Study of anomalous Spin/CP components on the Higgs coupling to W pairs with the ATLAS detector

Ricardo Filipe Coelho Barrué

Thesis to obtain the Master of Science Degree in

Engineering Physics

Supervisor(s): Prof. Patricia Conde Muíño

Prof. Rui Alberto Serra Ribeiro dos Santos

Examination Committee

Chairperson: Prof. Jorge Manuel Rodrigues Crispim Romão

Supervisor: Prof. Patricia Conde Muíño

Member of the Committee: Prof. Michelle Gallinaro

April 2020

To those who never stopped wondering...

Acknowledgments

Despite being the sole author of this thesis, it wouldn't have been possible without the help of many people and it's to those people that I would like to extend my deepest appreciation.

First of all, to my supervisors, Rui Santos, for the immensely useful discussions and the great ideas, and Patricia Muiño, for all the opportunities, for being a true mentor, and for believing in my potential, even when I didn't believe in it myself.

To my family, specially my parents and stepparents, for always believing in me and for allowing me to pursue my wildest dreams, and my brother, Bruno, for being my best friend and companion throughout the development and writing of this thesis. To Marta, for her endless love and support, even across an entire ocean. To all my friends, in Odivelas, Lisbon, in other countries and even other continents, for your support in good and bad times. To Dr. Ana Botas, for helping me make sense of the weird mess that was my head, which allowed me to live and work to my full potential.

To professor Paulo Gonçalves, for sparking my love of particle physics. To my colleagues at LIP, for all the helpful and interesting discussions, which have influenced both the ideas and the development of this work. To all my colleagues in the ATLAS analysis, specially Brian Moser, for all the patience in helping a young student understand the intricacies of such a complex analysis, and for all the thoughtful discussions and ideas.

This work has been partially funded by Fundação para a Ciência e Tecnologia, via the projects CERN/FIS-PAR/0008/2017 and CERN/FIS-PAR/0002/2019.

Resumo

Componentes anómalas de Spin/CP no vértice HWW constituiria evidência de física para lá do Standard Model. Mais especificamente, componentes anómalas CP-ímpar, que seriam uma nova fonte de violação de CP, que ajudaria a explicar o valor da assimetria bariónica do universo. Este trabalho estuda a sensibilidade em ATLAS a estas componentes, com uma luminosidade integrada $L = 139\ \mathrm{fb}^{-1}$ a $\sqrt{s} = 13$ TeV. Foca-se no processo $WH \to \ell\nu b\bar{b}, \ell = e, \mu$ no regime de alto momento, $p_{T_W} > 250$ GeV, onde a sensibilidade é maior.

Após otimização da seleção de eventos, a sensibilidade ao sinal do SM e a sinais com componentes anómalas é calculada via medições da força do sinal esperada, μ_{exp}, e da significância esperada, Z_{exp}. Para o sinal do SM, $\mu_{exp} = 1.00 \pm 0.41$ e $Z_{exp} = 2.52\sigma$. Para sinais com componentes anómalas CP-par e para um sinal com uma componente CP-ímpar com acoplamento grande, a força do sinal esperada é incompatível com o valor do SM, com $Z_{exp} > 5\sigma$. Para uma amostra com uma componente CP-ímpar com acoplamento pequeno, $\mu_{exp} = 1.54 \pm 0.42$ e $Z_{exp} = 4.04\sigma$, compatível com o valor medido no SM.

Um conjunto de observáveis angulares sensível às diferentes componentes anómalos é introduzido, mostrando que a assimetria na distribuição do observável $\cos\delta^+$tem um valor de $A(\cos\delta^+) = -0.319 \pm 0.12$ para a amostra com uma componente CP-ímpar com acoplamento pequeno e valor nulo para a amostra do SM e fundos, melhorando a sensibilidade.

Palavras-chave: bosão de Higgs, Spin/CP, violação de CP, teoria de campo efetiva, observáveis angulares

Abstract

Anomalous Spin/CP components in the HWW vertex would be evidence of physics beyond the Standard Model. More specifically, the measurement of anomalous CP-odd components would entail a new source of CP-violation that could, perhaps, explain the observed baryonic asymmetry of the Universe. This work studies the ATLAS sensitivity these components, with an integrated luminosity $L = 139\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 13$ TeV. It targets the $WH \to \ell\nu b\bar{b}, \ell = e, \mu$ process in the boosted regime, $p_{T_W} > 250$ GeV, where the sensitivity is higher.

After optimization of the event selection, the sensitivity to the SM signal and to signals with additional anomalous components was computed, by measuring the expected signal strength, μ_{exp}, and the expected significance, Z_{exp}. For the Standard Model (SM) signal, $\mu_{exp} = 1.00 \pm 0.41$ and $Z_{exp} = 2.52\sigma$. For signals with anomalous CP-even components, and also for a signal with an anomalous CP-odd component with large coupling, the expected signal strengths are not compatible with their SM value, with $Z_{exp} > 5\sigma$. For a signal with an anomalous CP-odd component with smaller coupling, $\mu_{exp} = 1.54 \pm 0.42$ and $Z_{exp} = 4.04\sigma$, compatible with the SM measurement.

A set of angular observables sensitive to the different anomalous components is introduced, showing that the asymmetry in the $\cos\delta^+$observable has a value of $A(\cos\delta^+) = -0.319{\pm}0.12$ for the signal with an anomalous CP-odd component with smaller coupling, and a zero value for SM sample and backgrounds, thus improving the sensitivity.

Keywords: Higgs boson, Spin/CP, CP violation, effective field theory, angular observables

Contents

List of Tables

List of Figures

Glossary

LO Leading Order.

SM Standard Model.

CP Charge-Parity.

ATLAS A Toroidal LHC Apparatus.

LHC Large Hadron Collider.

BR Branching Ratio.

ggF Gluon-gluon fusion.

VBF Vector boson fusion.

BSM Beyond the Standard Model.

2HDM Two Higgs doublet models.

CMS Compact Muon Solenoid.

QCD Quantum Chromodynamics.

ID Inner Detector.

LAr Liquid Argon.

MS Muon Spectrometer.

MC Monte Carlo.

PDF Parton Distribution Function.

NNLO Next to next to Leading Order.

NLO Next to Leading Order.

Chapter 1

Introduction

The Standard Model (SM) is the presently accepted theory in particle physics. Its formulation was done during the 1960s and 1970s and it has been one of the most successfully tested theories in physics ever since. The Higgs boson was discovered in 2012 and was the last SM particle to be discovered, almost fifty years after its prediction, representing an experimental confirmation of the electroweak symmetry breaking mechanism, by which the W and Z bosons acquire mass. After the observation of Higgs boson production and of its coupling to third generation fermions and to W and Z bosons, we are now entering an era of precision measurements of its properties, similar to the precision measurements of the Z boson properties made in the 90s at the Large Electron Positron collider. Any significant deviation from the SM predictions may be a signal of physics beyond it, but so far none have been observed. Thus, it becomes very important to improve the available analysis methods, in order to increase our sensitivity. An interesting area to look for deviations is in the interaction between the Higgs and the W boson, a direct probe into the electroweak symmetry breaking mechanism.

This work concerns the study of the tensor structure of the interaction between the Higgs boson and W boson pairs (HWW), namely the sensitivity to anomalous (not predicted in the SM) Spin/CP components in the interaction vertex, with a focus on CP-odd components. A measurement of such components would entail the existence of CP violation in the Higgs sector and would help physicists understand why the universe contains so much more matter than antimatter. Experimental studies of the Spin and CP properties of the particle discovered in 2012 showed that it was consistent with the Higgs boson predicted in the SM. Furthermore, several studies of the tensor structure of this interaction have shown no significant deviation with regards to the SM.

In this work, the HWW interaction is studied by analyzing the associated production of a W boson and a Higgs boson in the high transverse momentum (boosted) regime, targeting the Higgs decay to a pair of b quarks and the decay of the W boson to an electron or muon and a neutrino. The main objectives for this work are the optimization of the analysis procedure, the study of its sensitivity to the SM signal, and the study of its sensitivity to anomalous Spin/CP components. The latter is done in two ways: first, by calculating the modifications to the expected signal strength, defined as the ratio of the measured signal yield to that of the SM; second, by extending that measurement with a set of angular observables. This work was done in connection with the first ATLAS analysis of associated production of a Higgs and a W or Z boson in the boosted regime - boosted $VH(H \rightarrow b\bar{b})$ analysis - in which the author of this work had an active participation. The contributions to the analysis were the studies of the tagging strategy of the SM Higgs boson in the boosted regime and the optimization of the event selection, both of which are included in this thesis.

In Chapter 2, the theoretical background is presented. In Chapter 3, a description of the state of the art is done, including experimental studies of the Spin/CP properties of the Higgs boson and of the tensor structure of the HWW interaction vertex and a phenomenological study of angular observables to disentangle between the SM component and different anomalous components in the vertex. In Chapter 4, the LHC and the ATLAS detector are described, including a detailed description of each of its subdetectors and also of the trigger system. In Chapter 5, the techniques used to generate events and simulate the interaction of stable particles with the detector are presented, followed by a description of how the events are reconstructed. In Chapter 6, a description of the analysis is done, including also the specifics on the data and simulated samples used in this work, a description of the event selection and the optimization studies carried out, finalizing with the study of the sensitivity of the developed analysis to anomalous Spin/CP components. In Chapter 7, a set of conclusions are drawn and ideas for future work are presented.

Chapter 2

Theoretical background

In this chapter, the Standard Model is presented, starting from a description of its particle content and of its mathematical formulation. Next, the electroweak symmetry breaking mechanism is introduced, along with some of the properties of the Higgs boson, followed by the description of the tensor structure of the SM HWW interaction vertex. It finalizes with a brief description of some of the limitations of the SM and an overview of the effective field theory techniques used in this work to parametrize modifications to the vertex.

2.1 Standard Model

The Standard Model (SM) is the presently accepted theory to describe all known elementary particles and their interactions via the electromagnetic, weak and strong forces [1]. It is written in the language of Quantum Field Theory, where particles are interpreted as local excitations (or *quanta*) of fields. Furthermore, it is formulated as a gauge theory, in which interactions happen via the exchange of other particles called gauge bosons. In the SM, particles are characterized by their charge, mass and spin and are divided in three broad categories: the fermions, which have spin $1/2$; the gauge bosons, which have spin 1; and the Higgs boson, which has spin 0. Figure 2.1 contains a schematic representation of the SM particle content.

The fermions are organized in three generations of quarks and leptons with increasing mass. Each of the generations contains an up- and down-type quark with charges $Q = +2/3$ and $Q = -1/3$, a charged lepton, with charge $Q = -1$, and a neutrino, with charge $Q = 0$.

In the SM, there are five gauge bosons: the gluon, the photon, the $W^{\pm}$ bosons and the Z boson. The gluon is massless, neutral and is the boson of the strong interaction, which acts on particles with *color* charge, such as quarks or even gluons themselves. The photon is massless, neutral and is the boson of the electromagnetic interaction, which acts on electrically charged particles. The $W^{\pm}$ bosons are charged, while the Z boson is neutral. These three particles are massive and are the bosons of the weak interaction, which acts on particles with weak isospin charge and is responsible for phenomena

[1] The problem of formulating a quantum theory of gravitation is one of the most important unsolved problems in modern physics yet it is vastly outside of the scope of this work, so no further mention of it will be done.

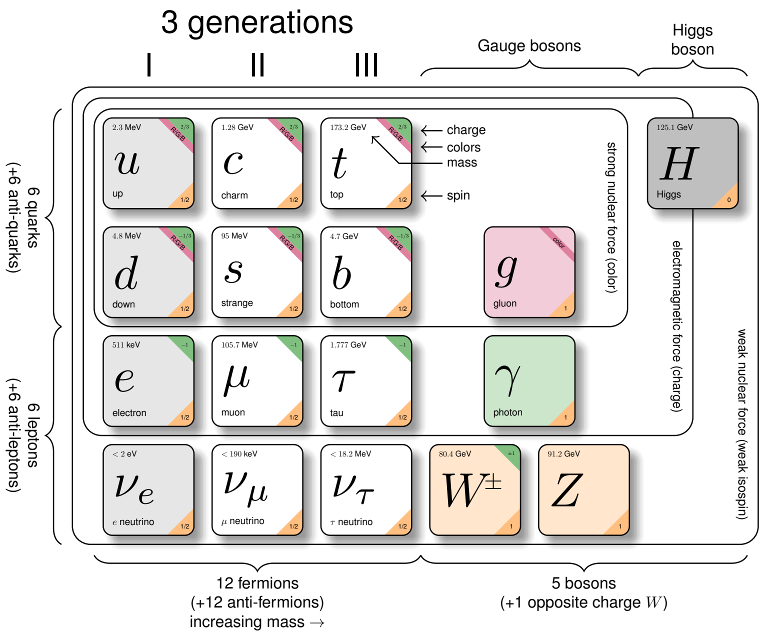


Figure 2.1: Schematic representation of the particle content of the Standard Model.

such as nuclear β decays. The non-zero masses of the W and Z bosons justify the short range of the weak interaction, when compared to the infinite range of the electromagnetic interaction.

The Higgs boson is the quanta of the Higgs field, responsible for the electroweak symmetry breaking mechanism, which gives mass to the W and Z bosons and to the elementary charged fermions. This mechanism was experimentally confirmed with the discovery of a Higgs-like particle in 2012 a mass of approximately 125 GeV [1, 2], later confirmed to be consistent with the SM Higgs boson [3, 4].

A fundamental feature of the SM with estabilished experimental confirmation are antiparticles, which have the same mass as their particle analogues, but with opposite charge and quantum numbers.

Mathematical formulation of the Standard Model

As mentioned previously, the SM is mathematically formulated as a gauge theory. In these theories, the Lagrangian is required to be invariant under local (gauge) transformations defined by a symmetry group. The SM Lagrangian is invariant under transformations of the symmetry group

$$SU(3)_C \times SU(2)_L \times U(1)_Y, \tag{2.1}$$

where C is the *color* charge, L is the left-handed component of weak isospin and Y is the hypercharge. $SU(3)_C$ is the symmetry group of the strong interaction [5], and $SU(2)_L \times U(1)_Y$ is the symmetry group of the electroweak interaction [6], a unification of the electromagnetic and weak interactions. The focus of this work is the study of the interaction between the Higgs boson and W bosons so no further description of the formalism of the strong interaction will be done, although some of its properties are described in Section 5.2.3, on jet reconstruction.

In the electroweak theory, B_μ is the gauge field of the $U(1)_Y$ symmetry and W_μ^a $(a = 1, 2, 3)$ are the gauge fields of the $SU(2)_L$ symmetry. Matter fields are minimally coupled to the gauge fields via the covariant derivative, D_μ, shown in Equation 2.2, following the conventions of [7].

$$D_\mu = \left(\partial_\mu - ig_2 T_a W^a_\mu - ig_1 \frac{Y}{2} B_\mu\right), \tag{2.2}$$

where g_2 and g_1 are the $SU(2)$ and $U(1)$ coupling constants, respectively. $T^a = \frac{1}{2}\tau^a$, where τ^a are the Pauli matrices. The field strength tensors are given in Equations 2.3 and 2.4.

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2\, \epsilon_{abc}\, W^b_\mu W^c_\nu \tag{2.3}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.4}$$

where ϵ^{abc} is the antisymmetric tensor in three dimensions. The Lagrangian of the electroweak sector of the SM for a generic fermion field, ψ, is given in Equation 2.5.

$$\mathcal{L}_{EW} = -\frac{1}{4}\left(W^a_{\mu\nu} W^{\mu\nu}_a + B_{\mu\nu} B^{\mu\nu}\right) + i\overline{\psi}\not{D}\psi, \tag{2.5}$$

where $\psi^\dagger$ denotes the transconjugate field, $\overline{\psi} = \psi^\dagger\gamma^0$ and $\not{D} = D_\mu\gamma^\mu$ (γ^μ are the Dirac matrices). The term in parenthesis is the kinetic term for the gauge fields and the other term encodes the interaction between matter and gauge fields. Mathematically, particles and antiparticles are connected by a combined charge and parity/mirror (CP) transformation, and it was postulated that the electromagnetic, weak and strong interactions were CP-symmetric, i.e. invariant under a CP transformation, until an experiment done in the 1960s proved the existence of CP violation in weak interactions [8].

An issue with the Lagrangian in Equation 2.5 is that it does not include mass terms for the gauge bosons of the type $\frac{1}{2}m_V^2 V_\mu V^\mu (V = W, Z)$. These terms are not invariant under an $SU(2)$ transformation so they can't be added by hand. Furthermore, mass terms for the fermions of the type $-m_f\overline{f}f$ are also not invariant under an $SU(2)$ transformation. In Quantum Electrodynamics, the theory of electromagnetic interactions based on the $U(1)_Q$ symmetry group, a mass term for the photon is also not invariant under $SU(2)$ transformations, but this doesn't pose a problem, since the photon is experimentally estabilished as a massless particle and one can simply set $m_A = 0$. One needs, then, to generate gauge boson and fermion masses in a way that doesn't violate $SU(2)_L \times U(1)_Y$ symmetry, while leaving the photon as a massless particle. This occurs via the electroweak symmetry breaking mechanism, also called the Brout-Englert-Higgs mechanism [9, 10, 11, 12], explained in the next section.

2.1.1 Electroweak symmetry breaking mechanism

The principle behind the electroweak symmetry breaking mechanism is that of spontaneous symmetry breaking, which consists of introducing in the theory a complex field, for which the potential has a non-zero vacuum expectation value and is symmetric for high energies. As the energy in the field decreases, it enters one of infinite stable minima of the potential, which are no longer invariant under the original symmetry, said to be *spontaneously broken*. This is what happens in the electroweak sector of the SM in which the $SU(2)_L \times U(1)_Y$ symmetry is broken, with the vacuum retaining the $U(1)_Q$ symmetry of

Quantum Electrodynamics.

Massless bosons only have two degrees of freedom, corresponding to their transverse polarizations, while massive bosons have three degrees of freedom, from the additional longitudinal polarization. Thus, in order to generate masses for three of the gauge bosons, one needs at least three new degrees of freedom. This is accomplished by introducing a doublet of complex scalar (spin 0) fields, defined in Equation 2.6.

$$\Phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix} \tag{2.6}$$

The most general Lagrangian containing Φ and invariant under $SU(2)$ transformations is given in Equation 2.7, and is divided into a kinetic term, $(D^\mu \Phi)^\dagger (D_\mu \Phi)$ and a potential term, $V(\Phi)$.

$$\begin{aligned} \mathcal{L}_S &= (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi) \\ V(\Phi) &= \mu^2 \left(\Phi^\dagger \Phi\right) + \lambda \left(\Phi^\dagger \Phi\right)^2 \end{aligned} \tag{2.7}$$

In this equation, μ^2 is identified as the mass term and λ as the self-coupling. A condition is that the self-coupling should be positive, for the potential to be bounded from below and for the vacuum to be stable. For $\mu^2 > 0$, $V(\Phi)$ is always positive and the minimum of the potential occurs for $\left(\Phi^\dagger \Phi\right) = 0$. In this case, the system doesn't develop a vacuum expectation value and no symmetry breaking occurs. For $\mu^2 < 0$, the potential has the shape of a 'mexican hat' (see Figure 2.2) and one of the components develops a vacuum expectation value, leading to the desired symmetry breaking.

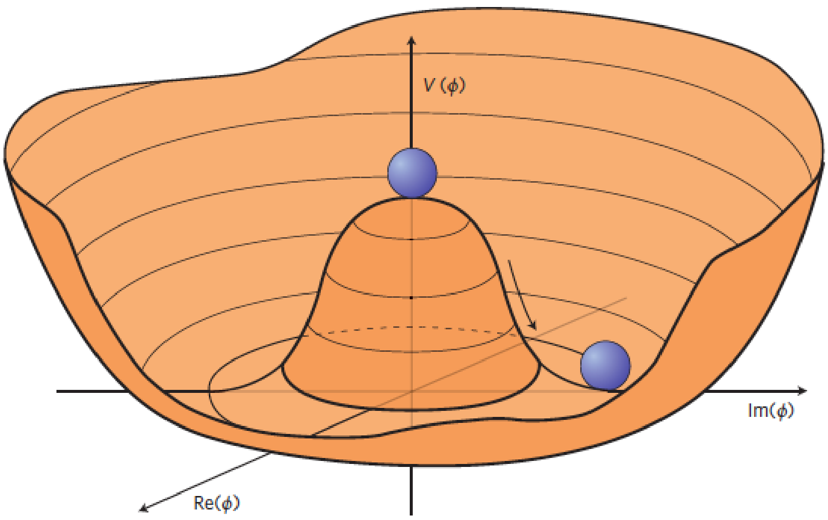


Figure 2.2: Mexican hat potential, with a vacuum expectation value

In this case, there are an infinite number of vacuum states for which the value of the potential is given by Equation 2.8, in which we identify v as the vacuum expectation value.

$$\frac{dV}{d\Phi} = 0 \Leftrightarrow v = \sqrt{\frac{-\mu^2}{\lambda}} \tag{2.8}$$

From the infinite number of vacuum states, the vacuum configuration with $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3 = v$ is chosen, which gives the vacuum state described in Equation 2.9.

$$\langle\Phi\rangle_0 \equiv \langle 0|\Phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.9}$$

The ϕ fields are then written as first order perturbations around their vacuum expectation values and after a gauge transformation where only physical particles are left in the Lagrangian, one gets the state written in Equation 2.10, where $H(x)$ is a scalar field called the Higgs field, mentioned in the introduction of this section.

$$\Phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.10}$$

Replacing this state back in $\mathcal{L}_S$, the term with the covariant derivative becomes as written in Equation 2.11.

$$|D_\mu\Phi|^2 \equiv (D^\mu\Phi)^\dagger (D_\mu\Phi) = \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{8}g_2^2 (v+H)^2 \left|W_\mu^1 + iW_\mu^2\right|^2 + \frac{1}{8}(v+H)^2 \left|g_2 W_\mu^3 - g_1 B_\mu\right|^2 \tag{2.11}$$

From this, one defines new fields, corresponding to the physical $W_\mu^\pm$ and Z_μ bosons and the photon, A_μ, given in Equation 2.12. Expanding the term with the covariant derivative gives the result shown in Equation 2.13.

$$W_\mu^\pm = \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}}, \quad Z_\mu = \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_1^2 + g_2^2}}, \quad A_\mu = \frac{g_1 W_\mu^3 + g_2 B_\mu}{\sqrt{g_1^2 + g_2^2}} \tag{2.12}$$

$$|D_\mu\Phi|^2 = \frac{1}{2}(\partial_\mu H)^2 + \left(\frac{1}{4}g_2^2 v^2\right)\left(W_\mu^+ W^{-\mu}\right) + \left(\frac{1}{8}v^2\left(g_1^2 + g_2^2\right)\right) Z_\mu Z^\mu + 0 \times A_\mu A^\mu + [...] \tag{2.13}$$

Identifying terms bilinear in the fields, $m_W^2 W_\mu^+ W^{-\mu}$, $\frac{1}{2}m_Z^2 Z_\mu Z^\mu$, $\frac{1}{2}m_A^2 A_\mu A^\mu$, one obtains for the masses of the gauge bosons

$$m_{W^\pm} = \frac{1}{2}g_2 v, \quad m_Z = \frac{1}{2}v\sqrt{g_1^2 + g_2^2}, \quad m_A = 0$$

From the relation between the W boson mass and the Fermi decay constant G_μ (extracted from muon decay), the numeric value of v can be extracted, as shown in Equation 2.14.

$$m_W \approx \left(\frac{\sqrt{2}g_2^2}{8G_\mu}\right)^{1/2} \Rightarrow v = \frac{1}{\left(\sqrt{2}G_\mu\right)^{1/2}} \approx 246 \text{ GeV} \tag{2.14}$$

One can use the same doublet to generate fermion masses in a way that doesn't violate $SU(2) \times U(1)$ symmetry, via Yukawa interactions. This leads to mass terms for the fermions of the type $m_f = \frac{y_f v}{\sqrt{2}}$, where y_f is called the Yukawa coupling of fermion f, leading also to interactions between the Higgs field and the fermions proportional to their mass. The Lagrangian of the electroweak sector of the SM after inclusion of the Higgs doublet, is written in Equation 2.15.

$$\mathcal{L} = -\frac{1}{4}\left(W^a_{\mu\nu}W^{\mu\nu}_a + B_{\mu\nu}B^{\mu\nu}\right) + i\overline{\psi}\not{D}\psi + \left(D^\mu\Phi\right)^\dagger\left(D_\mu\Phi\right) - V\left(\Phi\right) + \overline{\psi}_{L_i}y_{ij}\Phi\psi_{R_j} \tag{2.15}$$

Higgs boson production cross-sections and decay branching ratios

As mentioned previously, the Higgs boson is the *quanta* of the Higgs field. Its production cross-sections (σ) in proton-proton collisions at a center-of-mass energy $\sqrt{s} = 13$ TeV and its decay branching ratios (BR), both for a Higgs mass of 125 GeV, are shown in Tables 2.1 and 2.2, respectively.

Production Channel	σ (pb)
Gluon-gluon fusion (ggF)	48.58
Vector boson fusion (VBF)	3.78
WH	1.37
ZH	0.88
$t\bar{t}H$	0.51
$b\bar{b}H$	0.49

Table 2.1: Higgs boson production cross-sections in proton-proton collisions at $\sqrt{s} = 13$ TeV, for a Higgs mass of 125 GeV. Source: [13].

Decay Channel	$BR(\%)$
$b\bar{b}$	58.24
WW^*	21.37
gg	8.19
$\tau^+\tau^-$	6.27
$c\bar{c}$	2.89
ZZ^*	2.62
$\gamma\gamma$	0.23
$Z\gamma$	0.15
$\mu^+\mu^-$	0.02

Table 2.2: Higgs boson decay branching ratios for a Higgs mass of 125 GeV. Source: [13].

2.1.2 HWW vertex

This work studies the interaction between the Higgs boson and pairs of W bosons (HWW), encoded in the tensor structure of the HWW interaction vertex. From the expansion of $|D_\mu\Phi|^2$ in Equation 2.11, one identifies the term $\left(\frac{1}{2}g_2^2 v\right)\left(HW^+_\mu W^{-\mu}\right) = \left(g_2 m_W\right)\left(HW^+_\mu W^{-\mu}\right)$, leading to a coupling $g_{HWW} = g_2 m_W$. The SM HWW interaction vertex is written in Equation 2.16.

$$i\Gamma^{\mu\nu}_{HWW} = ig_{HWW}g^{\mu\nu} = i\left(g_2 m_W\right)g^{\mu\nu}, \tag{2.16}$$

where $g^{\mu\nu} = diag\left(+,-,-,-\right)$ is the Minkowski metric.

In order to extract information on this vertex, several production and decay channels can be studied. At production level, one can study either WH associated production or VBF production (both schematized in Figure 2.3), while at decay level, one can study the $H \rightarrow WW^*$ decay. Different channels will have advantages and disadvantages, and will also require different reconstruction and selection criteria, which will in turn lead to different systematics.

VBF production has the advantage of having a very distinct topology, containing two light quarks which are reconstructed as two jets very far apart in the detector. Despite this, this process is not well

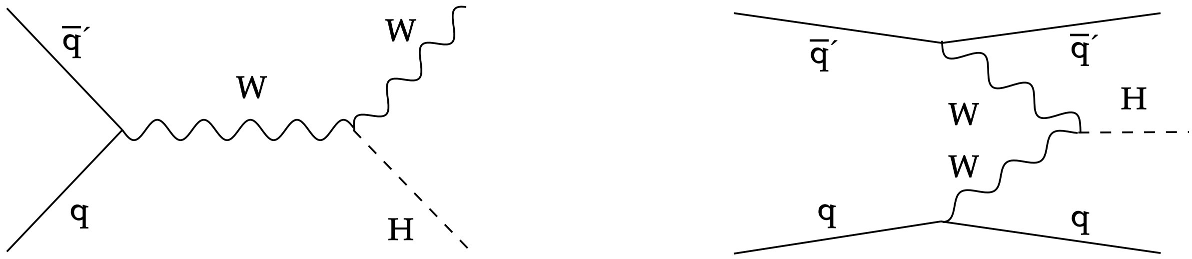


Figure 2.3: Leading order (LO) Feynman diagrams for WH associated production (left) and VBF production (right)

suited to study the HWW vertex independently of the HZZ vertex, due to the fact that it is very hard to disentangle between W and Z boson fusion, which have the same final state topology.

WH associated production has the advantage of allowing to probe the HWW vertex independently of the HZZ vertex, since one can identify the charge of the vector boson from the total charge of its decay products. The main disadvantage is that, even when targeting the $H \to b\bar{b}$ decay, the product of the cross-section and branching ratio is very low, $\sigma \times BR = 0.80\,\mathrm{pb}$ ($= 2.2\,\mathrm{pb}$ for VBF).

For the $H \to WW^*$ decay, targeting the ggF production channel leads to $\sigma \times BR = 10.38\,\mathrm{pb}$, one order of magnitude larger than other decay and production channels. Nonetheless, it is very hard to do full kinematic reconstruction of the event since, if one targets looks for events where at least one of the W bosons decays to quarks, the contamination from background processes becomes too large. On the other hand, if one targets events where the two W bosons decay to a charged lepton and a neutrino, it becomes impossible to reconstruct the Higgs mass and identify the transverse momenta of each of the neutrinos.

2.1.3 Limitations of the Standard Model

Despite its tremendous success in predicting and describing experimental observations, there are still a number of them for which the SM does not have an answer: it predicts massless neutrinos, disproven by the discovery of neutrino oscillations [14, 15, 16], which implies the neutrino have masses, with the origin of their masses still unknown; it does not have any natural candidate for dark matter, which interacts very weakly via the SM forces, but is thought to account for around 80% of the matter in the universe; of relevance for this work, it does not explain the observed imbalance between matter and antimatter in the Universe, the so-called baryonic asymmetry. This imbalance requires not only the existence of CP violation [17], but also a larger amount than that predicted in the SM [18], which appears in the quark sector, from complex phases in the Cabbibo-Kobayashi-Maskawa matrix, which relates weak eigenstates (participate in the weak interaction) with mass eigenstates (solutions of the equations of motion). In order to explain this imbalance, it is necessary to look for additional sources of CP violation beyond the SM.

There are several theories beyond the Standard Model (BSM) with additional sources of CP violation, such as two Higgs doublet models (2HDM) [19], which extend the SM with an additional complex Higgs doublet, allowing for a source of CP violation in the potential. Current experimental results have shown no evidence of new particles being discovered so it may happen that, if these models are realized in Nature, the energies at which new particles are produced are higher than those experimentally accessible today. Nonetheless, there may be some small effect at the energies explored today, e.g. through loops involving these new particles. An approach to studying these effects is effective field theory techniques, described below.

2.1.4 Effective Field Theory

Effective field theory techniques [20] are a model-independent approach to parametrizing modifications to the SM, which consist of extending the SM Lagrangian with higher-dimensional (mass dimension > 4) operators built from combinations of SM fields, supressed by increasing powers of the new physics scale, Λ. These operators come from integrating the various degrees of freedom of a possibly unknown model and can be consistently matched to operators in theories with new particles at higher masses, such as 2HDM [21]. Grzadkowski et al. [22] derived a set of 59 independent dimension-6 operators that can be added to the SM Lagrangian [2], called the 'Warsaw basis'. Of interest to this work are two dimension-6 operators with different CP properties: one CP-even, which doesn't change sign when a CP transformation operates on it, and one CP-odd, which changes sign when a CP transformation operates on it, both described in Equations 2.17 and 2.18.

$$\mathcal{O}_{WW} = \frac{g_2^2 b_{WW}}{4\Lambda^2}\Phi^\dagger\Phi W^a_{\mu\nu}W^{a\mu\nu} \quad \text{(CP-even)} \tag{2.17}$$

$$\tilde{\mathcal{O}}_{WW} = \frac{g_2^2 c_{WW}}{4\Lambda^2}\Phi^\dagger\Phi W^a_{\mu\nu}\tilde{W}^{a\mu\nu}, \quad \text{(CP-odd)} \tag{2.18}$$

where $\tilde{W}^{\mu\nu}$ is the dual field tensor, given by $\frac{1}{2}\epsilon^{\mu\nu\rho\sigma}W_{\rho\sigma}$, where $\epsilon^{\mu\nu\rho\sigma}$ is the (CP-odd) Levi-Civita tensor, and b_{WW} and c_{WW} are the Wilson coefficients, that parametrize the relative strength of the different operators. These operators are then added to the SM Lagrangian, leading to the modified HWW interaction vertex in Equation 2.19, following the parametrization of Ref. [23].

$$i\Gamma^{\mu\nu}_{HWW}(k_1,k_2) = i(g_2 m_W)\left[g^{\mu\nu}\left(1+a_W-\frac{b_{W1}}{m_W^2}(k_1\cdot k_2)\right)+\frac{b_{W1}}{m_W^2}\,k_1^\nu k_2^\mu+\frac{c_W}{m_W^2}\epsilon^{\mu\nu\rho\sigma}k_{1_\rho}k_{2_\sigma}\right], \tag{2.19}$$

where $(1+a_W)$ is a scaling factor to allow possible rescalings of the SM contribution. $b_{W1} = \frac{2m_W^2 b_{WW}}{\Lambda^2}$ and $c_W = \frac{2m_W^2 c_{WW}}{\Lambda^2}$ are functions of the Wilson coefficients, defined in a way to remove Λ from the set of independent variables, and k_1 and k_2 are the gauge boson momenta. Here, 2 anomalous components are identified, 1 CP-even with coupling b_{W1} and 1 CP-odd, with coupling c_W. These lead to modifications to the cross-sections of processes such as VBF or WH production, and also to the shape of kinematic distributions, such as the transverse momenta, which are pushed to higher values, due to the additional

[2] The only dimension 5 operator that is gauge-invariant is connected to neutrino oscillations, since it violates lepton number conservation, so no further mention of them will be made.

momentum factors in the vertex. Thus, analysis targeting higher transverse momentum regimes will, in principle, have increased sensitivity to these components.

Chapter 3

State of The Art

In 2012, the ATLAS and CMS collaborations announced the discovery of a Higgs-like particle at a mass of around 125 GeV [1, 2]. In order to confirm if the discovered particle was the SM Higgs boson, it was necessary to measure several of its properties such as its Spin and CP numbers, predicted to be $J^{CP} = 0^{++}$, and the strength of its interaction with other SM particles.

This chapter contains an overview of some of the recent studies relevant to the topic of this thesis and is divided in an experimental and phenomenological section. In the experimental section, it includes the latest studies of $H \to WW^*$ decays, the observation and measurement of VH associated production $(V = W, Z)$, studies of the spin and CP properties of the Higgs boson, and studies of the tensor structure of the HVV interaction vertex. The phenomenological section includes an overview of the latest results on global effective field theory fits to SM and Higgs data, along with the results of a study of angular observables to distinguish between different anomalous components in the HVV vertex.

3.1 Experimental studies

3.1.1 Study of the $H \to WW^*$ decay

The ATLAS Collaboration performed measurements of the Higgs decay into pairs of W bosons in ggF and VBF production [24] as well as in associated VH production [25], with an integrated luminosity $L = 36.1\ \mathrm{fb}^{-1}$ at a center-of-mass energy $\sqrt{s} = 13$ TeV. The signal strengths were measured to be $\mu_{ggF} = 1.10 \pm 0.21$ and $\mu_{VBF} = 0.62 \pm 0.36$ for ggF and VBF production, and $\mu_{WH} = 2.3^{+1.2}_{-1.0}$ and $\mu_{ZH} = 2.9^{+1.9}_{-1.3}$ for associated WH and ZH production. These measurements are consistent with the SM predictions within uncertainties, with the ggF measurement having a significance of 6.0σ, higher than the 5σ discovery threshold, while the other measurements have a significance lower than 5σ.

The CMS Collaboration performed a similar measurement, combining the three production channels mentioned above in a single analysis [26], with an integrated luminosity $L = 35.9\ \ \mathrm{fb}^{-1}$ at $\sqrt{s} = 13$ TeV. A fit was made to the $H \to WW^*$ signal strength, combining the four production mechanisms described previously and assuming their relative proportions were those predicted in the SM, measuring a signal strength $\mu = 1.28^{+0.18}_{-0.17}$, with a significance of 9σ. In the same analysis, further measurements were carried out in the framework of Simplified Template Cross-Sections, where one cross-section was measured for each production mechanism, for a generator level Higgs rapidity $y_H < 2.5$. The obtained cross-sections were then used to place constraints on the signal strength modifiers μ_f, which scale fermion-induced production mechanisms such as ggF, and μ_V, which scales vector boson-induced production mechanisms such as VBF. The results were $\mu_f = 1.37^{+0.21}_{-0.20}$ and $\mu_V = 0.78^{+0.60}_{-0.57}$, and were used to place constraints on coupling modifiers in the k-framework, which parametrize the signal strength as shown in Equation 3.1.

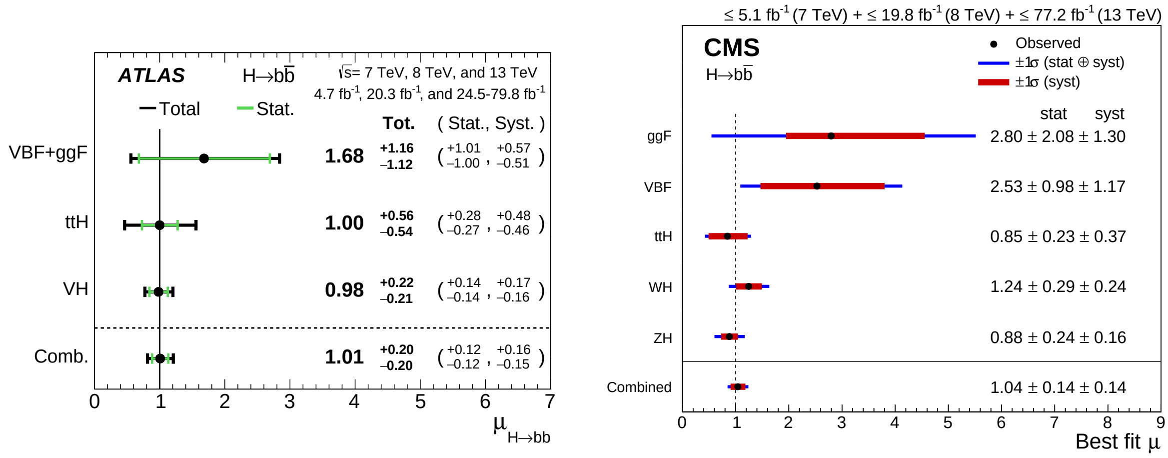
$$\mu = k_i^2 \frac{k_V^2}{k_H^2}, \tag{3.1}$$

where $k_i = k_f$ for fermion-induced production modes, $k_i = k_V$ for vector-boson-induced production modes and $k_H = k_H(k_f, k_V)$ is the global Higgs width modifier. The results were $k_f = 1.52^{+0.48}_{-0.41}$ and $k_V = 1.10 \pm 0.08$, compatible with their SM predictions, $k_V = k_f = 1$, within uncertainties.

3.1.2 Observation of VH associated production

In 2018, the ATLAS and CMS collaborations performed studies of the Higgs decay to b quarks in VH associated production at $\sqrt{s} = 13$ TeV [27, 28] with 79.8 fb^{-1} and 77.2 fb^{-1} of integrated luminosity, respectively, and targeting the decay of the vector bosons to $Z \to \nu\nu$, $W \to \ell\nu$ or $Z \to \ell\ell$, with $\ell = e, \mu$. A set of event selection cuts were imposed to reduce contamination from the main backgrounds, such as $t\bar{t}$ and associated production of a vector boson and b- or c-jets, and control regions were defined, in order to reduce the uncertainty on their normalization. To increase the sensitivity to the Higgs boson signal, these analyses resorted to multivariate analysis methods, which combine the kinematic information of the events to increase signal and background separation. These methods were successfully validated in a measurement of the $VZ(Z \to b\bar{b})$ signal strength, which agreed with the SM prediction. As a cross-check, cut-based analyses of the $VH(H \to b\bar{b})$ process were also carried out, using the invariant mass of the b-tagged jet pair as input to the signal extraction fit, and the results were shown to be compatible. The results of these analyses were combined with those from analyses done in Run 1 of the LHC at $\sqrt{s} = 7$ and 8 TeV and the signal strength measurements were compatible with their SM values within uncertainties, with significances very close to 5σ. These results were further combined with results of previous searches for the Higgs boson decaying into a pair of b quarks in VBF and $t\bar{t}H$ production, leading to the first observation of the Higgs decay to b quarks, with a significance above 5σ. The signal strengths obtained for the different production mechanisms and for their combination are shown in Figure 3.1 and were compatible with the SM within the uncertainties.

In the same analysis, the ATLAS Collaboration combined the results of the $VH(H \to b\bar{b})$ search with other searches for VH associated production in the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ decay channels done with a total integrated luminosity $L = 80 \text{ fb}^{-1}$. The signal strengths obtained for the different decay channels and for their combination are shown in Figure 3.2 and were compatible with their SM values. The observed significance of this measurement was larger than 5σ, corresponding to the first



(a) (b)

Figure 3.1: $H \to b\bar{b}$ signal strengths for the combination of Run 1 and Run 2 data in the ATLAS (left) and CMS (right) analysis. Source: [27] (left) and [28] (right)

observation of the associated VH production process.

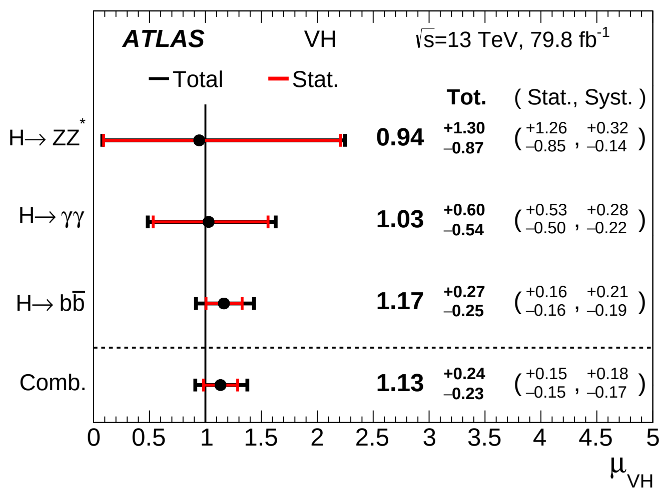


Figure 3.2: VH signal strengths for the combination of Run 2 searches in the ATLAS $VH(H \to b\bar{b})$ analysis. Source: [27]

3.1.3 Higgs Spin-CP studies

After the end of Run 1 of the LHC, the ATLAS and CMS collaborations performed measurements of the spin and CP quantum numbers of the Higgs-like particle discovered in 2012 [3, 4]. These were done in the $H \to WW^* \to e\nu\mu\nu$, $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ decay channels at $\sqrt{s} = 7$ and 8 TeV. This study compared the SM Higgs boson spin-CP hypothesis, $J^P = 0^+$, with three others: $J^P = 0^-$, which corresponds to a CP-odd (pseudoscalar) state; $J^P = 0^+_h$, which corresponds to a BSM spin-0 CP-even state; and $J^P = 2^+$, which corresponds to a graviton-like spin-2 state. The Landau-Yang theorem [29, 30] affirms that a spin-1 assignment is not possible, due to the observed decay of this particle to a pair of photons. A set of observables sensitive to the different spin-CP assignments was developed and fitted to the data under the several hypothesis, excluding all the tested alternative hypothesis in favour of the SM one at a confidence level higher than 99.9%. These results confirmed that the particle discovered in 2012 was the SM Higgs boson and motivated the Nobel Prize for Peter Higgs and François Englert in 2013.

3.1.4 HWW tensor structure studies

After the measurement of the spin and CP quantum numbers of the Higgs boson, one of the next steps was to measure precisely the strength of its interaction with the SM particles, as well as the tensor structure of those interactions. The latter is generally done using effective field theory techniques in a number of different ways: either by constraining directly the couplings at the level of the Lagrangian (as is generally done by the ATLAS Collaboration) or at the level of the scaterring amplitudes (as is generally done by the CMS Collaboration). Despite this apparent duality, it is possible to map results from one formalism to another.

The CMS collaboration performed a study of the tensor structure of the HVV interaction, with $VV = WW, ZZ, Z\gamma, \gamma\gamma$ and gg [31], using 35.9 fb^{-1} of integrated luminosity at $\sqrt{s} = 13$ TeV, with the Higgs boson decaying to a pair of τ leptons, in several production channels, but focusing mainly on VBF production. Anomalous interactions were parametrized by the amplitude in Equation 3.2.

$$A(HVV) = \left[a_1^{VV} + \frac{k_1^{VV} q_1^2 + k_2^{VV} q_2^2}{\left(\Lambda_1^{VV}\right)^2}\right] m_{V1}^2 \epsilon_{V1}^* \epsilon_{V2}^* + a_2^{VV} f_{\mu\nu}^{*(1)} f^{*(2)\mu\nu} + a_3^{VV} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2)\mu\nu}, \tag{3.2}$$

where q_i, ϵ_{Vi} $(i = 1, 2)$ are the four-momentum and polarization vector of the vector boson, respectively, and m_{V1} is the pole mass of the vector boson. $f^{*(i)\mu\nu}$ and $\tilde{f}^{*(i)\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} f_{\rho\sigma}^{*(i)}$ are the vector boson's field strength tensor and dual field strength tensor, respectively. The Leading Order (LO) SM couplings for W and Z bosons are assumed equal, $a_1^{WW} = a_1^{ZZ}$, as well as the couplings of the anomalous components, $a_2^{WW} = a_2^{ZZ}, a_3^{WW} = a_3^{ZZ}$. The measured quantities were the effective cross-section ratios and phases, defined as:

$$f_{ai} = \frac{|a_i|^2 \sigma_i}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + \cdots}, \phi_{ai} = arg\left(\frac{a_i}{a_1}\right), \tag{3.3}$$

where σ_i is the cross-section calculated with $a_i \neq 0$ and all others set to zero. $\phi_{a_i} = 0, \pi$ for the Lagrangian to be Hermitian.

The matrix element likelihood approach [32, 33] was used to build two discriminants sensitive to the tensor structure of the HVV interaction and the results from this analysis are combined with those from the $H \to ZZ^* \to 4l$ analysis, assuming that the HWW and HZZ couplings are the same. The results, shown in Table 3.1, were compatible with their SM predictions.

The ATLAS Collaboration performed measurements of the $VH(H \to b\bar{b})$ process as a function of the vector boson transverse momentum, p_{T_V}, in the Simplified Template Cross Section framework [34], using the same dataset as in the $VH(H \to b\bar{b})$ analysis previously described [27], but extending it

Parameter	Observed / (10^{-3})		Expected / (10^{-3})	
	68% CL	95% CL	68% CL	95% CL
$f_{a3}\cos(\phi_{a3})$	0.00 ± 0.27	$[-92, 14]$	0.00 ± 0.23	$[-1.2, 1.2]$
$f_{a2}\cos(\phi_{a2})$	$0.08^{+1.04}_{-0.21}$	$[-1.1, 3.4]$	$0.0^{+1.3}_{-1.1}$	$[-4.0, 4.2]$
$f_{\Lambda 1}\cos(\phi_{\Lambda 1})$	$0.00^{+0.53}_{-0.09}$	$[-0.4, 1.8]$	$0.00^{+0.48}_{-0.12}$	$[-0.5, 1.7]$
$f_{\Lambda 1}^{Z\gamma}\cos(\phi_{\Lambda 1}^{Z\gamma})$	$0.0^{+1.1}_{-1.3}$	$[-6.5, 5.7]$	$0.0^{+2.6}_{-3.6}$	$[-11, 8.0]$

Table 3.1: Observed and expected limits on effective cross-sections and phases from the combination of the $H \to \tau\tau$ and $H \to ZZ^* \to 4l$ channels. Source: [31]

with differential measurements. The measured cross-sections were compatible with their SM values within the uncertainties and were used to place constraints on the Wilson coefficients, c_i, of 5 CP-even dimension-6 operators, using known (polynomial) relations between the Simplified Template Cross Section signal strengths and those coefficients. These are based on LO predictions in the Strong Interacting Light Higgs basis, including the interference term between the SM and anomalous amplitudes, $\propto \frac{c_i}{\Lambda^2}$, and also the anomalous quadratic term, $\propto \frac{c_i^2}{\Lambda^4}$. In order to perform these constraints, the five measured signal strengths were fit simultaneously, once for each of the operators of interest, setting the other coefficients to zero. Fits in which the quadratic terms are neglected were also carried out (named 'interference only'). The results were compatible with their SM predictions within uncertainties and it was seen that for the measurements neglecting quadratic terms, the limits are tighter.

3.2 Phenomenological studies

3.2.1 Global effective field theory fits

Ellis *et al.* [35] performed a global fit to a dataset which included: electroweak and diboson precision data from the Large Electron-Positron collider at CERN, e.g. Z-pole observables; Higgs production data from Run 1 and 2 of the LHC at $\sqrt{s} = 7$, 8 and 13 TeV and also a single bin of the differential cross-section of the $pp \to WW \to e\nu\mu\nu$ process at $\sqrt{s} = 13$ TeV, corresponding to a transverse momentum of the leading (highest transverse momentum) lepton larger than 120 GeV. This study took into account 20 dimension-6 CP-even operators in the Strong Interacting Light Higgs and Warsaw basis, neglecting CP-odd operators, and the expansion in the Wilson coefficients was done up to first order, which corresponds to keeping only the interference terms. Two different Wilson coefficients were extracted for each operator: one where the coefficients for all the other operators were set to zero, and other where all the coefficients were allowed to vary simultaneously. The bounds on the coefficients showed no evidence of BSM physics. Different constraints on UV-complete models such as the Type-I 2HDM were also derived from the obtained coefficients. Studies similar to the this one for CP-odd operators are still a long ways of reaching similar levels of precision, mainly due to the fact that CP-odd components are expected to introduce very small changes to the inclusive cross-sections.

3.2.2 HWW tensor structure determination using angular observables

A study of angular observables sensitive to anomalous components in the HVV vertex was performed in Ref. [23], taking into account three dimension-6 operators relevant to this interaction, two CP-even and one CP-odd. It exploits the fact that the different anomalous Spin/CP components in the HVV vertex will have different effects on the angular distributions of the decay products of the vector boson. This study was done in the VH production channel, with $H \to b\bar{b}$ and $W \to \ell\nu$ or $Z \to \ell\ell$, at a center-of-mass energy of $\sqrt{s} = 14$ TeV and focused on the boosted regime, not only to maximize sensitivity to anomalous components but also to reduce the QCD-induced backgrounds. In this regime, the b-quarks from the Higgs decay are very collimated and standard jet reconstruction and b-tagging techniques become ineffective, requiring the use of jet substructure techniques, such as the mass-drop technique [36] to reconstruct the Higgs boson. Detector simulation was done using `Delphes` [37] with a modified version of the CMS detector parameters.

It was observed that, as expected, the anomalous components push the transverse momentum distribution of the W and Higgs bosons to higher values (motivating the use of boosted reconstruction techniques) and that the selection cuts lead to a better acceptance for BSM physics. Also, it shows that, since the interference between the SM component and the anomalous CP-odd component contributes very little to the cross-section, its variation after selection cuts was much smaller for a non-zero CP-odd component than for a non-zero CP-even component, for the same value of the coupling. The set of angular observables defined in this work was built to allow distinction between SM and anomalous components and also between different anomalous components, with the added feature of being linearly sensitive to the coupling of the CP-odd component. The definition of these variables is the following:

$$\cos\theta^* = \frac{\mathbf{p}_{l_1}^{(V)} \cdot \mathbf{p}_V}{|\mathbf{p}_{l_1}^{(V)}||\mathbf{p}_V|} \tag{3.4}$$

$$\cos\delta^+ = \frac{\mathbf{p}_{l_1}^{(V)} \cdot (\mathbf{p}_h \times \mathbf{p}_V)}{|\mathbf{p}_{l_1}^{(V)}||\mathbf{p}_h \times \mathbf{p}_V|} \tag{3.5}$$

$$\cos\delta^- = \frac{\left(\mathbf{p}_{l_1}^{(h-)} \times \mathbf{p}_{l_2}^{(h-)}\right) \cdot \mathbf{p}_V}{|\mathbf{p}_{l_1}^{(h-)} \times \mathbf{p}_{l_2}^{(h-)}||\mathbf{p}_V|} \tag{3.6}$$

$$\Delta\phi^{lV} = \Delta\phi\left(\mathbf{p}_{l_1}^{(V)}, \mathbf{p}_V\right), \tag{3.7}$$

where p_{l_1} corresponds to the momentum of the charged lepton and p_{l_2} to the momentum of the neutrino, in WH production, and corresponding to the momenta of the leading and subleading charged leptons in ZH production. $\mathbf{p}_l^{(V)}$ is the momentum of the lepton in the vector boson rest frame and $\mathbf{p}_l^{(h-)}$ is the 3-momentum of the lepton in the rest frame of a Higgs boson with $\mathbf{p} = -\mathbf{p}_h$. A likelihood analysis was performed in order to quantify the luminosity required to exclude the BSM hypothesis at the 95% confidence level and at 3σ level, for different anomalous components. For the SM supplemented with an anomalous CP-odd component with coupling $\tilde{c}_W = 0.1$ (see Equation 2.19), it showed that the strongest discriminants were the angular variables $\cos\delta^+$and $\cos\delta^-$, allowing exclusion of this hypothesis at 95% confidence level with 100 fb^{-1} of luminosity.

Using the fact that the angular observables were defined to be linearly sensitive to the coupling of the CP-odd component, an asymmetry parameter based on the $\cos\delta^\pm$ variables was calculated to probe its value. For W^+ events (tagged using the sign of the decay lepton), it was defined as shown in Equation

3.8.

$$A\left(\cos\delta^{\pm}\right)=\frac{\sigma\left(\cos\delta^{\pm}>0\right)-\sigma\left(\cos\delta^{\pm}<0\right)}{\sigma\left(\cos\delta^{\pm}>0\right)+\sigma\left(\cos\delta^{\pm}<0\right)}, \tag{3.8}$$

defining minus this quantity for W^- events. For the SM supplemented with a CP-odd component $(a_W = 0, c_W = 0.1)$, the value of the asymmetry calculated using $\cos\delta^+$after including detector effects and applying all selection cuts was $A(\cos\delta^+) = 0.315$. They also verified that for the SM case, the value of the asymmetry is compatible with 0, concluding that a significant non-zero measurement of this asymmetry most likely comes from BSM CP-odd components in the vertex.

Chapter 4

Experimental Apparatus

In order to test the Higgs sector and search for BSM physics, it is necessary to build very powerful colliders, with enough energy to produce Higgs bosons or other unknown particles that may exist at higher masses, and also build dedicated detectors that are able to efficiently reconstruct and measure the products of those collisions. This chapter describes the experimental setting in which this work was developed: the Large Hadron Collider (LHC), which accelerates and collides protons at very high energies, and the ATLAS experiment, used to detect and measure the products of these collisons.

4.1 Large Hadron Collider

The Large Hadron Collider [38] is the most powerful particle accelerator in the world and is based at CERN (European Organization for Nuclear Research), in Geneva, Switzerland. It is a 27 kilometer ring under the French-Swiss border, which uses a set of radiofrequency cavities to accelerate two beams of protons in opposite directions, up to an energy of 6.5 TeV each. The proton beams have a bunched structure (as opposed to being continuous), and are nominally composed of around 2800 bunches with 1.15×10^{11} protons each, which are bent around the ring by a series of 8.3 T dipoles and focused by quadrupoles and other multipole magnets. After reaching the desired energy, a set of magnets directs the proton beams onto each other, colliding them at a frequency f_{coll} = 40 MHz in four points, where dedicated experiments such as ATLAS record the events, i.e. the result of all the proton-proton collisions that occur in a single bunch crossing.

Luminosity and center-of-mass energy

An accelerator is mainly defined by two quantities: its instantaneous luminosity and center-of-mass energy. The instantaneous luminosity is related to the rate of occurrence of any desired process and is defined in Equation 4.1.

$$\mathcal{L} = f_{coll} \frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \,\mathrm{cm}^{-2}\mathrm{s}^{-1}, \tag{4.1}$$

where $n_1, n_2 \approx 10^{11}$ are the number of particles in each bunch and $\sigma_x, \sigma_y \approx 0.2$ mm are the transverse dimensions of the beams, assuming they are Gaussian. The LHC was designed to run at an instantaneous luminosity of 10^{34} cm^{-2}s^{-1}.

In any physics analysis, the relevant quantity is the integrated luminosity, defined as $L = \int \mathcal{L}\, dt$, usually measured in $\text{fb}^{-1} = 10^{39}\,\text{cm}^{-2}$. The center-of-mass energy, $\sqrt{s}$, directly affects the cross-section of any specific process, σ, which will in turn affect the number of occurrences of a specific process over a time period, $N = L\sigma$.

The large instantaneous luminosities needed to study rare processes, such as Higgs boson production, require a large number of simultaneous proton-proton interactions in the center of the experiments, called pile-up. For the current LHC operating conditions, the number of simultaneous interactions, μ, is generally between 30 and 40. Table 4.1 contains the values of some of the operational parameters for the two LHC runs: center-of-mass energy; collected integrated luminosity; average number of simultaneous interactions; maximum instantaneous luminosity, $\mathcal{L}_{peak}$; and maximum number of simultaneous interactions, μ_{peak}.

LHC Run	Years	$\sqrt{s}$ (TeV)	L (fb^{-1})	$\langle\mu\rangle$	$\mathcal{L}_{peak}$ (cm^{-2} s^{-1})	μ_{peak}
Run 1	2011-2012	7 (2011), 8 (2012)	5.5 (7 TeV), 22.8 (8 TeV)	14.9	7.50 $\times 10^{33}$	35.9
Run 2	2015-2018	13	156.0	33.7	2.09 $\times 10^{34}$	78.5

Table 4.1: LHC operating runs and parameters.

4.2 The ATLAS detector

The ATLAS detector [39] is a general-purpose apparatus with a forward-backward symmetric cylindrical geometry and with an almost 4π coverage of the solid angle around the collision point. It possesses an onion-like structure with a magnet system and different subdetectors and is shown in Figure 4.1.

4.2.1 Detector coordinates

The reference system used in the ATLAS detector is a right-handed coordinate system with its origin on the interaction point, the x direction pointing at the center of the LHC, the y direction pointing upwards and the z (longitudinal) direction pointing along the beam pipe. One can also use a set of cylindrical coordinates: ϕ, the azimuthal angle in the transverse plane, and θ, the polar angle from the beamline, both defined in Figure 4.2.

The pseudorapidity variable, η, defined in Equation 4.2, is frequently used instead of the polar angle θ. The relevance of this variable comes from the fact that differences in η between particles are invariant under longitudinal boosts.

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{4.2}$$

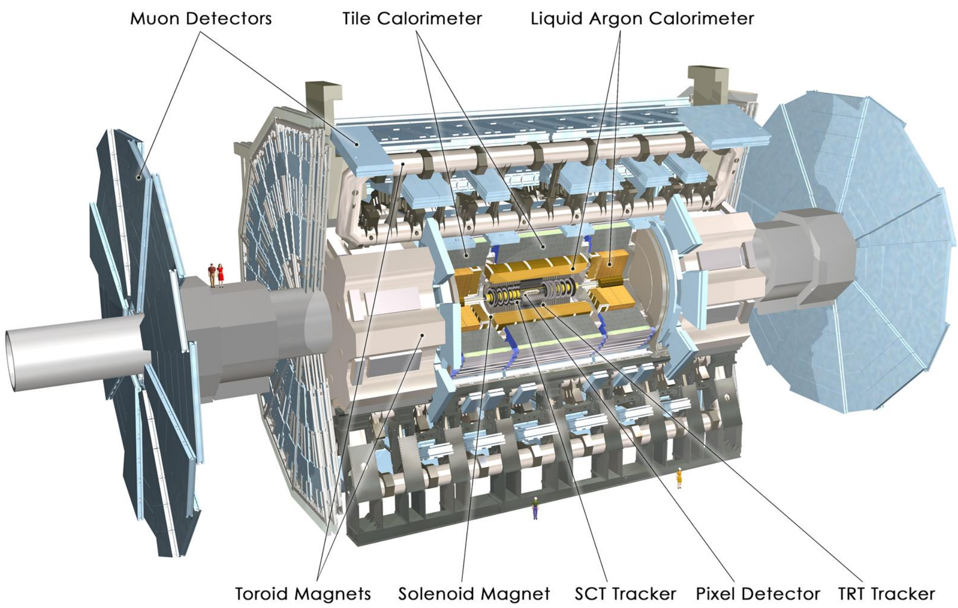


Figure 4.1: Schematic view of the ATLAS detector with its different subdetectors. Source: https://cds.cern.ch/images/CERN-GE-0803012-01

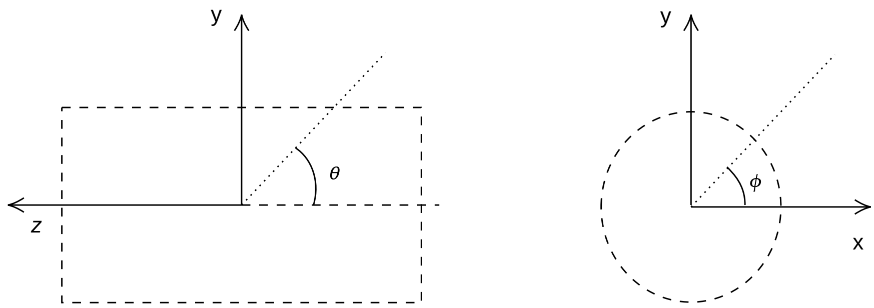


Figure 4.2: Schematic view of the coordinate system of the ATLAS detector

4.2.2 Magnet system

The ATLAS magnet system [40] is tasked with bending the trajectories of charged particles (via Lorentz force), allowing the measurement of their momenta, proportional to the curvature of the trajectory, and the sign of their charge, given by the direction of curvature. It comprises a central 2 T solenoid, placed between the tracker and the calorimeters, and a set of toroidal magnets outside of the calorimeters (from which ATLAS gets its name). There are eight toroids in the barrel region, with a magnetic field of 3.9 T, and two in the endcaps, with a magnetic field of 4.1 T. The toroids provide additional bending along ϕ to the trajectories of muons, allowing improved resolution in the measurement of their momenta.

4.2.3 Subdetectors

The ATLAS detector is divided in four distinct subdetectors: the Inner Detector [41], which reconstructs the trajectories of charged particles produced in the collision; the Liquid Argon [42] and Tile [43] calorimeters, which measure the energy of electrons, photons and hadrons; and the Muon Spectrometer [44], which assists in the precise reconstruction of muon trajectories.

Inner Detector

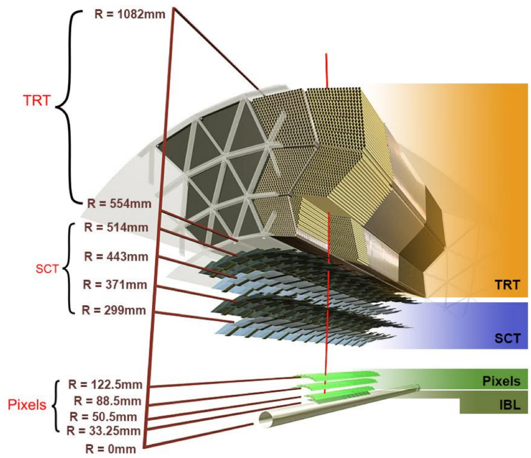


Figure 4.3: Cross-sectional view of the ATLAS Inner Detector barrel. Source: [45]

The Inner Detector (ID) covers a pseudorapidity range of $|\eta| < 2.5$. Its two innermost components are the Insertable B-Layer [46] and the Pixel Detector, four layers of very high-granularity networks of 2D silicon pixels. The next component in the structure is the Semiconductor Tracker, a series of orthogonal layers of silicon micro-strip detectors. The outermost component of the ID is the Transition Radiation Tracker, a series of straw (drift) tubes filled with Xenon, an inert gas.

When a charged particle passes through the silicon components in the Insertable B-Layer, Pixel detector or Semiconductor Tracker, it creates several electron-hole pairs. The electrons and holes are pushed to opposite sides by an applied electric field and are collected by electrodes, producing electric signals. The combination of signals from different detector planes are reconstructed into 3-dimensional points called hits. The hits are then used to build tracks and derive information on particle charge and momenta and also to reconstruct the several interaction vertices in the event, which helps with removing pile-up contamination.

When a charged particle passes through the Xenon inside the Transition Radiation Tracker tubes, electron-ion pairs are produced and collected by a wire inside the tube, generating an electric signal. This subdetector provides up to 38 additional hits per track to improve momentum resolution and do electron/pion discrimination. The latter is achieved by measuring transition radiation photons, emitted when a particle changes between media of different dielectric constant, with an intensity proportional to the Lorentz factor of the crossing particle.

A schematic description of the ID track parameters is given in Figure 4.4. The track resolution is between 100 $\mu\mathrm{m}$ and 300 $\mu\mathrm{m}$ for the transverse impact parameter, d_0, and between 100 $\mu\mathrm{m}$ and 1.5 mm for the longitudinal impact parameter, $z_0 \sin\theta$, decreasing with the transverse momentum and increasing with the pseudorapidity of the particle. The resolution in transverse momentum is $\sigma_{p_T}/p_T = 0.05\% \oplus 1\%$ (from Ref. [39]), where $\oplus$ represents sum in quadrature.

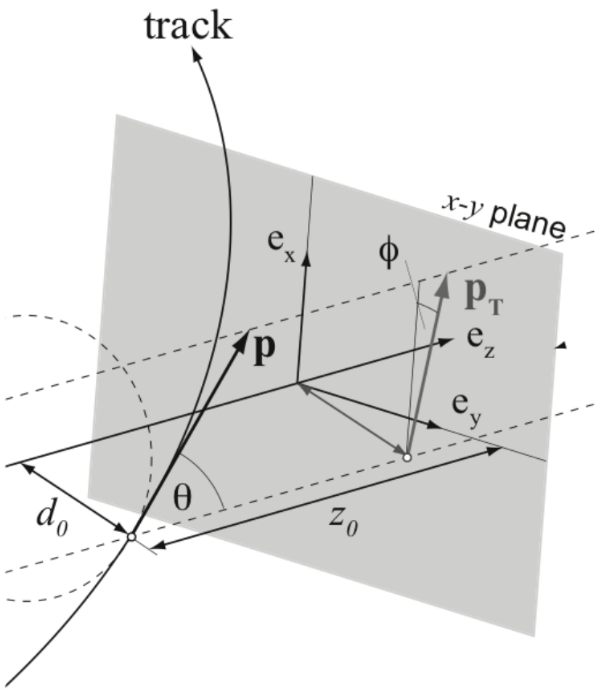


Figure 4.4: Schematic representation of the perigee parameters of an Inner Detector track. Source: https://indico.nbi.ku.dk/event/758/contributions/5022/attachments/1688/2373/trackingnote.pdf

Calorimeters

Calorimeters are designed to measure the energy of the particles that result from the collision. These are broadly divided in two categories: electromagnetic calorimeters, which measure the energies of electrons and photons, and hadronic calorimeters, which measure the energy of hadrons. Electrons and photons interact electromagnetically and produce compact showers (driven by Brehmsstrahlung and pair production), which are contained in a relatively short distance (around $20\ \mathrm{cm}$). Hadrons interact mainly through the strong force and produce very diffuse showers (driven by nuclear interactions), needing larger calorimeters to be fully contained. ATLAS uses sampling calorimeters, composed of layers of passive materials, in which the particles interact, interleaved with layers of active material, where their energy is measured. The layout of the ATLAS calorimeters is schematized in Figure 4.5.

The ATLAS electromagnetic calorimeter uses Liquid Argon (LAr) technology, with layers of steel and LAr in an accordion-like structure. It is divided in two sections: the LAr barrel, for $|\eta| < 1.4$, and the LAr electromagnetic endcaps, for $1.4 < |\eta| < 3.2$. The energy resolution for electrons is given in Equation 4.3 (from Ref. [39]).

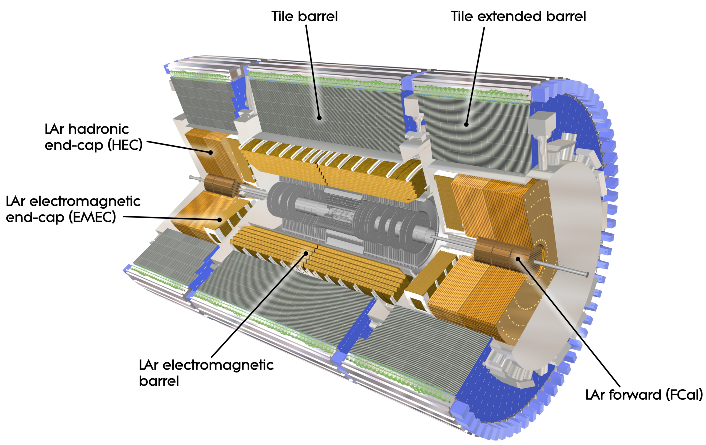


Figure 4.5: A cross-sectional view of the ATLAS calorimeters. Source: [47]

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E/GeV}} \oplus 0.7\% \tag{4.3}$$

The ATLAS hadronic calorimeter is composed of two calorimeters: the Tile calorimeter and the LAr hadronic endcaps. The Tile calorimeter covers the $|\eta| < 1.7$ region and is composed of layers of iron and plastic scintillating tiles. When a charged particle crosses a tile, the tile emits scintillation light with an intensity proportional to the particle's energy. This light is then guided to photomultiplier tubes via wavelength shifting optical fibers aluminized at one end, in order to maximize light collection efficiency at the other end. The LAr Hadronic endcaps cover the range $1.5 < |\eta| < 3.2$, with parallel layers of copper and LAr. Due to several factors, such as the large value of invisible energy from nuclear excitations or the fact that hadronic showers are much wider and stochastic than electromagnetic showers, the hadronic calorimeter's resolution for jets, given in Equation 4.4 (from Ref. [39]), is worse than that of the electromagnetic calorimeter for electrons.

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E/GeV}} \oplus 3\% \tag{4.4}$$

In the $3.2 < |\eta| < 4.8$ region, there is the LAr Forward Calorimeter, which performs electromagnetic and hadronic calorimetry, and is composed of layers of copper and LAr or tungsten and LAr.

Muon Spectrometer

The Muon Spectrometer (MS) covers the range $|\eta| < 2.7$ and has a dual function: measure very precisely the trajectory of muons and also assist triggering on events with muons. The main technology used are gaseous detectors, where a muon ionizes the gas, creating electron-ion pairs, which drift to opposite directions by means of an applied electric field. The collected charge produces an electric signal which is then measured and reconstructed as a series of hits, which are then combined to create tracks that can be further combined with ID tracks for improved muon momentum resolution. A schematic of the MS can be seen in Figure 4.6.

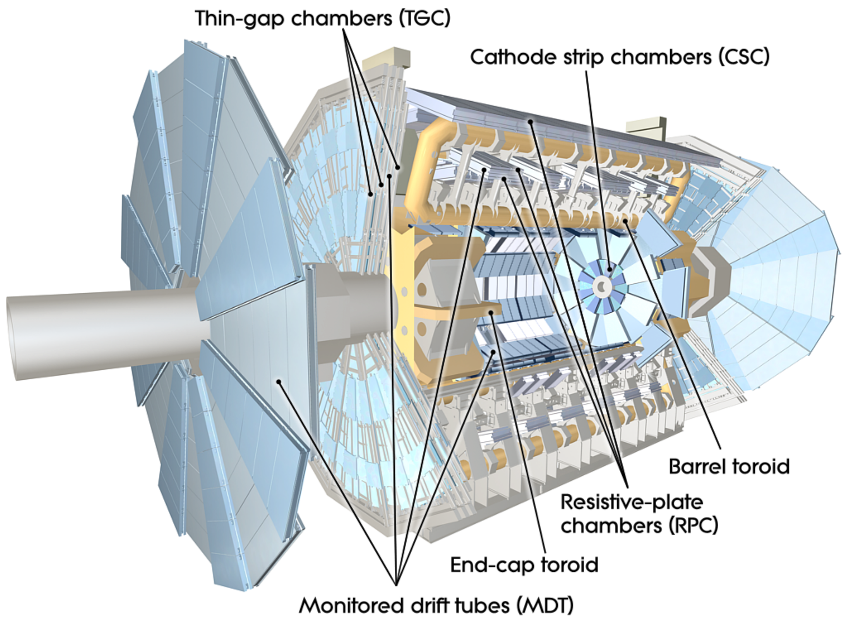


Figure 4.6: Cross-section view of the muon system. Source: [48]

For precision measurements, there are Cathode Strip Chambers in the range $2.0 < |\eta| < 2.7$ (due to high rates), and Monitored Drift Tubes elsewhere, providing very accurate measurements of the trajectory of passing muons, with a spatial resolution between 60μm and 80μm. For triggering purposes, there are Resistive Plate Chambers in the barrel, and Thin Gap (Wire) Chambers in the endcaps.

4.2.4 Trigger and data acquisition

The extremely high collision rate at the LHC, along with data readout and physical storage limitations, makes impossible storing every event. Because of this, the several detector components are connected to a two-level trigger system, which decides in real-time whether or not to record an event in permanent storage for further analysis. The trigger system selects about one thousand events out of the forty million events that occur each second. It is divided in two levels: the Level 1 (L1) trigger and the High Level Trigger (HLT). A schematic representation of the ATLAS Trigger system during Run 2 can be found in Figure 4.7.

The L1 trigger is hardware-based and consists of four main components: L1Calo, L1Muon, L1Topo and the Central Trigger Processor. L1Calo and L1Muon use low-granularity data from the calorimeters and the MS to identify signatures of objects such as electrons, jets or muons. These run simplified fast algorithms on dedicated hardware processors which sit very close to the detector. L1Topo allows the possibility of combining information from the different objects identified in L1Calo and L1Muon to build topological variables such as the angular distance between objects. The Central Trigger Processor combines the L1Calo, L1Muon and L1Topo information, compares it to a predefined menu of signatures of interest based on the existence of high transverse momentum objects, and decides whether or not to accept the event. If the event is accepted, an *L1Accept* signal and LHC timing information is sent to the detector read-out drivers, sending also the information of the position of the identified objects to the HLT.

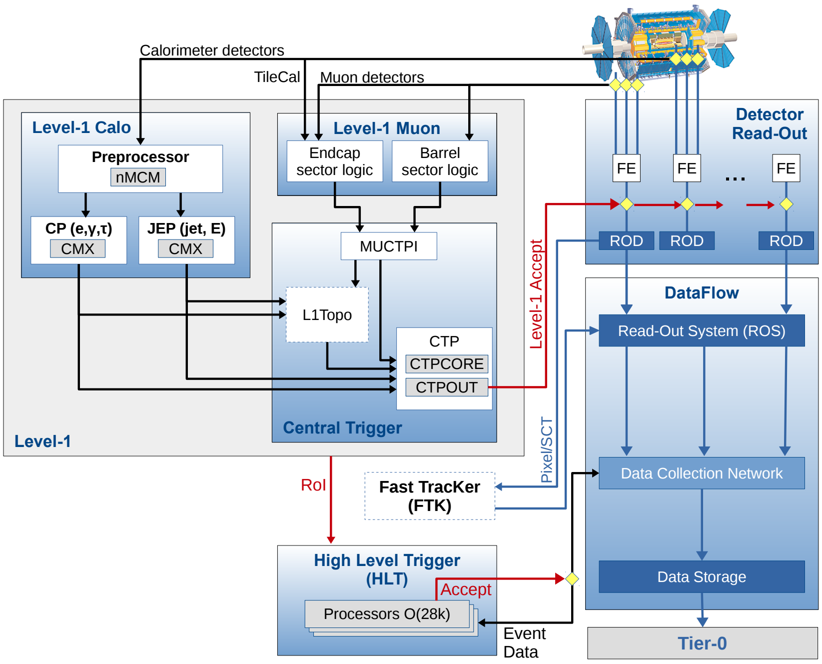


Figure 4.7: Schematic representation of the ATLAS Trigger system in Run 2 of the LHC. In this period, the FTK was still in comissioning. Source: [49]

The L1 trigger takes a decision within a 2.5 μs latency, reducing the rate from 40 MHz to 100 KHz.

The HLT is a software-based trigger, running sequences of algorithms (chains) to verify signatures identified by the L1 trigger or reject them as early as possible. It can run in a region-of-interest-based approach, reconstructing only a small region of the detector around the L1 objects, or do full event reconstruction, using algorithms similar to those used at analysis-level. Furthermore, it uses improved calibrations with respect to those used in the L1 trigger. If the event is accepted, an *Accept* signal is sent to the data collection network, which records it to permanent data storage. This decision is taken in a time frame of approximately 300 ms, reducing the rate from 100 KHz to 1 KHz.

The trigger system allows the possibility of applying a prescale factor N to individual L1 items or HLT chains, which means that it only accepts a fraction $1/N$ of the events which pass the requirements of the targeted L1 item or HLT chain.

Data reconstruction and data quality

Collected data are uniquely identified by their event and run numbers, where a run corresponds to a period of hours of data-taking or approximately one LHC fill. Part of the collected events is used to derive calibrations for the various subdetectors or monitor data quality, among other things. These are reconstructed immediately after storage, and several calibration and data quality monitoring algorithms

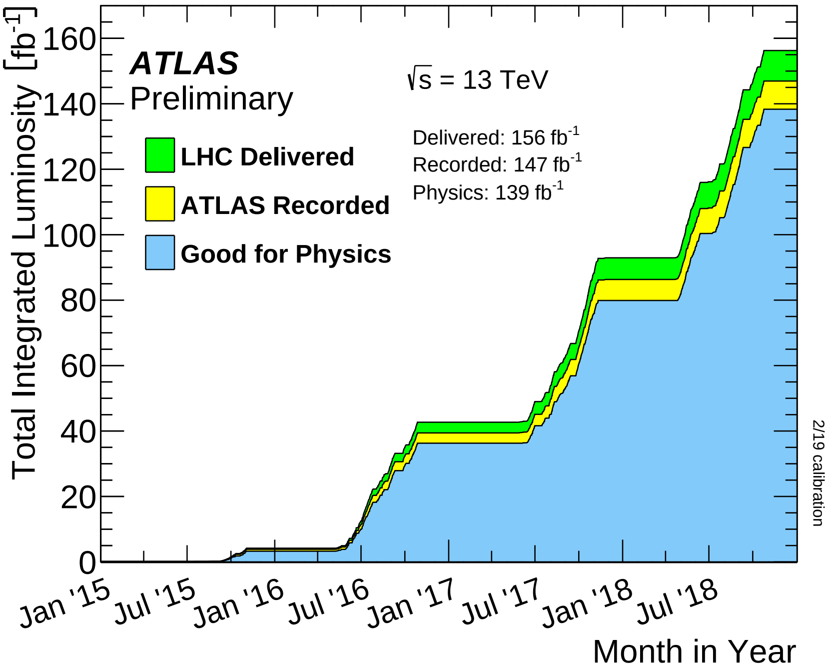


Figure 4.8: Evolution of the recorded ATLAS luminosity during Run 2 of the LHC. Source: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2

are run over it. If any problem is found, flags are set and encoded in a database. Using that database, a list of the usable runs is compiled, called the Good Runs List, which allows an easy and automatic way to avoid events which fail data quality requirements. The derived calibration constants are applied in the full run reconstruction that happens approximately 24 hours after the data-taking. For the 2018 data-taking period, the ATLAS detector had a data-taking efficiency larger than 98%.

Chapter 5

Simulation and event reconstruction

This chapter describes generally the principles behind event generation, detector simulation, and event reconstruction, i.e. how one goes from the results of a collision, measured as electric signals from different subdetectors to physics objects, such as electrons, muons or jets, with a definite energy and position.

5.1 Event simulation

Simulated events are the backbone of a large number of studies of the Higgs sector or BSM physics. Phenomenologists use them to study novel reconstruction or selection techniques such as jet substructure, and to guide experimentalists on where to look for possible signals of new physics. Experimentalists use them to develop an analysis and, most importantly, to calculate the statistical significance of a measurement, by comparing theoretical predictions with data. Event simulation can be divided in two parts: the simulation of the desired process, followed by simulation of the interaction of the resulting stable particles with a detector such as ATLAS.

5.1.1 Monte Carlo event generation

Monte Carlo (MC) event samples are used to simulate the expected shape and normalization of the different signal and background processes for a desired luminosity. Event generation is the process of obtaining the type and kinematic distributions of all the stable particles resulting from a collision, collectively called the final state. A schematic view of the event generation process is shown in Figure 5.1.

For LHC experiments such as ATLAS or CMS, event generation involves simulating proton-proton interactions. Due to the QCD factorization theorem [50], it is possible to write the proton-proton cross-section as a product of two parts: a short range part, which concerns the high energy interaction between a pair of proton constituents, called partons [51]; and a long range part, which takes into account the distribution of partons inside the proton.

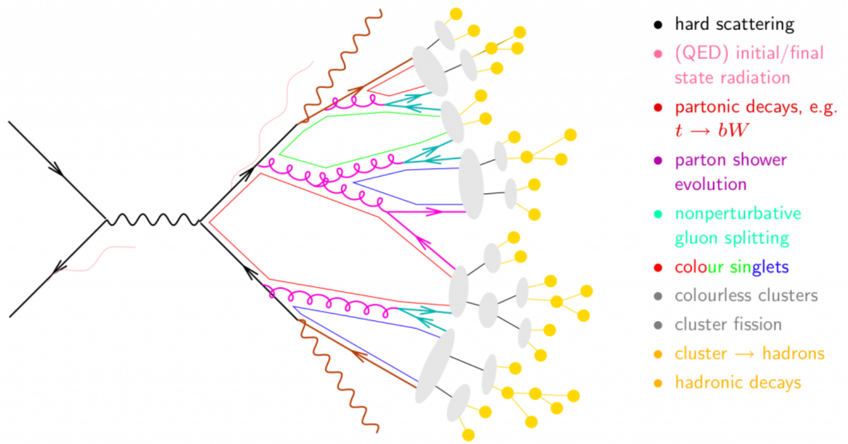


Figure 5.1: Schematic representation of the event generation process in a hadron-hadron collision. Underlying event generation not included.

The first step uses an interaction Lagrangian or matrix element calculated at a fixed order in perturbation theory to calculate differential cross-sections and simulate the kinematics of the high energy parton interaction, called the hard scattering. This can be done with generators such as `Powheg` [52, 53, 54] or `Madgraph5_aMC@NLO` [55]. The distribution of the partons inside the proton is modelled by parton distribution functions (PDF), $f_j(x)$. These give the probability that parton j carries fraction x of the momentum of the parent proton and are determined from data, with many different PDF sets available. The full proton-proton cross-section formula is given in Equation 5.1.

$$\sigma_{pp\to X}(s, \mu_F, \mu_R) = \sum_{a,b} \int dx_1\, dx_2\, d\Phi_{FS}\; f_a(x_1, \mu_F)\, f_b(x_2, \mu_F)\, \hat{\sigma}_{ab\to X}(\hat{s}, \mu_F, \mu_R)\,, \tag{5.1}$$

where $\hat{\sigma}_{ab\to X}$ is the partonic cross-section, $\hat{s}$ is the partonic center of mass energy, μ_F is the factorization scale and μ_R is the renormalization scale.

Next, the simulation of the underlying event is carried out and is added on top of the hard scattering simulation. This is followed by the simulation of the evolution of the resulting partons in an iterative process of parton radiation and splitting called parton shower. This is followed by the simulation of the hadronization process, by which low-energy partons combine to form hadrons, as a result of the confinement property of QCD. This is a low-energy process, which means that it can't be simulated or calculated exactly and one has to resort to models, with different generators using different models. The most common generators used for parton showering and hadronization are `Pythia` [56] and `Herwig` [57, 58]. The event generation process is finalized by modelling the decays of unstable hadrons, which can be done using specialized software such as `EvtGen` [59].

The simulated events contain global information, such as their generator weight or center of mass energy, as well as the information of the final state particles, such as their mass, transverse momentum and the (η, ϕ) coordinates with regards to the interaction point.

In order to approximate as accurately as possible what is seen in the experiments, it is also very important to generate pile-up events, which are then combined with the generated events of interest.

5.1.2 Detector simulation

Detector simulation is limited mainly by the time scale of the study, since the time to produce a sample is proportional to the desired accuracy. This leads to different choices in the tools used for this purpose.

For experimental studies, it is very important that the detector simulation is done as well as possible, in order to obtain an accurate representation of what will be seen in data. For this, very complex models of the detector based on the `Geant4` [60] software are used, which simulate very accurately the interaction of the particles with each of the detector components. This corresponds to simulating the Inner Detector and Muon Spectrometer hits and the calorimeter energy deposits, which are then converted to a computer-readable representation of the detector response. The ATLAS event simulation infrastructure is defined in Ref. [61] and is used to generate all of the SM signal and background samples used in this study.

For phenomenological studies, one can use simpler and faster detector simulations in order to have reasonable estimations of the kinematic distributions of the final state particles. One example of such software, used in this work, is `Delphes` [37], which takes the generator-level event record and simulates the interaction of the final state particles with the detector, using parametrizations of its response as a function of the particle's type and its kinematic quantities. Detector simulation in `Delphes` starts with the propagation of stable particles in a magnetic field parallel to the beam direction, followed by modelling of their interaction with the ID, taking into account its acceptance and resolution. This is followed by modelling the interaction of the particles with the electromagnetic and hadronic calorimeters - each represented by a single layer of cells in the (η, ϕ) plane - taking into account parameters such as the cell (η, ϕ) segmentation, energy resolution and the percentage of energy deposited in each of the calorimeters by the different particles. It finalizes by combining electromagnetic and hadronic calorimeter cells with the same (η, ϕ) coordinates into calorimeter towers.

5.2 Object/event reconstruction

Event reconstruction consists in taking detector signals and reconstructing physics objects such as electrons, muons or jets, with a definite energy and position. This process is required both in data and simulated samples. This section introduces the principles behind the reconstruction of several objects relevant to this analysis and the differences between ATLAS and `Delphes` reconstruction. It also introduces the set of object reconstruction and selection criteria used in this analysis to suppress background and pile-up contamination, and also to improve resolution on quantities such as the large-R jet invariant mass.

For ATLAS MC samples, the set of simulated detector responses is given as input to the same set of reconstruction and pile-up removal algorithms that run on data, in order to approximate them as much as possible.

`Delphes` parametrizes the efficiency and resolution of reconstruction algorithms as a function of particle type and kinematics. For this work, a modified version of the default ATLAS card was implemented and the main differences with regards to the default card will be mentioned when relevant.

5.2.1 Tracks

In ATLAS data and MC samples, the reconstruction of tracks starts with clustering ID hits to form a set of 3D space points, which are then combined together to form a track and fitted with a combinatorial Kalman filter, applying track quality cuts [1]as well as solving any possible ambiguities.

In `Delphes`, the efficiency and resolution of track reconstruction algorithms is parametrized as a function of particle type and kinematics. Particles with transverse momentum lower than 100 MeV or absolute pseudorapidity larger than 2.5 are removed and do not lead to a reconstructed track. This removes particles that have a very small momentum and end up looping inside the ID structure never reaching the calorimeters, or particles that fall outside the ATLAS ID acceptance. `Delphes` allows selecting and combining the tracks from all charged particles to build a set of tracks similar to ATLAS ID tracks.

Vertices

In ATLAS data and MC samples, the position of the several interaction vertices in an event is obtained by extrapolating tracks from the ID to the LHC beamline. The primary vertex is defined as the vertex with the largest value of the sum of the squared transverse momenta of the incoming tracks, and is generally associated to the hard scattering event.

`Delphes` simply records a list of the positions of the different interaction vertices, from the event generator record.

5.2.2 Electrons and muons

In ATLAS data and MC samples, electrons are reconstructed by matching an ID track to a reconstructed cluster in the LAr calorimeter. Other particles or jets can mimic this signature, so it is necessary to reject these fake electron candidates as much as possible. This is done by identification algorithms which use a likelihood-based method [62], taking as input information such as the energy over momentum ratio of the candidate particle, the number of high threshold Transition Radiation Tracker hits or shower shape variables. Three working points, with increasing background rejection, are defined: Loose, Medium and Tight. The Tight working point is used in this analysis, with an average identification efficiency of 80% and a fake/background rejection factor, defined as the inverse of the background efficiency, of approximately 1200.

Muon reconstruction starts by taking hits in the muon chambers to build MS tracks. These can be combined to ID tracks using an outside-in approach, i.e. extrapolating the MS tracks and matching them to ID tracks, and fitting them as whole to improve momentum resolution. Furthermore, muons can also be reconstructed using only MS tracks or even from a combination of ID tracks and calorimeter deposits,

[1] At least 7 hits in the Pixel and Strip detector, no more than one of the hits in the Pixel detector being shared by multiple tracks, no more than one missing hit in the Pixel detector and no more than two missing hits in the Strip detector

in the region where there is a small gap in the muon chambers. around $\eta = 0$. Muon track parameters, such as the χ^2 of the track fit, are given as input to muon identification algorithms, where 3 working points are defined, similar to electron identification algorithms. In this analysis, only muons built from combined MS+ID tracks which pass the *Medium* identification working point are taken into account. This working point corresponds to an efficiency of approximately 96% and background rejection factors between 500 and 1000, increasing with the transverse momentum of the muon candidate. [63]

Leptons produced in decays of particles such as W or Z bosons, tend to have low hadronic activity around them, contrary to those produced in the decay of a b or c quark, for instance. The presence of isolated leptons can then be used to distinguish between these processes. To identify isolated leptons, one calculates the energy deposited in a cone around the lepton, using tracks or calorimeter-based objects as input.

The lepton selection requirements for the ATLAS data and MC samples used in this analysis are shown in Table 5.1. These define what will be called in following sections as a *signal lepton*.

	Electron	Muon
Identification	LH Tight	Medium
d_0^{sig}	< 5	< 3
$\lvert\Delta z_0 \sin\theta\rvert$	< 0.5 mm	< 0.5 mm
p_T	> 27 GeV	> 25 GeV
$\lvert\eta\rvert$	< 2.47	< 2.5
Isolation	$\sum_{\text{TopoClusters},\Delta R<0.2} E_T < \max\left(0.015 \times p_T^e\,[\text{GeV}], 3.5\text{ GeV}\right)$	$\sum_{\text{tracks},\Delta R<0.2} p_T < 1.25$ GeV

Table 5.1: Signal lepton selection requirements. LH corresponds to the likelihood-based multivariate discriminant used for electron reconstruction.

$d_0^{\text{sig}} = \frac{d_0}{\sigma_{d_0}}$ is the transverse impact parameter significance and $|\Delta z_0 \sin\theta|$ is the longitudinal distance between the object and the primary vertex. TopoClusters [47] are three dimensional collections of topologically connected energy deposits, which can combine information from the LAr and Tile calorimeters,

In `Delphes`, after emulating tracking efficiency, a Gaussian smearing of the 4-momentum of electrons and muons is performed, parametrizing the resolution of the different subdetectors. The transverse momentum and pseudorapidity requirements are implemented as for the ATLAS samples, but electron isolation is calculated from calorimeter towers (instead of TopoClusters), and muon isolation is calculated from the set of ID-like tracks described in Section 5.2.1. Electrons or muons whose tracks have a longitudinal impact parameter $|\Delta z_0 \sin\theta| > 0.1$ mm are removed. The requirements transverse impact parameter significance and electron and muon identification are not replicable, due to not having a complete description of the interaction of the particles with the various subdetectors. Thus, `Delphes` samples don't take into account fake electron candidates.

The neutral component of pile-up leads to an increase in calorimeter-based variables, such as electron isolation, with regards to their expected value. This component can only be removed on average depending on the energy deposited in the calorimeters. In data, ATLAS and `Delphes` MC samples, the neutral component of pile-up is estimated and removed using the area method [64]. This consists in using a grid-based method to extract the average pile-up energy density, ρ, which is then used to correct the electron isolation, using the expression in Equation 5.2.

$$I \to I - \frac{\rho \pi R^2}{p_T^e}, \tag{5.2}$$

where I is the isolation and R is the radius of the isolation cone.

5.2.3 Jets

Stable hadrons in an event form streams, called jets, which carry a large fraction of the energy and momentum from the parent partons and show up in the detectors as clusters of tracks and energy deposits. In order to compare analytic predictions of high-energy QCD in terms of partons with experimental results in terms of stable particles and energy depositions, a common, unambiguous and rigorous definition of jet is necessary, called a jet algorithm. In order to give finite cross-sections at any order of calculation in QCD, these should be infrared and collinear safe, which means that the jets defined by the algorithm should not change if a soft (low energy) or collinear parton is emitted.

The most commonly used class of infrared and collinear safe jet algorithms are sequential recombination algorithms, which build jets by iteratively combining the closest pairs of objects. The generalization of these algorithms is done in Ref. [65], where two distances are defined: d_{ij}, the distance between each pair of input objects i and j, and d_{iB}, the distance between the object i and the proto-jet direction, both defined in Equation 5.3.

$$d_{ij} = \min\left(k_{t_i}^{2p}, k_{t_j}^{2p}\right) \Delta R_{ij}^2 \qquad\qquad d_{iB} = k_{t_i}^{2p} R^2, \tag{5.3}$$

where k_t is the transverse momentum of the particle with regards to the axis of the proto-jet, R is the jet radius, a parameter equivalent to the cone radius, and ΔR_{ij} is the angular distance in the (η, ϕ) plane, defined in Equation 5.4.

$$\Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \tag{5.4}$$

The three main algorithms used today correspond to different values of p: $p = -1$ corresponds to the anti-k_t algorithm, which starts by combining the objects with the highest transverse momentum, leading to jets whose shape is not modified by soft radiation such as that coming from pile-up; $p = 1$ corresponds to the k_t algorithm, which combines first the objects with lowest transverse momentum, making it very susceptible to pile-up contamination; $p = 0$ corresponds to the Cambridge/Aachen algorithm which combines geometrically nearby objects, since $d_{ij} = \Delta R_{ij}^2$. All of the reconstruction algorithms above are implemented in `FastJet` [66], used for jet reconstruction in ATLAS and `Delphes` samples.

Large-R jets

b quarks from the $H \rightarrow b\bar{b}$ or $Z \rightarrow b\bar{b}$ decay have an angular separation inversely proportional to the energy of the mother particle due to Lorentz boost, following roughly Equation 5.5.

$$\Delta R \approx \frac{2m}{p_T} \tag{5.5}$$

In the high transverse momentum (boosted) regime, these decay products are very collimated and it becomes difficult to disentangle them using standard techniques, such as anti-k_t, $R = 0.4$ jet reconstruction. In this work, the boosted Higgs boson is identified by reconstructing it as a large-R jet, in order to fully contain its decays, and then searching for substructure inside it.

Due to their large size, large-R jets suffer from severe pile-up and underlying event contamination, which is proportional to the jet area [67]. Several techniques, generally called *grooming* techniques, have been developed in order to remove this contamination. The one used in this work, called trimming [68], takes the constituents of the large-R jet and reconstructs subjets with the k_T algorithm with a smaller radius, R_{sub}, rejecting any subjets with a transverse momentum $p_T < f_{cut}\, p_T^{jet}$, where p_T^{jet} is the transverse momentum of the large-R jet. A schematic representation of the trimming algorithm is shown in Figure 5.2.

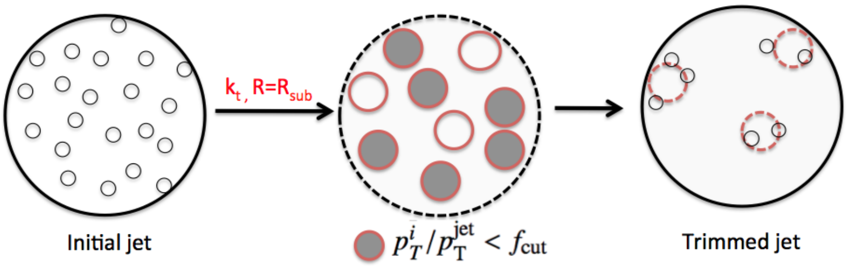


Figure 5.2: Schematic representation of the trimming algorithm. Source: [69]

In this work, large-R jets are reconstructed with the anti-k_t algorithm with a radius parameter $R = 1.0$. Furthermore, they undergo a *trimming* procedure with $R_{sub} = 0.2$ and $f_{cut} = 0.05$. Only large-R jets with an absolute pseudorapidity $|\eta| < 2.0$ and a transverse momentum $p_T > 250$ GeV are considered in this analysis. The transverse momentum requirement corresponds to an angular distance between the b-quarks of $\Delta R \approx 1.0$, using Equation 5.5, and the pseudorapidity requirement is made to ensure that the entire large-R jet is inside the tracker acceptance, in order to use the *combined mass* definition given in Equation 5.6 for ATLAS data and MC samples.

In ATLAS data and MC samples, large-R jets are reconstructed from TopoClusters corrected to point to the primary vertex, possible due to the fact that one can derive location and direction information from them. The input TopoClusters are calibrated with the Local Cell Weighting method, which aims to compensate for signal losses in clustering, energy loss in inactive material and the non-compensating nature of the calorimeters, which leads to a different response for electromagnetically and hadronically interacting particles with the same energy. In order to improve the resolution on the large-R jet invariant mass, a *combined mass* definition described in Equation 5.6, which combines tracker- and calorimeter-based measurements, is used. The weights for the tracker- and calorimeter-based terms are chosen in order to minimize the variance in the values of the mass.

$$m_{comb} = w_{calo} \times m_{calo} + w_{tracker} \times m_{tracker} \times \frac{p_T^{calo}}{p_T^{tracker}} \tag{5.6}$$

A set of MC-based jet energy and mass scale corrections are applied in order to approximate these quantities to their truth-level scale. These are derived in simulated events, by matching the reconstruction-level jets to truth-level jets on which the same trimming has been performed. About 20% of large-R jets from a $H \to b\bar{b}$ decay contain reconstructed muons, from the semileptonic decays of B hadrons. These carry most of their energy outside the calorimeter, leading to a degradation of the large-R jet invariant mass resolution, which can be improved by taking into account the energy carried by these muons. The method used in this analysis is based on the method described in [70], which starts by taking all reconstructed muons with $p_T > 10$ GeV, $|\eta| < 2.7$ and Medium identification working point. These are matched to track jets inside the large-R jet using a transverse momentum dependent criteria, $\Delta R\,(\mathrm{muon, track\,jet}) < \min\,(0.4, 0.04 + 10/p_T^{\mu})$. The energy of the matched muons is added to the calorimeter mass (after subtraction of energy loss in the calorimeter), the track-assisted mass is corrected, and the jet mass is recalculated, leading to a 6% improvement in resolution.

In `Delphes`, the input collection are calorimeter towers and the trimming procedure is applied, using the parameters described above. The jet mass is calculated using calorimeter objects only with no further corrections or calibrations applied.

Large-R jet substructure

One of the ways to study large-R jet substructure is by reconstructing smaller radius jets, match them to the large-R jet, and using the matched jets to separately identify the decay products of the mother particle using techniques such as b-tagging, described in Section 5.2.4. For this, it is useful to reconstruct jets with a radius inversely proportional to the transverse momentum of the particles, in order to be able to resolve them as well as possible across a large transverse momentum range. This is accomplished by variable-radius jet reconstruction [71], obtained by modifying the anti-k_t algorithm as described in Equation 5.7.

$$d_{ij} = \min\left(k_{t_i}^{2p}, k_{t_j}^{2p}\right)\Delta R_{ij}^2 \qquad\qquad d_{iB} = k_{t_i}^{2p} R_{eff}^2,\ R_{eff} = \frac{\rho}{k_{t_i}},\ R_{min} < R_{eff} < R_{max}, \tag{5.7}$$

where ρ, R_{min} and R_{max} are user-defined parameters. The latter two need to be defined so that the algorithm does not output jets with infinitesimally small radius, or a single jet spanning the entire detector.

Track jets

Track jets are used in this analysis to identify the b-quarks from the $H \to b\bar{b}$ decay and to categorize the events. These have an advantage in b-tagging with regards to calorimeter jets, in that they allow better removal of pile-up contamination, since one can use as input only tracks that are matched to the primary vertex. Furthermore, they allow a direct connection to the tracks used by the b-tagging algorithms. This leads to the track jet being much more aligned with the B hadron direction than a similarly built calorimeter jet. For this analysis, only track jets with $p_T > 10$ GeV, $|\eta| < 2.5$ and more than two tracks are considered

In ATLAS data and MC samples, they are built from ID tracks with $p_T > 0.5$ GeV, $|\eta| < 2.5$ and $|z_0 \sin\theta| < 3\mathrm{mm}$. For its reconstruction there are two possibilities: fixed-radius anti-k_t with $R = 0.2$ or variable-radius anti-k_t with $\rho = 30$ GeV, $R_{min} = 0.02$ and $R_{max} = 0.4$. These are matched to the large-R jet by ghost association, a method described in [67], which consists in adding the track jet 4-vectors with zero mass (*ghosts*) to the event, and rerunning large-R jet reconstruction. Ghost association is used instead of simple geometrical matching, $\Delta R\,$(track jet-large-R jet) < 1.0, since it is less ambiguous.

In `Delphes`, track jets are built from the set of ID-like tracks described in Section 5.2.1, using as input tracks with $|\Delta z_0 \sin\theta < 0.1\,\mathrm{mm}$ and $p_T > 0.1$ GeV. They are reconstructed using fixed-radius anti-k_t with $R = 0.2$ and are matched to the large-R jet by simple geometrical matching, $\Delta R\,$(track jet-large-R jet) < 1.0. This is done because variable radius reconstruction and ghost association are not implemented in `Delphes` and their implementation would fall out of the scope of this thesis.

Small-R jets

The set of anti-k_t, $R = 0.4$ calorimeter jets, called small-R jets, is used in this analysis to categorize the events. In order to be considered in the analysis, these should have $p_T > 30$ GeV and $|\eta| < 4.5$.

In ATLAS data and MC samples, they are reconstructed from TopoClusters and are split in two subsets: central jets, with $|\eta| < 2.5$ and forward jets, with $2.5 < |\eta| < 4.5$. Central jets with $p_T < 60$ GeV are required to have a Jet Vertex Tagger score greater than 0.59, in order to supress pile-up contamination. The Jet Vertex Tagger is a multivariable likelihood-based discriminant to suppress jets arising from pile-up, targeting the 20 GeV $< p_T < 60$ GeV kinematic range, expected to be the most sensitive to this contamination. Small-R jets from non-collision background processes are removed by a jet cleaning tool, applying standard ATLAS procedures. These undergo a jet energy scale correction.

In `Delphes` samples, small-R jets are built from calorimeter towers, and no further corrections and calibrations are applied, except the removal of the neutral pile-up contribution, described below.

In ATLAS and `Delphes` samples, a correction to the small-R jet momenta is applied to remove neutral pile-up contribution, based on the method described in Ref. [64] (similar to the one applied for electron isolation). The correction is proportional to the jet area [67] and momentum and is defined in Equation 5.8.

$$p_{jet} \to p_{jet} - \rho \times A_{jet}, \tag{5.8}$$

where ρ is the average pile-up energy density and A is the jet area.

5.2.4 b-tagging

B hadrons have a larger lifetime than hadrons containing light quarks, travelling longer distances before decaying, producing a vertex a few hundred μm away from the main interaction point which can be resolved by the ID, called a secondary vertex. This characteristic experimental signature is schematized in Figure 5.3 and is the base of a specific class of techniques used to identify b-jets, called b-tagging algorithms.

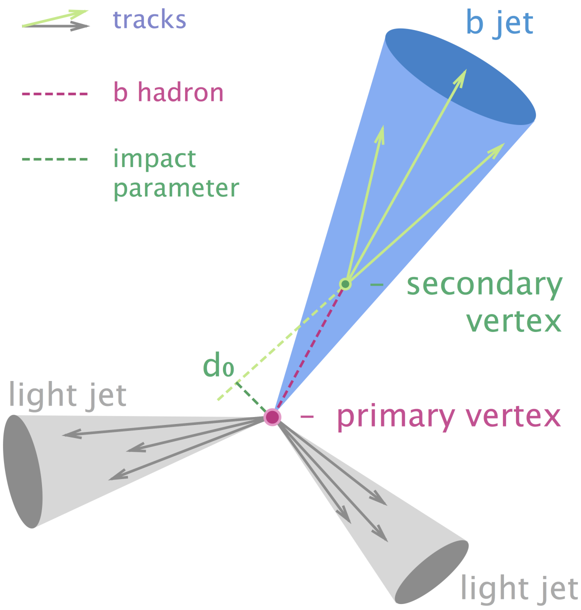


Figure 5.3: Schematic representation of a b-jet. Source: https://en.wikipedia.org/wiki/B-tagging

In ATLAS data and MC samples, the b-tagging process starts by exclusively selecting tracks within a cone around the jet axis, that are used as input to three low-level algorithms: IP3D, which exploits the longitudinal and transverse impact parameters, (usually higher for tracks from B hadron decays); SV1, which tries to reconstruct one secondary vertex for each jet; and JetFitter which tries to fit a line from the primary to the secondary vertex and beyond. The output of these low-level algorithms is combined in a high-level algorithm, called MV2. This work uses track jets as input for b-tagging, with the MV2c10 tagger in the 70% working point, corresponding to an average b-tagging efficiency (fraction of b-jets correctly identified) of 70%, derived in a non-allhadronic (dileptonic+semileptonic) $t\bar{t}$ sample for variable-radius track jets, with a c-jet (light jet) rejection of 9 (304). Events in which two variable-radius jet axes are inside the same cone, are removed, as recommended by the ATLAS Flavour Tagging group, in order to avoid poorly understood track-jet association for these jets.

The tagging and mistagging efficiencies (fraction of non b-jets identified as b-jets) are measured in data to correct the simulation, by applying data-to-MC scale factors. These are derived by the ATLAS Flavour Tagging group in a sample generated with `Powheg` and `Pythia8` and are applied for all jets that pass b-tagging, modifying the event weight in a multiplicative way. Inefficiency scale factors are applied to the jets that don't pass b-tagging, and are a function of the efficiency scale factors and the efficiencies themselves. Their definition is given in Equation 5.9.

$$\text{ineff. SF}\left(p_T\right) = \frac{1 - \text{SF}\left(p_T\right) \times \epsilon\left(p_T, |\eta|\right)}{1 - \epsilon\left(p_T, |\eta|\right)} \tag{5.9}$$

Since the scale factors depend on the shower model, i.e. generator and settings, the scale factor for an alternative generator, such as Sherpa, is obtained by multiplying the scale factor derived with the `Powheg+Pythia` sample with the ratio of the efficiencies obtained with the `Powheg+Pythia` sample and the sample produced with the alternative generator.

The efficiency maps, as well as the efficiency and inefficiency scale factors are stored in a Calibration Data Interface (CDI) file, which is read by the b-tagging tool in the analysis code.

In `Delphes`, given that secondary vertex identification is not possible, one uses an angular distance-based approach instead, where a jet becomes a b-jet candidate if a generator-level b quark is found within a $\Delta R \leq 0.3$ distance from the jet axis. The b-tagging efficiency is parametrized by the user, which has the possibility to also define the mistagging efficiencies. The efficiencies were extracted from the official ATLAS CDI file from July 2019 for bins of p_T and $|\eta|$ with no further scale factors applied to the events.

5.2.5 Missing transverse energy

At the LHC, the colliding beams have a negligible total transverse momentum, which should lead to a zero value for the missing transverse energy (MET/$E_{\mathrm{T}}^{\mathrm{miss}}$), defined as the norm of the vector sum of the transverse momentum of the final state particles. A non-zero value of the missing transverse energy allows infering the production of particles that cross the detector without interacting, such as neutrinos, or even possible dark matter particles.

In ATLAS data and MC samples, the missing transverse energy is built by summing the transverse momentum of all the reconstructed objects as well as an additional *soft* term, which corresponds to energy depositions not associated with any object. Being a simple energy sum, it is very susceptible to pile-up contamination and it has a very poor resolution, between 5 and 25 GeV, increasing with the sum of the transverse energy in the event [72].

In `Delphes`, the missing transverse energy is built in a similar fashion, the main difference being that there is no *soft* term in the calculation.

Neutrino Reconstruction

The value of the missing transverse energy can be used to reconstruct the full 4-momentum of the neutrino and, consequently, of the W boson, which can then be used to calculate quantities such as its rapidity. The rapidity of a particle, defined in Equation 5.10, is a function of its energy, E, and longitudinal momentum, p_z, and for $p \gg m$, it is equal to the pseudorapidity.

$$y = \frac{1}{2} \ln \left[\frac{E + |p_z|}{E - |p_z|}\right] \qquad \to \eta, \text{ when } p \gg m \tag{5.10}$$

The rapidity of the W boson can be used to calculate the difference in rapidity between the W boson and the Higgs candidate, studied in Section 6.4.3 as a variable to discriminate between signal and background. Furthermore, the full W boson 4-momentum can be used to build angular observables sensitive to anomalous Spin/CP components in the HWW vertex, such as those studied in Section 6.6.2.

To reconstruct the 4-momentum of the neutrino, one has to calculate its longitudinal component. This can be done by constraining the lepton+neutrino system to have the W boson mass, i.e. $(p_\ell^\mu + p_\nu^\mu)^2 = m_W^2$, neglecting off-shell effects and the W boson width. Identifying $p_{T_\nu} \equiv E_{\mathrm{T}}^{\mathrm{miss}}$, and $m_\ell = m_\nu = 0$, since the lepton and neutrino have masses much smaller than the W mass, this becomes equivalent to solving Equation 5.11.

$$p_{z_\nu} = \frac{p_{z_\ell} \times t}{2p_{T_\ell}^2} \pm \frac{E_\ell \times \sqrt{t^2 - \left(2 \times p_{T_\ell} \times E_{\mathrm{T}}^{\mathrm{miss}}\right)^2}}{2p_{T_\ell}^2} \tag{5.11}$$

$$t = m_W^2 + 2 \times \left(\mathbf{p_{T}}_\ell \cdot \mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}\right) = m_W^2 + 2 \times p_{T_\ell} \times E_{\mathrm{T}}^{\mathrm{miss}} \times \cos\left(\Delta\phi_{\ell, E_{\mathrm{T}}^{\mathrm{miss}}}\right) \tag{5.12}$$

where $\Delta\phi_{\ell, E_{\mathrm{T}}^{\mathrm{miss}}}$ is the difference between the azimuthal angle of the lepton and of the missing transverse energy. This equation can have an imaginary solution due to the finite resolution of the detector, or in background events with no neutrino, in which a non-zero missing transverse energy comes from detector effects. If that is the case, the imaginary part of the solution is neglected. In all other cases, there are are two possible solutions from the plus and minus signs in Equation 5.11. The solution chosen is the one that minimizes the difference in longitudinal boosts of the reconstructed W boson and the Higgs candidate, $|\beta_W^z - \beta_H^z|$, with the longitudinal boost of particle X defined as $\beta_X^z = \frac{p_{z_X}}{\sqrt{p_{z_X}^2 + m_X^2}}$ [2]. A truth-level study (before detector simulation) carried out in Ref. [73] shows that this method has a true reconstruction efficiency - fraction of events where it reconstructs the correct neutrino longitudinal momentum - of 67%. This is higher than other methods, such as choosing the solution with the lowest value of p_{z_ν}, which has a 57% true reconstruction efficiency.

5.2.6 Overlap removal

An overlap removal procedure is applied to the ATLAS samples to avoid double-counting between the reconstructed leptons and jets. It follows the steps below:

- τ-electron: if $\Delta R(\tau, e) < 0.2$, the τ lepton is removed
- τ-muon: if $\Delta R(\tau, \mu) < 0.2$, the τ lepton is removed, with the exception that, if the τ lepton has $p_T >$ 50 GeV and the muon is not reconstructed from a combination of MS and ID tracks, the τ lepton is not removed
- e-μ: if a muon reconstructed from a combination of MS and ID tracks shares an ID track with an electron, the electron is removed. If a muon identified using ID+calorimeter information shares an ID track with an electron, the muon is removed.

[2] This definition differs from the usual definition $\beta_X^z = \frac{p_{z_X}}{E}$, but was chosen due to consistency with definitions used in previous analyses from the group

- e-small-R jet: if $\Delta R\,(e, \text{jet}) < 0.2$, the jet is removed. For any remaining jets, if $\Delta R\,(e, \text{jet}) < 0.4$, the electron is removed.
- μ-small-R jet: if $\Delta R\,(\mu, \text{jet}) < 0.2$ or the muon ID track is ghost-associated to the jet, then the jet is removed if it has less than three associated ID tracks with $p_T > 500$ MeV or both of the following conditions are met:
  - $p_T^{\mu}/p_T^{\text{jet}} > 0.5$
  - $p_T^{\mu}/\left[\sum_{p_T > 500\text{MeV}} p_T^{\text{jet tracks}}\right] > 0.7$
    * for any remaining jets, if $\Delta R\,(\mu, \text{jet}) < 0.4$, the jet is removed
- τ-small-R jet: if $\Delta R\,(\tau, \text{jet}) < 0.2$, the jet is removed
- e-large-R jet: if $\Delta R\,(e, \text{large-}R\text{ jet}) < 0.2$, the large-$R$ jet is removed
- small-R jet-large-R jet: since small-R jets are only used for event categorization, no object removal is done, but only small-R jets with a $\Delta R\,(\text{small-R jet}, \text{large-}R\text{ jet}) > 1.0$ are taken into account in categorization

`Delphes` performs an approximation of overlap removal, finding uniquely identified muons, electrons, large-R jets and small-R jets.

Chapter 6

Analysis

Anomalous Spin/CP components in the HWW interaction vertex modify the signal expected from the SM, by increasing the cross-section of associated WH production and the acceptance of this channel at higher momenta. The ATLAS sensitivity to these components is thus expected to increase, when targeting higher transverse momenta of the W and Higgs bosons. The goal of this Chapter is to explain the methods used to determine the sensitivity to anomalous components of an analysis of the WH process in the boosted regime with an integrated luminosity $L = 139.9\ \mathrm{fb}^{-1}$ of proton-proton collisions at a center of mass energy $\sqrt{s} = 13$ TeV collected by the ATLAS detector during Run 2 of the LHC. In Section 6.1 an overview of the analysis and a description of the main backgrounds is done, followed by a description of the used data and simulated samples and of the signal extraction methodology in Sections 6.2 and 6.3, respectively. In Section 6.4, the event selection and categorization for this analysis are presented, as well as the studies of the boosted Higgs tagging strategy and of variables to be used in event selection, carried out in the context of the ATLAS boosted $VH(H \to b\bar{b})$ analysis mentioned in Chapter 1. Section 6.5 shows the sensitivity of the developed analysis to the SM boosted Higgs boson signal, via measurements of the expected signal strength and significance. Section 6.6 contains a study of the sensitivity to anomalous components, divided in two parts: a study of the modifications to the expected signal strength and significance and the study of angular observables sensitive to these components, motivated by the fact that these modify the spin correlations between the final state particles.

6.1 Analysis Overview

This analysis targets events where the W boson has a transverse momentum larger than 250 GeV and decays to an electron or muon and a neutrino [1], with the Higgs boson decaying to a pair of b-quarks.
This final state, schematized in Figure 6.1, is defined by the presence of missing transverse energy from the neutrino and an isolated electron or muon with high transverse momentum from the decay of the W boson, which allows efficient triggering in the presence of a large multijet background. The Higgs

[1] τ leptons are not directly taken into account since they are unstable, decaying into an electron or muon and an additional neutrino about 34% of the times.

boson is reconstructed as a large-R jet with two b-tagged subjets, since the very large Lorentz boost causes the b-quarks from the decay to become very collimated, which leads to a drop in efficiency if the Higgs is identified using two *resolved* b-jets, as was done for previous studies of this channel [27, 28]. There are several other SM processes with the same or very similar final states as the signal events, but with much larger cross-sections, called backgrounds, such as associated production of a W boson with jets (W+jets), top quark pair production $(t\bar{t})$, single top quark production in the Wt channel and WZ production.

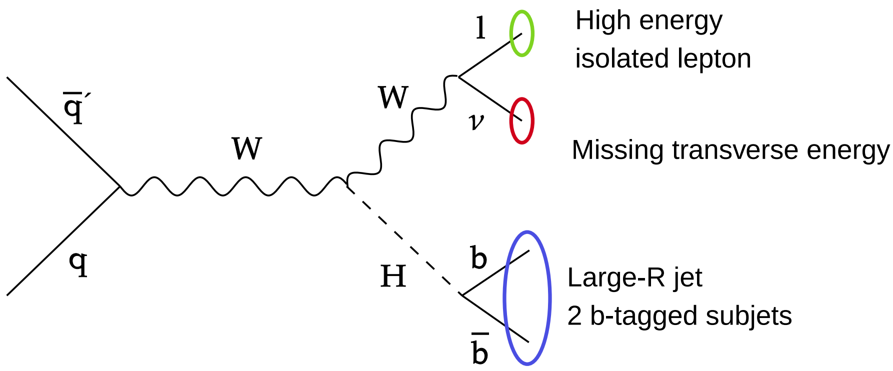


Figure 6.1: Schematic representation of the event topology of boosted $WH \to \ell\nu b\bar{b}$ events

6.1.1 Backgrounds

W+jets production has a cross-section approximately 10^5 times larger than WH production, with approximately one third of those events having a leptonically decaying W. If, in addition, the event has two collimated b-jets, the final state topology becomes the exact same as that of signal events and, even though the probability for this to happen is very small, the very large cross-section makes this process one of the biggest backgrounds.

$t\bar{t}$ production has a cross-section approximately 10^3 times larger than signal, with about 92% of the events containing two b-jets in the final state (from the $t \to bW$ decay) and around 45% of those events containing one of the W bosons decaying leptonically and the other to jets (*semileptonic* decay channel), schematized in Figure 6.2. Since the two b-quarks are produced in approximately opposite directions due to kinematics, the probability that these are collimated enough to be reconstructed inside a single large-R jet is small. Nonetheless, at least one of the quarks from the hadronically decaying W boson can end up inside the large-R jet along with one of the b-quarks. Given that the branching ratio of $W \to \bar{c}s$ is approximately 30%, and that the probability of mistagging a c-quark as a b-quark is approximately 10%, a non-negligible fraction of these events can be identified as signal.

Single top production in the Wt channel has a cross-section approximately 350 times larger than signal, and in around 20% of the events one of the W bosons decays leptonically and the other decays to jets. The decay topology is similar to that of $t\bar{t}$ events, but with only one b-jet in the final state. The presence of a leptonically decaying and a hadronically decaying W boson and a b-quark from the

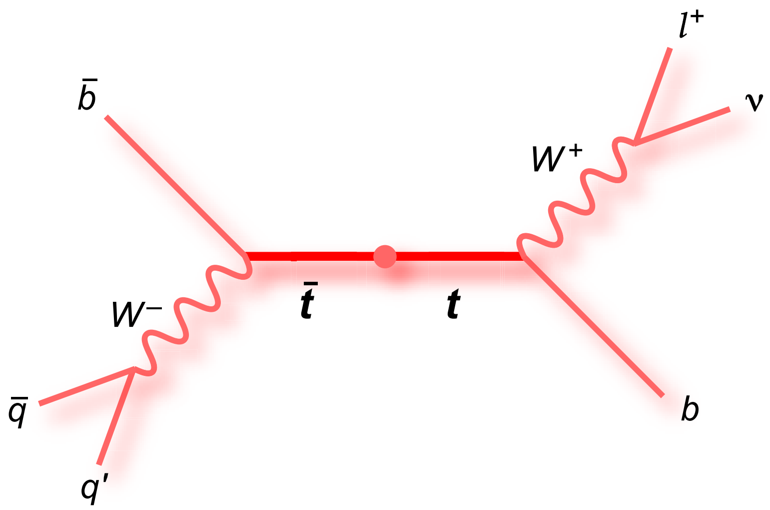


Figure 6.2: Schematic of the $t\bar{t}$ production process in the semileptonic decay topology. Source: https://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html

top quark decay, coupled with the non-negligible value of the c-quark mistagging efficiency makes this process approximately the third most important background.

WZ production has a cross-section approximately 250 times larger than signal, with approximately 23% of these events having a leptonically decaying W and a pair of quarks in the final state, from the Z boson decay. Of those events, 22% correspond to a Z boson decaying to a pair of b-quarks, which amounts to about 5% of the total WZ events, leading to a number of events in the signal region similar to that expected for the signal. If the Z boson is produced with a high transverse momentum, the two b-quarks can be reconstructed inside the large-R jet, making the final state topology exactly the same as of signal events. Since these are indistinguishable, $WZ \to \ell\nu b\bar{b}$ is called an irreducible background.

QCD-induced multijet production has a cross-section approximately 10^7 times larger than signal. In the final state, there can be a jet faking a lepton or a lepton from semileptonic decays of hadrons and missing transverse energy from resolution or acceptance effects. These events can also have two b-jets, one b-jet and one mistagged c- or light jet or even two mistagged c- or light jets. Fake lepton candidates are largely removed by lepton identification algorithms, while leptons from hadron decays are removed by lepton isolation and high transverse momentum criteria. Missing transverse energy from detector effects such as the ones mentioned above tends to have lower values than those characteristic of a neutrino from a boosted W boson, and a large fraction of these events are removed by requiring a large value of the missing transverse energy. All of the criteria mentioned above make this a negligible background.

6.2 Data and simulated samples

ATLAS data and MC samples are divided by periods, a, d and e, which correspond to different data-taking conditions, such as the detector state or the pile-up distributions, as described below.

Data description

The proton-proton collision data used in this analysis were collected by the ATLAS detector during the Run-2 of the LHC, corresponding to an integrated luminosity of 139.9 $\pm$ 2.4 fb^{-1} [74] of proton-proton collisions at a center-of-mass energy $\sqrt{s} = 13$ TeV. The total integrated luminosity corresponding to each of the periods, as well as the average number of collisions per bunch crossing ($\langle\mu\rangle$) is given in Table 6.1.

Period	Years	$\langle\mu\rangle$	Integrated Luminosity (fb^{-1})
a	2015-2016	18.5	36.1
d	2017	37.8	43.8
e	2018	36.1	59.9

Table 6.1: Integrated luminosities and average number of collisons per bunch crossing for the ATLAS data-taking periods during Run 2 of the LHC.

Simulated samples

Standard Model samples

The samples of SM signal and backgrounds were generated as described below for a $\sqrt{s} = 13$ TeV, using the Geant4-based ATLAS Simulation Infrastructure [61] for the detector description, as mentioned in Chapter 5. This process takes into account the different data-taking conditions in each period, and the samples are identified by $mc16$ and the period letter. Given that ATLAS MC samples are produced before or during a given period using an estimation of the true pile-up distribution, differences between the expected and the measured pile-up distribution have to be taken into account. This is done by measuring the true pile-up distribution in data after the end of a data-taking period, and deriving a set of weights to correct pile-up distributions at analysis level.

$qq \rightarrow WH \rightarrow \ell\nu b\bar{b}; \ell = e, \mu, \tau$ samples were generated using Powheg MiNLO [75] interfaced to Pythia8 [56] for parton showering and hadronization, with NNPDF3.0 NNLO parton distribution functions (PDFs) [76] and the AZNLO tune [77]. This sample is normalized to the best available production cross section, calculated at NNLO (QCD) and NLO (EW) with the Higgs boson mass fixed to 125 GeV and the $H \rightarrow b\bar{b}$ branching fraction fixed to 58%, according to the recommendations of the LHC Higgs Cross Section Working Group [13].

$t\bar{t}$ samples were generated with the Powheg-BOX framework [54] with NNPDF3.0 PDFs and interfaced with Pythia8 with NNPDF2.3 PDFs [78] and the A14 tune [79] for parton showering and hadronization. These were generated with a filter requiring that at least one of the W bosons decays leptonically, due to the negligible acceptance of this analysis to $t\bar{t}$ events in which both W bosons decay hadronically. This sample is normalized using cross-sections calculated at NNLO+NNLL in QCD.

Single top quark production in the Wt channel is generated without the aforementioned filter and using the same parameters as for the $t\bar{t}$ sample, and it is normalized to cross sections calculated at NLO in QCD. EvtGen [59] is used to generate the decays of B hadrons in the signal, $t\bar{t}$ and single top samples.

Semi-leptonic $qq \to VV$ samples $(V = W, Z)$ were generated using `Sherpa 2.2.1` [80] interfaced with NNPDF3.0 PDFs. These were generated at NLO accuracy for $VV + 0j$ and $VV + 1j$ final states and were combined with multi-leg LO matrix elements to obtain the $VV + 2, 3j$ final states. In order to reduce the MC statistical uncertainty in the irreducible $WZ \to \ell\nu b\bar{b}$ background, additional samples with this decay were generated and combined with the inclusive $WZ \to \ell\nu q\bar{q}$ sample, adjusting the event weights in order to not have double counting of overlapping events. In addition to the quark-induced diboson samples, semi-leptonic loop-induced $gg \to VV$ samples are generated using `Sherpa 2.2.2` with LO accurate matrix elements for $VV+0j$ and $VV+1j$ final states interfaced with NNPDF3.0 PDFs.

W+jets samples were generated with `Sherpa 2.2.1` [80] using the NNPDF3.0 PDFs with dedicated parton shower tuning developed by the `Sherpa` authors. This sample is normalized using cross-sections calculated at NNLO in QCD. This analysis is done in a phase space corresponding to high transverse momentum of the W boson and considering final states with b-tagged jets. It is therefore important to make sure that the MC statistics in this phase space are large enough to provide a robust MC prediction to be compared to data. To achieve this aim, the W+jets samples were generated in slices of $\max(H_T, p_{T_W})$, where H_T is the scalar sum of the transverse momenta of all jets in the event. Different filters are applied at generator-level to further split the samples, depending on the flavour composition of the jets produced in association with the W boson. The samples with $\max(H_T, p_{T_W}) > 500$ GeV are not separated by flavour composition and are used in the tagging efficiency studies described in Section 6.4.3.

All of these samples include the effect of pile-up, achieved by overlaying minimum bias events simulated using `Pythia8` with the A3 tune [79] and interfaced with NNPDF2.3 PDFs.

Table 6.2 contains the ATLAS simulated sample information for the relevant processes described above, such as the cross-sections, branching ratios and the number of generated events for each period. It should be noted that ATLAS samples are generated including electrons, muons and τ leptons.

Process	$\sigma \times BR$ [pb]	N_{events} $(mc16a)$	N_{events} $(mc16d)$	N_{events} $(mc16e)$
Signal				
$qq \to WH \to l^+\nu b\bar{b}$	0.1646	4 M	4 M	6.6 M
$qq \to WH \to l^-\nu b\bar{b}$	0.1045	2 M	2 M	3.3 M
Backgrounds				
W+jets, $W \to l\nu$	183600×0.325	489 M	598 M	836 M
$t\bar{t}$ non-fully-hadronic	831.76×0.543	175 M	205 M	261 M
Single-top Wt	71.7×1	20 M	48 M	66 M
$qq \to WZ \to l\nu qq$ (with $Z \to b\bar{b}$ extension)	50.3×0.227	7 M(6 M)	36 M(7 M)	12 M(10 M)

Table 6.2: ATLAS MC samples generated for the SM signal and background processes and the cross section times branching ratio $\sigma \times BR$ used to normalize the different processes at $\sqrt{s} =$ 13 TeV, where $l = e, \mu, \tau$. The last columns show the number of simulated events for each of the periods. Branching ratios correspond to the decays shown. For the $WZ \to \ell\nu q\bar{q}$ sample, the numbers in brackets represent the number of additional events generated with the $Z \to b\bar{b}$ decay.

Beyond Standard Model samples

BSM signal samples for the studies of the sensitivity to anomalous Spin/CP components in the HWW interaction vertex were generated at LO in QCD for $\sqrt{s} = 14$ TeV with Madgraph5_aMC@NLO [55], using the automatic decay width calculation feature and CTEQ6L1 PDFs [81] . The input model was a FeynRules [82] implementation of the extended interaction vertex defined in Equation 2.19. These samples were generated with the W boson decaying to electrons or muons (no τ leptons) and two filters were applied on the generator-level lepton, $p_T > 20$GeV and $|\eta| < 3$ (with an efficiency of $\approx 80\%$), due to the negligible acceptance of the analysis to this phase space. The Madgraph-level events were then interfaced with Pythia8 with the CTEQ6L1 PDFs, for generation of multi-parton interactions, underlying event, parton showering, hadronization and hadron decays.

Given that, when this work was being carried out, no ATLAS samples with anomalous components were available (and their production would take longer than time necessary to complete this work), detector simulation was done using using Delphes3 [37] with a modified version of the default ATLAS card described in Section 5.2. Just before running detector simulation, these samples were then overlaid with a sample of pile-up events generated with Pythia8, with $\langle\mu\rangle = 50$. The several interaction vertices are spread in the spatial component along the direction of the beam pipe and in the time component according to Gaussian distributions. The standard deviations are $53\ \mu\mathrm{m}$ and 160 ps for the spatial and time component, respectively.

These samples were then normalized to the target luminosity, using cross-sections calculated at LO for $\sqrt{s} = 13$ TeV with the aforementioned filters, and multiplied by a k-factor, given in Equation 6.1, to take into account higher order effects in the production vertex. This factor is calculated from the SM cross-section calculation at NNLO (QCD) + NLO (EW), given in Ref. [13].

$$k = \frac{\sigma_{\text{NNLO QCD + NLO EW}}}{\sigma_{\text{LO}}} = \frac{178.49}{151.4} = 1.18 \tag{6.1}$$

Table 6.3 shows the anomalous components (and couplings) in the different samples, the (filtered) cross-sections and the number of generated events for each of them. The uncertainty on the cross-sections is derived in Madgraph5_aMC@NLO and comes mainly from PDF uncertainties. It should be noted that the change in cross-section does not depend on the sign of c_W.

Couplings	$\sigma \times BR$ [fb]	N_{events}
$a_w = 0$ (SM-like)	151.4 $\pm$ 0.4	50000
$a_w = 0, b_{w1} = 0.05$	211.7 $\pm$ 0.7	50000
$a_w = 0, b_{w1} = 0.1$	275.1 $\pm$ 0.9	50000
$a_w = 0, c_w = 0.05$	154.1 $\pm$ 0.4	50000
$a_w = 0, c_w = 0.1$	159.9 $\pm$ 0.5	50000

Table 6.3: BSM samples and cross section times branching ratio (BR) used to normalise the different processes at 13 TeV. The last column shows the number of simulated events for each of the samples. b_{w1} (c_w) is the coupling of the CP-even (CP-odd) component, defined in Equation 2.19. The uncertainty on the cross-sections is derived in Madgraph5_aMC@NLO and comes from PDF uncertainties.

Luminosity weighting

To normalize the samples to a desired luminosity, L, their generator weight is multiplied by a luminosity weight, w_L, defined as a function of the number of generated events, N_{events}, cross-sections, σ, and the desired luminosity, as shown in Equation 6.2.

$$w_L = \frac{L\sigma}{N_{events}} \tag{6.2}$$

6.3 Statistical analysis methodology

In order to measure the signal yield, a binned likelihood fit is performed to the distribution of the invariant mass of the leading (highest transverse momentum) large-R jet in all the analysis regions simultaneously. The implementation is done using the `WorkSpaceMaker` framework, which builds upon `RooFit` [83], `RooStats` [84] and `HistFactory` [85], and was used in the ATLAS observation of VH production and of the $H \to b\bar{b}$ decay [27].

Since the base of this analysis is a counting experiment, the likelihood is built from a product of Poisson probability terms, given in Equation 6.3.

$$\mathcal{L}(\mu) = \prod_{j \in \text{bins}} \text{Poisson}\left(N_j | \mu s_j + b_j\right) = \prod_{j \in \text{bins}} \frac{\left(\mu s_j + b_j\right)^{N_j}}{N_j!} e^{-(\mu s_j + b_j)}, \tag{6.3}$$

where N_j is the number of data events in bin j, s_j and b_j are the expected signal and background yields derived from MC samples, and μ is the signal strength, the parameter of interest to be extracted.

Finite MC statistics for the background processes naturally lead to an imperfect representation of the background distributions, contributing to the uncertainty. This is addressed by introducing γ-parameters in each bin which modify the expected background yields as $b_j \to \gamma_j b_j$ and are constrained by a product of Gaussian distributions, described in Equation 6.4.

$$\mathcal{L}_{\text{MCStat}}\left(\vec{\gamma}\right) = \prod_{j \in \text{bins}} \text{Gauss}\left(\beta_j | \gamma_j \beta_j, \sqrt{\gamma_j \beta_j}\right), \tag{6.4}$$

where $\beta_j = \frac{1}{\sigma_{rel}^2}$ and σ_{rel} is the relative statistical uncertainty on the expected total background yield.

Systematic uncertainties are associated with the nature of the measurement apparatus, assumptions made and the models used to generate the MC samples. These can be divided in two groups: (detector) uncertainties such as those on the large-R jet energy or mass, and modelling uncertainties, that take into account uncertainties from the different generators. These are parameterized as a set of nuisance parameters, $\vec{\theta}$, that modify the expected signal and background yields. Their effect is usually determined from *auxiliary measurements* and implemented in the likelihood via a set of Gaussian terms, with mean zero and variance of unity, described in Equation 6.5. These terms enable constraining the effect of each nuisance parameter to the certainty with which its nominal value is known.

$$\mathcal{L}_{\text{Syst}} = \prod_{\theta \in \vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2} \tag{6.5}$$

The full likelihood is written in Equation 6.6.

$$\mathcal{L}(\mu, \vec{\gamma}, \vec{\theta}) = \prod_{j \in \text{bins}} \text{Poisson}\left(N_j | \mu s_j(\vec{\theta}) + \gamma_j b_j(\vec{\theta})\right) \times \prod_{j \in \text{bins}} \text{Gauss}\left(\beta_j | \gamma_j \beta_j, \sqrt{\gamma_j \beta_j}\right) \times \prod_{\theta \in \vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2} \tag{6.6}$$

The signal strength and its uncertainty are extracted by a maximum-likelihood procedure, profiling over the systematic uncertainties. The *p-value*, p, quantifies the compatibility between the background-only hypothesis $(\mu = 0)$ and the data. It is derived from the log-likelihood ratio test-statistic, defined in Equation 6.7.

$$q_\mu = 2 \ln \left(\frac{\mathcal{L}(\mu, \hat{\hat{\vec{\gamma}}}, \hat{\hat{\vec{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\vec{\gamma}}, \hat{\vec{\theta}})} \right), \tag{6.7}$$

where $\left(\hat{\mu}, \hat{\vec{\gamma}}, \hat{\vec{\theta}}\right)$ are the values of the parameters that maximize the likelihood and $\left(\hat{\hat{\vec{\gamma}}}, \hat{\hat{\vec{\theta}}}\right)$ are the values of the parameters that maximize the likelihood for a fixed μ. The asymptotic approximation of a non-central χ^2 distribution for the probability density function of the q_μ, described in Ref. [86] is used, leading to the identity $p_\mu = \sqrt{q_\mu}$. The *p-value* is given by $p = p_0 = \sqrt{q_0}$ and is extracted by comparing the expected signal and background distributions (or the observed data distributions) with a pseudo-data set matching the null (background-only) hypothesis, also called the Asimov dataset.

The statistical significance of the result, Z, is quoted in terms of (Gaussian) standard deviations as $Z = \Phi^{-1}(1-p)$, where Φ is the inverse of the Gaussian cumulative distribution function. Since the determination of the likelihood distribution is a computationally intensive process, one can approximate the statistical significance by the median (statistical) significance, Z_{med}, defined in Ref. [86] as $Z_{med} = \sqrt{2\,(s+b)\,ln\,(1+s/b) - s}$, where s and b are the expected signal and background yields, respectively (derived from MC samples).

Background normalization factors

The fitting procedure mentioned above also allows the determination of a set of scale factors for the different backgrounds, in order to match the expected and the observed background yields. These are extracted from data for all backgrounds simultaneously, using dedicated control regions or the sidebands of the invariant mass distribution, where no signal is expected, and are left as free parameters in the fit. To be noted that these are not the γ parameters described above.

6.4 Event Selection

Before any event selection, a set of conditions is applied to data, to remove events which may have been poorly reconstructed. First, the event should have happened in a period in which all ATLAS subdetectors were operational and providing good quality data, checked by means of a Good Runs List [2](described in Section 4.2.4) and, in addition, it should have a primary vertex.

Events are recorded with single electron triggers in the electron channel and with $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers in the muon channel, due to its high acceptance for the typically large values of the missing transverse energy from a boosted W boson, compared to the relatively low acceptance of the L1 muon trigger (approximately 80%). The selected triggers are applied to data and ATLAS MC samples, after reconstruction. The triggers used in this analysis are presented in Tables 6.4 and 6.5 and were shown to have an efficiency of 100% in the targeted phase-space.

Trigger name	Period	Threshold [GeV]
HLT_e24_lhmedium_L1EM20VH	2015	24
HLT_e60_lhmedium	2015	60
HLT_e120_lhloose	2015	120
HLT_e26_lhtight_nod0_ivarloose	2016-2018	26
HLT_e60_lhmedium_nod0	2016-2018	60
HLT_e140_lhloose_nod0	2016-2018	140

Table 6.4: Single electron triggers used during the 2015, 2016, 2017 and 2018 data collection periods. L1EM20VH: level 1 calibrated at the electromagnetic scale with a threshold of 20 GeV and a veto on events with hadronic calorimeter depositions above a variable threshold. lhloose: Loose ID likelihood required. lhmedium: Medium ID likelihood required. lhtight: Tight ID likelihood required. ivarloose: Variable cone loose isolation requirement.

Trigger name	Period	Threshold [GeV]
HLT_xe70	2015	70
HLT_xe90_mht_L1XE50	2016	90
HLT_xe110_mht_L1XE50	2016	110
HLT_xe110_pufit_L1XE55	2017	110
HLT_xe110_pufit_xe70_L1XE50	2018	110

Table 6.5: $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers used during the 2015, 2016, 2017 and 2018 data collection periods, seeded using calorimeter-based MET triggers (calibrated at the electromagnetic scale) with a threshold of 50 GeV(L1XE50) or 55 GeV(L1XE55).

Events are required to have exactly one lepton which passes the criteria defined in Table 5.1, and $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV.

The transverse momentum of the W boson, p_{T_W}, is a measure of the energy scale of the events that can be unambiguously reconstructed for signal and backgrounds, since all of them have a leptonically decaying W boson. It is obtained from the vectorial sum of the missing transverse energy and the transverse momentum of the lepton, and the events are required to have $p_{T_W} > 250$ GeV.

In order to select events consistent with a boosted $H \to b\bar{b}$ decay, at least one trimmed large-R jet

[2]data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml

with $p_T > 250$ GeV and $|\eta| < 2.0$, containing at least two matched track jets with $p_T > 10$ GeV and $|\eta| < 2.5$ is required. Throughout this work, subjets are defined as the set of track jets matched to the leading large-R jet. The matching can be done either by ghost-association - used in the studies of the boosted Higgs boson tagging strategy and event selection optimization shown in Section 6.4 - or by geometrical matching, $\Delta R\,(\text{track jet-large-}R\text{ jet}) < 1.0$ - used in the study of the sensitivity to anomalous Spin/CP components, shown in Section 6.6.

6.4.1 Studies of boosted Higgs boson tagging strategy

Since this is the first ATLAS physics analysis targeting the $WH \to \ell\nu b\bar{b}$ process in the boosted regime, it requires a study of the techniques developed in ATLAS to identify boosted $H \to b\bar{b}$ decays. The studies presented in this section aim at determining the optimal track jet reconstruction technique and which subjets to check for b-tagging. These studies use ATLAS MC samples from period e, described in Table 6.2, corresponding to a luminosity of 59 fb^{-1} . Only events which pass the trigger, lepton, missing transverse energy and large-R jet criteria described previously are taken into account. The median significance, Z_{med}, defined in Section 6.3, is used as a figure of merit to compare between the different possibilities.

The first number to verify is the fraction of events in which the angular distance (defined in Equation 5.4) between the truth-level Higgs boson and the leading large-R jet is smaller than its radius, $\Delta R\left(H^{\text{truth}}, \text{large}-\text{R jet}\right) < 1.0$. The $\Delta R\left(H^{\text{truth}}, \text{large}-\text{R jet}\right)$ distribution is shown in Figure 6.3, where it can be seen that, for a non-negligible fraction of events, around 20%, this does not happen. This is observed in the three MC periods, and can be due to the Higgs boson being produced outside the large-R jet acceptance. Around 1% of the events with $\Delta R\left(H^{\text{truth}}, \text{large}-\text{R jet}\right) > 1.0$ end up in the signal region. In what follows, signal events are required to have $\Delta R\left(H^{\text{truth}}, \text{large}-\text{R jet}\right) < 1.0$ to ensure that the efficiencies are calculated for jets that contain a Higgs boson and not for large-R jets which may come from pile-up.

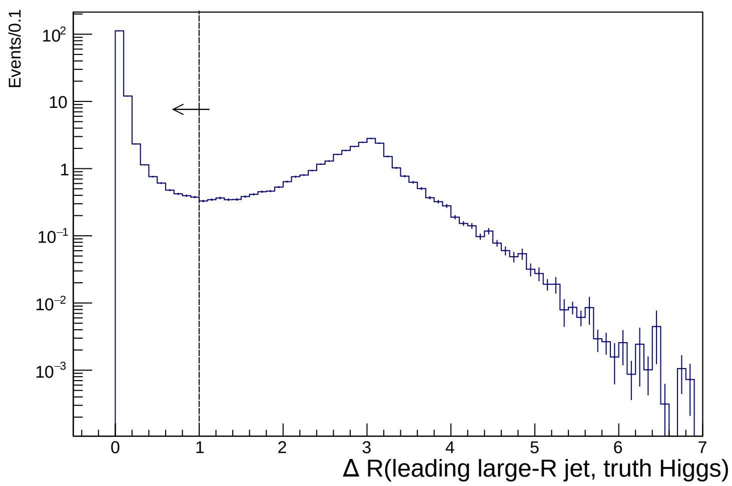


Figure 6.3: Angular separation between the truth-level Higgs boson and the leading large-R jet. A dashed line is drawn in $\Delta R\left(H^{\text{truth}}, \text{large}-\text{R jet}\right) = 1.0$.

Fixed-radius vs. variable-radius track jet reconstruction

The two possible track jet reconstruction techniques available in ATLAS are fixed-radius or variable-radius reconstruction, both described in Section 5.2.3. The use of fixed-radius and variable-radius track jets in boosted $H \to b\bar{b}$ tagging in ATLAS are described in Ref. [70] and [87], respectively. In the latter, variable-radius track jets are shown to have an improvement with regards to fixed-radius track jets, but since it was performed on the very high p_T regime typical of BSM scenarios, $(p_T^H > 1$ TeV$)$, it is necessary to perform a similar study in the energy regime relevant to this analysis, 250 GeV $< p_T^H <$ 600 GeV.

The distribution of the large-R jet invariant mass (calculated using the combined mass definition given in Section 5.2), for signal events that pass large-R jet selection is shown in Figure 6.4. It can be seen that this distribution contains a peak at low values, from events in which at least one of the B hadrons does not end up inside the large-R jet. This can be confirmed in Figure 6.5a, which shows the same distribution for events with a single subjet and is largely dominated by the low mass peak. Given that this peak corresponds to events where a Higgs boson can't be identified, the results derived in this study are constrained to the Higgs boson mass window, $70 < m_J < 150$ GeV.

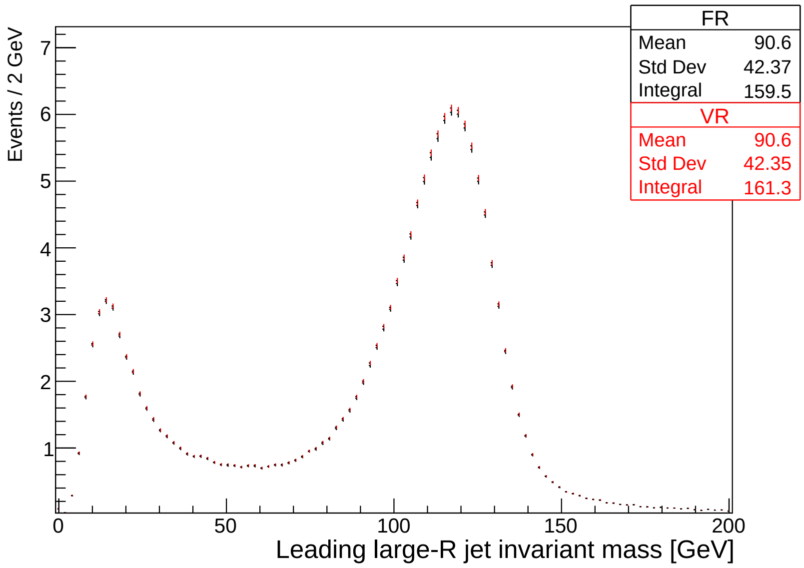
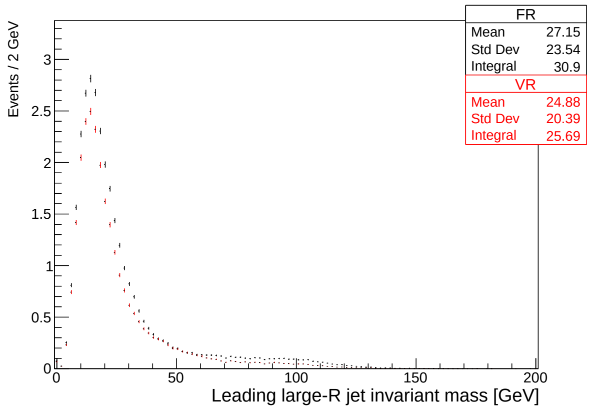


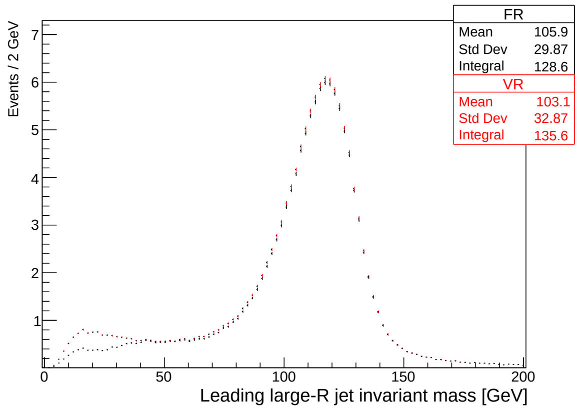
Figure 6.4: Leading large-R jet invariant mass (calculated using the combined mass definition) for signal events, after large-R jet selection.

To compare the two techniques, three values are calculated in signal events:

- ϵ_{2sub}, defined in Equation 6.8 as the fraction of events with at least one large-R jet which have at least two subjets
- the double b-labelling fraction, f_{2blab}, defined as the fraction of events with at least one large-R jet and two subjets which have at least two b-labelled subjets, where a b-labelled subjet is one for which $\Delta R\,(\text{truth B-hadron, subjet}) \leq 0.3$ and $p_T > 5$ GeV
- the double b-tagging efficiency, ϵ_{2tag}, defined in Equation 6.10 as the fraction of events with at



(a) 1 subjet



(b) $\geq$ 2 subjets

Figure 6.5: Leading large-R jet invariant mass distribution for signal events with one subjet (left) and at least two subjets (right)

least one large-R jet, two subjets and two b-labelled subjets in which the two leading subjets are b-tagged with MV2c10 in the 70% efficiency working point

$$\epsilon_{2sub} = \frac{N\,(\geq 1 \text{ large-}R \text{ jet}, \geq 2 \text{ subjets})}{N\,(\geq 1 \text{ large-}R \text{ jet})} \tag{6.8}$$

$$f_{2blab} = \frac{N\,(\geq 1 \text{ large-}R \text{ jet}, \geq 2 \text{ subjets}, \geq 2 \text{ b-labelled subjets})}{N\,(\geq 1 \text{ large-}R \text{ jet}, \geq 2 \text{ subjets})} \tag{6.9}$$

$$\epsilon_{2tag} = \frac{N\,(\geq 1 \text{ large-}R \text{ jet}, \geq 2 \text{ subjets}, \geq 2\ b\text{-labelled subjets}, 2\ b\text{-tagged subjets})}{N\,(\geq 1 \text{ large-}R \text{ jet}, \geq 2 \text{ subjets}, \geq 2\ b\text{-labelled subjets})}, \tag{6.10}$$

The results for ϵ_{2sub}, f_{2blab} are summarized in the first two lines of Table 6.6, where it can be seen that both techniques have a similar reconstruction performance. The value of ϵ_{2tag} is given in the last line of the same table, showing that the b-tagging efficiencies for fixed-radius track jets are larger than those for variable-radius track jets. Figure 6.6 shows the distribution of the double b-tagging efficiency in signal events, as a function of the truth-level transverse momentum of the Higgs boson, from which one can conclude that not only is the double b-tagging efficiency for fixed-radius track jets larger than that for variable-radius track jets, but that it also decreases with the truth-level transverse momentum of the Higgs boson.

	Fixed-radius	Variable-radius
ϵ_{2sub} (%)	97.9	98.9
f_{2blab} (%)	87.6	89.7
ϵ_{2tag} (%)	49.6	43.5

Table 6.6: ϵ_{2sub}, f_{2blab} and ϵ_{2tag} in signal events, for fixed- and variable-radius reconstruction, as defined in Equations 6.8, 6.9 and 6.10.

The next step is to calculate the Higgs tagging efficiency, ϵ (defined in Equation 6.11), for signal,

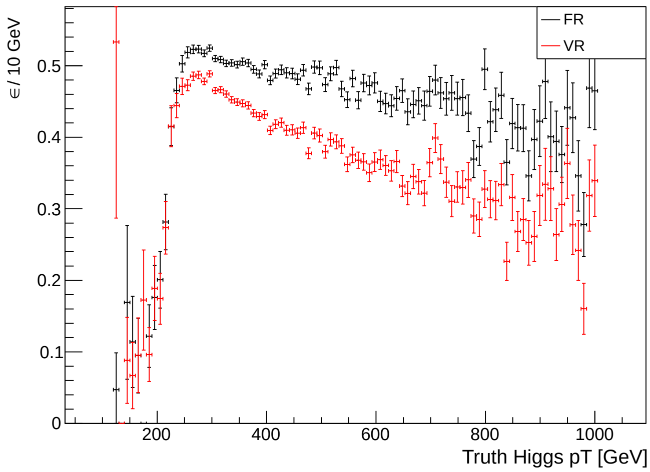


Figure 6.6: Distribution of the double b-tagging efficiency in signal events, ϵ_{2tag}, defined in Equation 6.10, as a function of the truth-level transverse momentum of the Higgs boson.

which corresponds to the fraction of events with at least one large-R jet which pass the criteria for identification of a boosted Higgs boson - at least two subjets and two b-tags in the two leading subjets. The two main backgrounds in this channel, $t\bar{t}$ production and $W+$jets production, are taken into account using the $t\bar{t}$ sample and the jet-flavour inclusive $W+$jets sample with the high-p_{T_W} filter described in Section 6.2. For backgrounds, the requirement $\Delta R\left(H^{\text{truth}}, \text{large} - \text{R jet}\right) < 1.0$ is not applied, since in these events no Higgs boson is present. In order to remove $t\bar{t}$ contamination, signal and background events with b-tagged track jets which are not matched to the leading large-R jet are removed (this selection condition is motivated in Section 6.4.2).

$$\epsilon = \frac{N\,(\geq 1 \text{ large-}R \text{ jet}, \geq 2 \text{ subjets}, 2\ b\text{-tagged subjets})}{N\,(\geq 1 \text{ large-}R \text{ jet})} \tag{6.11}$$

Table 6.7 contains the Higgs tagging efficiency in signal events, background rejection factors for the mentioned backgrounds (defined as the inverse of the tagging efficiency) and the median significance for fixed- and variable-radius reconstruction. From it, one can infer that fixed-radius track jets have a slightly higher tagging efficiency (around 4%), even though the reconstruction efficiency is approximately 2% larger for variable-radius track jets, explained by the higher b-tagging efficiencies of fixed-radius track jets. One can also see that, while having a lower tagging efficiency in signal, the background rejection is higher when using variable-radius track jets, which is expected, since a lower b-tagging efficiency is generally associated with a larger c- and light jet rejection. These results lead to the conclusion that the best choice is to use variable-radius track jets, which has a median significance approximately 4% higher when compared to fixed-radius track jets.

The distribution of the leading large-R jet invariant mass and of the transverse momentum of the truth-level Higgs boson for signal events which have passed the identification criteria of a boosted Higgs boson are shown in Figure 6.7. In this figure it can be seen that this analysis covers the Higgs transverse momentum range 250 GeV $< p_T^H < 600$ GeV, as expected.

	Fixed-radius	Variable-radius
ϵ (%)	43.7	39.1
$t\bar{t}$ rejection factor	100.0	142.9
W+jets rejection factor	184.3	219.5
Z_{med}	0.861	0.898

Table 6.7: Higgs tagging efficiencies, rejection factors for $t\bar{t}$ and W+jets production and median significance for fixed-radius and variable-radius track jets.

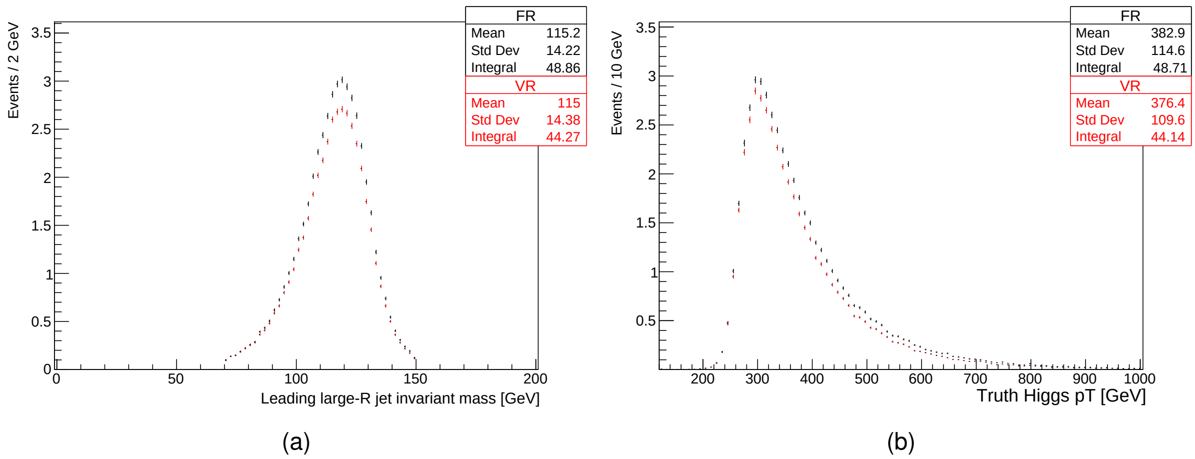


Figure 6.7: Leading large-R jet invariant mass (left) and truth-level Higgs transverse momentum (right) in signal events that pass all the Higgs tagging criteria.

b-tagging strategy

For this analysis, there are two possibilities for the b-tagging strategy: requiring exactly two b-tags either in the two leading subjets or in all the subjets. The study to evaluate the best option, shown below, is done using variable-radius track jets. Ref. [87] claims that for a non-negligible number of boosted $H \to b\bar{b}$ events reconstructed with variable-radius track jets, the subleading B hadron is reconstructed inside the sub-sub-leading subjet (third highest transverse momentum). Thus, requiring two b-tags in all the subjets is expected to accept more signal and background events.

The Higgs tagging efficiency, background rejection factors and (median) statistical significance are calculated for the two b-tagging strategies in the same conditions as the previous study and are shown in Table 6.8. In this table, it can be seen that the Higgs tagging efficiency in signal is higher when checking all of the subjets, but that the background rejection factors are larger when checking only the two leading subjets. It was verified that, even when requiring exactly two b-tags in all the subjets, variable-radius reconstruction is still favored with regards to fixed-radius reconstruction. The value of the significance leads to the conclusion that the best choice is to select events in which the 2 leading subjets are b-tagged, which leads to a median significance approximately 15% higher when compared to requiring 2 b-tags in all the subjets.

This study concludes that the optimal strategy for identifying the boosted Higgs boson is to use variable-radius reconstruction for the track jets, and select events in which the two leading subjets are b-tagged.

	Two leading subjets	All subjets
ϵ (%)	39.1	42.6
$t\bar{t}$ rejection factor	142.9	83.3
W+jets rejection factor	219.5	144.5
Z_{med}	0.898	0.766

Table 6.8: Comparison of signal efficiencies, background rejection factors and statistical significances for both tagging strategies using variable-radius track jets

6.4.2 Event categorization

Only events that pass the selection criteria defined previously are taken into account in this analysis.

Given that the sensitivity to anomalous components is expected to increase with the targeted transverse momenta of the W boson, the selected events are divided in two p_{T_W} regions: a medium p_{T_W} region with 250 GeV $< p_{T_W} < 400$ GeV , and a high p_{T_W} region with $p_{T_W} > 400$ GeV , chosen to match the Simplified Template Cross Section binning recommendations in Ref. [13].

The number of b-tagged track jets with $p_T > 10$ GeV which are not matched to the leading large-R jet is used to categorize the events into a signal region and a top control region. The signal region contains events with zero additional b-tagged track jets outside the large-R jet and has a high signal purity. The control region contains at least one additional b-tagged track jet not matched to the large-R jet and is very pure in $t\bar{t}$ events, since the b-quark from the leptonically decaying top quark is generally reconstructed outside the large-R jet, as schematized in Figure 6.8.

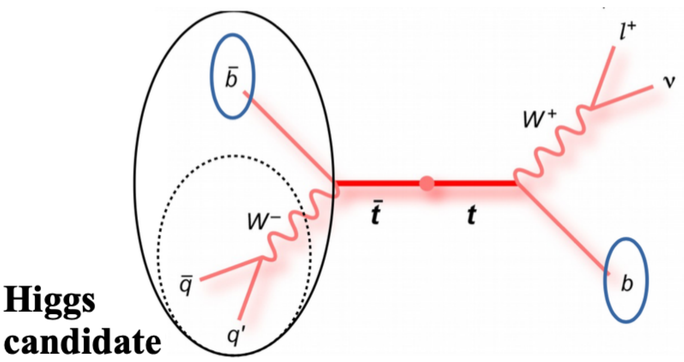


Figure 6.8: Schematic of the $t\bar{t}$ decay.

The number of small-R jets, i.e. anti-k_t $R = 0.4$ calorimeter jets with $p_T > 30$ GeV, which have $\Delta R\,(\text{small-R jet, large-}R\text{ jet}) > 1.0$ is used to split the events into high purity and low purity signal regions. The high purity signal region contains zero additional small-R jets, while the low purity signal region contains at least one additional small-R jet. The two regions have different values of the signal to background ratio, approximately 4.5% (1.1%) in the high (low) purity signal region.

The low mass peak observed in Figure 6.5 does not vanish when requiring two b-tags in the two leading subjets. Furthermore, ongoing work by the ATLAS Collaboration has shown that the combined mass calibration for large-R jets with invariant mass smaller than 50 GeV should not be applied. Due to this, a lower limit of 50 GeV on the leading large-R jet invariant mass is applied.

6.4.3 Event selection optimization studies

This section contains a study on the use of variables such as the transverse mass of the W boson, m_{T_W} and the difference in rapidity between the W boson and the Higgs candidate (leading large-R jet), $\Delta y(W,H)$, as variables to improve signal and background separation. In the following, bins where the expected number of signal events is at least 5% larger than the expected number of background events do not show a data point (blinding), in order to avoid biasing the selection. For these studies, the expected statistical significance, Z_{exp}, defined in Section 6.3, is used as a figure of merit and no systematic uncertainties are taken into account.

Study of an upper limit in the W boson transverse mass

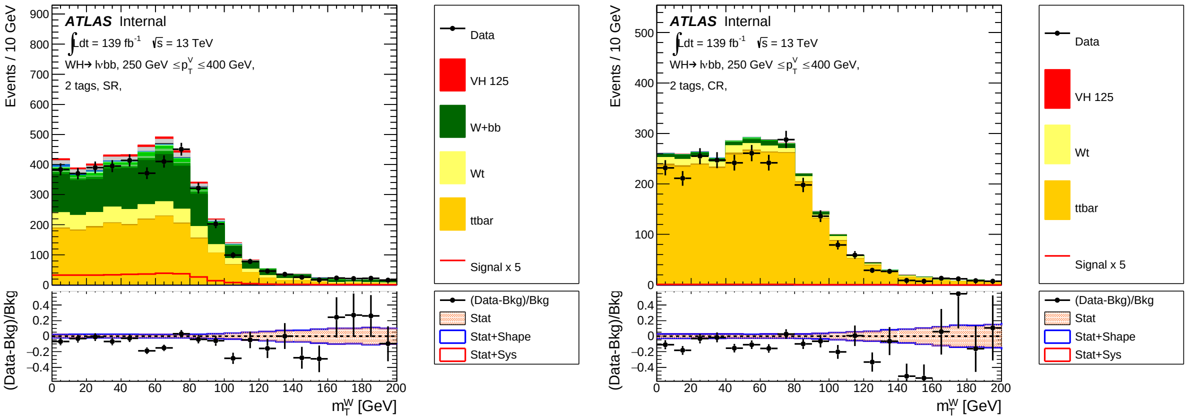
The W boson transverse mass, m_{T_W} , is defined in Equation 6.12.

$$m_{T_W}^2 = 2p_{T_\ell} E_{\mathrm{T}}^{\mathrm{miss}} \left(1 - \cos\left(\Delta\phi_{\ell, E_{\mathrm{T}}^{\mathrm{miss}}}\right)\right), \tag{6.12}$$

where $\Delta\phi_{\ell, E_{\mathrm{T}}^{\mathrm{miss}}}$ is the difference between the azimuthal angle of the lepton and of the missing transverse energy. In events with a real W boson, the distribution of this quantity has an endpoint at $m_{T_W} = m_W \approx 80$ GeV, degraded only by effects of detector resolution or acceptance. The previous study of this channel, done in a lower energy regime [27], includes an upper limit of $m_{T_W} < 120$ GeV, claimed to remove some of the $t\bar{t}$ contamination. To evaluate if a selection condition based on this variable is still useful for the boosted analysis, the modelling of this variable in the top control region was studied, as well as the expected significance for different values of the upper limit. The full Run-2 ATLAS data and MC samples were used, without taking into account the splitting of the signal region into high and low purity regions.

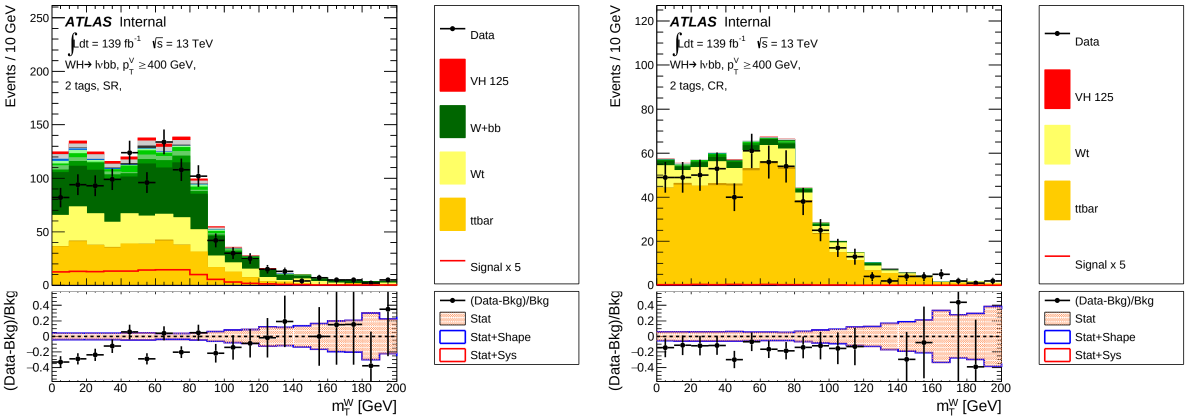
To study the modelling of this variable, data-MC comparisons in the signal and control regions defined in Section 6.4.2 (without the separation of the signal region) were done, as shown in Figures 6.9 and 6.10. From these figures, and excluding the top background normalization factor that will be fixed by the global fit, it can be seen that the ratio between data and simulation in the top control region fluctuates around unity with the difference most likely covered by systematics, similar to what happens in the signal region for 250 GeV $< p_{T_W} < 400$ GeV . The slope at lower values of m_{T_W} in the $p_{T_W} > 400$ GeV signal region, shown in Figure 6.10a can be explained by the small number of events in this region. These lead to the conclusion that this variable is sufficiently well modelled for signal and backgrounds, and can be studied further.

In order to test the effect of an upper limit on this variable, samples were generated with the value of the cut used in the study mentioned above, $m_{T_W} < 120$ GeV, and two values around it: $m_{T_W} < 100$ GeV and $m_{T_W} < 130$ GeV. The expected significance for the different values is shown in Table 6.9, where it can be seen that an upper limit in the transverse mass of the W boson does not show any improvement and will not be used in this analysis.



(a) (b)

Figure 6.9: Distribution of reconstructed W boson transverse mass in the signal region (left) and control region (right) for 250 GeV $< p_{T_W} < 400$ GeV . Only backgrounds that add up to more than 5% of the total background composition are shown in the labels. $VH125$ corresponds to the WH signal.



(a) (b)

Figure 6.10: Distribution of reconstructed W boson transverse mass in the signal region (left) and control region (right) for $p_{T_W} > 400$ GeV. Only backgrounds that add up to more than 5% of the total background composition are shown in the labels. $VH125$ corresponds to the WH signal.

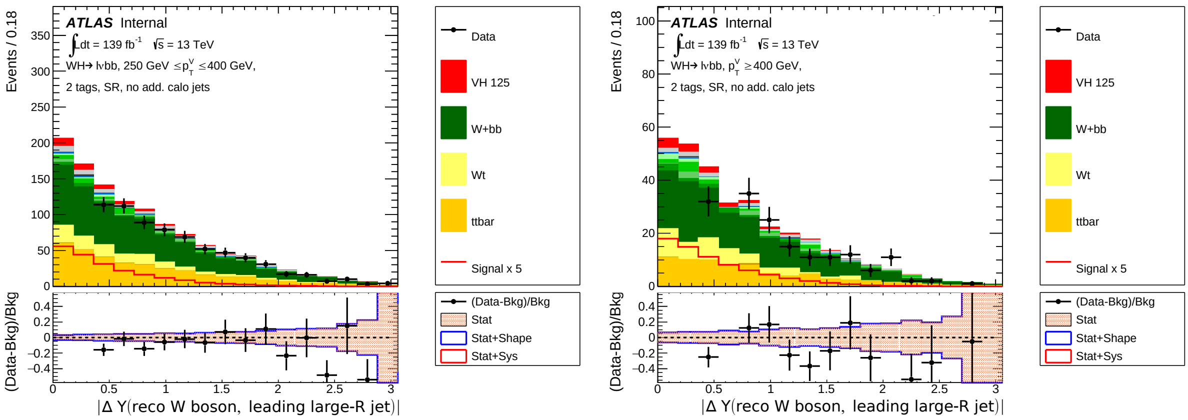
Upper limit	Z_{exp}
No limit	1.955
100 GeV	1.866
120 GeV	1.947
130 GeV	1.951

Table 6.9: Expected significances for the different values of the upper limit in the transverse mass of the W boson.

Study of an upper limit in the absolute difference in rapidity between the W and the Higgs boson

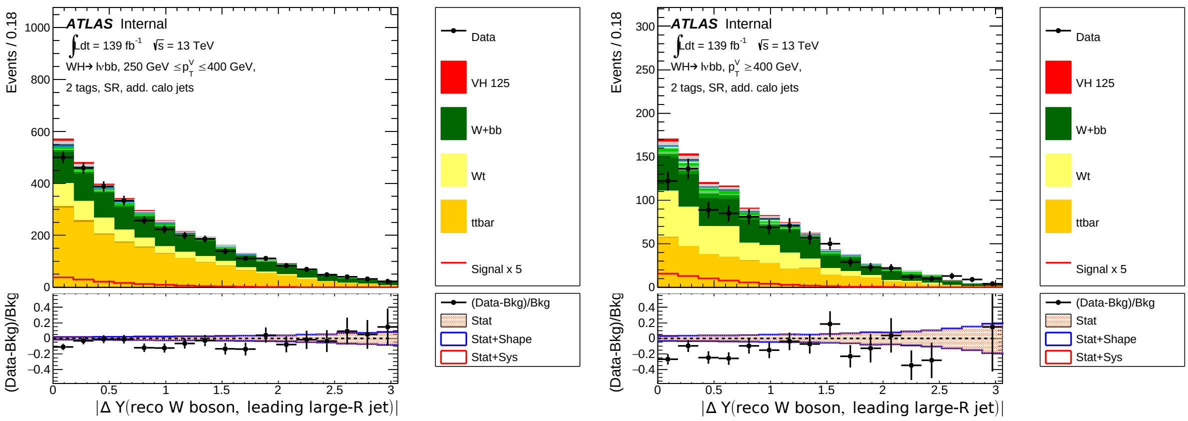
The absolute difference between the rapidity of the reconstructed W boson and the Higgs boson candidate, $|\Delta y(W, H)|$, was previously studied and used to discriminate between signal and backgrounds in the paper that led to the observation of the $H \to b\bar{b}$ decay and of VH associated production [27]. Since that paper was focused on a lower energy regime, the use of this variable to improve signal and background separation in the boosted regime needs to be demonstrated. This study is done with the event categorization into high and low purity signal regions and top control region, using the full Run-2 ATLAS data and MC samples.

The modelling of this variable has been studied by performing data-MC comparisons, in the high low purity signal regions, and the top control region, as shown in Figures 6.11, 6.12 and 6.13, respectively. In these figures, apart from the background normalization factors that will be fixed by the global fit, a good agreement between data and MC is observed, with the differences most likely covered by systematic uncertainties.



(a) (b)

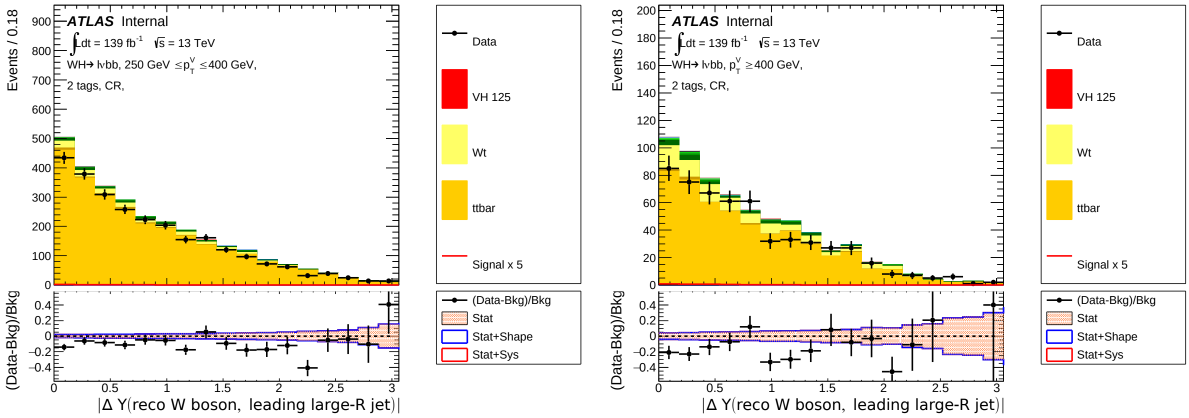
Figure 6.11: Distribution of $|\Delta y(W,H)|$ in the high purity signal region, for 250 GeV $< p_{T_W} <$ 400 GeV (left) and $p_{T_W} > 400$ GeV (right). Only backgrounds that add up to more than 5% of the total background composition are shown in the plot labels. $VH125$ corresponds to the WH signal.



(a) (b)

Figure 6.12: Distribution of $|\Delta y(W,H)|$ in the low purity signal region, for 250 GeV $< p_{T_W} < 400$ GeV (left) and $p_{T_W} > 400$ GeV (right). Only backgrounds that add up to more than 5% of the total background composition are shown in the plot labels. $VH125$ corresponds to the WH signal.

Two benchmark limits of $|\Delta y(W,H)| < 1.0$ and $|\Delta y(W,H)| < 1.5$ were tested, to evaluate a possible gain in the expected significance. Both of these limits increase the expected significance by approximately 5%. To harmonize with the $Z \to \ell\ell$ channel in the ATLAS analysis for which this study was carried out (boosted $VH, H \to b\bar{b}$ analysis, mentioned in Chapter 1), an upper limit of $|\Delta y(W,H)| < 1.4$ is



(a) (b)

Figure 6.13: Distribution of $|\Delta y(W,H)|$ in the control region, for 250 GeV $< p_{T_W} < 400$ GeV (left) and $p_{T_W} > 400$ GeV (right). Only backgrounds that add up to more than 5% of the total background composition are shown in the plot labels. $VH125$ corresponds to the WH signal.

implemented, after verifying that the gain in significance is similar to those obtained with $|\Delta y(W,H)| < 1.0$ and $|\Delta y(W,H)| < 1.5$. The results for $|\Delta y(W,H)| < 1.4$ are given in Table 6.10, showing a 4.3% increase in the expected significance, Z_{exp}.

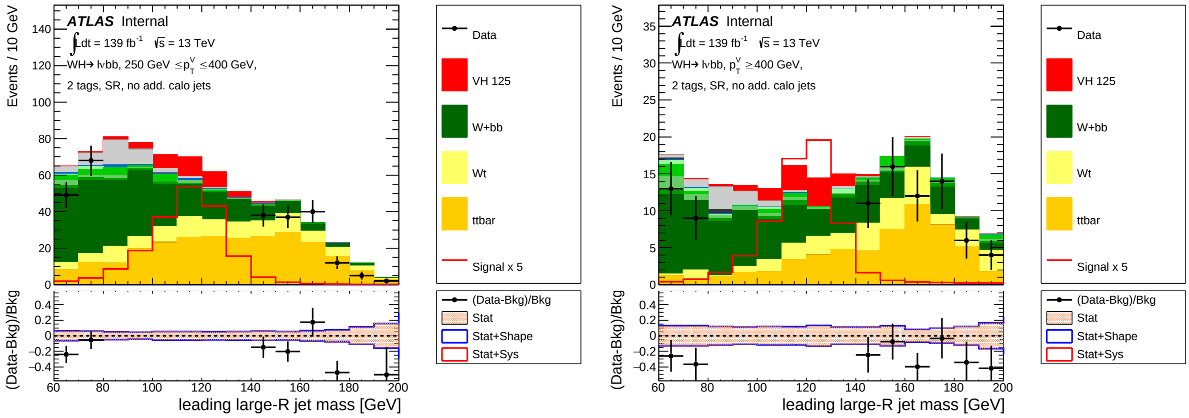
	Total (pre-cut)	**Total (post-cut $|\Delta y(W,H)| < 1.4$)**	**Difference (%)**
Signal yields	97.53 $\pm$ 9.88	91.94 $\pm$ 9.59	-5.7
Background yields	5913.05 $\pm$ 76.90	4695.9 $\pm$ 68.53	-20.6
Z_{med}	1.26	1.34	6.3
Z_{exp}	2.56	2.67	4.3

Table 6.10: Yields for signal and background in the signal region and fit significances before and after the requirement of an upper limit $|\Delta y(W,H)| < 1.4$. Statistical uncertainties on the yields are shown.

The Higgs candidate mass distributions after imposing this selection condition are shown in Figure 6.14, where it can be seen that it does not introduce any significant distortion or sculpting, and that data-MC agreement is maintained.

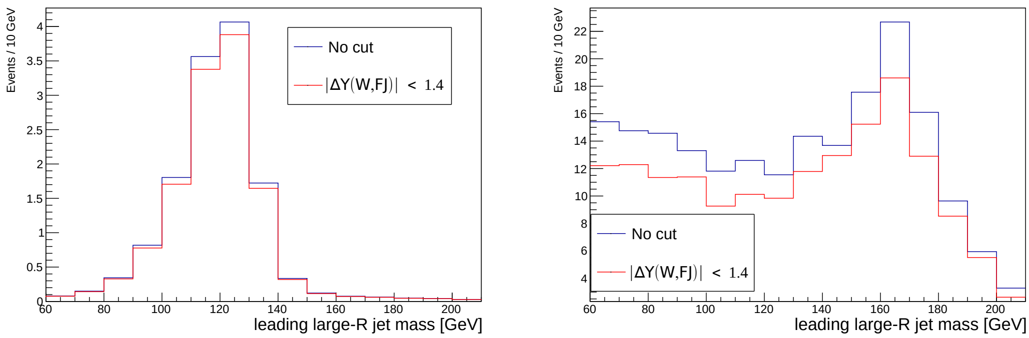
In order to verify that this selection criteria is not removing only events which can also be reconstructed using *resolved* techniques such as those used in the previous study, the Higgs candidate mass distribution for signal and backgrounds in the high purity signal region for $p_{T_W} > 400$ GeV , before and after the selection condition, is shown in Figure 6.15. In this figure, it can be seen that this condition also removes a significant fraction of background events in this p_{T_W} region. Due to this, it is be used in this analysis.

The final event selection for this analysis is summarised in Table 6.11.



(a) (b)

Figure 6.14: Leading large-R jet invariant mass distribution after the $|\Delta y(W,H)| < 1.4$ condition in the high purity signal region, for 250 GeV $< p_{T_W} < 400$ GeV (left) and $p_{T_W} > 400$ GeV (right)



(a) (b)

Figure 6.15: Leading large-R jet invariant mass distribution in the high purity signal region for $p_{T_W} > 400$ GeV , for signal (left) and backgrounds (right).

Trigger	Single lepton (e sub-channel)	$E_{\mathrm{T}}^{\mathrm{miss}}$ (μ sub-channel)
Leptons	Exactly 1 signal lepton	
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 30 GeV	
p_{T_W}	$p_{T_W} > 250$ GeV	
Large-R jet	≥ 1 large-R jet, $p_{\mathrm{T}} > 250$ GeV, $\lvert\eta\rvert < 2$	
Track-Jets	≥ 2 track-jets matched to the leading large-R jet (subjets), $p_{\mathrm{T}} > 10$ GeV, $\lvert\eta\rvert < 2.5$	
b-jets	leading two subjets must be b-tagged	
m_J	> 50 GeV	
$\Delta y(W,H)$	$\lvert\Delta y(W,H)\rvert < 1.4$	
p_{T_W} regions	250 GeV $< p_{T_W} < 400$ GeV and $p_{T_W} > 400$ GeV	
Signal regions	0 b-tagged track-jets not matched to the large-R jet	
	0 add. small-R jets (high purity)	≥ 1 add. small-R jets (low purity)
Control region	≥ 1 b-tagged track-jets not matched to the large-R jet	

Table 6.11: Event selection and categorization strategy. The objects reported in this table are defined as in Section 5.2.

6.5 Sensitivity to the SM boosted Higgs boson signal

To determine the sensitivity to the SM signal, the expected signal strength (μ_{exp}) and significance (Z_{exp}) for 139 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV are calculated using the fit described in Section 6.3, without taking into account systematic uncertainties. The expected signal strength is $\mu_{exp} = 1.00 \pm 0.41$ with an expected significance, $Z_{exp} = 2.52\sigma$, not enough to constitute evidence for the independent observation of boosted WH production. Taking into account all of the systematic uncertainties, one obtains $\mu_{exp} = 1.0 \pm 0.62$ and $Z_{exp} = 1.91$, with the uncertainty being dominated by the low statistics in data, followed by uncertainties on large-R jet quantities [88].

6.6 Study of the sensitivity to anomalous Spin/CP components

6.6.1 Signal strength-based studies

This section describes the method used to study the sensitivity of the analysis developed in the previous section to modifications in the signal strength induced by anomalous components, using the ATLAS SM signal and background samples described in Table 6.2 and the BSM signal samples described in Table 6.3.

The reconstruction of the BSM samples is done using `Delphes` and follows very closely the one for the ATLAS SM samples, except that fixed-radius track jets are used, since variable-radius reconstruction is not implemented. After detector simulation and event reconstruction, event selection and categorization criteria similar to those defined in Table 6.11 are applied to these samples, the only difference being that track jets are matched to the large-R jet by requiring ΔR (track jet, large-R jet) < 1.0, since ghost association is not available in Delphes. Additional ATLAS SM signal and background samples were generated with fixed-radius track jets, ΔR-based track jet to large-R jet association and fixed-radius μ-in-jet correction, $\Delta R\,(\mathrm{muon, track\,jet}) < 0.2$ and are used in the following.

In order to verify if the sensitivity to anomalous components is higher in the boosted regime, two values are calculated for SM and BSM signals (using `Delphes` samples): the signal acceptance, defined as the fraction of events that pass the event selection criteria, and the fraction of the events in the signal region with $p_{T_W} > 400$ GeV , $f_{\text{high}\,p_{T_W}}$. For the Delphes SM sample, the acceptance and $f_{\text{high}\,p_{T_W}}$ are approximately 1.24% and 26.8%, respectively. It should be noted that these are optimistic values for the signal acceptance when compared to what will be seen in the ATLAS analysis, due to the signal efficiency with variable-radius jets being lower than that for fixed-radius jets, as was seen in Section 6.4.1.

Table 6.12 shows the ratio of the cross-section, acceptance and $f_{\text{high}\,p_{T_W}}$ between BSM and SM samples. In it, one can see that the ratio is larger than one in samples with anomalous components, and that the ratios of the acceptance and $f_{\text{high}\,p_{T_W}}$ are higher for a sample with a CP-odd component than for a sample with a CP-even component, for the same value of the coupling. This supports the conclusion that analyses in the regime of high transverse momentum of the W boson are more sensitive to anomalous components.

Sample	σ/σ_{SM}	Acc/Acc_{SM}	$f_{\text{high } p_{T_W}}/f_{\text{high } p_{T_W}\ SM}$
$a_w = 0, b_{w1} = 0.05$	1.40	1.25	1.45
$a_w = 0, b_{w1} = 0.1$	1.82	1.67	1.54
$a_w = 0, c_w = 0.05$	1.02	1.29	1.61
$a_w = 0, c_w = 0.1$	1.06	1.99	1.69

Table 6.12: Ratio of the cross-section, acceptance (Acc) and $f_{\text{high } p_{T_W}}$ between a Delphes SM-like sample $(a_w = 0)$ and BSM samples with different anomalous components. b_{w1} is the coupling of the CP-even component and c_w is the coupling of the CP-odd component.

In order to determine the expected signal strength for BSM signals with anomalous components, the signal yields in the SM sample reconstructed using full ATLAS simulation (mentioned previously) are scaled by a factor $f = N_{BSM}/N_{SM}$, extracted from Delphes samples for each of the analysis regions. This assumes that anomalous components do not change the shape of the large-R jet invariant mass distribution, and that the ratio between the acceptances for BSM and SM signals are similar for $\sqrt{s} = 13$ TeV and $\sqrt{s} = 14$ TeV.

The values of the expected signal strength and significance for 139 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV for the different BSM signals, obtained with the procedure described in Section 6.3 and not taking into account systematic uncertainties, are given in Table 6.13, with the SM values also shown for comparison purposes. In the second and third lines of this table, it can be seen that the expected signal strengths for samples with an anomalous CP-even component are much larger than the one obtained for the SM sample, showing that this analysis is sensitive even to small values of the coupling. In the fifth line, one can also see that this analysis is sensitive to a signals with an anomalous CP-odd component with a large coupling, $c_w = 0.1$. This happens because, even though the change in the cross-sections induced by these components are very small, the modification in the acceptance is larger than that for CP-even ones, as can be seen in Table 6.12. Finally, in the third line, it can be seen that, for a CP-odd component with a small value of the coupling, $c_w = 0.05$, the expected signal strength is very similar to that obtained for the SM sample (albeit with a larger significance). This means that the signal strength measurements alone are not sensitive to CP-odd components with small couplings, and that additional observables are necessary.

Signal samples	μ_{exp}	$Z_{exp}(\sigma)$
SM	1.00 $\pm$ 0.41	2.52
SM + $b_{w1} = 0.05$	2.60 $\pm$ 0.44	6.49
SM + $b_{w1} = 0.1$	6.82 $\pm$ 0.51	15.90
SM + $c_w = 0.05$	1.54 $\pm$ 0.42	4.04
SM + $c_w = 0.1$	3.05 $\pm$ 0.45	7.78

Table 6.13: Expected signal strengths and significances for the SM signal and for BSM signals, in which the SM HWW vertex is supplemented with a set of (anomalous) components. b_{w1} is the coupling of the CP-even component and c_w is the coupling of the CP-odd component.

6.6.2 Angular observable studies

As seen in Table 6.13, anomalous CP-even and CP-odd components in the HWW vertex lead to modifications to the expected signal strength. However, using only the signal strength, it is not possible to disentangle the contributions from different anomalous components, for which other observables are necessary.

This section explores a set of angular observables, first proposed in Ref. [23], to disentangle the SM component from different BSM components as well as to distinguish between BSM components. It starts with a definition of the observables, followed by a truth-level study demonstrating the usefulness of these observables to distinguish between the SM signal and the BSM signals. Next, a study is carried out to determine if the conclusions of the truth-level analysis hold after taking into account full detector simulation, event reconstruction and selection, as well as the effect of backgrounds. It finalizes with a set of data-MC comparison plots in the top control region using ATLAS data and MC samples, in order to verify if these observables can be used in future ATLAS analyses.

Definition

According to Ref. [23], the following observables are defined:

$$\cos\theta^* = \frac{\mathbf{p}_\ell^{(W)} \cdot \mathbf{p}_W}{|\mathbf{p}_\ell^{(W)}||\mathbf{p}_W|} \quad (6.13) \qquad \Delta\phi^{lW} = \Delta\phi\left(\mathbf{p}_\ell^{(W)}, \mathbf{p}_W\right) \quad (6.14)$$

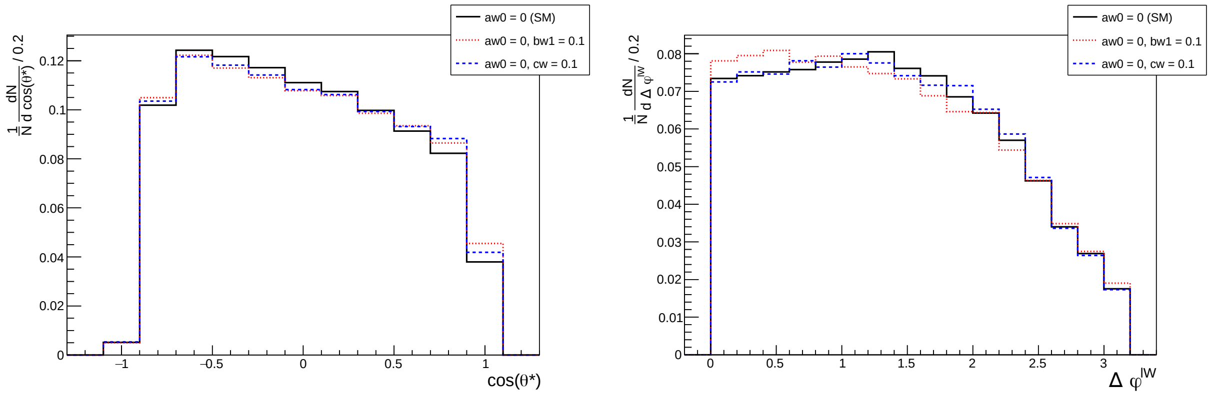
$$\cos\delta^+ = \frac{\mathbf{p}_\ell^{(W)} \cdot (\mathbf{p}_H \times \mathbf{p}_W)}{|\mathbf{p}_\ell^{(W)}||\mathbf{p}_H \times \mathbf{p}_W|} \quad (6.15) \qquad \cos\delta^- = \frac{\left(\mathbf{p}_\ell^{(H-)} \times \mathbf{p}_\nu^{(H-)}\right) \cdot \mathbf{p}_W}{|\mathbf{p}_\ell^{(H-)} \times \mathbf{p}_\nu^{(H-)}||\mathbf{p}_W|}, \quad (6.16)$$

where $\mathbf{p}_\ell^{(W)}$ is the 3-momentum of the electron or muon in the W boson rest frame and $\mathbf{p}_\ell^{(H-)}$ ($\mathbf{p}_\nu^{(H-)}$) is the 3-momentum of the electron or muon (neutrino) in the rest frame of a Higgs boson with $\mathbf{p} = -\mathbf{p}_H$. All other 3-momenta are defined in the lab frame.

Truth-level analysis

A truth-level analysis (before detector simulation) of these observables is done by taking the generator-level event record and extracting the kinematic information of the Higgs and W bosons just before decay, as well as the kinematic information of the lepton and neutrino from the W boson decay. These distributions are normalized to unit area, in order to remove any effects of normalization and compare only their shape.

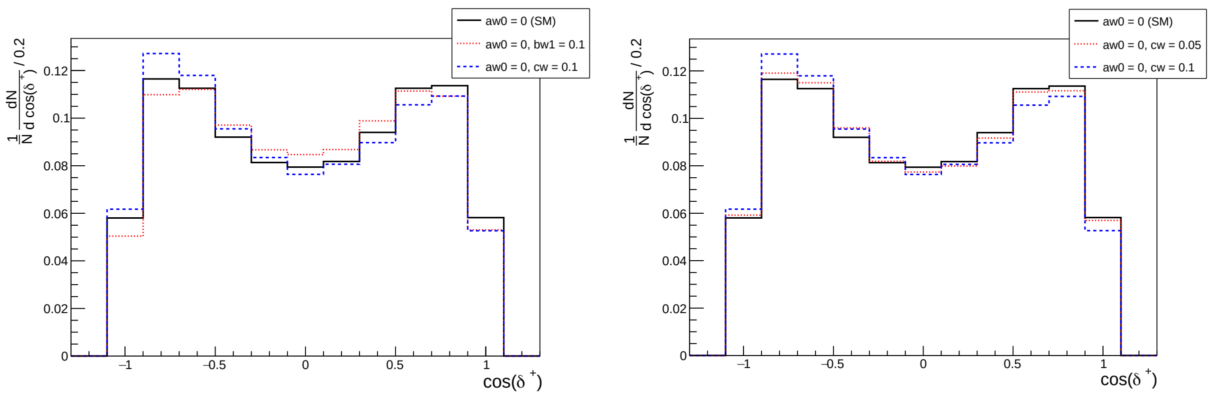
Figure 6.16 shows the truth-level $\cos\theta^*$and $\Delta\Phi^{lW}$distributions for the SM sample (with $a_w = 0$ and BSM samples with anomalous CP-even and CP-odd components, with couplings $b_{w1} = 0.1$ and $c_w = 0.1$, respectively. In Figure 6.16a, one can see that the $\cos\theta^*$distribution is shifted to higher values in samples with anomalous components, when compared to the SM sample, showing similar behaviour for samples with CP-even and CP-odd components, making it an observable suited to disentangle between SM and BSM components. In Figure 6.16b, one can see that the $\Delta\Phi^{lW}$distribution has a similar shape for the SM sample and the sample with an additional CP-odd component, showing differences with respect to the sample of with a CP-even component. Therefore, this variable can potentially be used to disentangle between the two.



(a) (b)

Figure 6.16: Shape comparison plots for the observables $\cos\theta^*$(left) and $\Delta\Phi^{lW}$(right) for the SM (solid black) sample, SM+CP-even component (dashed red) and SM+CP-odd component (dashed blue).

Figure 6.17a shows the truth-level distribution for the observable $\cos\delta^+$for the same samples as in Figures 6.16. One can see that the distribution of $\cos\delta^+$is symmetric for the SM sample and for a BSM sample with a CP-even component, but is asymmetric for a BSM sample the with a CP-odd component. From this, it can be inferred that the asymmetry is due to the existence of a CP-odd component. Figure 6.17b shows the truth-level distribution of the same observable for the SM sample, $a_w = 0$, and two BSM samples with CP-odd components with increasing values of the coupling, $c_w = 0.05$ and $c_w = 0.1$. In this figure, it can be seen that the asymmetry increases with the increase of the coupling, making $\cos\delta^+$an observable suited for extraction of its value.



(a) (b)

Figure 6.17: Shape comparison plots for the observable $\cos\delta^+$, comparing SM (solid black), SM+CP-even component (dashed red) and SM+CP-odd component (dashed blue) (left) and the SM with SM+CP-odd component with different couplings (dashed red and dashed blue - right)

Full event reconstruction analysis

The goal of this section is to see if, when taking into account the backgrounds, full event reconstruction and selection criteria, the conclusions of the truth-level analysis still hold, or if the angular observables lose some of their discriminating power.

A proper comparison between the different signals and backgrounds requires that `Delphes` and full ATLAS detector simulation and event reconstruction be as similar as possible. Since a lot of the intricacies of full ATLAS simulation are not replicable in `Delphes`, a reweighting of the `Delphes` samples is carried out, such that the acceptance and reconstruction effects from ATLAS samples are transferred to `Delphes`. For this, generator-level ATLAS SM signal samples are converted to a `Delphes`-readable format, processed, and the event selection and categorization criteria are applied. These are compared with equivalent samples processed with the full ATLAS simulation and reconstruction software (described in Section 6.6.1). The weight distributions are identified as the ratio between the full ATLAS simulation and `Delphes` distributions of a certain variable (after normalizing them to unit area). These can be calculated for each of the analysis regions or summing over them. This procedure expectedly ensures that the ATLAS and Delphes distributions for the chosen variable match perfectly (closure) and that the distributions of the angular observables become more similar.

After testing a set of variables such as the transverse momentum of the reconstructed W boson or the transverse momentum of the large-R jet, the choice is to do the reweighting based on the transverse momentum of the signal lepton in each of the analysis regions, since this combination gives the best post-reweighting agreement (closure) between the reweighted `Delphes` and full simulation distributions for the angular observables. In Figure 6.18, the distributions of the transverse momentum of the signal lepton before and after reweighting in the high purity signal region for 250 GeV $< p_{T_W} < 400$ GeV are shown. One can see that `Delphes` consistently underestimates the lepton transverse momentum given by full ATLAS simulation and that the reweighting procedure can correct this effect. Similar conclusions can be drawn for the $p_{T_W} > 400$ GeV region.

Figures 6.19 and 6.20 show the distributions of $\cos\theta^*$and $\cos\delta^+$in the high purity signal region for 250 GeV $< p_{T_W} < 400$ GeV , before and after reweighting. In these figures, one can see that the reweighting procedure improves the agreement between ATLAS and Delphes for these two variables and that, after reweighting, the agreement between the shapes of the angular observable distributions in ATLAS and Delphes simulation shows deviations lower than 20% (except for the first bin in the $\cos\theta^*$distribution, compatible with unity within 3 standard deviations). Similar conclusions can be drawn for the $p_{T_W} > 400$ GeV region, but with larger fluctuations around unity, most likely due to the small number of events in this region.

Following the derivation of the weights, these are applied to the BSM samples to correct for the acceptance differences, and a comparison of the distributions of the angular observables for different values of the couplings of the anomalous components $(b_{w1} = 0.1, c_w = 0.05, c_w = 0.1)$ is carried out, after applying the event selection and categorization criteria defined in Table 6.11, but using fixed-radius track jets and ΔR-based track jet to large-R jet association (described in the introduction of Section 6.6). This comparison is done for the high purity signal region and $p_{T_W} > 250$ GeV and is shown in

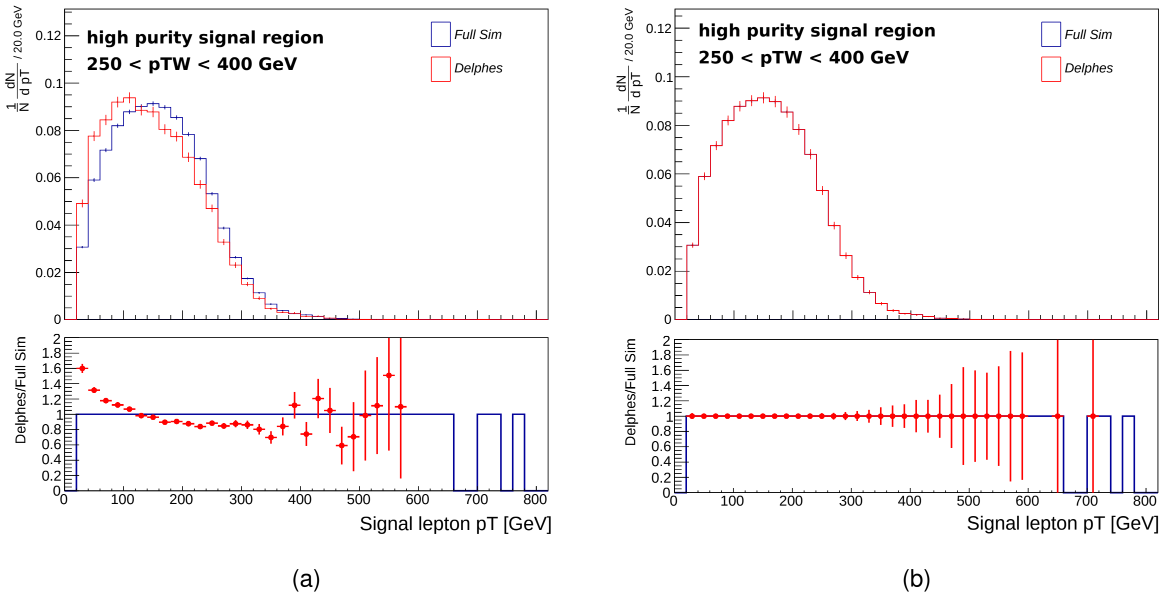


Figure 6.18: Shape comparison of the distributions of signal lepton transverse momentum in the high purity signal regions for 250 GeV $< p_{T_W} < 400$ GeV , before reweighting (left) and after reweighting (right). *Full Sim* refers to ATLAS detector simulation and event reconstruction. Uncertainties shown are of statistical nature.

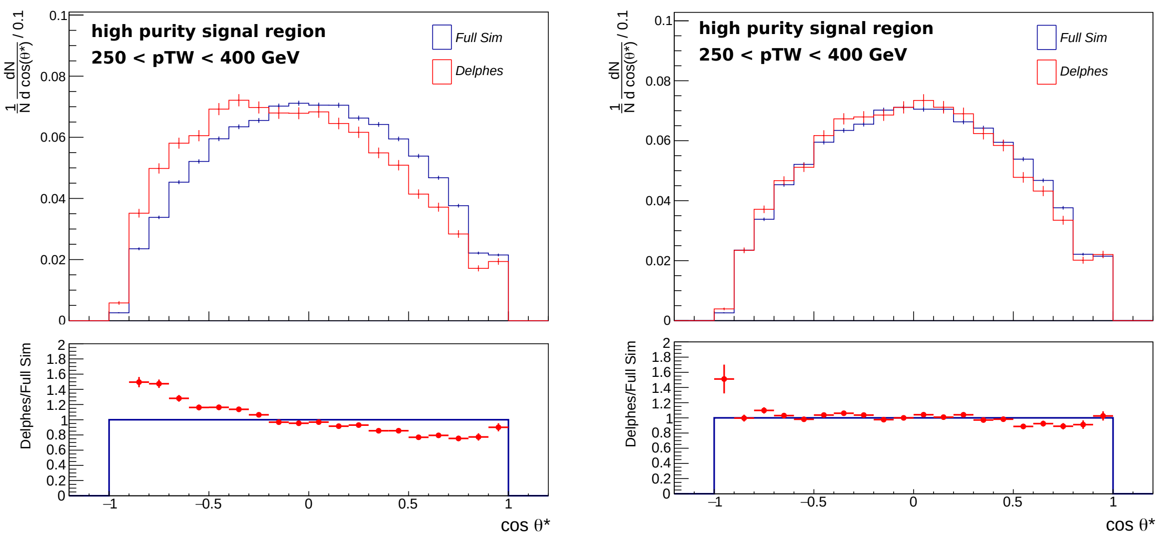


Figure 6.19: Shape comparison of the distributions of $\cos\theta^*$in the high purity signal regions for 250 GeV $< p_{T_W} < 400$ GeV , before reweighting (left) and after reweighting (right). *Full Sim* refers to ATLAS detector simulation and event reconstruction. Uncertainties shown are of statistical nature.

Figures 6.21 6.22 and 6.23. In each of the figures, two different comparisons are made: on the left side, a shape-only comparison of the distributions for the different signals and also the background, in which the distributions are normalized to unit area; on the right side, a comparison of the distributions for the different signal summed with the background, comparing also the background-only distribution.

In Figure 6.21a, it can be seen that for the observable $\cos\theta^*$, the SM signal distribution is symmetric around 0 and the BSM signal distributions are skewed to higher values, in agreement with the truth-level

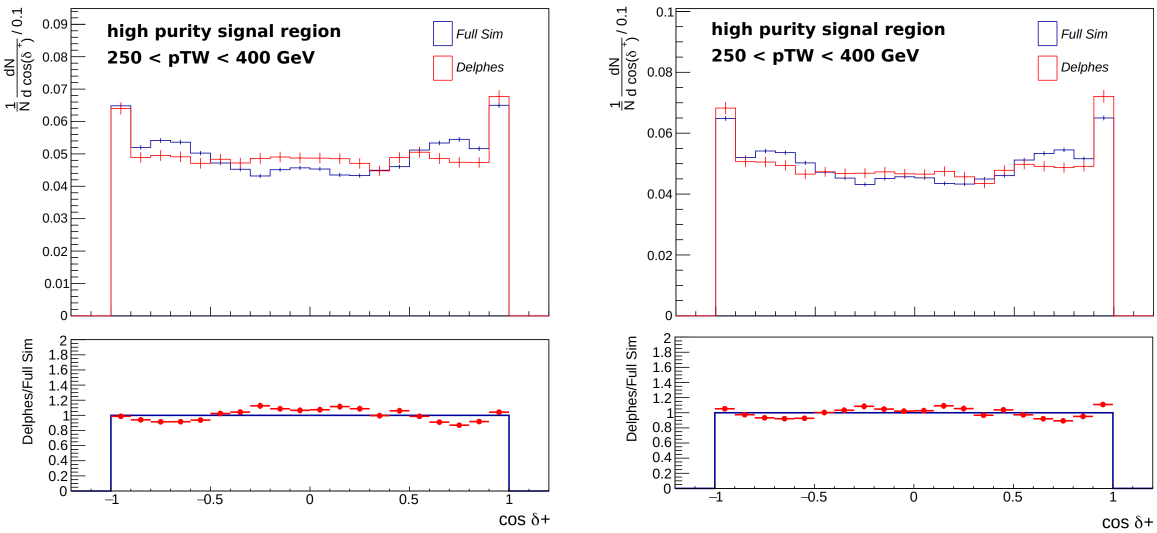
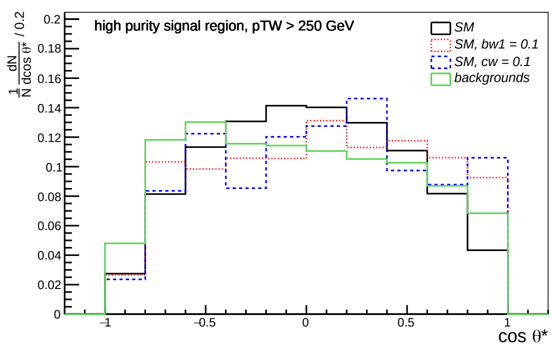
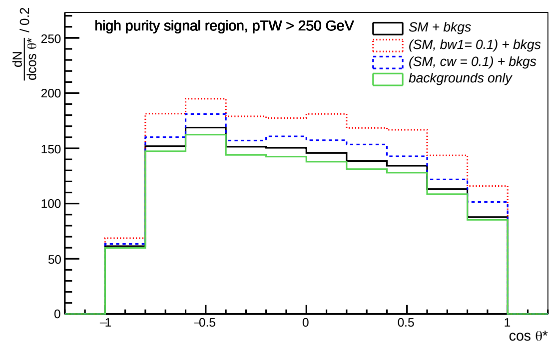


Figure 6.20: Shape comparison of the distributions of $\cos\delta^{+}$in the high purity signal regions for 250 GeV $< p_{T_W} < 400$ GeV , before reweighting (left) and after reweighting (right). *Full Sim* refers to ATLAS detector simulation and event reconstruction. Uncertainties shown are of statistical nature.

analysis. Furthermore, it can also be seen that the background distributions are skewed to lower values. This leads to the conclusion that, even after event selection and taking backgrounds into account, this variable is able to distinguish between SM signal and backgrounds and BSM signals.



(a)

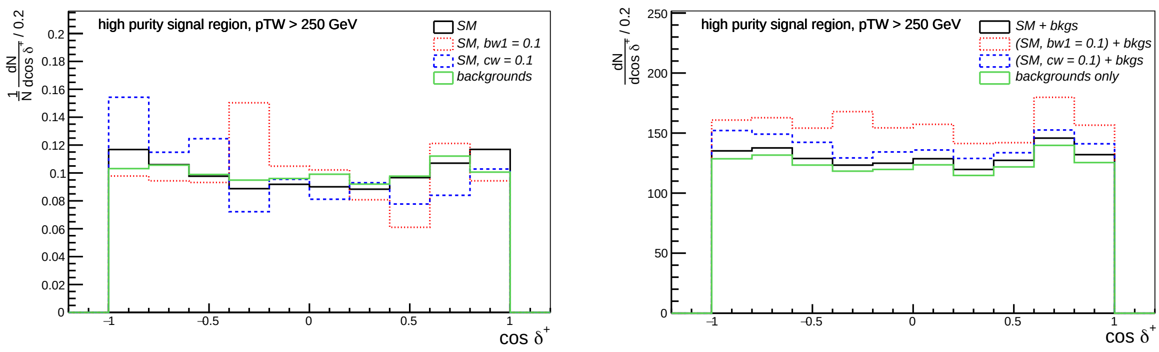


(b)

Figure 6.21: Comparison of $\cos\theta^{*}$distributions for the SM (solid black), SM+CP-even component (dashed red), SM+CP-odd (dashed blue) and backgrounds (green). Left: unit area normalization (shape-only). Right: cross-section normalization with backgrounds.

In Figure 6.22a, it can be seen that, for the observable $\cos\delta^{+}$, the distribution for the SM signal is symmetric around 0 and for the BSM sample with a CP-odd component, it is skewed to the left, in agreement with the truth-level analysis, the distribution for backgrounds also being symmetric around 0. For the BSM sample with a CP-even component, it is asymmetric, in disagreement with the truth-level analysis, but containing large fluctuations, which is most likely due to the small number of generated events for these samples and the fact that the acceptance is smaller for this sample than for the BSM sample with a CP-odd component (as can be seen in Table 6.12), thus leading to larger fluctuations.

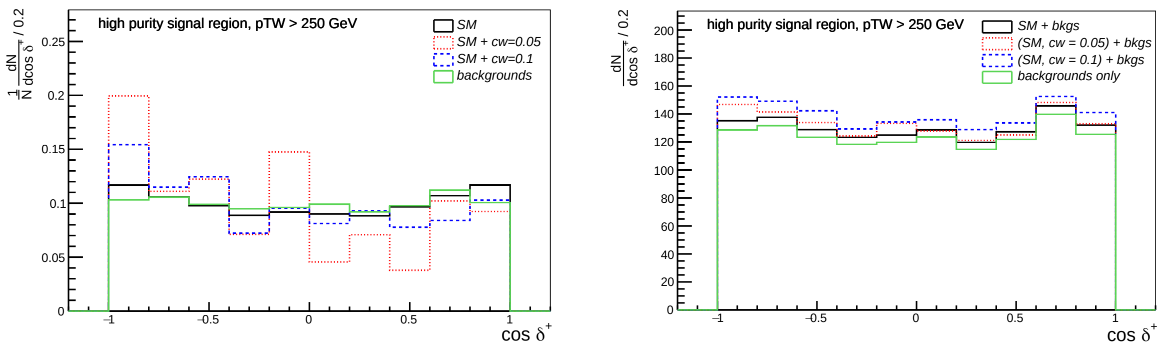
In Figure 6.23a, it can be seen that the distribution of $\cos\delta^{+}$for the BSM sample with a CP-odd



(a) (b)

Figure 6.22: Comparison of $\cos\delta^{+}$distributions for the SM (solid black), SM+CP-even component (dashed red), SM+CP-odd (dashed blue) and backgrounds (green). Left: unit area normalization (shape-only). Right: cross-section normalization with backgrounds.

component with a lower value of the coupling, $c_w = 0.05$, seems more asymmetric than that for the sample with the higher value of the coupling, $c_w = 0.1$, contradicting the conclusion of the truth-level analysis. This is believed to be due to large fluctuations, similar to what is seen in Figure 6.22a.



(a) (b)

Figure 6.23: Comparison of $\cos\delta^{+}$distributions for the SM (solid black), and SM+CP-odd component with increasing values of the coupling (dashed red and dashed blue) and backgrounds (green). Left: unit area normalization (shape-only). Right: cross-section normalization with backgrounds.

In Figure 6.21b, 6.22b and 6.23b, it can be seen that, even when taking into account the backgrounds, these variables still show visible differences between samples with anomalous components, which could be extracted with a global fit to their distributions.

The asymmetry in the observable $\cos\delta^{+}$, defined in Equation 6.17, shown to be sensitive to the existence of anomalous components, is calculated for the SM signal and background samples, as well as for the different BSM samples (assuming perfect background subtraction), after detector simulation, event reconstruction, selection and categorization, and is shown in Table 6.14. The calculated values of the asymmetry show that this observable is symmetric for the SM signal and backgrounds, while being asymmetric for samples with anomalous components, despite the large statistical uncertainties in these samples. This leads to the conclusion that this observable can potentially increase the sensitivity of this analysis to CP-odd components with small couplings with regards to the sensitivity obtained from signal strength measurements alone. Nonetheless, contrary to what is expected from the truth-level analysis, this observable is not symmetric in samples with anomalous CP-even components due to statistical fluctuations. Given that the equivalent generated luminosity for each of the BSM samples is larger than 139 fb^{-1}, similar or worse fluctuations are expected for the Run-2 integrated luminosity, leading to the conclusion that higher instantaneous luminosities are needed to clearly distinguish between CP-even and CP-odd anomalous components using these observables. To improve the sensitivity, a global fit that combines the information from several variables could be used.

$$A(\cos\delta^+) = \frac{N(\cos\delta^+ > 0) - N(\cos\delta^+ < 0)}{N(\cos\delta^+ > 0) + N(\cos\delta^+ < 0)} \tag{6.17}$$

Samples	Asymmetry
Backgrounds	0.003 $\pm$ 0.028
SM	-0.002 $\pm$ 0.133
SM + $b_{w1} = 0.05$	0.142 $\pm$ 0.087
SM + $b_{w1} = 0.1$	-0.081 $\pm$ 0.055
SM + $c_w = 0.05$	-0.319 $\pm$ 0.112
SM + $c_w = 0.1$	-0.123 $\pm$ 0.082

Table 6.14: Asymmetry of the $\cos\delta^+$distribution in the high purity signal region for $p_{T_W} > 250$GeV, for the ATLAS SM signal and backgrounds (described in Table 6.2) as well for different BSM signals (described in Table 6.3). Statistical uncertainties shown.

Background modelling of the angular distributions

A data-MC comparison of the distributions of the different angular observables in the top control region for each of the p_{T_W}regions is carried out, in order to verify if these are properly modelled in the ATLAS simulation. The comparison plots are shown in Figures 6.24 and 6.25 for the observables $\cos\theta^*$and $\cos\delta^+$, respectively. In these figures, it can be seen that the data-MC ratio floats around unity, with most of the differences likely covered by statistical fluctuations.

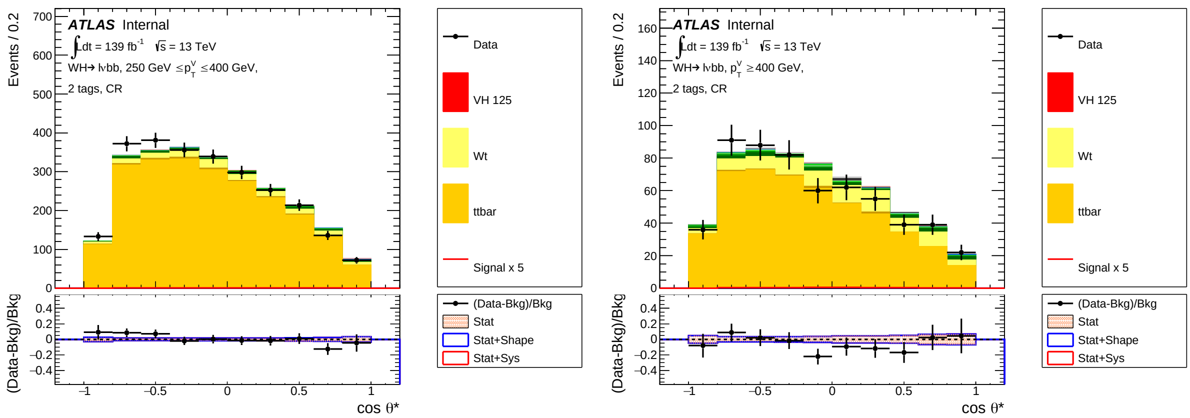


Figure 6.24: Data-MC comparison plots for the $\cos\theta^*$angular variable in the top control region for medium (left) and high (right) p_{T_W}regions

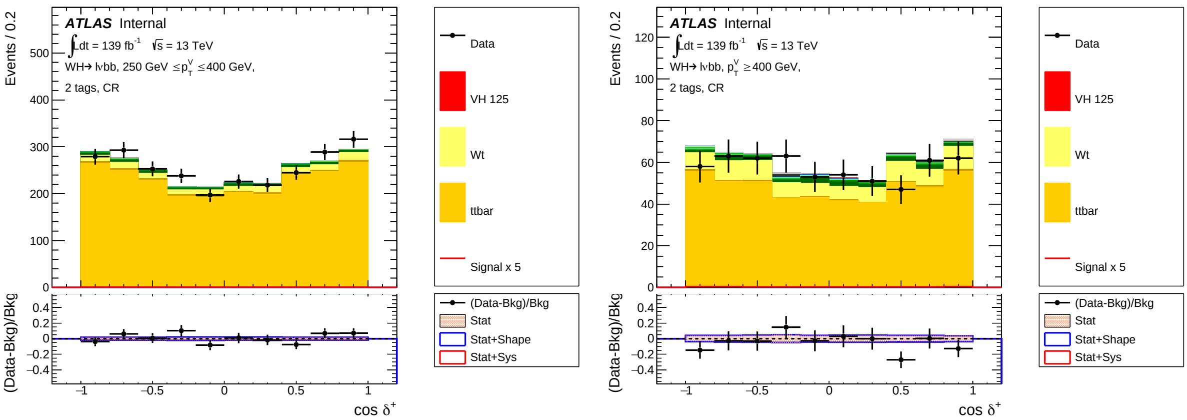


Figure 6.25: Data-MC comparison plots for the $\cos\delta^{+}$angular variable in the top control region for medium (left) and high (right) p_{T_W} regions

Chapter 7

Conclusions

This work explored the sensitivity of $WH \to \ell\nu b\bar{b}$ associated production channel in the boosted regime, $p_{T_W} > 250$ GeV, to anomalous Spin/CP components in the HWW vertex, with the full LHC Run-2 dataset, corresponding to a total integrated luminosity, $L = 139.9 \pm 2.4\ \mathrm{fb}^{-1}$ at a center-of-mass energy of $\sqrt{s} = 13$ TeV. It targets the decay of the W boson to an electron or muon and a neutrino and the decay of the Higgs boson to a pair of b-quarks. Given that, in this regime, these b-quarks are very collimated, the Higgs boson is identified using a trimmed anti-k_t, $R = 1.0$ jet with a substructure compatible with a $H \to b\bar{b}$ decay.

Since current experimental results indicate that the Higgs boson behaves very closely to its SM predictions, this analysis started by optimizing the identification strategy of the b-quarks from the SM $H \to b\bar{b}$ decay. The results demonstrated that the optimal strategy was to reconstruct variable-radius track jets, match them to the leading large-R jet using ghost-association, and require b-tagging on the two matched track jets with the highest transverse momentum, using MV2c10 in the 70% efficiency working point.

In order to improve signal and background separation, two discriminating variables were studied: the transverse mass of the reconstructed W boson, m_{T_W} , and the difference in rapidity between the reconstructed W boson and the Higgs boson candidate, $\Delta y(W,H)$. For the former, there was no increase in the expected significance from a global fit to all the analysis regions, for any of the tested values. For the latter, an upper cut of $|\Delta y(W,H) < 1.4|$ was implemented, leading to a 4.3% increase in the expected significance. The Higgs identification strategy and event selection studies presented in this thesis contributed to the search for boosted $H \to b\bar{b}$ decays in the WH associated production channel, to be published soon by the ATLAS collaboration [88].

The sensitivity to the SM $WH \to \ell\nu b\bar{b}$ signal was obtained. Taking only statistical uncertainties into account, the expected signal strength is $\mu_{exp} = 1.00 \pm 0.45$, corresponding to an expected significance $Z_{exp} = 2.52\sigma$, while taking into account the full set of experimental and modelling systematics, $\mu_{exp} = 1.0 \pm 0.62$ and $Z_{exp} = 1.91$, with the uncertainty being dominated by the low data statistics, followed by uncertainties on large-R jet quantities. The significance is not high enough to claim observation of boosted SM $WH(H \to b\bar{b})$ production, which was expected, given that the high energy regime targeted by this analysis contains less than 2% of the total SM production cross-section.

The sensitivity of the $WH \rightarrow \ell\nu b\bar{b}$ channel to anomalous components in the HWW vertex was studied using simulated samples in which the SM vertex is supplemented by CP-even or CP-odd anomalous components, with a set of benchmark values of its couplings, and taking only statistical uncertainties into account. The addition of a CP-odd anomalous component leads to cross-sections between 2% to 6% larger than that obtained in the SM, while the addition of a CP-even anomalous component leads to cross-sections between 40% and 82% larger. The signal acceptance was shown to be larger for signals with anomalous components than for the SM signal, and larger for signals with an anomalous CP-odd component than for samples with an anomalous CP-even component, for the same value of the coupling.

For signals with anomalous CP-even components, $\mu_{exp} = 2.60 \pm 0.44$ with an expected significance $Z_{exp} = 6.49$ and $\mu_{exp} = 6.82 \pm 0.51$, with an expected significance $Z_{exp} = 15.90$ (for increasing values of the coupling), which are not compatible the values expected for the SM signal, showing that this channel is sensitive to these components using only signal strength measurements. For a signal with an anomalous CP-odd component with a large coupling, the expected signal strength is $\mu_{exp} = 3.05 \pm 0.45$, corresponding to an expected significance of $Z_{exp} = 7.78$, which is also not compatible with the value expected for the SM signal, demonstrating that this channel is sensitive to anomalous CP-odd components with large values, using only measurements of the signal strength. For signal with an anomalous CP-odd component with small coupling, $\mu_{exp} = 1.54 \pm 0.42$, $Z_{exp} = 4.04$, which is compatible with the value obtained for the SM signal, leading to the conclusion that this channel is not sensitive to anomalous CP-odd components with small couplings using only signal strength measurements. These are optimistic results, since systematic uncertainties are not taken into account. Another conclusion is that the signal strength is not enough to distinguish between anomalous components with different CP properties, and that other observables are required.

A study of angular observables proposed in Ref. [23] to distinguish not only between the backgrounds, the SM signal and signals with anomalous components but also between different anomalous components was carried out. It showed that, when taking into account full detector simulation and event reconstruction, as well as the full of event selection and categorization criteria, the observable $\cos\theta^*$can distinguish between the SM signal and backgrounds and signals with anomalous components. It showed that, before detector simulation and event reconstruction, the observable $\cos\delta^+$is symmetric for the SM signal and backgrounds, as well as for BSM signals with anomalous CP-even components, and that it is asymmetric for signals with anomalous CP-odd components, with the asymmetry increasing with the value of the coupling. This study also showed that, after taking into account full detector simulation and event reconstruction, this observable is symmetric for the SM signal and backgrounds, but asymmetric for BSM samples with anomalous CP-even and CP-odd components, concluding that it has the potential to increase the sensitivity to anomalous CP-odd components with small couplings, but that higher luminosities than the ones currently accumulated for the LHC Run 2 are necessary to distinguish between anomalous components with different CP properties. The modelling of these variables was also studied in the control regions of the analysis, comparing MC and data, observing a very good agreement, and demonstrating that they can be used in a future ATLAS analysis.

The next steps in this study would be to include these observables in the next iteration of ATLAS analyses, using the full Run 2 luminosity with improved calibrations and simulations, and also explore methods to extract the value of the coupling using these variables, such as the calculation of asymmetry parameters, or multivariable methods to combine information from the signal strength and the different angular observables studied.

Bibliography

[1] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, Sep 2012. ISSN 0370-2693. doi: 10.1016/j.physletb.2012.08.020. URL http://dx.doi.org/10.1016/j.physletb.2012.08.020.

[2] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, Sep 2012. ISSN 0370-2693. doi: 10.1016/j.physletb.2012.08.021. URL http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[3] ATLAS Collaboration. Study of the spin and parity of the Higgs boson in Diboson decays with the ATLAS detector. *The European Physical Journal C*, 75(10), Oct 2015. ISSN 1434-6052. doi: 10.1140/epjc/s10052-015-3685-1. URL http://dx.doi.org/10.1140/epjc/s10052-015-3685-1.

[4] CMS Collaboration. Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. *Physical Review D*, 92(1), Jul 2015. ISSN 1550-2368. doi: 10.1103/physrevd.92.012004. URL http://dx.doi.org/10.1103/PhysRevD.92.012004.

[5] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. Advantages of the color octet gluon picture. *Physics Letters B*, 47(4):365 – 368, 1973. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(73)90625-4. URL http://www.sciencedirect.com/science/article/pii/0370269373906254.

[6] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. doi: 10.1103/PhysRevLett.19.1264. URL https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[7] A. Djouadi. The anatomy of electroweak symmetry breaking. Tome I: The Higgs boson in the Standard Model. *Physics Reports*, 457(1-4):1–216, 2008. ISSN 03701573. doi: 10.1016/j.physrep.2007.10.004.

[8] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. Evidence for the 2π Decay of the K_2^0 Meson. *Phys. Rev. Lett.*, 13:138–140, Jul 1964. doi: 10.1103/PhysRevLett.13.138. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[9] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964. doi: 10.1103/PhysRevLett.13.321. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964. doi: 10.1103/PhysRevLett.13.585. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[11] P. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132 – 133, 1964. ISSN 0031-9163. doi: https://doi.org/10.1016/0031-9163(64)91136-9. URL http://www.sciencedirect.com/science/article/pii/0031916364911369.

[12] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. doi: 10.1103/PhysRevLett.13.508. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[13] L. H. C. S. W. Group. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. Oct 2016. doi: 10.23731/CYRM-2017-002. URL https://cds.cern.ch/record/2227475.

[14] Super-Kamiokande Collaboration. Measurement of the Flux and Zenith-Angle Distribution of Upward Throughgoing Muons by Super-Kamiokande. *Physical Review Letters*, 82(13):2644–2648, Mar 1999. ISSN 1079-7114. doi: 10.1103/physrevlett.82.2644. URL http://dx.doi.org/10.1103/PhysRevLett.82.2644.

[15] SNO Collaboration. Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. *Physical Review Letters*, 89(1), Jun 2002. ISSN 1079-7114. doi: 10.1103/physrevlett.89.011301. URL http://dx.doi.org/10.1103/PhysRevLett.89.011301.

[16] KamLAND Collaboration. First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance. *Physical Review Letters*, 90(2), Jan 2003. ISSN 1079-7114. doi: 10.1103/physrevlett.90.021802. URL http://dx.doi.org/10.1103/PhysRevLett.90.021802.

[17] A. D. Sakharov. Violation of CP invariance,C asymmetry, and baryon asymmetry of the universe. *Soviet Physics Uspekhi*, 34(5):392–393, may 1991. doi: 10.1070/pu1991v034n05abeh002497. URL https://doi.org/10.1070%2Fpu1991v034n05abeh002497.

[18] P. Huet and E. Sather. Electroweak baryogenesis and standard model CP violation. *Physical Review D*, 51(2):379–394, Jan 1995. ISSN 0556-2821. doi: 10.1103/physrevd.51.379. URL http://dx.doi.org/10.1103/PhysRevD.51.379.

[19] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J. P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Physics Reports*, 516(1-2):1–102, Jul 2012. ISSN 0370-1573. doi: 10.1016/j.physrep.2012.02.002. URL http://dx.doi.org/10.1016/j.physrep.2012.02.002.

[20] I. Brivio and M. Trott. The standard model as an effective field theory. *Physics Reports*, 793:1–98, Feb 2019. ISSN 0370-1573. doi: 10.1016/j.physrep.2018.11.002. URL http://dx.doi.org/10.1016/j.physrep.2018.11.002.

[21] H. Bélusca-Maïto, A. Falkowski, D. Fontes, J. C. Romão, and J. P. Silva. Higgs EFT for 2HDM and beyond. *The European Physical Journal C*, 77(3), Mar 2017. ISSN 1434-6052. doi: 10.1140/epjc/s10052-017-4745-5. URL http://dx.doi.org/10.1140/epjc/s10052-017-4745-5.

[22] B. Grzadkowski, M. Iskrzyński, M. Misiak, and J. Rosiek. Dimension-six terms in the Standard Model Lagrangian. *Journal of High Energy Physics*, 2010(10), Oct 2010. ISSN 1029-8479. doi: 10.1007/jhep10(2010)085. URL http://dx.doi.org/10.1007/JHEP10(2010)085.

[23] R. M. Godbole, D. J. Miller, K. A. Mohan, and C. D. White. Jet substructure and probes of CP violation in Vh production. *Journal of High Energy Physics*, 2015(4), Apr 2015. ISSN 1029-8479. doi: 10.1007/jhep04(2015)103. URL http://dx.doi.org/10.1007/JHEP04(2015)103.

[24] ATLAS Collaboration. Measurements of gluon–gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s}$=13TeV with the ATLAS detector. *Physics Letters B*, 789:508–529, Feb 2019. ISSN 0370-2693. doi: 10.1016/j.physletb.2018.11.064. URL http://dx.doi.org/10.1016/j.physletb.2018.11.064.

[25] ATLAS Collaboration. Measurement of the production cross section for a Higgs boson in association with a vector boson in the $H \to WW^* \to \ell\nu\ell\nu$ channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physics Letters B*, 798:134949, Nov 2019. ISSN 0370-2693. doi: 10.1016/j.physletb.2019.134949. URL http://dx.doi.org/10.1016/j.physletb.2019.134949.

[26] CMS Collaboration. Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at s=13TeV. *Physics Letters B*, 791:96–129, Apr 2019. ISSN 0370-2693. doi: 10.1016/j.physletb.2018.12.073. URL http://dx.doi.org/10.1016/j.physletb.2018.12.073.

[27] ATLAS Collaboration. Observation of H$\to$ $b\bar{b}$ decays and VH production with the ATLAS detector. *Physics Letters B*, 786:59–86, Nov 2018. ISSN 0370-2693. doi: 10.1016/j.physletb.2018.09.013. URL http://dx.doi.org/10.1016/j.physletb.2018.09.013.

[28] CMS Collaboration. Observation of Higgs Boson Decay to Bottom Quarks. *Physical Review Letters*, 121(12), Sep 2018. ISSN 1079-7114. doi: 10.1103/physrevlett.121.121801. URL http://dx.doi.org/10.1103/PhysRevLett.121.121801.

[29] L. D. Landau. On the angular momentum of a system of two photons. *Dokl. Akad. Nauk SSSR*, 60 (2):207–209, 1948. doi: 10.1016/B978-0-08-010586-4.50070-5.

[30] C. N. Yang. Selection Rules for the Dematerialization of a Particle into Two Photons. *Phys. Rev.*, 77: 242–245, Jan 1950. doi: 10.1103/PhysRev.77.242. URL https://link.aps.org/doi/10.1103/PhysRev.77.242.

[31] CMS Collaboration. Constraints on anomalous HVV couplings from the production of Higgs bosons decaying to τ lepton pairs. *Physical Review D*, 100(11), Dec 2019. ISSN 2470-0029. doi: 10.1103/physrevd.100.112002. URL http://dx.doi.org/10.1103/PhysRevD.100.112002.

[32] K. Kondo. Dynamical Likelihood Method for Reconstruction of Events with Missing Momentum. I. Method and Toy Models. *Journal of the Physical Society of Japan*, 57(12):4126–4140, 1988. doi: 10.1143/JPSJ.57.4126. URL https://doi.org/10.1143/JPSJ.57.4126.

[33] K. Kondo. Dynamical Likelihood Method for Reconstruction of Events with Missing Momentum. II. Mass Spectra for 2→2 Processes. *Journal of the Physical Society of Japan*, 60(3):836–844, 1991. doi: 10.1143/JPSJ.60.836. URL https://doi.org/10.1143/JPSJ.60.836.

[34] ATLAS Collaboration. Measurement of VH, $H \to b\bar{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector. *Journal of High Energy Physics*, 2019(5), May 2019. ISSN 1029-8479. doi: 10.1007/jhep05(2019)141. URL http://dx.doi.org/10.1007/JHEP05(2019)141.

[35] J. Ellis, C. W. Murphy, V. Sanz, and T. You. Updated global SMEFT fit to Higgs, diboson and electroweak data. *Journal of High Energy Physics*, 2018(6), Jun 2018. ISSN 1029-8479. doi: 10.1007/jhep06(2018)146. URL http://dx.doi.org/10.1007/JHEP06(2018)146.

[36] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam. Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider. *Physical Review Letters*, 100(24), Jun 2008. ISSN 1079-7114. doi: 10.1103/physrevlett.100.242001. URL http://dx.doi.org/10.1103/PhysRevLett.100.242001.

[37] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014(2), Feb 2014. ISSN 1029-8479. doi: 10.1007/jhep02(2014)057. URL http://dx.doi.org/10.1007/JHEP02(2014)057.

[38] L. Evans and P. Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001–S08001, aug 2008. doi: 10.1088/1748-0221/3/08/s08001. URL https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001.

[39] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003–S08003, aug 2008. doi: 10.1088/1748-0221/3/08/s08003. URL https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003.

[40] ATLAS Collaboration. *ATLAS Magnet System: Technical Design Report*. ATLAS Technical Design Reports. CERN, Geneva, 1997. URL http://cds.cern.ch/record/338080.

[41] ATLAS Collaboration. *ATLAS Inner Detector: Technical Design Report*. ATLAS Technical Design Reports. CERN, Geneva, 1997. URL https://cds.cern.ch/record/331063.

[42] ATLAS Collaboration. *ATLAS Liquid-Argon Calorimeter: Technical Design Report*. ATLAS Technical Design Reports. CERN, Geneva, 1996. URL https://cds.cern.ch/record/331061.

[43] ATLAS Collaboration. *ATLAS Tile Calorimeter: Technical Design Report*. ATLAS Technical Design Reports. CERN, Geneva, 1996. URL https://cds.cern.ch/record/331062.

[44] ATLAS Collaboration. *ATLAS Muon Spectrometer: Technical Design Report*. ATLAS Technical Design Reports. CERN, Geneva, 1997. URL https://cds.cern.ch/record/331068.

[45] ATLAS Collaboration. The ATLAS Inner Detector commissioning and calibration. *The European Physical Journal C*, 70(3):787–821, Aug 2010.

[46] ATLAS Collaboration. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010. URL https://cds.cern.ch/record/1291633.

[47] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *The European Physical Journal C*, 77(7), Jul 2017. ISSN 1434-6052. doi: 10.1140/epjc/s10052-017-5004-5. URL http://dx.doi.org/10.1140/epjc/s10052-017-5004-5.

[48] E. Diehl. ATLAS Muon Detector Commissioning. In *Particles and fields. Proceedings, Meeting of the Division of the American Physical Society, DPF 2009, Detroit, USA, July 26-31, 2009*, Oct 2009. URL http://cds.cern.ch/record/1213055.

[49] A. Ruiz-Martinez. The Run-2 ATLAS Trigger System. Technical Report ATL-DAQ-PROC-2016-003, CERN, Geneva, Feb 2016. URL https://cds.cern.ch/record/2133909.

[50] J. C. Collins, D. E. Soper, and G. Sterman. Factorization of Hard Processes in QCD, 2004.

[51] F. Close. The quark parton model. *Reports on Progress in Physics*, 42(8):1285–1335, aug 1979. doi: 10.1088/0034-4885/42/8/001. URL https://doi.org/10.1088%2F0034-4885%2F42%2F8%2F001.

[52] P. Nason. A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms. *Journal of High Energy Physics*, 2004(11):040–040, Nov 2004. ISSN 1029-8479. doi: 10.1088/1126-6708/2004/11/040. URL http://dx.doi.org/10.1088/1126-6708/2004/11/040.

[53] S. Frixione, P. Nason, and C. Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *Journal of High Energy Physics*, 2007(11):070–070, Nov 2007. ISSN 1029-8479. doi: 10.1088/1126-6708/2007/11/070. URL http://dx.doi.org/10.1088/1126-6708/2007/11/070.

[54] S. Alioli, P. Nason, C. Oleari, and E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, 2010(6), Jun 2010. ISSN 1029-8479. doi: 10.1007/jhep06(2010)043. URL http://dx.doi.org/10.1007/JHEP06(2010)043.

[55] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), Jul 2014. ISSN 1029-8479. doi: 10.1007/jhep07(2014)079. URL http://dx.doi.org/10.1007/JHEP07(2014)079.

[56] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015. ISSN 0010-4655. doi: 10.1016/j.cpc.2015.01.024. URL http://dx.doi.org/10.1016/j.cpc.2015.01.024.

[57] M. Bähr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Plätzer, P. Richardson, M. H. Seymour, A. Sherstnev, and et al. Herwig++ physics and manual. *The European Physical Journal C*, 58(4):639–707, Nov 2008. ISSN 1434-6052. doi: 10.1140/epjc/s10052-008-0798-9. URL http://dx.doi.org/10.1140/epjc/s10052-008-0798-9.

[58] J. Bellm, S. Gieseke, D. Grellscheid, S. Plätzer, M. Rauch, C. Reuschle, P. Richardson, P. Schichtel, M. H. Seymour, A. Siódmok, and et al. Herwig 7.0/Herwig++ 3.0 release note. *The European Physical Journal C*, 76(4), Apr 2016. ISSN 1434-6052. doi: 10.1140/epjc/s10052-016-4018-8. URL http://dx.doi.org/10.1140/epjc/s10052-016-4018-8.

[59] D. J. Lange. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152 – 155, 2001. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(01)00089-4. URL http://www.sciencedirect.com/science/article/pii/S0168900201000894. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[60] Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(03)01368-8. URL http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[61] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *The European Physical Journal C*, 70(3):823–874, Sep 2010. ISSN 1434-6052. doi: 10.1140/epjc/s10052-010-1429-9. URL http://dx.doi.org/10.1140/epjc/s10052-010-1429-9.

[62] ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *Journal of Instrumentation*, 14(12):P12006–P12006, Dec 2019. ISSN 1748-0221. doi: 10.1088/1748-0221/14/12/p12006. URL http://dx.doi.org/10.1088/1748-0221/14/12/P12006.

[63] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV. *The European Physical Journal C*, 76(5), May 2016. ISSN 1434-6052. doi: 10.1140/epjc/s10052-016-4120-y. URL http://dx.doi.org/10.1140/epjc/s10052-016-4120-y.

[64] M. Cacciari and G. P. Salam. Pileup subtraction using jet areas. *Physics Letters B*, 659(1-2):119–126, Jan 2008. ISSN 0370-2693. doi: 10.1016/j.physletb.2007.09.077. URL http://dx.doi.org/10.1016/j.physletb.2007.09.077.

[65] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008. ISSN 1029-8479. doi: 10.1088/1126-6708/2008/04/063. URL http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[66] M. Cacciari, G. P. Salam, and G. Soyez. FastJet user manual. *The European Physical Journal C*, 72(3), Mar 2012. ISSN 1434-6052. doi: 10.1140/epjc/s10052-012-1896-2. URL http://dx.doi.org/10.1140/epjc/s10052-012-1896-2.

[67] M. Cacciari, G. P. Salam, and G. Soyez. The catchment area of jets. *Journal of High Energy Physics*, 2008(04):005–005, Apr 2008. ISSN 1029-8479. doi: 10.1088/1126-6708/2008/04/005. URL http://dx.doi.org/10.1088/1126-6708/2008/04/005.

[68] D. Krohn, J. Thaler, and L.-T. Wang. Jet trimming. *Journal of High Energy Physics*, 2010(2), Feb 2010. ISSN 1029-8479. doi: 10.1007/jhep02(2010)084. URL http://dx.doi.org/10.1007/JHEP02(2010)084.

[69] ATLAS Collaboration. Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Journal of High Energy Physics*, 2013(9), Sep 2013. ISSN 1029-8479. doi: 10.1007/jhep09(2013)076. URL http://dx.doi.org/10.1007/JHEP09(2013)076.

[70] ATLAS Collaboration. Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV. *The European Physical Journal C*, 79(10), Oct 2019. ISSN 1434-6052. doi: 10.1140/epjc/s10052-019-7335-x. URL http://dx.doi.org/10.1140/epjc/s10052-019-7335-x.

[71] D. Krohn, J. Thaler, and L.-T. Wang. Jets with variable R. *Journal of High Energy Physics*, 2009 (06):059–059, Jun 2009. ISSN 1029-8479. doi: 10.1088/1126-6708/2009/06/059. URL http://dx.doi.org/10.1088/1126-6708/2009/06/059.

[72] ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 78(11), Nov 2018. ISSN 1434-6052. doi: 10.1140/epjc/s10052-018-6288-9. URL http://dx.doi.org/10.1140/epjc/s10052-018-6288-9.

[73] R. Costa Batalha Pedro. Search for the Higgs boson at ATLAS/LHC in WH associated production and decay to b quark pairs, Apr 2017. URL https://cds.cern.ch/record/2304102. Presented 09 Nov 2017.

[74] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. Technical Report ATLAS-CONF-2019-021, CERN, Geneva, Jun 2019. URL http://cds.cern.ch/record/2677054.

[75] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano. HW$^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *Journal of High Energy Physics*, 2013

(10), Oct 2013. ISSN 1029-8479. doi: 10.1007/jhep10(2013)083. URL http://dx.doi.org/10.1007/JHEP10(2013)083.

[76] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, and et al. Parton distributions for the LHC run II. *Journal of High Energy Physics*, 2015(4), Apr 2015. ISSN 1029-8479. doi: 10.1007/jhep04(2015)040. URL http://dx.doi.org/10.1007/JHEP04(2015)040.

[77] A. Collaboration. Measurement of the transverse momentum distribution of Z/γ^* bosons in proton-proton collisions at $\sqrt{s} = 7$ TeVwith the ATLAS detector. *Physics Letters B*, 705(5):415–434, Nov 2011. ISSN 0370-2693. doi: 10.1016/j.physletb.2011.10.018. URL http://dx.doi.org/10.1016/j.physletb.2011.10.018.

[78] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and et al. Parton distributions with LHC data. *Nuclear Physics B*, 867(2):244–289, Feb 2013. ISSN 0550-3213. doi: 10.1016/j.nuclphysb.2012.10.003. URL http://dx.doi.org/10.1016/j.nuclphysb.2012.10.003.

[79] A. Collaboration. ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014. URL https://cds.cern.ch/record/1966419.

[80] E. Bothmann, G. Singh Chahal, S. Höche, J. Krause, F. Krauss, S. Kuttimalai, S. Liebschner, D. Napoletano, M. Schönherr, H. Schulz, and et al. Event generation with Sherpa 2.2. *SciPost Physics*, 7(3), Sep 2019. ISSN 2542-4653. doi: 10.21468/scipostphys.7.3.034. URL http://dx.doi.org/10.21468/SciPostPhys.7.3.034.

[81] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung. New Generation of Parton Distributions with Uncertainties from Global QCD Analysis. *Journal of High Energy Physics*, 2002(07):012–012, Jul 2002. ISSN 1029-8479. doi: 10.1088/1126-6708/2002/07/012. URL http://dx.doi.org/10.1088/1126-6708/2002/07/012.

[82] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks. FeynRules 2.0 — A complete toolbox for tree-level phenomenology. *Computer Physics Communications*, 185(8):2250–2300, Aug 2014. ISSN 0010-4655. doi: 10.1016/j.cpc.2014.04.012. URL http://dx.doi.org/10.1016/j.cpc.2014.04.012.

[83] W. Verkerke and D. Kirkby. The RooFit toolkit for data modeling, 2003.

[84] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf. The RooStats Project, 2010.

[85] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. (CERN-OPEN-2012-016), Jan 2012. URL https://cds.cern.ch/record/1456844.

[86] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), Feb 2011. ISSN 1434-6052. doi: 10.1140/epjc/s10052-011-1554-0. URL http://dx.doi.org/10.1140/epjc/s10052-011-1554-0.

[87] ATLAS Collaboration. Variable Radius, Exclusive-k$_T$, and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS. Technical Report ATL-PHYS-PUB-2017-010, CERN, Geneva, Jun 2017. URL http://cds.cern.ch/record/2268678.

[88] ATLAS Collaboration. Measurement of the associated production of a Higgs boson decaying to b quarks with a vector boson at high transverse momentum in pp collisions at 139 fb^{-1} with the ATLAS detector (in preparation). Technical Report ATL-COM-PHYS-2019-1125, CERN, Geneva, Sep 2019.