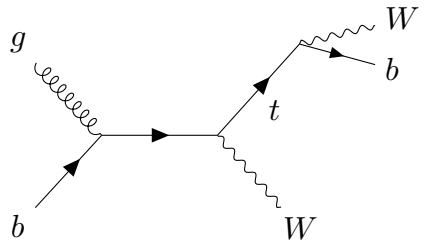
(a) $q\bar{q}$-induced $W{+}bb$ and $Z{+}bb$ production with gluon splitting [116, 117]



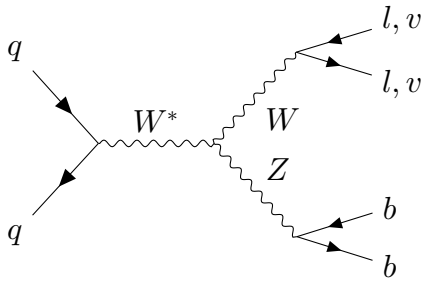
(b) gg-induced $t\bar{t}$ production



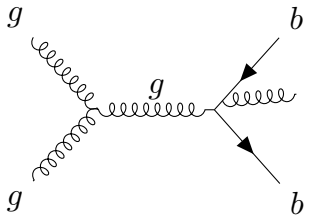
(c) Single-top s-channel production [118]. Rotation by $-90°$ yields a t-channel diagram.



(d) $W{+}t$ production [119]



(e) Diboson production via a virtual W boson



(f) QCD multijet flavour creation (s-channel)

Figure IV.2: Selection of leading-order diagrams for background processes of the VH search. The diagrams shown do not cover all possible production modes. Time runs from left to right.

$t\bar{t}$ Top-quark pairs almost exclusively decay into a pair of b quarks and a pair of W bosons. These can subsequently decay into leptons or quarks, see figure IV.2(b). If one charged lepton is produced, the $t\bar{t}$ events resemble WH events, but they often feature additional jets from the decay of one of the W bosons. If both W bosons decay leptonically, no additional jets are expected, and $t\bar{t}$ is a background to ZH production. If either the additional jets or a lepton is missed during the reconstruction, $t\bar{t}$ is an irreducible background of WH.

Single t The production of a single top quark has a lower cross section than $t\bar{t}$ pair production. However, s- and t-channel production (figure IV.2(c)) in leading order produce a final state with one b-jet, a W boson and an additional jet, which can either be a second b-jet or be misidentified as

one. $W+t$ production in figure IV.2(d) produces one b-jet and two W bosons. Depending on how the bosons decay, this creates final states similar to $WH \to \ell\nu b\bar{b}$. This background is practically irrelevant for a two lepton final state.

Diboson The production of two vector bosons like in figure IV.2(e) can lead to a final state that is identical to $WH \to \ell\nu b\bar{b}$ or $ZH \to \ell\bar{\ell}b\bar{b}$ if one of the Z bosons decays as $Z \to b\bar{b}$. It can partly be separated from $H \to b\bar{b}$ production because $m_Z < m_H$, but due to the limited resolution of the calorimeters the distributions overlap slightly. Because of the high similarity to the VH signal, this process can be used as a benchmark for the $H \to b\bar{b}$ measurement: in section VI 2, the signal strength parameter of diboson production is measured with the methods that are developed for the VH final state in this thesis.

QCD Multijet Gluon collisions like in figure IV.2(f) or the radiation of gluons from other objects can produce a pair of b-jets. Unlike the other backgrounds, these processes are entirely mediated by QCD interactions. They therefore have very high cross sections, in particular, because the LHC is essentially a gluon collider (section III 1.1 and figure III.4). The ratio of Higgs production to $b\bar{b}$ production in QCD, for example, is [64, 121]:

$$\frac{\sigma(ggH)}{\sigma(bb)} \approx \frac{21\,\text{pb}}{200\,\mu\text{b}} \approx 1.0 \times 10^{-7}.$$

By restricting the analysis to VH production, this ratio even falls to 5×10^{-10}. By requiring leptons in the final state, most of the QCD multijet events can be suppressed, though. Nevertheless, since leptons are occasionally produced in the vicinity of jets, a sizeable number of QCD multijet events is selected in the VH analysis. Furthermore, hadrons from the jets can be misidentified as electrons. In this case, an all-hadronic final state like in figure IV.2(f) can also be selected in the VH analysis. For this reason, QCD multijet events with an object identified as electron (■) are more likely than events with muons (■), which is visible in figure IV.1(a).

The W+jets and Z+jets simulations are categorised according to the flavours[1] of the jets created in association with the vector bosons. This is achieved by searching the truth records of the Monte Carlo generators for hadrons overlapping with the jet. Depending on whether these hadrons contain b, c or lighter quarks, b, c or light labels are assigned to the jets. The flavour categories used in figure IV.1 are $W+b := W + (bb + bc + bl)$, $W+c := W + (cc + cl)$ and $W+l$ for the rest. In events with three jets, the labelling is done using the jets with the highest transverse momenta. In figures where specific flavour combinations such as $W+bl$ need to be shown separately, these are taken out from the inclusive categories $W+b$ and $W+c$. In the ATLAS $H \to b\bar{b}$ publication [1], $W+cl$ is always singled out because on top of a multi-leg process, strange-gluon scattering $sg \to W+c$ can also produce this final state. However, an ATLAS measurement indicates that the contribution of sg scattering is less than 28 % of the total $W+cl$ production. It is therefore not plotted separately in this thesis[2].

Table IV.1 lists the cross sections of the background processes and the Monte Carlo programs used to simulate them. The choice of Monte Carlo programs is dictated by the availability of high-statistics simulations provided by ATLAS. `PowHeg` is used for simulating top processes, but for t-channel production of single top quarks, `AcerMC` was found to better describe the data. Both programs generate `NLO` matrix elements using `CTEQ10` (`NLO`) PDFs [127]. Parton showers and hadronisation are carried out by `Pythia6` based on the Perugia2011C tunes [128] with `CTEQ6L1` [129] (`LO`) PDFs.

[1] "Jet flavour" = flavour of the quark or hadron sparking the jet.
[2] See appendix B 1 for details on the fraction of sg scattering.

Table IV.1: Cross sections times branching ratios and Monte Carlo programs used for estimating backgrounds of the VH search ($\sqrt{s} = 8\,\text{TeV}$). If specific final states are listed, the branching ratios quoted are only for these states. If no final state is given, an inclusive simulation is used (*i.e.* $BR \equiv 1$). l denotes a charged lepton including τ. "Py" = Pythia. More details regarding the signal cross sections can be found in table IV.3.

Process	Simulation	$\sigma \times BR$		Remarks
$gg/qg/qq \to b\bar{b}/bb$ (QCD)	Data-driven	~200	μb	
	LO Multi-leg			
$W \to l\nu$ + jets	Sherpa 1.4.1 [106]	12.1	nb	
$Z/\gamma* \to l\bar{l}$ + jets	Sherpa 1.4.1 [106]	1.2	nb	$m_{ll} > 40\,\text{GeV}$
	NLO			
$t\bar{t}$	PowHeg+Py6 [122]	252.9	pb	
t (t-chan)	AcerMC+Py6 [123]	87.8	pb	
tW	PowHeg+Py6 [122, 124]	22.4	pb	
t (s-chan)	PowHeg+Py6 [122, 125]	5.6	pb	
WW	PowHeg+Py8 [122, 126]	52.4	pb	
$WZ \to l\nu\, q\bar{q}$	PowHeg+Py8 [122, 126]	4.9	pb	
$WZ \to q\bar{q}\, l\bar{l}$	PowHeg+Py8 [122, 126]	1.6	pb	$m_{ll} > 20\,\text{GeV}$
$ZZ \to l\bar{l}\, q\bar{q}$	PowHeg+Py8 [122, 126]	1.2	pb	$m_{ll} > 20\,\text{GeV}$
	LO			
$WH \to l\nu\, b\bar{b}$	Pythia8 [43]	87.5	fb	
$ZH \to l\bar{l}\, b\bar{b}$	Pythia8 [43]	19.4	fb	

The W+jets and Z+jets backgrounds are simulated with Sherpa, which also uses CTEQ10 NLO PDFs [127]. This version of Sherpa is a multi-leg generator at LO (see section III 4.3), which also carries out the parton showers and hadronisation to achieve consistent matching.

Diboson processes are also simulated with PowHeg and CTEQ10 (NLO) PDFs. In contrast to the generators used for top quark events, the diboson generators are interfaced to Pythia8 with CTEQ6L1 (LO) PDFs for parton shower and hadronisation. The Pythia8 showers were tuned to data by ATLAS ("AU2 tune") [112].

The Higgs signal is simulated at LO with Pythia8, but NLO shape corrections and NNLO cross-section corrections are applied, see section IV 2.1.

Even though the background normalisations of W+jets, Z+jets and $t\bar{t}$ are measured in data, the inclusive cross-section predictions are corrected using k-factors: to NNLO for V+jets [130], to NNLO+NNLL for $t\bar{t}$ [131], single-top t-channel [132], s-channel [133] and Wt [134].

QCD multijet production is entirely estimated from data. Although it is possible to simulate QCD jet production with Monte Carlo generators, almost none of these events enter the $VH \to Vb\bar{b}$ phase space of interest for this thesis. This is because isolated leptons with sufficient momentum are required during the event selection, which is very rare in QCD multijet events. Therefore, obtaining sufficient Monte Carlo statistics for the multijet background is not possible. Nevertheless, it is an important background due to the high cross section (table IV.1). A data-driven estimation is therefore inevitable.

Table IV.2: Object selection cuts for the $VH \to Vb\bar{b}$ analysis. IP = impact parameter. Calorimeter and stand-alone muons are not included in the medium and tight leptons.

Loose Leptons	e	μ		
		Combined	Calorimeter	Standalone
$\|\eta\|$	< 2.47	< 2.7	< 0.1	2.5–2.7
$E_{\mathrm{T}}, p_{\mathrm{T}}$ / GeV	> 7	> 7	> 20	> 7
Transverse IP / mm		< 1	< 1	
Longitudinal IP / mm		< 10	< 10	
$\mathrm{Iso}_{\mathrm{Track}}$	< 0.1	< 0.1	< 0.1	
Quality	VeryLoose LH	Tight MuID + Standard track requirements		

Medium Leptons	e	μ
Quality	Loose e	Loose μ + Combined
$E_{\mathrm{T}}, p_{\mathrm{T}}$ / GeV	> 25	> 25
$\|\eta\|$	< 2.47	< 2.5

Tight Leptons	e	μ
Quality	Medium e + VeryTight LH	Medium μ
$\mathrm{Iso}_{\mathrm{Track}}$	< 0.04	< 0.04
$\mathrm{Iso}_{\mathrm{Calo}}$	< 0.04	< 0.04

Jets	Signal	Veto
E_{T} / GeV	> 25 ($1^{\mathrm{st}} > 45$)	> 30
$\|\eta\|$	< 2.5	2.5–4.5
$\|\mathrm{JVF}\|$	> 0.5	

IV 1.1 Selection of Objects

Not all objects reconstructed in ATLAS can be used for the VH analysis. Several cuts need to be applied to separate misidentified objects or objects from pileup collisions from those that originate from the hard collision. These cuts are shown in table IV.2.

Most care is taken for leptons. This is because isolated leptons are the main handle to suppress backgrounds from QCD-only interactions in the $VH \to Vb\bar{b}$ analysis. Since τ leptons are not targeted because of the more involved τ identification as well as trigger requirements, see section III 3.1, the term "lepton" will refer only to e and μ in the following – the leptons that can be reconstructed directly. Leptonically decaying τ leptons contribute to the final states analysed in this thesis, and are therefore included in the simulations, but the analysis is not optimised to separate such events from events with prompt electrons or muons. Because of additional neutrinos produced in the τ decays, only a few τ events are actually selected in the VH analysis.

The leptons are sorted into three categories: loose, medium and tight, see again table IV.2. Loose leptons must have a transverse momentum of at least 7 GeV, and must be isolated. This means that the total transverse momentum of all tracks reconstructed in a cone around the lepton is required to be at

least 10 times lower than the lepton momentum:

$$\text{Iso}_{\text{Track}} = \frac{\sum_{\Delta R<0.2} p_{\text{T}}^{\text{Tracks}}}{p_{\text{T}}^{\text{Lepton}}} < 0.1 \tag{IV.1}$$

This suppresses leptons reconstructed in the vicinity of hadrons and jets, which mostly occur in QCD multijet events. Loose electrons are further required to pass a set of standard ATLAS electron quality cuts, to be inside the barrel calorimeter region ($|\eta| < 2.47$) and to pass the `Very Loose Likelihood` identification (section III 3.1, page 44).

Loose muons are reconstructed from the combined, stand-alone and calorimeter-tagged muons described in section III 3.1. The last two sources of muons are used in regions with low acceptance: stand-alone muons in the forward regions, which are not covered by the tracking detectors, and calorimeter-tagged muons in the centre of the detector where support structures are installed. This is reflected by the η cuts in table IV.2. If a track in the inner detector is matched to the track of a muon, it is required to be associated to the primary vertex within $0.1\,\text{mm}$ in transverse and $10\,\text{mm}$ in beam direction. This suppresses muons from pileup collisions. For stand-alone muons, this is not possible, though.

Medium leptons are loose electrons or loose combined muons with $p_{\text{T}} > 25\,\text{GeV}$. These can trigger the recording of an event because the trigger thresholds for electron and muon triggers are $p_{\text{T}} > 24\,\text{GeV}$. Therefore, at least one medium lepton is required in all events of the $VH \to Vb\bar{b}$ analysis because it must be ensured that one of the leptons actually triggered the recording of an event.

Tight leptons are medium leptons with additional isolation requirements: the track isolation, equation (IV.1), is tightened to 0.04, and a calorimeter-based isolation is required in addition. This isolation is similar to the track isolation, but energy clusters in the calorimeter are compared to the momentum of the lepton:

$$\frac{\sum_{\Delta R<0.3} E_{\text{T}}^{\text{Calo}}}{p_{\text{T}}^{\text{Lepton}}} < 0.04 \tag{IV.2}$$

For electrons, the likelihood identification requirement is also tightened to `Very Tight Likelihood` to achieve better background suppression.

Jets in the VH analysis are sorted into two categories: signal and veto. Signal jets are required to have a transverse momentum of more than $25\,\text{GeV}$, and to be in the barrel region of the detector ($|\eta| < 2.5$). The leading jet in p_{T} is even required to have $p_{\text{T}} > 45\,\text{GeV}$. Jets in the forward regions ($2.5 < |\eta| < 4.5$) with a transverse momentum of more than $30\,\text{GeV}$ are labelled "veto jet". If veto (*i.e.* forward) jets are found in a collision, an event is discarded for two reasons: firstly, b-tagging cannot be applied to forward jets. It is therefore not possible to determine if the forward jets originate from a $H \to b\bar{b}$ decay. Secondly, many background processes produce numerous jets in forward direction. Discarding these events therefore reduces the number of background events.

Signal jets are further required to have a "Jet Vertex Fraction (JVF)" of at least $50\,\%$. The JVF compares the scalar sum of track momenta from tracks associated to the primary vertex to the scalar sum of all tracks associated to the jet:

$$\text{JVF} = \frac{\sum_{A} p_{\text{T}}}{\sum_{\text{Tracks}} p_{\text{T}}}, \qquad \text{where} \quad A = \{t \in \text{Tracks} \mid t \text{ associated to primary vertex}\}.$$

Cuts on the jet vertex fraction suppress jets from pileup collisions because tracks in pileup jets are not associated to the main primary vertex. Pileup jets are discarded from the collection of candidate jets, and also subtracted from $E_{\mathrm{T}}^{\mathrm{miss}}$.

Overlap Removal

As described in section III 3.3, the ATLAS reconstruction algorithms reconstruct objects in different domains separately. Electrons are, for example, reconstructed as jets because both electrons and jets deposit energy in the calorimeter. For the VH analysis, the following overlap removal is applied after the object selection:

1. If both a calorimeter-tagged muon and a combined muon are reconstructed at a distance of $\Delta R < 0.1$, the calorimeter-tagged muon is discarded since momenta of combined muons are measured with much higher accuracy.

2. If a jet and an electron are found at a distance of $\Delta R < 0.4$, the electron is kept, and the jet is discarded. This is because almost all electrons are reconstructed as jets, but the opposite case is rare: the isolation requirements and the electron likelihood identification suppress the majority of quark or gluon jets that are misreconstructed as electron candidates. An example of an electron candidate that is discarded during overlap removal can be seen in the event display on page 49. The event is labelled as a one-electron event, even though two electrons are reconstructed.

3. If a jet and a muon are found at a distance of $\Delta R < 0.4$, the muon is kept if the number of tracks associated to the jet is $N_{\mathrm{trk}} \leq 3$. For $N_{\mathrm{trk}} \geq 4$, the jet is kept. This is based on a study of simulated events indicating that three or fewer tracks are caused by muons that radiate a photon. Such photons can convert into an electron/positron pair, which leads to the detection of up to three tracks and an energy cluster in the EM calorimeter. More than three tracks in vicinity to a muon, however, are usually caused by real jets, in which a muon was produced. The former muons are of interest for the $VH \to Vb\bar{b}$ analysis because these include muons from prompt W or Z bosons. The latter muons are usually caused by semileptonic decays within the jet, and are therefore not indicative of a VH process. Hence, the overlap removal is based on the track multiplicity. If a muon from a W/Z decay incidentally overlaps with a quark or gluon jet, the muon will be discarded incorrectly. In VH events, this is, however, rare because leptons and jets are usually emitted in opposite directions, compare again the event display in figure III.12 on page 49.

4. If an electron and a muon are identified at a distance of $\Delta R < 0.2$, the muon is kept, and the electron is removed. This happens when the muon radiates a photon that does not convert into an e^+e^- pair.

IV 1.2 Event Selection and Categorisation of Events

At this stage, all objects for the reconstruction of VH events are available. Depending on the decay of the vector boson, the analysis splits into three orthogonal categories, also called channels: zero, one and two charged leptons. These correspond to the decays

$$Z \to \nu\bar{\nu}, \qquad W \to \ell\nu \quad \text{and} \quad \qquad Z \to \ell\bar{\ell},$$

where ℓ denotes e, μ, the leptons directly detectable in ATLAS. Table IV.3 lists the Higgs production cross sections and branching ratios in these channels. The channel with the highest signal cross section is

Table IV.3: Cross sections and branching ratios for associated Higgs production at NNLO(QCD)+NLO(EW) accuracy for $m_H = 125.09\,\text{GeV}$. $l = e, \mu$ because only channels with visible leptons are used in the analysis. $\text{BR}(H \to b\bar{b}) = 57.5\,\%$. Higgs cross sections and branching ratios from [41], branching ratios for decays of the vector bosons from [135]. Uncertainties due to scale variations and PDF/α_s have been added linearly to account for correlations.

Process	σ / fb	BR[a]	$\sigma \times$ BR / fb	Used
$q\bar{q} \to WH \to l\nu b\bar{b}$	701 ± 20	0.125 ± 0.004	87.5 ± 3.5	✓
$q\bar{q} \to ZH \to \nu\bar{\nu}b\bar{b}$	420 ± 19	0.1150 ± 0.0032	48.3 ± 2.6	
$\to l\bar{l}b\bar{b}$		0.0421 ± 0.0012	17.7 ± 0.9	✓
$gg \to ZH \to \nu\bar{\nu}b\bar{b}$	39 ± 11	0.1150 ± 0.0032	4.5 ± 1.3	
$\to l\bar{l}b\bar{b}$		0.0421 ± 0.0012	1.7 ± 0.5	✓

[a] In contrast to the ATLAS publication [1], decays to τ leptons are not included in the $W/Z \to l\nu/ll$ branching ratio because these events would only be selected if the τ decays to charged leptons, *i.e.* in 35 % of the τ decays. Moreover, event kinematics in such events are distorted due to an additional neutrino. It is therefore unlikely that these events will contribute significantly to the $VH \to Vb\bar{b}$ analysis.

the 1-lepton channel, and is covered in this thesis with most detail. The channel with the second-highest cross section is the 0-lepton channel, but it is not covered in this work because no lepton triggers can be used. It is included in the $VH \to Vb\bar{b}$ Run-1 publication [1], though. The 2-lepton channel has the lowest signal cross section and branching ratios, but it is quite "clean": since the two leptons originate from Z decays, most of the backgrounds such as QCD multijet and W+jets production are efficiently suppressed by requiring two leptons with the same flavour, opposite charges and an invariant mass compatible with m_Z. The only relevant background processes passing this selection are $t\bar{t}$ and Z+jets production, as is visible in figure IV.1.

Events are categorised into the 1-lepton channel if exactly one tight lepton and no other loose, medium or tight leptons are found. This is because any additional lepton, regardless of the quality of reconstruction, is indicative of a background process. Choosing the requirements for the signal lepton very tight and for the veto leptons very loose leads to the highest background suppression.

Events are categorised into the 2-lepton channel if two loose leptons are found, of which at least one needs to pass the medium requirement. This selection is much looser in comparison to the 1-lepton definition to maximise the signal acceptance. Despite the rather loose lepton definition, background levels are not very high in this channel since a Z-boson selection is applied. All events that do not fall in one of these two categories are discarded.

Apart from categorising events based on the lepton multiplicity, the $VH \to Vb\bar{b}$ analysis also has jet-multiplicity and b-tag-multiplicity categories and categories for the boost of W or Z (p_{T}^V). This results in the categories two or three jets, one or two b-tags (one or both of the two leading[3] jets is b-tagged) and low or high p_{T}^V.

Figure IV.3 shows that background compositions and signal fractions differ strongly in these categories. This facilitates the measurement of background normalisations, and creates a few high-sensitivity categories. The main features of each region are:

[3] "Leading jet" denotes the jet with the highest transverse momentum, p_{T}. The next jet is the subleading jet, *etc.*

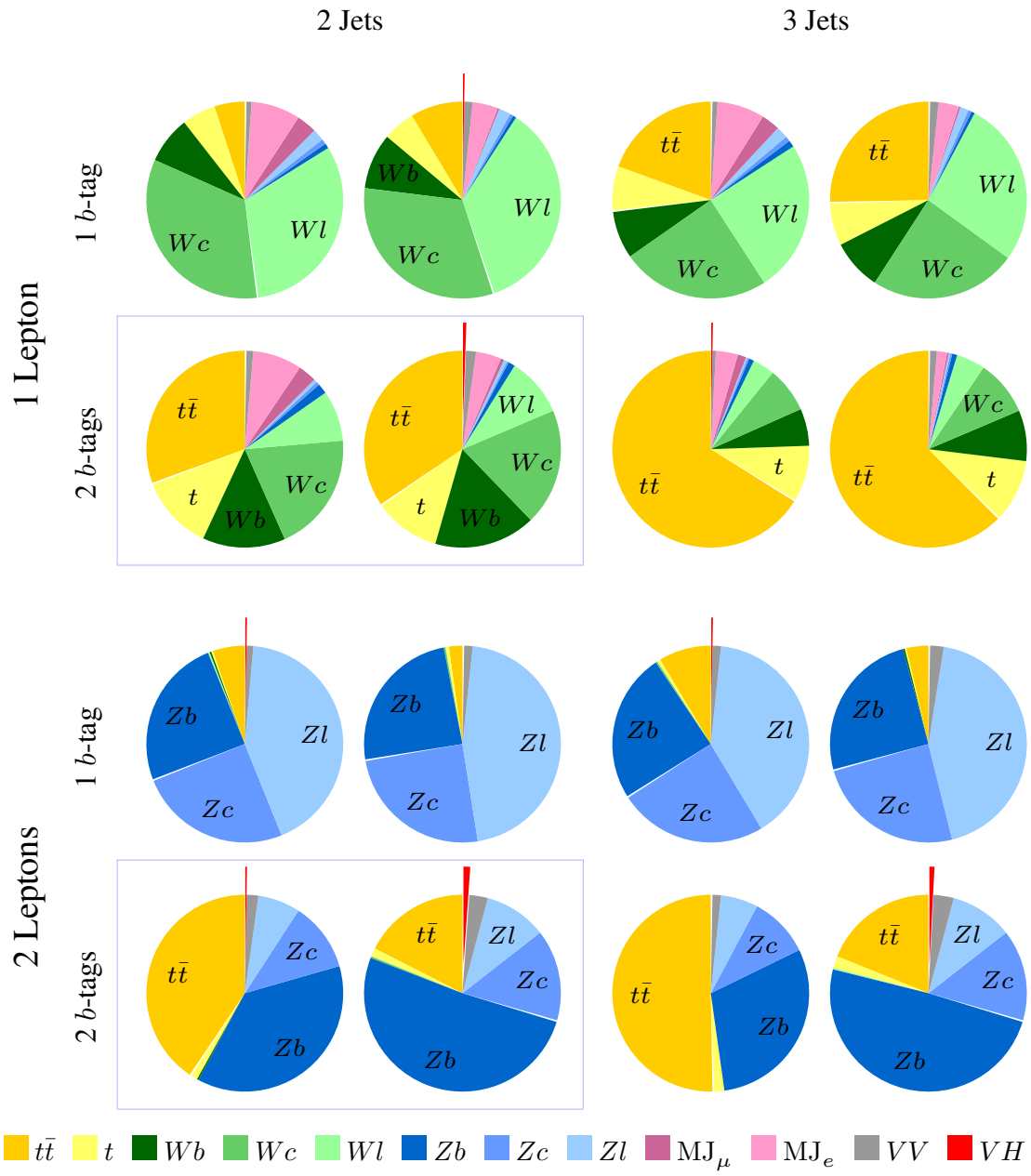


Figure IV.3: Signal regions and background control regions of the $VH \to Vb\bar{b}$ search. Each of the regions displayed is split into p_{T}^{V} < 120 GeV (left) and p_{T}^{V} > 120 GeV (right). Backgrounds are drawn counter-clockwise in the order shown in the legend. The VH signal is slightly displaced for better visibility. The main signal region is 2 jets, 2 b-tags, and is indicated by frames.

2 jets, 2 b-tags Main signal region (indicated by frames). Most of the $H \to b\bar{b}$ decays (■) fall into this region. Main backgrounds are $t\bar{t}$ production (■) in both channels, W+jets (■) in the 1-lepton channel and Z+jets (■) in the 2-lepton channel.

3 jets, 2 b-tags Top control region. A pair of top quarks most likely decays into a pair of b quarks, but if one of the two W bosons produced in this process decays leptonically and the other hadronically, an isolated lepton and jets from the W decay are found in addition. The 3- and 4-jet regions are therefore dominated by $t\bar{t}$ events. The 3-jet region, however, is sufficient to measure the $t\bar{t}$ normalisation.

The 3-jet region is also more interesting than the 4-jet region because $VH \to Vb\bar{b}$ events fall into this region when processes such as the radiation of gluons create an additional jet. This is the case for about 25 % of the VH signal. The top control region therefore also serves as a secondary signal region. Events with more than three jets can be discarded because the probability of two or more additional jets in a $VH \to Vb\bar{b}$ event is low. Comparing the 3-jet regions of the 1- and 2-lepton channels in figure IV.3 shows that $t\bar{t}$ events are less likely to occur in the 2-lepton channel. This is because the requirement of two leptons with same flavour, opposite charge and $m_{\ell\ell} \approx m_Z$ is rarely fulfilled in $t\bar{t}$ decays.

1b-tag region Light-flavour control region. With relaxed b-tagging, backgrounds with lighter jet flavours such as W+l (■), W+c (■), Z+l (■) and Z+c (■) are enhanced. This region can therefore be used to measure the normalisation of these backgrounds. Creating an additional region without b-tagged jets is not necessary because light-flavour backgrounds are abundant in the 1b-tag region.

Low-/High-p_{T}^V region The transverse momentum of the vector boson in VH events, p_{T}^V, is reconstructed from its decay products:

$$p_{\mathrm{T}}^V = \begin{cases} \left| \left(\vec{p}_l + \vec{E}_{\mathrm{T}}^{\mathrm{miss}} \right)_\perp \right| & \text{1 lepton channel} \\ \left| \left(\vec{p}_{l_0} + \vec{p}_{l_1} \right)_\perp \right| & \text{2 lepton channel} \end{cases} \tag{IV.3}$$

This quantity is only well defined for events with vector bosons, but will be denoted p_{T}^V for all events in the VH analysis. If lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$, or the two leptons, respectively, do not originate from the same particle, this observable is the sum of non-related, mostly randomly distributed objects. This results in a background distribution peaking close to zero and falling rapidly.

For $VH \to Vb\bar{b}$ events, p_{T}^V is not only the transverse momentum of the vector boson, but it is also an estimate for the transverse momentum of the Higgs boson. Due to transverse momentum conservation, Higgs boson and vector boson recoil from each other. Since a strong recoil is rare because it requires high q^2 in the collision, the p_{T}^V distribution falls rapidly, but it falls even more rapidly for backgrounds without vector bosons. This is used to suppress background processes, and is especially helpful in the 1-lepton channel, where requiring $p_{\mathrm{T}}^V > 120\,\mathrm{GeV}$ almost completely suppresses the QCD multijet background with muons (MJ_μ ■), see figure IV.3. Hadrons that are misreconstructed as electrons cannot be suppressed entirely, though. MJ_e (■) therefore remains non-negligible, but it is low in comparison to other backgrounds.

The downside of the high-p_{T}^V region is low data statistics. The split at 120 GeV is a compromise between high purity and sufficient data statistics.

A set of kinematic cuts listed in table IV.4 is applied to all regions to further suppress backgrounds, or to ensure that the simulations yield a good description of the data. The cut on the angular distance between the b-jets, $\Delta R(b,b)$, is a modelling cut: except for very high transverse momenta of the vector

Table IV.4: Preselection cuts for the 1- and 2-lepton channels. All quantities but angles in GeV.

	1 Lepton Channel	2 Lepton Channel	
Leptons	1 × tight	1 × ≥ medium + 1 × ≥ loose	
$\Delta R(\text{jet}_1, \text{jet}_2)$	> 0.7	> 0.7	if $p_{\text{T}}^V < 200$
H_{T}	> 180		if $p_{\text{T}}^V < 120$
$E_{\text{T}}^{\text{miss}}$	> 20		if $p_{\text{T}}^V > 120$
m_{ll}		71 – 121	

boson ($p_{\text{T}}^V > 200\,\text{GeV}$), the jets of the Higgs system are always well separated, that is, $\Delta R \gg 0.7$. Therefore, a minimal distance of $\Delta R > 0.7$ can safely be required. This improves the description of the data by the V+jets simulations because events with overlapping jet cones are suppressed. For high p_{T}^V, this cut removes signal events, though. Therefore, the cut is not applied in this region. Instead, a systematic uncertainty for the description of m_{bb} (low $\Delta R \mathrel{\hat{=}}$ low m_{bb}) takes modelling uncertainties into account.

The H_{T} and $E_{\text{T}}^{\text{miss}}$ cuts in the 1-lepton channel are applied to reduce the QCD multijet background, and thereby improve the quality of the background description. The observable H_{T} in figure IV.4 is the scalar sum of the transverse momenta of jets, leptons and $E_{\text{T}}^{\text{miss}}$:

$$H_{\text{T}} \equiv E_{\text{T}}^{\text{miss}} + \sum_{\text{jets, leptons}} p_{\text{T}}. \tag{IV.4}$$

H_{T} can be as low as 95 GeV in multijet events because jets and signal leptons are always required to have at least $p_{\text{T}} = 25\,\text{GeV}$ (45 GeV for the leading jet). For QCD multijet events with two jets and one lepton, H_{T} is at least

$$H_{\text{T}} \geq E_{\text{T}}^{\text{miss}} + 95\,\text{GeV} = 115\,\text{GeV},$$

but the $E_{\text{T}}^{\text{miss}}$ requirement was removed for figure IV.4. For VH events, H_{T} is usually larger than 130 GeV. This is because without any transverse boost, H_{T} can be estimated as follows:

$$\langle H_{\text{T}}(p_{\text{T}}^V = 0)\rangle = 2 \cdot \frac{m_H}{2} \frac{\int_0^{\pi/2} \sin(\theta)\,\mathrm{d}\theta}{\pi/2} + 2 \cdot \frac{m_W}{2} \frac{\int_0^{\pi/2} \sin(\theta)\,\mathrm{d}\theta}{\pi/2} = 130.8\,\text{GeV}$$

For this estimation, it is assumed that $p_{\text{T}}^V \approx 0$ and that daughter particles in the mother particle's rest system are back-to-back with a momentum of $m/2$ each. Since the direction of the daughter particles is isotropic in the rest system, momenta have to be averaged over all angles as shown in the equation. If the assumption of vanishing transverse momenta for W and H is not fulfilled, that is, W or H are boosted in transverse direction, H_{T} increases. Hence, most signal events are found at $H_{\text{T}} \geq 200\,\text{GeV}$ as shown in figure IV.4. Requiring $H_{\text{T}} > 180\,\text{GeV}$ is therefore a good compromise between signal efficiency and QCD multijet suppression. When $E_{\text{T}}^{\text{miss}} > 20\,\text{GeV}$ is applied in addition, the level of multijet background shown in figure IV.4 is reduced a bit more such that the total multijet contribution is well below 20 % of the total backgrounds in all categories of the VH analysis.

In the 2-lepton channel, only Z-boson cuts and the $\Delta R(b, b)$ modelling cut are applied. The event selection might hence seem quite loose. This is because it is designed to be used in conjunction with a multivariate classification: a restrictive preselection would limit the potential of the multivariate methods because all signal events discarded in the preselection phase are lost for further analysis. This will be demonstrated in more detail in section IV 3.3.

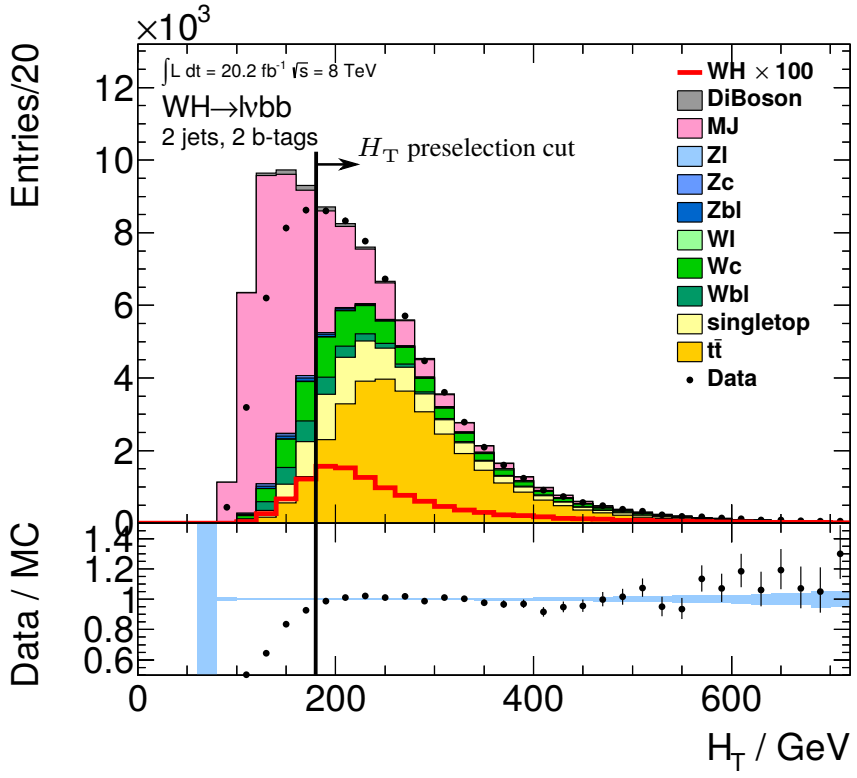
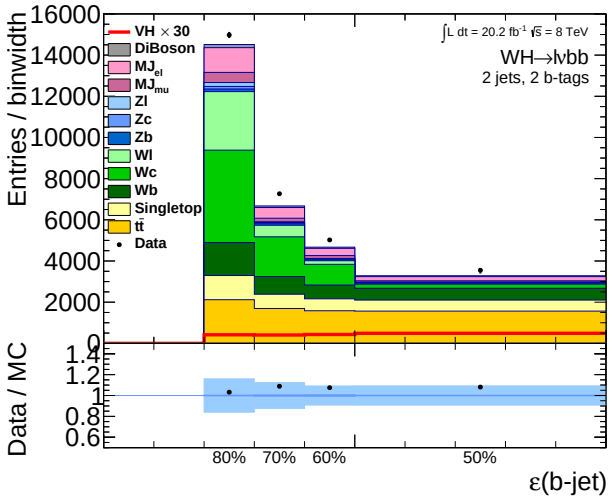


Figure IV.4: Distribution of H_{T}, equation (IV.4), in the 1-lepton channel *before* the full event selection from table IV.4 is applied. Only the object selection and exactly one tight lepton are required. Because of the missing event selection, it is *not* expected that the simulations and the data-driven QCD multijet estimate describe the distribution observed in data.
The region of low H_{T} is completely dominated by QCD multijet events (■), which are suppressed at higher values of H_{T}. Most of the WH signal (■) is found at $H_{\mathrm{T}} \gtrapprox 200\,\mathrm{GeV}$. Since a decent description of the data is achieved starting from $H_{\mathrm{T}} = 180\,\mathrm{GeV}$, H_{T} is always required to be larger than $180\,\mathrm{GeV}$. The blue band in the lower inset shows the statistical uncertainties of the background simulations. A band with systematic uncertainties is not shown because these are derived only for the phase space with the full event selection.

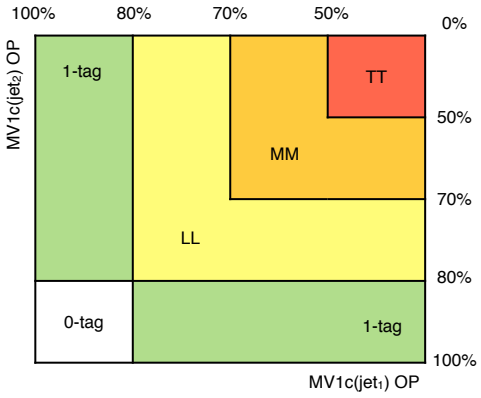
IV 1.3 b-Tagging at Multiple Working Points

A unique feature of the ATLAS $H \to b\bar{b}$ analysis is that b-tagging is used at multiple working points. The b-tagging algorithm `MV1c` (see section III 3.4) was designed with the requirements of the $H \to b\bar{b}$ search in mind, and was therefore calibrated for the simultaneous use of b-tagging efficiencies of 80, 70, 60 and 50 %. The background rejection power of `MV1c` ranges from 29 for light and 6 for charm jets at 80 % b-jet efficiency up to 1 400 (l) and 26 (c) at 50 % b-jet efficiency [136]. Charm rejection is more important for the $VH \to Vb\bar{b}$ analysis because the signature of charm jets is more similar to b-jets than for light jets. As shown in figure IV.3, $W{+}l$ or $Z{+}l$ events are not the dominant backgrounds, but b-tagging algorithms are normally trained to suppress light flavour jets. `MV1c` is trained to suppress light *and* charm jets, and therefore achieves the highest charm rejection on Run-1 data in ATLAS.

The fact that `MV1c` can be used at multiple working points is called "continuous b-tagging" in ATLAS. This is an unfortunate term because the b-jet selection efficiency *cannot* be varied continuously: calibrations only exist for the four working points listed previously.



(a) Background composition versus `MV1c` working point for the 1-lepton channel. Normalised to bin width.



(b) VH b-tagging categories [1].

Figure IV.5: Use of b-tagging for the measurement of background normalisations and for background suppression. **(a)** Flavour composition of the backgrounds for the available working points of the `MV1c` b-tagging algorithm. Backgrounds with light and charm jets are strongly suppressed with tight b-tagging. The simulations slightly underestimate the background cross sections or b-tagging efficiency, but this difference is covered by the uncertainties shown by the blue band. A full measurement of normalisations is carried out during the profile-likelihood fit discussed in sections IV 5 and IV 6. **(b)** Categories of the $VH \to Vb\bar{b}$ analysis: the $2b$-tag regions `LL`, `MM`, `TT` are used as signal regions, the $1b$-tag region is used as a background control region, and the $0b$-tag region is not used.

The availability of multiple working points allows for additional freedom in the use of b-tagging for the VH search: since the flavour composition of background processes depends on the b-tagging working point, see figure IV.5(a), background normalisations can be measured with higher accuracy. To achieve this, the $2b$-tag region of the $VH \to Vb\bar{b}$ analysis is split into three more categories: `LL` (loose+loose) for events where both jets pass b-tagging at $80\,\%$ efficiency, `MM` for medium with $70\,\%$ efficiency and `TT` for tight with $50\,\%$ efficiency. If the two leading jets pass different working points, the event is assigned to the lowest category passed by both jets. All b-tagging-related categories are shown in figure IV.5(b). Events with three jets are categorised based on the two leading jets in p_{T}.

Apart from a higher sensitivity to background normalisations, b-tagging at multiple working points is also useful for better background reduction. Figure IV.5(a) shows that the amount of b-jet-dominated processes ($t\bar{t}$, t, $W{+}b$), as well as $H \to b\bar{b}$, is nearly constant in each bin, but charm- and light-jet-dominated backgrounds are suppressed almost entirely at $50\,\%$ b-tagging efficiency. Previous $H \to b\bar{b}$ analyses, however, were only using b-tagging at the $70\,\%$ working point, which has a total $H \to b\bar{b}$ efficiency of $49\,\%$ because two jets have to be tagged. At the $50\,\%$ working point, the total signal efficiency would only be $25\,\%$. With b-tagging at multiple working points, both high background rejection and high signal statistics can be achieved: `TT` has $25\,\%$ efficiency and low background levels, `MM` has $(70\,\%)^2 - (50\,\%)^2 = 24\,\%$ efficiency and intermediate background levels, and `LL` has $15\,\%$ of the $H \to b\bar{b}$ events but high background levels. A simultaneous fit of all three therefore yields much higher sensitivity than any choice of a single b-tagging working point. The remaining $36\,\%$ of the $H \to b\bar{b}$ events fall into the 0- and $1b$-tag regions dominated by $W{+}l$ and $Z{+}l$, and cannot be identified because of the very high background levels. A dedicated $60\,\%$ category is not created: it would contain only about $11\,\%$ of the signal events, and is therefore merged into the $70\,\%$ category in order to ensure sufficient data statistics. Also due to low data and signal statistics, `MM` and `TT` are merged in the 2-lepton channel, where $\sigma \cdot BR$ is about five times lower than in the 1-lepton channel as shown in table IV.3.

IV 2 Modelling of Signal and Background Processes

Monte Carlo simulations cannot be expected to model the data perfectly: whereas kinematic relations in each event can be calculated exactly because of energy and momentum conservation, inclusive and especially differential cross-section predictions can have high uncertainties. Jet multiplicities or flavour fractions in $V+$jets events, for example, are observables with significant uncertainties, see section III 4.3. If predicted distributions are found to differ from distributions in data, corrections have to be applied by reweighting the Monte Carlo events. The corrections necessary for the $VH \to Vb\bar{b}$ simulations as well as a jet-energy resolution correction will be discussed in the following.

IV 2.1 Signal Simulations

As discussed in section II 5.2, the VH state mainly has two production modes: the $q\bar{q} \to VH$ process can produce both the WH and ZH state. The reaction $gg \to ZH$, however, can only produce a Z boson because the initial state is neutral. This process is only included when simulating VH production at `NNLO` accuracy. With the `LO` generator `Pythia8`, this is not possible.

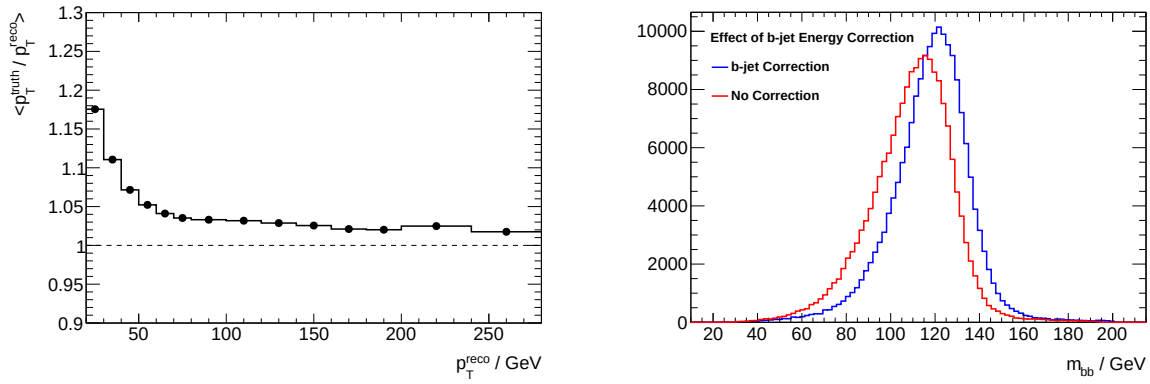
A simple method to correct this lack of events would be to scale the inclusive Monte Carlo cross section, that is, to apply a k-factor that accounts for the `LO-NNLO` difference. Although this is works well for the inclusive $q\bar{q} \to VH$ cross section, it does not correctly predict the kinematics of $gg \to ZH$. This process is therefore taken into account by a `PowHeg` simulation [122, 137] interfaced to `Pythia8`. The reason for differences in kinematic distributions is illustrated by figure II.11 on page 25: since the Z boson is produced by a quark loop, the cross section changes when the top quark threshold is reached: if the energy available in the collision is larger than $2 \times m_{\text{top}}$, two top quarks can significantly contribute to the loop, which increases the cross section. This changes the transverse momentum spectrum in $gg \to ZH$ events, but not in $q\bar{q} \to ZH$. Therefore, applying k-factors to the LO predictions is not sufficient.

Furthermore, NLO electroweak effects in the VH process affect the transverse momentum distribution of the Higgs boson. Radiation of photons or contributions of fermion loops as in the diagrams in figure II.12 on page 26 are examples of such higher-order processes: a comparison of LO predictions with NLO predictions obtained with the `HAWK` [138] program was used to derive correction factors in dependence of the transverse momentum of the Higgs boson, which are applied by reweighting the `Pythia8` predictions.

IV 2.2 b-Jet Energy Correction

Since the invariant mass of two b-jets is one of the most important observables to suppress background processes, the mass resolution of the Higgs peak, figure IV.1, should be as high as possible. A Monte-Carlo-driven jet energy correction is therefore applied to signal and background simulations, as well as data events, because an improved jet energy resolution also leads to a higher dijet mass resolution. The following effects limit the jet energy resolution:

- Calorimeter resolution, section III 2.2. The intrinsic calorimeter resolution is the lower limit for the jet energy resolution. It is minimised by applying the standard ATLAS energy calibration, but apart from minimising calibration untertainties no further reduction is possible.
- Neutrino production in the particle cascade of the jet. Neutrinos escape undetected and lower the measured jet energy.
- Muon production. If a muon is produced in a b decay ($b \to cW \to c\mu\nu$), or in a similar c decay, the muon is detected in the muon system, but not in the calorimeter. This reduces the measured jet energy.



(a) Correction function for the reconstructed transverse momentum of b-jets (p_T^{Reco}).

(b) Invariant mass of the two b-jets after the preselection for simulated WH events.

Figure IV.6: Effect of the b-jet energy correction ("pTReco"). **(a)** All b-jet transverse momenta are corrected with a p_{T}-dependent factor. **(b)** This improves the resolution of the invariant mass of two b-jets by 8 %.

- Out-of-cone leakage. Charged particles, especially with low momentum, are deflected by the magnetic field in ATLAS. If their momentum is low enough ($p_{\mathrm{T}} \lesssim 0.7\,\mathrm{GeV}$[4]), particles leave the cone of the jet clustering algorithm, and are therefore not attributed to the jet.

The b-jet energy correction aims to correct these effects. The energies of b-jets are corrected in two stages:

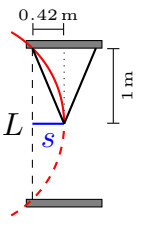
1. If muons are detected inside a jet, the muon four-momenta are added to the four-momenta of the jet. The small amount of energy that such muons typically deposit in the calorimeter is subtracted from the jet energy to prevent double counting. This corrects for muons from semileptonic b or c decays. The neutrino, which is produced in such decays, however, is not corrected for.

   Semileptonic b decays producing electrons need not be corrected for because the energy of such electrons is detected in the EM calorimeter, and therefore attributed to the jet. The energy measurement is therefore only weakly distorted.

2. The transverse energy of the jet is corrected using the function in figure IV.6(a). It mostly corrects for out-of-cone leakage, but also applies a p_{T}-independent scale correction. This accounts for residual miscalibrations of b-jet energies and for the energy that is carried by neutrinos. The correction function was derived by comparing the energy of truth jets to the reconstructed energies.

These corrections improve the resolution of the Higgs peak in m_{bb} by 8 to 14 %, depending on the category of the VH analysis. Figure IV.6(b) shows a comparison of WH events with and without this correction. The narrowed Higgs peak leads to higher sensitivity.

[4] Particles with low momenta are deflected outside of the jet cone. The minimal transverse momentum for this to happen can be estimated using the sagitta s. A track originating from the interaction point and touching the edge of the jet cone can be described by the red circle in the sketch on the right. The corresponding sagitta s is equal to the radius of the jet cone at the beginning of the calorimeter, $s \approx \tan(0.4) \cdot 1\,\mathrm{m} = 0.42\,\mathrm{m}$. The sagitta has to be related to the dashed baseline L, which is two times the distance between interaction region and calorimeter, and this yields the momentum using the following formula:



$$p_{\mathrm{T}} \approx \frac{0.3}{8}\frac{L^2 B}{s} = 0.7\,\mathrm{GeV}, \qquad \text{with } [p_{\mathrm{T}}] = \mathrm{GeV},\ L = 2\,\mathrm{m},\ B = 2\,\mathrm{T}.$$

The b-jet energy correction is applied to all b-tagged jets in data and simulations, but it is derived solely from the signal simulations. Two systematic uncertainties are added to account for the possibility that the signal simulations overestimate the achievable energy resolution: firstly, the jet energy scale is varied by $\pm 0.4\,\%$ (`SysJetBE`). It accounts for the uncertainty of the offset correction shown in figure IV.6(a).

Secondly, an uncertainty on the momentum resolution of b-jets is added (`SysBJetReso`) [139]:

$$\frac{\sigma(p_{\mathrm{T}}^{b})}{p_{\mathrm{T}}^{b}} = 0.4 * \left(\frac{0.99}{\sqrt{p_{\mathrm{T}}}} \oplus 0.12 \right), \quad [p_{\mathrm{T}}] = \mathrm{GeV} \tag{IV.5}$$

It is not a coincidence that this equation resembles the typical resolution function for calorimeters, compare, for example, equation (III.7) on page 39. When deriving the uncertainty of the b-jet energy correction, the calorimeter resolution function was chosen as the underlying model. The parameters in equation (IV.5) were obtained from fits to Monte Carlo simulations of the jet energy. The model neglects the noise term a/E, and is parametrised in p_{T}, but the parameters are similar to the expected resolution of the liquid argon + tile calorimeters discussed in section III 2.2, which is $\frac{0.52}{\sqrt{E}} \oplus 3\,\%$.

The momentum resolution uncertainty is estimated by scaling the transverse momentum of each jet with a random number obtained from a normal distribution, where σ of the normal distribution is given by equation (IV.5). Since the measured calorimeter energies are also normally distributed, this increases the intrinsic resolution by about $40\,\%$. This is because the widths are roughly equal, and widths for the product of Gaussian distributions add in quadrature: $1\,\sigma \oplus 1\,\sigma \approx 1.41\,\sigma$.

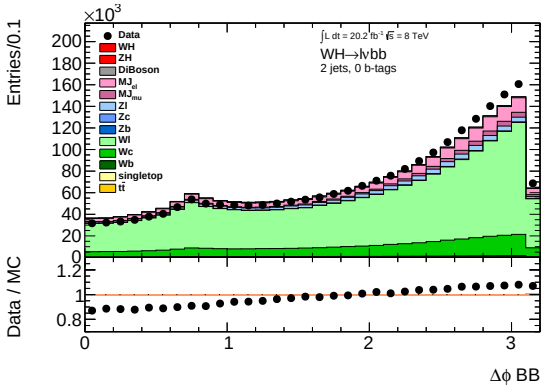
In the 2-lepton channel of the ATLAS $H \to b\bar{b}$ publication [1], a kinematic fit was applied to further reduce the jet energy resolution. It uses a Breit-Wigner constraint for the invariant mass of the Z boson, Gaussian constraints for the transverse momentum of ZH, transfer functions to relate true to reconstructed jet momenta and a prior for the typical jet p_{T} spectrum in ZH events. Although this correction is more powerful than $p_{\mathrm{T}}^{\mathrm{Reco}}$ shown in figure IV.6(a), it was not re-implemented in this work. An about $8\,\%$ better mass resolution can be expected [140], but the focus in this work is on the 1-lepton channel, which has higher sensitivity. In this work, the b-jet energy correction $p_{\mathrm{T}}^{\mathrm{Reco}}$ is therefore applied to all events. Yet more powerful methods such as multivariate jet energy regression [141] or particle flow [142] were not yet commissioned by ATLAS for Run-1 data.

IV 2.3 Angular Distance Between Jets in V+jets Simulations

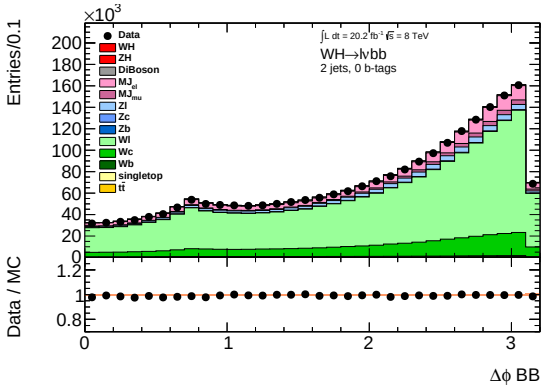
The Sherpa simulations used to model the V+jets background wrongly predict the angular distance between the two leading jets. Figure IV.7(a) shows that the simulations predict a surplus of events with small angular distances and a deficit of events with large distances. The root cause of the mismodelling is not known. It may be a misconfiguration of matching, shower or hadronisation parameters. More recent versions of Sherpa show a reduction of this effect [143], but these are not available in ATLAS for Run-1 data.

The mismodelling is not confined to the $\Delta\phi$ distribution shown in figure IV.7(a). It also affects the ΔR distribution because of $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. Furthermore, it affects p_{T}^{V} because the (transverse) recoil of a vector boson against the jets is low if the angle between the jets is large. This is visible in the p_{T}^{W} distribution in figure IV.7(c), which shows a deficit of W+jets events at low transverse momenta. Finally, the mismodelling of the angular distance is also connected with a misdescription of the dijet invariant masses, m_{jj}. This is because small distances between the jets correspond to low invariant masses. Considering that m_{jj} and p_{T}^{V} are two of the most powerful observables for separating $VH \to Vb\bar{b}$ from background processes, the mismodelling must be corrected for.

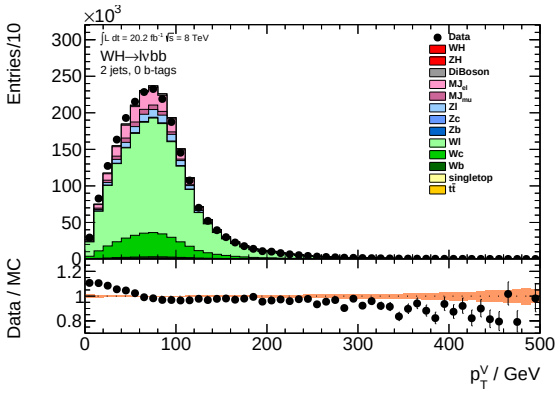
The mismodelling is most pronounced in the $\Delta\phi$ distribution of W+l events with two jets. After



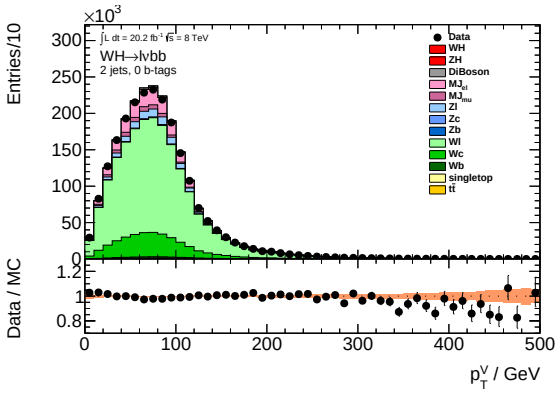
(a) $\Delta\phi$ distribution of the two leading jets in W+jets events



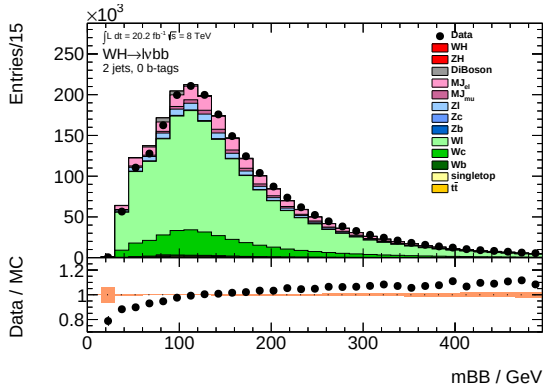
(b) $\Delta\phi$ distribution after $\Delta\phi$ correction



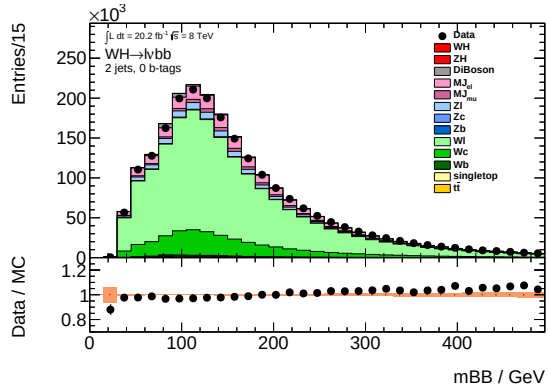
(c) p_{T}^{W} distribution in W+jets events



(d) p_{T}^{W} distribution after $\Delta\phi$ correction



(e) m_{jj} distribution of the two leading jets in W+jets events



(f) m_{jj} distribution after $\Delta\phi$ correction

Figure IV.7: W+jets events after application of the preselection detailed in section IV 1.2. No b-tagging applied. **Left**: **(a)** The Sherpa simulations (■) fail to describe the angle between the leading jets. The mismodelling of the angular distribution also affects other kinematic variables like **(c)** the transverse momentum of the W boson, p_{T}^{W}, and **(e)** the invariant mass of the leading jets m_{jj}. **Right**: Distributions after application of a correction. The correction improves not only **(b)** the $\Delta\phi$ distribution, but also **(d)** and **(f)**.
NB: The peak in the $\Delta\phi$ distributions at 0.7 is explained by the selection cut $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} > 0.7$.

b-tagging, where $W+l$ events are strongly reduced in favour of $W+b$ events, the effect is not visible. Therefore, the $\Delta\phi$ distributions for $W+l$ and $W+c$ events are reweighted to match the data using polynomials of 2nd to 4th degree. These are determined by fits to the ratio of data and Monte Carlo predictions such as the one shown in figure IV.7(a). These fits are carried out separately for the 2- and 3-jet region and for $p_{\mathrm{T}}^{W} \gtrless 120\,\mathrm{GeV}$. 50 and 150 % of the applied correction is used to estimate a systematic uncertainty for the $W+l$ and $W+c$ correction. No correction is applied to $W+b$ events since the mismodelling vanishes when b-tagging is applied. Nevertheless, a systematic uncertainty is estimated by using the $W+l$ and $W+c$ reweighting for $W+b$ events.

Figures IV.7(b), IV.7(d) and IV.7(f) show distributions after application of the correction: the $\Delta\phi$ distribution is fully corrected (by definition), but also p_{T}^{W} and m_{jj} improve. Residual differences are covered by systematic uncertainties.

A similar strategy as for $W+$jets events is pursued to correct $Z+l$ events. Instead of 2nd- to 4th-degree polynomials, linear corrections proved to be sufficient, though. Systematic uncertainties are again estimated by applying 50 and 150 % of the correction for $Z+l$ events.

In contrast to $W+$jets events, the ATLAS $H\to b\bar{b}$ group chose to correct the heavier $Z+$jets flavours by reweighting p_{T}^{Z} instead of $\Delta\phi$. p_{T}^{Z} is reweighted using the function $a\ln(p_{\mathrm{T}}^{Z})+b$, which is again fitted to the ratio of data and Monte Carlo predictions. This correction is used for both $Z+b$ and $Z+c$ events in all regions. Nevertheless, a $\Delta\phi$ uncertainty is estimated by applying the $\Delta\phi$ correction for $Z+l$ events to $Z+c$ and $Z+b$ events. It is unlikely that the misdescription in the $Z+$jets samples has a different cause than for the $W+$jets samples. Since the mismodelling is visible in $\Delta\phi$, p_{T}^{Z} and m_{bb}, the root cause is likely identical. The systematic uncertainty for the p_{T}^{Z} correction is again estimated by applying 50 and 150 % of the correction to the $Z+$jets samples.

IV 2.4 Truth-Flavour Tagging at Multiple Working Points

Strictly speaking, truth-flavour tagging is not a correction, but it is a weight-based technique to reduce statistical uncertainties of Monte Carlo simulations when b-tagging is used. The size of Monte Carlo samples is often limited by computational resources, and when b-tagging is applied, backgrounds with high fractions of light jets are highly suppressed. This leads to low Monte Carlo statistics in samples dominated by light-flavour jets because the probability of b-tagging a light jet is about 3 to 0.07 % [136], depending on the b-tagging working point. Even though strong light-jet suppression is desirable for a high-sensitivity analysis, it also implicates that most Monte Carlo events with light jets are discarded. For a $V+ll$ sample, even at the loose working point with 80 % b-tagging efficiency, only $(3\,\%)^2 = 0.09\,\%$ of the available Monte Carlo statistics would be used.

With truth-flavour tagging, the fraction of events passing the selection can be increased. In this work, it is applied to all $V+ql$ events in the $V+$jets samples as well as to WW simulations. Truth-flavour tagging works by searching the truth records of the Monte Carlo generators for hadrons that overlap with a reconstructed jet. If truth hadrons with b and c quarks or truth τ leptons with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ are found, the jet is labelled "true b", "true c" or "true τ". The probability to b-tag such jets is retrieved from the b-tagging calibration information in dependence of jet transverse momentum and η [96, 144]. Finally, the jet is treated as if it passed the b-tagging cut, and the event weight is modified according to the probability of this jet actually passing the b-tagging cut. In this way, the amount of jets passing the b-tagging cuts is simulated correctly, but a larger fraction of the Monte Carlo statistics is used than for direct b-tagging.

In this analysis, one jet is always required to be b-tagged directly, and the other jet, usually a light jet, is "truth tagged". This means that $V+ql$ events from the 1b-tag region are also used in the 2b-tag regions, but 0b-tag events are not used. This suppresses any bias and mismodelling observed in the 0b-tag region such as the $\Delta\phi$ mismodelling.

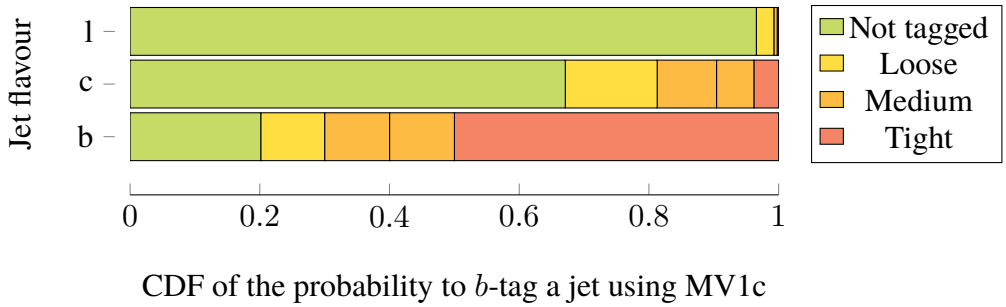


Figure IV.8: Cumulative distribution functions of the probability to b-tag a jet using `MV1c`. Data from [136]. The 60 % b-tagging working point is indicated by the line in the medium region.

In the ATLAS $VH \to Vb\bar{b}$ analysis, the use of truth tagging is complicated by the fact that multiple b-tagging working points are used and that this information is also used to train multivariate classifiers. This will be discussed in section IV 3.3. When using truth tagging, it is therefore not sufficient to only alter event weights because the "truth tagged" jets also need to be assigned plausible b-tagging scores in order to be recognised as b-jets by the multivariate classifiers. Otherwise, the background rejection power of the multivariate classifiers would be overestimated in simulations.

This problem is solved by calculating b-tagging scores based on the cumulative distribution functions (CDF) of `MV1c`. For each jet, these are extracted from the ATLAS b-tagging calibration as a function of the jet transverse momentum and jet η. Figure IV.8 shows CDFs for b, c and light jets integrated over p_{T} and η. The bottom graph for b-jets, for example, shows that there is a probability of 20 % that a b-jet is not tagged; 10 % that it passes loose, but not medium; 20 % that it passes medium, but not tight *etc*. The b-tagging scores assigned during truth tagging need to reproduce this distribution. For truth tagging, the lighter jet flavours are more interesting, though: to "truth b-tag" a charm jet, the event weight is multiplied by $\sim$0.33, the sum of the `L`, `M` and `T` probabilities. The b-tagging score assigned to the jet is determined by drawing a random number between 0.67, the start of the loose region, and 1. Based on the outcome, the `MV1c` score corresponding to `L`, `M`, `T` or to the 60 % working point is assigned to the jet. Although the 60 % working point is not used for categorising events, it is used as an input for the multivariate analysis.

IV 2.5 Top Transverse Momentum Distribution

In a previous version of the ATLAS VH analysis[5], the distribution of p_{T}^{W} in $t\bar{t}$ events predicted by `PowHeg` was found to disagree with the p_{T}^{W} distribution in data: the predicted spectrum is too hard. This does not mean, though, that `PowHeg` wrongly predicts the transverse momentum of W. Using equation (IV.3), one finds:

$$p_{\mathrm{T}}^{W} \equiv \left| \left(\vec{p}_{\mathrm{lep}} + \vec{E}_{\mathrm{T}}^{\mathrm{miss}} \right)_{\perp} \right| .$$

This definition is based on the assumption that one charged lepton and one neutrino are produced by a single W boson, which is valid for WH events in the 1-lepton channel. In $t\bar{t}$ events, however, this observable is not the transverse momentum of a W boson because multiple W bosons and neutrinos are produced. A misdescription of p_{T}^{W} therefore means that $E_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\mathrm{lepton}}$ or both are described wrongly. This can be caused, for example, by a misdescription of the top transverse momenta. Indeed, the transverse momentum distribution of top quarks from `PowHeg` simulations was found to be too hard in other ATLAS analyses [145].

[5] Without multivariate classifiers for the conference EPS [4].

In the 2-lepton channel, a different effect was observed: p_{T}^{Z} was found too soft. Again, the observable p_{T}^{Z} does not describe the momentum of a Z boson in $t\bar{t}$ events:

$$p_{\mathrm{T}}^{Z} \equiv \left| \left(\vec{p}_{\mathrm{lep}_0} + \vec{p}_{\mathrm{lep}_1} \right)_{\perp} \right| .$$

If the transverse momenta of top quarks are predicted too high, angles between the leptons increase because the leptons are emitted more collinear with the top quarks, which are usually well separated. Larger angles between the leptons lead to lower values of p_{T}^{Z}.

Both observations are therefore compatible with the top transverse momenta being predicted too high. To correct for this effect in the $VH \to Vb\bar{b}$ analysis, the top quarks are reconstructed using the Monte Carlo truth data, the average $p_{\mathrm{T}}^{\mathrm{top}}$ is calculated, and this distribution is corrected according to the findings of an ATLAS top analysis with 7 TeV data [145]. 50 and 150 % of the correction is again used to estimate systematic uncertainties of this correction.

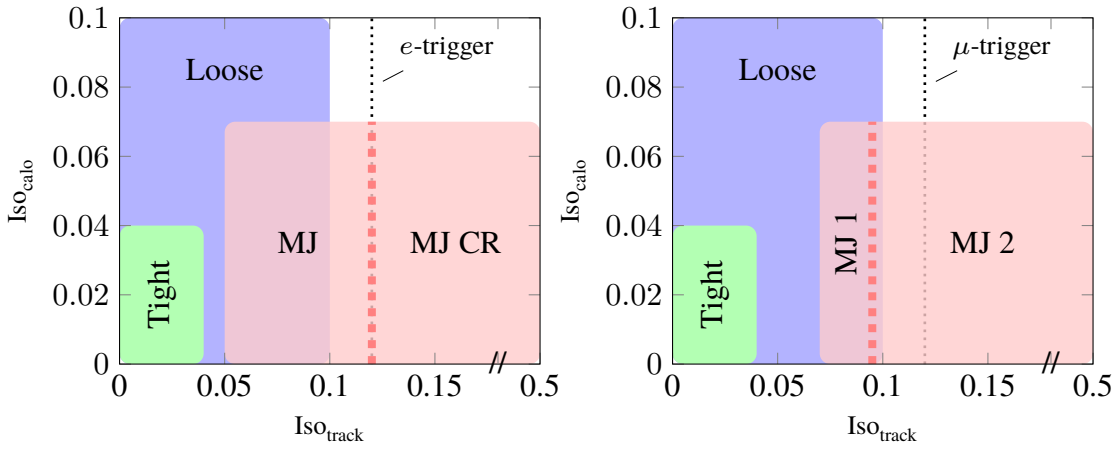
IV 2.6 QCD Multijet Estimation

Among the backgrounds for the $VH \to Vb\bar{b}$ analysis, the QCD multijet background is unique because it is produced with very high cross section ($\approx 200\,\mu\mathrm{b}$), and because it has to be estimated from data due to the very specific topology required in the VH analysis: the production of an isolated lepton is very unlikely in QCD multijet events, but only these events are compatible with the VH topology. The preselection described in section IV 1 aims to reduce the multijet background as far as possible, and only the part that cannot be suppressed without suppressing the Higgs signal is estimated from data.

This estimation relies on inverting the track-isolation requirement, see equation (IV.1) on page 67, because this selects leptons that are produced in the vicinity of other hadronic activity such as leptons from jets or misidentified hadrons. Leptons from electroweak processes, in contrast, are usually isolated from such activity. The underlying assumption is that QCD multijet events with poorly isolated leptons are similar to QCD multijet events with well-isolated leptons, and that the former can therefore be used to estimate the contribution of the latter in the signal region of the VH analysis.

The ATLAS lepton triggers, however, suppress poorly isolated leptons, and distort their transverse momentum spectrum, and therefore complicate the estimation of the multijet background: WH one-electron events are selected using an electron trigger, which is sensitive starting from $p_{\mathrm{T}} \approx 24\,\mathrm{GeV}$. To reduce the event rate due to the high cross section of multijet events, a track isolation of ≤ 0.12 is required on trigger level. This trigger is complemented with a 60 GeV electron trigger without isolation requirements. Both triggers achieve a combined efficiency of 0.97 ± 0.01 [146]. WH one-muon events are recorded using a 24 GeV trigger that also requires a track isolation of less than 0.12. This trigger is complemented with a non-isolated muon trigger starting to be sensitive at 36 GeV. These achieve a combined efficiency of 0.78 ± 0.01.

When inverting the track isolation, the turn-on of the non-isolated lepton triggers severely distorts the transverse momentum distribution of the leptons. These triggers therefore cannot be used to estimate the multijet templates. Therefore, the triggers with isolation, which are fully efficient at $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, have to be used, but the trigger track isolation also distorts the templates. Figure IV.9 shows the track- and calorimeter-isolation variables for leptons in the VH analysis. The tight category (■) comprises well-isolated leptons, which are the signal leptons of the 1-lepton channel. Loose leptons (■), where no multijet suppression is necessary, have a relaxed isolation requirement of ≤ 0.1, which is chosen such that the influence of the track isolation on trigger level is eliminated. "Multijet leptons" (■) are poorly isolated, and affected by the trigger-isolation cuts shown by the dotted line. There is a gap between the tight and the multijet regions in order to ensure that only a negligible amount of events from electroweak



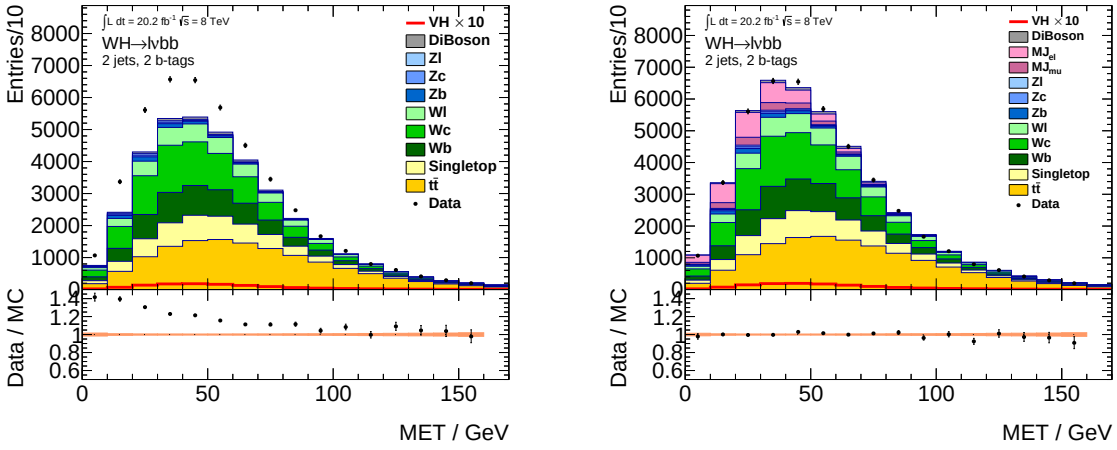
(a) Electrons **(b)** Muons

Figure IV.9: Isolation variables for electrons and muons. The blue region corresponds to loose leptons as defined in section IV 1. The green region is the signal region with very tight isolation cuts for the 1-lepton channel. The QCD multijet template is taken from the pink region (MJ) with an inverted track-isolation cut. There is no calorimeter-isolation requirement for loose leptons, but it is required for multijet template leptons. The lepton track isolation required by ATLAS triggers (dotted line) distorts the multijet templates.

processes such as $W+$jets and $t\bar{t}$ is present in the multijet template. This contribution is estimated with Monte Carlo simulations, and it is subtracted from the data events found in the multijet regions to avoid double counting. Since "multijet electrons", that is, predominantly misidentified hadrons, are produced in different processes than "multijet muons", which are real muons, the multijet template regions start at 0.05 for electrons and 0.07 for muons.

Since events with poorly isolated leptons are used to estimate the multijet contribution of events with well-isolated leptons, extrapolation uncertainties from the 'multijet' into the 'tight' region are derived as follows: the multijet-dominated regions for both electron and muon channel are split in two halves. It is likely that electrons with better isolation ("MJ" in figure IV.9(a)) are more similar to tight electrons than electrons with worse isolation (MJ ControlRegion). Shape differences between these two regions are therefore measured, and applied as shape uncertainties for electron-multijet events, which are moved from the MJ region to the tight region. The reason that the multijet control region MJ CR is populated at all is that track reconstruction on trigger level is more superficial than the offline track reconstruction. A small fraction of QCD multijet events therefore passes the trigger selection despite of the track-isolation cut at 0.12.

In the one-muon channel, the trigger-isolation requirement has an even stronger impact on the template statistics. Since only semileptonic decays can produce poorly isolated muons, the data statistics in the muon multijet template are much lower than in the electron channel. The trigger isolation, however, reduces the event rate so strongly that all poorly isolated muons with track-isolation values from 0.07 to 0.5 have to be used for the multijet template. The template is nevertheless split in two halves at 0.095 to estimate shape uncertainties due to the inversion of the track isolation. The splitting point for muons is closer to the tight region because the trigger and calorimeter isolation for muons is much more efficient than for electrons. To estimate the extrapolation uncertainty for one-muon multijet events, the difference between 'MJ 1' and 'MJ 2' is calculated, but the sum of the two regions is used as the nominal template. Because of the low template statistics in 'MJ 1' and 'MJ 2', the shape differences have to be smoothed when systematic uncertainties are evaluated.



(a) Electroweak backgrounds and data before QCD multijet estimation

(b) Multijet templates and electroweak backgrounds after a fit to data

Figure IV.10: Multijet estimation in the region with 2 jets and 2 b-tags. Inverting the track-isolation requirement for the leptons yields the multijet template. Its normalisation is subsequently measured in the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution, filling the gap between the data and electroweak backgrounds from simulations. The orange band shows the statistical uncertainty of the background estimates.

In conjunction with inverting the track-isolation requirement, the calorimeter isolation has to be relaxed from 0.4 to 0.7. This is necessary since good calorimeter isolation and poor track isolation are mutually exclusive. Relaxing the calorimeter isolation, however, also alters the template shapes. To estimate the influence of this choice, the shape difference to a multijet template with tight calorimeter isolation of 0.4 is measured, and applied as a systematic uncertainty to the nominal template. These shape uncertainties also have to be smoothed because statistics in the template with tight calorimeter isolation are low.

Normalisation of the Multijet Templates

The multijet templates obtained from events with poorly isolated leptons must be normalised to the QCD multijet cross section in the tight region. Since there are no cross-section predictions in dependence of the isolation variables, such normalisations have to be measured in data. The distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$, for example, shows good separation between electroweak processes and multijet events, as can be seen in figure IV.10. The gap between data and simulation in figure IV.10(a) is caused by QCD multijet events, which are not included in electroweak Monte Carlo predictions. $E_{\mathrm{T}}^{\mathrm{miss}}$ shows good separation because QCD multijet events have no on-shell W bosons, which could produce neutrinos with sufficient transverse momentum. This is also the reason for requiring $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$ during the preselection, see table IV.4. The normalisations of the multijet templates are determined in fits to $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions, separately for the 0-, 1- and 2b-tag categories, for the 2- and 3-jet regions, and for electron and muon events. The sum of all electroweak backgrounds is allowed to float in these fits in order to remove dependencies on electroweak cross-section predictions. The measured multijet normalisation factor is kept, but the electroweak normalisation factor is reverted to one because the full determination of the electroweak normalisations is carried out during the profile-likelihood fit. The $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions after the electron and muon multijet fits in the 2-jet regions is shown in figure IV.10(b).

In principle, the multijet normalisations can also be extracted from the distribution of the transverse W mass. Since fits to m_{T}^{W} yield different normalisation factors than for $E_{\mathrm{T}}^{\mathrm{miss}}$, the difference between these

measurements is assigned as a normalisation uncertainty for the multijet backgrounds, separately in each of the categories mentioned in the previous paragraph.

Extrapolating Multijet Events from the 1b-tag to the 2b-tag Region

The statistical uncertainties of the multijet templates are high due to the track isolation enforced by lepton triggers, due to the calorimeter-isolation cut applied, but especially due the use of b-tagging. Without further treatment, the statistical uncertainties of the multijet templates would be higher than the statistical uncertainties of the data, especially in the tight region. The QCD multijet background would therefore dominate the statistical uncertainties of the VH analysis. Since no triggers with higher statistics for poorly isolated leptons are available, and since the calorimeter isolation is already relaxed, the only way to further increase the template statistics is to relax the b-tagging requirement.

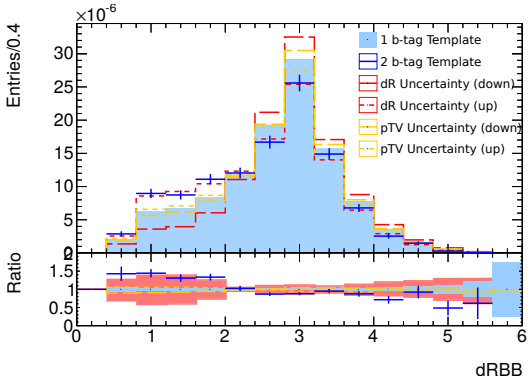
In previous versions of the ATLAS $H \to b\bar{b}$ analysis [2–4], where no multivariate classifiers were used, the b-tagging requirement was lifted entirely: the multijet template was obtained from the 0b-tag region. With this strategy, the statistical uncertainty of the multijet template is negligible, but event kinematics change strongly:

- b-tagging changes the transverse momenta of the jets. b-tagging is more efficient for high-momentum jets. Therefore, the jet p_{T} spectra are harder after b-tagging and too soft in the multijet template.
- A significant fraction of $b\bar{b}$ events is produced by gluon splitting. Lifting the b-tagging requirement increases the contribution of other jet production modes, which leads to a larger angular distance between the jets. Multijet templates without b-tagging therefore have larger invariant masses of jet pairs, which is an important observable for the $H \to b\bar{b}$ search.
- b-tagging favours jets with semileptonic decays. Leptons in or close to jets are predominantly produced by bottom to charm and charm to light decays. When b-tagging is applied, the angular distance between jets and leptons therefore changes.

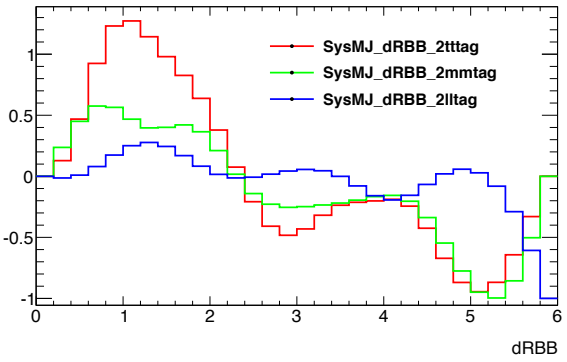
A QCD multijet template with distortions in the distributions of these three observables due to the lifting of the b-tagging requirement cannot be used in conjunction with a multivariate analysis because multivariate classifiers are sensitive to such differences.

Therefore, a better multijet estimation strategy was devised by only partly lifting the b-tagging requirement: instead of using events without b-tags, events with at least one b-tagged jet are used. In this case, the distributions used to train the multivariate classifiers are more similar to the distributions in the 2b-tag region, but differences are still measurable. To mitigate these residual differences, all observables used in the multivariate analysis as well as others such as the lepton momentum were compared for 1- and 2b-tag multijet events. For one-muon events, no significant differences were found. This is because most of the jets in these events are indeed b-jets or c-jets because the dominant source of such muons are semileptonic decays in heavy-flavour jets.

This is different for electrons because the dominant source of these are misidentified hadrons. In electron multijet templates, significant differences between the 1- and 2b-tag regions were found, the largest in the distance between the two leading jets, $\Delta R(b,b)$: the 1b-tag template predicts larger angles than the 2b-tag template. This is shown in figure IV.11(a). Correction functions for the 1b-tag template are shown in figure IV.11(b). These corrections are larger for tight b-tagging because the bias due to the more aggressive b-tagging is stronger: if both jets are b-tagged, gluon splitting is enriched, which leads to small distances between jets. If only one jet is b-tagged, gluon splitting is not enriched, and other processes contribute, which do not necessarily produce jets with small angular distance.



(a) $1b$-tag–$2b$-tag comparison for $\Delta R(b,b)$. 2 jets, MM region.



(b) $\Delta R(b,b)$ correction functions for the $1b$-tag template.

Figure IV.11: Comparison of multijet templates in the 1- and $2b$-tag region. **(a)** The shaded $1b$-tag template predicts higher values of $\Delta R(b,b)$ than the $2b$-tag template (blue markers). Differences are shown as red and orange uncertainty bands. **(b)** Functions used to correct the difference in $\Delta R(b,b)$. The corrections depend on the b-tagging working point because of changes in the background composition.

After the application of $\Delta R(b,b)$ corrections, the $1b$-tag templates (solid histogram in figure IV.11(a)) match the $2b$-tag distributions, the markers in figure IV.11(a). Correcting $\Delta R(b,b)$ also corrects differences observed in the dijet-mass distributions (*cf.* list above), but a small disagreement in p_{T}^{W} and $p_{\mathrm{T}}^{\mathrm{Lepton}}$ remains. Details are shown in figure B.1 on page 208. This difference in p_{T}^{V} arises because b-tagging favours events with higher transverse momenta of leptons (*cf.* list above), which coincide with higher p_{T}^{W}. Therefore, p_{T}^{W} is also reweighted after applying the $\Delta R(b,b)$ correction. This also corrects the disagreement observed in the $p_{\mathrm{T}}^{\mathrm{Lepton}}$ distribution. The corrections were derived by calculating the ratio of the distributions in the $2b$-tag region over the distributions in the $1b$-tag region. Since the accuracy of such correction factors is limited by the statistical uncertainties of the $2b$-tag templates (uncertainties of the dark blue markers), the distributions of the correction factors were smoothed. Two systematic uncertainties are assigned: a $\Delta R(b,b)$ uncertainty shown as a red band in figure IV.11(a) and a p_{T}^{W} uncertainty shown in orange. These were found to cover all remaining differences between the 1- and $2b$-tag templates.

Since the multivariate classifiers have access to b-tagging information, a strategy similar to truth tagging described in section IV 2.4 has to be applied for events in the multijet templates. Since there is no truth information in data, the 2-D b-tagging response of the leading versus the subleading jet in $2b$-tag multijet events was measured. This is a two-dimensional distribution with 16 bins, and can therefore be determined from the $2b$-tag template with sufficient statistical precision. This 2-D distribution is used to predict plausible b-tagging scores for the non-b-tagged jet in $1b$-tag multijet events: depending on the b-tagging score of the b-tagged jet, the CDF for the b-tagging score of the other jet is retrieved from this distribution. Similar to truth tagging, a b-tagging score is selected by sampling random points on this CDF. Systematic uncertainties of these CDFs are low because they are directly taken from data. Nevertheless, separate normalisation uncertainties in each of the LL, MM and TT regions are assigned to the QCD multijet templates.

Due to the double reweighting and assigning of b-tagging scores, the multivariate analysis cannot discriminate the multijet templates with one b-tag from templates with two b-tags. Only because of this, multivariate classifiers can be used in the 1-lepton channel, where the largest gains are expected from

using multivariate classifiers. In comparison to previous versions of the $H \to b\bar{b}$ analysis, this version of the multijet estimation was most extensively studied, assumptions were most thoroughly backed with systematic uncertainties, and the templates are as close to events with two truly b-tagged jets as possible. Because of the conservative uncertainties and since the event selection is designed to keep the fraction of QCD multijet below $20\,\%$ in comparison to other backgrounds, it will not need further special attention in subsequent chapters. Terms like "simulations" or "Monte Carlo" may therefore be understood as "simulations plus a small contribution of a data-driven QCD multijet estimate".

Other Corrections

Several corrections also being applied to the simulations were not discussed here. Among these are the matching of pileup levels in simulations to the levels observed in data, correction of trigger, lepton identification and b-tagging efficiencies, suppression of events with detector or reconstruction errors *etc*. Since these follow standard ATLAS recommendations, and are common to all ATLAS analyses, they are of minor interest here. Details can be found in the ATLAS $H \to b\bar{b}$ publication [1] and references therein.

IV 3 A Multivariate Analysis for the Detection of a $H \to b\bar{b}$ Signal

In this section, the first multivariate $H \to b\bar{b}$ search conducted by the ATLAS collaboration will be introduced [1]. The use of boosted decision trees as multivariate classifiers improved the expected significance of the ATLAS VH analysis from $1.65\,\sigma$[6] (dijet mass analysis [4]) to $2.5\,\sigma$ (multivariate analysis [1]), which amounts to an improvement of the sensitivity by $52\,\%$. The dijet mass analysis was updated with insights obtained from the multivariate analysis, which increased its expected significance to $1.9\,\sigma$ [1]. Using multivariate classifiers was therefore still $32\,\%$ better than using the improved cut-based analysis.

IV 3.1 Introduction to Boosted Decision Trees

Boosted Decision Trees (`BDTs`) are a relatively simple but powerful machine-learning method to classify data, which are represented by a set of observables. In high-energy physics, the most common application for BDTs is to classify events into the categories "signal" and "background".

Decision trees label data by following a hierarchic list of decisions such as a hierarchical set of cuts to select events. Therefore, every cut analysis such as the ones in the previous $H \to b\bar{b}$ publications [3, 4] can be implemented as a decision tree. The resulting trees would typically be several levels deep, and have many "background" leaves because each selection cut discards events until a few signal-enriched leaves, the signal regions of the analysis, remain. Depending on the number of selection cuts, the depth of such trees can grow to 10 to 20 levels.

`TMVA` [147], the "Toolkit for Multivariate Data Analysis with ROOT [120]" provides a framework to grow `binary decision trees` based on training data represented by a set of input variables. A binary decision tree is a tree with exactly two branches per node as in figure IV.12(b). To grow a tree, `TMVA` scans the distributions of all input variables for cuts that increase the purity of signal or background events, thereby separating the former from the latter. The `Gini index` is used to determine whether a cut increases the separation because it measures if one of the classes is in the majority:

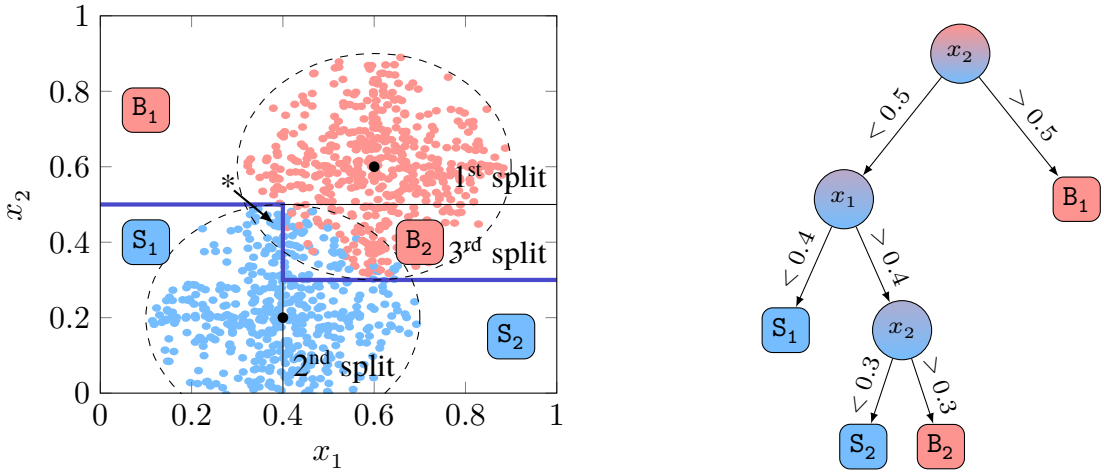
$$G = p \cdot (1 - p), \qquad \text{where} \quad p = \frac{n_S(x_i)}{n_S(x_i) + n_B(x_i)}$$

$n_S(x_i), n_B(x_i)$ are the (weighted) numbers of signal and background events found after applying a cut at the point x_i. Without separation, that is, $n_S = n_B$, one finds $G = 0.25$; at full separation one finds $G = 0$. Multiple splitting points in the distributions of all input observables are tested, and the split yielding the highest separation gain is selected to split the sample of events. The separation gain is the difference of the Gini index in the parent node and the sum of the Gini indices in the child nodes, weighted by the relative fraction of events:

$$\text{Separation Gain} = G_{\text{parent}} - G_{\text{left}} \cdot \frac{N_{\text{left}}}{N_{\text{total}}} - G_{\text{right}} \cdot \frac{N_{\text{right}}}{N_{\text{total}}}$$

When a split is implemented, the procedure is repeated within the new nodes until a stopping condition such as a maximum depth is reached. Each leaf of the tree is labelled "signal" or "background", depending on which class is in the majority in the leaf. Note that if one class completely dominates the training sample, the splitting will predominantly create nodes labelled with the majority class. In high-energy

[6] The significance for an exclusive measurement on $8\,\mathrm{TeV}$ data is not reported in the ATLAS note. $p_0 = 0.05$ (expected) is reported for the combined 7 and $8\,\mathrm{TeV}$ dataset, though. The significance is obtained from p_0 using equation (IV.24) from section IV 5.3.



(a) Simple 2-D classification problem

(b) Corresponding binary decision tree

Figure IV.12: Decision-tree classification with two variables. **(a)** Possible set of splits, which each minimise the Gini index. The blue line is the resulting decision boundary. The region marked with * is "overfitted", which will be discussed in the text. **(b)** shows the decision tree that implements the decision boundary.

physics, this is often the case since a few signal events have to be found in large numbers of background events. The signal cross section is therefore usually increased during the training to match the cross section of the total background. This leads to similar fractions of signal and background nodes, and therefore prevents trees that almost only have background leaves.

In the example in figure IV.12, two classes with equal numbers of training examples are distributed in the x_1, x_2-plane. A tree could be grown as follows: the first split could be implemented at $x_2 = 0.5$, creating a signal-free region B_1 with optimal Gini index $G = 0$. This is indicated by lines in the figure. A subsequent split in x_1 creates a background-free region S_1, again with $G = 0$. The order of these two splits could of course be reversed, since the Gini index is zero both times. Decision trees are therefore inherently unstable in the sense that small fluctuations in the training sample can lead to very different trees being grown. To categorise the remaining space, the next split could be implemented using x_2, creating a background-free region S_2 and a region with non-optimal separation B_2. The corresponding decision tree is shown in figure IV.12(b). The blue line in figure IV.12(a) is the `decision boundary` defined by this tree, which separates the two classes.

The optimal decision boundary would be a line connecting the points where the circles intersect. The tree in figure IV.12(b) crudely approximates this line using a step function in x_1 and x_2. With more splits, the accuracy of this approximation would increase. However, the quality of the inference of the optimal boundary depends on the statistics of the training sample. If the same point clouds as in figure IV.12(a) would be represented with lower statistics, the decision boundary would likely be distorted more.

In principle, `TMVA` could derive a classical cut selection by iteratively scanning all variables and growing a large decision tree. However, to reach human-level performance, an automated procedure has to overcome a lack of information: a classical cut analysis is guided by (external) physics knowledge, whereas a classifier for automated learning has to be inferred from the training data alone. It therefore needs higher training statistics and high-enough flexibility to derive a sophisticated cut selection and to overcome the lack of physics intuition.

When training decision trees, two extreme cases have to be balanced: versatile trees with large depth and robust trees with limited depth. The former achieve optimal separation on the training sample because

the decision boundary can be approximated with arbitrary precision. They might, however, perform only poorly on datasets that are statistically independent from the training sample. This is because a deep tree rebuilds all features of the training sample, but the underlying distributions are not approximated correctly: after many splits, the number of training events available in each node is low. The training algorithm will then base the splits on statistical fluctuations in the training data. In such cases, the resulting decision boundary, although seemingly optimal on the training sample, can be far from the truly optimal decision boundary. This is called `overfitting`. An overfitted region is marked with '*' in figure IV.12(a): although background events are expected in this region, it is classified as signal because a fluctuation in the training sample leads to an absence of background events.

Trees with very limited depth are robust because they are unlikely to overfit the training sample. Since only a few splits are possible, the number of events in each node is large, which mitigates the impact of statistical fluctuations. Such trees are more likely to `underfit`, though: the optimal decision boundary is only crudely approximated because only a few splits are possible.

The tree in figure IV.12 is an example of the latter class: the depth is limited to three consecutive splits, leaving region B_2 poorly separated. At the same time, however, it overfits region '*'. This, however, is a feature of the training sample, and cannot be avoided without higher training statistics.

A shortcoming of a tree-growing algorithm as described here is that the splitting decision does not involve any planning for future splits. It can therefore not easily handle the situation that two variables depend[7] on each other in the sense that a split in one of them has to be introduced to profit from a subsequent split in the other variable. This is the case in figure IV.12(a). The optimal cuts in such cases would be found using a 2-D scan. The computation time would, however, increase as $\binom{k}{2} = \frac{k!}{2!\cdot(k-2)!} = \frac{k\cdot(k-1)}{2} = \mathcal{O}(k^2)$ with the number of variables k. With fast 1-D scans, the splitting can only be optimal if the tree is very deep: in figure IV.12, the diagonal would only be found after many more splits.

The computational challenge to derive close-to-optimal splits with 1-D scans while controlling overfitting can be overcome by `boosting`. Any classifier can be boosted, in principle, but this technique is predominantly used for decision trees. For the ATLAS $H \to b\bar{b}$ analysis, the algorithm `adaptive boosting (AdaBoost)` [148] is used:

`AdaBoost`:

1. Fit a simple classifier $h(\boldsymbol{x})$ such as a decision tree to the training data $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$. $\boldsymbol{x}_i$ is the vector of input variables, and y_i is the desired response for each event i.
2. Determine the fraction of misclassified events ε.
3. Apply the `boost weight` α to the weights of all misclassified events:

$$w_i \to w_i \cdot \alpha, \quad \alpha = \frac{1-\varepsilon}{\varepsilon}, \qquad \forall i \in \{\text{misclass.}\}$$

4. Adjust all weights such that the sum of weights remains constant.
5. Repeat starting at step 1.
6. Finally, combine the classifiers:

$$y(\boldsymbol{x}) = \frac{1}{T}\sum_{j=0}^{T-1} \ln(\alpha_j) \cdot h_j(\boldsymbol{x}), \qquad T = \#\,\text{trees} \tag{IV.6}$$

[7] "Dependent" and "correlated" are different: "Dependent" is used for variables that have to be used in the same tree to achieve optimal separation. The variables in figure IV.12(a) depend on each other because the optimal decision boundary is the diagonal connecting the intersection of the circles. "Correlated" is used if variables have correlation coefficients significantly deviating from zero. Dependent variables may improve the separation if used correctly in the same tree. Yet correlated variables might not increase the separation at all, for example, in the extreme case of 100 % correlation: the variables express identical information.

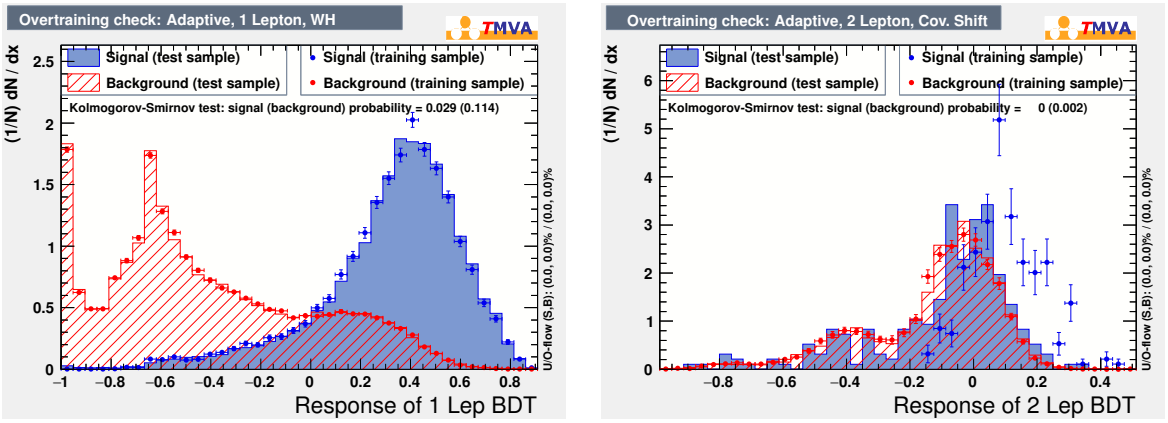
In each iteration, it can be assumed that $\varepsilon < 0.5$ because the label of a leaf is determined by the class that is in the majority, unless $n_S = n_B$. Therefore, $\alpha > 1$, and misclassified events are assigned a higher weight in comparison to correctly classified events.

In example IV.12(a), the only events whose weights will be increased are the signal events in region B_2. The next tree being grown will therefore predominantly separate events in B_2, and in a sense "correct" the mistakes of the previous tree. This procedure is repeated until it converges, combining the classification results of the trees according to equation (IV.6). $y(\boldsymbol{x})$ is the combined boosted classification result, h_j is the hypothesis of the j^{th} tree, and $\boldsymbol{x}$ is the vector of input variables. For signal/background binary decision trees, the hypothesis is 1 for signal and -1 for background leaves. Note that close to convergence one finds: $\ln(\alpha_j) \to 0$. This is because when the training approaches the optimal signal/background separation, newly grown trees cannot improve the classification result. This means that the misclassification probability ε approaches 0.5, which is equivalent to random guessing. Thus, $\alpha_j \searrow 1 \Rightarrow \ln(\alpha_j) \searrow 0$. This is the point where the training can be stopped.

Boosting proved particularly useful for "weak learners", that is, simple classifiers such as small but robust decision trees with only two to four levels. Being robust against overfitting due to their simplicity, their tendency to underfit is mitigated by the boosting procedure. Boosting also solves the problem of modelling dependencies between variables. Instead of performing 2-D scans (k-D scans for multiple dependent variables) and growing large trees, 1-D scans and boosting suffice. This is visible in figure IV.12(a): the optimal decision boundary would be a diagonal connecting the points where the circles intersect, but the first tree only finds a crude approximation. The next tree will, however, focus on the poorly classified events in B_2. The following tree will again focus on a different point along the diagonal, *etc*. In this way, the diagonal is much better represented without using deep trees or performing a 2-D scan of x_1 and x_2. With enough trees and high training statistics, boosted decision trees can therefore approximate any function in any k-dimensional hyperplane as a step function. Nonetheless, a coordinate transformation that aligns the optimal decision boundary with one of the axes in figure IV.12(a) would lead to much faster convergence. Therefore, uncorrelated variables should be preferred as input variables for BDTs.

Boosting also increases the robustness of a classifier: the majority of events is classified at the beginning of the training using a few trees with large weights. Since by construction the boost weights decrease in each iteration, outliers are classified in late iterations and thus contribute only *small* corrections to the general classification result. By construction, this limits the impact of outliers.

A different way to look at boosted decision trees is to picture the response y as a step function in k dimensions; k being the number of input variables: for a given data vector $\boldsymbol{x_i}$, the function assigns $y(\boldsymbol{x_i}) \in [-1, 1]$. The points, at which the step function jumps to a new value, are decision boundaries of trees, such as the dark blue line in figure IV.12(a): crossing this boundary is equivalent to the transition $y(\boldsymbol{x_i})$ to $y(\boldsymbol{x_i} + \boldsymbol{\delta}) = y(\boldsymbol{x_i}) \pm 2\ln(\alpha_j)$, where α_j is the boost weight of the tree whose decision boundary is being crossed. Each split implemented by any of the trees therefore corresponds to an edge in the k-dimensional space. The location of the edges is determined by the splitting algorithm, and indirectly by the boosting procedure. The size of the steps when crossing boundaries is determined by the boost weight $\ln(\alpha_j)$ of the corresponding tree. Early trees with large boost weights will cover large parts of the phase space, introduce large steps, and approximate the general features of the underlying distributions. Late trees with low boost weights will apply small corrections, and cover more narrow regions.



(a) Normal Training, 1-Lepton Analysis

(b) Training with Forced Overfitting

Figure IV.13: Overfitting check for BDTs trained with 1-lepton and 2-lepton events for the ATLAS $H \to b\bar{b}$ analysis. **(a)** Standard training with only small differences between the training sample (markers) and test sample (histogram). The BDTs model the Monte Carlo distributions well, and are therefore able to discriminate signal (blue) from background events (red) even on the test sample that is not analysed during the training. **(b)** Training with strong overfitting, which was provoked by providing low signal statistics. The separation power on the training sample is overestimated, and it does not generalise to the test sample (blue histogram). This training is a special configuration that is not used to search for VH events, but to detect a "covariate shift", that is, mismodelling. This will be discussed in section IV 4.

IV 3.2 Overfitting and Regularisation

Overfitting, in high-energy physics also called "overtraining", occurs when classification algorithms are trained on a dataset that is too small to correctly infer the underlying distributions. Instead, the classification algorithms rebuild random fluctuations in the training dataset such as the region marked with the symbol '*' in figure IV.12(a). Overfitting of a training sample can be detected by splitting the training data into a training and a test sample. The test sample is not used during the training, but it is used to evaluate the performance of the classifier after the training has converged. Since it is statistically independent, it shows whether the classification performance generalises to independent datasets. Figure IV.13 shows such a test for BDTs trained with events from the $VH \to Vb\bar{b}$ phase space. The response of the BDTs to signal (blue) and background events (red) is shown for both the training sample (filled histograms) and the test sample (points with error bars). In figure IV.13(a), the results agree well with only a few bins deviating more than the $1\,\sigma$ statistical uncertainty. This classifier, as well as all others used in the $VH \to Vb\bar{b}$ analysis show almost no signs of overfitting. For figure IV.13(b), overfitting was provoked in the 2-lepton channel by using a very small signal sample. On the training sample it looks like the classifier can separate signal and backgrounds, but on the test sample one sees that this does not generalise to statistically independent datasets.

Since overfitting a training sample can lead to poor performance on independent samples, it should be reduced using methods of `regularisation`. Regularisation enforces simpler models that tend to overfit less, and therefore generalise better to new data. All model parameters that limit the flexibility of the decision trees enforce regularisation. The training parameters for the $VH \to Vb\bar{b}$ decision trees are listed in table IV.5. Three of them regularise the models:

Table IV.5: Hyperparameters for training the BDTs used in the ATLAS $H \to b\bar{b}$ publication [1].

`TMVA` Parameter	Value	Description
`AdaBoostBeta`	0.15	β as in equation (IV.7)
`NTrees`	200	Number of trees trained
`MaxDepth`	4	Maximum depth of decision trees
`nCuts`	100	Granularity of scanning for possible splits
`nEventsMin`	100	Minimum number of events in a tree node

AdaBoostBeta The boosting can be regularised by introducing the "learning rate" β:

$$\alpha \to \alpha^{\beta}, \qquad \text{with } \beta < 1. \tag{IV.7}$$

This reduces the boost weights such that the convergence of the algorithm is slower, but the probability of overfitting the training sample decreases: with decreased boost weights, the training is more robust with respect to outliers because these are assigned lower boost weights. The learning rate of $\beta = 0.15$ in table IV.5 is considerably lower than for classic `AdaBoost` with $\beta = 1$.

If viewing BDTs as a method that implements a k-dimensional step function, a reduction of the learning rate corresponds to a reduced step size when crossing the decision boundary of a tree. This allows for a more fine-grained sampling of the phase space.

MaxDepth As discussed before, deeper trees are more likely to overfit the data because with every split the number of training events is reduced. Reducing the depth of the trees hence enforces simpler models. The depth of four is a compromise between complexity and robustness.

nEventsMin Split decisions can be regularised by requiring a minimum number of training events in each BDT node. This suppresses splits that are implemented based on features with high statistical uncertainty. For the VH BDTs, 100 training events are required in each BDT node; corresponding to about 1 ‰ of the high-p_{T}^{V} training sample. A split is not implemented if the number of events falls below this value.

Two more parameters not related to regularisation are listed in the table: the parameter `nCuts` determines how many points are scanned for split decisions: each variable is segmented into `nCuts` equidistant intervals, and the Gini index is calculated at each boundary. Choosing `nCuts` low speeds up the training, but it leads to a less accurate determination of the optimal cut position. Since this parameter was chosen to be 100 for the ATLAS $VH \to Vb\bar{b}$ analysis, the value range of certain variables such as transverse momenta or invariant masses had to be limited before using these in the training. A disadvantage of this method is that energy and momentum conservation laws cannot be inferred correctly for events where at least one variable exceeds one of the thresholds. It therefore has to be ensured that these events are irrelevant for the analysis at hand, for example, because the truncated regions are background-dominated. An alternative strategy will be discussed in section V 1.1, which is to use a transformation that does not violate energy and momentum conservation.

The parameter `nTrees` in table IV.5, determines the number of trees being trained. Since by construction the boost weights decrease with each new tree, the training converges to an optimal point, after which training new trees does not increase the separation. The optimal number of trees was found to be about 200 by testing if adding a new tree increases the separation on the test sample.

Although overfitting can be reduced by appropriate regularisation, all classifiers are by construction models being fit to the training sample. Statistical evaluations of such classifiers using the training sample will therefore always be biased towards better performance. Since the statistical evaluation of Monte Carlo simulations is a central step in each analysis in high-energy physics, this bias must be avoided. Any statistical analysis of the simulations is therefore always carried out on the test sample. If a single classifier is used, this, however, reduces the Monte Carlo statistics: the training sample is not available for statistical tests after the training.

The full Monte Carlo statistics can be retained, though, if *two* independent classifiers with identical model parameters are used. The training sample of one classifier is used as the test sample of the other and *vice versa*. This also eliminates biases due to overfitting. In order to use such classifiers on data, a reproducible splitting strategy needs to be devised. For the $VH \to Vb\bar{b}$ analysis, all data and Monte Carlo samples are split into even and odd event numbers, which creates two completely orthogonal samples of almost the same size.

Table IV.6: Comparison of event selection cuts for the ATLAS cut analysis [3, 4], the ATLAS $H \to b\bar{b}$ MVA baseline proposal and the H_{T}-based preselection developed for the ATLAS Run-1 publication [1]. Ranges indicated with "$\to$" vary in steps with p_{T}^{V}.

		ATLAS cut analysis	Proposed for MVA	H_{T}-based
p_{T}^{V}		0-90|90-120|120-160|160-200|200-∞	120 to ∞	–
$\Delta R(b,b)$		0.7[a] to 1.4 $\to$ 3.4	0.7[a] to ∞	0.7[a] to ∞
b-tagging		70 %	70 %	80 %
$E_{\mathrm{T}}^{\mathrm{miss}}$	1L	25 $\to$ 50 to ∞	25 to ∞	20 to ∞
m_{T}^{W}	1L	40[b] to 120	40[c] to ∞	–
H_{T}	1L	–	–	180[c] to ∞
m_{ll}	2L	83 to 99	71 to 121	71 to 121

[a] if $p_{\mathrm{T}}^{V} < 200\,\mathrm{GeV}$ [b] if $p_{\mathrm{T}}^{V} < 160\,\mathrm{GeV}$ [c] if $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$

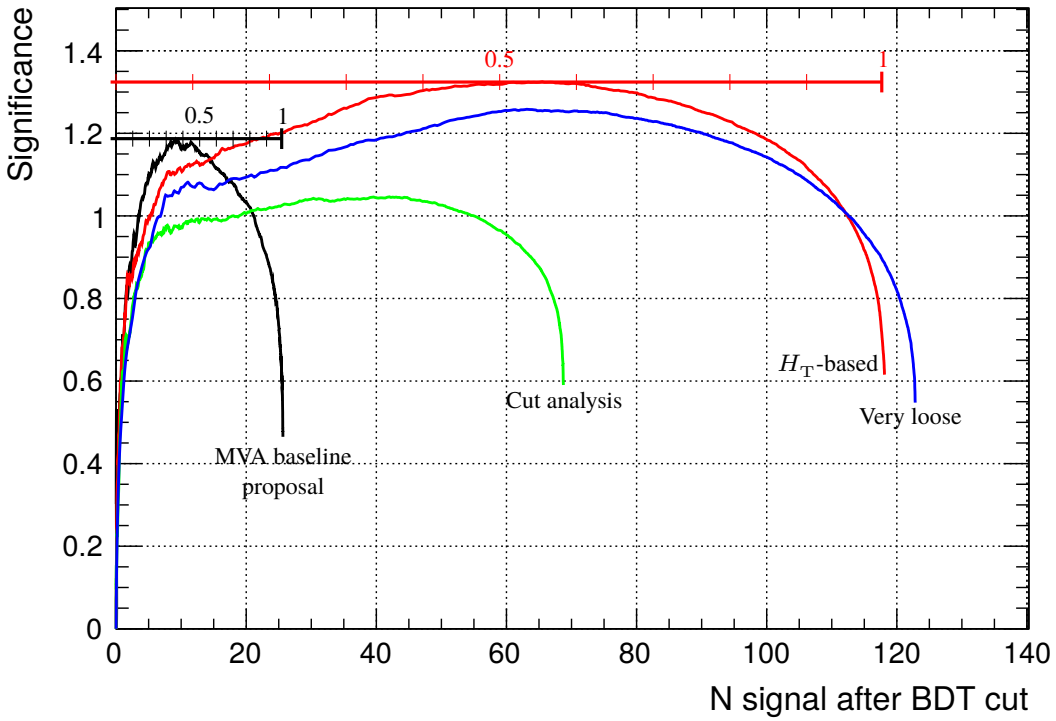


Figure IV.14: Comparison of preselection strategies. (■) MVA baseline proposal, (■) cut analysis preselection and (■) H_{T}-based preselection for the 1-lepton channel, 2 b-jets (70 %). The curves show the significance obtained by selecting only events that exceed a certain BDT classifier score. Scanning through the minimal required score yields the curves. The signal efficiency and highest significance for the two most important preselections are indicated at the top of the curves. The rightmost curve (■) shows a very loose preselection with only b-tagging and object selection applied, which was used to find the optimal preselection.

IV 3.3 Applying BDTs to the Search for $H \to b\bar{b}$ Decays

Some aspects of applying BDTs to the VH analysis have been discussed before because they are independent of the analysis at hand: overfitting, regularisation, model parameters and ensuring that the Monte Carlo simulations model the data well enough to train a classifier. Now, two aspects that are specific to the $VH \to Vb\bar{b}$ analysis will be discussed[8]: the preselection of events entering the training and the choice of discriminative variables.

A. Preselection of Events

To introduce multivariate classifiers to the $VH \to Vb\bar{b}$ search, a new preselection had to be derived. This is because the approach of separating signal from background is different when using an MVA (MultiVariate Analysis) instead of using a classical cut analysis. Recalling that the inference of the underlying distributions cannot be guided by physics intuition, but entirely has to come from training examples, large simulated datasets are needed. However, if sufficient training data are available, a multivariate selection can outperform a standard selection by applying much more complicated cuts. Yet in order to achieve this, the preselection of events entering the training should be kept to a minimum. Pre-emptively applying selection cuts may impede the determination of sophisticated multivariate selections. This can be shown by comparing the ATLAS cut analysis preselection [4], a preselection initially proposed for the MVA and the final H_{T}-based preselection listed in table IV.6. Their impact on the training of classifiers is shown in figure IV.14 for the most sensitive region of the VH analysis with one lepton and two jets. The figure shows the performance of BDT classifiers, which only differ in the preselection applied before the training. The discovery significance[9] is plotted versus the number of $VH \to Vb\bar{b}$ events that can be expected to pass the selection in the dataset of 2012, with b-tagging at 70 % efficiency. Each point on the curves corresponds to a working point that is selected by cutting on the classifier response $y(\boldsymbol{x})$: very high background suppression and low signal efficiency when requiring high y on the left and *vice versa* on the right. The optimal working point is usually at about 50 % signal efficiency. With higher signal efficiencies, the background suppression is too low, therefore the significance decreases. For low signal efficiency, the statistical uncertainty of the number of observed signal events is too high, which also decreases the significance. The jitter of the curves in regions where less than 20 signal events are expected is also caused by statistical uncertainties of the Monte Carlo samples. It should be noted that this is not the final significance of the VH analysis. Nonetheless, measuring this simple MVA-supported significance can guide the choice of the best preselection strategy.

The preselection for the ATLAS cut analysis [3, 4] (left column in table IV.6) draws its selection power from cuts in the $\Delta R(b,b)$, m_{T}^{W}, p_{T}^{V} and $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions. Used with BDTs, this preselection, however, leads to low discovery significances because it pre-emptively applies cuts, which BDTs could implement better: BDTs would create more fine-grained step functions, and can much better model the dependence between p_{T}^{V} and $\Delta R(b,b)$, which is only roughly captured in five steps by the cut-analysis selection. Therefore, loosening of the preselection cuts, but keeping $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$, was proposed (MVA baseline proposal, middle column in table IV.6, left graph in figure IV.14). The figure shows that this preselection allows for much higher significance, but it is far from being optimal. The best preselection is the H_{T}-based preselection introduced in section IV 1 because the cuts of the MVA baseline proposal – although relaxed in comparison with the cut analysis – still remove a large fraction of the $H \to b\bar{b}$ signal. At the optimal

[8] It is clear that the optimal model parameters depend both on the preselection and on the input variables. Yet the strategy for optimising the model parameters is independent of final state, preselection and input variables. The model parameters listed in table IV.5 are optimised for the preselection and variables discussed in the following.

[9] For details see section IV 5.3, page 108.

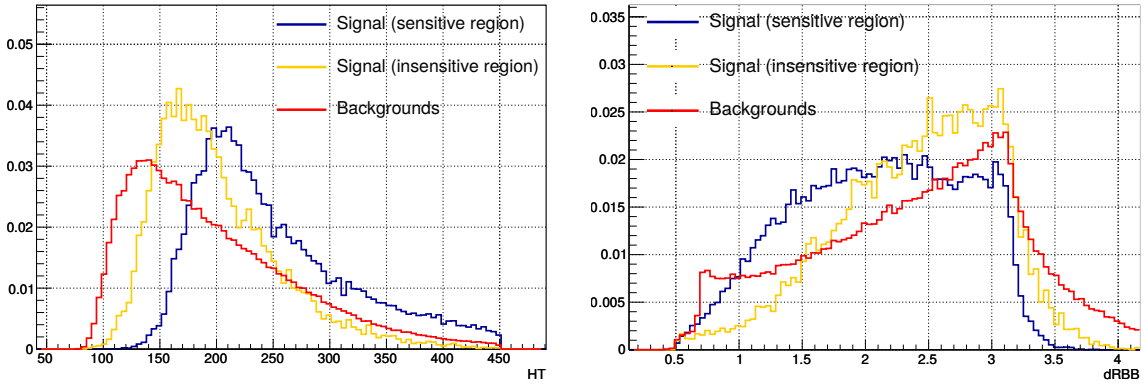


Figure IV.15: Usage of BDT classifiers to identify high-sensitivity regions to guide the derivation of preselection cuts. The figures show distributions of (left) H_{T} and (right) $\Delta R(b,b)$[a] for (■) signal events in the high-sensitivity phase space, (■) signal events in the low-sensitivity phase space and (■) the total background. All distributions are normalised to unit area. Preselection cuts not affecting high-sensitivity events can be applied without negative impact on the analysis. More distributions can be found in appendix B 3, page 209.

[a] The edge in the background distribution at $\Delta R(b,b) = 0.7$ is due to an oversight: the $\Delta R(b,b)$ modelling cut is applied for background events with $p_{\mathrm{T}}^{V} < 200\,\mathrm{GeV}$, but not to signal events. It does not change the conclusions, though: there are virtually no signal events in this region, and for $p_{\mathrm{T}}^{V} > 200\,\mathrm{GeV}$ the cut is not applied, anyway.

working points, one can expect about 10 signal events in the $8\,\mathrm{TeV}$ dataset with the baseline proposal, but ~ 65 signal events with the H_{T}-based preselection. The latter therefore outperforms all alternatives.

If only statistical uncertainties would have to be considered, the best strategy would be to not apply a preselection, and delegate the background suppression entirely to BDTs. However, if systematic uncertainties are taken into account, it is beneficial to restrict the phase space to a well-understood region, in which the Monte Carlo generators make reliable predictions. Ideally, such narrowing of the phase space would not discard any signal events. However, a sizeable number of signal events falls into regions with very high backgrounds.

Many of these can be discarded without reducing the sensitivity of the analysis, but a method to find such regions is necessary. One method is to train BDTs on a dataset with a very loose preselection, and using a large set of variables that could potentially be used for applying a preselection [7]. In the VH case, these candidates are:

- Angular variables: $\Delta R(b,b)$, $\Delta\eta(b,b)$, $\Delta\Phi(V,bb)$, $\Delta\Phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$, $\min(\Delta\Phi(\ell,b))$
- Transverse momenta: p_{T}^{V}, $p_{\mathrm{T}}^{b_1}$, $p_{\mathrm{T}}^{b_2}$, p_{T}^{ℓ}
- Invariant masses: m_{bb}, m_{T}^{W}
- Others: $E_{\mathrm{T}}^{\mathrm{miss}}$, H_{T}

The very loose preselection is the basic object selection from section IV 1 with loosened transverse momentum cuts, with signature selection cuts (2 jets, 1/2 leptons) and b-tagging. The classifier trained in this way is by no means optimal for extracting the $H \to b\bar{b}$ signal strength – its response for simulations and data might even be different – but it shows which regions of the phase space spanned by the candidate variables have high sensitivity. The performance of this classifier is shown by the rightmost curve in figure IV.14. The significance is lower than for the H_{T}-based classifier because input variables for the latter are optimised for discrimination power instead of being candidates for a preselection, and because the phase space covered by the former is considerably larger. For better performance, this classifier would require more flexibility, that is, less regularisation, but maximal performance is not necessary to derive a preselection.

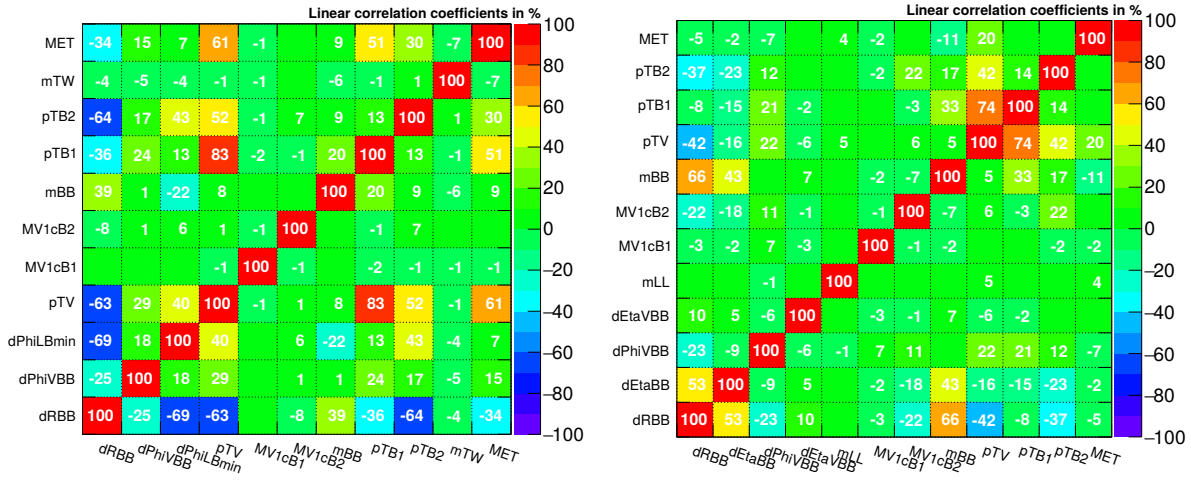
This preselection-study classifier guided the derivation of the H_{T}-based preselection. It was used to split the phase space into a signal-enriched region, that is, events passing the $50\,\%$ working point, and a background-enriched region. Figure IV.15 shows this for two variables: H_{T}, which suppresses the QCD multijet background, and $\Delta R(b,b)$, which can be used to improve the Monte Carlo description. The figures also show the inclusive distribution for background events without using any classifier. The histogram of H_{T} on the left suggests that the H_{T} cut should be placed at $H_{\mathrm{T}} > 150\,\mathrm{GeV}$, which would avoid any loss of high-sensitivity signal events. This cut was, however, tightened to $180\,\mathrm{GeV}$ as described in section IV 1.2 to restrict the analysis to a phase space with low QCD multijet background, compare again figure IV.4 on page 73. This means that a small fraction of high-sensitivity signal events is lost, but background uncertainties are reduced, and the description of the data improves.

The histogram of $\Delta R(b,b)$ on the right of figure IV.15 shows that $\Delta R(b,b) > 0.7$ can be applied with only a small loss of high-sensitivity signal events. This cut improves the Monte Carlo description because jet cones will overlap at this distance, which distorts the jet clustering. Details of parton shower and matching implementation are not negligible in this region, either. Results for other variables can be found in appendix B 3 starting on page 209. The only other acceptable cuts are $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$ and cuts on the transverse momenta of the b-jets, which are therefore part of the H_{T}-based preselection. Other cuts from the cut analysis or MVA baseline proposal were removed. Because of this, a larger number of signal events is retained, which increases the sensitivity of the multivariate classifiers.

Using the H_{T}-based preselection, one can expect 118 $WH \to \ell\nu b\bar{b}$ signal events in the data of 2012 instead of 25 for the MVA baseline proposal as shown by figure IV.14. At the same time, the QCD multijet background remains at a controllable level. The H_{T} cut is even superior to other QCD multijet cuts such as $E_{\mathrm{T}}^{\mathrm{miss}}$ or m_{T}^{W} used in the other preselections in table IV.6 because QCD multijet events predominantly contain soft jets and leptons. While $E_{\mathrm{T}}^{\mathrm{miss}}$ is only sensitive to neutrinos, and m_{T}^{W} is sensitive to $E_{\mathrm{T}}^{\mathrm{miss}}$ + lepton, H_{T} is sensitive to $E_{\mathrm{T}}^{\mathrm{miss}}$ + lepton + jets.

The H_{T}-based preselection differs from the alternatives in table IV.6 in one more respect: less restrictive b-tagging. With b-tagging at multiple working points (section IV 1.3), the phase space covered by the MVA could be enlarged to b-jet efficiencies of $80\,\%$, retaining $(80\,\%)^2 = 64\,\%$ of all di-b-jet events. This has to be compared to $49\,\%$ at $70\,\%$ b-tagging efficiency. Since the analysis is split into multiple b-tagging categories, the enlarged phase space does not lead to an increase of background levels: the sum of the `MM` and `TT` categories is identical to an analysis with $70\,\%$ b-tagging preselection, but a new category is introduced: `LL`. In figure IV.14, the H_{T}-based preselection is shown with $70\,\%$ b-tagging, though, to facilitate the comparison with the other alternatives. With $80\,\%$ b-tagging, the preselection retains 161 signal events.

The derivation of a preselection for the 2-lepton channel was carried out with the same strategy as has been described for the 1-lepton channel, but the results are simpler: only a loose cut on $m_Z = m_{\ell\ell}$, the signature selection (2 same-flavour, oppositely charged leptons, 2 b-jets), as well as the $\Delta R(b,b)$ modelling cut are applied. Other preselection cuts are not necessary because requiring two isolated leptons and two b-jets yields an event sample with only Z+jets and $t\bar{t}$ backgrounds.



(a) 1-lepton channel **(b)** 2-lepton channel

Figure IV.16: Correlation coefficients of the discriminating variables in simulated $VH \to Vb\bar{b}$ events in the 2-jet regions. **(a)** 13 pairs of variables have correlation coefficients $>30\,\%$, 5 have high correlations $>60\,\%$. **(b)** Correlations are lower in the 2-lepton channel: 8 pairs $>30\,\%$, and 2 pairs $>60\,\%$.

B. Selection of Variables

Several variables provide information to separate $VH \to Vb\bar{b}$ events from background events. The most powerful variables like m_{bb}, p_{T}^{V} and $\Delta R(b,b)$ were already identified by the ATLAS cut analyses. To complement these, a bottom-up technique was used. From a pool of possible discriminating variables, one at a time was added to the set of input variables, and its impact on the discovery significance of the resulting classifiers was measured. The most promising variable was added to the baseline set, and the procedure was repeated with further variables from the pool. The resulting training variables are shown in table IV.7. Nine kinematic variables such as momenta, masses and angles between final-state objects were selected for the 1-lepton channel. For the 2-lepton channel, 10 kinematic variables were selected.

Besides kinematic variables, the training algorithm also has access to the response of the b-tagging algorithm for each of the two leading jets. The quasi-continuous response of `MV1c` is transformed into a step function that can take four discrete values. These indicate if a jet passes the 80, 70, 60 or $50\,\%$ working points. Details on b-tagging can be found in section III 3.4. By adding this information to the set of input variables, the BDTs can essentially split the phase space into 16 b-tagging categories: four working points for each of the two jets. In this way, decision boundaries can be adapted to the event kinematics in dependence of b-tagging-related effects such as changes of background flavour compositions or the momentum dependence of b-tagging. The classifiers were found to indeed adapt to the momentum dependence: for high classifier outputs, b-tags on the leading jet ($p_{\mathrm{T}} > 45\,\mathrm{GeV}$) are always tight, but b-tags on the subleading jet with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ are often $60\,\%$ and medium, occasionally also loose.

Selecting discriminating variables by adding more and more candidates to the pool of input variables has a potential downside: this method detects if a variable improves the discrimination power, but it does not detect whether this variable encodes information optimally. This can lead to training variables with high correlations, which is the case for the $VH \to Vb\bar{b}$ input variables as shown in figure IV.16. High correlations are unfavourable, because implementing decision boundaries in such settings requires a high number of degrees of freedom, that is, cuts.

Table IV.7: Discriminating variables of the ATLAS $VH \to Vb\bar{b}$ analysis [1].

	1 Lepton	2 Leptons
p_{T}^{V}	•	•
$E_{\mathrm{T}}^{\mathrm{miss}}$	•	•
p_{T}^{j1}	•	•
p_{T}^{j2}	•	•
m_{bb}	•	•
$\Delta R(b,b)$	•	•
$\Delta\phi(V,bb)$	•	•
m_{T}^{W}	•	
$\min \Delta\phi(l,b)$	•	

	1 Lepton	2 Leptons
$\lvert\Delta\eta(b,b)\rvert$		•
$\lvert\Delta\eta(V,bb)\rvert$		•
m_{ll}		•
MV1c(b_1)	•	•
MV1c(b_2)	•	•
3-jet events		
p_{T}^{j3}	•	•
m_{bbj}	•	•

This is illustrated in figure IV.17, which shows two variables with high correlation. An exemplary decision boundary based on a grid spaced in steps of 0.1 in x_1 is shown. To implement this decision boundary, 24 cuts are necessary. If the ellipsoids were rotated by 45°, the correlation would vanish, and only two cuts would suffice to enclose the blue ellipsoid, yielding almost better separation than for the decision boundary shown in the figure. With high correlations as shown in the figure, the decision boundary can only be implemented with deep or numerous trees after multiple iterations of boosting. If training statistics are a limiting factor, strong regularisation has to be applied, which suppresses deep trees. The inference of the optimal decision boundary might therefore be prevented if multiple correlated variables are used.

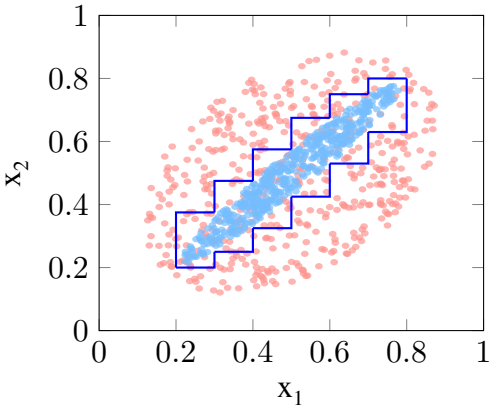


Figure IV.17: Implementation of a decision boundary for strongly correlated variables.

The downside of high correlations can alternatively be understood by estimating how much information is contained in the input variables. The example in figure IV.17 can be decomposed into a variable with low separation power (along the diagonal) and one with high separation power (counterdiagonal). The information contained in the pair of variables can almost fully be represented by the single variable in direction of the counterdiagonal. The pair therefore expresses only one "unit of information" with two variables. This is relevant when training statistics do not suffice to populate a k-dimensional training phase space. If the number of variables can be reduced by representing the same information with $u = k - n_{\mathrm{noisy}}$ variables, the phase space shrinks. The difference u is the number of usable units of information. Statistical uncertainties therefore have less influence on the decision boundaries. A set of variables with vanishing correlations that has $u = k$ will be derived in chapter V.

IV 4 Excursion: Covariate Shift and Multivariate Monte Carlo Modelling Checks

Before proceeding to the statistical analysis of the data, a short detour is taken here in order to look into how machine learning can help to study the quality of the Monte Carlo simulations. These studies were not conducted for the ATLAS Run-1 publication [1], but it makes sense to include these here because they are a natural extension of the modelling studies described in section IV 2. The multivariate modelling checks described in the following were used to check the Lorentz-invariant MVA (chapter V), but this MVA uses the same simulations as were used for the Run-1 publication.

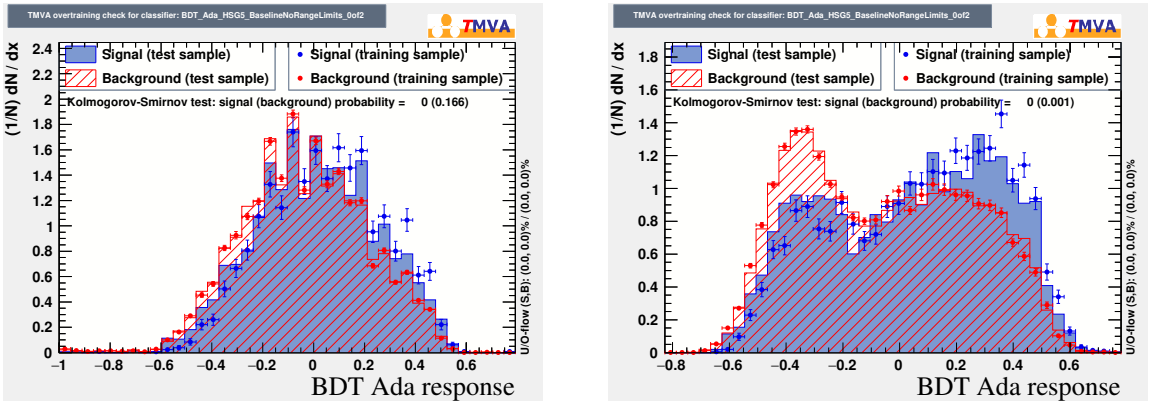
Apart from overfitting, there is a second form of bias that can affect the reliability of the classifiers used for the VH search: the fact that Monte Carlo simulations do not perfectly model the data. Classifiers are usually trained on simulations, and will therefore not make optimal predictions when used to classify data events. In machine learning, this would be called "covariate shift" [149], in high-energy physics "mismodelling". Both terms refer to the situation that distributions in the training domain differ from distributions in the target domain, that is, data. In high-energy physics, such differences are mitigated by reweighting the simulations[10], where necessary, or by discarding poorly described regions of the phase space. All remaining differences need to be addressed by assigning modelling uncertainties to the Monte Carlo simulations.

Covariate shifts can be detected using machine learning: a classifier can be trained to separate the training domain from the target domain, that is, to separate simulations from data events. If the classifier succeeds in separating the two classes, a covariate shift is present. The advantage of this method is that it not only checks the modelling of single variables, but also if the dependencies of multiple variables are predicted correctly. Such tests should be performed after known deficits of the simulations are corrected because they would otherwise only detect the known deficits.

Figure IV.18 shows two classifiers, which have been trained to separate data events from simulations in the $VH \to Vb\bar{b}$ phase space. The BDTs slightly overfit the data sample (labelled "signal") because the data statistics are considerably lower than the Monte Carlo statistics. The model parameters were not optimised for this kind of training, hence the overfitting, but an optimal training is anyway not necessary: the information that can be obtained from the test samples (filled histograms) is sufficient to conclude that a good description of the data is achieved in the 1-lepton channel, figure IV.18(a). The minimal shift is caused by slightly different b-tagging efficiencies between data and simulations, which are covered by the b-tagging uncertainties.

In figure IV.18(b), a covariate shift is detected, though, at least for a part of the phase space. The figure shows a classifier trained on the least sensitive region of the VH analysis: 2 leptons, 2 jets, low p_{T}^{V}. Based on this outcome, events with classifier scores larger than 0.2 were studied in more detail. It was found that observables such as the angles between the b-jets and the invariant mass of the b-jets are not described correctly by the Monte Carlo simulations. These are used by the covariate-shift classifiers to separate data from Monte Carlo events. This misdescription is caused by a deficit of $Z+$jets events in the Monte Carlo simulations: the $Z+$jets backgrounds have to be scaled up by about $16\,\%$ to match the

[10] Interestingly, Shimodaira shows in his paper on domain adaptation [149] that a likelihood function for predictive inference under covariate shift should be weighted by "the ratio of the density function of the covariate in the population to that in the observations". The "covariate" is an input variable, and the "population" corresponds to data events, whose distributions one wants to predict. The "observations" correspond to the training examples. In high-energy physics, this corresponds to simulations, contrary to what the term Shimodaira uses seems to imply. The weighting prescription quoted above is therefore: simulated events should be weighted by the ratio of the distribution in data over the distribution in simulations. This is identical to the reweighting used to correct Monte Carlo distributions that is normally used in high-energy physics.



(a) 1-lepton analysis, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$

(b) 2-lepton analysis, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$

Figure IV.18: BDTs trained to separate data (blue) from Monte Carlo events (red). Poor separation is a sign of good Monte Carlo description because it is not possible to separate data from Monte Carlo events. Figure **(b)** points to a misdescription in the 2-lepton category, which is caused by a deficit of Z+jets events in the simulations. Further discussion in the text.

data. This scaling changes the background composition because the amount of Z+jets is increased in relation to $t\bar{t}$ events. Since Z+jets and $t\bar{t}$ events have different event kinematics, the distributions of several observables such as angles between jets are changed.

Covariate-shift classifiers for all categories of the VH analysis – and also for the 2-lepton channel after an approximate normalisation correction for Z+jets – are shown in appendix B 4 on page 211. No further poorly described categories could be detected. Although a covariate shift in the category 2 leptons, low p_{T}^{V}, 3 jets was visible before the normalisation correction, there is no significant separation after the correction. This demonstrates that the misdescription is caused by incorrect normalisations. The only misdescribed category remains 2 leptons, low p_{T}^{V}, 2 jets, but the separation decreases after the normalisation correction. This is because an overall normalisation correction by $16\,\%$ applied to all Z+jets events is not sufficient: the full correction, which has to be done separately for 2/3 jets, in dependence of the jet flavours and including b-tagging uncertainties, is only achieved with the profile-likelihood fit discussed in the following sections.

The detection of a covariate shift using BDTs thus proved to be a powerful and easy-to-implement method to check the quality of the Monte Carlo description: it successfully detected the shape differences due to the deficit of Z+jets events, and did not indicate further problems. Throughout this thesis, an inclusive $16\,\%$ normalisation correction is therefore applied to Z+jets samples for all plots of the 2-lepton channel, but it is *not* applied to inputs to the profile-likelihood fit. This allows for an all-in-one determination of the optimal scale factors during the profile-likelihood fit.

IV 5 Statistical Analysis of the Data

IV 5.1 A Likelihood Model for Extracting a Signal Strength

The statistical model that is used to extract the Higgs signal in this thesis is a binned likelihood[11] model, which is implemented with the `RooFit/RooStats` [120, 150, 151] framework. This discussion follows [152, 153].

For analyses such as the $VH \to Vb\bar{b}$ search, the observed number of data events can be described by the expected numbers of signal and background events S and B:

$$n = \mu S + B$$

The signal strength μ is a parameter to scale the amount of signal events. It facilitates hypothesis testing: setting $\mu = 1$, the signal+background hypothesis is tested; setting $\mu = 0$, the background-only hypothesis is tested. For the likelihood models discussed here, μ is the parameter of interest, and in this thesis, it represents the signal strength of the Standard Model $VH \to Vb\bar{b}$ process.

For n events observed in data, the likelihood function L given μ, reads:

$$\begin{aligned} L(\text{data} \mid \mu) &= \text{Poisson}(n \mid \mu S + B) \cdot \left[\prod_{e=1}^{n} \frac{\mu S\, f_S(x_e) + B\, f_B(x_e)}{\mu S + B} \right] \\ \text{Poisson}(n \mid \mu S + B) &= \frac{(\mu S + B)^n e^{-(\mu S + B)}}{n!} \end{aligned} \tag{IV.8}$$

This likelihood is the product of two terms:

- A Poisson likelihood term to obtain n events given that $\mu S + B$ events are expected.
- For each event e, the probability density of measuring x_e (*e.g.* a BDT score) given the relative fractions of signal and background: the functions $f_{S,B}$ are probability density functions for the observable x, and f_S is for signal, f_B for background events. The expression $B \cdot f_B(x_e)\,\mathrm{d}x$, for example, denotes the expected number of background events in an interval of range $\mathrm{d}x$ around x_e. Using the 1-lepton $WH \to \ell\nu b\bar{b}$ BDT in figure IV.13(a) on page 91, typical values would be $f_S\,\mathrm{d}x = 0.1\,\%$, $f_B\,\mathrm{d}x = 10\,\%$ in the leftmost bin with $x_e = -1$, and $f_S\,\mathrm{d}x = 10\,\%$, $f_B\,\mathrm{d}x = 1\,\%$ at the signal peak at $x_e = 0.4$.

For given data, the likelihood function only depends on μ. Finding the maximum of the likelihood function yields the most probable value of μ, the maximum-likelihood estimator (`ML`). In practice, this is achieved by minimising the negative logarithmic likelihood function (`NLL`):

$$\begin{aligned} -\ln L(\mu) &= -n \ln(\mu S + B) + (\mu S + B) + \ln(n!\,) - \sum_{e=1}^{n} \ln \left[\frac{\mu S\, f_S(x_e) + B\, f_B(x_e)}{\mu S + B} \right] \\ &= (\mu S + B) + \ln(n!\,) - \sum_{e=1}^{n} \ln \left[\mu S\, f_S(x_e) + B\, f_B(x_e) \right] \end{aligned} \tag{IV.9}$$

The computationally difficult $n!$, as well as other constants, can be ignored for the maximisation of the likelihood function. Ignoring these shifts the log-likelihood, but it does not affect the location of the maximum. Obtaining the PDFs (in this chapter: Probability Density Functions) of the signal and

[11] Following the conventions in the HistFactory user's guide [152], "likelihood" will denote a probability that depends on data. A likelihood therefore quantifies if the data are compatible with the model being tested, whereas probability just denotes the chance of an event to occur.

background distributions, however, is challenging because the functional form of the signal/background distributions is usually not known. The PDFs are therefore approximated by histograms of certain observables such as the BDT classification score $y(\boldsymbol{x})$, which are estimated using simulated events.

Denoting the Poisson-distributed bin contents with 'ν_i', the likelihood of a histogram reads:

$$\begin{aligned} L(n_{\text{tot}} \mid \mu) &= \text{Poisson}(n_{\text{tot}} \mid \mu S + B) \cdot \left[\prod_{i \,\in\, \text{bins}} \frac{\mu \nu_i^{\text{sig}} + \nu_i^{\text{bkg}}}{\mu S + B} \right] \\ &= \mathcal{N} \prod_{i \,\in\, \text{bins}} \text{Poisson}(n_i \mid \mu \nu_i^{\text{sig}} + \nu_i^{\text{bkg}}), \\ \text{where} \qquad f_S(x) &\to \frac{\nu_i^{\text{sig}}}{S \Delta_i} \quad \text{and} \quad f_B(x) \to \frac{\nu_i^{\text{bkg}}}{B \Delta_i}. \end{aligned} \tag{IV.10}$$

Δ_i is the width of the i^{th} bin. The normalisation factor $\mathcal{N}$ is constant and can be removed when computing the NLL. Following the convention in [152], L is said to depend on 'variables' and 'parameters'. Variables are values that are measured in the experiment such as n, the number of observed events, whereas parameters (Greek letters) are adjusted to maximise the likelihood. If the dataset is fixed, variables also have a fixed value: the n_i are constant, but parameters like μ can be adjusted to maximise the likelihood.

Generalisation for Multiple Channels and Monte Carlo Samples

The probability density function (IV.10) can be generalised to multiple channels/categories and multiple background sources by multiplying the PDFs of all channels, and by summing the bin contents of all signal/background samples:

$$\begin{aligned} L(n_c, x_e \mid \phi_p) &= \prod_{c \,\in\, \text{channels}} \left[\text{Poisson}(n_c \mid \nu_c) \cdot \prod_{e=1}^{n_c} f_c(x_e \mid \phi_p) \right] \\ f_c(x_e \mid \phi_p) &= \frac{\nu_{ci_e}}{\nu_c} \qquad \text{with} \quad \nu_c = \sum_{i \,\in\, \text{bins}} \nu_{ci} \\ \nu_{ci}(\phi_p) &= \sum_{s \,\in\, \text{samples}} \phi_{cs}\, \sigma_{csi} \end{aligned} \tag{IV.11}$$

The variables n_c are the observed numbers of data events in the channels c, the parameters ϕ_p are a set of normalisation factors for the simulated samples, and ν_{ci_e} is the expected number of events in the bin where the data event 'e' is observed. The values ν_{ci_e} are normalised using ν_c, the total number of events predicted for each channel c. This number is the sum of all bin contents ν_{ci} in category c, which, in turn, are obtained by summing the predictions of all samples s. The last line contains normalisation factors ϕ_{cs}, which can be used to scale the raw predicted bin contents σ_{csi}, separately for each channel c and sample s.

In the log-likelihood function, the combination of different channels corresponds to summing the logarithmic PDFs. In comparison to equation (IV.10), the likelihood model in equation (IV.11) is more powerful because samples can be normalised or scaled using the additional parameters ϕ, separately or simultaneously across multiple samples and categories.

IV 5.2 Parametrisation of Systematic Uncertainties with Nuisance Parameters

Systematic uncertainties of the Monte Carlo predictions have to be translated into uncertainties of the predicted bin contents. This is achieved by making the likelihood function depend on additional parameters, the `nuisance parameters`. One is usually not interested in measuring these parameters, hence the name, but these can affect the measurement of the parameter(s) of interest. Variations of nuisance parameters change the predicted bin contents, and thereby represent systematic uncertainties in the likelihood. With nuisance parameters highlighted in bold font, the likelihood reads:

$$L\left(n_c, x_e, \boldsymbol{a_p} \mid \phi_p, \boldsymbol{\alpha_p}, \boldsymbol{\gamma_i}\right) = \prod_{c \in \text{channels}} \left[\text{Poisson}(n_c \mid \nu_c) \cdot \prod_{e=1}^{n_c} f_c(x_e \mid \phi_p, \boldsymbol{\alpha_p}, \boldsymbol{\gamma_{i_e}})\right] \cdot \prod_{p \in \mathcal{C}} \boldsymbol{g_p}(\boldsymbol{a_p} \mid \boldsymbol{\alpha_p}) \tag{IV.12a}$$

$$f_c(x_e \mid \phi_p, \boldsymbol{\alpha_p}, \boldsymbol{\gamma_{ci_e}}) = \frac{\nu_{ci_e}}{\nu_c} \quad \text{with} \quad \nu_c = \sum_{i \in \text{bins}} \nu_{ci}(\phi_p, \boldsymbol{\alpha_p}, \boldsymbol{\gamma_{ci}}) \tag{IV.12b}$$

$$\nu_{ci}(\phi_p, \boldsymbol{\alpha_p}, \boldsymbol{\gamma_{ci}}) = \boldsymbol{\lambda}\, \boldsymbol{\gamma_{ci}} \sum_{s \in \text{samples}} \phi_{cs}\, \boldsymbol{\eta_{cs}}(\boldsymbol{\alpha_p})\, \sigma_{csi}(\boldsymbol{\alpha_p}) \tag{IV.12c}$$

This likelihood has new terms. These are:

- **External constraints**: The $\boldsymbol{g_p}(\boldsymbol{a_p} \mid \boldsymbol{\alpha_p})$ terms are external constraints for the nuisance parameters $\boldsymbol{\alpha_p}$, where $\boldsymbol{p}$ is a source of uncertainty. $\boldsymbol{g_p}$ is the likelihood of finding the parameter $\boldsymbol{\alpha_p}$ at the value $\boldsymbol{a_p}$. The values $\boldsymbol{a_p}$ are usually known, or at least estimated, from external measurements or predictions. All nuisance parameters, for which such knowledge is available, are collected in the set of constrainable parameters $\mathcal{C}$.

  Example: constraining the luminosity scale factor $\boldsymbol{\lambda} = \int\mathscr{L} / \int\mathscr{L}_0$. Since the luminosity of the data sample is not known with arbitrary precision, the likelihood function must allow some freedom to adjust the normalisation of the Monte Carlo samples. $\boldsymbol{\lambda}$ in equation (IV.12c) therefore scales all samples simultaneously, but $\boldsymbol{\lambda}$ must be constrained to a reasonable range backed by external measurements. Ideally, this would be done by setting the constraining PDF $\boldsymbol{g}_{\text{lumi}}(\mathscr{L} \mid \lambda)$ to the full likelihood function of an external luminosity measurement, but this would lead to very complicated likelihood functions since the same would have to be done for other nuisance parameters. Instead, one includes a simplified representation of the external luminosity measurement: assuming such a measurement is normally distributed, the maximum-likelihood value $\int\mathscr{L}_0$ and its uncertainty are sufficiently represented by a Gaussian distribution centred around $\int\mathscr{L}_0$ with $\sigma = 68\,\%\text{CL}(\int\mathscr{L})$. The parameter $\boldsymbol{\lambda}$ adds one degree of freedom to the likelihood model, which can be used to adjust the total Monte Carlo normalisation, but the external constraint restricts such normalisation corrections to a few percent. If $\boldsymbol{\lambda}$ significantly deviates from one, the total likelihood falls because a tension with the external luminosity measurement is observed. In this way, additional information is propagated into the $VH \to Vb\bar{b}$ likelihood, thus making it a more powerful model.
- **Statistical uncertainties** of the Monte Carlo simulations, $\boldsymbol{\gamma_{ci}}$. If the statistical uncertainty of the Monte Carlo simulations – or data-driven prediction of the bin content – is not negligible in a given bin, a $\boldsymbol{\gamma}$ parameter is added to include this knowledge in the likelihood model. These parameters allow for limited variations of the predicted bin contents. The magnitude of such variations is determined by adding a constraint term $\boldsymbol{g}_{\text{bin }\boldsymbol{ci}}(\boldsymbol{\nu_{ci}} \mid \text{MC stat.unc.})$ for each bin with a $\boldsymbol{\gamma}$ parameter. Although several samples may contribute to a bin, a single parameter reflecting the total statistical uncertainty of each bin is sufficient.

Example: bin with very high sensitivity. In the VH analysis, the background suppression in the rightmost bins of a BDT response can be so high that the statistical uncertainty of the Monte Carlo simulations is non-negligible. A $\boldsymbol{\gamma}$ parameter is therefore assigned when the relative uncertainty falls below $5\,\%$.

- **Normalisation uncertainties**, $\boldsymbol{\eta_{cs}(\alpha_p)}$. These are similar to the normalisation factors $\boldsymbol{\phi}$ except for the fact that the $\boldsymbol{\eta}$ are constrained by $\boldsymbol{g_p}$-terms, whereas the $\boldsymbol{\phi}$ vary freely. The $\boldsymbol{\eta}$ parameters are used to scale the normalisation of a sample within predefined confidence intervals.

  Example: the multijet normalisation uncertainties. Before the profile-likelihood fit, the normalisations of the multijet templates are measured in data, and the uncertainties of these measurements have to be reflected in the likelihood model using the constraint terms. The $\boldsymbol{\phi}$ parameters, by contrast, are used for the main backgrounds $t\bar{t}$ and $V+$jets. They are not constrained since there are no auxiliary measurements (in the VH phase space), and hence their normalisations are directly inferred from the data.
- **Shape uncertainties**. Not only the normalisations of the samples, but also the contents of single bins can vary due to systematic uncertainties, which leads to changes in the predicted distributions of observables. Therefore, the predicted bin contents $\boldsymbol{\sigma_{csi}}$ in equation (IV.12c) depend on the nuisance parameters $\boldsymbol{\alpha_p}$. Shape changes are also constrained with $\boldsymbol{g_p}$-terms to keep the magnitude of the changes consistent with knowledge obtained from external measurements.

  Example: the jet energy resolution (JER). Variations of the jet energy resolution affect all jet-energy-dependent variables such as $p_{\mathrm{T}}^{\mathrm{jet}}$ or m_{bb}. This leads to shape changes in all distributions that depend on such variables, particularly in an observable like the BDT classification score.

The power of the constraint terms is to incorporate knowledge from auxiliary measurements, as well as theoretical predictions, into the likelihood model: without constraint terms, the nuisance parameters $\boldsymbol{\alpha}$ could be adjusted to any value that maximises the likelihood function. If it is known, however, that parameters like luminosity, normalisations or detector calibrations are likely to be found in a certain region, this information should be reflected by the likelihood model.

Ideally, to correctly incorporate all available knowledge, the full likelihood functions of all external measurements would be inserted. In practice, the constraint terms are much simpler: only the main results like central value and uncertainty are included to simplify the computations. For some systematic uncertainties such as the detector acceptance for $VH \to Vb\bar{b}$ events, no auxiliary measurements exist. In this case, the constraint terms can be seen as a kind of "prior belief" obtained from Monte Carlo predictions. Two kinds of constraint functions are used:

- **Gaussian**: Most systematic uncertainties are assumed to be normally distributed, and are thus constrained with a Gaussian PDF centred at zero and with a standard deviation of one. After fits to data, most $\boldsymbol{\alpha}$ are therefore distributed around zero, and fluctuate mostly between -1 and 1.
  The Gaussian constraint is used for all shape uncertainties and also for normalisation uncertainties. If an uncertainty should affect multiple samples, categories or bins simultaneously, a single constraint term $\boldsymbol{g_p(a_p \mid \alpha_p)}$ and a single parameter $\boldsymbol{\alpha_p}$ are assigned to all of them. In this way, information about a nuisance parameter can be measured in one region, and be transferred to others.
- **Poisson**: The number of Monte Carlo events in each bin fluctuates according to the Poisson distribution, and consequently cannot be constrained with Gaussian terms. The constraint term for $\boldsymbol{\gamma}$-parameters is therefore $g_p = \mathrm{Poisson}(m_i \mid \gamma_i \tau_i)$, where m_i is the number of events predicted in bin i, given $\tau_i = (\nu_i/\delta_i)^2$. This is the squared ratio of the relative Monte Carlo uncertainty δ_i and of the initially predicted bin content ν_i. The γ_i start at one and can fluctuate up and down during the likelihood maximisation.

Interpolation of Shape and Normalisation Changes

Similar to not including the full likelihood function for external constraints, systematic uncertainties are represented only approximately in a likelihood model. The impact of systematic uncertainties on the terms in the likelihood function such as predicted bin contents or normalisations is parametrised using three points: the nominal prediction and two predictions obtained by setting the uncertainty to $\pm 1\,\sigma$. These points are denoted with $\alpha = 0, \pm 1$. To allow for smooth transitions between these points, the bin contents $\sigma_{\text{csi}}(\alpha)$ in equation (IV.12c) are interpolated linearly between these three values of α.

For normalisation uncertainties like the η_{cs} in equation (IV.12c), a hybrid interpolation strategy with polynomial *inter*polation and exponential *extra*polation is used:

$$I(\alpha; I_0, I_+, I_-) = \begin{cases} (I_+/I_0)^{\alpha} & \alpha \geq 1 \\ 1 + \sum_{i=0}^{5} a_i \alpha^i & |\alpha| < 1 \\ (I_-/I_0)^{-\alpha} & \alpha \leq -1 \end{cases} \tag{IV.13}$$

The value I is the model prediction, that is, the normalisation factor of a sample for a particular value of α. I_0 is the initial value, $I_\pm$ the up and down variations caused by the uncertainty.

If an extrapolation is required ($|\alpha| > 1$), the exponential strategy ensures that normalisations cannot turn negative. This could alternatively be solved by using a log-normal constraint term for the nuisance parameter, and using linear extrapolation for the normalisation. With exponential extrapolation and Gaussian constraint, though, the same effect is achieved. This can be seen by writing equation (IV.13) as $I = \alpha'$ with $\alpha' = \mathrm{e}^{\alpha \ln (I_\pm/I_0)}$. Since α is normally distributed, and the other terms are constant, α' is log-normally distributed. The extrapolation prescription therefore simplifies to a linear extrapolation in a log-normally distributed variable. The advantage of using exponential extrapolations is that both normalisation- and shape-related nuisance parameters can be constrained using Gaussian distributions. It also facilitates the interpretation of the post-fit values of the nuisance parameters because both shape and normalisation uncertainties should be distributed normally.

Nonetheless, the exponential strategy has one disadvantage: it has a discontinuous first derivative at $\alpha = 0$, which poses a problem for the likelihood maximisation unless $I_+/I_0 = I_0/I_-$. Therefore, a polynomial interpolation is employed in the regions $|\alpha| < 1$. The six coefficients $a_i(I_+, I_0, I_-)$ are determined from six boundary conditions: the function as well as the first and the second derivatives have to be continuous at the points $\alpha = \pm 1$.

IV 5.3 Statistical Tests for Detecting a $H \to b\bar{b}$ Signal: The Profile-Likelihood Ratio

Likelihood models such as the one described by equations (IV.12a) to (IV.12c) can be used to detect signs of $H \to b\bar{b}$ decays in data. This can be done by testing hypotheses about μ, in particular the null hypothesis H_0 ($\mu = 0$) and the signal+background hypothesis H_1 ($\mu = 1$). Being able to reject H_0 in favour of H_1 shows that the data are compatible with the presence of a new process such as $H \to b\bar{b}$ decays. This does not prove the existence of $H \to b\bar{b}$ decays, though: other processes that lead to similar observations can also lead to a rejection of the null hypothesis.

Alternatively, one can try to exclude H_1. Being unable to exclude the signal hypothesis can be taken as indication for a Higgs-like process provided that it can be shown with simulations that the analysis is sensitive enough for an exclusion. Both approaches are followed in this work. First, however, a strategy must be established to measure μ with statistical and systematic uncertainties using the maximum-likelihood approach: the profile-likelihood ratio.

The profile-likelihood ratio is defined as:

$$\lambda(\mu) = \frac{L\left(\mu, \hat{\hat{\boldsymbol{\theta}}}\right)}{L\left(\hat{\mu}, \hat{\boldsymbol{\theta}}\right)}. \tag{IV.14}$$

Following Cowan et al.'s [154] notation, parameters with the symbol '^' are maximum-likelihood estimators. $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ in the denominator are the signal strength and the vector of nuisance parameters that yield the global maximum of the likelihood for the given data. The maximum can be found by minimising $-L$ or $-\ln L$ with gradient-descent minimisers such as `MINUIT` [155]. In the numerator, the same likelihood function is used, but μ is taken as a parameter of the function λ. This means that μ is fixed and *not* being optimised during the likelihood maximisation. Since μ cannot be changed by the minimiser, the likelihood in the numerator is either lower or equal to the one in the denominator, hence $0 \leq \lambda \leq 1$.

If a specific value of μ is chosen by the user, the nuisance parameters $\boldsymbol{\theta}$ in the numerator are re-optimised *depending* on μ. Therefore, the likelihood in the numerator is a conditional likelihood. The notation $\hat{\hat{\boldsymbol{\theta}}}$ denotes conditional maximum-likelihood estimators – in this case depending on μ – whereas in the denominator the maximum-likelihood estimators are unconditional.

The profile-likelihood ratio in equation (IV.14) is a special case of a composite statistical hypothesis test described by Wilks [156]. Wilks shows that twice the negative logarithmic ratio of two likelihoods is distributed as χ^2, that is

$$-2\ln\left(\lambda(\boldsymbol{\theta})\right) = \chi^2_k + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \quad \approx \quad \frac{1}{2^{k/2}\,\Gamma\left(k/2\right)} x^{k/2-1} \mathrm{e}^{-x/2}. \tag{IV.15}$$

$\boldsymbol{\theta}$ are the parameters of the likelihoods, χ^2_k is a χ^2 distribution with k degrees of freedom, n is the number of data events. Wilks' result is valid if two conditions are met: firstly, the likelihood in the numerator is required to be a special case of the likelihood in the denominator. The PDF in the numerator must result from the PDF in the denominator by releasing[12] k parameters of the model in the denominator. Secondly, the data need to be 'from a population with a distribution characterized by values of the $\boldsymbol{\theta}$' [156], the model parameters. This means that the parameters need to provide sufficient flexibility that the model can describe the data. For the profile-likelihood ratio, this can be tested by inverting Wilks' result: releasing one parameter of the model, one can test if $-2\ln(\lambda)$ is χ^2_1 distributed by computing the corresponding probability. If all nuisance parameters are released one by one, and each is found to be χ^2_1 distributed as well as consistent with expectations, one can conclude that the nuisance parameters are able to describe the data. This will be tested in detail in section V 2.

This result can be used to compute the uncertainties of the parameter of interest: releasing μ, the PDF of $-2\ln(\lambda(\mu))$ can be measured by maximising the likelihood in the numerator for all values of μ. Integrating this PDF to both sides of the value of maximum likelihood $\hat{\mu}$ yields the $1, 2$ and $3\,\sigma$ uncertainties of $\hat{\mu}$. The corresponding confidence intervals are defined by the central 68.3, 95.4 and $99.7\,\%$ quantiles of the cumulative density function (CDF). Since $-2\ln(\lambda(\mu))$ should be χ^2_1 distributed, which is the distribution of a squared, normally distributed random variable, these quantiles are defined by the points $-2\ln(\lambda) = 1, 4, 9$, that is, the 1, 2 and $3\,\sigma$ quantiles of the underlying normal distribution.

[12] Releasing in this context means that the parameters are released from the fit, and hence are not inferred from data. Instead, their value can be chosen by the user.

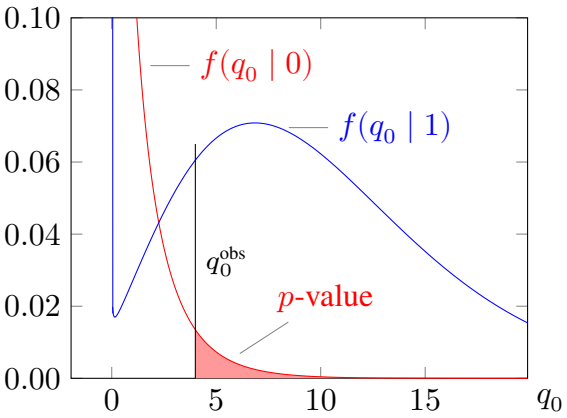


Figure IV.19: Asymptotic distribution of the test statistic q_0, equation (IV.19), which is used to compute p_0. (■) background-only data, (■) signal+background data. With signs of a signal one obtains higher values of q_0, which corresponds to low p_0 because p_0 is always computed using $f(q_0 \mid 0)$.

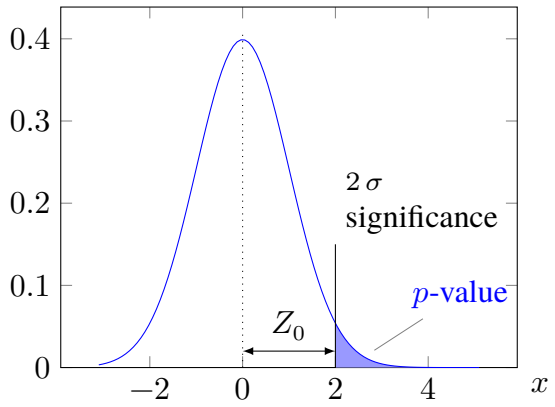


Figure IV.20: Relation of p-value and discovery significance Z_0: in a normal distribution with unit variance, the discovery significance is the lower bound of the upper quantile that equals the desired p-value.

This further allows to give an approximate expression for $-2\ln(\lambda(\mu))$ as shown by Wilks [156] and Wald [157]:

$$-2\ln(\lambda(\mu)) \approx \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \tag{IV.16}$$

$\hat{\mu}$ is the maximum-likelihood value of the signal strength, a normally distributed observable that fluctuates around the true signal strength in data, μ', with a standard deviation of σ. n is again the size of the data sample, and σ has to be estimated from the uncertainties of $\hat{\mu}$.

p-value and Discovery Significance

If the distribution of $-2\ln(\lambda)$ is known, one can perform hypothesis tests. To test, for example, the background-only hypothesis H_0, the p-value of the χ^2 distributions has to be computed. The p-value is the probability to obtain data that are equally or more incompatible with a hypothesis than the measured outcome. As shown in figure IV.19, p_0 is computed by integrating the distribution of a suitable test statistic based on λ from the measured value q_0^{obs} over all values that are even less compatible with the hypothesis:

$$p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_0)\,\mathrm{d}q_0. \tag{IV.17}$$

Here, q_0 is a test statistic for H_0 based on equation (IV.15), and it quantifies the compatibility of the data with the background hypothesis. The PDF $f(q_0)$ is its distribution. If a very low p-value is obtained, one can conclude that q_0 does not follow the distribution $f(q_0)$, and the null hypothesis should be rejected based on the measured outcome. In particle physics, it was agreed that the null hypothesis H_0 needs to be rejected with $p_0 \leq 2.8{\times}10^{-7}$ or "five standard deviations" (see next page) to claim a discovery.

For the discovery of a signal, that is, rejection of the background hypothesis H_0, Cowan et al. [154] propose the following test statistic:

$$q_0 = \begin{cases} -2\ln\lambda(\mu=0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \tag{IV.18}$$

If the maximum-likelihood value of the signal strength, $\hat{\mu}$, is larger than zero, q_0 is equivalent to Wilk's likelihood ratio test evaluated at $\mu = 0$, equation (IV.15). If $\hat{\mu}$ is smaller than zero, q_0 vanishes. This means that a downward fluctuation of $\hat{\mu}$ is not regarded as contradictory to the background-only hypothesis. This is because an incidental lack of events in data should not lead to the exclusion of the background hypothesis – or worse, be misinterpreted as the exclusion of a signal.

Since $0 \leq \lambda \leq 1$, one has $q_0 \geq 0$. The asymptotic PDF of this test statistic is [154]:

$$f(q_0 \,|\, \mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\mathrm{e}^{-\frac{1}{2}\left(\sqrt{q_0}-\frac{\mu'}{\sigma}\right)^2} \tag{IV.19}$$

μ' is the signal strength assumed in the data, σ is the uncertainty of measurements of this signal strength, and Φ in the first term is the CDF of a Gaussian distribution with unit variance. The δ term arises because a one-sided test is performed: q_0 is set to zero if the measured signal strength $\hat{\mu}$ is smaller than zero, but the probability to obtain negative values of $\hat{\mu}$ must be reflected in the CDF, and is thus placed exactly at zero. The CDF of the test statistic is found by integrating its asymptotic distribution, equation (IV.19). The result is:

$$F(q_0 \mid \mu') = \Phi\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right), \tag{IV.20}$$

where Φ is again the CDF of a Gaussian distribution with unit variance, μ' is the signal strength assumed in data, and $\sqrt{q_0}$ is always real.

When using p_0, one tests the assumption that the data show no signs of a signal, that is, $\mu' = 0$. Probability density and CDF therefore simplify to:

$$f(q_0 \mid 0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\mathrm{e}^{-q_0/2}, \quad \text{and} \tag{IV.21}$$

$$F(q_0 \mid 0) = \Phi\left(\sqrt{q_0}\right). \tag{IV.22}$$

For the PDF one finds a mixture of a δ distribution at zero and half of a χ^2 distribution with one degree of freedom, Wilks' result from equation (IV.15). The CDF simply is an integral over a Gaussian distribution.

Since $p_0 = 1 - F(q_0 \mid 0)$ according to equation (IV.17), using equation (IV.22) yields:

$$p_0 = 1 - \Phi\left(\sqrt{q_0}\right). \tag{IV.23}$$

The distribution of q_0 and the p-value for a given q_0^{observed} are shown in figure IV.19. Background-only data most likely yield small q_0 and large p-values (■), whereas data with signs of a signal often yield large q_0 (■). This yields small p-values because these are always computed using $f(q_0 \mid 0)$, that is, assuming no signs of a signal in the data. The largest-possible p-value is 50 %, which is reached when $\hat{\mu} \leq 0$.

p-values can be converted into a more intuitive quantity, the `discovery significance`, Z_0. It maps the measured p-values to p-values of a Gaussian distribution with unit variance, see figure IV.20:

$$Z_0 = \Phi^{-1}(1 - p_0). \tag{IV.24}$$

Here, Φ^{-1} is the inverse CDF of a Gaussian distribution, the function that maps the non-shaded area in figure IV.20 to the abscissa. The critical p-value in high-energy physics, 2.8×10^{-7}, is equivalent to the integral over the Gaussian distribution from $5\,\sigma$ to infinity.

The discovery significance can be computed from q_0 by inserting equation (IV.23) into equation (IV.24). One obtains the simple expression:

$$Z_0 = \sqrt{q_0}. \tag{IV.25}$$

Discovery significance and p_0 can not only be calculated with the observed data, but also with simulations to estimate the sensitivity of an analysis. For this, the likelihood model is fit to the `Asimov dataset`. This is a dataset generated from the signal+background prediction, which simulates the expected distribution of the data. The difference to the nominal Monte Carlo prediction is that the statistical uncertainties of the Asimov dataset are equal to the expected uncertainties of the data. If the plain Monte Carlo prediction was used, estimates of the sensitivity would be too optimistic because usually, one has higher statistics in the Monte Carlo samples than in the data. The signal strength in the Asimov dataset, μ', can be chosen freely. In this way, the result of the H_0 tests can be simulated for any value of μ' that might be encountered in data. Typically, μ' is set to one to obtain the `expected significance` of the Standard Model prediction. In section V 1, this will be used as a metric to compare the Lorentz-invariant MVA to the VH MVA from the current chapter.

Example: Discovery Significance for a Simple Counting Experiment

Using the results from this section, one can estimate the discovery significance for a simple counting experiment, for which one takes the background rate as known, and signs of a signal are quantified by trying to reject the null hypothesis. If one finds n data events while expecting a total number of $\mu S + B$, the likelihood is just one Poisson term:

$$L(n \mid \mu) = \frac{(\mu S + B)^n}{n!} \cdot \mathrm{e}^{-(\mu S + B)}$$

The profile-likelihood test statistic for testing H_0 is:

$$q_0 = \begin{cases} -2 \ln \frac{L(0)}{L(\hat{\mu})} & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

In the numerator, one fixes μ to zero, whereas in the denominator, μ is measured in data by maximising the likelihood. This yields the ML value $\hat{\mu}$, and if it is non-negative, the observed q_0 is:

$$\begin{aligned} q_0^{\text{obs}} &= -2 \ln \frac{B^n / n! \cdot \mathrm{e}^{-B}}{(\hat{\mu} S + B)^n / n! \cdot \mathrm{e}^{-(\hat{\mu} S + B)}} = -2 \ln \left(B^n / (\hat{\mu} S + B)^n \cdot \mathrm{e}^{\hat{\mu} S} \right) \\ &= -2 \left(n \ln \left(\frac{B}{\hat{\mu} S + B} \right) + \hat{\mu} S \right) \qquad = 2 \left(n \ln \left(1 + \frac{\hat{\mu} S}{B} \right) - \hat{\mu} S \right) \end{aligned}$$

If the data statistics are large enough, one can use equation (IV.25) to compute the discovery significance. Using an Asimov dataset, one can further estimate the median expected significance. In this case, the Asimov dataset with $\mu' = 1$ would by definition yield a count of $n = S + B$ data events, and a fit to this dataset would by definition converge to $\hat{\mu} = \mu' = 1$. Inserting these into q_0^{obs} yields:

$$\begin{aligned} \text{median}[Z_0 \,|\, \mu' = 1] = \sqrt{q_0^{\text{Asimov}}} &= \sqrt{2 \left((S + B) \ln \left(1 + \frac{S}{B} \right) - S \right)} && \text{if } B \gtrsim 5; \qquad \text{(IV.26)} \\ &= \sqrt{2 \left((S + B) \left(\frac{S}{B} - \frac{1}{2} \frac{S^2}{B^2} + \ldots \right) - S \right)} && \text{if } S < B; \\ &= \frac{S}{\sqrt{B}} \left(1 + \mathcal{O}(S/B) \right), && \text{if } S \ll B. \qquad \text{(IV.27)} \end{aligned}$$

Equation (IV.26) is the standard approximation for the discovery significance that is used in ATLAS. The asymptotic formulae used to derive it are, however, only valid with sufficient data statistics. This is shown in figure IV.21, which compares Monte Carlo measurements of the median significance with the asymptotic formula from equation (IV.26) (blue line) and the widely used approximation $S/\sqrt{B}$ from equation (IV.27) (red dotted line). The asymptotic formula usually overestimates the significance, but for $B \gtrsim 5$ (dashed line) or large S, it is quite accurate. The exact median significance jumps because the number of observed events n is discrete. $S/\sqrt{B}$ is only accurate if $S \ll B$ holds such that the logarithm in equation (IV.26) is sufficiently approximated by the first two terms of its Taylor series. Otherwise, $S/\sqrt{B}$ strongly overestimates the significance. In this work, asymptotic significances are therefore always computed using equation (IV.26).

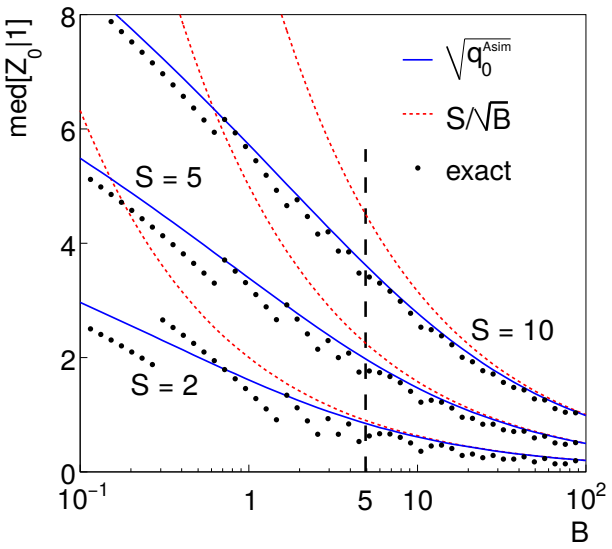


Figure IV.21: Asymptotic vs. exact discovery significance for counting experiments where B is well known and signs of the signal S are searched for in data. Based on [154].

Exclusion Limits

Instead of trying to reject the null hypothesis by calculating the corresponding p-value, one can also select a fixed p-value, and find the highest μ that is still compatible with the data. This defines an upper limit on μ. The p-value is usually set to $5\,\%$, corresponding to a `confidence level` (`CL`) of $95\,\%$. All values of μ with p-values lower than $5\,\%$ are said to be "excluded at $95\,\%$ confidence level". If the signal+background hypothesis $\mu = 1$ falls into the exclusion range, this can be taken as indication that the signal is not visible in the data. By construction, the probability of false exclusion is $5\,\%$, though. If an analysis is sensitive enough to exclude a signal, but an exclusion is not possible given the data, this can be taken as an indication for the presence of a new process.

A simple method to estimate the upper limit on μ is to assume that $\hat{\mu}$ follows a Gaussian distribution centred at the true value of the signal strength, μ', with a variance of σ^2. The $95\,\%$ confidence level is obtained by finding the $95\,\%$ quantile of the corresponding PDF and solving for μ^{up}:

$$0.95 = \int_{-\infty}^{\mu^{\mathrm{up}}} \mathrm{Gauss}(\mu \,|\, \hat{\mu}, \sigma)\, \mathrm{d}\mu \,. \tag{IV.28}$$

For a Gaussian distribution with unit variance, the $95\,\%$ quantile is reached at 1.645. The upper limit can therefore be estimated using:

$$\mu^{\mathrm{up}} \approx \hat{\mu} + 1.645\,\sigma. \tag{IV.29}$$

For a more thorough calculation that correctly takes into account the underlying PDF of μ, Cowan et al. [154] propose the test statistic

$$q_\mu = \begin{cases} 0 & \hat{\mu} > \mu\,, \\ -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu\,. \end{cases} \tag{IV.30}$$

High values of q_μ again indicate incompatibility of the data with the value of μ being tested. In contrast to equation (IV.18), this test statistic vanishes when $\hat{\mu} \geq \mu$ instead of $\hat{\mu} < 0$ because the observed value of the signal strength, $\hat{\mu}$, cannot be excluded by an upper limit. That means that q_0 tests for implausible values of μ to the left and q_μ to the right of $\hat{\mu}$.

The asymptotic PDF of this test statistic reads [154]:

$$f(q_\mu \,|\, \mu') = \Phi\left(\frac{\mu' - \mu}{\sigma}\right)\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}\mathrm{e}^{-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2}, \tag{IV.31}$$

where μ' is again the signal strength assumed in data, σ is the standard deviation of measuring the signal strength, and Φ is the CDF of a Gaussian distribution with unit variance. The δ term again arises because a one-sided test is performed, and the right-hand term converges to the distribution $\frac{1}{2}\chi^2_1$ when $\mu = \mu'$. The p-value, the probability of obtaining a test statistic higher than the observed value, q_μ^{obs}, is [154]:

$$p_\mu = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu \,|\, \mu')\, \mathrm{d}q_\mu, \tag{IV.32}$$

$$= 1 - \Phi\left(\sqrt{q_\mu} - \frac{\mu - \mu'}{\sigma}\right). \tag{IV.33}$$

Asymptotic distributions for q_μ for different values of μ' are shown in figure IV.22. To obtain the 95 % CL upper limit on μ given $\hat{\mu}$, one needs to find the value of μ where $q_\mu^{\text{obs}} = q_\mu^{95}$, which is the value of q_μ that defines the p-value of 5 % shown by the area shaded in red in figure IV.22. This can only be achieved numerically because σ in equations (IV.16), (IV.29) and (IV.33) depends on μ. One therefore has to update μ iteratively using the distribution of p_μ from equation (IV.33):

$$\sigma_i(\mu_i) = \frac{\mu_i - \hat{\mu}}{\sqrt{q_{\mu,i}}} \tag{IV.34a}$$

$$q_{\mu'}^{95}\left(\frac{\mu_i'}{\sigma_i(\mu_i)}\right) = q_{\mu',i} \approx \frac{(\mu_i' - \hat{\mu})^2}{\sigma_i^2(\mu_i)} \tag{IV.34b}$$

$$\mu_{i+1} = \mu_i + \gamma_i(\mu_i - \mu_i') \tag{IV.34c}$$

In each iteration, $\sigma_i(\mu)$ is computed directly from the profile-likelihood ratio $q_\mu = -2\ln(\lambda)$ using equation (IV.34a), which follows from the Wald approximation, equation (IV.16). Given this estimate for σ, one can numerically find the value of $q_{\mu'}$ that yields a p-value of 95 % according to equation (IV.34b). This also yields an estimate of the signal strength in data, μ', which is used to correct μ according to equation (IV.34c). However, since σ in equation (IV.34b) is computed from $-2\ln(\lambda(\mu))$ instead of $-2\ln(\lambda(\mu'))$, one needs to update μ using equation (IV.34c), and repeat the whole computation starting with a new estimate of σ from equation (IV.34a). γ_i is a damping factor that reduces the step size when updating μ. When a precision of $\frac{|\mu-\mu'|}{\mu} < 0.005$ is reached, the process is stopped because equation (IV.33) essentially becomes $p_\mu \approx 1 - \Phi(\sqrt{q_\mu}) = 5\,\%$.

As for the discovery significance, the test statistic q_μ can also be calculated from a fit to the Asimov dataset. This yields `expected limits`, which can be used to measure the exclusion power of an analysis.

The expected changing of exclusion limits due to statistical and systematic uncertainties can be estimated by shifting the expected limit up or down by one (or N) standard deviations of $\hat{\mu}$. This is evident from equation (IV.29) when substituting $\hat{\mu}$ with $\hat{\mu} \pm N \cdot \sigma$. As before, a more accurate way is to repeat the limit calculation numerically using equations (IV.34a) to (IV.34c) and substituting $\hat{\mu} \to \hat{\mu} \pm N \cdot \sigma$ therein. This strategy will be used in this thesis.

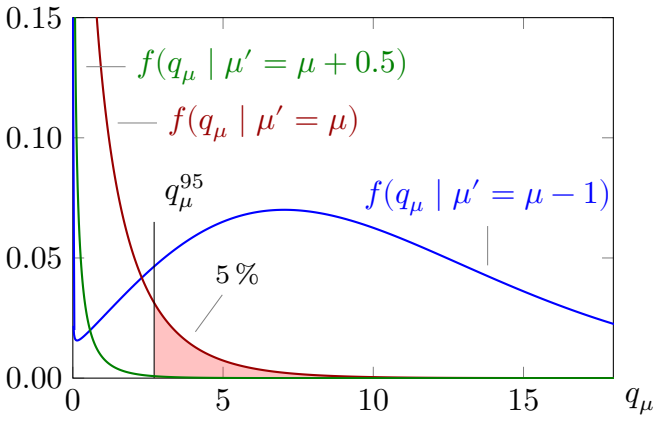


Figure IV.22: Asymptotic distribution of the test statistic q_μ, equation (IV.31), for different values of the signal strength in data (μ'). If the value μ being tested is *higher* than the value in data, q_μ is often significantly larger than zero (■), because q_μ is used to *exclude high* values of μ. If μ is *lower* than μ', q_μ is mostly close to zero (■), because low values of μ are only rarely excluded by this test. If $\mu = \mu'$ (■), q_μ has the same $\frac{1}{2}\chi^2_1$ distribution as p_0 (figure IV.19).
To find the upper limit at 95 % confidence level, one has to find a value of μ such that $q_\mu^{\text{obs}} = q_\mu^{95}$, where a p-value of 5 % is reached.

With the test statistic in equation (IV.30), negative limits on μ would be possible. If one wants to build a model that does not allow for a negative signal strength, the test statistic can be modified to:

$$\tilde{q}_\mu = \begin{cases} 0 & \hat{\mu} > \mu \\ -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \end{cases} = \begin{cases} 0 & \hat{\mu} > \mu\,, \\ -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu},\hat{\boldsymbol{\theta}}(\hat{\mu}))} & 0 \leq \hat{\mu} \leq \mu\,, \\ -2\ln\frac{L(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0,\hat{\boldsymbol{\theta}}(0))} & \hat{\mu} < 0\,. \end{cases} \tag{IV.35}$$

Although the ML estimator of the signal strength, $\hat{\mu}$, can be negative because of statistical fluctuations, the likelihood ratio test is always conducted with the conditional likelihood evaluated at $\mu \geq 0$. In other words, it is conducted against the background hypothesis plus a variable amount of (a positive) signal. This is the test statistic that will be used for upper limits in chapter VI.

The CL_s Method

In high-energy physics, exclusion limits are usually not quoted using p_{s+b} directly. The reason for not doing this is that such tests quantify the compatibility of the data with the signal *plus background* hypothesis. What is, however, desired, is a statement about the compatibility of the data with the signal-*only* hypothesis. Except for a background-free experiment, this is impossible, though. With limits based on p_{s+b}, an exclusion of the $s + b$ hypothesis can occur in cases where the measured data yield fewer events than expected from background predictions. This may be misinterpreted as an exclusion of a signal although it is a statement only about the combined hypothesis. Especially when an analysis is not sensitive to the signal being searched for, this can lead to surprising results: a 95 % confidence limit, by construction, falsely excludes the $s + b$ hypothesis in 5 % of the experiments, regardless of the sensitivity of the experiment. This also holds for the test statistic $\tilde{q}_\mu$ in equation (IV.35).

To obtain what Read [158] calls an "approximate confidence in the signal hypothesis", the confidence in the signal+background hypothesis is normalised to the confidence in the background hypothesis (Read et al. [159, 160]):

$$\mathrm{CL}_s = \mathrm{CL}_{s+b}/\mathrm{CL}_b \tag{IV.36}$$

Note that unlike Read et al., who use $CL_i = p_i(X \leq X^{\text{obs}})$ with the test statistic $X = -2\ln\frac{L(s+b)}{L(b)}$,

ATLAS uses:

$$p_\mu = \int_{q_{\text{obs}}}^{\infty} f(q_\mu \mid \mu)\, \mathrm{d}q_\mu \quad \equiv CL_{\mu s+b}, \text{ and}$$
$$1 - p_b = \int_{q_{\text{obs}}}^{\infty} f(q_\mu \mid 0)\, \mathrm{d}q_\mu \quad \equiv CL_b.$$

With this definition, equation (IV.36) can be rewritten, and be used to exclude the "signal hypothesis" at "95 % confidence level" if

$$\mathrm{CL}_s = \frac{p_\mu}{1 - p_b} \leq 5\,\%.$$

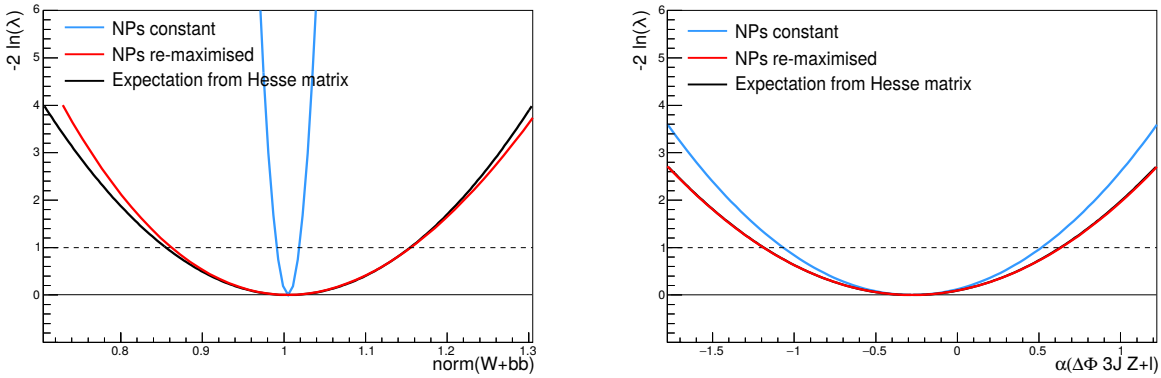
Since the denominator is less or equal to unity, CL_s is usually larger than $\mathrm{CL}_{\mu s+b}$, and therefore yields more conservative limits. If CL_s would be applied to figure IV.22 with $\mu = 1$, one would not search the value of μ where the red p-value is 5 %, but the value where it is 5 % times the area under the blue curve from q_μ^{95} to infinity. This defines $q_\mu^{95(\mathrm{CL}_s)}$, which is to the right of q_μ^{95}, and therefore yields a higher limit. Despite its name, CL_s does hence not define a confidence level because the coverage probability is larger than the stated 95 %. In the special case where the data are identical to the background prediction, $p_b = 50\,\%$, CL_s is twice as large as $\mathrm{CL}_{\mu s+b}$, and yields a limit at 97.5 % confidence level. The benefit of this prescription is that experiments with low sensitivity to μ, that is $p_{\mu s+b} \approx 1 - p_b$, cannot yield a limit of $\mu^{\text{up}} = 0$ even if an extreme downward fluctuation is observed in data. This avoids spurious exclusions of a signal if an experiment is not sensitive enough. Therefore, all q_μ^{95} in this thesis are always based on probabilities normalised according to the CL_s prescription.

IV 5.4 Systematic Uncertainties and the Profile-Likelihood Ratio

So far, the impact of systematic uncertainties on the uncertainties of μ did not receive any special attention. In this section, it will be explained briefly how these translate into uncertainties of μ. Two nuisance parameters of the $VH \to Vb\bar{b}$ likelihood model (section IV 6) will be used as real-world examples for parameters with large and small systematic uncertainties. These are studied after fitting the likelihood model to ATLAS data, which will be validated in more detail in section V 2.

Figure IV.23 shows two profile-likelihood ratios computed according to equation (IV.15). The 1, 2 and 3 σ uncertainties of the parameters are therefore found at the points where the curve reaches 1, 4 and 9. The 1 σ interval is indicated by the dashed line. The profile-likelihood ratios were obtained by maximising the unconditional likelihood in the denominator, releasing the nuisance parameter shown on the axis and re-maximising the conditional likelihood in the numerator for several values of the respective nuisance parameters. The likelihood ratios or log-likelihood differences are plotted against the nuisance parameters. The red curve in figure IV.23 shows the profiles with re-maximisation of the conditional likelihood, and the blue curve shows a profile where only the released nuisance parameter changes while all other parameters remain constant. The black curve is an approximation, which will be discussed at the end of this section.

Figure IV.23(a) shows a parameter with large systematic uncertainties: the normalisation of W+bb backgrounds in the VH analysis. This normalisation cannot be determined accurately because it is anti-correlated to the $t\bar{t}$ normalisation: if W+bb is reduced by a few percent, $t\bar{t}$ increases, and the total background prediction therefore remains nearly constant. Small changes in the W+bb to $t\bar{t}$ ratio will therefore yield a likelihood that is only slightly lower than the global maximum. This means that the numerator and the denominator in the profile-likelihood ratio (IV.14) are almost equal, regardless of the W+bb normalisation. This is because the $t\bar{t}$ normalisation parameters as well as other nuisance



(a) Logarithmic likelihood ratios for a nuisance parameter with large systematic uncertainties.

(b) Logarithmic likelihood ratios for a nuisance parameter with small systematic uncertainties.

Figure IV.23: Profile-likelihood method applied to two nuisance parameters to illustrate the impact of systematic uncertainties. The figures show logarithmic likelihood ratios where the likelihood in the numerator is (■) re-maximised in dependence of the nuisance parameter, (■) not re-maximised, and (■) where the profile is approximated using the Hesse matrix in the minimum of $-2\ln(\lambda)$; see next page.

parameters in the numerator are allowed to compensate changes of norm$(W+bb)$. In figure IV.23(a), this yields the wide red parabola: only strong changes of the $W+bb$ normalisation lead to a notable difference in the likelihoods, and therefore the $W+bb$ normalisation has a large uncertainty of about $10\,\%$.

The blue parabola in figure IV.23(a) shows a likelihood scan without systematic uncertainties: the uncertainty of the $W+bb$ normalisation therefore falls to $\sim 1\,\%$. The parabola is much narrower because the nuisance parameters in the numerator are *not* allowed to compensate for changes of the $W+bb$ normalisation. This is equivalent to an experiment where the $W+bb$ normalisation is measured in data, and all other backgrounds are assumed to be fixed. In this case, only the statistical uncertainty of the data limits the precision of the $W+bb$ measurement.

It is the fact that nuisance parameters are allowed to compensate changes of the parameter of interest (POI) that translates systematic uncertainties implemented in the likelihood model into uncertainties of the POI. If a nuisance parameter can compensate changes of the POI, it also means that the associated systematic uncertainty has a similar impact on the predicted distributions as the signal that is in the focus of the search. The more nuisance parameters can compensate changes of the POI, the higher is the total uncertainty of the POI.

Figure IV.23(b) shows a nuisance parameter with very small systematic uncertainties: an uncertainty for the simulation of the azimuthal angle between jets ($\Delta\Phi$) for $Z+l$ events in the 3-jet region. This is a shape uncertainty that only has a small impact on the predicted distributions. Furthermore, there are no nuisance parameters in the likelihood model that can effectively compensate shape changes associated with this nuisance parameter. Whether other nuisance parameters are re-optimised (red) or remain unchanged (blue) therefore has almost no effect on the profile-likelihood ratio. Consequently, the blue and red parabolas are similar, and the total uncertainty of the nuisance parameter is almost the same as the statistical uncertainty. If this nuisance parameter was a POI, its systematic uncertainties would be negligible.

The power of the profile-likelihood method is that it can compute the simultaneous impact of numerous systematic uncertainties. It automatically computes whether a systematic uncertainty has a high or low impact on the POI because it measures if the effects on the high-sensitivity bins are significant or negligible. It also takes into account external knowledge via the constraint terms and constraints obtained

from fits to the data. This means that nominally large uncertainties can be reduced if the likelihood model can infer the value of certain nuisance parameters from the data, that is, constrain them. If the value of a nuisance parameter cannot be constrained using the data, the constraint terms at least ensure that its impact is consistent with external knowledge such as auxiliary measurements.

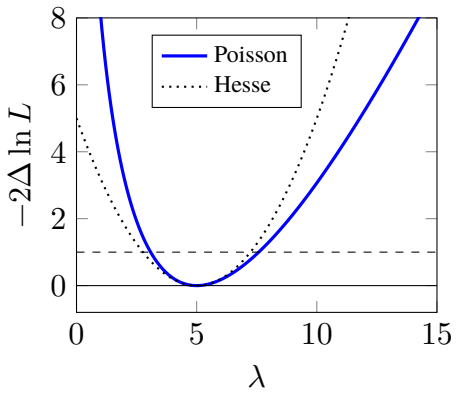


Figure IV.24: Distribution of a log-likelihood difference for a Poisson-distributed nuisance parameter if $n = 5$ events have been measured.

The likelihood profiles in figure IV.23 are all nearly parabolic. This is not a coincidence because most terms in the likelihood function are Gaussian distributions or at least nearly normally distributed. The log-likelihood is therefore quadratic in the nuisance parameters, which is consistent with Wilks' result, equation (IV.15). Deviations from the parabolic shape can occur for parameters where the constraint term is not dominant or absent, that is, parameters that are inferred from data. Nevertheless, measurements of such observables are often normally or nearly normally distributed because of the central limit theorem [161]: the sum of a large number of independent random variables converges to a Gaussian distribution. Nuisance parameters where the parabolic assumption does not hold entirely are normalisation uncertainties (η parameters) and the γ parameters: normalisation uncertainties have slightly asymmetric likelihood profiles. This is because normalisations are often products of several positive and independent random factors, which renders them approximately log-normally distributed[13]. A log-normal constraint should therefore be used, but in practice, one uses the Gaussian constraint with exponential extrapolation and polynomial interpolation, which have an approximately log-normal effect, see page 106. Since both the constraints and the underlying distributions of normalisation uncertainties are only approximately log-normal, the log-likelihood profiles are also only approximately parabolic. Finally, γ nuisance parameters associated with the statistical uncertainty of the simulations follow the Poisson distribution, which in the log-likelihood becomes $n \ln \lambda - \lambda - n!$, see blue curve in figure IV.24. Although this distribution is asymmetric, the central part can be approximated quite well using the parabola labelled "Hesse", whose parameters one can compute by calcuating the second derivative of the Poisson log-likelihood.

Given that most profiles are parabolic, a parabolic approximation can be used to calculate the uncertainties of all nuisance parameters: instead of evaluating the full profile likelihood for each nuisance parameter as shown for the red curves in figure IV.23, one can compute the matrix of 2^{nd} derivatives in the minimum of the NLL. This is done with `Minuit`'s algorithm `HESSE`, and yields the black parabolas shown in figure IV.23. It is much faster than finding the crossing point of the log-likelihood curve with the line at one (`MINOS` algorithm), and simultaneously yields the parabolas for all parameters. In most cases, `HESSE` uncertainties are quite accurate – at least for the $\pm 1\,\sigma$ interval. Figures IV.23(a) and IV.24, however, show that the approximation is not accurate for asymmetric distributions, especially outside the $1\,\sigma$ interval. The `HESSE` approximations will therefore be used to validate the likelihood model in chapter V, but for the measurement of μ, a strategy like `MINOS` will be used in order to obtain the most accurate uncertainties.

[13] More accurately, one can show that the central limit for the product $Y = \prod_i X_i$ of several independent positive random variables X_i has the property [162]:

$$\ln Y = \xi G + \frac{1}{2}\xi^2 F.$$

G is a random variable with a semicircle distribution, and F is a uniformly distributed random variable, which are both centred at zero and independent. Therefore, the log-normal assumption is not entirely correct.

IV 6 The Likelihood Model for the $VH \to Vb\bar{b}$ Search

Table IV.8: Categories and distributions used in the $VH \to Vb\bar{b}$ likelihood fit. The likelihood model is based on the ATLAS $H \to b\bar{b}$ publication [1]. In the 2-lepton channel, the `MM` and `TT` regions are merged because of low data statistics. One further categorisation is not shown: all categories are split into a 2-jet and a 3-jet category, leading to eight classifiers in total.

Category		1 Lepton		2 Leptons	
		$p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$	$p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$	$p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$	$p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$
1 b-tag		MV1c	MV1c	MV1c	MV1c
2 b-tags	`LL`	BDT 1	BDT 2	BDT 3	BDT 4
	`MM`	BDT 1	BDT 2	BDT 3	BDT 4
	`TT`	BDT 1	BDT 2		

Table IV.8 shows the categories of the $VH \to Vb\bar{b}$ analysis that are represented in the VH likelihood model. In both the 2- and the 3-jet region, events with two b-tagged jets are represented by 10 histograms of BDT scores. The categories with only one b-tagged jet are represented by histograms of the `MV1c` b-tagging score. Jets that are not b-tagged are not used. More details on the b-tagging categories can be found in section IV 1.3.

Since event kinematics differ depending on the jet and lepton multiplicity and depending on $p_{\mathrm{T}}^{V} \lessgtr 120\,\mathrm{GeV}$, different classifiers are trained for each category as shown in table IV.8. Each category is covered by a pair of classifiers with identical model parameters but orthogonal training samples to remove any bias due to overfitting. The table shows only four pairs of classifiers, but four more pairs are trained for the 3-jet region. All classifiers are trained on the sum of the `LL`, `MM` and `TT` b-tagging categories because the BDTs have access to the b-tagging information, and can therefore autonomously create b-tagging categories as discussed in section IV 3.3. In spite of the inclusive training, the likelihoods of the `LL`, `MM` and `TT` regions are evaluated separately to measure the normalisations of the V+jets backgrounds. This is necessary because the BDTs are trained to separate signal from background events, but not to disentangle different background classes. The response to various background classes is often similar, especially for different V+jets flavours. Without splitting the histograms into b-tagging categories, it would be challenging to measure and constrain background flavour fractions. In the 2-lepton channel, where data statistics are lower than in the 1-lepton channel, `MM` and `TT` are combined into a single category.

The likelihood model includes systematic uncertainties due to detector resolution and calibration uncertainties and due to imperfections of the simulations. The simulation and modelling uncertainties are listed in table IV.9. These are identical to the ones used in the ATLAS $H \to b\bar{b}$ publication [1]. Many of these originate from the limited precision of parton-shower models or cross-section predictions. In that sense, they are "theoretical" uncertainties, that is, they are not obtained from external measurements, and not constrained by external measurements. Constraint terms for these nuisance parameters therefore do not represent confidence intervals, but are similar to a kind of prior belief. If Monte Carlo generators make conflicting predictions for observables such as jet multiplicities or flavour fractions, systematic uncertainties are added to the likelihood model by comparing the predicted observables for different Monte Carlo programs. The nominal prediction corresponds to $\alpha = 0$, and the conflicting predictions are represented by $\alpha = 1$ or -1. Since it is often unknown which generator should be preferred over the other(s), or since multiple generators may make acceptable predictions, all post-fit values of α between -1 and 1 are acceptable for the "theoretical" nuisance parameters in table IV.9.

Table IV.9: Systematic uncertainties of the signal and background modelling. The symbol "S" denotes shape uncertainties, others are implemented as normalisation uncertainties using η parameters.

Interpretation	Impact	Name in likelihood model
	Signal	
Cross section (scale)	1 % ($q\bar{q}$), 50 % (gg)	`TheoryQCDScale`
Cross section (PDF)	2.4 % ($q\bar{q}$), 17 % (gg)	`TheoryPDF`
Branching ratio	3.3 %	`TheoryBRbb`
Acceptance (scale)	1.5 % – 3.3 %	`TheoryAcc_J2`
3-jet acceptance (scale)	3.3 % – 4.2 %	`TheoryAcc_J3`
p_{T}^{V} shape (scale)	S	`TheoryVPtQCD`
Acceptance (PDF)	2 % – 5 %	`TheoryAccPDF`
p_{T}^{V} shape (NLO EW correction)	S	`TheoryVHPt`
Acceptance (parton shower)	8 % – 13 %	`TheoryAccPS`
	$t\bar{t}$	
3/2-jet ratio	20 %	`ttbarNorm_J3`
High/low-p_{T}^{V} ratio	7.5 %	`ttbarHighPtV`
Top-quark p_{T}, m_{bb}, $E_{\mathrm{T}}^{\mathrm{miss}}$ shapes	S	`TopPt, Ttbar{MBB,Met}Cont`
	W+jets	
Wl normalisation, 3/2-jet ratio	10 %	`WlNorm, WlNorm_J3`
Wcl, W+hf 3/2-jet ratio	10 %	`W{cl,hf}Norm_J3`
Wbl/Wbb ratio	35 %	`WblWbbRatio`
Wbc/Wbb, Wcc/Wbb ratio	12 %	`W{bc,cc}WbbRatio`
$\Delta\phi$, p_{T}^{V}, m_{bb} shapes	S	`WDPhi, WPtV, WMbb`
	Z+jets	
Zl normalisation, 3/2-jet ratio	5 %	`ZlNorm, ZlNorm_J3`
Zcl 3/2-jet ratio	26 %	`ZclNorm_J3`
Z+hf 3/2-jet ratios	20 %	`ZbbNorm_J3`
Zbc/Zbb, Zbl/Zbb, Zcc/Zbb ratios	12 %	`Z{bc,bl,cc}ZbbRatio`
$\Delta\phi$, p_{T}^{V}, m_{bb} shapes	S	`ZDPhi, ZPtV, ZMbb`
	Single top	
Cross section	4 % (s-, t-channel), 7 % (Wt)	`stop{s,t,Wt}Norm`
Acceptance (generator)	3 % – 52 %	`{S,T,Wt}Chan<generator>`
m_{bb}, $p_{\mathrm{T}}^{b_2}$ shapes	S	`WtChan<generator>, TChanPtB2`
	Diboson	
Cross section and acceptance (scale)	3 % – 29 %	`VVJetScalePtS`
Cross section and acceptance (PDF)	2 % – 4 %	`VVJetPDFAlphaPt`
m_{bb} shape	S	`VVMbb_{WW,WZ,ZZ}`
	Multijet	
1-lepton channel normalisation	2 % – 60 %	`MJ{El,Mu}Norm_*`
Template variations, reweighting	S	`MJEl{CaloIso,DR,PtV,TrkIso}` `MJMuTrkIso`
2-lepton channel	–	

The modelling uncertainties for all signal and background processes will now be discussed in more detail:

Signal samples Since the signal strength measures the ratio of the observed to the predicted signal cross sections [39], any uncertainties in these predictions will have a strong impact on μ. Therefore, all uncertainties of cross sections, PDFs (in the following: parton distribution functions) and branching ratios are included in form of η nuisance parameters assigned to the signal samples. Table IV.9 shows the relative normalisation changes encountered when setting the nuisance parameter to $\pm 1\,\sigma$.

The choice of factorisation and renormalisation scales, the parton-shower model and the PDFs used in the Monte Carlo simulations affect the event kinematics and therefore the acceptance of $VH \to Vb\bar{b}$ events. The impact of scale variations is estimated with a `POWHEG` simulation with varied scales; parton-shower uncertainties are estimated by replacing the `Pythia8` showers with `HERWIG` showers, and PDF uncertainties are estimated by exchanging the PDF set. All resulting changes of the acceptance are measured, and parametrised in the likelihoods using the η parameters shown in the table.

Most of these uncertainties have low impact on the distributions of observables, and are therefore sufficiently described by global VH normalisation changes or by normalisation changes that increase the amount of events in the 2-jet region while decreasing it in the 3-jet region or *vice versa*. Scale changes, however, affect the p_{T}^{V} distribution, which is why a shape uncertainty is assigned. A shape uncertainty is also assigned to reflect the impact of NLO electroweak corrections (see section IV 2.1).

Top-pair samples Normalisation changes of background samples such as $t\bar{t}$ and V+jets have a large impact on the measured Higgs signal. Therefore, these are assigned normalisation uncertainties, which often differ between analysis categories. In the 2-jet region, the normalisation of the top-pair background is, for example, described using two sample normalisation factors ϕ, separately in the 1- and 2-lepton channel. This is because the kinematics of the top pairs are different in these channels: in the 2-lepton channel, almost all top quarks decay leptonically, $t \to l\nu b$, and all visible objects are detected. In the 1-lepton channel, the $t\bar{t}$ background is a mixture of fully leptonic decays where one lepton is missed (limited acceptance, τ decay) and events where one top quark decays hadronically, but the resulting light-quark jets are missed.

The `POWHEG+Pythia` $t\bar{t}$ cross-section predictions were compared to several other Monte Carlo generators. The cross-section ratios between the 3- and the 2-jet region were found to be as high as $20\,\%$. A corresponding η uncertainty is therefore assigned to 3-jet top events. A similar normalisation uncertainty is introduced for the high-p_{T}^{V} region: the high-p_{T}^{V} top normalisation is allowed to differ from the one measured in the low-p_{T}^{V} region by assigning a normalisation uncertainty with $\eta = 1 \pm 0.075$.

All regions in the 1- and 2-lepton channel share a common, unconstrained top normalisation factor. The global normalisation is measured in all jet-multiplicity and p_{T}^{V} categories simultaneously, but local normalisation corrections for specific regions are allowed by assigning η parameters. The following scheme shows the top normalisation strategy:

$$\phi_{\text{2j, low } p_{\mathrm{T}}} = \phi_{\text{top}} \qquad \phi_{\text{2j, high } p_{\mathrm{T}}} = \phi_{\text{top}} \cdot \eta_{\text{h}p_{\mathrm{T}}}(7.5\,\%)$$
$$\phi_{\text{3j, low } p_{\mathrm{T}}} = \phi_{\text{top}} \cdot \eta_{\text{3j}}(20\,\%) \qquad \phi_{\text{3j, high } p_{\mathrm{T}}} = \phi_{\text{top}} \cdot \eta_{\text{h}p_{\mathrm{T}}}(7.5\,\%) \cdot \eta_{\text{3j}}(20\,\%)$$

The η parameters are the ratios of the normalisation in a specific region to the global normalisation factor shared by all regions. All nuisance parameters that implement a similar scheme are therefore called "ratio" in table IV.9. The η parameters are defined such that ϕ_{top} is the normalisation in the main signal region: the 2-jet region.

The comparison of different $t\bar{t}$ simulations further revealed differences in the distributions of several observables that are used in the MVA. Shape uncertainties for the variables m_{bb}, $E_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T}}^{\mathrm{top}}$ are therefore parametrised in the likelihood. The distribution of $p_{\mathrm{T}}^{\mathrm{top}}$ is corrected on truth level (*cf.* section IV 2.5), and half of the correction is applied as systematic uncertainty. The other shape uncertainties listed in the table are estimated by reweighting reconstructed quantities, separately for the 2-/3-jet and the low-/high-p_{T}^{V} regions. For each of these uncertainties, a nuisance parameter is assigned, which causes simultaneous shape variations in all regions.

W+jets The W+jets normalisations require a more sophisticated uncertainty model because the jets in W+jets events occur in all jet flavours, unlike for $t\bar{t}$ pairs. W+bb is the most important jet flavour combination for the measurement of μ, and is therefore measured directly. This means that the normalisation of the W+bb sample floats with an unconstrained normalisation factor ϕ_{W+bb}. The W+cc, W+bc and W+bl flavours float simultaneously with ϕ_{W+bb}, but are allowed to differ from the W+bb normalisation by assigning ratio uncertainties using η parameters with a $1\,\sigma$ impact of 12, 12 and 35 %. The magnitude of these uncertainties was determined by comparing `Sherpa` to `ALPGEN`. The W+cl normalisation is taken as independent from the other flavour components, and measured in data, mostly in the 1b-tag region. This is because sg scattering contributes to W+c production. W+l events are not floating freely. They are allowed to vary around the NLO cross-section predictions by 10 %. This uncertainty was estimated from normalisation measurements in the 0b-tag region. Like in the $t\bar{t}$ case, the normalisations in the 3-jet region are allowed to differ from the ones measured in the 2-jet region by introducing additional 3-jet η parameters, separately for W+l, W+cl and the heavier flavours.

Shape uncertainties for W+jets simulations originate from the $\Delta\phi$ correction discussed in section IV 2.3 and from modelling uncertainties for the p_{T}^{V} and m_{bb} distributions. These were obtained by comparing the `Sherpa` predictions with `POWHEG+Pythia8`, `aMC@NLO+HERWIG` and `ALPGEN`, by varying the renormalisation and factorisation scales, as well as the PDFs. The comparisons yielded differences regarding the jet multiplicity of up to 10 % and differences in the distributions of p_{T}^{V} and m_{bb}. The corresponding shape uncertainties are represented by the nuisance parameters `WMbb` and `WPtV`.

Z+jets The Z+jets samples are assigned similar systematic uncertainties as the W+jets samples because both are simulated with the same version of `Sherpa`: the Z+l normalisation is taken from an NLO Monte Carlo prediction, and checked in data, the Z+bb and Z+cl normalisations are measured directly, other flavours are coupled to Z+bb using η parameters whose $1\,\sigma$ uncertainties are derived from Monte Carlo comparisons. The 3-jet normalisations are coupled to their 2-jet counterparts using additional η parameters as shown in table IV.9. Most of the uncertainties are lower in the Z+jets case, only the 3/2 jet ratios for Z+cl and heavier flavours are higher.

Shape uncertainties similar to the W+jets uncertainties are assigned for $\Delta\phi$ and p_{T}^{V} as explained in section IV 2.3. An uncertainty for m_{bb} was derived by comparing the simulated m_{bb} spectrum to the spectrum measured in data. This is possible because the 2-lepton region is less contaminated with $t\bar{t}$ and QCD multijet backgrounds than the 1-lepton region, which yields an almost pure sample of Z+jets events. The signal region $100\,\mathrm{GeV} < m_{bb} < 150\,\mathrm{GeV}$ was excluded from this comparison, but the uncertainty is extrapolated into this region. Further comparisons of `Sherpa` with `ALPGEN` only showed differences that are covered by the uncertainties derived using the data.

Single t The single-top backgrounds are normalised to NNLO cross-section predictions, and the theoretical uncertainties are assigned as normalisation uncertainties [163]. The samples do not float

freely during the fit. By comparing to other Monte Carlo generators such as ACERMC, MC@NLO, POWHEG+HERWIG, acceptance uncertainties were derived. These vary from $5\,\%$ for Wt production, $\sim 20\,\%$ for s-channel production and up to $52\,\%$ for t-channel production in 2-jet events with $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$. For this reason, acceptance uncertainties are assigned separately for s-, t- and Wt-channel production, separately for 2/3 jets and separately for low/high p_{T}^{V}.

The Monte Carlo comparisons also revealed differences in the spectra predicted for m_{bb} and $p_{\mathrm{T}}^{b_2}$, the transverse momentum of the subleading b-jet. Shape uncertainties are therefore assigned for these observables.

Diboson The normalisations of the POWHEG diboson simulations are corrected to NLO in QCD using MCFM. The samples do not float freely, but normalisation uncertainties due to changes in renormalisation and factorisation scales and the choice of PDFs are assigned separately for 2/3 jets, low/high p_{T}^{V} and for the 1-/2-lepton channel. The dijet-mass distribution in $Z \to b\bar{b}$ processes is affected by the parton shower and hadronisation model, which was revealed by comparing POWHEG+Pythia8 with HERWIG. Corresponding shape uncertainties were therefore assigned for all three diboson channels.

Multijet The derivation of the multijet template was discussed in section IV 2.6, and all shape corrections and assumptions used there are backed by shape uncertainties. These are MJElTrkIso, MJMuTrkIso for the track isolation inversion, separately for different categories, an uncertainty covering the choice of the calorimeter isolation for electrons (MJElCaloIso) and the extrapolation of 1 b-tag events to the 2 b-tag region for electrons (MJElDR, MJELPtV). For muons, no significant shape differences were observed, but the statistics of the muon multijet template are lower than for electrons.

Since the multijet background is fitted into the gap between the electroweak predictions and the distributions observed in data, separately for all jet- and b-tag multiplicities, normalisation uncertainties are also assigned separately. These uncertainties were derived by comparing fits in $E_{\mathrm{T}}^{\mathrm{miss}}$ to fits in m_{T}^{W}, which yield different normalisation factors. The multijet normalisation is measured simultaneously for LL, MM and TT, but uncertainties are assigned separately for each region to cover flavour-fraction uncertainties of the multijet "truth tagging" procedure, see section IV 2.6. In the signal region with one electron and two jets, the normalisation uncertainties amount to 11, 14 and $22\,\%$ for the LL, MM and TT regions. In the muon channel, uncertainties are even larger: 28, 42 and $60\,\%$. This is because the overall multijet background with muons is smaller, and because the templates have lower statistical precision. In the 3-jet regions, the normalisation nuisance parameters for LL, MM, TT were merged into a single nuisance parameter since their impact on the $t\bar{t}$ normalisation, the dominant background in this region, is negligible. The large number of independent normalisation nuisance parameters for the QCD multijet background prevents any extrapolation of normalisation measurements across categories. Therefore, normalisation uncertainties cannot be constrained significantly during the profile-likelihood fit, which means that the QCD multijet estimate is very conservative.

Table IV.10: Summary of experimental systematic uncertainties applied to all simulated events. The last column lists the names of the nuisance parameters in the likelihood model.

Jet energies		
Energy resolution		
• all jets	1 NP	`JetEResol`
• b-jets	1 NP	`BJetReso`
Jet energy scale (JES)		
• Calibration	6 NPs	`JetNP[1-6]`
• MC modelling	2 NPs	`Jet{EtaModel,EtaStat}`
• Fast/full simulation difference	1 NP	`JetNonClos`
• Pileup dependence	4 NPs	`Jet{NPV,PilePt,PileRho,Mu}`
b-jets		
• Detector response for b-jets	1 NP	`JetFlavB`
• Energy loss in B decays due to μ, ν	1 NP	`JetBE`
light/g components in non-b-jets		
• Fraction	4 NPs	`JetFlavComp_*`
• Difference in response	4 NPs	`JetFlavResp_*`
b-tagging		
b-jet	10 NPs	`BTagB[0-9]Effic`
c-jets, τ leptons	15 NPs	`BTagC[0-14]Effic`
light jets	10 NPs	`BTagL[0-9]Effic`
MC generator dependence	4 NPs	`BTag[BC]{Pythia8,Sherpa}`
Other sources		
$E_{\mathrm{T}}^{\mathrm{miss}}$	2 NPs	`MET{Scale,Reso}SoftTerms`
Luminosity	1.9 %	`LUMI_2012`
Lepton efficiency	2 NPs	`{Muon,Elec}Effic`
Lepton isolation	1 NP	`LeptonIso`
Electron energy	2 NPs	`Elec{E,EResol}`
Muon momentum	2 NPs	`MuonEResol{ID,MS}`

Experimental Uncertainties

Experimental uncertainties describe how well particles can be reconstructed, and how well the response of the ATLAS detector to certain kinds of particles is understood. Reconstruction and identification efficiencies for all relevant particles have been measured by ATLAS, and their uncertainties are estimated following standard ATLAS recommendations. Table IV.10 lists these uncertainties, and how they are reflected in the likelihood model. Almost all are shape uncertainties that either are implemented with event weights (reconstruction & identification efficiencies), or are obtained by varying reconstructed quantities such as energies or momenta and re-running the varied events through the analysis chain. The most important sources of uncertainties are:

Jet energy resolution (JER) The JER is the dominant uncertainty for reconstructing the invariant mass of the $b\bar{b}$ system. It is caused by the limited resolution of the calorimeters. It has been measured by ATLAS [75], and is parametrised in dependence of η and p_{T}. The parametrisations follow the calorimeter resolution functions discussed in equations (III.6) and (III.7) on the pages 38 and 39. To estimate the resolution uncertainty, jet energies in all simulated events are randomised according to the measured calorimeter uncertainties.

In addition to the resolution uncertainty common to all jets, b-jets are assigned an extra uncertainty in the $VH \to Vb\bar{b}$ analysis because a dedicated b-jet energy correction is applied to improve the resolution of m_{bb}, see equation (IV.5) on page 77. The uncertainty of this correction is not reflected by the standard ATLAS uncertainties, and was therefore estimated separately [139].

Jet energy scale (JES) Jets have to be calibrated after the reconstruction [164]. The uncertainties of this calibration are parametrised in bins of the jet transverse momentum and pseudorapidity. Since systematic uncertainties of the jet calibration are affected by various sources that often impact large regions of the phase space simultaneously, they are represented in a large covariance matrix. This matrix can be diagonalised, yielding eigenvectors and eigenvalues that represent the main effects of the calibration uncertainties. The uncertainties with the five largest eigenvalues (`JetNP[1-5]`) are added as nuisance parameters to the VH likelihood model, and the corresponding eigenvectors define the magnitude of the impact on the observables. All other components with small eigenvalues are summed in quadrature, and added as a 6$^{\text{th}}$ nuisance parameter (`JetNP6_rest`). Other jet-energy uncertainties such as pileup dependence, Monte Carlo modelling and jet-flavour dependence are also included, but their impact on the $H \to b\bar{b}$ analysis is low.

***b*-tagging** b-tagging is a crucial tool for the $H \to b\bar{b}$ analysis. Since four operating points from 50 to 80 % are used, and since the b-tagging calibration is p_{T}-, η- and jet-flavour-dependent, b-tagging uncertainties are also represented as eigenvalues and eigenvectors of large covariance matrices. Details on the b-tagging efficiency measurements can be found in [96, 144][14].

The $VH \to Vb\bar{b}$ analysis uses 10, 15 and 10 nuisance parameters for b-, c- and light jets (out of a total of 24, 16 and 48 eigenvectors). Four more nuisance parameters cover differences in b-tagging efficiencies between different Monte Carlo generators. All b-tagging-related uncertainties are uncertainties of the b-tagging probability of single jets, and can therefore be represented by adjusting event weights in dependence of the properties of each jet.

$E_{\mathrm{T}}^{\mathrm{miss}}$ The calculation of the missing transverse energy has to take into account all objects reconstructed in the detector. If systematic uncertainties for these objects are evaluated, for example, by changing the

[14] `MV1c` is not documented in these publications, but the methods for calibrating `MV1c` are identical to the methods used for `MV1`.

jet momenta, $E_{\mathrm{T}}^{\mathrm{miss}}$ is also recalculated. Therefore, $E_{\mathrm{T}}^{\mathrm{miss}}$ is affected by all experimental uncertainties that change an event observable. However, in addition to the reconstructed objects, calorimeter clusters that are not associated to any object also contribute to $E_{\mathrm{T}}^{\mathrm{miss}}$– the "MET Soft Terms". Two systematic uncertainties (`MET{Scale/Reso}SoftTerms`) parametrise the energy scale and resolution of these soft terms [93]. These are mostly important for the 1-lepton channel because $E_{\mathrm{T}}^{\mathrm{miss}}$ is used to detect leptonic W decays.

Luminosity The luminosity uncertainty is the only experimental uncertainty implemented as a normalisation change. All Monte Carlo samples are normalised to the integrated luminosity of $20.2\,\mathrm{fb}^{-1}$, but are allowed to vary according to the luminosity uncertainty derived from beam-separation scans performed in November 2012 [113]. The uncertainty amounts to $1.9\,\%$. The corresponding nuisance parameter simultaneously scales all samples in all categories. At the time of the $H \to b\bar{b}$ publication [1], a preliminary uncertainty of $2.8\,\%$ was used, while the Lorentz-invariant analysis in chapter V uses the updated uncertainty. The impact of this change is negligible, though.

Leptons Lepton uncertainties have vanishing impact on the VH analysis because lepton momenta can be measured with much higher accuracy than b-jet momenta. The latter therefore limit the VH analysis. Furthermore, the lepton identification efficiencies can be measured with high accuracy such that the corresponding systematic uncertainties for the VH analysis are negligible. These would only be relevant if τ leptons were used.

Smoothing and Pruning of Systematic Uncertainties

Most of the experimental systematic uncertainties in table IV.10 vary measured quantities such as jet momenta. This causes events to disappear from the Monte Carlo sample because they might fail one of the selection cuts, which the original event passes. Similarly, events that previously were discarded might pass all cuts, and appear as new events when systematic uncertainties are estimated. Finally, events can also migrate to different categories. Although precisely these effects need to be estimated, the raw effect of the systematic uncertainty can be overshadowed by statistical fluctuations. If systematic uncertainties are estimated on simulations with very high statistics, the statistical uncertainties are negligible, and the impact of the systematic uncertainty can be measured easily. However, if the statistical power of the simulations in certain categories is low because, for example, b-tagging suppresses nearly all events of a light-jet simulation, the shape differences need to be smoothed. Otherwise, statistical uncertainties of the simulations would be taken into account multiple times: by assigning γ parameters and by re-sampling events when trying to estimate systematic uncertainties. Smoothing is applied to all systematic uncertainties that are not weight or normalisation changes, and it reduces noise in the likelihood model. The smoothing algorithm developed for the ATLAS $VH \to Vb\bar{b}$ analysis will be described in more detail in chapter V because in this chapter it will be compared to an improved smoothing algorithm developed for this thesis.

Apart from smoothing systematic uncertainties, noise is also reduced by pruning systematic uncertainties. This is because some experimental uncertainties have little to no impact on the $VH \to Vb\bar{b}$ measurement beyond statistical fluctuations. Therefore, shape and normalisation uncertainties are pruned from the likelihood model if their impact on the likelihood function is insignificant. An uncertainty is pruned if at least one of the following conditions is met:

- Normalisation uncertainties lead to variations $< 0.5\,\%$. This is checked separately for each region and Monte Carlo sample such that uncertainties might be removed only in single categories.

- Both the up and down variations for a shape uncertainty change a distribution in the same direction. In this case, finding the minimum of the NLL is an ill-posed, degenerate problem. It is reasonable to prune such uncertainties because this effect mostly occurs for uncertainties whose impact is overshadowed by statistical fluctuations.
- Shape uncertainties only introduce differences of $< 0.5\,\%$ in all bins of a distribution.
- If an up-variation was pruned, the down variation is pruned as well (and *vice versa*).
- If a sample in a given region amounts to $< 1\,\%$ of the total background, a shape or normalisation uncertainty is pruned if one of the following conditions is true:
  - The uncertainty changes the shape or event count of a specific background sample by $< 2\,\%$, and the region is background-dominated ($< 2\,\%$ of the signal).
  - The uncertainty changes the shape or event count of the *total* background in the region by $< 0.5\,\%$.

CHAPTER V

Beyond the ATLAS $H \to b\bar{b}$ Analysis

V 1 A Lorentz-Invariant MVA

In this section, an alternative approach to unambiguously select a set of variables for a multivariate analysis that describes the VH topology will be discussed. In contrast to the bottom-up approach, that is, increasing the number of input variables until the performance saturates as was done for the ATLAS MVA in chapter IV, a top-down approach is proposed. The key idea of the top-down approach is to find a set of variables that contains the *full* kinematic information to describe a VH system, and to reduce it to a set that only contains the *usable* information while avoiding any redundancies. In this way, one arrives at the minimal set that optimally encodes all exploitable kinematic information about the VH system.

There are two reasons for seeking a minimal set without redundancies: firstly, variables without discriminative power introduce noise into the training. Training events need to populate a higher-dimensional phase space, and are thus distributed more sparsely. Though BDTs are quite robust against variables that do not provide discriminative power, such variables will not be ignored completely, and occasionally affect split decisions because of random fluctuations in the training sample. This can have negative effects on the classification results. Secondly, and more importantly, using such variables can increase the systematic uncertainties of the classification result. Since systematic uncertainties are not evaluated during the training of classifiers, negative effects due to useless variables might only be measurable using the profile-likelihood fit.

V 1.1 Lorentz-Invariant Quantities for a Full Description of the VH System

In the most important categories of the $VH \to Vb\bar{b}$ analysis, the 2-jet regions, one expects four objects in the final state. The kinematic information of such events can therefore be described with 16 parameters because each object is fully described by a four-vector. An MVA with more than 16 kinematic variables therefore cannot provide better separation than an MVA with 16 (well-chosen) variables.

There are multiple ways to choose the 16 variables. A natural choice in high-energy physics are Lorentz-invariant quantities because these are independent of the reference frame. With four final-state objects, 10 Lorentz-invariant quantities can be derived that describe the kinematics *inside* the VH system. Six more parameters describe the VH system from the outside, that is, its orientation and its boost. A possible set of parameters therefore is:

- 4 four-vector norms, the masses of the four final-state objects:

$$\langle 1|1\rangle\,,\ \langle 2|2\rangle\,,\ \langle 3|3\rangle\,,\ \langle 4|4\rangle$$

- 6 inner products of distinct four-vectors:

$$\langle 1|2\rangle\,,\ \langle 1|3\rangle\,,\ \langle 1|4\rangle\,,\ \langle 2|3\rangle\,,\ \langle 2|4\rangle\,,\ \langle 3|4\rangle$$

- 3 angles describing the orientation of the VH system
- 3 parameters describing the momentum or boost of the VH system in relation to a reference frame

The Lorentz invariants are obtained by calculating the inner products of all possible combinations of two four-vectors. Since these describe the *inner* kinematics of the VH system, they are especially useful at proton-proton colliders. Other parametrisations would depend on the boost and orientation of the VH system, which cannot be predicted: the longitudinal boost depends on the fractional momenta of the colliding partons. The transverse boost depends on the physics of confinement, the underlying event and possibly also the radiation of gluons, against which the VH system might recoil. This was discussed in section III 1.1 and section III 4. For 10 of the 16 variables, all these uncertainties are therefore irrelevant.

The Lorentz-invariant parameters can be interpreted either as invariant masses or as Lorentz-invariant angles between two objects. This is illustrated by writing down the inner product of a sum of four-vectors $C^\mu \equiv A^\mu + B^\mu$, that could, for example, describe a decay $C \to A + B$:

$$\begin{aligned} C_\mu C^\mu &= (A_\mu + B_\mu)\,(A^\mu + B^\mu) && \text{(V.1)}\\ &= A_\mu A^\mu + 2\,A_\mu B^\mu + B_\mu B^\mu \\ \Leftrightarrow\quad A_\mu B^\mu &= \tfrac{1}{2}(\, C_\mu C^\mu - A_\mu A^\mu - B_\mu B^\mu \,) && \text{(V.2)}\\ &= \tfrac{1}{2}(\quad M^2 \quad - \quad m_A^2 \quad - \quad m_B^2 \quad) \end{aligned}$$

$\langle A|A\rangle$ and $\langle B|B\rangle$ can be identified with the squared masses of the final-state particles, m_A^2 and m_B^2. If A and B have a common ancestor, $\langle C|C\rangle = M^2$ can be identified with the ancestor's mass. If A and B are four-vectors of b-jets (b_i) or of light leptons (ℓ_i), the masses of the final-state particles in equation (V.2) can be neglected. The invariant masses of the heavy particles in a VH system are therefore represented by:

$$\langle b_1|b_2\rangle \approx \tfrac{1}{2}m_H^2, \qquad \text{(V.3a)}$$
$$\langle \ell_1|\,\nu\,\rangle \approx \tfrac{1}{2}m_W^2, \qquad \text{(V.3b)}$$
$$\langle \ell_1|\ell_2\rangle \approx \tfrac{1}{2}m_Z^2. \qquad \text{(V.3c)}$$

If A and B do not have a common ancestor, one can write the four-vectors in terms of angular coordinates:

$$\begin{aligned} \langle\,1\,|\,2\,\rangle &= \begin{pmatrix} p_{\mathrm{T}1}\cosh(\eta_1) \\ p_{\mathrm{T}1}\ \cos(\phi_1) \\ p_{\mathrm{T}1}\ \sin(\phi_1) \\ p_{\mathrm{T}1}\sinh(\eta_1) \end{pmatrix} \cdot \begin{pmatrix} p_{\mathrm{T}2}\cosh(\eta_2) \\ p_{\mathrm{T}2}\ \cos(\phi_2) \\ p_{\mathrm{T}2}\ \sin(\phi_2) \\ p_{\mathrm{T}2}\sinh(\eta_2) \end{pmatrix} && \text{(V.3d)}\\ &= p_{\mathrm{T}1}p_{\mathrm{T}2}\big[\quad \cosh(\eta_1)\cosh(\eta_2) - \quad \cos(\phi_1)\ \cos(\phi_2) \\ &\qquad\qquad - \ \sin(\phi_1)\ \sin(\phi_2) - \sinh(\eta_1)\sinh(\eta_2)\,\big] \\ &= p_{\mathrm{T}1}p_{\mathrm{T}2}\,[\,\cosh(\eta_1-\eta_2) - \cos(\phi_1-\phi_2)\,] \end{aligned}$$

Although equation (V.3d) depends on transverse momenta and angular distances, which are both not Lorentz invariant, the combination of the two (p_{T}-weighted angles or angle-weighted p_{T}) is Lorentz invariant. The parametrisation of the four-vectors used in equation (V.3d) follows from equation (III.4) on page 35.

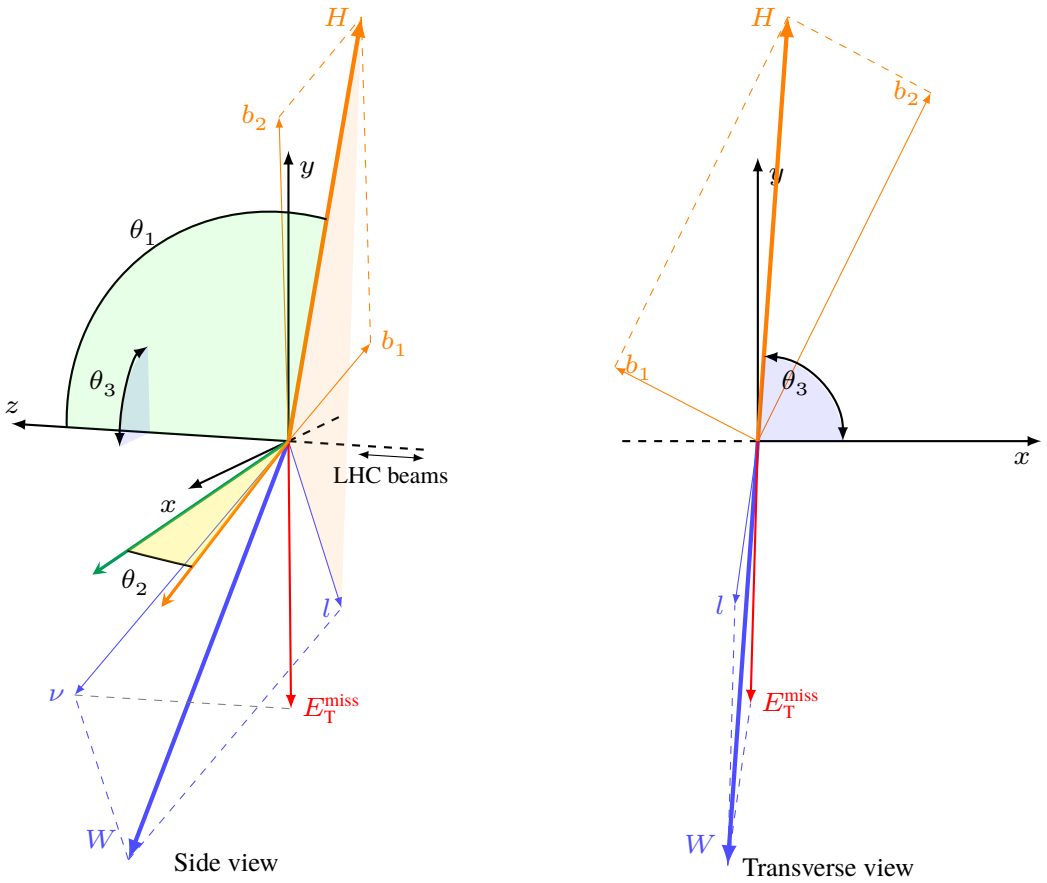


Figure V.1: WH event of the high-p_{T}^{V} category in the laboratory system, with $p_{\mathrm{T}}^{W} = 160\,\mathrm{GeV}$. Three angles describing the orientation of the WH system are marked: θ_1 measures the inclination of the Higgs momentum vector with respect to the beam axis (green arc). θ_2 (yellow arc between the normal vectors of the planes) measures the angle between the Higgs-beam-plane (green) and the Higgs-lepton-plane (orange). θ_3 (blue arc in the transverse view) measures the ϕ-coordinate of the Higgs momentum vector.

The three angles describing the orientation of the VH system depend on the reference frame, and can therefore be defined in multiple ways. Figure V.1 shows momentum vectors for a typical WH event, where three angles have been defined with respect to the laboratory system. The first angle θ_1 is the angle between the Higgs momentum vector and the LHC beams. The Higgs momentum vector is estimated as the sum of the b-jet momentum vectors: $\vec{H} = \vec{b}_1 + \vec{b}_2$. The second angle θ_2 measures the inclination of the plane spanned by the Higgs boson and the lepton[1] against the plane spanned by the Higgs boson and the beams. The inclination can be measured using the normal vectors of these planes, which are also shown in the figure. The third angle θ_3 describes rotations around the z-axis, and is thus equivalent to the angle ϕ in ATLAS coordinates. The description of the VH system is completed by three parameters describing the movement or boost of the VH system, which can be obtained by summing the momentum vectors of all final-state objects.

Not all of these parameters contain useful information, though:

- The four masses of the final-state objects are smaller than what calorimeters or muon system can resolve. These parameters are therefore dominated by noise and do not provide useful information.

[1] In the 2-lepton channel, the lepton with higher transverse momentum is chosen.

- LHC collisions are isotropic in ϕ-direction[2]. The angle θ_3, the ϕ-coordinate of $\vec{H}$, therefore does not provide any information about the VH process.
- Not all boost parameters are useful, either:
  - Longitudinal: The longitudinal boost of the VH system is determined by the parton distribution functions. The boost provides little insights into the VH process because the PDFs are determined by properties of the protons. The average boost, however, varies with the PDFs involved in a collision. The average boost in gg-collisions is low, whereas collisions of quarks and sea quarks are boosted along the direction of the valence quark. More details on PDFs can be found in section III 1.1.
  - Transverse: The two parameters of the transverse boost of the VH system depend on the underlying event, gluon emissions and the physics of confinement. Because of the ϕ-symmetry of the LHC's collisions, the angular component of the transverse boost does not provide usable information. The magnitude of the transverse boost, however, indicates if the $bb\ell\nu$ or $bb\ell\ell$ system is at rest (in transverse direction).

Due to the above restrictions, a sufficient description of the VH system in LHC collisions can be achieved with six Lorentz invariants, two angles and up to two boost parameters. Whether the boost parameters are useful, depends on the final state. This will be discussed further in the next sections.

3-Jet Region

Since events in the 3-jet region have five objects in the final state, four more parameters can be added to the set of discriminating variables. Given that there are five objects, the additional four parameters can all be expressed in terms of Lorentz invariants. One possible set of Lorentz-invariant parameters is:

$$\langle j_1|j_3\rangle\,, \quad \langle j_2|j_3\rangle\,, \quad \langle l_1|j_3\rangle\,, \quad \langle l_2,\nu\,|j_3\rangle\,. \tag{V.4}$$

Again, not all parameters are useful. The angles between the third jet and the leptons are not of interest because gluons, the main cause of an additional jet, are not radiated from leptons. Gluon radiation, however, is the dominant process for moving $VH \to Vb\bar{b}$ events into the 3-jet region: about 30 % of all signal events are found there. Therefore, the four-vector products between the third jet and each of the two b-tagged jets are added to the set of training variables in the 3-jet region. Their distributions are shown in appendix C 1.1. Low values, that is, Lorentz-invariant angles, indicate that the additional jet is close to the b-tagged jets j_1 or j_2, whereas high values indicate large separation. As can be expected if gluon radiation is the main cause of additional jets in $H \to b\bar{b}$ events, the signal distribution peaks at lower angles than for background events. The figures, however, also show that the signal distribution extends to large angles. This is expected because there are always two b-jets: if the additional jet is close to one of them, it is usually far from the other jet. Such a topology can easily be analysed using BDTs, given that both Lorentz invariants are accessible at the same time.

The ATLAS VH analysis also included the mass of all three jets to enhance the discriminative power in the 3-jet region. Since this variable is Lorentz invariant, and the parametrisation in equation (V.4) is not unique, it was tested to add the three-jet mass to the set of training variables for the Lorentz-invariant BDTs. It was found that it is strongly correlated with the three four-vector products that represent the Lorentz-invariant angles between the jets: $\langle j_1|j_2\rangle\,, \langle j_1|j_3\rangle\,, \langle j_2|j_3\rangle$. This indicates that the information expressed by this variable is already accessible to the classifiers. Classifiers with access to the three-jet

[2] The fact that the beams have a small crossing angle and are not exactly aligned with the z-axis is not measurable in the VH analysis.

mass were consequently achieving only insignificantly higher separation ($\ll 1\,\%$). The three-jet mass is therefore not used in this thesis.

Transformation of the Lorentz Invariants

Since the Lorentz invariants are inner products of four-vectors, the range of possible values can be very large: from close to zero for an almost vanishing four-vector product to $\mathcal{O}(10^{10}\,\mathrm{MeV}^2)$[3]. When scanning input variables for optimal splits, the tree-growing algorithm of `TMVA` tests a fixed number of points distributed on a grid across the value range of a variable. For the ATLAS BDTs, 100 points are tested (p. 92), which was increased to 200 points for the Lorentz-invariant BDTs.

Given that the range of possible values for the Lorentz invariants is large, sampling at only 200 points would lead to poor split decisions. The signal distribution, however, does not extend over the full range of possible values. The background-dominated regions are therefore compressed by mapping the value range of each Lorentz invariant into the interval $[0, 1]$ using the following transformation:

$$x' = \frac{x}{c + x}, \qquad c = \text{median}[x^{VH}] \approx 10^{10}\,\mathrm{MeV}^2$$

x is one of the Lorentz-invariant variables, x^{VH} is its distribution for signal events. By choosing the constant c equal to the median of the signal distribution, the most relevant range of x is mapped to about 0.5. If $x \approx c$, x' varies quickly with x. This means that the signal is found towards the centre of the interval, and it is rather spread out, or "zoomed". If $x \ll c$ or $x \gg c$, x' varies only weakly with x. Background-dominated regions are therefore compressed towards the boundaries of the interval $[0, 1]$, where large variations of x have only a small effect on x'. This aids the BDT splitting algorithm because the signal region can be scanned with high granularity while the compressed, background-dominated regions are scanned with low granularity. The reduction of the granularity in the tails of the spectra prevents fine-grained splitting of background regions with low Monte Carlo statistics or high systematic uncertainties. Finally, this transformation eliminates the need to truncate the distributions of input variables, which allows for an undistorted inference of energy and momentum conservation laws, unlike for the ATLAS BDTs as discussed in section IV 3.2.

b-tagging Information

b-tagging information is part of the ATLAS training variables, see section IV 3.3 on page 98. The response of `MV1c` to both the leading and subleading jet in p_T is available to the classifiers. Since kinematic distributions depend on the flavour of the jets, inclusion of this information yields more sensitive classifiers. b-tagging information is therefore also part of all sets with Lorentz-invariant variables. Since b-tagging information, by itself, is not kinematic information about the VH system, it is neither correlated to the other input variables nor is it redundant. It is also independent of the specifics of the 1- or 2-lepton channels, and can therefore be included in both channels without affecting the selection of the optimal kinematic variables.

[3] ATLAS measures momenta, masses and energies in MeV. Recall also that the Lorentz invariants have units of squared masses. $(125\,\mathrm{GeV})^2 = 1.56{\times}10^{10}\,\mathrm{MeV}^2$.

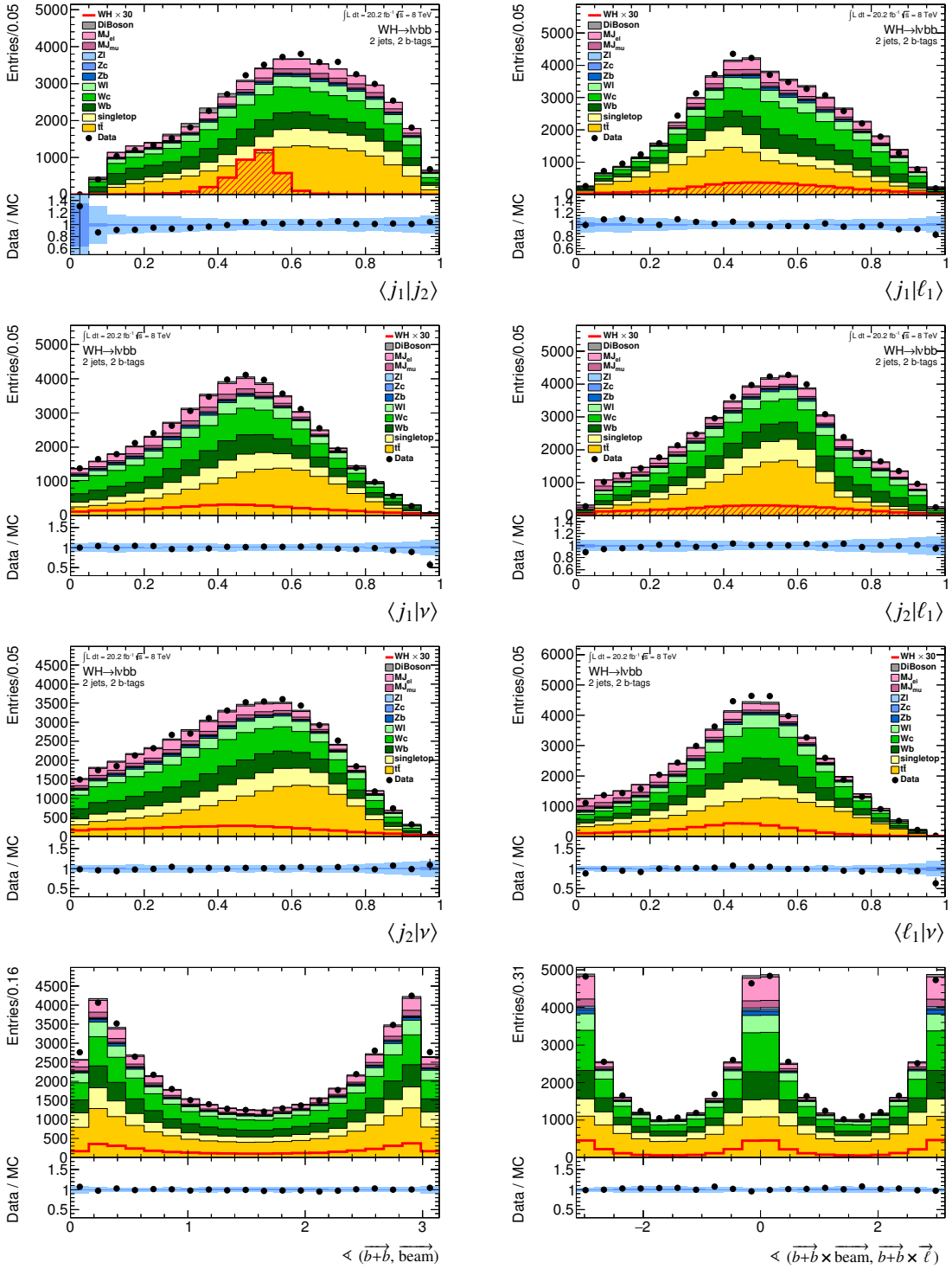


Figure V.2: Kinematic variables of the Lorentz-invariant MVA in the 1-lepton channel after application of the MVA preselection outlined in section IV 1. Shape corrections from section IV 2 are applied, normalisation corrections for the background simulations are not applied except for the multijet template. The dark blue band in the Data/MC plot shows the statistical uncertainty of the background simulations. The larger, light blue band shows the quadratic sum of the systematic uncertainties. Normalisation uncertainties are not shown because they are added in the likelihood model.

V 1.2 Details on the 1-Lepton Channel: Estimation of the Longitudinal Neutrino Momentum

Figure V.2 shows histograms of all kinematic variables used in the 2-jet region of the 1-lepton analysis: the six Lorentz invariants after application of the transformation and two angles. Including information on the longitudinal or transverse boost of the WH system did not significantly improve the separation. This will be shown in section V 1.5. A peculiarity of the computation of the Lorentz invariants in the 1-lepton channel is that not all four-vectors can be determined fully. The reason is that the longitudinal momentum of the neutrino cannot be measured, see section III 3.2. This is also illustrated in figure V.1: the neutrino momentum vector is shown in blue, but the event reconstruction algorithms can only find the red momentum vector, $E_{\mathrm{T}}^{\mathrm{miss}}$, which is the projection of the neutrino momentum onto the transverse plane.

The longitudinal component can be approximated, though, at least in WH events. One has to assume that the available energy in the WH rest system is distributed equally among the four final-state particles:

$$\langle E'_\nu \rangle = \tfrac{1}{4} m_{WH}. \tag{V.5}$$

Primed quantities denote quantities in the WH rest system. Equation (V.5) is a decent approximation because momentum conservation will lead to $\vec{p}'_H = -\vec{p}'_W$, and the rest masses of the final-state particles are negligible. Therefore, $E' = p'$ holds, and the energies of W and H are therefore equal. When W and H decay, their energies are each distributed among two particles with low masses. Consequently, each particle will on average carry $\frac{1}{4}$ of the total energy. Note that m_{WH} is not the sum of the W and H rest masses but the invariant mass, or full energy, of the WH system. The expected z-momentum in the laboratory frame can be obtained by applying a Lorentz transformation that boosts the WH rest system along the z-axis, and by inserting equation (V.5):

$$\langle p_\nu^z \rangle = \gamma \, \langle p_\nu'^z \rangle + \beta\gamma \, \langle E'_\nu \rangle = \beta\gamma \, \langle E'_\nu \rangle = \frac{p_{WH}^z}{m_{WH}} \, \langle E'_\nu \rangle = \tfrac{1}{4} p_{WH}^z. \tag{V.6}$$

This computation is simplified by the fact that the neutrino momenta in the WH rest system are distributed isotropically: the probability of the neutrino moving into the forward direction is the same as of moving into the backward direction. Therefore, the *average* neutrino z-momentum in the WH rest system vanishes: $\langle p_\nu'^z \rangle = 0$.

Finally, the total z-momentum of the WH system needs to be determined. This requires knowledge of the neutrino z-momentum, which is a circular dependency. The WH z-momentum can, however, be approximated using the three observable particles: summing the three observable z-momenta should account for about $\frac{3}{4}$ of the total WH z-momentum. Hence,

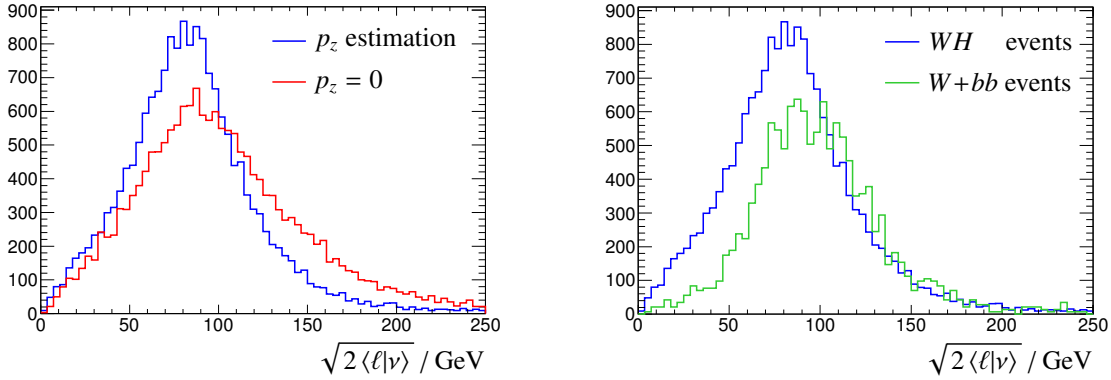
$$p_{WH}^z \approx \frac{4}{3} \left(p_\ell^z + p_{j_1}^z + p_{j_2}^z \right). \tag{V.7}$$

Using the Lorentz transformation (V.6) and inserting the WH z-momentum (V.7) yields the average z-momentum of the neutrino:

$$\langle p_\nu^z \rangle \approx \frac{1}{3} \left(p_\ell^z + p_{j_1}^z + p_{j_2}^z \right). \tag{V.8}$$

This is a plausible result because the equation states that the neutrino z-momentum is approximately the average z-momentum of the three observable final-state particles.

Often, a more simple approach is pursued, that is, setting the neutrino z-momentum to zero. Figure V.3(a) compares these two approaches. The figure demonstrates that the approximation in equa-



(a) W mass with/without p_ν^z estimation in WH events **(b)** W mass hypothesis for WH vs. background events

Figure V.3: Invariant W mass calculated from four-vector products using equation (V.3b) after the standard MVA preselection. Estimating the longitudinal neutrino momentum significantly improves the mass peak. **(a)** Comparison of W mass hypotheses in WH events. The blue curve with estimation of the neutrino p_z shows a more narrow W peak compared to the red curve where $p_z \equiv 0$. **(b)** Applying the p_z estimation to WH events (blue) and $W+$jets background events (green). The mass hypothesis is not accurate for background events because assumptions about the WH system fail. This makes $\langle \ell|\nu \rangle$ a useful variable for a multivariate analysis.

tion (V.8) is superior to the simple approach: the figures show the W mass, obtained from $\langle \ell|\nu \rangle \approx \frac{1}{2} m_W^2$, compare equation (V.3b) on page 128. The distributions should therefore peak at 80 GeV, which is the case for the blue distribution with neutrino momentum estimation. The red distribution without this estimation is shifted to higher masses, and has a broader peak. This is explained by equation (V.3d): if the WH system acquires a sizeable boost in beam direction, the angle η_i of the visible lepton and the neutrino four-vector will either both be positive or both be negative. When ignoring the neutrino z-momentum, that is, setting $\eta_\nu = 0$, the hyperbolic cosine will always grow because it is an even function. $\langle \ell|\nu \rangle$ will therefore also grow.

The assumptions used in the above estimation are only valid for the WH system, and only when looking at many events. For background events, some assumptions will fail, which leads to p_ν^z being estimated wrongly. This, in turn, yields a wrong estimate of the W mass, promoting $\langle \ell|\nu \rangle$ to an observable that discriminates plausible WH kinematics from the kinematics of $W+$jets events. This is illustrated in figure V.3(b), which shows the reconstructed W mass for both $WH \to \ell\nu b\bar{b}$ and $W+bb$ events. Even though the final state is identical, and both samples contain leptonic W decays, assuming that momenta in $W+bb$ events are distributed as in WH events leads to a shifted and broadened W peak.

V 1.3 Details on the 2-Lepton Channel: Separating Top from ZH Events

As mentioned in section V 1.1, a sufficient description of the VH system is achieved with six Lorentz invariants, two angles and a varying number of boost parameters. Classifiers for the 2-lepton channel were found to require the inclusion of one boost parameter: the magnitude of the transverse boost.

The Lorentz-invariant variables[4] describe the kinematics inside the ZH system while angles and boost parameters describe the ZH system from the outside, for example, from the laboratory system. If the $bb\ell\ell$ system was always at rest in transverse direction, and if the longitudinal boost was irrelevant for background suppression, no boost parameters would be necessary. Indeed, the ZH system is usually at rest in transverse direction if no other particles are produced. By contrast, $t\bar{t}$ events selected in the 2-lepton channel will contain two neutrinos. These leave the detector in the vicinity of the visible leptons (*cf.* figure IV.2(b) on page 63). Since these are invisible to the detector, missing transverse momentum can be observed in such events. This also means that the transverse momentum of the $bb\ell\ell$ system does not vanish, and this resembles a ZH system that is not at rest.

To augment the Lorentz-invariant variables with this information, the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ could be added to the set of input variables. $E_{\mathrm{T}}^{\mathrm{miss}}$ should vanish in ZH events, but not in $t\bar{t}$ events. However, $E_{\mathrm{T}}^{\mathrm{miss}}$ does not vanish in ZH events, either, because of calorimetric uncertainties and neutrinos produced inside jets. However, the same information can be expressed using the transverse boost of the ZH system, which is a superior observable because its uncertainties are lower. This is because all final-state objects are visible, and because the lepton momenta can be measured accurately. The transverse boost will also not vanish entirely because of the underlying event or gluon radiation, but its uncertainties are lower than for $E_{\mathrm{T}}^{\mathrm{miss}}$. $E_{\mathrm{T}}^{\mathrm{miss}}$ is affected by pileup activity, calorimeter uncertainties and the $E_{\mathrm{T}}^{\mathrm{miss}}$ calibration.

Figure V.4 shows the BDT input variables for the 2-jet region of the 2-lepton channel. The distribution of the transverse boost, p_T^{VH}, is shown in the top row. ZH and $Z+$jets events peak close to zero as expected; $t\bar{t}$ events have non-vanishing p_T^{VH}[5]. The magnitude of the transverse boost vector thus helps to discriminate ZH events from $t\bar{t}$ events.

The Z+jets backgrounds in figure V.4 have been scaled up by 16 % to correct for the deficit of background events that was also detected using the multivariate modelling checks discussed in section IV 4. The distributions such as the Z mass in the third row, $\langle \ell_1 | \ell_2 \rangle$, show that this is a sufficient approximation for the 2-jet region. The final and more sophisticated determination of the background scale factors will be carried out using the profile-likelihood fit, which will be discussed in section V 2.

One may ask if the transverse boost could also improve the training of classifiers in the 1-lepton channel. If additional objects are not detected, the $bb\ell\nu$ system can also have a non-zero transverse momentum. However, distortions of the transverse kinematics due to undetected objects also affect $E_{\mathrm{T}}^{\mathrm{miss}}$. Since $E_{\mathrm{T}}^{\mathrm{miss}}$ is the transverse component of the neutrino four-vector, the information about p_{T}^{VH} is already encoded in the neutrino four-vector, and thus in the Lorentz invariants. Additional transverse boost information is therefore only necessary in the 2-lepton channel.

[4] This name describes a set of variables that contains Lorentz invariants, angles to describe the orientation of the VH system, b-tagging information and possibly also the boost.

[5] As for p_{T}^{V}, the name p_T^{VH} is only well defined for VH events. In $t\bar{t}$ events, this variable expresses the transverse momentum of two jets and two leptons.

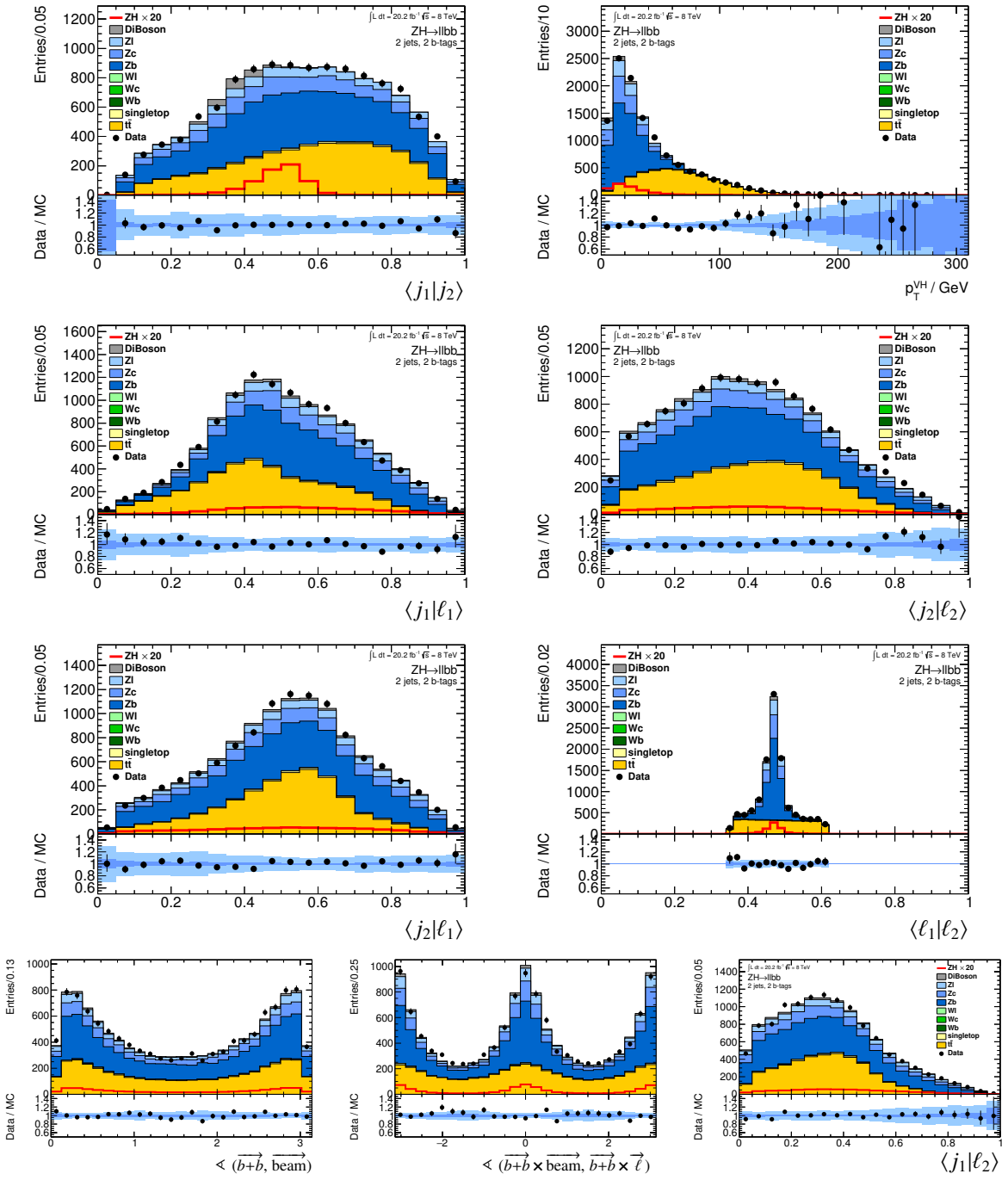


Figure V.4: Kinematic variables of the Lorentz-invariant MVA in the 2-lepton channel after application of the MVA preselection outlined in section IV 1. Shape corrections from section IV 2 are applied. The Z+jets backgrounds are scaled up by 16 %. The dark blue band in the Data/MC plot shows the statistical uncertainty of the background simulations. The larger, light blue band shows the quadratic sum of the systematic uncertainties. Normalisation uncertainties are not shown because they are added in the likelihood model. The particular shape of $\langle \ell_1 | \ell_2 \rangle$ is explained by the event selection: $m_Z = \sqrt{2 \langle \ell_1 | \ell_2 \rangle}$ is truncated during the event selection because no signal events are expected in the side bands of the invariant mass distribution.

V 1.4 Better BDT Training With Gradient Boosting

The configuration of the boosted decision trees discussed in chapter IV was optimised for the 16 variables of the ATLAS $H \to b\bar{b}$ analysis[6]. In optimising the BDT training for the Lorentz-invariant variables, it was found that a different boosting algorithm yields better performance. This algorithm is gradient boosting [165]. Gradient-boosted trees will be called 'BDTG' in the following. Gradient boosting is based on a differentiable loss function that is minimised using a gradient-descent algorithm. High losses indicate poor separation of signal and background, low losses correspond to good separation.

Gradient Boosting:

1. Find a constant that minimises the loss L:

$$F_0(\boldsymbol{x}) = \arg\min_{\gamma} \sum_{i=1}^{N} L(y_i, \gamma)$$

   F_0 is the starting value to construct a classifier, $\boldsymbol{x}$ is the vector of input variables, y_i is the target for each event i. Note that the input variables are not scanned during the first step.

2. In each iteration m, compute pseudo-residuals for all events i:

$$r_{im} = -\frac{\partial L(y_i, F_m(\boldsymbol{x}_i))}{\partial F_m(\boldsymbol{x}_i)}.$$

   $F_m(\boldsymbol{x}_i)$ is the prediction of the classifier for event i after m training cycles.

3. Fit a classifier h_{m+1} to the pseudo-residuals, that is, to the training set $\{(\boldsymbol{x}_i, r_{im})\}_{i=1}^{N}$.

4. Improve the prediction model according to

$$F_{m+1}(\boldsymbol{x}) = F_m(\boldsymbol{x}) + \gamma_{m+1} h_{m+1}(\boldsymbol{x}) \ .$$

   The optimal step size γ is given by minimising the loss:

$$\gamma_{m+1} = \arg\min_{\gamma} \sum_{i=1}^{N} L\Big(y_i, F_m(\boldsymbol{x}_i) + \gamma h_{m+1}(\boldsymbol{x}_i)\Big) \ . \tag{V.9}$$

5. Repeat starting at step 2 until residuals do not decrease, any more.

For decision trees, it is common to compute the step size separately for each leaf of the tree. This algorithm minimises the loss, and thus the difference between the targets y_i and the responses $F_m(\boldsymbol{x}_i)$. By fitting trees to the mistakes (residuals) of the previous classifier, each update of the classifier with a new tree reduces the loss a bit further, until no further improvements can be achieved. The negative gradient of the loss function in the second step shows the direction of steepest descent, that is, the direction where L falls most rapidly.

The approach of using trees to predict the residuals of a loss function is conceptually different from adaptive boosting: Adaptively boosted BDTs always predict a class label, -1 or 1, which is good for classification problems. Gradient-boosted BDTs can predict any value in the value set of L, provided that L can be differentiated. The algorithm can therefore be used for regression, ranking and classifications problems, with any loss function that is differentiable.

[6] See table IV.5 on page 92 for the training parameters and table IV.7 on page 99 for the ATLAS variables.

As adaptively boosted BDTs, gradient-boosted trees can be regularised. This is achieved with similar strategies as for adaptively boosted BDTs: limiting the number of splits or the depth of the trees or requiring a minimal number of events in a leaf. A reduction of the learning rate is also possible by reducing the step size in equation (V.9):

$$\gamma_m \to \beta\gamma_m, \quad \text{with } \beta \in [0, 1].$$

Lower values of the learning rate or "shrinkage" β lead to slower convergence because the minimum of the loss function is approached in smaller steps. The benefit of lower learning rates is higher robustness. Approaching the minimum slowly and constantly re-estimating the gradient samples the region close to the minimum of the loss function with higher precision. This is because close to the minimum, the residuals r_{im} of most events will be very small. New trees will therefore implement only small corrections, and events with small residuals will more often be sorted into the same leaf. This prevents leaves with high statistical uncertainties and therefore numerically unstable gradients.

Slightly randomising the gradient computation further improves robustness. This can be achieved with `bagging`. When bagging is used, each gradient calculation is performed only on a randomly chosen subset of the full training data. This prevents converging into pseudo-minima, which may be caused by statistical fluctuations of the training sample. Instead, minima that are stable with respect to re-sampling from the set of training events are approached, leading to more robust estimates of the underlying distributions. This prevents overfitting the training sample.

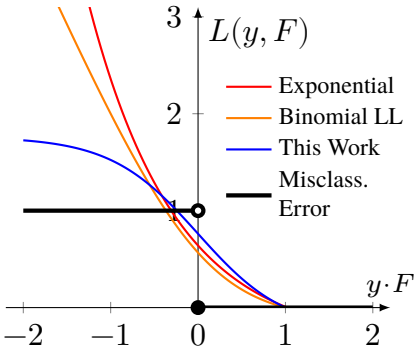


Figure V.5: Loss functions for binary classification. Shifted to pass through (1,0).

The biggest advantage of gradient boosting in comparison to adaptive boosting is that the loss function can be adapted to the problem at hand. For the Lorentz-invariant analysis, the `TMVA` default, binomial log-likelihood loss (orange),

$$L(y, F) = \sum_{i=1}^{N} \ln\left(1 + \mathrm{e}^{2 y_i F(\boldsymbol{x}_i)}\right),$$

was replaced with a less aggressive loss function (blue):

$$L(y, F) = \sum_{i=1}^{N} \frac{2 \cdot y_i}{1 + e^{2 \cdot y_i F(\boldsymbol{x}_i)}}\,. \tag{V.10}$$

The scalar y_i is the target, $\boldsymbol{x}_i$ is the vector of all input variables, and $F(\boldsymbol{x}_i)$ is the response of the classifier for an event i. All loss functions approximate the naïve loss function for binary classification, the misclassification error $\Theta(-y_i \cdot F(\boldsymbol{x}_i))$, which is shown in black: the loss is zero if the prediction of the trees is correct, that is, has the same sign as the target, and it is one if the signs differ. Its derivative is constant, though, and it can therefore not be used for a gradient computation. Adaptive boosting, which was used in chapter IV, was shown [166] to be equivalent to gradient boosting with an exponential loss function (red):

$$L(y, F) = \sum_{i=1}^{N} e^{-y_i F(\boldsymbol{x}_i)}.$$

The default loss functions of `TMVA`, binomial log-likelihood loss for gradient boosting and especially the exponential loss for adaptive boosting, have the shortcoming of heavily penalising misclassified

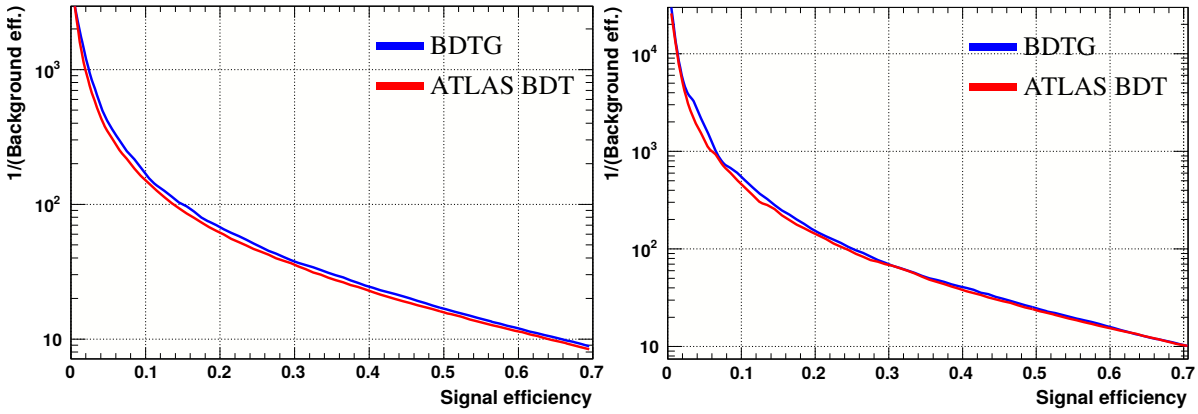
Table V.1: Hyperparameters for the training of gradient-boosted decision trees used with the Lorentz-invariant variables. If two parameters are listed, the first is for the 1-lepton channel, the second for the 2-lepton channel, which has lower training statistics, and therefore profits from stronger regularisation.

`TMVA` Parameter	Value	Description
`Shrinkage`	0.5 / 0.05	Step size along the gradient
`NTrees`	800	Number of trees trained
`MaxDepth`	5 / 4	Maximum depth of decision trees
`nCuts`	200 / 500	Granularity of scanning for possible splits
`BaggedSampleFraction`	0.6 / 0.7	Fraction of randomly selected events per training step
`MinNodeSize`	5 %	Minimum fraction of total training events in a BDT node

events because they grow fast towards -1. With training data where the classes are difficult to separate, that is, signal and background events are overlapping frequently, this leads to steeper gradients (when implemented as gradient boosting) or larger boost weights (when implemented as AdaBoost). This is a disadvantage when a few misclassified background events are found close to correctly classified signal events. Due to their high weights or large residuals, such background events define how trees in the signal region are grown. This may be undesired because the discovery significance scales as $\frac{S}{\sqrt{B}}$, whereas the Gini index scales as $\frac{SB}{S+B} \approx B$ (if S is large). The training algorithm will therefore aggressively try to introduce splits in the signal region even if the impact of a few background events on the discovery significance is low. This can lead to higher noise in the regions of very high classifier scores that determine the sensitivity of a search. However, in such high-sensitivity regions, robustness might outweigh aggressiveness. The section to follow next will show that this is indeed the case for the VH analysis.

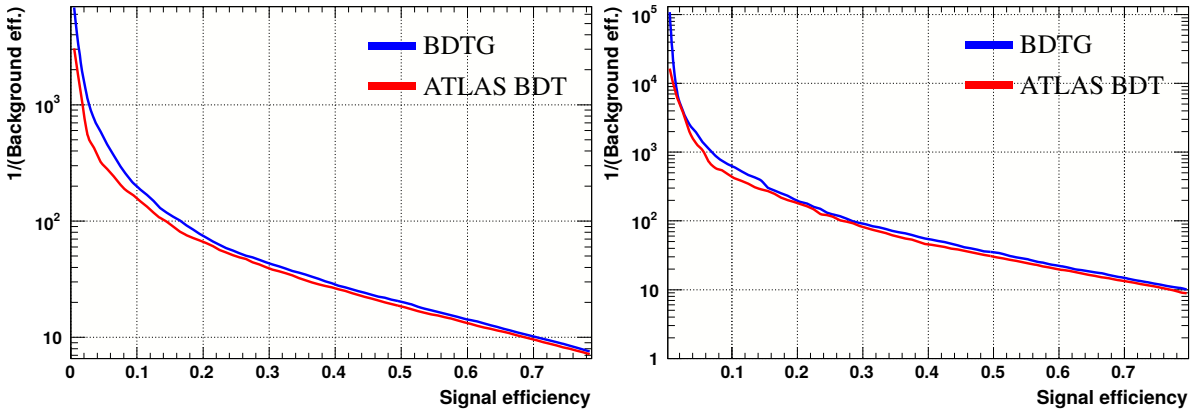
Gradient boosting with the less steeply rising loss in equation (V.10) was therefore chosen as the baseline boosting algorithm for the Lorentz-invariant MVA. Given this choice, other training parameters were optimised separately for the 1- and 2-lepton channel. The results are shown in table V.1. The parameters used for the ATLAS BDTs can be found in table IV.5 on page 92. Apart from using gradient boosting, the optimised trees use lower learning rates, bagging, and in the 1-lepton channel also deeper trees. The fact that the parameters were optimised separately for each channel particularly improves the training in the 2-lepton channel because the lower training statistics in this channel require more regularisation to build reliable models. In the 1-lepton channel, which has higher training statistics, increasing the tree depth while maintaining strong regularisation allows for more complicated models without increased overfitting. The background rejection in both channels increases due to these changes, and there is no reason to use identical model parameters for the two channels as was done for the ATLAS analysis. Apart from the reduced learning rate, regularisation for the BDTG training is achieved by bagging and by requiring 5 % of the full training sample in each BDT node. The regularisation for the BDTG training is therefore much stronger than for the ATLAS BDTs, but this is compensated by training more trees: 800 instead of 200.

Figure V.6 shows the background rejection for the gradient-boosted trees and for classifiers that were trained with Lorentz-invariant variables but adaptive boosting. In all relevant signal regions, 1 and 2 leptons, low and high p_{T}^{V}, the gradient-boosted trees reach higher significance than the ATLAS configurations. One should also note that the advantage of gradient boosting is only visible in the regions with low signal and low background efficiency, that is, in regions where a less aggressive loss function and better regularisation are expected to yield improvements.



(a) 1 lepton, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$

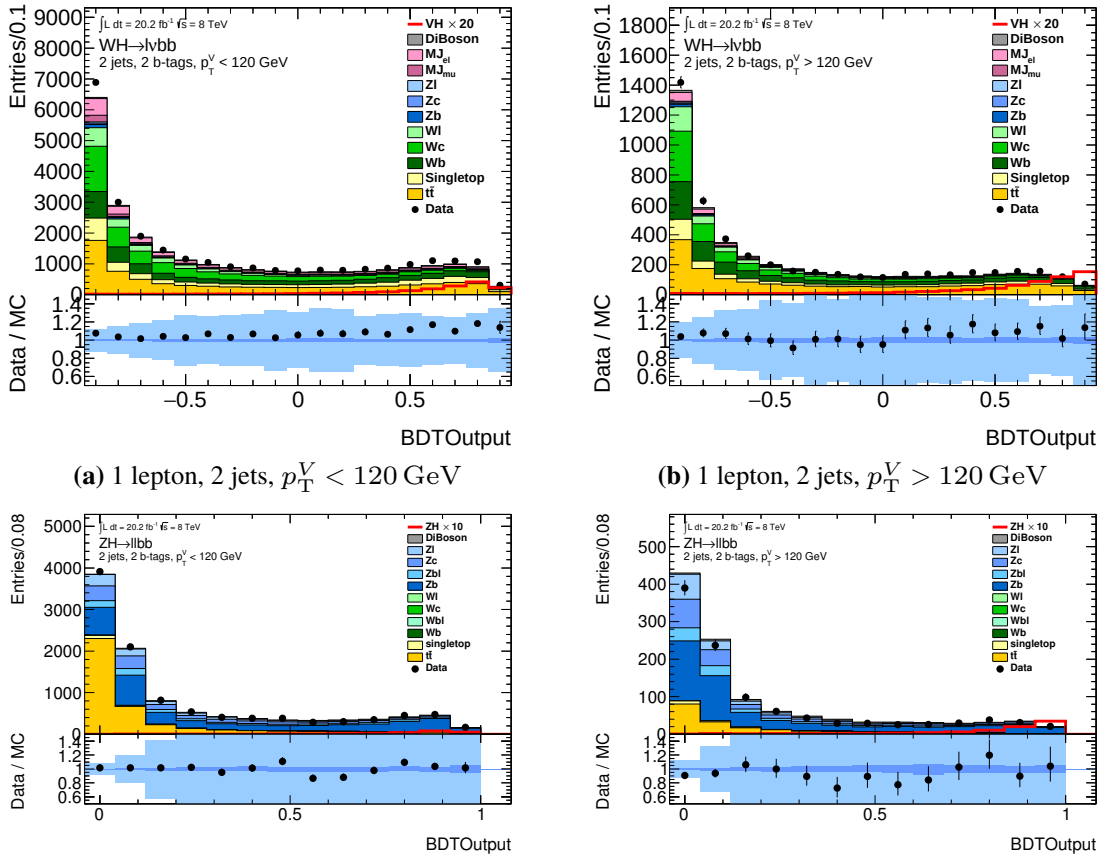
(b) 1 lepton, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$



(c) 2 leptons, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$

(d) 2 leptons, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$

Figure V.6: Background rejection versus signal efficiency of Lorentz-invariant BDT classifiers in events with two jets. The gradient-boosted BDTs with the loss function from equation (V.10) (BDTG) reach higher significance than BDTs trained with the setup used in the ATLAS $H \to b\bar{b}$ publication [1] in all four categories.



(a) 1 lepton, 2 jets, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$

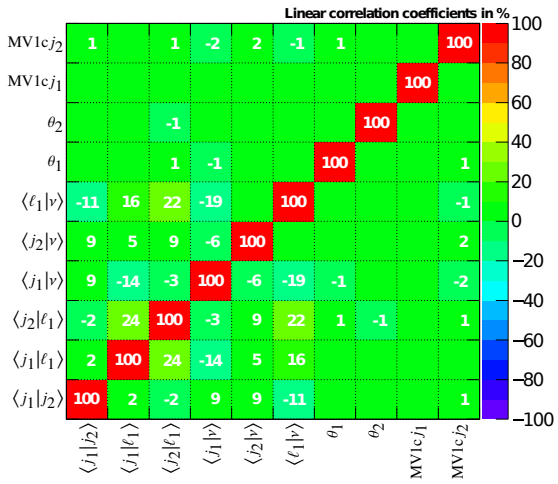
(b) 1 lepton, 2 jets, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$

(c) 2 leptons, 2 jets, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$

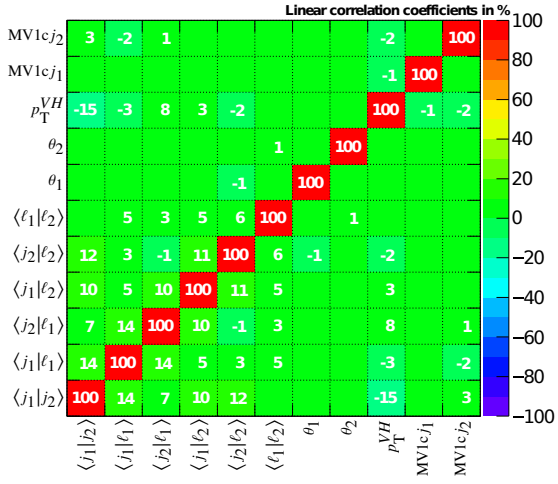
(d) 2 leptons, 2 jets, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$

Figure V.7: BDT classifiers with Lorentz-invariant variables and gradient boosting for 1- and 2-lepton events with two jets. The preselection outlined in section IV 1 and shape corrections from section IV 2 are applied. Data uncertainties are marked at the 68 % confidence level for Poisson distributions, and are therefore asymmetric. The dark blue bands (■) in the bottom plots show statistical uncertainties of the background simulations, and the light blue band (■) shows the quadratic sum of systematic uncertainties. Normalisation uncertainties are not shown because they are implemented in the likelihood fit. Normalisation corrections are only applied for the QCD multijet and $Z+$jets backgrounds as before. BDT scores in the 2-lepton channel are mapped into the interval $[0, 1]$. This does not change the performance of the classifiers.

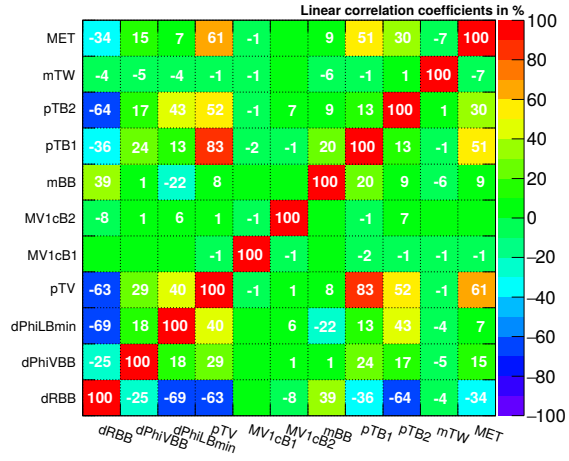
The response of the BDTG classifiers to data and simulations in the 2-jet region are shown in figure V.7. The less sensitive 3-jet region is shown in appendix C 1.2 on page 214. $H \to b\bar{b}$ simulations are superimposed in red and enlarged for better visibility. The figures show that the BDTs successfully separate signal and background events, and that the data distributions are well modelled by the Monte Carlo simulations. A detailed validation of systematic uncertainties, where correlations and constraints are also taken into account, will be discussed in section V 2.



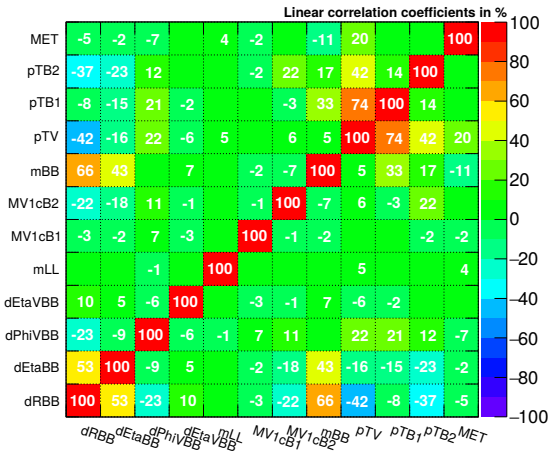
(a) 1 lepton, Lorentz-invariant variables



(b) 2 leptons, Lorentz-invariant variables



(c) 1 lepton, ATLAS variables



(d) 2 leptons, ATLAS variables

Figure V.8: Linear correlation coefficients of BDT input variables for **(top)** the Lorentz-invariant variables and (**bottom**) the ATLAS $H \to b\bar{b}$ variables. High correlation coefficients indicate redundant information, which requires higher model complexity when using BDTs. The sets with Lorentz-invariant variables avoid correlations, and therefore describe VH events with one variable less in both the 1- and 2-lepton channel.

Table V.2: Comparison of Lorentz-invariant variables and ATLAS $H \to b\bar{b}$ variables. Where applicable, similarities ($\approx$) or equivalences ($\Leftrightarrow$) between the sets are indicated. Redundancies in the ATLAS variables explain why the Lorentz-invariant set contains one variable less in both channels.
For the Lorentz-invariant variables, the only difference between the 1- and the 2-lepton channel is the inclusion of p_{T}^{VH} in the latter. By definition, all other variables in the Lorentz-invariant set represent the same information in the 1- and 2-lepton channels. The strategy to include the b-tagging score `MV1c` is identical for all configurations.

Lorentz inv.				ATLAS $H \to b\bar{b}$		
	1 L	2 L			1 L	2 L
$\langle\ j_1 \mid j_2\ \rangle$	•	•	$\Leftrightarrow$	m_{bb}	•	•
				$\lvert\Delta\eta(b,b)\rvert$		•
				$\Delta R(b,b)$	•	•
$\langle\ j_1 \mid l_1\ \rangle$	•	•	$\approx$	$\min \Delta\phi(l,b)$	•	
$\langle\ j_2 \mid l_1\ \rangle$	•	•				
$\langle\ j_1 \mid \nu/l_2 \rangle$	•	•				
$\langle\ j_2 \mid \nu/l_2 \rangle$	•	•				
$\langle\ l_1 \mid \nu/l_2 \rangle$	•	•	$\approx$	m_{T}^{W}	•	
				m_{ll}		•
θ_1	•	•				
θ_2	•	•	$\approx$	$\Delta\phi(V,bb)$	•	•
				$\lvert\Delta\eta(V,bb)\rvert$		•
MV1c(j_1)	•	•	$=$	MV1c(b_1)	•	•
MV1c(j_2)	•	•	$=$	MV1c(b_2)	•	•
p_T^{VH}		•	$\Leftrightarrow$	$E_{\mathrm{T}}^{\mathrm{miss}}$	•	•
				p_{T}^{V}	•	•
				p_{T}^{j1}	•	•
				p_{T}^{j2}	•	•
3-jet Events						
$\langle\ j_1 \mid j_3\ \rangle$	•	•		p_{T}^{j3}	•	•
$\langle\ j_2 \mid j_3\ \rangle$	•	•		m_{bbj}	•	•

Table V.3: Discovery significance of different multivariate classifiers for the combination of the 1- and 2-lepton channels. The significance is calculated from the profile-likelihood ratio using equation (IV.26), p. 110. The first row shows the expected significance without inclusion of systematic uncertainties; the second row includes these with all nuisance parameters set to zero. `BDTG` denotes classifiers trained with the set-up in table V.1, and `Ada` denotes the ATLAS $H \to b\bar{b}$ BDT set-up with adaptive boosting (table IV.5). The Lorentz-invariant classifier used in this work is shown in the leftmost column, and the classifier used for the ATLAS $H \to b\bar{b}$ publication is shown in the rightmost column.

	Lorentz		Lorentz + γ_z		ATLAS	
	BDTG	Ada	BDTG	Ada	BDTG	Ada
Z_0^{stat}	2.45	2.37	2.42	2.38	2.60	2.41
Z_0^{sys}	1.95	1.88	1.92	1.90	1.97	1.86
Difference	-20.6%	-20.6%	-20.3%	-20.1%	-24.3%	-22.7%

V 1.5 Comparison of Lorentz-Invariant and ATLAS VH Classifiers

At this point, the Lorentz-invariant $H \to b\bar{b}$ analysis can be compared to the ATLAS $H \to b\bar{b}$ MVA, that is, BDTG training with Lorentz-invariant variables is compared to adaptive boosting with a selection of standard kinematic observables. Table V.2 lists both sets of input variables, and also indicates similarities or equivalences among certain subsets. For three subsets of the ATLAS $H \to b\bar{b}$ set, there is just a single variable in the Lorentz-invariant set, whereas the opposite holds for only one subset in the Lorentz-invariant collection. This suggests that the ATLAS variables are (at least partly) redundant. An analysis of linear correlation coefficients shows that this is indeed the case, see figure V.8. The ATLAS set contains several variables with high correlation coefficients. By contrast, the Lorentz-invariant set has no redundant variables. This not a coincidence, it is by design: the derivation starts from 16 orthogonal and uncorrelated parameters.

The low correlations are a first reason to prefer the Lorentz-invariant variables over the ATLAS variables. Even though boosted decision trees deal well with correlated variables, correlations require higher model complexity. This was shown in figure IV.17 on page 99. If strong regularisation is used during the training because the training sample is not large enough for complex models, the extraction of information from correlated variables might be inhibited. Since the BDTG models are both more complex and use stronger regularisation than the ATLAS BDT model, removing correlations among input variables facilitates the extraction of information.

A second reason to prefer the Lorentz-invariant variables is that many ATLAS variables depend on the movement or boost of the VH system. All angular distances in ϕ and η, for example, vary with the boost of VH. The pseudorapidity differences are nearly Lorentz invariant, though, at least with respect to boosts along the z-axis. For the Lorentz-invariant variables, such dependencies are by definition minimal because the *inner*, Lorentz-invariant kinematics of the VH system are separated from the *outer* description.

The third reason to prefer the Lorentz-invariant classifiers is the impact of systematic uncertainties on the expected discovery significance. This significance is an important indicator for the sensitivity of a search. Discovery significances for different classifiers are shown in table V.3. Significances in the first row (Z_0^{stat}) only include the statistical uncertainties of the data. In the second row, systematic uncertainties are included as well. Several configurations are compared in the table because the Lorentz-invariant and the ATLAS analysis differ in two main points: the boosting algorithm and the input variables.

To measure the impact of these changes separately, both the Lorentz-invariant (left) and the ATLAS

variables (right) have been used with both training algorithms. The classifiers selected for this work are shown in the leftmost column. These use gradient boosting, the six Lorentz-invariant variables (eight in the 3-jet region), two angles, b-tagging information and the transverse boost in the 2-lepton channel, as shown in table V.2. Furthermore, the middle column shows a set of Lorentz-invariant variables identical to the leftmost sets except for including the longitudinal boost of the VH system. As discussed at the beginning of section V 1.1, this parameter is sensitive to the difference between gg and $q\bar{q}$ collisions.

The following conclusions can be drawn from table V.3:

- The Lorentz-invariant variables express all relevant kinematic information. Although 7 to 8 of the 16 parameters to describe the VH kinematics are removed from the set of variables, the statistical significances of the Lorentz-invariant classifiers are comparable to the significance of the ATLAS classifiers in the rightmost column.

- Including the longitudinal boost of the $bb\ell(\nu/\ell)$-system in the set of Lorentz-invariant training variables yields no improvement – although this variable separates $t\bar{t}$ from VH events. This shows that the other observables are sufficient to separate the two classes, and including the boost mostly adds noise to the training. The set "Lorentz $+\,\gamma_z$" was therefore not pursued any further.

- The systematic uncertainties relevant for the $VH \to Vb\bar{b}$ topology decrease the discovery significance by about $20\,\%$ for Lorentz-invariant classifiers and by 23 to $24\,\%$ for the ATLAS classifiers. This shows that the Lorentz-invariant variables are more robust with respect to systematic uncertainties. This is a decisive advantage in an analysis where statistical uncertainties are equally large as systematic uncertainties.

- For all configurations, gradient boosting is better than adaptive boosting. It was therefore chosen as the baseline training algorithm for all classifiers in this chapter. This improvement comes at the only cost of a longer training time because more trees have to be trained to compensate the lower learning rates and stronger regularisation.

The Lorentz-invariant classifiers selected for this work reach an expected discovery significance of $1.95\,\sigma$. This is $5\,\%$ higher than for the standard ATLAS classifiers. They are also the second best classifiers in the table. One might be tempted to draw the conclusion that the best classifier for the VH analysis is the classifier with ATLAS variables and BDTG training in the second-to-last column. The significance of this configuration is $1\,\%$ higher than for the Lorentz-invariant classifiers.

However, the impact of systematic uncertainties is also higher. This is, in fact, the largest degradation observed for all configurations. This is a reason not to choose this classifier because the expected significance as shown in table V.3 only measures the sensitivity of an analysis for an idealised case: it is assumed that the simulations perfectly model the data, and that all nuisance parameters remain at zero. The data are ignored entirely in the computation of the expected significance. It is clear that a high sensitivity to systematic uncertainties already on the Asimov dataset can lead to more severe degradations when the imperfect simulations are fitted to the data. The data distributions are, in fact, *expected* to differ from the Monte Carlo predictions – this is the main reason to add systematic uncertainties to the likelihood model. The sensitivities of the different classifiers therefore also have to be compared after fitting the likelihood model to the data. After fits to the data, the $VH \to Vb\bar{b}$ likelihood model will, for example, take into account the normalisation corrections that have to be applied to the Z+jets simulations. The ranking of the classifiers is therefore likely to change, especially if these are highly sensitive to systematic uncertainties. Although the BDTG training with ATLAS variables achieves the highest expected significance, its sensitivity will likely degrade most strongly due to systematic uncertainties.

Evaluating the performance of the classifiers after fits to data requires that the likelihood model is able to describe the data given the systematic uncertainties. That this is indeed the case was validated *before* deriving the final performance estimates of the classifiers. This discussion will be postponed to section V 2, though, because it is unrelated to the choice of boosting algorithms or input variables, which is the subject of the current section.

Comparison of the Classifiers After Fits to Data

To compare the classifiers, four likelihood models have been fit to data. Categories, preselection, simulations, data and systematic uncertainties for all models are identical[7]. The models only differ in the classification of events: two models use input variables with Lorentz-invariant observables, one with gradient and one with adaptive boosting, two models use the ATLAS $H \to b\bar{b}$ variables (again with BDTG and Ada boosting). For each model, the point of maximum likelihood is found, and the uncertainty of the signal strength μ is calculated using the profile-likelihood ratio. For details on the profile-likelihood ratio see section IV 5.3.

For each model, the total uncertainty was decomposed into different sources. This is achieved by fixing a set of nuisance parameters to their maximum-likelihood values and re-evaluating the profile likelihood. The resulting uncertainty of μ is lower than the total uncertainty because the corresponding nuisance parameters cannot be readjusted when the profile-likelihood ratio is evaluated. The quadratic difference of the total and the reduced uncertainty yields the impact of the fixed nuisance parameter(s). It should be understood that fixing a nuisance parameter is not identical to removing it from the likelihood model entirely. This is because pulling nuisance parameters can be necessary to describe the data. A good example is the $Z+$jets normalisation correction: fixing the $Z+$jets nuisance parameters to their maximum-likelihood values means to measure systematic uncertainties with a fixed, but optimal $Z+$jets normalisation. This quantifies the impact of the $Z+$jets normalisation uncertainties. Entirely removing the $Z+$jets normalisation nuisance parameters yields a likelihood model that is unable to describe the data, and can therefore not be used to conduct any measurement.

The decomposed uncertainties of the four classifiers are compared in figure V.9. The Lorentz+BDTG classifier selected for this work is the baseline, and the relative change of systematic uncertainties when an alternative classifier is used is shown as bars. The ATLAS Run-1 classifier is shown in red, and the two other classifiers demonstrate the effects of changing the training variables or changing the boosting algorithm. A more detailed decomposition of uncertainties can be found in appendix C 2.

On the very left, the total uncertainty is shown, which is decomposed into a statistical and a systematic component, the second and third set of bars. These show that the Lorentz-invariant classifier with gradient boosting outperforms all alternatives, especially the ATLAS Run-1 classifier. If the Lorentz-invariant classifier had been used for the Run-1 analysis, the total uncertainty in the channels with charged leptons would have reduced by $10\,\%$, and systematic uncertainties even by $16\,\%$. From the statistical uncertainties one can conclude that the gradient boosting configuration used in this thesis is superior to the ATLAS $H \to b\bar{b}$ configuration. With better BDT training, more information can be extracted from the data, and

[7] The response of all nuisance parameters was checked for all four likelihood models to prevent overconstraints or pulls of single nuisance parameters due to random fluctuations: if a nuisance parameter is found strongly constrained or pulled for only one model, this indicates a fluctuation. Such a fluctuation should not affect the comparison of the classifiers. An overconstraint in the 3-jet region with low p_{T}^{V} was, for example, observed for the jet energy resolution uncertainty, but only for one of the likelihood models. This region has no sensitivity to μ, but it can be used to (over)constrain the jet energy resolution with negative effects on other regions. To remove the impact of this overconstraint, this nuisance parameter was decoupled from the other regions in all four models. For three of the models, this change has a negligible effect, but for the model with the fluctuation, the overconstraint is removed. This ensures that the comparison of the classifiers is on equal footing.

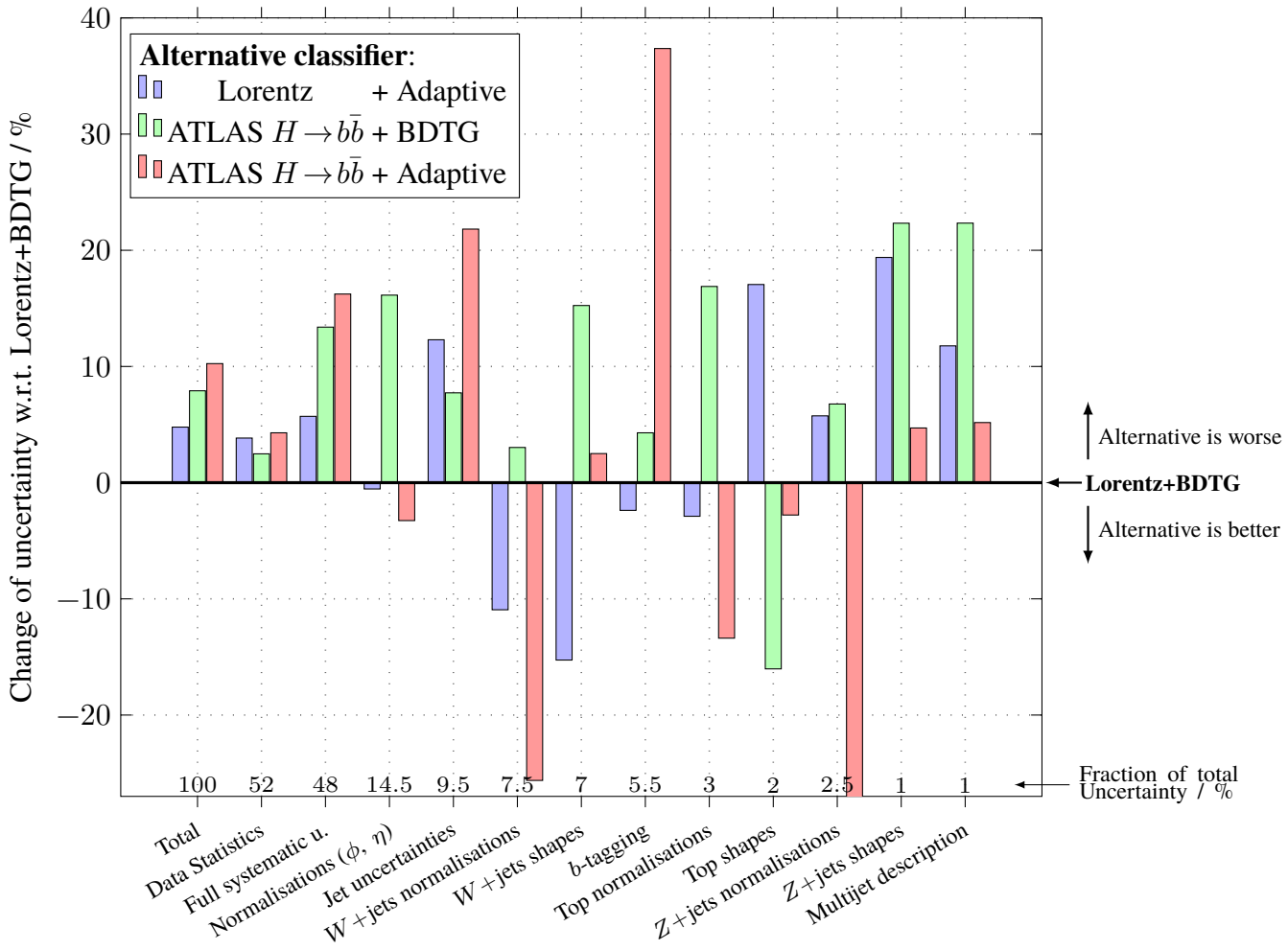


Figure V.9: Comparison of uncertainties affecting the measurement of the signal strength μ for different BDT classifiers. The Lorentz-invariant classifier for this thesis (baseline) is compared to three alternative configurations for different classes of systematic uncertainties listed in order of importance. Their relative impact is shown at the bottom. The bars show how much uncertainties change when the Lorentz+BDTG classifier is replaced with one of the three alternatives. The classifier developed for the ATLAS $H \to b\bar{b}$ publication [1] is shown in red (■). A table with absolute uncertainties and further sources of uncertainties can be found in appendix C 2.

therefore the impact of limited data statistics is less severe. The figure shows that *both* BDTG classifiers, regardless of the input variables, have lower uncertainties than classifiers with adaptive boosting.

However, the separation power of the BDTs is less important than their sensitivity to systematic uncertainties: whereas statistical uncertainties change on the level of a few percent, systematic uncertainties differ by up to 16 %. Since statistical and systematic uncertainties contribute equally to the total uncertainty, the third set of bars in figure V.9 is of highest interest for the choice of the best classifier. It shows that both classifiers with Lorentz-invariant variables are superior to the classifiers with ATLAS $H \to b\bar{b}$ variables (■,■), and among the Lorentz-invariant classifiers the gradient-boosted classifier is the best.

This result may seem to contradict the results obtained using Asimov datasets, that is, from the expected significance in table V.3. In this test, the ATLAS $H \to b\bar{b}$ classifier was only 5 % worse than

Lorentz+BDTG whereas now it is $10\,\%$ (total) and $16\,\%$ (systematic). This difference arises because two different tests are carried out: the Asimov datasets are idealised data, whereas in figure V.9, the performance for likelihood models that actually have to adapt to the observed data is tested. Hence, the latter is the more relevant result. Furthermore, the fits to data in figure V.9 confirm the observation from Asimov fits that the two classifiers with ATLAS variables have the highest sensitivity to systematic uncertainties. The effect is just more pronounced in fits to data.

To understand why the Lorentz-invariant classifiers with gradient boosting are the best choice, figure V.9 shows a more fine-grained decomposition of the systematic uncertainties. The fourth to 11^{th} entries in the figure show that classifiers with Lorentz-invariant variables are particularly robust with respect to shape uncertainties: 'Jet uncertainties', '$W+$jets shapes', 'b-tagging', '$Z+$jets shapes' and 'Multijet description' all are classes of shape uncertainties where Lorentz-invariant classifiers are better than classifiers with ATLAS variables. In most cases, this is true for both gradient (baseline) and adaptive boosting (■), but gradient boosting usually performs best. The only exception where classifiers with ATLAS $H \to b\bar{b}$ variables (■,■) are clearly the better choice with respect to shape uncertainties are the top shape uncertainties. These, however, only amount to $2\,\%$ of the total uncertainty.

The BDTs with ATLAS $H \to b\bar{b}$ variables, particularly with adaptive boosting that uses the more aggressive loss function with a high sensitivity to outliers (■), are the best choice if one is interested in constraining normalisation uncertainties. This is visible for the class 'Normalisations (ϕ, η)', which contains all normalisation uncertainties for all samples, but also when the normalisation uncertainties (ϕ, η) are separated into $W+$jets, $t\bar{t}$ and $Z+$jets. This indicates that the profile-likelihood fits use differences in the response of the ATLAS classifiers to separate $W+$jets from $t\bar{t}$ and $Z+$jets from $t\bar{t}$. The ATLAS variables make explicit use of jet transverse momenta and angles between reconstructed objects, and angular variables strongly depend on the boost of the VH system. Since these distributions differ for $t\bar{t}$, $W+$jets and $Z+$jets, the classifier outputs also differ. Measuring background normalisations, however, is not the primary goal of using such classifiers: it is to separate the signal from backgrounds. The disadvantage of reduced normalisation uncertainties is a higher sensitivity to shape uncertainties because the measurement of background normalisations is only possible when the BDTs are sensitive to shape differences between the backgrounds.

The Lorentz-invariant variables avoid explicit dependencies on variables such as transverse momenta and Lorentz-*variant* angles, which have visibly different distributions for different backgrounds. Furthermore, they avoid correlated input variables. When correlations among the input variables are low, detector uncertainties affect smaller subsets of input variables. This leads to smaller changes of the classifier response due to such uncertainties. Additionally, the transformation used to compress the dynamic range of the Lorentz invariants into the interval $[0, 1]$ suppresses shape changes that affect only background-dominated regions[8]. Lastly, the less aggressive boosting for the BDTG classifier makes it less sensitive to particularities of certain backgrounds. Instead of learning features (outliers) that make certain backgrounds unique, more general shape differences between the signal and *all* backgrounds are learned, and used to separate only these two classes. Therefore, the BDT response of Lorentz+BDTG is similar for all background processes. This means that background normalisations cannot be measured with the same accuracy as for the ATLAS classifiers, but shape changes also have a lower impact.

One can, in fact, design a Lorentz-invariant classifier that is more sensitive to the difference between $W+$jets and $t\bar{t}$ by including the boost along the z-axis. This is because $t\bar{t}$ events are produced mostly in gluon collisions and $W+$jets events in quark-antiquark collisions (see *e.g.* section III 1.1). Although normalisation uncertainties reduce when using such a classifier, it was shown in table V.3 for the classifiers "Lorentz + γ_z" that this does not improve the sensitivity of the search. The reason is that by including the

[8] For more on the transformation see page 131.

boost, the classifiers also become more sensitive to shape uncertainties.

Given that normalisations can be measured externally, for example, in control regions, and given that many of the dominant uncertainties are shape uncertainties such as the jet uncertainties, the $W+$jets description and the b-tagging uncertainties, the Lorentz-invariant classifiers are the best choice for the $VH \to Vb\bar{b}$ analysis. Moreover, background normalisation uncertainties can also be reduced by obtaining higher data statistics or by obtaining better Monte Carlo predictions. This means that there are three independent ways of mitigating the only disadvantage of the Lorentz-invariant classifiers. The Run-2 version of the ATLAS $VH \to Vb\bar{b}$ analysis [44] indeed uses a larger data sample and `NLO` $V+$jets simulations, and introduced a control region to better measure $W+$jets normalisations. It could further benefit from a fourth layer of pixel detectors, IBL, which is closer to the interaction region, and therefore improves the b-tagging performance by almost a factor of two. The classifiers, however, are nearly identical to the ATLAS Run-1 classifiers from chapter IV apart from adding a few more discriminating variables. As one can expect from the results obtained in this chapter, the Run-2 analysis is now limited by systematic uncertainties. It reached an expected significance of three standard deviations. Lorentz-invariant classifiers therefore are an interesting option for future updates of the Run-2 analysis.

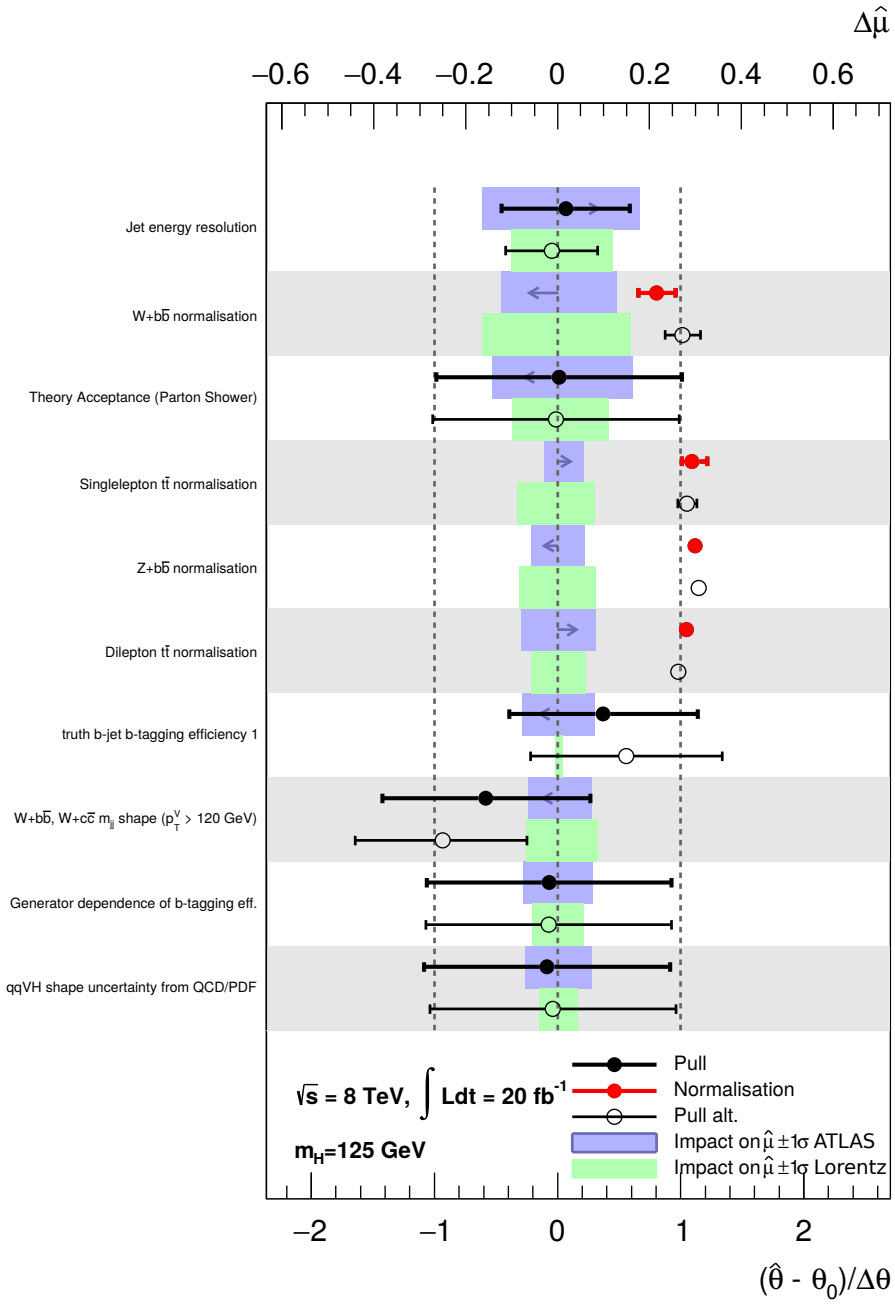


Figure V.10: Dominant systematic uncertainties for the measurement of the signal strength $\hat{\mu}$. Each bar represents the impact on the observed value of μ if the corresponding systematic uncertainty is varied by $\pm 1\sigma$ (**top axis**). Arrows indicate the $+1\sigma$ variation. (**Upper bars**, blue) ATLAS BDTs, (**Lower bars**, green) Lorentz-invariant BDTs.
Measured nuisance-parameter pulls are overlaid with black markers (**bottom axis**). The $\pm 1\sigma$ thresholds are indicated by dashed lines. Red nuisance parameters are unconstrained sample normalisation factors like the ϕ factors in equation (IV.12c), page 104. These should be distributed around one whereas other nuisance parameters are constrained to zero using Gaussian constraint terms.

Ranking of Nuisance Parameters

Figure V.10 shows the ten nuisance parameters with the highest impact on $\hat{\mu}$ (from both the Lorentz-invariant and the ATLAS search). The coloured bars show the impact on $\hat{\mu}$ when pulling a nuisance parameter by $\pm 1\,\sigma$ (top axis), and the black points show the maximum-likelihood values after the fit to data as pulls (bottom axis). Arrows in the blue boxes indicate the impact of the $+1\sigma$ variation. If the jet energy resolution, for example, would be pulled to $\alpha_{\text{JER}} = +1\sigma$, that is, the pull marker is found at the dashed line at $+1$ (bottom axis), $\hat{\mu}$ would increase by ~ 0.2 (top axis).

The red nuisance parameters are normalisation factors. They should be distributed around one, and they are not constrained. A detailed discussion of the pulls will be deferred to section V 2, where the likelihood model is validated, because for the current comparison the impact on μ (coloured bars) is of higher interest.

Comparing the impact on $\hat{\mu}$ confirms the findings from figure V.9: the Lorentz-invariant classifiers are significantly less sensitive to the dominant shape uncertainties such as the jet energy resolution and b-tagging-related or m_{bb}-related shape uncertainties. The only uncertainties that have higher impact on the Lorentz-invariant classifiers compared to the ATLAS classifiers are the background normalisations of the three dominant backgrounds $W{+}bb$, $t\bar{t}$ and $Z{+}bb$.

The signal acceptance uncertainty deserves special attention because it cannot be constrained in fits to data. It encodes the impact of the parton-shower model on the signal predictions, and is implemented as a normalisation uncertainty for the signal samples. Higher acceptance leads to lower $\hat{\mu}$ and *vice versa*. This is explained by the definition of μ:

$$N = \mu\, S + B.$$

The number of observed events N is given by the data, and B as well as N are constant after the likelihood maximisation. μ is a free parameter, and will therefore vary oppositely to S, which varies in dependence of the parton-shower model. Since the data only constrain the sum $\mu S + B$, the acceptance uncertainty cannot be measured in data. These properties are correctly reflected in the likelihood model: the $+1\sigma$-variation of the parton shower uncertainty leads to a reduction of $\hat{\mu}$ and *vice versa*. Further, the pulls of the nuisance parameter are zero, and there are no constraints from the fit to data. The signal acceptance uncertainties are identical for both the Lorentz-invariant and ATLAS classifiers. Nonetheless, the impact on $\hat{\mu}$ is lower for the Lorentz-invariant classifiers. This is because the Lorentz-invariant classifiers have higher sensitivity in the high-p_{T}^{V} region, where the parton-shower acceptance uncertainty is minimal. The total impact of this uncertainty is therefore lower for the Lorentz-invariant MVA.

This concludes the discussion of the Lorentz-invariant classifiers. It was shown that the Lorentz-invariant classifiers are superior to the ATLAS classifiers used for the $VH \to Vb\bar{b}$ Run-1 result [1] because they would have reduced the total uncertainty of the signal strength $\hat{\mu}$ by $10\,\%$ in the channels with charged leptons. Systematic uncertainties would have reduced by $16\,\%$. Lorentz-invariant classifiers are also an interesting option for future updates of the $VH \to Vb\bar{b}$ search because these will be limited by systematic uncertainties where the Lorentz-invariant classifiers perform best.

Before measuring the $VH \to Vb\bar{b}$ signal strength in ATLAS data of 2012, two flaws of the ATLAS likelihood have to be addressed. These became evident when the model was adapted to the Lorentz-invariant classifiers. The necessary modifications of the likelihood model change some uncertainties of μ, which is why all results in this thesis, also from the current section, were derived with corrected likelihood models. Nevertheless, it made sense to postpone the discussion of these changes because the choice of classifiers is largely independent of the modifications to the likelihood model.

V 2 Improvement of the Likelihood Model

When using the profile-likelihood approach, it must be ensured that the likelihood model is able to describe the data given the set of systematic uncertainties. Otherwise, any measurement of observables is meaningless. Since all simulations used to derive the likelihood functions rely on approximations, the model must possess sufficient degrees of freedom, that is, systematic uncertainties, to account for differences between the predictions and the data. The validation of the likelihood model in this section ensures that this is the case. The focus will be on two aspects: firstly, to ensure that systematic uncertainties are parametrised properly, and secondly, to ensure that the likelihood model can use these to describe the data.

V 2.1 Smooth Parametrisations of Systematic Uncertainties

All ATLAS analyses share a set of systematic uncertainties, which can be assigned to two classes: uncertainties of efficiencies or probability distributions (*e.g.* reconstruction efficiency) and uncertainties of event observables (*e.g.* momentum of a particle). The former can be estimated by changing the weights assigned to Monte Carlo events, the latter are estimated by changing observables in each event. The changes of event weights are usually small, which means that the distributions of observables change smoothly because the same events enter the distributions with only slightly altered weights.

However, when event observables are changed, events can migrate between different categories or pass/fail selection cuts, which they do not pass/fail in the nominal analysis. The goal is to parametrise the resulting shape changes of the probability distributions, but these might be overshadowed by statistical uncertainties. Statistical fluctuations therefore should be removed when estimating shape changes, or at least mitigated. To achieve this, the normalised shape difference of the samples with and without systematic uncertainty is calculated bin by bin:

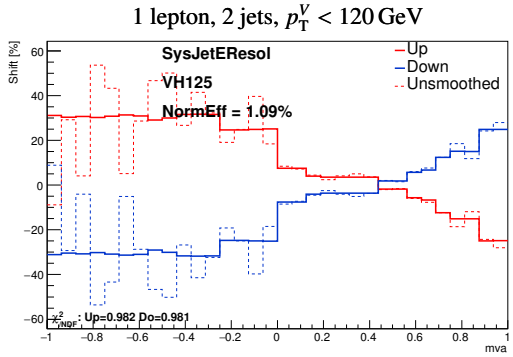
$$\Delta_{\text{rel}} = \frac{x_{\text{Sys}} - x_{\text{Nominal}}}{x_{\text{Nominal}}}. \tag{V.11}$$

x is an observable such as the BDT classifier score. To separate the impact of the systematic uncertainty from statistical fluctuations, the shape difference is subsequently smoothed. After smoothing, the distribution representing the systematic uncertainty is reconstructed from the nominal distribution using the smoothed relative difference:

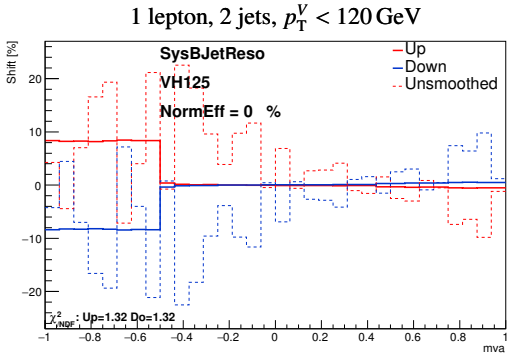
$$x_{\text{Sys}}^{\text{smooth}} = x_{\text{Nominal}} \cdot (1 + \Delta_{\text{rel}}^{\text{smooth}}).$$

Aggressive smoothing methods reduce statistical fluctuations, but the effect of a systematic uncertainty might be represented wrongly when distributions are “oversmoothed”. Weak smoothing might not suppress statistical fluctuations sufficiently. The smoothing algorithm therefore decides how accurately systematic uncertainties are described by the likelihood model.

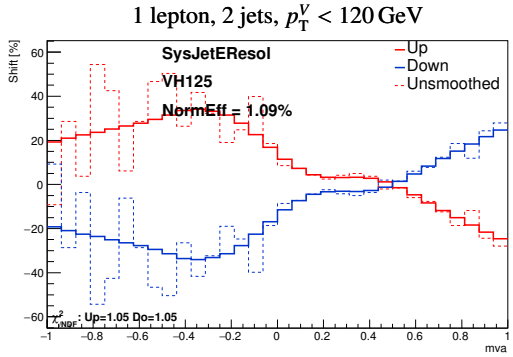
The performance of the ATLAS $H \to b\bar{b}$ smoothing algorithm was validated, especially for the dominant systematic uncertainties of the Lorentz-invariant $VH \to Vb\bar{b}$ analysis. It was found that the ATLAS smoothing algorithm can only correctly approximate the impact of systematic uncertainties whose distributions can be represented with monotonous step functions. Other distributions are oversmoothed. This is visible in figure V.11: the systematic uncertainty in figure V.11(a) is represented well, but the distribution in figure V.11(b) is oversmoothed. This is because the ATLAS $VH \to Vb\bar{b}$ smoothing was designed assuming that the normalised difference is always monotonous – for all systematic uncertainties, processes, classifiers and categories. This assumption fails often for the Lorentz-invariant classifiers, but also for some official ATLAS histograms used for the Run-1 $H \to b\bar{b}$ publication [1]. In figure V.11(a),



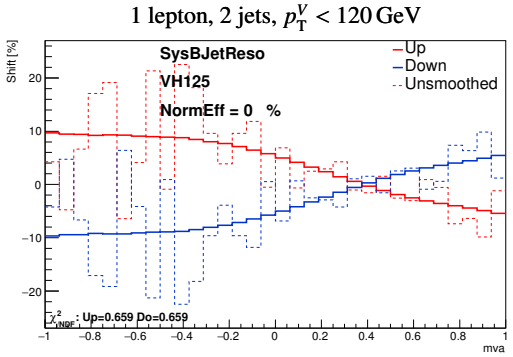
(a) Energy resolution uncertainty for all jets with ATLAS smoothing



(b) Energy resolution uncertainty for b-jets with ATLAS smoothing



(c) Energy resolution uncertainty for all jets with polynomial smoothing



(d) Energy resolution uncertainty for b-jets with polynomial smoothing

Figure V.11: Impact of systematic uncertainties before and after smoothing, illustrated using the BDT response to WH events. The figures show the difference of a histogram with systematic uncertainties applied with respect to the nominal distribution. The dashed lines show the impact of the systematic uncertainty without any smoothing. **Top**: ATLAS smoothing. **Bottom**: Polynomial smoothing.

for example, the jet energy resolution uncertainty for WH events vanishes completely.

Popular smoothing algorithms to replace the ATLAS $H \to b\bar{b}$ smoothing like calculating the mean or median of three adjacent bins or the more sophisticated `353QH` [167] commonly used in `ROOT` did not produce superior results because these do not take into account the statistical uncertainties of the bins. Yet these uncertainties are very relevant because bin contents of similar magnitude are subtracted from each other when calculating the normalised difference, see equation (V.11). This leads to significant uncertainties in numerous bins, which needs to be taken into account when smoothing is applied.

A method to take such uncertainties into account while at the same time lifting the false assumption of monotonous distributions is to fit functions to the normalised difference. The monotonous step functions used by ATLAS were therefore replaced with polynomials fitted to the normalised difference. Results are shown in figures V.11(c) and V.11(d): both shapes are reproduced accurately and without statistical fluctuations. Nonetheless, it must be ensured that the polynomials do not overfit the data: with sufficient degrees of freedom, a polynomial can follow any distribution – in the worst case it follows random noise. Therefore, the `F-test` [168] is employed to choose the lowest-order polynomial that reliably represents the impact of a systematic uncertainty. F is the ratio of two independent, χ^2-distributed variables U_i,

normalised by the number of degrees of freedom:

$$F_{a,b} = \left(\frac{U_a}{\text{ndf}_a}\right) / \left(\frac{U_b}{\text{ndf}_b}\right)$$

F can hence be used to compare the fit quality of two models. Be a a simple model with p_a parameters, and b a more complex model with $p_b > p_a$ parameters. Model a should further be a special case of b; in the sense that b can be made identical to a if a set of parameters in b is chosen correctly. For polynomials, this is the case if b is of higher order than a. If both models are fit to data, and if model a captures the underlying distributions equally well as b, any additional degrees of freedom in b should not improve the quality of the fit. The squared sum of residuals of a will be larger than for b, but a difference in χ^2-values, U, only occurs because of statistical fluctuations, and therefore it will itself be χ^2-distributed:

$$\chi^2_a = \chi^2_b + U, \quad \text{where } U \sim \chi^2_{\Delta\text{ndf}}.$$

'Δndf' is the difference in degrees of freedom between the two models. In such a case, the F-statistic is:

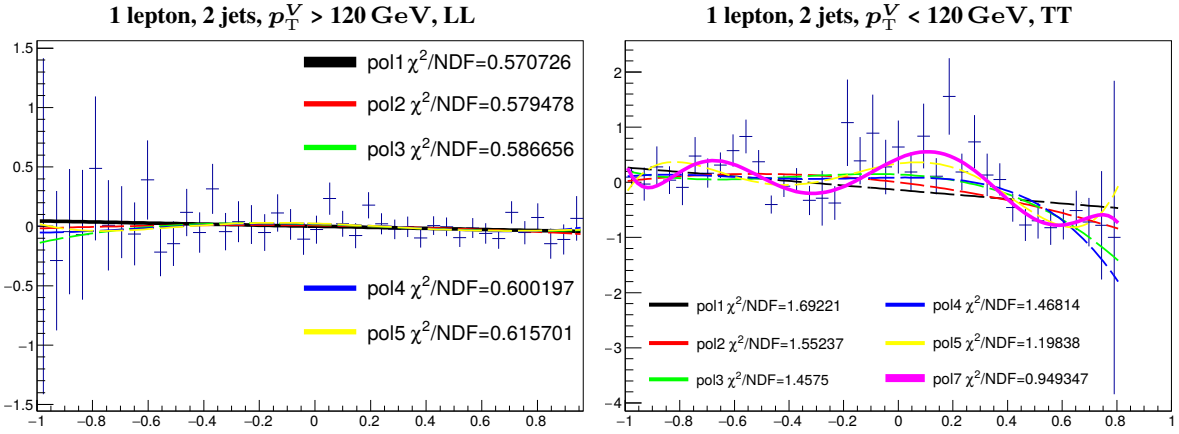
$$F_{a,b} = \frac{\frac{\chi^2_a - \chi^2_b}{\Delta\text{ndf}}}{\frac{\chi^2_b}{\text{ndf}_b}} = \frac{\frac{U}{\Delta\text{ndf}}}{\frac{\chi^2_b}{\text{ndf}_b}}.$$

Knowing the distribution of F, one can test if the more complex model b is better suited to describe the data. The corresponding hypothesis test is:

- $\boldsymbol{H}_0$ $F_{a,b}$ is F-distributed. Both models describe the data equally well. Choosing the more complex model is not necessary.
- $\boldsymbol{H}_1$ $F_{a,b}$ is not F-distributed. Model b provides a superior description of the data. The more complex model should be chosen.

The probability of H_0 can be computed by evaluating the cumulative F-distribution. The critical value for rejecting the null hypothesis, that is, accepting the more complicated polynomial, was set to $p(H_0) < 30\,\%$. There is no guideline for choosing the critical value, but setting it to 20 and $40\,\%$ did not yield significantly different results.

Using the F-test to choose the best polynomial is a way to regularise the models used to represent the shape changes caused by systematic uncertainties. Unless statistical evidence is found that a complex model provides a superior description of the shape, it is assumed that the shape change can be represented using a simple polynomial such as a straight line. The baseline model therefore is a 1^{st}-order polynomial, and polynomials up to 10^{th} order are tested. If H_0 for the straight line is rejected in favour of a more complex polynomial, the process is repeated with this polynomial as the baseline model until no further significant F-values are found, or until a polynomial of 10^{th} order is reached. Since the histograms for smoothing systematic uncertainties were configured to have 50 bins, this ensures that the fits to extract the coefficients of the polynomials are never underdetermined. For the $VH \to Vb\bar{b}$ likelihood models, the polynomials are usually of low orders up to 3^{rd}. More complex polynomials up to 6^{th} order are chosen rarely, and higher orders were chosen only in single instances. Figure V.12 shows two examples: a systematic uncertainty with negligible impact in (a) and one of the rare uncertainties that requires a 7^{th}-order polynomial in (b). The first five polynomials that were tested in each instance are shown as dashed lines, and the best-fitting polynomial is shown with a bold solid line. The figures demonstrate that the complexity of the model is successfully adapted to the shape changes caused by the systematic uncertainties, and that statistical fluctuations are efficiently reduced.



(a) Jet energy resolution uncertainty for WH events.

(b) Jet energy pileup-correction uncertainty for the muon multijet estimate.

Figure V.12: Smoothing of systematic uncertainties with polynomials. The figures show the relative difference as in equation (V.11), which describes the impact of a systematic uncertainty on the classification result. The F-test is used to find the polynomial that best approximates the distributions. **(a)** Noisy systematic uncertainty with insignificant shape changes. A straight line is the best approximation. **(b)** Systematic uncertainty with complex shape changes. A 7^{th}-order polynomial is found to be the best approximation.

χ^2 tests were used to rate the performance of polynomial smoothing and ATLAS $VH \to Vb\bar{b}$ smoothing. The statistical compatibility between the smoothed and the raw shape was tested for both smoothing algorithms in all categories, for all uncertainties and for all signal/background samples. Polynomial smoothing outperforms the ATLAS method in all categories, although the ATLAS method *also* minimises a χ^2-metric: a χ^2-test is used to determine whether the step function should remain constant or jump to a new value. However, polynomial smoothing uses superior baseline models: by lifting the assumption of monotonously rising or falling distributions, shapes can be represented more accurately, and with fewer parameters.

Apart from the F-test, a second method of regularisation is employed for polynomial smoothing: after selecting the best polynomial, the parameters of the polynomial are tested for being compatible with zero. If all parameters are compatible with zero within the $1\,\sigma$ uncertainty, the systematic uncertainty is removed from the respective category. This reduces the number of noisy uncertainties being propagated into the likelihood model, but it also ensures that all significant shape changes are represented correctly. This does not mean that a systematic uncertainty is removed completely, though: smoothing and pruning of systematic uncertainties are applied separately per category, per background/signal sample and per source of systematic uncertainty. Although shape changes might be insignificant in single categories, they might be dominant in others. The method introduced here fully takes this into account.

Polynomial smoothing is used for all likelihood models and all comparisons in this thesis. Its impact on the uncertainties of μ was compared with the ATLAS smoothing method: it was found that depending on the category of the analysis and the classifiers used, uncertainties of μ increase or decrease by 0.05 to 0.1. Given that the total uncertainty is about 0.6, this impact is not negligible, but it is limited because the most important Monte Carlo samples – VH, $t\bar{t}$ and partly also V+jets – provide rather high statistics, and therefore are less affected by the details of the smoothing algorithm. Nevertheless, the ATLAS smoothing method occasionally fails also for high-statistics simulations such as WH shown in figure V.11(b).

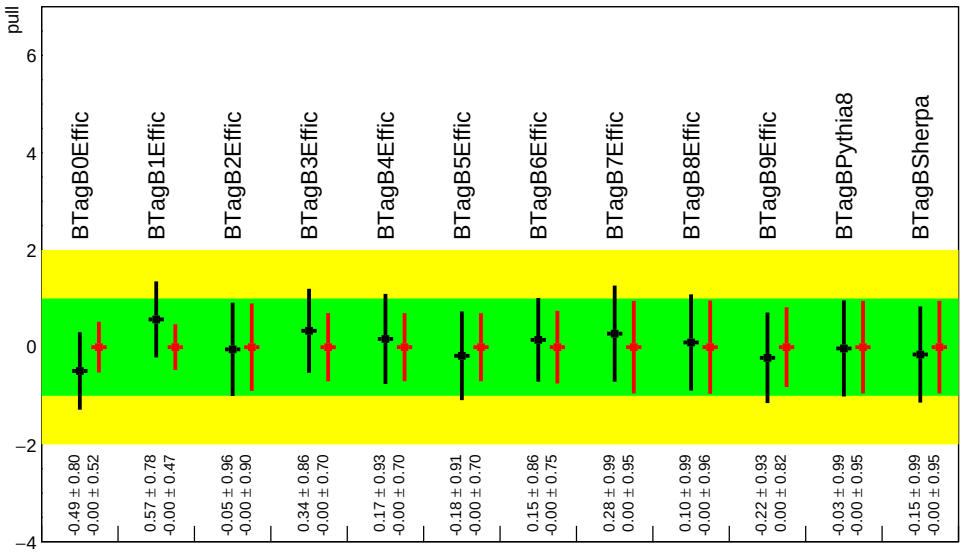


Figure V.13: Nuisance parameters of b-tagging uncertainties. These are shown as pulls, that is, the post-fit values of the nuisance parameters divided by the pre-fit uncertainties. The green and yellow bands show the one- and two-standard-deviation uncertainties of the b-tagging calibration. The error bars show the post-fit one-standard-deviation uncertainty. (■) Fit to data (■) Fit to a dataset generated from simulations ("Asimov dataset").

V 2.2 Analysis of Nuisance-Parameter Pulls

The analysis of nuisance-parameter pulls is even more important than the smoothing algorithm: the nuisance parameters provide the degrees of freedom to fit a likelihood model to the data taking into account that the simulations usually cannot perfectly reproduce the distributions in data. To judge if the likelihood model is able to describe the data given the systematic uncertainties, the response of all nuisance parameters must be checked because $-2\ln(\lambda)$ should be χ^2-distributed when releasing a nuisance parameter. This is usually done by analysing pulls, that is, the post-fit values of the nuisance parameters divided by the input uncertainties. As described in section IV 5.3, the post-fit values of the nuisance parameters are the values that maximise the likelihood, and the post-fit uncertainties are obtained by evaluating the matrix of second derivatives of the profile-likelihood ratio.

Figure V.13 shows an example of pulls that can be regarded as well-behaved, in this case b-tagging uncertainties. The full set of nuisance-parameter pulls can be found in appendix C 4.3. In the present example, two likelihood fits are compared in the same panel: a fit to data (left, ■) and a fit to the Asimov dataset (right, ■). For the interpretation of the pulls, the following points need to be considered:

- Pulls should be distributed around zero, and uncertainties should normally be close to one.
- If the fit can infer the value of a nuisance parameter from the data, uncertainties are lower than one. This is because the uncertainties result from two terms in the likelihood function: the constraint term, a Gaussian term centred at zero with a standard deviation of one, and the impact on the bin contents constrained by the data. If an uncertainty has measurable impact on the bin contents, and the data have low-enough statistical uncertainties in these bins, the latter term becomes relevant, and the uncertainty of the nuisance parameter shrinks.
- Whether a fit model is sensitive to certain nuisance parameters, can be determined using the fit to the Asimov data. If nuisance-parameter uncertainties reduce to less than one in the Asimov fit, the likelihood model uses the (Asimov) data to measure the nuisance parameter. It can therefore be expected that such a measurement is also possible with real data. In the example of figure V.13,

the first two nuisance parameters, which have the strongest impact on b-tagging efficiencies, can be inferred from data because Asimov uncertainties are about half of the input uncertainties. Other b-tagging nuisance parameters can be inferred only with much lower sensitivity because they also have a lower impact on the likelihood. The Asimov fit, however, only tests the sensitivity for an idealised data sample that exactly follows the nominal prediction. Asimov uncertainties are therefore usually lower than in fits to data (as in the present example).

- Post-fit uncertainties in fits to real data that are considerably larger than $1\,\sigma$ indicate ill-defined uncertainties: the upper bound on the post-fit uncertainty is given by the constraint term in the likelihood, which yields an uncertainty of exactly $1\,\sigma$. If the post-fit uncertainty is larger, this is a sign of a numerical instability in the minimum of the log-likelihood that leads to an inaccurate estimation of the Hesse matrix (*e.g.* because the log-likelihood profile is asymmetric). Such uncertainties need to be reparametrised, unless the excess over $1\,\sigma$ is small, and the corresponding systematic uncertainty is negligible for the measurement of μ. Only a few of the 177 post-fit uncertainties documented in appendix C 4.3 are slightly larger than one, but all of these are irrelevant for the measurement of μ.
- Nuisance parameters can be pulled or constrained more than expected from the Asimov fit for different reasons:
  1. Poorly parametrised systematic uncertainty.
  2. Statistical fluctuation: both the auxiliary measurements and the values inferred from the data may occasionally fluctuate. This can lead to significant pulls or constraints.
  3. No prior knowledge: some uncertainties need to be estimated without external measurements. Especially when comparing Monte Carlo generators, it is unclear which Monte Carlo generator models the data best, for example, with respect to the flavour fractions of b, c, l jets in V+jets events. Pulls must therefore be expected.
  4. Missing flexibility of the likelihood model: if an uncertainty relevant for a measurement is not included in the likelihood model, other nuisance parameters with a similar effect on the bin contents might be pulled unjustifiably to cover the differences between data and simulations.

Depending on which of these four effects causes significant pulls, different actions are advisable: case 1 needs to be excluded by analysing the histograms that parametrise the systematic uncertainty. The result of such checks is the correction of the smoothing algorithm described in section V 2.1. Pulls caused by case 2 can be detected by determining the value of the nuisance parameter separately in different categories. If only one category shows a large pull, the pull is likely a fluctuation. The category that is causing the pull should be decoupled from the other categories, which eliminates the impact of the fluctuation on other categories. Parameters falling into class 3 are known in advance. Pulls can be accepted, unless they are very high ($\gg 1\,\sigma$). Case 4 remains if none of the other explanations are plausible. This case should be avoided because measurements conducted with such a likelihood model are likely unreliable. If case 4 is the only plausible explanation for a pull, the likelihood model needs to be extended in order to provide the missing degrees of freedom to describe the data.

Given figure V.13, it is justified to state that b-tagging efficiencies in the Lorentz-invariant VH analysis are modelled well. The likelihood model is only weakly sensitive to most of the nuisance parameters, and therefore the calibration uncertainties provided by ATLAS are directly translated into uncertainties of μ. Only for the two dominant b-tagging nuisance parameters uncertainties reduce to $0.8\,\sigma$. The nuisance parameters are also slightly pulled, but well within the $1\,\sigma$ interval of the ATLAS measurements. Both the constraints and the pulls are justified because the $VH \to Vb\bar{b}$ analysis relies heavily on b-tagging, but it probes only a specific region of phase space (two Higgs b-jets recoiling against a vector boson), whereas ATLAS calibrations are provided for a much larger phase space. It is therefore not surprising that pulls and post-fit uncertainties slightly deviate from the ATLAS calibration.

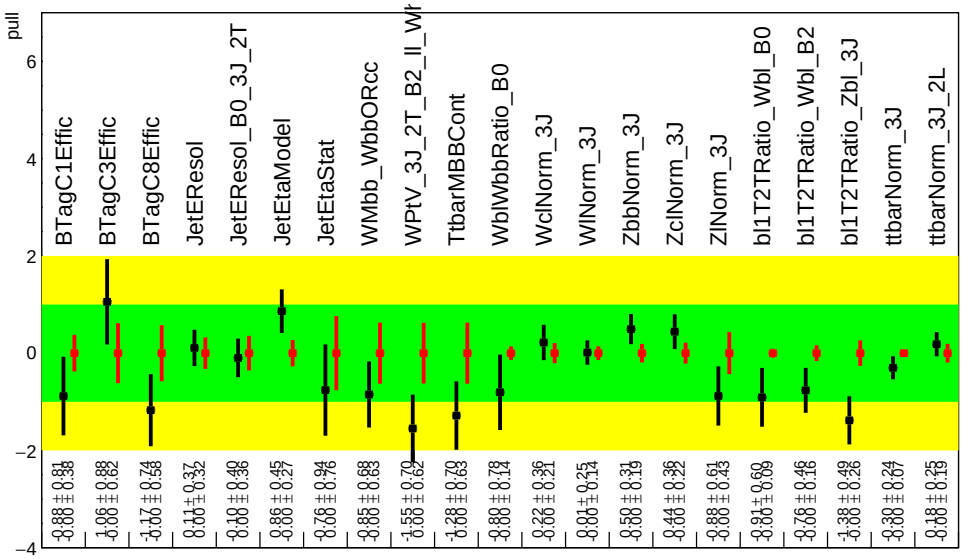


Figure V.14: All nuisance parameters of the Lorentz-invariant analysis that are pulled by more than $\pm 0.75\,\sigma$, or constrained to less than $0.4\,\sigma$. (■) Fit to data (■) Fit to a dataset generated from simulation ("Asimov dataset").

The post-fit values of nuisance parameters should not be interpreted as measurements in their own right. This is because the nuisance-parameter formalism only approximates a full measurement: instead of adding the full likelihood function for a thorough measurement of external parameters, the main results of external measurements are approximated in the $VH \to Vb\bar{b}$ likelihood model by computing the $\pm 1\,\sigma$ impact on the VH distributions and by adding a Gaussian constraint term to the VH likelihood. Most likely, the phase space covered by the $VH \to Vb\bar{b}$ analysis is different from the phase space probed by external measurements. Hence, if post-fit values of nuisance parameters differ from zero, this is not a disagreement with external results. Only when strong pulls ($\gg \pm 1\,\sigma$) are observed, this should be taken as an indication that the VH likelihood model detects a tension between the present data and the external knowledge.

In order to find unjustified pulls or unreasonable constraints, all nuisance parameters of the VH analysis that are either pulled by more than $0.75\,\sigma$, or whose post-fit uncertainties are below $0.4\,\sigma$, have been compiled in figure V.14. There is no convention to set the thresholds to $0.75\,\sigma$ and $0.4\,\sigma$. Larger pulls or stronger constraints also do not necessarily indicate deficits of the likelihood model. These thresholds ensure, however, that neither strong pulls nor strong constraints are missed. The Asimov uncertainties in figure V.14 show that the likelihood model is sensitive to most of these "questionable" nuisance parameters. This is expected because pulls are unlikely when a likelihood model is not sensitive to the values of certain nuisance parameters: the constraint term would dominate the likelihood, and the pull would therefore be close to zero[9]. The figure shows that only 14 out of 177 parameters[10] are pulled outside of the interval $\pm 0.75\,\sigma$, and only five are pulled to more than $\pm 1\,\sigma$. This means that only in very few cases the likelihood fit significantly deviates from external knowledge. Nine out of 177 nuisance parameters are constrained to less than $0.4\,\sigma$. Many of these are simulation uncertainties such as jet multiplicities that can be measured in data. It is expected, in fact, desired that the fit constrains these (item 3 on the previous page).

[9] High sensitivity does not mean that the impact on μ is also high. The impact on μ is entirely determined by how strongly an uncertainty changes the bin contents in the bins with the highest S/B. The sensitivity to a nuisance parameter is determined by how strongly the corresponding systematic uncertainty changes bins where the statistical uncertainties of the data are low.

[10] The γ parameters are not counted because these are not sensitive to deficits of the likelihood model: each γ parameter only affects a single bin and is therefore not sensitive to shape and normalisation differences.

In the following, all nuisance parameters in figure V.14 will be discussed:

`BTagC*Effic` These nuisance parameters describe the probability of b-tagging charm jets or hadronically decaying τ leptons. Since the ATLAS $H \to b\bar{b}$ analysis was the first to use the b-tagging algorithm `MV1c`, whose main purpose is to suppress charm backgrounds, one can expect a high sensitivity to the calibration of charm b-tagging probabilities. This is supported by the Asimov uncertainties.

3 of the 15 `BTagC` nuisance parameters are pulled close to the $1\,\sigma$ level, yet they are consistent with the calibration. Such pulls are acceptable, especially because the likelihood model is unable to strongly constrain the nuisance parameters. The post-fit uncertainties remain close to the input uncertainty, and are therefore translated into uncertainties of μ without significantly reducing the impact of the underlying uncertainty. The same pulls were also observed in the ATLAS $H \to b\bar{b}$ analysis [1].

`JetEResol` The jet energy resolution uncertainty is constrained to 37 %. A strong constraint is expected for two reasons: firstly, the $H \to b\bar{b}$ analysis strongly depends on the jet energy resolution. The Higgs peak in the invariant mass of two b-jets is the most powerful discriminator. Consequently, the BDT classifiers are very sensitive to changes of the jet energy resolution. The fact that the uncertainty in the Asimov fit agrees with the uncertainty obtained in the fit to data shows that the measured constraints are also consistent with predictions. Secondly, other ATLAS analyses also observed constraints of this nuisance parameter, to about 50 % of the input uncertainty. This indicates that the uncertainties of the ATLAS jet energy resolution parametrisation are more conservative than necessary. A constraint of the jet energy resolution parameter is therefore acceptable.

`JetEResol_B0_3J_2T` This nuisance parameter is the same as the previous one, but it is restricted to the region with 3 jets, 2 b-tags, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$ – a region with very low sensitivity[11]. The pull in this region deviates from the pull in the rest of the $VH \to Vb\bar{b}$ phase space for one of the three likelihood models that were compared to the Lorentz+BDTG configuration in section V 1.5, likely due to a statistical fluctuation. This leads to an overconstraint of the jet energy resolution parameter. The impact of this overconstraint is removed by decoupling the nuisance parameter in this category from the nuisance parameter common to all other categories. In order to compare the four classifiers on an equal basis, this decorrelation is also introduced in all other likelihood models, see also footnote 7 on page 146.

For the Lorentz+BDTG configuration in figure V.14, both nuisance parameters agree. This shows that this exception would not be necessary for the Lorentz-invariant analysis used for the main result of this thesis, and it supports the hypothesis that the pull in the other fit configuration is caused by a statistical fluctuation.

`JetEtaModel/-EtaStat` These nuisance parameters describe uncertainties of jet energy calibrations: such calibrations are provided in bins of η, and these two uncertainties have a small, η-dependent impact on jet energies. Therefore, they change the shapes of jet-related observables in all simulations on the level of 1 %, but there is no indication that the pulls are unreasonable. Both uncertainties have a vanishing impact on μ.

`WMbb_B0_WbbORcc` This is a shape uncertainty for the modelling of m_{bb} for the $W{+}bb$ and $W{+}cc$ backgrounds in the region $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$. As discussed in section IV 6, the Monte Carlo predictions for the shape of m_{bb} and p_{T}^{V} differ depending on the generator used. No measurements with data have been conducted, though: this nuisance parameter therefore falls in the category “no prior knowledge”, see item 3 in the list on page 157, and any post-fit value between -1 and 1 is reasonable.

[11] `B0` stands for the first p_{T}^{V} bin, starting at 0 GeV, `B2` stands for the second p_{T}^{V} bin, starting at 120 GeV. `B1` is not used, any more. It is reserved for an intermediate bin in the ATLAS cut analysis [3, 4].

`WPtV_3J_2T_B2_ll_Whf` As for the previous nuisance parameter, Monte Carlo predictions for p_{T}^{V} differ in W+jets events. Pulls of the `WPtV` uncertainty can therefore be expected. However, the p_{T}^{W} nuisance parameter for the 3-jet region was initially pulled to $< -1.5\,\sigma$, outside of the credible interval obtained from Monte Carlo comparisons. Further studies revealed that there is a discrepancy between the region 3 jets, 2 loose b-tags, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$, '`3J_2T_B2_ll`' and all other 3-jet regions: the pulls are -1.6 ± 0.7 in this region vs. -0.7 ± 0.6 in all other regions. Since the strong pull is confined to one region only, a statistical fluctuation is the most likely cause. Therefore, the nuisance parameter in this particular 3-jet category was decoupled from the rest of the 3-jet regions to avoid a bias due to the fluctuation.

`TtbarMBBCont` This nuisance parameter represents uncertainties regarding the predicted m_{bb} distributions in $t\bar{t}$ events, which is estimated by comparing Monte Carlo generators as described on page 119. As for the last two nuisance parameters, pulls can be expected because no external measurements are available. Contrary to the previous nuisance parameter, a statistical fluctuation is likely not the cause of the pull: a clear shape difference between data and Monte Carlo predictions is observed, see appendix C 3 on page 217. By pulling the nuisance parameter, this shape difference is corrected, and this pull is therefore required to correctly describe the data.

`WblWbbRatio_B0, bl1T2TRatio_*` These nuisance parameters are associated to flavour-fraction uncertainties for the Sherpa simulations. These will be discussed in detail in the next section.

`*Norm_3J` As discussed in section III 4.3 of the introduction, the VH analysis relies on cross-section predictions in two jet-multiplicity bins: 2 jets and 3 jets. Since cross-section uncertainties of the 3/2-ratio are larger than for simpler matrix elements without additional jets, '`3J`' normalisation nuisance parameters are assigned to the 3-jet predictions of most backgrounds. As for other Monte-Carlo-only nuisance parameters, no external measurements are fed into the profile-likelihood fit. Jet multiplicities have of course been measured, but not specifically in the $VH \to Vb\bar{b}$ phase space. As shown by the Asimov fits, the likelihood model is very sensitive to such cross-section differences. That means that these can be measured in the $VH \to Vb\bar{b}$ phase space, and it was therefore decided to allow for strong constraints of the 3/2-ratios. All post-fit values are found within the plausible range predicted by Monte Carlo generators. The magnitude of the pre-fit 3-jet normalisation uncertainties can be found in table IV.9 on page 118.

V 2.3 The Treatment of $V+bl$ Normalisations

The use of MV1c and b-tagging at multiple working points (section IV 1.3) improved the sensitivity of the $VH \to Vb\bar{b}$ cut analysis from $1.6\,\sigma$ [4] to $1.9\,\sigma$ [1]. However, with more powerful b-tagging, the $VH \to Vb\bar{b}$ analysis is also more sensitive to the modelling of b-tagging probabilities and flavour fractions in $V+$jets simulations. Since these flavour fractions cannot be predicted accurately (section III 4.3), normalisation uncertainties for all flavour categories have to be added to the likelihood model. These are estimated by comparing predictions of different Monte Carlo generators and changing parton shower and hadronisation models. In the VH analysis, these normalisation uncertainties are as high as $35\,\%$ (ratio of $W+bl$ to $W+bb$ events). The full list of uncertainties can be found in table IV.9 on page 118. No external measurements are available for the $V+$jets flavour fractions. They therefore have to be constrained by fits to data. With the ATLAS likelihood model, the measurement of many flavour fractions is carried out in the $1b$-tag region. However, it is assumed that this measurement can be extrapolated to the $2b$-tag region without any uncertainties. This turned out to be false during the validation of the Lorentz-invariant likelihood models as will be shown in the following.

In likelihood fits using the Lorentz-invariant MVA, two nuisance parameters were found to be pulled significantly – more than expected from comparing various simulations: the $W+bl$ and $Z+bl$ flavour fractions. The ratio of $W+bl$ to $W+bb$ events was found to be pulled to -1.29, and the ratio of $Z+bl$ to $Z+bb$ events to -1.91. With prior uncertainties of $35\,\%$ and $11\,\%$, respectively, and by applying the exponential normalisation extrapolation $(I_-/I_0)^{-\alpha}$ for negative values of α[12], this is equivalent to normalisation factors of

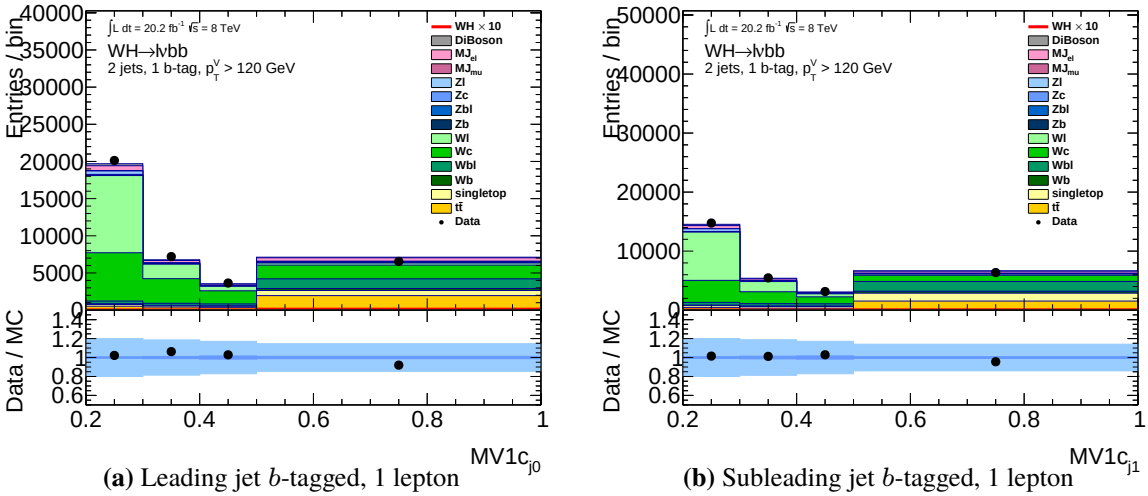
$$\eta_{W+bl} = 0.65^{1.29} \approx 0.57 \quad \text{and} \quad \eta_{Z+bl} = 0.88^{1.91} \approx 0.78\,.$$

The data would favour even lower normalisation factors, but the constraint terms prevent further reduction of these two $V+bl$ normalisations. For this reason, other nuisance parameters such as b-tagging uncertainties or flavour fractions of other $V+$jets flavours are pulled when the ATLAS likelihood model is used. This is a sign of tensions between data and simulations that the likelihood model cannot account for because certain systematic uncertainties have been overlooked.

The cause of the $V+bl$ normalisation reduction was found in the $1b$-tag region, see figures V.15 and V.16. The figures show the response of MV1c for both the leading and the subleading jet. Simulations predict an excess of events in the rightmost bin, whereas predictions in other bins are correct or slightly low. The rightmost bins would thus require decreased normalisation factors, whereas other bins require constant or increased normalisations. The rightmost bins are populated by events in which one jet passes the tight b-tagging requirement ($50\,\%$) while the other jet is not b-tagged. The dominant background process in these bins is $V+bl$, whose normalisation can only be constrained in these bins when the ATLAS likelihood model is used. This is because in the 1-lepton channel, $W+bl$ is almost completely absent from other bins (figure V.15), and $Z+bl$ in the 2-lepton channel is relatively low in the first three bins (figure V.16). An excess of Monte Carlo events in the rightmost bin therefore leads to a significant reduction of the $V+bl$ flavour fractions, whereas other $V+$jets normalisations are increased, especially $Z+bb$. Figures V.15 and V.16 only show the regions with the highest sensitivity, but a similar excess of simulated events is visible in all other $1b$-tag categories of the analysis. This is shown in appendix C 4.1 on page 218.

While an excess of Monte Carlo events is visible in the $1b$-tag region, no such excess is visible in the signal regions with two b-tagged jets. On the contrary, an *increase* of background normalisations is

[12] See equation (IV.13) on page 106.



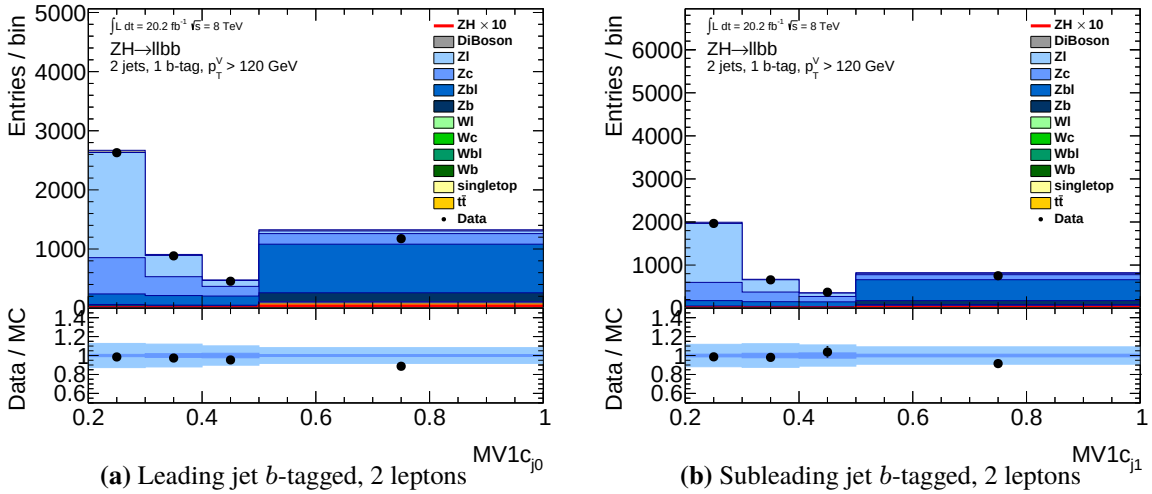
(a) Leading jet b-tagged, 1 lepton **(b)** Subleading jet b-tagged, 1 lepton

Figure V.15: b-tagging response for events with one b-tagged jet in the 2-jet region, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$, 1-lepton channel. The bins correspond to the four working points of the `MV1c` b-tagging algorithm. From left to right: 80, 70, 60 and 50 %. 1b-tag distributions show an excess of Monte Carlo events in the rightmost bin. One of the dominant backgrounds in this bin is $W+bl$ (■). A similar excess is observed in other regions of the analysis. See figure V.16 and appendix C 4.

necessary to describe the data in the 2b-tag region, especially in the 2-lepton channel where the 16 % correction was applied to all plots shown previously. `MV1c` distributions with two b-tagged jets are shown in appendix C 4.2. The excess of events in the 1b-tag region can be caused by several effects:

1. The b-tagging efficiency of the 50 % working point is overestimated in simulations. This option is unlikely, though: ATLAS provides b-tagging calibrations with detailed uncertainties, which are all propagated into the likelihood model as b-tagging nuisance parameters. The blue band in the bottom plot in figure V.15 shows the quadratic sum of these uncertainties. Since these b-tagging uncertainties for both b- and light jets are neither constrained nor pulled strongly, the (single jet) b-tagging calibration is probably not the cause of the misdescription.
2. The probability that both jets in $V+bl$ and partly also $V+cl$ events are b-tagged is underestimated in simulations. $V+bl$ events should normally be observed in the 1b-tag region, but some will migrate into the 2b-tag region if the light jet is misidentified as a b-jet. If the probability of such "mistags" is lower in simulations than in data, a deficit of $V+bl$ events in the 2b-tag region and an excess in the 1b-tag region will be observed.
   Contrary to the previous case, this effect is not fully covered by b-tagging uncertainties. These are derived for single light/charm/b-jets but not for whole events where *two* jets are tagged. b-tagging uncertainties are applied separately for each jet, and no interdependency between the jets is assumed. However, if the response of `MV1c` for a given jet depends on other jets in the same event, the single-jet b-tagging uncertainties can be insufficient to describe the 2b-tag probability in $V+bl$ events.
3. The cross section of $V+bl$ events is overestimated in the Sherpa simulations. This would produce an excess of Monte Carlo events in the rightmost bin of the 1b-tag region, but *also* in the 2b-tag region. A misdescription of cross sections or flavour fractions is therefore likely not the cause of the excess.

Depending on the cause of the excess, the impact on the analysis is different: for cases 1 and 2, an excess in the 1b-tag region would lead to a deficit of events in the 2b-tag region. For case 3, an excess in the



(a) Leading jet b-tagged, 2 leptons **(b)** Subleading jet b-tagged, 2 leptons

Figure V.16: b-tagging response for events with one b-tagged jet in the 2-jet region, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$, 2-lepton channel. The bins correspond to the four working points of the `MV1c` b-tagging algorithm. From left to right: 80, 70, 60 and 50 %. $1b$-tag distributions show an excess of Monte Carlo events in the rightmost bin. The dominant background in this bin is $Z{+}bl$ (■). Z+jets backgrounds are scaled up by 16 % as in previous figures. A similar excess is observed in other regions of the analysis: see figure V.15 and appendix C 4.

$1b$-tag region would be observed in conjunction with an excess also in the $2b$-tag region. The flaw of the ATLAS likelihood model is that it cannot handle all three cases: the flavour-fraction uncertainties are designed to describe case 3, and ATLAS b-tagging uncertainties should cover case 1, but case 2 cannot be described.

For completeness, one should establish that other background processes are most likely not the cause of the excess:

- $t\bar{t}$ events (■) significantly contribute to the rightmost bin in the 1-lepton analysis (figure V.15), but are irrelevant in the 2-lepton channel (figure V.16). Furthermore, the fraction of $t\bar{t}$ events is different depending on whether the leading or subleading jet in p_{T} is b-tagged (figures V.15(a) and V.15(b)). Nevertheless, the excess is of similar magnitude in both distributions.
- Single-top events (■) could cause an excess in the 1-lepton channel, but are absent in the 2-lepton channel. It is therefore unlikely that these are the only source of the excess. They could, however, be affected by a b-tagging calibration problem such as case 2 in the list above because they also have a 'lepton+bl' signature.
- The diboson backgrounds (■) do not impact normalisations because their contribution is very low.
- QCD multijet events (■) are absent in the 2-lepton channel. In the 1-lepton channel, multijet events with a single b-tagged jet are directly taken from data[13]. Hence, their b-tagging efficiencies cannot be calibrated wrongly.

[13] This is an improvement with respect to previous versions of the ATLAS $H \to b\bar{b}$ analyses, which were taking *all* multijet events from the $0b$-tag region, see section IV 2.6.

Nuisance Parameters to Account for the $V+bl$ Excess

Although case 2 from the list above seems to be the most likely cause of the excess, the ATLAS likelihood model treats the $V+bl$ normalisations in both b-tag categories as correlated. That means that it can correctly treat case 3, global changes of the $V+bl$ cross section, and partly case 1 provided that the official ATLAS b-tagging uncertainties are large enough. Yet if case 2 is the cause of the excess, the $V+bl$ flavour fractions will be measured wrongly: since the number of observed events in the $1b$-tag region is about 10 times larger than in the $2b$-tag region, the likelihood model gives higher precedence to the normalisations measured in the $1b$-tag region. (460 000 vs. 45 000 events in the 1-lepton channel. Event yields can be found in appendix D). Any measurement of normalisations in the $1b$-tag region will therefore override conflicting measurements in the $2b$-tag region. Yet if case 1 or 2 are the cause of the excess, a reduction of $V+bl$ events in the $1b$-tag region should lead to an increase of events in the $2b$-tag region. However, fits using the ATLAS likelihood model will do the opposite, and pull *both* normalisations down.

To create a likelihood model that can handle all cases, the ATLAS likelihood model was extended with nuisance parameters allowing for migrations between the 1- and $2b$-tag regions. The essence of a migration is that events that disappear in the $1b$-tag region appear in the $2b$-tag region, which is not what happens with the ATLAS likelihood model. The implementation of these nuisance parameters is similar to the 3- to 2-jet ratio nuisance parameters:

$$\phi_{\text{2T, low } p_{\text{T}}^V} = \phi_{Wbb} \cdot \eta_{\texttt{WblWbbRatio\_B0}} \qquad \phi_{\text{1T, low } p_{\text{T}}^V} = \phi_{Wbb} \cdot \eta_{\texttt{WblWbbRatio\_B0}} \cdot \eta_{\texttt{bl1T2TRatio\_Wbl\_B0}}$$
$$\phi_{\text{2T, high } p_{\text{T}}^V} = \phi_{Wbb} \cdot \eta_{\texttt{WblWbbRatio}} \qquad \phi_{\text{1T, high } p_{\text{T}}^V} = \phi_{Wbb} \cdot \eta_{\texttt{WblWbbRatio}} \cdot \eta_{\texttt{bl1T2TRatio\_Wbl}}$$

This scheme defines the $W+bl$ sample normalisations $\phi_{\text{<region>}}$ as the product of three factors. They float synchronously with the total $W+$jets normalisation factor ϕ_{Wbb}. This factor represents global $W+$jets cross-section changes. The second factor, $\eta_{\texttt{WblWbbRatio}}$, is a flavour-fraction uncertainty, which allows for global adjustments of the $W+bl$ normalisations, but a Gaussian constraint term couples these to $W+bb$ within a credible interval ($1\,\sigma$) of 1.00 ± 0.35. This accounts for parton shower, hadronisation or flavour-fraction differences that change the total number of $W+bl$ events, regardless of the b-tagging category. The newly added third factor is a normalisation uncertainty named `bl1T2TRatio*` (blue), which is sensitive to migrations between the b-tagging categories. Only with these uncertainties, all three possible causes of the excess can be described. Similar uncertainties were added for $Z+bl$ events, as well as single-top production. This is because these events will be affected in a similar way if the excess is caused by the b-tagging calibration for 'bl' events.

Since flavour-fraction uncertainties in the ATLAS $H \to b\bar{b}$ analysis from chapter IV were found to depend on p_{T}^V in the 1-lepton analysis, the low- and high-p_{T}^V flavour fractions were decoupled for the ATLAS Run-1 publication [1]. In the 2-lepton channel, the same was done for the 2- and 3-jet regions. The same scheme was therefore implemented also for the new nuisance parameters; leading to a total of five new nuisance parameters: two for $W+bl$ and $Z+bl$ each, one for single-top production. For single-top production, separate uncertainties were tested but there is no indication that such separate treatment is necessary, particularly because single-top production is only relevant in the 1-lepton channel, and here it is small.

In principle, the $1b$-tag $\leftrightarrow$ $2b$-tag uncertainties could also be implemented with nuisance parameters that have a simultaneous effect on both the 1- and $2b$-tag region. In this case, $\eta_{1\text{T}}$ and $\eta_{2\text{T}}$ would have to be anti-correlated to describe the migration of events between the 1- and the $2b$-tag regions. The reason this was not done is twofold:

Firstly, it is known that the Monte Carlo description of the Sherpa simulations is better when more jets are b-tagged: the $0b$-tag regions show clear signs of mismodelling, and corrections had to be applied

(see section IV 2.3). The $0b$-tag regions were therefore not included in the profile-likelihood fit. In the $2b$-tag regions, no such mismodelling could be detected. It is therefore assumed that the $2b$-tag region should be preferred over the $1b$-tag region. However, due to the 10 times higher statistics in the $1b$-tag region, the ATLAS likelihood model measures $V+bl$ flavour fractions virtually only in the $1b$-tag region. Assigning η parameters only to the $1b$-tag regions gives higher precedence to the $V+bl$ flavour-fraction measurements from the $2b$-tag regions.

Secondly, assigning these uncertainties only to the $1b$-tag region yields a more conservative likelihood model: with more emphasis on the $2b$-tag regions, the $V+bl$ flavour fractions acquire larger uncertainties – both due to lower data statistics and because $V+bl$ events are not the dominant background after requiring two b-tags. If backgrounds are constrained less, uncertainties of μ become larger. This kind of likelihood model better reflects the fact that before doing the measurement it is unknown whether normalisations in the 1- and $2b$-tag regions should be correlated or anti-correlated. It can handle both cases, and will not need any adjustments if the data or $V+$jets Monte Carlo predictions are exchanged.

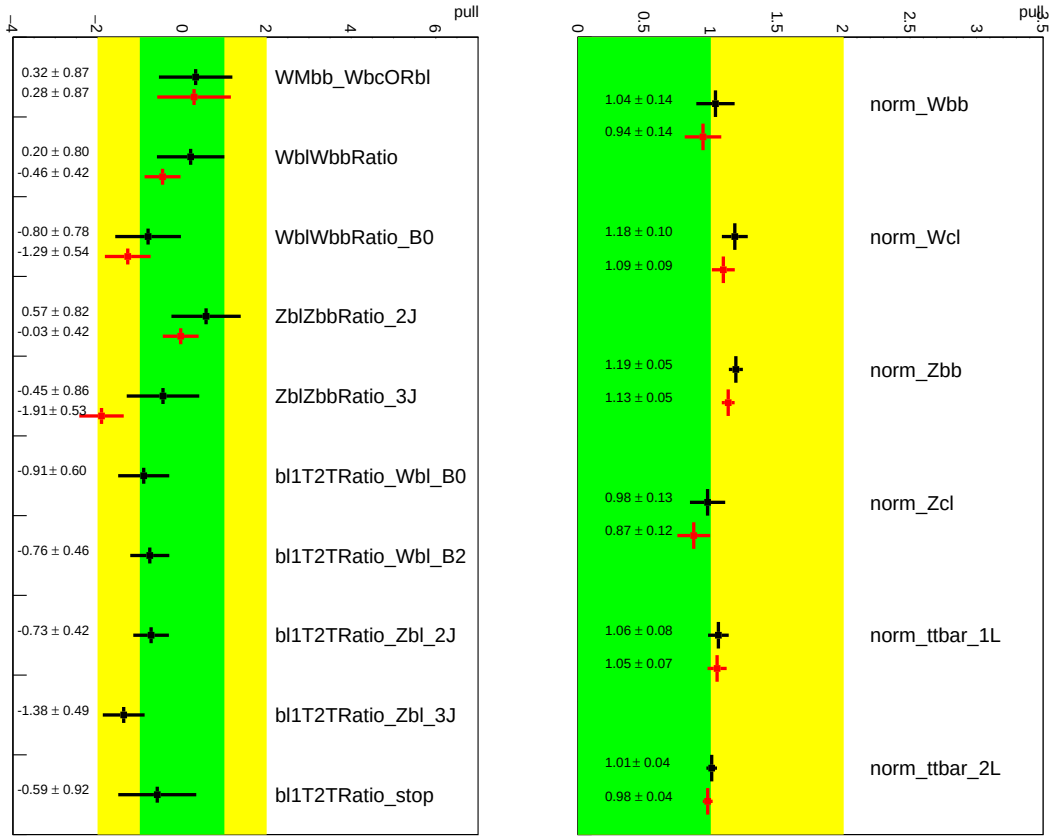
It is also unknown how large the new uncertainties should be. The initial flavour-fraction uncertainties were obtained by Monte Carlo comparisons on truth events without any use of b-tagging. The uncertainties assigned to the $1b$-tag flavour fractions were therefore chosen slightly larger than the "truth" flavour-fraction uncertainties: 1.0 ± 0.4 for $W+bl$, 1.0 ± 0.2 for $Z+bl$ and 1.0 ± 0.1 for the single-top s-, t- and Wt channel.

A decisive advantage of the augmented likelihood model is that it can be used for all possible causes of the excess that were listed before: in case of a b-tagging-related problem (case 1 and 2), the $1b$-tag normalisations will move opposite to the $2b$-tag normalisations. The `1T2TRatio` nuisance parameters will therefore be pulled down, whereas the flavour-fraction uncertainties common to both b-tagging categories will remain close to zero or increase slightly. In case of a global cross-section problem, case 3, the `1T2TRatio` nuisance parameters will remain close to zero, and $\eta_{\texttt{WblWbbRatio}}$ or $\eta_{\texttt{ZblZbbRatio}}$ will be pulled instead.

Figure V.17 shows results from fits to data. The bl-related nuisance parameters are shown in (a), and the global normalisation factors are shown in (b). Fit results with the ATLAS likelihood model (■) are compared with the improved likelihood model (■). In the fit with the improved model, all `bl1T2TRatio` nuisance parameters are pulled down, and global normalisations as well as flavour fractions move up. This indicates a problem with the simulation or calibration of b-tagging, case 1 or case 2, and rules out a global cross-section problem, case 3. Incidentally, the case that is entirely ruled out is the only case that the ATLAS likelihood model could have treated correctly.

One further finds that *all* pulls of the `bl1T2TRatio*` uncertainties for $V+$jets simulations are incompatible with zero within post-fit uncertainties. Instead, they are compatible with $\alpha \approx -1$ in *all* regions. This clearly proves that the mismodelling is present in all regions, and it suggests that the underlying cause is independent of whether $W+$jets, $Z+$jets, two jets or three jets are simulated. The single-top extrapolation nuisance parameter `bl1T2TRatio_stop` is consistent with both zero and one within post-fit uncertainties. It is therefore unclear if the misdescription is caused by the Sherpa simulations or a problem of the `MV1c` calibration.

Since the ATLAS likelihood model was augmented with additional degrees of freedom to accommodate a clear misdescription of the data, pulls on other nuisance parameters reduce: flavour fractions such as `WblWbbRatio` or `ZblZbbRatio` move up, which is expected if events migrate between the 1- and $2b$-tag regions. More importantly, the flavour fractions are also compatible with the pre-fit uncertainties obtained from Monte Carlo comparisons when the likelihood model is augmented. Moreover, the global normalisation factors in figure V.17(b) become more consistent with expectations: $Z+$jets backgrounds had to be scaled up by $16\,\%$ in chapters IV and V, for example in figure V.7 on page 141. A normalisation



(a) Normalisation uncertainties for bl-events **(b)** All normalisation factors

Figure V.17: **(a)** Normalisation uncertainties and **(b)** normalisation factors for the Lorentz-invariant $VH \to Vb\bar{b}$ analysis using the ATLAS likelihood model (bottom ■) and an improved likelihood model with additional normalisation uncertainties (top ■).
Nuisance parameters, which have no counterpart in the ATLAS likelihood model, were introduced to model the excess of Monte Carlo events in the rightmost bins of `MV1c`.

correction for W+jets events is also necessary as shown in the same figure. The improved likelihood model supports this observation, adjusting Z+bb to 1.19 instead of 1.13, Z+cl to 0.98 instead of 0.87, W+bb to 1.04 instead of 0.94 and W+cl to 1.18 instead of 1.09. $t\bar{t}$ normalisations are not affected, which is justified if the misdescription is confined to bl-events. Furthermore, the uncertainty of all normalisation factors and all flavour fractions increases. This is a consequence of choosing the more conservative way to correct the likelihood model, and it is justified because normalisations are evidently hard to determine given that different regions of the analysis yield conflicting normalisation factors. Lastly, the improved likelihood model reduces pulls of b-tagging uncertainties. Figures C.11(a) to C.11(c) in appendix C 4.3 show that 15 pulls of b-tagging uncertainties decrease, whereas only three increase notably. This indicates that the deficits of the ATLAS likelihood model lead to erroneous pulls of these parameters because no other degrees of freedom exist to describe the mismodelling. This is an expected behaviour given that reducing the normalisation of V+bl events synchronously in both b-tagging categories is the wrong approach if events migrate between the categories.

The results obtained with the improved likelihood model also explain why one-electron events with low p_{T}^{V} were removed from the ATLAS $H \to b\bar{b}$ analysis [1]. The synchronous reduction of V+bl

normalisations due to the excess in the $1b$-tag region leads to a deficit of background events in the $2b$-tag region. Although numerous nuisance parameters are pulled to compensate, the deficit cannot be eliminated entirely because none of the systematic uncertainties of the ATLAS model can sufficiently describe migrations between the 1- and $2b$-tag categories. The resulting deficit of events is most apparent in the regions '2 jets, 1 lepton, low p_{T}^{V}' and '3 jets, 2 leptons'. These are the regions with the strongest pulls in figure V.17(a), especially when using the ATLAS likelihood model. The deficit of V+jets events affects the measurement of the signal strength: $t\bar{t}$ events are constrained from the 3-jet region, and can therefore not be scaled up significantly. Other background normalisations are inferred mostly from the $1b$-tag region because its statistical power is higher than that of the $2b$-tag region. Therefore, only one freely floating Monte Carlo sample remains in the $2b$-tag region: the VH signal. If the mismodelling in the $1b$-tag region pulls backgrounds in the $2b$-tag region down, the Higgs signal has to be increased to compensate. Given that the tensions observed with the ATLAS likelihood model were most severe in the 1-lepton channel with low p_{T}^{V}, it was decided to remove events with one isolated electron from this region. This reduces the statistical power of the $1b$-tag region by $50\,\%$, and therefore also reduces the pull of the `WblWbbratio_B0` nuisance parameter. Furthermore, it reduces pulls of the global normalisations shown in figure V.17(b). However, tensions in other regions are not corrected by this decision. Even though $50\,\%$ of the events in the low-p_{T}^{V} region were missing, `WblWbbratio_B0` was still found to be pulled to -0.9 ± 0.6, which is more than for the Lorentz-invariant analysis that uses the *full* low-p_{T}^{V} region. `ZblZbbratio_J3` was even found to be pulled to -2.0 ± 0.6 [169], but no actions were taken by ATLAS because this region has a low impact on μ.

In this thesis, the decision to remove one-electron events is questioned: events with one electron and low p_{T}^{V} are not the cause of the mismodelling. Removing these admittedly reduces its impact, but only in one region of the analysis. Given that all `1T2TRatio` nuisance parameters in figure V.17(a) are pulled consistently across lepton multiplicity, p_{T}^{V} and jet-multiplicity categories, one must conclude that the underlying problem affects all regions simultaneously. Reducing the statistical power of a single region is not the correct approach.

To finalise the likelihood model for the Lorentz-invariant MVA, further studies were conducted in order to test if similar uncertainties could be necessary for other V+jets flavour combinations, but no indication was found. This is because, firstly, heavier-flavour events such as $V{+}cc$, $V{+}bc$ are negligible in the $1b$-tag region. Their normalisation is therefore predominantly measured in the $2b$-tag region, and no migration uncertainty is necessary. Secondly, events with only light flavour jets are of low importance in the $2b$-tag region, even negligible for b-tagging tighter than `LL`, see for example figure C.7. Thirdly, it was tested if $V{+}cl$ migration uncertainties are necessary, but the result was negative: $V{+}cl$ is the only flavour category similar to $V{+}bl$, where a heavy-flavour jet occurs in combination with a light-flavour jet. A likelihood model with additional $1b$-tag $\leftrightarrow$ $2b$-tag ratios for $V{+}cl$ events was created to test for signs of mismodelling. Nuisance-parameter pulls for this likelihood model are compared to the baseline model for this thesis and to the ATLAS model in appendix C 4.4 on page 227. It is found that pulls for events other than $V{+}cl$ change only marginally. Moreover, three of four $V{+}cl$ nuisance parameters are found to be consistent with zero. Only one is found at 0.9 ± 0.7, but this pull likely occurs because of an anti-correlation with the corresponding $V{+}bl$ pull. The two nuisance parameters describe the same problem and can thus be represented by a single nuisance parameter, preferably the $V{+}bl$ nuisance parameter. Finally, the impact on μ due to $V{+}cl$ parameters was found to be marginal as well. $V{+}cl$ ratios are therefore not necessary to describe the data.

Table V.4: Impact of the $V+bl$ 1T $\to$ 2T extrapolation uncertainties on the signal strength in different categories of the Lorentz-invariant analysis.

	1 Lepton	2 Leptons	1+2 Leptons
low p_{T}^{V}	$+0.4$ -0.3	$+1.0$ -1.0	
high p_{T}^{V}	$+0.2$ -0.2	$+0.1$ -0.1	
Combination	$+0.12$ -0.12	$+0.29$ -0.27	$+0.11$ -0.11

The impact of the $1b$-tag $\leftrightarrow$ $2b$-tag ratio parameters on μ can be quantified by fixing the parameters while carrying out likelihood ratio tests as for figure V.9. The quadratic difference of uncertainties is shown in table V.4. The full contribution to the uncertainty of μ is 0.11. This is not negligible because the total uncertainty of μ is about 0.6, but in comparison to other uncertainties it is of minor importance. Not including the nuisance parameters, however, yields a likelihood model that is unable to describe the data, which is why one-electron events were removed from the ATLAS likelihood fit.

By constructing likelihood models with separate parameters of interest such as μ_{1L} and μ_{2L}, or even separate parameters for each p_{T}^{V} category, the impact on different regions was measured. The largest uncertainties are obtained in the low-p_{T}^{V} regions, which is consistent with the observations of the ATLAS $H \to b\bar{b}$ analysis. It should be understood that the uncertainties in table V.4 are not showing the bias of μ that would be encountered if the ATLAS likelihood model was used to fit the data. The table only shows uncertainties that are introduced by correcting the $V+bl$ excess.

On one hand, uncertainties of μ increase as shown in the table, but on the other hand the improved likelihood model corrects for the modelling problem. Improving the b-tagging calibrations, external measurements of flavour fractions or Monte Carlo predictions would be preferable, but all these remedies are beyond the scope of this thesis. Instead, the presence of the modelling problem is acknowledged, and properly reflected in the likelihood model at the cost of slightly higher uncertainties of μ. Nonetheless, all results from section V 1 hold: the comparison of the classifiers was done with the corrected likelihood model. The Lorentz-invariant classifiers should therefore be preferred over the standard ATLAS classifiers.

One might wonder why the mismodelling of the `MV1c` distributions was not detected with the multivariate modelling checks introduced in section IV 4. This is because no $1b$-tag events enter the BDT training[14]. The misdescription of normalisations and migrations affecting this region could therefore not be detected.

The repetition of the $VH \to Vb\bar{b}$ analysis on Run-2 data [44] is most likely not affected by the modelling problem because all critical regions are excluded from the analysis: $1b$-tag events are ignored entirely, as well as all events with $p_{\mathrm{T}}^{V} < 150\,\mathrm{GeV}$ (0- and 1-lepton channel) or with $p_{\mathrm{T}}^{V} < 75\,\mathrm{GeV}$ (2-lepton channel). Furthermore, b-tagging is used only with a single working point, and the Sherpa generator, the b-tagging algorithm and its calibration have been updated.

This concludes the validation of the likelihood model. In validating the likelihood model for the Lorentz-invariant classifiers, two deficits of the ATLAS likelihood model were uncovered, which also affect the ATLAS Run-1 publication [1]. In order to conduct measurements with the Lorentz-invariant classifiers, the smoothing algorithm and the treatment of $V+bl$ normalisations were corrected, which

[14] Except for the QCD multijet estimate, for which great care is taken that these $1b$-tag events actually resemble $2b$-tag events.

enables the inclusion of events with one isolated electron in the low-p_{T}^{V} region. These had been excluded from the ATLAS analysis. All known deficits of the ATLAS likelihood model are now corrected for, and since the analysis of pulled and constrained nuisance parameters in section V 2.2 did not reveal any further problems, the likelihood model can now be used for measurements in ATLAS data recorded in 2012.

CHAPTER VI

Measurement of the $VH \to Vb\bar{b}$ Signal Strength

The likelihood model introduced in section IV 6 and validated in section V 2 has been used with the Lorentz-invariant MVA from section V 1 to measure the signal strength of $VH \to Vb\bar{b}$ decays in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ recorded by ATLAS in 2012. Contrary to the ATLAS VH analysis described in chapter IV, events with one isolated electron and $p_{\text{T}}^{V} < 120\,\text{GeV}$ are included in the measurement. Distributions of the BDT discriminant after the profile-likelihood fit for the two most sensitive regions of the Lorentz-invariant analysis are shown in figure VI.1.

For a Standard Model Higgs boson with $m_H = 125\,\text{GeV}$, the signal strength measured in the data is[1]:

$$\mu_{VH} = 1.3^{+0.7}_{-0.6}.$$

Measurements with separate parameters of interest in the 1-lepton and 2-lepton channels yield:

$$\mu_{WH} = 1.6^{+0.8}_{-0.7} \qquad \text{and} \qquad \mu_{ZH} = 0.7^{+1.1}_{-1.0}.$$

Computing the χ^2 probability for the compatibility of the separate and the combined measurement yields $46\,\%$, which indicates that the measurements are well compatible despite the difference in the central values. The results are also fully compatible with the Standard Model prediction, which is within 0.6 standard deviations of the measured outcome.

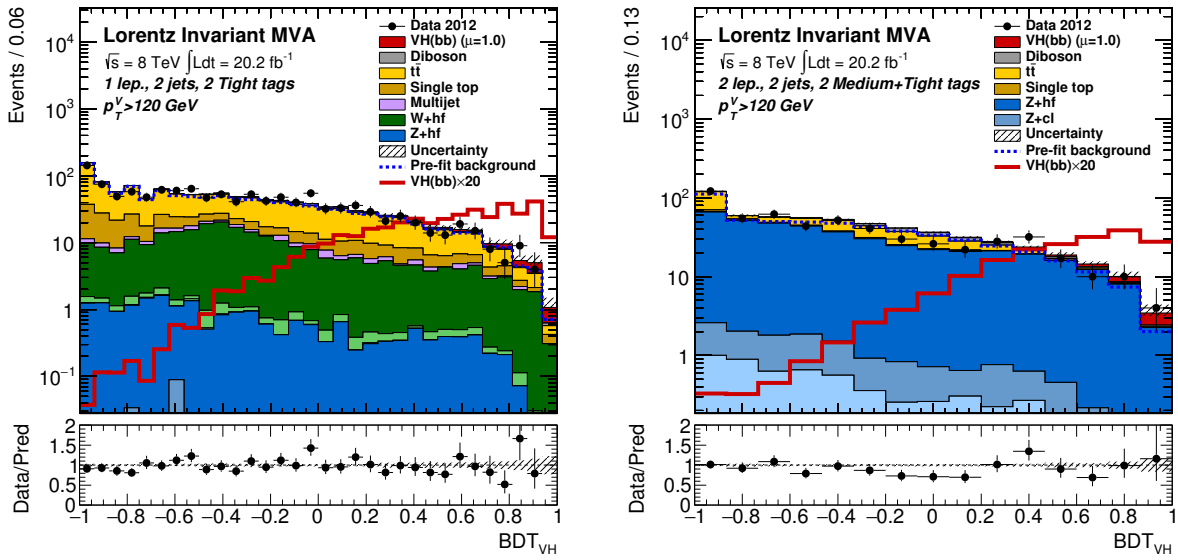
Figure VI.2 shows the signal strength measurements with uncertainties separated into statistical and systematic uncertainties. The $VH \to Vb\bar{b}$ analysis is limited by both statistical and systematic uncertainties, which are of equal size except for the 2-lepton channel, where statistical uncertainties are higher than systematic uncertainties.

An excess over the background-only hypothesis is found. For a Higgs boson with $m_H = 125\,\text{GeV}$, p-value and significance are:

$$p_0 = 8.17 \times 10^{-3},$$
$$Z_0 = 2.40\,\sigma.$$

p_0 is the probability that a null model without signal events would yield a signal strength equal or higher than the observed value of μ. The null hypothesis is rejected with a confidence of $99.2\,\%$ or $2.4\,\sigma$. The

[1] Uncertainties are rounded according to the PDG rounding conventions [13].



(a) 1 lepton, 2 jets, high p_{T}^{V}, TT b-tagging

(b) 2 leptons, 2 jets, high p_{T}^{V}, MM+TT b-tagging

Figure VI.1: BDT-classifier distributions after the profile-likelihood fit in the most sensitive regions of the Lorentz-invariant analysis. The Higgs signal (■) is shown both as an overlay magnified by 20 and as a histogram scaled to the Standard Model expectation.

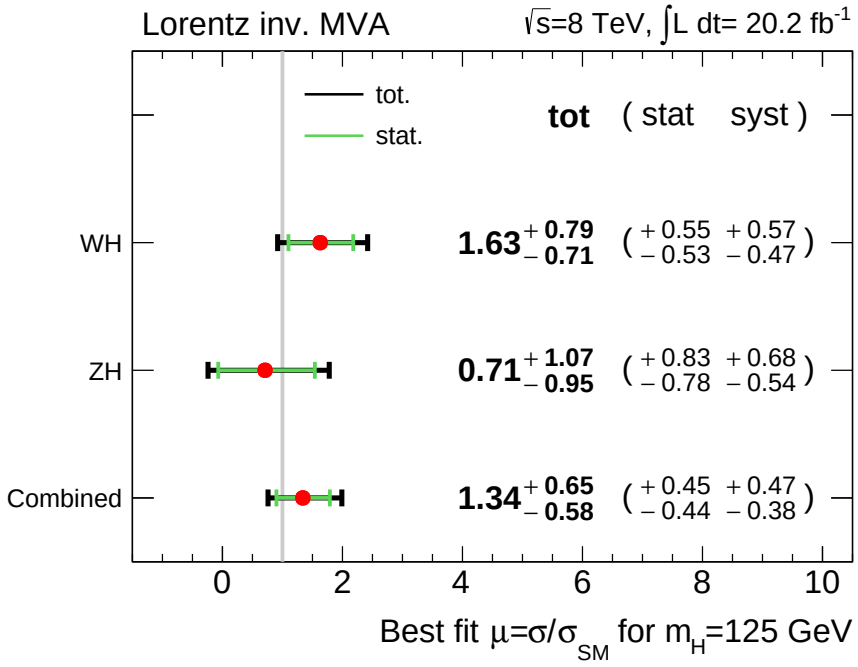
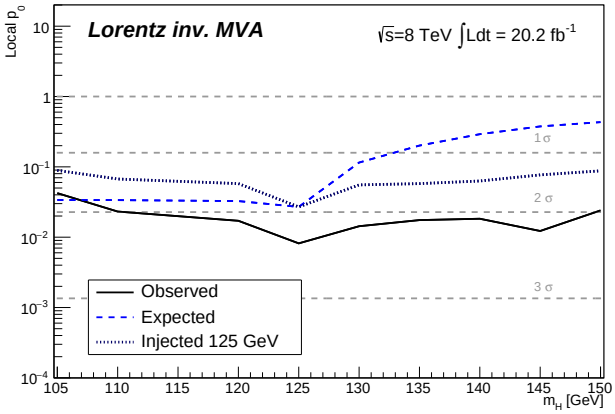
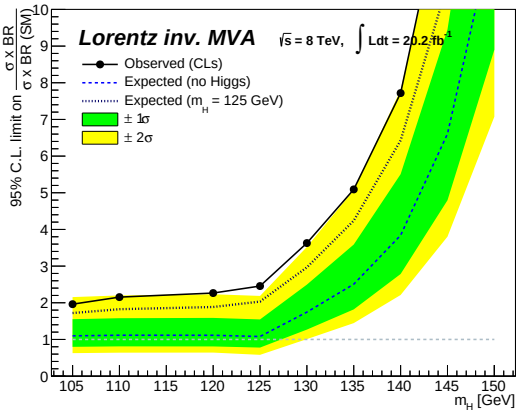


Figure VI.2: Signal strength obtained with Lorentz-invariant multivariate classifiers in the search for associated Higgs production with $H \to b\bar{b}$ decays. The dataset consists of $20.2\,\mathrm{fb}^{-1}$ of proton-proton collisions recorded by ATLAS in 2012 at a centre-of-mass energy of $8\,\mathrm{TeV}$. The Higgs mass is assumed to be $125\,\mathrm{GeV}$. The WH and ZH measurements are obtained from a combined fit in the 1- and 2-lepton channels that uses two parameters of interest: μ_{WH} in the 1-lepton channel and μ_{ZH} in the 2-lepton channel.



(a) p_0 values of the Lorentz-invariant $H \to b\bar{b}$ search.



(b) 95 % $\mathrm{CL_s}$ upper limits on the signal strength μ for the Lorentz-invariant $H \to b\bar{b}$ search.

Figure VI.3: p_0 values, discovery significance and 95 % $\mathrm{CL_s}$ upper limits obtained with the Lorentz-invariant $H \to b\bar{b}$ search. The figures show the results from combined measurements in the 1- and 2-lepton channels for Higgs production in association with a vector boson. The data used are $20.2\,\mathrm{fb}^{-1}$ of proton-proton collisions recorded by the ATLAS collaboration in 2012 at a centre-of-mass energy of $8\,\mathrm{TeV}$.

excess is therefore not large enough to claim evidence for $H \to b\bar{b}$ decays. $p_0 = 2.69 \times 10^{-2}$ and $1.93\,\sigma$ were expected from simulations. Since the observed signal strength μ is higher than the Standard Model expectation, the measured significance is also higher than expected, and the measured p_0 is lower.

Figure VI.3(a) shows p_0 for Higgs boson masses from 105 to 150 GeV. The black line shows the observed p_0, and the dashed blue line shows the expected distribution of p_0, predicted by Asimov datasets, for various Higgs masses. The dotted line shows the expectation for a Higgs mass of 125 GeV. The figure shows that the analysis is optimised for a Higgs mass of 125 GeV, but due to the low mass resolution of the di-b-jet system, Higgs bosons with lower masses would also be detected with similar sensitivity. The loss in sensitivity for higher Higgs-boson masses is primarily caused by the lower probability of $H \to b\bar{b}$ decays: $H \to W^+W^-$ decays are more probable than $H \to b\bar{b}$ decays when $m_H > 135\,\mathrm{GeV}$, see section II 5.3.

Although evidence for $H \to b\bar{b}$ decays cannot be claimed, the signal hypothesis is not rejected, either. Figure VI.3(b) shows the 95 % confidence level upper limit on μ for the production of a VH state with subsequent $H \to b\bar{b}$ decay. At $m_H = 125\,\mathrm{GeV}$, one finds:

$$\mu_{VH} \leq 2.46.$$

Upper limits of 1.08 (null hypothesis) and 2.03 (SM hypothesis) were expected from simulations. These expectations are indicated by dashed lines in the figure. Coloured bands show the impact on the limits if μ would fluctuate by ± 1 and ± 2 standard deviations. The simulations suggest that the current analysis should have nearly sufficient sensitivity to exclude the signal hypothesis at the 95 % confidence level if $H \to b\bar{b}$ decays were absent. Yet the observed upper limits are found to be higher by about two standard deviations of μ. This result is consistent with the non-zero discovery significance. The steep rise of the upper limits for $m_H \gtrsim 130\,\mathrm{GeV}$ is explained by the rapidly falling $H \to b\bar{b}$ branching fraction, similar as for p_0.

In comparison with the first multivariate $WH \to \ell\nu b\bar{b}$ analysis developed in Bonn [7], the exclusion power was almost doubled: the expected upper limit μ^{up} reduced from 2.06 to 1.08, and the expected

discovery significance increased from $1.04\,\sigma$ to $1.93\,\sigma$ – with an identical dataset. This is the result of introducing "continuous b-tagging" in conjunction with a more sophisticated likelihood model, of adding the $ZH \to \ell\bar{\ell}b\bar{b}$ channel, of redefining the preselection to give freedom to multivariate classifiers, and finally of optimising BDT classifiers with Lorentz-invariant variables for each of the two channels in order to reduce systematic uncertainties.

Figures VI.4 and VI.5 show event displays of the two data events that were assigned the highest classifier scores in the 1- and 2-lepton channels. The 1-lepton event in figure VI.4 is the most signal-like of the four events observed in the second-to-last bin of the 1-lepton classifier output in figure VI.1(a). After the fit to data, simulations predict 2.1 signal events ($\mu = 1$) and 3.9 background events in this bin, that is, a chance of 1:1.9 that this is a $WH \to \ell\nu b\bar{b}$ event. The 2-lepton event in figure VI.5 is the most signal-like of the four events observed in the rightmost bin of the 2-lepton classifiers shown in figure VI.1(b). Here, 1.4 signal events and 2.0 background events, a $ZH \to \ell\bar{\ell}b\bar{b}$ likelihood of 1:1.4, are predicted. The event shown in figure III.12 on page 49 to illustrate secondary vertices in b-jets has the third-highest classifier score in the 1-lepton analysis, and therefore also falls in the bin with a signal likelihood of 1:1.9.

VI 1 Comparison with ATLAS Results

The results of the ATLAS Run-1 analysis [1] cannot be compared directly to the results of the Lorentz-invariant analysis because the likelihood models differ: firstly, ATLAS results are obtained from a simultaneous fit to 0+1+2 lepton data recorded at $\sqrt{s} = 7$ and $8\,\mathrm{TeV}$, but in this work, only the channels with charged leptons are analysed using $8\,\mathrm{TeV}$ data. Secondly, the ATLAS likelihood has two flaws that were discussed in section V 2. It does not correctly take into account an excess of simulated events in the $1b$-tag region. It therefore overestimates μ, and underestimates normalisation uncertainties, which is why events with one isolated electron in the low-p_{T}^{V} region were removed for the ATLAS Run-1 publication. Table VI.1 therefore compares the Lorentz-invariant MVA both with the official ATLAS results in the middle column and with a projection of results that should have been obtained if the ATLAS analysis had been conducted only for the 1+2-lepton channels with $8\,\mathrm{TeV}$ data.

This projection was obtained as follows: the official inputs to the ATLAS Run-1 $VH \to Vb\bar{b}$ profile-likelihood fit were rebinned from 1 000 to 32 (1L) and 15 bins (2L) to match the binning of the Lorentz-invariant analysis shown in figure VI.1. In this way, differences due to histogram binning are mitigated. The resulting histograms were used to create an ATLAS-style $H \to b\bar{b}$ likelihood model without the improvements introduced in sections V 2.1 and V 2.3, and a simultaneous fit of the 1- and 2-lepton channels with $8\,\mathrm{TeV}$ data was carried out. Since it is known that this likelihood model has no parameters to account for the mismodelling of the `MV1c` distribution, events with one electron and low p_{T}^{V} were removed as for the official ATLAS result. This reduces the impact of the excess in the $1b$-tag region, but does not entirely remove the bias towards higher values of μ. Since the ATLAS likelihood model is known to underestimate uncertainties from extrapolating the $V{+}bl$ normalisations from the 1- to the $2b$-tag region, the uncertainties shown in table V.4 were added in quadrature to the uncertainties obtained from the profile-likelihood fit. This yields approximately correct results because the $V{+}bl$ migration uncertainties are uncorrelated with all but the $V{+}bl$ flavour-fraction uncertainties, which are not the dominating uncertainties of the VH analysis. Table VI.1 also has one column that quantifies the impact of not removing one-electron events, but this column only shows the relative change of μ.

Although the third column in table VI.1 does not show officially approved ATLAS results, one can conclude that the Lorentz-invariant analysis and the ATLAS Run-1 inputs are in good agreement, despite the deficits of the ATLAS likelihood model discussed in section V 2.

Table VI.1: Comparison of the signal strength obtained with the Lorentz-invariant $VH \to Vb\bar{b}$ analysis (left) and the official ATLAS analysis (middle). Since the official ATLAS result does not separately quote an 8 TeV result, ATLAS inputs for the Run-1 $VH \to Vb\bar{b}$ profile-likelihood fit were re-analysed using the ATLAS likelihood model (right). Since this likelihood model was shown to not describe the excess of $V+bl$ events in the 1b-tag region (section V 2.3), events with one electron and low p_{T}^{W} were removed from the ATLAS histograms, and uncertainties have been adjusted accordingly. The impact of keeping these events is shown in the rightmost column.

Analysis	Lorentz invariant	ATLAS [1]	Run 1 VH inputs[a]	
Dataset	8 TeV	7+8 TeV	8 TeV	
Fit configuration	1+2 L	0+1+2 L	1+2 L	
Electrons with low p_{T}^{W}	Included	Excluded	Excluded	Included
1 L	$1.6^{+0.8}_{-0.7}$	$1.2^{+0.7}_{-0.6}$	$1.6^{+0.9}_{-0.8}$	+29 %
2 L	$0.7^{+1.1}_{-1.0}$	$0.9^{+0.9}_{-0.9}$	$0.8^{+1.0}_{-0.9}$	+ 6 %
1+2 L	$1.3^{+0.7}_{-0.6}$		$1.3^{+0.8}_{-0.7}$	+26 %

[a] Not an ATLAS result. See discussion in the text.

In the 1-lepton channel, results of the Lorentz-invariant analysis and results obtained from official ATLAS histograms even agree to the first decimal place after applying the PDG rounding recommendations [13]. Evidently, the removal of one-electron events in the low-p_{T}^{W} region sufficiently reduces the bias towards higher values of μ such that the measured signal strength is identical. The last column shows that this would not be the case if the low-p_{T}^{W} region was fully included as in the Lorentz-invariant analysis. The official ATLAS 1-lepton signal strength in the middle column is considerably lower because it includes the 7 TeV measurement [4] of $\mu_{7\,\mathrm{TeV}}^{1L} = -2.5 \pm 2$. Uncertainties also differ, which is caused by three effects: 7 TeV data are included, the 0-lepton channel is fit simultaneously with 1+2 leptons, and no $V+bl$ normalisation uncertainties are included. The inclusion of the 0-lepton channel and the 7 TeV data yields additional constraints for background normalisations, and thus also reduces uncertainties in the 1- and 2-lepton channel.

In the 2-lepton channel, results are similar, but they do not agree as closely as in the 1-lepton channel. The 2-lepton ATLAS signal strength can be expected to be slightly higher because $V+bl$ migration uncertainties also affect the signal strength in the 2-lepton channel, but no extra treatment such as excluding a part of the low p_{T}^{V} region is applied. The measurement of the bias in the rightmost column only quantifies the impact of excluding events from the 1-lepton channel. The 2-lepton result obtained from ATLAS inputs further includes a kinematic likelihood fit, which improves the resolution of m_{bb}. These two differences explain the lower uncertainties, and why central values differ slightly. The official ATLAS 2-lepton result differs even more because it is again obtained from a simultaneous fit with the 0-lepton channel and with 7 TeV data. The 7 TeV result is $\mu_{7\,\mathrm{TeV}}^{2L} = 0.6 \pm 4.0$.

Finally, the combined 1+2 lepton results are found to agree. This can be seen as an a-posteriori justification for excluding one-electron events with low p_{T}^{W}. Although the wrong treatment of migrations between the 1- and 2b-tag region is not confined to this region, and the reduction of the statistical power in this region is an unreasonable remedy, it nonetheless leads to similar results as with the Lorentz-invariant analysis, which covers the full $VH \to Vb\bar{b}$ phase space. A comparison with an official ATLAS result for the combination of the 1- and 2-lepton channels is not possible because such a combination was not published.

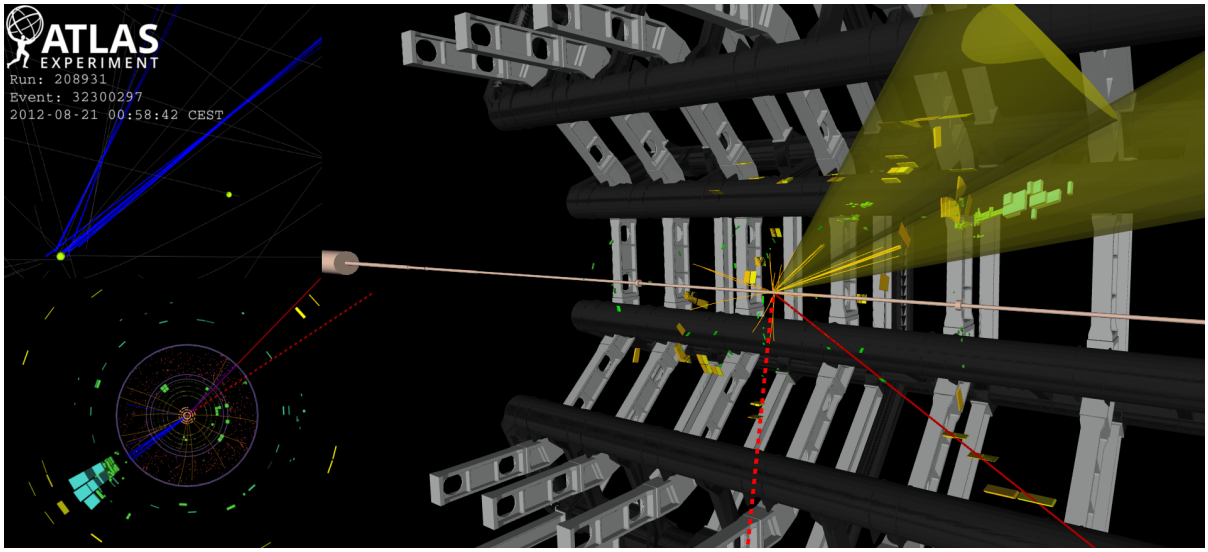


Figure VI.4: $WH \to b\bar{b}\mu\nu$ candidate event with the highest signal probability recorded by ATLAS in the data of 2012. A pair of b-jets (yellow cones) recoils off a muon and $E_{\mathrm{T}}^{\mathrm{miss}}$ (red tracks), which are compatible with a W decay. In the transverse view, tracks assigned to the b-jets are shown in blue. Several tracks have significant impact parameters. The event is categorised as 1 lepton, 2 jets, high p_{T}^{V}, TT b-tagging, and has the highest classifier score observed in the 1-lepton channel: 0.997. This corresponds to one signal event in 1.9 background events.

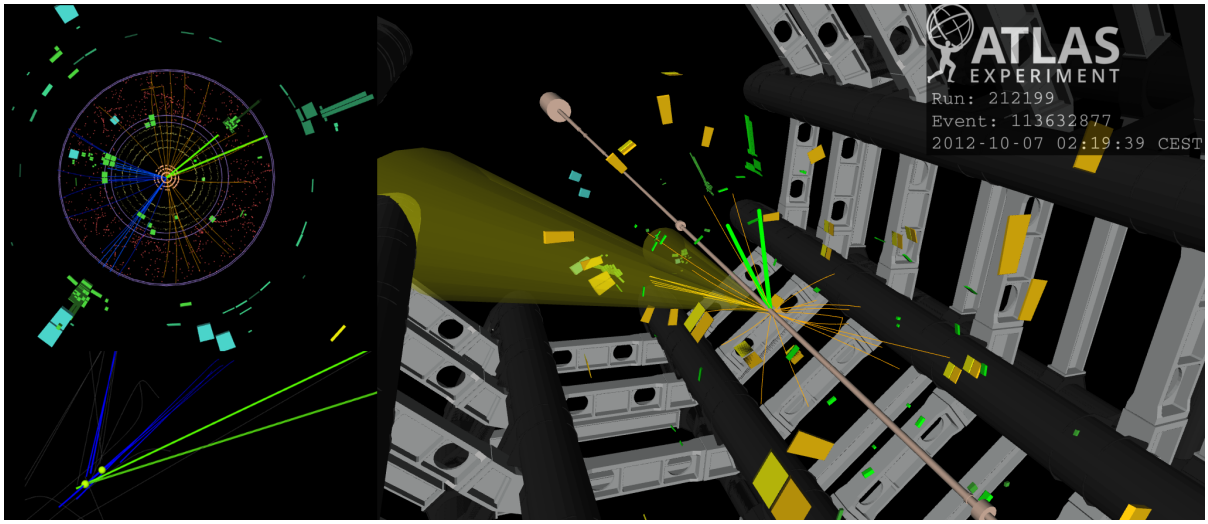


Figure VI.5: $ZH \to b\bar{b}e^{-}e^{+}$ candidate event with the highest signal probability recorded by ATLAS in the data of 2012. A pair of b-jets (yellow cones) recoils off an electron-positron pair (green tracks), which is compatible with being produced in a Z decay. In the transverse view, tracks assigned to the b-jets are shown in blue. Several tracks have significant impact parameters. One of the electron tracks in the transverse view is short because the $b\bar{b}e^{-}e^{+}$-system is boosted in forward direction, see main panel. Therefore, the electron does not reach the barrel calorimeters, and is detected in the endcap calorimeters. The event is categorised as 2 leptons, 2 jets, high p_{T}^{V}, TT b-tagging, and achieves the highest classifier score observed in the 2-lepton channel: 0.959. This corresponds to one signal event in 1.4 background events.

VI 2 Measurement of the Diboson Process

The likelihood model developed for searching $H \to b\bar{b}$ decays can be used to measure one more Standard Model process: the production of two vector bosons, of which one decays as $Z \to b\bar{b}$. If the other vector boson (either W or Z) decays leptonically, such events have the same signature as associated Higgs production with $H \to b\bar{b}$ decays. Cross sections, branching ratios and event kinematics differ since $m_Z < m_H$, and since the spin of Z is one instead of zero. However, by training boosted decision trees with Lorentz-invariant variables to select VZ instead of VH decays, the VH analysis strategy can be applied to the diboson process, and in this way be validated by interpreting data and simulations differently.

The diboson reactions are about eight times more likely than VH production. Cross sections times branching ratios are compared in table VI.2, which follow from tables IV.1 and IV.3 by applying the following branching-ratio corrections [135]:

$$\frac{\mathrm{BR}(Z \to b\bar{b})}{\mathrm{BR}(Z \to q\bar{q})} = \frac{0.15}{0.70} \qquad \text{and} \qquad \frac{\mathrm{BR}(V \to e, \mu \quad)}{\mathrm{BR}(V \to e, \mu, \tau)} = \frac{2}{3}.$$

These account for the fact that the diboson cross sections in table IV.1 include $Z \to q\bar{q}$ decays and final states with τ leptons. The WZ process is about eight, the ZZ process about nine times more likely than associated Higgs production. These are inclusive ratios, but due to differences in acceptance, these ratios also depend on b-tagging and p_{T}^V, see table VI.3. The dependence on p_{T}^V arises because the recoil of W/Z is different in VZ events: if W/Z recoil against Z instead of H, they recoil against a lighter particle. This leads to lower transverse momenta in diboson events. The largest cross-section difference is therefore observed in the low p_{T}^V regions. Furthermore, the VZ/VH ratios depend on the b-tagging category. This is because the $Z \to b\bar{b}$ branching fraction is 15 % instead of 58 % for $H \to b\bar{b}$. Since $Z \to c\bar{c}$ is much more likely than $H \to c\bar{c}$, decays to charm quarks contaminate the 80 % b-tagging category in the diboson analysis but not in the $H \to b\bar{b}$ analysis. Unlike the Higgs boson, the Z decays to c quarks with almost the same probability as to b quarks: to up-type quarks with about 12 % and to down-type quarks with 15 % branching fraction. In the LL b-tagging region, $Z \to c\bar{c}$ decays therefore increase the VZ/VH ratio to about 9 in the 1-lepton channel and to 12 in the 2-lepton channel. This effect wears off in the TT b-tagging region, where almost only $Z \to b\bar{b}$ decays contribute.

To conduct the diboson measurement, only two changes with respect to the $H \to b\bar{b}$ analysis have to be implemented: firstly, instead of training the BDTs to separate Higgs simulations from backgrounds, the BDTs are trained to separate diboson simulations from other backgrounds. These "backgrounds" also include the VH process. Secondly, the diboson normalisation nuisance parameter is promoted to a parameter of interest, whereas the Higgs normalisation is demoted to a nuisance parameter. The constraint

Table VI.2: Cross sections and braching ratios for associated Higgs production and diboson production (from tables IV.1 and IV.3).

	$\sigma \times$ BR
$WH \to l\nu b\bar{b}$	88 fb
$WZ \to l\nu b\bar{b}$	700 fb
$ZH \to l\bar{l}\, b\bar{b}$	19 fb
$ZZ \to l\bar{l}\, b\bar{b}$	171 fb

Table VI.3: VZ to VH ratios in dependence of b-tagging category and lepton channel. MM and TT are merged in the 2-lepton channel.

	VZ/VH	LL	MM	TT
1 L	low p_{T}^V	9.4	3.6	2.6
1 L	high p_{T}^V	6.5	2.4	1.7
2 L	low p_{T}^V	11.6	6.4	
2 L	high p_{T}^V	5.5	3.0	

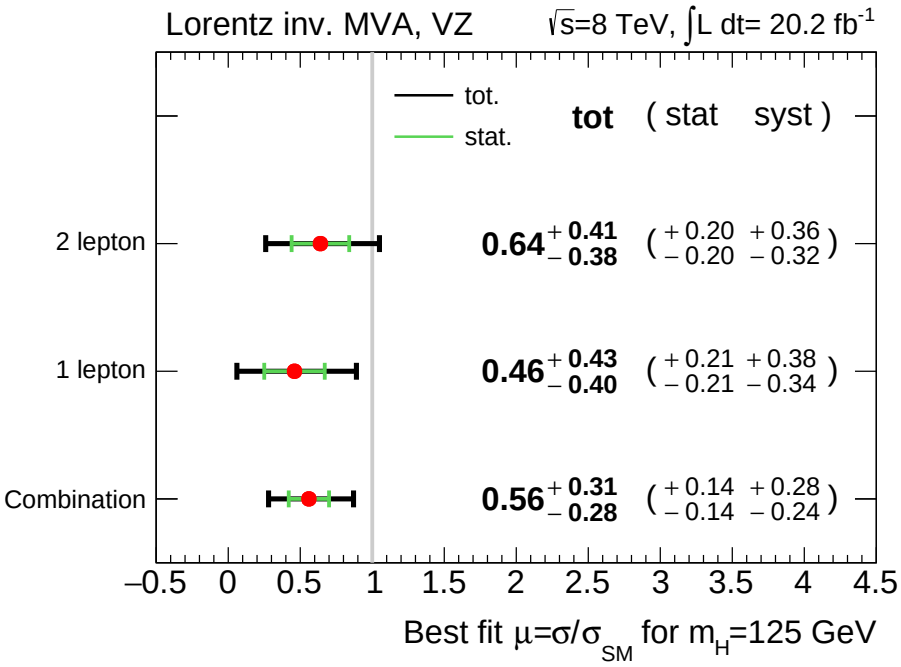


Figure VI.6: Signal strength of the Lorentz-invariant multivariate analysis trained to select the diboson process.

term for $H \to b\bar{b}$ was chosen conservatively as a $50\,\%$ normalisation uncertainty because $H \to b\bar{b}$ decays have not been observed with a significance of five standard deviations.

The Lorentz-invariant MVA and the selection of events are not changed, even though the selection could be optimised for the diboson processes. One such optimisation would be to adjust the transverse-momentum cuts for b-jets. Since the transverse momenta in Z decays are on average lower, the cut $p_{\mathrm{T}}^{j_1} > 45\,\mathrm{GeV}$ removes almost all Z decays where the Z has low transverse boost: from $m_Z = 91.2\,\mathrm{GeV}$ [135] one finds $p^{j_1} = 45.6\,\mathrm{GeV} \Rightarrow p_{\mathrm{T}}^{j_1} \leq 45\,\mathrm{GeV}$ if Z is at rest. On one hand, this significantly lowers the sensitivity of the diboson analysis, but on the other hand, it restricts the phase space to the same region as the $H \to b\bar{b}$ analysis. The latter is more important for a sanity check, which is why the selection cuts were not optimised. The same holds for an optimisation of the MVA: it would be possible, but it is not advisable because the VV analysis serves as a cross check of the Lorentz-invariant VH analysis.

Figure VI.7 on page 180 shows the response of BDTs trained to enhance the diboson process. The red overlay shows $VH \to Vb\bar{b}$ events as before, and the diboson signal is shown in grey. For a diboson classifier, the signal should peak at high classifier scores whereas the Higgs signal should be assigned low scores. The figure shows that this is indeed the case.

A likelihood model similar to the one described in section IV 6, but with diboson BDTs and μ_{VZ} as parameter of interest was used to measure the diboson signal strength in data. The result is:

$$\mu_{VZ} = 0.56^{+0.31}_{-0.28}.$$

The Standard Model prediction is within 1.4 standard deviations of the value of maximum likelihood. As for the VH measurement, a measurement with separate parameters of interest for the 1- and 2-lepton channels has been conducted, see figure VI.6. The results are:

$$\mu_{WZ}^{1L} = 0.5^{+0.4}_{-0.4} \quad \text{and} \quad \mu_{ZZ}^{2L} = 0.6^{+0.4}_{-0.4}.$$

The negative logarithmic-likelihood values of the separate and the combined measurement differ by 0.05. Performing a likelihood-ratio test using equation (IV.15) on page 107 yields a compatibility between the separate and the combined measurement of 75 %.

The measurement of μ_{VZ} is almost completely dominated by systematic uncertainties, in contrast to the VH measurement. This is due to the higher cross sections, which lead to higher statistical significance. Moreover, in contrast to the $H \to b\bar{b}$ case, the 2-lepton channel has higher sensitivity than the 1-lepton channel because the ratio of VZ to VH events is higher in the 2-lepton channel, see table VI.3. The highest-ranking systematic uncertainties are modelling uncertainties for the m_{bb} distribution in V+jets events, the measurement of W+bb, Z+bb and Z+cl normalisations and the cross-section and acceptance uncertainties of the diboson signal[2]. Pulls and normalisations of the diboson analysis and their impact on μ^{VZ} are shown in figure VI.8. They are consistent with the pulls observed in the $VH \to Vb\bar{b}$ analysis.

The discovery significance of the diboson signal is

$$Z_0^{VZ} = 2.0\,\sigma.$$

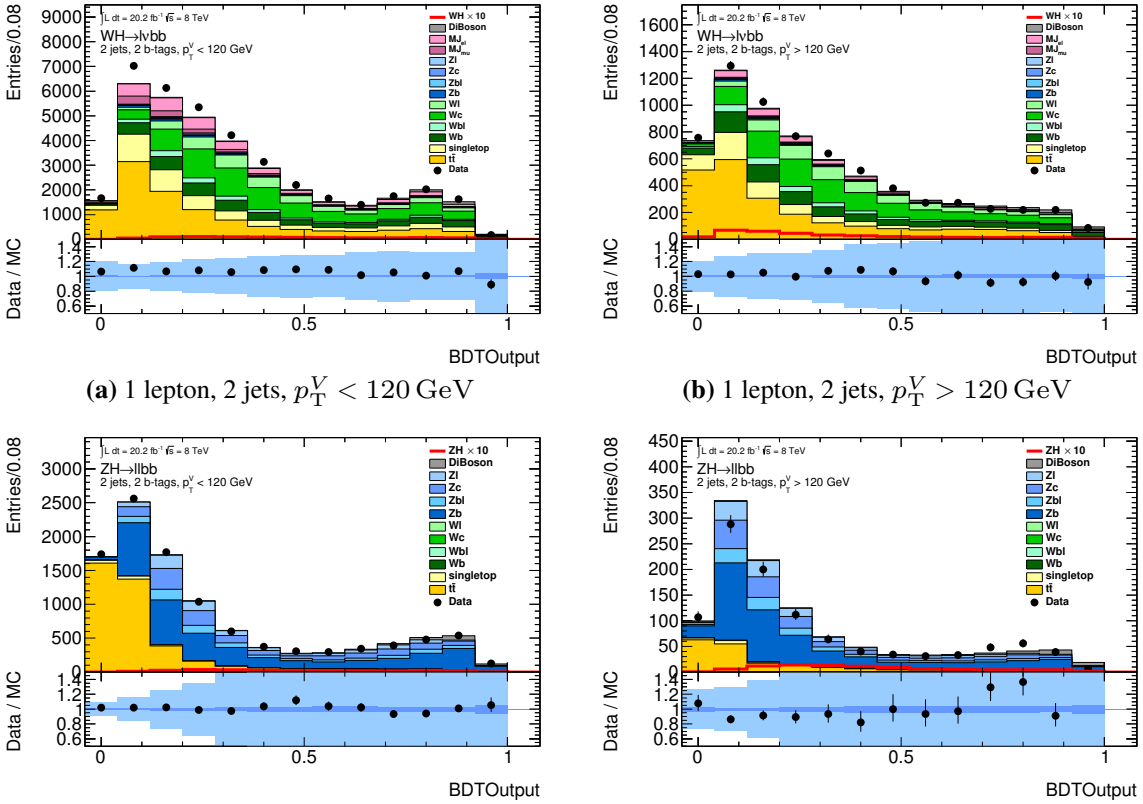
$3.5\,\sigma$ would have been expected from Monte Carlo predictions, but since μ_{VZ} is lower than the Standard Model prediction, the discovery significance is also lower.

VI 2.1 Comparison with ATLAS Diboson Results

The ATLAS Run-1 VZ analysis [1] finds $\mu_{VZ}^{1L} = 0.80^{+0.39}_{-0.35}$ and $\mu_{VZ}^{2L} = 0.67^{+0.25}_{-0.24}$. As for the VH analysis, these values cannot be compared directly:

- The ATLAS measurement combines the 7 and 8 TeV measurements whereas this work only uses the 8 TeV data. Small changes in central values and lower statistical uncertainties are therefore expected.
- The ATLAS likelihood model does not include extrapolation uncertainties for the measurement of V+bl normalisations in the 1b-tag region, see section V 2.3. The ATLAS likelihood model therefore underestimates the systematic uncertainty of background normalisations, and the mismodelling of the rightmost bin of the `MV1c` distributions is not treated.
- This explains the small difference of the measured μ_{VZ}^{2L}, and why the ATLAS uncertainties are lower.
- A second reason for lower uncertainties in the ATLAS result is the 0-lepton channel. It is included in the ATLAS fit, and slightly reduces background normalisation uncertainties.
- In the 1-lepton channel, a larger difference of central values is observed. However, a larger difference can also be expected: due to the mismodelling of the rightmost bin of the `MV1c` distributions, the ATLAS likelihood model does not include electron events with $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$. This region, however, has the highest contribution of WZ events, up to nine times more than WH events. The Lorentz-invariant analysis therefore probes different events than the ATLAS analysis, and can thus be expected to yield a different central value.
- The uncertainty in the 1-lepton channel is similar to the ATLAS result, but this is the result of two competing effects: on one hand, the higher signal statistics reduce the statistical uncertainty in the Lorentz-invariant analysis. On the other hand, this channel is most affected by the uncertainty of W+bl normalisations, which increases the systematic uncertainties of the Lorentz-invariant analysis. The net result is an uncertainty of the same magnitude.

[2] Technically, one can also interpret normalisation uncertainties as statistical uncertainties because they fall with growing datasets. In this case, the measurement is also limited by statistical uncertainties.



(a) 1 lepton, 2 jets, $p_{\mathrm{T}}^V < 120\,\mathrm{GeV}$ **(b)** 1 lepton, 2 jets, $p_{\mathrm{T}}^V > 120\,\mathrm{GeV}$

(c) 2 leptons, 2 jets, $p_{\mathrm{T}}^V < 120\,\mathrm{GeV}$ **(d)** 2 leptons, 2 jets, $p_{\mathrm{T}}^V > 120\,\mathrm{GeV}$

Figure VI.7: BDT classifiers with Lorentz-invariant variables and gradient boosting trained to enrich the diboson process (■). The Higgs signal is overlaid in red (■), and is mostly flat as expected when training to select diboson events. The preselection outlined in section IV 1 and shape corrections from section IV 2 are applied. Normalisation corrections for the background simulations are not applied except for the multijet template and for increasing the Z+jets normalisations by 16 % as before. The dark blue band (■) in the data/MC plot shows the statistical uncertainty of the background simulations. The light blue band (■) shows the quadratic sum of systematic uncertainties excluding normalisation uncertainties.

The diboson analysis can be seen as a sanity check for the VH analysis. It yields a result that is well compatible with Standard Model predictions: $\mu_{VZ}^{\mathrm{SM}} = 1$ is within $1.4\,\sigma$ or within the 92 % confidence interval of μ_{VZ}, and the p-value for rejecting the Standard Model hypothesis is $p_1 = 7.8\,\%$. Furthermore, differences to the ATLAS result are understood. The diboson sanity check therefore does not give rise to concerns regarding the $H \to b\bar{b}$ measurement.

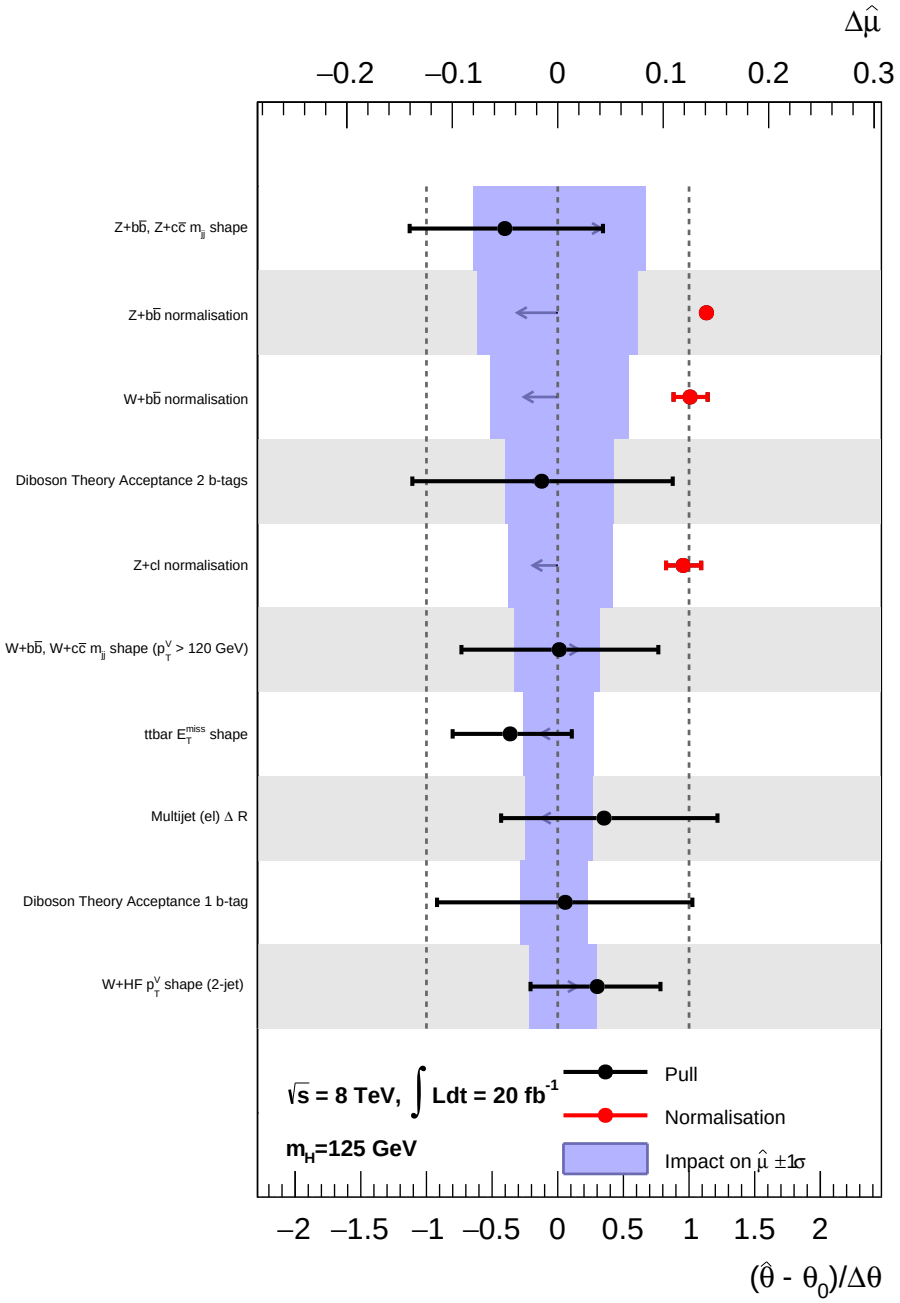


Figure VI.8: Nuisance-parameter pulls and impact on μ^{VZ} for the diboson analysis. The coloured bars show the $\pm 1\,\sigma$ impact on μ^{VZ} (top axis). The post-fit values of the nuisance parameters are overlaid with black markers (bottom axis). Arrows indicate the direction of the $+1\,\sigma$ uncertainties.

CHAPTER VII

Summary

In this thesis, two ways to search for the $H \to b\bar{b}$ decay of the Standard Model Higgs boson are presented. These focus on Higgs production in association with vector bosons that decay to charged leptons with $20.2\,\mathrm{fb}^{-1}$ of ATLAS data recorded in 2012 at a centre-of-mass energy of $\sqrt{s} = 8\,\mathrm{TeV}$.

The aspects discussed in chapter IV were important contributions to successfully implementing the first ATLAS analysis to use machine learning for $VH \to Vb\bar{b}$ searches. The use of boosted decision-tree classifiers for selecting $VH \to Vb\bar{b}$ events improved the expected significance of the $H \to b\bar{b}$ search by $32\,\%$ in comparison with the best ATLAS dijet mass analysis on the same dataset [1]. This analysis is also the first ATLAS analysis to use multivariate b-tagging at five working points simultaneously, which improves the sensitivity to $b\bar{b}$ final states by $15\,\%$. The sensitivity of the multivariate analysis therefore increased to $2.5\,\sigma$. With this, the ATLAS analysis is the most sensitive $VH \to Vb\bar{b}$ search published using data from LHC's Run 1. The corresponding search by the CMS collaboration reached an expected discovery significance of $2.1\,\sigma$ [6].

The analysis discussed in chapter V improves two important aspects of the ATLAS Run-1 search in final states with charged leptons: the first aspect is the multivariate analysis. The concept of BDT classifiers with Lorentz-invariant variables is applied to the 1- and 2-lepton channels, which leads to a reduction of systematic uncertainties by $16\,\%$. The total uncertainty of μ is reduced by $10\,\%$. It is also shown that systematic uncertainties reduce because the classifier responds less to shape changes in the distributions of kinematic observables. This reduces the impact of systematic uncertainties, and makes the classifier less sensitive to differences between background processes. Although such classifiers can be used less well for the measurement of background normalisations, this is an advantage for future $H \to b\bar{b}$ searches because background normalisations can be measured in background control regions, especially with larger datasets. The Lorentz-invariant classifiers therefore provide an interesting alternative for future $VH \to Vb\bar{b}$ searches, which will be limited by systematic uncertainties. This is because the data statistics recorded by the ATLAS experiment are already larger than for the LHC's Run 1, and are still increasing.

The second aspect that is improved is the statistical model. The ATLAS likelihood model is corrected to be able to describe an excess of simulated events with one b-tagged jet. This extension permits analysing the full $WH \to \ell\nu b\bar{b}$ and $ZH \to \ell\bar{\ell}b\bar{b}$ phase space, in contrast to the ATLAS Run-1 analysis, which excluded one-electron events with low p_{T}^{V}. The results obtained in this thesis explain why tensions that lead to the exclusion of such events were observed, and it is shown how to include these correctly. It is further shown that the signal strength measurements for the ATLAS Run-1 publication [1] yield similar values in spite of deficits of the likelihood model because the regions with the largest tensions were not analysed.

Finally, an excess of 2.4 standard deviations over the Standard Model background prediction is observed in the ATLAS data recorded in 2012. The ratio of the measured signal strength over the Standard Model prediction is $\mu_{VH} = 1.3^{+0.7}_{-0.6}$, and is fully consistent with the Standard Model.

VII 1 Outlook

Despite the large branching fraction of $58\,\%$, $H \to b\bar{b}$ remains a challenging decay, which has not been observed yet. With growing datasets, statistical uncertainties of $VH \to Vb\bar{b}$ searches will decrease, but systematic uncertainties will be the limiting factor for a discovery of this process at the level of five standard deviations. The concepts studied in this thesis may guide future $H \to b\bar{b}$ searches at the LHC. Lorentz-invariant classifiers are expected to outperform the standard classifiers also at the higher centre-of-mass energy achieved during the LHC's Run 2. The higher energy will increase the boost of the VH system and also of objects inside this system, but Lorentz-invariant quantities will not depend on this. Such classifiers might therefore perform even better than on Run-1 data if boost-related uncertainties are relevant for the classification result.

Moreover, with IBL installed, the sensitivity of b-tagging in ATLAS increased. A way to further improve the ATLAS Run-2 $VH \to Vb\bar{b}$ analysis is therefore to re-introduce b-tagging at multiple working points. However, this kind of b-tagging is not commissioned yet for Run-2 data. This thesis shows how to apply it in a multivariate analysis, and also shows possible problems that may arise in the determination of flavour fractions in V+jets events when very powerful b-tagging is used.

Five interesting aspects could not be considered in this work:

1. Events with $ZH \to b\bar{b}\nu\bar{\nu}$ decays, the "0-lepton channel", were not studied. Since the branching fraction of $Z \to \nu\bar{\nu}$ in comparison to $Z \to \mu^+\mu^- + e^+e^-$ is three times higher, this channel can improve the sensitivity significantly. However, of the 16 parameters necessary to fully describe the ZH system, four are unknown in the 0-lepton channel, instead of one parameter in the 1-lepton channel. Nonetheless, many of the strategies used to derive a minimal set of Lorentz-invariant discriminating variables would still apply.

2. Multivariate regression methods can be employed to improve the resolution of reconstructed quantities such as b-jet energies [141]. This would be a meaningful extension of the dijet-mass resolution correction discussed in section IV 2.2.

3. In recent years, deep neural networks evolved as a more powerful alternative to boosted decision trees. These outperform BDTs when several 'low-level observables' are used. These are observables that are processed less by reconstruction algorithms such as the energy distribution of a jet in the calorimeter, and are therefore less structured and harder to use for a multivariate training. Unlike for high-level observables such as the Lorentz invariants, where physicists select the best way to present information to the classifier, the training algorithms autonomously have to infer the best way to combine the information from low-level observables. It is, however, unclear if gains from more powerful classifiers outweigh the most likely higher susceptibility to systematic uncertainties. The results from section V 1 indicate that systematic uncertainties instead of the sensitivity of the classifiers might be the limiting factor for future $H \to b\bar{b}$ analyses. Nevertheless, the potential of deep neural networks should be explored.

4. Neural-network classifiers are an interesting option for a second reason because they can be trained to be insensitive to certain quantities [170, 171]. Such "pivotal quantities" can, for example, be event observables: a neural network can be trained to leave the invariant mass of two jets undistorted, which can be used to study an undistorted resonance peak. This is usually not possible because if a classifier is used to suppress background events, it distorts most observables. A more interesting application in light of searches limited by systematic uncertainties, however, is to make a classifier independent of the dominant systematic uncertainty. In general, the discriminative power of the

classifier decreases when this technique is applied, but the loss of (statistical) sensitivity might be outweighed by reduced systematic uncertainties. This is especially interesting with growing datasets because the loss of discriminative power can be compensated with higher data statistics.

5. Analyses in high-energy physics often face the problem that classifiers are trained on Monte Carlo simulations that do not accurately predict distributions in data. The training domain differs from the target domain ("covariate shift"), which causes the classifiers to perform worse on data because the distributions in data are not predicted correctly. Classifiers can be made robust against differences between training and target domain by using domain adaptation [172]. One technique of domain adaptation, for example, is to mix data events and simulations during the training. A first step in this direction was done by trying to detect a covariate shift using multivariate classifiers in section IV 4, but machine-learning literature indicates more potential in this field.

6. Similar to domain adaptation, classifiers can be trained entirely without (truth) labels [173]. This means that it is possible to train a classifier on data events, which by definition do not provide truth information. This allows suppressing backgrounds that can hardly be simulated such as the QCD multijet background *without* having to estimate these events with data-driven extrapolations. Such extrapolations might be biased, which was shown for the VH analysis in section IV 2.6. These biases would be learned by classifiers that are trained with the standard strategies. Data-driven extrapolations with corresponding uncertainties would still be necessary to evaluate the performance of the classifiers trained without truth labels, but their response would be biased less in the first place.

It is not a coincidence that the largest potential for future improvements of the $VH \to Vb\bar{b}$ analysis seems to be connected with machine learning: manual optimisations of the $VH \to Vb\bar{b}$ analysis in all three lepton channels have been performed from 2011 to 2015, reducing expected exclusion limits from $\mu < 4$ [2], to 1.9 [3] to 1.3 [4]. However, only the use of machine learning for signal selection and b-tagging improved the limit to 0.8 [1], below the Standard Model prediction. The dataset was *not* growing when machine learning was introduced.

After reducing the systematic uncertainties of BDT classifiers in the 1- and 2-lepton channels by another $16\,\%$ in this thesis, there seems to be little room for further improvements, unless novel strategies from the field of machine learning are explored.

Bibliography

[1] 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] (cit. on pp. 1, 61, 64, 69, 74, 77, 86, 87, 92, 94, 99, 100, 117, 124, 140, 147, 151, 152, 159, 161, 164, 166, 168, 174, 175, 179, 183, 185, 207).

[2] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a b-quark pair with the ATLAS detector*, Phys. Lett. B **718** (2012) 369, arXiv: 1207.0210 [hep-ex] (cit. on pp. 1, 84, 185).

[3] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to bottom quarks with the ATLAS detector*, ATLAS-CONF-2012-161, 2012, URL: https://cds.cern.ch/record/1493625 (cit. on pp. 1, 61, 84, 87, 94, 95, 159, 185).

[4] 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*, ATLAS-CONF-2013-079, 2013, URL: https://cds.cern.ch/record/1563235 (cit. on pp. 1, 61, 80, 84, 87, 94, 95, 159, 161, 175, 185).

[5] CMS Collaboration, *Search for the standard model Higgs boson decaying to bottom quarks in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Lett. B **710** (2012) 284, arXiv: 1202.4195 [hep-ex] (cit. on p. 1).

[6] CMS Collaboration, *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*, Phys. Rev. D **89** (2014) 012003, arXiv: 1310.3687 [hep-ex] (cit. on pp. 1, 183).

[7] J. Therhaag, *Search for the $H \to b\bar{b}$ decay mode of the Standard Model Higgs boson in associated production with a vector boson in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS experiment*, BONN-IR-2014-01, PhD thesis: U. Bonn, Phys. Inst., 2014, URL: http://inspirehep.net/record/1381417/files/3631.pdf (cit. on pp. 1, 61, 96, 173).

[8] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579 (cit. on p. 3).

[9] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264 (cit. on pp. 3, 20).

[10] A. Salam, "Weak and Electromagnetic Interactions", *Proceedings of the 8th Nobel symposium*, (Lerum, Sweden, 19th–25th May 1968), ed. by N. Svartholm, Conf. Proc. C680519, Almqvist & Wiskell, 1968 367, URL: http://inspirehep.net/record/53083 (cit. on p. 3).

[11] G. 't Hooft and M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. **B44** (1972) 189 (cit. on pp. 3, 51).

[12] H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, Phys. Lett. **47B** (1973) 365 (cit. on p. 3).

[13] C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. **C40**.10 (2016) 100001 (cit. on pp. 4, 38, 50, 51, 171, 175).

[14] D. Galbraith, *UX: Standard Model of the Standard Model*, 2012, URL: http://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model/ (visited on 17/05/2017) (cit. on p. 4).

[15] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*, New York, NY: Wiley, 2008, ISBN: 978-3527406012, URL: https://cds.cern.ch/record/111880 (cit. on pp. 6, 50).

[16] M. E. Peskin and D. V. Schroeder, *An introduction to Quantum Field Theory; 1995 ed.* Boulder, CO: Westview, 1995, ISBN: 978-0201503975, URL: https://cds.cern.ch/record/257493 (cit. on p. 6).

[17] A. Salam and J. C. Ward, *On a gauge theory of elementary interactions*, Il Nuovo Cimento (1955-1965) **19**.1 (1961) 165, ISSN: 1827-6121 (cit. on p. 7).

[18] S. F. Novaes, "Standard model: An Introduction", *Particles and fields. Proceedings, 10th Jorge Andre Swieca Summer School*, (Sao Paulo, Brazil, 6th–12th Feb. 1999), 1999 5, arXiv: hep-ph/0001283 [hep-ph] (cit. on pp. 7, 10, 17, 18).

[19] P. Skands, "Introduction to QCD", *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales (TASI 2012)*, (Boulder, Colorado, 4th–29th June 2012), 2013 341, arXiv: 1207.2389 [hep-ph] (cit. on pp. 7, 50, 51, 54, 56–58).

[20] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, Phys. Rept. **457** (2008) 1, arXiv: hep-ph/0503172 [hep-ph] (cit. on pp. 7, 202, 205).

[21] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531, [,648(1963)] (cit. on p. 14).

[22] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652 (cit. on p. 14).

[23] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*, Phys. Rev. **127** (3 1962) 965 (cit. on p. 16).

[24] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321 (cit. on p. 16).

[25] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12**.2 (1964) 132, ISSN: 0031-9163 (cit. on p. 16).

[26] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508 (cit. on p. 16).

[27] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Phys. Rev. **145** (1966) 1156 (cit. on p. 16).

[28] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585 (cit. on p. 16).

[29] T. Kibble, *Symmetry breaking in non-Abelian gauge theories*, Phys. Rev. **155** (1967) 1554 (cit. on p. 16).

[30] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv: 1503.07589 [hep-ex] (cit. on pp. 19, 22, 202).

[31] Y. Nambu and G. Jona-Lasinio, *Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I*, Phys. Rev. **122** (1961) 345 (cit. on p. 20).

[32] Y. Nambu and G. Jona-Lasinio, *Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. II*, Phys. Rev. **124** (1961) 246 (cit. on p. 20).

[33] 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] (cit. on pp. 22, 23).

[34] 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], cit. on p. 22.

[35] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. B **726** (2013) 120, arXiv: 1307.1432 [hep-ex] (cit. on p. 22).

[36] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC*, Phys. Lett. B **726** (2013) 88, arXiv: 1307.1427 [hep-ex] (cit. on p. 22), Erratum: Phys. Lett. B **734** (2014) 406.

[37] R. S. A. of Sciences, *The Nobel Prize in Physics 2013*, Press release, 2013, URL: https://www.nobelprize.org/nobel_prizes/physics/laureates/2013/press.html (visited on 04/09/2017) (cit. on p. 22).

[38] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Phys. Lett. **B579** (2004) 149, arXiv: hep-ph/0307206 [hep-ph] (cit. on p. 25).

[39] S. Heinemeyer et al., eds., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, 2013, arXiv: 1307.1347 [hep-ph] (cit. on pp. 24–27, 119).

[40] ATLAS and CMS Collaborations, *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] (cit. on pp. 24, 27, 28).

[41] D. de Florian et al., eds., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 2016, arXiv: 1610.07922 [hep-ph] (cit. on pp. 24, 27, 69).

[42] M. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, Phys. Rev. **D68** (2003) 073003, arXiv: hep-ph/0306234 [hep-ph] (cit. on p. 26).

[43] T. Sjöstrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on pp. 26, 50, 57, 65).

[44] ATLAS Collaboration, *Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector*, JHEP **12** (2017) 024, arXiv: 1708.03299 [hep-ex] (cit. on pp. 26, 149, 168).

[45] CMS Collaboration, *Evidence for the Higgs boson decay to a bottom quark-antiquark pair*, 2017, arXiv: 1709.07497 [hep-ex] (cit. on p. 26).

[46] ATLAS Collaboration, *Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector*, Phys. Rev. D **92** (2015) 012006, arXiv: 1412.2641 [hep-ex] (cit. on p. 26).

[47] ATLAS Collaboration, *Measurement of fiducial differential cross sections of gluon-fusion production of Higgs bosons decaying to $WW^* \to e\nu\mu\nu$ with the ATLAS detector at $\sqrt{s} = 8$ TeV*, JHEP **08** (2016) 104, arXiv: 1604.02997 [hep-ex] (cit. on p. 26).

[48] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, Phys. Rev. D **91** (2015) 012006, arXiv: 1408.5191 [hep-ex] (cit. on p. 26).

[49] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **04** (2015) 117, arXiv: 1501.04943 [hep-ex] (cit. on p. 26).

[50] CMS Collaboration, *Observation of the Higgs boson decay to a pair of τ leptons*, Submitted to Phys. Lett. B.: CERN, 2017, arXiv: 1708.00373 [hep-ex], URL: http://cds.cern.ch/record/2276465 (cit. on p. 26).

[51] ATLAS Collaboration, *Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, Phys. Rev. D **90** (2014) 112015, arXiv: 1408.7084 [hep-ex] (cit. on p. 26).

[52] ATLAS Collaboration, *Search for Higgs boson decays to a photon and a Z boson in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector*, Phys. Lett. B **732** (2014) 8, arXiv: 1402.3051 [hep-ex] (cit. on p. 27).

[53] ATLAS Collaboration, *Search for the Standard Model Higgs boson decay to $\mu^+\mu^-$ with the ATLAS detector*, Phys. Lett. B **738** (2014) 68, arXiv: 1406.7663 [hep-ex] (cit. on p. 27).

[54] J. Ellis and T. You, *Updated Global Analysis of Higgs Couplings*, JHEP **06** (2013) 103, arXiv: 1303.3879 [hep-ph] (cit. on p. 28).

[55] P. Mouche, *Overall view of the LHC. Vue d'ensemble du LHC* (2014), General Photo, URL: https://cds.cern.ch/record/1708847 (cit. on p. 30).

[56] O. S. Brüning et al., eds., *LHC Design Report. 1. The LHC Main Ring*, CERN-2004-003-V-1, CERN-2004-003, Geneva, 2004, URL: https://cdsweb.cern.ch/record/782076 (cit. on p. 29).

[57] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 30, 34).

[58] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 30).

[59] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005 (cit. on p. 30).

[60] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002 (cit. on p. 30).

[61] A. Valishev et al., "Tevatron End-of-Run Beam Physics Experiments", *Proceedings, 3rd International Conference on Particle accelerator (IPAC 2012)*, (New Orleans, USA, 2nd–25th May 2012), Conf. Proc. C1205201, 2012 2128 (cit. on p. 31).

[62] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) 2518, arXiv: 1302.4393 [hep-ex] (cit. on p. 31).

[63] A. D. Martin et al., *Parton distributions for the LHC*, Eur.Phys.J.C **63** (2009) 189, eprint: 0901.0002 (cit. on pp. 32, 33, 58).

[64] W. J. Stirling, *Standard Model cross sections as a function of collider energy, with 125 GeV Higgs*, private communication, 2012, URL: http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html (visited on 11/07/2016) (cit. on pp. 32, 64).

[65] J. Pequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: https://cds.cern.ch/record/1095924 (cit. on p. 34).

[66] JabberWok, *Pseudorapidity2*, Wikipedia-en, CC BY-SA 3.0, 2007, URL: https://commons.wikimedia.org/w/index.php?curid=5970403 (visited on 23/06/2017) (cit. on p. 34).

[67] ATLAS Collaboration, *ATLAS Publication Committee Home Page*, Restricted to ATLAS, 2017, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/PubComHome (visited on 23/06/2017) (cit. on p. 34).

[68] ATLAS Collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-018, 2015, URL: https://cds.cern.ch/record/2037683 (cit. on p. 36).

[69] J. Pequenao, "Computer generated image of the ATLAS inner detector", 2008, URL: https://cds.cern.ch/record/1095926 (cit. on p. 36).

[70] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. J. C **70** (2010) 787, arXiv: 1004.5293 [hep-ex] (cit. on p. 36).

[71] J. Pequenao, "Computer Generated image of the ATLAS calorimeter", 2008, URL: https://cds.cern.ch/record/1095927 (cit. on p. 37).

[72] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*, tech. rep., CERN, 1996, URL: https://cds.cern.ch/record/331061 (cit. on p. 38).

[73] ATLAS Collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, Eur. Phys. J. C **70** (2010) 1193, arXiv: 1007.5423 [hep-ex] (cit. on p. 38).

[74] ATLAS Collaboration, *The ATLAS Tile Calorimeter*, tech. rep. ATL-TILECAL-PROC-2015-002, CERN, 2015, URL: https://cds.cern.ch/record/2004868 (cit. on p. 39).

[75] ATLAS Collaboration, *Jet energy resolution in proton–proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. C **73** (2013) 2306, arXiv: 1210.6210 [hep-ex] (cit. on pp. 39, 123).

[76] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, Eur. Phys. J. C **74** (2014) 3130, arXiv: 1407.3935 [hep-ex] (cit. on pp. 40, 44).

[77] 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] (cit. on p. 41).

[78] ATLAS Collaboration, *Search for Higgs boson production via weak boson fusion and decaying to $b\bar{b}$ in association with a high-energy photon using the ATLAS detector*, ATLAS-CONF-2016-063, 2016, URL: https://cds.cern.ch/record/2206201 (cit. on p. 41).

[79] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, JINST **9** (2014) P09009, arXiv: 1406.7690 [hep-ex] (cit. on p. 42).

[80] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in High Pile-Up LHC Environment*, ATLAS-CONF-2012-042, 2012, URL: https://cds.cern.ch/record/1435196 (cit. on p. 42).

[81] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton–proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment*, ATLAS-CONF-2010-069, 2010, URL: https://cds.cern.ch/record/1281344 (cit. on p. 42).

[82] J. Pequenao, "Event Cross Section in a computer generated image of the ATLAS detector.", 2008, URL: https://cds.cern.ch/record/1096081 (cit. on p. 43).

[83] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data*, Eur. Phys. J. C **77** (2017) 195, arXiv: 1612.01456 [hep-ex] (cit. on pp. 43, 44).

[84] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton–proton collision data*, ATLAS-CONF-2014-032, 2014, URL: https://cds.cern.ch/record/1706245 (cit. on p. 43).

[85] ATLAS Collaboration, *Description and Performance of the Electron Likelihood Tool at ATLAS using 2012 LHC Data*, internal report ATL-COM-PHYS-2013-378, CERN, 2013, URL: https://cds.cern.ch/record/1537410 (cit. on p. 44).

[86] ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. C **75** (2015) 303, arXiv: 1412.7086 [hep-ex] (cit. on p. 45).

[87] 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] (cit. on p. 45).

[88] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, URL: https://cds.cern.ch/record/1099735 (cit. on p. 45).

[89] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C **73** (2013) 2304, arXiv: 1112.6426 [hep-ex] (cit. on pp. 45, 46).

[90] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. B **641** (2006) 57, arXiv: hep-ph/0512210 (cit. on p. 45).

[91] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 45).

[92] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 17, arXiv: 1406.0076 [hep-ex] (cit. on p. 46).

[93] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction in proton–proton collisions at $\sqrt{s} = 7$ TeV with ATLAS*, Eur. Phys. J. C **72** (2012) 1844, arXiv: 1108.5602 [hep-ex] (cit. on pp. 47, 124).

[94] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton–Proton Collisions recorded in 2012 at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2013-082, 2013, URL: https://cds.cern.ch/record/1570993 (cit. on p. 47).

[95] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex] (cit. on p. 48).

[96] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046, 2014, URL: https://cds.cern.ch/record/1741020 (cit. on pp. 48, 79, 123).

[97] G. P. Lepage, *A new algorithm for adaptive multidimensional integration*, Journal of Computational Physics **27** (1978) 192 (cit. on p. 50).

[98] F. James, *Monte Carlo theory and practice*, Reports on Progress in Physics **43**.9 (1980) 1145, URL: http://stacks.iop.org/0034-4885/43/i=9/a=002 (cit. on p. 50).

[99] A. Buckley et al., *General-purpose event generators for LHC physics*, Phys. Rept. **504** (2011) 145, arXiv: 1101.2599 [hep-ph] (cit. on pp. 50, 55).

[100] X.-G. Wu, S. J. Brodsky and M. Mojaza, *The Renormalization Scale-Setting Problem in QCD*, Prog. Part. Nucl. Phys. **72** (2013) 44, arXiv: 1302.0599 [hep-ph] (cit. on p. 52).

[101] J. C. Collins and D. E. Soper, *Parton distribution and decay functions*, Nuclear Physics B **194**.3 (1982) 445, ISSN: 0550-3213 (cit. on p. 54).

[102] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nuclear Physics B **126**.2 (1977) 298, ISSN: 0550-3213; V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438, [Yad. Fiz.15,781(1972)]; 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, [Zh. Eksp. Teor. Fiz.73,1216(1977)], cit. on pp. 54, 55.

[103] T. Sjöstrand, "Monte Carlo Generators", *High-energy physics. Proceedings, European School*, (Aronsborg, Sweden, 18th June–1st July 2006), 2006 51, arXiv: hep-ph/0611247 [hep-ph] (cit. on p. 55).

[104] T. Kinoshita, *Mass singularities of Feynman amplitudes*, J. Math. Phys. **3** (1962) 650 (cit. on p. 56).

[105] T. D. Lee and M. Nauenberg, *Degenerate Systems and Mass Singularities*, Phys. Rev. **133** (1964) B1549 (cit. on p. 56).

[106] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph]; S. Höche et al., *QCD matrix elements and truncated showers*, JHEP **05** (2009) 053, arXiv: 0903.1219 [hep-ph]; T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph]; S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph], cit. on pp. 52, 53, 56, 57, 65.

[107] G. Corcella et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **01** (2001) 010, arXiv: hep-ph/0011363 (cit. on pp. 56, 57).

[108] M. H. Seymour, *A Simple prescription for first order corrections to quark scattering and annihilation processes*, Nucl. Phys. **B436** (1995) 443, arXiv: hep-ph/9410244 [hep-ph] (cit. on p. 56).

[109] M. H. Seymour, *Matrix element corrections to parton shower algorithms*, Comput. Phys. Commun. **90** (1995) 95, arXiv: hep-ph/9410414 [hep-ph] (cit. on p. 56).

[110] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv: hep-ph/0603175 (cit. on p. 57).

[111] ATLAS Collaboration, *Measurement of underlying event characteristics using charged particles in pp collisions at $\sqrt{s} = 900$ GeV and 7 TeV with the ATLAS detector*, Phys. Rev. D **83** (2011) 112001, arXiv: 1012.0791 [hep-ex] (cit. on p. 58).

[112] ATLAS Collaboration, *Example ATLAS tunes of Pythia8, Pythia6 and Powheg to an observable sensitive to Z boson transverse momentum*, ATL-PHYS-PUB-2013-017, 2013, URL: https://cds.cern.ch/record/1629317; ATLAS Collaboration, *ATLAS tunes of Pythia 6 and Pythia 8 for MC11*, ATL-PHYS-PUB-2011-009, 2011, URL: https://cds.cern.ch/record/1363300, cit. on pp. 58, 65.

[113] 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], cit. on pp. 31, 58, 124.

[114] S. Agostinelli et al., *Geant4 – a simulation toolkit*, Nucl. Instr. Meth. Phys. Res. A **506**.3 (2003) 250, ISSN: 0168-9002 (cit. on p. 58).

[115] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [hep-ex] (cit. on p. 59).

[116] ATLAS Collaboration, *Measurement of differential production cross-sections for a Z boson in association with b-jets in 7 TeV proton–proton collisions with the ATLAS detector*, JHEP **10** (2014) 141, arXiv: 1407.3643 [hep-ex] (cit. on p. 63).

[117] J. Campbell and R. K. Ellis, *Next-to-leading order corrections to $W + 2$ jet and $Z + 2$ jet production at hadron colliders*, Phys. Rev. D **65** (11 2002) 113007 (cit. on p. 63).

[118] ATLAS Collaboration,
Evidence for single top-quark production in the s-channel in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector using the Matrix Element Method,
Phys. Lett. B **756** (2016) 228, arXiv: 1511.05980 [hep-ex] (cit. on p. 63).

[119] ATLAS Collaboration, *Measurement of the production cross-section of a single top quark in association with a W boson at 8 TeV with the ATLAS experiment*, JHEP **01** (2016) 064,
arXiv: 1510.03752 [hep-ex] (cit. on p. 63).

[120] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework",
New computing techniques in physics research V. Proceedings, 5th International Workshop, AIHENP '96, (Lausanne, Switzerland, 2nd–6th Sept. 1996), Nucl. Instrum. Meth. A389, 1997 81,
URL: https://doi.org/10.1016/S0168-9002(97)00048-X (cit. on pp. 62, 87, 102).

[121] ATLAS Collaboration, *Measurement of dijet cross sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector*, JHEP **05** (2014) 059,
arXiv: 1312.3524 [hep-ex] (cit. on p. 64).

[122] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*,
JHEP **11** (2004) 040, arXiv: hep-ph/0409146; 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]; S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*,
JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph], cit. on pp. 65, 75.

[123] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*,
Comput. Phys. Commun. **184** (2013) 919, arXiv: hep-ph/0405247 [hep-ph] (cit. on p. 65).

[124] E. Re,
Single-top Wt-channel production matched with parton showers using the POWHEG method,
Eur. Phys. J. **C71** (2011) 1547, arXiv: 1009.2450 [hep-ph] (cit. on p. 65).

[125] S. Alioli et al.,
NLO single-top production matched with shower in POWHEG: s- and t-channel contributions,
JHEP **09** (2009) 111, [Erratum: JHEP02,011(2010)], arXiv: 0907.4076 [hep-ph]
(cit. on p. 65).

[126] T. Melia et al., *W+W-, WZ and ZZ production in the POWHEG BOX*, JHEP **11** (2011) 078,
arXiv: 1107.5051 [hep-ph]; P. Nason and G. Zanderighi,
W^+W^-, WZ and ZZ production in the POWHEG-BOX-V2, Eur. Phys. J. **C74**.1 (2014) 2702,
arXiv: 1311.1365 [hep-ph], cit. on p. 65.

[127] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024,
arXiv: 1007.2241 [hep-ph] (cit. on pp. 64, 65).

[128] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*,
Phys. Rev. **D82** (2010) 074018, arXiv: 1005.3457 [hep-ph] (cit. on p. 64).

[129] 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] (cit. on p. 64).

[130] K. Melnikov and F. Petriello,
*Electroweak gauge boson production at hadron colliders through O(alpha(s)**2)*,
Phys. Rev. **D74** (2006) 114017, arXiv: hep-ph/0609070 [hep-ph] (cit. on p. 65).

[131] M. Czakon, P. Fiedler and A. Mitov,
Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$,
Phys. Rev. Lett. **110** (2013) 252004, arXiv: 1303.6254 [hep-ph] (cit. on p. 65).

[132] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. **D83** (2011) 091503, arXiv: 1103.2792 [hep-ph] (cit. on p. 65).

[133] N. Kidonakis, *NNLL resummation for s-channel single top quark production*,
Phys. Rev. **D81** (2010) 054028, arXiv: 1001.5034 [hep-ph] (cit. on p. 65).

[134] N. Kidonakis,
Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-,
Phys. Rev. **D82** (2010) 054018, arXiv: 1005.4451 [hep-ph] (cit. on p. 65).

[135] K. A. Olive et al., *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001
(cit. on pp. 69, 177, 178).

[136] ATLAS Collaboration, *BTaggingBenchmarks*, Restricted to ATLAS, 2016, URL:
https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTaggingBenchmarks
(cit. on pp. 73, 79, 80).

[137] S. Alioli et al., *NLO Higgs boson production via gluon fusion matched with shower in POWHEG*,
JHEP **04** (2009) 002, arXiv: 0812.0578 [hep-ph] (cit. on p. 75).

[138] A. Denner et al., "EW corrections to Higgs strahlung at the Tevatron and the LHC with HAWK",
Proceedings, 21st International Europhysics Conference on High energy physics (EPS-HEP 2011), (Grenoble, France, 21st–27th July 2011), PoS EPS-HEP, 2011 235,
arXiv: 1112.5258 [hep-ph] (cit. on p. 75).

[139] $H \to b\bar{b}$ working group in ATLAS,
Invariant Mass Studies for the $H \to b\bar{b}$ Measurements for HCP,
internal report ATL-COM-PHYS-2012-1451, CERN, 2012,
URL: https://cds.cern.ch/record/1482360 (cit. on pp. 77, 123).

[140] H. Ritter, *Improvement of the Higgs mass resolution in the $ZH \to l^+l^-b\bar{b}$ channel using a Kinematic Fit with constraints*, BONN-IB-2014-14, Universität Bonn, 2014,
URL: https://www.hep1.physik.uni-bonn.de/results/theses (cit. on p. 77).

[141] E. Schopf, *Improvement of the Higgs Mass Resolution in the Decay Channel $H \to b\bar{b}$ in Associated Production with a W Boson at the ATLAS Detector Using Multivariate Regression*,
BONN-IB-2014-07, Universität Bonn, 2014,
URL: https://www.hep1.physik.uni-bonn.de/results/theses (cit. on pp. 77, 184).

[142] ATLAS Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 77).

[143] S. Höche et al., *QCD matrix elements + parton showers. The NLO case*, JHEP **2013**.4 (2013) 1,
ISSN: 1029-8479 (cit. on p. 77).

[144] ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004, 2014, URL: https://cds.cern.ch/record/1664335 (cit. on pp. 79, 123).

[145] ATLAS Collaboration, *Measurements of normalized differential cross-sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. D **90** (2014) 072004, arXiv: 1407.0371 [hep-ex] (cit. on pp. 80, 81).

[146] ATLAS Collaboration, *Search for the Standard Model Higgs boson in associated production with a vector boson and decaying to bottom quarks with the ATLAS detector*, internal report ATL-COM-PHYS-2012-1081, CERN, 2012, URL: https://cds.cern.ch/record/1463051 (cit. on p. 81).

[147] A. Hoecker et al., "TMVA - Toolkit for Multivariate Data Analysis", *Proceedings, 11th International Workshop on Advanced computing and analysis techniques in physics research*, (Amsterdam, Netherlands, 23rd–27th Apr. 2007), PoS ACAT, 2007 040, arXiv: physics/0703039 [PHYSICS] (cit. on p. 87).

[148] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, J. Comput. Syst. Sci. **55**.1 (1997) 119, ISSN: 0022-0000, URL: http://dx.doi.org/10.1006/jcss.1997.1504 (cit. on p. 89).

[149] H. Shimodaira, *Improving predictive inference under covariate shift by weighting the log-likelihood function*, Journal of Statistical Planning and Inference **90**.2 (2000) 227, ISSN: 0378-3758 (cit. on p. 100).

[150] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", *Proceedings of Computing in High Energy and Nuclear Physics*, (La Jolla, California, 24th–28th Mar. 2003), eConf C0303241, 2003 MOLT007, arXiv: physics/0306116 [physics] (cit. on p. 102).

[151] L. Moneta et al., "The RooStats Project", *Proceedings of the 13th International Workshop on Advanced computing and analysis techniques in physics research*, (Jaipur, India, 22nd–27th Feb. 2010), PoS ACAT, 2010 057, arXiv: 1009.1003 [physics.data-an] (cit. on p. 102).

[152] K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep. CERN-OPEN-2012-016, CERN, 2012, URL: https://cds.cern.ch/record/1456844 (cit. on pp. 102, 103).

[153] W. Verkerke, "Guide to parameterized likelihood analyses", Restricted to ATLAS, 2012, URL: svn.cern.ch/reps/atlasgrp/Physics/StatForum/FrequentistRecommendation/doc/profiling_recommendations_draft.pdf (visited on 25/07/2017), Draft (cit. on p. 102).

[154] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, [Erratum: Eur. Phys. J.C73,2501(2013)], arXiv: 1007.1727 [physics.data-an] (cit. on pp. 107–109, 111, 112).

[155] 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 (cit. on p. 107).

[156] S. S. Wilks,
The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses,
Ann. Math. Statist. **9**.1 (1938) 60 (cit. on pp. 107, 108).

[157] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society **54**.3 (1943) 426, ISSN: 00029947 (cit. on p. 108).

[158] A. L. Read, *Presentation of search results: The CL(s) technique*, J. Phys. **G28** (2002) 2693 (cit. on p. 113).

[159] T. Junk, *Confidence level computation for combining searches with small statistics*,
Nucl. Instrum. Meth. **A434** (1999) 435, arXiv: hep-ex/9902006 [hep-ex] (cit. on p. 113).

[160] A. L. Read, "Modified frequentist analysis of search results (The CL(s) method)",
Workshop on confidence limits, (CERN, Geneva, Switzerland, 17th–18th Jan. 2000), 2000 81,
URL: http://cds.cern.ch/record/451614 (cit. on p. 113).

[161] R. J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*,
Manchester physics series, Chichester: Wiley, 1989 49, ISBN: 0-471-92295-1 (cit. on p. 116).

[162] K.-P. Ho, *Central Limit for the Product of Free Random Variables*, 2011,
arXiv: arXiv:1101.5220 (cit. on p. 116).

[163] N. Kidonakis, "Differential and total cross sections for top pair and single top production",
Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012), (Bonn, Germany), 2012 831, arXiv: 1205.3453 [hep-ph] (cit. on p. 120).

[164] ATLAS Collaboration, *Jet energy measurement and systematic uncertainties using tracks for jets and for b-quark jets produced in proton–proton collisions at* $\sqrt{s} = 7$ *TeV in the ATLAS detector*,
ATLAS-CONF-2013-002, 2013, URL: https://cds.cern.ch/record/1504739
(cit. on p. 123).

[165] J. H. Friedman, *Greedy function approximation: A gradient boosting machine*,
Ann. Statist. **29**.5 (2001) 1189 (cit. on p. 137).

[166] J. Friedman, T. Hastie and R. Tibshirani, *Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors)*, Ann. Statist. **28**.2 (2000) 337
(cit. on p. 138).

[167] J. H. Friedman, "Proceedings of the 1974 CERN School of Computing", ed. by G. R. Macleod,
Godøysund, Norway, 1974 292, URL: https://cds.cern.ch/record/186223
(cit. on p. 153).

[168] B. L. Welch, *The Significance of the Difference Between two Means When the Population Variances are Unequal*, Biometrika **29**.3-4 (1938) 350 (cit. on p. 153).

[169] F. Ahmadov et al., *Supporting Document for the Search for the bb decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector*,
internal report ATL-COM-PHYS-2014-051, CERN, 2014,
URL: https://cds.cern.ch/record/1645654 (cit. on p. 167).

[170] I. J. Goodfellow et al., *Generative Adversarial Networks*, 2014, arXiv: 1406.2661 [stat.ML]
(cit. on p. 184).

[171] G. Louppe, M. Kagan and K. Cranmer, *Learning to Pivot with Adversarial Networks* (2016),
arXiv: 1611.01046 [stat.ME] (cit. on p. 184).

[172] M. Chen, K. Q. Weinberger and J. C. Blitzer, "Co-training for Domain Adaptation", *Proceedings of the 24th International Conference on Neural Information Processing Systems*, (Granada, Spain), ed. by J. Shawe-Taylor et al., NIPS'11, USA: Curran Associates Inc., 2011 2456, ISBN: 978-1-61839-599-3, URL: http://papers.nips.cc/paper/4433-co-training-for-domain-adaptation (cit. on p. 185).

[173] E. M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: Learning from mixed samples in high energy physics* (2017), arXiv: 1708.02949 [hep-ph] (cit. on p. 185).

[174] G. Degrassi et al., *Higgs mass and vacuum stability in the Standard Model at NNLO*, JHEP **08** (2012) 098, arXiv: 1205.6497 [hep-ph] (cit. on p. 204).

[175] U. Baur et al., *The Charm content of W + 1 jet events as a probe of the strange quark distribution function*, Phys. Lett. **B318** (1993) 544, arXiv: hep-ph/9308370 [hep-ph] (cit. on p. 207).

[176] W. J. Stirling and E. Vryonidou, *Charm Production in Association with an Electroweak Gauge Boson at the LHC*, Phys. Rev. Lett. **109** (2012) 082002, arXiv: 1203.6781 [hep-ph] (cit. on p. 207).

[177] ATLAS Collaboration, *Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **05** (2014) 068, arXiv: 1402.6263 [hep-ex] (cit. on p. 207).

APPENDIX A

Notes on the Introduction

A 1 Notes on the Discussion of Gauge Invariance

The Neutral Part of the Lepton Lagrangian

Writing out the R, L shorthands for the photon terms, the neutral part of the Lagrangian becomes

$$\begin{aligned}
\mathscr{L}^{\text{Neutral}}_{\text{Leptons}} = & -\left[\frac{g}{2}\sin(\theta_W) \qquad \left(-\bar{\nu}_L\gamma^\mu\nu_L - \bar{\ell}_L\gamma^\mu\ell_L - 2\bar{\ell}_R\gamma^\mu\ell_R + \bar{\nu}_L\gamma^\mu\nu_L - \bar{\ell}_L\gamma^\mu\ell_L\right)\right] A_\mu \\
& -\left[\frac{g'}{2}(-\sin(\theta_W))\left(\bar{L}\gamma^\mu Y L + \bar{R}\gamma^\mu Y R\right) + \frac{g}{2}\cos(\theta_W)\left(\bar{L}\gamma^\mu 2T_3 L\right)\right] Z_\mu \\
= & -g\sin(\theta_W)\,\bar{\ell}\gamma^\mu\ell\,A_\mu \\
& -\frac{g}{2\cos(\theta_W)}\left[-\sin^2(\theta_W)\left(\bar{L}\gamma^\mu Y L + \bar{R}\gamma^\mu Y R\right) + (1-\sin^2(\theta_W))\left(\bar{L}\gamma^\mu 2T_3 L\right)\right] Z_\mu \\
= & -g\sin(\theta_W)\,\bar{\ell}\gamma^\mu\ell\,A_\mu \\
& -\frac{g}{2\cos(\theta_W)}\left[\bar{L}\gamma^\mu 2T_3 L - \sin^2(\theta_W)\left(\bar{L}\gamma^\mu 2T_3 L + \bar{L}\gamma^\mu Y L + \bar{R}\gamma^\mu Y R\right)\right] Z_\mu \\
= & -g\sin(\theta_W)\,\bar{\ell}\gamma^\mu\ell\,A_\mu \\
& -\frac{g}{2\cos(\theta_W)}\sum_{i=\nu_L,\ell_L,\ell_R}\bar{\psi}_i\gamma^\mu\left[2T_3\,P_L - \sin^2(\theta_W)\,(2T_3+Y)\right]\psi_i\,Z_\mu \\
= & -g\sin(\theta_W)\,\bar{\ell}\gamma^\mu\ell\,A_\mu \\
& -\frac{g}{2\cos(\theta_W)}\sum_{i=\nu,\ell_L,\ell_R}\bar{\psi}_i\gamma^\mu(g_V - g_A\gamma^5)\psi_i\,Z_\mu
\end{aligned} \tag{A.1}$$

$T_3 = \frac{1}{2}\tau_3$ is the operator for the 3rd component of isospin, which acts only on the $SU(2)$ doublets. For the $U(1)$ singlets, it is understood to be zero. For the Z^0 terms, the *V-A* structure of weak interactions has been emphasised by inserting $P_L L = \mathbb{1}L$ into the left-handed currents.

The Full Standard Model Lagrangian

$$\begin{aligned}
\mathscr{L}_{\text{boson}} &= \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}(\partial^\mu G^\nu_a - \partial^\nu G^\mu_a)(\partial_\mu G^a_\nu - \partial_\nu G^a_\mu) \\
&\quad - \frac{1}{2}W^+_{\mu\nu}W^{-\,\mu\nu} + m_W^2 W^+_\mu W^{-\,\mu} - \frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu \\
&\quad + \frac{1}{2}\partial_\mu H \partial^\mu H - \frac{1}{2}m_H^2 H^2 \\
&\quad + \boxed{ggg} + \boxed{gggg} \\
&\quad + \boxed{W^+W^-A} + \boxed{W^+W^-AA} + \boxed{W^+W^-Z} + \boxed{W^+W^-ZZ} + \boxed{W^+W^-AZ} \\
&\quad + \boxed{W^+W^-W^+W^-} \\
&\quad + \boxed{HHH} + \boxed{HHHH} \\
&\quad + \boxed{W^+W^-H} + \boxed{W^+W^-HH} + \boxed{ZZH} + \boxed{ZZHH}
\end{aligned} \tag{A.2}$$

Terms in boxes are shorthands for interactions predicted by the Standard Model. Note that all weak boson vertices are always vertices with one W^+ and W^- boson, since the third component of weak isospin is conserved.

$$\begin{aligned}
\mathscr{L}_{\text{Fermion}} &= \sum_{\ell=e,\mu,\tau} \bar{\ell}(\mathrm{i}\gamma_\mu\partial^\mu - m_\ell)\ell + \sum_{\nu=\nu_e,\nu_\mu,\nu_\tau} \bar{\nu}(\mathrm{i}\gamma_\mu\partial^\mu)\nu + \sum_{q=u,\ldots,t} \bar{q}(\mathrm{i}\gamma_\mu\partial^\mu - m_q)q \\
&\quad + \boxed{\bar{\ell}\ell A} + \boxed{\bar{q}qA} \\
&\quad + \boxed{\bar{\nu}_\ell \ell W^+} + \boxed{\bar{\ell}\nu_\ell W^-} + \boxed{\bar{u}d'W^+} + \boxed{\bar{d}'uW^-} \\
&\quad + \boxed{\bar{\ell}\ell Z} + \boxed{\bar{\nu}_\ell\nu_\ell Z} + \boxed{\bar{q}qZ} \\
&\quad + \boxed{\bar{\ell}\ell H} + \boxed{\bar{q}qH}
\end{aligned} \tag{A.3}$$

A 2 A Note on the Higgs Mass

As mentioned in the introduction (see *e.g.* the section II 3.2 on the Higgs mechanism), the mass of the Higgs boson is not predicted by the theory. In fact, it is the only free parameter concerning the Higgs boson, as charge (0), spin and parity (0^+) are fixed by the requirement that the field is an electrically neutral scalar. The mass therefore governs all properties of the Higgs, since it unambiguously determines λ and therefore all couplings to the Higgs boson.

The mass has been measured by the ATLAS and CMS collaborations in 2015 [30]. The result is $m_H = (125.09 \pm 0.24)\,\text{GeV}$. The value of 125 is quite interesting, because it is very close to the theoretically tolerable minimum; at the same time it is also low enough so that many of the possible decay channels become accessible at the LHC. If the mass was $\gtrsim 160\,\text{GeV}$, decays into W bosons would dominate, and the search for $H \to b\bar{b}$ decays would be even more challenging than it is already now — it might even be pointless.

The theoretical constraints limiting the Higgs mass will be discussed now, summarising a more detailed discussion by Djouadi [20]:

Unitarity Due to its coupling to vector bosons, the Higgs boson affects the scattering amplitudes of these bosons, *e.g.* in WW scattering:

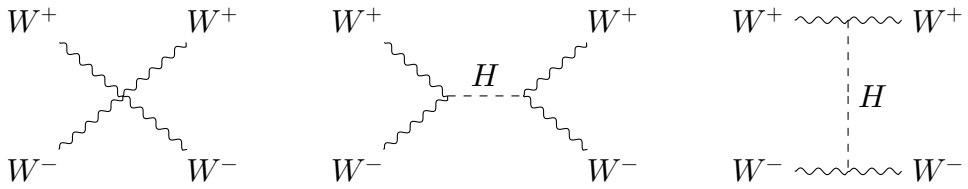


Figure A.1: Selection of diagrams contributing to WW scattering.

The amplitude of such scattering processes at high energies ($s \gg m_W^2$) can be written as

$$A(WW \to WW) \to \frac{1}{v^2}\left(s + t - \frac{s^2}{s - m_H^2} - \frac{t^2}{t - m_H^2}\right).$$

s, t are the Mandelstam variables. Due to the Higgs-mediated diagrams, the total amplitude depends on the Higgs mass. High Higgs masses might render the scattering amplitude so high that unitarity is violated, that is, the probabilities of all possible outcomes exceed unity. This is shown by Djouadi using the optical theorem, which relates the total scattering cross section to the amplitude in forward direction: $\sigma_{\text{tot}} = \frac{1}{s}\Im A_{\text{fwd}}$. Expanding the amplitude in terms of partial waves using Legendre polynomials, one obtains the $J = 0$ partial wave

$$a_0 = -\frac{m_H^2}{8\pi v^2}.$$

Unitarity requires that $|\Re a_l| < \frac{1}{2}$ which limits the Higgs boson mass to $m_H \lesssim 870\,\text{GeV}$. Taking into account other boson scattering amplitudes, this bound lowers to $m_H \lesssim 710\,\text{GeV}$.

This argument breaks down for very high Higgs masses: since the Higgs self-coupling $\lambda = \frac{m_H^2}{2v^2}$ grows with the Higgs mass, perturbation theory cannot be applied to calculate radiative corrections for this result. Given that the Higgs mass is well below this theoretical bound, this is, however, not an issue.

Triviality The quartic Higgs coupling has higher-order radiative corrections, some of which are depicted in figure A.2:

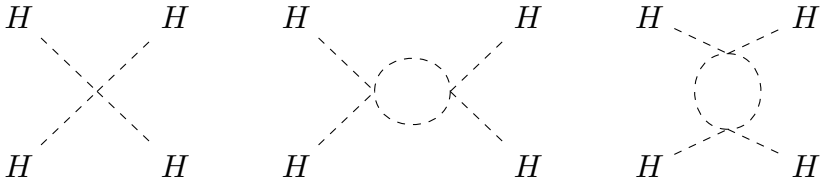


Figure A.2: Quartic Higgs coupling and subset of radiative-correction diagrams.

As all couplings in the Standard Model, λ varies with the energy scale probed. Taking the EW symmetry breaking scale $\lambda(v^2)$ as reference point, the coupling evolves as

$$\lambda(Q^2) = \frac{\lambda(v^2)}{1 - \frac{3}{4\pi^2}\lambda(v^2)\ln\left(\frac{Q^2}{v^2}\right)}$$

If the energy probed is much less than the EW symmetry breaking scale ($Q^2 \ll v^2$), the denominator goes to positive infinity, implying a vanishing coupling, thus vanishing interaction. The theory is called trivial, since it does not describe an interaction, any more. This does not set a bound on the

Higgs mass, but the triviality argument is relevant for the next point.

Perturbativity If the energy is much higher than the EW scale ($Q^2 \gg v^2$), the coupling can become infinitely large. This happens at the "Landau pole"

$$\Lambda_C = v\,\mathrm{e}^{\frac{4\pi^2}{3\lambda}} = v\,\mathrm{e}^{\frac{4\pi^2 v^2}{m_H^2}} \,.$$

The location of this pole depends on the Higgs mass. The Standard Model is valid only below the cut-off scale determined by the pole Λ_C, where the coupling is small, and where perturbation theory can be applied. There are two solutions to avoid the breakdown of perturbation theory:

- Triviality: If one imposes triviality (*i.e.* $\lambda = 0$), the problem vanishes, but the Higgs boson becomes massless. This is clearly not a helpful solution.
- New physics: New phenomena could have an effect on the coupling, which could make it finite. The corresponding bound on the Higgs mass, where new processes can be expected is shown as the upper red band in figure A.3. Given that the Higgs mass is much lower than this bound, perturbativity is also not an issue.

Vacuum Stability Not only loops with Higgs bosons as in figure A.2 contribute to the running of λ, but also fermions play a role due to their Yukawa coupling to the Higgs boson. Since the top quark is the heaviest fermion with the strongest Yukawa coupling, it plays the most important role, but other quarks and the weak bosons also contribute. The evolution of λ can be approximated by

$$\lambda(Q^2) = \lambda(v^2) + \frac{1}{16\pi^2}\left[-12\frac{m_t^4}{v^4} + \frac{3}{16}\left(2g^2 + (g^2 + g'^2)^2\right)\right]\ln\left(\frac{Q^2}{v^2}\right).$$

Depending on the value of the top quark mass, λ could become negative, which would lead to an unbounded potential without a stable vacuum state. The "Mexican hat" of figure II.4(b) (p. 15) would transform into a potential that falls to negative infinity, if only high enough Q^2 are involved. This constrains λ and thus the Higgs mass. The corresponding bound is shown versus the cut-off scale as the green band in figure A.3. The figure shows that the observed Higgs mass is not far from this bound: at energies of the order of $1\,000\,\mathrm{TeV}$, λ should become negative. There might still be a stable vacuum state (different from v) if the potential *e.g.* contained ϕ^6 terms or if new physics affected the evolution of λ, but this is beyond the scope of this work. Although figure A.3 is slightly outdated because the precision of the top quark mass measurement has increased, the result that λ may become negative remains valid.

This work will neither have an impact on the measurement of the Higgs boson mass, nor will it probe the Yukawa coupling (or the mass) of the top quark. However, in light of the interesting fact that the Higgs and top masses are so close to the stability bound, it might provide helpful insights into the properties of the Higgs boson. Considering that the universe may in fact be long-lived but only metastable, that is, a small but non-zero probability exists that the vacuum Higgs field transitions to a state with lower potential energy [174], one might ask if contributions of new physics processes could prevent this. New particles might change the evolution of λ, which means that the vacuum stability calculations would have to be revised. If such particles affected λ, they would also couple to the Higgs, and might therefore change its decay branching ratios and production cross sections. Since the aim of this work is to measure cross section times branching ratio, it indirectly probes for new physics. Aside from proving that the Higgs couples to b quarks, this is one of the main motivations to search for $H \to b\bar{b}$ decays.

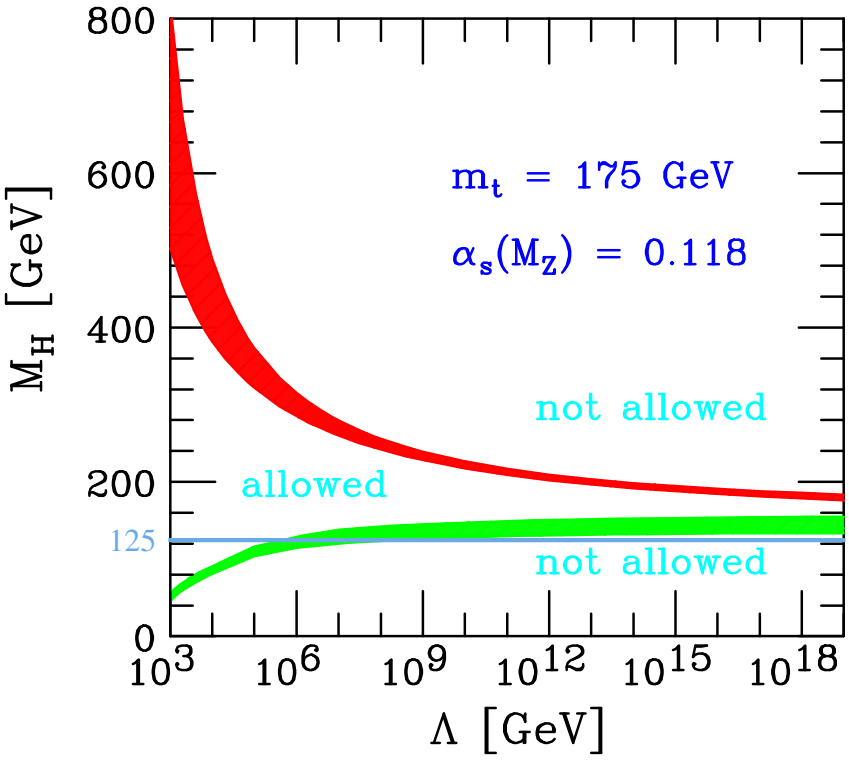


Figure A.3: Theoretical bounds on the Higgs mass with uncertainties [20]. The upper bound results from requiring perturbativity, that is, λ remains small; the lower bound results from vacuum stability, that is, λ stays positive. The bands indicate the theoretical uncertainties. The measured Higgs mass is indicated by the blue line.

A 3 Useful Relations for Four-Vectors

$$p^\mu = \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} \qquad p^\mu = \begin{pmatrix} E \\ p_\mathrm{T}\cos(\phi) \\ p_\mathrm{T}\sin(\phi) \\ p\ \ \cos(\theta) \end{pmatrix}$$

p_T, η, ϕ coordinates for a massless particle:

$$p^\mu = \begin{pmatrix} p_\mathrm{T}\cosh(\eta) \\ p_\mathrm{T}\ \ \cos(\phi) \\ p_\mathrm{T}\ \ \sin(\phi) \\ p_\mathrm{T}\sinh(\eta) \end{pmatrix} \tag{A.4}$$

This parametrisation follows from:

$$\begin{aligned} \eta &= \frac{1}{2}\ln\left(\frac{p+p_{\mathrm{L}}}{p-p_{\mathrm{L}}}\right) \\ &= \frac{1}{2}\ln\left(\frac{1+p_{\mathrm{L}}/p}{1-p_{\mathrm{L}}/p}\right) \\ &= \operatorname{artanh}\left(\frac{p_{\mathrm{L}}}{p}\right) \\ \Leftrightarrow \quad \tanh(\eta) &= \frac{p_{\mathrm{L}}}{p} \\ \Leftrightarrow \quad p_{\mathrm{L}} &= p\tanh(\eta) \end{aligned}$$

Using the definition of p and the above result:

$$\begin{aligned} p^2 &= p_{\mathrm{L}}^2\tanh^{-2}(\eta) = p_{\mathrm{L}}^2 + p_{\mathrm{T}}^2 \\ \Leftrightarrow \quad p_{\mathrm{L}}^2 &= \frac{p_{\mathrm{T}}^2}{\tanh^{-2}(\eta)-1} \\ \Leftrightarrow \quad p_{\mathrm{L}}^2 &= \frac{p_{\mathrm{T}}^2\sinh^2(\eta)}{\cosh^2(\eta)-\sinh^2(\eta)=1} \\ \Leftrightarrow \quad p_{\mathrm{L}} &= p_{\mathrm{T}}\sinh(\eta) \end{aligned}$$

This can be combined with the previous result on p_{L}:

$$\begin{aligned} p_{\mathrm{L}} &= p_{\mathrm{T}}\sinh(\eta) = p\tanh(\eta) \\ \Leftrightarrow \quad p &= p_{\mathrm{T}}\cosh(\eta) \end{aligned}$$

Inserting these two results in the energy and p_z components of a four-vector, equation (A.4) is obtained.

APPENDIX B

Notes on the ATLAS $H \to b\bar{b}$ Analysis

B 1 The Contribution of sg Scattering in $W+cl$ Events

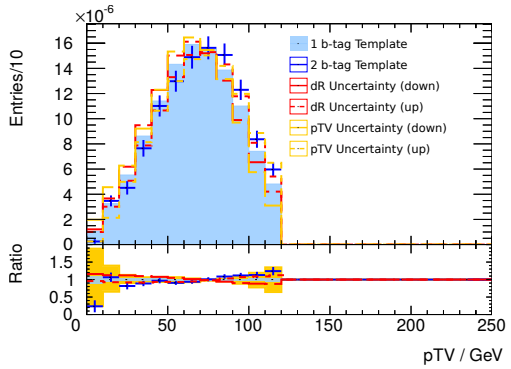
In the ATLAS $H \to b\bar{b}$ publication [1] $W+cl$ events is plotted separately from other $W+$jets events with charm hadrons because on top of gluon radiation like in the other $W+$jets processes, there is an additional process to produce this final state: strange-gluon scattering $sg \to W+c$.

Event though it is difficult to quantify the exact fraction of sg scattering in $W+cl$ events, it is certainly not the dominating process: most measurements of sg scattering focus on $W+1$ jet final states but this analysis only retains events with at least two jets. sg scattering for $W+1$ jet events is expected to be 5% [175] of the total cross-section for the Tevatron, 10 to 12 % for LHC [176]. Still, for this process to appear in the $H \to b\bar{b}$ selection, there needs to be an additional light or gluon jet. The best estimate for LHC can be obtained from an ATLAS measurement of $W+c$ production with one additional jet [177]: comparing same sign (SS) with opposite sign final states (OS), an **an upper limit** of the sg scattering contribution can be obtained. This is because same sign final states cannot be created in sg scattering since the charge of the s-quark needs to be conserved: $s^{-1/3}g \to W^- c^{+2/3}$ or its charge conjugate are the only possible reactions. The measurement yields

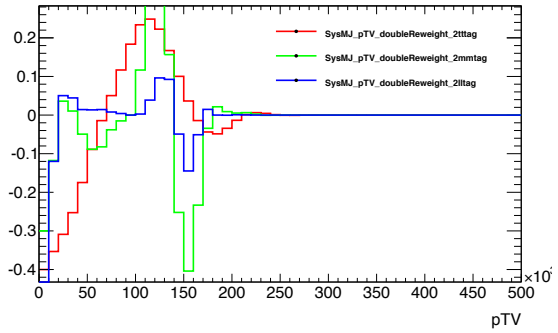
$$\frac{(W+cx)^{sg\ \text{scat}}}{(W+cx)^{\text{total}}} \leq \frac{(W+cx)^{\text{OS}}}{(W+cx)^{\text{total}}} = 28\%.$$

Combining this with the 10 to 12 % expectation for $W+1$ jet events does not implicate a strong necessity to plot the $W+cl$ contribution separately.

B 2 Extrapolating the 1b-tag QCD Multijet Template to the 2b-tag Region



(a) 1b-tag– 2b-tag comparison for p_{T}^{V} in events with 2 jets after applying the $\Delta R(b,b)$ correction



(b) Correction functions for the 1b-tag template in the 2-jet region in dependence of b-tagging working point

Figure B.1: Comparison of the multijet template in the 1 and 2b-tag region. (a) the shaded 1b-tag template predicts lower values of p_{T}^{V} than the 2b-tag template. The differences between the templates are shown as red and orange uncertainty bands. The central value of the 1b-tag template is corrected to match the 2b-tag template and half of the shown uncertainty band is assigned as systematic uncertainty. The correction functions are shown in (b). With tighter b-tagging, the corrections become larger because the bias on the kinematic distributions is stronger.
The correction rapidly changes at $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$ because the multijet template in the high-p_{T}^{V} region is normalised separately from the low-p_{T}^{V} region. The multijet normalisation measurement partly corrects the p_{T}^{V} dependence which is why the correction is positive at $p_{\mathrm{T}}^{V} \gtrsim 120\,\mathrm{GeV}$ but negative at $p_{\mathrm{T}}^{V} \gtrsim 150\,\mathrm{GeV}$ to cancel the overcorrection due to the normalisation correction.

B 3 A BDT-driven Determination of Preselection Cuts

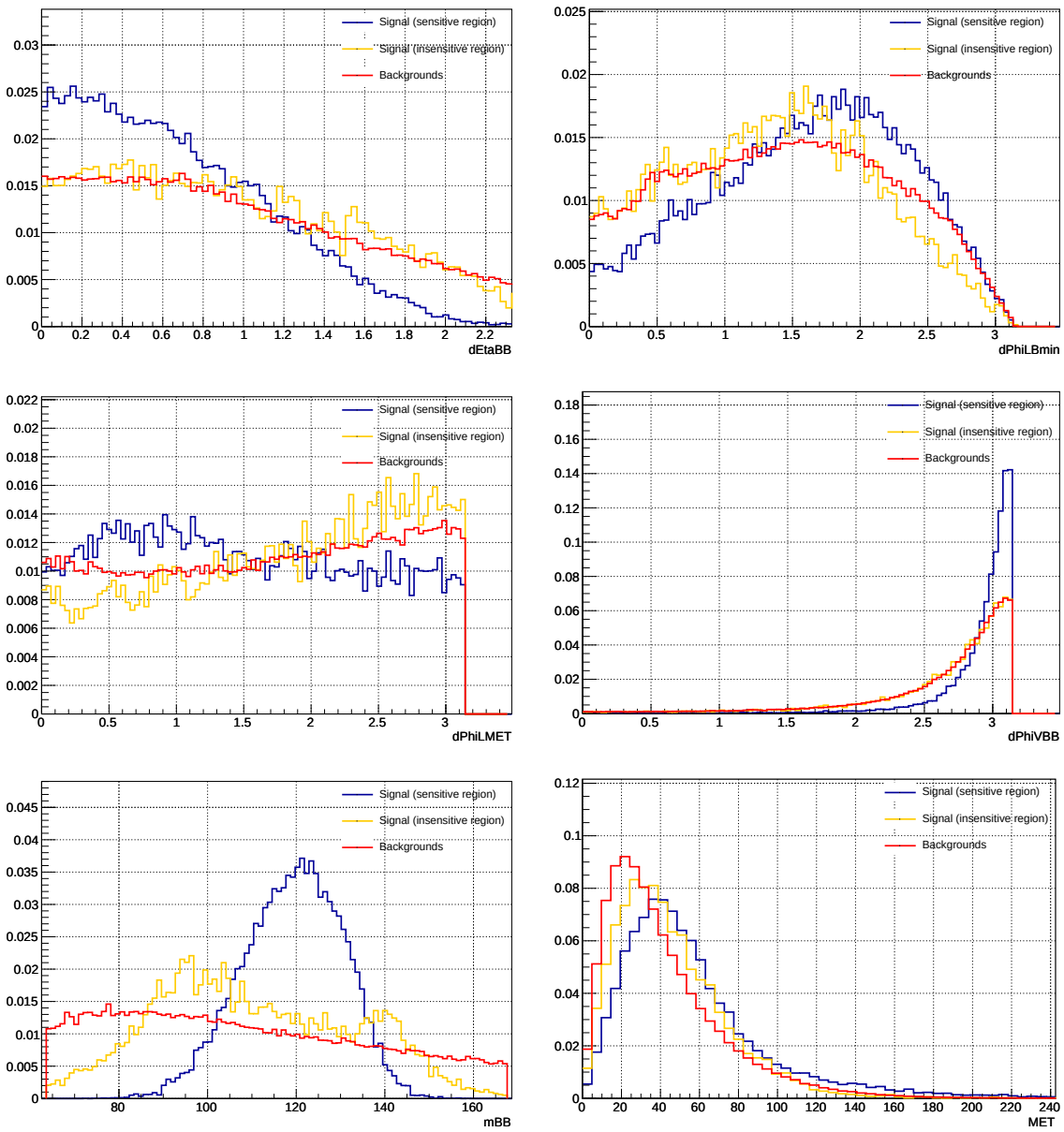


Figure B.2: Usage of a BDT classifier to derive preselection cuts that retain regions of high-sensitivity. The BDT is trained events where only topology selection and b-tagging is applied. It uses a large set of candidates for preselection variables. The figures show the distributions of H_{T} and $\Delta R(b,b)$ for a phase space with high S/B (50 % working point), low S/B and all backgrounds. The distributions show that lower cuts on H_{T} and $\Delta R(b,b)$ can be applied without suppressing signal events in regions with high S/B.
The only usable preselection cut is $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$, which is indeed applied in the analysis. The angular variables show only low separation, the sidebands of m_{bb} are necessary to measure the normalisation of backgrounds.

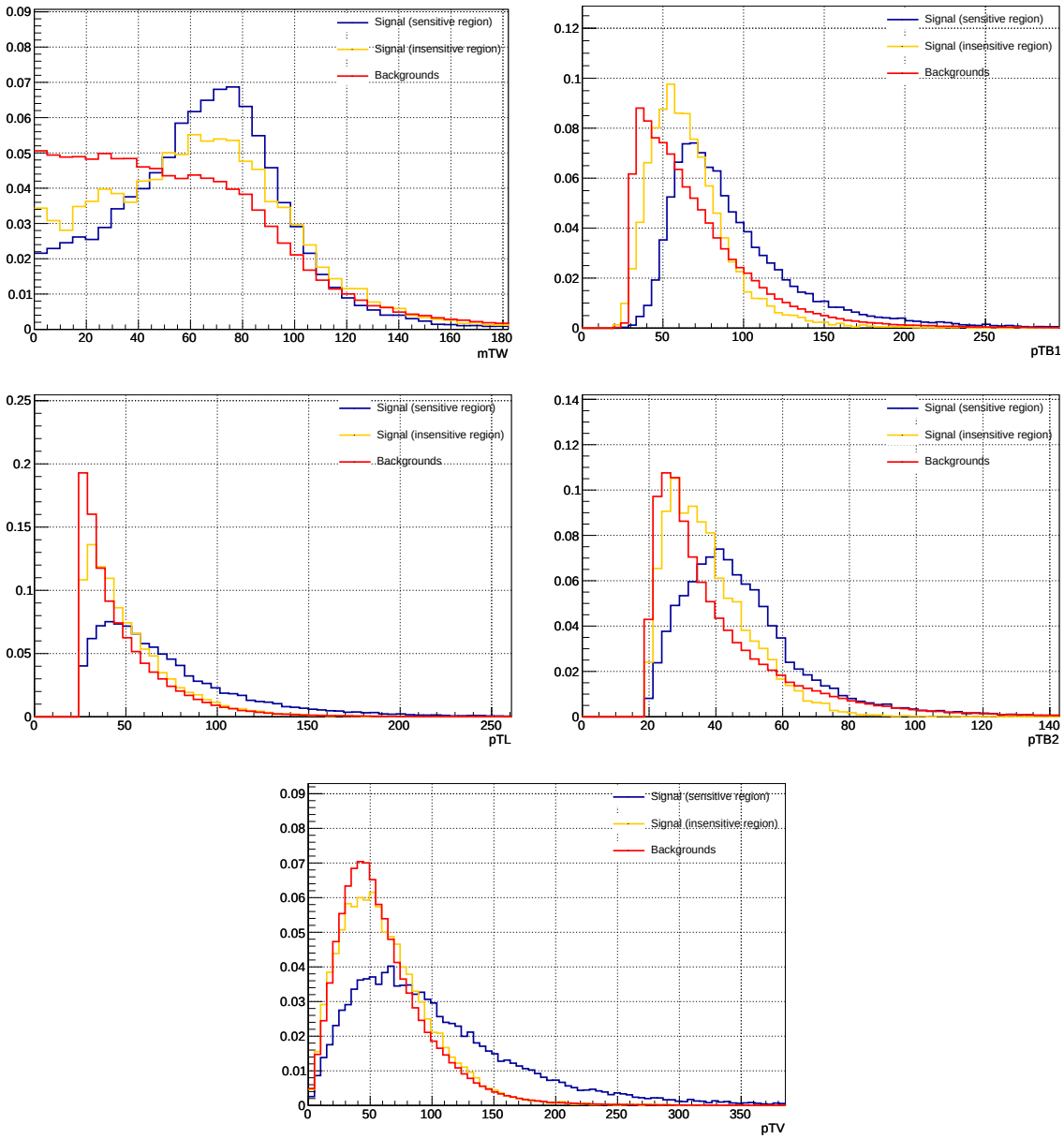
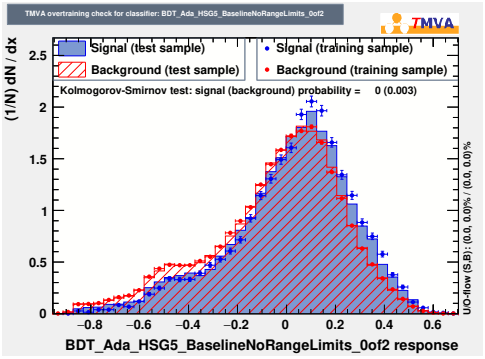
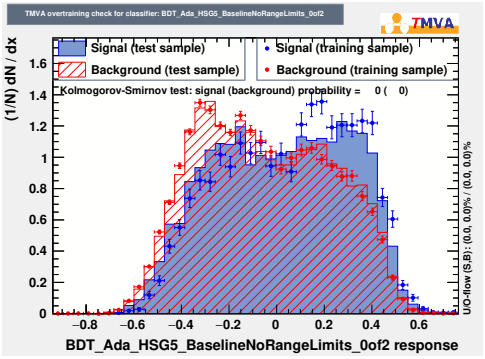


Figure B.3: Usage of a BDT classifier to derive preselection cuts that retain regions of high-sensitivity. The BDT is trained events where only topology selection and b-tagging is applied. It uses a large set of candidates for preselection variables. The figures show the distributions of H_{T} and $\Delta R(b,b)$ for a phase space with high S/B (50 % working point), low S/B and all backgrounds. The distributions show that lower cuts on H_{T} and $\Delta R(b,b)$ can be applied without suppressing signal events in regions with high S/B.
Usable preselection cuts are $p_{\mathrm{T}}^{b_1} > 45\,\mathrm{GeV}$, $p_{\mathrm{T}}^{b_2} > 20\,\mathrm{GeV}$, which are indeed applied in the analysis. Cuts on m_{T}^{W} and p_{T}^{V} are not advisable and were thus removed.

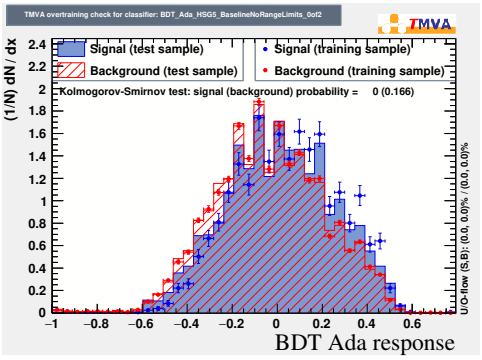
B 4 Multivariate Modelling Checks and Covariate Shift



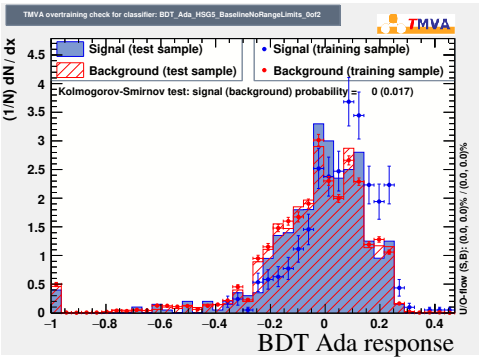
(a) 1-lepton analysis, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$



(b) 2-lepton analysis, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$

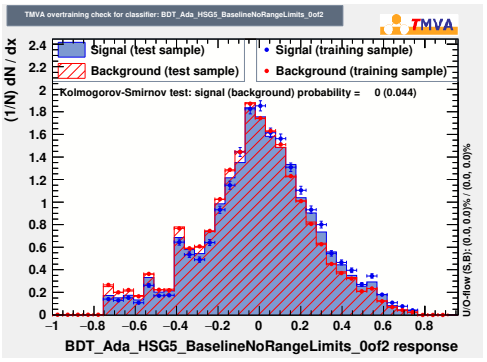


(c) 1-lepton analysis, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$

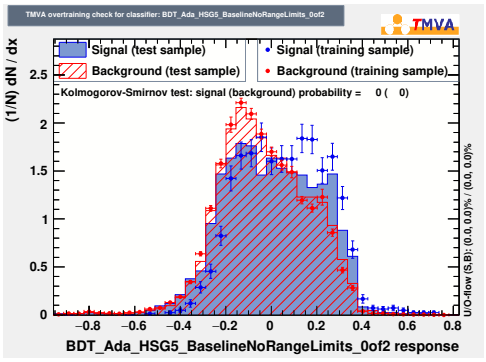


(d) 2-lepton analysis, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$

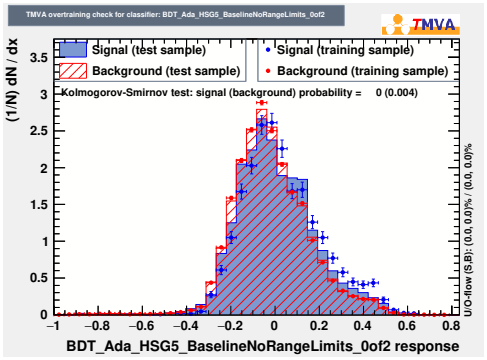
Figure B.4: BDTs trained to separate data (blue) from Monte Carlo events (red). The poor separation of the two classes in the test sample (filled histograms) shows that the BDTs are not able to detect deficits in the Monte Carlo simulation.



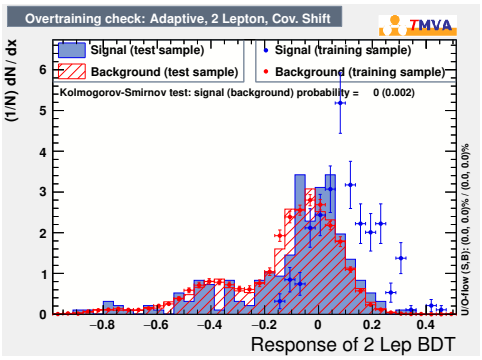
(a) 1-lepton analysis, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$, 3 jets



(b) 2-lepton analysis, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$, 3 jets



(c) 1-lepton analysis, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$, 3 jets



(d) 2-lepton analysis, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$, 3 jets

Figure B.5: BDTs trained to separate data (blue) from Monte Carlo events (red). The poor separation of the two classes in the test sample (filled histograms) shows that the BDTs are not able to detect deficits in the Monte Carlo simulation.

APPENDIX C

Notes on the Lorentz-Invariant Analysis

C 1 Plots of the 3-Jet Regions

C 1.1 Lorentz Invariants

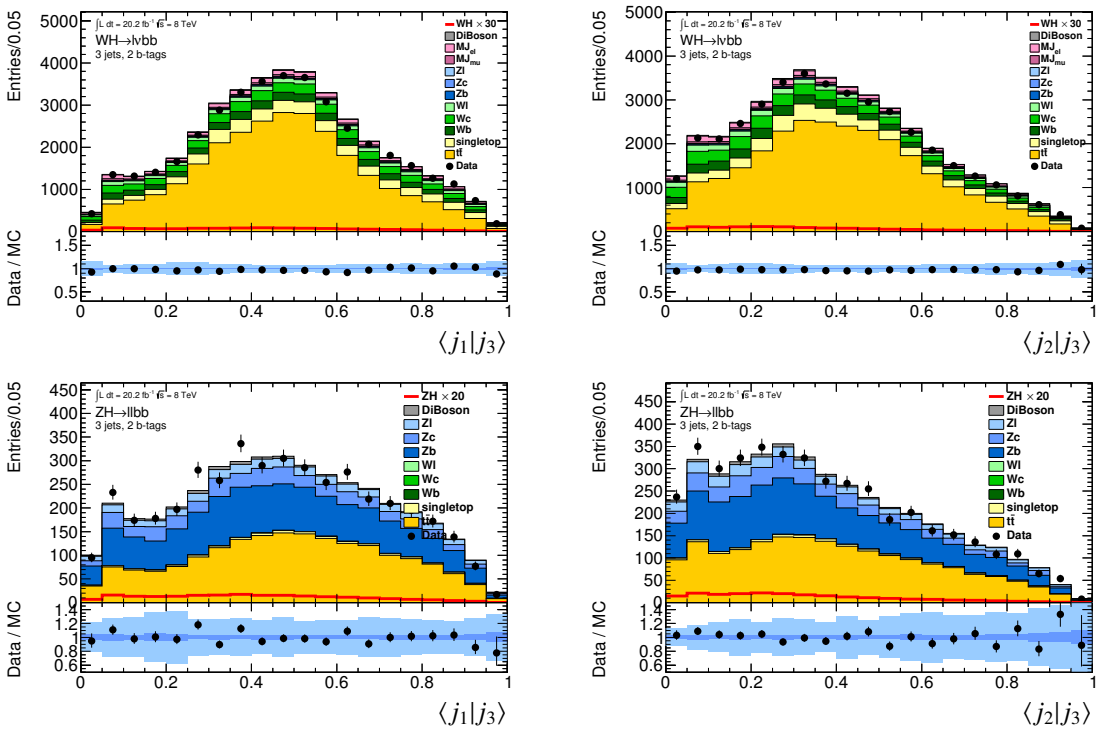
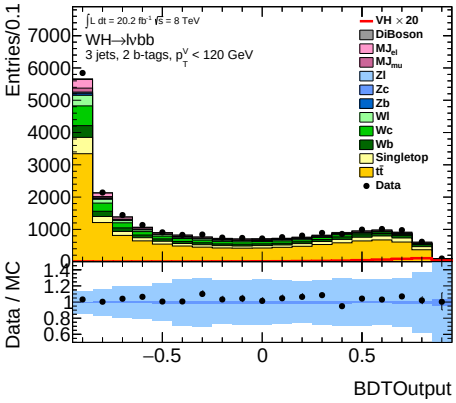
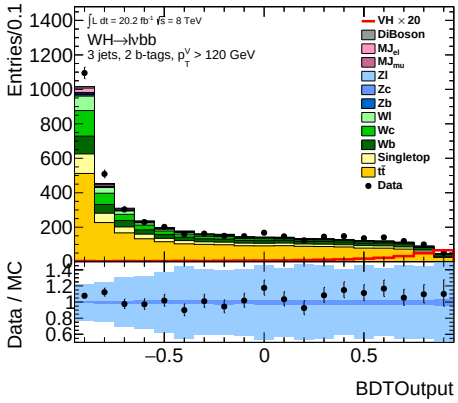


Figure C.1: Additional Lorentz invariants used in the 3-jet region. These can be interpreted as a Lorentz-invariant angle between jets. **Left**: Inner product of leading and third-leading momentum jet. **Right**: Inner product of subleading and third-leading momentum jet. Low values indicate more collinear jets.

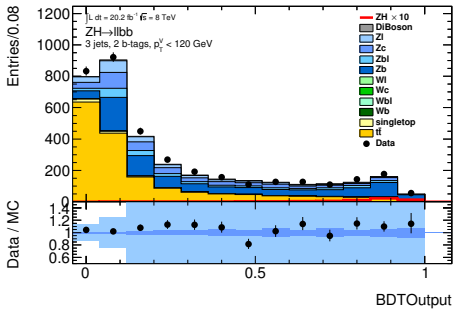
C 1.2 BDT Classifiers



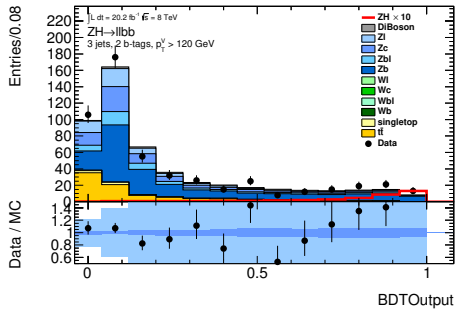
(a) 1 lepton, 3 jets, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$



(b) 1 lepton, 3 jets, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$



(c) 2 leptons, 3 jets, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$



(d) 2 leptons, 3 jets, $p_{\mathrm{T}}^{V} > 120\,\mathrm{GeV}$

Figure C.2: BDT classifiers with Lorentz-invariant variables and gradient boosting in the 3-jet region. These classifiers are fed into the likelihood model outlined in section IV 6. The preselection outlined in section IV 1 and shape corrections from section IV 2 are applied. Normalisation corrections for the background simulations are not applied except for the multijet fit and the scaling up the $Z+$jets backgrounds by 16 % as before. The dark blue band in the data/MC plot shows the statistical uncertainty of the background simulations. The larger band in light blue shows the quadratic sum of the systematic uncertainties. Normalisation uncertainties are not shown because they are added in the likelihood model.

C 2 Breakdown of Systematic Uncertainties of the Lorentz-Invariant MVA in Comparison to the ATLAS Approach

Notes for table C.1:

- "Floating normalisations" denote normalisation factors without constraint terms: $W+bb$, $Z+bb$, $W+cl$, $Z+cl$ and $t\bar{t}$.
- "Normalisation uncertainties" denote η parameters as described in section IV 5.2, page 104. These are constrained with Gaussian terms.
- "Normalisations" denote both: factors and uncertainties.

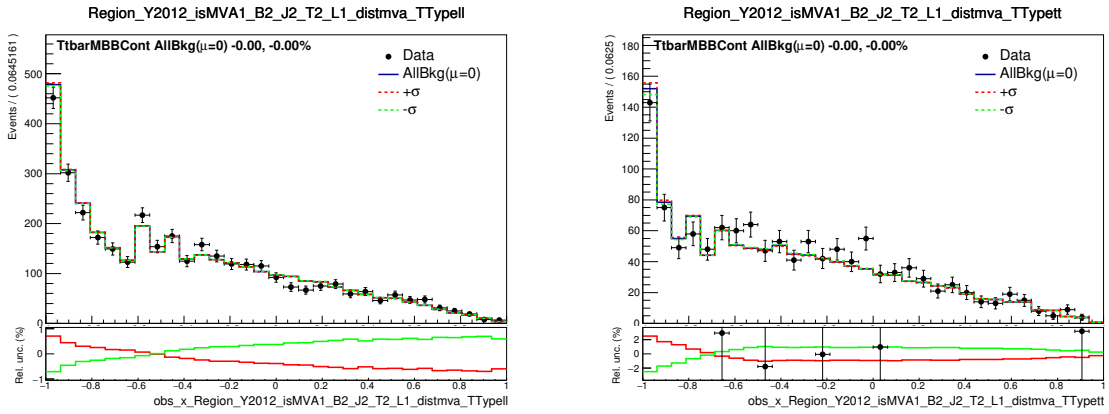
Table C.1: Breakdown of uncertainties of the signal strength μ. Tabular version of figure V.9, page 147. The sets of nuisance parameters are not orthogonal, i.e. nuisance parameters may be found in several of the sets listed here. The Lorentz invariant classifier with gradient boosting serves as a baseline, for other classifiers the mean relative increase/decrease of uncertainties with respect to the first column is shown.
Since the 1 lepton uncertainties dominate, the test was simplified by not retraining the 2 lepton classifiers in the two middle columns.

	Lorentz invariants			ATLAS $H \to b\bar{b}$		
Uncertainty on μ	BDTG	Adaptive		BDTG	Adaptive	
Total	$^{+0.65}_{-0.58}$	+4.8 %	$^{+0.68}_{-0.61}$	+7.9 %	$^{+0.71}_{-0.62}$	+10 %	$^{+0.72}_{-0.64}$
Statistics (data)	$^{+0.45}_{-0.44}$	+3.8 %	$^{+0.47}_{-0.45}$	+2.5 %	$^{+0.46}_{-0.45}$	+4.3 %	$^{+0.47}_{-0.46}$
Statistics (MC)	$^{+0.09}_{-0.09}$	+0.0 %	$^{+0.09}_{-0.09}$	+16 %	$^{+0.11}_{-0.10}$	−18 %	$^{+0.08}_{-0.07}$
Full systematics	$^{+0.47}_{-0.39}$	+5.7 %	$^{+0.50}_{-0.41}$	+13 %	$^{+0.54}_{-0.43}$	+16 %	$^{+0.55}_{-0.44}$
Normalisations	$^{+0.25}_{-0.22}$	−0.5 %	$^{+0.24}_{-0.22}$	+16 %	$^{+0.29}_{-0.25}$	−3.3 %	$^{+0.24}_{-0.21}$
Normalisations (floating part)	$^{+0.21}_{-0.18}$	−5.9 %	$^{+0.19}_{-0.18}$	+10 %	$^{+0.23}_{-0.20}$	−18 %	$^{+0.17}_{-0.15}$
Normalisations (modelling uncert.)	$^{+0.16}_{-0.15}$	+3.5 %	$^{+0.17}_{-0.16}$	+26 %	$^{+0.21}_{-0.19}$	−9.8 %	$^{+0.15}_{-0.14}$
W+jets	$^{+0.23}_{-0.20}$	−8.8 %	$^{+0.20}_{-0.19}$	+7.7 %	$^{+0.24}_{-0.21}$	−6.4 %	$^{+0.21}_{-0.19}$
W+jets normalisations	$^{+0.18}_{-0.16}$	−11 %	$^{+0.15}_{-0.15}$	+3.0 %	$^{+0.18}_{-0.16}$	−26 %	$^{+0.13}_{-0.12}$
W+jets modelling	$^{+0.17}_{-0.16}$	−15 %	$^{+0.14}_{-0.14}$	+15 %	$^{+0.20}_{-0.18}$	+2.5 %	$^{+0.17}_{-0.16}$
Jets & MET	$^{+0.22}_{-0.16}$	+12 %	$^{+0.25}_{-0.18}$	+10 %	$^{+0.25}_{-0.17}$	+22 %	$^{+0.27}_{-0.20}$
Jets	$^{+0.22}_{-0.16}$	+12 %	$^{+0.25}_{-0.18}$	+7.7 %	$^{+0.24}_{-0.17}$	+22 %	$^{+0.27}_{-0.19}$
Top	$^{+0.16}_{-0.14}$	+8.3 %	$^{+0.17}_{-0.15}$	+3.1 %	$^{+0.18}_{-0.13}$	−6.0 %	$^{+0.15}_{-0.13}$

	Lorentz invariants			ATLAS $H \to b\bar{b}$			
Uncertainty on μ (cont.)	BDTG	Adaptive		BDTG		Adaptive	
Top normalisation	$^{+0.12}_{-0.10}$	−2.9 %	$^{+0.11}_{-0.10}$	+17 %	$^{+0.15}_{-0.11}$	−13 %	$^{+0.10}_{-0.08}$
Top modelling	$^{+0.09}_{-0.09}$	+17 %	$^{+0.11}_{-0.10}$	−16 %	$^{+0.08}_{-0.07}$	−2.8 %	$^{+0.09}_{-0.09}$
b-tagging	$^{+0.17}_{-0.13}$	−2.4 %	$^{+0.16}_{-0.12}$	+4.3 %	$^{+0.17}_{-0.13}$	+37 %	$^{+0.23}_{-0.17}$
b-tagging (b-jets)	$^{+0.10}_{-0.07}$	−3.9 %	$^{+0.10}_{-0.06}$	+9.9 %	$^{+0.12}_{-0.07}$	+84 %	$^{+0.18}_{-0.13}$
b-tagging (c-jets)	$^{+0.08}_{-0.07}$	−9.0 %	$^{+0.07}_{-0.06}$	+8.7 %	$^{+0.08}_{-0.07}$	+22 %	$^{+0.10}_{-0.08}$
b-tagging (light-jets)	$^{+0.07}_{-0.06}$	−7.6 %	$^{+0.06}_{-0.06}$	−8.6 %	$^{+0.06}_{-0.06}$	+0.5 %	$^{+0.07}_{-0.06}$
Signal Theory	$^{+0.20}_{-0.09}$	−2.3 %	$^{+0.20}_{-0.08}$	+25 %	$^{+0.24}_{-0.12}$	+69 %	$^{+0.29}_{-0.17}$
Jet Energy Resolution	$^{+0.17}_{-0.11}$	+2.9 %	$^{+0.18}_{-0.11}$	−4.5 %	$^{+0.17}_{-0.10}$	+31 %	$^{+0.21}_{-0.15}$
Z+jets	$^{+0.13}_{-0.13}$	+8.4 %	$^{+0.14}_{-0.14}$	+12 %	$^{+0.15}_{-0.15}$	−15 %	$^{+0.12}_{-0.11}$
Z+jets normalisations	$^{+0.09}_{-0.10}$	+5.7 %	$^{+0.10}_{-0.10}$	+6.8 %	$^{+0.10}_{-0.11}$	−34 %	$^{+0.06}_{-0.07}$
Z+jets modelling	$^{+0.07}_{-0.07}$	+19 %	$^{+0.09}_{-0.08}$	+22 %	$^{+0.09}_{-0.08}$	+4.7 %	$^{+0.08}_{-0.07}$
bl 1T $\to$ 2T extrapolation	$^{+0.11}_{-0.11}$	+1.9 %	$^{+0.10}_{-0.11}$	+26 %	$^{+0.14}_{-0.14}$	−21 %	$^{+0.08}_{-0.09}$
Multijet	$^{+0.07}_{-0.06}$	+12 %	$^{+0.08}_{-0.06}$	+22 %	$^{+0.09}_{-0.06}$	+5.2 %	$^{+0.07}_{-0.06}$
Single top	$^{+0.07}_{-0.05}$	−16 %	$^{+0.05}_{-0.05}$	−93 %	$^{0.00}_{0.00}$	+18 %	$^{+0.07}_{-0.06}$
Luminosity uncertainty	$^{+0.06}_{-0.03}$	+0.4 %	$^{+0.05}_{-0.03}$	+0.9 %	$^{+0.06}_{-0.03}$	+36 %	$^{+0.07}_{-0.05}$
Diboson	$^{+0.03}_{-0.02}$	−9.7 %	$^{+0.02}_{-0.02}$	−1.9 %	$^{+0.03}_{-0.02}$	−8.1 %	$^{+0.03}_{-0.02}$
Lepton ident./reconstr.	$^{+0.02}_{-0.01}$	−51 %	$^{+0.01}_{-0.01}$	−15 %	$^{+0.02}_{-0.01}$	−18 %	$^{+0.01}_{-0.01}$
MET	$^{+0.01}_{-0.01}$	+157 %	$^{+0.02}_{-0.02}$	+550 %	$^{+0.05}_{-0.04}$	+114 %	$^{+0.02}_{-0.01}$

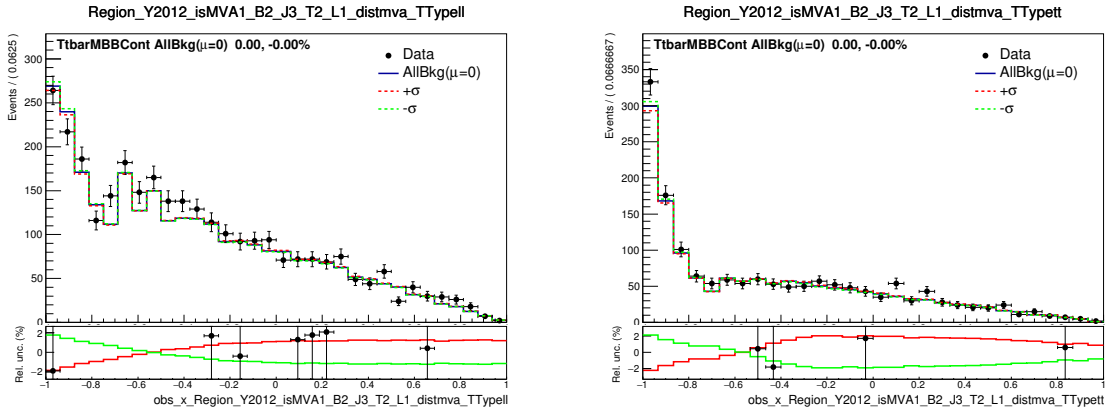
C 3 Details on the Validation of the Likelihood Model

C 3.1 Invariant Mass of b-jets for $t\bar{t}$ Events



(a) 2 jets, 2 ll b-tags

(b) 2 jets, 2 tt b-tags



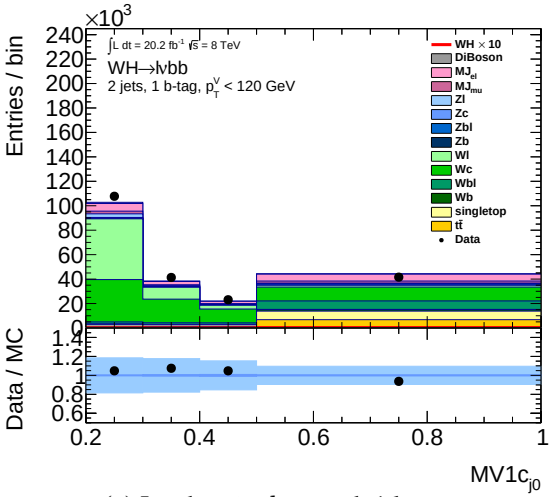
(c) 3 jets, 2 ll b-tags

(d) 3 jets, 2 tt b-tags

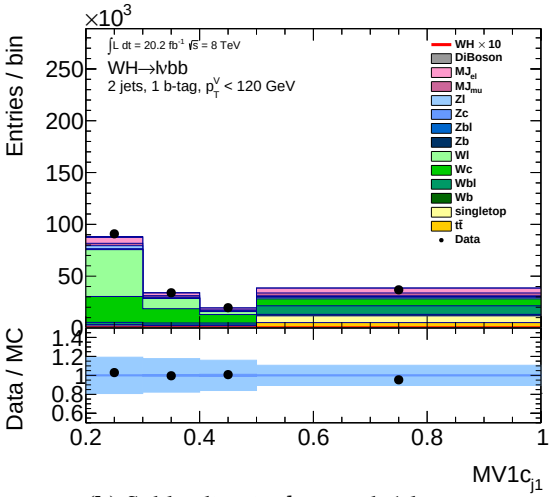
Figure C.3: Systematic uncertainty of the m_{bb} distribution for $t\bar{t}$ events. The histograms show data and the total background prediction before the profile-likelihood fit. Predicted distributions of m_{bb} depend on the Monte Carlo generator being used. Differences between the predictions were assigned as systematic uncertainty `TtbarMBBCont`, indicated by the coloured lines. **Top**: In the 2-jet region `POWHEG+Pythia` predicts too many $t\bar{t}$ events in the most background-like region (classifier output of -1 to -0.6), but too many at higher classifier scores. **Bottom**: In the 3-jet region, the effect is reversed: the generator predicts too few events. If the systematic uncertainty is pulled towards -1σ, both effects are corrected.

C 4 Response of MV1c

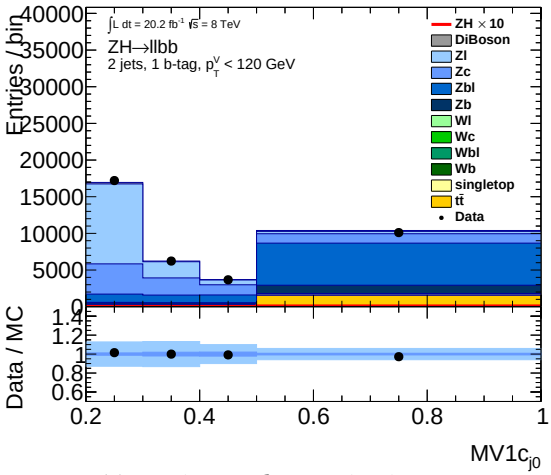
C 4.1 1b-tag Region



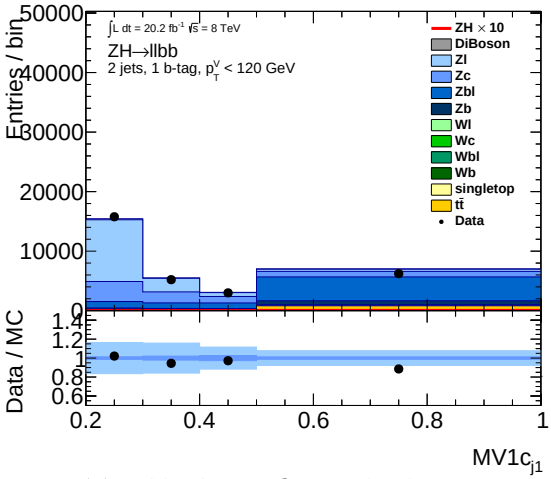
(a) Leading jet b-tagged, 1 lepton



(b) Subleading jet b-tagged, 1 lepton



(c) Leading jet b-tagged, 2 leptons



(d) Subleading jet b-tagged, 2 leptons

Figure C.4: b-tagging score of the b-tagged jet for 2-jet events, $p_{\mathrm{T}}^{V} < 120\,\mathrm{GeV}$. The bins correspond to the four working points of the `MV1c` b-tagging algorithm. From left to right: 80, 70, 60 and 50 %. 1b-tag distributions show an excess of Monte Carlo events in the rightmost bin.

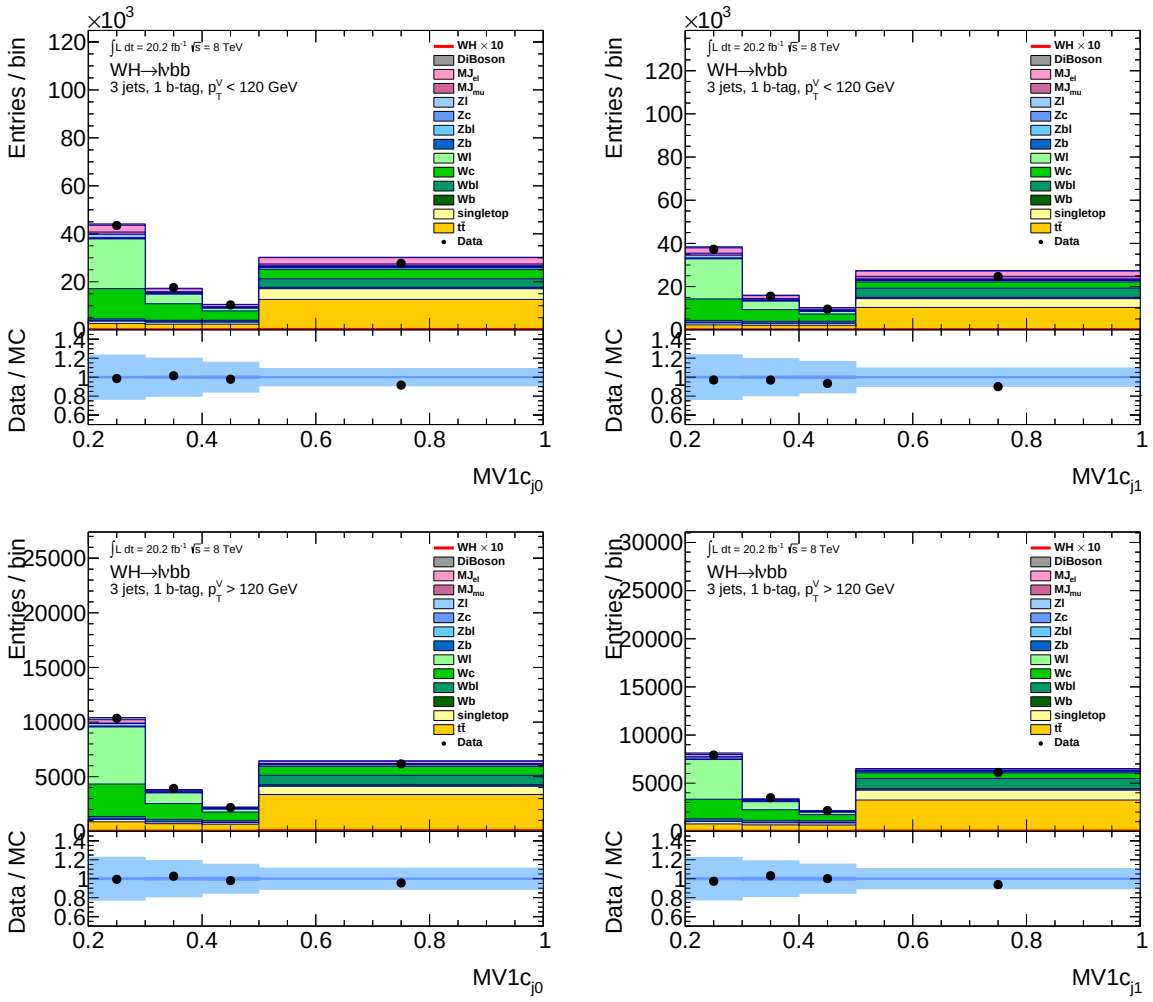


Figure C.5: b-tagging score of the b-tagged jet for 3-jet events, 1-lepton channel. The bins correspond to the four working points of the `MV1c` b-tagging algorithm. From left to right: 80, 70, 60 and 50 %. 1b-tag distributions show an excess of Monte Carlo events in the rightmost bin.

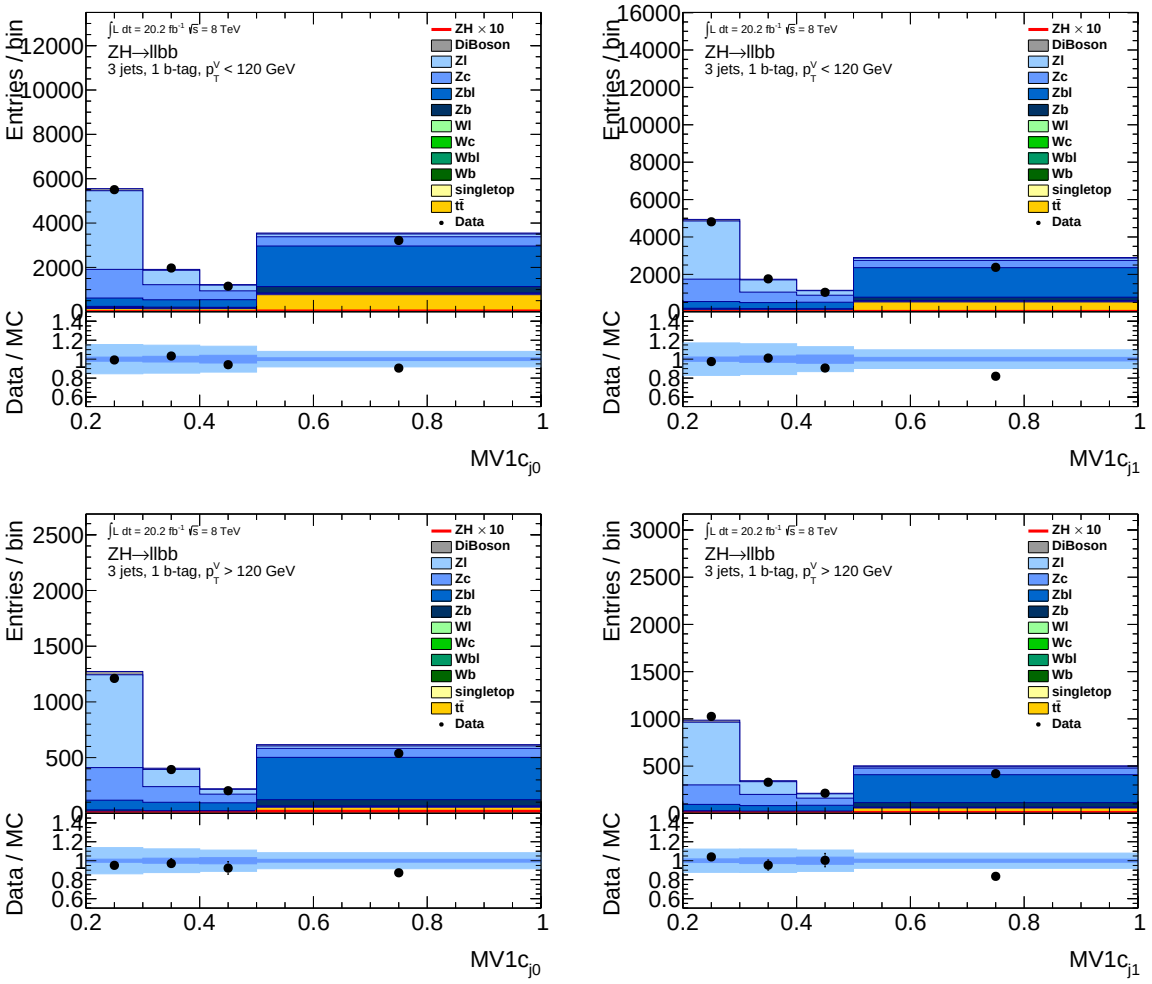
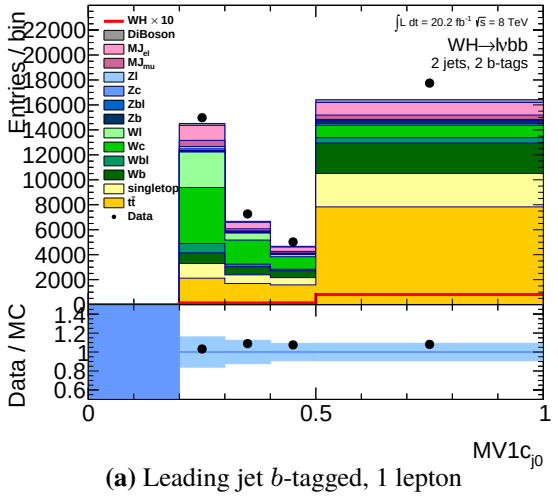
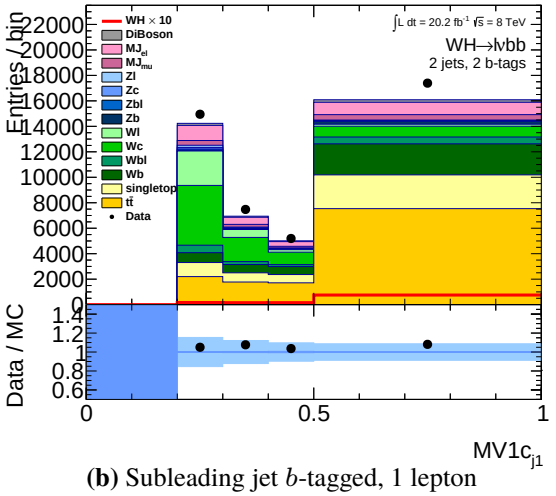


Figure C.6: *b*-tagging score of the *b*-tagged jet for 3-jet events, 2-lepton channel. The bins correspond to the four working points of the `MV1c` *b*-tagging algorithm. From left to right: 80, 70, 60 and 50 %. 1*b*-tag distributions show an excess of Monte Carlo events in the rightmost bin.

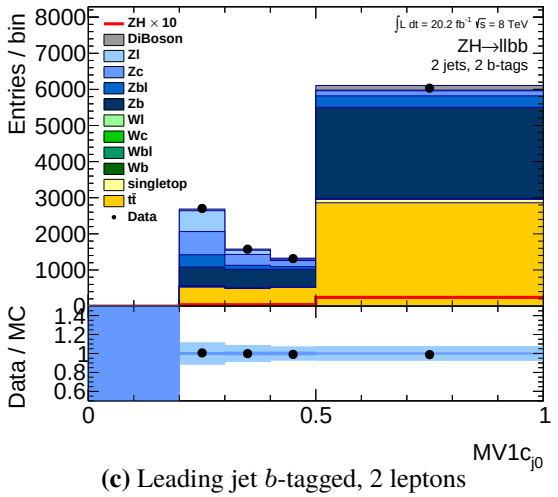
C 4.2 2b-tag Region



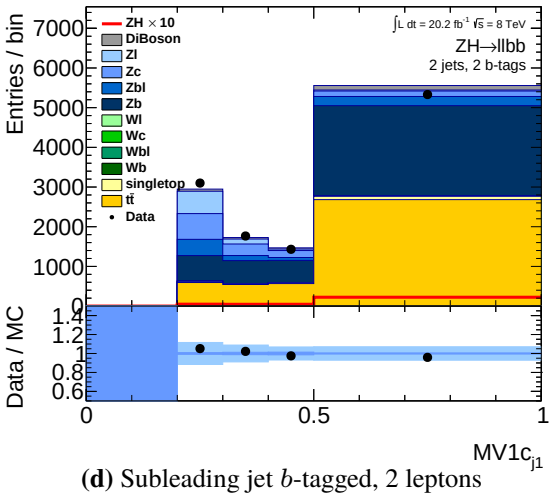
(a) Leading jet b-tagged, 1 lepton



(b) Subleading jet b-tagged, 1 lepton

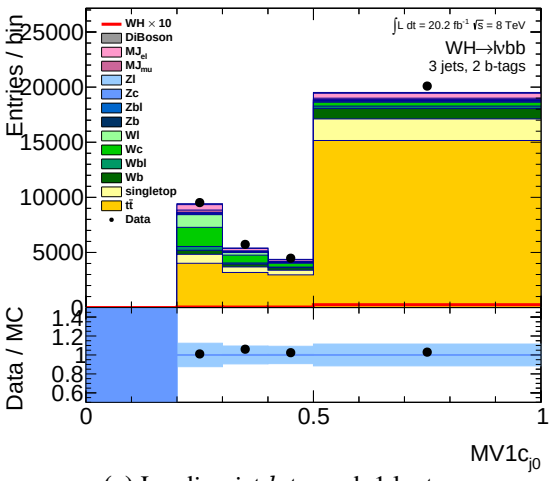


(c) Leading jet b-tagged, 2 leptons

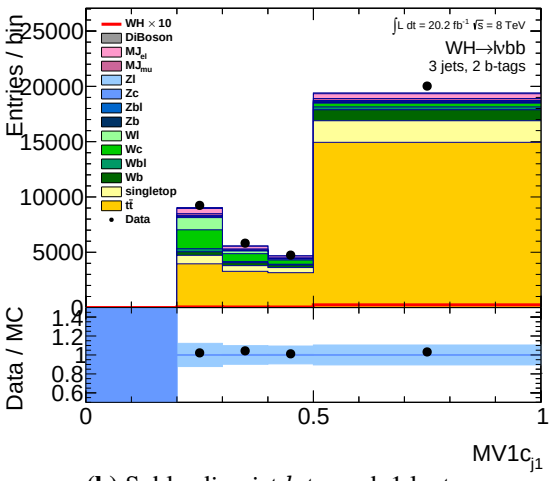


(d) Subleading jet b-tagged, 2 leptons

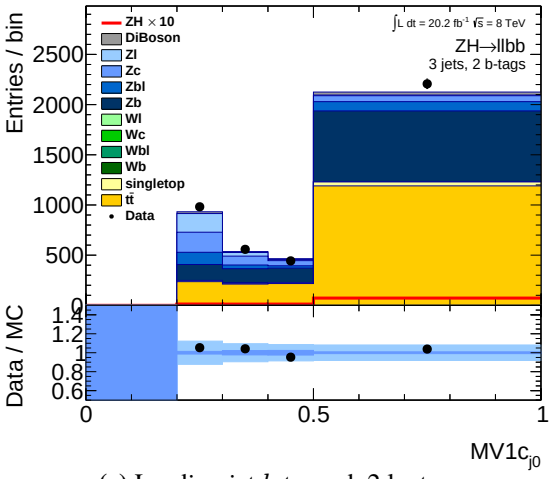
Figure C.7: b-tagging scores for 2-jet events with 2 b-tags, inclusive in p_{T}^{V}. The bins correspond to the four working points of the `MV1c` b-tagging algorithm. From left to right: 80, 70, 60 and 50 %. 1b-tag distributions show an excess of Monte Carlo events in the rightmost bin. 2b-tag distributions show no such excess.



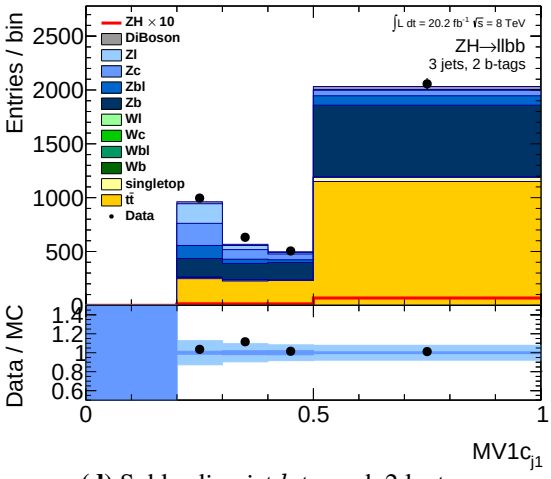
(a) Leading jet b-tagged, 1 lepton



(b) Subleading jet b-tagged, 1 lepton



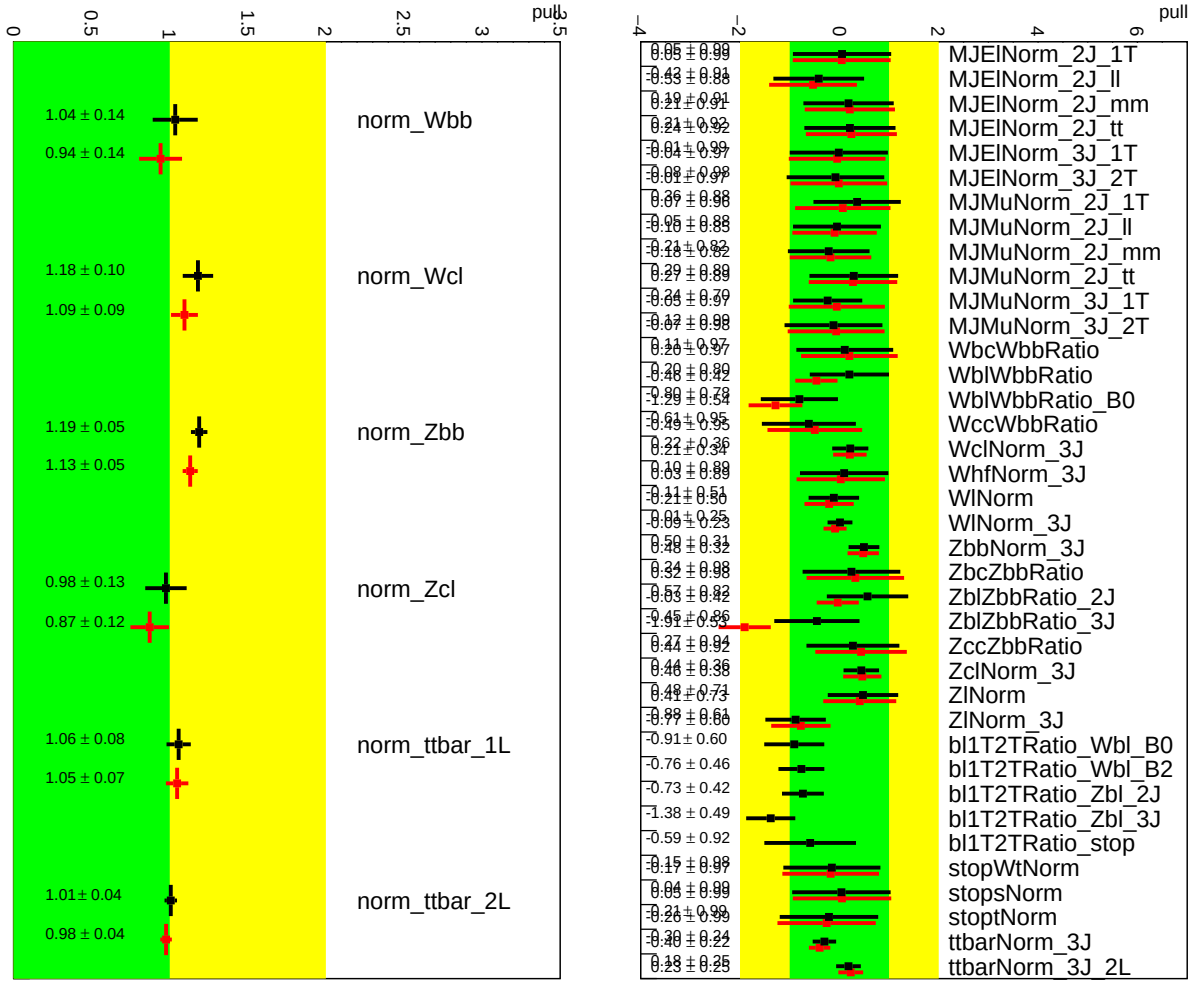
(c) Leading jet b-tagged, 2 leptons



(d) Subleading jet b-tagged, 2 leptons

Figure C.8: b-tagging scores of the two leading jets for 3-jet events with 2 b-tags, inclusive in p_{T}^{V}. The bins correspond to the four working points of the `MV1c` b-tagging algorithm. From left to right: 80, 70, 60 and 50 %. 1b-tag distributions show an excess of Monte Carlo events in the rightmost bin. 2b-tag distributions show no such excess.

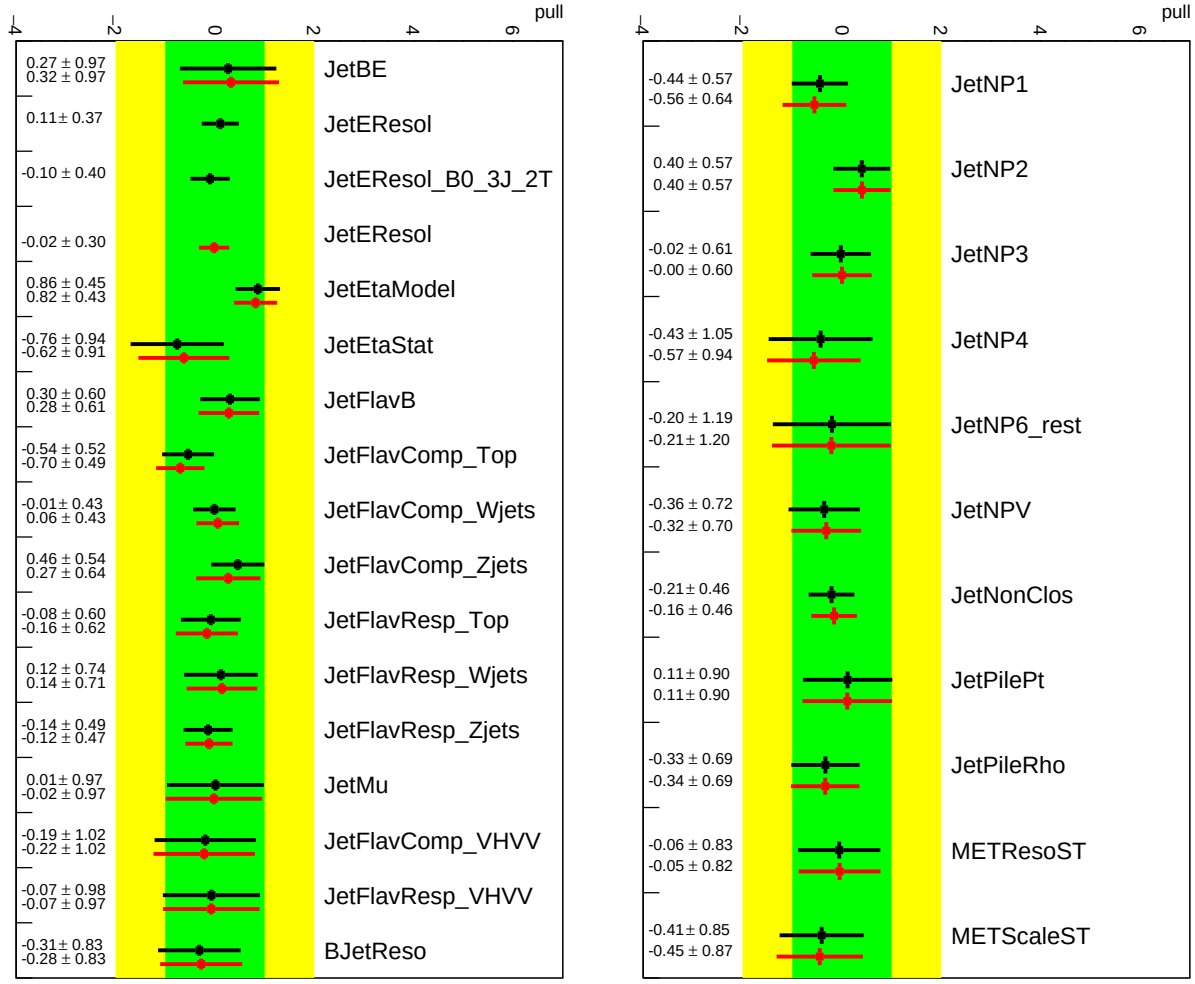
C 4.3 Nuisance-Parameter Pulls of the Improved Likelihood Model



(a) Normalisation factors

(b) Normalisation uncertainties

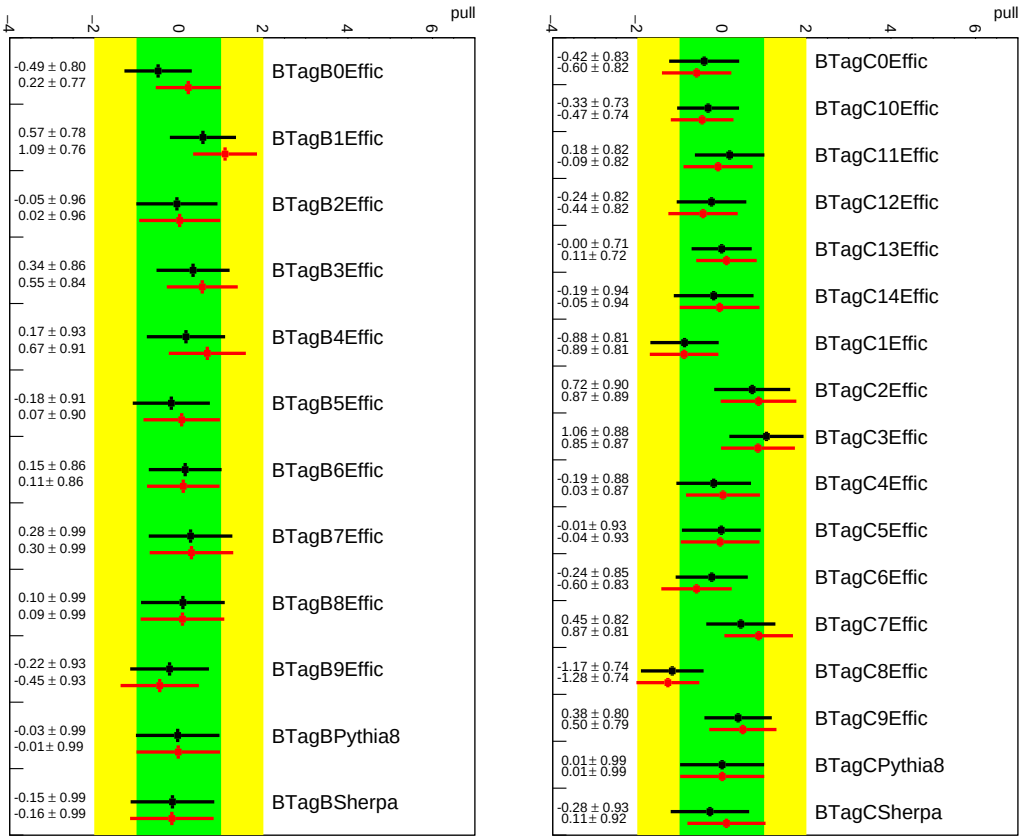
Figure C.9: Comparison of (**black**) improved likelihood model and (**red**) ATLAS likelihood model. Nuisance parameters in the black model without counterparts in the red model have been introduced to mitigate deficits of the ATLAS model.



(a) Jet uncertainties 1

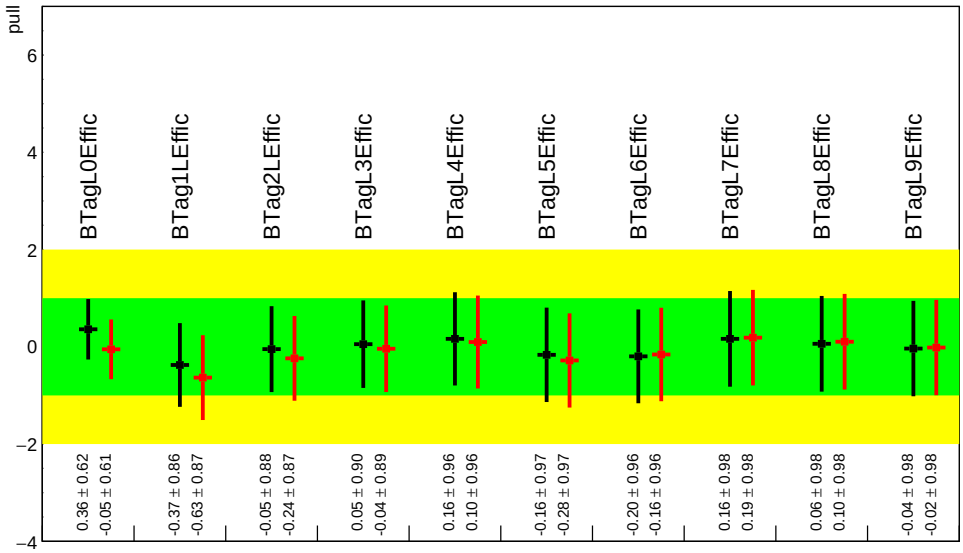
(b) Jet uncertainties 2

Figure C.10: Comparison of (**black**) improved likelihood model and (**red**) ATLAS likelihood model. Nuisance parameters in the black model without counterparts in the red model have been introduced to mitigate deficits of the ATLAS model.



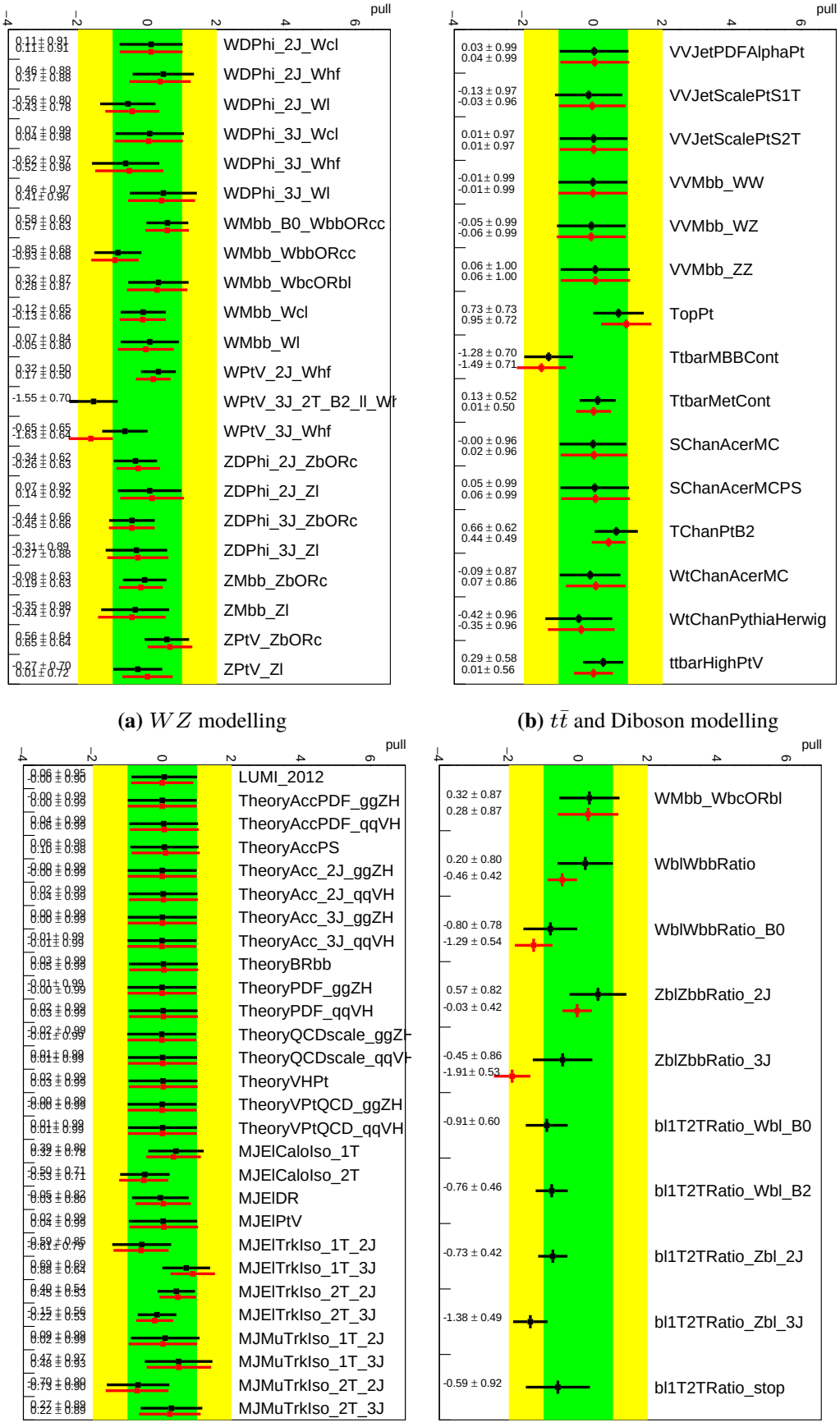
(a) b-tagging for b-jets

(b) b-tagging for c-jets



(c) b-tagging for light jets

Figure C.11: Comparison of (**black**) improved likelihood model and (**red**) ATLAS likelihood model. Nuisance parameters in the black model without counterparts in the red model have been introduced to mitigate deficits of the ATLAS model.



(a) WZ modelling

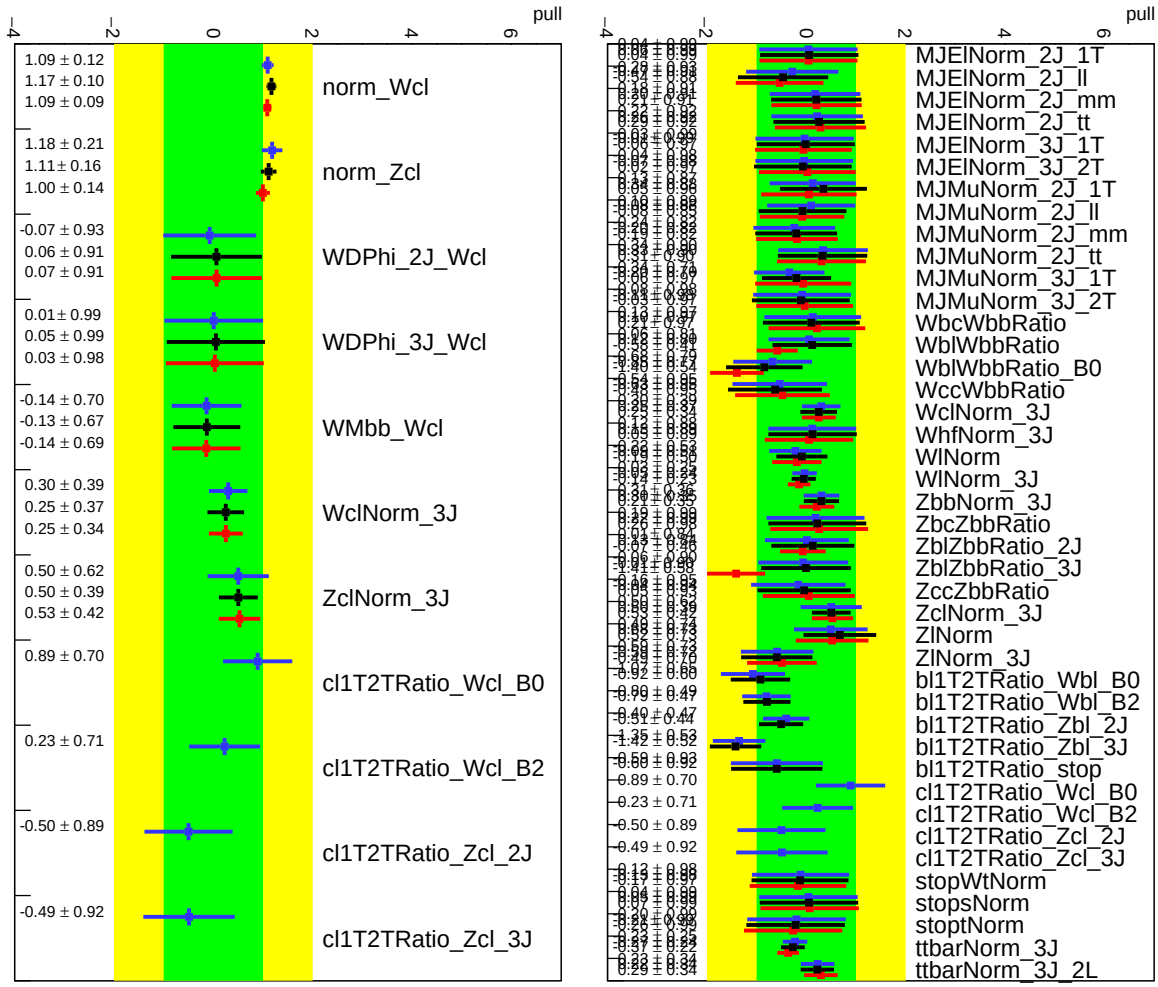
(b) $t\bar{t}$ and Diboson modelling

(c) Other uncertainties

(d) Summary of bl-related uncertainties

Figure C.12: Comparison of (**black**) improved likelihood model and (**red**) ATLAS likelihood model. Nuisance parameters in the black model without counterparts in the red model have been introduced to mitigate deficits of the ATLAS model.

C 4.4 Necessity of an Extension of the Likelihood Model for cl Events



(a) Nuisance parameters connected to cl events

(b) All normalisation uncertainties

Figure C.13: **blue**: Likelihood model with bl and cl uncertainties **black**: Likelihood model with bl uncertainties **red**: ATLAS likelihood model
(a) The introduction of additional extrapolation uncertainties for cl events only marginally affects other cl-related nuisance parameters. Except for `cl1T2TRatio_Wcl_B0` the cl extrapolation uncertainties are consistent with zero. **(b)** All other nuisance parameters for background normalisation uncertainties are consistent with the likelihood model that was extended with only bl uncertainties. It can therefore be concluded that additional cl $1b$-tag to $2b$-tag ratio uncertainties are not necessary to describe the data.

APPENDIX D

Event Yields

The following pages show event yields in all categories of the 1- and 2-lepton channels after the profile-likelihood fit.

Table D.1: Post-fit yields for 1-lepton and 2- or 3-jet events after the MVA selection. The uncertainties are the full post-fit errors including all nuisance parameters with priors, floating normalizations, and the correlations deduced from the data. The difference in the data and the post-fit simulation is compared to σ which is obtained from the sum in simulation as the sum in quadrature of the post-fit and Poissonian error.

Sample	1-tag		2L-tag		2M-tag		2T-tag	
	$p_{\mathrm{T}}^V < 120$	$p_{\mathrm{T}}^V > 120$	$p_{\mathrm{T}}^V < 120$	$p_{\mathrm{T}}^V > 120$	$p_{\mathrm{T}}^V < 120$	$p_{\mathrm{T}}^V > 120$	$p_{\mathrm{T}}^V < 120$	$p_{\mathrm{T}}^V > 120$
VH	98.9	37.1	27.5	10.9	43.4	17.0	48.2	19.2
VV	3 648.1	920.2	190.0	54.1	115.6	30.7	92.8	25.0
$t\bar{t}$	21 709.5	6 400.3	3 539.1	844.8	4 469.0	885.0	4 221.7	732.1
s-top	25 189.7	3 529.9	2 123.7	327.2	1 774.3	252.2	1 364.0	177.2
W+l	120 376.8	23 186.5	2 731.4	577.8	174.9	31.5	4.3	0.3
W+cl	139 808.0	21 992.5	5 188.2	868.3	853.5	133.0	39.2	5.5
W+hf	23 278.0	5 653.5	3 178.2	653.8	2 100.3	402.3	1 132.0	237.8
Z+l	8 160.2	1 268.8	194.2	35.4	13.3	2.6	0.2	0.0
Z+cl	2 948.0	368.8	116.7	17.0	22.0	3.3	0.9	0.2
Z+hf	4 465.9	517.8	335.4	41.8	238.4	28.4	153.9	20.3
MJ_e	30 639.0	3 160.4	1 561.9	128.3	769.1	71.5	385.9	31.4
MJ_μ	14 645.1	217.4	701.9	19.7	286.8	6.9	145.0	1.5
Total	394 967.0	67 253.0	19 888.1	3 579.0	10 860.6	1 864.4	7 588.0	1 250.5
Data	394 974 $\pm$ 629	67 271 $\pm$ 259	19 872 $\pm$ 141	3 533 $\pm$ 59	10 847 $\pm$ 104	1 917 $\pm$ 44	7 621 $\pm$ 87	1 221 $\pm$ 35
				3-jet				
VH	38.7	18.2	8.0	4.3	12.2	6.8	13.8	7.6
VV	2 056.4	646.6	90.9	33.7	39.4	14.5	25.2	11.8
$t\bar{t}$	37 593.9	10 919.9	6 151.3	1 386.0	7 949.2	1 436.5	8 097.8	1 199.6
s-top	12 942.5	2 628.3	1 192.1	263.0	1 084.2	214.3	875.0	149.5
W+l	46 496.3	11 296.6	999.0	267.0	63.6	14.9	2.0	0.2
W+cl	50 729.2	10 901.9	1 766.8	406.5	279.1	61.0	10.9	2.7
W+hf	9 597.1	3 870.1	1 210.6	494.4	774.5	253.3	413.0	147.7
Z+l	3 532.8	585.0	81.5	15.4	6.0	1.2	0.0	0.0
Z+cl	1 479.4	204.3	53.5	9.3	9.0	1.4	0.3	0.0
Z+hf	2 174.6	296.6	151.5	25.7	115.8	21.0	81.3	14.3
MJ_e	14 535.1	859.1	673.4	74.4	343.7	34.8	172.4	16.0
MJ_μ	4 895.2	124.4	282.0	10.3	130.5	3.6	59.6	0.9
Total	186 071.0	42 351.1	12 660.6	2 989.9	10 807.1	2 063.2	9 751.3	1 550.4
Data	186 115 $\pm$ 431	42 293 $\pm$ 206	12 605 $\pm$ 112	3 007 $\pm$ 55	10 833 $\pm$ 104	2 037 $\pm$ 45	9 765 $\pm$ 99	1 581 $\pm$ 40

Table D.2: Post-fit yields for 2-lepton and 2- or 3-jet events after the MVA selection. The uncertainties are the full post-fit errors including all nuisance parameters with priors, floating normalizations, and the correlations deduced from the data. The difference in the data and the post-fit simulation is compared to σ which is obtained from the sum in simulation as the sum in quadrature of the post-fit and Poissonian error.

Sample	1-tag		2L-tag		2M+T-tag	
	$p_{\mathrm{T}}^{V} < 120$	$p_{\mathrm{T}}^{V} > 120$	$p_{\mathrm{T}}^{V} < 120$	$p_{\mathrm{T}}^{V} > 120$	$p_{\mathrm{T}}^{V} < 120$	$p_{\mathrm{T}}^{V} > 120$
VH	27.9	9.9	7.5	2.7	25.7	9.4
VV	774.7	131.6	65.6	11.5	121.7	21.3
$t\bar{t}$	3 481.3	216.5	939.4	45.5	3 050.2	136.5
s-top	347.1	40.5	39.1	4.9	85.2	10.7
W+l	105.1	7.5	2.1	0.2	0.6	0.0
W+cl	95.4	10.1	3.8	0.4	0.6	0.1
W+hf	59.1	10.8	3.7	0.5	4.1	0.5
Z+l	28 283.5	4 070.8	608.5	98.5	37.3	5.6
Z+cl	13 058.9	1 647.8	462.8	63.7	70.9	9.5
Z+hf	21 193.5	2 732.7	2 002.5	276.9	2 969.9	408.8
Total	67 426.5	8 878.1	4 135.0	504.7	6 366.2	602.4
Data	67 414 $\pm$ 260	8 889 $\pm$ 94	4 196 $\pm$ 65	503 $\pm$ 22	6 374 $\pm$ 80	555 $\pm$ 24
			3-jet			
VH	9.9	5.2	2.0	1.2	6.6	3.8
VV	338.1	94.4	19.6	6.2	25.1	7.6
$t\bar{t}$	1 963.5	166.1	413.3	24.1	1 312.4	66.4
s-top	122.6	18.5	14.1	2.6	35.1	6.2
W+l	30.7	3.9	0.7	0.1	0.0	0.0
W+cl	33.3	4.4	1.2	0.2	0.2	0.0
W+hf	28.8	4.6	1.4	0.6	0.9	0.3
Z+l	8 440.4	1 869.9	174.0	42.8	10.5	2.3
Z+cl	4 945.2	958.1	163.7	34.0	23.9	5.0
Z+hf	5 962.9	1 194.8	577.7	129.7	866.5	191.4
Total	21 875.4	4 319.9	1 367.8	241.5	2 281.3	283.1
Data	21 834 $\pm$ 148	4 333 $\pm$ 66	1 357 $\pm$ 37	253 $\pm$ 16	2 310 $\pm$ 48	270 $\pm$ 16

List of Figures

II.1 Particles described by the Standard Model . . . 4
II.2 Feynman diagram showing the scattering of fermions via photon exchange. . . . 6
II.3 Interaction vertices of QCD. . . . 11
II.4 Illustration of Spontaneous Symmetry Breaking (SSB) . . . 15
II.5 Higgs- and Vectorboson Vertices . . . 19
II.6 Interaction vertex between fermions and the Higgs boson. . . . 21
II.7 ATLAS Higgs Discovery . . . 23
II.8 Dominant Higgs production modes at LHC . . . 24
II.9 Higgs production cross sections . . . 25
II.10 NLO (QCD) diagrams for VH production . . . 25
II.11 NNLO (QCD) diagrams for VH production . . . 25
II.12 NLO (EW) diagrams . . . 26
II.13 Higgs decay branching fractions in dependence of m_H . . . 27
II.14 ATLAS and CMS Higgs results at the end of the LHC's Run 1 . . . 28

III.1 Overview of LHC and SPS . . . 30
III.3 Parton distribution functions for protons . . . 32
III.4 Production cross sections at the LHC . . . 32
III.2 A gluon temporarily splitting to quarks. . . . 32
III.5 The ATLAS Detector . . . 34
III.6 Pseudorapidity . . . 34
III.7 Inner detectors of ATLAS . . . 36
III.8 ATLAS Calorimeters . . . 37
III.9 ATLAS EM Calorimeter . . . 38
III.10 ATLAS Tile Calorimeter . . . 38
III.11 Interaction of particles with ATLAS . . . 43
III.12 Event display of a $WH \to \ell\nu b\bar{b}$ candidate event with secondary vertices . . . 49
III.13 NLO Feynman diagram with one loop . . . 50
III.14 Scale dependence of the strong coupling constant α_s . . . 51
III.15 Stages necessary to simulate a proton-proton collision . . . 53
III.16 Gluon radiation in the final state. . . . 55
III.17 Double-counting problem for multi-leg generators . . . 56
III.18 String model of hadronisation used by `Pythia` . . . 57

IV.1 Invariant mass of b-tagged jets. . . . 62
IV.2 Leading-order diagrams of background processes . . . 63
IV.3 Signal regions and background control regions of the $H \to b\bar{b}$ search . . . 70
IV.4 Distribution of H_T before the preselection . . . 73

IV.5 Use of b-tagging for background normalisation/suppression . . . 74
IV.6 b-jet energy correction . . . 76
IV.7 $\Delta\phi$ correction for W+jets events . . . 78
IV.8 Cumulative distribution functions of the b-tagging probability for MV1c . . . 80
IV.9 Isolation variables for electrons and muons . . . 82
IV.10 Multijet Fit . . . 83
IV.11 Correction of the MJ template in $\Delta R(b,b)$ for events with one b-tag . . . 85
IV.12 Illustration of a decision-tree classification . . . 88
IV.13 Overfitting check for the ATLAS $H\to b\bar{b}$ MVA . . . 91
IV.14 Optimisation of $H\to b\bar{b}$ preselection . . . 94
IV.15 Using BDT classifiers to guide the derivation of preselection cuts . . . 96
IV.16 Correlation coefficients of discriminating variables in simulated $VH\to Vb\bar{b}$ events with 2 jets . . . 98
IV.17 Implementation of a decision boundary for strongly correlated variables. . . . 99
IV.18 BDTs trained to separate data from Monte Carlo events . . . 101
IV.19 Illustration of the test statistic q_0 . . . 108
IV.20 Relation of p-value and significance . . . 108
IV.21 Asymptotic vs. exact discovery significance for counting experiments where B is well known and signs of the signal S are searched for in data . . . 111
IV.22 Illustration of the test statistic q_μ . . . 113
IV.23 Profile-likelihood method applied to two nuisance parameters . . . 115
IV.24 Log-likelihood difference for a Poisson-distributed nuisance parameter . . . 116

V.1 Angles to describe VH events . . . 129
V.2 Input variables for the Lorentz-invariant MVA in the 1-lepton channel . . . 132
V.3 Invariant W mass from four-vector products. . . . 134
V.4 Input variables for the Lorentz-invariant MVA in the 2-lepton channel . . . 136
V.5 Loss functions for binary classification . . . 138
V.6 Background rejection of Lorentz-invariant BDT classifiers in the 2-jet region . . . 140
V.7 Lorentz-invariant BDT classifiers for the 2-jet region . . . 141
V.8 Linear correlation coefficients of MVA input variables . . . 142
V.9 Comparison of uncertainties of the signal strength μ . . . 147
V.10 Nuisance-parameter ranking . . . 150
V.11 Comparison of smoothing algorithms for systematic uncertainties . . . 153
V.12 Smoothing of systematic uncertainties with polynomials . . . 155
V.13 Well-behaved nuisance parameters . . . 156
V.14 Nuisance parameters with the highest pulls or strongest constraints . . . 158
V.15 b-tagging response for 1b-tag events . . . 162
V.16 b-tagging response for 1b-tag events . . . 163
V.17 Nuisance-parameter pulls for background normalisations . . . 166

VI.1 Post-fit BDT-classifier distributions for the most sensitive regions . . . 172
VI.2 Signal strength of the Lorentz-invariant VH search . . . 172
VI.3 p_0 values, discovery significance and 95 % $\mathrm{CL_s}$ upper limits for the Lorentz-invariant MVA 173
VI.4 1-lepton event with the highest signal probability . . . 176
VI.5 2-lepton event with the highest signal probability . . . 176
VI.6 Signal strength of the diboson analysis . . . 178

VI.7 BDT classifiers for diboson production . . . 180
VI.8 Nuisance-parameter pulls and impact on μ^{VZ} for the diboson analysis . . . 181

A.1 Selection of diagrams contributing to WW scattering. . . . 203
A.2 Quartic Higgs coupling and subset of radiative-correction diagrams. . . . 203
A.3 Theoretical Higgs-Mass Bounds . . . 205

B.1 Correction of the MJ template in p_{T}^{V} for events with 1b-tag . . . 208
B.2 BDT-driven preselection cuts . . . 209
B.3 BDT-driven preselection cuts . . . 210
B.4 BDTs trained to separate data from Monte Carlo events . . . 211
B.5 BDTs trained to separate data from Monte Carlo events . . . 212

C.1 Additional Lorentz invariants used in the 3-jet region . . . 213
C.2 BDT classifiers in the 3-jet region . . . 214
C.3 Systematic uncertainy for m_{bb} in $t\bar{t}$ events . . . 217
C.9 Comparison of improved and ATLAS likelihood model . . . 223
C.10 Comparison of improved and ATLAS likelihood model . . . 224
C.11 Comparison of improved and ATLAS likelihood model . . . 225
C.12 Comparison of improved and ATLAS likelihood model . . . 226
C.13 Normalisation uncertainties for a likelihood model with cl 1b-tag to 2b-tag ratios . . . 227

List of Tables

II.1 Higgs decay channels predicted for $m_H = 125.09\,\text{GeV}$ . . . 27

III.1 Subsystems and number of readout channels of the ATLAS detector. . . . 40

IV.1 Cross sections times branching ratios and Monte Carlo programs used for estimating backgrounds of the VH search . . . 65
IV.2 Object selection cuts . . . 66
IV.3 Cross sections and branching ratios for associated Higgs production . . . 69
IV.4 Preselection cuts for the 1- and 2-lepton channels . . . 72
IV.5 Hyperparameters for training the BDTs used in the ATLAS $H \to b\bar{b}$ publication . . . 92
IV.6 Comparison of event selection cuts for ATLAS cut analysis, ATLAS MVA baseline proposal and ATLAS Run-1 publication . . . 94
IV.7 Discriminating variables of the ATLAS $VH \to Vb\bar{b}$ analysis . . . 99
IV.8 $VH \to Vb\bar{b}$ likelihood model based on the ATLAS $H \to b\bar{b}$ publication . . . 117
IV.9 Systematic uncertainties of signal and background modelling . . . 118
IV.10 Summary of experimental systematic uncertainties . . . 122

V.1 Hyperparameters for the training of gradient-boosted decision trees used with the Lorentz-invariant variables . . . 139
V.2 Comparison of Lorentz-invariant variables and ATLAS $H \to b\bar{b}$ variables . . . 143
V.3 Comparison of multivariate classifiers . . . 144
V.4 Impact of the $V{+}bl$ 1T $\to$ 2T extrapolation uncertainties . . . 168

VI.1 Comparison with ATLAS Results . . . 175
VI.2 Cross sections and braching ratios for associated Higgs production / diboson production 177
VI.3 VZ to VH ratios in dependence of b-tagging category and lepton channel . . . 177

C.1 Breakdown of systematic uncertainties of μ . . . 215

D.1 Post-fit yields for 1-lepton and 2- or 3-jet events in the MVA selection. . . . 230
D.2 Post-fit yields for 2-lepton and 2- or 3-jet events in the MVA selection. . . . 231