

Searching for the Higgs boson in the $b\bar{b}$ decay channel with the ATLAS experiment

Maria Inês A. J. Ochoa de Castro
University College London


Submitted to University College London in fulfilment
of the requirements for the award of the
degree of **Doctor of Philosophy**

February 6, 2015

Declaration

I, Maria Inês A. J. Ochoa de Castro confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Inês Ochoa

_____

Abstract

The discovery of the Higgs boson by the ATLAS and CMS experiments is one of the main results of Run 1 of the Large Hadron Collider. However, clear evidence for the Higgs boson decay to a pair of b -quarks has not been observed and is crucial to establish the nature of the new found particle. The work presented in this thesis focuses on the search for the Higgs boson in the $VH(b\bar{b})$ channel, where it is produced in association with a leptonically decaying vector boson (W, Z), and decays to a pair of b -quarks.

Prior to the start of LHC operations, the challenges posed by a pp collider to a $H \rightarrow b\bar{b}$ search motivated the development of jet substructure techniques. The boosted regime plays a vital role in the sensitivity of a $VH(b\bar{b})$ search and the topologies where the decay products merge can be recovered by implementing a substructure-based selection. The sensitivity of such an approach in a $VH(b\bar{b})$ search is studied using ATLAS pp collision data, at a centre-of-mass energy of $\sqrt{s} = 7$ TeV. It was found that the sensitivity in the boosted region of the $VH(b\bar{b})$ channel in Run 1 is already fully exploited by the resolved approach. The mass of the Higgs boson, the energies and luminosities delivered and the good performance of anti- k_t jets resulted in little or no gain, at this stage, from performing a jet substructure analysis.

The final ATLAS $VH(b\bar{b})$ Run 1 result is presented. The systematic uncertainties related to the $W + b\bar{b}$ process are estimated and discussed. As an irreducible background to this search, the description of $W + b\bar{b}$ events plays an important role on the final obtained sensitivity.

Finally, in preparation for Run 2 and future colliders, the potential benefits from jet substructure techniques are reviewed at different centre-of-mass energies in the context of a boosted $WH(b\bar{b})$ search. A detailed study of the signal significance as a function of the boost of the system reveals that the region of highest sensitivity is already fully exploited by the resolved reconstruction. A substructure approach is only beneficial in events with boosts greater than 600 GeV, outside the phase-space region of maximum significance.

Acknowledgements

I would like to thank everyone who has supported me since I moved to London to start this PhD. Firstly, my supervisor Jon Butterworth for his guidance and for transmitting so much enthusiasm towards high energy physics; Adam Davison for his patience and essential insights into all aspects of my work; everyone in my office and in the UCL HEP group for such a welcoming work environment; Tim Scanlon, for his indispensable help towards the end of this PhD; and everyone from CERN and other institutions with whom I've had the chance to work along with.

I would like to mention my good friends Buga, Rita, Gil, Nuno, Susana, for never letting me feel too far away from Coimbra. I am also very grateful to all the friends I made during my time at UCL, in particular Cristóvão Vilela, for helping me find a second home in London.

I acknowledge the support from Fundação para a Ciência e Tecnologia, through the grant SFRH/BD/68487/2010. I am also thankful to the Science & Technology Facilities Council for travel and accommodation support during my visits to Geneva.

Finally, I would like to express my gratitude towards my family, especially my parents and my sister Constança, for their love and encouragement.



Preface

The number of people involved in the construction and operation of the Large Hadron Collider and its experiments reflects the complexity of modern high energy physics. The ATLAS experiment brings together 3000 scientists to ensure a successful detector operation and data analysis. Consequently, research work in this field is essentially team work, and it is important to clarify what contents of this thesis are my own.

Chapters 2 to 4 present the general theoretical and experimental basis of the work developed. My contribution to the operation of the ATLAS experiment has been associated with data production for physics analyses with the LHC computing Grid.

The materials presented in Chapters 5 and 6 were produced by me, in close collaboration with Adam Davison. The plots in Figures 5.2, 5.3 and 6.3 were produced by Adam, with software I contributed to.

The analysis presented in Chapter 7 is a product of the work of a large and enthusiastic group, with a core team of approximately 15 people, of whom I was one. My contributions to the result presented in this Chapter and previous ATLAS $VH(b\bar{b})$ results were mostly related to the estimation of systematic uncertainties on background processes such as top-pair, single-top and $W + b\bar{b}$ production. The plots and numbers in Section 7.3.1 of this Chapter were entirely produced by me. Insights from the working group were vital, in particular from the conveners Heather Gray and Giacinto Piacquadio.

The feasibility study presented in Chapter 8 has been performed with essential guidance from Jonathan Butterworth and Tim Scanlon.

Throughout the work presented in this document I had the supervision of Jonathan Butterworth.

Contents

List of Figures	15
List of Tables	21
1. Introduction	23
2. Theoretical framework	25
2.1. The Standard Model of Particle Physics	25
2.2. The Brout-Englert-Higgs Mechanism	28
2.3. The Higgs boson	31
2.4. $VH, H \rightarrow b\bar{b}$ channel	33
2.4.1. Background processes	35
2.5. Boosted VH production	37
2.6. Monte-Carlo event generators	38
3. The ATLAS detector	43
3.1. Large Hadron Collider	43
3.2. ATLAS	45
3.2.1. Inner-Detector	47
3.2.2. Calorimetry	48
3.2.3. Muon Spectrometer	50
3.3. Trigger	51
3.4. Particle Identification	52
3.4.1. Electrons	52
3.4.2. Muons	53
3.4.3. Jets	53
3.4.4. b -tagging	55
3.4.5. Missing Transverse Energy	56

3.5. Detector Simulation	57
3.6. Preparations for Run 2	58
4. Jet algorithms and jet grooming	59
4.1. Jet Algorithms	60
4.2. Jet substructure	61
4.2.1. Splitting and filtering	62
4.2.2. Pruning	64
4.2.3. Trimming	65
4.2.4. N -subjettiness	65
4.3. Performance of grooming techniques in ATLAS	66
5. Jet mass scale validation with W-boson hadronic decays	69
5.1. Jet energy scale for small- R jets	69
5.2. Jet energy and mass scale for large- R jets	70
5.3. Hadronic W -boson with boosted WH analysis	71
5.4. Jet mass scale validation with hadronic W -bosons	73
6. Boosted $WH(b\bar{b})$ searches	77
6.1. Boosted $WH(b\bar{b})$ analysis at 7 TeV	77
6.1.1. Event Selection	77
6.1.2. MC simulation	78
6.1.3. Results	79
6.2. Optimization of jet substructure algorithms	82
6.2.1. Background rejection	84
6.2.2. Resolution of the mass peak	85
6.3. Summary	86
7. Run 1 $VH(b\bar{b})$ search	91
7.1. Analysis Strategy	91
7.1.1. Object and Event Selection	93
7.2. Background composition	95
7.3. Systematic uncertainties	98
7.3.1. $W + b\bar{b}$ background	100
7.3.2. Summary	108
7.4. Statistical Analysis	108

7.5. Results	110
8. Prospects for 14 TeV and higher centre-of-mass energies	113
8.1. Introduction	114
8.2. Event Generation	115
8.3. Event Selection	118
8.4. Signal Efficiency	120
8.5. Physics at 14 TeV and higher energies	122
8.6. Results	127
8.7. Conclusions	134
9. Conclusions	137
A. Optimization of jet substructure algorithms for $H \rightarrow b\bar{b}$	139
B. Modelling of $W + b\bar{b}$ background	141
C. Prospects for 14 TeV and higher centre-of-mass energies	147
C.1. Invariant mass distribution of Higgs boson candidates	147
C.2. Jet veto optimization	148
Bibliography	153

List of Figures

2.1.	Illustration of a Higgs field potential in two dimensions.	30
2.2.	Feynman diagrams representing examples of tree-level Higgs boson production channels at the LHC	34
2.3.	Feynman diagram representing the tree-level VH production and the Higgs boson decay to two b -quarks at leading order.	35
2.4.	Feynman diagram representing $q\bar{q}'$ tree-level production of a W -boson and a pair of b -quarks.	36
3.1.	Schematic representation of the ATLAS detector.	46
3.2.	Schematic representation of the Inner Detector components in the barrel.	49
3.3.	Schematic representation of the calorimeter components in the barrel, end-caps and forward region.	50
3.4.	Schematic representation of the muon spectrometer components in the barrel.	51
3.5.	Schematic view of a B hadron decay inside a jet.	56
3.6.	Light-flavor rejection as a function of b -tagging efficiency for the MV1 algorithm.	57
4.1.	Schematic representation of the splitting/filtering algorithm.	63

4.2.	Mean of the uncalibrated jet mass distribution before and after grooming, using Cambridge/Aachen split and filtered jets in the range $200 \leq p_T < 300$ GeV.	67
4.3.	Invariant mass of the highest- p_T Cambridge/Aachen split and filtered jet in Herwig $Z \rightarrow qq$ events and Powheg+Pythia dijet events.	68
5.1.	Invariant mass distribution of a Cambridge/Aachen split and filtered jet, in events where a W -boson decays leptonically, with $p_T > 200$ GeV	73
5.2.	Jet mass distribution of hadronic W -boson candidates in events where a leptonic W has been identified, using Cambridge/Aachen jets after splitting and filtering and anti- k_t jets after trimming.	76
5.3.	Results of fit to subtracted jet mass distribution in data and Monte Carlo, for anti- k_t jets after trimming.	76
6.1.	Invariant mass distribution of the Higgs candidate using the standard jet collection and the jet substructure method.	81
6.2.	Expected and observed 95% CL upper limits on the WH cross-section using the standard jet collection and the jet substructure method for the identification of the Higgs boson.	81
6.3.	Invariant mass distributions of the reconstructed Cambridge/Aachen jet and anti- k_t dijet system in signal events with $m_H = 120$ GeV, $t\bar{t}$ events and $W + b\bar{b}$ events.	83
6.4.	Invariant mass of the reconstructed Higgs candidate with the resolved approach and several substructure-based methods.	88
6.5.	Invariant mass of the Higgs candidate in the $WH(b\bar{b})$ simulation sample, reconstructed with the resolved approach and several substructure-based methods.	89
7.1.	$p_T(W)$ distributions comparing data and simulation (after global fit).	96
7.2.	$p_T(Z)$ distributions comparing data and simulation (after global fit).	97

7.3. $\Delta\varphi(j_1, j_2)$ distribution in the 1-lepton 2-jet 0-tag region, before a dedicated correction.	99
7.4. $m_{b\bar{b}}$ distribution in $W + b\bar{b}$ events as predicted by different MC generators and their ratio with respect to SHERPA.	102
7.5. $m_{b\bar{b}}$ distribution in $W + b\bar{b}$ events as predicted by SHERPA simulated events where the B hadrons are originating exclusively from the matrix element (SHERPA BB) and from either matrix element or parton shower (SHERPA B-filter).	103
7.6. Distribution of jet multiplicity in $W + b\bar{b}$ events as predicted by different MC generators and their ratio with respect to SHERPA.	104
7.7. $p_T(W)$ spectrum in $W + b\bar{b}$ events as predicted by different MC generators and their ratio with respect to SHERPA.	104
7.8. $\Delta\varphi(j_1, j_2)$ distribution in $W + b\bar{b}$ events as predicted by different MC generators and their ratio with respect to SHERPA.	107
7.9. $m_{b\bar{b}}$ distributions comparing data and simulation (after global fit). . .	109
7.10. Observed and expected 95% CL upper limits on the VH production cross section with respect to the SM value, for the combination of lepton channels and datasets.	111
8.1. Cross-sections of SM and Higgs production processes as predicted as a function of the centre-of-mass energy for a pp collider.	114
8.2. Distribution of the logarithm of the proton momentum fraction x carried by the incoming partons and their relative rate, in WH events.	116
8.3. Differential cross-section times branching ratio for $t\bar{t}$ and $W + b\bar{b}$ events with respect to the signal $WH(b\bar{b})$, as a function of $p_T(W)$, normalized to same area.	118
8.4. Signal efficiency with resolved and substructure selection approaches.	121
8.5. Pseudo-rapidity of the muon in $WH(b\bar{b})$ events.	122

8.6. Complementary cumulative distribution of signal events uniquely selected by the substructure-based approach and by the resolved approach.	123
8.7. $W(\mu\nu)H(b\bar{b})$ differential cross-section with respect to $p_T(W)$, times branching ratio and selection efficiency.	123
8.8. Number of anti- k_t $R = 0.4$ jets with $p_T > 20$ GeV and $ \eta < 5.0$ in signal and background events, at $\sqrt{s} = 8, 14, 33, 100$ TeV.	124
8.9. Number of Cambridge/Aachen $R = 1.2$ split and filtered jets with $p_T > 150$ GeV and $ \eta < 5.0$ in signal and background events, at $\sqrt{s} = 8, 14, 33, 100$ TeV.	125
8.10. Dijet invariant mass after the resolved selection and jet veto, for signal, $t\bar{t}$ and $W + b\bar{b}$ events.	127
8.11. Invariant mass of the Cambridge/Aachen jet after the substructure selection and jet veto, for signal, $t\bar{t}$ and $W + b\bar{b}$ events.	128
8.12. Dijet invariant mass after the resolved selection and invariant mass of the Cambridge/Aachen jet after the substructure selection, with the jet veto, for signal, $t\bar{t}$ and $W + b\bar{b}$ events at $\sqrt{s} = 14$ TeV.	129
8.13. Dijet invariant mass after the resolved selection and jet veto, for signal, $t\bar{t}$ and $W + b\bar{b}$ events, at $\sqrt{s} = 14$ TeV and in different bins of $p_T(W)$	130
8.14. Invariant mass of the Cambridge/Aachen jet after the substructure selection and jet veto, for signal, $t\bar{t}$ and $W + b\bar{b}$ events, at $\sqrt{s} = 14$ TeV and in different bins of $p_T(W)$	131
8.15. Signal-to-background ratio as a function of $p_T(W)$, at $\sqrt{s} = 8, 14, 33, 100$ TeV.	132
8.16. S/\sqrt{B} as a function of $p_T(W)$, at $\sqrt{s} = 8, 14, 33, 100$ TeV, considering a total integrated luminosity of 3000 fb^{-1}	133
8.17. S/\sqrt{B} as a function of $p_T(W)$, at $\sqrt{s} = 14$ TeV, considering a total integrated luminosity of 150 and 300 fb^{-1}	134
A.1. Distribution of N -subjettiness ratio τ_{21} for anti- k_t $R = 1.5$ trimmed jets.	140

B.1. Distribution of the invariant mass of leading B hadrons matched to leading jets, as predicted by different MC generators.	141
B.2. Distribution of jet multiplicity in $W + b\bar{b}$ events as predicted by the baseline aMC@NLO sample and by samples where the μ_R and μ_F parameters have been varied.	142
B.3. Distribution of relative yields in each $p_T(W)$ bin, in $W + b\bar{b}$ events, as predicted by aMC@NLO samples where the μ_R and μ_F parameters have been varied.	143
B.4. Distribution of jet multiplicity and $p_T(W)$ in $W + b\bar{b}$ events as predicted by SHERPA simulated events where the B hadrons are originating exclusively from the matrix element (SHERPA BB) and from either matrix element or parton shower (SHERPA B-filter)	144
B.5. Distribution of jet multiplicity (top) and $p_T(W)$ (bottom) in $W + b\bar{b}$ events as predicted by the baseline aMC@NLO sample and by aMC@LO.	145
B.6. $\Delta\varphi(j_1, j_2)$ distribution in $W + b\bar{b}$ events as predicted by the baseline aMC@NLO sample and by samples where the μ_R and μ_F parameters have been varied.	146
C.1. Invariant mass distribution of the Higgs boson candidate using the resolved and substructure-based selections, at $\sqrt{s} = 8, 14, 33, 100$ TeV.	147

List of Tables

2.1. Gauge bosons in the Standard Model. [1]	26
2.2. The Standard Model leptons and corresponding electric charge and mass.	27
2.3. The Standard Model quarks and corresponding electric charge and mass.	27
2.4. Inclusive cross-section values of the $W^\pm H$ process at the LHC centre-of-mass energies of 7, 8 and 14 TeV, for a Higgs boson mass of $m_H = 125$ GeV.	35
2.5. Inclusive cross-section values of the ZH process at the LHC centre-of-mass energies of 7, 8 and 14 TeV, for a Higgs boson mass of $m_H = 125$ GeV.	36
3.1. Design resolutions of the ATLAS sub-detectors.	47
6.1. List of Monte-Carlo samples used for signal and background simulation.	79
6.2. Values of S/B and S/\sqrt{B} for the best performing substructure methods.	84
6.3. Values of the figures of merit as calculated for the resolved and substructure reconstruction methods.	86
7.1. Event categories used in the 1-lepton analysis (8 TeV dataset).	93
7.2. Selection cuts applied in the 1-lepton channel, for both dijet mass and MVA analyses (8 TeV dataset).	95
7.3. MC simulation samples for signal and background processes.	98

7.4. MC samples used in the estimation of $W + b\bar{b}$ modelling uncertainties at particle-level.	101
7.5. MC samples used in the estimation of $W + b\bar{b}$ modelling uncertainties at detector-level.	106
8.1. Parameters used in the generation of $WH(b\bar{b})$, $t\bar{t}$ and $W + b\bar{b}$ processes with aMC@NLO.	116
8.2. Cross-section values calculated with aMC@NLO, at different centre-of-mass energies, including QCD corrections at NLO.	118
C.1. Optimization of jet veto at $\sqrt{s} = 8$ TeV, considering the (di-)jet invariant mass window of 90 to 140 GeV.	149
C.2. Optimization of jet veto at $\sqrt{s} = 14$ TeV, considering the (di-)jet invariant mass window of 90 to 140 GeV.	150
C.3. Optimization of jet veto at $\sqrt{s} = 33$ TeV, considering the (di-)jet invariant mass window of 90 to 140 GeV.	151
C.4. Optimization of jet veto at $\sqrt{s} = 100$ TeV, considering the (di-)jet invariant mass window of 90 to 140 GeV.	152

Chapter 1.

Introduction

“I want to play music that draws a picture of the world and its space.”

— Jimi Hendrix

After over two years of successful operation, the Large Hadron Collider (LHC) has provided immense amounts of proton and heavy ion collisions at unprecedented energies. Understanding nature at a fundamental level can only be accomplished through numerous technological and scientific advances, realized by large collaborations of scientists and engineers. The discovery of the long-sought Higgs boson has changed the landscape of high energy physics and provided one of the first benchmarks of the LHC era. However, many phenomena remain unexplained, and expectations for the next years of LHC operation and for other cutting-edge particle physics experiments are very high.

The work presented in this thesis is highly focused on one of the (as of yet) unobserved signatures of the Standard Model Higgs boson: its decay to a pair of bottom-quarks. Before its discovery, the possible existence of a boson with a mass in the region of 115-125 GeV was explored by introducing jet substructure techniques, whose scope has in the meantime been extended to other searches for new physics. The performance of these techniques in the context of a $H \rightarrow b\bar{b}$ search is an important topic of this work.

This document will begin with a brief discussion of the theoretical background to a Higgs boson search in a hadronic collider, in Chapter 2. The experimental setup of the ATLAS detector is described in Chapter 3. The definition of hadronic jets is

given a special emphasis on Chapter 4, where jet algorithms and jet substructure techniques are introduced.

Chapter 5 demonstrates the validation of the ATLAS jet mass scale by reconstructing the hadronically-decaying W -boson mass peak, taking advantage of jet substructure techniques.

The data collected during 2011 at a centre-of-mass energy of 7 TeV provided the first opportunity to test the sensitivity of the substructure approach to a light Higgs boson decaying to a pair of b -quarks. These results are presented in Chapter 6.

Chapter 7 presents the result of the ATLAS Higgs boson search in the $b\bar{b}$ decay mode and produced in association with a vector boson. The $VH(b\bar{b})$ result is produced with the entirety of the data collected during the first run of the LHC.

In preparation for the next LHC run at centre-of-mass energies of 13, 14 TeV and possible future collider machines, a particle-level study is presented in Chapter 8. The potential usefulness of jet substructure techniques in the boosted $WH(b\bar{b})$ channel is revisited.

Finally, a summary and conclusion of the preceding studies are presented in Chapter 9.

Chapter 2.

Theoretical framework

“All the good music has already been written by people with wigs and stuff.”

— Frank Zappa

During the twentieth century, major achievements in theoretical and experimental physics led to the development of Quantum Field Theories describing the Electromagnetic, Weak and Strong interactions. Together with Gravity, these are the forces governing all known fundamental particles.

The unification of quantum electrodynamics (QED) and weak theory, by Glashow, Salam and Weinberg [2–4], laid the theoretical ground for the formulation of quantum chromodynamics (QCD) [5–8]. The resulting Standard Model of Particle Physics (SM), combining three of the four fundamental forces in nature, is a mathematical framework which has provided numerous predictions confirmed by experimental observations.

In this chapter, basic formulations of the SM are introduced and the Brout-Englert-Higgs mechanism [9–13] is briefly explained.

2.1. The Standard Model of Particle Physics

Elementary particles and their interactions are described by the laws of quantum mechanics and special relativity. A theory of quantized fields relies on symmetry principles to consistently describe interactions. In other words, the invariance of the dynamical properties of a system under a continuous symmetry transformation

translates into the conservation of a physical property, as stated by Noether's theorem. The principle of energy conservation, for example, is in this way incorporated through the time translation symmetry.

Gauge theories are described by a Lagrangian that remains invariant under a continuous group of local¹ transformations. Different mathematical configurations of the fields will therefore result in equivalent observable physical states. A theory with predictive power must fulfill this requirement.

The Standard Model is a gauge quantum field theory which is invariant under transformations governed by the $SU(3)_c \times SU(2)_L \times U(1)_Y$ groups. The $SU(3)_c$ symmetry group is associated to the QCD sector of the Lagrangian, which describes the interactions between color-charged particles (quarks and gluons). The $SU(2)_L \times U(1)_Y$ symmetry corresponds to the electroweak sector, with the weak isospin T_3 and weak hypercharge Y for the corresponding generators. This sector incorporates the interaction of the photon with electrically charged particles, as well as the W^\pm weak couplings to left-handed particles and Z^0 couplings to right and left-handed particles.

The SM postulates that elementary particles are divided into two groups: gauge bosons and fermions, listed in Table 2.1, Tables 2.2 and 2.3, respectively. The gauge bosons, responsible for mediating the interactions, are a consequence of the gauge invariance built into the model.

Table 2.1.: Gauge bosons in the Standard Model. [1]

Boson	Interaction	Electric Charge (Q)	Mass (GeV)
gluon (g)	strong	0	0
photon (γ)	electromagnetic	0	0
W^\pm	weak	± 1	80.385 ± 0.015
Z^0	weak	0	91.1876 ± 0.0021

Each term in the SM Lagrangian can be represented by a Feynman rule, comprising propagators and interaction vertices (each with an amplitude proportional to the

¹Local transformations are dependent on the space-time coordinates of the system.

Table 2.2.: The Standard Model leptons and corresponding electric charge and mass. The antiparticles are implicit, with opposite sign electric charges. Neutrino masses are larger than zero for at least two generations. [1]

Leptons			
	Particle	Electric Charge (Q)	Mass (GeV)
1 st Generation	ν_e	0	$< 10^{-9}$
	e	-1	0.5×10^{-3}
2 nd Generation	ν_μ	0	$< 10^{-9}$
	μ	-1	106×10^{-3}
3 rd Generation	ν_τ	0	$< 10^{-9}$
	τ	-1	1.78

Table 2.3.: The Standard Model quarks and corresponding electric charge and mass. The antiparticles are implicit, with opposite sign electric charges. [1]

Quarks			
	Particle	Electric Charge (Q)	Mass (GeV)
1 st Generation	u	$+2/3$	2×10^{-3}
	d	$-1/3$	5×10^{-3}
2 nd Generation	c	$+2/3$	1.28
	s	$-1/3$	95×10^{-3}
3 rd Generation	t	$+2/3$	173.3
	b	$-1/3$	4.18

coupling constant of the interaction). The full calculation of the cross-section associated to a given process involves real and virtual corrections to the leading order (LO) diagram². This is mathematically equivalent to performing the calculation with a perturbative series, expanding on the coupling constant parameter. However, convergence of the series is not guaranteed, since the perturbative regime is not always valid. A relevant example to this work is the case of the strong force at energies below the GeV scale.

The prediction of the 3rd generation of fermions and W^\pm and Z^0 masses are good examples of the internal consistency of the SM. However, certain phenomena are not incorporated in the theory, and in reality it only attempts to describe a very small fraction of the Universe. It is not the aim of this work to give a thorough description of the SM of Particle Physics, but rather to motivate the search and study of the Higgs boson.

2.2. The Brout-Englert-Higgs Mechanism

In the same way the massless photon arises in QED by requiring local gauge invariance, when the theory is extended to include the weak force, a set of massless gauge bosons arise from the $SU(2) \times U(1)$ sector: W_i^μ and B^μ .

However, this is not in agreement with what is observed in nature, i.e., with the short-range characteristic of the nuclear force, fundamentally different from electromagnetism. Furthermore, the introduction in the Lagrangian of mass terms associated with these fields, while in agreement with the interaction range, would break local gauge invariance.

By introducing a complex scalar field that transforms as a doublet under $SU(2)$, $\phi = (\phi^+, \phi^0)$, consisting of four real degrees of freedom, the resulting Lagrangian terms corresponding to the couplings of ϕ to the gauge fields are³:

$$\mathcal{L}_\phi = (\partial^\mu \phi^\dagger + ig_W W^\mu \cdot T \phi^\dagger + \frac{i}{2} g'_W Y B^\mu \phi^\dagger)(\partial_\mu \phi - ig_W W_\mu \cdot T \phi - \frac{i}{2} g'_W Y B_\mu \phi) - V(\phi^\dagger, \phi), \quad (2.1)$$

²Additional orders of calculation are referred to as next-to-leading order (NLO), next-to-next-to-leading order (NNLO), etc.

³ g_W and g'_W are the $SU(2) \times U(1)$ coupling constants.

where one can define the potential $V(\phi^\dagger, \phi)$ with the form:

$$V(\phi^\dagger, \phi) = \lambda(\phi^\dagger\phi)^2 - \mu^2\phi^\dagger\phi. \quad (2.2)$$

This potential has the shape shown in Figure 2.1, if both λ and μ^2 are positive. The circle of degenerate minima correspond to $|\phi| = \sqrt{\mu^2/(2\lambda)} = v/\sqrt{2}$, where v is the vacuum expectation value⁴.

A particular direction in the $SU(2)_L$ space is chosen and the symmetry is broken by defining an expectation value of the vacuum state as

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (2.3)$$

This is equivalent to allowing the electrically-neutral component of the doublet to acquire a vacuum expectation value, which defines the scale of the electroweak symmetry breaking. This ground state remains invariant when the 3rd component of the weak isospin and the weak hypercharge operators act on it, corresponding to the conservation of the electric charge generator $Q = T_3 + Y$. Such a choice of ground state breaks both $SU(2)_L$ and $U(1)_Y$ gauge symmetries, while remaining invariant under $U(1)_{em}$, originating the massless photon. This type of continuous symmetry breaking, where the Lagrangian remains invariant but the ground state is altered, is known as spontaneous symmetry breaking.

Fluctuating $\phi(x)$ around the ground state (as shown in Equation 2.4), four degrees of freedom are introduced, the real fields $\xi_i(x)$ and $H(x)$. A gauge transformation of the following form can be done:

$$\phi(x) \rightarrow U(\xi)\phi(x) = U(\xi)e^{i(\xi_a(x)T^a - \xi_3(x)Y)/v} \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix}, \quad (2.4)$$

$$T \cdot W^\mu \rightarrow UT \cdot W^\mu U^{-1} + \frac{i}{g_W}(\partial^\mu U)U^{-1}. \quad (2.5)$$

⁴This is the simplest potential configuration with a non-zero vacuum expectation value.

Consequently, by fixing the gauge appropriately, the $\xi_i(x)$ fields are absorbed into the corresponding transformations of the gauge bosons, creating a longitudinal polarization which allows them to be massive. The only field remaining, $H(x)$ corresponds to a massive scalar boson, the Higgs boson, an excitation of the Higgs field $\phi(x)$ around its ground state.

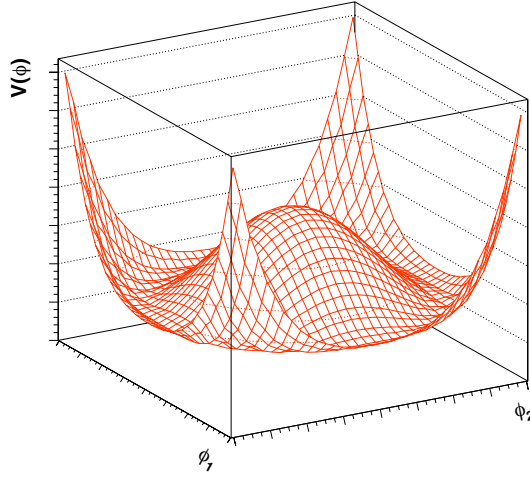


Figure 2.1.: Illustration of a Higgs field potential in two dimensions.

Rewriting Equation 2.1 in terms of $\phi(x)$ as it fluctuates around the vacuum, the first term comes out as:

$$\mathcal{L}_M = \frac{v^2}{8} [(g_W W_\mu^3 - g'_W B_\mu)(g_W W^{3\mu} - g'_W B^\mu) + 2g_W^2 W_\mu^- W^{+\mu}], \quad (2.6)$$

using the definition $W^\pm = (W^1 \mp iW^2)/\sqrt{2}$. The first term is not diagonal in W_μ^3 and B_μ , and so can be redefined in terms of two independent fields Z_μ and A_μ , such that:

$$A_\mu = \cos \theta_W B_\mu + \sin \theta_W W_\mu^3, \quad (2.7)$$

$$Z_\mu = -\sin \theta_W B_\mu + \cos \theta_W W_\mu^3. \quad (2.8)$$

It is now clear that the photon and Z bosons are the mass eigenstates corresponding to the mixing of the fields W_μ^3 and B_μ . From Equation 2.6, the mass terms for the

W and Z can be identified as

$$M_W = \frac{1}{2}vg_W, \quad (2.9)$$

$$M_Z = \frac{1}{2}v\sqrt{g_W^2 + g_W'^2}. \quad (2.10)$$

The vacuum expectation value of the Higgs field can finally be extracted from the relation between M_W and the Fermi coupling constant G_F : $v \approx 246$ GeV.

The strength of this theory was clear once neutral currents were observed in 1973 by Gargamelle [14], and the W^\pm and Z -bosons discovery established by UA1 [15, 16] and UA2 [17, 18] at CERN, in 1983.

Additionally, Lagrangian terms can be defined of the form $G_f(\bar{f}_L\phi f_R + \bar{f}_R\phi^\dagger f_L)$ where the fermions and the Higgs doublet are coupled. These correspond to fermion mass terms, which make it possible for quark and charged lepton masses to be accommodated. The couplings G_f between the fields are, however, arbitrary, in contrast to what happens for the gauge bosons, and therefore the model can't predict fermion masses. Nonetheless, as in the gauge boson case, the couplings to the fermions are still proportional to their masses, which allows for these predictions to be tested experimentally.

To conclude, the Higgs boson mass can be extracted from the fluctuations around the ground state of Equation 2.2, as $m_H^2 = 2\lambda v^2$. As the parameter λ is unknown, so is the Higgs boson mass. For many decades, the existence of the Higgs boson was the missing piece of the puzzle of electroweak symmetry breaking through the Brout-Englert-Higgs mechanism. Lower and upper bounds were defined by theoretical arguments (see, for example, Ref. [19]). Furthermore, electroweak precision data was used to place constraints on m_H , where a preferred value of 94 GeV was obtained, with the exclusion of masses greater than 152 GeV at 95% confidence level [20].

2.3. The Higgs boson

The existence of a Higgs field can be proved by the detection of its excitation, the Higgs boson. The search for the Higgs boson was one of the main motivations for

the construction of the LHC and was also a key component of the physics programs at other colliders, such as LEP and the Tevatron.

On the 4th of July, 2012, the two general-purpose experiments at the LHC (ATLAS and CMS) announced the discovery of the Higgs boson, more than 40 years after the formulation of the Brout-Englert-Higgs mechanism and its prediction. Both experiments later published the observation of the new particle at a mass of approximately 125 GeV with a significance of 5 to 6 standard deviations each [21, 22]. The Tevatron experiments (D0 and CDF) also reported a combined excess of events with respect to the null hypothesis, with 3 standard deviations at a mass of $m_H = 125$ GeV [23].

At a pp collider, the Higgs boson can be produced through four different production mechanisms. For each one, an example tree-level diagram is shown in Figure 2.2. The most abundant production mechanisms at the LHC are gluon-gluon and vector-boson fusion, followed by associated production with a vector boson ($V = W, Z$) and associated production with a $t\bar{t}$ pair. The Higgs particle preferred decay channels are heavily dependent on its mass, and therefore many different searches were performed by both ATLAS and CMS collaborations to target the different mass points that were excluded until the discovery.

Throughout this work, the focus will be on a SM Higgs particle with a mass of 125 GeV (or 120 GeV, for results produced before the discovery). A Higgs particle with such a mass decays mainly to a pair of b -quarks, followed by the WW^* , gg , $\tau\bar{\tau}$, ZZ^* and $\gamma\gamma$ channels. Evidence for the Higgs boson is present so far on the individual $H \rightarrow \gamma\gamma$, $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$, $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ and $H \rightarrow \tau^-\tau^+$ decay channels [24–27]. The ATLAS mass measurement performed in the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ channels obtains $m_H = 125.5 \pm 0.2(\text{stat})_{-0.6}^{+0.5}(\text{sys})$ GeV [28]. When combining all of the above mentioned searches, the observed signal strength is $\mu = 1.30 \pm 0.12(\text{stat})_{-0.11}^{+0.14}(\text{sys})$ [29].

Evidence for the coupling of the Higgs particle to fermions is currently present on the $\tau\bar{\tau}$ channel, where a signal strength of $\mu = 1.4_{-0.4}^{+0.5}$ has been measured in ATLAS [27], and on the combined search result published by the Tevatron, with a sensitivity dominated by the $b\bar{b}$ decay channel. Furthermore, the coupling to fermions is already indirectly probed by constraints on the $gg \rightarrow H$ production channel and on the $H \rightarrow \gamma\gamma$ decay channel, which proceed through a fermion-loop (see Figure

2.2). The $H \rightarrow b\bar{b}$ search results in the associated vector boson production channel are the subject of Chapter 7.

Since the discovery, several measurements have been performed in ATLAS and CMS to understand the properties of the boson, such as its spin and couplings [29,30]. The results are compatible with the SM expectations within the uncertainties, deferring the confirmation of the boson's SM nature to the next run of the LHC. Given the scope of this work, extensions to the SM Higgs sector will not be covered.

2.4. $VH, H \rightarrow b\bar{b}$ channel

A Higgs boson with an approximate mass of 125 GeV will decay dominantly to a pair of b -quarks, given the phase-space constraints, the color factor and the b -quark mass. The branching ratio has been calculated to be $BR(H \rightarrow b\bar{b}) = 58.3_{-1.8}^{+4.4}(\%)$, including N³LO QCD and NLO EW corrections [31,32].

Despite the large fraction of Higgs bosons decaying to $b\bar{b}$, in a hadron machine such as the LHC a clear observation of such a signature is not possible due to the enormous cross-section of QCD multijet processes that are backgrounds to this search.

Instead of searching inclusively for a pair of b -quarks, a common strategy is to search for the associated production of a Higgs and a vector boson (Figure 2.3), also known as Higgs-strahlung. When the vector boson decays leptonically, this process has a clearer experimental signature, which helps reducing backgrounds and triggering events. A search on this channel can combine three different final-states, according to the vector boson and its decay mode: $W^\pm \rightarrow \ell^\pm \nu$, $Z \rightarrow \ell\bar{\ell}$ and $Z \rightarrow \nu\bar{\nu}$.

The dominant VH diagrams involve a virtual vector boson splitting into a pair of vector and scalar bosons. The ZH channel has an extra contribution from gg -initiated channels, which amounts to approximately 4% (6%) of the ZH production at the LHC at $\sqrt{s} = 8$ TeV ($\sqrt{s} = 14$ TeV) [33,34].

The NNLO QCD corrections to these processes have been derived in Ref. [35] and have a small impact in the total cross-section, while the NLO QCD corrections increase the LO cross-section value by 30% at the LHC. Electroweak corrections at NLO are also applied, reducing the cross-section by less than 10% [36]. The resulting

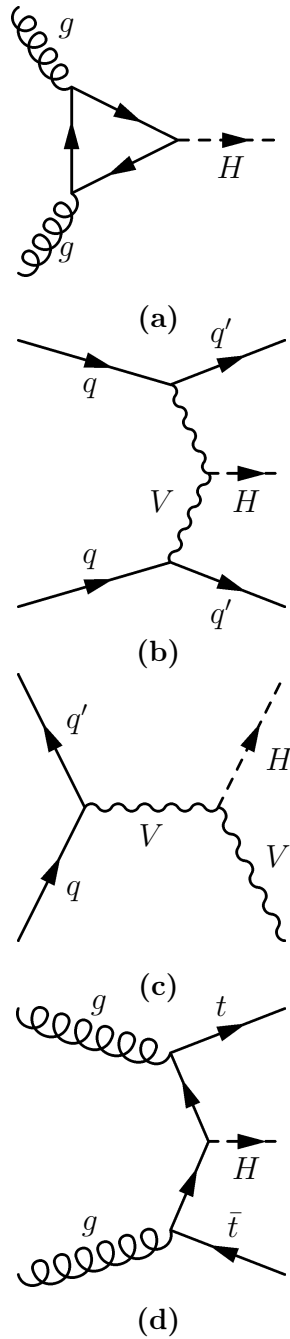


Figure 2.2.: Feynman diagrams representing examples of tree-level Higgs boson production channels at the LHC: a) gluon-gluon fusion, b) vector-boson fusion, c) association with vector-boson and d) association with a top and an anti-top quark.

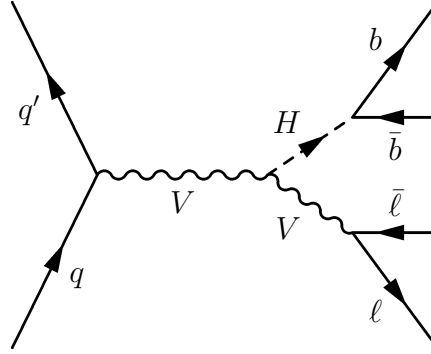


Figure 2.3.: Feynman diagram representing the tree-level VH production and the Higgs boson decay to two b -quarks at leading order.

Table 2.4.: Inclusive cross-section values of the $W^\pm H$ process at the LHC centre-of-mass energies of 7, 8 and 14 TeV, for a Higgs boson mass of $m_H = 125$ GeV. [37]

WH	central value (pb)	(PDF + α_S) (%)	scale (%)
$\sqrt{s} = 7$ TeV	0.58	± 2.6	± 0.9
$\sqrt{s} = 8$ TeV	0.70	± 2.3	± 1.0
$\sqrt{s} = 14$ TeV	1.50	± 3.8	+0.3/-0.6

cross-section values at centre-of-mass energies of 7, 8 and 14 TeV are shown in Tables 2.4 and 2.5.

An alternative strategy for the $H \rightarrow b\bar{b}$ search is to look for the associated production of a Higgs boson with a $t\bar{t}$ pair. The latest sensitivity study for this channel in ATLAS has been reported in Ref. [38]. Searches for $H \rightarrow b\bar{b}$ in the vector-boson fusion production channel and in association with a b -quark are also ongoing in ATLAS.

2.4.1. Background processes

The SM processes that constitute an irreducible background to this search are $W + b\bar{b}$ (shown at tree level in Figure 2.4) and $Z + b\bar{b}$ production. The main focus of this work will be on the $W + b\bar{b}$ process, but the differences and similarities to $Z + b\bar{b}$ will be addressed when relevant.

Table 2.5.: Inclusive cross-section values of the ZH process at the LHC centre-of-mass energies of 7, 8 and 14 TeV, for a Higgs boson mass of $m_H = 125$ GeV. [37]

ZH	central value (pb)	(PDF + α_S) (%)	scale (%)
$\sqrt{s} = 7$ TeV	0.34	± 2.7	± 2.9
$\sqrt{s} = 8$ TeV	0.42	± 2.5	± 3.1
$\sqrt{s} = 14$ TeV	0.88	± 3.7	+2.7/-1.8

The $W + b\bar{b}$ leading order diagram has a single possible configuration, with a $q\bar{q}'$ initial state, while $Z + b\bar{b}$ can be produced through $q\bar{q}$ and gg initial states. The gg -initiated processes, taking $q = u, d, c, s^5$, contribute to the $W + b\bar{b} + 1$ parton and $Z + b\bar{b} + 1$ parton final states, hence introducing a production channel which does not exist at LO. This effect is more pronounced in the $W + b\bar{b}$ case, and especially at LHC energies, where there is an increase of a factor 2-3 in the LO cross section [39].

Calculations in NLO QCD of $W + b\bar{b}$ (4FNS), $W + b$ inclusive and Wbj (5FNS) cross-sections with a finite b -quark mass have been compared to the measurements performed by D0, CDF and ATLAS. There is a discrepancy of 1.5σ and 2.8σ in the ATLAS and CDF fiducial cross-section measurements, respectively, both found to be larger than the corresponding theoretical calculations [40, 41]. The measurement of this SM process is very important to constrain uncertainties on the $VH(b\bar{b})$ result, such as those associated with the rate of gluons splitting to $b\bar{b}$.

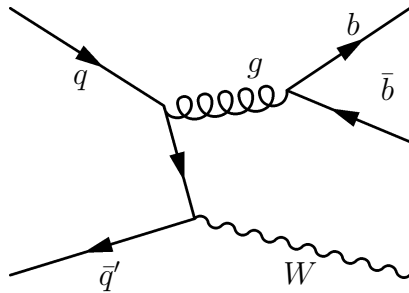


Figure 2.4.: Feynman diagram representing $q\bar{q}'$ tree-level production of a W -boson and a pair of b -quarks.

Top pair production constitutes one of the largest backgrounds to the $WH(b\bar{b})$ search, due to its relatively large cross-section and to a signature which easily fakes the signal one. The top-quark decays before hadronizing, through the $t \rightarrow Wb$ channel

⁵This corresponds to the four-flavor-number-scheme (4FNS). In the five-flavor-number-scheme (5FNS), the bottom-quark is included in the possible initial state quarks.

with a branching ratio very close to 1. The final state topology of a $t\bar{t}$ event depends therefore on the W decay products. Semi-leptonic and dileptonic configurations (when either one or both W 's decay to leptons) are the topologies that contribute to the $VH(b\bar{b})$ background. The first ATLAS and CMS combination of the inclusive $t\bar{t}$ cross-section measurement resulted in $\sigma_{t\bar{t}} = 173.3 \pm 2.3(\text{stat}) \pm 7.6(\text{sys}) \pm 6.3(\text{lumi})$ pb, for a top-quark mass of 172.5 GeV [42]. Due to the large abundance of top pairs produced in pp collisions, this measurement is already limited by systematic uncertainties with only a subset of the $\sqrt{s} = 7$ TeV dataset. At 8 TeV, the $t\bar{t}$ cross-section has been calculated to be $252.89^{+6.39}_{-8.64}(\text{scale})^{+7.58}_{-7.33}(m_t) \pm 11.67(\text{PDF} + \alpha_s)$ pb, considering the same top mass value and NNLO QCD corrections. At 14 TeV, this calculation yields $984.50^{+23.21}_{-34.69}(\text{scale})^{+27.14}_{-26.29}(m_t) \pm 41.31(\text{PDF} + \alpha_s)$ pb.

Although they have smaller cross-sections, the three single-top channels also have an impact on the $VH(b\bar{b})$ channel. The main contribution to the highest sensitivity regions comes from the Wt channel, given its final-state topology is very similar to both $t\bar{t}$ and $VH(b\bar{b})$.

Diboson production, WZ , ZZ with $Z \rightarrow b\bar{b}$ are another irreducible background, with the feature of having a $b\bar{b}$ invariant mass spectrum peaking close to 125 GeV. Therefore, these processes are useful as a ‘‘standard candle’’ for the validation of the analysis procedure.

Finally, QCD events contaminate the lower sensitivity regions, particularly in the VH channel, where the contamination arises from jets of particles with lepton-like signatures (for example, semi-leptonic decays of heavy flavor quarks). This background is estimated through data-driven techniques, using samples enriched in QCD events.

2.5. Boosted VH production

The VH production channel is a promising way to observe a Higgs boson decaying to a pair of b -quarks at the LHC. However, the level of backgrounds is still limiting. At a centre-of-mass energy of 14 TeV, the $t\bar{t}$ event rate becomes overwhelming due to the increase in gluon initial states with respect to $\sqrt{s} = 8$ TeV. From 8 to 14 TeV,

while the total signal cross-section is expected to increase by a factor ~ 2 , the top pair production increases by a factor 4.

The level of $t\bar{t}$ contamination is the main motivation to focus on a specific phase-space region where the V and the H are produced with a large boost and back-to-back in the plane transverse to the collision. The $t\bar{t}$ cross-section decreases rapidly with increasing transverse momentum of the vector boson V , and QCD processes become negligible in this regime. Furthermore, the triggering and detection of particles with a large boost is more efficient, since they will be produced in regions with higher detector acceptance. This approach was first suggested in [43]. In such a topology, the b -tagging efficiency and corresponding c - and light-quark rejection become the main experimental limitation to a successful $t\bar{t}$ reduction.

A consequence of the kinematics of this region is that the decay products of a boosted Higgs boson will be very collimated. Their angular separation scales approximately as $2m_H/p_T$. As the two b -jets merge, their identification and measurement becomes increasingly difficult, which would undermine the advantages of the boosted regime. To address this problem, a jet grooming technique was first proposed in [43]. Instead of identifying the decay products of the Higgs boson as two isolated jets, the principle is to deconstruct a large jet fully containing the decay and to identify its main features and components, making it possible to discriminate between jets of different origins. This is the concept of jet grooming (or jet substructure), which will be explained in detail in Chapter 4.

2.6. Monte-Carlo event generators

Monte-Carlo event generators (MC) are essential tools for a hadron collider physics program, given the high complexity of the event topologies. On one hand, experimental data can be used to help constrain phenomenological models, while on the other hand simulations are essential for the understanding of the detector and for physics analyses involving many different background processes.

To achieve high precision at the energies being probed at the LHC, it is important to include higher order corrections to EW and QCD processes in the simulations: Feynman diagrams with loops and real emissions which are not always analytically

available and need to be numerically approximated. This is more relevant for strong interactions, where the number of radiated particles tends to be larger with respect to the electroweak case, due to the relatively large coupling constant and to the gluons being able to self-interact.

The partonic cross-section of an arbitrary event at the LHC can be written as:

$$d\sigma_{ab \rightarrow N} = \int_{x_1} \int_{x_2} f_a(x_1, \mu_F) f_b(x_2, \mu_F) \int_{\Phi_N} d\Phi_N |\mathcal{M}_{ab \rightarrow N}(\Phi_N; \mu_F, \mu_R)|^2 dx_1 dx_2, \quad (2.11)$$

where \mathcal{M} is the matrix element for the $ab \rightarrow N$ process and Φ_N is the phase-space element. The renormalization scale μ_R results from the process of renormalization applied in quantum field theories to avoid the inherent ultraviolet divergences [44]. $f_{a,b}(x, \mu_F)$ are the Parton Distribution Functions (PDF's) for each of the incoming partons. When the colliding particles are hadrons, given their bound-state nature, their composition needs to be parameterized in terms of their constituents and the fraction of momentum x they carry, as a function of the squared momentum transfer Q^2 of the interaction. PDF's are extracted from fits to collider and fixed-target experimental data and are frequently a major source of uncertainty on LHC analyses.

A fundamental QCD feature that is used in event generation is factorization, i.e., to consider different stages of the event separately, neglecting interference effects, under the approximation that they occur at sufficiently different time scales. This is already clear from the structure of Equation 2.11, and it introduces a factorization scale, μ_F . According to this approach, the matrix element of a typical hard-scattering event or decay can be calculated at a fixed order in QCD (reflecting the fact that truncating higher-order elements is feasible when the scale α_s is small) while softer processes such as parton showers (described by logarithmically enhanced terms in the perturbative series) are approximated through MC techniques. Outside the scope of perturbation theory, phenomenological models can be used for hadronization and interactions involving the remaining constituents of the protons (known as “underlying event”).

The different event generators relevant to this work are now briefly described:

Parton Shower (PS) Parton shower algorithms generate the dominant QCD radiation terms associated to a given tree-level process, to all orders in perturbation

theory. A Monte-Carlo method associates a probability for a real emission to occur, using an approximation for the real cross-section which is valid only in the soft/collinear limit. In other words, the leading logarithmically enhanced terms are kept, and hence the PS are said to have leading-log (LL) or next-to-leading log (NLL) accuracy. The PS can be combined with a model for hadronization, by truncating its evolution at a scale Q_{had} (typically ~ 1 GeV), where the hadronization takes place. Examples of PS programs are PYTHIA [45, 46], HERWIG [47] and HERWIG++ [48, 49].

Matrix Element Correction (MEC) Given that the first (usually the hardest) emission in a PS is only accurate in the soft/collinear regime, it can be replaced by an exact calculation of the leading-order diagram of the real emission. This improves the accuracy at large angles but does not reach full next-to-leading order accuracy. HERWIG, HERWIG++ and PYTHIA have included MEC in their implementations.

Next-to-Leading Order (NLO) + PS An exact fixed order calculation of the matrix element at next-to-leading order can be made, including real and virtual corrections. The first real emission now has full NLO accuracy, and all softer/smaller-angle radiation is handled by matching a parton shower. This includes initial state and final state radiation. Two methods are currently available to perform this matching: POWHEG [50, 51] and MC@NLO [52].

Matrix Element (ME) + PS Events up to a high multiplicity of real emissions can be generated with tree-level accuracy by combining matrix elements with different number of emitted partons. A separation scale Q_{cut} is defined, below which the radiation is handled by the PS. This is also known as multi-jet merging, or multi-leg. The hardest emission is as accurate as in the MEC or NLO+PS methods, with the advantage that further radiation is also LO accurate, while it's only covered by the PS in the other methods. Virtual corrections are however not included. SHERPA [53] and ALPGEN [54] are examples of ME+PS generators.

MENLOPS A combination of the previous methods can be implemented by generating events with one extra emission with NLO+PS accuracy and using multi-jet merging for events with more than one real emissions [55, 56].

Multi-jet merging at NLO The state of the art calculations for a specific process can be taken advantage of by consistently combining NLO calculations with varying real emission multiplicity with the PS. One example of such a method is MEPS@NLO, implemented within the most recent versions of SHERPA [57].

After the generation of the hard process, showering, hadronization and particle decays, the events can be analyzed with physical meaning. However, for comparisons to data, the interaction of the particles throughout the detector needs to be simulated. A description of detector-level objects in the ATLAS detector is given in Chapter 3.

Chapter 3.

The ATLAS detector

“You’ve got to learn your instrument. Then, you practice, practice, practice. And then, when you finally get up there on the bandstand, forget all that and just wail.”

— Charlie Parker

The European Organization for Nuclear Research (CERN) is one of the largest centres for scientific research in the world, joining the efforts of thousands of scientists and engineers from all continents. CERN operates a network of particle accelerators, among them the LHC, which is the highest energy collider ever built. The LHC [58] consists of a 27 km accelerator ring, approximately 100 m below the French-Swiss border, near Geneva.

Two beams of protons or heavy ions are accelerated at the LHC, circulating in adjacent rings and in opposite directions. These beams are made to collide in four different locations around the ring, where particle detectors are positioned: ATLAS and CMS, both general purpose detectors; LHC***b*** and ALICE, the first focusing on *b*-physics and the latter on QCD measurements in dense heavy-ion environments. Brief descriptions of the LHC, the ATLAS detector and its performance are given in this Chapter.

3.1. Large Hadron Collider

The design energy that a proton can reach at the LHC is 7 TeV, corresponding to a total *pp* centre-of-mass energy of $\sqrt{s} = 14$ TeV. This is achieved by accelerating the protons through a chain of linear and circular accelerators, the LHC being

the last stage, where the protons enter with an energy of 450 GeV. Protons are then accelerated and kept at the designated energy by passing through alternating electrical fields in radio frequency cavities around the ring, which will group the protons in bunches. The beams are bent and kept in their circular trajectories through the use of extremely powerful superconducting magnets, operated at a temperature of 1.9 K.

In the year 2010, the LHC accelerated and collided protons at an energy of 3.5 TeV per beam, already breaking the record for the highest-energy beams ever achieved in an accelerator. The Run 1 of the LHC pp program produced collisions at centre-of-mass energies of 7 TeV during 2010 and 2011, and 8 TeV during 2012.

The characteristics of the beam, such as its transverse area, the number of bunches and the number of protons per bunch, together with their revolution frequency around the LHC ring, will define the collider's instantaneous luminosity. Determining the delivered luminosity is crucial to establishing the relation between the rate of a measured process and its cross-section. The LHC Run 1 has reached a maximum instantaneous luminosity of $7.73 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ in August 2012, with proton bunches being collided in intervals of 50 ns. An integrated luminosity of 5 fb^{-1} was recorded by ATLAS at $\sqrt{s} = 7 \text{ TeV}$, during 2010 and 2011, while a total of approximately 21 fb^{-1} was recorded at 8 TeV in 2012. The relative uncertainty on the luminosity measurement is 1.8% for the 2011 dataset and 3.6% for 2012 [59].

For each bunch crossing, multiple pp interactions are likely to occur, accompanying the rarer and more interesting hard interactions where there is a large exchange of energy. These additional interactions are referred to as pile-up, specifically “in-time pile-up”. Interactions between protons in preceding and subsequent bunch crossings can contaminate the measurement, and this effect is referred to as “out-of-time pile-up”. During the $\sqrt{s} = 7 \text{ TeV}$ run, the average number of interactions per crossing was approximately 9, increasing to 20 during the $\sqrt{s} = 8 \text{ TeV}$ run. At the start of operations in 2010 the pile-up was lower, with the number of interactions per crossing averaging at 3. As protons are composite objects, while the hard scatter can be described simply as an interaction between a pair of quarks and/or gluons, the underlying event (see Section 2.6) can produce additional particles which will further disguise the signature of the main interaction.

The LHC is currently on a long shutdown, in preparation for Run 2, which will start in 2015 and produce collisions at the nominal energies, achieving and potentially exceeding the design instantaneous luminosities.

3.2. ATLAS

The ATLAS detector [60,61], “A Toroidal LHC Apparatus”, is one of the experiments designed to measure proton-proton and heavy-ion collisions at the LHC, probing a vast range of physical phenomena¹.

A typical hard interaction will produce a set of particles that can have a wide variety of properties. Information about the fundamental physics mechanism describing the interaction can be extracted by identifying these particles, or their decay products, and measuring their energy and momentum.

The ATLAS detector is therefore designed so as to be able to observe many different fundamental and composite particles, covering a broad energy and angular range. As is shown in Figure 3.1, ATLAS is a large and complex system, composed of many sub-detectors covering a large solid angle around the interaction point. It is approximately 44 m long and 25 m in diameter, and its components are generally divided into a cylindrical barrel around the beam axis and perpendicular end-caps at both ends. The collision vertices and charged particle tracks are sampled by the Inner-Detector, immersed in a magnetic field so as to allow the measurement of the momentum of the particles. Surrounding it are the calorimeter systems, where most particles will deposit all their energy. Finally, the muon spectrometer and an associated magnet system sit outside the calorimeters. The design resolutions of each sub-detector are shown in Table 3.1.

The ATLAS coordinate system is a right-handed Cartesian system defined with the z axis pointing along the beamline, the y pointing upwards and x towards the centre of the LHC. The azimuthal angle φ is defined on the plane transversal to the beamline, while the polar angle θ covers the angular distance between the y axis and the beamline. A convenient transformation often used in hadron colliders is the

¹Proton collisions will be the subject of this work, but the general descriptions also apply to heavier ions.

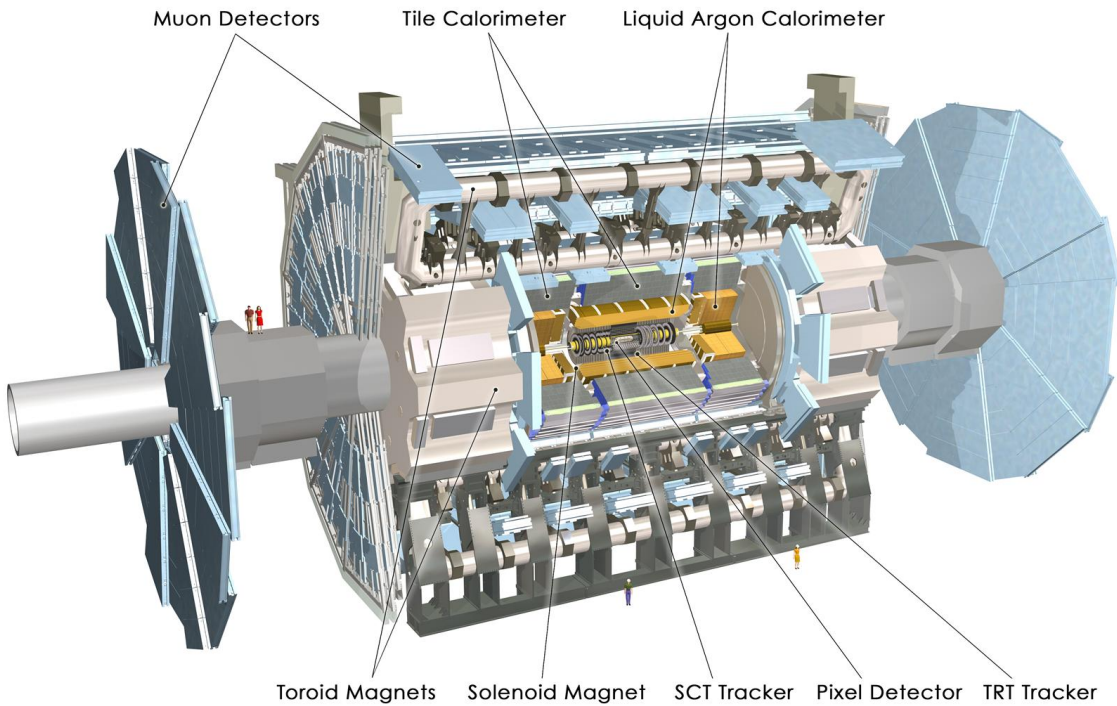


Figure 3.1.: Schematic representation of the ATLAS detector. [60]

following:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right], \quad (3.1)$$

where the pseudo-rapidity η replaces θ to describe the angle of a particle with respect to the beamline, ranging from zero when the particle travels perpendicularly to it, and infinity when it is parallel. In the massless particle limit, the variable η is equivalent to the relativistic rapidity, which is defined as $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$, in natural units, where E is the energy of the particle and p_z its momentum along the beamline. Throughout this work, a “central” particle or object will be one with an absolute pseudo-rapidity of less than 2.5, and “forward” one with pseudo-rapidity greater than 2.5, limits which are related to the geometry of the sub-detectors, as will be described in the following Sections.

Table 3.1.: Design resolutions of the ATLAS sub-detectors.

Sub-detector	Resolution
Tracking (Inner-Detector)	$\sigma_{p_T}/p_T = 0.05\%p_T \oplus 1\%$
EM calorimeter	
barrel and end-caps	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 3.5\%$
Hadronic calorimeter	
barrel and end-caps	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$
Muon Spectrometer	$\sigma_{p_T}/p_T = 10\%$ at 1 TeV

3.2.1. Inner-Detector

The Inner-Detector (ID) is the first sub-detector encountered by particles produced in beam collisions. It consists of three components, as depicted in Figure 3.2: the Pixel detector, the Semiconductor tracker (SCT) and the Transition Radiation tracker (TRT). The purpose of the ID is to detect electrically charged particles as they travel through the detector, providing position information with enough granularity for a track to be identified. A superconducting solenoid magnet outside the ID provides a 2 T magnetic field aligned with the beam axis, curving the trajectories and allowing for momentum and charge sign measurement.

The Pixel detector is composed of three layers of small semi-conductive pixels, parallel to the beamline in the barrel and arranged perpendicularly in the end-caps. It extends from a radius of 4.5 cm to 24 cm outwards from the beamline, and covers the pseudo-rapidity region of $|\eta| < 2.5$. It is the most finely segmented of the ID components, with a resolution of $10 \times 115 \mu\text{m}$ in $(R - \phi) \times z$ in the barrel and $(R - \phi) \times R$ in the end-caps.

The SCT is composed of pairs of silicon micro-strips, its layers also concentric in the barrel and transverse to the beam axis in the end-caps. The micro-strips have a resolution of $17 \times 530 \mu\text{m}$ in $(R - \phi) \times z$ and are arranged in four layers covering a radius of 0.5 m. The SCT η coverage is equivalent to that of the Pixel detector.

Finally, the TRT consists of several gas drift tubes with 4 mm of diameter and containing a gold plated tungsten wire in the centre. The tubes are oriented along the beam axis in the barrel and disposed in a fan-layout in the end-caps, covering the region $|\eta| < 2.0$. They have an intrinsic drift-time resolution of 120-130 μm and together provide typically 30 position measurements per track. Besides recording the position of particles in the azimuthal direction, the TRT is capable of detecting the transition radiation produced by relativistic charged particles when crossing the different mediums inside and outside the tubes, enabling discrimination between electrons and heavier pions.

Particle hits in the Pixel detector, SCT and TRT provide position measurements, and the combination of hits in the different radial layers of the sub-detectors can be used to reconstruct the particle's trajectory. Information on the charged particle tracks in a collision can be used to derive the position of interaction vertices. The first layer of the Pixel detector, also called *b*-layer, provides precise and crucial information for vertexing. The primary vertex, where the hard scattering is most likely to have originated from, is the reconstructed vertex which has the largest sum of squared track p_T , and a minimum requirement on the number of good-quality tracks associated to it. Its position can be measured with a resolution of 30 μm in the transverse plane, and 50 μm in the longitudinal plane [62].

Transverse (d_0) and longitudinal (z_0) impact parameters of a given track with respect to the collision vertex are useful variables for the identification of long-lived particles that decay inside the ID volume, as will be discussed further in Subsection 3.4.4, and can be measured with resolutions of approximately 10 and 80 μm , respectively.

3.2.2. Calorimetry

The calorimetry system in ATLAS is designed to contain the development of electromagnetic and hadronic showers originating in the interaction of particles with the medium. The energy measurement is performed by sampling the energy loss in the dense absorber materials with the instrumented active material. Hadronic showers will typically penetrate further in the material and produce broader shower shapes than those dominated by the electromagnetic (EM) interaction. Therefore,

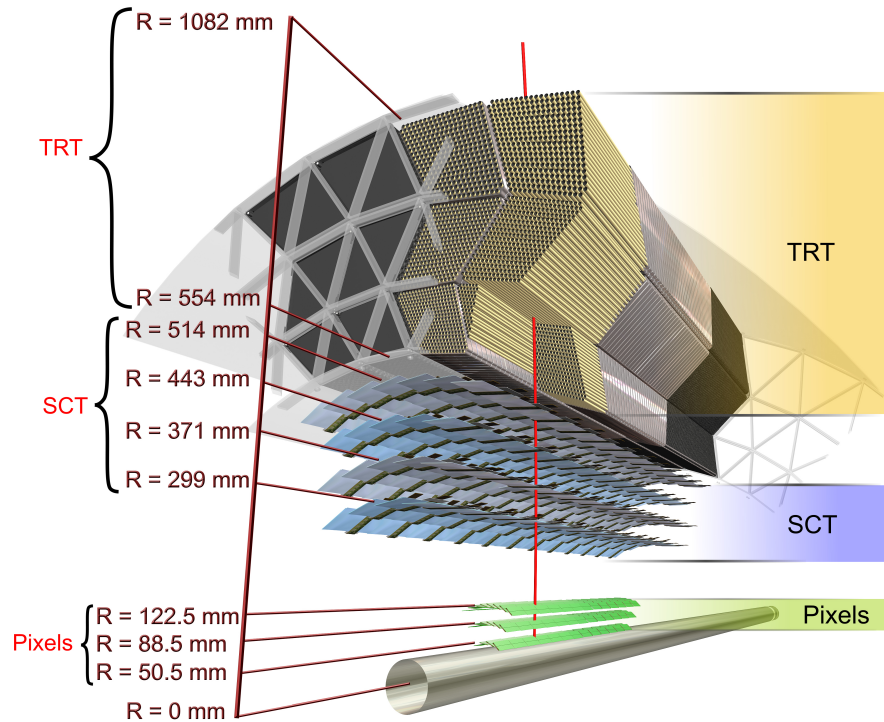


Figure 3.2.: Schematic representation of the Inner Detector components in the barrel. [60]

the systems optimized to contain the EM interaction are located closer to the beam axis, after the ID and solenoid systems.

The calorimeters have complete azimuthal coverage and a pseudo-rapidity acceptance of $|\eta| < 4.9$, extending to an outer radius of 4.25 m. Both hadronic and EM systems are divided in barrel (EM and TileCal), end-caps (EM and HEC, covering $1.3 < |\eta| < 3.2$) and forward regions (FCal, $3.1 < |\eta| < 4.9$), as shown in Figure 3.3.

The EM sampling calorimeter uses lead plates alternated with liquid argon (LAr), in an accordion shaped geometry. The barrel section complements the ID measurements with excellent angular resolution, of typically 0.025×0.025 in $\eta \times \phi$. A presampler is located just before the EM calorimeter, covering $|\eta| < 1.8$, to measure the energy of particles which start showering before reaching the calorimeter.

The hadronic barrel calorimeter (TileCal) is composed of alternating steel and plastic scintillating tiles, with a maximum angular granularity of 0.1×0.1 in $\eta \times \phi$. The end-caps are composed of copper as absorber and LAr as active material.

The forward calorimeters, subject to an extreme radiation environment given their proximity to the beam, consist of copper/tungsten and LAr modules.

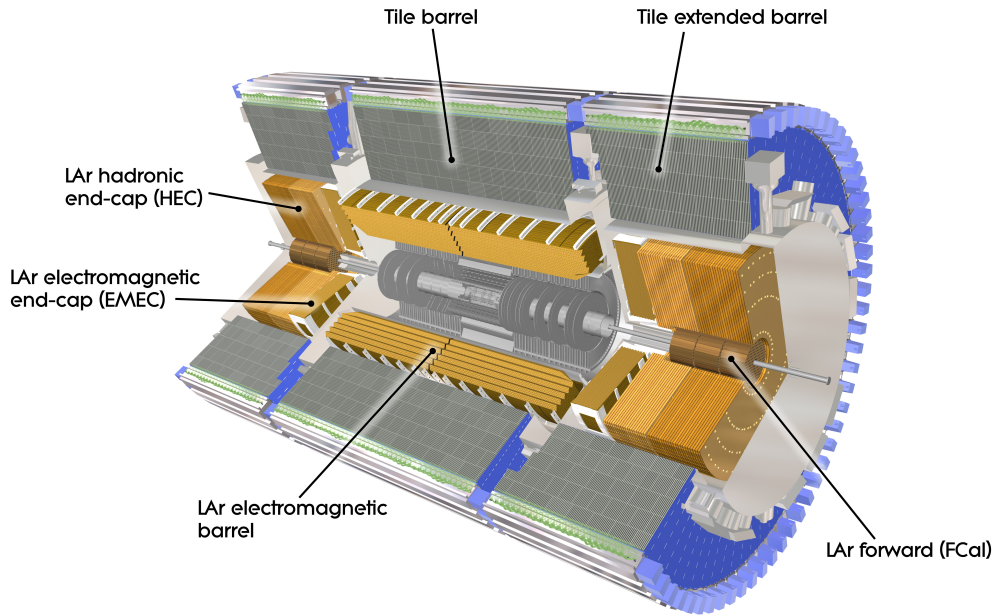


Figure 3.3.: Schematic representation of the calorimeter components in the barrel, end-caps and forward region. [60]

3.2.3. Muon Spectrometer

Outside the system of calorimeters, reaching an outer radius of 12 m, are the ATLAS muon systems. Similarly to the ID, the muon spectrometer (MS) registers the position of charged particles passing through it. The associated magnetic system drove the design of the ATLAS detector, and consists of a barrel and two end-cap superconducting air-core toroid magnets, as can be seen in Figure 3.4. The high-precision tracking and triggering of muons is performed with different technologies of multi-wire chambers: Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC) for tracking, Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) for triggering. The chambers are arranged in three cylindrical layers aligned with

the beam axis in the barrel and in three perpendicular disks in the end-cap regions. Position measurements in η and ϕ are made up to $|\eta| = 2.7$.

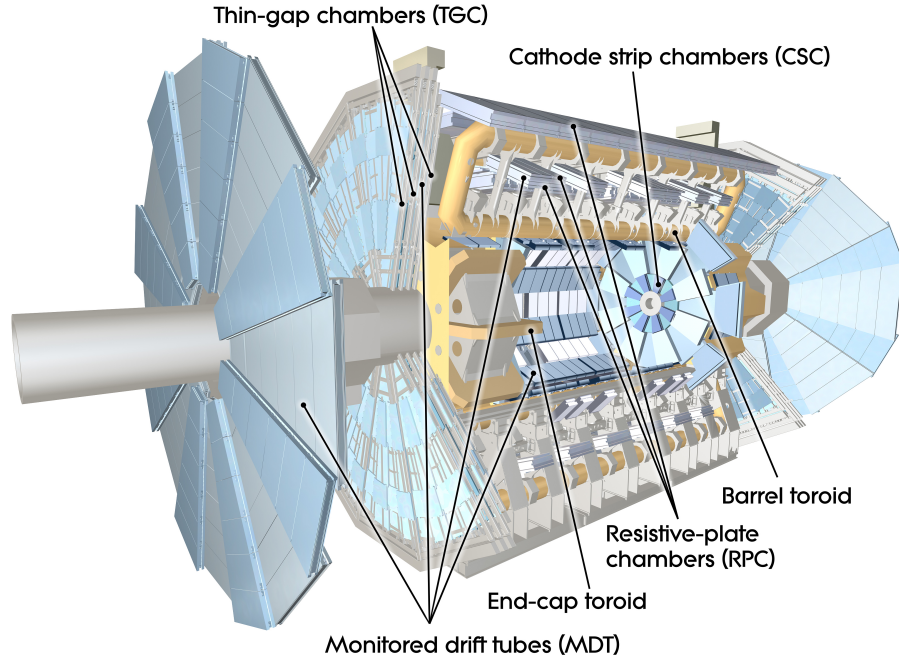


Figure 3.4.: Schematic representation of the muon spectrometer components in the barrel. [60]

3.3. Trigger

During Run 1, the LHC has collided protons at a rate of approximately 20 MHz and will reach the design crossing rate of 40 MHz on Run 2. However, such rates are too large for permanent storage and analysis. Furthermore, the interesting physics processes do not occur at this frequency, given their relatively small cross-sections. Therefore, a trigger and data acquisition system are in place to rapidly and reliably select events of interest, reducing the event rate by a factor 10^5 .

The trigger system is divided into three stages. The first level (L1) performs a hardware-based scan of the event, so as to identify potential regions of interest (RoI) in the calorimeters and muon systems. The RoI correspond to cones in $\eta - \phi$, z where basic features compatible with signatures of high p_T leptons, photons, jets, total or missing transverse energy in the event are identified. The second level of

the trigger system (L2), applicable to those events passing the L1 requirements, has access to all the detector information related to the RoI and implements basic reconstruction algorithms. The final step is then the event filter (EF), which uses complex reconstruction algorithms that profit from offline calibration to make a final decision on whether to permanently store the event or not.

3.4. Particle Identification

The ATLAS sub-detectors can measure energy and momentum. Stable particles that travel a measurable distance in the ATLAS systems can also be identified. The following summarizes the techniques employed to identify and reconstruct electrons, photons, muons and jets.

3.4.1. Electrons

The identification and reconstruction of electrons can be performed with a combination of sub-detectors in the pseudo-rapidity range $|\eta| < 2.5$, and with calorimeter information in the forward pseudo-rapidity range.

As a charged particle traverses the different media in the TRT, alternating between the gaseous tubes and the radiator material between them, transition radiation will be emitted. The electron, with a larger characteristic Lorentz factor than charged hadrons, will emit more photons and consequently the number of TRT hits above a certain threshold can be used as a discrimination variable. Electrons originating from photon conversions can be rejected by requiring that their tracks have a hit on the b -layer.

The EM calorimeter will collect most of electron's energy. Due to its longitudinal and lateral granularity, it is also able to determine the impact point. The energy deposits in the barrel are selected by identifying clusters of energy with associated tracks in the ID. In contrast, the identification of a photon cluster requires that there is no track matching. An EM cluster is built by grouping cells within a fixed-size window, positioned in the local maxima of transverse energy. The four-momentum

of the reconstructed electron is computed with the energy information from the EM cluster and the η, ϕ coordinates from the matched track.

Combining shower shape variables, track and track-cluster matching quality, TRT information, hits on the b -layer, and other variables, electrons can be identified with increasing degrees of purity, and corresponding increasing rejection factors against photon pair production and hadronic jets. The absolute energy scale of electrons can be measured in ATLAS with an uncertainty at the sub-percent level [63].

3.4.2. Muons

Muons, as minimum ionizing particles, typically traverse the ID and the calorimeters without great loss of energy, and their transverse momentum and charge sign can be measured by the Muon Spectrometer and the Inner Detector, covering a region with $|\eta| < 2.7$. To a lesser extent, to recover acceptance losses, the calorimeter can also be used to identify and reconstruct muons. The highest muon purity is achieved by combining tracking information measured independently in the ID and in the MS, defining “combined” muons. When such information is not available, other types of muons can be defined, albeit with lower purity: “stand-alone”, with tracking information from the MS; “segment-tagged”, in cases where the ID information can be combined with hits in the first chambers of the MS; and “calorimeter-tagged”, combining ID and a calorimeter energy deposit.

The muon momentum scale for combined muons in the range $5 \leq p_T \leq 100$ GeV is extremely well measured in ATLAS, with an uncertainty not larger than 0.2%, and a relative resolution of 2-4% [64]. High- p_T muons are subject to relative momentum resolutions of up to 10%.

3.4.3. Jets

Hadronic showers created in the electromagnetic and hadronic calorimeters can be clustered into structures referred to as jets. In addition to calorimeter clusters, other objects with a four-momentum representation, such as tracks, can be used as detector-level inputs to jet clustering algorithms, which will be further explained in Chapter 4.

Calorimeter jets are built from energy depositions in the hadronic and electromagnetic calorimeters. An incoming particle will deposit energy in the calorimeter cells, longitudinally and laterally with respect to its direction of motion. According to the total energy deposited in each cell and the overall expected noise, a seed cell is found and an iterative procedure adds the neighboring cells if their energy is above a certain threshold. The resulting three-dimensional set of cells is referred to as a topological cluster (or topo-cluster) [65], and is classified as hadronic or electromagnetic depending on their shape, longitudinal and lateral depth and energy density. The total energy is measured by assuming the electromagnetic interaction as its origin, corresponding to the so-called EM scale, and can be calibrated to the hadronic scale through a process known as local cluster weighting [66], based on single pion interactions. This approach corrects for the calorimeter's different response to EM and hadronic showers. Further calibration procedures are described in Chapter 5.

Non-collision backgrounds affect the quality and purity of calorimeter signals. Their contamination is removed through jet quality criteria, so that the jets from the hard scatter can be distinguished. A variety of phenomena can give origin to these false signals, e.g., calorimeter noise, proton collisions with residual gas in the beam pipe or cosmic ray muons, and an event containing any of these effects is discarded.

Tracks can also be used as input to jet algorithms, defining the so-called track jets. Similarly to the jet quality criteria applied on the calorimeter signals, tracks are selected based on the number of ID hits, transverse momentum and impact parameters, to minimize the inclusion of tracks which do not originate from the primary vertex.

A useful quantity known as jet vertex fraction (JVF) can be used to reject jets originating from pile-up events [67]. JVF quantifies the fraction of the scalar summed p_T of the tracks in the jet that originate from the primary vertex. With the 2012 dataset, the JVF cut was optimized to efficiently select typically more than 90% of the jets originating from the primary vertex, figure which varies with the number of reconstructed vertices, and resulting in a pile-up rejection close to 100% [68].

3.4.4. *b*-tagging

Hadrons that contain *b* or *c*-quarks can be distinguished from hadrons composed of lighter quarks by their relatively long lifetimes and by their leptonic decay signatures. This is especially true for *B* hadrons, which have lifetimes of the order of 1.5 ps. When produced with enough transverse momentum (at least ~ 20 GeV), their average flight length will be of a few mm, usually decaying before reaching the Inner Detector. Such features can be explored to develop techniques to identify jets that contain *B* hadrons.

The key inputs to the *b*-tagging algorithms developed in ATLAS are the charged particle tracks reconstructed in the Inner Detector, which are spatially matched to calorimeter jets with a p_T -dependent condition. Variables related to the impact parameter of the tracks, to the reconstructed secondary vertex where the decay occurs, and to the reconstruction of the topological decay chain [69] are used to discriminate between heavy flavor and light jets (see Figure 3.5). One of the most sophisticated algorithms, the MV1 tagger, implements a neural network that combines track, secondary vertex and decay chain information, taking correlations between the variables into account [70]. An alternative to MV1 is the MV1c tagger, which is trained specifically against a charm jet background and therefore achieves a higher discrimination between *b* and *c* originated jets. Both MV1 and MV1c are used in this work.

The performance of a *b*-tagging algorithm can be characterized by its efficiency to tag jets that originate from *b*-quarks and the corresponding rejection rates for *c* and light jets. The MV1 tagger is implemented by selecting an efficiency working point, typically 70%, which translates into charm and light mis-tagging rates of $\sim 20\%$ and $\sim 1\%$, respectively. The light-jet rejection as a function of the *b*-tagging efficiency is shown in Figure 3.6. The *b*-tagging efficiencies measured in Monte-Carlo simulation as a function of the jet transverse momentum are calibrated to the values observed in data. The efficiencies for tagging *b*-jets, *c*-jets and light jets are measured in $t\bar{t}$ events, D^{*+} events and inclusive jet samples, respectively [70, 71]. The systematic uncertainties impacting the measurements on *b*-jets cover jet p_T values between 20 and 300 GeV and are of the order of 2% in the intermediate p_T range, being considerably larger at higher and lower values of p_T .

The identification of isolated b -jets in a moderate p_T range is well understood. However, the performance of these techniques worsens in dense environments such as boosted $H \rightarrow b\bar{b}$ decays, where the jets can overlap and their tracks can become very collimated. Due to the additional activity surrounding a given jet, track-jet matching, for example, becomes more ambiguous. Dedicated b -taggers trained specifically to handle high occupancy environments are currently being validated in ATLAS [72, 73], and will certainly be of use in a future boosted $VH(b\bar{b})$ analysis.

b -tagging algorithms identify jets with B hadron content, but provide no information on the number of such hadrons in the jet. Novel techniques to double b -tag jets are under development, and could also have a significant impact in the work described in this thesis [74].

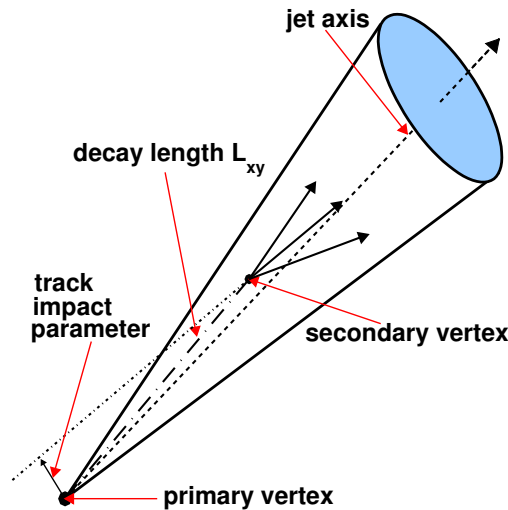


Figure 3.5.: Schematic view of a B hadron decay inside a jet. [73]

3.4.5. Missing Transverse Energy

Stable and non-interacting particles, such as neutrinos, will escape the ATLAS volume undetected. Before the collision, the momentum of the incoming partons is essentially limited to the z direction. Due to conservation of momentum, a vectorial sum of all the calorimeter energy depositions can therefore be used to infer the transverse energy of escaping particles. E_T^{miss} is calculated as the negative of the vector sum of all reconstructed objects in the event (after calibration) and any remaining unmatched

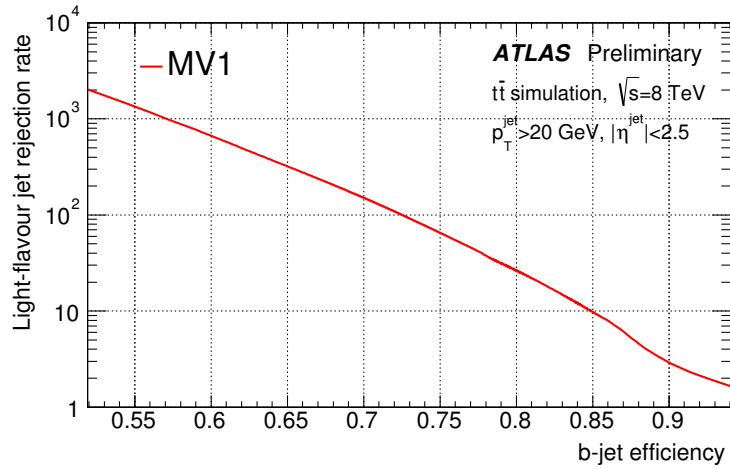


Figure 3.6.: Light-flavour rejection (defined as the inverse of the mis-tagging rate) as a function of b -tagging efficiency for the MV1 algorithm, as measured in simulated $t\bar{t}$ jets. [70]

calorimeter deposits and tracks. Other sources of missing transverse energy are detector inefficiencies and resolution, which lead to mis-measurement of the true transverse energy of the event objects.

3.5. Detector Simulation

Predictions from Monte-Carlo event generators, as described in Chapter 2, can be compared with ATLAS data by simulating the interactions of particles with a detailed model of the detector apparatus, including reconstruction and identification performances. A full simulation of the ATLAS detector response [60] can be performed with Geant4 [75, 76], implementing an accurate model of the detector geometry, relative alignment of the sub-systems, material composition and physics models for the interaction of particles with matter.

When performing the full simulation, the computing time per event becomes a real limitation. This is especially true when the cross-sections of interest are only a very small fraction of the total background processes and where very precise background modelling and estimation of systematic uncertainties is necessary. The required statistical precision of the Monte-Carlo samples can be achieved by using a fast simulation of the detector, ATLFAST-II [77]. In this case, the response of

each calorimeter cell is parameterized and tuned against data, whereas the ID and MS responses are still fully simulated. ATLFAST-II reduces the computing time by more than one order of magnitude and, although with a lower degree of accuracy, its results have been validated against the full simulation, where general reasonable agreement has been found. Throughout this work, full and ATLFAST-II simulations are used.

3.6. Preparations for Run 2

The Run 2 of the LHC will begin with pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, to be eventually increased to the nominal energy of $\sqrt{s} = 14$ TeV. With a bunch spacing of 25 ns, the collider will reach the nominal instantaneous luminosity of up to $3 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, with an estimated average number of interactions per bunch crossing of 50.

The ATLAS detector was upgraded during the 2012-2013 shutdown to be able to cope with the luminosity increase. Among the improvements are the installation of a new silicon pixel layer at a radius of approximately 30 mm from the beamline, the so-called Insertable B-layer (IBL) [78], which will require important optimizations of tracking, vertexing and b -tagging algorithms that will greatly impact the sensitivity of search channels with b -jets in the final state, such as $VH(b\bar{b})$.²

Upgrades to the trigger system are crucial to prevent the degradation of the detector's performance with the large increase in the interaction rate. Sophisticated trigger methods were developed to achieve a high event reduction while maintaining the sensitivity to low p_T objects. Examples of trigger improvements are the addition of a track trigger at the L1 level and the implementation of hardware parallelized track finding (FTK) [79].

²Initial studies performed with b -tagging algorithms based on impact parameter and secondary vertex reconstruction indicate an improvement of 100% in the rejection of light jets, using $t\bar{t}$ events without pile-up. [78]

Chapter 4.

Jet algorithms and jet grooming

“In my music, I’m trying to play the truth of what I am. The reason it’s difficult is because I’m changing all the time.”

— Charles Mingus

Interactions mediated by the strong force are characterized by a coupling that decreases with smaller distances (or higher energies), while increasing for larger distances, to a point where perturbation theory no longer holds. In the small-distance regime, as quarks and gluons are separated, QCD radiation is produced in the form of gluons and quark-antiquark pairs. This is the process of fragmentation, and it is the origin of jets as objects produced in particle accelerators. As the energy scale decreases and the coupling becomes stronger, quarks and gluons eventually experience the property of confinement and consequent hadronization into colorless objects.

Due to the inherent stochastic nature of jets, algorithms and conventions are necessary to define them. This chapter describes some of the most commonly used algorithms in ATLAS and other HEP experiments to define hadronic jets.

At the detector level, the basic constituents of jets are typically calorimeter topo-clusters, or alternatively Inner Detector tracks (for the charged component only), both already described in Chapter 3. Throughout this work, the detector-level jets used are all calorimeter-jets, unless otherwise stated. In Monte-Carlo outputs, (truth) jets can be defined instead with stable particles typically with a lifetime of at least 10 ps and excluding muons and neutrinos.

4.1. Jet Algorithms

Given the complexity of the final state where a jet is produced, the criteria to group particles together to form a jet is somewhat arbitrary and depends on the choice of algorithm. The resultant 4-vector is also ambiguous, and a recombination scheme to assign a total momentum to the jet is necessary. Throughout this work, the 4-vector sum of the jet constituents is used (with the constituents assumed massless, in the case of calorimeter-based jets).

Historically, the development of algorithms to define jets has gone through a vast evolution. The use of cone algorithms was introduced in the context of e^+e^- experiments [80], where a jet would be formed with all the constituents within the volume of a cone, with a base radius defined according to the topology of the event (e.g., the energy spatial distribution). The final 4-vector of the jet would then be equivalent to the geometrical centre of the cone. This approach, while intuitive, is not infra-red safe, i.e., is sensitive to collinear and soft emissions, which means that the final jet topology of the event would be dependent on the parton shower model and would be susceptible to event-by-event fluctuations. Another class of jet algorithms was developed to address this problem: the sequential recombination algorithms, pioneered by the JADE Collaboration [81, 82]. One basic difference between these and cone algorithms is that jet finding is done in several clustering steps, which means that these algorithms associate a clustering sequence to the event.

The most widely used jet algorithm in ATLAS is a sequential recombination algorithm, the anti- k_t [83], chosen for its fast and efficient clustering as well as good physics performance. Other common algorithms are the k_t [84] and the Cambridge/Aachen [85].

Considering particles p_i and p_j , the grouping of these into a jet using a sequential recombination algorithm will depend on a distance definition, d_{ij} :

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \quad (4.1)$$

where ΔR_{ij} is the angular distance between particles i and j , defined in terms of rapidity and azimuthal angles¹. In hadron colliders, an additional distance measure

¹ $\Delta R_{ij} = \sqrt{(\Delta y)^2 + (\Delta \phi)^2}$

is defined as $d_{iB} = p_{Ti}^{2p}$, associated to the distance between particle i and the beam. If d_{iB} is the smallest, particle i is recombined as part of the beam-jet. The constant R is a parameter of the jet algorithm and takes different values for different experiments, and also for different analysis (e.g., resonance searches vs dijet spectrum measurement). The exponent p can take the values $-1, 0, 1$ to correspond to the anti- k_t , Cambridge/Aachen and k_t jet definitions.

From Equation 4.1 one can see that, by definition, jets created with the anti- k_t algorithm start from a combination of the higher transverse momentum constituents. As the algorithm progresses, the wider-angle emissions with lower p_T are combined with the core jet, and the result is a circular jet in the rapidity-azimuthal (y, ϕ) plane. Contrasting with this definition are the Cambridge/Aachen and the k_t algorithms, which follow a QCD-like structure when recombining the constituents into a jet. This means that softer and collinear branchings are merged first. In both cases, the jet boundaries are not circular in the (y, ϕ) plane, and this is a property which can make measurement and calibration less straightforward than in the anti- k_t case.

The choice of the optimal jet definition is heavily dependent on the purpose of the analysis. One thing to note, however, is the presence of initial state QCD radiation (ISR), multiple parton interactions (MPI, or underlying event - UE) and pile-up collisions along with any hard scattering event of interest. This will limit, for example, the extent to which a large- R algorithm would be useful, since there will be a trade-off between the inclusion of all the relevant final state radiation (FSR) and the rejection of contamination from the processes mentioned.

The problem of identifying the relevant constituents of the jet, as well as distinguishing jets originating from different particles or processes is addressed by jet substructure techniques that are described in the next section.

4.2. Jet substructure

Jets with a merging history which follows QCD probability principles can be deconstructed with the purpose of identifying those branches which are more likely related to QCD radiation than, for example, a heavy boson decay.

In this section, several ways of analysing the substructure of a jet are introduced and briefly discussed. One of the goals is to explore differences between jets which have originated from a two or three-body decay of a high- p_T massive particle and typical high- p_T jets formed from gluon and light quarks. It has also been shown that these techniques have advantages when dealing with high levels of pile-up and UE, improving the resulting jet mass resolution. All the grooming techniques and jet shapes described in this section have been validated in ATLAS with the 2010 and 2011 datasets [86, 87].

4.2.1. Splitting and filtering

The use of the Cambridge/Aachen algorithm with the splitting and filtering technique is one example of a jet grooming approach. This technique was originally developed for the reconstruction of a low mass Higgs boson decaying to a pair of b -quarks in a boosted configuration [43], and is therefore used throughout this work.

As introduced in Chapter 2, the boosted regime is crucial for a Higgs boson search in the $H \rightarrow b\bar{b}$ channel, and in these conditions the decay products of the Higgs boson won't necessarily be resolved by standard sized jets (in ATLAS, this corresponds to $R = 0.4, 0.6$). The idea of this approach is to use a large- R Cambridge/Aachen jet (in this case, $R = 1.2$) and to perform the splitting and filtering technique to extract the constituents that correspond to the decay products, and at the same time reject those jets which are more likely to be only QCD in nature, and not Higgs boson decay products.

The choice of the R parameter is related to the mass of the Higgs candidate for which this approach was initially developed ($m_H = 120$ GeV) and to the nominal LHC run conditions ($\sqrt{s} = 14$ TeV).

The splitting/filtering procedure partially reverses the Cambridge/Aachen algorithm described above, and applies a set of criteria to choose jets which are more likely to have originated from a heavy particle decay. This approach, also known as mass-drop algorithm, depicted in Figure 4.1, works as follows:

1. Start by considering a Cambridge/Aachen jet j . Undo the last step of clustering, and take the two resulting subjects as j_1 and j_2 , where $m_{j_1} > m_{j_2}$.

2. Test the mass-drop and the asymmetry condition, i.e.:

$$\mu = \frac{m_{j_1}}{m_j} < 0.67, \quad (4.2)$$

$$y = \frac{\min(p_{T,j_1}^2, p_{T,j_2}^2)}{m_j^2} \Delta R_{j_1, j_2}^2 > 0.09. \quad (4.3)$$

3. If both conditions are met, perform the filtering: re-run the Cambridge/Aachen algorithm on the constituents of j_1 and j_2 with $R = \min(0.3, R_{sub}/2)$, where R_{sub} is the distance between the two subjects j_1 and j_2 . Keep the resulting two or three leading subjects as the filtered jet. If any of the conditions fail, take j_1 as j and return to step 1.

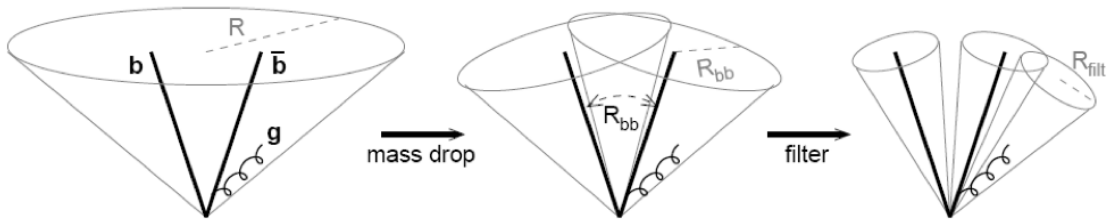


Figure 4.1.: Schematic representation of the splitting/filtering algorithm. [43]

In context of the Higgs boson search, the b -tagging of the two p_T leading subjects is an additional requirement. The purpose of the mass-drop condition is to identify the splitting in which the boson decay has occurred, i.e., where the subjects added together result in a massive jet. When there is a large mass-drop, the subject is less likely to correspond to a radiated gluon. The symmetry in momentum is required to reject configurations in which the selected subjects correspond to QCD radiation. On a candidate jet with more than two subjects, the three leading ones are kept so that any extra radiation from the b 's is included, to take into account the $b\bar{b}g$ final state.

For the studies described in Chapter 6 with the 2011 dataset, the first step of the algorithm was defined with an extra condition: if the unclustering resulted in subjects within $\Delta R_{ij} < 0.3$ of each other, then the original jet would be discarded because of detector resolution limitations.

The advantage of using a Cambridge/Aachen jet for this procedure is that this algorithm allows for the angular scale of the problem to be dynamic and therefore to be optimized on a jet-by-jet basis.

4.2.2. Pruning

The pruning technique [88] was developed for general searches, with the purpose of being applicable to many different mass ranges. It is not aimed at any particular resonance search, but at distinguishing a QCD jet from any other type of jet, while improving its mass resolution. This is done by removing any soft and large-angle recombination from the substructure of the jet in question.

The initial jet finding for the pruning procedure can be performed with any algorithm, and therefore the anti- k_t is usually used for its advantages such as its circular configuration on the (y, ϕ) plane. The constituents of each jet are then reclustered with a sequential recombination algorithm with a meaningful substructure (k_t or Cambridge/Aachen). On each step of the algorithm, $i, j \rightarrow p$, the softer constituent is discarded if the following conditions are met:

$$z = \frac{\min(p_T^i, p_T^j)}{p_T^p} < z_{cut} \quad (4.4)$$

$$\Delta R_{ij} > R_{cut} \quad (4.5)$$

After following the full procedure on all the constituents, the resulting jet is the pruned jet. This approach does not define subjets of the original jets.

The parameters z_{cut} and R_{cut} can be tuned to the signal being studied, but it has been shown that these vary slowly with m/p_T and therefore do not need fine-tuning for the technique to offer significant improvement. Typical values for these parameters are $z_{cut} = 0.05, 0.1$ and $R_{cut} = 0.1, 0.2, 0.3$. Throughout this work, the pruning algorithm is implemented with difference choices for z_{cut} and R_{cut} , with the initial clustering of the constituents performed with the Cambridge/Aachen algorithm.

4.2.3. Trimming

The trimming technique [89] was originally developed for the reconstruction of jets coming from light partons in general, as opposed to jets coming from heavy particle decays. The intention is to achieve a decontamination of the soft components of the jet, which are likely due to pile-up, MPI and ISR, while keeping all the relevant contributions, including FSR.

This algorithm is based on the observation that pile-up, MPI and ISR are softer than FSR, and also that the spatial overlap between them is negligible. Starting with an untrimmed jet of any kind, but with a relatively large radius as to include all the relevant radiation, the trimming algorithm proceeds as follows:

1. Within each initial untrimmed jet, recluster its constituents with the k_t algorithm and a characteristic radius R_{sub} smaller than the original jet radius.
2. From the produced subjets, discard those with a transverse momentum p_T smaller than a fraction f_{cut} of a hard scale characteristic of the event kinematics (for example, the p_T of the original untrimmed jet or the scalar sum of the transverse momenta in the event).
3. Define the final trimmed jet with the remaining subjets.

In the ATLAS trimming implementation, the hard scale used to define a subjet p_T cut-off is the transverse momentum of the original untrimmed jet, and several parameter configurations are considered and validated, for example, $f_{cut} = 0.01, 0.03, 0.05$ and $R_{sub} = 0.2, 0.3$.

4.2.4. N -subjettiness

N -subjettiness, denoted τ_N , is a jet shape that can be effective in discriminating boosted hadronically-decaying objects from QCD jets [90]. It is based on the number of subjets in a given jet to be able to identify two-prong or three-prong decays that might originate from boosted W , Z , H -bosons or top-quarks. A boosted QCD jet will acquire mass mainly through a single hard parton undergoing large angle soft splittings. Therefore, N -subjettiness exploits the specific pattern of energy flow and energy deposition of a decaying boosted object to differentiate it from a QCD jet.

This jet shape variable is defined as follows. Identifying the initial jet with any algorithm, one identifies N candidate subjets using the exclusive k_t algorithm (where every particle is assigned to one of N final subjets). Then, N -subjettiness is calculated as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\}, \quad (4.6)$$

where the summation in k runs over the constituent particles of the jet, $p_{T,k}$ are their transverse momenta and $\Delta R_{J,k}$ are the angular separation between each constituent and a candidate subjet J . d_0 is a normalization factor defined as $d_0 = \sum_k p_{T,k} R_0$, where R_0 is the radius used in the original jet finding algorithm.

This definition quantifies how well a jet can be described by N -subjets, tending to larger values if a $N + 1$ configuration would be more appropriate. However, the best discriminating variable is the ratio τ_N/τ_{N-1} . For example, W jets will have typically smaller τ_2/τ_1 values than QCD jets. While τ_2 will be small for both W and QCD jets, since most of the radiation will be aligned with two subjets, QCD jets with small τ_2 typically have small τ_1 as well, and τ_1 should be relatively large for a W jet.

N -subjettiness is typically calculated for anti- k_t trimmed jets. It is worth noting that this jet shape is not meaningful for Cambridge/Aachen split/filtered jets, since the filtering already constrains the number of subjets.

4.3. Performance of grooming techniques in ATLAS

The mass, transverse momentum and other kinematic variables related to Cambridge/Aachen and anti- k_t jets which have been subject to the grooming techniques described in the previous section have been validated on inclusive jet samples with 2010 ATLAS data (taking advantage of the low pile-up conditions), and also samples enriched in boosted W , Z -bosons and top-quark pairs with 2011 ATLAS data [86, 87]. The Monte-Carlo description of these quantities has shown reasonably good agreement with data, where both LO and NLO generators perform well.

One challenge that analyses in ATLAS have to overcome is the high event pile-up. How a reconstructed jet quantity is affected by the number of primary vertices in the event, N_{PV} , can guide the choice for the appropriate jet algorithm and grooming technique. Figure 4.2 shows the dependence of the mean jet mass $\langle m^{jet} \rangle$ with N_{PV} , for Cambridge/Aachen split and filtered jets. It has been shown in [87] that while trimming and filtering reduce the slope seen for ungroomed jets, pruning does not have the same effect. The susceptibility of the N -subjettiness ratios $\langle \tau_{21} \rangle$ and $\langle \tau_{32} \rangle$ to pile-up was also tested, and found to be fairly independent for some of the parameter sets.

The impact of these techniques on the reconstructed mass spectrum can help the identification of boosted hadronic decays: on one hand, the signal peak is enhanced due to the rejection of pile-up, UE and ISR radiation (see Figure 4.3). On the other hand, the jet mass originating from QCD splittings is usually shifted to lower masses after grooming, which improves the background rejection.

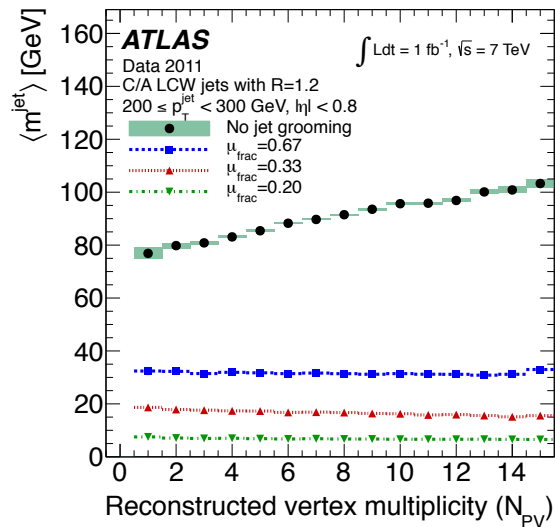


Figure 4.2.: Mean of the uncalibrated jet mass distribution before and after grooming, for jets in the central region as a function of the reconstructed vertex multiplicity, using Cambridge/Aachen split and filtered jets in the range $200 \leq p_T < 300$ GeV. [87]

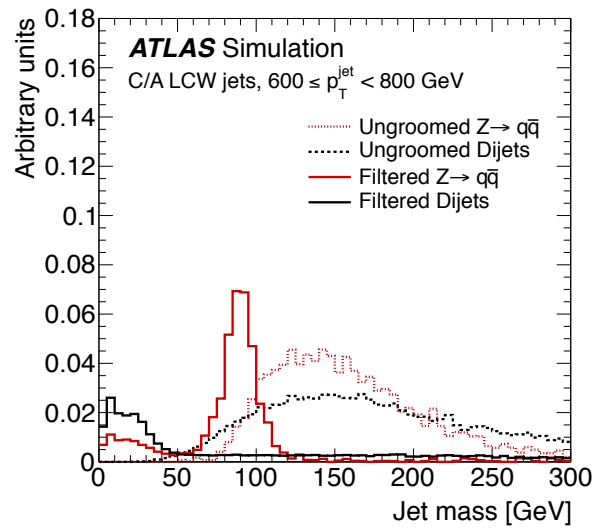


Figure 4.3.: Invariant mass of the highest- p_T Cambridge/Aachen split and filtered jet in Herwig $Z \rightarrow qq$ events and Powheg+Pythia dijet events, in the range $600 \leq p_T < 800$ GeV. The groomed distributions are normalized to the ungroomed distributions, which are normalized to unit area. [87]

Chapter 5.

Jet mass scale validation with W -boson hadronic decays

“What about a blues in W , in the key of W .”

— Rahsaan Roland Kirk

The accuracy and precision of energy and mass measurements in the ATLAS calorimeters have to be established to perform an analysis based on calorimeter jets. This Chapter briefly describes the calibration of the jet energy scale for small- R jets ($R = 0.4, 0.6$) and the jet mass scale for large- R groomed jets ($R \geq 1.0$).

5.1. Jet energy scale for small- R jets

The energy of a jet as measured in the calorimeters does not necessarily correspond to the true energy of the set of stable particles that originated it: this is due to several detector effects, such as dead material, leakage, non-compensation and particle reconstruction efficiencies. Therefore, an energy calibration must be performed to improve the accuracy of all jet related measurements and their uncertainties [91].

Calorimeter jets are reconstructed at the electromagnetic energy scale (EM), which correctly reproduces the energy originated in the electromagnetic showers. The associated electronic signal is calibrated to the EM-scale with test-beam measurements with electrons, and the measured energy in the EM calorimeters is finally corrected using the Z -boson invariant mass in $Z \rightarrow e^+e^-$ events. As described in Chapter 3, by

using information such as the shower shape and the energy density in the calorimeter, a local hadronic calibration to the calorimeter clusters can be performed, where the corrected energy accounts for the hadronic component of the shower. Jets whose energy has been corrected to this scale are referred to as LC jets.

For the purpose of this work, EM and LC jets are used. On top of these calibrations, jet energy scale (JES) corrections are derived [92]. For the case of anti- k_t jets with radius parameters $R = 0.4$ and $R = 0.6$, JES calibration is performed in several steps. Firstly, an offset correction is applied to the energy of the jet, derived from Monte-Carlo, to subtract the pile-up contribution. This is determined as a function of the number of primary vertices (N_{PV}) and average number of interactions (μ), and is based on the area of the jet [93]. The energy and rapidity are then corrected to the Monte-Carlo truth-level components of the jet, and the jet direction is adjusted so as to point towards the primary vertex, instead of the detector's nominal origin. Finally, *in-situ* corrections and uncertainties are derived from data, to address the remaining differences between data and simulation, using events where one jet is balanced to a well known physics object, such as a Z or a photon [94].

The resulting detector-level jets are referred to as EM+JES or LC+JES. The calibrations are validated and systematic uncertainties associated to them through several *in-situ* methods, using e.g. dijet, $Z + \text{jet}$ or $\gamma + \text{jet}$ samples. Each method is applicable to a different kinematic range, and the combination allows for a very complete understanding of jets in ATLAS.

Additionally, a Global Sequential Calibration (GSC) can also be applied on top of the EM+JES calibrated jets. The goal of GSC is to improve the jet energy resolution by sequentially correcting the average energy response for effects caused by jet properties such as track multiplicity or jet width¹. This results in a lower dependence of the jet response on the quark or gluon nature of the jet.

5.2. Jet energy and mass scale for large- R jets

Large- R jets subject to grooming procedures are built from locally calibrated clusters, and are also subject to a JES calibration [87]. The correction is derived from a

¹Defined as the p_T weighted average ΔR of the tracks in the jet.

PYTHIA dijet sample and is applied similarly to what was described for small- R jets, with the exception that the offset energy correction related to pile-up is not applied.

The jet mass scale (JMS) needs an extra calibration procedure, given that the jet mass is essential for any boosted hadronically-decaying heavy particle search.

The Monte-Carlo correction to the jet mass response is able to restore uniformity across the energy and y range within a level of 3%, as shown in [87]. Similarly to what is done with the JES, this procedure is validated with different methods, and systematic uncertainties are estimated. One method is related to track-jets and takes advantage of the fact that detector effects on these jets are uncorrelated with the corresponding calorimeter jets. Another method is to use boosted W -bosons decaying hadronically. In ATLAS, there is an abundance of top-quark pairs ($t\bar{t}$) being produced, and these events provide a rich sample of hadronic W -boson decays. If produced with significant boost, the W -boson decay products can be contained in a single large radius jet, and the resulting jet mass spectrum can be compared between data and MC. This study is described in detail in the next Sections.

5.3. Hadronic W -boson with boosted WH analysis

A validation of the jet mass scale has been performed using W -boson decays in $t\bar{t}$ events. This validation was first performed in the context of the boosted $WH(b\bar{b})$ search, where it is fundamental to reconstruct the mass of a large- R groomed jet and to understand how well it is modelled by Monte-Carlo generators.

Figure 5.1 shows one of the first peak structures observed in ATLAS with jet substructure techniques, corresponding to boosted W -bosons decaying hadronically in $t\bar{t}$ events. It shows the mass distribution for a split/filtered Cambridge/Aachen jet after a basic $WH(b\bar{b})$ event selection, excluding the b -tagging requirement on the subjets, as is summarized in the following list:

1. Events are required to pass single-lepton triggers with a transverse momentum threshold of 20 GeV for electrons and 18 GeV for muons.
2. At least one isolated lepton (electron or muon) must be identified and reconstructed in the central region of the detector with $p_T > 25$ GeV.

3. The missing transverse energy (E_T^{miss}) is required to be greater than 25 GeV, to reduce contamination from events without a large real E_T^{miss} , but one originating for example from instrumental effects or leptons escaping detection.
4. A leptonic W -boson candidate is constructed from the selected lepton and E_T^{miss} . Its transverse mass² is required to be greater than 40 GeV, which helps reducing the QCD multijet background.
5. The leptonic W -boson candidate must have $p_T > 200$ GeV.
6. One Cambridge/Aachen split/filtered jet must be reconstructed with $p_T > 180$ GeV and $|\eta| < 2.5$.
7. The W and the Cambridge/Aachen jet must be back-to-back: $\Delta\phi(W, \text{jet}) > 1.2$.

Additionally, data quality requirements are imposed to select periods of data taking when the accelerator and detector conditions were stable. Non-collisions sources or detector noise are reduced by rejecting events where jets do not pass certain quality criteria, as described in Reference [95].

The jet mass is measured with a subset of 2011 ATLAS data, and compared to MC simulation, where the main processes are $t\bar{t}$ generated with MC@NLO [52] interfaced with HERWIG [47] and JIMMY [96] for parton showers; W +jets with ALPGEN [54] with HERWIG and JIMMY; and WW production with HERWIG and JIMMY. The MC samples are normalized to the highest order cross-section calculation available.

A jet energy calibration derived from simulated QCD jets is applied, but no correction for the specific environment of these events is applied. No corrections are applied to MC simulated events, and at this stage no systematic uncertainties were evaluated. Despite this, it can be seen that the mass distribution is reasonably well described by MC simulation.

The distribution in Figure 5.1 shows that the W -boson mass can be used for controlling systematics in an analysis where jet substructure techniques are employed. It is not expected that the mean value of this distribution matches the world-average for the W -boson mass, due to pile-up, underlying event and detector effects. What is of interest is the comparison of this mean between data and simulation, and also

²Defined as $m_T = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos \Delta\phi(\ell, E_T^{\text{miss}}))}$, where E_T^{miss} is assumed to correspond to the neutrino's transverse energy.

a comparison of the corresponding resolution. This first result, reported in [97], encouraged a dedicated jet mass validation using hadronic W decays in $t\bar{t}$ enriched samples, where these comparisons are used to derive uncertainties and where the associated systematic variations are included. This work is detailed in the next Section.

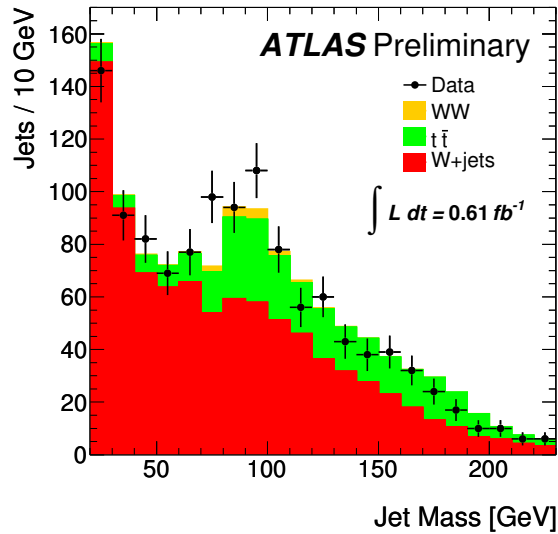


Figure 5.1.: Invariant mass distribution of a Cambridge/Aachen split and filtered jet, in events where a W -boson decays leptonically, with $p_T > 200$ GeV. Data points correspond to a small subset of 2011 ATLAS data, at a centre-of-mass energy of 7 TeV.

5.4. Jet mass scale validation with hadronic W -bosons

The study described in this Section is intended as proof of principle of a technique to determine how well the ATLAS detector simulation models the response of the calorimeter to the mass of jets.

This particular technique is applicable to a specific type of jet, with an invariant mass close to 80 GeV, and in a particular p_T range. As mentioned before, this is not the only way to constrain the JMS uncertainty, and more studies are reported in [87]. A combination of the available methods could potentially extend the applicability of the derived uncertainties.

The $t\bar{t}$ signature considered corresponds to approximately 15% of the total $t\bar{t}$ production cross-section: one of the W -bosons decays leptonically to a muon and a neutrino, while the other decays to quarks. The leptonic signature is used to distinctively trigger the event and the hadronic products to probe the JMS uncertainty.

To enhance the selection of $t\bar{t}$ events and suppress other processes, the selection described in the previous chapter is improved with additional conditions: events are required to have at least one anti- k_t $R = 0.4$ jet with $p_T > 20$ GeV and separated from the large- R jet by $\Delta R > 1.0$. At least one of these jets is required to be tagged as a b -jet (using the MV1 algorithm at the 70% efficiency working point).

The hadronic W -boson is identified with one of two algorithms, which will be compared³:

- Cambridge/Aachen split/filtered jets, with $R = 1.2$. The parameters of the splitting/filtering are the ones described in Chapter 4.
- anti- k_t trimmed jets, with a R -parameter of 1.0. The trimming parameters are $R_{sub} = 0.3$ and $f_{cut} = 0.05$.

The baseline MC samples used in this study are the same as in the previous Section. Event and object corrections are applied to improve data to MC agreement: the average number of interactions per bunch crossing in MC is adjusted to the conditions in data; object scale factors are applied to match the efficiencies in simulation to those measured in data (b -tagging, muon identification); muon momentum smearing factors and in-situ corrections to jet energy measurement are applied.

The jet mass distributions resulting from this selection can be seen in Figure 5.2, for the two implemented jet algorithms. The plots shown in this Section correspond to the full 2011 dataset. This study was repeated with the full 2012 dataset and the results are included here for completeness.

The multi-jet contribution (shown in purple) is due to jets that produce lepton-like signatures. It is estimated through a data-driven method, by inverting the isolation requirements on the muon and fitting the E_T^{miss} distribution, taking as template the difference between data and MC.

³No dedicated optimization of the grooming algorithms was performed for this analysis.

To extract the jet mass scale, a fit is performed to the peak in both data and MC. Assuming the signal has a Voigtian shape and the subtracted background ($t\bar{t}$, single-top and WW) a quadratic polynomial shape, the fit function is defined as follows:

$$f(m) = Ax^2 + Bx + C + D\text{Voigt}(x - \mu; \sigma, \gamma) \quad (5.1)$$

The multi-jet and W +jets backgrounds are subtracted to simplify the background modelling. The mean of the Voigtian distribution is μ , which is the parameter of interest for the jet mass scale measurement. σ is the width of the Gaussian contribution and γ is the width of the Lorentz contribution. For the jet mass scale measurement, the free parameters are A , B , C , D , μ and σ , while γ is fixed to the world-average W -boson width [1].

The results of the fit can be seen in Figure 5.3 for the Cambridge/Aachen split/filtering case, using the full 2011 dataset. For Cambridge/Aachen jets, the obtained $\mu_{data}/\mu_{MC} - 1$ value is $(-0.6 \pm 1.0(\text{stat})_{-1.7}^{+1.9}(\text{syst}))\%$. For anti- k_t trimmed jets the value is $(+0.5 \pm 1.2(\text{stat}) \pm 2.7(\text{syst}))\%$. Repeating the same study with the 2012 dataset allows for the reduction of statistical and systematic uncertainties. The final scales obtained are $(+0.27 \pm 0.58(\text{stat})_{-0.87}^{+0.75}(\text{syst}))\%$ for Cambridge/Aachen jets and $(-0.22 \pm 0.58(\text{stat})_{-0.76}^{+0.71}(\text{syst}))\%$ for anti- k_t trimmed jets. A good agreement between data and MC has been found for the jet mass scale. The modelling of the $t\bar{t}$ hard process and parton shower, and the knowledge of the large- R jet p_T scale are some of the sources of systematic error. The dominant contribution to the final systematic uncertainty is related to the modelling of initial and final state radiation in the $t\bar{t}$ process, for both the 2011 and 2012 datasets.

The extraction of the jet mass scale was also performed in bins of jet $|\eta|$, jet ϕ , beam μ and event N_{PV} , where no statistically significant dependence was found.

On the 2012 dataset, this measurement was used to extract not only the jet mass scale, but also the jet mass resolution. For this purpose, the fit described above is repeated, with σ as parameter of interest. In contrast with the μ extraction, this measurement is more strongly affected by statistical uncertainties, and stronger correlations with the signal normalization and background shape modelling are observed. The final resolution measurement obtained is $(-4.4 \pm 9.1(\text{stat})_{-8.5}^{+13.5}(\text{syst}))\%$ for the

Cambridge/Aachen jets and $(+0.1 \pm 6.7(\text{stat})_{-10.9}^{+11.6}(\text{syst}))\%$ for the anti- k_t trimmed jets. Among the largest systematics affecting these results are the hard process modelling (by comparing POWHEG to other MC prediction, of which the comparison to ALPGEN provides the largest variation), the JES on the Cambridge/Aachen case, and the fit linearity on the anti- k_t trimmed case.

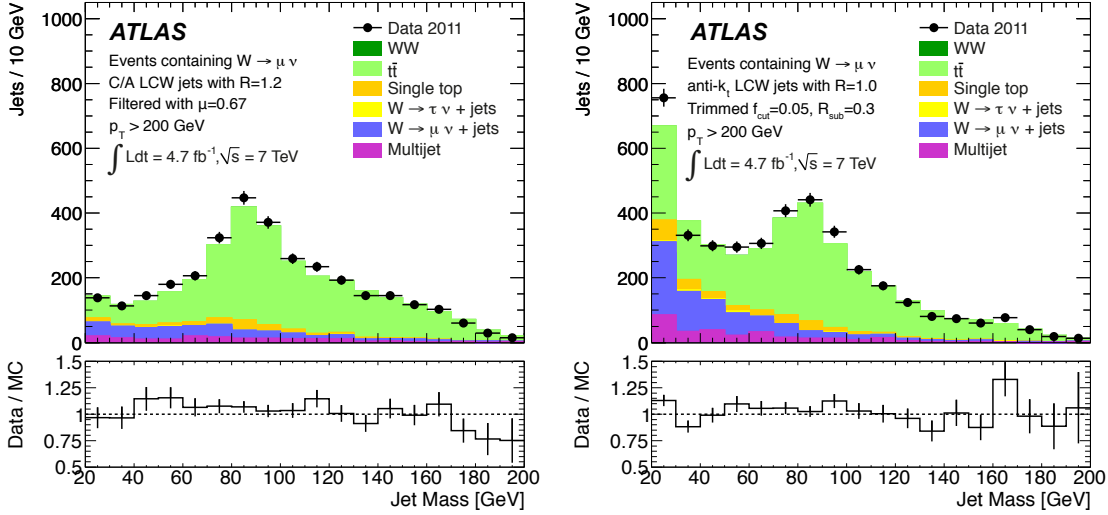


Figure 5.2.: Jet mass distribution of hadronic W -boson candidates in events where a leptonic W has been identified, using Cambridge/Aachen jets after splitting and filtering (left) and anti- k_t jets after trimming (right). [87]

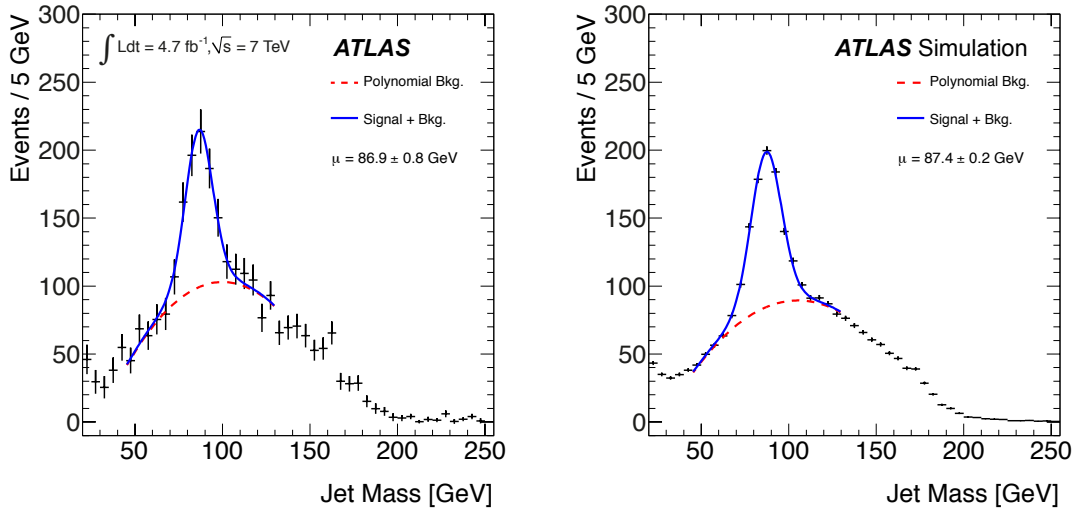


Figure 5.3.: Results of fit to subtracted jet mass distribution in data (left) and Monte Carlo (right), for Cambridge/Aachen jets after splitting and filtering. [87]

Chapter 6.

Boosted $WH(b\bar{b})$ searches

“Playing ‘bop’ is like playing Scrabble with all the vowels missing.”

— Duke Ellington

In this Chapter, $H \rightarrow b\bar{b}$ searches in the boosted regime are presented. The focus is on the WH production channel, but the main conclusions can be qualitatively extended to all $VH(b\bar{b})$ final states. Using the ATLAS 7 TeV dataset, the jet substructure approach was tested and compared to the 7 TeV $VH(b\bar{b})$ publication [98], in order to establish its feasibility and sensitivity.

The jet grooming technique used throughout this Chapter for the identification of the Higgs boson candidate is the split and filtering algorithm applied on Cambridge/Aachen jets ($R = 1.2$), unless otherwise stated.

6.1. Boosted $WH(b\bar{b})$ analysis at 7 TeV

6.1.1. Event Selection

The identification of the W -boson through its leptonic decay follows a similar procedure to the one detailed in Chapter 5, with the addition of a lepton veto, to reduce the contribution of dileptonic $t\bar{t}$ events, and spatial overlap cuts to avoid double-counting of objects and take into account electrons and muons that produce jet-like signatures or vice-versa [98].

The Higgs candidate selection is performed only in events where a W -boson with transverse momentum greater than 200 GeV is reconstructed, and consists of the following requirements:

- Exactly one Cambridge/Aachen $R = 1.2$ jet passing the splitting/filtering procedure, with $p_T > 180$ GeV and $|\eta| < 2.5$, which will be referred to as signal jet.
- The W and signal jet must be back-to-back, $\Delta\phi(W, \text{signal jet}) > 1.2$.
- The p_T leading and sub-leading subjects in the signal jet are b -tagged, using the MV1 tagger at an efficiency working point of 70%.

The standard Higgs candidate selection described in [98] was performed for comparison: two central anti- k_t $R = 0.4$ jets are required, both b -tagged with the same tagging algorithm and with $p_T > 45, 25$ GeV. This approach will be referred to as “resolved” or “standard” in the following discussion.

The contamination of $t\bar{t}$ events can be greatly reduced by applying a jet veto, which targets the semi-leptonic decay configuration. Events are rejected if anti- k_t jets with $p_T > 20$ GeV and $|\eta| < 4.5$ are found outside the volume of the signal jet (i.e. $\Delta R(\text{anti-}k_t, \text{signal jet}) > 1.2$). As the boost of the $t\bar{t}$ system increases, the region of phase-space enhanced by the selection cuts corresponds to events where the b and W hadronic decay products ($W \rightarrow cs$) merge into a single large jet, and where a bc pair passes the b -tagging selection, given the relatively high charm mis-tag rate.

The contamination of jets originating from pile-up interactions is dealt with by considering only anti- k_t jets with a JVF greater than 0.75. This cut is valid for jets with $p_T > 20$ GeV and $|\eta| < 4.5$, on the 2011 dataset. Other jet quality criteria are applied, as described in Chapter 3. There is no procedure analogous to the JVF for large- R jets, however, the MC samples used to derive the JES correction include pile-up events. Furthermore, the jet grooming itself is already a powerful tool to reduce the contribution from pile-up and underlying events.

6.1.2. MC simulation

This study was performed with signal and background samples generated at a centre-of-mass energy of 7 TeV and subject to the full GEANT4 ATLAS detector simulation.

Table 6.1.: List of Monte-Carlo samples used for signal and background simulation.

MC Generator	Process
POWHEG + HERWIG++	$WH, m_H = 120 \text{ GeV}$
SHERPA	$W + b\bar{b}$
ALPGEN + HERWIG	$W + l$
ALPGEN + HERWIG	$W + c$
ALPGEN + HERWIG	$W + c\bar{c}$
MC@NLO + HERWIG	$t\bar{t}$
MC@NLO + HERWIG	WZ

The Monte-Carlo samples are listed in Table 6.1. The W decays to leptons (e, μ, τ) are considered, as well as the hadronic decays in the case of $t\bar{t}$ (excluding cases where both W 's decay hadronically). The diboson sample consists of a leptonically decaying W and a Z -boson decaying to a pair of b -quarks. In the signal sample, the $b\bar{b}$ decay mode of the Higgs boson is considered.

To maximize the available statistics in the region of phase-space this study focuses on, a generation filter was applied to all samples, requiring the reconstruction of at least one Cambridge/Aachen jet with $p_T > 150 \text{ GeV}$, a central lepton (e, μ) with $p_T > 15 \text{ GeV}$, and missing transverse energy greater than 15 GeV . An additional filter is present in the signal and $W + b\bar{b}$ samples, requiring the transverse momentum of the W to be greater than 100 GeV . Potential generation biases on the variables of interest were investigated and found to be negligible.

The choice of MC generators for the simulation of the $W + \text{jets}$ background is not perfectly consistent: as a consequence of the limited size of the ALPGEN $W + b\bar{b}$ simulation available in ATLAS, this process was generated with SHERPA with up to two additional partons. A POWHEG + PYTHIA $W + b\bar{b}$ sample was also included for comparison.

6.1.3. Results

The invariant mass of the Higgs candidate ($m_{b\bar{b}}$) is calculated using the resolved and the jet substructure methods, as shown in Figure 6.1. The $m_{b\bar{b}}$ distributions have

distinct shapes, related to the characteristics of the jet(s) used in the reconstruction. While Cambridge/Aachen jets limit the high mass region due to their cone size ($R = 1.2$), the resolved reconstruction limits the mass at low values, due to the lower bound on the angular separation between the anti- k_t jets.

The main background contributions originate from $t\bar{t}$ and $W + b\bar{b}$ processes. All processes are normalized to the highest-order available cross-section, except $W + b\bar{b}$: given the large uncertainty on its total normalization (e.g., predictions from SHERPA and POWHEG differ by a factor ~ 4 , with POWHEG strongly overshooting data), a scale factor is estimated from data. Using the $m_{b\bar{b}} < 80$ GeV sideband of the jet substructure distribution, the scale factor was determined to be 1.45 ± 0.5 and applied to the SHERPA prediction. The choice of SHERPA instead of POWHEG is motivated by the observed data to simulation agreement and by the $W + \text{jets}$ study with 2010 ATLAS data [99,100], given that no differential cross-section measurement of $W + b$ jets had been performed. Furthermore, it follows from SHERPA's higher accuracy in the prediction of extra jet multiplicities that the jet veto should in principle be better described. The results presented in this Section were however cross-checked with both generators, and the conclusions with respect to the performance of each approach were found to remain valid.

In the mass window 110 – 130 GeV of a 120 GeV Higgs boson, a signal-to-background ratio (S/B) of approximately 16% is obtained for the resolved approach, and 23% for the substructure-based analysis. This $\sim 40\%$ gain is however less significant when calculated for a 125 GeV mass Higgs: a higher mass implies that the phase-space for the boosted regime where the jet substructure techniques may become helpful is reduced.

Profile likelihood fits¹ were performed using the resolved and substructure-based invariant mass distributions as discriminants. The background processes considered were $W + b\bar{b}$ and $t\bar{t}$, and the systematic uncertainties limited to a 10% prior on their normalizations. Figure 6.2 shows the expected and observed 95% confidence level [101, 102] upper limits on the WH production cross-section obtained with the resolved and substructure analyses. The sensitivities of the two approaches are equivalent for most mass points (approximately $12 \times \text{SM}$ at 125 GeV), with the exception that for higher masses the relative sensitivity of the Cambridge/Aachen

¹The statistical procedure used to obtain the results presented in this Section is explained in more detail in Section 7.4.

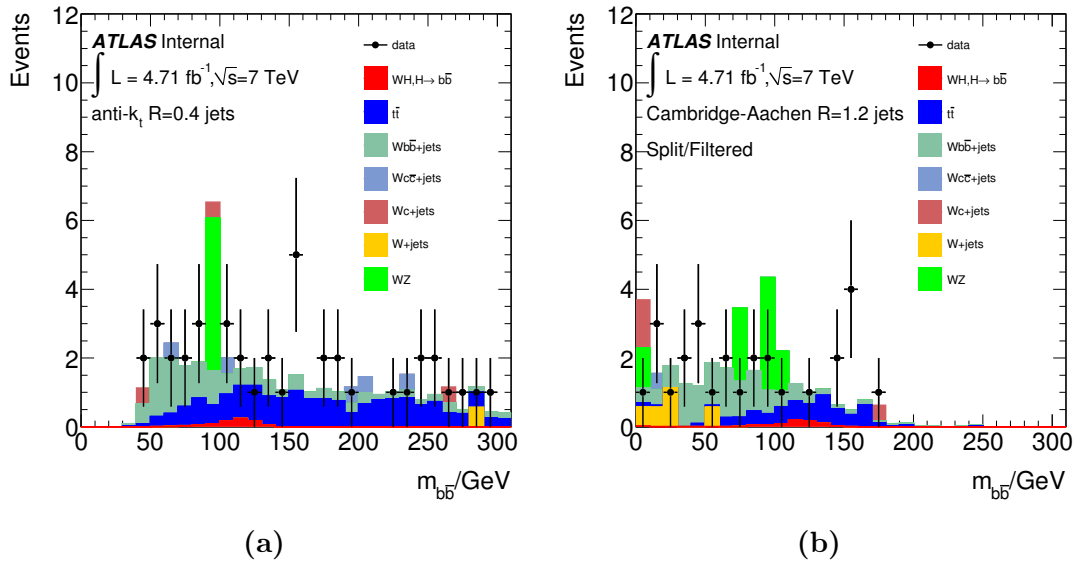


Figure 6.1.: Invariant mass distribution of the Higgs candidate using a) the standard jet collection and b) the jet substructure method. The signal simulation sample corresponds to a Higgs boson mass of 120 GeV, and the data plotted to 4.71 fb^{-1} of integrated luminosity.

selection decreases: as the Higgs boson becomes heavier, the contribution of the boosted regime to the overall cross-section decreases. Assuming a luminosity of 20 fb^{-1} and only statistical uncertainties, the limits approach $5\text{-}6 \times \text{SM}$ at 125 GeV, still far from having sensitivity to the SM predicted cross-section.

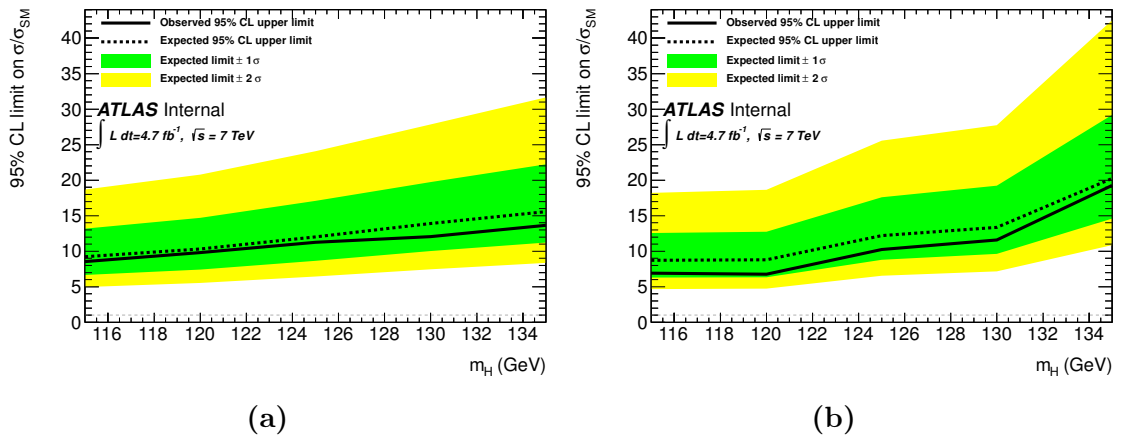


Figure 6.2.: Expected and observed 95% CL limits on the WH cross-section using a) the standard jet collection and b) the jet substructure method for the identification of the Higgs boson.

An overlap should exist between the events selected by each of the reconstruction methods. Considering only $t\bar{t}$ and $W + b\bar{b}$ backgrounds and the signal sample at $m_H = 120$ GeV, it was observed that in the Higgs boson mass window, the overlap between the substructure-based and resolved approaches is much larger in the signal than in the background processes, as can be seen in Figure 6.3. Quantitatively, inside a mass window of 86-134 GeV, 60% of the signal events are reconstructed by both approaches, while that is only the case for 29% of $t\bar{t}$ and 44% of $W + b\bar{b}$ events. Furthermore, the resolved approach has a higher signal efficiency (already expected given the available $p_T(W)$ phase-space at this energy).

Given the different signal and background acceptances associated to each approach, the possibility of increasing the signal significance by combining exclusive sets of events was exploited. Combining events uniquely reconstructed by the substructure-based approach with events reconstructed by the resolved approach, for example, or considering only events which are reconstructed by both approaches, are possibilities which were tested. It was found that, although an increase in the signal-to-background ratio can be obtained, the signal over the square root of the background (S/\sqrt{B}) and the expected 95% CL upper limit do not change significantly.

6.2. Optimization of jet substructure algorithms

Despite having been developed specifically for the boosted Higgs case, the splitting and filtering technique was compared to other available jet substructure algorithms, in terms of signal efficiency and background rejection in the context of a $WH(b\bar{b})$ search, for a Higgs mass of $m_H = 125$ GeV and at a centre-of-mass energy of $\sqrt{s} = 8$ TeV. The signal sample was generated with PYTHIA 8. For simplicity, the background processes considered are only $t\bar{t}$ and $W + b\bar{b}$, the latter generated with POWHEG+PYTHIA². The simulated samples used in this study are not subject to event generation filters to enhance the boosted phase-space.

²As can be deduced from the MC samples used throughout this Chapter, there is no clear indication of which generator performs better in describing $W + b\bar{b}$ events. This issue will be one of the focuses of the next Chapter's discussion.

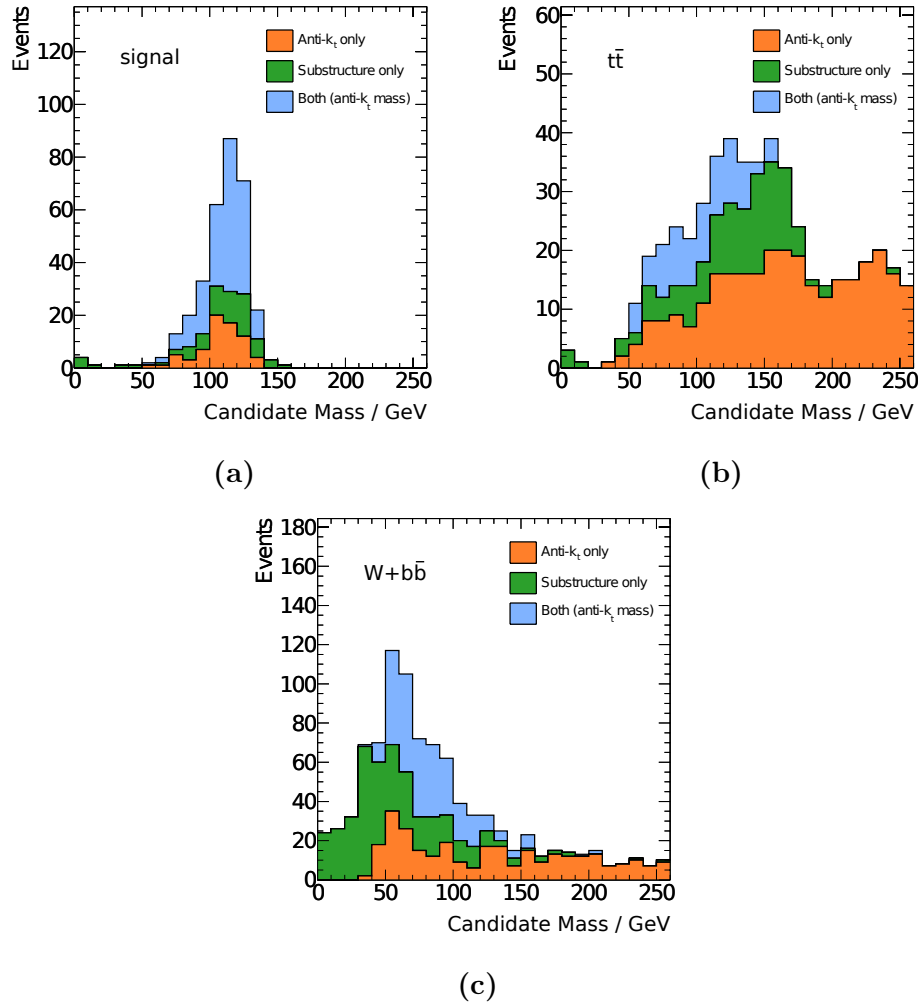


Figure 6.3.: Invariant mass distributions of the reconstructed Cambridge/Aachen jet and anti- k_t dijet system in (a) signal events with $m_H = 120$ GeV, (b) $t\bar{t}$ events and (c) $W + b\bar{b}$ events. The exclusively resolved and substructure regions are shown, as well as the set of events which are reconstructed by both methods. In these cases, the invariant mass of the dijet anti- k_t system is considered.

Given the wide use of trimming and pruning algorithms in ATLAS, particularly in boosted top resonance searches, these techniques are validated and their calibration available, making them good candidates for this study.

The splitting and filtering procedure was tested in Cambridge/Aachen jets of $R = 1.2$ (default) but also $R = 1.3, 1.5$. The trimming procedure was applied to anti- k_t jets of $R = 1.0, 1.3, 1.5$ with a scan of the parameters f_{cut} and R_{sub} . The pruning algorithm was also tested on anti- k_t jets with $R = 1.0, 1.3$ and with several values of z_{cut} and R_{cut} . Finally, the impact of the N -subjettiness substructure variable

was studied. To establish a baseline, the standard anti- k_t $R = 0.4$ selection is also evaluated. The selection of the baseline events is equivalent to the one described in the previous Section but with an important addition: an upper cut on the angular separation between the two leading jets is imposed, $\Delta R(j_1, j_2) < 1.4$, according to the ATLAS $VH(b\bar{b})$ analysis reported in Ref. [103].

The optimization study was performed by selecting events where the large- R groomed jet is overlapping in the (ϕ, η) plane with two b -tagged anti- k_t $R = 0.4$ jets. This requirement is imposed by the technical difficulty of deriving b -tagging information for all the jet algorithms tested. Such an approach constrains the comparison to topologies where the resolved reconstruction of the two b -jets is still possible, but allows for the study of the background rejection and invariant mass resolution with respect to the anti- k_t $R = 0.4$ reconstruction. Also for technical reasons, there is no p_T and η calibration applied on the large- R jets.

6.2.1. Background rejection

The values of S/B and S/\sqrt{B} were evaluated for each jet algorithm, jet radius size and set of grooming parameters, in the 110-140 GeV mass window. The best performing cases (in comparison to the resolved case) are listed in Table 6.2. The statistical uncertainty is no larger than 10% for S/B and of the order of 20% for S/\sqrt{B} , dominated by the $t\bar{t}$ sample statistical uncertainty. Figure 6.4 shows the resulting invariant mass distributions.

Table 6.2.: Values of S/B and S/\sqrt{B} for the best performing substructure methods. A mass window of 110-140 GeV is considered.

Method	S/\sqrt{B}	S/B (%)
Resolved (two anti- k_t $R = 0.4$ jets)	0.56	10.57
Cambridge/Aachen splitting and filtering, $R = 1.2$	0.48	12.54
anti- k_t $R = 1.0$ pruned ($z_{cut} = 0.15$, $R_{cut} = 0.4$)	0.43	13.27
anti- k_t $R = 1.0$ trimmed ($f_{cut} = 0.05$, $R_{sub} = 0.3$)	0.46	14.80
anti- k_t $R = 1.0$ trimmed ($f_{cut} = 0.05$, $R_{sub} = 0.2$)	0.42	15.22
anti- k_t $R = 1.5$ trimmed ($f_{cut} = 0.03$, $R_{sub} = 0.3$), $\tau_{21} < 0.4$	0.52	12.11

As can be concluded from the numbers and distributions presented, reconstructing the Higgs candidate with an appropriate jet size and substructure technique can provide very strong background rejection. Among the configurations that were tested, the Cambridge/Aachen splitting and filtering approach achieves a competitive S/B and S/\sqrt{B} . The trimmed anti- k_t jets were found to have the best signal-to-background ratio, recovering the 40% improvement with respect to the resolved approach, seen in the previous Section with Cambridge/Aachen jets at a mass point of 120 GeV. A similar gain is obtained by using the pruning technique. It was observed that increasing the radius of the large- R jet generally damages the S/B : moving to a larger radius can double the signal efficiency, but dramatically limit the power to reject $t\bar{t}$. In the particular case of anti- k_t trimmed jets with $R = 1.5$, this loss is recovered by cutting on the N -subjettiness variable τ_{21} (shown in Figure A.1), which is effective against 3-prong topologies. These results, while encouraging, are limited by the statistical power of the available $t\bar{t}$ sample.

Overall, the signal significance obtained in the resolved case (here approximated by S/\sqrt{B}) could not be matched with any of the techniques used. As mentioned before, the main selection optimization that justifies this observation is the upper $\Delta R(j_1, j_2)$. In fact, tightening this cut further reproduces the improvements in S/B observed with the jet substructure techniques.

6.2.2. Resolution of the mass peak

To finalize the performance assessment of jet definitions for $H \rightarrow b\bar{b}$, the sharpness of the signal peak was compared. The figures of merit are chosen based on Reference [104], where the quality of the peak is quantified by measuring two different values:

- The smallest mass window that contains 68.2% of the Higgs boson candidates after the full event selection, $f(1)$.
- The inverse of the number of reconstructed Higgs boson candidates in a window of width $w = 1.25\sqrt{m_H} \text{ GeV}$ with respect to the total number of events in the sample, $f(2)$.

These quantities are designed to quantify the quality of the mass peak through its height and width. In the first case, different values for the fraction considered in the window revealed the same conclusions. In the second case, the window has a

width of 13.75 GeV, of the same order as the $m_{b\bar{b}}$ resolution in ATLAS. The signal mass distributions considered in this study are shown in Figure 6.5 and the obtained values for the figures of merit are shown in Table 6.3.

Table 6.3.: Values of the figures of merit as calculated for the resolved and substructure reconstruction methods.

Method	$f(1)$ (GeV)	$f(2)$
Resolved (two anti- k_t $R = 0.4$ jets)	30	3.85
Cambridge/Aachen splitting and filtering, $R = 1.2$	40	4.05
anti- k_t $R = 1.0$ pruned ($z_{cut} = 0.15$, $R_{cut} = 0.4$)	40	4.55
anti- k_t $R = 1.0$ trimmed ($f_{cut} = 0.05$, $R_{sub} = 0.3$)	40	4.36
anti- k_t $R = 1.0$ trimmed ($f_{cut} = 0.05$, $R_{sub} = 0.2$)	40	4.27
anti- k_t $R = 1.5$ trimmed ($f_{cut} = 0.03$, $R_{sub} = 0.3$), $\tau_{21} < 0.4$	40	4.35

The dijet anti- k_t $R = 0.4$ approach revealed the best performance, especially with the fixed width measure, closely followed by the Cambridge/Aachen split and filtered jets. The trimmed approaches performed very similarly: although with small selection efficiency, the reconstructed mass peaks have a relatively small width. The anti- k_t pruned jets showed a slightly worse performance, where its larger width is not compensated by the peak height. In summary, the evaluation of the figures of merit listed above revealed that while the anti- k_t $R = 0.4$ approach achieves the minimum values for both, all six reconstruction approaches perform quite similarly, especially taking into account the lack of pile-up corrections on the large- R jets.

6.3. Summary

Two distinct approaches for the reconstruction of a boosted Higgs boson were presented in this Chapter, in the context of a $WH(b\bar{b})$ search with the 7 TeV ATLAS dataset. It has been shown that using jet substructure techniques on the boosted regime of the $WH(b\bar{b})$ channel, specifically split and filtered Cambridge/Aachen jets, does not result in a sensitivity gain when compared to the standard (resolved) approach. The S/B improvement initially observed in simulation for a Higgs boson mass of $m_H = 120$ GeV is diluted when moving to the 125 GeV mass value: a higher

boost is needed in this case so that the $b\bar{b}$ decay is contained within the radius of the Cambridge/Aachen jet, and the already small phase-space available at 7 TeV becomes even more of a restriction. Both approaches revealed similar sensitivities, both being far from probing the SM values.

MC studies were performed as an attempt to optimize the jet grooming techniques validated in ATLAS for the purpose of the $H \rightarrow b\bar{b}$ reconstruction. The groomed large- R jets can be very powerful in rejecting the $t\bar{t}$ background, and also reducing the $W + b\bar{b}$ contamination. Despite this, it was observed that there is no obvious large improvement to be obtained from these techniques at this stage. However, the limited scope of this study must be taken into account, as it does not consider events with merged decay products.

When the LHC centre-of-mass energy increases to 14 TeV, the jet substructure approach is expected to have greater relevance, as will be explored in Chapter 8. This can be in the context of a combination with the resolved approach at lower boosts, or by relying only on the boosted region. At 8 TeV, the opening of the boosted phase-space is not expected to be significant enough for the resolved approach to limit the sensitivity of the search. There are, however, performance and optimization aspects that can already be addressed with Run 1 data, such as the jet veto (complicated by the combination of two jet algorithms) or the b -tagging of subjets, essential for a substructure-based $H \rightarrow b\bar{b}$ search. Both these issues are helpful for dealing with $t\bar{t}$, a reducible contamination. On the other hand, it should also be stressed that there are large theoretical uncertainties on the $W + b\bar{b}$ background, as will be shown on the next Chapter, which deserve a dedicated study with substructure-based techniques.

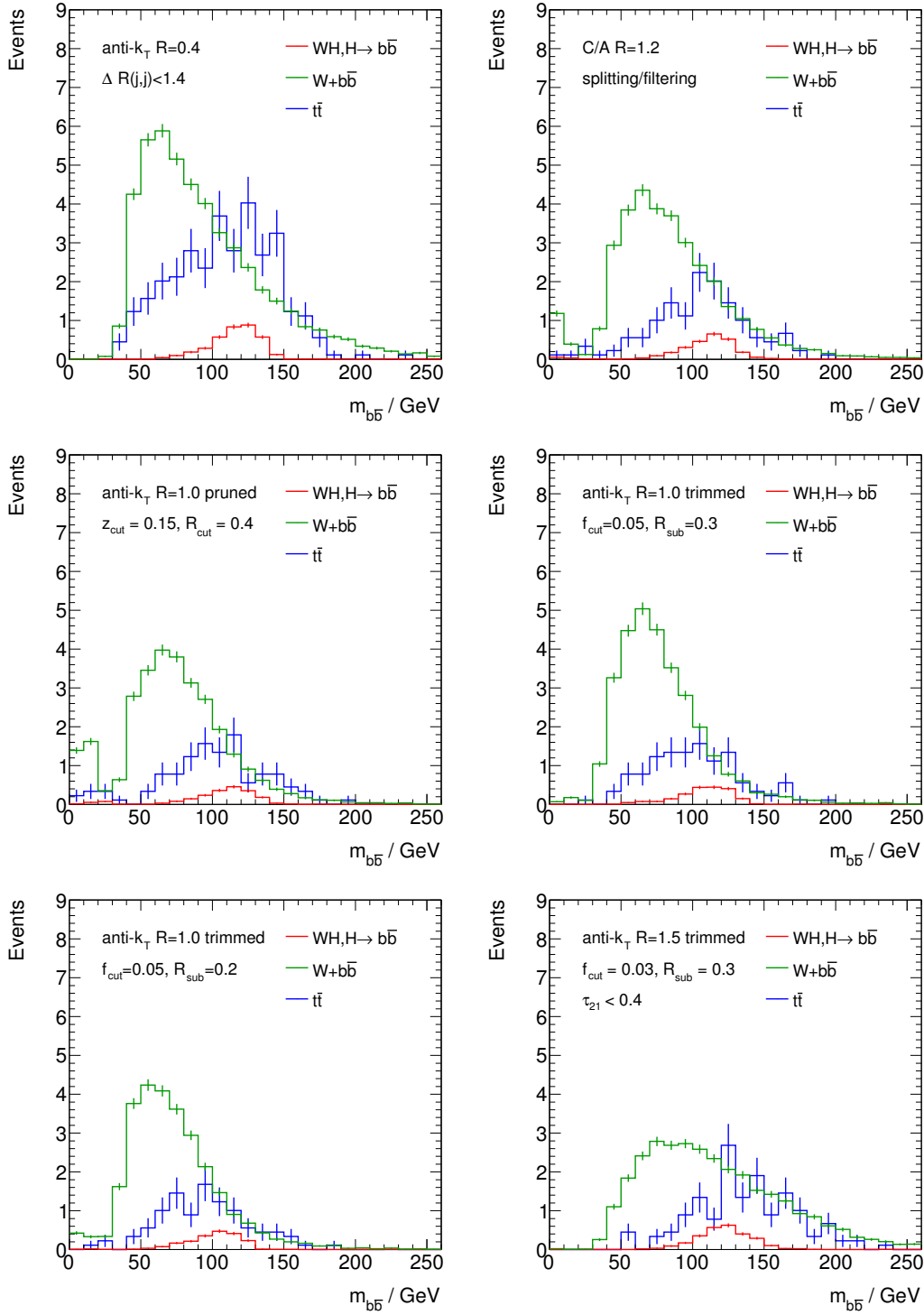


Figure 6.4.: Invariant mass of the reconstructed Higgs candidate with the resolved approach and several substructure-based methods. $m_H = 125$ GeV is considered in the signal sample generation. The QCD next-to-leading order cross-sections were used, and the samples normalized to a luminosity of 10 fb^{-1} .

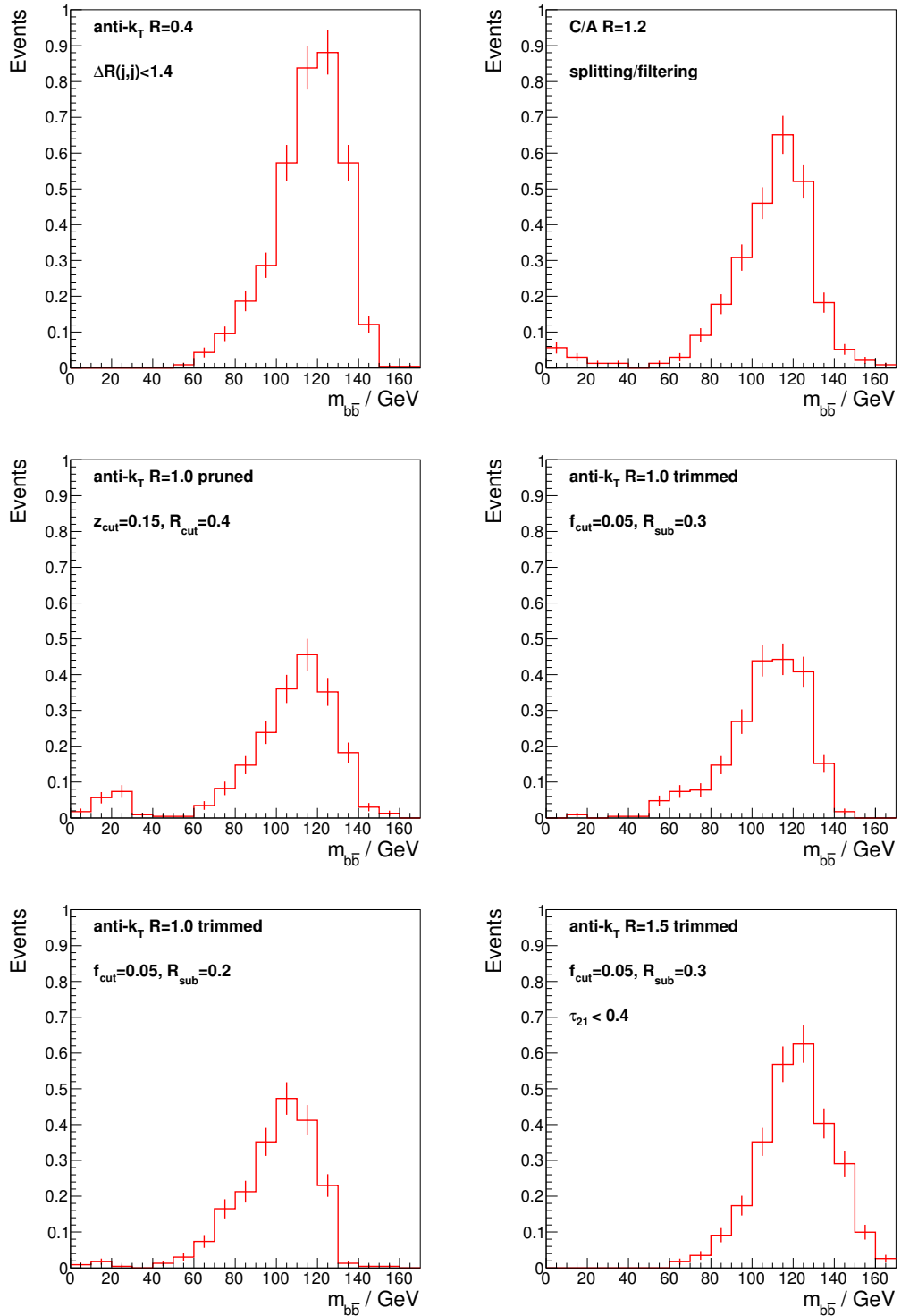


Figure 6.5.: Invariant mass of the Higgs candidate in the $WH(b\bar{b})$ simulation sample, reconstructed with the resolved approach and several substructure-based methods. $m_H = 125$ GeV is considered in the sample generation. The QCD next-to-leading order cross-sections were used, and the samples normalized to a luminosity of 10 fb^{-1} .

Chapter 7.

Run 1 $VH(b\bar{b})$ search

“I start in the middle of a sentence and move both directions at once.”

— John Coltrane

The observation of the $H \rightarrow b\bar{b}$ decay is crucial to establish the nature of the Higgs boson, constraining its width and couplings. The CDF and D0 collaborations have reported an excess of 2.8σ on their $VH(b\bar{b})$ search [105], and CMS has observed an excess of 2.1σ [106], for $m_H = 125$ GeV. The ATLAS Run 1 result on the $VH(b\bar{b})$ search [107] is presented in this Chapter, with a focus on the $WH(b\bar{b})$ channel.

7.1. Analysis Strategy

The $VH(b\bar{b})$ search combines three final-state topologies, according to the W/Z decay modes, as shown in Figure 2.3. In order to target each final-state as efficiently as possible, the analysis combines three channels, according to the number of reconstructed charged leptons (e^\pm, μ^\pm , referred to from now on as signal leptons). The resulting 0-, 1- and 2-lepton channels are composed dominantly by the $Z \rightarrow \nu\bar{\nu}$, $W \rightarrow \ell\nu$ and $Z \rightarrow \ell\bar{\ell}$ signal decay modes respectively, with a 1-2% contribution of $Z \rightarrow \ell\bar{\ell}$ to 1-lepton and $W \rightarrow \ell\nu$ to 0-lepton, when a lepton escapes undetected or unidentified. In the following Sections, the 1-lepton channel will be described, and comparisons to the 0- and 2-lepton channels will be made when relevant.

The vector-boson transverse momentum $p_T(V)$, is reconstructed from the signal lepton(s) and (or) E_T^{miss} information, depending on the channel: for the 0-lepton

channel, a veto on the number of leptons is applied and $p_T(Z)$ identified with the total missing transverse energy; on the 1-lepton channel the transverse momentum of the W corresponds to the vector sum of E_T^{miss} and the signal lepton p_T ; finally, on the 2-lepton channel the $p_T(Z)$ can be reconstructed with the two selected signal leptons (same flavor, opposite signs). The search is done inclusively in terms of $p_T(V)$ ¹, however the characteristics of the boosted regime are taken advantage of by performing the analysis in bins of $p_T(V)$.

The reconstruction of the Higgs candidate is performed with two anti- k_t $R = 0.4$ jets, which are identified as b -jets with the MV1c b -tagging algorithm. In addition to the categorization in terms of $p_T(V)$, each lepton channel is further split into different regions according to the number of jets (two or three) and number of b -tagged jets (one or two). The different event categories are listed in Table 7.1 (for example., 2-jet 2 loose tags, etc.).

In summary, the approach taken in this analysis is to define meaningful event regions with the available kinematic, topological and b -tagging information, so as to be able to measure and constrain the different background processes from data, and take advantage of regions with higher signal sensitivity. The background composition in the 1-lepton regions is described in Section 7.2.

The ATLAS $VH(b\bar{b})$ search is performed with two different approaches: one for which the final discriminant is the invariant mass of the Higgs candidate ($m_{b\bar{b}}$, referred to as dijet mass analysis) and one where the final discriminant is provided by a multivariate analysis (MVA) using several event and object variables. The latter provides the highest expected sensitivity and is therefore used for the main result, while the first is intended as a cross-check. To interpret the result, a likelihood function is constructed from the discriminant variable in the regions defined in all three channels and its maximization is performed through a “global fit”.

The full Run 1 dataset is used, however the dijet mass and MVA analyses here described only apply to the 8 TeV dataset. The 7 TeV analysis (based on the dijet mass approach) is described in Reference [108]. Some of the main differences are the jet calibration and the use of MV1 instead of MV1c as b -tagging algorithm.

¹With the exception of the 0-lepton channel, due to E_T^{miss} trigger limitations at lower energies.

Table 7.1.: Event categories used in the 1-lepton analysis (8 TeV dataset). Signal jets are defined in the next Section, as well as the b -tagging efficiency working points.

Variable	Regions
$p_T(V)$	dijet mass: 0-90, 90-120, 120-160, 160-200, > 200 (GeV) MVA: 0-120, > 120 (GeV)
number of signal jets	2, 3
number of b -tagged jets	1, 2
b -tagging efficiency	2 loose, 2 medium, 2 tight

7.1.1. Object and Event Selection

Events in the 1-lepton channel are selected if they pass a single-lepton trigger with a p_T threshold of 24 GeV and a track isolation criteria, or a higher p_T threshold trigger with no track isolation requirement (60 GeV for electrons, 36 GeV for muons). This reflects the high pile-up conditions of the 8 TeV dataset. Acceptance in the muon channel is limited by the lack of trigger-chamber coverage in some regions, and these cases are recovered by using E_T^{miss} triggers in addition to single-lepton triggers.

Electrons and muons can be identified with different degrees of purity (referred to as loose, medium and tight). In the 1-lepton channel, one tight electron or muon must be reconstructed in the event, and a veto on additional loose leptons is applied. Tight leptons are defined as having a p_T greater than 25 GeV and being reconstructed in the region $|\eta| < 2.5$ of the detector. Tight electrons and muons are also required to be isolated in the calorimeter (and in the Inner Detector): no more than 4% of the total lepton energy (p_T) should be associated to clusters (tracks) in a cone of radius 0.3 (0.2) centered around the lepton.

The jets considered in this analysis have $p_T > 20$ GeV and $|\eta| < 4.5$. The JVF cut already introduced in Chapter 3 to handle pile-up jets in the 7 TeV dataset was optimized for the 8 TeV conditions: at least 50% of the scalar p_T sum of tracks associated to the jet needs to be associated with the primary vertex. Moreover, only jets with $p_T < 50$ GeV and $|\eta| < 2.4$ are subject to this cut [68].

Two classes of jets can be defined, which will be referred to as signal and forward jets. Forward jets are those reconstructed in the $2.5 < |\eta| < 4.5$ region of the detector,

with $p_T > 30$ GeV. Events with forward jets are vetoed, which helps reduce $t\bar{t}$ contamination.

Signal jets, on which the event categorization is based, are defined as having $p_T > 20$ GeV and $|\eta| < 2.5$. The Higgs candidate is reconstructed from the two signal jets with highest transverse momentum, where an extra requirement of $p_T > 45$ GeV is applied on the leading one. On the dijet mass analysis, a series of $\Delta R(j_1, j_2)$ cuts is applied to the two leading signal jets, as listed in Table 7.2². The lower bounds of these cuts are optimized to reject combinatorial backgrounds and $V+$ jets events, where the jets might be collimated if they originate from the same QCD vertex. The upper bound is placed to reduce $t\bar{t}$ events, particularly at high $p_T(V)$. These cuts are dependent on the $p_T(V)$ bin considered and will have an effect, for example, on the flavor composition of the $t\bar{t}$ events selected in each category. The variable H_T corresponds to the scalar sum of the p_T of the two signal jets, the signal lepton and E_T^{miss} . It is therefore a measure of the hardness of the event, and can be used to reduce QCD multijet events.

Overlap removal is performed as described in Chapter 6, between loose leptons and signal jets, with the modification that loose muons can be kept when overlapping with a jet with low track multiplicity, in which case the jet is discarded. Additionally, criteria to select overlapping electrons and muons are defined.

Jets are identified as b -jets with different MV1c efficiency working points, which are also used to define event categories: 50% (tight, T), 70% (medium, M), 80% (loose, L). Consequently, 2 b -tag regions are split into Tight, Medium and Loose exclusive regions, according to the MV1c weight associated to each of the leading signal jets. The 1 b -tag event category is formed by events with only one b -tagged signal jet, tagged with the loose working point (the most inclusive one). The advantage of splitting 2 b -tag regions according to the tagging efficiency is related to the different charm and light rejection factors associated with each: looser b -tagging is equivalent to lower c and light rejection, which provides control regions for the $V+$ jets events where there are no real bottom-quarks. Additionally, categories with zero b -tagged jets are used for control and validation of $V+$ jets processes, but not used to build the likelihood function.

²This is not the case for the MVA-based approach, where a looser set of cuts is applied. The goal is to train a MVA technique that is able to extract information from the supplied variables, and one of the input distributions is $\Delta R(j_1, j_2)$.

Table 7.2.: Selection cuts applied in the 1-lepton channel, for both dijet mass and MVA analyses (8 TeV dataset).

Variable	dijet mass					MVA	
$p_T(W)$ (GeV)	0-90	90-120	120-160	160-200	>200	0-120	> 120
$\Delta R(j_1, j_2)$	0.7-3.4	0.7-3.0	0.7-2.3	0.7-1.8	<1.4	>0.7 ($p_T(W) < 200$ GeV)	
$m_T(W)$ (GeV)	<120					-	
H_T (GeV)	>180		-			> 180	-
E_T^{miss} (GeV)	-		>20	>50		-	> 20

A Boosted Decision Tree (BDT) [109, 110] is used to perform the MVA analysis, trained to separate the $VH(b\bar{b})$ signal against the total background. Variables such as $m_{b\bar{b}}$, $p_T(V)$, E_T^{miss} , $p_T(j_1)$, $p_T(j_2)$ are used as input to the BDT, as well as angular variables, b -tagging information on each jet and other kinematic variables.

7.2. Background composition

The main backgrounds to the 1-lepton channel search are $t\bar{t}$ and W +jets processes, as already described in Chapter 2. However, diboson and single-top processes, especially Wt , also contribute to the highest sensitivity regions. The MC generators used to simulate each process are listed in Table 7.3.

The dependence of the background composition with $p_T(W)$, number of jets and b -tagging efficiency can be seen in Figure 7.1 for 1-lepton channel events, after the dijet mass analysis selection. The regions with higher signal-to-background ratio are those with two signal jets and higher b -tagging efficiency, where $W + b\bar{b}$ and $t\bar{t}$ are dominant. In the 2-jet regions, the $t\bar{t}$ contributes mainly at low $p_T(W)$, with $W + b\bar{b}$ becoming more relevant for higher $p_T(W)$, and with a contribution from single-top. In the 3-jet region, the behaviour is similar, but with a more dominant $t\bar{t}$ fraction. Multijet QCD events are characteristic of the lower $p_T(W)$ bins. The flavor composition of the two signal jets in $t\bar{t}$ events varies with the $p_T(W)$ bin considered: as the boost of the system increases, the fraction of bc events increases until it becomes dominant, where the c originates from the W decay and is mis-tagged as a b .

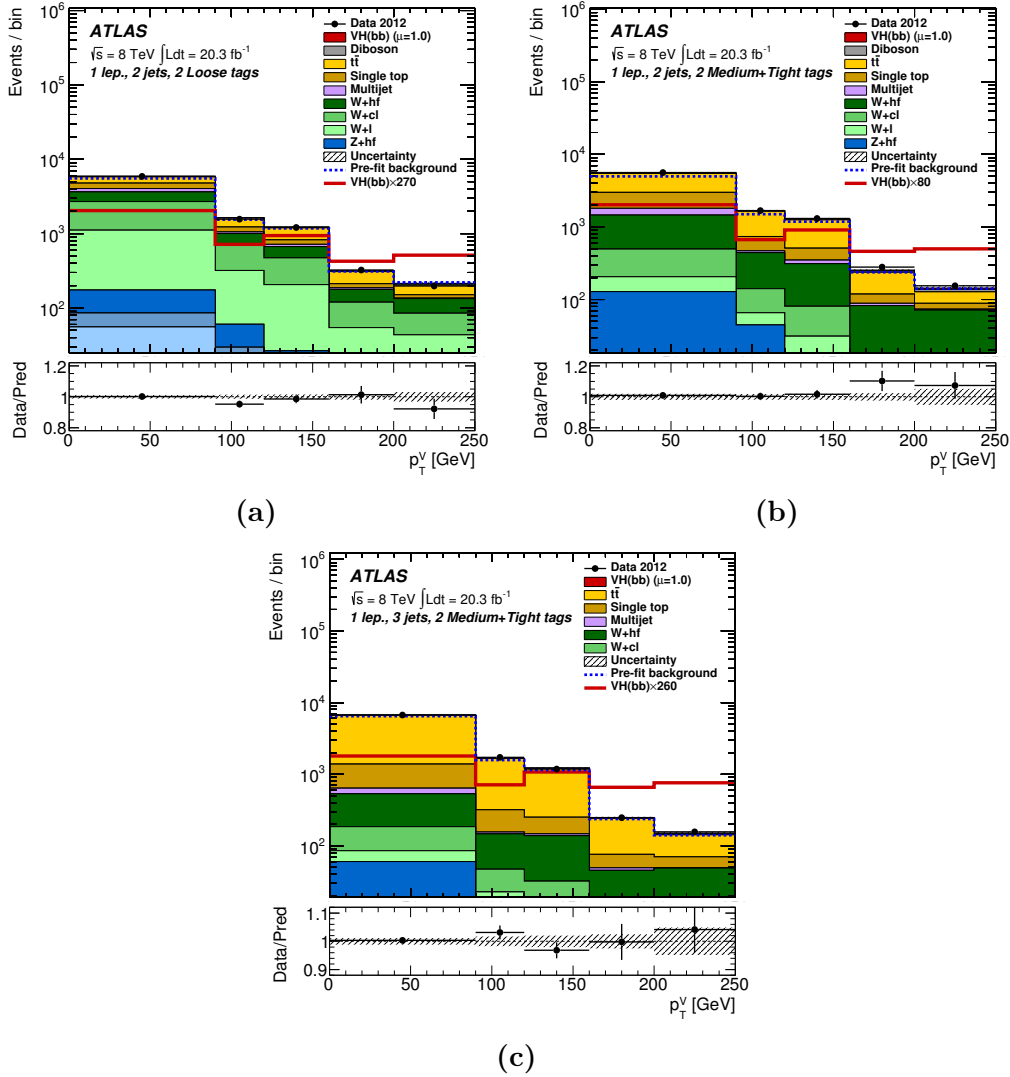


Figure 7.1.: $p_T(W)$ distributions comparing data and simulation (after global fit) in the 1-lepton regions with (a) 2 jets and 2 Loose tags, b) 2 jets and 2 Medium + Tight tags and c) 3 jets and 2 Medium + Tight tags. [107].

In the case of 2-lepton events, $Z + b\bar{b}$ processes are dominant over all the $p_T(Z)$ range, and the 0-lepton channel is subject to contributions from all the different backgrounds, but mainly $Z + b\bar{b}$ and $t\bar{t}$, as can be seen in Figure 7.2.

The V +jets processes are labelled in simulation according to the flavor of the two leading (signal) jets: Vbb , Vbc , Vcc , Vcl and Vl , where l stands for light ($l = g, u, d, s$). The flavor is determined by ΔR matching of the jet with its B hadron content, which if nonexistent is then checked for C hadron content and, failing that,

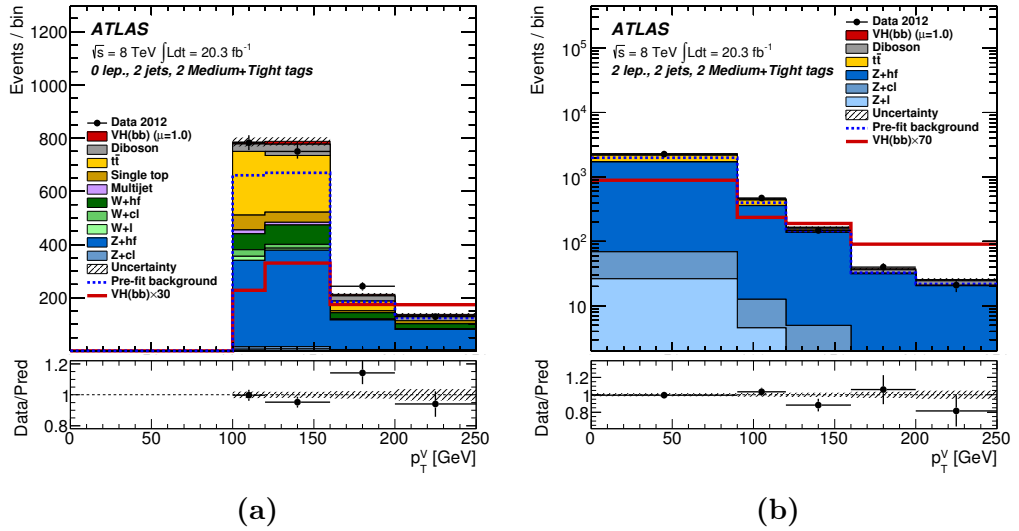


Figure 7.2.: $p_T(Z)$ distributions comparing data and simulation (after global fit) in the regions with (a) 0 leptons, 2 jets and 2 Medium + Tight tags, b) 2 leptons, 2 jets and 2 Medium + Tight tags. [107].

the jet is labelled as light. A cone of 0.4 radius centered on the jet axis is used for the matching criteria, and only hadrons with $p_T > 5$ GeV are considered.

The variety of the background composition among channels and categories is taken advantage of by the global fit, which is capable of determining and constraining the yields of each background process. This is particularly true for $t\bar{t}$ and V +jets, for which the data contains a powerful amount of information. The total yields of $t\bar{t}$, V + heavy flavor jets³ and Vcl are fully extracted from data, with no prior normalization constraint. Events labelled as Vcl are floated separately from Vb events given the different nature of the leading order diagrams contributing to each final state.

The power to constrain $t\bar{t}$ in the 1-lepton channel comes mainly from the 3-jet 2-tag regions. Vcl events have as control regions the 1-tag and 2 Loose tags regions⁴. While $Z + b\bar{b}$ is easily extracted from 2-lepton categories, constraints on $W + b\bar{b}$, on the other hand, do not originate from a particular region, but rather from the interplay with the other background processes in each region, as a single pure region in $W + b\bar{b}$ events does not exist in this analysis. The normalization of the remaining

³ V + heavy flavor corresponds to $b\bar{b}$, bc , bl and cc event labels, and will be denoted as $Vb = Vbb + Vbc + Vbl + Vcc$, or alternatively $V + h.f.$

⁴The sensitivity of the MVA analysis decreases by 4% if the 1-tag region is not considered.

Table 7.3.: MC simulation samples for signal and background processes. Additional generators and configurations are used to derive systematic uncertainties.

Process	Generator	Comments
WH and ZH	PYTHIA 8	$q\bar{q}$ initial states
ZH	POWHEG + PYTHIA 8	gg initial state
W/Z +jets	SHERPA	massive c and b quarks
$t\bar{t}$	POWHEG + PYTHIA	-
single-top Wt	POWHEG + PYTHIA	-
single-top s	POWHEG + PYTHIA	-
single-top t	ACERMC + PYTHIA	-
diboson	POWHEG + PYTHIA 8	-

smaller backgrounds is allowed to float within the theoretical uncertainties associated to their cross-sections, calculated including the highest-order corrections available.

7.3. Systematic uncertainties

It is important to understand how the signal and background processes behave in each fit region and how accurately the event categorization is modelled by simulation. Since the selected events are divided in terms of jet multiplicity and $p_T(V)$ bins, it is essential to assess how well these quantities are understood and predicted for each process. This is also true for the $m_{b\bar{b}}$ shape in each region and, in the case of the MVA approach, for other input distributions used to extract the final BDT. Finally, when labeling Vb processes by the leading jet flavors, an uncertainty needs to be estimated for how the split into bb , bc , bl and cc events is performed in simulation.

Therefore, modelling systematic uncertainties are estimated for each process, based on data to simulation comparisons when possible, otherwise on MC based studies. In the latter case, the prediction of the baseline MC is tested against other predictions, and the largest deviation is introduced in the fit as a systematic uncertainty. The fit is then allowed to extract the parameter value that best suits the data, within its previously defined uncertainty, and according to the other parameters the model is dependent on.

The basic modelling uncertainties considered in the fit are estimated on the following variables:

- Jet multiplicity, translated into an uncertainty on the ratio of number of events in the 2-jet and 3-jet regions (3-to-2-jet ratio);
- Flavor composition, where the relative yields of the bc , bl , cc components with respect to the bb rate are assigned an uncertainty;
- $p_T(V)$ and $m_{b\bar{b}}$ distributions, where for each a shape uncertainty is derived.

Uncertainties on variable shapes are introduced in the fit as an uncertainty band, with the alternative shape symmetrized around the nominal shape.

Other modelling uncertainties are related to corrections to the simulation. By performing data-to-MC comparisons, event reweighting functions are derived when deemed necessary to correct for mis-modelling in the simulation. In the 1-lepton channel 0-tag 2-jet region, the $\Delta\varphi(j_1, j_2)$ distribution is one such example: a slope is observed on the data/MC ratio, translating into an excess of close-by jets in simulation, and a corresponding lack of back-to-back jets. This is clearly shown in Figure 7.3, and the same effect is seen in Z +jets events in the 2-lepton channel.

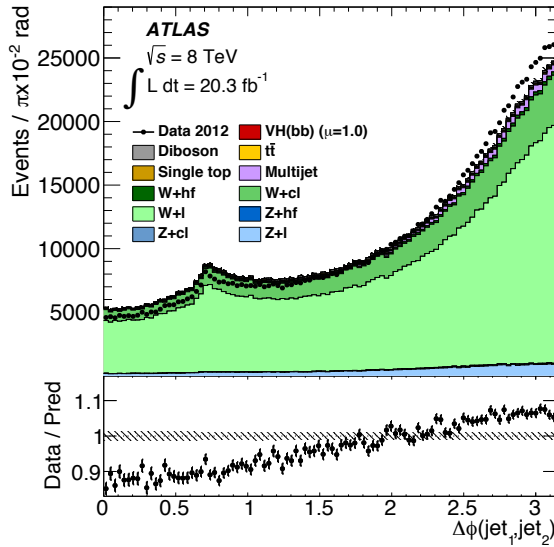


Figure 7.3.: $\Delta\varphi(j_1, j_2)$ distribution in the 1-lepton 2-jet 0-tag region, before the correction. [107].

Given the background composition in this region, a correction is applied to the Wcl and Wl labelled events only, generated with SHERPA. Correcting for this slope improves the data-to-MC agreement on other variables, for example $p_T(W)$, where SHERPA originally predicts a higher rate at high $p_T(W)$ and a lower rate at low $p_T(W)$ than observed in data. This can be expected, given that an energy configuration where the vector boson is produced with high p_T is more likely to produce jets in a similar direction (and opposite to the W 's). Therefore, depleting the high $p_T(V)$ region in the simulation naturally depletes the low $\Delta\varphi(j_1, j_2)$ region, bringing data and MC to better agreement. It is not possible to conclude from data if the correction applied in cl and l events is necessary also in $W+$ h.f. Consequently, no reweighting is applied and an uncertainty is assigned to the shape of the $\Delta\varphi(j_1, j_2)$ distribution, of the size and shape of the correction to Wcl and Wl events.

A similar disagreement in $\Delta\varphi(j_1, j_2)$ is observed in the 0-tag region of the 2-lepton channel, concerning the Zl background process, for which a correction is derived. The $p_T(Z)$ distribution in Zc and Zb simulated events is observed to be mismodelled in the 2-tag region and a reweighting procedure is also applied based on the data-to-MC agreement. Finally, a correction to the generator-level p_T distribution of top-quarks is applied to $t\bar{t}$ events, based on an unfolded measurement described in Reference [111].

7.3.1. $W + b\bar{b}$ background

In contrast to what happens for the $t\bar{t}$ background (and $Z+b\bar{b}$ in the 2-lepton channel), the $W + b\bar{b}$ process has limited available control regions in the $VH(b\bar{b})$ analysis. This poses a limitation on the understanding of this process in the phase-space of interest, since data-driven techniques can't be used. Systematic uncertainties are then derived based on MC studies at particle-level (i.e., before reconstruction to detector-level), since at this stage a larger number of MC generators are available as well as larger samples. The uncertainties on the flavor composition are an exception, for which case the study is performed at detector level. The generators used for the estimation of the uncertainties at particle-level are listed in Table 7.4.

The baseline generator used to simulate the W +jets (and Z +jets) background is SHERPA⁵, where the matrix element is inclusive in the flavor of jets in the final state.

⁵The SHERPA generator version used is 1.4.1.

Table 7.4.: MC samples used in the estimation of $W + b\bar{b}$ modelling uncertainties at particle-level.

MC sample	Matrix Element Process
SHERPA	$W + \text{jets}$ with B hadron filter (“baseline”)
SHERPA	$W + b\bar{b}$
ALPGEN + HERWIG	$W + b\bar{b}$
POWHEG + PYTHIA 8	$W + b\bar{b}$
aMC@NLO + HERWIG++	$W + b\bar{b}$

Events with B or C hadrons are filtered out to specific samples (again, considering only hadrons with $p_T > 5$ GeV), while the remaining are filtered to a light sample. As a consequence, b -jets originate not only from the matrix element but also from the parton shower (e.g., gluon splitting to a pair of b -quarks). The generation combines matrix elements with parton multiplicities of 0, 1, 2 and 3. In contrast, aMC@NLO [112, 113], POWHEG⁶ and ALPGEN generate events with two b -quarks originating explicitly from the matrix element. This is calculated at NLO in QCD in aMC@NLO and POWHEG, and at LO in ALPGEN, merging parton multiplicities of 0, 1, 2 and 3. To establish a common ground for comparison with the other generators, a $W + b\bar{b}$ SHERPA sample is generated where a pair of b -quarks is produced in the matrix element. This sample is used for the estimation of the modelling uncertainties by comparison to the other generators, instead of the baseline B-filtered SHERPA sample.

The rate at which gluon splitting to $b\bar{b}$ occurs, a phenomenon strongly enhanced by collinear factors in QCD, has very few handles in data. Theoretically, different scale choices for the calculation of this rate produce results which can be different by more than a factor 2. Experimentally, the available measurements are statistically limited and indicate that simulation under-predicts this rate, as reported by the CMS Collaboration in Reference [114]. Therefore, a dedicated uncertainty related to the rate of gluon splitting is derived. It is defined by comparing the predictions of the two SHERPA simulations, assessing the impact of different gluon splitting rates on the jet multiplicity and $m_{b\bar{b}}$ shape (specifically, a rate of zero and the one predicted by SHERPA, and considering half of the variation as the uncertainty).

⁶POWHEG is generated with associated event weights, based on the generated W -boson p_T spectrum.

The total uncertainty on the $m_{b\bar{b}}$ shape associated to bb -labelled events has therefore two sources: one related to the shape variation seen from changing the MC program or accuracy of calculation, and one from changing the rate of gluon splitting. The resulting distributions are shown in Figures 7.4 and 7.5, inclusively in $p_T(W)$.

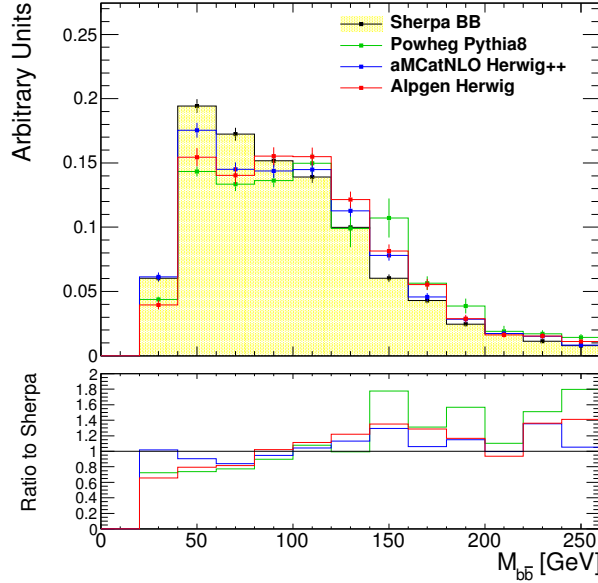


Figure 7.4.: $m_{b\bar{b}}$ distribution in $W + b\bar{b}$ events as predicted by different MC generators and their ratio with respect to SHERPA. Events correspond to the 2-jet bb region (dijet mass selection). Normalized to same area.

The systematic uncertainty on the $m_{b\bar{b}}$ shape is derived from the ALPGEN to SHERPA ratio, characterized by a second order polynomial that describes the low mass discrepancy and the resulting offset at higher masses. The requirement of the two real b -jets is the origin of this disagreement, associated to the discrepancies also seen in the invariant mass distribution of the B hadrons matched to the jets (see Figure B.1).

In total, an effect of $\pm 23\%$ at 50 GeV and $\pm 28\%$ at 200 GeV is obtained with the derived uncertainty and applied to Wb events. The effect of the gluon splitting rate, floated between two arbitrary (and conservative) values, has a similar shape effect on the $m_{b\bar{b}}$ distribution, but with a smaller magnitude.

The impact of using an alternative MC generator on the number of jets and on the $p_T(W)$ distribution is shown in Figures 7.6 and 7.7, respectively. The effect of missing higher order corrections in QCD was estimated by varying the factorization

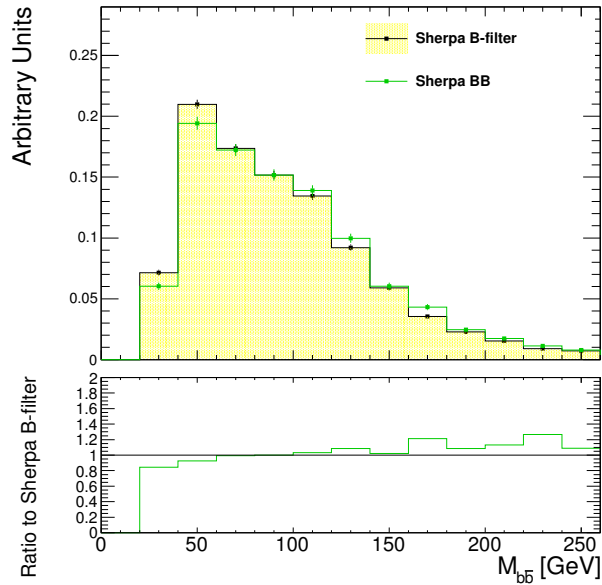


Figure 7.5.: $m_{b\bar{b}}$ distribution in $W + b\bar{b}$ events as predicted by SHERPA simulated events where the B hadrons are originating exclusively from the matrix element (SHERPA BB) and from either matrix element or parton shower (SHERPA B-filter). Events correspond to the 2-jet bb region (dijet mass selection). Normalized to same area.

and renormalization scales⁷ in the aMC@NLO sample (by factors 0.5 and 2) and the PDF set choice (baseline CT10 [115], MSTW2008NLO [116] and NNPDF2.3 [117]). The effect of these variations on the jet multiplicity is negligible (see Figure B.2), but is likely to be an underestimate, given that the scale choice is not propagated to the parton shower. The only sizable effect on the $p_T(W)$ distribution occurs at high p_T , where the scale variations can have an impact of order 10-20%, as shown in Figure B.3.

From the jet multiplicity in each generator the 3-to-2-jet ratio can be extracted for bb -labelled events. The uncertainty on the yield extrapolation between these regions is derived from the relative difference in the ratios observed in each MC with respect to the baseline SHERPA. The discrepant behaviour seen in POWHEG's 3-jet bin is potentially related to a problem in the generation. The origin of the discrepancy in POWHEG is not understood at the time of writing. A 10% uncertainty on the 3-to-2-jet ratio is derived by comparing SHERPA and aMC@NLO, and applied

⁷The central scale value is defined on an event-by-event basis as $\mu_F = \mu_R = (m_W^2 + p_T(W)^2 + m_b^2 + (p_T^2(b) + p_T^2(\bar{b}))/2)^{\frac{1}{2}}$.

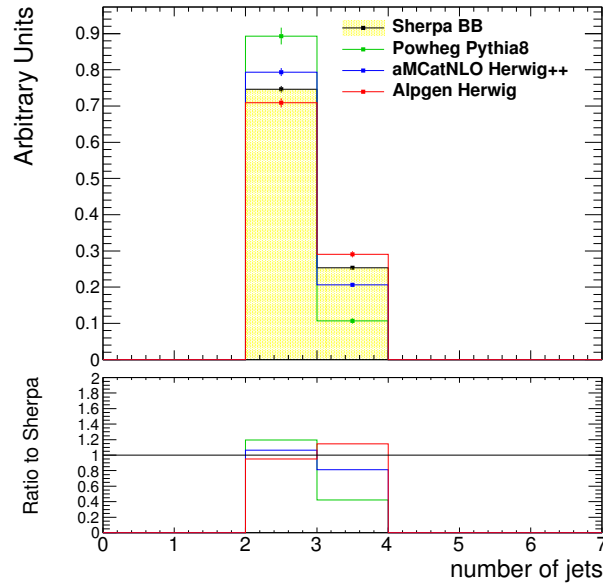


Figure 7.6.: Distribution of jet multiplicity in $W + b\bar{b}$ events as predicted by different MC generators and their ratio with respect to SHERPA. Events correspond to the $b\bar{b}$ region (dijet mass selection). Normalized to same area.

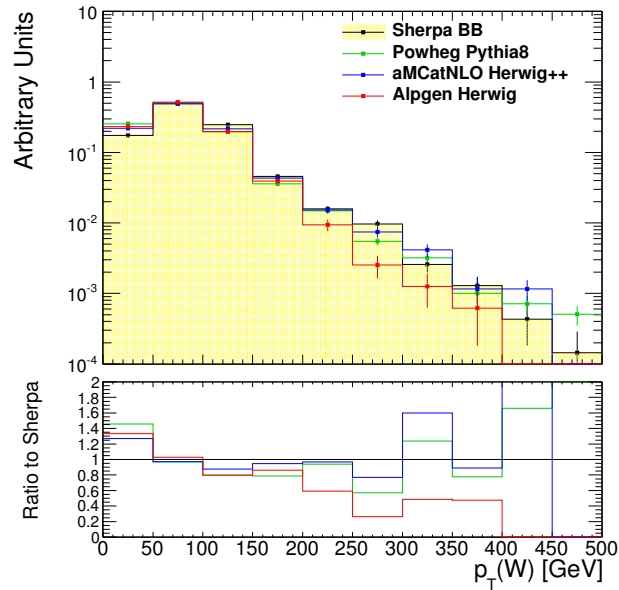


Figure 7.7.: $p_T(W)$ spectrum in $W + b\bar{b}$ events as predicted by different MC generators and their ratio with respect to SHERPA. Events correspond to the 2-jet $b\bar{b}$ region (dijet mass selection). Normalized to same area.

to Wb labelled events. The impact of different gluon splitting rates is found to be negligible (see Figure B.4).

The behaviour of the $p_T(W)$ distribution in 2-jet events is observed to be very different between ALPGEN and SHERPA. A comparison between aMC@NLO generation and an equivalent with no NLO QCD corrections, reveals that the impact on the $p_T(W)$ spectrum is similar to the difference seen between ALPGEN and SHERPA (Figure B.5). Given that higher order corrections are important in high $p_T(W)$ regimes, the aMC@NLO to SHERPA ratio is used to derive the shape uncertainty in this case. This uncertainty amounts to $\pm 9\%$ at $p_T(W) = 50$ GeV and $\pm 23\%$ at $p_T(W) = 200$ GeV.

Within the processes labelled $W + \text{h.f.}$, the separation into bb , bc , bl and cc is heavily dependent on the details of the simulation: the matrix element calculation, the parton shower details and their matching, as well as on the procedure to combine different jet multiplicities. To avoid double-counting of events, SHERPA implements the CKKW scheme, where the progression of the parton shower is truncated to avoid regenerating events already described by the matrix element calculation [118, 119]. ALPGEN implements a similar procedure for light quarks, but lacks a specific treatment of the b and c -quarks, which are generated in dedicated samples and need to be treated to avoid overlap with similar events in the light samples. A dedicated procedure referred to as HFOR (for heavy flavor overlap removal) [120] is applied in ATLAS based on the ΔR between the heavy flavor quarks, to determine if they are likely to have originated from the parton shower evolution or from the matrix element hard emissions, in principle removing double-counted configurations. It is worth noting, however, that there is currently no assessment of the systematic uncertainties related to this procedure. Other differences between ALPGEN and SHERPA implementations are related to the parton shower, the PDF set in the matrix element (CTEQ6L1 [121] and CT10, respectively) and also the treatment of heavy quarks in the calculation: ALPGEN considers b and c -quarks as massless, while SHERPA includes their mass effects.

The relative rates of each flavor component with respect to $b\bar{b}$ are compared between SHERPA and ALPGEN, at detector-level, with the samples listed in Table 7.5. It was observed that ALPGEN produces 20-40% lower rates of events with heavy flavor jets, while a 10% increase in the rate of ll events is observed with respect to SHERPA's prediction. Uncertainties of 35%, 12%, 12% are assigned to the ratios of bl/bb , bc/bb and cc/bb , respectively.

Table 7.5.: MC samples used in the estimation of $W + b\bar{b}$ modelling uncertainties at detector-level. The version generator versions are SHERPA 1.4.1 and ALPGEN.

MC generator	Comments
SHERPA	W +jets with B hadron filter (“baseline”)
SHERPA	W +jets with C hadron filter (“baseline”)
SHERPA	W +jets with B and C hadron vetoes (“baseline”)
ALPGEN + HERWIG	$W + b\bar{b}$
ALPGEN + HERWIG	$W + c\bar{c}$
ALPGEN + HERWIG	$W + c$
ALPGEN + HERWIG	$W + l$

How the systematic uncertainties are propagated between fit regions needs to be carefully defined for each case, considering the sources of uncertainty, the (lack of) knowledge about the validity of the extrapolations being made and the right number of degrees of freedom for the model to be flexible and avoid biases. In the case of the $m_{b\bar{b}}$ shape systematic, independent uncertainties are assigned to $Wbb + Wcc$, Wcl , Wl , $Wbl + Wbc$, according to the different leading order diagrams contributing to each final state. The uncertainty on the $Wbb + Wcc$ $m_{b\bar{b}}$ shape is allowed to float independently between low and high $p_T(W)$ regions. The $p_T(W)$ shape uncertainty is correlated among the Wb processes and allowed to float independently in the 2-jet and 3-jet regions.

The nature of the $\Delta\varphi(j_1, j_2)$ discrepancy was investigated using the same set of simulated samples. Varying the renormalization and factorization scales or the PDF set used in the matrix element revealed that these modifications do not have the power to impact the distribution by the amount seen in the data-to-MC comparison (as is shown in Figure B.6). Comparisons to the other available MC samples, however, show how the different choice of generator affects the $\Delta\varphi$ between the leading jets in $W + b\bar{b}$ events, as can be seen in Figure 7.8. The low $\Delta\varphi(j_1, j_2)$ region of soft-collinear radiation is in principle well described by parton showers, while the wide-angle region should be more sensitive to higher order corrections in the matrix element calculation. The ratio observed between SHERPA and ALPGEN seems to suggest that replacing SHERPA by ALPGEN as baseline for the V +jets processes should be a reasonable approach to improve the general analysis modelling. This was in fact tested, but seen to reproduce the same problem in the 0-tag and 1-tag regions of the 1-lepton

channel, which as mentioned before, are dominated by cl and ll events and not bb . How the behaviour is correlated among final states is subject to many uncertainties, including the effect of the HFOR procedure mentioned before. It has been reported in [57] that the MEPS@NLO version of SHERPA may improve the description in this region. Further studies with the MEPS@NLO SHERPA version are in progress, which should help shed light on the behaviour of the $\Delta\varphi(j_1, j_2)$ (and associated $p_T(W)$) modelling. Among these studies, the $W + \text{jets}$ cross-section measurement by ATLAS [122] has revealed modelling issues in angular distributions in $W + \text{jets}$ events, which are compatible to the ones observed in this analysis.

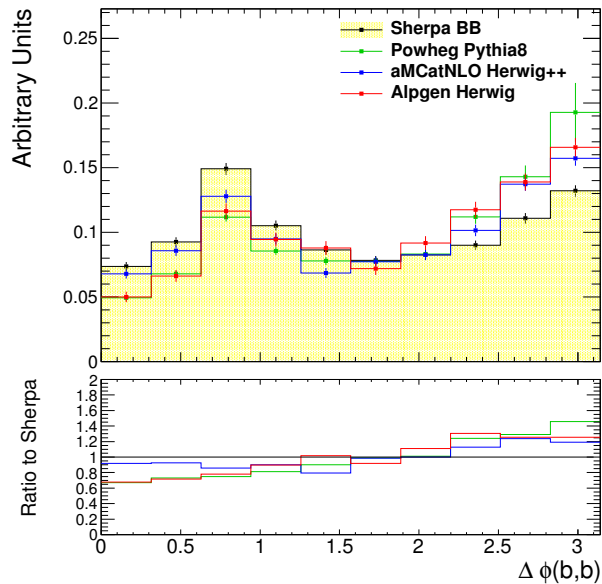


Figure 7.8.: $\Delta\varphi(j_1, j_2)$ distribution in $W + b\bar{b}$ events as predicted by different MC generators and their ratio with respect to SHERPA. Events correspond to the 2-jet bb region (dijet mass selection). Normalized to same area.

The shape uncertainties described in this Section are derived for the dijet mass analysis and found to be enough to cover the variations seen on the BDT output when comparing MC generators. Systematic uncertainties from experimental sources, as well as theoretical uncertainties related to the signal and other background processes are detailed in the main publication, Reference [107].

7.3.2. Summary

Overall, large discrepancies between $W + b\bar{b}$ MC samples are observed in the variables relevant to the $VH(b\bar{b})$ search. The associated set of uncertainties are estimated conservatively, as to properly take into account the current theoretical and experimental limitations associated to the modelling of $W +$ h.f. processes. While many features of the MC generators can be explored and understood through studies like the one presented in this Section, data measurements are essential to truly be able to assess the uncertainties associated to these processes, and should be one of the focus of Standard Model Run 1 and Run 2 data analyses. Measuring inclusive $W + b$ and $W + b\bar{b}$ processes can help improve the performance of the available MC generators, but also the background description in analyses such as the $VH(b\bar{b})$. In particular, a measurement of the gluon splitting rate in ATLAS could be very profitable to a $H \rightarrow b\bar{b}$ search.

7.4. Statistical Analysis

Once the discriminant variables are understood in each event category, a statistical analysis is built to quantify the level of agreement between the measured data and a given model, e.g. the presence of a SM Higgs boson. For this purpose, a likelihood function is defined from the Poisson distribution of data events, expected signal and background yields in each bin of the final discriminant. The variables used as input are the $m_{b\bar{b}}$ distribution or the BDT discriminant in each event category of the dijet mass or the MVA analyses, respectively. Both are complemented by the MV1c distribution of the b -tagged jet in 1 b -tag events, in each $p_T(W)$ interval. The likelihood function is defined as follows:

$$L(\mu, \boldsymbol{\theta}) = \prod_{i=1}^N \frac{(\mu s_i + b_i)^{n_i}}{n_i!} \exp(-\mu s_i - b_i), \quad (7.1)$$

where s_i and b_i are the expected signal and background yields in bin i , and n_i the measurement. The parameter of interest is μ , the signal strength which represents the amount of VH present with respect to the SM expectation. The signal and background expectation in each bin will be dependent on the systematic uncertainties, effect which is described by the addition of nuisance parameters, $\boldsymbol{\theta}$. The prior

knowledge of the constraints on each nuisance parameter is added to the likelihood function as multiplicative Gaussian or log-normal terms. By maximizing the likelihood (or its logarithm), the optimal value for the floating normalizations and the systematic uncertainties is extracted, the latter regulated by their constraints around the estimated central value. The $m_{b\bar{b}}$ variables in the highest sensitivity regions of the dijet mass analysis are shown in Figure 7.9, for each lepton channel, after the global fit.

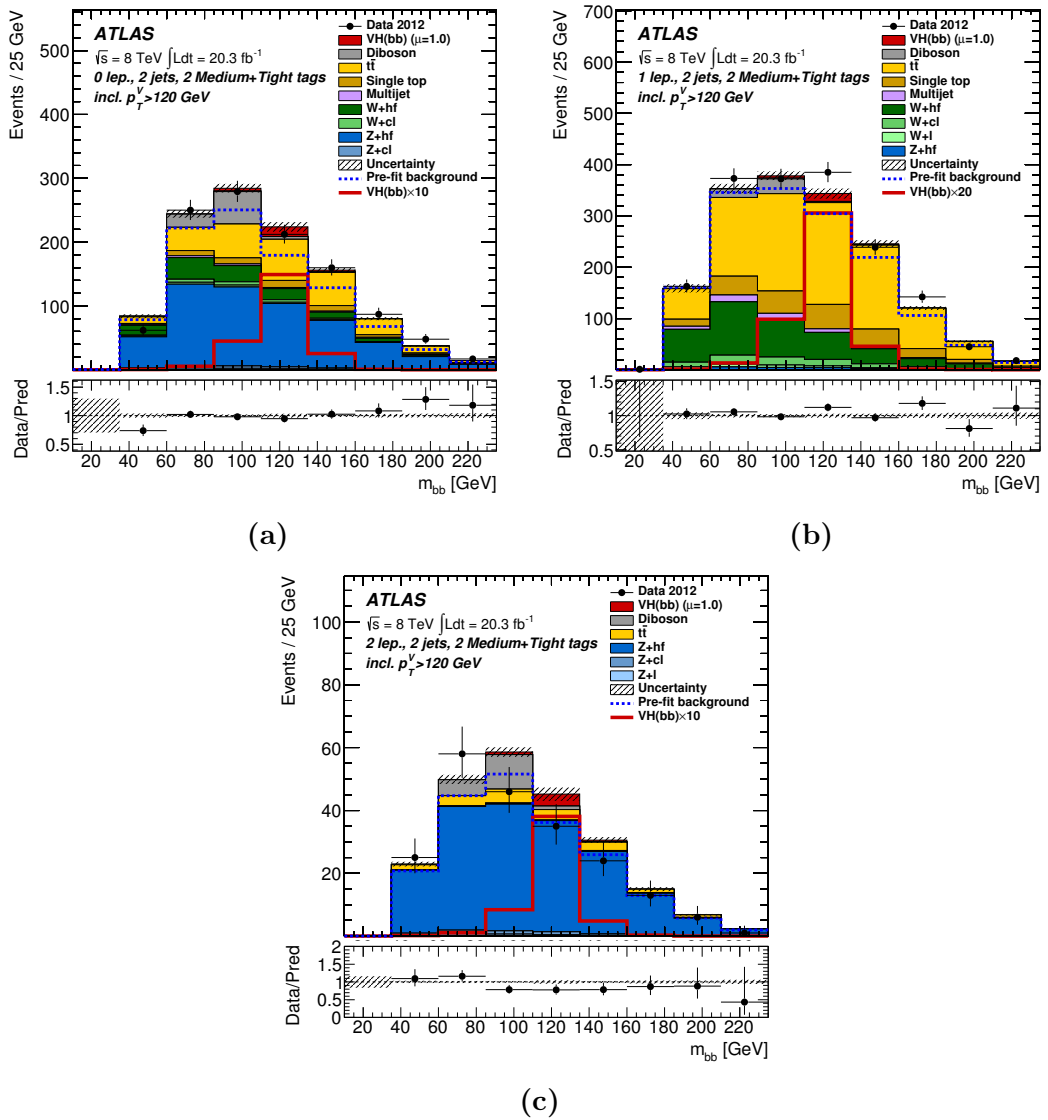


Figure 7.9.: $m_{b\bar{b}}$ distributions comparing data and simulation (after global fit) in the regions with 2 jets and 2 Medium + Tight tags of the (a) 0-lepton, (b) 1-lepton and (c) 2-lepton channels. [107].

A test statistic is defined to measure the compatibility of a hypothesis with the measured data,

$$q_\mu = -2 \ln \frac{L(\mu, \hat{\boldsymbol{\theta}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \quad (7.2)$$

which relates the conditional likelihood $L(\mu, \hat{\boldsymbol{\theta}})$, maximized for a given μ , and the nominal likelihood $L(\hat{\mu}, \hat{\boldsymbol{\theta}})$ maximized with respect to all parameters, with $0 \leq \hat{\mu} \leq \mu$. The test statistic is evaluated for the background only and background plus signal hypotheses, where the distribution of possible outcomes for each is calculated. The p -value associated to the measured test statistic, defined as the probability to obtain a result at least as unlikely under a given hypothesis, is used to calculate the CLs confidence interval. A discovery significance can be defined by measuring the compatibility of the measured test statistic with $\mu = 0$, the null hypothesis.

7.5. Results

An excess is observed in the 8 TeV dataset with respect to the background-only hypothesis. The 95% CL upper limit on the VH production cross section is shown in Figure 7.10, for the combination of the 7 TeV and 8 TeV datasets (the latter using the BDT approach). At the mass point of $m_H = 125$ GeV, the observed limit is $1.2 \times \text{SM}$, while the expected (in the absence of signal) is 0.8 times the SM value. The probability of this result or another more signal-like to be obtained without the presence of signal is 8%, which corresponds to an observed excess with a significance of 1.4σ , whereas the expected significance is 2.6σ . The value of μ is fitted for the $m_H = 125$ GeV mass point, combining both datasets and lepton channels, using the MVA approach:

$$\mu = 0.51 \pm 0.31(\text{stat.}) \pm 0.24(\text{syst.}). \quad (7.3)$$

By analyzing the 8 TeV and 7 TeV datasets separately, the following values of the signal strength are obtained with the MVA approach: $\mu = 0.65 \pm 0.32(\text{stat.}) \pm 0.26(\text{syst.})$ and $\mu = -1.6 \pm 1.2(\text{stat.}) \pm 0.9(\text{syst.})$, respectively. The uncertainty on the obtained μ is dominated by systematic uncertainties such as the normalization of $W + b\bar{b}$ and $t\bar{t}$, the Wcc and Wbb $m_{b\bar{b}}$ shapes and the ratio of Wbl to Wbb at high $p_T(W)$.

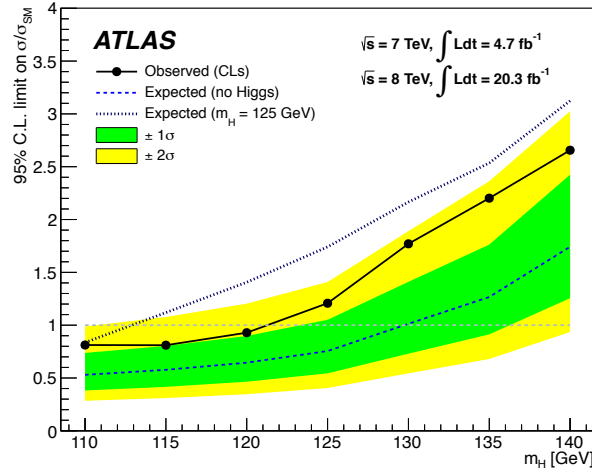


Figure 7.10.: Observed and expected 95% CL upper limits on the VH production cross section with respect to the SM value, for the combination of lepton channels and datasets. [107]

The analysis of the 8 TeV dataset with the dijet mass approach results in a measurement of the signal strength of $\mu = 1.23 \pm 0.44(\text{stat.}) \pm 0.41(\text{syst.})$. The statistical compatibility of this result with the value obtained with the MVA approach is of approximately 8%, with a correlation factor of 0.67.

The procedure is validated by measuring with the same procedure the yield of $VZ, Z \rightarrow b\bar{b}$ production. This process has the same final state as $VH(bb)$ and a $m_{b\bar{b}}$ spectrum peaking not far from the 125 GeV mass Higgs boson, therefore providing a very good testing ground for the fit model. The observed VZ signal strength is found to be consistent with the SM expected value: $\mu_{VZ} = 0.74 \pm 0.09(\text{stat.}) \pm 0.14(\text{syst.})$, as obtained with the combined 7 and 8 TeV dataset (the latter performed with the MVA approach). The VZ process is therefore observed with a significance of 4.9σ , with an expected significance of 6.3σ .

While the result is encouraging, the $H \rightarrow b\bar{b}$ search remains inconclusive after the LHC Run 1 and is only expected to be measurable with Run 2 energies and luminosities.

Chapter 8.

Prospects for 14 TeV and higher centre-of-mass energies

“I think music needs to be of its time and speak to that time.”

— Dave Holland

Preparations for Run 2 are in progress in ATLAS and new performance challenges are expected. In the meantime, the physics potential beyond LHC energies, which will ultimately depend on what is learned from Run 2, is under study. One of the planned upgrades to the LHC is the High Luminosity LHC (HL-LHC), where the discovery potential is extended by increasing the luminosity by a factor 10. Another post-LHC scenario that is being developed is the High Energy LHC (HE-LHC) [123], which involves introducing stronger superconducting magnets in the accelerator, increasing the achievable centre-of-mass energy of pp collisions to 33 TeV. Alternatively, the technology for a very High Energy LHC (VHE-LHC), which requires a new longer tunnel, is also under investigation [124], with collision energies of the order of 100 TeV, a powerful probe for Beyond the Standard Model (BSM) phenomena.

This Chapter presents a particle-level study of boosted WH production with a Higgs boson decaying to a pair of b -quarks, and the sensitivity of this channel with two different selection approaches, exploring centre-of-mass energies of $\sqrt{s} = 8, 14, 33$ and 100 TeV. The boosted phase-space is defined as the region where a W -boson decaying to leptons is reconstructed with a transverse momentum greater than 200 GeV.

8.1. Introduction

The TeV scale can potentially hold a new spectra of particles or phenomena that can help shed light on the Higgs sector and on open questions beyond the SM, such as the nature of dark matter or the asymmetry of matter and antimatter. The physics program of the LHC, its planned upgrades and future accelerators can probe this scale and perform BSM searches and precision measurements.

The b -quark and vector-boson couplings to the Higgs boson are examples of measurements that can be performed with higher energies and luminosities. However, the growth of the Higgs production cross-sections, although dramatic, is accompanied by a similar or even greater growth of background rates, as shown in Figure 8.1. As the cross-sections increase with \sqrt{s} , reflecting the evolution in the partonic luminosities, the gluon-initiated processes such as top-pair production are heavily enhanced. Consequently, in a HE-LHC scenario, the $t\bar{t}H$ cross-section exceeds the VH production rates.

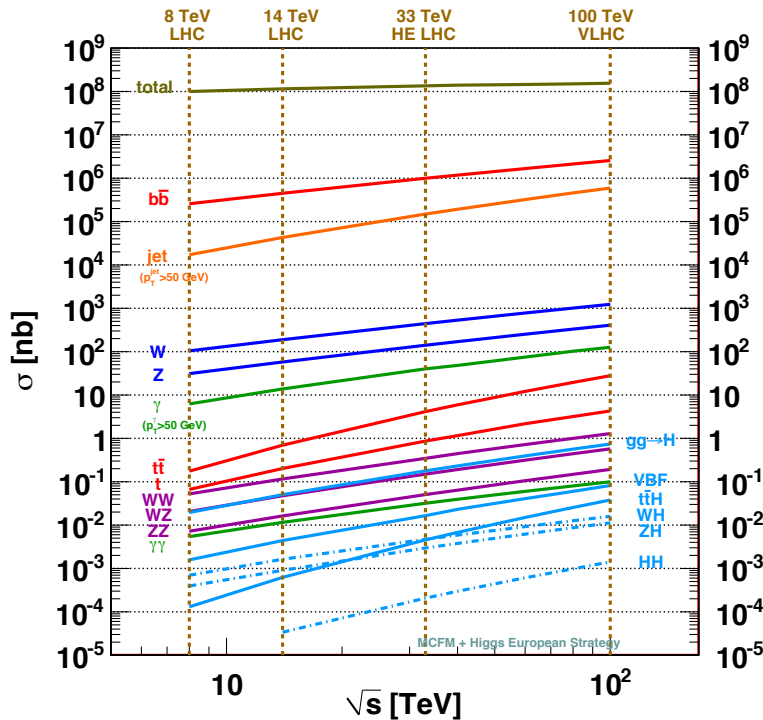


Figure 8.1.: Cross-sections of SM and Higgs production processes as predicted as a function of the centre-of-mass energy for a pp collider. [125]

The collision environment changes significantly as the centre-of-mass energy increases from 8 to 100 TeV. The phase-space for the transferred energy Q increases, while for a given scale, lower regions of proton momentum fractions x are accessible. Furthermore, the EW coupling constants increase for higher Q^2 , while α_S only slightly decreases. Additionally, the phase-space for extra radiation originating from a hard scatter is also larger, resulting in significant rates of gluon, photon and W/Z radiation in each event, generally translating into higher jet multiplicities (for a given p_T threshold).

Pile-up conditions are expected to become more extreme in a HL-LHC scenario and double-parton scattering also becomes considerably more important, i.e., the probability for two hard scattering events to occur in the same pp collision is no longer negligible. The study of these features is outside the scope of this work, and the event generation described in the following Section does not include such phenomena. However, it has been shown in ATLAS that jet grooming and jet area subtraction techniques are able to fully recover hadronic W -boson and top-quark mass peaks in events with an average number of interactions per bunch crossing of up to 200. Track-based strategies are also under development, which are valid for both small- R and large- R jets [68].

8.2. Event Generation

A particle-level study of boosted $WH(b\bar{b})$ production is performed, considering semi-leptonic $t\bar{t}$ and $W + b\bar{b}$ processes as backgrounds. The calculation of the matrix elements is performed with `amc@NLO`, including NLO corrections in QCD. The description of the processes is improved by matching the NLO calculation with a parton shower program, in this case `HERWIG++`, which also includes models for the underlying event and hadronization. The renormalization and factorization scales are dynamically defined as the sum of the transverse masses of all final state particles and partons. Other relevant parameters and physical quantities used in the calculation are listed in Table 8.1. The decays of the top-quark, W and H -bosons are simulated using `MadSpin` [126], considering the $t \rightarrow Wb$, $W \rightarrow \mu\nu$ and $H \rightarrow b\bar{b}$ decay modes, with branching ratios of 1.0, 0.11 and 0.58, respectively.

Table 8.1.: Parameters used in the generation of $WH(b\bar{b})$, $t\bar{t}$ and $W + b\bar{b}$ processes with aMC@NLO.

Parameter	Value
Z -boson mass	91.19 GeV
G_F	$1.166 \times 10^{-5} \text{ GeV}^{-2}$
$\alpha_S(M_Z)$	0.118
$\alpha_{EW}(M_Z)$	1/132.5
H -boson mass	125 GeV
b -quark mass	4.70 GeV
t -quark mass	174.30 GeV

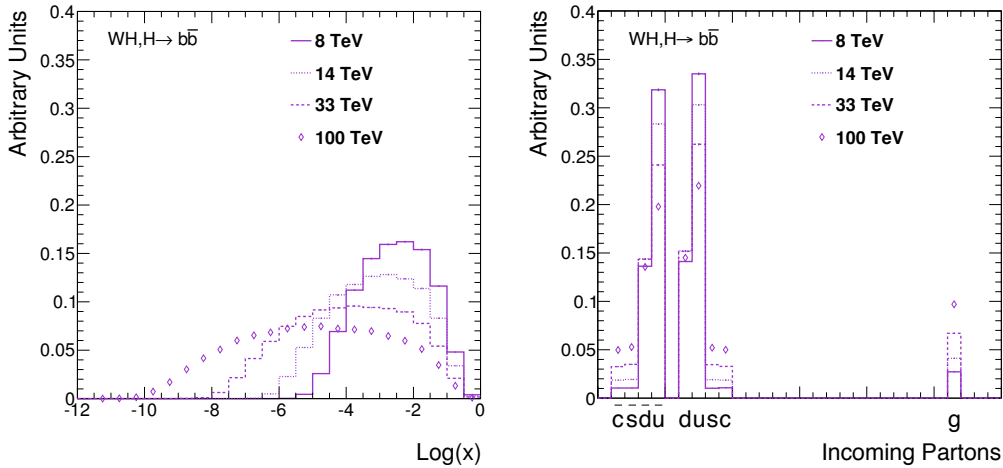


Figure 8.2.: Distribution of the logarithm of the proton momentum fraction x carried by the incoming partons (left) and their relative rate (right), in WH events, at different centre-of-mass energies. The transverse momentum of the reconstructed W is required to be larger than 200 GeV and no other selection cuts are applied. Normalized to same area.

The CT10 PDF set is used in the NLO calculation and in the parton shower, in the 4FNS. The distribution of the proton momentum fractions carried by the incoming partons and their relative rates are shown in Figure 8.2, in WH events. Higher centre-of-mass energies probe smaller x regions, where gluons and sea quarks dominate, and which are currently subject to large PDF uncertainties.

To increase the statistical power of the samples in the boosted region of phase-space, generation filters are implemented prior to the parton shower: the base-

line signal and $W + b\bar{b}$ samples are produced by requiring that the transverse momentum of the W is greater than 100 GeV, while on $t\bar{t}$ events the leptonically-decaying top (or anti-top) quark is required to have a p_T greater than 100 GeV. The highest boost region of background events is again improved by combining samples with stricter generation cuts: $p_T(W) > 400$ GeV for the $W + b\bar{b}$ process and $p_T(\text{top or anti-top-quark}) > 300$ GeV for $t\bar{t}$. Further cuts are applied at generation level when calculating the NLO matrix element: a minimum p_T cut of 7 GeV is applied to jets, clustered with the anti- k_t algorithm with $R = 0.4$, and limited to the pseudo-rapidity range of $|\eta| < 5.0$.

The cross-section values calculated with `amc@NLO` for the signal and background processes are listed in Table 8.2, inclusive in W -boson or top-quark transverse momentum. The cut of $p_T > 100$ GeV on the generated W selects 28% to 35% of the signal phase-space, and 3 to 9% of the $W + b\bar{b}$ cross-section, increasing with \sqrt{s} . The baseline filter on $t\bar{t}$ events selects 50 to 60% of its total cross-section.

The total rate of $t\bar{t}$ events is corrected from NLO to NNLO in QCD by scaling the cross-section by a factor of 1.25, according to the values quoted in Chapter 2. This correction assumes a uniform enhancement of the cross-section as a function of the top-quark p_T , and is therefore a conservative estimation of the expected behaviour¹.

The distributions in Figure 8.3 compare the evolution of $t\bar{t}$ and $W + b\bar{b}$ cross-sections to the signal cross-section, as a function of the reconstructed W -boson p_T (as defined in the next Section). The fact that the $p_T(W)$ spectrum falls less rapidly for signal than for background processes can be observed, especially in the case of $t\bar{t}$ events. However, it can also be seen that the signal-to-background gain in moving to higher boosts decreases as the centre-of-mass energy increases. A possible explanation is that a harder $p_T(W)$ spectrum is observed in background events as the centre-of-mass energy increases, due to the increase in the relative contribution of gluon-initiated diagrams.

¹ATLAS measurements have found that the top-quark is predicted by several simulation programs to have a p_T spectrum harder than that found in data. [127]

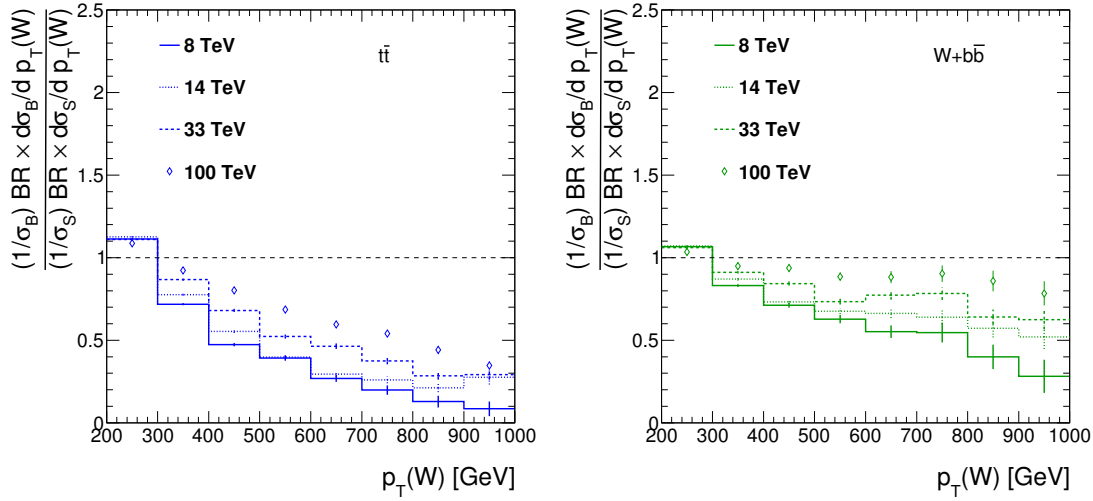


Figure 8.3.: Differential cross-section times branching ratio for $t\bar{t}$ (left) and $W + b\bar{b}$ events (right) with respect to the signal $WH(b\bar{b})$, as a function of $p_T(W)$, normalized to same area and at different centre-of-mass energies. The transverse momentum of the reconstructed W is required to be larger than 200 GeV and no other selection cuts are applied.

Table 8.2.: Cross-section values calculated with `amc@NLO`, at different centre-of-mass energies, including QCD corrections at NLO. Numerical uncertainties of less than 1% are associated to these numbers.

\sqrt{s} / TeV	Cross-Section / pb		
	WH	$t\bar{t}$	$W + b\bar{b}$
8	0.72	188	444
14	1.55	744	914
33	4.45	4.3×10^3	2.7×10^3
100	15.23	2.8×10^4	9.1×10^3

8.3. Event Selection

A W candidate is found by identifying a muon with $p_T > 20$ GeV and $|\eta| < 3.0^2$, as well as missing transverse energy greater than 20 GeV. Only events where the transverse momentum of the W is greater than 200 GeV are considered. Two jet algorithms are included in this study: anti- k_t $R = 0.4$ and Cambridge/Aachen

²Extended to < 4.0 for centre-of-mass energies of 33 and 100 TeV, as explained in the following Section.

$R = 1.2$ split and filtered jets. Additionally, a modified version of the split and filtered procedure is tested on Cambridge/Aachen jets, where the mass-drop condition is not imposed.

The simulated signal and background events are subject to two sets of event cuts, along the lines described throughout this thesis. In the resolved approach, the following requirements are initially applied:

- At least two anti- k_t $R = 0.4$ jets with $p_T > 20$ GeV;
- The absolute pseudo-rapidity of the jets should be less than 3.0 (4.0);
- The ΔR distance between the two leading anti- k_t jets is limited to be less than 1.4;
- Each of the two leading jets is matched to a B hadron.

The substructure-based approach is studied by applying the following requirements:

- At least one Cambridge/Aachen split and filtered jet with $p_T > 180$ GeV;
- The absolute pseudo-rapidity of the Cambridge/Aachen split and filtered jet should be less than 3.0 (4.0);
- The two subjets with highest p_T in the leading Cambridge/Aachen split and filtered jet are matched to a B hadron each.

The jet selection is performed using different pseudo-rapidity ranges according to the centre-of-mass energy: $|\eta| < 3.0$ is considered at 8 and 14 TeV, while the interval $|\eta| < 4.0$ is considered at 33 and 100 TeV, a choice which is explained in the next Section.

The geometrical matching of jets or subjets to B hadrons identified in the event is performed with a ΔR condition on their overlap, chosen to be less than 0.4 or 0.3, respectively, i.e., the size of the jet objects themselves. If more than one hadron overlaps, the closest one is chosen, and the matching proceeds with the remaining hadrons. Only hadrons with $p_T > 5$ GeV are considered. If a jet or subjet is not matched to a B hadron, an additional check is performed with charm hadrons, so as to simulate the large charm-quark mis-tag rate present in experimental techniques to b -tag jets. If both matching conditions fail, the jet is labelled as light.

To fully take into account the possible background contaminations in this search, an important experimental aspect to be considered is the b -tagging efficiency and corresponding mis-tag rates. Therefore, events where a $b\bar{b}$ pair is found are weighted by a typical double b -tagging efficiency of $(0.7)^2$, whereas bc events are weighted by 0.14, where a mis-tag efficiency of 20% is assumed. Finally, bl events are also taken into account, with a much smaller weight of 0.007, by considering a light mis-tag rate of 1%. Furthermore, although the requirement of two b -tagged jets reduces most of the W +jets background to $W + b\bar{b}$ events, it was observed in Chapter 7 that a contribution from $W + c\bar{c}$ events is not negligible. Based on the yields obtained in the ATLAS result of [108], the $W + b\bar{b}$ process is scaled by a factor of 1.25 to approximately consider a $W + c\bar{c}$ contamination.

8.4. Signal Efficiency

The evolution of the signal efficiency with the resolved and substructure methods as a function of $p_T(W)$ is shown in Figure 8.4, considering the selection cuts described in the previous Section and (di)jet invariant masses inside a window $90 < m_H < 140$ GeV³. Additionally, efficiencies for events uniquely reconstructed by each of the approaches are shown in dashed lines. A drop in the resolved efficiency is observed when $p_T(W)$ exceeds 500-600 GeV, reflecting the increasing probability for the $b\bar{b}$ pair to be emitted with an angular separation smaller than 0.4, failing to be reconstructed by two anti- k_t $R = 0.4$ jets. On the other hand, approximately 10% of events are missed by the substructure approach throughout the entire $p_T(W)$ range, mostly due to the momentum balance condition of the splitting algorithm or the p_T requirement on the large- R jet.

Limiting the pseudo-rapidity of the objects to the range $|\eta| < 3.0$ has a strong impact in the signal efficiency at 33 and 100 TeV. In fact, as partons with smaller values of x are probed, larger asymmetries in the momentum fractions carried by the initial-state partons are more likely to occur, which effectively boost the system to higher values of $|\eta|$. Such an effect can be clearly observed in Figure 8.5, where the fraction of WH events with a muon in the “forward” region ($|\eta| > 3.0$) is 1,4,15,30%

³The mass window is chosen based on the shape of the (di)jet invariant mass distribution in signal events, see Figure C.1. It selects almost 90% of signal events in the resolved case, 75-80% in the substructure case.

for $\sqrt{s} = 8, 14, 33, 100$ TeV, respectively. It can be concluded from these numbers that tracking in the forward region is an important aspect to take into consideration when designing detectors for future higher energy colliders. Throughout this study, the pseudo-rapidity cuts applied on muons and jets are therefore extended to ± 4.0 , for $\sqrt{s} = 33, 100$ TeV.

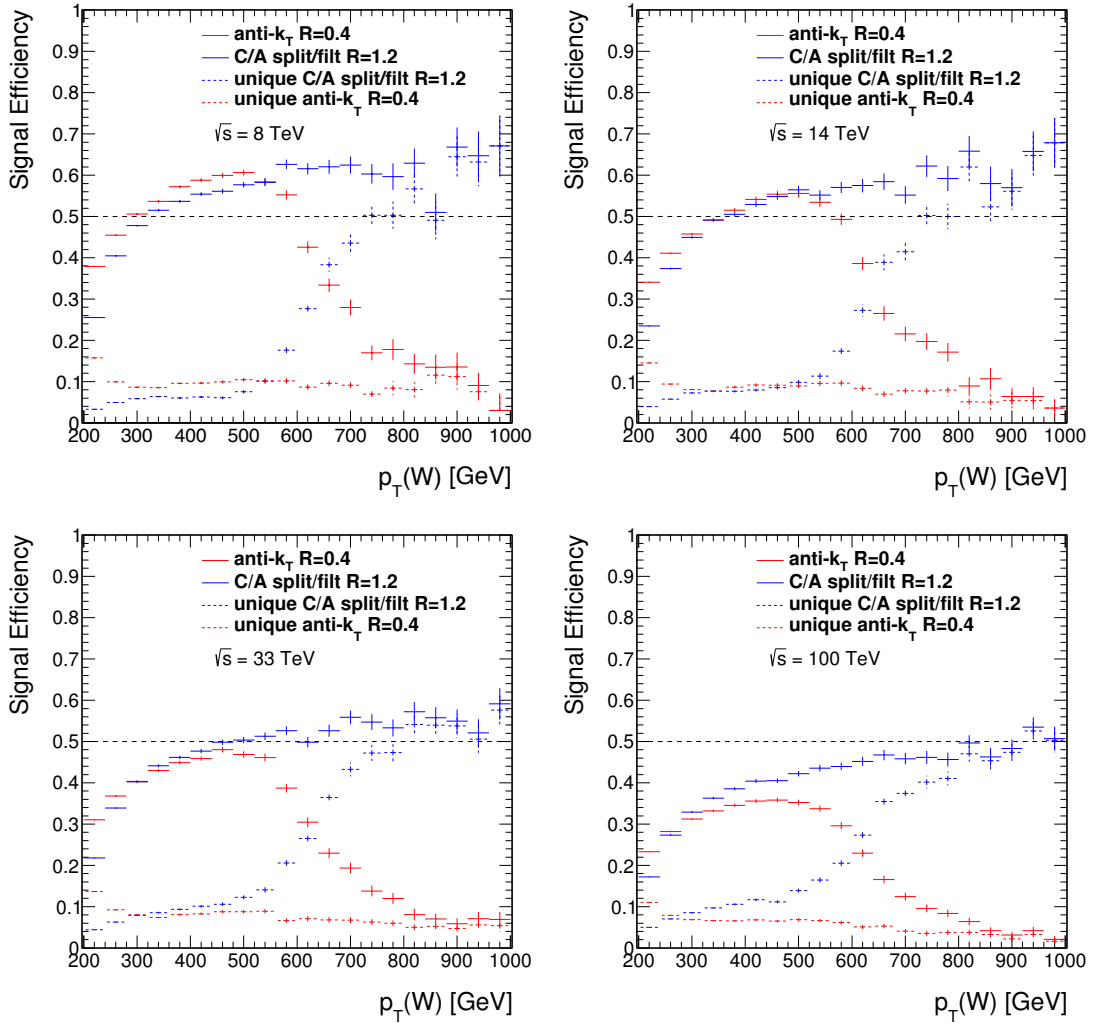


Figure 8.4.: Signal efficiency with resolved and substructure selection approaches, at $\sqrt{s} = 8, 14, 33, 100$ TeV. The cuts outlined in Section 8.3 are applied. Baseline events are defined by requiring $p_T(W) > 200$ GeV and a lepton pseudo-rapidity within ± 5.0 .

Complementary cumulative distributions (CCDF) can be defined to quantify the efficiency of each selection method in a range of $p_T(W)$ greater than a given value. The CCDF's are shown in Figure 8.6 for events uniquely selected by the

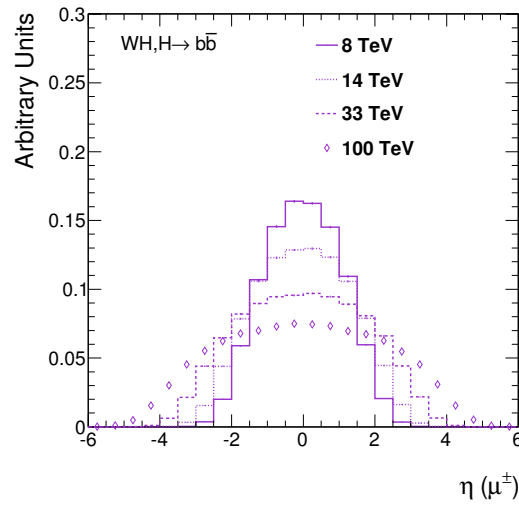


Figure 8.5.: Pseudo-rapidity of the muon in $WH(b\bar{b})$ events, at $\sqrt{s} = 8, 14, 33, 100$ TeV. The transverse momentum of the reconstructed W is required to be larger than 200 GeV and no other selection cuts are applied. Normalized to same area.

substructure approach and for all resolved events. It can be concluded from these distributions that approximately 10% of signal events are being missed by the resolved approach at 8 TeV, in the region where $p_T(W) > 200$ GeV. This number increases to approximately 30% at 100 TeV. The contribution of the substructure analysis to the signal efficiency steeply increases in the region $500 < p_T(W) < 700$ GeV, where the resolved approach begins to fail.

Finally, the obtained WH differential cross-sections with respect to the W transverse momentum, times branching ratio and selection efficiency, are shown in Figure 8.7, for each of the selection approaches.

8.5. Physics at 14 TeV and higher energies

The performance of each selection method in terms of background rejection, and the evolution of signal-to-background ratios and significance with $p_T(W)$ are important figures-of-merit to conclude on the feasibility of the $VH(b\bar{b})$ channel and on the usefulness of substructure techniques. Consequently, limitations on the number of jets in an event are imposed in order to reduce contamination originating from $t\bar{t}$ and $W + b\bar{b}$ processes.

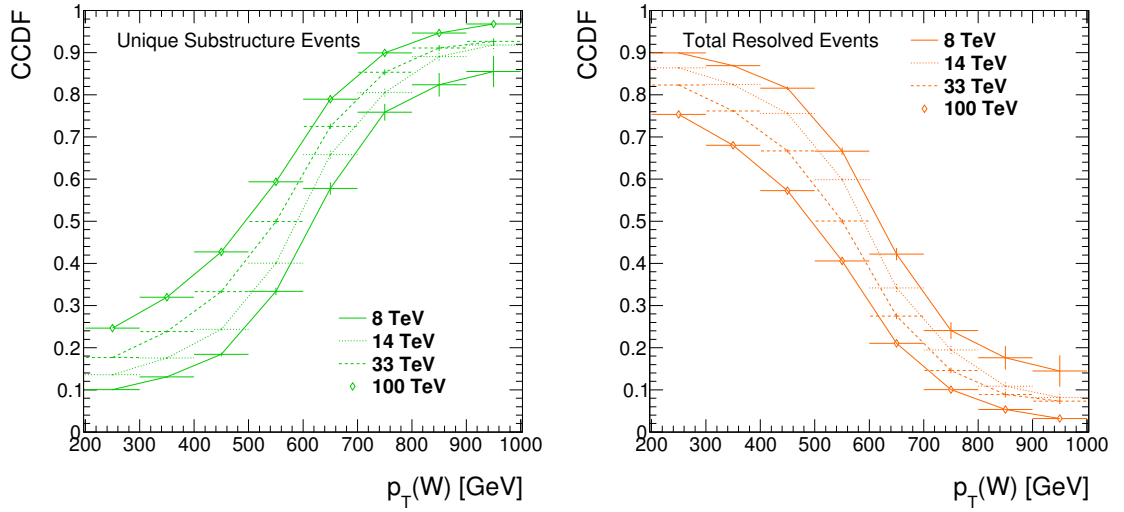


Figure 8.6.: Complementary cumulative distribution (CCDF) of signal events uniquely selected by the substructure-based approach (left) and by the resolved approach (right), with respect to all events selected by both techniques, as a function of $p_T(W)$ and at different centre-of-mass energies. The cuts outlined in Section 8.3 are applied.

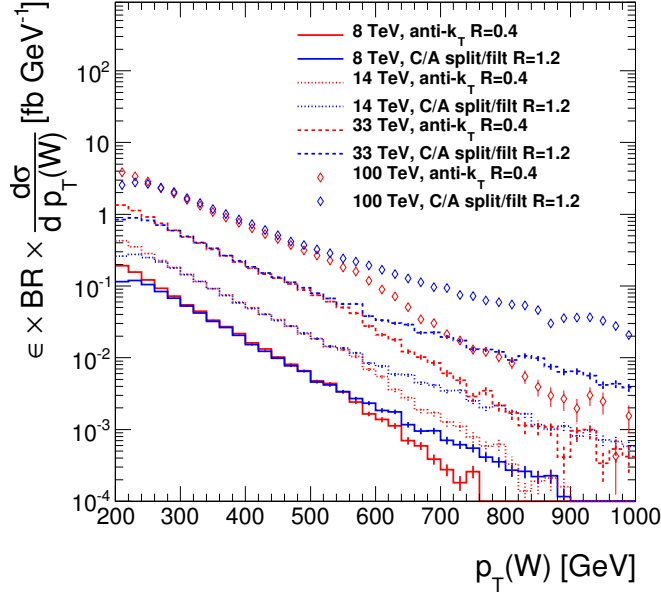


Figure 8.7.: $W(\mu\nu)H(b\bar{b})$ differential cross-section with respect to $p_T(W)$, times branching ratio and selection efficiency, using the resolved and substructure approaches, at $\sqrt{s} = 8, 14, 33, 100$ TeV.

The distribution of the number of anti- k_t $R = 0.4$ and Cambridge/Aachen $R = 1.2$ split and filtered jets in signal and background processes and their evolution as a

function of \sqrt{s} is shown in Figures 8.8 and 8.9. The average number of jets in an event shifts to higher values as the total energy increases, with high- p_T groomed large- R jets showing higher robustness with respect to \sqrt{s} .

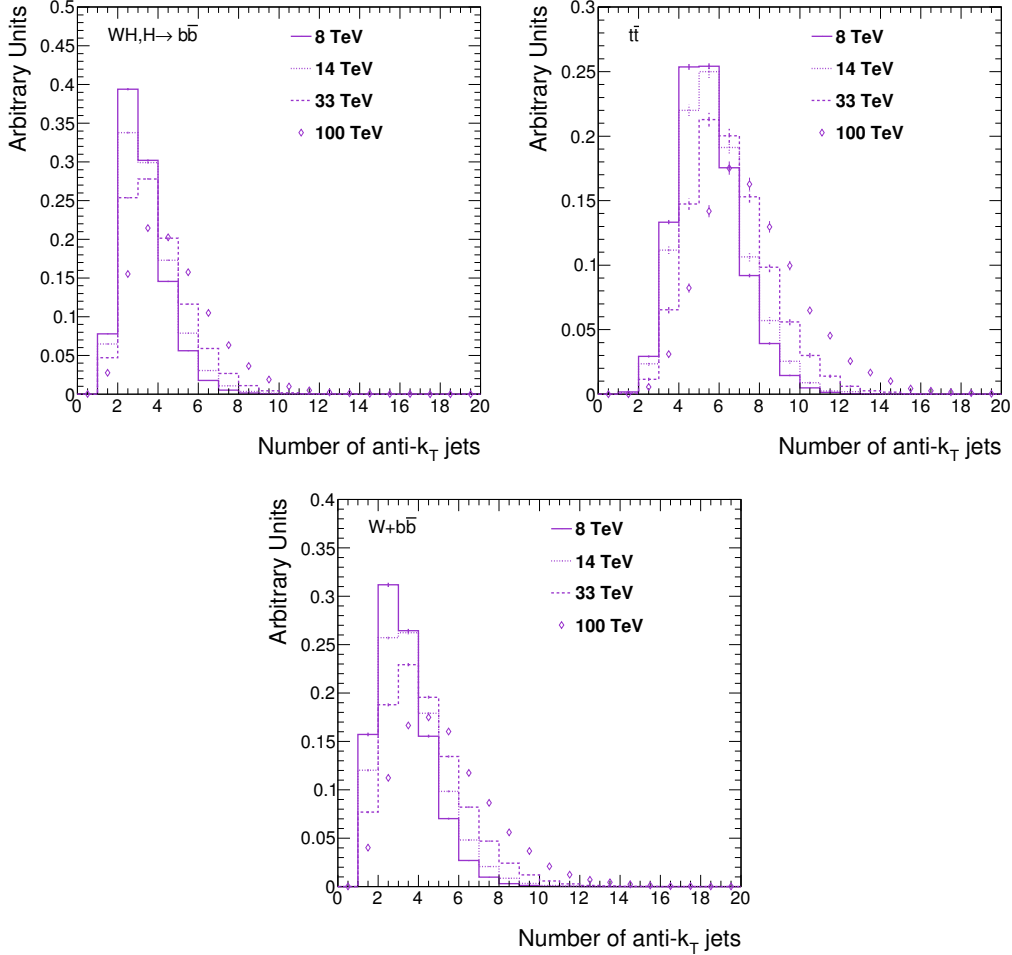


Figure 8.8.: Number of anti- k_t $R = 0.4$ jets with $p_T > 20$ GeV and $|\eta| < 5.0$ in signal and background events, at $\sqrt{s} = 8, 14, 33, 100$ TeV. The transverse momentum of the reconstructed W is required to be larger than 200 GeV and no other selection cuts are applied. Normalized to same area.

A veto on the number of anti- k_t jets in different pseudo-rapidity regions is defined to target mainly the top background. The optimization of the veto as a function of S/\sqrt{B} for both reconstruction approaches is described in Appendix C.2, from where the following configuration is defined:

- A maximum of three anti- k_t jets with $p_T > 20$ GeV and $|\eta| < 5.0$ are allowed in an event.

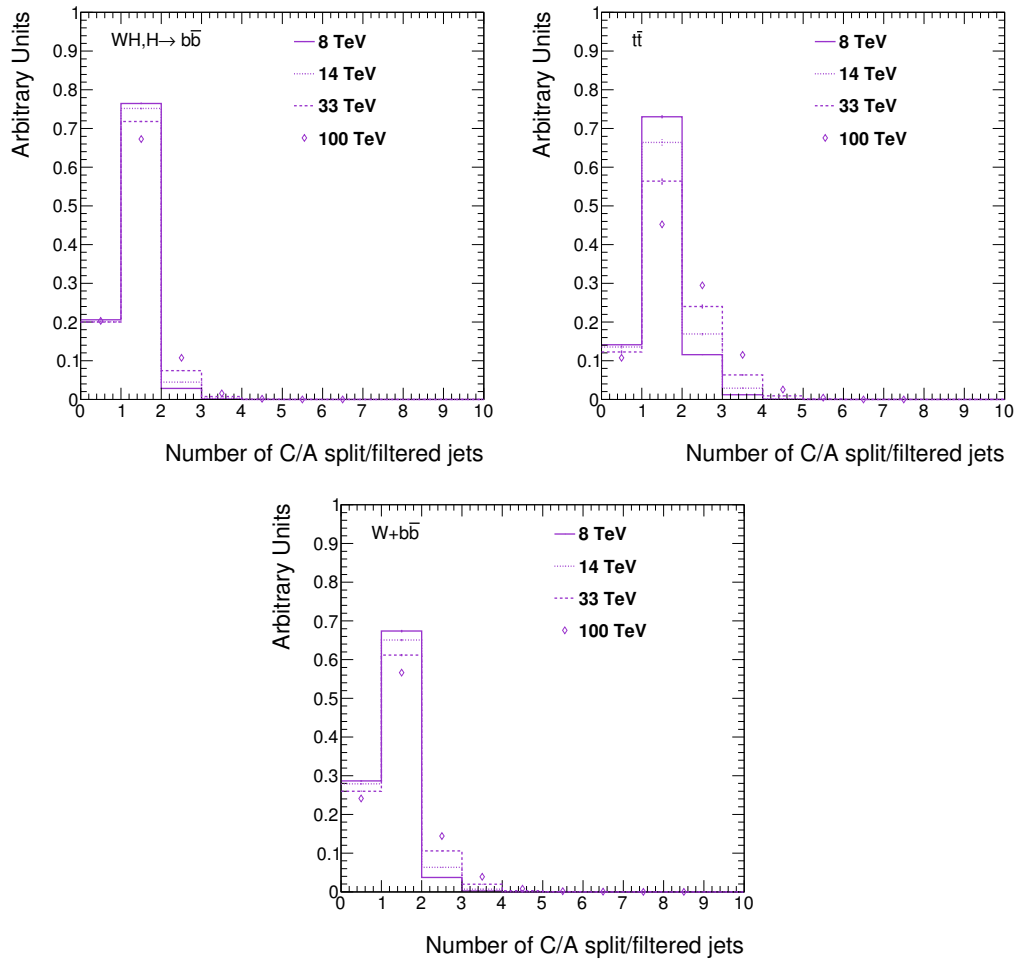


Figure 8.9.: Number of Cambridge/Aachen $R = 1.2$ split and filtered jets with $p_T > 150$ GeV and $|\eta| < 5.0$ in signal and background events, at $\sqrt{s} = 8, 14, 33, 100$ TeV. The transverse momentum of the reconstructed W is required to be larger than 200 GeV and no other selection cuts are applied. Normalized to same area.

- The sub-subleading anti- k_t jet is required to be “forward” (in the region $|\eta| > 3.0(4.0)$) or have low transverse momentum (defined as less than 10% of $p_T(W)$).

The chosen jet veto rejects roughly 30-40% of signal events in the Higgs boson mass window at 8 TeV and 60% at 100 TeV, with both reconstruction approaches. It is however extremely effective in reducing the $t\bar{t}$ contamination in the mass window down to 8-10% at 8 TeV and 2-4% at 100 TeV. The efficiency for $W + b\bar{b}$ events is different for each method, ranging from 50 to 60% at 8 TeV and 20 to 30% at 100 TeV,

with the substructure approach providing the best rejection. When tackling the $W + b\bar{b}$ background, the rejection obtained by the substructure approach originates mainly from the mass window cut, given that the events that do survive the selection typically have a Cambridge/Aachen jet with a low invariant mass.

The resulting (di)jet invariant mass distributions are shown in Figures 8.10 and 8.11 for the resolved and substructure approaches, respectively. The flavor composition in $W + b\bar{b}$ events has a small contribution from bl and bc , adding up to a total of the order of 4% after the substructure selection and approximately 10% after the resolved approach. On the other hand, the $t\bar{t}$ contamination originates heavily from bc -labelled events: the resolved method selects approximately 60% of bc events and 30% of bb , whereas the substructure approach selects 30% and 60%, respectively.

The contribution of each flavor component in $t\bar{t}$ events to the resulting mass distribution is shown in Figure 8.12, at $\sqrt{s} = 14$ TeV. The total $t\bar{t}$ background is problematically peaking in the same region as the signal, particularly in the resolved case. Although this is true also for the substructure selection, in this case this peak originates mainly from bc -labelled events, whereas the bb contribution has a distinct shape which extends to the low mass region. In fact, the shape of the total $t\bar{t}$ background in the substructure case is very susceptible to the exact veto definition, given the different event configurations that originate bb and bc -labelled events. It is clear from these distributions that the improvement in b -tagging techniques and their level of charm-quark rejection are absolutely vital to reduce the $t\bar{t}$ contribution in the mass window of the Higgs boson.

The evolution of the invariant mass distributions with $p_T(W)$ can be seen in Figures 8.13 and 8.14, at $\sqrt{s} = 14$ TeV. Despite the limited statistics, it is observed that in the resolved analysis, the shapes of the background processes are kinematic in origin, and heavily dependent on the boost of the system. A more precise determination of the invariant mass shapes of the background processes at such high boosts and small selection efficiencies would require considerable computation time.

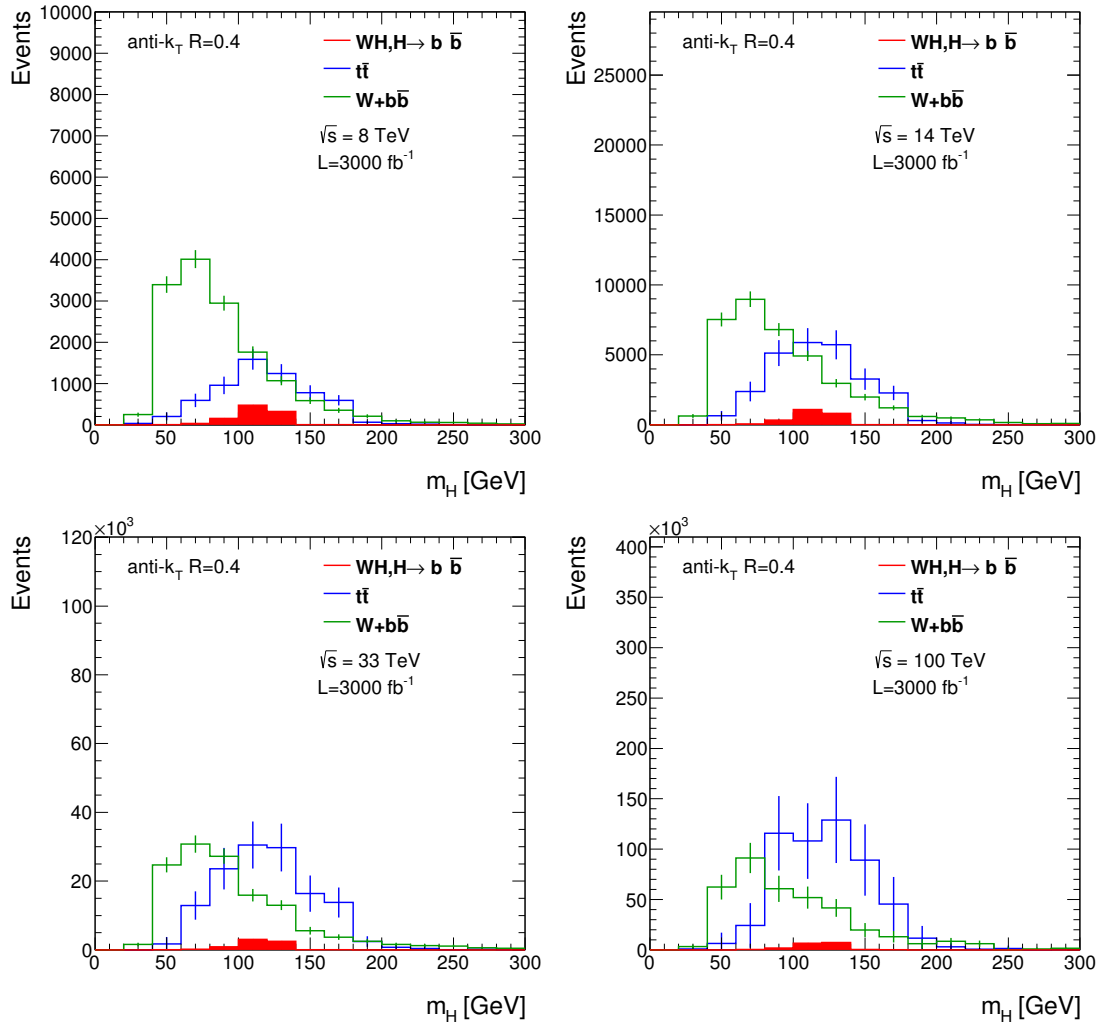


Figure 8.10.: Dijet invariant mass after the resolved selection and jet veto, for signal, $t\bar{t}$ and $W + b\bar{b}$ events in the muon channel at $\sqrt{s} = 8, 14, 33, 100 \text{ TeV}$.

8.6. Results

An estimation of the signal sensitivity in each centre-of-mass energy is presented in this Section, assuming different integrated luminosity scenarios, and comparing the resolved and substructure approaches.

The signal-to-background ratios in bins of $p_T(W)$ are shown in Figure 8.15, calculated in the Higgs boson mass window. It is observed that the substructure method achieves the highest S/B in the $200 < p_T(W) < 400$ range, outperforming the resolved approach. However, as the boost of the system becomes higher, a

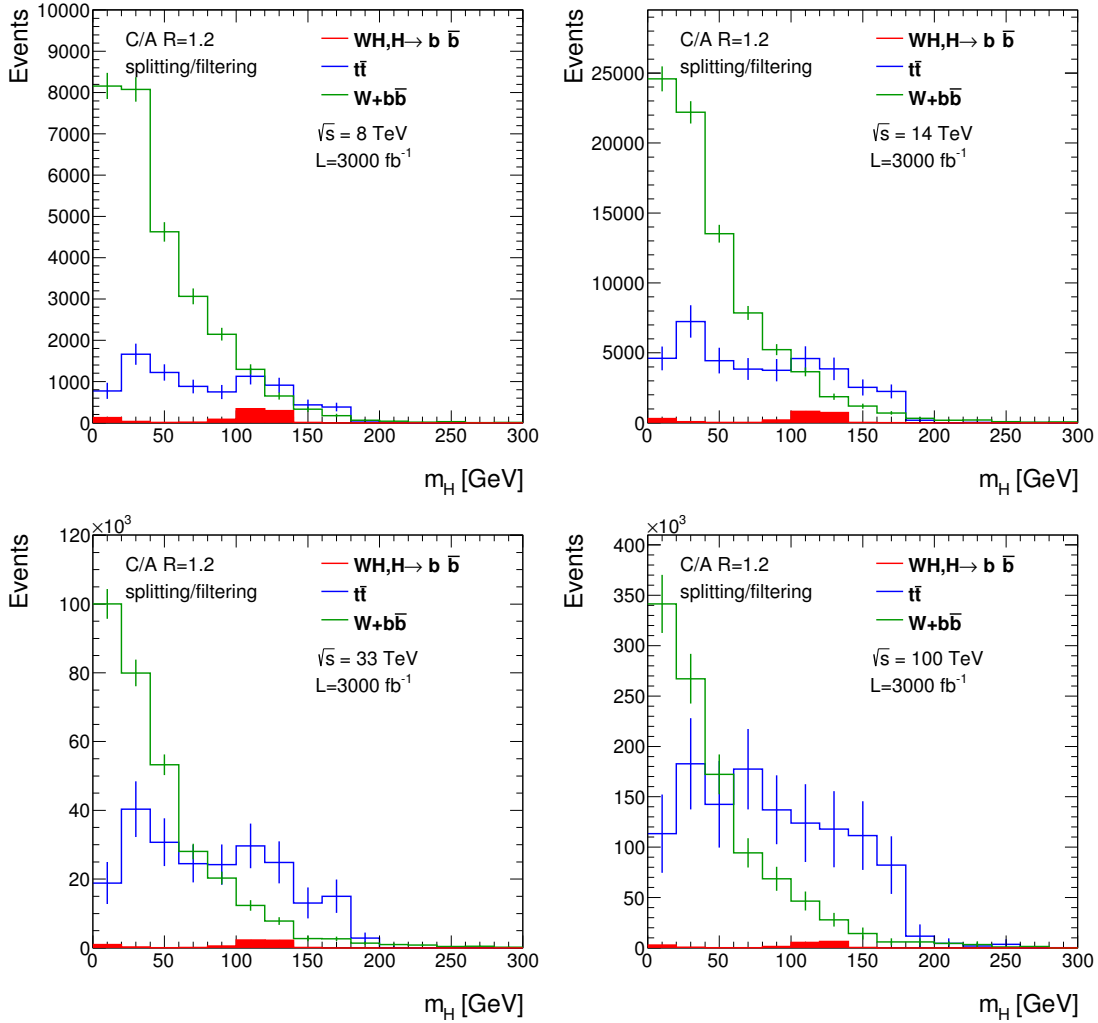


Figure 8.11.: Invariant mass of the Cambridge/Aachen jet after the substructure selection and jet veto, for signal, $t\bar{t}$ and $W + b\bar{b}$ events in the muon channel at $\sqrt{s} = 8, 14, 33, 100$ TeV.

gradual loss of background rejection power results in a decreasing S/B as a function of $p_T(W)$.

In the case of the resolved approach, given the drop in signal efficiency observed for values of $p_T(W)$ greater than 500 GeV, a similar decrease in S/B could have been expected. This drop is however accompanied by an increase in background rejection power, and the S/B therefore increases and stabilizes at higher $p_T(W)$, within the statistical uncertainties.

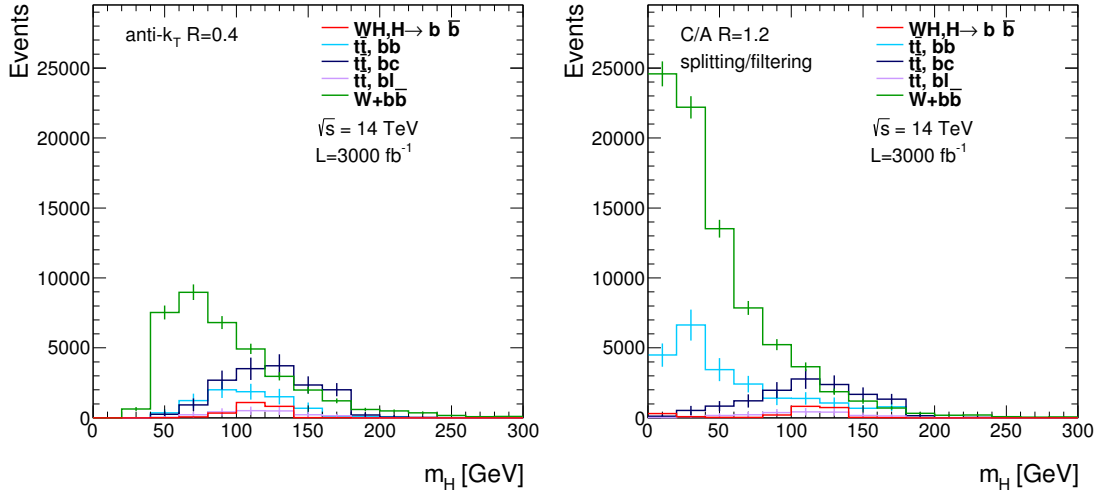


Figure 8.12.: Dijet invariant mass after the resolved selection (right) and invariant mass of the Cambridge/Aachen jet after the substructure selection (left), with the jet veto, for signal, $t\bar{t}$ and $W + b\bar{b}$ events in the muon channel at $\sqrt{s} = 14$ TeV. The $t\bar{t}$ background is split into components according to the flavor of the leading jets/subjets.

For both cases, the exact evolution of the signal-to-background ratios with $p_T(W)$ depends on the definition of the jet veto. As a function of \sqrt{s} , the signal-to-background ratios decrease, as expected, according to the growth of the $t\bar{t}$ event rate with respect to the signal.

The signal significance in the inclusive analysis of Chapter 7 has its maximum in the event region with $p_T(W) > 200$ GeV. How the significance evolves within this range was, however, unclear up to this point. To address this question, the S/\sqrt{B} is calculated in bins of $p_T(W)$, as shown in Figure 8.16 (correcting for additional acceptance originating from the electron channel). It is observed that the most significant event region corresponds to the range $200 < p_T(W) < 400$ GeV, and that higher boosts do not help in achieving a higher signal significance. This observation suggests that the great advantage in boosting the VH system consists in reducing the combinatorial background and the large $t\bar{t}$ contribution, achieved by requiring a transverse momenta of the order of the Higgs boson mass. Higher p_T regions are not beneficial to the signal significance due to the extremely small signal cross-section.

The observed evolution of the signal significance as a function of $p_T(W)$ is not very encouraging with respect to improving the sensitivity of the $VH(b\bar{b})$ analysis. Another conclusion that can be drawn is that relying solely on a substructure

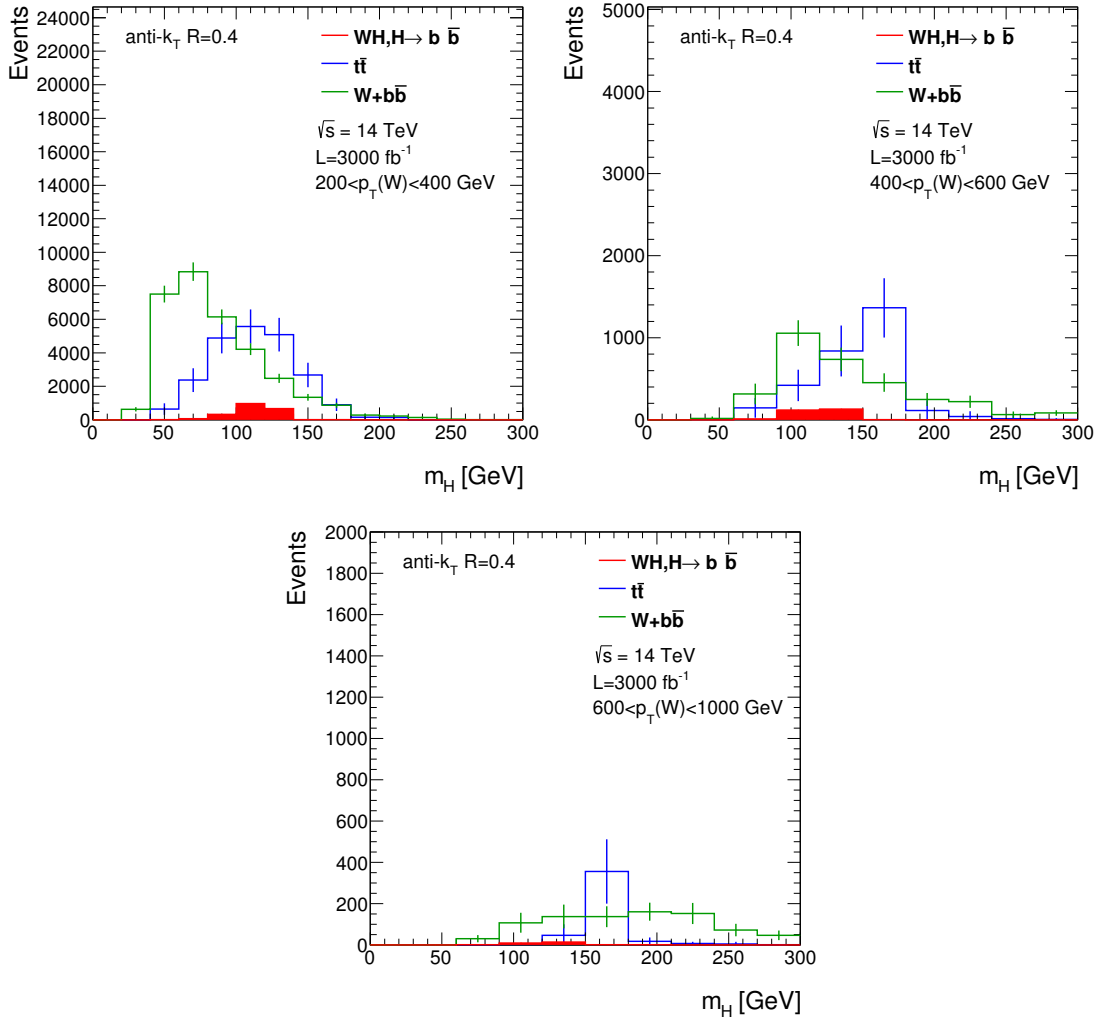


Figure 8.13.: Dijet invariant mass after the resolved selection and jet veto, for signal, $t\bar{t}$ and $W + b\bar{b}$ events in the muon channel, at $\sqrt{s} = 14$ TeV and in different bins of $p_T(W)$.

approach for this analysis does not increase its potential. There is an indication that the significance obtained with the substructure selection may exceed the one obtained with the resolved selection at extremely high boosts, but the statistical error does not allow to make a more accurate determination of the potential gain at this point. Such an observation would be in agreement with the observed evolution of signal efficiency with Cambridge/Aachen jets. However, it also corresponds to the range where S/\sqrt{B} is smaller. A combination of events reconstructed by the resolved approach with events uniquely reconstructed by the substructure method shows an improvement of the order of 1% on the significance of the inclusive boosted region.

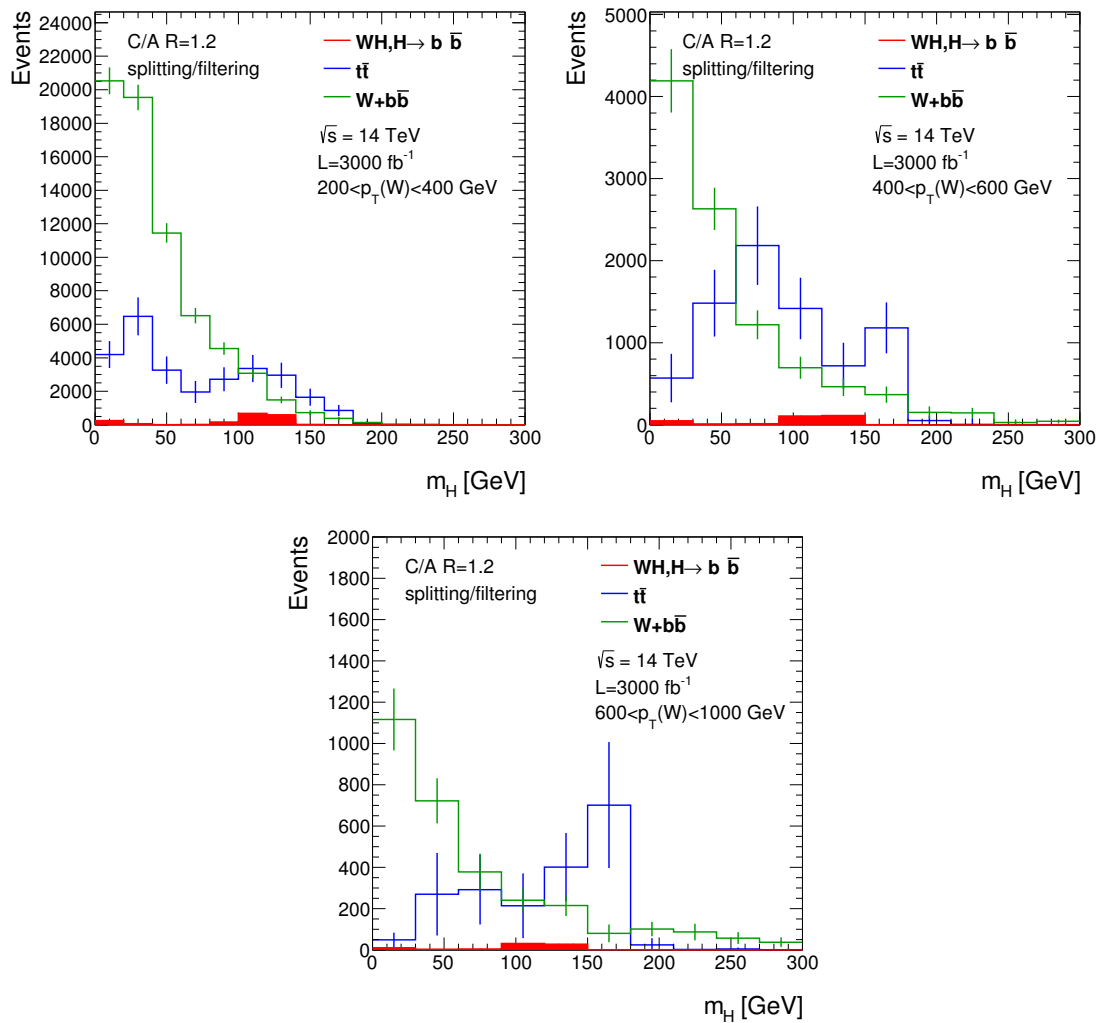


Figure 8.14.: Invariant mass of the Cambridge/Aachen jet after the substructure selection and jet veto, for signal, $t\bar{t}$ and $W+b\bar{b}$ events in the muon channel, at $\sqrt{s} = 14$ TeV and in different bins of $p_T(W)$.

Figure 8.17 shows the signal significances at $\sqrt{s} = 14$ TeV with total integrated luminosities of 150 and 300 fb $^{-1}$, relevant quantities to be expected from Run 2 of the LHC. Assuming that similar sensitivities are obtained from the WH and ZH production channels, the latter in the 2-lepton final-state, a sensitivity per channel of 3.5σ would provide a combined significance large enough to establish the observation of the $H \rightarrow b\bar{b}$ decay mode. At 14 TeV, according to the simple particle-level study presented in this Chapter, such an observation could in principle be claimed with a total integrated luminosity of approximately 75 fb $^{-1}$. At Run 1 energies, approximately 125 fb $^{-1}$ of data would provide the same significance.

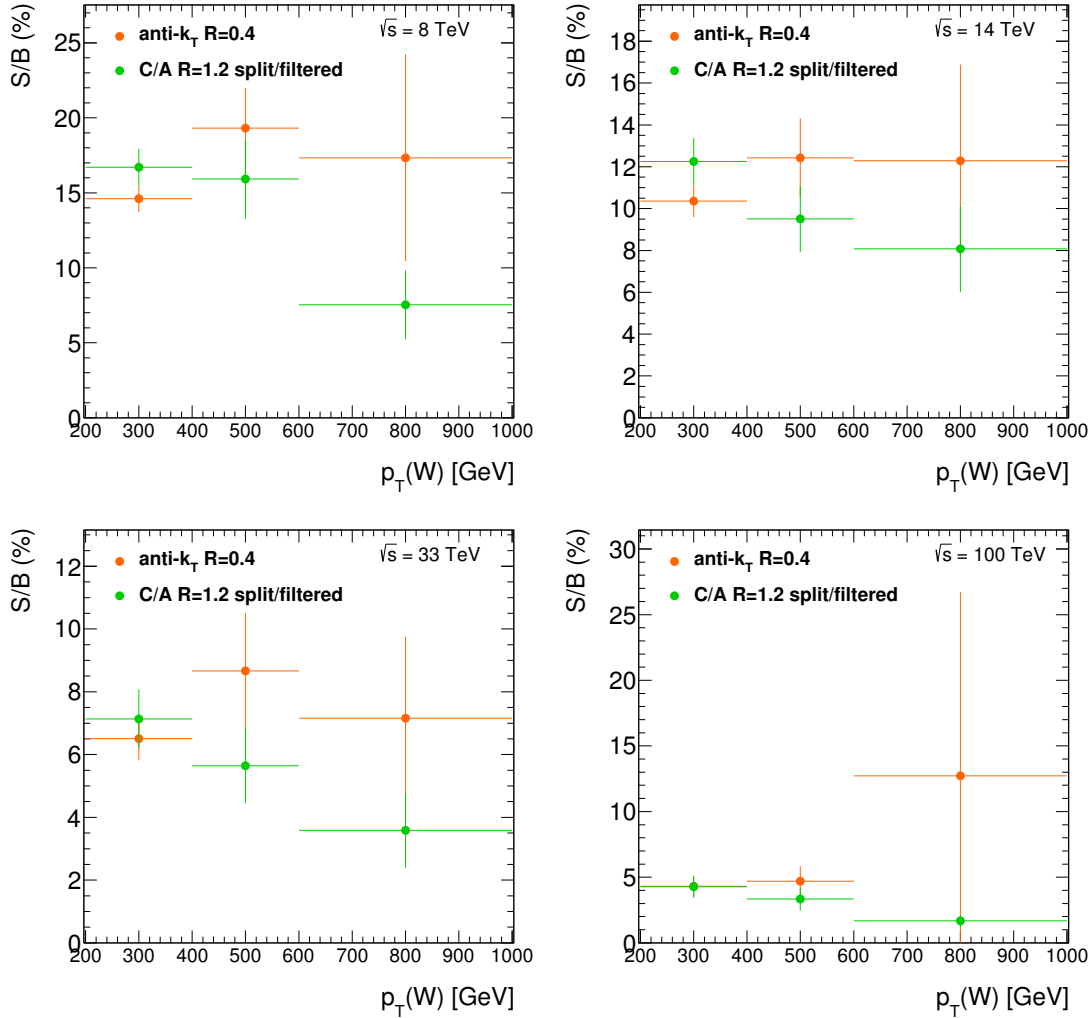


Figure 8.15.: Signal-to-background ratio as a function of $p_T(W)$, at $\sqrt{s} = 8, 14, 33, 100$ TeV.

Such estimates are naturally too optimistic, and the impact of high pile-up and systematic uncertainties considerably deteriorates these predictions. On the other hand, improvements from the new IBL sub-detector and new b -tagging algorithms were also not considered. Finally, the addition of the 0-lepton channel, while experimentally challenging, will further improve the final $VH(b\bar{b})$ sensitivity.

A complete particle-level study should also take into account the contribution of the single-top Wt channel and WZ production where the Z -boson decays to a pair

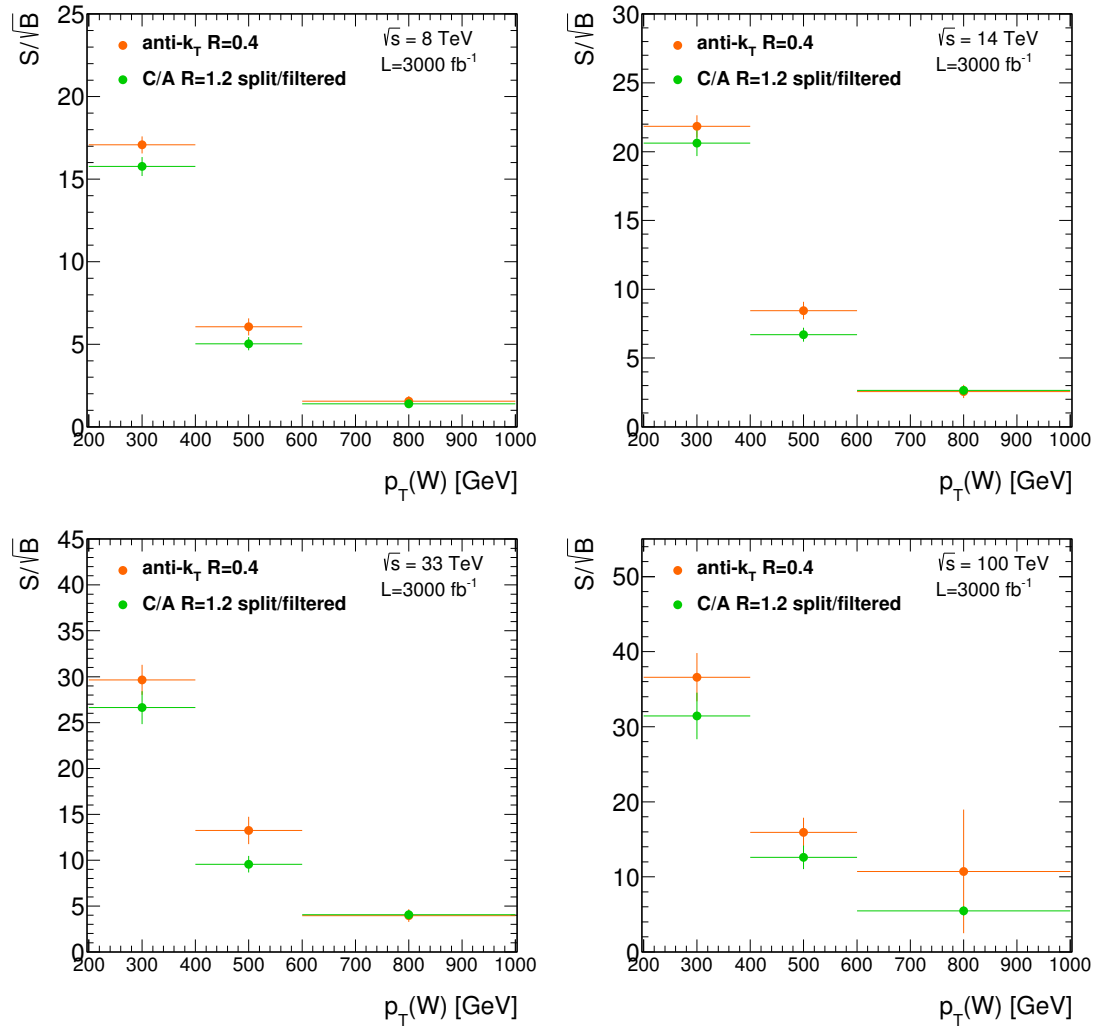


Figure 8.16.: S/\sqrt{B} as a function of $p_T(W)$, at $\sqrt{s} = 8, 14, 33, 100$ TeV, considering a total integrated luminosity of 3000 fb^{-1} .

of b -quarks⁴. The Wt process has a signature very similar to top-pair production and a harder $p_T(W)$ spectrum than other single-top channels, resulting in a small but non-negligible contribution of Wt events in the boosted region. Furthermore, contrary to the s - and t -channels, the resulting mass shape tends to have a peak structure in the same region as the signal. The WZ process has a cross-section approximately 5 times larger than the signal, with a well defined Z resonance peak which, assuming good jet mass resolution, is possible to separate from the Higgs mass peak. Its $p_T(W)$ spectrum is also softer than the signal and its contribution is

⁴At 8 TeV, the single-top and VV processes contribute up to 20% of the total background to the boosted region of the WH channel, whereas QCD multijet events are negligible. [107]

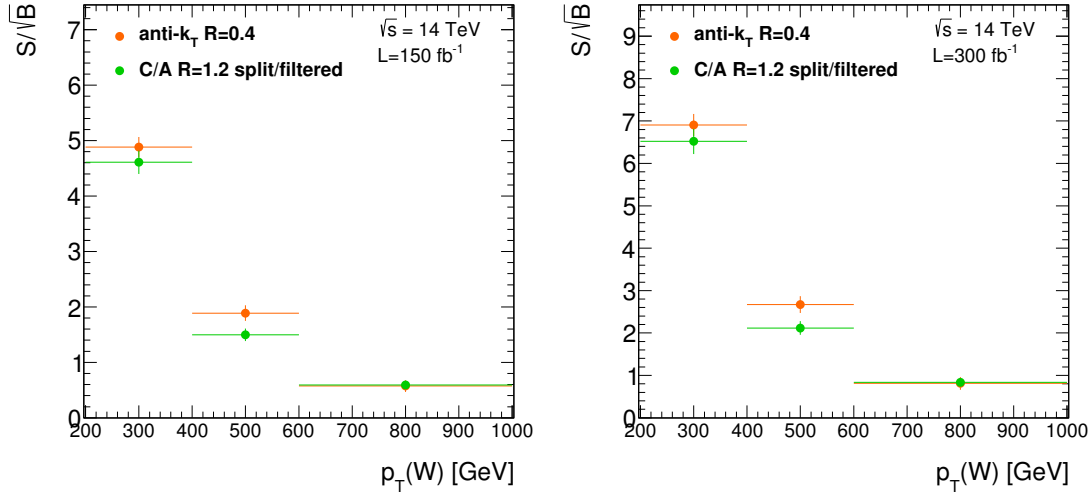


Figure 8.17.: S/\sqrt{B} as a function of $p_T(W)$, at $\sqrt{s} = 14$ TeV, considering a total integrated luminosity of 150 and 300 fb^{-1} .

therefore expected to be relatively small as the boost of the system increases. The performance of the substructure approach in terms of rejection of these processes is expected to be very similar to the $t\bar{t}$ and $W + b\bar{b}$ cases: as in the resolved case, the jet mass resolution and charm-quark rejection rates are essential features.

8.7. Conclusions

A feasibility study of a $WH(b\bar{b})$ search at a pp collider has been performed, exploring centre-of-mass energies of $\sqrt{s} = 8, 14, 33, 100$ TeV and Higgs candidate selections with and without jet substructure techniques. A Standard Model Higgs boson of mass 125 GeV is considered.

The jet substructure technique used was the splitting and filtering algorithm, applied on Cambridge/Aachen $R = 1.2$ jets. The mass drop condition in this algorithm, which is theoretically motivated, was also investigated in this study, and was found not to impact either the signal selection efficiency or the background rejection rates.

A study of the signal significance as a function of the W -boson transverse momentum reveals that the most sensitive region of the analysis corresponds to the $p_T(W)$ bin of 200-400 GeV. In such a regime, the resolved analysis performs well,

in that a good signal efficiency is still achieved. The region where the substructure approach recovers signal events is not as significant: the signal event rates are low and translate into a small and non-competitive S/\sqrt{B} , conclusion which holds true for all the studied centre-of-mass energies.

It can be concluded that the most significant region of the boosted phase-space does not profit from replacing the resolved selection with jet substructure techniques to reconstruct the Higgs boson. An alternative and interesting approach to improve the analysis sensitivity, compatible with the current ATLAS $VH(b\bar{b})$ strategy, would be to combine the power of substructure variables in the current MVA. Most importantly, b -tagging is the central issue in a $VH(b\bar{b})$ analysis, especially given that the $t\bar{t}$ contamination has been observed to be coming mainly from mis-tagged charm jets. The reduction of mis-tagging rates would therefore be a very important improvement to this search.

New physics searches, e.g., involving Higgs-pair production processes [128], or measurements of the Higgs boson differential cross-section with respect to its p_T , would also greatly benefit from developments on substructure and b -tagging techniques.

The physics of 33 TeV and 100 TeV colliders has interesting features, distinct from the Run 1 and Run 2 environments of the LHC. Constraining of proton parton distribution functions has the utmost importance, with the added contribution of electroweak corrections. On the experimental side, forward tracking capabilities could be relevant to many searches at these energies. It should also be noted that at 33 and 100 TeV, the $t\bar{t}H$ cross-section exceeds the VH cross-section. A dedicated study of signal efficiency and background rejection would be of value to establish the most sensitive channel to study the specific coupling of the Higgs boson to bottom-quarks. Boosted topologies, in particular, can have a great impact in this production channel, through the reduction of combinatorial backgrounds.

Chapter 9.

Conclusions

“Jazz is a very democratic musical form. It comes out of a communal experience. We take our respective instruments and collectively create a thing of beauty.”

— Max Roach

A new era of electroweak breaking precision measurements can be anticipated from the ATLAS and CMS physics programs for the Run 2 of the LHC. Among the most challenging and high-priority analysis is the search for a Higgs boson decaying in the $b\bar{b}$ decay mode. In this thesis, several aspects associated to a $VH(b\bar{b})$ search were presented, preceding and following the Higgs boson observation by ATLAS and CMS in 2012 [21, 22].

Anticipating the limiting background rates to this search, it was suggested in a previous study [43] that exploring and optimising events where the V and the Higgs boson are produced with large transverse momenta, the so-called boosted topology, could improve the discovery potential in this channel. To take full advantage of such a regime, jet substructure techniques were proposed and developed. Many other techniques have been developed since then and are currently widely used in searches and measurements involving massive high- p_T objects such as W , Z -bosons and t -quarks.

A $WH(b\bar{b})$ search in the boosted regime was performed with the 7 TeV ATLAS dataset, employing jet substructure techniques for the reconstruction of a Higgs boson with masses of 120 GeV and 125 GeV. A Monte-Carlo based optimization of jet grooming algorithms was carried out, and the obtained sensitivities were compared to the result of the ATLAS (resolved) analysis. It was observed that, while the boost is essential, the substructure approach did not improve the search sensitivity.

The mass of the discovered Higgs boson, at the edge of applicability of substructure methods, associated with the good performance of anti- k_t jets and optimization of resolved selection cuts, contributed to this conclusion.

The result of the ATLAS Run 1 $VH(b\bar{b})$ search was presented [107]. The features of the boosted regime are taken advantage of by categorizing events in bins of $p_T(V)$, establishing control regions to model $t\bar{t}$ and $V + \text{jets}$ processes with a data-driven approach, and applying an upper angular separation cut on the two anti- k_t jets. The treatment of systematic uncertainties related to the production of W -bosons and heavy flavor jets was described, highlighting the large limitations introduced in this search by an incomplete knowledge of Standard Model processes. The compatibility between the observed data and a Higgs boson with a mass of 125 GeV was measured and an excess with a significance of 1.4σ was observed.

In preparation for a centre-of-mass energy of 14 TeV, and taking into account the worsening of pile-up conditions and background levels with respect to the already difficult Run 1 $VH(b\bar{b})$ search, strategies to improve the sensitivity of the boosted regime are necessary. Therefore, a particle-level study was performed to revisit the potential of substructure techniques in the $WH(b\bar{b})$ channel. It was concluded that the high sensitivity of the boosted regime originates from a specific region, where $200 < p_T(W) < 400$ GeV. In such conditions, the performance of anti- k_t jets is still very good. Although greater boosts result in a severe drop in signal efficiency, only recovered by jet substructure techniques, the signal significance of such event regions is small. It can thus be concluded that the potential for improving the sensitivity of this search does not necessarily involve jet grooming algorithms or substructure variables. It was nonetheless suggested that more sophisticated approaches can be developed to exploit the information contained in these variables. Finally, the importance of developing b -tagging algorithms with strong charm rejection was emphasized, as a strategy to tackle the growing contribution of $t\bar{t}$ events in the Higgs boson mass window. Developments on both substructure and b -tagging techniques would also be relevant in new physics searches and measurements involving the Higgs boson.

Appendix A.

Optimization of jet substructure algorithms for $H \rightarrow b\bar{b}$

Jet shape variables can be employed in addition to jet grooming algorithms to achieve greater background rejection. As described in Chapter 4, N -subjettiness ratios can provide discrimination between jets with different prong structures. The τ_{21} ratio as calculated for anti- k_t trimmed jets with $R = 1.5$ is shown in Figure A.1 for WH , $W + b\bar{b}$ and $t\bar{t}$ events.

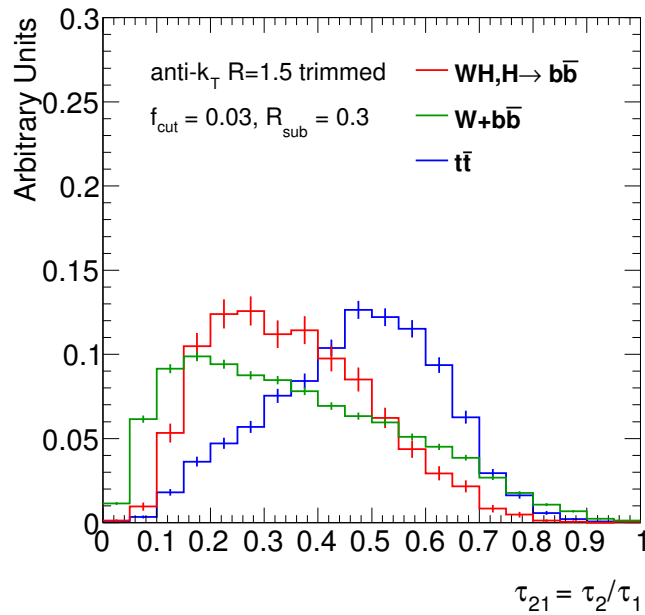


Figure A.1.: Distribution of N -subjettiness ratio τ_{21} for anti- k_t $R = 1.5$ trimmed jets. $m_H = 125$ GeV is considered in the sample generation. The distributions are normalized to the same area.

Appendix B.

Modelling of $W + b\bar{b}$ background

The modelling of the $W + b\bar{b}$ process, an irreducible background to the $H \rightarrow b\bar{b}$ search in the WH production channel, can be studied through MC-to-MC comparisons. This Appendix shows additional comparisons between the available MC generators, which were useful to understand and motivate the choice of systematic uncertainties associated to the W +jets backgrounds, as described in Chapter 7.

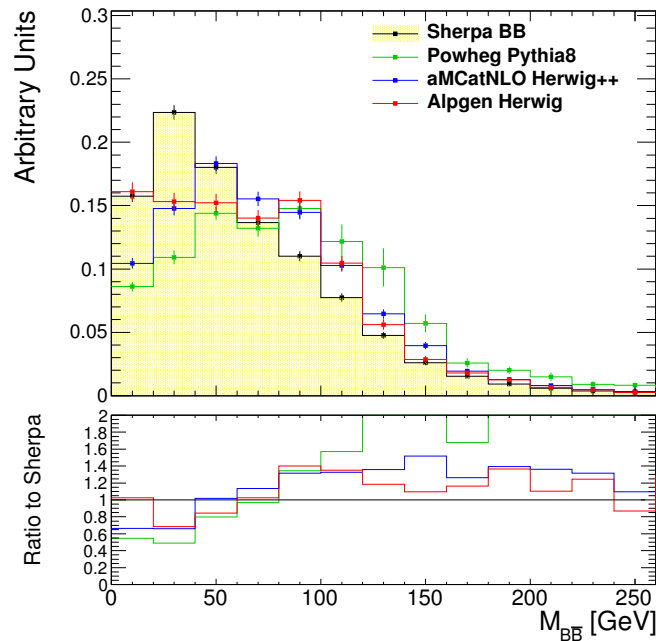


Figure B.1.: Distribution of the invariant mass of leading B hadrons matched to leading jets, as predicted by different MC generators and their ratio with respect to SHERPA. Events correspond to the 2-jet bb region (dijet mass selection). Normalized to same area.

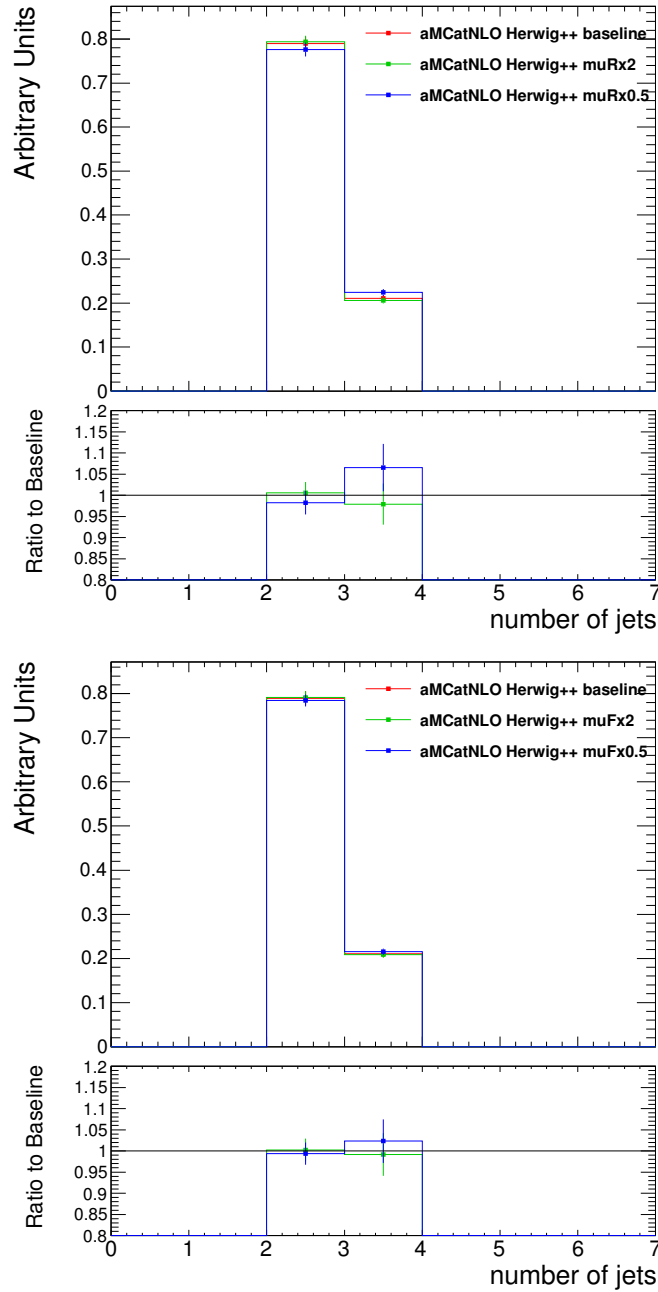


Figure B.2.: Distribution of jet multiplicity in $W + b\bar{b}$ events as predicted by the baseline aMC@NLO sample and by samples where the μ_R (top) and μ_F (bottom) parameters have been varied (independently). Events correspond to the bb region (dijet mass selection). Normalized to same area.

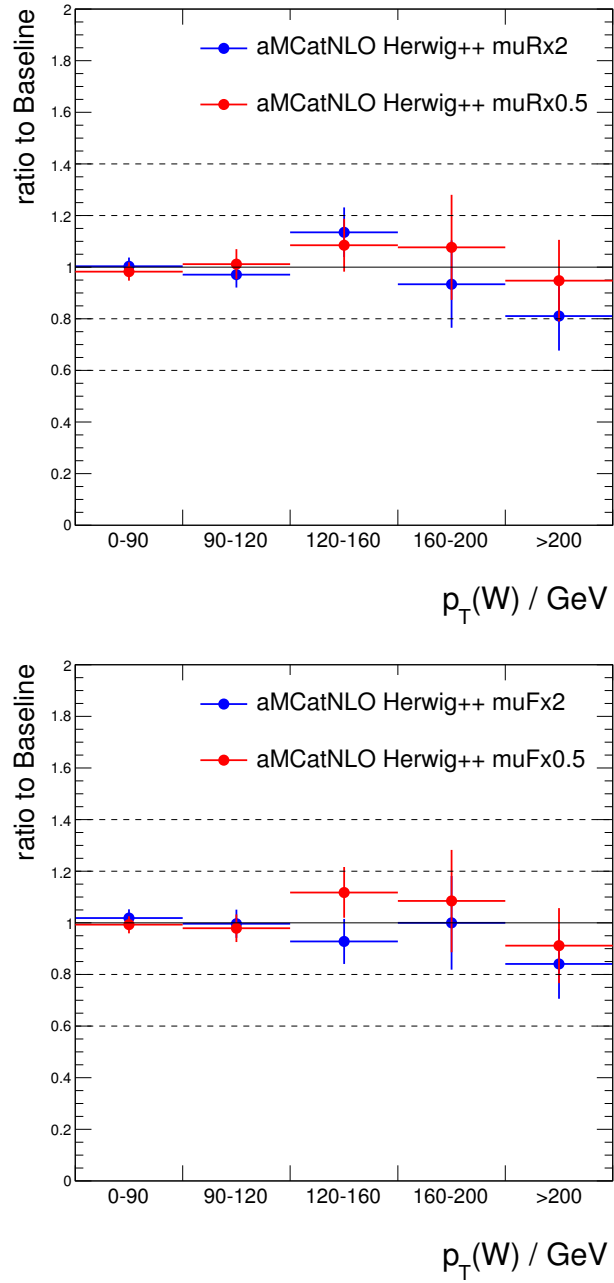


Figure B.3.: Distribution of relative yields in each $p_T(W)$ bin, in $W + b\bar{b}$ events, as predicted by aMC@NLO samples where the μ_R (top) and μ_F (bottom) parameters have been varied (independently), with respect to the baseline aMC@NLO. Events correspond to the 2-jet bb region (dijet mass selection).

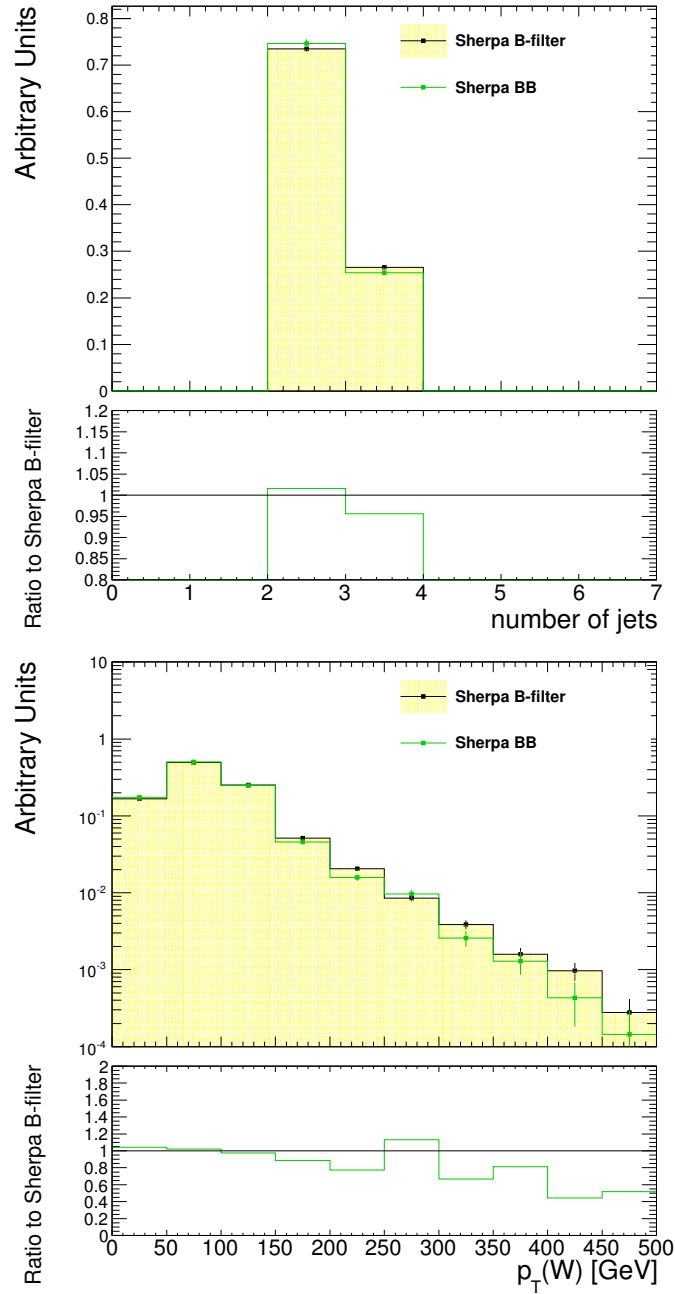


Figure B.4.: Distribution of jet multiplicity (top) and $p_T(W)$ (bottom) in $W + b\bar{b}$ events as predicted by SHERPA simulated events where the B hadrons are originating exclusively from the matrix element (SHERPA BB) and from either matrix element or parton shower (SHERPA B-filter). Events correspond to the 2-jet $b\bar{b}$ region (dijet mass selection). Normalized to same area.

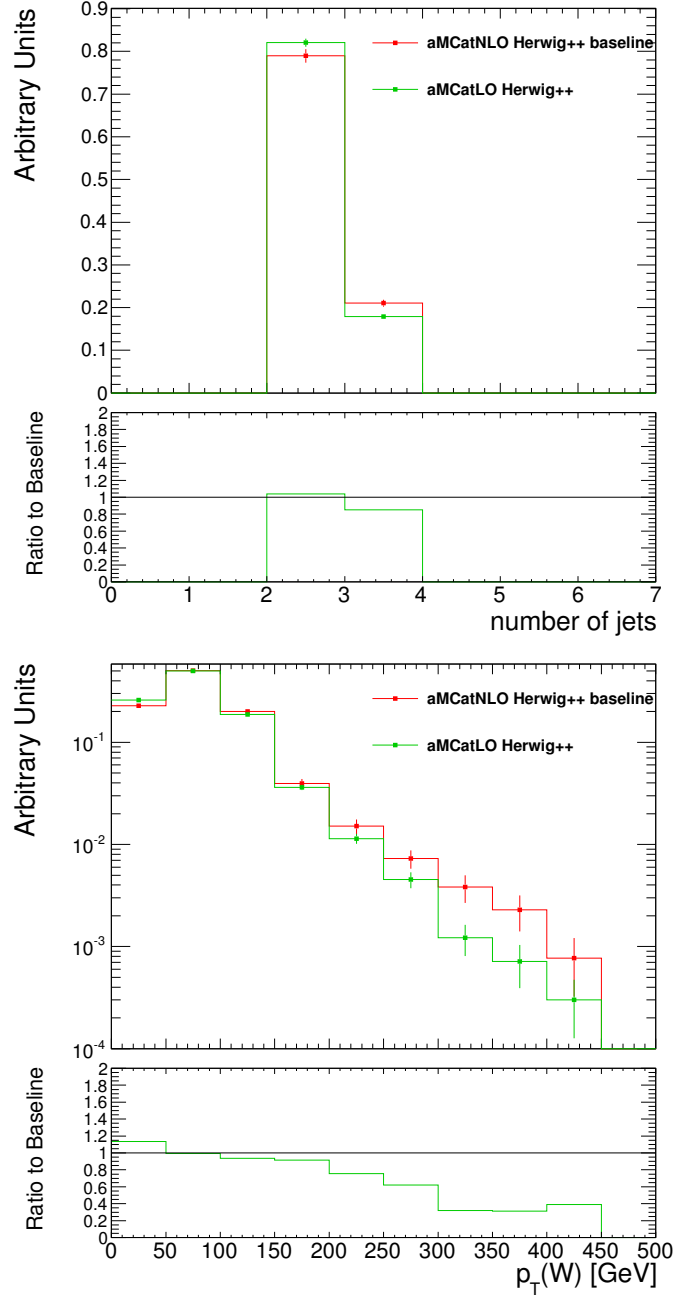


Figure B.5.: Distribution of jet multiplicity (top) and $p_T(W)$ (bottom) in $W + b\bar{b}$ events as predicted by the baseline aMC@NLO sample and by aMC@LO. Events correspond to the 2-jet $b\bar{b}$ region (dijet mass selection). Normalized to same area..

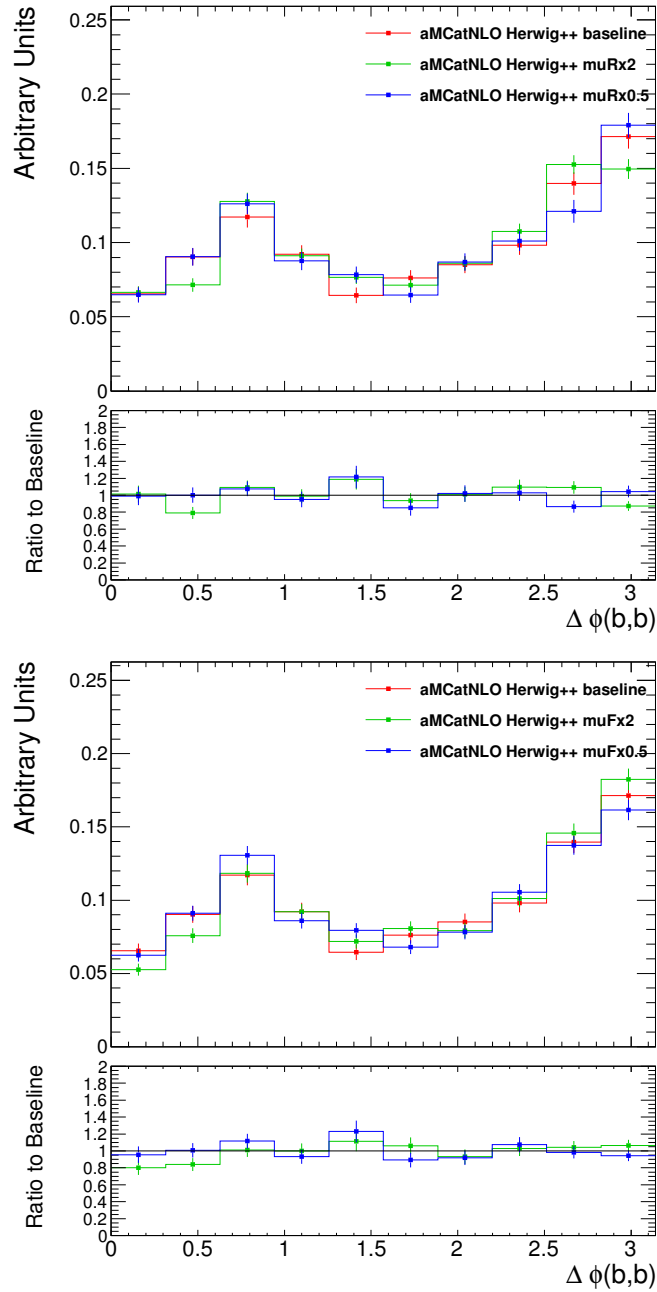


Figure B.6.: $\Delta\varphi(j_1, j_2)$ distribution in $W + b\bar{b}$ events as predicted by the baseline aMC@NLO sample and by samples where the μ_R (top) and μ_F (bottom) parameters have been varied (independently). Events correspond to the 2-jet bb region (dijet mass selection). Normalized to same area.

Appendix C.

Prospects for 14 TeV and higher centre-of-mass energies

C.1. Invariant mass distribution of Higgs boson candidates

The results of the study presented in Chapter 8 are produced considering a Higgs boson mass window of $90 < m_H < 140$ GeV, motivated by the shape of the invariant mass distributions shown in Figure C.1.

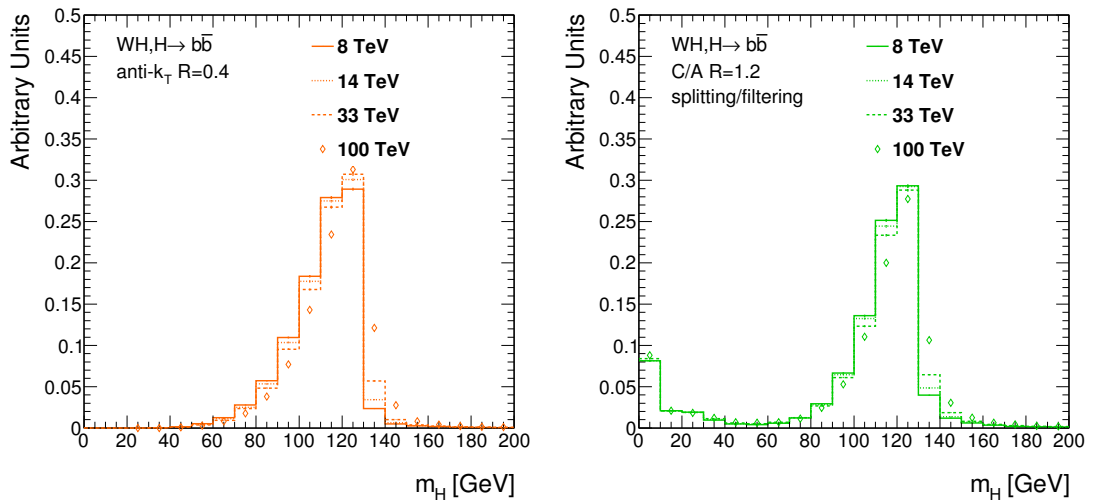


Figure C.1.: Invariant mass distribution of the Higgs boson candidate using the resolved (left) and substructure-based selections (right), at $\sqrt{s} = 8, 14, 33, 100$ TeV.

C.2. Jet veto optimization

A fundamental difference between the $t\bar{t}$ process and the WH signal is the number of final state objects originating from the hard process. The anti- k_t jet multiplicity after the resolved and substructure selection is therefore used to discriminate between top and signal events, considering jets with $p_T > 20$ GeV and $|\eta| < 5.0$. Furthermore, jets in $t\bar{t}$ events will typically be more central, given that they originate from a top-quark decay, whereas in the WH case they will correspond to QCD radiation. Therefore, a minimum absolute pseudo-rapidity cut is imposed on the sub-subleading anti- k_t jet. Only jets with a transverse momentum above a given threshold are, however, considered for the pseudo-rapidity veto. Such an approach is more realistic with respect to the expected levels of underlying event and pile-up contamination, despite decreasing the $t\bar{t}$ rejection power. An event-by-event p_T threshold was tested in this study, dependent on the transverse momentum of the reconstructed W -boson.

Tables C.1 to C.4 show the impact of each cut on the signal and background efficiencies, signal-to-background ratio and signal significance.

Table C.1.: Optimization of jet veto at $\sqrt{s} = 8$ TeV, considering the (di-)jet invariant mass window of 90 to 140 GeV.

Event cut	$\epsilon(WH)$ (%)	$\epsilon(W + b\bar{b})$ (%)	$\epsilon(t\bar{t})$ (%)	S/B (%)	S/\sqrt{B}
Resolved selection					
no jet veto	100	100	100	3.8	7.9
number of jets < 5	97	92	58	5.8	9.6
number of jets < 4	88	80	26	9.6	11.7
number of jets < 3	58	47	6	17.9	13.0
number of jets < 4 (3 rd jet: $ \eta > 3.0$)	60	49	6	17.1	13.0
number of jets < 4 (3 rd jet: $ \eta > 3.0$ or $p_T < 0.1p_T^W$)	69	59	9	14.9	13.0
number of jets < 4 (3 rd jet: $ \eta > 3.0$ or $p_T < 0.2p_T^W$)	83	75	20	10.8	12.2
Substructure selection					
no jet veto	100	100	100	3.7	7.2
number of jets < 5	95	87	50	6.3	9.1
number of jets < 4	83	67	22	10.6	11.0
number of jets < 3	52	36	4	20.1	12.0
number of jets < 4 (3 rd jet: $ \eta > 3.0$)	54	37	5	19.3	12.0
number of jets < 4 (3 rd jet: $ \eta > 3.0$ or $p_T < 0.1p_T^W$)	63	46	8	16.2	11.8
number of jets < 4 (3 rd jet: $ \eta > 3.0$ or $p_T < 0.2p_T^W$)	77	60	16	12.3	11.4

Table C.2.: Optimization of jet veto at $\sqrt{s} = 14$ TeV, considering the (di-)jet invariant mass window of 90 to 140 GeV.

Event cut	$\epsilon(WH)$ (%)	$\epsilon(W + b\bar{b})$ (%)	$\epsilon(t\bar{t})$ (%)	S/B (%)	S/\sqrt{B}
Resolved selection					
no jet veto	100	100	100	2.0	9.1
number of jets < 5	95	89	49	3.6	11.9
number of jets < 4	83	71	21	6.4	14.8
number of jets < 3	52	38	4	13.6	17.0
number of jets < 4 (3 rd jet: $ \eta > 3.0$)	55	41	5	12.6	16.9
number of jets < 4 (3 rd jet: $ \eta > 3.0$ or $p_T < 0.1p_T^W$)	65	52	8	10.5	16.8
number of jets < 4 (3 rd jet: $ \eta > 3.0$ or $p_T < 0.2p_T^W$)	79	66	16	7.5	15.6
Substructure selection					
no jet veto	100	100	100	1.9	8.2
number of jets < 5	92	78	39	4.0	11.5
number of jets < 4	78	56	16	7.0	14.1
number of jets < 3	46	26	3	15.5	16.0
number of jets < 4 (3 rd jet: $ \eta > 3.0$)	49	28	4	14.4	16.0
number of jets < 4 (3 rd jet: $ \eta > 3.0$ or $p_T < 0.1p_T^W$)	58	36	6	11.5	15.6
number of jets < 4 (3 rd jet: $ \eta > 3.0$ or $p_T < 0.2p_T^W$)	71	48	12	8.4	14.7

Table C.3.: Optimization of jet veto at $\sqrt{s} = 33$ TeV, considering the (di-)jet invariant mass window of 90 to 140 GeV.

Event cut	$\epsilon(WH)$ (%)	$\epsilon(W + b\bar{b})$ (%)	$\epsilon(t\bar{t})$ (%)	S/B (%)	S/\sqrt{B}
Resolved selection					
no jet veto	100	100	100	1.1	12.3
number of jets < 5	91	81	39	2.4	17.2
number of jets < 4	75	59	17	4.2	20.6
number of jets < 3	42	30	3	9.3	23.2
number of jets < 4 (3 rd jet: $ \eta > 4.0$)	44	32	4	8.5	22.7
number of jets < 4 (3 rd jet: $ \eta > 4.0$ or $p_T < 0.1p_T^W$)	56	42	6	6.7	22.7
number of jets < 4 (3 rd jet: $ \eta > 4.0$ or $p_T < 0.2p_T^W$)	70	55	13	4.8	21.4
Substructure selection					
no jet veto	100	100	100	0.8	10.0
number of jets < 5	85	64	28	2.2	15.5
number of jets < 4	67	40	11	4.1	18.7
number of jets < 3	36	17	2	10.2	21.5
number of jets < 4 (3 rd jet: $ \eta > 4.0$)	37	18	2	9.0	20.6
number of jets < 4 (3 rd jet: $ \eta > 4.0$ or $p_T < 0.1p_T^W$)	48	25	4	6.4	19.8
number of jets < 4 (3 rd jet: $ \eta > 4.0$ or $p_T < 0.2p_T^W$)	61	34	8	4.8	19.2

Table C.4.: Optimization of jet veto at $\sqrt{s} = 100$ TeV, considering the (di-)jet invariant mass window of 90 to 140 GeV.

Event cut	$\epsilon(WH)$ (%)	$\epsilon(W + b\bar{b})$ (%)	$\epsilon(t\bar{t})$ (%)	S/B (%)	S/\sqrt{B}
Resolved selection					
no jet veto	100	100	100	0.6	15.5
number of jets < 5	77	60	24	1.6	23.3
number of jets < 4	57	40	9	3.1	27.6
number of jets < 3	28	18	1	6.6	28.5
number of jets < 4 (3 rd jet: $ \eta > 4.0$)	31	21	2	5.9	28.3
number of jets < 4 (3 rd jet: $ \eta > 4.0$ or $p_T < 0.1p_T^W$)	43	29	4	4.4	28.8
number of jets < 4 (3 rd jet: $ \eta > 4.0$ or $p_T < 0.2p_T^W$)	54	37	7	3.3	28.1
Substructure selection					
no jet veto	100	100	100	0.3	12.1
number of jets < 5	69	40	16	1.3	20.4
number of jets < 4	48	25	5	2.6	23.8
number of jets < 3	22	10	1	6.3	25.0
number of jets < 4 (3 rd jet: $ \eta > 4.0$)	25	12	1	5.5	24.8
number of jets < 4 (3 rd jet: $ \eta > 4.0$ or $p_T < 0.1p_T^W$)	35	16	2	3.7	24.1
number of jets < 4 (3 rd jet: $ \eta > 4.0$ or $p_T < 0.2p_T^W$)	43	21	4	3.0	24.3

Bibliography

- [1] Particle Data Group Collaboration, J. Beringer et al., *Review of Particle Physics (RPP)*, Phys.Rev. **D86** (2012) 010001.
- [2] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl.Phys. **22** (1961) 579–588.
- [3] S. Weinberg, *A Model of Leptons*, Phys.Rev.Lett. **19** (1967) 1264–1266.
- [4] A. Salam, *Weak and Electromagnetic Interactions*, Conf.Proc. **C680519** (1968) 367–377.
- [5] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, Phys.Rev.Lett. **30** (1973) 1343–1346.
- [6] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys.Rev.Lett. **30** (1973) 1346–1349.
- [7] D. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. 1*, Phys.Rev. **D8** (1973) 3633–3652.
- [8] H. D. Politzer, *Asymptotic Freedom: An Approach to Strong Interactions*, Phys.Rept. **14** (1974) 129–180.
- [9] P. W. Anderson, *Plasmons, Gauge Invariance, and Mass*, Phys.Rev. **130** (1963) 439–442.
- [10] G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys.Rev.Lett. **13** (1964) 585–587.
- [11] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys.Rev.Lett. **13** (1964) 508–509.
- [12] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*,

- Phys.Rev. **145** (1966) 1156–1163.
- [13] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys.Rev.Lett. **13** (1964) 321–323.
- [14] Gargamelle Neutrino Collaboration, F. Hasert et al., *Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment*, Phys.Lett. **B46** (1973) 138–140.
- [15] UA1 Collaboration, G. Arnison et al., *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, Phys.Lett. **B122** (1983) 103–116.
- [16] UA1 Collaboration, G. Arnison et al., *Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c² at the CERN SPS Collider*, Phys.Lett. **B126** (1983) 398–410.
- [17] UA2 Collaboration, M. Banner et al., *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, Phys.Lett. **B122** (1983) 476–485.
- [18] UA2 Collaboration, P. Bagnaia et al., *Evidence for $Z^0 \rightarrow e^+ e^-$ at the CERN anti-p p Collider*, Phys.Lett. **B129** (1983) 130–140.
- [19] M. S. Carena and H. E. Haber, *Higgs boson theory and phenomenology*, Prog.Part.Nucl.Phys. **50** (2003) 63–152, arXiv:hep-ph/0208209 [hep-ph].
- [20] LEP Electroweak Working Group (LEP EWWG), *Status of March 2012*, <http://lepewwg.web.cern.ch/LEPEWWG/>.
- [21] ATLAS Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].
- [22] CMS Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].
- [23] CDF Collaboration, D0 Collaboration, T. Aaltonen et al., *Higgs Boson Studies at the Tevatron*, Phys.Rev. **D88** (2013) no. 5, 052014, arXiv:1303.6346 [hep-ex].

- [24] ATLAS Collaboration, G. Aad et al., *Measurements of the properties of the Higgs-like boson in the two photon decay channel with the ATLAS detector using 25 fb⁻¹ of proton-proton collision data*, Tech. Rep. ATLAS-CONF-2013-012, CERN, Geneva, Mar, 2013.
- [25] ATLAS Collaboration, G. Aad et al., *Measurements of the properties of the Higgs-like boson in the four lepton decay channel with the ATLAS detector using 25 fb⁻¹ of proton-proton collision data*, Tech. Rep. ATLAS-CONF-2013-013, CERN, Geneva, Mar, 2013.
- [26] ATLAS Collaboration, G. Aad et al., *Measurements of the properties of the Higgs-like boson in the $WW^{(*)} \rightarrow \ell\nu\ell\nu$ decay channel with the ATLAS detector using 25 fb⁻¹ of proton-proton collision data*, Tech. Rep. ATLAS-CONF-2013-030, CERN, Geneva, Mar, 2013.
- [27] ATLAS Collaboration, G. Aad et al., *Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2013-108, CERN, Geneva, Nov, 2013.
- [28] ATLAS Collaboration, *Combined measurements of the mass and signal strength of the Higgs-like boson with the ATLAS detector using up to 25 fb⁻¹ of proton-proton collision data*, Tech. Rep. ATLAS-CONF-2013-014, CERN, Geneva, Mar, 2013.
- [29] ATLAS Collaboration, *Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb⁻¹ of proton-proton collision data*, Tech. Rep. ATLAS-CONF-2014-009, CERN, Geneva, Mar, 2014.
- [30] ATLAS Collaboration, *Study of the spin of the new boson with up to 25 fb⁻¹ of ATLAS data*, Tech. Rep. ATLAS-CONF-2013-040, CERN, Geneva, Apr, 2013.
- [31] S. Dittmaier, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, et al., *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, arXiv:1201.3084 [hep-ph].
- [32] J. Baglio and A. Djouadi, *Higgs production at the LHC*, JHEP **1103** (2011) 055, arXiv:1012.0530 [hep-ph].
- [33] C. Englert, M. McCullough, and M. Spannowsky, *Gluon-initiated associated production boosts Higgs physics*, Phys.Rev. **D89** (2014) 013013,

- arXiv:1310.4828 [hep-ph].
- [34] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak, and T. J. Zirke, *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, JHEP **1302** (2013) 078, arXiv:1211.5015 [hep-ph].
- [35] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Phys.Lett. **B579** (2004) 149–156, arXiv:hep-ph/0307206 [hep-ph].
- [36] M. Ciccolini, S. Dittmaier, and M. Kramer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, Phys.Rev. **D68** (2003) 073003, arXiv:hep-ph/0306234 [hep-ph].
- [37] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593 [hep-ph].
- [38] *Search for the Standard Model Higgs boson produced in association with top quarks and decaying to $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector at the LHC*, Tech. Rep. ATLAS-CONF-2014-011, CERN, Geneva, Mar, 2014.
- [39] F. Febres Cordero, L. Reina, and D. Wackerroth, *Associated Production of a W or Z Boson with Bottom Quarks at the Tevatron and the LHC*, PoS **RADCOR2009** (2010) 055, arXiv:1001.3362 [hep-ph].
- [40] ATLAS Collaboration, G. Aad et al., *Measurement of the cross-section for W boson production in association with b-jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **1306** (2013) 084, arXiv:1302.2929 [hep-ex].
- [41] CDF Collaboration, T. Aaltonen et al., *First Measurement of the b-jet Cross Section in Events with a W Boson in p anti-p Collisions at $s^{**}(1/2) = 1.96$ -TeV*, Phys.Rev.Lett. **104** (2010) 131801, arXiv:0909.1505 [hep-ex].
- [42] ATLAS Collaboration, *Combination of ATLAS and CMS top-quark pair cross section measurements using up to 1.1 fb^{-1} of data at 7 TeV*, Tech. Rep. ATLAS-CONF-2012-134, CERN, Geneva, Sep, 2012.
- [43] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet*

- substructure as a new Higgs search channel at the LHC*, Phys.Rev.Lett. **100** (2008) 242001, arXiv:0802.2470 [hep-ph].
- [44] M. Peskin and D. Schroeder, *An Introduction to Quantum Field Theory*. Addison-Wesley Publishing Company, 1995.
- [45] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026, arXiv:0603175 [hep-ph].
- [46] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput.Phys.Commun. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].
- [47] G. Corcella et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, JHEP **01** (2001) 010, arXiv:0011363 [hep-ph].
- [48] S. Gieseke, D. Grellscheid, K. Hamilton, A. Papaefstathiou, S. Platzer, et al., *Herwig++ 2.5 Release Note*, arXiv:1102.1672 [hep-ph].
- [49] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58** (2008) 639–707, arXiv:0803.0883 [hep-ph].
- [50] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **1006** (2010) 043, arXiv:1002.2581 [hep-ph].
- [51] 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].
- [52] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **0206** (2002) 029, arXiv:hep-ph/0204244 [hep-ph].
- [53] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph].
- [54] M. L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, arXiv:0206293 [hep-ph].

- [55] K. Hamilton and P. Nason, *Improving NLO-parton shower matched simulations with higher order matrix elements*, JHEP **1006** (2010) 039, arXiv:1004.1764 [hep-ph].
- [56] S. Hoche, F. Krauss, M. Schonherr, and F. Siegert, *NLO matrix elements and truncated showers*, JHEP **1108** (2011) 123, arXiv:1009.1127 [hep-ph].
- [57] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, JHEP **1304** (2013) 027, arXiv:1207.5030 [hep-ph].
- [58] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN, Geneva, 2004.
- [59] ATLAS Collaboration, G. Aad et al., *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, Eur.Phys.J. **C73** (2013) 2518, arXiv:1302.4393 [hep-ex].
- [60] ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, J. Instrum. **3** (2008) S08003.
- [61] ATLAS Collaboration, G. Aad et al., *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex].
- [62] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2010-069, CERN, Geneva, Jul, 2010.
- [63] ATLAS Collaboration, G. Aad et al., *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, arXiv:1407.5063 [hep-ex].
- [64] ATLAS Collaboration, G. Aad et al., *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, arXiv:1407.3935 [hep-ex].
- [65] ATLAS Collaboration, W. Lampl et al., *Calorimeter Clustering algorithms: Description and Performance*, Tech. Rep. ATLAS-LARG-PUB-2008-002, CERN, Geneva, 2008.
- [66] T. Barillari, E. Bergeas Kuutmann, T. Carli, J. Erdmann, P. Giovannini, K. J.

- Grahn, C. Issever, A. Jantsch, A. Kiryunin, K. Lohwasser, A. Maslennikov, S. Menke, H. Oberlack, G. Pospelov, E. Rauter, P. Schacht, F. Spanó, P. Speckmayer, P. Stavina, and P. Strízenec, *Local Hadronic Calibration*, Tech. Rep. ATL-LARG-PUB-2009-001-2. ATL-COM-LARG-2008-006. ATL-LARG-PUB-2009-001, CERN, Geneva, Jun, 2008.
- [67] *Pile-up subtraction and suppression for jets in ATLAS*, Tech. Rep. ATLAS-CONF-2013-083, CERN, Geneva, Aug, 2013.
- [68] ATLAS Collaboration, G. Aad et al., *Tagging and suppression of pileup jets*, Tech. Rep. ATL-PHYS-PUB-2014-001, CERN, Geneva, Jan, 2014.
- [69] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, Tech. Rep. ATLAS-CONF-2011-102, CERN, Geneva, Jul, 2011.
- [70] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, Tech. Rep. ATLAS-CONF-2014-046, CERN, Geneva, Jul, 2014.
- [71] ATLAS Collaboration, G. Aad et al., *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2014-004, CERN, Geneva, Feb, 2014.
- [72] ATLAS Collaboration, *Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector*, Tech. Rep. ATL-PHYS-PUB-2014-013, CERN, Geneva, Aug, 2014.
- [73] ATLAS Collaboration, *b-tagging in dense environments*, Tech. Rep. ATL-PHYS-PUB-2014-014, CERN, Geneva, Aug, 2014.
- [74] ATLAS Collaboration, *Identification and Tagging of Double b-hadron jets with the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2012-100, CERN, Geneva, Jul, 2012.
- [75] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.
- [76] ATLAS Collaboration, G. Aad et al., *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823–874, arXiv:1005.4568 [physics.ins-det].

- [77] ATLAS Collaboration, G. Aad et al., *The ATLAS calorimeter simulation FastCaloSim*, IEEE Nucl.Sci.Symp.Conf.Rec. **2010** (2010) 1–5.
- [78] M. Capeans, G. Darbo, K. Einsweiler, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet, *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19, CERN, Geneva, Sep, 2010.
- [79] A. Annovi, P. Catastini, V. Cavasinni, M. Dell’Orso, A. Dotti, et al., *Hadron collider triggers with high-quality tracking at very high event rates*, IEEE Trans.Nucl.Sci. **51** (2004) 391–400.
- [80] G. F. Sterman and S. Weinberg, *Jets from Quantum Chromodynamics*, Phys.Rev.Lett. **39** (1977) 1436.
- [81] JADE Collaboration, S. Bethke et al., *Experimental Investigation of the Energy Dependence of the Strong Coupling Strength*, Phys.Lett. **B213** (1988) 235.
- [82] JADE Collaboration, W. Bartel et al., *Experimental Studies on Multi-Jet Production in e^+e^- Annihilation at PETRA Energies*, Z.Phys. **C33** (1986) 23.
- [83] M. Cacciari, G. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, JHEP **0804** (2008) 063, arXiv:0802.1189 [hep-ph].
- [84] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber, *New clustering algorithm for multijet cross sections in e^+e^- annihilation*, Phys. Lett. **B269** (1991) 432–438.
- [85] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, arXiv:9907280 [hep-ph].
- [86] ATLAS Collaboration, G. Aad et al., *Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment*, JHEP **1205** (2012) 128, arXiv:1203.4606 [hep-ex].
- [87] ATLAS Collaboration, G. Aad et al., *Performance of jet substructure techniques for large- R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, JHEP **1309** (2013) 076, arXiv:1306.4945 [hep-ex].
- [88] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Recombination Algorithms and*

- Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, Phys.Rev. **D81** (2010) 094023, arXiv:0912.0033 [hep-ph].
- [89] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, JHEP **1002** (2010) 084, arXiv:0912.1342 [hep-ph].
- [90] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N -subjettiness*, JHEP **1103** (2011) 015, arXiv:1011.2268 [hep-ph].
- [91] ATLAS Collaboration, G. Aad et al., *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, Tech. Rep. CERN-PH-EP-2011-191, CERN, Geneva, 2011. arXiv:1112.6426 [hep-ex].
- [92] ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS 2011 data*, Tech. Rep. ATLAS-CONF-2013-004, CERN, Geneva, Jan, 2013.
- [93] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys.Lett. **B659** (2008) 119–126, arXiv:0707.1378 [hep-ph].
- [94] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s}=7$ TeV with the ATLAS detector*, Tech. Rep. arXiv:1406.0076. CERN-PH-EP-2013-222, CERN, Geneva, May, 2014.
- [95] ATLAS Collaboration, *Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data*, Tech. Rep. ATLAS-CONF-2012-020, CERN, Geneva, Mar, 2012.
- [96] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. **C72** (1996) 637–646, arXiv:9601371 [hep-ph].
- [97] 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 at the LHC*, Tech. Rep. ATLAS-CONF-2011-103, CERN, Geneva, Jul, 2011.
- [98] ATLAS Collaboration, G. Aad et al., *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. **B718** (2012) 369–390,

- arXiv:1207.0210 [hep-ex].
- [99] ATLAS Collaboration, G. Aad et al., *Measurement of the Transverse Momentum Distribution of W Bosons in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector*, Phys.Rev. **D85** (2012) 012005, arXiv:1108.6308 [hep-ex].
- [100] ATLAS Collaboration, G. Aad et al., *Study of jets produced in association with a W boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys.Rev. **D85** (2012) 092002, arXiv:1201.1276 [hep-ex].
- [101] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur.Phys.J. **C71** (2011) 1554, arXiv:1007.1727 [physics.data-an].
- [102] A. L. Read, *Presentation of search results: The $CL(s)$ technique*, J.Phys. **G28** (2002) 2693–2704.
- [103] 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*, Tech. Rep. ATLAS-CONF-2012-161, CERN, Geneva, Nov, 2012.
- [104] M. Cacciari, J. Rojo, G. P. Salam, and G. Soyez, *Quantifying the performance of jet definitions for kinematic reconstruction at the LHC*, JHEP **0812** (2008) 032, arXiv:0810.1304 [hep-ph].
- [105] CDF Collaboration, D0 Collaboration, T. Aaltonen et al., *Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron*, Phys.Rev.Lett. **109** (2012) 071804, arXiv:1207.6436 [hep-ex].
- [106] CMS Collaboration, S. Chatrchyan et al., *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*, Phys.Rev. **D89** (2014) 012003, arXiv:1310.3687 [hep-ex].
- [107] 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*, Tech. Rep. CERN-PH-EP-2014-214. arXiv:1409.6212 (submitted to JHEP), CERN, Geneva, Sep, 2014.

- [108] ATLAS Collaboration, G. Aad et al., *Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated W/ZH production with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2013-079, CERN, Geneva, Jul, 2013.
- [109] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.
- [110] Y. Freund and R. E. Schapire, *Experiments with a New Boosting Algorithm*, in *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 148–156. Morgan Kaufmann, Bari, Italy, 1996.
- [111] ATLAS Collaboration, G. Aad et al., *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. **D90** (2014) no. 7, 072004, arXiv:1407.0371 [hep-ex].
- [112] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407** (2014) 079, arXiv:1405.0301 [hep-ph].
- [113] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau, et al., *W and Z/γ^* boson production in association with a bottom-antibottom pair*, JHEP **1109** (2011) 061, arXiv:1106.6019 [hep-ph].
- [114] CMS Collaboration, V. Khachatryan et al., *Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV*, JHEP **1103** (2011) 136, arXiv:1102.3194 [hep-ex].
- [115] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, et al., *New parton distributions for collider physics*, Phys.Rev. **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].
- [116] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur.Phys.J. **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].
- [117] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, et al., *Parton distributions with LHC data*, Nucl.Phys. **B867** (2013) 244–289, arXiv:1207.1303 [hep-ph].

- [118] S. Catani, F. Krauss, R. Kuhn, and B. Webber, *QCD matrix elements + parton showers*, JHEP **0111** (2001) 063, arXiv:hep-ph/0109231 [hep-ph].
- [119] F. Krauss, *Matrix elements and parton showers in hadronic interactions*, JHEP **0208** (2002) 015, arXiv:hep-ph/0205283 [hep-ph].
- [120] Allwood-Spires et al., *Monte Carlo samples used for top physics*, Tech. Rep. ATL-PHYS-INT-2010-132, CERN, Geneva, Dec, 2010.
- [121] J. Pumplin, D. Stump, J. Huston, H. Lai, P. M. Nadolsky, et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **0207** (2002) 012, arXiv:hep-ph/0201195 [hep-ph].
- [122] ATLAS Collaboration, G. Aad et al., *Measurements of the W production cross sections in association with jets with the ATLAS detector*, arXiv:1409.8639 [hep-ex].
- [123] E. Todesco and F. Zimmermann, *Proceedings, EuCARD-AccNet-EuroLumi Workshop: The High-Energy Large Hadron Collider, Malta, Republic of Malta, 14 - 16 Oct 2010*, arXiv:1111.7188 [physics.acc-ph].
- [124] O. Dominguez and F. Zimmermann, *Beam Parameters and Luminosity Time Evolution for an 80-km VHE-LHC*, Tech. Rep. EuCARD-CON-2013-010, 2013.
- [125] CERN, *Higgs cross sections for European Strategy studies in 2012*, <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HiggsEuropeanStrategy2012>, 2012.
- [126] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **1303** (2013) 015, arXiv:1212.3460 [hep-ph].
- [127] ATLAS Collaboration, *Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum using the ATLAS detector in $\sqrt{s} = 8$ TeV proton-proton collisions*, Tech. Rep. ATLAS-CONF-2014-057, CERN, Geneva, Sep, 2014.
- [128] D. Wardrope, E. Jansen, N. Konstantinidis, B. Cooper, R. Falla, et al., *Non-resonant Higgs pair production in the $b\bar{b}b\bar{b}$ final state at the LHC*,

arXiv:1410.2794 [hep-ph].