Charming Higgs bosons

Constraint on the Higgs-charm coupling from a search for Higgs boson decays to charm quarks with the ATLAS detector

Marko Stamenkovic

ISBN: 978-94-6421-656-1

Charming Higgs bosons - Constraint on the Higgs-charm coupling from a search for Higgs boson decays to charm quarks with the ATLAS detector

Thesis, Universiteit van Amsterdam (UvA), Amsterdam

Cover design by Samuel Pasquier

Printed in the Netherlands by Ipskamp

This work has been performed at the National Institute for Subatomic Physics (Nikhef) which is financed by the Netherlands Organisation for Scientific Research (NWO).

Charming Higgs bosons

Constraint on the Higgs-charm coupling from a search for Higgs boson decays to charm quarks with the ATLAS detector

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 11 maart 2022, te 14.00 uur

door

Marko Stamenkovic

geboren te Bossonnens

Promotiecommissie:

Promotor:	prof. dr. W. Verkerke	Universiteit van Amsterdam
Copromotor:	dr. T.A. du Pree	Nikhef
Overige leden:	prof. dr. M.H.M. Merk	Universiteit Maastricht
	dr. M. Delmastro	Laboratoire d'Annecy de Physique des Particules
	prof. dr. E.L.M.P. Laenen	Universiteit van Amsterdam
	dr. F. Filthaut	Radboud Universiteit Nijmegen
	prof. dr. F.L. Linde	Universiteit van Amsterdam
	dr. J.M. Sonneveld	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

For Alice,

May you always witness the beauty of this universe...

Contents

1 Introduction

Particles physics aims at describing and measuring the properties of fundamental particles. The Standard Model (SM) provides a description of these elementary particles and their interaction. Elementary particles are classified in two types: the fermions, which compose all the visible matter in the universe, and the bosons, which mediate the interactions between elementary particles. In the SM, the fermions are further classified in three different generations. The first generation of particles is composed by the electron, the electron neutrino, the up and the down quarks. In addition, there are two heavier versions of each particle, which have the same properties except for their mass, as illustrated in Fig. 1.1 for the charged fermions. The large differences observed in the masses of elementary particles as well as the existence of exactly three generations of particles remain unexplained to this day. The main interaction that distinguishes between the three generations of particles is the interaction of elementary particles with the Higgs boson, which is proportional in strength to their masses.

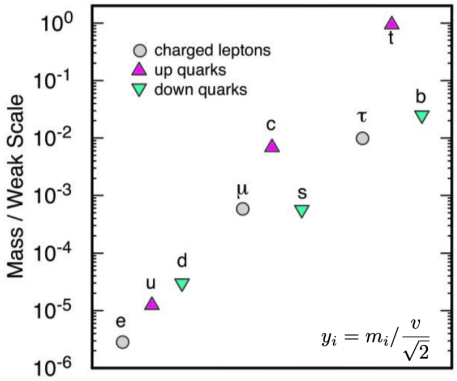


Figure 1.1: Masses of charged fermions ordered according to the three different generations [1].

The precise measurement of the interactions of the Higgs boson with elementary particles is intrinsically a probe for new physics, which could unravel a more complex structure in our understanding of nature. In particular, the existence of new massive particles interacting with the Higgs boson could significantly modify its properties, which would in turn affect the interaction of the Higgs boson with particles of the Standard Model. The discovery of physics beyond the Standard

Model and the precise measurement of the properties of the Higgs boson are among the major goals of the LHC programs at CERN.

The first collisions in the Large Hadron Collider (LHC) at the European Organisation for Nuclear Research (CERN) were recorded on the 6th of December 2009 at a center of mass energy of $\sqrt{s} = 900$ GeV. Since then, the LHC delivered an increasing number of collisions which were recorded by the ALICE, ATLAS, CMS and LHCb physics experiments. In particular, the ATLAS experiment measured a record-breaking data set of proton-proton collisions, summarised in Table 1.1, which resulted in more than 1000 publications. Among these publications, the discovery of the Higgs boson [2] was announced in 2012.

Periods	2011	2012	2015-2018
$\sqrt{s}$	7 TeV	8 TeV	13 TeV
Luminosity	5.1 fb^{-1}	21.3 fb^{-1}	139.0 fb^{-1}

Table 1.1: Integrated luminosity recorded by the ATLAS experiment for the different periods of data taking and corresponding center of mass energy.

So far, the ATLAS and CMS experiments successfully measured the interaction of the Higgs boson to gauge bosons as well as to charged particles classified as the third generation of fermions. In addition, the collaborations recently announced evidences of the interaction with muons, which are part of the second generation. All these measurements were found compatible with the theory prediction. Nevertheless, the interaction of the Higgs boson with the second generation of quarks remains unobserved yet. In particular, there is no direct constraint on the Higgs boson coupling to charm quarks.

Charm quarks are the heaviest particles of the second generation of fermions and the predicted decay rate of the Higgs boson to a pair of charm quark-antiquarks is 2.89%. Attempts to directly measure this interaction were performed by the LHCb [3], ATLAS [4] and CMS [5] collaborations and the measurements are summarised in Table 1.2. The constraint prior to the measurement described in this thesis on the signal strength of the $H \to c\bar{c}$ decay is an observed upper limit on the signal strength of 70 times the prediction of the SM at 95% confidence level.

Experiment	Observed (expected) limit at 95% CL	$\mathcal{L}$ [fb^{-1}]	$\sqrt{s}$ [TeV]
LHCb	7900 (6400) $\times$ SM	2	8
ATLAS	110 (150) $\times$ SM	36	13
CMS	70 (37) $\times$ SM	36	13

Table 1.2: Observed and expected upper limits on the signal strength of the $H \to c\bar{c}$ decay obtained by the LHCb, ATLAS and CMS experiments.

The work presented in this thesis aims at performing a measurement of the

$H \to c\bar{c}$ process using the data collected by the ATLAS experiment with an integrated luminosity of $\mathcal{L} = 139$ fb^{-1} at a center of mass energy of $\sqrt{s} = 13$ TeV. The highlights are the usage of the most recent charm tagging algorithms, designed to identify signatures of c-quarks in the detector, as well as a new reconstruction strategy of the Higgs boson, which allows to simultaneously measure the $H \to b\bar{b}$ and $H \to c\bar{c}$ processes for the first time. The methodology of the analysis is validated by the first measurement of the W and Z boson decays to at least one c-quark, using charm tagging algorithms. In addition, the measurement of the $H \to c\bar{c}$ process results in the most stringent constraint on this rare process, which allows to set a direct constraint on modifications of the Higgs boson coupling to charm quarks for the first time. Finally, the universality of the Higgs boson coupling to b- and c-quarks is probed.

Thesis structure and own contributions

Chapter 2 introduces the theory of the Standard Model, including specific theory models that predict enhanced couplings of the Higgs boson with charm quarks. Chapter 3 describes the ATLAS detector as well as its performance during the second run of the Large Hadron Collider.

Chapter 4 describes the reconstruction of physics object based on the measurements obtained from the ATLAS detector. I have contributed to the algorithms used to identify the flavour of the heavy hadrons within jets (flavour tagging). In particular, I have defined the c-tagging working point and I have worked on the c-tagging calibrations, including the post-processing of the efficiency scale factors and the implementation of the software required to use these calibrations in various analyses. In addition, I have implemented the truth tagging method in the context of c-tagging. As a member of the ATLAS flavour tagging group, I have implemented, validated and delivered all the calibrations files used for b-tagging in the ATLAS collaboration since 2017.

Chapter 5 focuses on the search for the $H \to c\bar{c}$ process with the ATLAS detector. As one of the main analyser, I have significantly contributed to this measurement. I have implemented and validated the software of the $H \to c\bar{c}$ analysis within the framework used by the $H \to b\bar{b}$ analysis performed by the ATLAS collaboration. I have proposed the flavour tagging paradigm using c-tagging together with a b-tag veto. I have optimised the flavour tagging working point and I have designed the analysis strategy, which results in the analysis of the $H \to c\bar{c}$ measurement being statistically independent of the $H \to b\bar{b}$ measurement. In addition, I have optimised the event selection of the 0-, 1- and 2-lepton channels. I have also implemented the truth tagging method. I have implemented and studied the kinematic fit used in the 2-lepton channel to perform jet energy corrections. I have implemented the first version of the fit model with systematic uncertainties and I have implemented the software used to derive the theory uncertainties, for both the normalisation and the shape uncertainties. I have studied the fit model, particularly for the 2-lepton

channel and also for the combination of the 0-, 1- and 2-lepton channels. Finally, I have studied the parametrisation used for the interpretation of the measurement in terms of coupling modifiers of the Higgs boson interaction with charm quarks and I have contributed to the combination of the $H \to b\bar{b}$ and $H \to c\bar{c}$ processes.

Chapter 6 presents a comparison of the $H \to c\bar{c}$ result with the previous measurements performed by the LHCb, ATLAS and CMS collaborations. I have performed the comparison with the previous result of the ATLAS collaboration.

Chapter 7 describes future prospects for the $H \to c\bar{c}$ measurement. I have contributed to the extrapolation to the High-Luminosity LHC, in particular the extrapolation of the combined measurement of the $H \to b\bar{b}$ and $H \to c\bar{c}$ processes.

The final chapter concludes the thesis with a discussion of the results.

2 Theory introduction

The aim of particle physics is to describe the fundamental constituents of the universe, the elementary particles, and the interactions between them, the forces. The most successful description of all experimental data collected so far is embodied in the Standard Model (SM) of particle physics, which provides a unified picture in which the forces between particles are themselves described by particles. By collecting and analysing data, one of the main objectives of experimental particle physicists is to test this theory with ever increasing precision, searching for phenomena in the data that are not yet predicted and that could shed a new light on our current understanding.

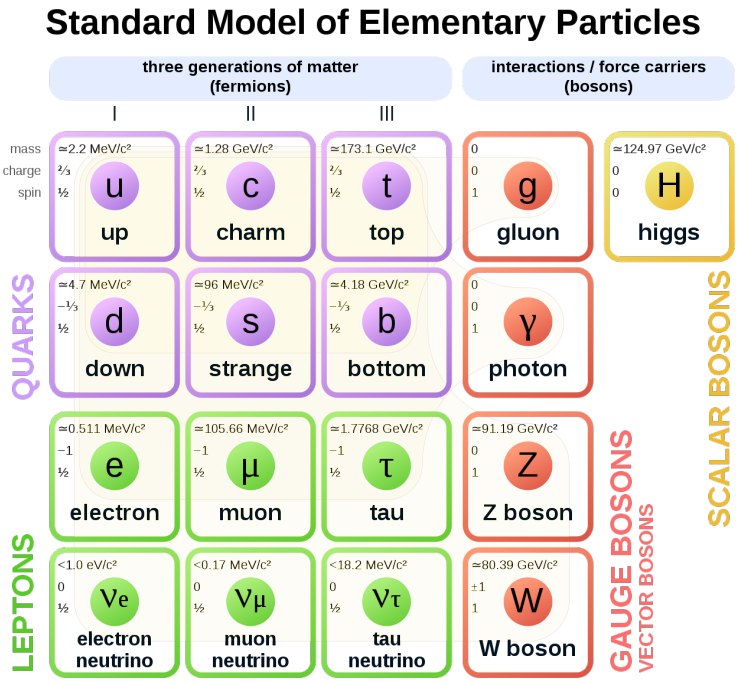


Figure 2.1: Standard model of elementary particles: the 12 fundamental fermions and 5 fundamental bosons.

Elementary particles are classified in two types: fermions and bosons. In total, there are 12 elementary fermions and their interactions are mediated by bosons. There are 3 fundamental interactions in the SM: the electromagnetic, strong and weak interaction, which are respectively mediated by the photon, the gluon and the W and Z bosons. On a macroscopic level, gravity is described by general relativity, however, in order to be included in the SM, a quantum theory of gravity that describes the gravitational interaction between particles is yet to be found. Finally, there is one additional boson, the Higgs boson, which interacts with elementary particles proportionally to their mass. All these particles, shown in Fig. 2.1, are observed experimentally and the main aim of the SM is to provide a mathematical description of these particles as well as their interactions.

2.1 The Standard Model

In the SM, the electron, the electron neutrino, the up-quark and the down-quark, shown for example in the left most columns of Fig. 2.1, are classified as the first generation of particles. These particles are hypothesised to be elementary and represent the basic building blocks of the visible matter in the universe. Nevertheless, when particle interactions are studied at the high energies of particle colliders, a new level of complexity is observed. For each elementary particle of the first generation, there exist two copies which differ in their masses. For example, the charm quark of the second generation of quarks is the heavier version of the up quark with a mass of $m_c \approx 400\ m_u$ and the top quark of the third generation is an even heavier copy with $m_t \approx 58000\ m_u$. Apart from the differences in the masses, which have physical consequences, the up, charm and top quarks are subject to the same interactions.

Particles of the the SM are commonly described by a Lagrangian density, where symmetries play a fundamental role. Elementary particles are represented by excitations on associated field $\psi(x)$ in 4-dimensional space-time and their motion are described by the Lagrangian density:

$$\mathcal{L}(x) = \mathcal{L}(\psi, \partial_\mu \psi). \tag{2.1}$$

Symmetries are a fundamental concept in the SM and are linked to conservation laws by the theorem of Noether. For example, the invariance under an external symmetry such as time and space translations results in the conservation of physical properties as the energy and the momentum, respectively. Relativistic fermions that obey the conservation of energy and momentum are described by the Dirac Lagrangian density:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \tag{2.2}$$

where γ^μ are the Dirac gamma matrices. The equations of motion are obtained by

solving the Euler-Lagrange equations, resulting in the Dirac equation:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0. \tag{2.3}$$

One important consequence is that for each of the twelve fermions, there exists an antiparticle with the exact same mass but the opposite charge.

In this representation of the Lagrangian, the motion of each elementary fermion is successfully described in the absence of any interaction. Interactions are incorporated using two types of continuous internal symmetries: global and local symmetries. Global symmetries are transformations applied identically on all points of space-time, while local symmetries are transformations that can differ from point to point in space-time. The invariance of the Lagrangian density under a global symmetry yields a conserved current, for example the conservation of the electric charge in particle interactions. The invariance of the Lagrangian density under a local symmetry, also known as gauge symmetry, results in the existence of particles of spin 1 known as gauge bosons, that mediate the different interactions through the exchange of a conserved quantity.

2.1.1 Quantum Electrodynamics - QED

Electromagnetic interactions are described by Quantum Electrodynamics (QED), which is defined by requiring the Lagrangian density to be invariant under phase translations of the fermionic fields, also known as the $U(1)$ symmetry group. The Lagrangian is described by:

$$\mathcal{L}_{\text{QED}} = \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{EM}} = \sum_f \bar{\psi}_f (i\gamma^\mu D_\mu - m_f)\psi_f - F_{\mu\nu}F^{\mu\nu}, \tag{2.4}$$

where f corresponds to electrically charged fermions, $D_\mu = \partial_\mu - ieA_\mu$ is the covariant derivative, A_μ denotes a gauge boson that mediates the electromagnetic interaction, the photon, and the Lagrangian density of this field is described using the electromagnetic tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The conserved quantity associated with the local gauge transformations of the $U(1)_{\text{QED}}$ symmetry group is the electric charge, which explains that the photon only directly interacts with electrically charged particles. The addition of a mass term to the photon $m^2 A^\mu A_\mu$ Lagrangian density is only invariant under phase shifts for $m = 0$, thus explaining that photons are massless.

2.1.2 Quantum Chromondynamics - QCD

Strong interactions are described by Quantumchromodynamics (QCD), which is invariant under the symmetries of the local gauge group $SU(3)_c$. The conserved quantum number is the color charge, which can be either red, blue or green. There are many experimental evidences that quark exist, however, free quarks have

never been observed. This is explained by the phenomena of colour confinement, which states that coloured objects are always confined and that no object with non-zero colour charge can propagate as free particles. As a consequence, quarks are observed as jets of colourless composite particles, i.e mesons and baryons.

The interactions between quarks are obtained by requiring the Lagrangian density to be invariant under phase translations of the $SU(3)_c$ symmetry group, and the resulting Lagrangian can be expressed as:

$$\mathcal{L}_{\text{QCD}} = \sum_q \bar{\psi}_q (i\gamma^\mu D_\mu - m_q)\psi_q - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a, \tag{2.5}$$

where q correspond to the sum over the quarks. The invariance under $SU(3)_c$ is guaranteed by using the covariant derivative:

$$D_\mu = \partial_\mu + i g_s \frac{\lambda^a}{2} G^a_\mu, \tag{2.6}$$

where g_s is the coupling constant of the strong interaction, λ^a are the Gell-Mann matrices and G^a_μ denotes 8 fields that mediate the strong interaction, the gluons, each corresponding to one generator of the $SU(3)$ symmetry. The Lagrangian density of these fields is described using the strong interaction tensor:

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu, \tag{2.7}$$

where f^{abc} are the structure constants of the $SU(3)$ group. For the same reason as photons, gluons are massless particles. Due to the fact that gluons always carry a color charge, they also confine to colourless objects and therefore, unlike photons, gluons do not propagate over macroscopic distances.

2.1.3 Electroweak interaction

QED and QCD share a number of common features: both interactions are mediated by massless neutral spin-1 bosons and the interaction vertices have the same form. The weak interaction, which is responsible for β decays and nuclear fusion, differs in almost all respects: the interaction is mediated by massive bosons, of which the $W^\pm$ boson carries an electric charge that allows for charge transfer in particle interactions.

The weak interaction is invariant under the symmetries of the local gauge group $SU(2)_L \times U(1)_Y$. In order to define the Lagrangian density invariant under $SU(2)_L$ symmetries, the fermionic fields are split according to their left- and right-handed components using the projection operators P_L and P_R:

$$\psi = P_L\psi + P_R\psi = \frac{1-\gamma^5}{2}\psi + \frac{1+\gamma^5}{2}\psi = \psi_L + \psi_R. \tag{2.8}$$

The charge associated with the $SU(2)_L$ symmetry group is the weak isospin I_3. Left- and right-handed fermions have a different weak isospin, respectively $I_3 = \pm\frac{1}{2}$ and $I_3 = 0$. The quantity associated with the $U(1)$ symmetry group is the weak hypercharge Y_W, which is related to the electric charge Q of the $U(1)_{\text{QED}}$ group as:

$$Y_W = 2(Q - I_3). \tag{2.9}$$

	Particles	Type	Q	I_3	Y_W
Q_L	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}$	Left-handed quarks	$\frac{2}{3}$ or -$\frac{1}{3}$	$\frac{1}{2}$ or -$\frac{1}{2}$	$\frac{1}{3}$
U_R	u_R,c_R,t_R	Right-handed up-type quarks	$\frac{2}{3}$	0	$\frac{4}{3}$
D_R	d_R,s_R,b_R	Right-handed down-type quarks	-$\frac{1}{3}$	0	-$\frac{2}{3}$
L_L	$\begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}$	Left-handed leptons	0 or -1	$\frac{1}{2}$ or -$\frac{1}{2}$	-1
L_R	e_R,μ_R,τ_R	Right-handed leptons	-1	0	-2

Table 2.1: Quantum numbers of the left-handed (doublets) and right-handed (singlets) fermions for the weak interaction.

The weak interaction that affects the fermions, listed in Table 2.1 with their respective quantum numbers, is described by the Lagrangian density:

$$\mathcal{L}_{\text{EW}} = \sum_f \bar{\psi}_{L,f} i\gamma^\mu D_\mu \psi_{L,f} + \sum_f \bar{\psi}_{R,f} i\gamma^\mu D_\mu \psi_{R,f} - \frac{1}{4} W_a^{\mu\nu} W^a_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu}. \tag{2.10}$$

The invariance under $SU(2) \times U(1)$ transformations is guaranteed using the covariant derivative:

$$D_\mu = \partial_\mu - i\frac{g_1}{2} Y_W B_\mu - i\frac{g_2}{2} I_j W^j_\mu, \tag{2.11}$$

where g_1 and g_2 are the coupling constants of the $U(1)$ and $SU(2)$ groups, B_μ denotes a gauge boson with spin 1 that mediates the part of the weak interaction associated with the hypercharge, and $W^j_\mu, j = 1, 2, 3$ denotes 3 gauge bosons of spin 1 that couple to the weak isospin. The Lagrangian density that describes the free motion of these new fields is obtained from the field tensors:

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g_2 \epsilon^{abc} W^b_\mu W^c_\nu, \tag{2.12}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{2.13}$$

As for QED or QCD, the gauge bosons that mediate the weak interaction in Eq. (2.10) are massless. However, the weak interaction observed experimentally is short ranged and mediated by massive bosons. In addition, the split in left-handed and right-handed fermions breaks the invariance of the mass term in the electroweak Lagrangian. For these two reasons, a different mechanism is required to generate the masses of the gauge bosons and the fermions.

2.1.4 Particle masses and the Higgs boson

In order to guarantee the invariance of the Lagrangian while including massive particles, the electroweak symmetry is spontaneously broken through the Brout-Englert-Higgs mechanism [6, 7, 8]. The simplest solution is provided by adding two additional complex fields of spin 0. In order to generate the mass of the electroweak gauge bosons, one of the fields must be neutral ϕ^0 and the other one must be charged ϕ^+. The minimal model consists of these two scalar fields placed in a weak isospin doublet:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.14}$$

The Lagrangian density that describes its motion and interaction is invariant under transformations of the $SU(2)_L$ symmetry group:

$$\mathcal{L}_{\text{Higgs}} = D^\mu \phi^\dagger D_\mu \phi - V(\phi) = D^\mu \phi^\dagger D_\mu \phi - \mu^2(\phi^\dagger \phi) - \lambda(\phi^\dagger \phi)^2, \tag{2.15}$$

where $V(\phi)$ is the potential of the scalar field ϕ, also known as the Higgs potential. The potential $V(\phi)$, shown in Fig 2.2, is composed of two free parameters μ and λ. For $\mu^2 < 0$, the potential has an infinite set of minima satisfying:

$$\phi^\dagger \phi = \frac{1}{2}(\phi_1 + \phi_2 + \phi_3 + \phi_4) = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}, \tag{2.16}$$

where $v = \sqrt{-\frac{\mu^2}{\lambda}}$ is the vacuum expectation value of ϕ.

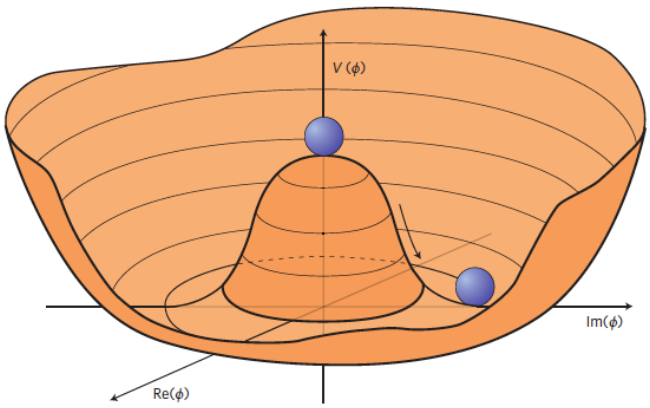


Figure 2.2: The potential $V(\phi)$ for the Higgs field and parameters $\mu^2 < 0$ and $\lambda > 0$.

After the electroweak symmetry breaking, the photon is required to be massless and therefore the neutral field ϕ^0 is required to have only one non-zero vacuum

expectation value,

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix}. \tag{2.17}$$

The field ϕ is expanded around this minimum and can be expressed as:

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v+h(x)\end{pmatrix}. \tag{2.18}$$

In order to mathematically describe the physical gauge fields observed in experiments, the Lagrangian density can be expressed as:

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} = &\frac{1}{2}\partial_\mu h \partial^\mu h + \frac{g_2^2}{2}(v+h)^2 W_\mu^+ W^{-\mu} + \frac{1}{8}\frac{g_2^2}{\cos\left(\theta_W\right)^2}(v+h)^2 Z_\mu Z^\mu \\ &+ m_\gamma A_\mu A^\mu + \frac{\mu^2}{2}h^2 - \lambda v h^3 - \frac{\lambda}{4}h^4,\end{aligned} \tag{2.19}$$

where $W_\mu^\pm$ is the field associated with the $W^\pm$ boson, Z_μ the field associated with the Z boson and A_μ the field associated with the photon. The physical fields are obtained from the gauge fields through the mixing angle θ_W:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2), \tag{2.20}$$

and

$$\begin{aligned}Z_\mu &= -\sin\left(\theta_W\right)B_\mu + \cos\left(\theta_W\right)W_\mu^3, \\ A_\mu &= \cos\left(\theta_W\right)B_\mu + \sin\left(\theta_W\right)W_\mu^3.\end{aligned} \tag{2.21}$$

The mass of each boson arising from the BEH mechanism, especially the W and Z bosons mediating the weak interaction, can be identified as:

$$m_W = \frac{v g_2}{2},\ m_Z = \frac{m_W}{\cos\theta_W},\ m_\gamma = 0. \tag{2.22}$$

The excitation of the physical field h is known as the Higgs boson. Its mass is defined as:

$$m_H = \sqrt{-\mu^2}. \tag{2.23}$$

In addition to generating the masses of the W, Z and Higgs bosons, the Brout-Englert-Higgs mechanism can be used to generate the masses of the fermions. Due to the non-vanishing expectation value of the scalar field, the interaction between the fermions and the field is expressed by a Lagrangian density that is invariant

under transformations of the $SU(2)_L \times U(1)_Y$ symmetry group:

$$\mathcal{L}_{\text{Yukawa}} = \sum_f -\frac{y_f v}{\sqrt{2}}\bar{\psi}_f\psi_f - \frac{y_f}{\sqrt{2}}\bar{\psi}_f\psi_f h, \tag{2.24}$$

where y_f is a fundamental coupling constant, known as the Yukawa coupling. It can be chosen to be consistent with the observed masses of particles:

$$y_f = \sqrt{2}\frac{m_f}{v}. \tag{2.25}$$

In the SM, the Yukawa coupling is the only interaction that distinguishes between the three generations of particles and is therefore the only part of the theory that is not flavour universal. For example, the main difference between the quarks u, c and t is the mass.

For the quark sector, one additional consequence is that the Yukawa interaction allows transitions between the different quark flavours. The strength of these flavour-changing interactions, mediated by the $W^{\pm}$ boson, is described by the Cabbibo-Kobayashi-Maskawa matrix:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{cd} & V_{td} \\ V_{us} & V_{cs} & V_{ts} \\ V_{ub} & V_{cb} & V_{tb} \end{pmatrix}. \tag{2.26}$$

These interactions are observed as flavour changing charged currents between the up type and down type quarks. This matrix is measured to be almost diagonal and the consequence is that the properties of the hadronic bound states formed from the quarks vary according to their flavour. For example, the favoured decay mode of c-hadrons is the $c \to s$ transition. For b-hadrons, the CKM favoured decay mode is the transition to a quark t but due its very large mass with respect to the other quarks, this mode is kinematically suppressed and b-hadrons tend predominantly to decay to c-hadrons. The physical consequence is that the average lifetime of b-hadrons is significantly longer than the lifetime of c-hadrons, which can directly decay through the favored $c \to s$ transition.

As formulated, the SM is a mathematical description of all the observed particles as well as their interactions. It provides a theoretical explanation to the observed physics properties of elementary particles such as the electric and colour charges, the isospin and the dynamic process through which particles acquire a mass. However, this description relies on a set of fundamental constants and their values are not predicted.

2.1.5 Fundamental constants

A fundamental constant is a degree of freedom of a theory that cannot be expressed in terms of other parameters. The numerical values themselves are not predicted

and are determined only from measurements. In order to describe particles and their interactions, the SM requires at least 18 fundamental constants:

- The coupling strengths of the electroweak and strong interactions: g_1, g_2 and g_s,
- The 9 masses of up and down quarks and charged leptons,
- The 3 mixing angles and 1 phase of the Cabbibo-Kobayashi-Maskawa matrix,
- The mass of the Higgs boson and the vacuum expectation value.

Additional fundamental constants related to the charge parity violation in the QCD interaction and the oscillation of neutrinos may exist, but they are not relevant for the work of this thesis and hence are not included in this list.

The majority of these fundamental constants are related to the Brout-Englert-Higgs mechanism. One of the major goals of modern particle physics is to measure these parameters and their direct and indirect effect on physics observables. Any deviations between different measurements are a sign of physics beyond the SM and could potentially shed a new light on our understanding of fundamental particles and their interactions. In that regard, the physics program of the LHCb experiment which targets measurements of the flavour structure of the SM is complementary to the physics program of the ATLAS and CMS experiments, which probe the properties of the Higgs boson and the mass structure of the SM.

2.2 Properties of the Higgs boson

In order to test the Brout-Englert-Higgs mechanism and measure the fundamental constants as well as the resulting predictions, the masses of the fermions are investigated:

$$\mathcal{L}_{\text{Yukawa}} = \sum_f -m_f \bar{\psi}_f \psi_f - \frac{m_f}{v} \bar{\psi}_f \psi_f h. \tag{2.27}$$

The first term represents fermions with a mass m_f and the second term represents the interaction of the fermions with the field associated with the Higgs boson. This interaction is proportional to the mass of the fermion and manifests itself in terms of a coupling to the Higgs boson.

In principle, the Higgs boson can interact with all massive particles of the SM. However, due to the proportionality of the coupling to the mass of the particles, the largest couplings are obtained for the most massive particles. The branching ratio, which quantifies the probability that the Higgs boson decays in a certain mode, is defined as:

$$B(H \to X\bar{X}) = \frac{\Gamma(H \to X\bar{X})}{\sum_i \Gamma(H \to X_i \bar{X}_i)}, \tag{2.28}$$

where Γ is the partial width of the decay of the Higgs boson to the particle X. In the case of the Higgs boson interaction with quarks, the partial width is expressed as:

$$\Gamma(H \to q\bar{q}) = 3 \times \frac{m_q^2 m_H}{8\pi v}, \tag{2.29}$$

where the factor 3 accounts for the different color charges. Given the measurement of the masses of the fermions, the mass of the Higgs boson and the vacuum expectation value, the production and decay rates of the Higgs boson can be predicted and therefore compared to the data collected in experiments. Similarly for leptons, the W and Z bosons, the partial width can be expressed as a function of the masses. The Higgs boson does not directly interact with the massless photons and gluons, however, the decay of a Higgs boson to photons and gluons is possible through virtual loops of massive particles.

For the charged leptons and quarks, the masses are determined from indirect experimental measurements, thus fixing the predicted Higgs branching fractions under the SM. The mass of the Higgs boson is determined to be $m_H = 125.09 \pm 0.21\ (\text{stat}) \pm 0.11\ (\text{syst})$ GeV by the ATLAS and CMS experiments [9] and the vacuum expectation value v is estimated from precise measurements of the mass of the W boson and its coupling to fermions.

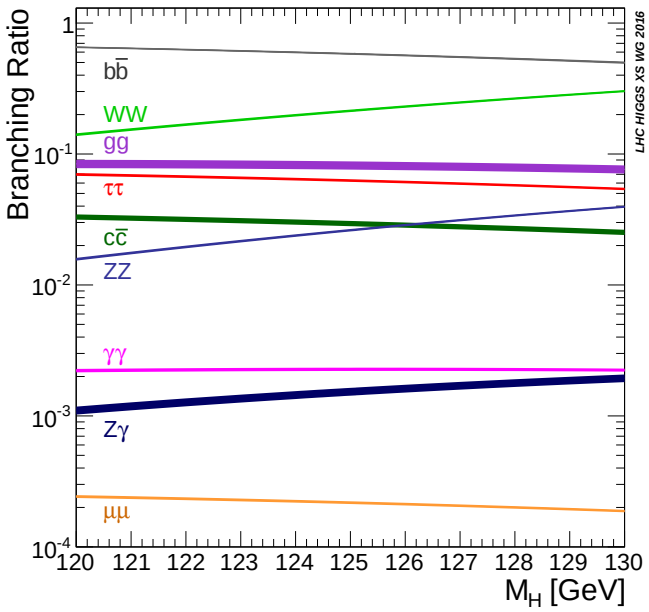
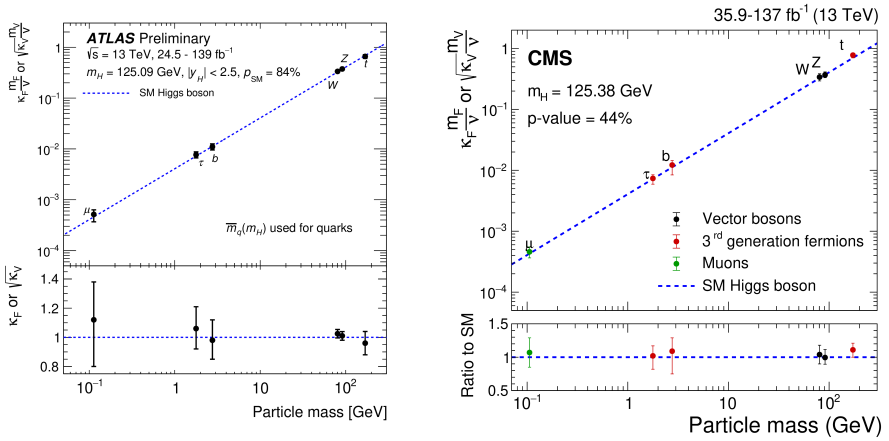


Figure 2.3: Higgs boson branching ratios and their uncertainties for the mass range around 125 GeV [10].

The largest branching ratio of the Higgs boson decays predicted by the SM, listed in Fig. 2.3, is the $H \to b\bar{b}$ decay, followed by the $H \to W^+W^-$ decay. Due to the fact that $m_t > m_H$, the decay of the Higgs boson to a pair of top quarks is suppressed. The Higgs coupling to top quarks can nevertheless be measured through the production mode, where a Higgs boson is produced in association to two top quarks, and no kinematic constraints exists. In addition, the Higgs boson can decay to massless particles, $H \to gg$ and $H \to \gamma\gamma$, through loops of virtual particles.



(a) ATLAS (b) CMS

Figure 2.4: Coupling-strength modifiers κ as a function of the particles masses measured by the ATLAS [11] and CMS [12] experiments. All measurements are found in agreement with the SM within the uncertainties ($\kappa = 1$).

The measurement of the Higgs boson interaction to massive particles is one of the most important challenges in modern particle physics. All interactions measured so far by the ATLAS and CMS experiments, shown in Fig. 2.4, are found to be in agreement with the SM prediction. Due to the proportionality of the coupling of the Higgs boson to the mass of particles, the interactions with the third generation fermions and the vector bosons were observed first. Furthermore, by exploiting the more accurate reconstruction of muons with respect to quarks, the detectors were already sensitive to the interaction with the muons of the second generation of particles. However, there is currently no experimental proof that the Higgs boson interacts with particles of the first generation nor with the second generation of quarks. This is striking, the heaviest particle in the second generation of fermions is the charm quark and the decay of the Higgs boson $H \to c\bar{c}$ represents 2.9% of the total decay rate, compared to a decay fraction of 0.02% of Higgs boson decay

to muon $H \to \mu^+\mu^-$.

2.3 Enhanced coupling to first and second generation of quarks

Despite the fact that the SM provides a description of fundamental particles as well as their interactions, it does not explain the existence of three generations of fermions nor the large differences observed in their masses. It also does not explain the presence of dark matter in the universe, which accounts for about 25% of its total content and could potentially be explained by a new type of massive particles.

Precision tests of the SM are intrinsically linked to searches for physics beyond the SM: higher-scale particles participating in loop corrections to the SM can have an impact on the measured cross-sections and kinematic observables. Any deviation of the predictions obtained with the SM would provide important insights for a new type of physics. Measuring the Higgs boson coupling to the lighter generations of fermions is therefore one of the main objectives of modern particle physics as any enhanced or decreased coupling could be a sign of a more general theory that could uncover a more complex structure in our understanding of the universe.

There are several theories involving additional Higgs bosons that predict an interaction between the Higgs boson and the lighter generations of fermions that is different from the SM. In these theories, the Higgs potential $V(\phi, \phi')$ is constructed from two scalar fields and chosen to satisfy the local gauge invariance requirement. These models, referred to as two-Higgs-doublet models [13], result in 5 Higgs bosons: three neutral bosons h, H and A and two charged bosons $H^{\pm}$. In these theories, the discovered Higgs boson with a mass of 125 GeV is identified as the boson h. These models require the addition of 6 free parameters: the masses of 4 types of Higgs bosons, the ratio of the vacuum expectation values $\tan\beta = \frac{v}{v'}$ and a mixing angle α between the h and H bosons.

One version of this theory considers that only fermions of the third generation acquire their mass through the Brout-Englert-Higgs mechanism and the interaction with the field ϕ that corresponds to the Higgs boson h. The observed masses of the two first generations of fermions can be generated by the addition of the new Higgs field ϕ'. The Yukawa interaction between the fermions and the field ϕ' are chosen to be:

$$Y'_{\mathrm{down-type}} = \begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & 0 & 0 \end{pmatrix}, \; Y'_{\mathrm{up-type}} = \begin{pmatrix} \times & \times & 0 \\ \times & \times & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{2.30}$$

where the non-zero values are chosen to correspond to the quark mixing observed in the V_{CKM} matrix. Due to the mixing between the bosons h and H, these Yukawa interactions also enhance the h boson decays to the two first generations with respect to the SM. In general, the two-Higgs-doublet models allow for flavour

changing neutral currents, where a fermion of a given flavour decays in a fermion of a different flavour while preserving the same electric charge throughout the process. These processes are not observed experimentally, which provides strong constraints on these theories. Within the current constraints, this model allows for a coupling of the Higgs boson to fermions enhanced by a factor 3 with respect to the SM [14].

Another class of theories predicts that the interaction of the Higgs boson of the SM with the first two generations of fermions is enhanced due to a mixing between the two neutral Higgs fields. In this specific two-Higgs-doublet model, the two scalar fields ϕ and ϕ' are chosen as:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h_1 \end{pmatrix}, \; \phi' = \frac{1}{\sqrt{2}} \begin{pmatrix} H^+ \\ h_2 + i h_3 \end{pmatrix}. \tag{2.31}$$

In this representation, the field ϕ is responsible for electroweak symmetry breaking and gives the mass to the fermions of the SM. The field ϕ' is chosen with no vacuum expectation value $v' = 0$. In the case of quarks, for the first Higgs field ϕ, the Yukawa couplings are chosen such that they correspond to the observed masses:

$$\begin{aligned} Y_{\mathrm{down-type}} &= \begin{pmatrix} y_d^{\mathrm{SM}} & 0 & 0 \\ 0 & y_s^{\mathrm{SM}} & 0 \\ 0 & 0 & y_b^{\mathrm{SM}} \end{pmatrix} = \frac{\sqrt{2}}{v} \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix}, \\ Y_{\mathrm{up-type}} &= \begin{pmatrix} y_u^{\mathrm{SM}} & 0 & 0 \\ 0 & y_c^{\mathrm{SM}} & 0 \\ 0 & 0 & y_t^{\mathrm{SM}} \end{pmatrix} = \frac{\sqrt{2}}{v} \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix}. \end{aligned} \tag{2.32}$$

On the other hand, the Yukawa matrices Y' for the second field ϕ' are not constrained by any measurements and are in principle arbitrary. Flavour changing neutral currents are avoided by spontaneous flavour violation [15], which means that the Yukawa matrices Y' have the same diagonal form as the Yukawa matrices Y and the up-type quarks are affected differently with respect to the down-type quarks:

$$\begin{aligned} Y'_{\mathrm{down-type}} &= \xi Y_{\mathrm{down-type}}, \\ Y'_{\mathrm{up-type}} &= \begin{pmatrix} y'_u & 0 & 0 \\ 0 & y'_c & 0 \\ 0 & 0 & y'_t \end{pmatrix}, \end{aligned} \tag{2.33}$$

where ξ is a proportionality constant and y' are the coupling constants to the second Higgs field ϕ' to the up-type quark generation. The physical Higgs fields

are obtained through the mixing of the two fields h_1 and h_2:

$$\begin{aligned} h &= \sin(\beta-\alpha)h_1 + \cos(\beta-\alpha)h_2, \\ H &= -\cos(\beta-\alpha)h_1 + \sin(\beta-\alpha)h_2. \end{aligned} \tag{2.34}$$

where $\beta-\alpha$ is the alignment angle, h is the Higgs boson observed at 125 GeV and H is a second Higgs boson. In the limit where $\cos(\beta-\alpha)=0$, the Higgs boson h corresponds exactly to the Higgs boson of the SM. However, in the case where the two Higgs bosons mix ($\cos(\beta-\alpha)\neq 0$), the interaction of the Higgs boson h with up-type quarks depends on the Yukawa coupling y':

$$\lambda^h_{\mathrm{q_u}\bar{\mathrm{q}}_\mathrm{u}} = y^{\mathrm{SM}}_{q_u}\sin(\beta-\alpha) + y'_{q_u}\cos(\beta-\alpha). \tag{2.35}$$

In this model, the interaction between the Higgs boson at 125 GeV and the up-type quarks of the first two generations can be enhanced while the interaction with the down-type quarks is proportional to the interaction within the SM and therefore relatively small due to the mass. This prediction is particularly interesting for the interaction of the SM Higgs boson with charm quarks, which is not yet experimentally observed and sensitive to this enhanced coupling.

The decay rate of the Higgs boson to charm quark $H\to c\bar{c}$ is sensitive to coupling enhancements $\lambda^h_{c\bar{c}}/\lambda^{h_{\mathrm{SM}}}_{c\bar{c}}$. This enhancement, shown in Fig. 2.5, depends on the coupling y'_c (labelled κ_c in the figure, which is not the same κ as in Fig 2.4) and the alignment angle $\cos(\beta-\alpha)$. Enhancement of the decay $H\to c\bar{c}$ by more than a factor of 3 are constrained by the measurement of the Higgs boson decay to photons $H\to\gamma\gamma$, where the Higgs boson is produced through the fusion of two gluons.

A different class of theories considers that other new particles can affect the mass of the particles of the SM. Their effect is parametrised using effective field theories, where higher order terms are added to the Lagrangian of the SM and expanded in a scale parameter Λ, which is typically chosen to be large, but should be below the energy scale of any physics beyond the SM. The Lagrangian becomes

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_k \frac{1}{\Lambda^k}\sum_i C_{ik}(\Lambda)O_{ik}, \tag{2.36}$$

where O are operators that contain the contribution of new physics and $C(\Lambda)$ are coefficients that depend on the energy scale of the new physics process. In these models, the additional Lagrangian can be chosen such that the Yukawa interaction of elementary fermions is modified:

$$y_q = \frac{m_q}{v} - \frac{v^2}{\Lambda^2}\frac{C_{qH}}{\sqrt{2}}. \tag{2.37}$$

For specific values of the parameters Λ and C_{qH}, these theories predict that the

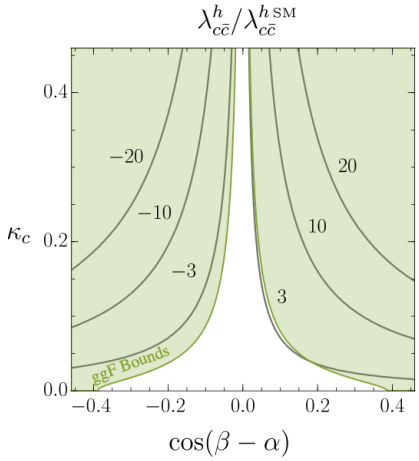


Figure 2.5: Enhance coupling of the Higgs boson of the SM with charm quarks $\lambda^h_{c\bar{c}}/\lambda^{h_{\mathrm{SM}}}_{c\bar{c}}$ as a function of the Yukawa couplings κ_c of the second Higgs field ϕ' and the alignment parameter $\cos(\beta-\alpha)$. The notation of the paper [15] is used and the κ_c parameter is not the same as the κ_i in Fig. 2.4 and generally the κ framework. The green bands are obtained from the experimental measurement of the SM Higgs boson measured in the gluon-gluon fusion (ggF) production mode and a decay to two photons $H \to \gamma\gamma$.

Higgs boson couples universally to the b- and c-quarks, i.e $y_c \approx y_b^{SM} = 4.5\, y_c^{SM}$ [16]. Other new physics scenarios, such as vector-like-quarks or heavy scalar fields, can also result in these type of terms in the effective Lagrangian. The existence of all these theoretical models motivates the measurement of the Higgs boson interaction with quarks from the first two generations.

2.4 Event simulation

To compare theory predictions and experimental measurements, theorists and experimentalists meet on a common ground: the cross-section σ of a particular process under study. On one side, the cross-section is computed by theorists starting from the Lagrangian for a given process, for example in the collision of a

particle a and b:

$$\sigma = \frac{1}{4E_a E_b |\vec{v}_a - \vec{v}_b|} \int \prod_f \left(\frac{d^3 p_f}{(2\pi)^3} \frac{1}{2E_f} |\mathcal{M}_{fi}|^2 (2\pi)^4 \delta^4 (p_a + p_b - \sum_f p_f) \right), \tag{2.38}$$

where f corresponds to the particles produced in the collision, E the energy, p the momenta and $|\mathcal{M}_{fi}|^2$ corresponds to the probability that a given process happens from an initial state with two particles colliding to produce a final state with N particles, also referred to as the matrix element squared. On the other side, experimentalists measure this cross-section at particle colliders:

$$\sigma = \frac{N_{\text{events}}}{\epsilon \cdot L_{\text{int}}}, \tag{2.39}$$

where N_{events} is the number events measured for a given process corrected for the phase-space acceptance ϵ and normalised by the integrated luminosity L_{int}.

In order to study the properties of the Higgs boson, events predicted by the theory are simulated and compared to collected data. In the ATLAS experiment, the Higgs bosons are produced in proton-proton collisions, illustrated in Fig. 2.6. The content of the protons are modelled by the probability density function to find a quark with a certain longitudinal momentum fraction.

During a collision, two constituents of each proton (quarks or gluons) undergo a hard scattering with a high momentum transfer that produces particles such as the Higgs boson. Based on the lifetime, these particles decay in a cascade that produces final state particles that are stable enough to be measured in the detector.

In the leptonic sector, the electrons and muons are measured in the detector. The quarks produced in the decay of the intermediary particles have a high energy and produce showers of partons until the energy of all the produced particles has decreased below a certain energy scale of the order of $\mathcal{O}(1)$ GeV, at which the strong force becomes the dominant interaction. At this point, due to confinement properties of QCD, the quarks hadronise and produce a collimated spray of hadrons that are reconstructed as jets. Due to the large transfer of energy, the intial and final partons can emit QCD radiation in the form of a gluon emissions. Depending if these gluons were emitted before or after the hard scattering, they are referred to as initial or final state radiations.

The remaining constituents of the proton that are not part of the hard scattering are referred to as the underlying event. These remnants also undergo QCD interactions. The other protons in the beam also interact and create additional vertices that are referred to as pileup and are recorded together with the hard scattering process as a single event.

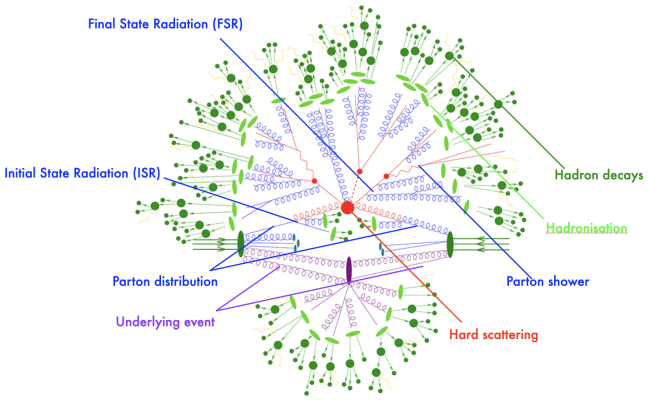


Figure 2.6: Illustration of a proton-proton collision.

2.4.1 Renormalisation

Solving exactly the equations of motion for interacting particles in the SM is generally not possible. Instead, calculations are performed within perturbation theory, using Feynman diagrams. At leading order in perturbation theory, the SM results in finite predictions. Nevertheless, at higher order, the predictions for physical observables diverge due to loop corrections, i.e shown in Fig 2.7.

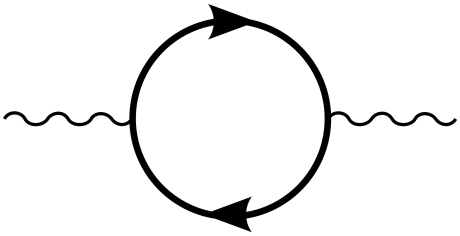


Figure 2.7: Example of a loop correction to a vertex in a Feynman diagram.

In order to obtain finite predictions, these divergences are renormalised depending on an arbitrary scale parameter [17, 18], that removes the divergences by expressing the the quantities of the theory, such as coupling and masses, in terms of measured finite quantities. The consequence is that physical quantities change their value as the newly introduced scale parameter changes. Including higher order contributions in the calculation reduces the dependency of predicted observable quantities on the scale parameter.

For example, in the context of analyses targeting quarks in the final states, the renormalisation of the strong coupling of the QCD interaction needs to be taken

into account. The strong coupling constant is defined as:

$$\alpha_s = \frac{g_s}{4\pi}, \tag{2.40}$$

which depends on the threshold Q

$$Q = \frac{\mu_R}{\lambda_{QCD}}, \tag{2.41}$$

where μ_R is a scale parameter and λ_{QCD} is the energy scale at which non-perturbative effects become dominant. Up to leading order, the strong coupling can be expressed as:

$$\alpha_s(Q) = \frac{12\pi}{(11n_c - 2n_f)\ln Q^2}, \tag{2.42}$$

where n_c and n_f are the number of quark colours and flavours respectively. Eq. (2.42) shows that the strong coupling weakens at high energies. At low energies, the coupling diverges as the non-perturbative effects become larger. One consequence is that, to describe parton distribution functions of the proton, empirical measurements are needed, as the processes that govern these distributions in the proton are firmly in the non-perturbative regime.

2.4.2 Factorisation and parton distribution function

The LHC does not collide free quarks, but protons, which are composite bound states of three strongly interacting quarks (uud), gluons, and sea quarks ($q\bar{q}$). Each constituent of the proton carries a fraction $x_i \in [0, 1]$ of the total momentum of the proton, $\vec{p_i} = x_i \vec{p}_{\text{proton}}$, assuming that the partons move collinear to the surrounding proton.

In order to calculate the probability that a proton-proton initiated process results in a final state f, the cross-section is expressed as the convolution of a non-perturbative, but universal, parton distribution function $f_{i/h}(x_i, \mu_f^2)$, describing the probability of finding a parton i with a longitudinal momentum fraction x_i inside a hadron h at a given energy scale Q, and the partonic cross-section of the hard-scattering process, $\hat{\sigma}_{i+j\to f}$, calculated perturbatively:

$$\sigma_{h_1+h_2\to f}(s, \mu_f, \mu_r) = \sum_{i,j} \int_0^1 dx_1 \int_0^1 dx_2 \, f_{i/h_1}(x_1, \mu_f^2) \, f_{j/h_2}(x_2, \mu_f^2) \, \hat{\sigma}_{i+j\to f}(\hat{s}, \mu_f, \mu_r) \tag{2.43}$$

where μ_r the renormalisation scale, $\hat{s}$ is the fraction of the center-of-mass energy s used in the partonic cross-section and μ_f is the energy scale at which the PDF is evaluated. The separate of the perturbative and non-perturbative part of the calculation, expressed as a convolution in Eq.(2.43), is possible through the

factorisation theorem [19].

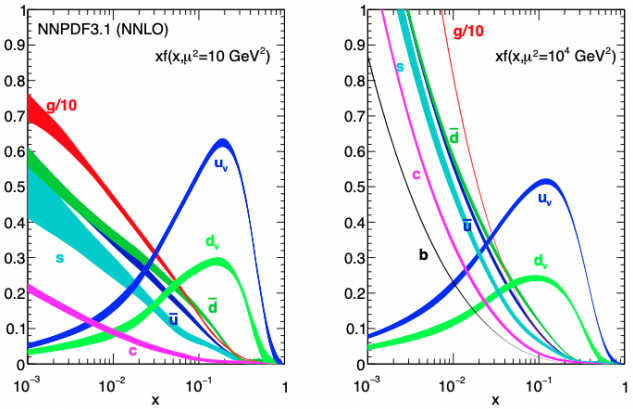


Figure 2.8: Parton distribution functions provided by the NNPDF collaboration and shown as a function of the fraction x, evaluated at $\mu_f^2 = 10$ GeV^2 (left) and $\mu_f^2 = 10^4$ GeV^2 (right) [20].

Due to the non-perturbative regime, the parton distribution functions are not calculable but need to be measured at an arbitrary scale μ_0 in well-understood processes, such as deep inelastic scatterings and $pp \to jets$. As the parton distribution functions are assumed to be universal, they can be extrapolated to other scales as a function of a factorisation scales [20, 21, 22]. The parton distribution functions, shown in Fig. 2.8, are compared for two different scales. The valence u quarks carry a larger fraction of the proton energy on average, twice more than the d quark due to the proton composition. At high energies, the effect of sea quarks and gluons is more important and the fraction carried by the valence quarks is on average reduced.

2.4.3 Monte-Carlo simulations

Simulated events of the Higgs boson production and relevant background processes are generated with Monte-Carlo techniques. The simulation process describes all components of the collision process previously mentioned and listed in Table 2.2: the hard scattering (HS), shower (S), hadronisation (H) and underlying event (UE). From a technical point of view, one of the most practical generators is the `Sherpa` generator which describes the full chain of physics processes in a single software package. Other generators are used in combinations to jointly simulate the full simulation chain, for example `Powheg+Pythia8`, where `Powheg` is used for the hard scattering and `Pythia8` is used for the parton shower and the hadronisation. Simulated event samples are used to calculate predictions of the cross-section of

a given process as well as the topology of the event. Finally, for each event, the detector response is simulated using the Geant4 [23, 24] package after the physics simulation in order to generate an event that can be compared to the observed data.

MC generator	HS	PS	UE	H	Accuracy
`Sherpa` [25]	×	×	×	×	NLO
`Powheg` [26]	×		×		NLO
`Madgraph5+aMC@NLO` [27]	×		×		NLO
`Pythia8` [28]		×	×	×	LO
`Herwig7` [29]		×	×	×	LO

Table 2.2: Overview of the Monte-Carlo generators which are used in this thesis. The various processes of the event simulations are listed for each generator: the hard scattering (HS), shower (S), hadronisation (H) and underlying event (UE)

3 The ATLAS experiment at the LHC

As the Higgs boson couples to elementary particles proportionally to their mass, a large number of Higgs bosons are produced in proton-proton collisions. The Large Hadron Collider (LHC) [30, 31, 32] located at CERN is the largest proton-proton collider in the world. Experiments which study the particles produced in the collisions are located at 4 interaction points: the ATLAS [33, 34], CMS [35, 36], LHCb [37] and ALICE [38] experiments.

Out of these, the ATLAS and CMS detectors are designed to discover higher-mass particles like the Higgs boson or hypothetical particles from theories beyond the SM, as well as to perform high precision measurements of particles like the top quark. Due to their varied physics program, these two detectors are often referred to as general-purpose detectors. On the other hand, the ALICE detector is dedicated to studies of heavy-ion collisions and focuses on high-density QCD states, while the LHCb detector is optimised to investigate hadrons composed of heavy quarks. This chapter concentrates on the ATLAS detector, which is used to perform the analysis presented in this thesis.

3.1 The ATLAS detector

The ATLAS detector is a cylindrical detector comprised of charged particle trackers, electromagnetic and hadronic calorimeters and a muon detector. Its design allows a broad physics program that aims at precision measurements of physical constants such as particle masses, couplings and cross-sections, but also the observation of rare processes such as the Higgs boson production in association with $t\bar{t}$-pair [39], light-by-light scattering [40] or the decay of heavy mesons such as $B_s^0 \to \mu^+\mu^-$ [41]. Between 2015 and 2018, the ATLAS experiment recorded proton-proton collisions delivered by the LHC with a total integrated luminosity of 139 fb^{-1} at a center-of-mass energy of $\sqrt{s} = 13$ TeV.

The ATLAS detector, shown in Fig. 3.1, is 44 m long, has a diameter of 25 m and weighs over 7000 tons. In order to measure the momentum of charged particles produced at the interaction point, the detector relies on four magnets: a 2 T central solenoid, an 8-coil barrel toroid that is cylindrically placed around the detector generating a peak field of 4 T and two other 8-coil magnets at the detector endcaps with a field of 4 T too.

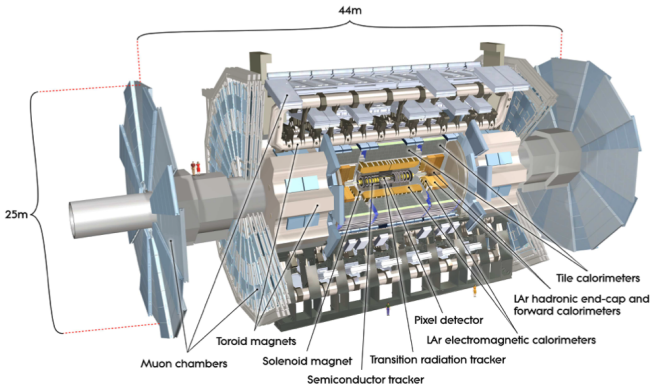


Figure 3.1: Illustration of the ATLAS detector [34].

The coordinate system of ATLAS is defined as right-handed and centered at the interaction point, with the beam axis chosen as the z-axis, the x-axis pointing towards the center of the LHC ring and the y-axis pointing upwards. A cylindrical coordinate system is used to define positions in the detector, where the azimuthal angle ϕ is defined on the $x - y$ plane and the polar angle θ is the angle from the z-axis. As the collisions happen on the z-axis, the energy of the interactions on the $x - y$ plane, referred to as the transverse plane, is conserved and physics quantities such as the momentum are projected on the transverse plane as $p_{\mathrm{T}} = p \sin(\theta)$.

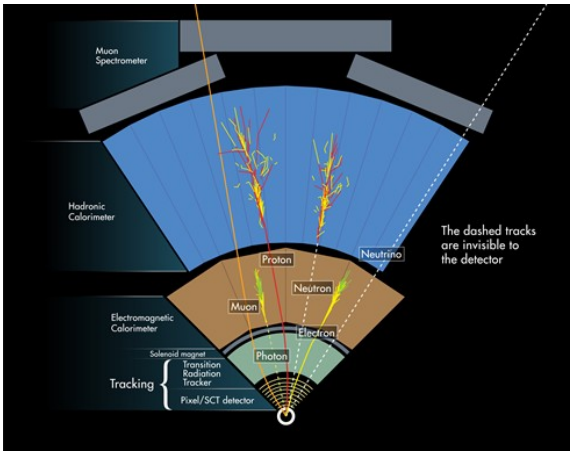


Figure 3.2: Illustration of the signatures of different particles inside the ATLAS detector [42].

In order to identify particles, the different sub-detectors are designed such that each type of particle has a combined signature that is unique, as shown in Fig. 3.2. Charged particles are measured in the tracker and, due to the magnetic field, their tracks are bent when they arrive to the calorimeters. Electrons are measured and stopped in the electromagnetic calorimeter, while hadrons such as protons are measured and stopped in the hadronic calorimeter. The neutral hadrons, such as the neutron, are distinguished from the charged hadrons due to their lack of tracks in the inner detector. Since for the muons, the main loss of energy occurs through ionisation, muons interact minimally with the calorimeters, and will generally traverse them with only a small loss of energy. To identify muons and to better measure their properties, large tracking detectors are installed as the outer layer of the detector. Photons do not leave tracks as they carry no electric charge, but can be measured in the electromagnetic calorimeter, unless they convert to a pair of e^+e^-, in which case the electrons generate reconstructable tracks. Neutrinos do not interact with the detector, but their presence can be inferred from the imbalance of energy in the transverse plane.

Detector component	Required resolution	η coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\%\ p_T \bigoplus 1\%$	$\lvert\eta\rvert < 2.5$	
EM calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \bigoplus 0.7\%$	$\lvert\eta\rvert < 3.2$	$\lvert\eta\rvert < 2.5$
Hadronic calorimetry			
Barrel and endcap	$\sigma_E/E = 50\%/\sqrt{E} \bigoplus 3\%$	$\lvert\eta\rvert < 3.2$	$\lvert\eta\rvert < 3.2$
Forward	$\sigma_E/E = 100\%/\sqrt{E} \bigoplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon detector	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV	$\lvert\eta\rvert < 2.7$	$\lvert\eta\rvert < 2.4$

Table 3.1: General performance of the ATLAS detector [34].

The ATLAS detector is designed to accurately measure the energy and momentum resolution of particles, as summarised in Table 3.1. In the tracker, the momentum is determined based on the bending of particles in the magnetic field. Particles with a high momentum are less bent and the relative uncertainty on the momentum of the tracks increases with the momentum. In the calorimeters, particles with a higher energy are more likely to create large showers and the relative uncertainty on the energy measurement decreases for particles with higher energy.

In the region $|\eta| < 2.5$, both the tracker and calorimeter measurement can be used. While signatures of jets initiated by quarks can be reconstructed up to $|\eta| < 4.9$ based on their deposit in the calorimeter, the identification of the flavour of the jet, which relies on tracking information can only be performed up to $|\eta| < 2.5$. The next sections describe the three sub-detectors of ATLAS: the inner detector, the calorimeters and the muon detectors in more details.

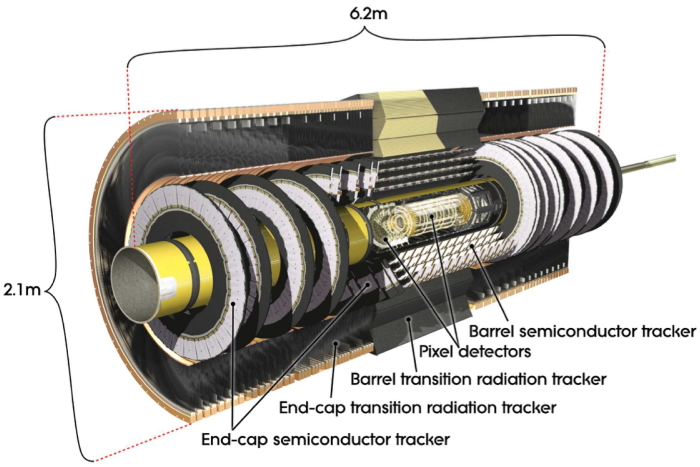


Figure 3.3: Cut-away view of the ATLAS inner detector.

3.1.1 Inner detector

The goal of the inner detector is to measure the tracks of the charged particles. It is a necessary component in order to distinguish between electrons and photons but also to reconstruct the sub-structure of jets and the secondary decay vertices that reveal information on the flavour of the quarks that initiated them. This is particularly relevant in the search of the $H \to c\bar{c}$ process, which requires the ability to identify c-quarks in the detector. The inner detector, shown in Fig. 3.3, is made of 4 sub-detectors: 3 silicon based detectors and 1 drift-tube detector. The 4 detector components from inside to out are the insertable b-layer, the pixel detector, the semi-conductor tracker and the transition radiation tracker, as illustrated on Fig. 3.4.

Each sub-detector is designed to achieve μm precision on the reconstruction of tracks. The best spatial resolution, listed in Table 3.2 for all components of the inner detector, is obtained by the insertable b-layer and the pixel detector, which are closest to the interaction point and have the highest granularity. Due to the importance of track reconstruction for identification of the flavour of heavy hadrons within jets, a detailed description of the sub-detectors of the inner tracker is given below.

Pixel detector and insertable B-layer The pixel detector is the innermost sub-detector of the ATLAS tracker. It consists of three layers of modules in the barrel region and three disks in each of the two end-cap structures to guarantee at least three space points over the full tracking pseudo-rapidity range. The detector consists of 1744 silicon pixel modules, with 47232 pixels per sensor, and the sensor

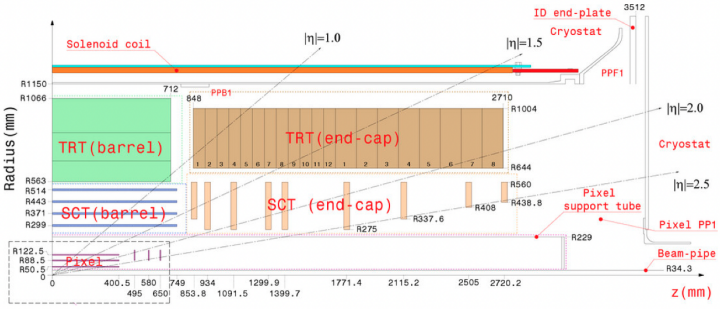


Figure 3.4: Plan view of a quarter-section of the ATLAS inner detector showing each of the major elements with its active dimensions [43]. The insertable B-layer is not represented on this diagram but is located just below the pixel detector, nearest to the interaction point.

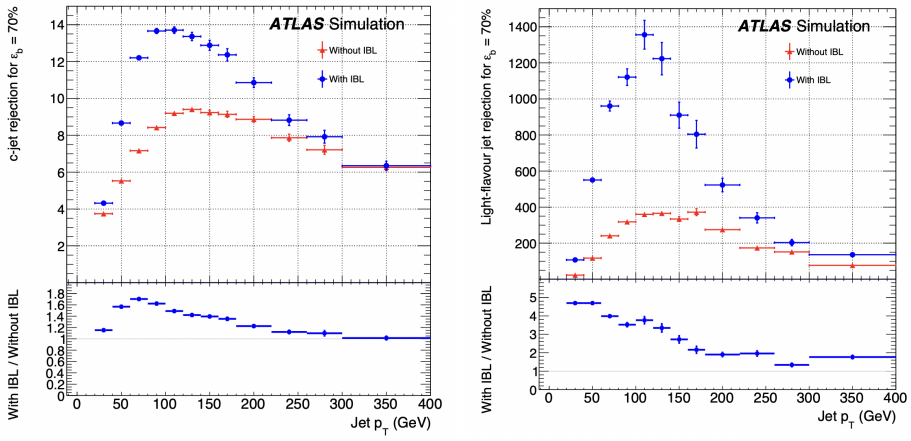
Sub-detector	Spatial resolution	
	$(R-\phi)$ plane	z-axis
Insertable b-layer	8.5 μm	47μm
Pixel detector	10 μm	115 μm
Semi-conductor tracker	17 μm	580 μm
Transition radiation tracker	130 μm	-

Table 3.2: Spatial resolution of each components of the inner detector [34].

area for each module is composed of silicon wafer readout pixels with a nominal size of 50×400 μm^2. The performance of the pixel detector is critical to the full realisation of the physics capabilities of ATLAS. In particular, this performance must be maintained to ensure good vertexing and flavour tagging.

With time, irreparable failures of modules appear in the pixel detector and in order to guarantee a robust and precise tracking, an additional fourth layer, referred to as the insertable B-layer [44], was added to the pixel detector between 2013 and 2014. The insertable B-layer is located between the beam-pipe and the pixel detector, closest to the interaction point and at a radial distance of 33mm from the center of the beam-pipe. The nominal size of the pixel modules is 50×250 μm. Due to its proximity to the interaction point and the smaller size of its pixel detector modules, the insertable B-layer significantly improves the impact parameter resolution along the $R-\phi$ plane for low p_{T} particles. The improvement benefits the secondary vertex reconstruction that is important for b- and c-hadron identification. This effect of the instertable B-layer on flavour tagging algorithms is shown in Fig. 3.5, where the rate of c-jets and light-jets mis-identified as b-jets is reduced by 60% and 300% respectively for jets with a momentum of $p_{\mathrm{T}} = 100$

GeV.



(a) c-jets rejection (b) light-jets rejection

Figure 3.5: Rate of c-jets and light-jets mis-identified as b-jets with and without the insertable B-layer [44].

Semi-conductor tracker The semi-conductor tracker is a silicon strip detector located after the pixel detector and consists of four cylindrical layers in the barrel region and nine disks in the end-cap region. The detector is composed of 4088 modules and each module consists of two strips with a mean pitch of 80 μm and tilted with respect to each other by 40 mrad, which allows to provide a two-dimensional measurement in the module plane. For charged particles originating in the beam-interaction region, the semi-conductor tracker measures four space-points. The semi-conductor tracker allows to establish a measurement of the z-coordinate along the strip length and improves spatial resolution in conjunction with the $R-\phi$ measurement.

Transition radiation tracker The transition radiation tracker is the outermost sub-detector of the ATLAS inner detector. Unlike the pixel and semi-conductor tracker, which are made of silicon, the transition radiation tracker is made of gas-filled proportional drift-tubes called straws. Inside the straw, there is a gold-tungsten anode wire of 30 μm of diameter. The configuration allows for tracking and particle identification using transition radiation. When a charged particle passes through the straw, it ionises the gas and the electrons drift towards the wire and the ions drift to the straw walls. When the electrons approach the wire, an avalanche of electrons is induced due to the large potential difference, amplifying the signal. Timing information is extracted from the straw hits, resulting in a

resolution of 130 μm, along the direction of the wire. Although this accuracy is lower than for the pixel detector and the semi-conductor tracker, the performance of the transition radiation tracker is compensated by the large number of hits measured. On average, a charged particle deposits 36 hits in this sub-detector, which allows to measure the curvature of charged particle tracks over a larger radial angle, therefore improving the p_{T} of particles.

3.1.2 Calorimeters

Calorimeters are used to measure the energy of particles, which is used in the reconstruction of both the electrons, photons and jets. Calorimeters are composed of layers of passive and active materials. The passive material interacts with the passing particles, which starts a showering process. The active material is used to measure the flux of particles produced in the shower, which is related to the energy of the particle that initiated the shower. The calorimeter used in the ATLAS experiment, shown in Fig 3.6, is composed of two detectors: an electromagnetic and a hadronic calorimeter. Both detectors are designed to measure the energy and stop the particles. The electromagnetic calorimeter is composed of lead and liquid argon and measures the energy of electrons and photons, while the hadronic calorimeter, which is composed of steel and scintillators, measures the energy of the hadrons that was not captured by the thiner but higher resolution electromagnetic calorimeter.

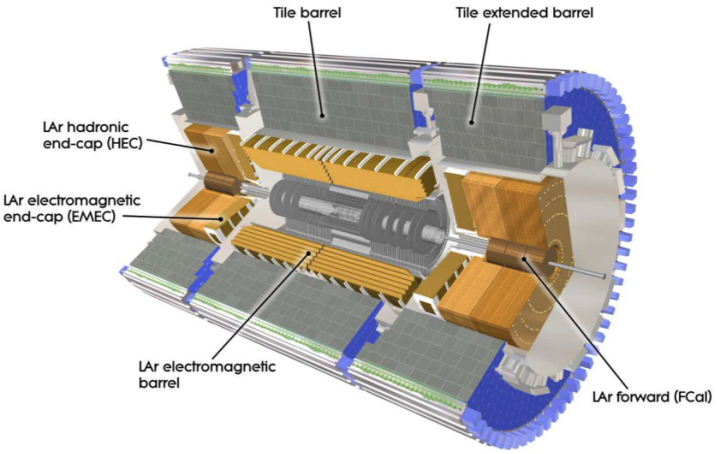


Figure 3.6: Cut-away view of the ATLAS calorimeters.

3.1.3 Muon detector

In the energy range of the LHC, most muons only lose relatively little energy via ionisation and are not stopped by the calorimeters. With a mean lifetime of $\tau \approx 2.2$ μs, muons leave the ATLAS detector before decaying. In order to measure the energy and tracks of the muons, a dedicated muon tracking system is located at the outermost layer of the ATLAS detector and is used to measure the tracks and momentum of muons. Most particles have decayed or have been stopped by the calorimeter before they reach this sub-detector. Muons are detected through the ionisation in the detector and their momentum is measured through the bending of the particles in the magnetic field produced by the toroid magnets. Muon momenta from 3 GeV to 3 TeV can be measured with the muon detector alone, while lower momentum tracks can be measured in conjunction with the inner detector information.

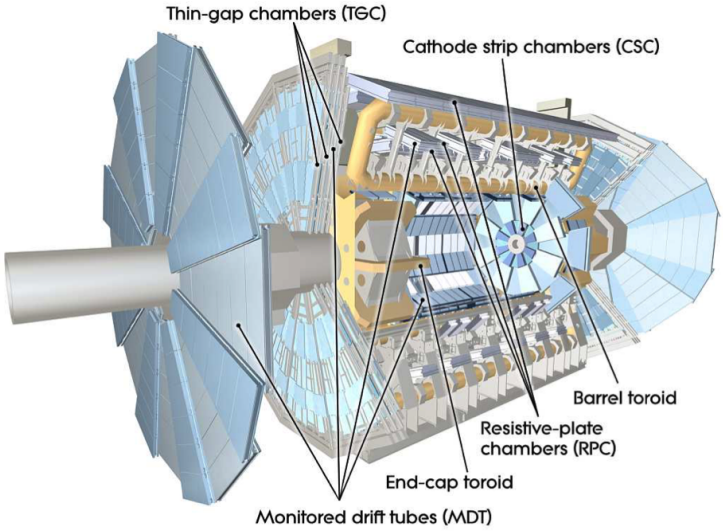


Figure 3.7: Cut-away view of the ATLAS muon detector.

The muon detector, shown in Fig. 3.7, is composed of four different sub-detectors, where two are used for track reconstruction and two for the trigger. The two sub-detectors dedicated to tracking consist of monitored drift tubes and cathode-strip chambers. As listed in Table 3.3, the monitored drift tubes have a better spatial resolution whereas the cathode-strip chambers have a better timing resolution.

The sub-detectors used to trigger on muons are the thin-gap chambers for the barrel and the resistive-plate chambers for the endcap. Similar to the track reconstructions sub-detectors, the thin-gap chambers have a better spatial resolution whereas the resistive-plates have a better time resolution. Compared to the tracking

sub-detectors, the sub-detectors used for triggering have a better timing resolution to allow fast triggering.

Sub-detector	Function	Resolution		
		z/R	ϕ	Time
Monitored drift tubes	Tracking	35 μm (z)	-	-
Cathode-strip chambers	Tracking	40 μm (R)	5 mm	7 ns
Resistive-plates chambers	Trigger	10 mm (z)	10 mm	1.5 ns
Thin-gap chambers	Trigger	2-6 mm (R)	3-7 mm	4 ns

Table 3.3: Spatial and time resolution for the sub-detectors of the muon detector [34].

3.2 Trigger and Data Acquisition

Inside the LHC, bunches of up to 10^{11} protons collide 40 million times per second to provided 13 TeV proton-proton collisions. For each event recorded by the ATLAS detector, tens of thousand of read-out channels are used, which amounts to about 1.5 MB of information to write to disk per event. Therefore, a 60 TB/s rate would be necessary to record all the events produced in the collisions, which is limited by the current capabilities of $\mathcal{O}(1)$ GB/s. In order to accommodate for the rate with which events can be written to disk, a trigger system is used to select only those events that contain interesting physics. A two-level trigger system is used, where the second level refines the decisions from the previous level by applying more sophisticated selection criteria.

The first level of the trigger is implemented with dedicated hardware, and is designed to select events based on the calorimeters and the muon detectors. This trigger level is capable of selecting events with electrons, muons, τ-jets, photons and missing transverse energy, and is designed to reduce the rate of events recorded to 75 thousand events per second. Events passing the first level trigger are temporarily stored in order to allow the second level trigger to reconstruct the events, and decide if it should be written to disk, based on properties of the event defined as a region of interest. The second level trigger, also known as the Event Filter, performs a full reconstruction of all events passed by the level-1 trigger using a large computing cluster close to the detector. Ultimately, 1000 events per second are written to disk and the full trigger system and data acquisition takes approximately 4 seconds, for the events that were finally selected.

4 Object reconstruction

The ATLAS detector consists of millions of modules that measure the hits and energy of particles. In order to analyse the physics properties of the Higgs bosons produced in the collisions, the information of the modules is combined to reconstruct the physics properties of the particle that was detected. Physics objects relevant for the $H \to c\bar{c}$, listed in Table 4.1, all have a momentum and a direction in the detector that needs to be reconstructed. The information from the detector is processed in two steps. First, hits in the tracker are combined to reconstruct the tracks and vertices of the particles. Secondly, the tracks, vertices and the clusters in the calorimeter are combined to reconstruct the particles trajectories and their kinematic properties.

Particle	ID	ECal	HCal	MS
Electron	×	×		
Muons	×			×
Jets initiated by quarks	×	×	×	
Missing energy	×	×	×	×

Table 4.1: List of particles relevant to the search for $H \to c\bar{c}$ reconstructed as physics objects from the different sub-detectors: inner detector (ID), electromagnetic calorimeter (ECal), hadronic calorimeter (HCal) and muon spectrometer (MS).

4.1 Tracks and vertices

The reconstruction of the trajectories of particles in the detector is fundamental for the particle identification. Track reconstruction algorithms consist of two stages: the process of finding track candidates, *the pattern recognition*, and the estimation of parameters to describe the trajectory, *the track fitting*. A charged particle deposits energy through ionisation in detector modules, which are read out as hits in the detector by the electronics. Pattern recognition algorithms aim at identifying the set of hits originating from a single particle. The challenge in the reconstruction is that other charged particles as well as detector noise can fake hits and therefore decrease the accuracy of the reconstruction. The identified hit

collection is fitted to determine the parameters that describe the trajectory of the particle.

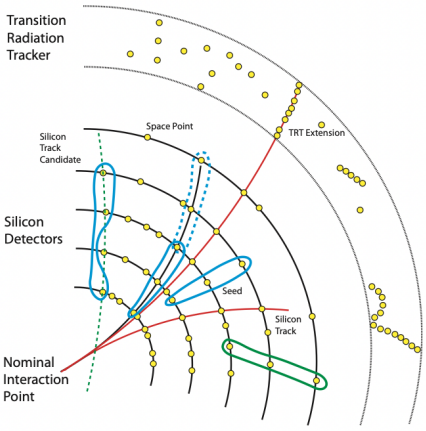
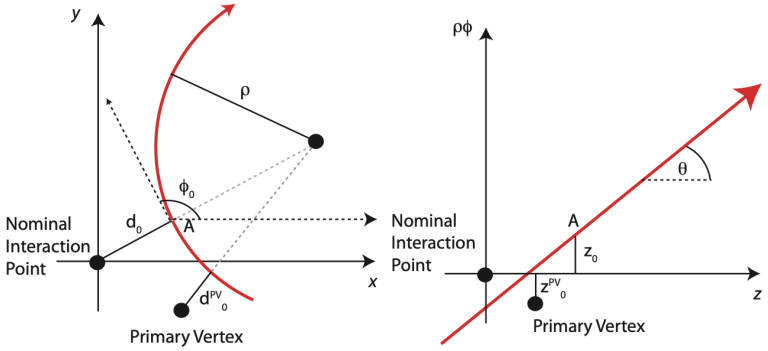


Figure 4.1: Illustration of the pattern recognition applied to tracks [45]. The hits in the detector are represented in yellow. The seeds reconstructed from combinations of three hits are shown in blue. The dashed blue line illustrates a case where two seeds correspond to the trajectory of the same particle. The green seed is rejected due to the inconsistency with the interaction point. Fully reconstructed tracks with and without the transition radiation tracker are shown in red. The greed dashed line represents a track fitted but inconsistent with the interaction point.

The primary sequence of the track reconstruction algorithm used in ATLAS [46, 47, 48] reconstructs tracks from the centre of the detector outwards, and is referred to as the inside-out sequence and illustrated in Fig. 4.1. The algorithm begins with finding seeds from groups of three hits in the silicon layers of the detector. The seeds are then used to build the trajectory to find hits while moving towards the outer edge of the detector. Once all the hits in the silicon trackers are found, hits in the transition radiation tracker are probed and matching hits are added to the silicon track. Finally, a fit is performed to estimate the parameters of the tracks. Once a track is found, the procedure is applied again on the remaining hits to find the other tracks in the event. Hits are sometimes attached to multiple tracks. In order to solve this ambiguity, the quality of the fitted track as well as their direction are used to identify a collection of hits that most likely correspond to a physical charged particle signature.

The trajectory of a charge particle travelling in a uniform magnetic field is parametrised by a helix, which depends on 5 parameters, also shown in Fig 4.2:

$$\tau = (d_0, z_0, \phi_0, \cot\theta, q/p_{\mathrm{T}}), \tag{4.1}$$



(a) $x - y$ plane (b) z-axis

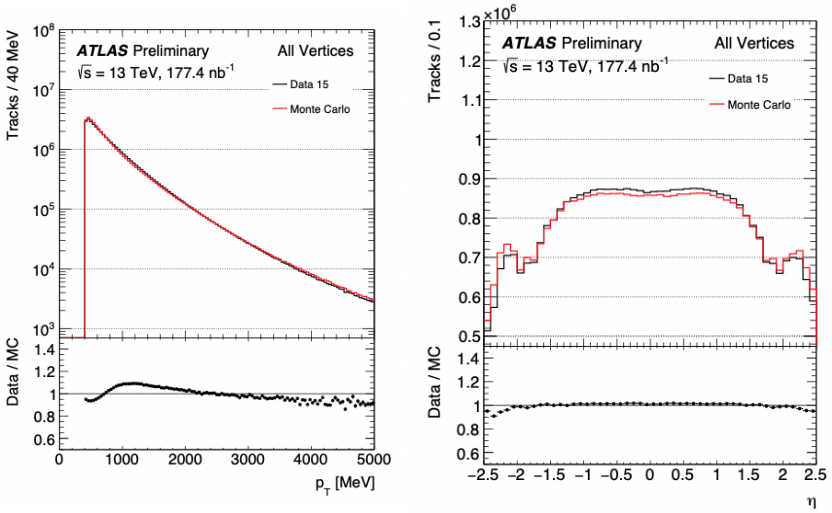
Figure 4.2: Illustration of the 5 parameters of the track of a particle [45].

where d_0 is the closest distance on the trajectory to the interaction point in the transverse plane, z_0 is the closest distance to the interaction point on the z-axis, known as the longitudinal impact parameter, ϕ_0 and θ are the transverse and polar angles and q/p_{T} is the curvature of the track which depends on the inverse transverse momentum with the sign defined by the electric charge q. A comparison of the data and simulation is shown in Fig. 4.3 for the p_{T} and η of tracks. The data is underestimated by the simulation by about 10% is observed for tracks with $p_{\mathrm{T}} = 1500$ MeV and a selection of $p_T > 500$ MeV is applied on the tracks to remove the effect of tracks reconstructed from combinatorial hits in the tracker. The data and the simulation are generally found in agreement for the track η in the barrel region and an overestimation of about 10% is found for tracks reconstructed in the region $|\eta| > 1.5$. This difference in the data and the simulation is attributed to the simulation of these observables, which is challenging in an environment with high pile-up.

Primary vertex reconstruction

Due to the fact that bunch crossing generate multiple proton-proton collisions, there are multiple vertices in each recorded event. In order to ensure that physics objects in an event originate from the same hard process, the vertex at which the Higgs boson is produced, also known as the primary vertex, needs to be reconstructed.

The vertex reconstruction [50, 51] is done iteratively starting from tracks that fulfill quality criteria on the momentum $p_{\mathrm{T}} > 400$ MeV and are compatible with the interaction point ($d_0 < 4$ mm, $\sigma_{d_0} < 5$ mm and $\sigma_{z_0} < 10$ mm). All the tracks are then fitted together in order to optimally find the vertex location and to quantify the compatibility of each track with the vertex that is calculated. The tracks incompatible with the best fitted vertex are discarded and are used for the reconstruction of other vertices in the event. The reconstruction is performed



(a) Track p_{T}

(b) Track η

Figure 4.3: Track p_{T} and η in a high pile-up environment [49]. Monte-Carlo is normalised to the data.

iteratively for all the remaining tracks until all the vertices are identified. As the hard scattering is expected to involve particles with higher energies than the ones from the underlying event and the pileup interactions, the primary vertex is identified as the vertex with the highest energy transfer, corresponding to the scalar sum of the absolute value of the momentum $\sum_i |p_{\mathrm{T},i}|^2$.

In Fig. 4.4, the data and the simulation is compared for the vertex reconstruction efficiency as a function the number of tracks. A lower efficiency is observed for vertices consisting of 2 tracks and the reconstruction efficiency gradually increases for vertices with more tracks.

4.2 Leptons

The ability to reconstruct leptons is important in many analyses performed by the ATLAS collaboration. Out of the three types of charged leptons, only the electron and the muons are stable enough to be measured and reconstructed directly. Due to its lifetime, the τ-lepton decays in the detector, either leptonically or fully hadronically, and is reconstructed indirectly based on its stable decay products measured in the event.

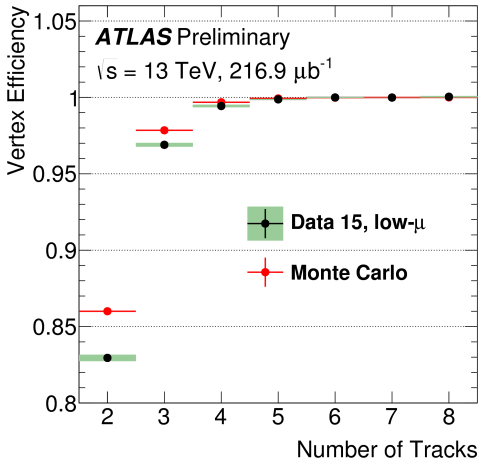


Figure 4.4: Vertex reconstruction efficiency as a function of the number of tracks [49].

4.2.1 Electrons

Electrons are reconstructed from energy deposits in the electromagnetic calorimeter. The reconstruction of electrons, shown in Fig. 4.5, requires at least one track associated to a cluster of energy. If more than one track is associated with the electron candidate, the track closest to the center of the energy cluster is selected.

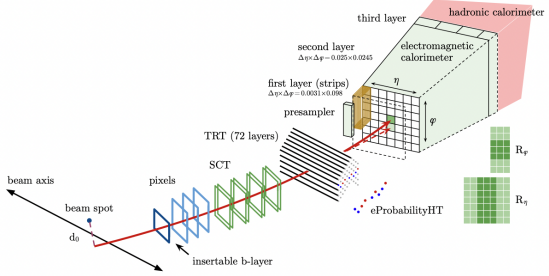


Figure 4.5: Schematic view of the electron reconstruction and identification [52].

The performance of the electron reconstruction is $> 95\%$ for a momentum $p_{\mathrm{T}} > 10$ GeV. The energy resolution as function of the simulated energy, shown in Fig. 4.6, improves for higher energies and is $\mathcal{O}(1\%)$ for electrons with 100 GeV. This resolution of the electrons energy is particularly useful to reconstruct resonant decays such as $Z \to e^+e^-$.

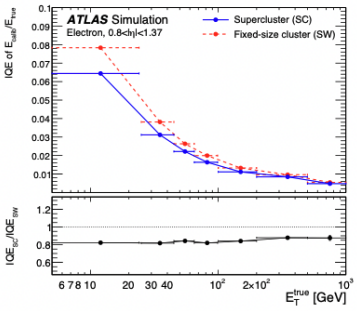
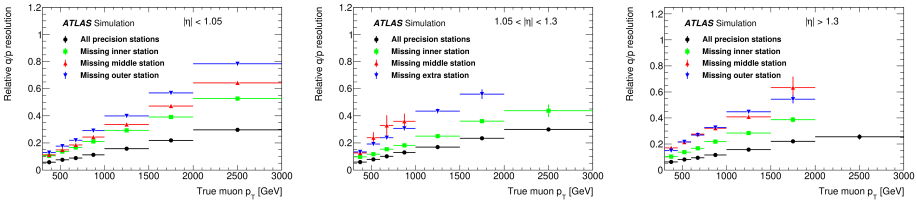


Figure 4.6: Electron energy resolution [53]. The supercluster is a variable size cluster in the electromagnetic calorimeter, as opposed to a fixed size cluster.

4.2.2 Muons

Muons are reconstructed based on the information provided by the inner tracking and the muon detector. The first step is to reconstruct the tracks independently in both detectors and then combine the information of both detectors to perform a full reconstruction. The majority of muons are reconstructed using an outside-in algorithm, where the tracks are extrapolated from the muon detector to the inner detector.



(a) Barrel (b) Barrel / endcap (c) Endcap

Figure 4.7: Resolution on the q/p of the tracks of the muons as a function of the muon p_{T} [54]. The precision stations are defined as the modules of the monitored drift tubes or cathode-stip chambers. Missing hits in the muon detector degrade the resolution q/p, missing an outer station is worse for the resolution than missing the inner station.

The resolution on the inverse momentum q/p, i.e the track curvature shown in Fig. 4.7, is of the order of $\mathcal{O}(5\%)$ for muons with $p_{\mathrm{T}} = 300$ GeV and degrades for higher momenta. This momentum resolution for muons is worse than the energy resolution for electrons at 300 GeV, due to the better performance of the calorimeter at high energies. The momentum scale calibration of the muon reconstruction is

performed on samples of $Z \to \mu^+\mu^-$ and $J/\psi \to \mu^+\mu^-$ events.

4.3 Jets

Quarks are reconstructed as collimated sprays of particles known as jets and energy deposits from the calorimeters as well as tracks from the inner detector are needed to reconstruct their signatures. The energy and position of each jet is determined from the signature in the calorimeters, while the flavour of the quark from which the jets originate is determined from the tracks reconstructed by the inner detector.

Jets are observed as topologically connected energy clusters in the calorimeters. The most common algorithm used in ATLAS is the anti-k_{T} [55, 56] algorithm, which is a recombination algorithm that groups energy clusters based on the distance between them d_{ij} and the transverse energy k_t:

$$d_{ij} = min(\frac{1}{k_{t,i}^2}, \frac{1}{k_{t,j}^2})\frac{\Delta R_{ij}^2}{R^2}, \tag{4.2}$$

where ΔR_{ij} is the angular separation between the two clusters and $R = 0.4$ is the cone size of the jet. For all available clusters, the distance d_{ij} is calculated and the closest clusters are combined. The procedure is then repeated for every cluster until the distance between two clusters is found to be larger than the distance to the beam line, at which point the procedure is stopped. The resulting object is referred to as a jet with a cone of radius $R = 0.4$. The energy resolution of the jets, shown in Fig. 4.8, improves for higher momenta and is of the order of $\mathcal{O}(10)\%$ for jets with a momentum $p_{\mathrm{T}} = 100$ GeV. In the context of the $H \to c\bar{c}$ measurement, jets are reconstructed based on this algorithm using the electromagnetic topological clusters.

4.3.1 Secondary vertices

Heavy hadrons originating from b- and c-quarks are produced at the primary vertex and decay in the detector at a secondary vertex. In order to distinguish jets originating from different heavy flavour quarks, the mass and lifetime properties of the hadrons are exploited by reconstructing secondary decay vertices.

Reconstruction algorithms consider all possible track pairs to construct 2-track vertices. Only those vertices that are consistent with a heavy hadron decay are retained. Vertices associated with photon conversion and material interactions are explicitly rejected. The remaining vertices are merged into a multi-track vertex if they are close in space.

In ATLAS, two algorithms are used to reconstruct secondary vertices: the SV algorithm [58], which targets the reconstruction of a single secondary vertex in the jet, and the JetFitter algorithm [59], which aims at reconstructing the decay chain of heavy hadrons with multiple vertices that are aligned. The performance of the

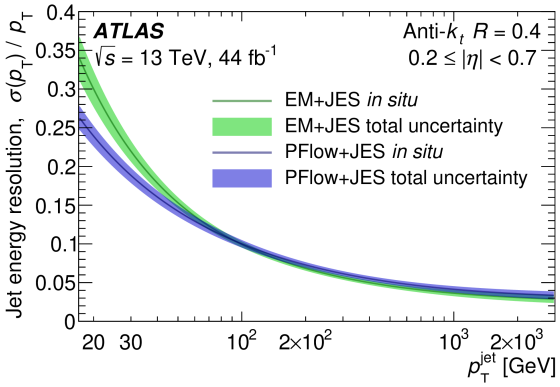


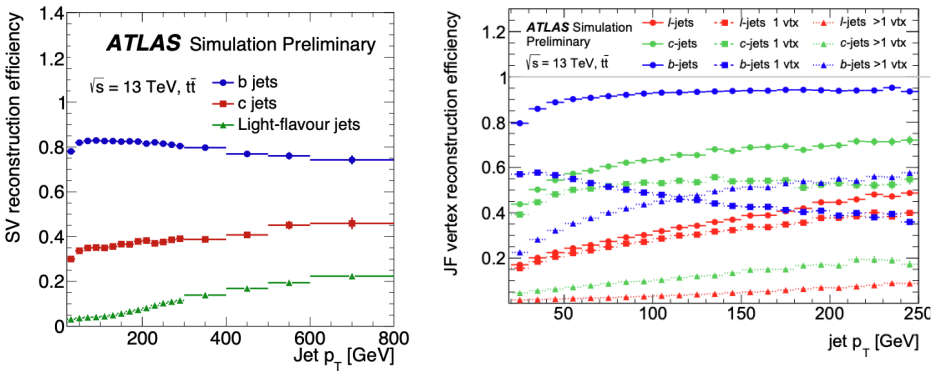
Figure 4.8: Jet energy resolution [57] as a function of p_{T}. The EM jets (green) are the jets used in the analysis presented in this thesis. The PFlow jets (blue) are a different algorithm that exploits the tracking information in addition to the calorimeter to reconstruct the energy of the jet.

two algorithms, shown in Fig. 4.9, is evaluated in terms of reconstruction efficiency. Due to the longer lifetime of b-hadrons, the reconstruction efficiency is highest for b-jets at about $80-90\%$. The efficiency for c-jets is about $30-40\%$ and increases as a function of the momentum. Light-flavour jets do not have a secondary vertex in the decay chain but can contain a fake vertex, which is reconstructed for $10-30\%$ of these jets and, due to the higher track density, increases at high momentum. The presence of one or multiple vertices within a jet, as well as the kinematic properties (mass, displacement, ...) of the vertex, are crucial tools to identify the flavour of the quark that initiated the jet.

4.4 Flavour tagging

Quark signatures in particle detectors are challenging to reconstruct. Unlike leptons, quarks are confined inside hadrons and are not observed as separate particles. Therefore, quarks are measured as a collimated spray of hadrons referred to as a jet. The identification of the hadron flavour within a jet is known as flavour tagging. Flavour tagging is a crucial component for any measurement involving quark signatures in particle detectors. In particular, the sensitivity to the $H \to c\bar{c}$ process is heavily reliant on the performance of charm hadron identification.

Due to their mass, quarks of different flavours have different signatures in the detector: the bottom and the charm quark hadronise respectively into a bottom-hadron (b-hadron) and a charm-hadron (c-hadron). The up, down and strange quarks hadronise into light hadrons which do not decay in the detector in most cases. The top quark decay is uniquely shorter than the time scale at which the



(a) SV (b) JetFitter

Figure 4.9: Secondary vertex reconstruction efficiency for two standard algorithms used in ATLAS: SV [58] and JetFitter [59].

strong force acts and decays via the weak force before it can hadronise. The τ-lepton, while not a quark, can also decay fully hadronically and its signature is reconstructed as a jet, and thus consitutes an important background in flavour tagging studies.

Species	Fragmentation $f(q \to X_q)$ [%]	Mass [MeV]	Lifetime [s]
B^+	40.4 ± 0.6	5279.29 ± 0.15	$(1.638 \pm 0.004) \times 10^{-12}$
B^0	40.4 ± 0.6	5279.61 ± 0.16	$(1.520 \pm 0.004) \times 10^{-12}$
B_s^0	10.3 ± 0.5	5366.79 ± 0.23	$(1.510 \pm 0.004) \times 10^{-12}$
Baryons	8.8 ± 1.2	$\mathcal{O}(5600)$	$\mathcal{O}(10^{-12})$
D^+	26.39 ± 1.39	1869.62 ± 0.20	$(1.040 \pm 0.007) \times 10^{-12}$
D^0	57.72 ± 2.41	1864.84 ± 0.17	$(4.101 \pm 0.015) \times 10^{-13}$
D_s^+	6.91 ± 0.45	1968.49 ± 0.34	$(5.00 \pm 0.07) \times 10^{-13}$
Λ_c^+	5.26 ± 0.31	2286.46 ± 0.14	$(2.00 \pm 0.06) \times 10^{-13}$

Table 4.2: Fragmentation, mass and lifetime of most frequent b- and c-hadrons produced in the hadronisation of bottom and charm quarks. Hadrons initiated by b-quarks are presented in the upper table, while the lower table represent hadrons initiated by c-quarks [60].

The fragmentation, mass and lifetime of the most frequent b- and c-hadrons produced in bottom and charm quark hadronisations are presented in Table 4.2. Due to the Cabbibo-Kobayashi-Maskawa suppresion, b-hadrons have a longer lifetime than c-hadrons, resulting in a flight length of b-hadrons in the detector that is longer than that of c-hadrons. The b-hadrons are also heavier than c-hadrons

due to the mass difference between bottom and charm quarks. The lifetime of light hadrons, typically of the order 10^{-8}s, is significantly longer than the b- and c-hadrons lifetime, these hadrons do therefore not decay in the detector. The τ-lepton, conversely, has properties similar to c-hadrons: its lifetime is $(2.903 \pm 0.005) \times 10^{-13}$s and its mass is 1776 ± 0.16 MeV.

In order to distinguish the flavour of the quark or lepton that initiated a jet, the lifetime and the mass are reconstructed from the final decay products. Quarks produced in the pp-collisions hadronise near the interaction point at a primary vertex (PV). Due to the energy of the collision, hadrons from heavy quarks are relativistic and travel in the detector before decaying at a secondary vertex (SV) into a cascade, resulting in stable final states particles. This process is illustrated in Fig. 4.10. The displacement between the PV and SV corresponds to the flight length of the hadron and depends on its lifetime. The sum of the energy of the decay products at the SV depends on the mass of the hadron.

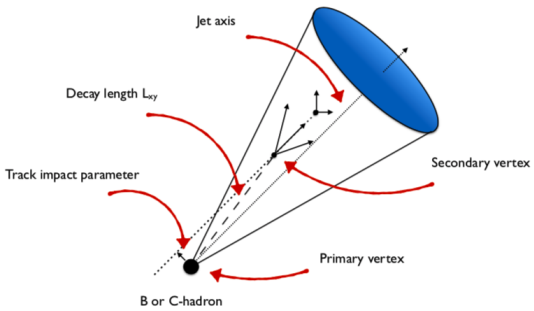


Figure 4.10: Sketch of a jet substructure initiated by a b- or c-hadron. The hadron is produced at a primary vertex and decays at a secondary vertex.

A diagram representing the various signatures of bottom, charm, light quarks and τ-leptons is shown in Fig. 4.11. The identification of b-hadrons within jets, also referred to as b-tagging, is possible through their longer lifetime and heavy mass. In the case of the c-hadrons, the physics properties are between the properties of b-hadrons and light hadrons. In addition, the hadronic decay of τ-leptons is similar to the signature of c-hadrons. In the simulation, the flavour of the jet is used as an input and jets are categorised depending on the true flavour: c-jets, b-jets, light-jets and τ-jets.

The b-tagging algorithms are a necessity for the physics program of the ATLAS collaboration. In the year of 2020, 40% of all ATLAS publications have used b-tagging. Consequently, this technology is continuously developed and studied [61, 62]. On the contrary, only 2 publications from the ATLAS collaboration have used c-tagging algorithms since the start of the experiment: the previous search of the $H \to c\bar{c}$ process [4] and the search for pair-produced third-generation squarks

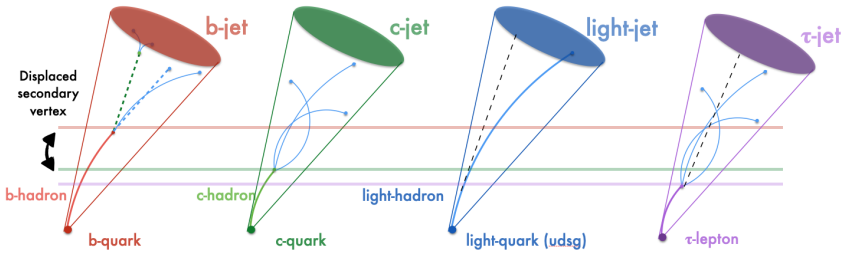


Figure 4.11: Diagram representing the various scenarios for bottom, charm and light-quark hadronisations as well as the τ-lepton and their decay products within jets.

decaying via charm quarks [63].

For this thesis, a new c-tagging approach is adopted, based on the latest improvements of b-tagging and is optimised for the sensitivity to the $H \to c\bar{c}$ search.

4.4.1 Charm tagging algorithm

All c-hadrons decay through the weak interaction, which results in a large amount of possible final states. The most important one of these, the inclusive decay modes of the D^+ and D^0 mesons, are presented in the Table 4.3. Generally, the most frequent decay modes happen through the $c \to s$ transition and include a $K^\pm$ or a K^0 meson. In order to identify c-jets, c-tagging algorithms focus on the common properties of the various decay modes which depend on both the mass and the lifetime of the decaying quark. These are: the existence of a SV, the invariant mass of the SV, the displacement of the SV with respect to the PV, the number of tracks and their impact parameter. In order to maximise the c-jets identification with respect to non c-jets backgrounds, a multivariate approach (MVA) is applied.

The MVA used for c-tagging in the ATLAS experiment utilises a deep neural network which takes as input the η and p_T of the jet as well as the output of dedicated heavy quark identification algorithms. These algorithms target different properties of b- and c-hadrons such as the impact parameter of the tracks (IP2D, IP3D) [64], the reconstructed SV (SV1) [58] and the decay chain of the b- and c-hadrons (JetFitter, JF) [59]. The complete list of input variables used by the MVA algorithm for c-tagging is presented in Table 4.4.

The neural network produces 3 output variables, corresponding to the probability of being a b-jet $p(b)$, c-jet $p(c)$ and a light-jet $p(l)$. In order to simplify the usage, these probabilities are combined into a one dimensional discriminant:

$$DL1_{c-\text{tagging}} = \frac{p(c)}{fp(b) + (1-f)p(l)}, \tag{4.3}$$

Species	Inclusive decay mode	Fraction [%]
D^+	$\rightarrow e^+$ semileptonic	(16.07 ± 0.30)
	$\rightarrow \mu^+$ anything	(17.6 ± 3.2)
	$\rightarrow K^-$ anything	(25.7 ± 1.4)
	$\rightarrow \bar{K^0}$ or K^0 anything	(61 ± 5)
	$\rightarrow \bar{K^*}(892)^0$ anything	(23 ± 5)
D^0	$\rightarrow e^+$ semileptonic	(6.49 ± 0.11)
	$\rightarrow \mu^+$ anything	(6.7 ± 0.6)
	$\rightarrow K^-$ anything	(54.7 ± 2.8)
	$\rightarrow \bar{K^0}$ or K^0 anything	(47 ± 4)
	$\rightarrow K^*(892)^-$ anything	(15 ± 9)

Table 4.3: Main inclusive decay modes for the two largest fragmentation modes of the c-quark hadronisation: D^+ and D^0 [60].

where the fraction f is a relative weight tuning the b-jet ($f = 1$) or light-jet ($f = 0$) rejection. It is important to note that the τ-jets are not considered in the output of the neural network. The training of the c-tagging algorithm is based on the b-tagging algorithm used by the ATLAS collaboration [61], using the same input variables as well as the same training set for both c-tagging and b-tagging.

Both the c- and b-tagging identification relies on the same set of physics properties of heavy hadrons. The result is that both algorithms are correlated and, without further treatment, a jet can be both b-tagged and c-tagged. In order to simplify the c-tagging usage and to address the overlap between c-tagging and b-tagging, c-tagging is defined with an additional b-tag veto. This veto allows the simultaneous usage of b- and c-tagging and simplifies the treatment of systematic uncertainties between the two orthogonal taggers. The b-tag veto is defined with a standard b-tagger used in the ATLAS collaboration known as $MV2c10$ [62]. The c-tagging procedure including a b-tagging veto is illustrated in Fig. 4.12.

The distribution of b- and c-tagging weights for the various jet flavours in simulated $t\bar{t}$ events are shown in Fig. 4.13. In the case of the b-tagger $MV2c10$, the region corresponding to high b-tagging scores is dominated by b-jets. The vertical line represents the b-tagging operating point: a jet is defined as "b-tag vetoed" if the b-tagging weight is below the operating point value. In the case of the c-tagger $DL1$, the region corresponding to high c-tagging scores is dominated by c-jets and τ-jets. For low values, b- and light-jets dominate the spectrum. A jet is c-tagged if it is b-tag vetoed and if its c-tagging weight is above the operating point represented by the vertical line in the c-tagging discriminant.

The performance is evaluated in terms of the tagging *efficiency* ϵ, which corresponds to the probability that a jet is tagged. The parameters ϵ_c,ϵ_b, ϵ_l and ϵ_τ quantify the performance of the flavour tagging algorithms and are shown in Table 4.5 for the c- and b-tagger. In the case of the c-tagger, the probability that

Input	Variable	Description
Kinematics	p_T	Jet p_T
	η	Jet $\lvert\eta\rvert$
IPD2D/IP3D	$\log(p_b/p_c)$	Likelihood ratio between the b-jet and light-jet hypotheses
	$\log(p_b/p_{\text{light}})$	Likelihood ratio between the b-jet and light-jet hypotheses
	$\log(p_c/p_{\text{light}})$	Likelihood ratio between the c-jet and light-jet hypotheses
SV1	$m(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion masses
	$f_E(\text{SV})$	Energy fraction of the tracks associated with the secondary vertex
	$N_{\text{TrkAtVtx}}(\text{SV})$	Number of tracks used in the secondary vertex
	$N_{\text{2TrkVtx}}(\text{SV})$	Number of two-track vertex candidate
	$L_{xy}(\text{SV})$	Transverse distance between the primary and secondary vertex
	$L_{xyz}(\text{SV})$	Distance between the primary and secondary vertex
	$S_{xyz}(\text{SV})$	Distance between the primary and secondary vertex divided by its uncertainty
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex
JetFitter	$m(\text{JF})$	Invariant mass of tracks from displaced vertices
	$f_E(\text{JF})$	Energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{JF})$	ΔR between the jet axis and vectorial sum of momenta of all tracks attached to displaced vertices
	$S_{xyz}(\text{JF})$	Distance between the primary and secondary vertex divided by its uncertainty
	$N_{\text{TrkAtVtx}}(\text{JF})$	Number of tracks of multi-prong displaced vertices
	$N_{\text{2TrkVtx}}(\text{JF})$	Number of two-track vertex candidates
	$N_{1-\text{trkvertices}}(\text{JF})$	Number of single prong displaced vertices
	$N_{\geq 2-\text{trkvertices}}(\text{JF})$	Number of multi-prong displaced vertices
	$L_{xy}(2^{\text{nd}}/3^{\text{rd}/})(\text{JF})$	Transverse distance between the 2^{nd} or 3^{rd} vertex from primary vertex
	$L_{xyz}(2^{\text{nd}}/3^{\text{rd}/})(\text{JF})$	Distance between the 2^{nd} or 3^{rd} vertex from primary vertex
	$m_{\text{Trk}}(2^{\text{nd}}/3^{\text{rd}/})(\text{JF})$	Invariant mass of the track associated with the 2^{nd} or 3^{rd} vertex
	$E_{\text{Trk}}(2^{\text{nd}}/3^{\text{rd}/})(\text{JF})$	Energy fraction of the track associated with the 2^{nd} or 3^{rd} vertex
	$f_E(2^{\text{nd}}/3^{\text{rd}/})(\text{JF})$	Fraction of charged jet energy the track associated with the 2^{nd} or 3^{rd} vertex
	$N_{\text{TrkAtVtx}}(2^{\text{nd}}/3^{\text{rd}/})(\text{JF})$	Number of tracks associated with the 2^{nd} or 3^{rd} vertex
	$Y_{\text{trk}}^{\text{min}}, Y_{\text{trk}}^{\text{max}}, Y_{\text{trk}}^{\text{avg}}(2^{\text{nd}}/3^{\text{rd}/})(\text{JF})$	Minimum, maximum and average track rapidity of track at the 2^{nd} or 3^{rd} vertex

Table 4.4: Input variables to the MVA algorithm used for c-tagging.

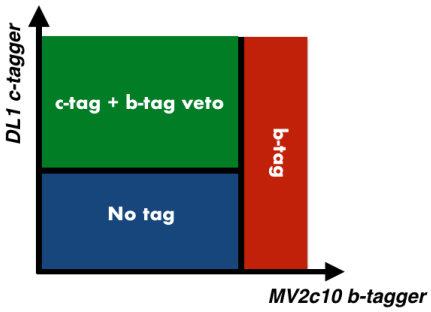


Figure 4.12: Flavour tagging paradigm use to c-tag and b-tag jets. For both taggers, a low score means a low probability the jet as the target particle.

a c-jet is c-tagged is 22%, which is significantly higher than the mis-tagging rate of b-jets and light-jets. Due to the similar mass and lifetime between the τ-lepton and c-hadrons, τ-jets have a high probability of being identified as c-jets. The τ-jets mis-tagging rate is therefore similar to the tagging efficiency of c-jets. The effect of the b-tag veto is to remove 70% of b-jets and 12% c-jets.

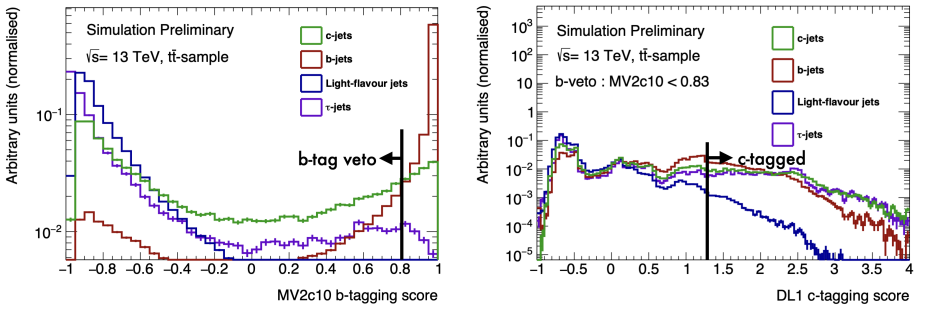
Efficiencies	b-tagger	c-tagger including b-tag veto
ϵ_c	12%	27%
ϵ_b	70%	8%
ϵ_l	0.3%	1.6%
ϵ_τ	2.5%	20%

Table 4.5: Efficiencies of the c-tagging and b-tagging working points. The performance is estimated on a 139fb^{-1} of simulated $t\bar{t}$ events generated with Powheg+Pythia8.

4.4.2 Calibration

Due to the imperfect description of the detector response and physics modelling in simulations, flavour tagging algorithms require a calibration with collision data. To obtain this calibration, the performance is evaluated in simulation and measured in the collision data by dedicated analyses that use data samples enriched in either c-, b- or light-jets. The resulting efficiencies are used to define scale factors that adjust the flavour tagging performance of the simulations to match that of the data.

The calibration is derived assuming a binary per-jet tagging output as depicted



(a) $MV2c10$ b-tagging score (b) $DL1$ c-tagging score

Figure 4.13: Normalised distributions of the c- and b-tagging outputs as a function of the tagging score: $DL1$ c-tagger (a) and $MV2c10$ b-tagger (b). The distributions are shown for c-jets, b-jets, light-jets and τ-jets.

in Fig. 4.14: *c-tagged* or *not tagged*. The corrections are derived on a $t\bar{t}$ sample simulated with POWHEG+PYTHIA8. Two scale factors are calculated to correct the tagging efficiency $SF_{\text{efficiency}}$ and inefficiency $SF_{\text{inefficiency}}$ in the case where the jet fails the tagging requirement, respectively defined as:

$$SF_{\text{efficiency,PwPy8}} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC,PwPy8}}}, \tag{4.4}$$

$$SF_{\text{inefficiency,PwPy8}} = \frac{1-\epsilon_{\text{data}}}{1-\epsilon_{\text{MC,PwPy8}}}, \tag{4.5}$$

where ϵ_{data} is the tagging efficiency on data and $\epsilon_{\text{MC,PwPy8}}$ the efficiency on the simulation sample.

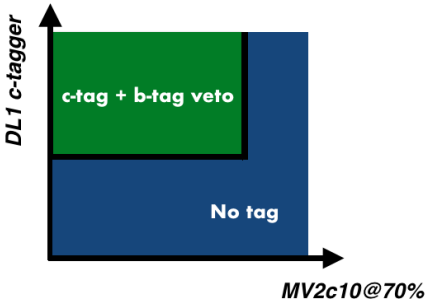
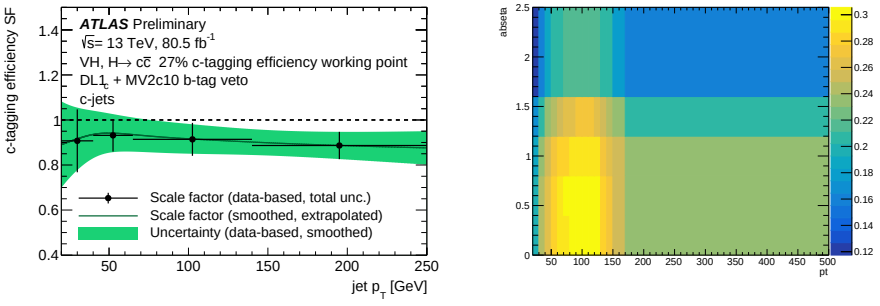


Figure 4.14: Sketch of the binary c-tagger. The operating point of the taggers that define the region are calibrated using collision data.

The calibrations are derived for each jet flavour with a fit to the data. For each calibration, a smoothing [65] is applied on the binned calibrations to reduce the statistical noise. The details of each calibration, as well a the correction procedure applied to account for the different hadronisation models used in the simulation are described in the following paragraphs.

c-jets calibrations For c-jets, the c-tagging efficiency is measured in single-lepton $t\bar{t}$ events [66]. In this method, a kinematic fit is applied to events with a single lepton and exactly four jets in order to select a pure sample of $t\bar{t}$ events in which the jets are produced in the hadronic decay of the W boson. About 25% of these jets originating from the W decay are expected to be c-jets. The c-tagging efficiency for c-jets in data is extracted with a likelihood fit which uses the number of candidate hadronic W decay in two jets. Efficiency scale factors for c-jets are then derived to correct the simulation as shown in Figure 4.15. The c-tagging uncertainty on c-jets efficiency scale factors is typically $\pm 10\%$ for $p_{\mathrm{T}}(jet) < 250$ GeV.



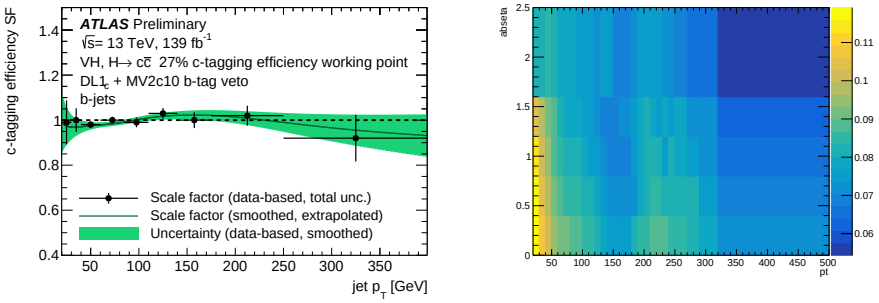
(a) Scale factor

(b) Simulation tagging efficiency

Figure 4.15: Flavour tagging calibration for c-jets: (a) efficiency scale factors (black points) and smoothed (green), (b) efficiency maps as a function of p_T and $|\eta|$ evaluated on a simulated $t\bar{t}$ sample.

b-jets calibrations For b-jets, a data sample of $t\bar{t}$ events in which both W bosons decay leptonically is used to calibrate the c-tagging efficiency [67]. The event selection used, based on a BDT trained on kinematic variables, leads to a sample very pure in b-jets. Probability density functions (PDFs) of the tagging discriminant are built in several regions in the 2D space of p_{T} of the leading and sub-leading jet, where the bin boundaries are defined as shown in Figure 4.16. These PDFs are used in a likelihood fit to derive the b-jet tagging efficiency scale factor from the data sample. The corresponding efficiency maps, binned as a function of (p_T,$|\eta|$), are evaluated on a simulated $t\bar{t}$ sample. The hadronisation model is simulated

with PYTHIA8. The c-tagging uncertainty on the efficiency scale factors for b-jets is typically a few percent for $p_{\mathrm{T}}(jet) < 250$ GeV.



(a) Scale factor

(b) Simulation tagging efficiency

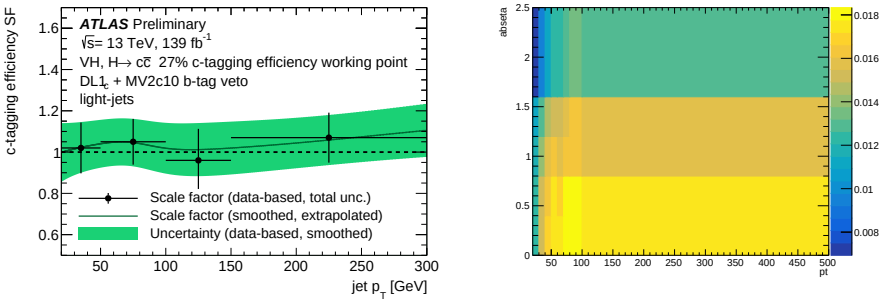
Figure 4.16: Flavour tagging calibration for b-jets: (a) efficiency scale factors (black points) and smoothed (green), (b) efficiency maps as a function of p_T and $|\eta|$ evaluated on a simulated $t\bar{t}$ sample.

Light-jets calibrations The c-tagging efficiency of light-jets is measured in data using a sample enriched in Z+jets events. The efficiency measurement is based on the "negative tag" method, which exploits an alternate control tagger, which is built on track and vertex based inputs where the signed quantities (e.g. track d_0 signed with respect to the jet axis) are "flipped" (i.e. the sign is inverted). This "flipped" version of the c-tagging algorithm exhibits a light flavour efficiency which is close to that of the nominal tagger, but with a substantially lower efficiency for heavy flavour jets, and thus allows to measure the light jet efficiency without large backgrounds. Further details on the negative tag calibration method are given in Ref. [68].

The measured c-mistagging efficiency scale factors for light-jets are shown in Fig. 4.17. The uncertainty is typically $\pm 15\%$ for $p_{\mathrm{T}}(jet) < 250$ GeV.

τ-jets calibrations The τ-jets calibrations are not derived from a fit to the collision data. Due to the similar physics properties between τ-jets and c-jets, the c-jets efficiency scale factors are used with an additional systematic uncertainty of 22% added to each p_{T} bin. This additional uncertainty has been derived in the context of the b-tagging mis-tagging rate on c-jets and τ-jets and the same treatment is adopted for c-tagging. The extrapolated efficiency scale factor is shown in Fig. 4.18.

MC/MC scale factors Flavour tagging algorithms and calibrations are derived on $t\bar{t}$ or Z+jets samples generated, where the hadronisation model is simulated with PYTHIA8.230. As the tagging algorithms depend on the lifetime of the



(a) Scale factor (b) Simulation tagging efficiency

Figure 4.17: Flavour tagging calibration for light-jets: (a) efficiency scale factors (black points) and smoothed (green), (b) efficiency maps as a function of p_T and $|\eta|$ evaluated on a simulated $t\bar{t}$ sample.

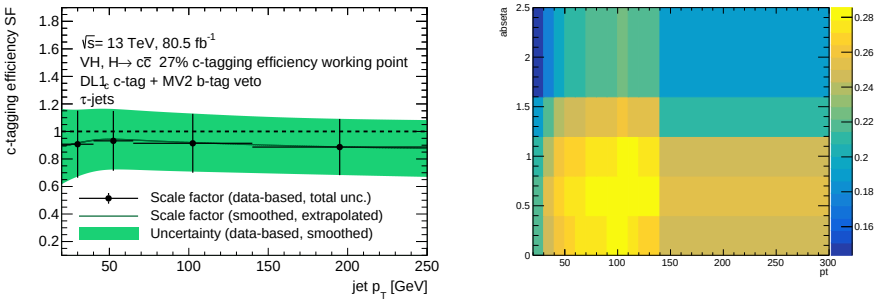
heavy hadrons, the modelling of the quark fragmentation to hadrons influences the tagging decision. Specifically, a different prediction of heavy quark abundances can directly impact the tagging efficiency. This is the case for the hadronisation model simulated with Sherpa2.2.1 and Herwig7.1.3, which is different than the one obtained with Pythia8.230. To ensure that c-tagging is calibrated for different Monte Carlo (MC) generators, a MC/MC correction is calculated on simulated $t\bar{t}$ events [69], which is defined as a ratio of c-tagging efficiencies:

$$SF_{\mathrm{MC/MC,Alt}} = \frac{\epsilon_{\mathrm{eff,PwPy8}}}{\epsilon_{\mathrm{eff,Alt}}}. \tag{4.6}$$

The result is presented in Fig. 4.19 for the two different alternative hadronisation simulations Sherpa2.2.1 and Herwig7.1.3. In addition, the uncertainty of the efficiency scale factors is presented as a reference. For b-jets, light-jets and τ-jets MC/MC scale factors, the differences are up to 10% for both simulations and these differences are within the stated uncertainty of the c-tagging efficiency calibrations.

In the case of c-jets simulated with Sherpa2.2.1, the correction is much larger than for the other scale factors, of order $\approx 30\%$, and also significantly larger than the stated uncertainty on the scale factor. This large correction is mostly due to the overestimation of the Λ_c abundance in Sherpa2.2.1 with respect to the Pythia8.230 prediction [70]. As the Λ_c lifetime is 5 to 10 times smaller than the typical D-meson lifetimes, this change in the relative abundances affects the average c-tagging performance significantly. The observed 30% difference between Sherpa2.2.1 and Pythia8.230 is corrected for in all Sherpa2.2.1 simulation samples.

The MC/MC scale factor is multiplied with the efficiency calibration derived on



(a) Scale factor

(b) Simulation tagging efficiency

Figure 4.18: Flavour tagging calibration for τ-jets: (a) efficiency scale factors (black points) and smoothed (green), (b) efficiency maps as a function of p_T and $|\eta|$ evaluated on a simulated $t\bar{t}$ sample.

data to obtain the final calibration:

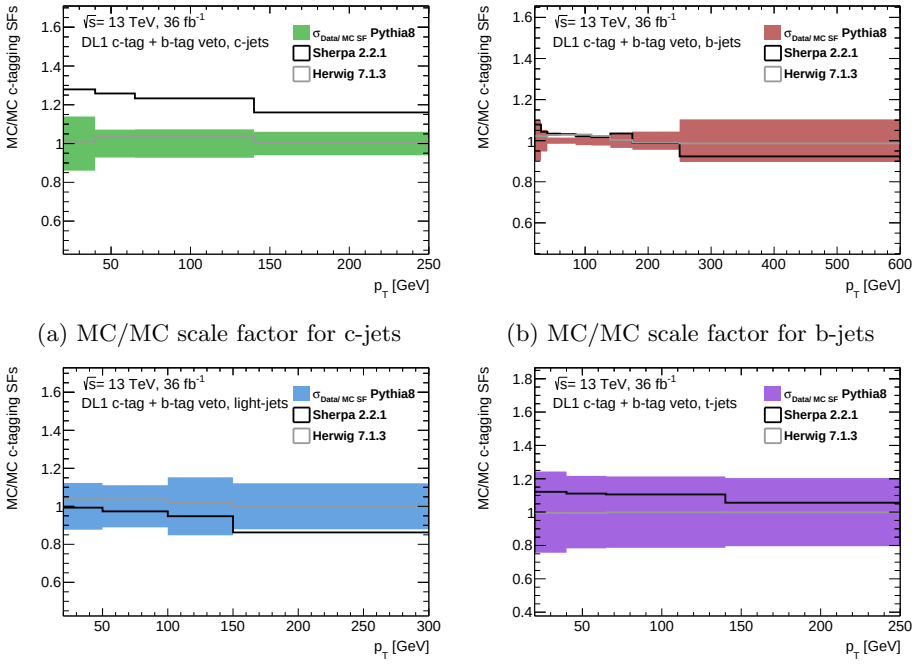
$$SF_{\text{eff,Alt}} = SF_{\text{eff,PwPy8}} \times \frac{\epsilon_{\text{eff,PwPy8}}}{\epsilon_{\text{eff,Alt}}} = \frac{\epsilon_{\text{eff,data}}}{\epsilon_{\text{eff,PwPy8}}} \times \frac{\epsilon_{\text{eff,PwPy8}}}{\epsilon_{\text{eff,Alt}}}. \tag{4.7}$$

4.4.3 Calibration Data Interface

In the ATLAS collaboration, the main reconstruction techniques used by multiple analyses are centralised. The aim is to avoid duplication of work and guarantee the same standardised implementation across all publications. In particular, flavour tagging is centrally developed and regularly updated to describe the latest data collected by the experiment. In addition to the research presented in this thesis, my main contribution to the ATLAS collaborative effort focused on the software used to deliver flavour tagging technologies to the ATLAS analysers also known at the Calibration Data Interface (CDI).

In addition to the development of tagging algorithms, the flavour tagging group calibrates the taggers for different jet reconstruction algorithms including different cone size. As a consequence, a large number of tagging algorithms and corresponding calibrations must be delivered to the collaboration. For Run 2, the implementation of 3 different b-taggers, calibrated across 3 different jet reconstructions algorithms, were delivered. Furthermore, for b-tagging 4 standardised operating points were delivered along with their different performance characteristics. These operating points can be used in two different ways, either as a binary tagger or as a pseudo-continuous tagger where all the jets are tagged in different efficiency bins. All taggers that were calibrated for the Run 2 of the LHC are presented in Table 4.6.

All tagger working points and calibrations are stored in a single file known as the CDI file. The inputs required to create the file are provided as text files used



(a) MC/MC scale factor for c-jets

(b) MC/MC scale factor for b-jets

(c) MC/MC scale factor for light-jets

(d) MC/MC scale factor for τ-jets

Figure 4.19: Flavour tagging MC/MC scale factors derived on a $t\bar{t}$ sample to account for the differences in the alternative simulation hadronisation with respect to Pythia8 for c-jets (a), b-jets (b), light-jets (c) and τ-jets (d).

by the building software. Each tagger is calibrated for all possible permutations of the jet collection, operating point and flavour. In total, 180 unique calibrations are produced and updated to include the data collected each year. Each update requires the build of a new file.

In order to distribute these results, the CDI file is made available to the ATLAS analysers, along with a standard software tool to retrieve the information of these files. The calibration delivery workflow is illustrated in Fig. 4.20.

As all the calibrations are stored in a single file, any error potentially affects a large number of analyses, and due to the large amount of calibrations, it is challenging to ensure that every input is validated and processed correctly. In addition, the time required to release a CDI file depends on the number of iterations between the groups in charge of calibrating the taggers and assembling the file.

In order to simplify the procedure, a new platform has been developed to submit the results in a framework located on the ATLAS Gitlab servers. The advantage

Tagger	Jet collection	Operating point	Flavour
$MV2c10$	AntiKt4EMTopoJets	60%	b-jets
$DL1$	AntiKt4EMPFlowJets	70%	c-jets
$DL1r$	AntiKtVR30Rmax4Rmin02TrackJets	77%	light-jets
		85%	τ-jets
		Pseudo-continuous	

Table 4.6: Summary of the b-tagging algorithms, jet collections and operating points calibrated by the flavour tagging group. Each tagger is calibrated for each possible combination of the jet collection, operating point and flavour.

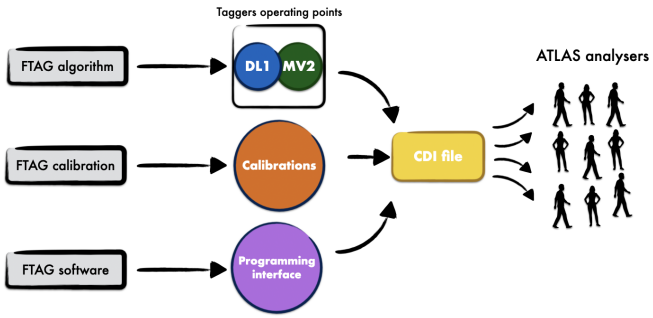


Figure 4.20: Diagram of the structure of the flavour tagging group. The flavour tagging group provides the tagger definition and calibrations that are then delivered to the ATLAS collaboration through a CDI file.

of this approach is that the validation is automated. Each submission of text inputs automatically triggers the build of the CDI file and performs an additional validation against the previous release. The automated validation emulates the analysis usage by creating jets with a certain p_{T} and η in order to retrieve the calibrations and the tagging information. Every large deviation is flagged and the content is scrutinised to avoid that errors are propagated. This workflow is illustrated in Fig. 4.21.

The advantage of this new procedure is that the time between the submission of the inputs and the build of the CDI file is 10 minutes. The release of the CDI file to the whole collaboration takes another 2 days, mostly related to the procedures required to install the file in a location accessible to the entire collaboration.

This new submission structure has been used to build 10 CDI files since 2017 and validate their content. These files have been used for all the ATLAS publications for the Run 2 and most notably for the observation of the $H \to b\bar{b}$ process [71] and the observation of the Higgs boson coupling to top quarks $t\bar{t}H$ [72].

While this interface is mainly used for b-tagging, the same software is used for c-tagging, when the c-tagging calibrations are added to the CDI file. In the case of the $H \to c\bar{c}$ search, several custom CDI files were built to include the c-tagging calibrations used for the measurement, where the ability to build custom CDI files significantly accelerated the progress of the analysis.

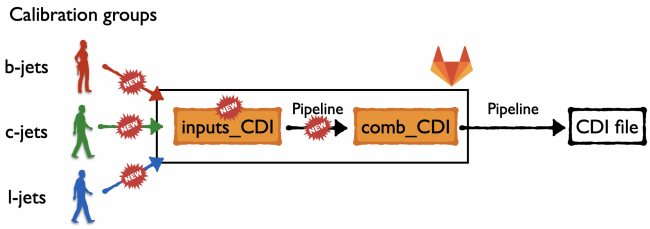


Figure 4.21: Diagram of the new workflow installed in 2017 in order to produce a CDI file. The submission of inputs is performed on an online interface stored on Gitlab and propagated automatically to the software framework which produces the CDI file. Both the inputs and CDI validation is triggered automatically.

4.4.4 Truth tagging

A consequence of the modest efficiency of c-tagging is that imposing a c-tagging requirement on multiple jets in the same event will result in rather low event selection probabilities. This is a particular concern for the $H \to c\bar{c}$ search, where 2 jets are required to be c-tagged: based on the performance in Table 4.5, only $\approx 5\%$ of events are reconstructed with two c-tagged jets. This translates in a loss of 95% of the available signal in both the data and simulation sample. In the case of the simulation, the result is a larger uncertainty due to the finite size of the simulation samples, which degrades the precision of the measurement.

In order to mitigate the effect of the c-tagging efficiency on the size of the simulated sample, a *truth tagging* approach is developed. Instead of placing a direct selection requirement on the c-tagging discriminant, events are weighted based on their probability to pass the c-tagging requirement. The weight is calculated using the tagging efficiency as a function of the jet transverse momentum p_T and pseudo-rapidity η. For an event with 2 jets, the probabilities to have 2 or 1 c-tagged jets are defined as follows:

$$TT_{\text{weight}}^{\text{2c-tag}} = \epsilon_1 \times \epsilon_2, \tag{4.8}$$

$$TT_{\text{weight}}^{\text{1c-tag}} = \epsilon_1 \times (1 - \epsilon_2) + (1 - \epsilon_1) \times \epsilon_2, \tag{4.9}$$

where ϵ_1 and ϵ_2 are the tagging efficiencies for the first and second jet. The advantage of this method is that no simulated events are discarded by the tagging selection and the full simulated sample is used. The efficiency maps used for truth tagging are presented in Sec. 4.4.2 in Fig. 4.15 to 4.17.

In Fig. 4.22, the truth and direct tagging methods are compared for simulated $t\bar{t}$ events for the p_{T} distribution of the c-jets with the highest p_{T} in the event. The events are required to have exactly one c-tagged jet. The truth and direct tagging distributions are in agreement within their uncertainty. Due to the 22% c-tagging efficiency on c-jets, the size of the simulated sample is about four times smaller for direct tagging in comparison to truth tagging. The resulting effect, shown in the middle of Fig. 4.22, is that the relative uncertainty for truth tagging is reduced by a factor two.

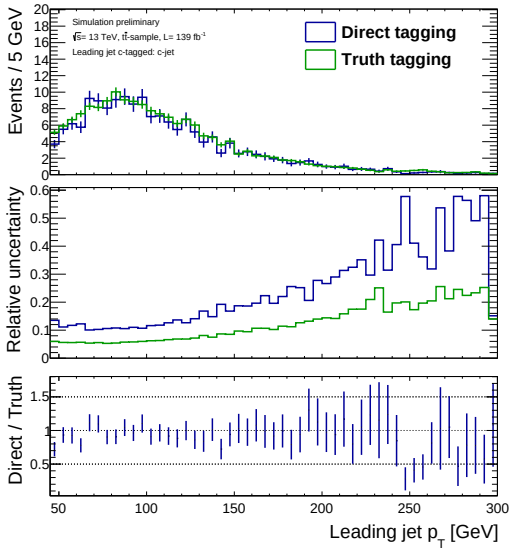


Figure 4.22: Truth and direct tagging comparison for a simulated $t\bar{t}$ sample as a function of the leading c-jet p_{T}. The events are required to have 1 c-tagged c-jet. The upper panel shows the truth and direct tagging distributions, the middle panel shows the relative uncertainty which depends on the number of simulated events, the lower panel shows the ratio of the direct and truth tagging distributions.

5 Search for $H \to c\bar{c}$ with the ATLAS detector

The measurement of the Higgs interaction with c-quarks is interesting for two main reasons. On one hand, an observation of this process in agreement with the SM would be a proof that the Higgs boson interacts with the c-quark. On the other hand, any deviation from this prediction would be a clear sign of physics beyond the SM. A direct probe of this interaction is possible through the measurement of the $H \to c\bar{c}$ process.

The challenge of this measurement lies in the small branching ratio of the $H \to c\bar{c}$ process, which is 2.89%, and the signature of the decay, which contains c-jets. Since various processes produce events with a similar signature via the electroweak or strong force, the number of background events produced by these processes is much larger than the expected number of $H \to c\bar{c}$ events. In addition, the modest efficiency to identify c-jets results in only a small fraction of the total signal produced that is identified as such.

The result of the measurement is quantified in terms of the signal strength of the $H \to c\bar{c}$ process:

$$\mu_{VH(\to c\bar{c})} = \frac{\sigma_{VH}^{obs} \times BR^{obs}(H \to c\bar{c})}{\sigma_{VH}^{SM} \times BR^{SM}(H \to c\bar{c})}, \tag{5.1}$$

where $\sigma \times BR$ is the cross-section times branching ratio of the $H \to c\bar{c}$ process. The signal strength quantifies the agreement between the observed collision data and the SM prediction.

Attempts to measure this interaction were performed by the LHCb [3], ATLAS [4] and CMS [5] collaborations and are shown in Table 5.1. The current best constraint on the $H \to c\bar{c}$ decay, using the data collected by the CMS experiment in 2015 and 2016 [5], is an observed upper limit of $\mu_{VH(\to c\bar{c})} < 70$ at 95% confidence level (CL) .

The aim of the analysis described in this thesis is to measure the $H \to c\bar{c}$ process using the data collected by the ATLAS experiment between 2015 and 2018. This data set is four times larger than the data sets previously used by the ATLAS and CMS collaborations and will result in a higher precision with respect to the previous measurements, which are statistically limited by the size of the recorded sample.

Experiment	Observed (expected) limit at 95% CL	$\mathcal{L}$	$\sqrt{s}$
LHCb	$\mu_{VH(\to c\bar{c})} < 7900$ (6400)	2 fb^{-1}	8 TeV
ATLAS	$\mu_{VH(\to c\bar{c})} < 110$ (150)	36 fb^{-1}	13 TeV
CMS	$\mu_{VH(\to c\bar{c})} < 70$ (37)	36 fb^{-1}	13 TeV

Table 5.1: Observed and expected upper limit on $H \to c\bar{c}$ decay from the LHCb, ATLAS and CMS experiments direct measurements.

5.1 VH production mode

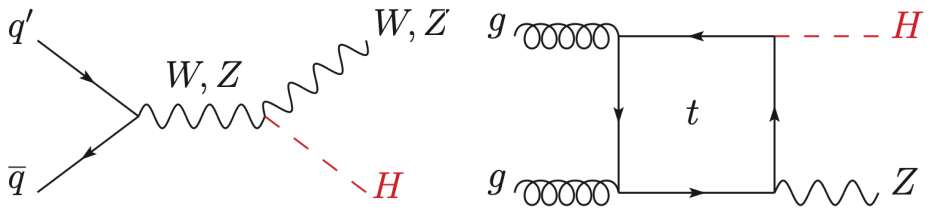
In order to maximise the sensitivity to the $H \to c\bar{c}$ process, the choice of the Higgs boson production mode is important. For the measured Higgs boson mass of $m_H = 125.09 \pm 0.24$ GeV [60], the different production modes and the relative contribution of each mode with respect to the total Higgs boson production are listed in Table 5.2. The most frequent production modes are the gluon-gluon fusion (ggF) and vector boson fusion (VBF) productions, which are responsible for 95% of the total Higgs bosons produced.

Name	Description	Cross-section [pb]	Fraction
ggF	Gluon-gluon fusion	$48.6^{+4.6\%}_{-6.7\%}$	88.2%
VBF	Vector boson fusion	$3.78^{+2.2\%}_{-2.2\%}$	6.8%
WH	Higgsstrahlung	$1.37^{+2.6\%}_{-2.6\%}$	2.5%
ZH	Associated with Z boson	$0.88^{+4.1\%}_{-3.5\%}$	1.6%
$t\bar{t}H$	In association with $t\bar{t}$	$0.50^{+6.8\%}_{-9.9\%}$	0.9%

Table 5.2: Higgs boson production modes and their fraction with respect to the total cross-section production at $\sqrt{s} = 13$ TeV for $m_H = 125$ GeV [10].

However, many processes have a signature similar to the $H \to c\bar{c}$ process. In particular, the QCD multijet process $g \to c\bar{c}$, where a gluon splits to a pair of c-quarks, has a similar final state as the ggF and VBF production modes but with a more frequent production, which degrades the sensitivity to detect the $H \to c\bar{c}$ process. In order to reduce the QCD multijet background, the production process in association with a gauge boson (VH), shown in the two Feynman diagrams in Fig. 5.1, is preferred: when considering only the semi-leptonic decays of the gauge bosons, the $VH(\to c\bar{c})$ signature is significantly different from that of QCD multijets final states.

An illustration of a $Z(\to e^+e^-)H(\to c\bar{c})$ event is presented in Fig. 5.2. The protons accelerated by the LHC collide at the centre of the ATLAS detector. From this collision, a Higgs boson is produced in association with a Z boson, which both decay near the pp-interaction point. The Z boson decays to an electron-positron



(a) Higgsstrahlung (b) Gluon-gluon initiated Higgsstrahlung

Figure 5.1: Leading order Feynman diagrams for the Higgs boson production in association with a vector boson VH. The WH production happens via Higgsstrahlung and the ZH can happen either via the Higgsstrahlung at tree level or via gluon-gluon interaction at loop level.

pair and the Higgs boson decays to a charm-anticharm quark pair, where both quarks hadronise to produce c-hadrons. Due to their lifetime, the c-hadrons fly away from the primary vertex of the collision and decay at a SV in light hadron particles that are measured in the detector. The collimated spray of particles is reconstructed as a jet.

In order to optimise the signal to background discrimination, the VH production mode is further categorised according to the number of charged leptons in the vector boson decay: 0-lepton, 1-lepton and 2-lepton. The list of decay modes of the vector boson considered in the analysis are listed in Table 5.3. In the case of the 0-lepton channel, the $Z \to \nu\bar{\nu}$ decay signature is reconstructed as missing energy in the transverse plane $E_{\mathrm{T}}^{\mathrm{miss}}$. In the case of the 1- and 2-lepton channels, the charged leptons considered are the electron and the muon because they are more cleanly identified than the τ-lepton, which is more complex to reconstruct due to its hadronic decay modes similar to jet signatures.

0-lepton	1-lepton	2-lepton
$Z \to \nu_e\bar{\nu}_e$	$W^+ \to e^+\nu_e$	$Z \to e^-e^+$
$Z \to \nu_\mu\bar{\nu}_\mu$	$W^+ \to \mu^+\nu_\mu$	$Z \to \mu^-\mu^+$
$Z \to \nu_\tau\bar{\nu}_\tau$	$W^- \to e^-\bar{\nu}_e$	
	$W^- \to \mu^-\bar{\nu}_\mu$	

Table 5.3: Leptonic decays of the vector boson in the VH production mode considered in the analysis. The events are categorised in 3 different channels with respect to their number of charged leptons.

The signatures of the 0-, 1- and 2-lepton channels in the ATLAS detector are shown in Fig. 5.3, selected from events recorded in 2015 and 2016. In the VH production mode, the Higgs boson recoils from the vector boson and the energy

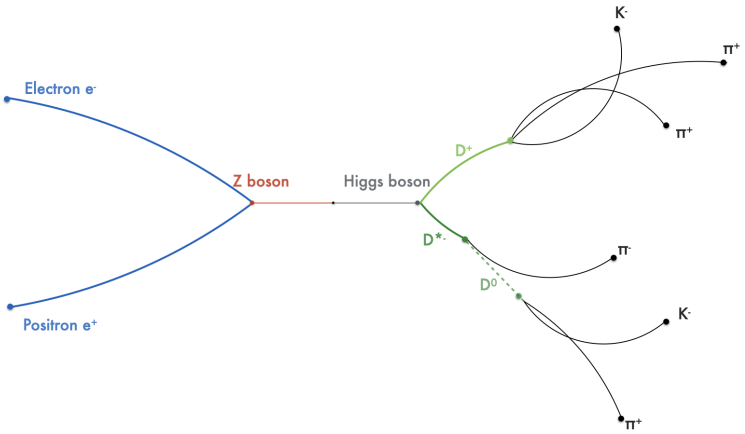


Figure 5.2: Sketch of the production of a $Z(\to e^+e^-)H(\to c\bar{c})$ event in the ATLAS experiment.

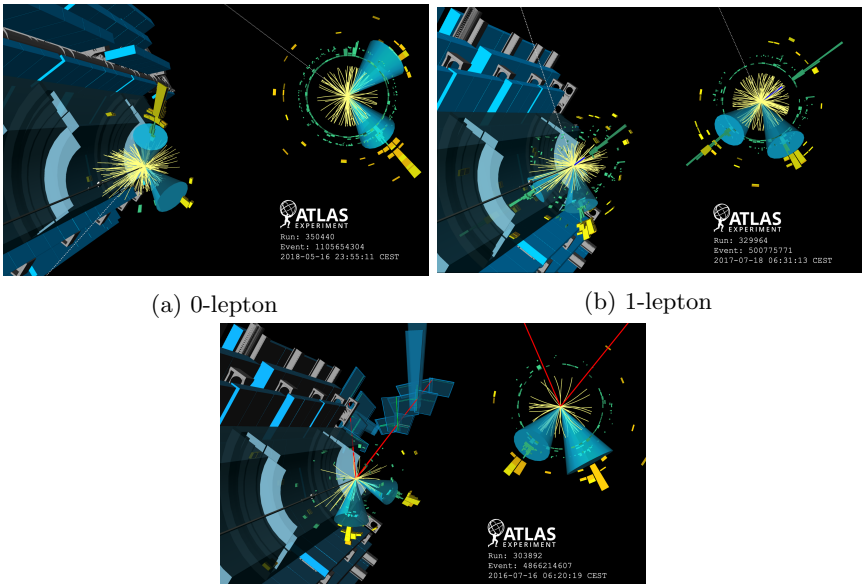
of the vector and Higgs boson is expected to be balanced on the transverse plane. Therefore, the typical signature of the VH production mode consists of jets recorded in an opposite direction from the leptons.

One additional example of a $ZH(\to c\bar{c})$ candidate event of the 2-lepton channel, shown in Fig. 5.4, is selected with 2 c-tagged jets and a mass of $m(c\bar{c}) = 127$ GeV. The energy deposits in the electromagnetic and hadronic calorimeter are shown respectively with blue and yellow bars. The two muons of the Z boson decay, shown in red, are selected with a mass of $m(l\bar{l}) = 88$ GeV. While performing research at CERN, I performed shifts for the inner detector in the ATLAS control room during data taking and this event was recorded in the night of the 23rd October 2018 during one of my shifts.

Due to the different decay modes of the vector boson targeted in each channel, the amount of Higgs boson produced in each channel differs across the channels. The relative amount of Higgs bosons produced per channel with respect to the total cross-section of the VH production mode is listed in Table 5.4. The 1-lepton channel has the largest contribution and represents 56% of the Higgs bosons. The 0-lepton corresponds to 33% and the 2-lepton to 11%. In order to achieve an optimal sensitivity to the $H \to c\bar{c}$ process, all three channels are considered in the analysis.

5.2 The VV process

In order to validate the study of the $H \to c\bar{c}$ decay, the processes $Z \to c\bar{c}$, $W \to cs$ and $W \to cd$, referred to as $W \to cq$, are also measured. Both the W



(a) 0-lepton (b) 1-lepton

(c) 2-lepton

Figure 5.3: Event displays of the 0-lepton (a), 1-lepton (b) and 2-lepton (c) channels recorded by the ATLAS detector in 2015 and 2016.

and Z production and decay modes have been studied at collider experiments and their measurement is expected to be in excellent agreement with the theory prediction. As the $H \to c\bar{c}$ process has never been measured and could potentially be sensitive to new physics, it is important to guarantee that the result can be interpreted in terms of physics effects and minimise systematic biases. As these biases would similarly affect the validation processes, the extent of any biases is experimentally verified from their measurement. In the analysis, the events are categorised depending on the number of c-tagged jets. Therefore, the 2 c-tags category is validated through the $Z \to c\bar{c}$ measurement and the 1 c-tag category is validated through the $W \to cq$ measurement.

The W and Z bosons are produced in association with a vector boson ($VV = VW + VZ$), similarly to the signal, and are reconstructed as a resonant peak in the invariant mass distribution of the two jets. The main difference of these processes compared to the $H \to c\bar{c}$ process is the invariant mass of the second boson, $m_Z = 91$ GeV and $m_W = 80$ GeV, compared to the invariant mass of the Higgs boson $m_H = 125$ GeV. In addition, all $VV(\to qq')$ production cross-sections are larger than the $VH(\to c\bar{c})$ cross-sections. A summary of the production cross-section for the VH, VZ and VW signals are listed in Table 5.5. The number of $Z \to c\bar{c}$ and $W \to cl$ events expected to be produced in the collisions are

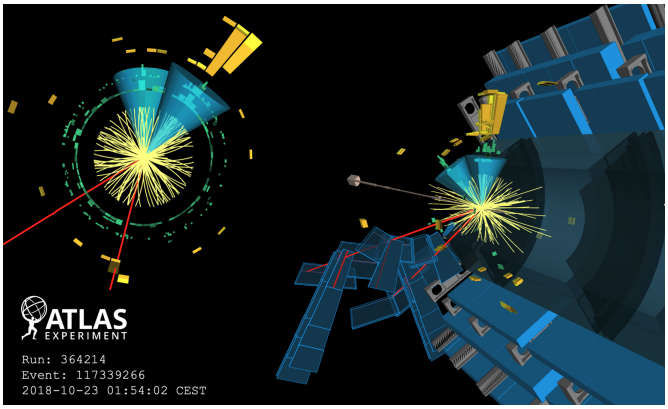


Figure 5.4: Candidate event for the $Z(\to \mu^+\mu^-)H(\to c\bar{c})$ process recorded on the 23rd of October 2018. The muons are shown in red and the mass of the Z boson is measured to be $m(\mu^+\mu^-) = 88$ GeV. The blue cones represent the jets from the decay of the cadidate Higgs boson with a mass $m(c\bar{c}) = 127$ GeV.

respectively $\mathcal{O}(150)$ and $\mathcal{O}(1500)$ times larger than the $H \to c\bar{c}$ signal.

5.3 Background processes

Several other processes have final states containing leptons and jets, which can be misidentified as $VH(\to c\bar{c})$ candidates. In order to extract the $H \to c\bar{c}$ signal from the data, each of these processes needs to be taken into account as a background process. Depending on their final states, these backgrounds affect the 0-, 1- and 2-lepton channels differently.

The dominant backgrounds are W+jets and Z+jets, where the W or Z boson decays leptonically and is produced in association with at least 2 jets. These backgrounds have the same signature as the $VH(\to c\bar{c})$ signal. The main difference with respect to the signal is that the invariant mass of the 2 jets system is not resonant and approximately uniformly distributed in the signal region. As this process is mediated by the electroweak force, its cross-section is $\mathcal{O}(10^6)$ times larger than the signal cross-section.

Two further backgrounds of interest with the same signature as the $VH(\to c\bar{c})$ signal are the $VH(\to b\bar{b})$ and $VH(\to \tau\bar{\tau})$ processes. Both backgrounds have the same invariant mass distribution as the signal and the production of the $H \to b\bar{b}$ and $H \to \tau\bar{\tau}$ processes are respectively 20 times and 2 times more abundant than the signal.

In addition, due to detector limitations, such as reconstruction inefficiencies and kinematic acceptance, other processes with a signature that differs from the signal

Channel	Process	Cross-section [fb]	Fraction
0-lepton	$qq \to ZH \to \nu\bar{\nu}c\bar{c}$	4.5	28%
	$gg \to ZH \to \nu\bar{\nu}c\bar{c}$	0.7	5%
1-lepton	$qq \to W^+H \to l^+\nu c\bar{c}$	5.4	34%
	$qq \to W^-H \to l^-\nu c\bar{c}$	3.5	22%
2-lepton	$qq \to ZH \to l^-l^+c\bar{c}$	1.5	9%
	$gg \to ZH \to l^-l^+c\bar{c}$	0.25	2%

Table 5.4: The 0-, 1- and 2-lepton channels of the VH production mode and their relative cross-section with respect to the total production cross-section for the leptonic decays considered. These numbers are obtained based on the `Powheg+Pythia8` calculation for the quark initiated processes and `Sherpa2.2` for the gluon initiated processes.

Channel	Process	$\sigma\times$ BR [fb]	$\frac{\sigma_{\text{channel}}\times\text{BR}_{\text{channel}}}{\sigma_{\text{VH}}\times\text{BR}_{\text{VH}}}$
VH	$qq/gg \to ZH(\to c\bar{c})$	8	1
	$qq \to WH(\to c\bar{c})$	13	
VZ	$qq/gg \to ZZ(\to c\bar{c})$	1236	152
	$qq \to WZ(\to c\bar{c})$	1961	
VW	$qq \to ZW(\to cl)$	5055	1453
	$qq/gg \to WW(\to cl)$	25465	

Table 5.5: cross-section multiplied by the branching ratio for the $H \to c\bar{c}$, $Z \to c\bar{c}$ and $W \to cl$ processes produced in association with a vector boson. In addition, the relative ratio with respect to the VH production mode is shown.

are also sometimes mistakenly reconstructed as $H \to c\bar{c}$ candidates. This is the case for the top-antitop quark pairs production ($t\bar{t}$), single top production and QCD multijet production. The main backgrounds for the $H \to c\bar{c}$ measurement are summarised in Table 5.6.

In the case of the 0-lepton channel, the three main backgrounds are the Z+jets, W+jets and $t\bar{t}$ processes. In the W+jets and $t\bar{t}$ processes, the leptonic decay of the W boson, where the lepton is lost due to detector acceptance, has the same signature as a 0-lepton event. In the case of 1-lepton, the main backgrounds are the W+jets and $t\bar{t}$ processes, where one of the W decays leptonically. In addition, the QCD multijet production is also a significant background in the 1-lepton channel due to semileptonic decays within c-jets. The 2-lepton channel has the simplest background composition and mainly consists of the Z+jets process, as well as the $t\bar{t}$ process, where both W bosons decay leptonically. All three channels are affected by both the single top, $VH(\to b\bar{b})$ and $VH(\to \tau\bar{\tau})$ backgrounds, but their relative

Proces	0-lepton	1-lepton	2-lepton
$Z + jets$	✓		✓
$W + jets$	✓	✓	
$t\bar{t}$	✓	✓	✓
Single top	✓	✓	✓
QCD multijets		✓	
$VH(\to b\bar{b})$	✓	✓	✓
$VH(\to \tau\bar{\tau})$	✓	✓	✓

Table 5.6: Main backgrounds for the $H \to c\bar{c}$ measurement.

contribution is much smaller than the other backgrounds. All backgrounds are simulated from theory predictions and are estimated from the data in dedicated control regions.

5.4 Data and Monte Carlo samples

In order to measure the $H \to c\bar{c}$ process, candidate events in the data are selected from the data set collected between 2015 and 2018 by ATLAS with $\sqrt{s} = 13$ TeV pp-collisions, which corresponds to an integrated luminosity of 139 fb^{-1}. The signal strength is extracted from a fit to the data using simulated Monte-Carlo (MC) events passing the same selection. Therefore, the signal, validation processes and backgrounds are simulated to model the conditions observed in the data.

Process	Generator	σ [pb]	Precision σ_{incl}
VH $H \to c\bar{c}$	Powheg+Pythia8	0.729×0.0289	NNLO(QCD) + NLO(EW)
$qq/qg/gg \to V(\to \nu\nu, l\nu, ll)V$	Sherpa	80	NLO
$V(\to \nu\nu, l\nu, ll)$+jets	Sherpa	33560	NNLO
$t\bar{t}$	Powheg+Pythia8	831	NNLO + NNLL
Single top	Powheg+Pythia8	299	NLO
VH $H \to b\bar{b}$	Powheg+Pythia8	0.729×0.582	NNLO(QCD) + NLO(EW)

Table 5.7: MC generators used to model the signal and the background processes, and corresponding cross-sections. The precision σ_{incl} gives the calculation order of the inclusive cross-section applied to the sample.

The MC event generators used to simulate the signal and background processes are listed in Table 5.7. In the case of the of the VH, $t\bar{t}$ and single top processes, separate programs are used to generate the hard scattering process and model the subsequent parton showering, hadronisation and the underlying event that accounts for additional inelastic pp-interactions in the event. The hard scattering is simulated using Powheg-Box v2. The parton shower, hadronisation and underlying event are modelled using Pythia8. The matrix element is generated

using the parton distribution function set NNPDF3.0NLO. In the case of the V+jets background and the diboson validation process, the entire physics simulation chain is performed by Sherpa2.2, with the exception of the underlying event that is simulated with Pythia8. For the Sherpa hard scattering simulation, the matrix element is generated using the parton distribution function set NNPDF3.0NNLO. To account for the processes where higher order calculations exist, the simulated samples are normalised to the latest cross-section calculations.

The choice of the MC generator is determined empirically, based on the previous measurements performed by the ATLAS collaboration. Alternative predictions are used for all the main backgrounds to estimate systematic uncertainties on these predictions.

5.5 Object selection

In order to select the 0-, 1- and 2-lepton final states in the data, information from all sub-detectors of the ATLAS experiment are combined to reconstruct physics signatures of leptons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$. To maximise the sensitivity to the $VH(\rightarrow c\bar{c})$ process, while reducing the mis-identification of other signatures, events are selected based on the reconstructed objects.

Primary Vertex

As the Higgs boson is produced through the hard scattering of protons, the vertex of the VH production is expected to have the highest scalar p_{T} sum among all reconstructed vertices. Vertices are reconstructed by combining at least two tracks measured in the Inner Detector that are measured to be consistent with the same origin. Each reconstructed event is required to have one primary vertex, which is selected as the one with the highest sum of transverse momentum $\Sigma|p_{\mathrm{T}}|^2$ [49].

Leptons

In order to reconstruct events with the 0-, 1- and 2-lepton signatures, the presence or absence of leptons is a key component. To avoid overlaps between the channels, events are categorised depending on the number of leptons satisfying a common selection, which is referred to as "V-loose-lepton" and which depends on the flavour of the lepton, and will be detailed further in the following sections. Furthermore, in the case of the 1- and 2-lepton channels, events are selected with additional requirements on the kinematics of the leptons, referred to as "W-signal-lepton" and "Z-signal-lepton", which are designed to reduce the rate of non-prompt leptons produced in the semileptonic decays of heavy hadrons or fake lepton signatures. These fake leptons are typically mis-reconstructed from jets with a high energy deposit in the electromagnetic calorimeter.

Electron Selection

For electrons, the *V-loose-lepton* selection is chosen to maximise the electron selection efficiency for the VH signal: the electron p_{T} is required to be greater than 7 GeV and the electromagnetic cluster in the range of $|\eta| < 2.47$. In order to select prompt electrons, electron candidates are required to pass a 93% efficiency operating point on a likelihood discriminant constructed from quantities measured in the tracker and calorimeters [53]. Jets mis-reconstructed as electrons are rejected by requiring the electron candidates to pass an isolation selection with a 99% efficiency operating point. The electrons produced in semileptonic decays of heavy hadrons within jets are further rejected by ensuring that the electron is consistent with the primary vertex through the impact parameter significance $\frac{d_0}{\sigma_{d_0}} < 5$ and track displacement $|\Delta z_0 \sin\theta| < 0.5$ mm. In addition to the *V-loose-lepton* electron selection, in the *W-signal-lepton* and *Z-signal-lepton* selections, electrons are required to have a $p_{\mathrm{T}} > 27$ GeV.

In order to reduce the QCD multijet background in the 1-lepton channel, the likelihood and isolation operating point used in the *W-signal-lepton* selection require a lower electron efficiency with a higher background rejection: electron candidates are required to pass the 80% efficiency operating point of the likelihood discriminant. The track isolation selection is based on the calorimeter only and is designed to give the highest background rejection at $p_{\mathrm{T}} > 100$ GeV, while maintaining 95% electron efficiency.

A summary of electron selection criteria, for *V-loose-lepton*, *W-signal-lepton* and *Z-signal-lepton* is listed in Table 5.8.

Electron Selection	p_{T}	η	Likelihood OP	$\frac{d_0}{\sigma_{d_0}}$	$\lvert\Delta z_0 \sin\theta\rvert$	Isolation OP
V-loose-lepton	>7 GeV	$\lvert\eta\rvert < 2.47$	93%	< 5	< 0.5 mm	99%
W-signal-lepton	>27 GeV	$\lvert\eta\rvert < 2.47$	80%	< 5	< 0.5 mm	95%
Z-signal-lepton	>27 GeV	$\lvert\eta\rvert < 2.47$	93%	< 5	< 0.5 mm	99%

Table 5.8: Electron selection criteria.

Muon Selection

For muons, a similar procedure as for electrons is applied and a summary of the *V-loose-lepton*, *W-signal-lepton* and *Z-signal-lepton* selections is shown in Table 5.9. In order to select prompt muons and reduce the selection of non-prompt muons and jets mis-reconstructed as a muon, a likelihood discriminant as well as an isolation criterium are used with a similar reconstruction efficiency as for electrons [73]. In addition, as muon signatures are reconstructed more precisely than electron signatures, the requirement on the track impact parameter significance is tightened, to ensure a higher compatibility with the primary vertex.

Muon Selection	p_{T}	η	Likelihood OP	$\frac{d_0}{\sigma_{d_0}}$	$\lvert\Delta z_0 \sin\theta\rvert$	Isolation OP
V-loose-lepton	>7 GeV	$\lvert\eta\rvert < 2.7$	98%	< 3	< 0.5 mm	99%
W-signal-lepton	>25 GeV	$\lvert\eta\rvert < 2.5$	96%	< 3	< 0.5 mm	96%
Z-signal-lepton	>27 GeV	$\lvert\eta\rvert < 2.5$	98%	< 3	< 0.5 mm	99%

Table 5.9: Muon selection criteria.

Jets

Reconstructed jets are another key component to measure the $VH(\rightarrow c\bar{c})$ signal. While the typical event signature of the signal is composed of 2 jets, initial and final state radiation can increase the number of jets in the final state. In order to maximise the signal reconstruction efficiency, multiple jets need to be reconstructed in different regions of the detector.

Each jet is reconstructed from topological calorimeter-cell clusters via the anti-k_t algorithm [74] with a radius parameter of $R = 0.4$. Jets are reconstructed in both the central ($|\eta| < 2.5$) and forward ($2.5 < |\eta| < 4.5$) regions of the detector and are respectively labelled "signal" and "forward" jets. In order to suppress jets arising from pileup, a likelihood-based discriminant, referred to as the the Jet Vertex Tagger (JVT [75]), is used at an operating point corresponding to 92% tagging efficiency. The full set of selection requirements is given in Table 5.10. The difference in the selection between the signal and forward jets is due to the η coverage of the inner detector, which is up to $|\eta| < 2.5$ and no track information is available for the forward jets.

Jet Category	Selection Requirements
Signal Jets	$p_{\mathrm{T}} \geq 20\,\mathrm{GeV}$ and $\lvert\eta\rvert < 2.5$ JVT> 0.59 for $\lvert\eta\rvert < 2.4$ and > 0.11 for $2.4 < \lvert\eta\rvert < 2.5$ if $p_{\mathrm{T}} < 120\,\mathrm{GeV}$
Forward Jets	$p_{\mathrm{T}} > 30\,\mathrm{GeV}$ $2.5 \leq \lvert\eta\rvert < 4.5$

Table 5.10: `AntiKt4EMTopoJets` selection requirements.

Hadronic τ-jets

Due to the similarity in the signatures of τ- and c-jets, τ-jets require a dedicated reconstruction in the analysis and a similar approach to the jet reconstruction is applied. Hadronic decays of the τ-lepton are characterised by having one or three charged particles (prongs) in addition to an unseen neutrino and more neutral pions that are detected as electromagnetic calorimeter depositions. The reconstruction of τ-jets starts with the anti-k_t algorithm with a size parameter value of $R = 0.4$ [76]. Charged tracks associated to the τ-lepton decay are required to be in the core region $\Delta R < 0.2$ around the τ-jet axis and the p_{T} is reconstructed as the sum

of these tracks. The following selection is applied to reconstructed τ-jets:

- $p_{\mathrm{T}} > 20$ GeV
- $|\eta| < 2.5$
- 1 or 3 tracks

In order to distinguish τ-jets from other flavour jets, a BDT based τ-jet identification is used with two operating points depending on the number of prongs: 60% τ-tagging efficiency for 3 prongs and 75% for 1 prong [77]. In the analysis, τ-tagged jets are separately counted from signal jets: for example, an event reconstructed with 1 signal jet and 1 τ-tagged jet does not satisfy the event selection criteria of at least 2 signal jets.

Missing transverse energy $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$

In the 0- and 1-lepton channel, the final states include neutrinos, which do not interact with the detector. The resulting signature of a proton-proton collision producing one more neutrino is missing energy in the transverse plane, $E_{\mathrm{T}}^{\mathrm{miss}}$, which is defined as the negative vector sum of the p_{T} of all physics objects in the event that were reconstructed (leptons, jets, ...). In addition to the energy from the reconstructed signatures in the event, the $E_{\mathrm{T}}^{\mathrm{miss}}$ also includes a track-based soft term, which is calculated from tracks not associated with any object in the event. In order to reduce the influence of pile-up collisions in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, only tracks associated with the primary vertex are used in this soft term.

Overlap Removal

In events with multiple physics signatures, each signature is reconstructed independently from each other and, in some cases, a signature can be reconstructed as two different objects. In order to avoid double counting and to ensure that every signature in the event is only used in the analysis with the best-fitting reconstruction hypothesis, an overlap removal procedure is applied as follows:

Electron-muon overlap: If a muon shares a track with an electron, the electron is removed, unless the muon is tagged in the calorimeter, in which case the muon is removed.

Jet-lepton overlap: If an electron or muon is reconstructed closer than $\Delta R < 0.2$ to a jet, the jet is removed. If an electron or muon is found between $0.2 < \Delta R < 0.4$ near a jet, the electron or muon is removed.

τ-jet-lepton overlap: If a τ-jet is reconstructed closer than $\Delta R \leq 0.2$ to an electron or muon, the τ-jet is removed.

τ-jet-jet overlap: If a τ-jet is reconstructed closer than $\Delta R \leq 0.2$ to a jet , the jet is removed.

Trigger Selection

In the data, candidate $VH(\to c\bar{c})$ events are recorded following the activation of one or more triggers, which are designed to identify events with charged leptons or missing transverse energy.

Events in the 0-lepton channel are selected using the $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers with the lowest energy threshold that is not prescaled. This lowest available threshold has increased over the years with increasing luminosity: 70 GeV for 2015 and 2016 and 110 GeV for 2017 and 2018. For the 1-lepton channel, events with muons are recorded using the same $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers as those of the 0-lepton channel, however for the electron channel the lowest single electron triggers in each data collection period are used, with thresholds ranging from 24 to 26 GeV. For the 2-lepton channel, events are selected using the same single electron triggers as are used for the 1-lepton channel. For 2-muon events, the lowest single muon triggers are used in each data collection period, with thresholds ranging from 20 to 26 GeV.

5.6 Event selection

The event selection is designed to maximise the sensitivity to the $VH(\to c\bar{c})$ signal, while maintaining a high sensitivity to the $VZ(\to c\bar{c})$ and $VW(\to cl)$ processes. As the events are categorised in three different channels, the strategy is to harmonise the Higgs boson reconstruction and, where possible, the selection criteria of the jets across all three channels. Beyond these common criteria, additional requirements on the lepton and/or $E_{\mathrm{T}}^{\mathrm{miss}}$ are applied in each channel to optimise the background suppresion.

5.6.1 Event selection common to 0-, 1- and 2-lepton channels

In the $VH(\to c\bar{c})$ process, initial or final state radiation increase the number of jets beyond the nominal 2 jets present at leading order. In order to maximise the signal efficiency, events are required to have at least 2 jets.

Due to the mass of the Higgs boson, the average jet p_{T} is 60 GeV for the jet originating from the Higgs boson, which is generally higher than for jets originating from radiation. Therefore, for events with multiple jets, the jets are ordered according to their p_{T}. The 2 jets in the event with the highest p_{T} are assumed to originate from the Higgs boson decay, which are referred to as the leading and sub-leading jets in the event. In order to further reduce the backgrounds, the leading jet in the event is required to have a $p_{\mathrm{T}} > 45$ GeV, which is optimal for the significance of both the signal and validation processes.

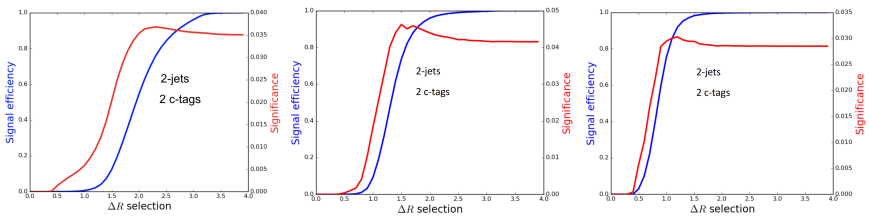
In addition, in the case of the Z+jets process, a fraction of the events are produced through quarks radiating both a Z boson and a gluon, which splits in two jets. Due to the soft collinear divergence in QCD, the jets are more difficult to model in the simulation, which also results in a larger disagreement between the

data and the simulation for lower $m(c\bar{c})$ values. In order to largely exclude this difficult-to-simulate region of phase space from the analysed event selection, the invariant mass $m(c\bar{c})$ is required to be larger than 50 GeV.

Furthermore, jets originating from the Higgs boson, a vector boson or produced via gluon splitting are expected to have a smaller ΔR between the leading jets compared to the backgrounds where the jets do not originate from the decay of a single particle. This is especially true for events with a high transverse momentum p_{T}^{V} of the vector boson, where the jets are expected to be more collimated for resonant decays. In order to suppress the background from $t\bar{t}$ and V+jets, an upper boundary on the selection applied to $\Delta R(\mathrm{jet1}, \mathrm{jet2})$ is optimised in different p_{T}^{V} regimes across the 0-, 1- and 2-lepton channels. In all selection optimisations considerations, the figure of merit is the statistical significance of the $VH(\to c\bar{c})$ signal, which is calculated using the $m(c\bar{c})$ distribution.

The result of the signal significance optimisation is shown in Fig. 5.5, where the signal efficiency and significance are calculated for different values of the upper boundary on the selection applied to $\Delta R(\mathrm{jet1}, \mathrm{jet2})$. The results are shown for the 2-lepton channel and the optimal values for the 0- and 1-lepton channels are found to be consistent. As a result of the optimisation study, three separate kinematic regimes have been identified that each have separate selection criteria:

- 75 GeV $< p_{\mathrm{T}}^{V} \leq$ 150 GeV: $\Delta R(\mathrm{jet1}, \mathrm{jet2}) \leq 2.3$
- 150 GeV $< p_{\mathrm{T}}^{V} \leq$ 250 GeV: $\Delta R(\mathrm{jet1}, \mathrm{jet2}) \leq 1.6$
- 250 GeV $< p_{\mathrm{T}}^{V}$: $\Delta R(\mathrm{jet1}, \mathrm{jet2}) \leq 1.2$.



(a) $75 < p_{\mathrm{T}}^{V} < 150$ GeV (b) $150 < p_{\mathrm{T}}^{V} < 250$ GeV (c) $p_{\mathrm{T}}^{V} > 250$ GeV

Figure 5.5: Upper selection on $\Delta R(\mathrm{jet1}, \mathrm{jet2})$ performed on 139 fb^{-1} with the signal region selection applied at the exception of the ΔR upper selection.

5.6.2 Selection specific to the 0-lepton channel

In the 0-lepton channel, the selection is designed to include events containing a $Z(\to \nu\bar{\nu})H(\to c\bar{c})$ signatures. All events are required to have exactly 0 leptons that satisfy the *V-loose-lepton* requirement. In order to reduce the background

from QCD multijet production, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is required to be larger than 150 GeV. In addition, the scalar jet p_{T} sum $\sum_j p_{\mathrm{T}}$ (j) is required to be greater than 120 GeV for events with 2 jets and greater than 150 GeV for events with 3 jets.

QCD multijet events can enter the selection of the 0-lepton channel due to loss of energy in the jet p_{T}, in which case the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the jet are aligned. In order to reduce this effect, events are required to pass the four following requirements:

- $|\Delta\Phi(E_{\mathrm{T,trk}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss}})| < 90$ °
- $|\Delta\Phi(\mathrm{jet1}, \mathrm{jet2})| < 140$ °
- $|\Delta\Phi(E_{\mathrm{T}}^{\mathrm{miss}}, H)| > 120$ °
- $\min[|\Delta\Phi(E_{\mathrm{T}^{\mathrm{miss}}}, \mathrm{jets})|] > 20$ ° for 2 jets, > 30 ° for 3 jets.

Furthermore, in order to reduce the effect of pile-up, the missing transverse energy reconstructed exclusively from tracks in the events, $E_{\mathrm{T,trk}}^{\mathrm{miss}}$, is required to be larger than 30 GeV.

5.6.3 Selection specific to the 1-lepton channel

In the 1-lepton channel, the selection is designed to include events containing a $W(\rightarrow l\nu)H(\rightarrow c\bar{c})$ signature. Events are required to contain exactly 1 *W-signal-lepton* and events containing additional *V*-loose-leptons are rejected. In order to reduce the background from QCD multijet production, only events with $p_{\mathrm{T}}^{V} > 150$ GeV are considered. In the electron channel, events are additionally required to have $E_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV. In order to reduce the $t\bar{t}$ background, the transverse mass of the W boson, m_{T}^{W}, is required to be smaller than 120 GeV.

5.6.4 Selection specific to the 2-lepton channel

In the 2-lepton channel, the selection is designed to include events containing a $Z(\rightarrow l\bar{l})H(\rightarrow c\bar{c})$ signature. Events are required to have exactly two leptons of the same flavour satisfying the *V-loose-lepton* criteria, and at least one lepton is required to have $p_{\mathrm{T}} > 27$ GeV, which corresponds to the *Z-signal-lepton* requirement. Events with any additional *V-loose-lepton* are discarded. In 2-muon events, both muons are required to have opposite charges and have kinematics consistent with that of a neutral Z boson decay. Due to the higher charge mis-identification rate for high p_{T} electrons, the opposite charge requirement is not applied to events with electrons.

In order to suppress non-resonant backgrounds such as $t\bar{t}$ and the QCD multijet production, the invariant mass reconstructed from the two leptons must be approximately consistent with the Z boson mass: $81 < m(ee, \mu\mu) < 101$ GeV.

The full event selection applied in the 0-lepton, 1-lepton and 2-lepton is listed in Table 5.11.

Common Selections	
Jets	≥ 2 signal jets
Leading jet p_{T}	> 45 GeV
$m(c\bar{c})$	> 50 GeV
ΔR(jet1, jet2)	$p_{\mathrm{T}}^{V} \in$ [75,150] GeV, $\Delta R \leq 2.3$ $p_{\mathrm{T}}^{V} \in$ [150,250] GeV, $\Delta R \leq 1.6$ $p_{\mathrm{T}}^{V} \in$ [250,∞] GeV, $\Delta R \leq 1.2$
0 Lepton	
Trigger	lowest un-prescaled $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers
Leptons	0 *V-loose-lepton* lepton
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 150 GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 30 GeV
Sum p_{T}	> 120 (2 jets), >150 GeV (3 jets)
$\lvert\min\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet})\rvert$	$> 20^{\circ}$ (2 jets) , $> 30^{\circ}$(3 jets)
$\lvert\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, h)\rvert$	$> 120^{\circ}$
$\lvert\Delta\phi$(jet1,jet2)$\rvert$	$< 140^{\circ}$
$\lvert\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$	$< 90^{\circ}$
1 Lepton	
Trigger	e channel: un-prescaled single electron μ channel: above see 0-lepton triggers
Leptons	1 *W-signal-lepton* lepton > 1 *V-loose-lepton* lepton veto
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 30 GeV (e channel)
m_{T}^{W}	< 120 GeV
2 Lepton	
Trigger	un-prescaled single lepton
Leptons	2 *V-loose-lepton* leptons (≥ 1 *Z-signal-lepton* lepton) Same flavor, opposite-charge for $\mu\mu$
$m(l\bar{l})$	$81 < m(l\bar{l}) < 101$ GeV

Table 5.11: Summary of the event selection in the 0-, 1- and 2-lepton channels.

5.7 Flavour tagging

The search for the $VH(\to c\bar{c})$ signal is similar to the measurement of the $VH(\to b\bar{b})$ process performed by the ATLAS collaboration [71]. For both analyses, events are selected with the same kinematic criteria and the main difference is in the flavour tagging selection. Due to the similarity of the two measurements, candidate events can be identified as part of both the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$ selection. In order to perform a combined measurement of the two processes, which allows for a better interpretation of the Higgs boson properties, this overlap will need to be addressed.

5.7.1 Flavour tagging categorisation

As the $VH(\to c\bar{c})$ process contains 2 c-jets in the final state, events are selected with 2 c-tagged jets. However, due to the low performance of the c-tagging algorithm, only a smaller fraction of signal events are identified with 2 c-tagged jets, while a larger fraction of the signal is selected with only 1 c-tagged jet. In order to increase the sensitivity to the signal process, events with 1 c-tagged jets are also considered in the analysis to recover part of the sensitivity lost by requiring 2 c-tagged jets. In the $VH(\to b\bar{b})$ analysis, events are required to have 2 b-tagged jets and, therefore, to remove the overlap between the two analyses, the c-tagging requirement is defined to include a b-tag veto.

Events with 2 jets In events with 2 jets, the Higgs boson candidate is reconstructed from the 2 jets. For the flavour tagging categorisation, illustrated in Fig. 5.6, each jet can be either b-tagged, c-tagged and not tagged. No event is simultaneously categorised in both the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$ signal regions. Events with 1 b-tagged jet are not considered in the $VH(\to b\bar{b})$ analysis but are included in the 1 c-tag category of the $VH(\to c\bar{c})$ analysis. The $VH(\to c\bar{c})$ events selected with 2 b-tagged jets are used in the $VH(\to b\bar{b})$ analysis, however, the sensitivity to the $VH(\to c\bar{c})$ process in the 2 b-tags category is recovered in the combination of the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$ analysis.

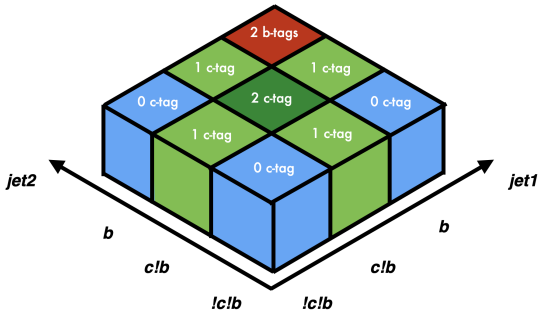
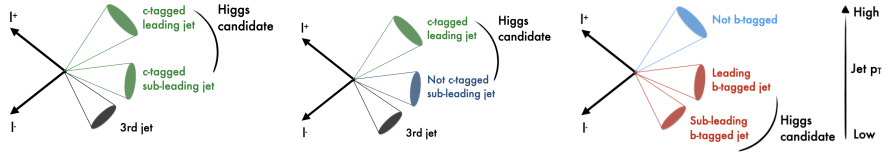


Figure 5.6: Illustration of the flavour tagging categorisation for events with 2 jets. Events with 2 b-tags are depicted in red, 2 c-tags in dark green, 1 c-tag in light green and 0 c-tag in blue.

Events with ≥ 3 jets In order to achieve the orthogonality between the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$ measurements, defining the c-tagging requirement with an additional b-tag veto is not always sufficient. Although the veto suffices to remove the overlap for events with exactly 2 jets, overlap still occurs in the categorisation of events with ≥ 3 jets. This is due to the strategy used to reconstruct the Higgs

boson in the two analyses, namely that events with 3 jets are possibly categorised as both 2 b-tags and 1 c-tag. This overlap between the two categories is shown in Fig. 5.7, where the sub-leading jet failing the c-tagging requirement is potentially b-tagged. A similar issue arises in events with 4 jets that can be categorised as 2 b-tags and 2 c-tags.



(a) 2 c-tags (b) 1 c-tag (c) 2 b-tags

Figure 5.7: Flavour tagging categorisation in 2 c-tags (a), 1 c-tag (b) and 2 b-tags (c) based on the strategies to reconstruct the Higgs boson used in the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$ analyses.

In the $VH(\to c\bar{c})$ analysis, to suppress backgrounds, the best sensitivity to the signal is achieved when the Higgs boson candidate is reconstructed from the 2 jets with the highest p_{T} in the event. Therefore, the categorisation is based on the number of c-tagged jets among the 2 leading jets. In the $VH(\to b\bar{b})$ analysis, the Higgs boson candidate is reconstructed from any 2 b-tagged jets in the event.

For events with 3 jets, the categorisation is illustrated in Fig. 5.8 for three different scenarios depending on the tagging information of the 3rd jet in the event. Events where the 3rd jet is not tagged are categorised in the same way as events with exactly 2 jets. Since the Higgs boson candidate in the $VH(\to c\bar{c})$ analysis is reconstructed from the 2 leading jets in the event, events are categorised irrespective of the 3rd jet being c-tagged. Events with 2 b-tagged jets are always used in the $VH(\to b\bar{b})$ analysis.

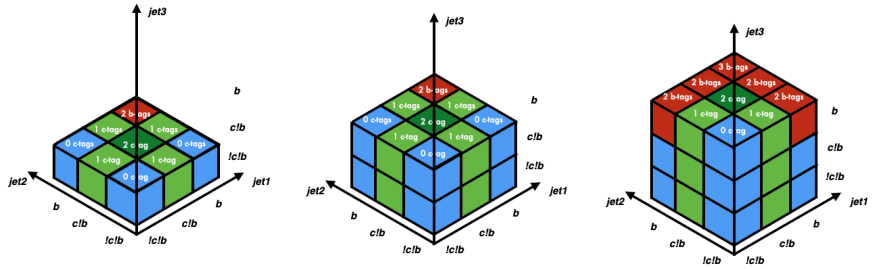
In order to remove any overlap between the categories, a b-tag veto is applied on all ≥ 3 jets in the event. In addition to the orthogonality between the two measurements, the advantage of this approach is to remove more $t\bar{t}$ background, which is generally larger in events with ≥ 3 jets.

Charm tagging working point optimisation

To achieve the best sensitivity to the $VH(\to c\bar{c})$ process, the c-tagging requirement is optimised as illustrated in Fig. 5.9. To define the b-tag veto in the c-tagging requirement, events are selected using the 70% b-tagging efficiency working point ($MV2c10 > 0.83$), identical to the $VH(\to b\bar{b})$ analysis.

As described in Section 4.4.1, the per-jet c-tagging requirement is defined as:

$$DL1_{c-\mathrm{tag}} = \frac{p(c)}{f_b p(b) + (1 - f_b) p(l)} > c, \tag{5.2}$$



(a) 3rd jet not b-tagged and not c-tagged

(b) 3rd jet c-tagged

(c) 3rd jet b-tagged

Figure 5.8: Illustration of the flavour tagging categorisation for events with 3 jets. Events with 2 b-tags are depicted in red, 2 c-tags in dark green, 1 c-tag in light green and 0 c-tag in blue.

which depends on two parameters: the fraction f_b and the operating point c. To optimise the c-tagging requirement, a two dimensional grid scan is performed on both parameters. The tagger, defined by these parameters, is then used to categorise $VH(\rightarrow c\bar{c})$ events in 1 and 2 c-tags categories. The sensitivity is estimated using the $m(c\bar{c})$ distribution and combined for both categories to obtain the total sensitivity to the $VH(\rightarrow c\bar{c})$ process.

The working point is optimised across the three channels and the sensitivity is evaluated with 2 figures of merit: 1) the expected limit on the $VH(\rightarrow c\bar{c})$ signal strength at 95% confidence level (CLs) for the 0- and 1-lepton channel, and 2) the significance using the signal s and background b,i.e $Z = \frac{s}{\sqrt{s+b}}$, for the 2-lepton channel. The result of the optimisation is shown in Fig. 5.10. The optimal working point is found to be consistent across all three channels and for both figures of merit and the parameter values used for flavour tagging are presented in Table 5.12. Due to the b-tag veto, which significantly reduces the abundance of b-jets, the best sensitivity to the $VH(\rightarrow c\bar{c})$ process is obtained when the c-tagger has a large discrimination power between c-jets and light-jets and, therefore, the optimal value for the fraction is low ($f_b = 8\%$). The performance and the calibration of the optimised c-tagging working point is discussed in Section 4.4.

c-tag + b-tag veto working point	$DL1 \geq 1.3$ $\quad f_b = 8\%$

Table 5.12: Parameter values of the flavour tagging working point obtained from the optimisation of the sensitivity to the $VH(\rightarrow c\bar{c})$ process.

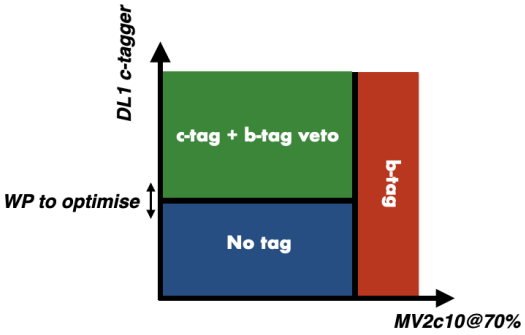
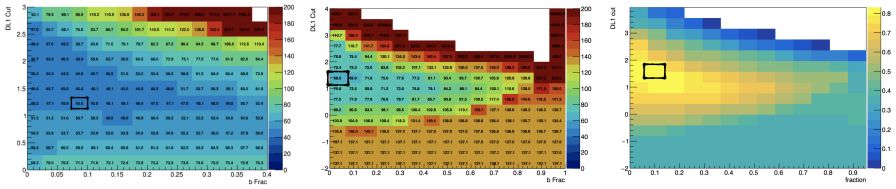


Figure 5.9: Per jet flavour tagging strategy. Jets are labelled b-tagged if they pass the 70% b-tagging efficiency working point, as defined by the $VH(\to b\bar{b})$ analysis. The c-tagging working point is subsequently optimised to achieve the best sensitivity to the $VH(\to c\bar{c})$ process.



(a) 0-lepton (b) 1-lepton (c) 2-lepton

Figure 5.10: Working point optimisation of the c-tagging requirement on the $VH(\to c\bar{c})$ analysis. The black rectangle indicates the optimal value across all three channels.

5.8 Truth Tagging

Due to the modest c-tagging efficiency, a c-tagging requirement on multiple jets in the same event results in small event selection probabilities. In events with 2 c-jets in the final state, such as the $VH(\to c\bar{c})$, $VZ(\to c\bar{c})$ and V+2 c-jets processes, the probability to enter the 1 and 2 c-tags categories is respectively about 40% and 5%. This results in a loss of 60% and 95% of the simulated events, respectively, when applying the c-tagging requirement. In addition, due to the low tagging efficiency on b-jets and light-jets, this effect is even larger for simulated samples with non-c-jets in the final state. The result is a larger statistical uncertainty of the simulated sample, which degrades the precision of the measurement.

In order to mitigate the effect of the c-tagging efficiency, the truth tagging

technique, described in Section 4.4.4, is applied on the simulated events. For an event with 2 jets, the weight is calculated using per-jet tagging efficiencies, calculated as a function of the p_{T} and η of the jet. The probabilities for an event to be selected in the 1 c-tag and 2 c-tags categories is defined as follow:

$$TT^{2c-tag}_{weight} = \epsilon_1 \times \epsilon_2, \tag{5.3}$$

$$TT^{1c-tag}_{weight} = \epsilon_1 \times (1 - \epsilon_2) + (1 - \epsilon_1) \times \epsilon_2, \tag{5.4}$$

where ϵ_1 and ϵ_2 are the tagging efficiencies for the first and second jet. For events with ≥ 3 jets, the probabilities to be selected in the 1 and 2 c-tag categories is calculated based on the 2 highest-p_{T} jets in the event and the additional b-tag veto is applied as a requirement on the b-tagging discriminant.

One of the main assumption of truth tagging is that the tagging efficiency can be calculated on a per-jet basis and the tagging probability of both jets is uncorrelated. However, in events where two jets are geometrically close to each other, part of the constituents of one jet can enter in the second jet and impact the reconstruction of both jets. This affects direct tagging but is not accounted for in truth tagging, where the efficiency is parametrised on a per-jet basis. This results in a disagreement between the direct and truth tagged distributions, which is shown in Fig 5.11. The agreement between direct and truth tagging is compared as a function of ΔR between the tagged jet and the closest jet in the event: the largest effect is observed on the tagging of c-jets with a nearby b- or c-jet.

In order to account for the difference between direct and truth tagging at small ΔR, an additional correction is derived from a 2^{nd} order polynomial fit to the ratio between direct and truth tagged events. This correction depends on the ΔR between two jets, as well as their flavours, and is applied on a per-jet basis as a correction to the tagging efficiency. The final weight from Eq. (5.3) and (5.4) becomes:

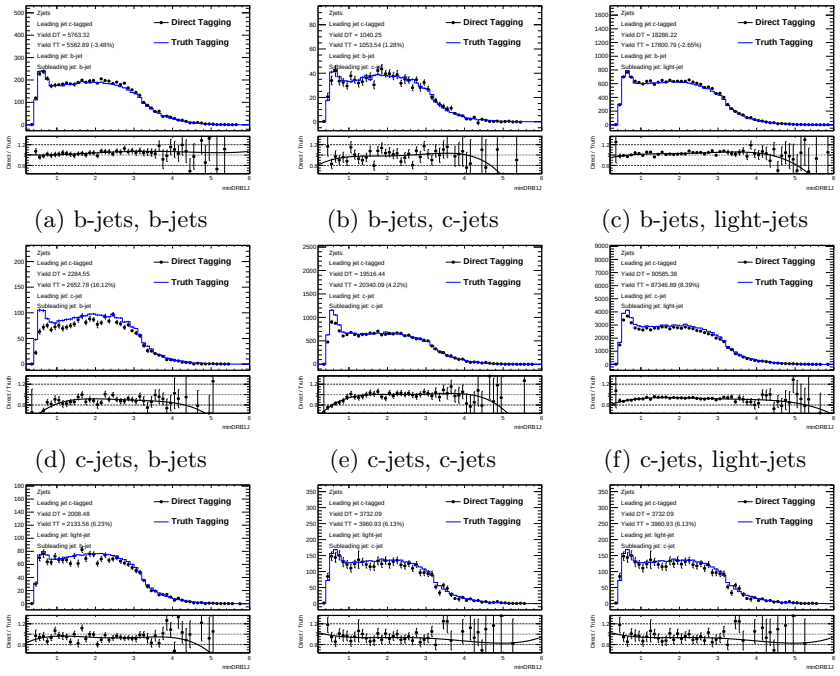
$$TT^{2c-tag}_{weight} = c_1\epsilon_1 \times c_2\epsilon_2, \tag{5.5}$$

$$TT^{1c-tag}_{weight} = c_1\epsilon_1 \times (1 - c_2\epsilon_2) + (1 - c_1\epsilon_1) \times c_2\epsilon_2, \tag{5.6}$$

where c_1 and c_2 are the ΔR dependent per-jet corrections.

The truth tag methodology is validated by comparing the direct and truth tagged $m(c\bar{c})$ distributions. The comparison is shown in Fig 5.12, before and after the ΔR correction, where the effect of the correction is mostly visible at low $m(c\bar{c})$. After the correction, the truth tagged distribution is consistent with the direct tagged distribution.

Furthermore, in order to test the agreement between the direct and truth tagged simulation samples, an additional validation procedure has been developed in the context of the $VH(\rightarrow c\bar{c})$ analysis. The validation consists of replacing the measured data set with the prediction obtained from the direct tagged distributions



(a) b-jets, b-jets (b) b-jets, c-jets (c) b-jets, light-jets

(d) c-jets, b-jets (e) c-jets, c-jets (f) c-jets, light-jets

(g) light-jets, b-jets (h) light-jets, c-jets (i) light-jets, light-jets

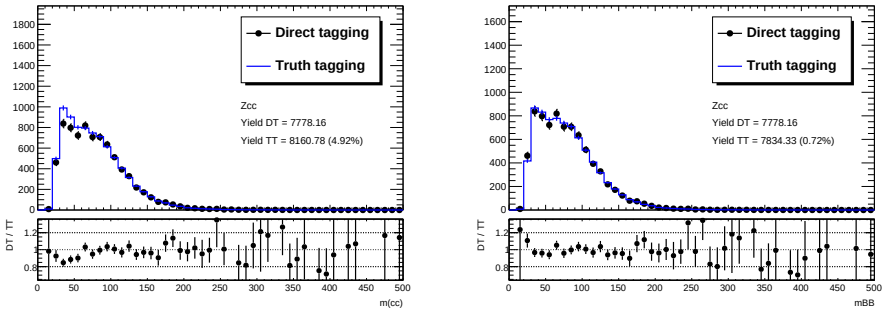
Figure 5.11: Direct and truth tagging compared as a function of ΔR between the tagged jet and the closest jet. In the caption, the left label is the flavour of the tagged jet and the right the flavour of the closest jet.

and performing a fit to this pseudo-data set using the truth tagged template. From this procedure, both methodologies are found to be consistent and no bias on the $VH(\to c\bar{c})$ signal strength is induced by the truth tagging method.

The truth tagging method is applied on all the simulated background samples and is not used for the signal samples. As the selection is optimised to select events from the $VH(\to c\bar{c})$ signal, the available simulated sample in the 1 and 2 c-tags categories after tagging is significantly larger than the background samples and less subject to statistical fluctuations. Altogether, the improvement on the final sensitivity obtained from the usage of truth tagging is 10%.

5.9 Jet energy corrections

In the $VH(\to c\bar{c})$ search, the invariant mass $m(c\bar{c})$ is estimated from 2 jets and depends on the reconstruction of the p_{T} of both jets. For a two-body decay, such



(a) Before correction (b) After correction

Figure 5.12: Direct and truth tagging comparison as a function of the di-jet invariant mass m($c\bar{c}$) for the $Z + cc$ background. The agreement is shown (a) before and (b) after the ΔR correction described in Eq. (5.5) and 5.6.

as the $H \to c\bar{c}$ process, the invariant mass is given by:

$$M_H = \sqrt{m_1^2 + m_2^2 + 2(E_1 E_2 - p_1 p_2 \cos\alpha)}, \tag{5.7}$$

where E is the energy, p the norm of the momentum and α the angle between the jets. The signal has a broad peak in the $m(c\bar{c})$ distribution on top of an overall smooth non-resonant background shape. Improvements of the jet reconstruction can directly improve the resolution of the $m(c\bar{c})$ signal distribution, which results in an increased sensitivity to the $VH(\to c\bar{c})$ signal.

Due to the decay chains of c-hadrons, some decay particles deposit their energy in the calorimeter beyond the cone radius of the jet reconstruction algorithm, which results in a lower reconstructed jet p_{T}. This affects any decay of c-hadrons and is mostly present at low jet p_{T}, where the decay products are less collimated. In addition, in the case of semi-leptonic decays of c-hadrons, the muons and neutrinos deposit little or no energy in the calorimeters and are therefore not accounted for in the energy reconstruction. In order to partially correct these energy losses and thereby improve the $m(c\bar{c})$ estimate, a muon-in-jet correction is applied in all three channels of the analysis. In the specific case of the 2-lepton channel, a kinematic fit correction is studied, which exploits the reconstruction of the Z boson decaying to leptons.

Muon-in-jet correction

Approximately 5% of c-jets contain semi-leptonic decays with a muon in the decay chain, which deposit little energy in the calorimeter. To correct for this loss of

energy, the muon four-momentum is added to the jet four-momentum:

$$\begin{pmatrix} E_{\text{jet,corrected}} \\ \vec{p}_{\text{jet,corrected}} \end{pmatrix} = \begin{pmatrix} E_{\text{muon-in-jet}} \\ \vec{p}_{\text{muon-in-jet}} \end{pmatrix} + \begin{pmatrix} E_{\text{jet}} \\ \vec{p}_{\text{jet}} \end{pmatrix}. \tag{5.8}$$

If more than one muon is found within the jet cone, only the muon closest to the jet axis is considered for the jet energy correction. The muon-in-jet selection criteria are:

- $p_{\text{T}} > 4$ GeV, $|\eta| < 2.7$
- $\Delta R(\text{jet}, \text{muon}) < \min(0.4, 0.04 + 10/p_{\text{T}}^{\mu})$.

The muon-in-jet correction is applied in the 0-, 1- and 2-lepton channels. The $m(c\bar{c})$ distribution after the correction is shown in Fig. 5.15 and the improvement on the $VH(\to c\bar{c})$ signal resolution is up to 6%.

Kinematic fit

Kinematic fitting is a mathematical procedure in which the physical properties of particle interactions are exploited to improve the measurement of a specific process. The physical information is provided in the form of a constraint that the process must satisfy. In the $VH(\to c\bar{c})$ analysis, the kinematic fit (KF) is used to improve the VH, VW and VZ measurements by exploiting the constraint that the total momentum of the system is conserved:

$$\vec{p}_V = \vec{p}_{H/W/Z}. \tag{5.9}$$

In the 0-lepton and 1-lepton channels, due to the presence of neutrinos, the conservation of the momentum is used to calculate the $E_{\text{T}}^{\text{miss}}$ and, therefore, the main constraint used in the KF is redundant and the method yields no significant improvement. In the 2-lepton channel, however, as all the final state objects are reconstructed, the KF is used to correct for the jet energy, which in turn improves the resolution of the $m(c\bar{c})$ distribution of the signal.

The KF is illustrated in Fig. 5.13 for signal $VH(\to c\bar{c})$ events of the 2-lepton channel. One advantage is the lepton energy resolution of $\mathcal{O}(1\%)$, which is more precise than the jet energy resolution $\mathcal{O}(10\%)$ and both are used as constraints. Furthermore, as the leptons originate from the decay of a Z boson, an additional constraint is provided by requiring the $m(l\bar{l})$ distribution to be compatible with the Z mass.

To improve the reconstruction, a fit is performed on each event with the parameters, listed in Table 5.13, which consist of the p_{T} of the jets from the Higgs candidate, the transverse momentum of the ZH system and the mass of the Z boson. For events with 3 jets, the p_{T} of the additional jet is also considered as a fit parameter. All parameters are estimated in a fit using a negative log-likelihood.

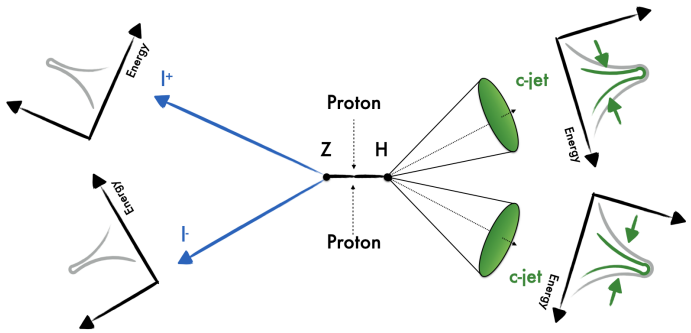


Figure 5.13: Illustration of the $VH(\to c\bar{c})$ kinematic fit: the energy reconstruction of the jets produced in the Higgs boson decay is improved using the better precision of the lepton energy reconstruction.

Parameter	Reconstructed	Kinematic fit	Constraint
Leading-jet p_{T} Subleading-jet p_{T} 3$^{\mathrm{rd}}$ jet p_{T}	$p_{\mathrm{T}}^{\mathrm{Jet}}$	$p_{\mathrm{T}}^{\mathrm{Fit,Jet}}$	$\sigma_{p_{\mathrm{T}}^{\mathrm{Jet}}}$
Lepton l^+ p_{T} Lepton l^- p_{T}	$p_{\mathrm{T}}^{\mathrm{Lepton}}$	$p_{\mathrm{T}}^{\mathrm{Fit,Lepton}}$	$\sigma_{p_{\mathrm{T}}^{\mathrm{Lepton}}}$
ZH transverse momentum	p_{X}^{ZH}, p_{Y}^{ZH}	$p_{\mathrm{X}}^{\mathrm{Fit},ZH}$, $p_{\mathrm{Y}}^{\mathrm{Fit},ZH}$	$\sigma_{p_{\mathrm{X}}^{ZH}}$, $\sigma_{p_{\mathrm{Y}}^{ZH}}$
Z boson mass	$m(l^+l^-)$	m_Z	Γ_Z

Table 5.13: Parameters used in the kinematic fit.

Each constraint is parametrised in the negative log-likelihood:

$$-2\ln L = \frac{\left(p_{\mathrm{X}}^{ZH} - p_{\mathrm{X}}^{\mathrm{Fit},ZH}\right)^2}{\sigma^2_{p_{\mathrm{X}}^{ZH}}} + \frac{\left(p_{\mathrm{Y}}^{ZH} - p_{\mathrm{Y}}^{\mathrm{Fit},ZH}\right)^2}{\sigma^2_{p_{\mathrm{Y}}^{ZH}}} + 2\ln\left((m_{ll}^2 - m_Z^2)^2 + m_Z^2\Gamma_Z^2\right) + \sum_{\text{leptons}} \frac{\left(p_{\mathrm{T}}^{\mathrm{Lepton}} - p_{\mathrm{T}}^{\mathrm{Fit,Lepton}}\right)^2}{\sigma_{p_{\mathrm{T}}^2}} - 2\sum_{\text{jets}}(\ln L_{\text{truth}} + \ln L_{\text{TF}}), \tag{5.10}$$

where the first two terms correspond to the constraint on the momentum of the ZH system, the third term the constraint on the leptons produced in the decay of a Z boson (which is parametrised as a Breit-Wigner distribution with $m_Z = 91.1876$

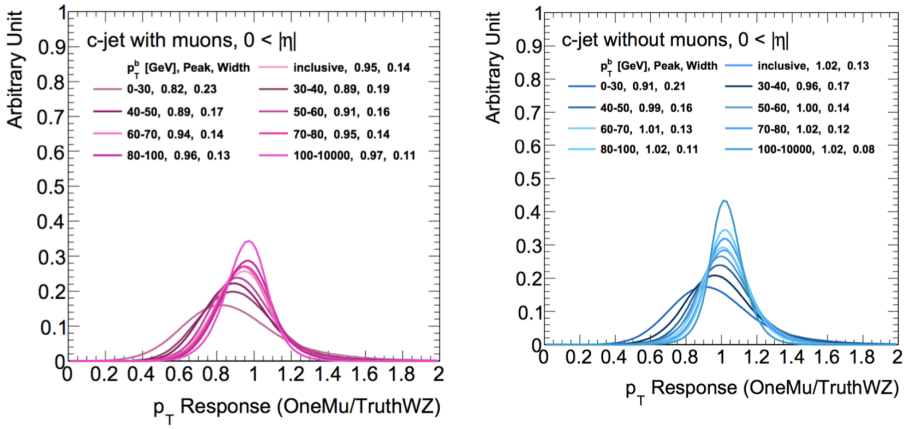
GeV) and $\Gamma_Z = 2.4952$ GeV), the fourth term the constraint on the lepton energy and the last term corresponds to the constraint on the jets.

Due to energy losses affecting the jets in the event, the likelihood for the jets is built from the true jet p_{T}:

$$-2\ln L_{\mathrm{jets}} = -2\sum_{\mathrm{jets}}(\ln L_{\mathrm{truth}} + \ln L_{\mathrm{TF}}), \tag{5.11}$$

where L_{truth} corresponds to the jet p_{T} likelihood, which is derived from a fit to the true simulated jet p_{T} to estimate the probability density function, and L_{TF} corresponds to the likelihood derived from a transfer function (TF), which estimates the detector response between the true simulated $p_{\mathrm{T}}^{\mathrm{True}}$ and the reconstructed $p_{\mathrm{T}}^{\mathrm{Reco}}$.

The TF, shown in Fig. 5.14, are derived from the detector response parametrised as the ratio of reconstructed momentum $\frac{p_{\mathrm{T}}^{\mathrm{Reco}}}{p_{\mathrm{T}}^{\mathrm{True}}}$ and a fit is performed to the distribution to estimate the probability density function that is used in the likelihood L_{TF}. In order to account for the energy losses due to the c-hadrons semi-leptonic decays, which are different from the energy loss of the fully hadronic decays, two sets of TF are derived depending on the presence or absence of a muon in the jet.



(a) Semi-leptonic

(b) Fully hadronic

Figure 5.14: Transfer functions derived on c-jets using the simulation sample $ZH(\to c\bar{c})$, (a) with muons in jets, (b) without muons in jets.

The improvement obtained using the KF is evaluated on the $m(c\bar{c})$ distribution. The standard jet calibration, which does not include a dedicated jet energy correction, is used as a reference to compare the resolution obtained with the muon-in-jet and KF corrections. The invariant mass distribution is fitted using a Bukin function and the resolution σ is chosen as the figure of merit.

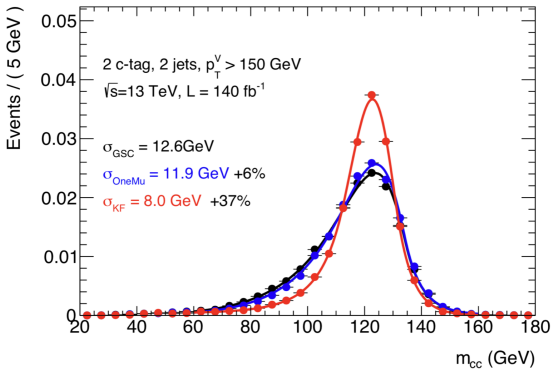


Figure 5.15: Invariant mass distribution of the Higgs candidate in the signal sample for no jet energy corrections (black), muon-in-jet (blue) and KF (red) corrections.

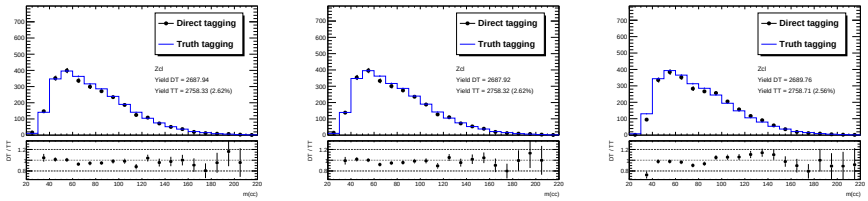
The resulting $m(c\bar{c})$ distribution obtained with the different jet energy corrections is shown on Figure 5.15. The relative improvement of the resolution for *KF* corrections is 37%.

Truth tagging and jet energy corrections

Jet energy corrections induce migration of events in the $m(c\bar{c})$ distribution. The effect on the $m(c\bar{c})$ of the backgrounds is expected to be randomly distributed and events migrate to lower or higher $m(c\bar{c})$ bins. In the truth tagging method, the weights are calculated from the jets without energy corrections, to be consistent with the flavour tagging calibrations. In order to ensure that jet energy corrections and the truth tagging method can be simultaneously applied on the events, the agreement between the direct and the truth tagged distributions is evaluated after applying the jet energy corrections.

The comparison between the truth and direct tagged $m(c\bar{c})$ distribution is shown in Fig. 5.16 for the events with no jet energy corrections, the muon-in-jet and KF corrections for the Zcl background in the 1 c-tag category of the 2-lepton channel. The direct and truth tagged distributions are in agreement for events with no energy corrections and muon-in-jet corrections. When the KF correction is applied, a difference is observed in the $100 < m(c\bar{c}) < 140$ GeV region. This discrepancy with respect to the direct tagged distribution is around the Higgs boson mass and can induce a bias in the $VH(\rightarrow c\bar{c})$ signal strength estimated from the fit to the data.

Even though the KF correction significantly improves the sensitivity of the 2-lepton channel, the final improvement on the $VH(\rightarrow c\bar{c})$ sensitivity, which is



(a) No jet energy correction (b) Muon-in-jet correction (c) Kinematic fit correction

Figure 5.16: Truth and direct tagging comparison in 2-lepton for three different jet energy corrections: no jet energy correction (a), muon-in-jet (b) and kinematic fit (c) corrections.

obtained from the combination of the 0-lepton, 1-lepton and 2-lepton channels, is of the order of $\mathcal{O}(5\%)$. This improvement is less than the improvement obtained through the usage of truth tagging in all three channels and, in order to avoid a bias, it is decided to not apply the KF correction in the 2-lepton channel. One promising solution to improve the agreement between the direct and truth tagged distribution is to parametrise the tagging efficiency using a neural network [78], which is trained on a per-event basis and is capable of taking into account the correlations between the jets in the events.

5.10 Signal and control regions

The Higgs boson is expected to peak at 125 GeV and sensitivity to the signal strength of the $VH(\to c\bar{c})$ process is achieved by using the $m(c\bar{c})$ distribution, which discriminates between the signal and the non-resonant background processes. The sensitivity to the $VH(\to c\bar{c})$ process is further improved by exploiting physics properties with discrimination power between the signal and the background: the jet multiplicity and the p_{T}^{V} distribution, which are used to categorise the events.

As the $VH(\to c\bar{c})$ signal is primarily produced with 2 jets in the final state, a first set of events is selected in a category with 2 jets. In the 0- and 1-lepton channels, an additional category of events with 3 jets is also considered. Due to the larger $t\bar{t}$ background contamination in these channels, events with ≥ 4 jets are discarded, as this category does not significantly improve the sensitivity in the 0- and 1-lepton channels. For the 2-lepton channel, where the $t\bar{t}$ contamination is lower, events are selected in an additional category with ≥ 3 jets. The sensitivity improvement obtained from the categorisation based on the jet multiplicity is $\approx 3\%$.

In the VH production mode, due to the production cross-section, the discrimination between signal and background increases at high momentum of the Higgs boson. In order to maintain high sensitivity and reduce the number of background

events, events are selected with $p_{\mathrm{T}}^{V} > 150$ GeV across all three channels. In addition, in the 2-lepton channel, which is subject to less background from $t\bar{t}$ and the QCD multijet processes, an additional category is added for events with $75 < p_{\mathrm{T}}^{V} < 150$ GeV. The sensitivity improvement obtained from this additional p_{T}^{V} category in the 2-lepton channel is $\approx 5\%$ on the total sensitivity of the three channels combined.

An additional split in the p_{T}^{V} categorisation at $p_{\mathrm{T}}^{V} = 250$ GeV is studied and the potential improvement on the sensitivity is up to 5%. However, this results in analysis categories with a small amount of data, which creates instabilities in the fit to the data. Since the addition of multiple categories significantly increases the complexity of the fit model in the treatment of systematic uncertainties, it is decided to simplify the fit model and keep the $p_{\mathrm{T}}^{V} \geq 150$ GeV merged category.

Altogether, events passing this selection are categorised in number of c-tags, jets and p_{T}^{V} to define the signal region (SR) of the $VH(\rightarrow c\bar{c})$ analysis. A total of 16 SRs, shown in Table 5.14, are combined to improve the sensitivity to the $VH(\rightarrow c\bar{c})$ process.

SR: 16 categories			
0-lepton	$p_{\mathrm{T}}^{V} > 150$ GeV	1 c-tag, 2 jets 1 c-tag, 3 jets	2 c-tags, 2 jets 2 c-tags, 3 jets
1-lepton	$p_{\mathrm{T}}^{V} > 150$ GeV	1 c-tag, 2 jets 1 c-tag, 3 jets	2 c-tags, 2 jets 2 c-tags, 3 jets
2-lepton	$75 < p_{\mathrm{T}}^{V} < 150$ GeV	1 c-tag, 2 jets	2 c-tags, 2 jets
	$p_{\mathrm{T}}^{V} > 150$ GeV	1 c-tag, ≥ 3 jets	2 c-tag, ≥ 3 jets

Table 5.14: Categorisation of the signal region for the 0-, 1- and 2-lepton channels.

5.10.1 Comparison of data and simulation

In order to estimate the expected contributions of the signal and background processes in the $m(c\bar{c})$ distribution, all the simulated processes are compared to the data and to each other. Resonant processes such as the $VH(\rightarrow c\bar{c})$ signal and the $VW(\rightarrow cq)$ and $VZ(\rightarrow c\bar{c})$ processes peak in the $m(c\bar{c})$ distribution around the mass of the respective boson. For the non-resonant V+jets background, the distribution is a continuous spectrum decreasing towards higher $m(c\bar{c})$ values. To describe the various production mechanisms contributing to the V+jets background, the events are grouped in three different production components according to the flavour of the jets: V+heavy flavours (Vhf), V+mixed flavours (Vmf) and V+light flavours (Vl).

For the $t\bar{t}$ background, the distribution is composed of a resonant and a non-resonant contribution, depending on the flavour of the two leading jets used to estimate the $m(c\bar{c})$ distribution. The resonant contribution is composed of jets originating from a W boson in the $t\bar{t}$ decay and peaks around the W boson mass. The non-resonant contribution is composed of a b-jet produced in the decay of one of the top quarks, together with another jet in the event, and the $m(c\bar{c})$ distribution is constant in the range $50 < m(c\bar{c}) < 200$ GeV.

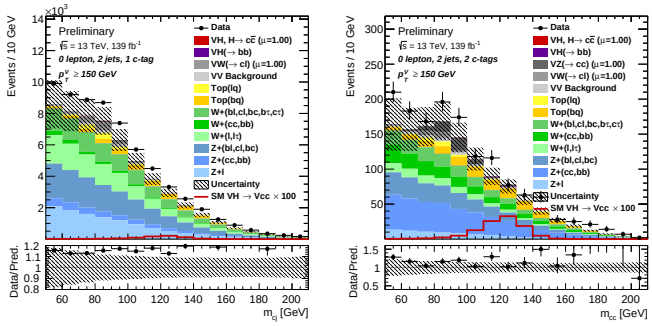
Label	Description	Process
Zhf	Z+heavy flavours	$Z+(c\bar{c}, b\bar{b})$
Zmf	Z+mixed flavours	$Z+(cl, bl, bc)$
Zl	Z+light flavours	$Z+(ll)$
Whf	W+heavy flavours	$W+(c\bar{c}, b\bar{b})$
Wmf	W+mixed flavours	$W+(cl, bl, bc, b\tau, c\tau)$
Wl	W+light flavours	$W+(ll, l\tau)$
Top(other)	W decay from $t\bar{t}$ background	$W(\to cs, cd, ud, us)$
Top(b)	b-jet + other jet in event from $t\bar{t}$ decays	-
$t\bar{t}$	$t\bar{t}$ process used in 2-lepton	-
VV background	VV decays without c-jets	$VV(\to b\bar{b}, ll, bl)$
$VH(\to b\bar{b})$	$VH(\to b\bar{b})$ process	$VH(\to b\bar{b})$

Table 5.15: Background processes separated by flavours of the 2 jets used to estimate the $m(c\bar{c})$ distribution.

The categorisation and the event selection are different for the 0-, 1- and 2-lepton channels and the background processes, summarised in Table 5.15, vary in abundances in the three channels.

0-lepton As the events in the 0-lepton channel are selected based on the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event, all the main background processes result in events with zero charged leptons or events where one of the leptons is not reconstructed in the acceptance of the detector. The $m(c\bar{c})$ distribution, shown in Fig. 5.17 for the categories with $p_{\mathrm{T}}^{V} > 150$ GeV and 2 jets, is mainly composed of Z+jets, W+jets and $t\bar{t}$. The simulated background is found to underestimate the data by $\approx 20\%$, with an increasing underestimation at higher values of $m(c\bar{c})$. The $VH(\to c\bar{c})$ signal peaks at $m(c\bar{c}) = 125$ GeV, with a resolution of approximately 10 GeV.

The proportion of the background components varies in the different categories of the SR. For the background composition, shown in Fig. 5.18 for all the SRs of the 0-lepton channel, the dominant processes in the 1 c-tag category are the Zmf, Zl, Wmf and Wl backgrounds. The category with 2 c-tags is subject to a larger contamination from the Zhf and Whf backgrounds. In the 3 jets category, the relative contribution of the $t\bar{t}$ processes is doubled with respect to the 2 jets



(a) 1 c-tag (b) 2 c-tag

Figure 5.17: Pre-fit data and simulation comparison for the 0-lepton channel in the categories with 2 jets and $p_{\mathrm{T}}^{V} > 150$ GeV: (a) 1 c-tag and (b) 2 c-tag.

category. For the $VW(\to cq)$ and $VZ(\to c\bar{c})$ processes, the largest fraction is found in the 1 and 2 c-tags categories, respectively.

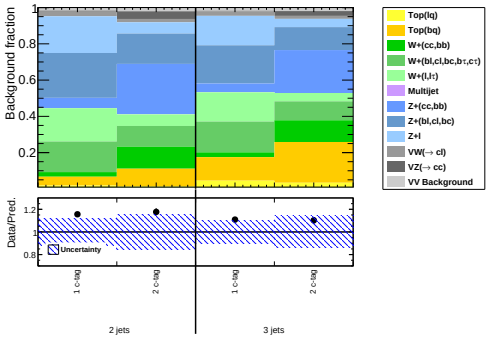
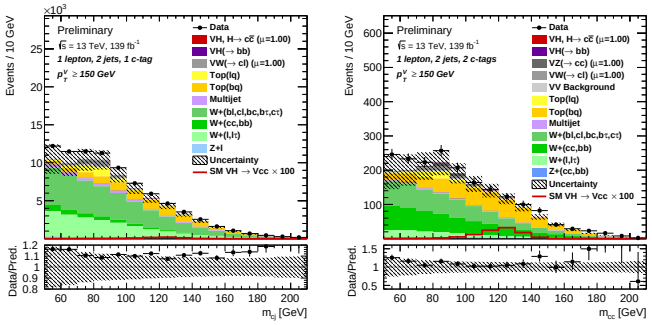


Figure 5.18: Background composition of the 0-lepton channel.

1-lepton In the 1-lepton channel, the main background contributions are the W+jets and $t\bar{t}$ processes. In addition, the 1-lepton channel is subject to a higher contamination from the QCD multijets process with respect to the 0- and 2-lepton channels. Since the simulation sample is small, this background is estimated with a data-driven method, where the lepton isolation selections are inverted to define a region enriched in events produced by the QCD multijets process. To estimate the normalisation, a fit is performed in this region, which is then extrapolated to the SR of the 1-lepton channel. The data and the prediction are compared for the

$m(c\bar{c})$ distribution, shown in Fig. 5.19. The prediction underestimates the data by 10%, with an increasing difference at higher values of $m(c\bar{c})$.



(a) 1 c-tag (b) 2 c-tag

Figure 5.19: Data and simulation comparison for the 1-lepton channel in the categories with 2 jets and $p_{\mathrm{T}}^{V} > 150$ GeV: (a) 1 c-tag and (b) 2 c-tag.

For the 4 SRs of the 1-lepton channel, the background composition, shown in Fig. 5.20, consists of the Wmf and Wl processes for both 1 and 2 c-tags categories and, in addition, the Whf process for the 2 c-tags category. In events with 3 jets, the dominant background is the $t\bar{t}$ process.

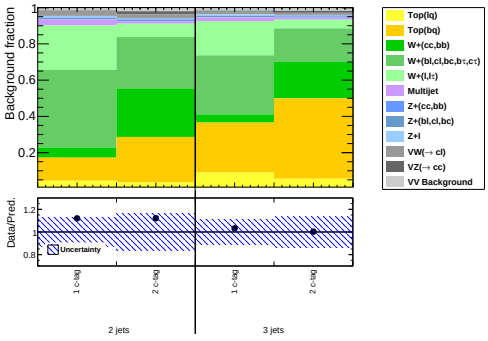
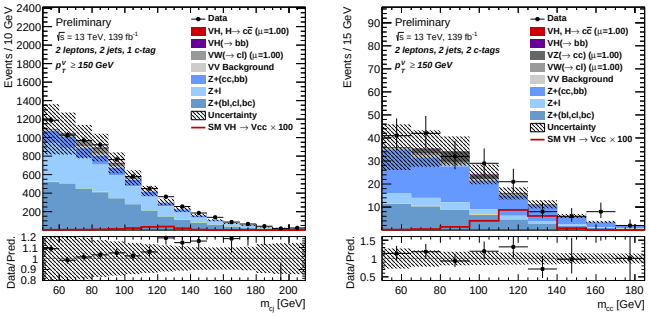


Figure 5.20: Background composition of the 1-lepton channel.

2-lepton The dominant background contribution in the 2-lepton channel is the Z+jets process, which results in the least diverse background composition of all the three channels. In the $m(c\bar{c})$ distribution, shown in Fig. 5.21, the prediction underestimates the data by $\approx 10\%$ with an increasing difference at high values of $m(c\bar{c})$.



(a) 1 c-tag (b) 2 c-tag

Figure 5.21: Data and simulation comparison for the 2-lepton channel in the categories with 2 jets and $p_{\mathrm{T}}^{V} > 150$ GeV: (a) 1 c-tag and (b) 2 c-tag.

For the 8 SRs of the 2-lepton channel, the background composition, shown in Fig. 5.22, is dominated by the Zmf process in the 1 c-tag category and the Zhf process in the 2 c-tags category. In the combined analysis of the 0-, 1- and 2-lepton channels, these categories with a large amount of Zmf and Zhf events are used to constrain the Z+jets background in the 0-lepton channel, providing overall a better control over one of the main background process of the analysis.

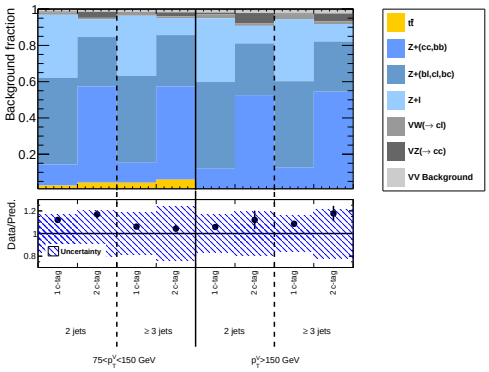


Figure 5.22: Background composition of the 2-lepton channel.

Sensitivity estimate

Due to the different background compositions of the $VH(\rightarrow c\bar{c})$ process in the three channels, as well as the different production cross-sections, the sensitivity among the categories varies. For each SR category of the $VH(\rightarrow c\bar{c})$ analysis,

the significance $\frac{s}{\sqrt{b}}$ is calculated and shown in Fig. 5.23. Despite the complex background composition, the best sensitivity is obtained in the 2 jets categories of the 0-lepton channel. Similarly in the 1-lepton channel, the categories with 2 jets are the most sensitive. In the 2-lepton channel, the category with $p_{\mathrm{T}}^{V} > 150$ GeV is generally more sensitive than the category with $75 < p_{\mathrm{T}}^{V} < 150$ GeV, while the sensitivity decreases for larger jet multiplicity.

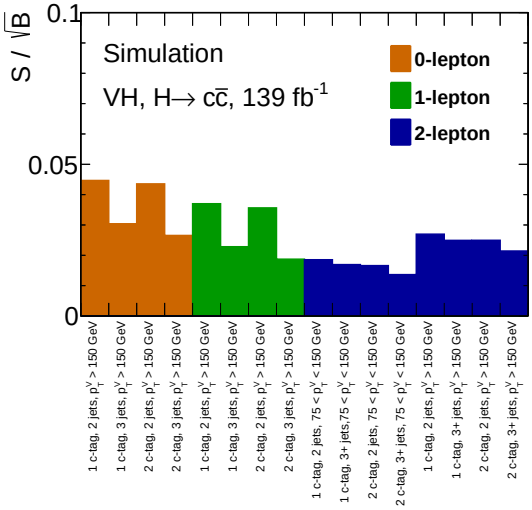


Figure 5.23: Sensitivity in each SR of the $VH(\to c\bar{c})$ analysis estimated with the significance $z = s/\sqrt{b}$ for the 0-, 1- and 2-lepton channels.

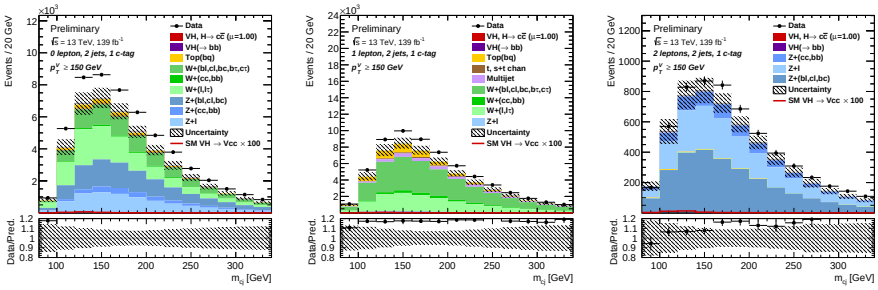
After combining the significance of all categories in each channel, the 0-lepton channel is the most sensitive, followed by the 2-lepton and the 1-lepton channels. Compared to the sensitivity of the 0-lepton channel, the combination of the 0-, 1- and 2-lepton channel is 50% more sensitive.

5.10.2 Control region definition

A control region (CR) is a data sample enriched in a background process and with a small contribution from the signal process. Together with the SRs, these CRs are used in the fit to provide constraints on these background directly from the data. To make a CR and avoid overlap with the SRs, the SR event selection is modified to ensure that the CR is statistically independent to the SR. For the $VH(\to c\bar{c})$ analysis, the main backgrounds are Vhf, Vmf, Vl and $t\bar{t}$ and therefore events are selected in dedicated CRs enriched in these backgrounds.

High ΔR CR Since the SR selections require small values of the ΔR between the jets and in order to constrain the Z+jets and W+jets background processes, a CR is defined by selecting events with higher values than the the ΔR selection used for the SR:

- 75 GeV $< p_{\mathrm{T}}^{V} <$ 150 GeV: $2.3 < \Delta R(\mathrm{jet1}, \mathrm{jet2}) < 2.5$
- 150 GeV $< p_{\mathrm{T}}^{V} <$ 250 GeV: $1.6 < \Delta R(\mathrm{jet1}, \mathrm{jet2}) < 2.5$
- $p_{\mathrm{T}}^{V} >$ 250 GeV: $1.2 < \Delta R(\mathrm{jet1}, \mathrm{jet2}) < 2.5$.

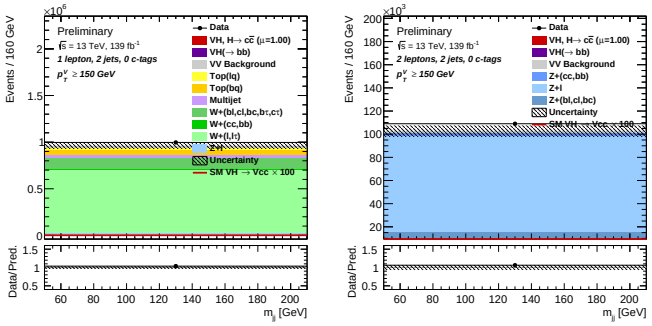


(a) 0-lepton (b) 1-lepton (c) 2-lepton

Figure 5.24: Example of the high ΔR CR for the category with 2 jets and $p_{\mathrm{T}}^{V} >$ 150 GeV.

The $m(c\bar{c})$ distribution, shown in Fig. 5.24, is estimated for events passing the high ΔR selection, using the same categorisation as for the SRs to define 16 CRs. As the ΔR selection in the SR is optimised to increase the signal sensitivity, the signal efficiency is low in the CR where the main Z+jets and W+jets background processes are constrained from the data. In addition, for the 0- and 1-lepton channels, the $t\bar{t}$ contamination is lower than in the SR. Finally, one of the main advantages of the high ΔR CR is that, due to the correlation of ΔR with $m(c\bar{c})$, the CR provides a better control over the high $m(c\bar{c})$ distribution, where the difference between the data and the prediction is larger.

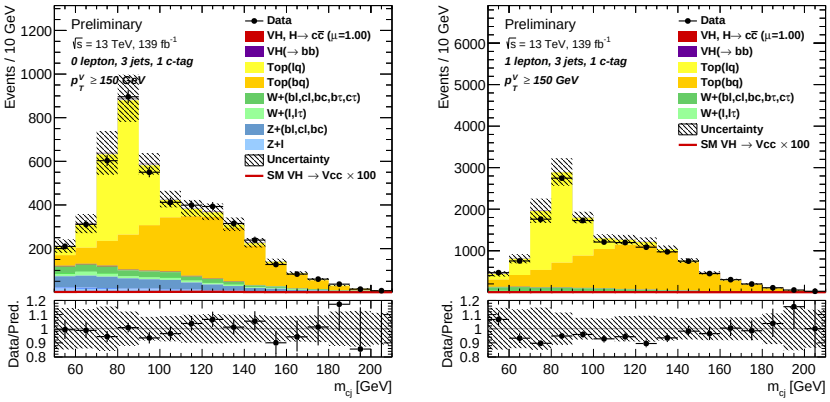
0 c-tag CR The 1 c-tag category of the SR is subject to a large contamination from the Zl and Wl processes. In order to provide a better constraint on these two processes, a selection with 0 c-tags, where both jets that are used to estimate the $m(c\bar{c})$ distribution fail the c-tagging requirement, is used to define a CR. Due to the size of the data sample in the 0 c-tag category, the $m(c\bar{c})$ distribution, shown in Fig. 5.25, is used as a single-bin CR to constrain merely the normalisation of the Zl and Wl background processes. This CR is only defined for the 1- and 2-lepton channels, which are purer in Wl and Zl with respect to the 0-lepton channel, and are categorised similarly to the SR, resulting in 6 additional CRs.



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

Figure 5.25: Example of the 0 c-tag control region for the categories with 2 jets and $p_{\mathrm{T}}^{V} > 150$ GeV.

Top CR One of the dominant backgrounds in the 0- and 1-lepton channels is the $t\bar{t}$ process, which consists of a resonant and a non-resonant part. In order to provide a better constraint on this background, a top CR is made using events with 3 jets and the requirement that the 3rd jet in the event is b-tagged. This selection guarantees there is no overlap between the CR and the SR, where the 3rd jet is b-vetoed. This CR is used only for the 1- and 2-lepton channels in the 1 c-tag 3 jets category, providing 2 additional CRs.



(a) 0-lepton (b) 1-lepton

Figure 5.26: Example of the top control region for the category with 3 jets and $p_{\mathrm{T}}^{V} > 150$ GeV.

The resulting $m(c\bar{c})$ distribution, shown in Fig. 5.26, is pure in $t\bar{t}$ events and the resonant and non-resonant contributions are distinguishable in the data. Since the

resonant $t\bar{t}$ background peaks at the W boson mass, it provides a better control of this distribution in the SR, where the $VW(\rightarrow c\bar{q})$ process is measured.

Top-$e\mu$ CR In order to constrain the $t\bar{t}$ contamination in the 2-lepton channel, an additional CR is made by selecting events with two leptons of a different flavour, one electron and one muon, which are typically produced in the leptonic decay of the two W bosons originating from the $t\bar{t}$ process. This CR, labelled $top - e\mu$ CR, has no overlap with the SR selection, where the two leptons are required to have the same flavour. In order to constrain the normalisation of the $t\bar{t}$ process, the $m(c\bar{c})$ distribution, shown in Fig. 5.27, is used as a single bin. Only the events selected in the 1 c-tag category are used, with similar categorisation in terms of jet multiplicity and p_{T}^{V} as the SR, resulting in 4 additional CRs.

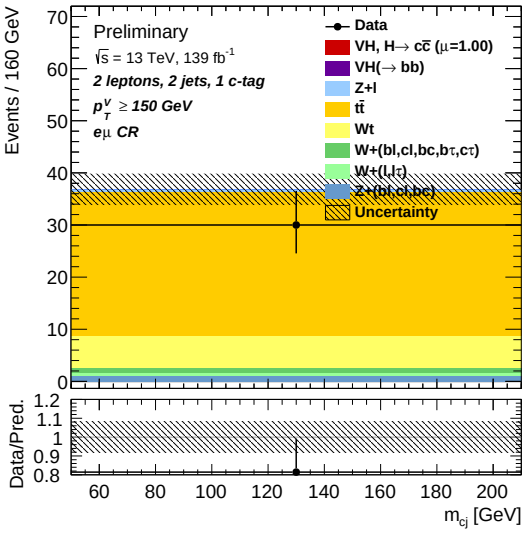


Figure 5.27: Example of the top-$e\mu$ control region for the 2-lepton channel in the category with 1 c-tag and 2 jets and $p_{\mathrm{T}}^{V} > 150$ GeV.

Summary of the CR definition In total, 28 additional CRs, summarised in Table 5.16, are used to constrain the main backgrounds in the data. These CRs are analysed together with the 16 SRs to measure the signal strength of the $VH(\rightarrow c\bar{c})$ process.

High ΔR CR: 16 categories			
0-lepton	$p_{\mathrm{T}}^{V} > 150$ GeV	1 c-tag, 2 jets 1 c-tag, 3 jets	2 c-tags, 2 jets 2 c-tags 3 jets
1-lepton	$p_{\mathrm{T}}^{V} > 150$ GeV	1 c-tag, 2 jets 1 c-tag 3 jets	2 c-tags, 2 jets 2 c-tags, 3 jets
2-lepton	$75 < p_{\mathrm{T}}^{V} < 150$ GeV	1 c-tag, 2 jets	2 c-tags, 2 jets
	$p_{\mathrm{T}}^{V} > 150$	1 c-tag, ≥ 3 jets	2 c-tags, ≥ 3 jets
0 c-tag CR: 6 categories			
1-lepton	$p_{\mathrm{T}}^{V} > 150$ GeV	0 c-tag, 2 jets 0 c-tag, 3 jets	-
2-lepton	$75 < p_{\mathrm{T}}^{V} < 150$ GeV	0 c-tag, 2 jets	-
	$p_{\mathrm{T}}^{V} > 150$	0 c-tag, ≥ 3 jets	-
Top CR: 2 categories			
0-lepton	$p_{\mathrm{T}}^{V} > 150$ GeV	1 c-tag, 3 jets	-
1-lepton	$p_{\mathrm{T}}^{V} > 150$ GeV	1 c-tag, 3 jets	-
Top-$e\mu$ CR: 4 categories			
2-lepton	$75 < p_{\mathrm{T}}^{V} < 150$ GeV	1 c-tag, 2 jets	-
	$p_{\mathrm{T}}^{V} > 150$	1 c-tag, ≥ 3 jets	-

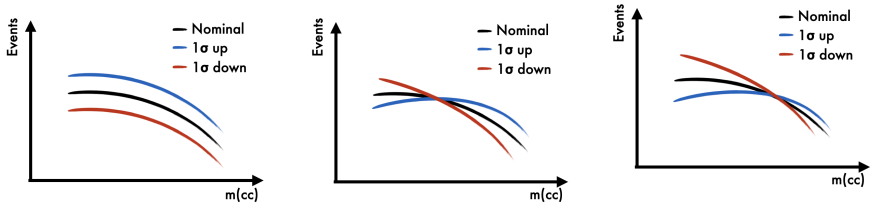
Table 5.16: Categorisation of the CRs in the 0-, 1- and 2-lepton channels.

5.11 Systematic uncertainties

The main sources of systematic uncertainties for the $VH(\to c\bar{c})$ analysis are the imperfect detection and reconstruction of the physics objects, the detector calibrations and the theory predictions of the physics processes.

Systematic uncertainties, as illustrated in Fig. 5.28, affect the $m(c\bar{c})$ distribution of the signal and the background processes in two ways: the normalisation and the shape. A normalisation uncertainty increases or decreases the number of events uniformly across all the bins of the $m(c\bar{c})$ distribution. A shape uncertainty modifies the number of events in specific values of the $m(c\bar{c})$ distribution, while maintaining the number of predicted events. Overall, this translates as a modification of the shape of the $m(c\bar{c})$ distribution. In addition, certain uncertainties modify both the normalisation and the shape of the process.

Two types of systematic uncertainties are considered in the analysis: experimental



(a) Normalisation (b) Shape (c) Normalisation and shape

Figure 5.28: Illustration of the effect of systematic uncertainties on the $m(c\bar{c})$ distribution.

uncertainties, which are due to detector effects, and modelling uncertainties, which stem from the imperfect prediction of the theory used to model both the number of events and the shape of the $m(c\bar{c})$ distribution.

5.11.1 Experimental systematic uncertainties

Experimental uncertainties are uncertainties on the reconstruction of the events. These errors quantify the precision of the reconstruction of each object as well as the total number of events. Each reconstruction is calibrated and experimental systematic uncertainties, listed in Table 5.17, are assigned on certain properties, such as the efficiency of the reconstruction and identification as well as the energy and resolution measurements. In the $VH(\rightarrow c\bar{c})$ analysis, the main uncertainties are uncertainties related to the reconstruction of jets and the flavour tagging.

Jet reconstruction uncertainties Due to the less precise reconstruction of jets with respect to the electrons and muons in the event, the dominant uncertainties of the $VH(\rightarrow c\bar{c})$ analysis are related to the jets. In particular, the scale and the resolution of the energy of the jets have a direct impact on the estimated $m(c\bar{c})$ distributions of the backgrounds and signal processes. As the signal is a resonant process in the $m(c\bar{c})$ distribution, a worse resolution of the energy results in a worse resolution of the $m(c\bar{c})$ distribution. For the non-resonant background processes, the $m(c\bar{c})$ distribution is uniformly distributed and a worse jet energy resolution does not impact the $m(c\bar{c})$ distribution significantly. As an example, these effects are shown in Fig. 5.29.

Flavour tagging uncertainties One of the major sources of uncertainty in the $VH(\rightarrow c\bar{c})$ analysis is the c-tagging, which is approximately 10% for each c-jet. These uncertainties affect the tagging efficiency and the effect is correlated between the 1 and 2 c-tags categories.

Object	Experimental uncertainty
$E_{\mathrm{T}}^{\mathrm{miss}}$ Electrons and muons	Trigger efficiency
$E_{\mathrm{T}}^{\mathrm{miss}}$	Energy scale uncertainty
$E_{\mathrm{T,Trk}}^{miss}$	Track-based soft term scale and resolution uncertainty
Electron and muon	Reconstruction, identification and isolation efficiency
	Energy scale and resolution
Muons	Track-to-vertex association efficiency
Jets	Energy scale and resolution
	b- and c-jets energy response
	Pile-up related uncertainty
	Flavour composition and response (quark / gluon induced)
	Flavour tagging efficiency
	Truth tagging ΔR correction and normalisation
τ-jets	Energy scale and resolution
	Flavour tagging efficiency
	Truth tagging ΔR correction and normalisation
Luminosity	Integrated luminosity measurement
Pile-up	Number of vertices from inelastic pp-collisions

Table 5.17: Experimental systematic uncertainties in the $VH(\to c\bar{c})$ analysis.

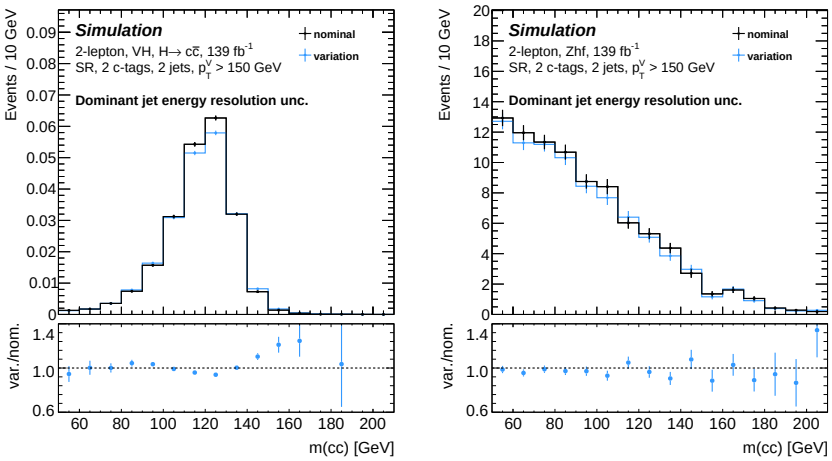
The variation of the c-tagging efficiency as a function of the p_{T} affects the $m(c\bar{c})$ distribution similarly for the $VH(\to c\bar{c})$ signal and the $Z + hf$ background processes, as illustrated in Fig 5.30. The overall effect is a variation of 20% of the number of predicted signal and background events in the category with 2 c-tagged jets.

In addition to the c-tagging uncertainties, the uncertainties related to the b-veto on the 3rd jet in the event are also accounted for. These uncertainties are smaller than the c-tagging uncertainties and are treated uncorrelated between the two tagging algorithms.

Truth tagging uncertainties One of the dominant systematic uncertainty of the $VH(\to c\bar{c})$ analysis originates from the truth tagging method. To account for the effect of the ΔR correction due to near-by jets and the residual disagreement between truth and direct tagged distributions, two additional systematic uncertainties are added.

The difference between the truth and direct tagged $m(c\bar{c})$ distributions is reduced with a ΔR correction that is applied on the Z+jets and W+jets backgrounds. The size of this correction to the $m(c\bar{c})$ distribution, shown in Fig.5.31, is added as a systematic uncertainty and the effect is up to 40% at low values of the $m(c\bar{c})$ distribution.

In addition, for all $m(c\bar{c})$ distributions using the truth tagging method, an



(a) $VH(\rightarrow c\bar{c})$ signal (b) Zhf background

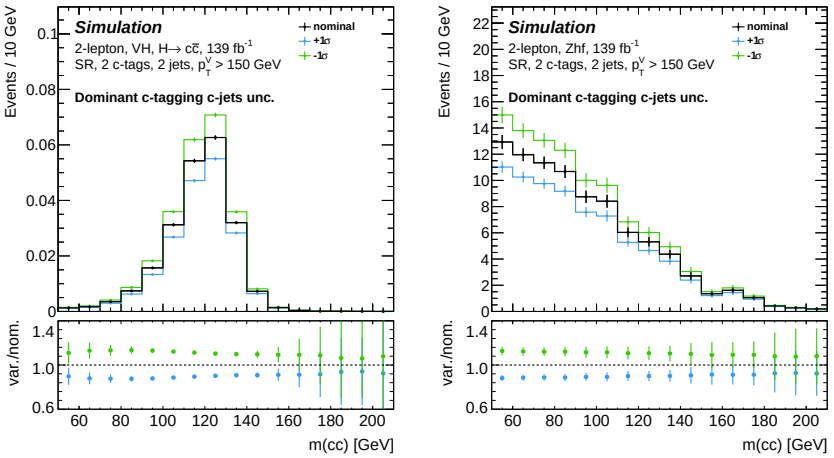
Figure 5.29: Example of the impact of the jet energy resolution uncertainty.

additional correction is added to account for differences between the number of events predicted by the truth and direct tagged $m(c\bar{c})$ distributions. The correction is calculated as the ratio of the number of events predicted by both methods and is applied on the normalisation of the truth tagged $m(c\bar{c})$ distributions. The size of the correction is used as a normalisation uncertainty and the effect is up to 10% in the categories with 2 c-tags, which is smaller than other dominant uncertainties from the flavour tagging or the truth tagging methods.

5.11.2 Theoretical uncertainties

In addition to the experimental systematic uncertainties, theoretical uncertainties are accounted for in the $VH(\rightarrow c\bar{c})$ analysis. The choice of the nominal MC generator used for the predictions of the signal and background processes is determined based on the precision of the cross-section calculations. However, due to the complexity required to simulate events as well as the different models that are used to describe the hadronisation and the parton shower, it is not guaranteed that the nominal generator describes the data the most accurately. In order to account for the uncertainty of the baseline prediction, the prediction obtained from the nominal generator is compared to the predictions derived from a set of alternative generators to quantify known limitations of the baseline prediction. The differences are used as systematic uncertainties.

To account for known limitations, alternative generators are considered with different choices for the factorisation and normalisation scales, the parton distribution function and the parton shower modelling. The $m(c\bar{c})$ distribution is estimated for



(a) $VH(\to c\bar{c})$ signal (b) Zhf background

Figure 5.30: Example of the c-tagging on c-jets uncertainty.

each alternative generator, shown in Table 5.18.

Process	Nominal	Alternative
$VH(\to c\bar{c})$, $VH(\to b\bar{b})$	`Powheg+Pythia8`	`Powheg+Herwig7`
VV	`Sherpa2.2`	`Powheg+Pythia8` `Sherpa2.2` internal weights (`MuF`,`MuR`)
Z+jets, W+jets	`Sherpa2.2`	`MadGraph5+Pythia8` `Sherpa2.2` internal weights (`MuF`,`MuR`)

$t\bar{t}$, single top
 	
Single top
	`Powheg+Pythia8`	`MadGraph5+aMC@NLO+Pythia8`
		`Powheg+Herwig7`
		`Powheg+Pythia8` + initial/final state radiations
		`Powheg+Pythia8` diagram substraction and removal

Table 5.18: The nominal and alternative MC generators used per physics process.

Shape uncertainties To account for the known limitations of the nominal prediction, alternative generators are used to derive systematic uncertainties. The $m(c\bar{c})$ distribution obtained with each alternative generator is compared to the $m(c\bar{c})$ distribution of the nominal generator and the difference is used as a systematic uncertainty.

These shape uncertainties are derived for all processes. Due to the large amount of categories and processes per category, deriving one uncertainty per process results in a large amount of systematic uncertainties. In order to reduce the number of systematic uncertainties, when possible, the uncertainties are derived in merged categories.

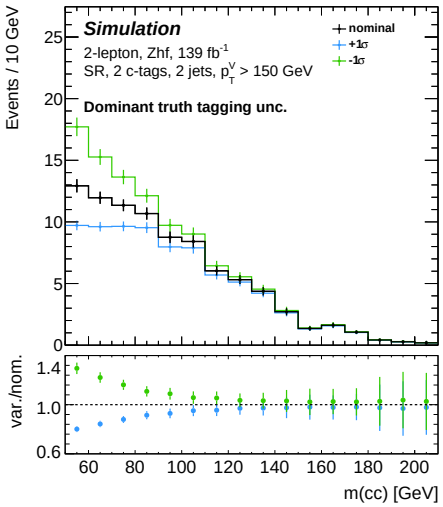
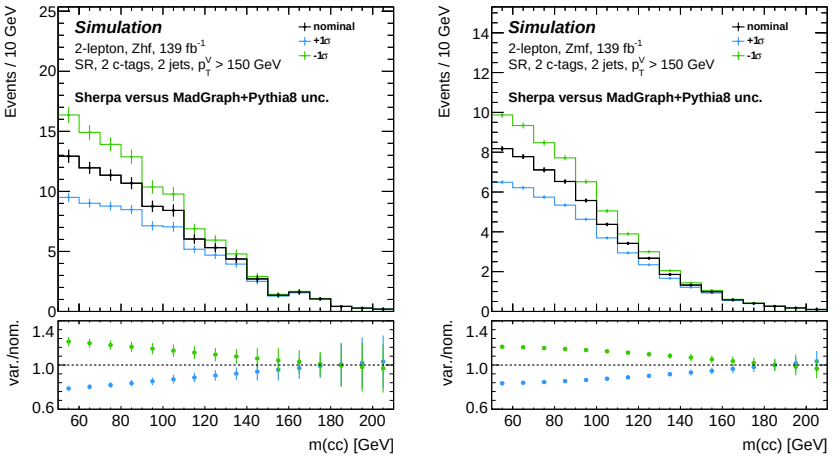


Figure 5.31: Example of the truth tagging uncertainty for the Zhf background.

For each alternative generator, the $m(c\bar{c})$ distribution is compared to the nominal prediction and the ratio of both generators is calculated for each process in each category. The ratio is then compared between the categories with different jet multiplicity, p_{T}^{V} and number of c-tags. If the ratio is found to be in agreement, the categories are summed and the ratio is re-calculated. Once all the categories with a similar ratio are merged, the ratio is fitted with a polynomial function that best describes the shape and the fitted function is applied to the $m(c\bar{c})$ distribution to derive the $\pm 1\sigma$ uncertainty on the shape.

The dominant shape uncertainty of the $VH(\rightarrow c\bar{c})$ analysis is the difference between the `MadGraph+Pythia8` and `Sherpa2.2` predictions of the $m(c\bar{c})$ distribution for the Z+jets and W+jets processes. This uncertainty, shown in Fig. 5.32 for the 2-lepton channel, affects both the shape and the normalisation of the $m(c\bar{c})$ distribution. The largest effect is observed at low values of the $m(c\bar{c})$ distribution, where the magnitude is up to 40%.

Normalisation uncertainties For the $VH(\rightarrow c\bar{c})$ signal, the theoretical uncertainties on the prediction of the branching fraction and the cross-section, shown in Table 5.19, are taken into account as an uncertainty on the total number of events predicted, also referred to as a normalisation uncertainty. The uncertainties related to the branching fraction $BR_{H\rightarrow c\bar{c}}$ arise from missing higher order calculations and the parametric uncertainties resulting from the quark mass estimation $m_c(m_c)$ and the precision on the strong force coupling $\alpha(m_Z)$. The theoretical uncertainties on the production cross-sections $\sigma_{pp\rightarrow ZH}$ and $\sigma_{pp\rightarrow WH}$ originate from different



(a) Zhf background (b) Zmf background

Figure 5.32: Example of the `Sherpa2.2` versus `MadGraph+Pythia8` modelling uncertainties.

sources: the QCD factorisation and normalisation scales, the missing higher order EW corrections, the limited size of the data from which the parton distribution function is extracted and the precision on the strong force coupling $\alpha(m_Z)$.

Source	Uncertainty
$BR_{H\to c\bar{c}}$	+5.53% / -1.99%
$\sigma_{pp\to ZH}$	+4.11% / -3.48%
$\sigma_{pp\to WH}$	+1.98% / -2.05%
$\sigma_{pp\to ZH} \times BR_{H\to b\bar{b}}$	+27% / -25%
$\sigma_{pp\to WH} \times BR_{H\to b\bar{b}}$	+25% / -23%

Table 5.19: Theoretical uncertainties on the prediction of the branching fraction and the production cross-section of the $VH(\to c\bar{c})$ process.

For the $VH(\to b\bar{b})$ background, to account for the precision on the branching fraction and production cross-section predictions, the result of the ATLAS measurement is taken as a normalisation uncertainty.

5.12 Statistical treatment

In order to measure the signal strength of the $VH(\rightarrow c\bar{c})$ process, a fit to the observed $m(c\bar{c})$ distribution is performed. The signal strength of the $VH(\rightarrow c\bar{c})$ process is defined as:

$$\mu_{VH(\rightarrow c\bar{c})} = \frac{\sigma^{obs}_{VH} \times BR^{obs}_{H\rightarrow c\bar{c}}}{\sigma^{SM}_{VH} \times BR^{SM}_{H\rightarrow c\bar{c}}}, \tag{5.12}$$

where $\sigma \times BR$ is the cross section times branching ratio of the $H \rightarrow c\bar{c}$ process. The signal strength quantifies the agreement between the observed data and the SM prediction. The aim of the fit to the data is to determine the value of the signal strength. The goodness of the fit is quantified by a likelihood ratio test, where the hypotheses are based on the simulation of all processes.

5.12.1 Likelihood Function

To define the likelihood function, the ATLAS experiment is assumed to be a counting experiment and the measured data is expected to follow a Poisson probability distribution. Furthermore, each bin of the $m(c\bar{c})$ distribution is statistically independent from the others. The likelihood as a function of the signal strength is then given by:

$$\mathcal{L}(\mu) = \prod_{i\in\text{bins}} \text{Pois}(N_i|\mu s_i + b_i) = \prod_{i\in\text{bins}} \frac{(\mu s_i + b_i)^{N_i}}{N_i!} e^{-(\mu s_i + b_i)}, \tag{5.13}$$

where i corresponds to each bin of the $m(c\bar{c})$ distribution, N_i is the number of data events measured and s_i and b_i are the expected signal and background yields.

Systematic uncertainties affect the $m(c\bar{c})$ distribution and are considered as nuisance parameters (NPs) $\vec{\theta}$ that modify the expected signal and background yields, $\{s_i, b_i\} \rightarrow \{s_i(\vec{\theta}), b_i(\vec{\theta})\}$. Available knowledge on the constraints on $\vec{\theta}$ is assumed to be expressable as a Gaussian distribution, and is included in the likelihood through an additional term $\mathcal{L}_{\text{Syst}}$:

$$\mathcal{L}_{\text{Syst}} = \prod_{\theta\in\vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2}. \tag{5.14}$$

In the fit, the effect of these NPs on the various s_i and b_i is correlated across all the bins of the $m(c\bar{c})$ distribution and, in order to correct for a specific shape or normalisation in the data, the NP can be adjusted to express a fraction or multiple of the total uncertainty. Pulling these parameters away from the nominal values deteriorates the goodness of the fit quantified by the profile likelihood ratio. Additional NPs that modify the yield of the process uniformly across all bins,

referred to as *floating normalisations*, are considered in the likelihood for the main background processes in the analysis. For these floating normalisations, no external knowledge is included in $\mathcal{L}_{\text{Syst}}$.

Furthermore, uncertainties due to the limited size of the simulation sample are taken into account by introducing γ-parameters γ_i in each bin of the $m(c\bar{c})$ distribution. These parameters are allowed to modify the expected background yield as $b_i(\vec{\theta}) \to \gamma_i b_i(\vec{\theta})$ in specific bins independently from the others, in accordance with the uncertainty that is expected from the finite size of the simulation samples. They are constrained by multiplying the likelihood with

$$\mathcal{L}_{\text{BkgStat}}(\vec{\gamma}) = \prod_{i \in \text{bins}} \text{Gauss}(\beta_i | \gamma_i \beta_i, \sqrt{\gamma_i \beta_i}), \text{with } \beta_i = \frac{1}{\sigma_{\text{rel}}^2}, \tag{5.15}$$

where σ_{rel} is the relative statistical uncertainty on the expected total background yield. Thanks to the truth tagging method, which allows to improve the statistical power of the simulated samples, these γ-parameters are reduced and their effect on the final fit is small.

The full likelihood is written as:

$$\mathcal{L}(\mu, \vec{\theta}, \vec{\gamma}) = \prod_{i \in \text{bins}} \text{Pois}(N_i | \mu s_i(\vec{\theta}) + \gamma_i b_i(\vec{\theta})) \times \prod_{\theta \in \vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2} \times \prod_{i \in \text{bins}} \text{Gauss}(\beta_i | \gamma_i \beta_i, \sqrt{\gamma_i \beta_i}). \tag{5.16}$$

Finally, a profile likelihood ratio test-statistic is used to estimate the p-value necessary for the calculation of the significance and limit on the signal strength:

$$q_\mu = \begin{cases} 2 \ln \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\gamma}}})}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}}, \hat{\vec{\gamma}})} & \mu \geq \hat{\mu} \\ -2 \ln \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\gamma}}})}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}}, \hat{\vec{\gamma}})} & \mu < \hat{\mu} \end{cases} \tag{5.17}$$

where $\{\hat{\mu}, \hat{\vec{\theta}}, \hat{\vec{\gamma}}\}$ are the values that maximise the likelihood and $\{\hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\gamma}}}\}$ the values that maximize $\mathcal{L}$ for a fixed value of μ, in this case the background only hypothesis which corresponds to $\mu = 0$. In order to probe p-values larger than 50%, which arise when $\mu \gg \hat{\mu}$ in the case of negative fluctuations in the data, the sign of the ratio is inverted.

The signal strengths of the $VH(\to c\bar{c})$ process and the $VW(\to cq)$ and $VZ(\to c\bar{c})$ processes are considered to be the parameters of interest: $\mu_{VH(\to c\bar{c})}$, $\mu_{VW(\to cq)}$ and $\mu_{VZ(\to c\bar{c})}$. Due to the wide range of sensitivity to these parameters of interest, the results are expressed in two different figures of merit, a limit obtained in a confidence interval of 95% for the $VH(\to c\bar{c})$ signal (CLs) [79], and the significance for the $VW(\to cq)$ and $VZ(\to c\bar{c})$ processes. The procedure is implemented within a statistical framework based on `HistFactory` [80] together with `RooFit` [81] and `RooStats` [82].

5.12.2 Fit Model

In order to measure the signal strength of the $VH(\to c\bar{c})$, $VW(\to cq)$ and $VZ(\to c\bar{c})$ processes, the signal and background processes as well as the systematic uncertainties are combined in a fit model. The goal of the fit model is to estimate the values of its parameters that are most probable for the observed data. In order to reduce the dependency on the physics simulation assumption for the background processes, the normalisation of each main background is estimated from the data in CRs.

Signal processes The signal strengths of the $VH(\to c\bar{c})$, $VZ(\to c\bar{c})$ and $VW(\to cq)$ processes are floating parameters determined in the fit to the data. For each process, a single parameter of interest is fitted across all the categories of the analysis. The signal strength is considered compatible with the SM if the value of the signal strength is equal to unity $\mu = 1$ within uncertainties. As the VZ, VW, $Z \to c\bar{c}$ and $W \to cq$ processes are electroweak processes that were previously measured [83], the measurement is expected to be in agreement with the SM, which would indicate that the measured $VH(\to c\bar{c})$ signal strength is not affected by any bias.

Background processes Each main background process, listed in Table 5.20, is assigned a floating normalisation in the fit model and the normalisation is determined from the data. In order to reduce the dependency on the modelling of the simulation, additional CRs, enriched in one of the background process, are fitted simultaneously with the SRs.

Process	Floating normalisation	Description
$Z + (cc, bb)$	$Z + hf$	Z+heavy flavours
$Z + (cl, bl, bc)$	$Z + mf$	Z+mixed flavours
$Z + (ll)$	$Z + lf$	Z+light flavours
$W + (cc, bb)$	$W + hf$	W+heavy flavours
$W + (cl, bl, bc, c\tau, b\tau)$	$W + mf$	W+mixed flavours
$W + (ll, l\tau)$	$W + lf$	W+light flavours
$t\bar{t}$ and single-top non-resonant	$top(b)$	(b-jet, other jet)
$t\bar{t}$ and single-top resonant	$top(other)$	$W \to qq'$
$t\bar{t}$	$t\bar{t}$	$t\bar{t}$ process in 2-lepton only

Table 5.20: Parametrisation of dominant background processes used as floating normalisations in the fit.

There are two possible treatments for the background in the fit model. The first approach consists of using one floating normalisation per category and consider each category as uncorrelated with the other ones. This approach is in principle

more conservative as it assumes that the simulation is not modelling the data correctly across different categories. The main disadvantage of this approach is that it significantly increases the number of free parameters in the fit and the number of systematic uncertainties that need to be evaluated. In addition, if two backgrounds have a similar shape in one category, it can create an ambiguity that can result in very different behaviours for each background in different categories, which is nonphysical.

The second approach consists of assigning a single floating normalisation that is correlated across various categories in the fit model. As an example in the specific case of the $VH(\to c\bar{c})$ analysis, a single floating normalisation for the $Z + hf$ process is used across the categories with 2 jets and with 3 jets. The assumption is that the $Z + hf$ process is the same process in both categories and any mis-modelling in the simulation should affect both categories similarly. In this approach, the category with the largest number of events has a larger weight in the determination of the floating normalisation. In order to account for residual differences between the number of events predicted in the categories that are treated as correlated, an additional normalisation uncertainty is added to the category with the lowest number of events and corresponds to a prior assumption on the relative difference between the two categories. This systematic uncertainty, referred to as an acceptance uncertainty, is derived by comparing the prediction of the nominal MC simulation and alternative generators on the ratio of the number of events in the two categories and the relative difference is taken as a normalisation uncertainty.

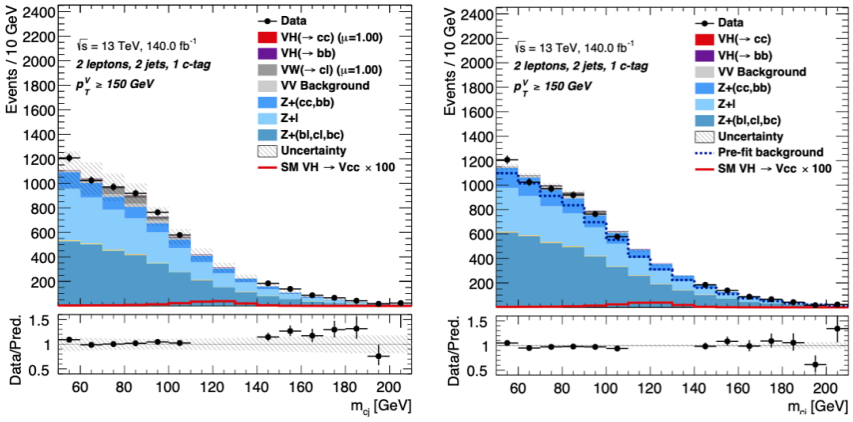
In the $VH(\to c\bar{c})$ analysis, the floating normalisations of the $Z + hf$, $Z + mf$, $W + hf$, $W + mf$, $top(b)$ and $top(other)$ processes are correlated across the different categories of the SR and CR as well as the different channels. Due to the high purity of 0 c-tag and top CRs enriched in the $Z + lf$, $W + lf$ and $t\bar{t}$ (2-lepton only) processes, one floating normalisation uncorrelated per category is used in the fit model. This treatment of each background in the fit model, summarised in Table 5.21, ensures that the background components are constrained by the data.

Blinding procedure In order to avoid the bias of the experimenter [16], a blinding procedure is applied throughout the analysis and in the fit model. The data in region of 110 GeV $< m(c\bar{c}) <$ 140 GeV is only looked at when the analysis is validated in detail.

The fit model is studied by fitting all the background contributions to the data in the entire range of the $m(c\bar{c})$ distribution, including the blinded bins, while fixing the signal strength of the three parameters of interest to $\mu_{\text{signal}} = 1$. As an illustration, the $m(c\bar{c})$ distribution of the 2-lepton channel is shown in Fig. 5.33 for the pre-fit and the post-fit results. Once the effect of systematic uncertainties is understood, the blinding of the data is removed and the result on the signal strength is measured.

$VH(\to b\bar{b})$	
$WH(\to b\bar{b})$ normalisation	27%
$ZH(\to b\bar{b})$ normalisation	25%
Diboson	
$WW/ZZ/WZ$ acceptance	10/5/12%
p_{T}^{V} acceptance	4%
N_{jet} acceptance	7 – 11%
Z**+jets**	
Z+hf normalisation	Floating
Z+mf normalisation	Floating
Z+lf normalisation	Floating
$Z + bb$ to $Z + cc$ ratio	20%
$Z + bl$ to $Z + cl$ ratio	18%
$Z + bc$ to $Z + cl$ ratio	6%
p_{T}^{V} acceptance	1 – 8%
N_{jet} acceptance	10 – 37%
High ΔR CR to SR	12 – 37%
0- to 2-lepton ratio	4 – 5%
W**+jets**	
W+hf normalisation	Floating
W+mf normalisation	Floating
W+lf normalisation	Floating
$W + bb$ to $W + cc$ ratio	4 – 10 %
$W + bl$ to $W + cl$ ratio	31 – 32 %
$W + bc$ to $W + cl$ ratio	31 – 33 %
$W \to \tau\nu(+c)$ to $W + cl$ ratio	11%
$W \to \tau\nu(+b)$ to $W + cl$ ratio	27%
$W \to \tau\nu(+l)$ to $W + l$ ratio	8%
N_{jet} acceptance	8 – 14%
High ΔR CR to SR	15 – 29%
$W \to \tau\nu$ SR to high ΔR CR ratio	5 – 18%
0- to 1-lepton ratio	1 – 6 %
Top quark (0- and 1-lepton)	
top(b) normalisation	Floating
top(other) normalisation	Floating
N_{jet} acceptance	7 – 9%
0- to 1-lepton ratio	4%
SR/top CR acceptance ($t\bar{t}$)	9%
SR/top CR acceptance (Wt)	16%
Wt / $t\bar{t}$ ratio	10%
Top quark (2-lepton)	
Normalisation	Floating
Multi-jet (1-lepton)	
Normalisation	20 – 100%

Table 5.21: Summary of the background modelling systematic uncertainties considered. The values given refer to the size of the uncertainty affecting the acceptance of each background. Where the size of an acceptance systematic uncertainty varies between analysis regions, a range is displayed. Uncertainties in the shapes of the $m(c\bar{c})$ distributions are not shown, but are taken into account for all backgrounds.



(a) Pre-fit (b) Post-fit

Figure 5.33: Illustration of the pre-fit and post-fit blinded $m(c\bar{c})$ distributions for the 2-lepton channel.

5.13 Fit to the data

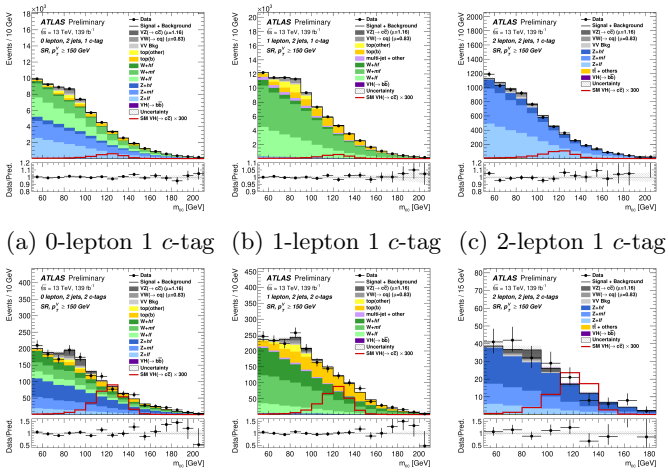
The fit to the data is performed on the $m(c\bar{c})$ distribution in a range that includes the mass of the W, Z and Higgs bosons, shown in Table 5.22. The extended fit range also provides additional constraining power on the main background processes.

	SR	0 c-tag CR	top $e\mu$ CR	High ΔR CR
Fitting range	50 GeV $< m(c\bar{c}) <$ 210 GeV			100 GeV $< m(c\bar{c}) <$ 350 GeV

Table 5.22: Range of the $m(c\bar{c})$ distribution used in the fit.

For low values of the $m(c\bar{c})$ distribution, the largest background component is V+jets, which is dominated by the process where the two jets are produced from a gluon decay. This background process is challenging to simulate and therefore, in order to mitigate the effect of this background, the lower boundary of the $m(c\bar{c})$ distribution is chosen to be 50 GeV so that it can be measured from the data. For high values of the $m(c\bar{c})$ distribution, the number of observed data events decreases and are subject to higher statistical fluctuations. To mitigate this effect, the upper boundary is chosen to be 210 GeV. A similar fitting range is used in the 0 c-tag, top and top $e\mu$ CRs. For the high ΔR CR, the fitting range is chosen to include events with even higher values of the $m(c\bar{c})$ distribution. All boundaries are always chosen considering the statistical power of the data sample.

For the SR, the bin widths are close to the experimental resolution of the



(a) 0-lepton 1 c-tag (b) 1-lepton 1 c-tag (c) 2-lepton 1 c-tag

(d) 0-lepton 2 c-tag (e) 1-lepton 2 c-tag (f) 2-lepton 2 c-tag

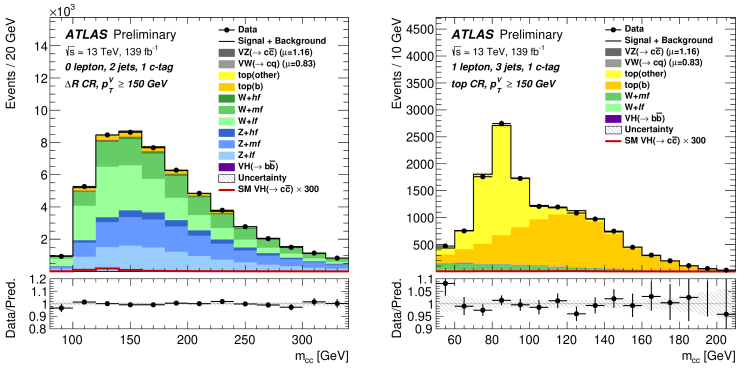
Figure 5.34: Example of the post-fit $m(c\bar{c})$ distributions for the SR in categories with 2 jets and $p_{\mathrm{T}}^{V} > 150$ GeV.

$VH(\to c\bar{c})$ signal of 10 GeV. Due to the contribution of the resonant W component in the top CR of the 0-lepton and 1-lepton channels, the binning is chosen to be 10 GeV for that region too. In the high ΔR CR, in order reduce the effect of statistical fluctuations, the binning is chosen to be twice the bin width of the SR: 20 GeV. For the 0 c-tag and top $e\mu$ CRs, the aim is to determine the normalisation of the main backgrounds and thus only a single bin is used.

As an illustration, the post-fit $m(c\bar{c})$ distribution for the 0-, 1- and 2-lepton channels is shown in Fig. 5.34 for 6 SRs. The data is in agreement with the simulation after the fit.

The post-fit distributions for the ΔR CR and top CR, as illustrated in Fig. 5.35, is also in agreement between the data and the simulation. These central regions provide a strong constraint on the main background processes of the analysis. In particular, the W boson originating from the $t\bar{t}$ process is measured precisely in the top CR. This background is present in the same $m(c\bar{c})$ range as the $VZ(\to c\bar{c})$ and $VW(\to cq)$ process and a precise constraint on it is necessary for the measurement to be able to disentangle the various contributions.

The sum of all the SR is shown in Fig. 5.36. In the lower part of the plot, where the background is subtracted from the data, the $VZ(\to c\bar{c})$ and $VW(\to cq)$ processes can be seen to peak in the 90 GeV region of the $m(c\bar{c})$ observable.



(a) ΔR CR (b) Top CR

Figure 5.35: Example of the post-fit $m(c\bar{c})$ distributions for the ΔR and top CRs.

5.14 Results of the measurement

The final result, obtained from the fit to the data for the three parameters of interest, is:

$$\begin{aligned} \mu_{VH(\to c\bar{c})} &= -9 \pm 10 \text{ (stat.)} \pm 12 \text{ (syst.)} \\ \mu_{VW(\to cq)} &= 0.83 \pm 0.11 \text{ (stat.)} \pm 0.21 \text{ (syst.)} \\ \mu_{VZ(\to c\bar{c})} &= 1.16 \pm 0.32 \text{ (stat.)} \pm 0.36 \text{ (syst.)}. \end{aligned} \tag{5.18}$$

The fitted signal strength of the three processes are compatible with the prediction of the SM. The $m(c\bar{c})$ distributions for events selected with 1 or 2 c-tagged jets, summed over all the channels after the background substraction, is shown in Fig. 5.37. No excess of the $VH(\to c\bar{c})$ process is found. For both the $VW(\to cq)$ and $VZ(\to c\bar{c})$ processes, an excess is found in the 1 and 2 c-tags categories. The observed (expected) significances of the $VW(\to cq)$ and $VZ(\to c\bar{c})$ signals are 3.8 (4.6) and 2.6 (2.2) standard deviations. For the first time at the LHC, decays of the W and Z bosons to a final state with at least one c-quark have been measured, using inclusive c-tagging algorithms.

As no excess is measured for the $VH(\to c\bar{c})$ process, an upper limit on the signal strength is set. This limit, shown in Fig. 5.38, restricts the observed (expected) signal strength $\mu_{VH(\to c\bar{c})}$ to be < 26 (31) at 95% confidence level. In order to estimate the contribution of each channel to the sensitivity, a fit where the parameters of interest are decorrelated across the different channels is performed. The channel with the highest sensitivity is the 0-lepton channel, followed by the 2-lepton and 1-lepton channels.

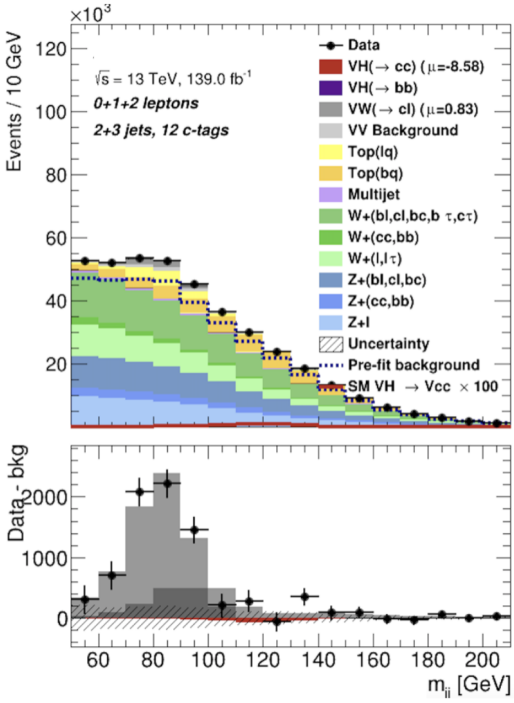


Figure 5.36: Sum of all the SR categories of the analysis.

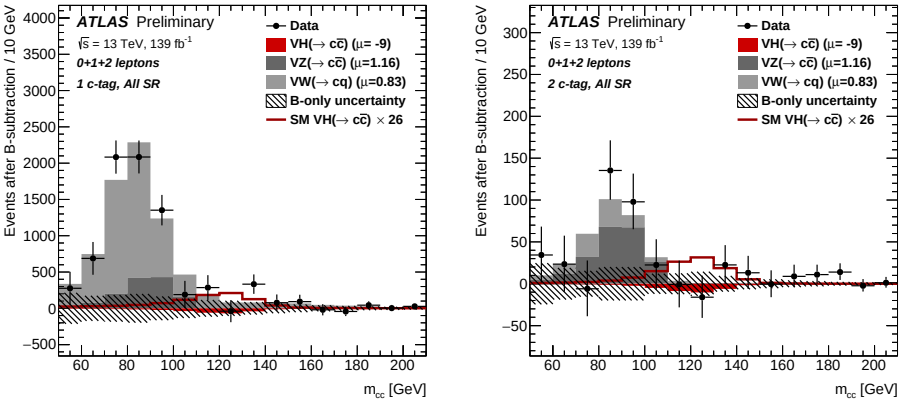
Breakdown of uncertainties

In order to understand the limitation on the sensitivity of the measurement, the effect of each systematic uncertainty on the signal strength is evaluated. The impact is obtained by fitting the data with and without varying each NPs and comparing the uncertainty on the signal strength $\sigma_{\hat{\mu}'}$ with the uncertainty from the nominal fit $\sigma_{\hat{\mu}}$ as

$$\text{impact} = \sqrt{\sigma^2_{\hat{\mu}} - \sigma^2_{\hat{\mu}'}}. \tag{5.19}$$

The impact of statistical uncertainty on the signal strength of the $VH(\to c\bar{c})$ process is similar to the impact of the total systematic uncertainty. Table 5.23 shows the breakdown of systematic uncertainties into categories of similar origin and their respective impact on the fitted $VH(\to c\bar{c})$ signal strength $\hat{\mu}$.

The largest source of uncertainty is the modelling of the main backgrounds: Z+jets, $t\bar{t}$ and W+jets processes. In addition, uncertainties related to the truth tagging method, as well as the size of the simulated sample, are among the dominant



(a) 1 c-tag (b) 2 c-tags

Figure 5.37: Result of the $m(c\bar{c})$ distribution after background substraction from the fit to data in the 1 c-tag (a) and 2 c-tags (b) categories.

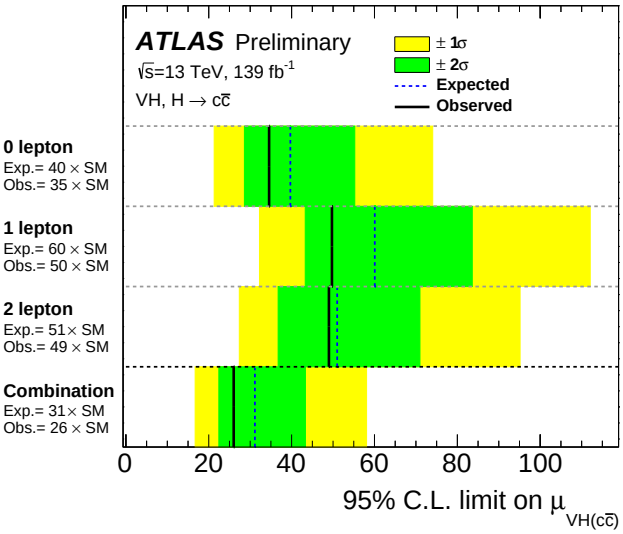


Figure 5.38: Observed and expected limits for the $VH(\to c\bar{c})$ signal at 95% CL, the result in the channels correspond to the results obtained from a combined fit with the channels POI decorrelated.

contributions to the total systematic uncertainty. Even though the uncertainty is large, the sensitivity gain due to the truth tagging method is approximately 10% with respect to the direct tagging method. The effect of these uncertainties could be mitigated in the future by using more advanced simulations for the main background processes and by simulating more events.

5.15 Interpretation of the result

The sensitivity to the $VH(\rightarrow c\bar{c})$ process has a large constraining power on any theory predicting an enhanced coupling of the Higgs boson to charm quarks. The signal strength measured in the analysis is interpreted within the κ framework [84] to set a constraint on a generic modified Yukawa couplings scenario. The coupling modifiers κ_i (where i represents the direct coupling of a particle i to the Higgs boson) modify the leading-order interaction of the Higgs boson production or decay, which affects the number of events produced.

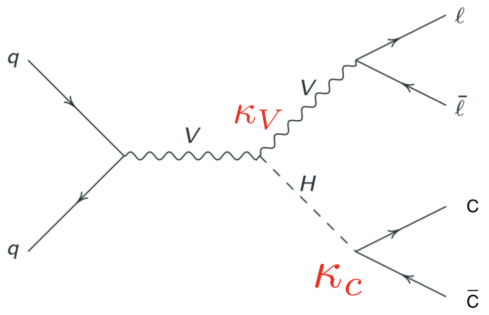


Figure 5.39: Feynman diagram of the $VH(\rightarrow c\bar{c})$ process for the 2-lepton channel, red represent the two vertices where the κ_V and κ_c coupling modifiers appear explicitly.

In the κ framework, the modifications to the couplings g_i of the Higgs boson are parametrised as:

$$g_i \rightarrow \kappa_i g_i, \tag{5.20}$$

where κ_i is a coupling modifier that can have any real value. In the scope of the $VH(\rightarrow c\bar{c})$ analysis, the κ_V and κ_c are the coupling modifiers that affect the production and the decay rate of the observed signal process, as shown in Fig. 5.39. The cross-section and the branching fractions are expressed as:

$$\sigma_{pp\rightarrow VH} \times BR_{H\rightarrow c\bar{c}} = \sigma^{SM}_{pp\rightarrow VH} \times BR^{SM}_{H\rightarrow c\bar{c}} \times \frac{\kappa_V^2 \kappa_c^2}{\kappa_H^2}, \tag{5.21}$$

Source of uncertainty		$\mu_{VH(\to c\bar{c})}$	$\mu_{VW(\to cq)}$	$\mu_{VZ(\to c\bar{c})}$
Total		15.3	0.24	0.48
Statistical		10.0	0.11	0.32
Systematics		11.5	0.21	0.36
Statistical uncertainties				
Data statistics only		7.8	0.05	0.23
Floating normalisations		5.1	0.09	0.22
Theoretical and modelling uncertainties				
$VH(\to c\bar{c})$		2.1	< 0.01	0.01
Z+jets		7.0	0.05	0.17
Top-quark		3.9	0.13	0.09
W+jets		3.0	0.05	0.11
Diboson		1.0	0.09	0.12
$VH(\to b\bar{b})$		0.8	< 0.01	0.01
Multi-Jet		1.0	0.03	0.02
Simulation statistics		4.2	0.09	0.13
Experimental uncertainties				
Jets		2.8	0.06	0.13
Leptons		0.5	0.01	0.01
$E_{\mathrm{T}}^{\mathrm{miss}}$		0.2	0.01	0.01
Pile-up and luminosity		0.3	0.01	0.01
Flavour tagging	c-jets	1.6	0.05	0.16
	b-jets	1.1	0.01	0.03
	light-jets	0.4	0.01	0.06
	τ-jets	0.3	0.01	0.04
Truth-flavour tagging	ΔR correction	3.3	0.03	0.10
	Residual non-closure	1.7	0.03	0.10

Table 5.23: Breakdown of contributions to the uncertainty in the fitted values of $\mu_{VH(\to c\bar{c})}$, $\mu_{VW(\to cq)}$ and $\mu_{VZ(\to c\bar{c})}$. In the case that the up and down systematic variations have different values, the mean of the absolute values is shown.

Since a large value of κ_c also modifies the width of the Higgs boson, this effect must be taken into account in the parametrisation, which is done as follows:

$$\kappa_H = \frac{\sum_i \Gamma_i^{SM} \kappa_i^2}{\Gamma_H^{SM}}. \tag{5.22}$$

The signal strength, parametrised as a function of the coupling modifiers, then becomes:

$$\mu_{VH(\to c\bar{c})} = \frac{BR_{H\to c\bar{c}} \times \sigma_{pp\to VH}(\kappa_i)}{BR^{SM}_{H\to c\bar{c}} \times \sigma^{SM}_{pp\to VH}} = \frac{\kappa_V^2 \kappa_c^2}{\kappa_H^2}. \tag{5.23}$$

Assuming that for all other massive particles, the coupling to the Higgs boson is as predicted by the SM, i.e. $\kappa_i = 1, \forall i \neq c$, the signal strength parametrisation in Eq. 5.23 reduces to:

$$\mu(\kappa_i) = \frac{\kappa_c^2}{BR^{SM}_{H\to c\bar{c}} \times \kappa_c^2 + (1 - BR^{SM}_{H\to c\bar{c}})}. \tag{5.24}$$

As can be seen in Fig. 5.40, for a small value of the signal strength, the constraint on κ_c scales as $\mu(\kappa_c) \approx \kappa_c^2$. For higher values of the signal strength, due to the modification to the width of the Higgs boson, the corresponding $\mu(\kappa_c)$ reaches an asymptotic limit $\mu(\kappa_c) \approx \frac{1}{BR^{SM}_{H\to c\bar{c}}}$. Therefore, any $\mu_{VH(\to c\bar{c})}$ signal strength above this threshold has no constraining power over κ_c.

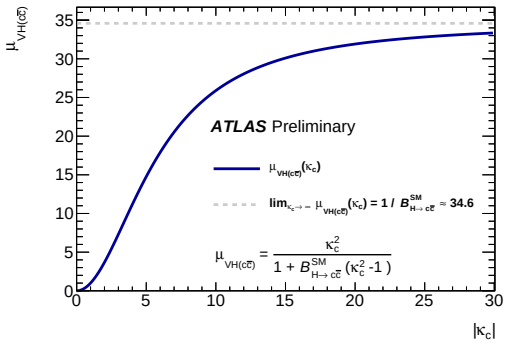


Figure 5.40: Parametrisation of the signal strength as a function of the charm quark coupling modifier $\mu(\kappa_c)$ in the context of a scenario where all other coupling modifiers are considered unity, see equation Eq. 5.24.

In order to constrain the κ_c coupling modifier, a profile likelihood scan is performed. The parameter of interest κ_c is defined following the Eq. 5.24 and the scan is performed for different fixed values of κ_c. The data is fitted for each value of κ_c and the resulting likelihood value is calculated. The statistical test is

defined as the negative log-likelihood ratio of the fitted likelihood and the maximal likelihood obtained for the best fitted $\hat{\kappa}_c$, from which the 95% confidence interval is extracted.

The result of the profile likelihood scan, shown in Fig. 5.41, is an observed (expected) limit of $\kappa_c < 8.5$ (12.4) at 95% confidence level. This result is the first direct constraint on the κ_c coupling modifier.

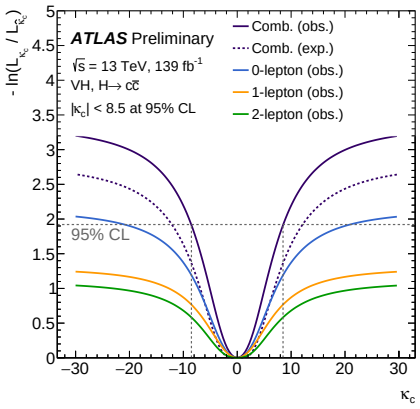


Figure 5.41: Profile likelihood scan for κ_c for the 0-lepton, 1-lepton, 2-lepton and combination of all three channels.

The direct constraint on κ_c obtained from the $VH(\to c\bar{c})$ analysis is visualised as a function of the masses of particles in Fig. 5.42 alongside the other coupling modifiers constraints obtained from analyses published by the ATLAS collaboration comparing the 68% CL uncertainties from the measurements to the 95% CL uncertainty from the $VH(\to c\bar{c})$ search. The observed κ_c value allows to exclude the universality of the interaction of the Higgs boson to the top quarks and c-quarks.

5.16 Universality of Higgs boson coupling to b- and c-quarks

The SM predicts that the coupling of the Higgs boson to b-quarks is stronger than the coupling to c-quarks. The difference is quantified by the ratio of the masses $m_b/m_c = 4.578 \pm 0.008$ [85]. The larger coupling to b-quarks results in a decay rate of the Higgs boson approximately 20 times more frequent to b-quarks than to c-quarks.

Nevertheless, as discussed in the Section 2.3, various new physics scenarios predict an enhanced coupling of the Higgs boson to c-quarks. In particular, models

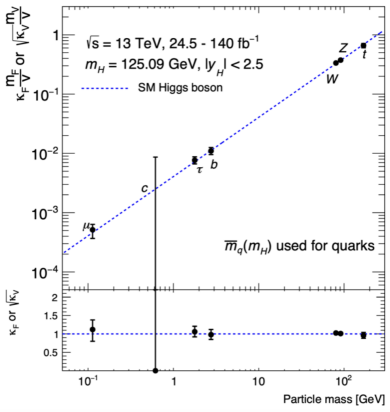


Figure 5.42: Coupling-strength modifier $\kappa_f m_f/v$ and $\sqrt{\kappa_V} m_V/v$ as a function of the mass of the fermions m_f and vector boson m_V. The constraint on the coupling modifier κ_c, corresponding to 68% CL, from the $VH(\to c\bar{c})$ analysis is illustrated along the other coupling modifiers that are obtained in a combined fit of analyses of the Higgs boson performed by the ATLAS collaboration.

predicting that the Higgs boson coupling to b-quarks is equal to the coupling to c-quarks are still experimentally possible. This equality, referred to as the universality of the Higgs boson coupling to b- and c-quarks, is described by multiplying the coupling modifier *kappa* and the coupling predicted by the SM y:

$$\kappa_b y_b = \kappa_c y_c. \tag{5.25}$$

The simultaneous measurement of the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$ processes is directly sensitive to the universality of the Higgs boson coupling to b- and c-quarks. The signal strength of both processes can be parametrised as a function of the coupling modifiers:

$$\mu_{VH(\to b\bar{b})} = \frac{\kappa_V^2 \kappa_b^2}{\kappa_H^2},\ \mu_{VH(\to c\bar{c})} = \frac{\kappa_V^2 \kappa_c^2}{\kappa_H^2}. \tag{5.26}$$

Both signal strengths depend on κ_V, the modification of the coupling of the Higgs boson to vector boson V, as well as κ_H, the modification to the total width of the Higgs boson. This dependency on modifications of the width of the Higgs boson implies that the signal strengths of the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ processes are sensitive to the effect of particles beyond the SM. Nevertheless, this effect cancels in the ratio of the signal strengths and this measurement is therefore a probe of the interaction of the Higgs boson to b- and c-quarks, independent of any assumptions

on the amount and type of new particles. The ratio of the signal strengths is parametrised as:

$$\frac{\mu_{VH(\to c\bar{c})}}{\mu_{VH(\to b\bar{b})}} = \frac{\kappa_c^2}{\kappa_b^2}. \tag{5.27}$$

The combined measurement of the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ allows the measure of this ratio for the first time and any value below the universality threshold constitutes a direct proof that the Higgs boson couples less to c-quarks than to b-quarks.

5.16.1 Overview of the $VH(\to b\bar{b})$ analysis

The $VH(\to b\bar{b})$ measurement performed by the ATLAS collaboration [86] is a direct probe of the Higgs boson coupling to b-quarks. Both analyses of the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ processes are similar by design, as shown in Table 5.24. The same simulation samples, object reconstruction and a similar event selection are used. One difference is the flavour tagging categorisation, where the $VH(\to b\bar{b})$ candidate events are selected with 2 b-tagged jets. Due to the strategy implemented in the $VH(\to c\bar{c})$ analysis, the SRs regions of the two analyses are statistically independent.

The major difference between the two analyses is the discriminant used in the fit to measure the signal strength. In the $VH(\to b\bar{b})$ measurement, the discriminant is a boosted decision tree trained on a set of variables including both kinematic variables related to the jets and the leptons in the events, among the most discriminating are the $m(b\bar{b})$, $\Delta R(b\bar{b})$, $p_{\mathrm{T}}(b_1)$ and $p_{\mathrm{T}}(b_2)$ distributions. The improvement on the sensitivity with respect to using the $m(b\bar{b})$ distribution is of the order of 50%.

	$VH(\to c\bar{c})$	$VH(\to b\bar{b})$
Luminosity	139 fb^{-1}	139 fb^{-1}
Simulation samples	Same	
Object reconstruction	Same	
Event selection	Same up to ΔR selection	
Flavour tagging	2 c-tags, 1 c-tags	2 b-tags
Jet multiplicity	2 jets, 3(+) jets	2 jets, 3(+) jets
Vector boson momentum	$p_{\mathrm{T}}^V > 150$ GeV 75 GeV $< p_{\mathrm{T}}^V < 150$ GeV	$p_{\mathrm{T}}^V > 250$ GeV 150 GeV $< p_{\mathrm{T}}^V < 250$ GeV 75 GeV $< p_{\mathrm{T}}^V < 150$ GeV
Discriminant	$m(c\bar{c})$	Boosted Decision Tree

Table 5.24: Comparison of the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$ analyses.

Due to the definition of the top CR and the 0 c-tag CR in the $VH(\to c\bar{c})$ measurement, a fraction of events with 2 b-tagged jets are used in the CRs. These

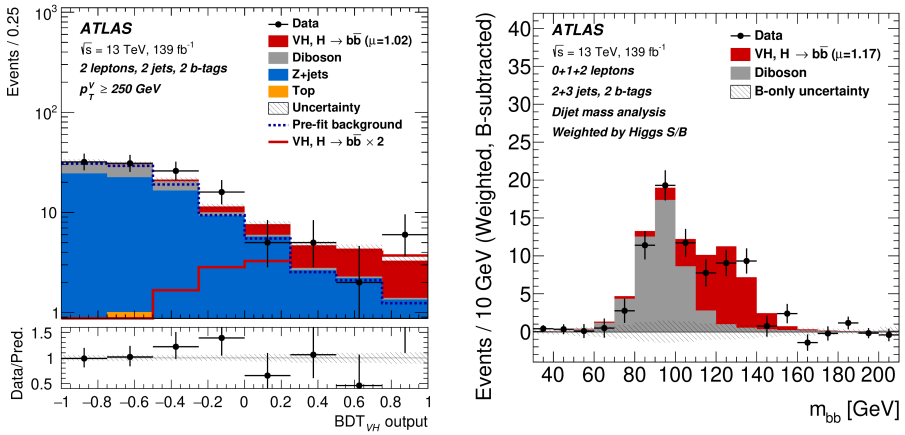


Figure 5.43: Boosted decision tree distribution used in the fit of the $VH(\to b\bar{b})$ analysis (left) and $m(b\bar{b})$ distribution after the substraction of the background processes (right) [86].

events are potentially also used in the measurement of the $VH(\to b\bar{b})$ process. The number of events with 2 b-tagged jets is about 3% in the top CR and 6% in the 0 c-tag CR. These events are mostly categorised in the CR of the $VH(\to b\bar{b})$ measurement and the effect on the combined measurement of the two processes is found to be $< 1\%$ on the signal strengths, which is negligible.

The final result obtained by the ATLAS collaboration is a direct observation of the $H \to b\bar{b}$ process, shown in Fig. 5.43. The measurement of the signal strength is

$$\mu_{VHbb} = 1.02 \pm 0.12 \text{ (stat.)} \pm 0.14 \text{ (syst.)}, \tag{5.28}$$

and corresponds to a significance of 6.7σ. The impact of the systematic uncertainty is comparable to the impact of the statistical uncertainty. The dominant source of uncertainty is the uncertainty from the theory prediction on the production, branching ratio and observable distributions of the $VH(\to b\bar{b})$ process.

5.16.2 Combined fit model

In order to measure the signal strength, as well as the coupling modifiers κ_b and κ_c, the combination is performed by fitting simultaneously all regions of both analyses. The parameters of interest are the signal strengths of the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ processes, which are correlated across all categories. Furthermore, the $VZ(\to c\bar{c})$ and $VW(\to cq)$ signal strengths are also parameters of interest in the $VH(cc)$ analysis regions. For the combination, the fit model is similar to the fit model used in the stand-alone analyses.

In the $VH(\to c\bar{c})$ fit model, the only difference with respect to the stand-alone measurement is that the overall normalisation uncertainties on the $VH(\to b\bar{b})$ background are removed, since the $VH(\to b\bar{b})$ signal strength is treated as a parameter of interest.

The experimental NPs related to the ATLAS detector are correlated across the two analyses, with one exception: the flavour tagging uncertainties. Due to a difference in the technical implementation of the flavour tagging uncertainties between the two analyses, the NPs related to flavour tagging do not represent the same underlying systematic uncertainty and these parameters cannot be correlated in the fit model. In principle, in the future, these NPs should be correlated as they originate from the same underlying physics. Due to the different strategies employed for the modelling of the background processes, the NPs related to the different predictions of MC generators are treated as uncorrelated too. The floating normalisations used to describe the main backgrounds are treated as uncorrelated as well.

The combined measurement of the two processes is studied with three different fits to the data, summarised in Table 5.25. The first measurement is a fit of the signal strength of both the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$ processes, which aims at providing a comparison to the stand-alone results. The second measurement is an interpretation of the combined result in terms of coupling modifiers κ_c and κ_b, where both coupling modifiers are accounted for in the modification of the total Higgs width:

$$\mu_{VH(bb)} = \frac{\kappa_b^2}{1 - BR_{H\to b\bar{b}} - BR_{H\to c\bar{c}} + BR_{H\to c\bar{c}}\kappa_c^2 + BR_{H\to b\bar{b}}\kappa_b^2} \tag{5.29}$$

$$\mu_{VH(cc)} = \frac{\kappa_c^2}{1 - BR_{H\to b\bar{b}} - BR_{H\to c\bar{c}} + BR_{H\to c\bar{c}}\kappa_c^2 + BR_{H\to b\bar{b}}\kappa_b^2}. \tag{5.30}$$

In addition, for the $VH(\to b\bar{b})$ process, the effect of the coupling modifier κ_b is included in the $gg \to ZH$, where the Higgs boson is radiated from a loop of virtual particles. In order to probe the universality of the Higgs boson coupling to b- and c-quarks, the third measurement is a fit parametrised as a function of the ratio κ_c/κ_b and κ_b. In this parametrisation, the modifications to the Higgs boson width originating from any type of massive particles beyond the SM is removed in the ratio.

	Parameters of interest	
Scenario 1	$\mu_{VH(\to c\bar{c})}$	$\mu_{VH(\to b\bar{b})}$
Scenario 2	κ_c	κ_b
Scenario 3	$\frac{\kappa_c}{\kappa_b}$	κ_b

Table 5.25: Different parameters of interest used in the combined fit of the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$.

5.16.3 Combined measurement of $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$

The result of the combination is a measured signal strength of

$$\begin{aligned} \mu_{VH(cc)} &= -9.1 \pm 15.3 \\ \mu_{VH(bb)} &= 1.06 \pm 0.18. \end{aligned} \tag{5.31}$$

A comparison between the standalone fits and the combination is provided in Table 5.26. Compared to the stand-alone measurement, the signal strength of the $VH(\to c\bar{c})$ process is 0.07σ lower and the signal strength of the $VH(\to b\bar{b})$ process is 0.2σ higher. This change is attributed to the $VH(\to c\bar{c})$ events mis-identified in the categories with 2 b-tagged jets and the $VH(\to b\bar{b})$ events in the categories with 2 c-tagged and 1 c-tagged jets. The signal strength of the combined fit is found in agreement with the stand-alone measurements.

	$\mu_{VH(\to b\bar{b})}$	$\mu_{VH(\to c\bar{c})}$	$\mu_{VW(\to cq)}$	$\mu_{VZ(\to c\bar{c})}$
$VH(bb)$	1.02 ± 0.18	-	-	-
$VH(cc)$	-	-8.6 ± 15.3	0.83 ± 0.24	1.16 ± 0.48
$VH(bb/cc)$ combination	1.06 ± 0.18	-9.1 ± 15.3	0.81 ± 0.23	1.11 ± 0.46

Table 5.26: Best fit values for $\mu_{VH(bb)}$, $\mu_{VH(cc)}$, $\mu_{VZ(cc)}$ and $\mu_{VW(cl)}$ in the individual $VH(cc)$ and $VH(bb)$ analyses and in the combination.

To evaluate the correlation between the two signal strengths, a likelihood scan is performed and shown in Fig 5.44. The correlation between the two signal strengths is of the order of -10%, which explains the inclination in the curves of the 68% and 95% confidence intervals. Due to the similar shape of both signals in the distributions used in the fit, an anti-correlation between the two signals is expected.

In addition, a likelihood scan is performed in two dimensions over the κ_c and κ_b parameters, as shown in Figure 5.45. In the figure, both the 68% and 95% confidence intervals are subject to an asymptotic limit in the diagonal. This effect is due to the impact of the κ_c and κ_b parameters in the modification of the total Higgs width. At certain values of κ_c and κ_b, each coupling modifier is compensating the effect of the other and therefore no constraining power can be achieved. In the likelihood scan for the observed measurement, the best fit value for the κ_b is negative. This is due to the contribution of the κ_b parameter in the production

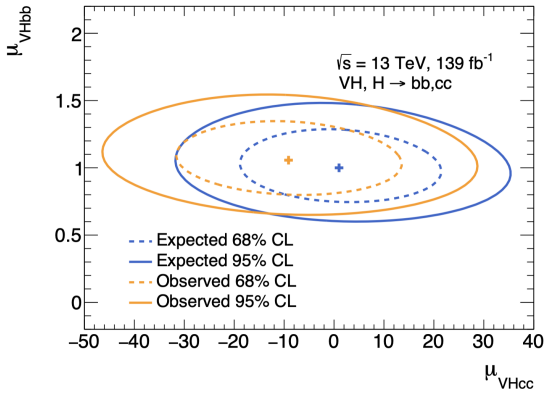


Figure 5.44: Profile likelihood scan of $\mu_{VH(\to b\bar{b})}$ and $\mu_{VH(\to c\bar{c})}$ in the combined fit.

of the $gg \to ZH$ process, which appears in the virtual loop. Measuring the ZH production is therefore important to determine the sign of the coupling modifiers.

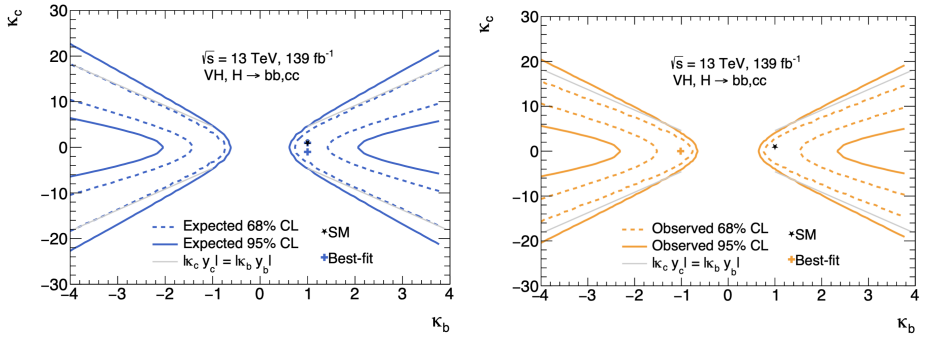


Figure 5.45: Expected (left) and observed (right) profile likelihood ratio scan of κ_c vs. κ_b in the combined fit.

Finally, a likelihood scan is performed to measure the ratio κ_c/κ_b, as shown in Figure 5.46. The result is a limit of at 95% confidence level on the ratio:

$$|\kappa_c/\kappa_b| < 4.5 \text{ (5.1)}. \tag{5.32}$$

This limit represents the direct constraint on the ratio of the Higgs coupling modifier to b- and c-quarks. The result is significantly more constraining on the coupling modifiers than a ratio of the stand-alone measurement, which would be closer to a limit value of 8. This is largely due to the fact that the contributions of the Higgs width cancel in the ratio. For the first time, the Higgs boson universality

to b- and c-quarks $m_b/m_c = 4.578 \pm 0.008$ is excluded at 95% CL and this result represents the first experimental proof that the Higgs boson coupling to c-quarks is weaker than the coupling to b-quarks, independent of any assumptions on the type and amount of particles from physics beyond the SM.

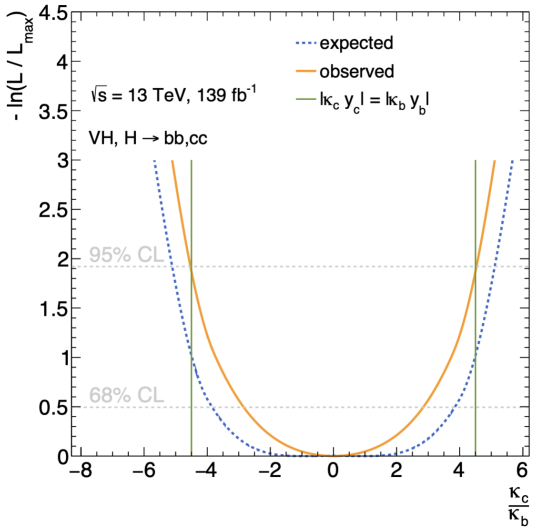


Figure 5.46: Profile likelihood scan of κ_c/κ_b in the combined fit.

6 Comparison with previous measurements

Prior to the search for the $VH(\to c\bar{c})$ process performed by the ATLAS collaboration with 139 fb^{-1}, the ATLAS, CMS and LHCb experiments reported limits on the signal strength of the $H \to c\bar{c}$ process, as shown in Fig. 6.1. All three measurements target the $VH(\to c\bar{c})$ production mode with different integrated luminosities. In order to understand how the latest measurement of the ATLAS collaboration compares to the previous measurements, a detailed comparison is presented.

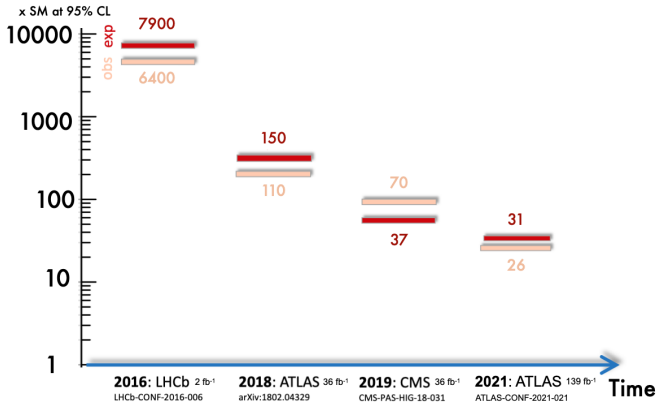


Figure 6.1: Comparison of the sensitivity of the ATLAS, CMS and LHCb experiments.

6.1 Comparison with the LHCb result

In 2016, the LHCb experiment reported a first direct limit on the signal strength of the $VH(\to c\bar{c})$ process of $\mu_{VH(\to c\bar{c})} < 6400$ at 95% confidence level [3]. The LHCb experiment is primarily a heavy flavour experiment. Its main purpose is to search for indirect evidence of new physics in charge-parity violation and in rare decays of b- and c-hadrons, by looking for the effects of new particles in processes that are predicted by the SM. The detector is designed to operate in the forward region,

$2 < \eta < 5$, and at an integrated luminosity 10 to 20 times lower than the ATLAS and CMS experiments. Therefore, LHCb analyses targeting measurements of the Higgs boson properties are challenging but could potentially play an important role in the future, when the experiment is designed to record larger data sets.

The search at the LHCb experiment is performed with data collected during Run 1 at a center of mass energy of $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 2 fb^{-1}. The analysis targets simultaneously the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ processes, which is possible thanks to a two-dimensional flavour tagging algorithm. The efficiency is 60% for b-jets, 20% for c-jets and 0.1% for light-jets. While the tagging efficiency is comparable to the performance achieved in the ATLAS experiment for b-jets and light-jets, the mis-identification of light-jets is 10 times lower. This superior performance is due to the vertex locator sub-detector of the LHCb experiment, which allows a precise reconstruction of secondary vertices, crucial for b- and c-tagging.

In comparison to the measurement of the ATLAS collaboration, the sensitivity of the LHCb experiment to the $VH(\to c\bar{c})$ process is limited by the lower luminosity. Furthermore, as the detector is operating in the forward region, the LHCb analysis targets events with at least one reconstructed electron or muon. Therefore, the LHCb experiment does not use the 0-lepton channel, which is the most sensitive channel in the ATLAS measurement. While the sensitivity of the LHCb measurement is 250 times less sensitive than the latest measurement of ATLAS, the detector is planned to undergo an important upgrade and recent projections indicate that a sensitivity of $\mu_{VH(\to c\bar{c})} < 50$ at 95% confidence level could be reached by the end of the high luminosity phase of the LHC. With further improvements in c-tagging and analysis techniques, the LHCb collaboration expects to reach a limit of $\mu_{VH(\to c\bar{c})} < 4$ at 95% confidence level [87].

6.2 Comparison with the previous ATLAS result

A previous search for the $VH(\to c\bar{c})$ process was performed by the ATLAS collaboration and published in 2018 [4]. The result was an expected limit on the signal strength $\mu_{VH(\to c\bar{c})} < 154$ at 95% confidence level. Several aspects of the analysis were improved and the main differences are listed in Table 6.1. The most notable change is the inclusion of the 0- and 1-lepton channels, which were found to significantly improve the sensitivity. In addition, the c-tagger was updated to the latest taggers available in the ATLAS collaboration and was optimised to account for the additional channels and the increased luminosity. The resulting c-tagging performance is 35% lower on the c-jets tagging efficiency but is 4 times better on the b-jets and light-jets rejection.

In order to understand the impact of each improvement on the sensitivity to the $VH(\to c\bar{c})$ process, a fit to the 2-lepton channel of the latest measurement is performed using the same data set as the previous publication, corresponding to an integrated luminosity of 36 fb^{-1}. Each improvement is incrementally added

	ATLAS 36 fb^{-1}	ATLAS 139 fb^{-1}
Channels	2-lepton	0-, 1-, 2-lepton
Flavour tagger	c-tagging	c-tagging + **b-tag veto**
Tagging categories	1 c-tag and 2 c-tags	
Flavour tagging performance		
c-jets	41%	27%
b-jets	25%	**8%**
light-jets	5%	**1.6%**
Jet multiplicity	$\geq$ 2 jets	**2 and $\geq$3 jets**
p_{T}^{V}	Same categorisation	
CRs	Top-$e\mu$	Top-$e\mu$, **Top, ΔR CR and 0 c-tags**
$VH(\rightarrow b\bar{b})$ treatment	Overlapping events	**Orthogonality**
$VH(\rightarrow b\bar{b})$ fraction	6%	**0.7%**
Calibrations	36 fb^{-1}	**80-139 fb^{-1}**

Table 6.1: Comparison between the analyses of the $VH(\rightarrow c\bar{c})$ process performed by the ATLAS collaboration at 36 fb^{-1} and 139 fb^{-1}. The changes between the two analyses are highlighted in a bold font.

to the setup. The result, shown in Table 6.2, indicates that when using the same categories as the previous publication, the current result is 38% more sensitive. This difference is largely explained by the improvement in the flavour tagging performance. With the inclusion of the 0- and 1-lepton channels, as well as the additional integrated luminosity, the limit on $\mu_{VH(\rightarrow c\bar{c})}$ is 5 times better than the previous publication.

	$VH(\rightarrow c\bar{c})$	$VW(\rightarrow cq)$	$VZ(\rightarrow c\bar{c})$
$ZH(\rightarrow c\bar{c})$ analysis at 36 fb^{-1}	$<$ 154$\times$ SM	-	-
Current 2-lepton channel fit at 36 fb^{-1}	$<$ 97$\times$ SM	-	-
Additional diboson POIs	$<$ 104$\times$ SM	1.0σ	0.6σ
Split in 2 jets and 3 $\geq$ jets categories	$<$ 98$\times$ SM	1.0σ	0.7σ
Additional CRs	$<$ 88$\times$ SM	1.1σ	0.8σ
Additional luminosity (up to 139 fb^{-1})	$<$ 54$\times$ SM	1.8σ	1.4σ
Additional 0- and 1-lepton channels	$<$ 31$\times$ SM	4.6σ	2.2σ

Table 6.2: Impact on the expected sensitivity of the various improvements used in the latest measurement of the $VH(\rightarrow c\bar{c})$ process with respect to the previous measurement from the ATLAS collaboration.

This improvement is also illustrated in Fig 6.2, where the result is extrapolated as a function of the integrated luminosity. The previous result is projected in the context of an optimistic scaling, where both the statistical and systematical uncertainties are reduced with the increased luminosity. The latest measurement of the $VH(\rightarrow c\bar{c})$ process is clearly shown to outperform this optimistic scaling.

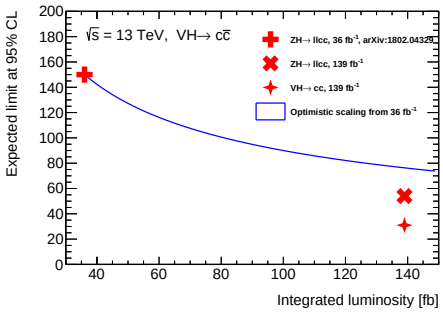
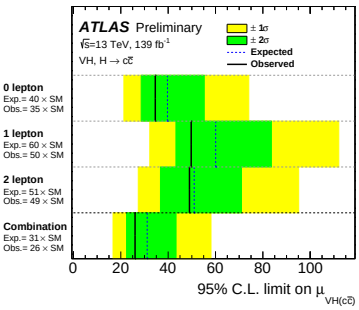


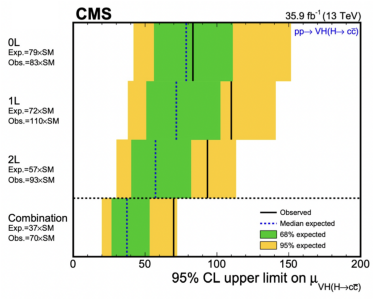
Figure 6.2: Comparison with the ATLAS result performed with 36 fb^{-1}, including a projection of the result as a function of the integrated luminosity. The optimistic scaling corresponds to improvements on both the statistical and systematical uncertainty that scale with the luminosity.

6.3 Comparison with the CMS result

The CMS experiment reported an observed (expected) limit on the signal strength of the $H \to c\bar{c}$ process $\mu_{VH(\to c\bar{c})} < 70$ (< 37) at 95% confidence level, using the 2016 dataset of 36 fb^{-1} [5]. Compared to the expected sensitivity obtained by the ATLAS experiment with a data set 4 times larger, the result of CMS is only 20% less sensitive, as shown in Fig. 6.3. Assuming that the sensitivity scales with statistical uncertainties only, four times more data should result in an improvement of 50%. In order to understand why the result of CMS is more sensitive than the result of ATLAS, a detailed comparison of both analyses is presented.



(a) ATLAS 139 fb^{-1}



(b) CMS 36 fb^{-1} [5]

Figure 6.3: Sensitivity to the $VH(\to c\bar{c})$ process for the ATLAS and CMS results.

The differences in the measurements of the $VH(\to c\bar{c})$ process performed by the ATLAS and CMS experiments are shown in Table 6.3. The most notable difference

is the discriminant used in the fit to the data. In the analysis performed by the CMS experiment, a machine learning algorithm is trained to distinguish the signal $VH(\to c\bar{c})$ from the background processes and this variable is used to perform the fit to the data. As multivariate analyses optimally exploit the correlation across different input variables, the result is an enhanced sensitivity to the $VH(\to c\bar{c})$ process.

Furthermore, the analysis of the CMS collaboration uses events categorised with 2 c-tagged jets and an additional veto on b-tagged jets is not considered, which implies that the SR of the analysis is not statistically independent from the one used in the $VH(\to b\bar{b})$ analysis. Finally, a difference is observed in the performance of the c-tagging algorithms used in the two experiments. For a similar efficiency on c-jets, the c-tagging performance of the ATLAS experiment is 2 times better in terms of b-jets and light-jets rejection. This is partially thanks to the insertable B-layer sub-detector of the ATLAS inner tracker, which improves the spatial resolution and the reconstruction of secondary vertices.

	CMS	ATLAS
Integrated luminosity	36 fb^{-1}	139 fb^{-1}
Channels	0-, 1-, 2-lepton	0-, 1-, 2-lepton
Flavour tagger	c-tagging	c-tagging + **b-tag veto**
Tagging categories	2 c-tags	**1 c-tag and 2 c-tags**
Flavour tagging performance		
c-jets	27%	27%
b-jets	17%	**8%**
light-jets	4%	**1.6%**
Discriminant	**Multivariate analysis**	$m(c\bar{c})$
Number of POIs	1	3

Table 6.3: Comparison between the analyses of the $VH(\to c\bar{c})$ process performed by the ATLAS collaboration at 36 fb^{-1} and 139 fb^{-1}. The differences between the two analyses are highlighted in a bold font.

In order to study the impact of these different strategies, a fit to the data collected by the ATLAS experiment at an integrated luminosity of 36 fb^{-1} is performed using the latest fit model of the measurement.

ATLAS vs CMS: 36 fb^{-1} comparison The analyses of the ATLAS and CMS experiments are compared both for a 36 fb^{-1} data set in terms of impact of uncertainties on the expected result. The breakdown of uncertainties, listed in Table 6.4, indicates that the total uncertainty of the analysis performed by the CMS collaboration is about 20% lower than the uncertainty obtained with the data set of ATLAS.

The main uncertainty difference comes from the systematic uncertainties, in

	ATLAS	CMS
Total uncertainty	**±23.2**	**±19.8**
Data stat	**±17.1**	**±17.2**
Data stat-only	±13.7	±13.2
Floating norms.	±7.7	±10.2
Modelling	**±12.0**	**±7.24**
Signal	±3.33	±3.75
Background	±11.5	±4.10
Simulation modelling	–	±4.65
MC stat	**±5.75**	**±1.2**

Table 6.4: Comparison of the impact of uncertainties on the expected results obtained by the ATLAS and CMS analyses using 36 fb^{-1}.

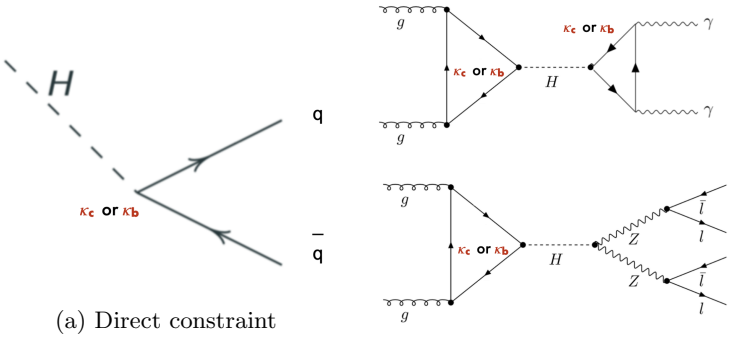
particular the uncertainties related to the background modelling and MC statistical uncertainties, which are respectively 2x and 5x smaller in the analysis by CMS. The difference in the background modelling uncertainty is explained by the fact that the analysis by CMS does not use alternative generators as a source of systematic uncertainties but relies more heavily on the usage of control regions to constrain the main backgrounds from the data. In addition, the usage of the category of events with 2 c-tagged jets only, which is limited by the statistical size of the data sample, is partially responsible for the reduced uncertainty. The difference in the uncertainty due to the size of the simulated sample is explained by the fact that the nominal generator of the V+jets sample used in the CMS analysis is `MadGraph+amc@NLO`, which has a next-to-leading-order precision with up to two additional partons and and has a leading-order precision for events with up to four additional partons. The nominal generator used in the ATLAS analysis for the same process is `Sherpa 2.2.1`, which has a next-to-next-to-leading-order precision for any number of additional partons. This generator provides more accurate predictions but is computationally more heavy and therefore results in a smaller number of simulated events.

Nevertheless, the statistical uncertainty due to the size of the data collected is similar for both analyses, indicating that without systematic uncertainties, both analyses have a similar sensitivity. The usage of categories with 1 and 2 c-tagged jets and the $m(c\bar{c})$ distribution to measure the signal strength has a similar statistical sensitivity to the usage of a category with 2 c-tagged jets and a multivariate analysis statregy. This comparison indicates that an even higher sensitivity to the $VH(\rightarrow c\bar{c})$ process could be achieved by a combination of both strategies: a multivariate analysis using categories with 1 and 2 c-tagged jets.

7 Prospects

The measurement of the $VH(\to c\bar{c})$ process results in an observed limit on the signal strength $\mu_{VH(\to c\bar{c})} < 26$ at 95% confidence level. In addition, a first direct constraint is set on the coupling modifier $|\kappa_c| < 8.5$ at 95% confidence level and, from a combination with the $VH(\to b\bar{b})$ process, a direct limit on the ratio $|\kappa_c/\kappa_b| < 4.5$ is obtained. While these results allow to exclude a variety of theories that predict an enhanced coupling of the Higgs boson to c-quarks, several improvements are possible both on the shorter and longer term.

7.1 Direct and indirect constraints on Higgs coupling modifiers



(a) Direct constraint

(b) Indirect constraint

Figure 7.1: Illustration of the constraint on the κ_b and κ_c coupling modifiers obtained from the direct $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ measurements (a), as well as indirectly through the virtual loops in the decay of $H \to \gamma\gamma$ and the production of the $H \to ZZ^*$ processes (b).

The combined measurement of the $VH(\to c\bar{c})$ and $VH(\to b\bar{b})$ process allows to probe the decay of the Higgs boson to heavy quarks, where coupling modifiers directly affect the branching ratio. These same coupling modifiers may also affect the production and the decay modes of particles interacting with the Higgs boson. In particular, the gluon-gluon fusion production of the Higgs boson is sensitive to coupling modifiers through loops of virtual particles. In the case of the $H \to \gamma\gamma$

process, the decay mode is sensitive to the effect of coupling modifiers in the decay. The production and decay of the Higgs boson, illustrated in Fig. 7.1b, are primarily sensitive to modifications of the Higgs boson interaction with the top quark, which is the most massive particle in the SM.

Due to the relatively high mass of the b- and c-quarks, virtual loops in the production and decay modes of the Higgs boson are also sensitive to the coupling modifiers κ_b and κ_c. In order to probe these effects of the coupling modifiers on loop-mediated production and decay processes, the precise measurements of the $H \to \gamma\gamma$ and $H \to ZZ^*$ processes can be used, which are determined at a precision of 10%. The production cross-section is parametrised as a function of the coupling modifiers κ_b and κ_c:

$$\begin{aligned}\sigma_{pp\to H}(\kappa_t,\kappa_b,\kappa_c) &= \sigma_{ggF}(\kappa_b,\kappa_c) + \sigma_{b\to H}(\kappa_b) + \sigma_{c\to H}(\kappa_c) + \sigma_{ttH} + \sigma_{other} \\ &= \kappa_t^2(\sigma_{ggF}^{tt} + \sigma_{ttH}) + \kappa_t\kappa_b\sigma_{ggF}^{tb} + \kappa_t\kappa_c\sigma_{ggF}^{tc} + \kappa_b\kappa_c\sigma_{ggF}^{bc} \\ &+ \kappa_b^2(\sigma_{ggF}^{bb} + \sigma_{bbH}) + \kappa_c^2(\sigma_{ggF}^{cc} + \sigma_{ccH}) + \sigma_{other}.\end{aligned} \tag{7.1}$$

In Eq. (7.1), the gluon-gluon fusion production of the Higgs boson as well as the production in association with top, bottom and charm quarks are affected by the coupling modifiers κ_t, κ_b and κ_c. Each of these production mode results in a different shape of the momentum of the Higgs boson and the predicted fraction of the production cross-section depends on the relative weights of each production cross-section. Therefore, the coupling modifiers κ_t, κ_b and κ_c affect the shape of the p_{T} distribution of the Higgs boson, as shown for κ_c in Fig. 7.2. By measuring the p_{T}^V distribution of the the $H \to \gamma\gamma$ and $H \to ZZ^*$ processes, the coupling modifiers can be constrained indirectly.

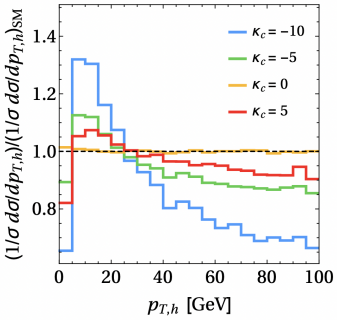
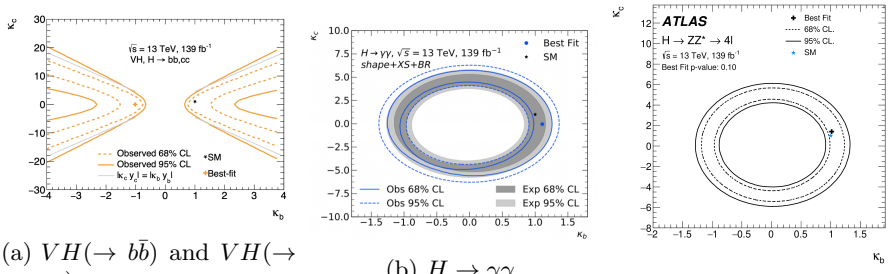


Figure 7.2: Effect of the coupling modifier κ_c on the distribution of the p_{T} of the Higgs boson, when produced through the gluon-gluon fusion mode [88].

In order to compare the direct constraints obtained from the combination of the

$VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ measurements with the indirect constraints obtained from the measurement of the p_{T}^{V} distribution of the $H \to \gamma\gamma$ and $H \to ZZ^*$ processes, the same set of assumptions is used: all coupling modifiers κ_i are assumed to be 1, with the exception of the κ_b and κ_c coupling modifiers; the coupling modifiers affect both the production and the decay rate of the Higgs boson; the Higgs boson is assumed to only couple to particles of the SM and the width of the Higgs boson is assumed to be the one predicted by the SM.

With these assumptions, the constraints on the coupling modifiers κ_b and κ_c obtained from direct and indirect measurements, shown in Fig.7.3, are found to be complementary. For the indirect constraints, the coupling modifiers κ_b and κ_c affect the total cross-section of the Higgs boson. Due to the fact that the $H \to \gamma\gamma$ and $H \to ZZ^*$ processes are measured at a precision level of 10%, the quadratic sum of the κ_b and κ_c is bound both from above and below. Since the indirect measurement cannot distinguish modifications of κ_b and κ_c, the constraint on the modifiers has the shape of a circle in the two-dimensional plane.



(a) $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$

(b) $H \to \gamma\gamma$

(c) $H \to ZZ^* \to l^+l^-l^+l^-$

Figure 7.3: Direct and indirect constraint on coupling modifiers κ_b and κ_c.

Due to the complementarity of the shape of the direct and indirect constraints, a combined fit of the $H \to \gamma\gamma$, $H \to ZZ^*$, $H \to b\bar{b}$ and $H \to c\bar{c}$ measurements will significantly improve the constraint on both κ_b and κ_c, as illustrated in Fig 7.4. This approach is currently further investigated by a collaboration between the Nikhef and Lawrence Berkeley National Laboratory groups.

7.2 Extrapolation at High-Luminosity LHC

In 2028, the LHC is planned to start a new phase with an increased instantaneous luminosity and an integrated luminosity of 3000 fb^{-1} is expected to be recorded within ten years. In addition, the ATLAS experiment will undergo an upgrade of the trigger, data acquisition system and the tracker, to be able to deal with the much larger particle flux and data rates. In order to understand the impact of these modifications on the sensitivity to the $VH(\to c\bar{c})$ process, a projection

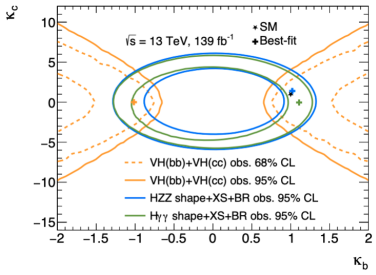


Figure 7.4: Superposition of the direct and indirect constraints obtained from the $H \to \gamma\gamma$, $H \to ZZ^*$, $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$.

study is performed by taking into account the increased luminosity, the increase in the center of mass energy of $\sqrt{s} = 14$ TeV and foreseen improvements on the systematic uncertainties.

Uncertainties	Reduction of uncertainty
$E_{\mathrm{T}}^{\mathrm{miss}}$	50%
Leptons	No improvement foreseen
Jets	No improvement foreseen
Flavour tagging	50%
Luminosity	42%
Signal modelling	50%
Background modelling	50%
MC statistic	Not considered
Truth tagging	Not considered

Table 7.1: Foreseen reductions of systematic uncertainties considered in the projection for the high luminosity phase of the LHC.

Based on the analysis approach presented in this thesis, the result on the signal strength of the $VH(\to c\bar{c})$ process obtained with a luminosity of 139 fb^{-1} is extrapolated to a luminosity of 3000 fb^{-1} [89]. The simulation of the signal and the background processes are scaled by the ratio of the integrated luminosities 3000 fb^{-1}/139 fb^{-1}. To account for the increase in the production cross-section at $\sqrt{s} = 14$ TeV, the signal and background processes are increased by 10%-18% [90]. Based on the ATLAS future performance studies presented in Ref. [91, 92], systematic uncertainties relevant to the analysis are foreseen to be reduced by the factors shown in Table 7.1.

When performing the combined measurement of the $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ processes, as shown in Fig. 7.5, the result of the projection is an expected limit on the signal strength $\mu_{VH(\to c\bar{c})} < 6.4$ at 95% confidence level for 3000 fb^{-1} of data.

The corresponding signal strength is:

$$\mu_{VH(\to c\bar{c})} = 1.0 \pm 3.2 = 1.0 \pm 2.0 \text{ (stat.)} \pm 2.5 \text{ (syst.)}. \tag{7.2}$$

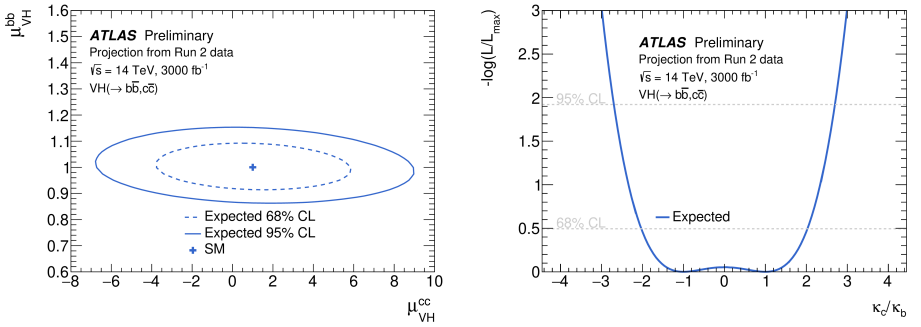
The size of the statistical uncertainty is foreseen to be reduced to an amount that is comparable to the size of the systematic uncertainty. The main sources of uncertainties, listed in Table 7.2, are still systematic uncertainties by a small margin, with dominant contributions from the theoretical uncertainties on the main background processes: Z+jets, $t\bar{t}$ and W+jets.

Source of uncertainty		$\Delta\mu_{VH(\to c\bar{c})}$
Total		3.21
Statistical		1.97
Systematics		2.53
Statistical uncertainties		
Data statistics only		1.59
Floating normalisations		0.95
Theoretical and modelling uncertainties		
$VH, H \to c\bar{c}$		0.27
Z+jets		1.77
Top-quark		0.96
W+jets		0.84
Diboson		0.34
$VH, H \to b\bar{b}$		0.29
Multi-Jet		0.09
Experimental uncertainties		
Jets		0.59
Leptons		0.20
$E_{\mathrm{T}}^{\mathrm{miss}}$		0.18
Pile-up and luminosity		0.19
Flavour tagging	c-jets	0.61
	b-jets	0.16
	light-jets	0.51
	τ-jets	0.19

Table 7.2: Breakdown of contributions to the uncertainty on $\mu_{VH(\to c\bar{c})}$ from a fit to an simulated dataset scaled to 3000 fb^{-1} and $\sqrt{s} = 14$ TeV.

This projected result can also be interpreted in terms of modifications of the Higgs boson coupling to charm quarks κ_c. The result is a limit on the coupling modifier $|\kappa_c| < 3.0$ at 95% confidence level. The measurement on the ratio is projected to be $|\kappa_c/\kappa_b| < 2.74$ at 95% confidence level. With the increased data

sample, the difference between the limit on the coupling modifier κ_c and the ratio of coupling modifiers κ_c/κ_b is reduced. This is due to the parametrisation of the signal strength as a function of the coupling modifiers. This happens because for values of the signal strength close to the prediction of the SM, the effect of the coupling modifier κ_c on the width of the Higgs boson is reduced and the signal strength scales similarly to $\mu_{VH(\to c\bar{c})} \approx \kappa_c^2$. Therefore, the ratio of the coupling modifiers κ_c/κ_b is closer to the value of κ_c, which assumes a width of the Higgs boson as predicted by the SM.



(a) Signal strength (b) Ratio of coupling modifiers

Figure 7.5: Extrapolation of the sensitivity to the combination of $VH(\to b\bar{b})$ and $VH(\to c\bar{c})$ at 3000 fb^{-1} and $\sqrt{s} = 14$ TeV.

When considering only the statistical uncertainty on the extrapolation of the signal strength, the sensitivity to the $VH(\to c\bar{c})$ process corresponds to an expected excess with a significance of 0.5 standard deviation. Taking into account the systematic uncertainties, the projected sensitivity corresponds to an expected excess with a significance of 0.3 standard deviations.

However, this result does not take into account any improvements on analysis techniques and represents a conservative projection. In the short-term, possible improvements are achievable with the usage of machine learning techniques to separate more efficiently the signal from the background processes. In the long term, improvements on the c-tagging performance and calibrations will further increase the sensitivity to the $VH(\to c\bar{c})$ process. In order to reduce the impact of systematic uncertainties, dedicated measurements of the main background processes, such as the V+jets process, in a phase-space separate from the one used to analyse the $H \to c\bar{c}$ signal process would provide a valuable information, which could result in a simulation of these important background processes describing more accurately the data.

For the ultimate high-precision measurement of the interaction of the Higgs boson with charm quarks, a future collider will be needed. Several colliders are considered: the High-Energy LHC (HE-LHC), the Future Circular Collider (FCC-

ee,eh,hh) [93, 94, 95], the Circular Electron Positron Collider (CEPC) [96], the International Linear Collider (ILC) [97, 98], the compact Linear Collider (CLIC) [99] and the Large Hadron electron Collider (LHeC) [100]. With the goal of precision measurements of the Higgs boson properties in mind, electron-positron colliders produce events in a cleaner environment with respect to the proton-proton colliders, for which events are subject to a large background of QCD processes. Due to the energy losses from synchrotron radiation, the reach of the center-of-mass energy of electron-positron colliders is usually lower than the one achieved at proton-proton colliders. The main production mode for the Higgs boson at electron-positron colliders is the ZH associated production process and the energies projected to be reached usually range from 250 GeV and 500 GeV, well above the energy threshold for the production of the Z and Higgs bosons. Based on studies of the physics program at future colliders, all future colliders [101] will be able to observe the $H \to c\bar{c}$ process and the relative precision on the signal strength that can be reached is of the order of 2-20%.

7.3 Multivariate analysis in $VH(\to c\bar{c})$

In the $m(c\bar{c})$ distribution used to measure the signal strength of the $VH(\to c\bar{c})$ process, the signal is a resonant process peaking at 125 GeV while the background distribution is a falling spectrum. However, additional kinematic variables also have discriminating power between the $VH(\to c\bar{c})$ signal and the background processes and the inclusion of these variables in the analysis potentially add some information.

In order to exploit the correlations between the $m(c\bar{c})$ distribution and the other kinematic variables of the $VH(\to c\bar{c})$ process, multivariate analyses are performed using a boosted decision tree and a neural network. The output of the multivariate analyses is an observable distribution with a higher discriminating power between the signal and the background processes and therefore a higher sensitivity. These studies were conducted on the 2-lepton channel of the $VH(\to c\bar{c})$ analysis together with Zhuoran Feng in the context of her master thesis at Nikhef.

The choice of the input variables, shown in Table 7.3, is based on the measurements of the $VH(\to b\bar{b})$ process performed by the ATLAS collaboration and the measurement of the $VH(\to c\bar{c})$ process performed by the CMS collaboration. The set of variables is composed of the invariant mass of the various particles in the event, the transverse momentum, the pseudo-rapidity and the azimuthal angle difference in the transverse plane. As the studies focus on the 2-lepton channel of the $VH(\to c\bar{c})$ measurement, the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ is not used. The ratio of transverse momentum $p_{\mathrm{T}}^{V}/p_{\mathrm{T}}^{H}$ is also not considered but both the variables p_{T}^{V} and p_{T}^{H} are used as inputs to the multivariate analyses. The kinematic variables added with respect to the ATLAS and CMS analyses are $m(lc)$, $m(VH)$, $\Delta\eta(c\bar{c})$, $\Delta\eta(VH)$ and $\Delta\phi(c\bar{c})$. These variables are found to provide a low discrimination power between the signal and the background processes and could potentially be

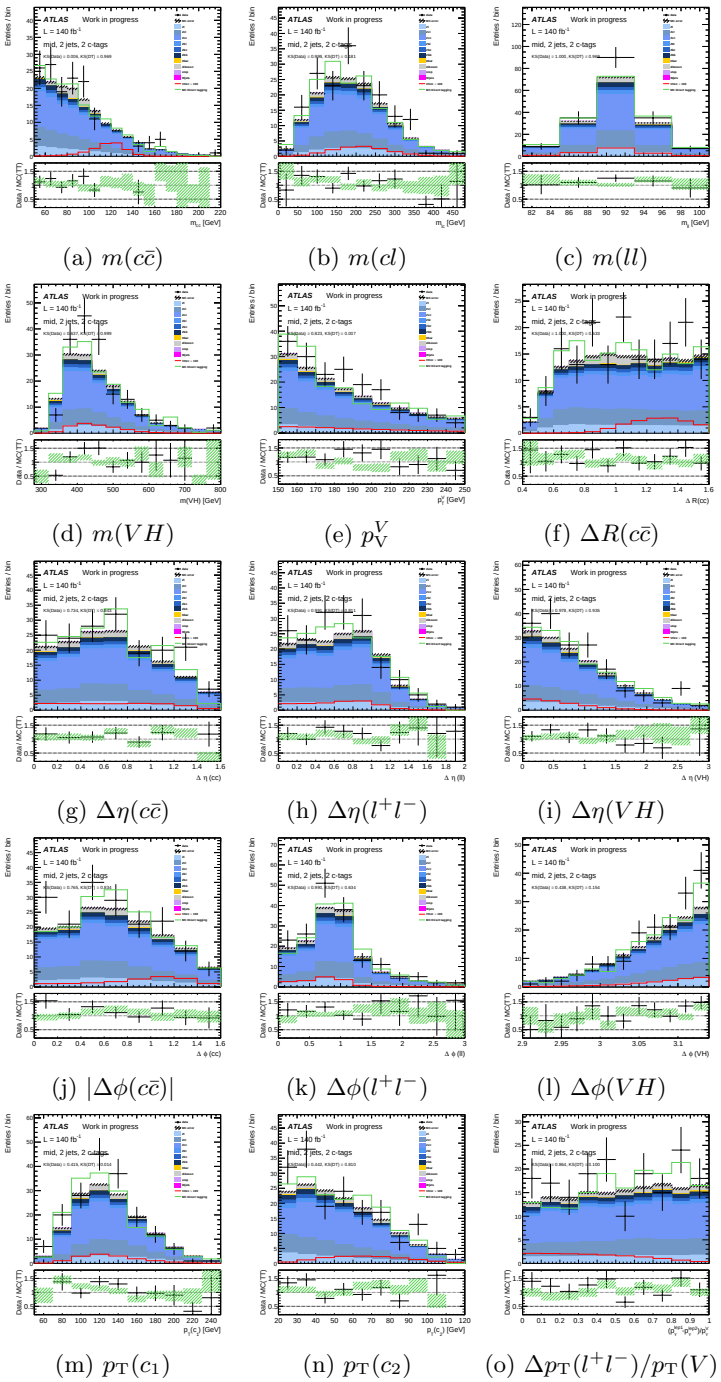
Variables	ATLAS $VH(\to b\bar{b})$	CMS $VH(\to c\bar{c})$	This study
$m(c\bar{c})$	×	×	×
$m(c\bar{c}j_3)$	×		×
$m(l^+l^-)$	×	×	×
$m(lc)$			×
$m(VH)$			×
$\Delta R(c\bar{c})$	×	×	×
$\Delta\eta(c\bar{c})$			×
$\Delta\eta(l^+l^-)$		×	×
$\Delta\eta(VH)$	×		×
$E_{\mathrm{T}}^{\mathrm{miss}}$	×		
$p_{\mathrm{T}}^{V}/p_{\mathrm{T}}^{H}$		×	
N_{jets}		×	
p_{T}^{H}		×	
p_{T}^{V}	×	×	×
p_{T}^{c1}	×	×	×
p_{T}^{c2}	×	×	×
p_{T}^{j3}	×	×	×
$\Delta\phi(VH)$	×	×	×
$\Delta\phi(c\bar{c})$			×
$\Delta\phi(l^+l^-)$		×	×
$\Delta p_{\mathrm{T}}(l^+l^-)/p_{\mathrm{T}}(V)$	×		×

Table 7.3: Variables used as inputs to the multivariate analyses.

removed in the future.

The multivariate analyses are trained on simulated events, where the signal is defined as the $VH(\to c\bar{c})$ process and the background is defined as all the other processes. In order to ensure that the output distribution of the multivariate analyses is describing the data accurately, each kinematic distribution is scrutinised and only the kinematic distributions where the simulation is found in agreement with the data are used as inputs. For all kinematic variables presented in Table 7.3, the simulation is consistent with the data within statistical uncertainties and their distributions are shown in Fig. 7.6.

Two types of machine learning algorithms are trained: a boosted decision tree and an artificial neural network. When trained on the same input variables, the two algorithms are expected to have a similar performance. The list of hyperparameters used for the training of the boosted decision tree and the neural network, summarised in Tables 7.4 and 7.5, are based on studies performed in the analysis of the $VH(\to b\bar{b})$ process. In the future, these parameters can be optimised further to achieve a higher sensitivity to the $VH(\to c\bar{c})$ process. Furthermore, similarly to the multivariate analysis used in the $VH(\to b\bar{b})$ measurement, the optimisation could be performed across the 0-, 1- and 2-lepton channels to find one set of hyper parameters optimising the sensitivity across all three channels.



(a) $m(c\bar{c})$ (b) $m(cl)$ (c) $m(ll)$

(d) $m(VH)$ (e) p_{V}^{V} (f) $\Delta R(c\bar{c})$

(g) $\Delta\eta(c\bar{c})$ (h) $\Delta\eta(l^{+}l^{-})$ (i) $\Delta\eta(VH)$

(j) $|\Delta\phi(c\bar{c})|$ (k) $\Delta\phi(l^{+}l^{-})$ (l) $\Delta\phi(VH)$

(m) $p_{\mathrm{T}}(c_1)$ (n) $p_{\mathrm{T}}(c_2)$ (o) $\Delta p_{\mathrm{T}}(l^{+}l^{-})/p_{\mathrm{T}}(V)$

Figure 7.6: Data and simulation comparison for the variables used in the multivariate analysis.

Hyper parameter	Value	Hyper parameter	Value
Boost type	Gradient boost	NTrees	200
Shrinkage	0.5	MaxDepth	4
SeparationType	GiniIndex	nCuts	100
PruneMethod	NoPruning	nEventsMin	5%

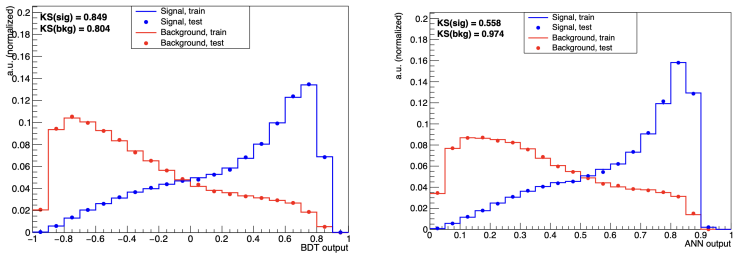
Table 7.4: The hyper parameters used in the boosted decision tree. The values are the same as the ones used in the boosted decision tree of the $VH(\to b\bar{b})$ analysis.

Hyper parameter	Value	Hyper parameter	Value
Hidden layers	4	Loss function	Binary cross entropy
Hidden layer nodes	250,125,60,30	Activation	SELU and sigmoid
OUtput nodes	1	Epochs	50
Batch size	1000	Dropout factor	0.2

Table 7.5: The hyper parameters used in the neural network, inspired by the boosted $VH(\to b\bar{b})$ analysis. The activation function is the Scaled Exponential Linear Unit (SELU) and sigmoid function $1/(1+e^{-x})$.

The output of the machine learning algorithms, illustrated in Fig. 7.7, discriminates the signal from the background: high values of the scoring observable correspond to signal-like events, while lower values correspond to events with characteristics of the background processes. Neither algorithm results in events with a maximum score of 1. The reason is that no events are identified as exclusively signal-like because the $VH(\to b\bar{b})$ process is treated as a background and no flavour tagging variables are used in the training. A possible improvement in the future could be achieved by treating the $VH(\to b\bar{b})$ process as a signal as well as including flavour tagging variables in the training of the algorithm. The algorithms are trained in categories based on the number of c-tagged jets, the jet multiplicity and the p_{T}^{V} categories. The categorisation is based on a preliminary categorisation used in the $VH(\to c\bar{c})$ analysis at the time of this study and includes an additional split at $p_{\mathrm{T}}^{V} = 250$ GeV. Events with $p_{\mathrm{T}}^{V} < 150$ GeV are not considered in this study.

In order to validate the training result, the performance of both algorithms is evaluated based on the significance of the score variable, compared to the significance obtained with the $m(c\bar{c})$ distribution. The result of the multivariate analyses, shown in Table 7.6, indicates that an improvement of 50% can be achieved with respect to using the $m(c\bar{c})$ distribution, when considering only the statistical uncertainty. The improvement is of similar level for the boosted decision tree and the artificial neural network, which validates both implementations, as similar results are expected when using the same input variables. As an additional validation, using a boosted decision tree in the measurement of the $VH(\to b\bar{b})$ results in a similar improvement of 50% on the sensitivity of the 2-lepton channel,



(a) Boosted decision tree

(b) Neural network

Figure 7.7: Example of the output distribution of the boosted decision tree and neural network for the category with 150 GeV $< p_{\mathrm{T}}^{V} <$ 250 GeV, 1 c-tagged jet and 3 jets or more in the event.

Category		$m(c\bar{c})$	BDT	ANN
150 $< p_{\mathrm{T}}^{V} <$ 250 GeV	2 jets, 1 c-tag	2.5	3.6 (+44%)	3.8 (+52%)
	2 jets, 2 c-tags	2.2	2.9 (+32%)	3.3 (+50%)
	$\geq$ 3 jets, 1 c-tag	2.1	3.1 (+48%)	3.0 (+43%)
	$\geq$ 3 jets, 2 c-tags	1.8	2.3 (+28%)	2.4 (+33%)
$p_{\mathrm{T}}^{V} >$ 250 GeV	2 jets, 1 c-tag	2.0	3.1 (+55%)	3.4 (+70%)
	2 jets, 2 c-tags	1.6	2.1 (+31%)	2.4 (+50%)
	$\geq$ 3 jets, 1 c-tag	1.6	2.4 (+50%)	2.2 (+38%)
	$\geq$ 3 jets, 2 c-tags	1.2	1.6 (+33%)	1.7 (+42%)
Total		5.4	7.7 (+43%)	8.1 (+50%)

Table 7.6: Signal significance (multiplied by a factor 100) using the $m(c\bar{c})$, boosted decision tree (BDT) and artificial neural network (ANN). The percentage represents the improvement with respect to $m(c\bar{c})$.

which also includes an optimisation of the hyper parameter space and takes into account systematic uncertainties.

In order to understand which variables provide the highest discrimination power between the signal and the background processes, the ranking is studied. The variables with the highest rankings are the variables related to the jets in the events: $m(c\bar{c})$, $\Delta R(c\bar{c})$, $\Delta\eta(VH)$, p_{T}^{V} and $p_{\mathrm{T}}(c_1)$. The variables related to the two leptons in the event usually rank low, as the main background Z+jets has a similar signature, when considering the leptons in the event. Additional improvements could be achieved by optimising the hyper parameters of both multivariate analyses.

Based on the expected sensitivity obtained in the projection for the High Luminosity LHC, the usage of multivariate analyses is expected to improve the sensitivity to the $VH(\to c\bar{c})$ process and could result in an expected excess of 1 standard deviation, when considering only the statistical uncertainty. This result

will be particularly of interest in the context of a future combination with the CMS and LHCb measurements, which could result in an excess of 0.5 to 1 standard deviations, taking into accounts the systematic uncertainties.

8 Conclusion

A direct measurement of the Higgs boson coupling to charm quarks through the $H \to c\bar{c}$ process is challenging for several reasons: the difficult identification of c-jets signatures in the detector, a large contamination of background processes with similar particles in the final state and the influence of sizable systematic uncertainties. In order to achieve a high sensitivity to the $H \to c\bar{c}$ process, these challenges have all been addressed.

From the measurement of the $H \to c\bar{c}$ process in the data, as well as the combined measurement of the $H \to b\bar{b}$ and $H \to c\bar{c}$ processes, several results are obtained:

- A limit on the signal strength of the $VH(\to c\bar{c})$ signal $\mu_{VH(\to c\bar{c})} < 26$ at 95% confidence level,
- A first direct constraint on the coupling modifier of the Higgs boson interaction with c-quarks $|\kappa_c| < 8.5$ at 95% confidence level,
- A first direct exclusion of the universality of the Higgs coupling to b- and c-quarks $|\kappa_c/\kappa_b| < 4.5$ at 95% confidence level,
- An observed excess of the $VZ(\to c\bar{c})$ process with a significance of 2.6σ,
- An observed excess of the $VW(\to cq)$ process with a significance of 3.8σ.

While the coupling of the Higgs boson to charm quarks is not observed yet, the limit on the signal strength is the most constraining limit on the direct coupling of the Higgs boson to charm quarks to date. For the first time, a direct limit on the coupling modifier $|\kappa_c| < 8.5$ is obtained, which allows to constrain any model predicting a large enhancement of the coupling of the Higgs boson to charm quarks. The combined measurement of the $H \to b\bar{b}$ and $H \to c\bar{c}$ processes allows to probe the ratio of the coupling modifiers κ_b and κ_c simultaneously. The main advantage of measuring the ratio is the cancellation of the effect of massive particles beyond the SM on the width of the Higgs boson, effectively allowing the probe of the interaction of the Higgs boson with b- and c-quarks only. The direct constraint $|\kappa_c/\kappa_b| < 4.5$ is an experimental proof that the coupling of the Higgs boson to c-quarks is weaker than to b-quarks, independent of any assumption on the type and number of new particles. This result allows thus to exclude any theory predicting a universal coupling of the Higgs boson to heavy quarks and represent a milestone in our understanding of the Higgs boson properties.

A projection of this measurement, extrapolated from the LHC to the High Luminosity LHC, indicates that a limit on the signal strength of $\mu_{VH(\to c\bar{c})} < 6.4$

at 95% confidence level is attainable with 3000 fb^{-1}. However, with the usage of machine learning techniques, the sensitivity to the $VH(\rightarrow c\bar{c})$ process might be improved by approximately 50%. When factoring in potential improvements in c-tagging algorithms and detector upgrades, a combination of the results of the ATLAS, CMS and LHCb experiments could result in a sensitivity of 0.5σ to 1σ by the end of the High Luminosity LHC. In our quest to further understand the laws of nature, the search for the $H \rightarrow c\bar{c}$ process will thus continue to represent an exciting probe for physics beyond the SM.

Bibliography

[1] Chris Quigg. Unanswered questions in the electroweak theory. *Annual Review of Nuclear and Particle Science*, 59(1):505–555, 2009.

[2] 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, 2012.

[3] LHCb collaboration. Search for $H^0 \to b\bar{b}$ or $c\bar{c}$ in association with a W or Z boson in the forward region of pp collisions. Sep 2016.

[4] ATLAS collaboration. Search for the decay of the higgs boson to charm quarks with the atlas experiment. *Phys. Rev. Lett.*, 120:211802, May 2018.

[5] CMS collaboration. Search for the standard model Higgs boson decaying to charm quarks. Technical Report CMS-PAS-HIG-18-031, CERN, Geneva, 2019.

[6] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[7] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[8] 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.

[9] ATLAS collaboration. Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{\text{s}} = 7$ and 8 TeV in the ATLAS experiment. 3 2015.

[10] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. Oct 2016. 869 pages, 295 figures, 248 tables and 1645 citations. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG.

[11] ATLAS collaboration. Combined measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment. Technical report, CERN, Geneva, Nov 2021. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-053.

[12] CMS collaboration. Evidence for Higgs boson decay to a pair of muons. *JHEP*, 01:148, 2021.

[13] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1–102, 2012.

[14] F. J. Botella, G. C. Branco, M. N. Rebelo, and J. I. Silva-Marcos. What if the masses of the first two quark families are not generated by the standard model Higgs boson? *Phys. Rev. D*, 94(11):115031, 2016.

[15] Daniel Egana-Ugrinovic, Samuel Homiller, and Patrick Roddy Meade. Higgs bosons with large couplings to light quarks. *Phys. Rev. D*, 100(11):115041, 2019.

[16] J. R. Klein and A. Roodman. Blind analysis in nuclear and particle physics. *Ann. Rev. Nucl. Part. Sci.*, 55:141–163, 2005.

[17] P. Butera, G. M. Cicuta, and E. Montaldi. Dimensional renormalization. *NATO Sci. Ser. B*, 5:1–13, 1974.

[18] Gerard 't Hooft. Dimensional regularization and the renormalization group. *Nucl. Phys. B*, 61:455–468, 1973.

[19] John C. Collins and Davison E. Soper. Parton Distribution and Decay Functions. *Nucl. Phys. B*, 194:445–492, 1982.

[20] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017.

[21] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.

[22] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.

[23] K. Amako et al. Geant4 and its validation. *Nucl. Phys. B Proc. Suppl.*, 150:44–49, 2006.

[24] ATLAS collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823–874, 2010.

[25] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[26] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[27] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011.

[28] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[29] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016.

[30] Oliver S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. LHC Design Report Vol.1: The LHC Main Ring. 6 2004.

[31] O. Buning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. LHC Design Report. 2. The LHC infrastructure and general services. 11 2004.

[32] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. LHC Design Report. 3. The LHC injector chain. 12 2004.

[33] ATLAS collaboration. ATLAS: Technical proposal for a general-purpose p p experiment at the Large Hadron Collider at CERN. 12 1994.

[34] ATLAS collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[35] CMS collaboration. CMS, the Compact Muon Solenoid: Technical proposal. 12 1994.

[36] CMS collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[37] LHCb collaboration. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[38] ALICE collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[39] ATLAS collaboration. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Phys. Lett. B*, 784:173–191, 2018.

[40] ATLAS collaboration. Observation of light-by-light scattering in ultraperipheral Pb+Pb collisions with the ATLAS detector. *Phys. Rev. Lett.*, 123(5):052001, 2019.

[41] ATLAS collaboration. Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector. *JHEP*, 04:098, 2019.

[42] Joao Pequenao and Paul Schaffner. How ATLAS detects particles: diagram of particle paths in the detector. Jan 2013.

[43] ATLAS collaboration. Expected performance of the ATLAS experiment: detector, trigger and physics. 2009.

[44] ATLAS collaboration. ATLAS Insertable B-Layer Technical Design Report. 9 2010.

[45] Heather Mary Gray. The Charged Particle Multiplicity at Center of Mass Energies from 900 GeV to 7 TeV measured with the ATLAS Experiment at the Large Hadron Collider. 2010.

[46] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, and E. Moyse. Concepts, Design and Implementation of the ATLAS New Tracking (NEWT). 3 2007.

[47] ATLAS collaboration. Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment. 3 2012.

[48] ATLAS collaboration. The Optimization of ATLAS Track Reconstruction in Dense Environments. 3 2015.

[49] ATLAS collaboration. Vertex Reconstruction Performance of the ATLAS Detector at "$\sqrt{s} = 13$ TeV". Technical Report ATL-PHYS-PUB-2015-026, CERN, Geneva, Jul 2015.

[50] G. Piacquadio, Kirill Prokofiev, and A. Wildauer. Primary vertex reconstruction in the ATLAS experiment at LHC. *J. Phys. Conf. Ser.*, 119:032033, 2008.

[51] CMS collaboration. Adaptive vertex fitting. *J. Phys. G*, 34:N343, 2007.

[52] ATLAS collaboration. Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. 6 2016.

[53] 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.

[54] ATLAS collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13$ TeV. 12 2020.

[55] ATLAS collaboration. Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV. 6 2015.

[56] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[57] ATLAS collaboration. Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 7 2020.

[58] ATLAS Collaboration. Secondary vertex finding for jet flavour identification with the ATLAS detector. Technical Report ATL-PHYS-PUB-2017-011, CERN, Geneva, Jun 2017.

[59] ATLAS Collaboration. Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC. Technical Report ATL-PHYS-PUB-2018-025, CERN, Geneva, Oct 2018.

[60] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[61] ATLAS collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(11):970, 2019.

[62] ATLAS collaboration. Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. *JHEP*, 08:089, 2018.

[63] ATLAS collaboration. Search for pair-produced third-generation squarks decaying via charm quarks or in compressed supersymmetric scenarios in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev. D*, 90(5):052008, 2014.

[64] ATLAS Collaboration. Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run. Technical Report ATL-PHYS-PUB-2017-013, CERN, Geneva, Jul 2017.

[65] ATLAS collaboration. Optimisation of the smoothing of b-jet identification efficiency and mistag rate simulation-to-data scale factors in ATLAS. Technical report, CERN, Geneva, 2020. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-004.

[66] ATLAS collaboration. Measurement of c-jet tagging efficiency in $t\bar{t}$ events using a likelihood approach. Technical Report ATL-COM-PHYS-2017-073, CERN, Geneva, Feb 2017.

[67] ATLAS collaboration. Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment. 2 2014.

[68] ATLAS collaboration. Calibration of light-flavour jet b-tagging rates on ATLAS proton-proton collision data at $\sqrt{s} = 13$˜TeV. 4 2018.

[69] ATLAS collaboration. Monte Carlo to Monte Carlo scale factors for flavour tagging efficiency calibration. Technical report, CERN, Geneva, May 2020. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-009.

[70] ATLAS collaboration. Studies on top-quark Monte Carlo modelling with Sherpa and MG5aMC@NLO. Technical Report ATL-PHYS-PUB-2017-007, CERN, Geneva, May 2017.

[71] ATLAS collaboration. Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector. 7 2020.

[72] ATLAS collaboration. Observation of higgs boson production in association with a top quark pair at the lhc with the atlas detector. *Physics Letters B*, 784:173–191, 2018.

[73] 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):292, 2016.

[74] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008.

[75] ATLAS collaboration. Tagging and suppression of pileup jets with the ATLAS detector. Technical Report ATLAS-CONF-2014-018, CERN, Geneva, May 2014.

[76] ATLAS collaboration. Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC. Technical Report ATL-PHYS-PUB-2015-045, CERN, Geneva, Nov 2015.

[77] ATLAS collaboration. Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment. Technical report, CERN, Geneva, Aug 2019. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-033.

[78] C. Badiali, F. A. Di Bello, G. Frattari, E. Gross, V. Ippolito, M. Kado, and J. Shlomi. Efficiency Parameterization with Neural Networks. 4 2020.

[79] A L Read. *Presentation of search results: the CLs technique*, volume 28. IOP Publishing, sep 2002.

[80] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. 2012.

[81] Wouter Verkerke and David Kirkby. The RooFit toolkit for data modeling, 2003.

[82] Lorenzo Moneta, Kevin Belasco, Kyle S. Cranmer, S. Kreiss, Alfio Lazzaro, Danilo Piparo, Gregory Schott, Wouter Verkerke, and Matthias Wolf. The RooStats Project. *PoS*, ACAT2010:057, 2010.

[83] J. Alcaraz et al. Precision Electroweak Measurements and Constraints on the Standard Model. 12 2010.

[84] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group. Jul 2013. Comments: 404 pages, 139 figures, to be submitted to CERN Report. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections.

[85] A. Bazavov et al. Up-, down-, strange-, charm-, and bottom-quark masses from four-flavor lattice QCD. *Phys. Rev. D*, 98(5):054517, 2018.

[86] ATLAS collaboration. Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector. *Eur. Phys. J. C*, 81(2):178, 2021.

[87] LHCb collaboration. Physics case for an LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era. 8 2018.

[88] Fady Bishara, Ulrich Haisch, Pier Francesco Monni, and Emanuele Re. Constraining Light-Quark Yukawa Couplings from Higgs Distributions. *Phys. Rev. Lett.*, 118(12):121801, 2017.

[89] ATLAS collaboration. Extrapolation of ATLAS sensitivity to $H \to b\bar{b}$ and $H \to c\bar{c}$ decays in VH production at the HL-LHC. Technical report, CERN, Geneva, Oct 2021. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-039.

[90] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2/2017, 10 2016.

[91] ATLAS collaboration. Expected performance of the ATLAS detector under different High-Luminosity LHC conditions. 2021.

[92] ATLAS collaboration. Expected performance of the ATLAS detector at the High-Luminosity LHC. 2019.

[93] A. Abada et al. FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1. *Eur. Phys. J. C*, 79(6):474, 2019.

[94] A. Abada et al. FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2. *Eur. Phys. J. ST*, 228(2):261–623, 2019.

[95] A. Abada et al. FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3. *Eur. Phys. J. ST*, 228(4):755–1107, 2019.

[96] Mingyi Dong et al. CEPC Conceptual Design Report: Volume 2 - Physics & Detector. 11 2018.

[97] Philip Bambade et al. The International Linear Collider: A Global Project. 3 2019.

[98] Keisuke Fujii et al. Tests of the Standard Model at the International Linear Collider. 8 2019.

[99] T. K. Charles et al. The Compact Linear Collider (CLIC) - 2018 Summary Report. 2/2018, 12 2018.

[100] J. L. Abelleira Fernandez et al. A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector. *J. Phys. G*, 39:075001, 2012.

[101] J. de Blas et al. Higgs Boson Studies at Future Particle Colliders. *JHEP*, 01:139, 2020.

Summary

Particle physics is at the centre of understanding the laws of Nature. Its aim is to describe the fundamental constituents of the universe, the elementary particles, and the interactions between them, the forces. Elementary particles are classified in two types: fermions, which compose all the visible matter in the universe, and bosons, which mediate the interactions between elementary particles. The Standard Model provides a description of elementary particles and their interaction. In total, there are 12 elementary fermions and 4 elementary gauge bosons that mediate the interactions between particles.

All these particles have fundamental properties: an electric charge, a color charge, a spin and a mass. In the Standard Model, the fermions are further classified in three different generations. The first generation is composed of the electron, the electron neutrino, and the up and the down quarks. In addition, there are two copies for each particle type, which have similar fundamental properties but are heavier, as illustrated in Fig. 8.1 for the charged fermions. In 2012, one additional particle was discovered, the Higgs boson, which interacts with elementary particles proportionally to their mass.

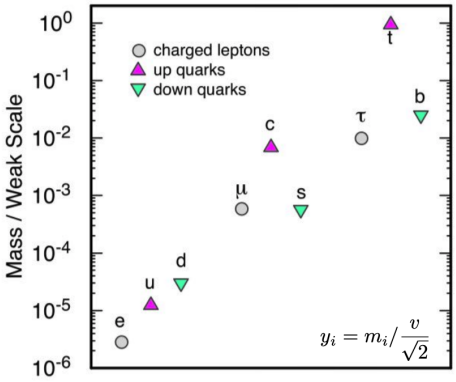


Figure 8.1: Mass of charged fermions ordered in the three different generations [1].

Despite the fact that the Standard Model provides a description of fundamental particles as well as their interactions, it does not explain the existence of exactly three generations of fermions, nor does it explain the large differences observed in their masses. It also does not explain the presence of dark matter in the universe.

The precise measurement of the interactions of the Higgs boson with elementary particles is intrinsically a probe for new physics, which could unravel a more complex structure in our understanding of Nature. In particular, the existence of new massive particles interacting with the Higgs boson could significantly modify its properties, which would in turn affect the interaction of the Higgs boson with other particles of the Standard Model. The discovery of physics beyond the Standard Model and the precise measurement of the properties of the Higgs boson are among the major goals of the Large Hadron Collider programs at CERN.

Since its discovery in 2012, the ATLAS and CMS experiments successfully measured the interaction of the Higgs boson to gauge bosons as well as to charged particles of the third generation of fermions. The first evidence of the Higgs boson interaction with muons was announced by the collaborations in 2020. All these measured interactions were found compatible with the prediction of the Standard Model. Yet, the interaction of the Higgs boson with the second generation of quarks remains unobserved. In particular, there is no direct constraint on the coupling of the Higgs boson to charm quarks and no experimental proof that this coupling is weaker than the coupling to the more massive bottom quarks, as predicted by the Standard Model since the charm quark is lighter than the bottom quark. A direct measurement of the Higgs boson coupling to charm quarks through the $H \to c\bar{c}$ process is challenging for several reasons: the identification of c-jets signatures in the detector is difficult, there is a large contamination of background processes with similar particles in the final state and the effect of systematic uncertainties is substantial.

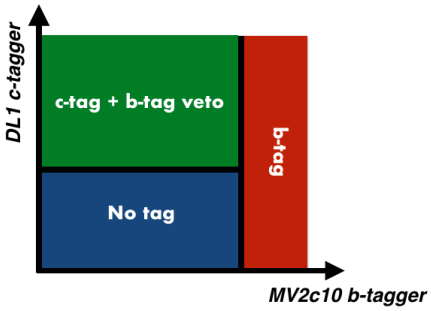


Figure 8.2: Flavour tagger used to c-tag and b-tag jets. For both taggers, a low score means a low probability the jet as the target particle.

This thesis gives a detailed account of the first direct determination of constraints on the coupling of the Higgs boson with charm quarks with the ATLAS detector. In order to reduce the contamination of background processes containing b-jets, the identification of c-quarks in the ATLAS detector is performed using a novel

c-tagging technique that includes a selection that explicitly removes b-tagged jets, shown in Fig 8.2. The performance of this new c-tagger, summarised in Table 8.1, is optimised to achieve the highest sensitivity to the $VH(\to c\bar{c})$ process.

Efficiencies	b-tagger	c-tagger including b-tag veto
ϵ_c	12%	27%
ϵ_b	70%	8%
ϵ_l	0.3%	1.6%
ϵ_τ	2.5%	20%

Table 8.1: Efficiencies of the c-tagging and b-tagging working points. The performance is estimated on a 139 fb^{-1} sample of simulated $t\bar{t}$ events generated with Powheg+Pythia8.

In the interest of further understanding the properties of the Higgs boson, events corresponding to the $H \to c\bar{c}$ process are selected in categories with 1 and 2 c-tagged jets, with an additional requirement to remove b-tagged jets, as shown in Fig. 8.3. This additional requirement, also applied on additional jets in the events, ensures that no events are categorised with 2 b-tagged jets in the $H \to c\bar{c}$ analysis and allows a combination with the measurement of the $H \to b\bar{b}$ process, which does not double count events. The combined measurement of the $H \to b\bar{b}$ and $H \to c\bar{c}$ processes is used to probe the ratio of the coupling modifiers κ_b and κ_c simultaneously. The main advantage of this joint measurement is to cancel the effect of massive particles beyond the Standard Model that interact with the Higgs boson and can modify its decay rates on the measurement, effectively allowing the probe enhancements of the interaction of the Higgs boson with bottom and charm quarks with minimal assumptions. To validate the measurement strategy, two processes of the Standard Model are measured simultaneously with the $H \to c\bar{c}$ process: the $W \to cs, cd$ and the $Z \to c\bar{c}$ processes.

The main result of the search for the $H \to c\bar{c}$ process, shown in Fig. 8.4 for the invariant mass distribution of the Higgs boson, is a limit on the production of the $VH(\to c\bar{c})$ signal smaller than 26 times the prediction of the Standard Model. While the interaction is not observed yet, this result represents the most sensitive limit on the direct coupling of the Higgs boson to charm quarks to date. In addition, for the first time, this result is used to set a direct constraint on enhancements of the coupling of the Higgs boson to charm quarks $|\kappa_c| < 8.5$, which allows to exclude theory models predicting large enhancements of the interaction of the Higgs boson with charm quarks. This direct constraint on κ_c is visualised as a function of the masses of particles in Fig. 8.5 alongside the other Higgs boson coupling constraints obtained from earlier analyses published by the ATLAS collaboration. The observed limits on κ_c allows to exclude the universality of the interaction of the Higgs boson to the top quarks and c-quarks.

In order to probe possible enhancements of the coupling of the Higgs boson with

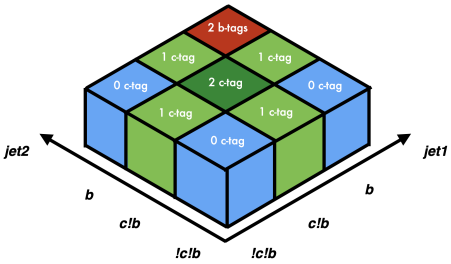
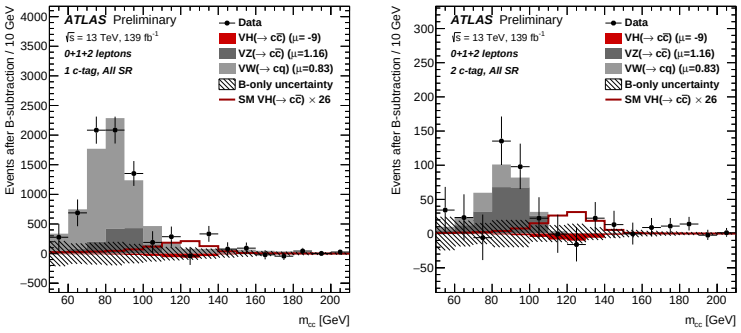


Figure 8.3: Illustration of the flavour tagging categorisation for events with 2 jets. Events with 2 b-tags are depicted in red, 2 c-tags in dark green, 1 c-tag in light green and 0 c-tag in blue.



(a) 1 c-tag (b) 2 c-tags

Figure 8.4: Result of the $m(c\bar{c})$ distribution after background substraction from the fit to data in the 1 c-tag (a) and 2 c-tags (b) categories.

bottom and charm quarks without any assumptions on the existence of new types of particles beyond the Standard Model, the ratio of the coupling modifiers κ_b and κ_c is measured. The result is a direct exclusion of the universality of the coupling of the Higgs boson to botton and charm quarks $|\kappa_c/\kappa_b| < 4.5$. This result is the first experimental proof that the Higgs boson couples less to charm quarks than to bottom quarks.

Based on the shown analysis approach, the result on the signal strength of the $H \to c\bar{c}$ process is extrapolated from the Large Hadron Collider to the High Luminosity Large Hadron Collider, which is planned to collect a data set 20 times larger than the data set used in this thesis. The extrapolated result predicts an expected limit on the signal strength $\mu_{VH(\to c\bar{c})} < 6.4$ at 95% confidence level. With the usage of machine learning techniques, the sensitivity to the $VH(\to c\bar{c})$

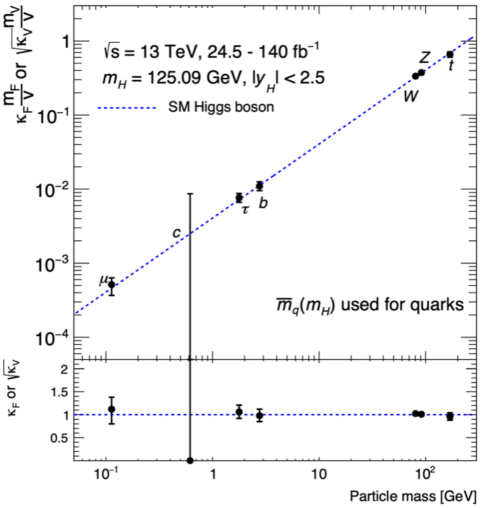


Figure 8.5: Coupling strength modifiers $\kappa_f m_f/v$ and $\sqrt{\kappa_V} m_V/v$ as a function of the masses of the fermions m_f and vector boson m_V. The constraint on the coupling modifier κ_c, corresponding to 68% CL, from the $VH(\rightarrow c\bar{c})$ analysis is illustrated along the other coupling modifiers that are obtained in a combined fit of analyses of the Higgs boson performed by the ATLAS collaboration [11].

process might be improved by approximately 50%. When considering potential improvements in c-tagging algorithms and detector upgrades, a combination of the results of the ATLAS, CMS and LHCb experiments could result in a sensitivity of 0.5σ to 1σ by the end of the High Luminosity Large Hadron Collider. An exciting phase of the Higgs physics program with the ATLAS detector has started and the increasing data set will allow to probe more precisely rare properties of the Higgs boson, in hope of unravelling a new layer of complexity in our understanding of Nature.

Samenvatting

De deeltjesfysica staat centraal bij het begrip van de natuurwetten. Het doel is om de fundamentele bestanddelen van het heelal, de elementaire deeltjes, en de interacties daartussen, de krachten, te beschrijven. Elementaire deeltjes worden ingedeeld in twee soorten: fermionen, die alle zichtbare materie in het universum vormen, en bosonen, die de interacties tussen elementaire deeltjes veroorzaken. Het standaardmodel geeft een beschrijving van elementaire deeltjes en hun interacties. In totaal zijn er 12 elementaire fermionen en 4 elementaire vectorbosonen.

Al deze deeltjes hebben fundamentele eigenschappen: een elektrische lading, een kleurlading, een spin en een massa. In het standaardmodel worden de fermionen verder ingedeeld in drie verschillende generaties. De eerste generatie bestaat uit het elektron, het elektron-neutrino en de up- en down-quarks. De tweede en derde generatie bestaan uit kopieën van de deeltjes uit de eerste generatie die vergelijkbare fundamentele eigenschappen hebben maar zwaarder zijn, zoals geïllustreerd in Figuur 8.6 voor de geladen fermionen. In 2012 werd nóg een deeltje ontdekt, het Higgs-deeltje, dat een interactie aangaat met elementaire deeltjes die evenredig is met hun massa.

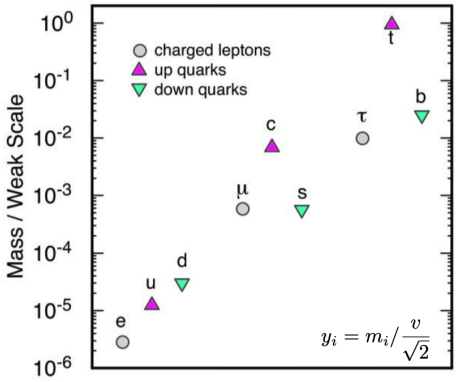


Figure 8.6: De massa's van de geladen fermionen gerangschikt in de drie verschillende generaties [1].

Ondanks het feit dat het standaardmodel een beschrijving geeft van de fundamentele deeltjes en hun interacties, verklaart het niet het bestaan van precies drie generaties fermionen, en evenmin de grote waargenomen verschillen in hun massa's.

De precieze meting van interacties van het Higgs-deeltje met elementaire deeltjes is van groot belang bij de zoektocht naar nieuwe fysica, die een oplossing zou kunnen bieden voor de tekortkomingen van het standaardmodel. Met name het bestaan van nieuwe massieve deeltjes die een interactie aangaan met het Higgs-deeltje zou de eigenschappen ervan aanzienlijk kunnen wijzigen, wat op zijn beurt de interactie van het Higgs-deeltje met andere deeltjes van het standaardmodel zou beïnvloeden. De ontdekking van fysica buiten het standaardmodel en de nauwkeurige meting van de eigenschappen van het Higgs-deeltje behoren daarom tot de belangrijkste doelen van het onderzoek aan de Large Hadron Collider bij CERN.

Sinds de ontdekking in 2012 hebben de ATLAS- en CMS-experimenten met succes de interacties tussen het Higgs-deeltje en bosonen gemeten, evenals interacties tussen het Higgs-deeltje en de geladen deeltjes van de derde generatie. Het eerste bewijs van de interactie van het Higgs-deeltje met muonen werd door beide collaboraties aangekondigd in 2020. Al deze gemeten interacties werden compatibel bevonden met de voorspellingen van het standaardmodel. De interacties tussen het Higgs-deeltje en de quarks van de tweede generatie zijn vooralsnog echter nog niet waargenomen. In het bijzonder is er tot op heden nog geen bovengrens gesteld aan de koppelingssterkte tussen Higgs-deeltjes en charm-quarks en is er nog geen experimenteel bewijs geleverd voor de voorspelling uit het standaardmodel dat deze koppeling zwakker is dan tussen Higgs-deeltjes en de zwaardere bottomquarks. Een directe meting van de Higgs-bosonkoppeling met charm-quarks via het $H \to c\bar{c}$-proces is om verschillende redenen een uitdaging: de identificatie van c-jets in de detector is moeilijk, er is een grote vervuiling door achtergrondprocessen met vergelijkbare deeltjes in de eindtoestand en het effect van systematische onzekerheden is aanzienlijk.

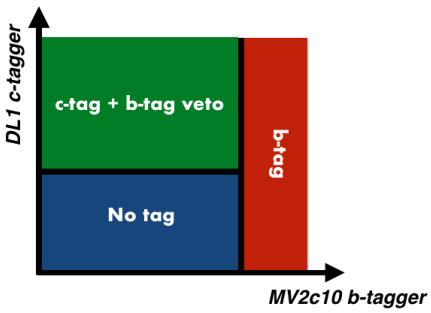


Figure 8.7: De flavour tagger identificeert b-jets en c-jets. Voor beide taggers betekent een lage score een lage waarschijnlijkheid dat de jet van een b-quark of c-quark komt.

Dit proefschrift geeft een gedetailleerd verslag van de eerste directe limieten die

zijn gesteld aan de koppeling tussen het Higgs-deeltje en charm-quarks met de ATLAS-detector. Om de relatieve bijdrage van achtergrondprocessen door b-jets te verminderen, wordt de identificatie van c-quarks in de ATLAS-detector uitgevoerd met behulp van een nieuwe c-tagging-techniek die expliciet b-quarks verwijdert. Figuur 8.7 toont de verschillende tags. De efficiëntie van deze nieuwe c-tagger, samengevat in Tabel 8.2, is geoptimaliseerd om de hoogste gevoeligheid voor het $VH(\rightarrow c\bar{c})$-proces te bereiken.

Efficiencies	b-tagger	c-tagger including b-tag veto
ϵ_c	12%	27%
ϵ_b	70%	8%
ϵ_l	0.3%	1.6%
ϵ_τ	2.5%	20%

Table 8.2: Efficiënties van de c-tagging en b-tagging selecties. De prestaties zijn bepaald aan de hand van 139 fb^{-1} gesimuleerde $t\bar{t}$ gebeurtenissen gegenereerd met POWHEG+PYTHIA8.

Om de eigenschappen van het Higgs-deeltje beter te begrijpen, worden gebeurtenissen die overeenkomen met het $H \rightarrow c\bar{c}$-proces ingedeeld in twee categorieën met respectievelijk 1 of 2 geïdentificeerde c-jets. Daarbij geldt als extra selectiecriterium dat alle geïdentificeerde b-jets worden verwijderd. Deze aanvullende eis, die ook van toepassing is op extra jets in de gebeurtenissen, zorgt ervoor dat er geen gebeurtenissen worden gecategoriseerd met 2 geïdentificeerde b-jets in de $H \rightarrow c\bar{c}$ analyse. Hierdoor wordt ook een combinatie met de meting van het $H \rightarrow b\bar{b}$ proces mogelijk gemaakt zonder dat gebeurtenissen dubbel worden meegeteld. Figuur 8.8 toont de selectiemethode. De gecombineerde meting van de $H \rightarrow b\bar{b}$ en $H \rightarrow c\bar{c}$ wordt gebruikt om de verhouding van de gewijzigde koppelingssterktes κ_b en κ_c gelijktijdig te bepalen. Het belangrijkste voordeel van deze gezamenlijke meting is dat eventuele interacties tussen het Higgs-deeltje en nog onontdekte massieve deeltjes, die de vervalskans van het Higgs-boson kunnen wijzigen, geen invloed uitoefenen op het eindresultaat. Hierdoor toetst de meting afwijkingen van de koppelingssterktes tussen het Higgs-boson en c- en b-quarks met minimale aannames. Om de meetstrategie te valideren, worden twee processen van het standaardmodel gelijktijdig gemeten met het $H \rightarrow c\bar{c}$ proces: de $W \rightarrow cs, cd$ en de $Z \rightarrow c\bar{c}$ processen.

Het belangrijkste resultaat van deze zoektocht naar het $H \rightarrow c\bar{c}$ proces, geïllustreerd in Figuur 8.9 met de invariante massaverdeling van het Higgs-deeltje, is een limiet op de productie van het $VH(\rightarrow c\bar{c})$ signaal: minder dan 26 maal de voorspelling van het standaardmodel. Hoewel de interactie nog niet is waargenomen, is dit resultaat de beste limiet op de directe koppeling van het Higgs-deeltje aan charm-quarks tot nu toe. Bovendien wordt dit resultaat voor de eerste keer gebruikt om een directe limiet te stellen op de koppelingssterkte tussen het Higgs-deeltje en

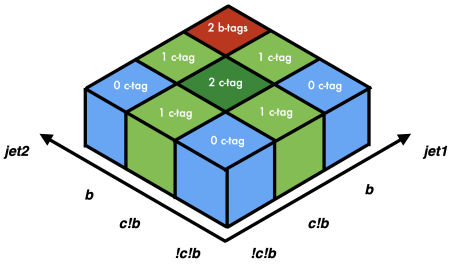
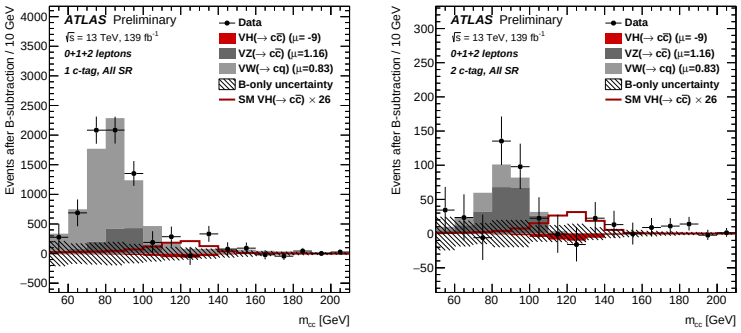


Figure 8.8: Illustratie van de categorisatie van b-jets en c-jets voor gebeurtenissen met 2 jets. Gebeurtenissen met 2 b-tags zijn afgebeeld in rood, 2 c-tags in donkergroen, 1 c-tag in lichtgroen en 0 c-tag in blauw.



(a) 1 c-tag (b) 2 c-tags

Figure 8.9: De $m(c\bar{c})$-verdeling na aftrek van de achtergronden ten opzichte van de fit op de data in de categorieën voor 1 c-tag (a) en 2 c-tags (b).

charm-quarks, $|\kappa_c| < 8.5$, wat het mogelijk maakt om theoretische modellen uit te sluiten waarin een grotere koppelingssterkte wordt voorspeld. In Figuur 8.10 wordt deze directe limiet op κ_c gevisualiseerd als een functie van de deeltjesmassa's, samen met limieten op andere koppelingssterktes van het Higgs-boson die zijn verkregen uit eerdere metingen van de ATLAS-collaboratie. De waargenomen limiet op κ_c maakt het mogelijk om uit te sluiten dat de interactie van het Higgs-deeltje met top-quarks en c-quarks hetzelfde is.

Om de koppeling van het Higgs-deeltje met bottom- en charm-quarks te onderzoeken zonder enige aannames over het bestaan van nieuwe deeltjes buiten het standaardmodel, is de verhouding van de koppelingssterktemodificatoren κ_b en κ_c gemeten. Deze meting sluit een gelijke koppeling tussen het Higgs-deeltje en bottom-quarks en tussen het Higgs-deeltje en charm-quarks uit: $|\kappa_c/\kappa_b| < 4.5$.

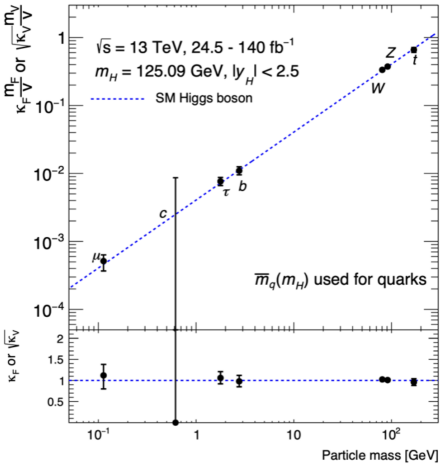


Figure 8.10: De koppelingssterktemodificatoren $\kappa_f m_f/v$ en $\sqrt{\kappa_V} m_V/v$ als functie van de massa's van de fermionen m_f en vectorbosonen m_V. De limiet op de koppelingssterktemodificator κ_c, overeenkomend met 68% CL, uit de $VH(\rightarrow c\bar{c})$-analyse wordt geïllustreerd met de andere koppelingssterktemodificatoren die zijn verkregen uit een gecombineerde fit van analyses van het Higgs-deeltje uitgevoerd door de ATLAS-collaboratie [11].

Dit resultaat is het eerste experimentele bewijs dat het Higgs-deeltje minder sterk koppelt aan charm-quarks dan aan bottom-quarks.

Op basis van de gebruikte analyse is het resultaat van de signaalsterkte van het $H \rightarrow c\bar{c}$-proces geëxtrapoleerd van de Large Hadron Collider naar de High Luminosity Large Hadron Collider, waarmee naar verwachting een dataset zal worden verkregen die 20 keer groter is dan de dataset die in dit proefschrift is gebruikt. Het geëxtrapoleerde resultaat voorspelt een limiet voor de signaalsterkte $\mu_{VH(\rightarrow c\bar{c})} < 6.4$ met een betrouwbaarheid van 95%. Met het gebruik van machine learning-technieken kan de gevoeligheid voor het $VH(\rightarrow c\bar{c})$-proces met ongeveer 50% worden verbeterd. Samen met mogelijke verbeteringen in c-tagging-algoritmen en detectorupgrades, zou een combinatie van de resultaten van de ATLAS-, CMS- en LHCb-experimenten uiteindelijk kunnen resulteren in een gevoeligheid van 0.5σ tot 1σ tegen het einde van de High Luminosity Large Hadron Collider. Kortom: er breekt een spannende tijd aan voor het Higgs-fysica programma binnen de ATLAS-collaboratie. De groeiende dataset zal het mogelijk maken om nauwkeuriger zeldzame eigenschappen van het Higgs-deeltje te onderzoeken, wat hopelijk leidt tot nieuwe, verbeterde inzichten in de natuur.

Acknowledgements

Hard to believe that this is the last chapter of this amazing PhD I had. I will try to keep this chapter shorter than my detector chapter. There is many people who deserve to be thanked for this adventure.

First, I would like to thank Tristan du Pree, your guidance and support shaped these 4 years. You always encouraged me to be creative, to work hard to solve problems and to keep the big picture in mind. I am very proud of the results we obtained together, especially the combination of the VHbb and VHcc analyses and the κ_c interpretation. These required forward thinking and is certainly one big lesson I learned during this PhD. I will always be grateful to you for supporting me for my ideas, reviewing my grants and research proposals but also celebrating the moments that deserved to be celebrated. I would also like to thank Wouter Verkerke, your time, guidance and support were necessary to finish this PhD. You were always available when I needed you and your advice helped me go through some difficult times, thank you for that.

I also want to thank Marcel Merk, Marco Delmastro, Eric Laenen, Frank Filthaut, Frank Linde and Jory Sonneveld for being in my thesis committee and reviewing my thesis. A special thank you to Marcel for being my C3 member during my PhD, thank you for following me during this thesis and making sure every thing was going well. Special thanks to Frank too, our discussions about flavour tagging and the regular coffee meetings during the pandemic were important for me. Special thanks to Marco too, your support in my job and grant applications really means a lot to me.

Many thanks to Heather Gray too, your support for my grant proposal, your new ideas to improve flavour tagging and your guidance throughout my visit were precious. These 6 months at LBL were among the best of my PhD and I will always be grateful to you for enabling it.

Next I would like to thank Hannah Arnold and Brian Moser. This PhD would have been very different without you. I learned so much with you and we went through all the highs and lows of this PhD together, but I always knew we would get to where we wanted because we were together. Thank you for always being there, I can only wish to every new PhD student to have you as companions, for both the laughs and the physics. I am very lucky to have you in my life.

The VHcc analysis would not have been possible without the amazing team we had, I would like to thank Andy Chisholm, Maria Mironova, Tom Neep, Antonio Costa, Guy Koren, Hadar Cohen, Jonathan Shlomi, Zijun Xu and Tao Wang. This was not easy but we all managed together. In particular, a very special

thank you to Maria, we solved the unsolvable together and your friendship and support throughout the end of the analysis was really necessary. Thanks! For my involvement in the flavour tagging of ATLAS, I would like to warmly thank Andrea Coccaro, Chris Pollard, Carlo Schiavi and Valerio Dao, releasing so many CDIs together is something I will never forget. Thank you for all your support, it was always a pleasure to work together. A special thanks to Andrea and Valerio, your support for my job search also means a lot to me. A very special thanks to Zhuoran, Cecile and Riley, supervising you as students was a privilege for me and working with you was always a great pleasure. I wish you all the best!

For all the great moments I had in Amsterdam, I would like to thank Rahul and Ash too, being your housemate was amazing. I would like to thank Bouke too, this pandemic was more bearable thanks to the dinners we had all together with Brian. I would also like to thank Marc, Broos and Karel, you were among my first friends in Amsterdam and I am glad we managed to see each other as much as possible. Many thanks to Bryan, Jordy and Kees too, especially that one night at Paradiso.

Thank you to all the people at Nikhef who made this PhD so special, first the ATLAS group: Alice, Annamika, Michiel, Edwin, Lydia, Flavia, Sascha, Pamela, Geoffrey, Nicolo, Peter, Marion, Clara, Federica, Alessio, Hella, Bob, Marcel and Ivo. Next I would also like to thank Michele, Brian, Patrick, Auke-Pieter, Stan, Niels and Jacco too. Many thanks to Martijn for all the help doing outreach. Special thanks to Maureen too, for all the coffees and other drinks together!

A huge thanks to the LBL team and all the great people I met in California. Special thanks to Miha, Kaja, Elodie and Marija, you made Berkeley feel like a second home and counting you as friends is really a privilege for me. In particular, Miha, working and talking about physics with you was among the best experience I had. Special thanks to Elodie too, you always had the best stories and I am so happy I got to know you. Special thanks to Mariel too, finally meeting you was a great experience! A special thanks to Marjorie, Tomohiro, Louis-Guillaume, Irina, Robin, Johannes, Joaquin, Cesar and Greg, working with you was amazing. And special thanks to all the LBL people who made this trip memorable: Aleksandra, Rebecca, Timon, Hongtao, Xiangyang, Zhicai, lil Jay, Elliot and Danny.

Getting to the end, I would also like to thank all my friends in Switzerland and Europe, I missed you all dearly during these 4 years. Special thanks to Gregoire first, we were already best friends for a long time but being in Amsterdam with you really made a huge difference. Many thanks to Loic, I always loved seeing you when I was back. Thanks to my friends from university: Morgan, Jordan, Scott, Ivan, Matthieu and Yann! Thanks to Camille and Guillaume too! Thanks to Sam for the thesis cover! A big thank you to May and Chloe! And a huge thanks to Marie, I wish I could celebrate this success with you but I know you always believed in me, I miss you.

Finally, a big thanks to my parents Predrag and Milojka, to my brother and sister Radivoje and Suzana, their partners, Charlotte and Stephane and my niece and nephew Alice and Elie, your support was unconditional and I needed it!