UNIVERSIDADE D
COIMBRA

Luis Felipe Falda Ulhoa Coelho

STUDY OF THE CP NATURE OF THE TOP-HIGGS COUPLING IN TTH EVENTS AT THE LHC

VOLUME 1

Dissertação no âmbito do Mestrado em Física Nuclear e de Partículas orientada pelo Professor Doutor José Ricardo Morais Silva Gonçalo e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Outubro de 2020

UNIVERSIDADE D
COIMBRA

FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Luis Felipe Falda Ulhoa Coelho

Study of the CP nature of the top-Higgs coupling in $t\bar{t}H$ events at the LHC

Thesis submitted to the
University of Coimbra for the degree of
Master in Nuclear and Particle Physics

Supervisors:
Prof. Dr. José Ricardo Morais Silva Gonçalo

Coimbra, 2020

Resumo

No Modelo Padrão (SM), é previsto que o bosão de Higgs seja uma partícula escalar e que suas interações não violem a simetria CP. Após a observação da produção do bosão de Higgs em associação com um par de quarks top ($t\bar{t}H$) pelas experiências ATLAS e CMS em 2018, a observação de uma componente ímpar à transformações de carga-paridade (CP) em um dos acoplamentos do bosão de Higgs constituiria uma importante descoberta de física além do SM (BSM). Recentemente, ATLAS e CMS procuraram essa componente em eventos $t\bar{t}H$ com o Higgs decaindo em dois fotões. No entanto, o acoplamento entre o Higgs e os fotões é induzido por loops e pode ser afetado por efeitos da nova física. Esta tese descreve o estudo, atualmente na colaboração ATLAS, da natureza CP do acoplamento de Yukawa do Higgs aos quarks top, por meio da análise de eventos $t\bar{t}H$ no canal de decaimento $H \to b\bar{b}$. Também fornece projeções deste estudo do Run 2 do Large Hadron Collider (LHC) ao LHC de alta luminosidade (HL-LHC).

A análise usa dados de colisão protão-protão coletados com o detector ATLAS durante o período do Run 2 do Grande Colisor de Hadrões (LHC) com uma energia de centro de massa de $\sqrt{s} = 13$ TeV e luminosidade integrada total de 139 fb^{-1}. São usados apenas eventos contendo um ou dois leptões no estado final do decaimento do par de quarks top. Em seguida, os eventos são classificados em regiões de acordo com o número de jatos e o número de b-jatos. Várias técnicas multivariadas foram usadas para melhorar a sensibilidade da análise, uma árvore da decisão reforçada (BDT) foi treinada para separar o sinal do fundo e outra para distinguir entre diferentes cenários de CP. Variáveis sensíveis ao CP, incluindo observáveis calculadas no referencial de laboratório e variáveis angulares calculadas em referenciais específicos, foram usadas no ajuste e no treinamento das BDTs. Um ajuste de verossimilhança é executado em todas as regiões de análise para restringir as previsões de fundo e reduzir as incertezas sistemáticas. O valor esperado para o ângulo de mistura CP é obtido desse ajuste.

Além disso, uma extrapolação da análise foi realizada nesta tese, a fim de fornecer projeções sobre a medição do ângulo de mistura de CP para o LHC de alta luminosidade (HL-LHC). Foram considerados diferentes cenários para a evolução das incertezas sistemáticas com o aumento esperado da luminosidade. O valor esperado para o ângulo de mistura de CP foi obtido para vários valores diferentes de luminosidade até 3000 fb^{-1}. Com as atuais incertezas sistemáticas, espera-se que a produção de ttH pura CP-ímpar seja excluída com 99.73% de nível de confiança (CL) apenas no final do projeto HL-LHC. A significância de exclusão CP-ímpar é representada em função da luminosidade para cada um dos cenários considerados, e os efeitos dos vários tipos de incertezas são avaliados.

Palavras-Chave: Bosão de Higgs, Produção associada à quarks top, Violação CP, HL-LHC, Yukawa

Abstract

In the Standard Model (SM), the Higgs boson is a scalar particle with no CP-violating interactions. After the observation of the Higgs boson production in association with a top quark pair ($t\bar{t}H$) by ATLAS and CMS in 2018, the measurement of an odd charge-parity (CP) component in one of the Higgs boson couplings would constitute an important discovery of physics beyond the SM (BSM). Recently, ATLAS and CMS searched for such a component in $t\bar{t}H$ events with the Higgs decaying into two photons. However, the coupling between the Higgs and the photons is loop-induced and could be modified by effects of new physics. This thesis describes the ongoing study of the CP nature of the Higgs Yukawa coupling to the top quarks by analyzing $t\bar{t}H$ events in the $H \to b\bar{b}$ decay channel, and provides projections of this search from the Large Hadron Collider (LHC) Run 2 to the High-Luminosity LHC (HL-LHC).

The analysis uses the full Run 2 dataset of proton-proton collision collected with the ATLAS detector at a center-of-mass energy of $\sqrt{s} = 13$ TeV and total integrated luminosity of 139 fb^{-1}. Only events containing either one or two leptons in the final state from the decay of the top quark pair are used in the analysis. Then, the events are classified into regions according to the number of jets and the number of b-tagged jets. Two sets of multivariate classifiers are utilized to improve the analysis sensitivity. One classifier targets the classification of signal against backgrounds and the other targets the separation between different CP scenarios. Several CP sensitive variables, including lab-frame observables and angular variables calculated in specific frames, were used in the fit and in the training of the BDTs. A profile likelihood fit is performed over all analysis regions to constrain the background predictions and reduce the systematic uncertainties. The expected value for the CP mixing angle is presented.

An extrapolation of the analysis is performed in order to provide projections

on the measurement of the CP mixing angle for the HL-LHC. Different scenarios for the evolution of the systematic uncertainties with the expected increase in the luminosity were considered. The expected value for the CP mixing angle was obtained for several different values of luminosity up to 3000 fb^{-1}. With the current systematic uncertainties, the pure CP-odd $t\bar{t}H$ production is expected to be excluded with 99.73% confidence level (CL) only at the end of the HL-LHC project. The CP-odd exclusion significance is represented as a function of the luminosity for each of the scenarios considered, and the effects of the various types of uncertainties are evaluated.

Keywords: Higgs boson, Top quark associated production, CP violation, HL-LHC, Yukawa

Acknowledgments

First, I would like to thank my advisor Ricardo Gonçalo for allowing me to participate in this project, for thorough reviews of this thesis, and his constant support and invaluable guidance during this year.

I would also like to thank Ana Luisa Carvalho and Emanuel Gouveia for the time devoted to answering my questions, helping me solve my problems with the TRExFitter software and for comments on presentations and this thesis.

I am thankful to Patricia Muiño, António Onofre, Orlando Oliveira, Filipe Veloso and Alexandre Lindote for their support and valuable advise through this year during meetings, seminars and more.

I would like to thank the jury members João Carvalho and Nuno Castro for taking the time to read my thesis and listen to my defense.

I am grateful for the financial support of Fundação para a Ciência e Tecnologia and LIP, through the CERN/FIS-PAR/0002/2019 grant.

I would also like to thank Jelena Jovicevic and Nedaa Asbah for their advise during the CERN summer initiative.

I am grateful to all my friends and colleagues from the University of Coimbra for the good company and support during the years at the university, specially Guilherme, Joaquim, Miguel, Diogo, Salah, Felipe and Afonso.

Most of all, I would like to express my very profound gratitude to my family and Heloisa for their constant support and motivation.

Contents

List of Tables

List of Figures

1

Introduction

The Standard Model (SM) of particle physics describes the known fundamental particles and their interactions. It is considered as one of the greatest scientific achievements in recent times, predicting several processes and observables that were later confirmed by experiments. The Higgs boson discovery [1, 2] at CERN by both the ATLAS and CMS collaborations in 2012 was a meaningful accomplishment for experimental and theoretical particle physics. It was the last particle of the SM to be observed and this discovery confirms the existence of an associated Higgs field that is responsible for the generation of the weak boson masses through the Electroweak Symmetry Breaking mechanism and, as a consequence, the fermion masses through Yukawa couplings. Since the discovery, the measured properties of the Higgs boson have shown a remarkable agreement with the SM predictions. However, despite its success, the SM does not describe all of the observed physical phenomena. In particular, it does not explain the baryon asymmetry in the Universe and neither provides a candidate for dark matter.

The observed asymmetry between matter and antimatter could be explained by introducing additional sources of charge-parity (CP) violation beyond the SM (BSM), one of them could come from the Higgs sector. Several BSM models allow for the Higgs boson to be a mixed state particle, resulting in a (CP-violating) Yukawa coupling with two components, one CP-even and one CP-odd. The mixing of these two components is dictated by an angle referred to as CP mixing angle. Since a mixed state is still not excluded by experimental data, it is crucial to determine the CP structure of the Yukawa couplings. The production of the Higgs boson in association with a top quark-antiquark pair ($t\bar{t}H$) is the best process for this measurement, since it provides a direct access to the large top quark Yukawa coupling. Unfortunately, the $t\bar{t}H$ production is a rare process, that was only observed in 2018 [3, 4] at the Large Hadron Collider (LHC). It accounts for only 1% of the total Higgs boson production cross section. For this reason, data

from the High Luminosity LHC (HL-LHC) project, which is a future upgrade to the LHC that aims to increase its luminosity, will be needed to obtain sensitivity for this measurement. The goal of the HL-LHC upgrade is to expand the LHC potential for new discoveries and precision in the measurements of rare processes.

This thesis describes an ongoing search for the $t\bar{t}H$ CP-odd component using the full Run 2 dataset of proton-proton collision collected during 2015 and 2016 with the ATLAS detector at a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to a total integrated luminosity of 139 fb^{-1}. The analysis targets the single lepton and dilepton channels of the Higgs decay mode to b-quarks, which is the Higgs boson decay mode with the largest branching fraction. This channel suffers from the presence of a large and badly modelled irreducible background from the production of top-quark pairs with two additional b-jets ($t\bar{t}b\bar{b}$), which makes the analysis even more challenging.

In addition, this thesis provides projections on the measurement of the CP mixing angle for the HL-LHC and gives predictions of when the pure CP-odd $t\bar{t}H$ production will be excluded with 99.73% confidence level (CL). The systematic uncertainties are expected to change during the running of the HL-LHC and it is a significant challenge to predict their values. Therefore, different scenarios were considered and the expected CP-odd exclusion was obtained for each one of these scenarios.

In Chapter 2 of this thesis, an overview of the theoretical formulation of the SM is presented, followed by a description of the Higgs mechanism and the Yukawa couplings. The Higgs boson production and decay modes at the LHC are also introduced, together with the experimental status of the Higgs boson.

Chapter 3 introduces CERN's LHC and the HL-LHC project. It also details the ATLAS detector, its components and the trigger and data acquisition system.

In Chapter 4, the state of the art regarding the study of $t\bar{t}H$ CP properties is presented with its theoretical motivations. The chapter starts with a description of the $t\bar{t}H$ production, followed by an explanation of CP sensitive variables used in this thesis, and by the experimental status of $t\bar{t}H$ CP-odd production.

Chapter 5 summarizes the generation of Monte Carlo samples and presents the signal and background samples used in the analysis, as well as the data samples. It also describes the reconstruction and identification of physical objects. In addition, a comparison between a fast detector simulation with the ATLAS full simulation is

also given, with the intention of validating the modelling of substructure variables of large-R jets.

Chapter 6 describes the analysis strategy. This chapter details the event selection and classification into regions, as well as the techniques applied for distinguishing between the CP-odd and CP-even $t\bar{t}H$ production and the background. It also explains the fit used to measure the expected CP mixing angle, the systematic uncertainties present in the analysis and a comparison between two different strategies that could be applied.

In Chapter 7, a discussion of the main results is presented. The chapter starts by describing the expected limit for the CP mixing angle. Then, the results of the study of the projected sensitivity of the HL-LHC to the CP mixing angle is presented. The analysis is extrapolated for higher values of luminosity and effects of different scenarios for the systematic uncertainties are considered.

Finally, in chapter 8, an overall summary of this thesis is given and conclusions are drawn.

2

The Standard Model of Elementary Particles

The SM is described with significant detail in this chapter, starting with an overview of its particle content in Section 2.1 and the underlying group structure, and the Lagrangian density that governs the particle fundamental interactions and describes the theory in Section 2.2. Particular emphasis is given to the Higgs mechanism of Electroweak Symmetry Breaking (Section 2.3), explaining how particles acquire mass through the interaction with the Higgs field (Section 2.4). Then, the production and decay channel of the Higgs boson and other experimental aspects of the Higgs production are presented in Section 2.5 and 2.6. Finally, a description of the discovery of the Higgs Boson and of its measured properties is given in Section 2.7.

2.1 Overview

The SM is a gauge field theory that describes the interactions of elementary matter with three of the four known fundamental forces: the electromagnetic force, the weak nuclear force, and the strong nuclear force (the gravitational force is not included). Fundamental particles in the SM are divided into fermions and bosons. Fermions are spin-1/2 particles that obey Fermi-Dirac statistics and compose all the known matter in our Universe. Bosons are integer spin particles that obey Bose-Einstein statistics. The gauge bosons, which are spin-1 bosons, act as force carriers and mediate the interactions between particles. All the particles described by the SM and their properties (mass, electric charge, color charge, spin) are summarized in Figure 2.1.

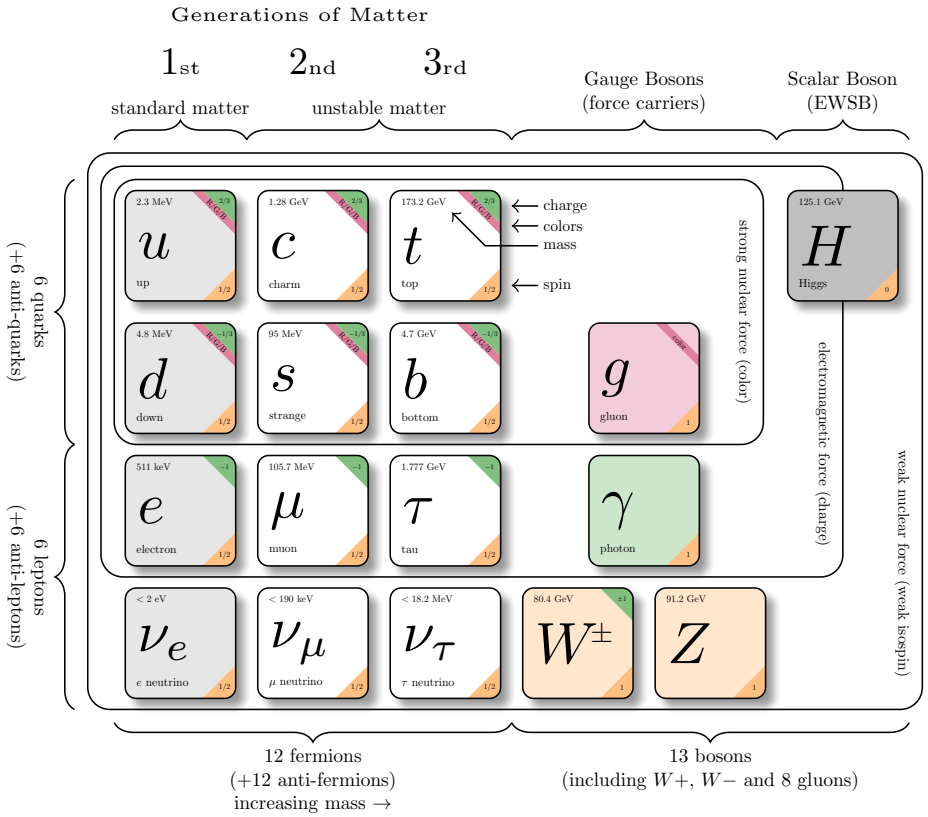


Figure 2.1: Particles and interactions of the SM. The diagram shows the three generations of fermions (quarks and leptons), the gauge vector bosons (gluons, photon, $W^{\pm}$, and Z), and the scalar Higgs boson. The masses are displayed in the left upper corner of each box, the electric and color charge are displayed in the right upper corner and the spin in the right lower corner. Image adapted from [5].

Quarks and Leptons are fermions, there are six flavors of quarks and six flavors of leptons. They are organized into three families, each family is constituted by two quarks and two leptons. There are in total 12 fermions, each associated with an antiparticle with the same mass but with opposite electric charge and quantum numbers.

Quarks are the fundamental constituents of hadrons, each of them carries one of the three color charges of Quantum Chromodynamics (QCD), named red, green and blue, and a fractional electric charge of Quantum Electrodynamics (QED). The different flavors of quarks are named as up (u), down (d), charm (c), strange (s), top (t) and bottom (b). As shown in Figure 2.1, the quarks in the top row

carry electric charge of $q = {}^{2}/_{3}|e|$, whereas the quarks in the bottom row have $q = {}^{-1}/_{3}|e|$.

Leptons carry no color charge, they are divided into the electrically charged leptons: the electron (e), muon (μ) and tau (τ), with electric charge $q = -|e|$, and their electrically neutral neutrino counterparts: the electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ).

The interaction between fermions occurs through the exchange of gauge bosons and each gauge boson mediates one interaction. The photon (γ), which is electrically neutral and massless, mediates the electromagnetic interaction and couples with all electrically charged particles (charged leptons, quarks, and charged gauge bosons), this interaction is described by quantum electrodynamics (QED).

The gluon (g) exists in eight different states, it is the mediator of the strong nuclear force and interacts with all particles that carry color charge (quarks and gluons). The gluons themselves carry color, therefore in addition to mediating the strong interaction, they also participate in it. This makes quantum chromodynamics (QCD), the theory that describes strong interactions, more complicated than QED.

The W^+, W^- and Z bosons are the mediators of the weak nuclear force. All quarks and leptons participate in weak interactions. There are two types of weak interactions: the charged one, mediated by the W^+ and the W^- bosons, and the neutral one, mediated by the Z boson. The W^+ and the W^- have a mass of 80.379 ± 0.012 GeV [6] and electric charge of $q = +|e|$ and $q = -|e|$, whereas the Z boson has a mass of 91.1876 ± 0.0021 GeV [6] and is electrically neutral. The significant masses of its mediators explain the extremely short range of the weak force, which is about 0.1% of the diameter of a proton. The weak interaction is also the only known interaction capable of changing the flavor of quarks and that violates charge-parity (CP) symmetry.

In addition to quarks, leptons and gauge boson, one electrically neutral scalar boson (spin-0) is needed to make the SM a consistent theory. This scalar boson is called Higgs boson and has a mass of 125 GeV. The Higgs field, from which excitations originate the Higgs boson, is related to a mechanism called Higgs mechanism [7–12]. This mechanism is responsible for giving rise to the non-zero masses of the weak interaction gauge bosons (W^+, W^- and Z) through a process called "Electroweak Symmetry Breaking" (discussed in Section 2.3).

2.2 The Standard Model Lagrangian

In a quantum field theory, every particle is described as an excitation of a quantum field that is defined over all spacetime. All the interactions of the fields are described by a Lagrangian density $\hat{\mathcal{L}}$, which is Lorentz invariant, i.e. is the same for every reference frame. Besides, the SM is a local gauge theory, which means that $\hat{\mathcal{L}}$ is invariant under certain local gauge transformations. The underlying gauge symmetry of the SM is $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, the product of the color symmetry groups $SU(3)_C$ that describes QCD and strong interactions, $SU(2)_L$ that describes the weak isospin interactions between the left-handed fermions, and $U(1)_Y$ that describes the weak hypercharge interactions, that differ between the left- and right-handed fermions. The weak and electromagnetic interactions are unified in the electroweak $SU(2)_L \otimes U(1)_Y$ group, which describes the Glashow-Weinberg-Salam electroweak theory.

According to Noether's theorem [13], to every differentiable symmetry generated by local actions there corresponds a conserved current and a conserved charge. The color charge, the weak isospin (I) and the weak hypercharge ($Y = 2(Q - I_3)$)[1] are the respective conserved charges associated with each symmetry group. The $SU(3)_C$ group is non-abelian[2], has coupling g_s and 8 Hermitian gauge fields, these are the 8 gluon fields represented as $G^i_\mu (i = 1,...,8)$. The $SU(2)_L$ group is non-abelian, has coupling g and 3 gauge fields $W^i_\mu (i = 1,2,3)$. The $U(1)_Y$ group is abelian, has coupling g' and one gauge field B_μ. The $SU(2)_L \otimes U(1)_Y$ symmetry is spontaneously broken due to the Higgs mechanism (explained in section 2.3), leading to the mixing of W^i_μ and the B_μ gauge fields to form the two charged $W^\pm$ bosons, the neutral Z boson of the weak interactions and the photon γ of the electromagnetic interactions. In this process the Z and $W^\pm$ acquire mass.

The standard model Lagrangian density can be written as the sum of the gauge, fermion, Higgs, and Yukawa terms:

$$\hat{\mathcal{L}}_{SM} = \hat{\mathcal{L}}_g + \hat{\mathcal{L}}_f + \hat{\mathcal{L}}_H + \hat{\mathcal{L}}_Y. \tag{2.1}$$

The first term, which is the gauge term, describes the self-interaction between the

[1] Q is the electric charge and I_3 is the projection of the weak isospin along the z-axis

[2] A group is called Abelian if all the generators of the group commute.

gauge fields. It is given by

$$\hat{\mathcal{L}}_g = -\frac{1}{4}\hat{G}^i_{\mu\nu}\hat{G}^{\mu\nu i} - \frac{1}{4}\hat{W}^i_{\mu\nu}\hat{W}^{\mu\nu i} - \frac{1}{4}\hat{B}_{\mu\nu}\hat{B}^{\mu\nu}, \tag{2.2}$$

where the field strengths for $SU(3)_C$, $SU(2)_L$, and $U(1)_Y$ are, respectively,

$$\hat{G}^i_{\mu\nu} = \partial_\mu \hat{G}^i_\nu - \partial_\nu \hat{G}^i_\mu - g_s f_{ijk}\hat{G}^j_\mu \hat{G}^k_\nu \quad i,j,k = 1,...,8 \tag{2.3}$$

$$\hat{W}^i_{\mu\nu} = \partial_\mu \hat{W}^i_\nu - \partial_\nu \hat{W}^i_\mu - g\epsilon_{ijk}\hat{W}^j_\mu \hat{W}^k_\nu \quad i,j,k = 1,2,3 \tag{2.4}$$

$$\hat{B}_{\mu\nu} = \partial_\mu \hat{B}_\nu - \partial_\nu \hat{B}_\mu, \tag{2.5}$$

with f_{ijk} and ϵ_{ijk} being the structure functions of $SU(3)_C$ and $SU(2)_L$. The fact that $SU(3)_C$ and $SU(2)_L$ are non-Abelian gives rise to the third terms of equations (2.3) and (2.4), resulting in self-interactions between the gauge fields. Since the $U(1)_Y$ group is abelian there is no self-interactions between photons, which is equivalent to saying that the photon carries no electric charge.

Table 2.1: Q, I_3 and Y quantum numbers of all Standard Model fermions.

Fermions	Q	I_3	Y
q_{mL}	$\begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	$\begin{pmatrix} 1/3 \\ -1/3 \end{pmatrix}$
u_{mR}, d_{mR}	$2/3, -1/3$	$0, 0$	$4/3, -2/3$
l_{mL}	$\begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}$	$\begin{pmatrix} -1 \\ -1 \end{pmatrix}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$
e_{mR}	0	-2	1

The second part of the Lagrangian of equation (2.1) describes the fermion fields and their interactions with the gauge fields. It is given by

$$\hat{\mathcal{L}}_f = \bar{q}_{mL} i\gamma^\mu D_\mu q_{mL} + \bar{l}_{mL} i\gamma^\mu D_\mu l_{mL} + \bar{u}_{mR} i\gamma^\mu D_\mu u_{mR} \tag{2.6}$$

$$+ \bar{d}_{mR} i\gamma^\mu D_\mu d_{mR} + \bar{e}_{mR} i\gamma^\mu D_\mu e_{mR} + \bar{\nu}_{mR} i\gamma^\mu D_\mu \nu_{mR}, \tag{2.7}$$

where $m = 1,2,3$ denotes the family, the subscripts L and R represent the left and right chirality components of the respective field and $\hat{D}_\mu$ is the gauge covariant derivative. Table 2.1 summarizes the quantum numbers Q, I_3 and Y for all the fermion fields. The left–handed fermions are weak isodoublets, while the

right–handed fermions are weak isosinglets:

$$q_{mL} = \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L \qquad l_{mL} = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \tag{2.8}$$

$$u_{mR} = u_R,\ c_R,\ t_R \qquad \nu_{mR} = \nu_{eR},\ \nu_{\mu R},\ \nu_{\tau R} \tag{2.9}$$

$$d_{mR} = d_R,\ s_R,\ b_R \qquad e_{mR} = e_R^-,\ \mu_R^-,\ \tau_R^-. \tag{2.10}$$

The covariant derivatives acting on the various fermion fields are given by

$$\hat{D}_\mu \hat{q}_{mL} = \left(\partial_\mu + \frac{ig_s}{2}\lambda_\alpha \hat{G}_\mu^\alpha + \frac{ig}{2}\tau_\beta \hat{W}_\mu^\beta + \frac{ig'}{6}\hat{B}_\mu\right)\hat{q}_{mL} \tag{2.11}$$

$$\hat{D}_\mu \hat{l}_{mL} = \left(\partial_\mu + \frac{ig}{2}\tau_\beta \hat{W}_\mu^\beta - \frac{ig'}{2}\hat{B}_\mu\right)\hat{l}_{mL} \tag{2.12}$$

$$\hat{D}_\mu \hat{u}_{mR} = \left(\partial_\mu + \frac{ig_s}{2}\lambda_\alpha \hat{G}_\mu^\alpha + \frac{2ig'}{3}\hat{B}_\mu\right)\hat{u}_{mR} \tag{2.13}$$

$$\hat{D}_\mu \hat{d}_{mR} = \left(\partial_\mu + \frac{ig_s}{2}\lambda_\alpha \hat{G}_\mu^\alpha - \frac{ig'}{3}\hat{B}_\mu\right)\hat{d}_{mR} \tag{2.14}$$

$$\hat{D}_\mu \hat{\nu}_{mR} = \partial_\mu \hat{\nu}_{mR} \tag{2.15}$$

$$\hat{D}_\mu \hat{e}_{mR} = \left(\partial_\mu - ig'\hat{B}_\mu\right)\hat{e}_{mR}, \tag{2.16}$$

where $\lambda^\alpha(\alpha = 1,...,8)$ are the Gell-Mann matrices and $\tau^\beta(\beta = 1,2,3)$ are the Pauli matrices. Since the right-handed fermions transform as isosinglets under $SU(2)_L$ they will not interact with the weak fields ($\tau_\beta \hat{W}_\mu^\beta$ are 2×2 matrices in $SU(2)_L$ space) and the $W^\pm$ gauge bosons will couple only to the left-handed fermions, as seen in equations (2.11 - 2.16). In addition, the quarks transform as color triplets[3] under $SU(3)_C$ and the leptons as color singlets, thus only the quarks will interact via strong interaction ($\lambda_\alpha \hat{G}_\mu^\alpha$ are 3×3 matrices in color space).

Another consequence of the left-handed fermions being $SU(2)_L$ doublets and right-handed fermions being singlets is that an explicit mass term for the fermions ($-m_f\bar{\psi}\psi$) is forbidden in the Lagrangian. This term mixes the left- and right-handed components and is not invariant under isospin transformations. Similarly, the mass terms for the gauge bosons (e.g. $-m_W^2 W_\mu^i W^{\mu i}$) are not invariant under gauge transformations. The mechanism that allows the construction of a

[3]The color indices have been suppressed for simplicity.

meaningful gauge theory, which can include mass terms for the gauge bosons and fermions, will be presented in the following sections together with the Higgs ($\hat{\mathcal{L}}_H$) and Yukawa ($\hat{\mathcal{L}}_Y$) terms of the Lagrangian.

2.3 Electroweak Symmetry Breaking

Up to this point, there are no mass terms in the Lagrangian. All fermions and gauge fields are massless. The gluons and photons are indeed massless. However, the electroweak vector bosons have been proved to be massive. Including a mass term for the weak bosons or for the fermions in the Lagrangian would violate the local $SU(2)_L \otimes U(1)_Y$ gauge invariance, which is needed to ensure that the theory is renormalizable. These masses are included by the Higgs mechanism without violating gauge invariance by introducing an $SU(2)_L$ Higgs doublet with weak isospin $I_3 = 1/2$ and hypercharge $Y = 1$:

$$\hat{\phi} = \begin{pmatrix} \hat{\phi}^+ \\ \hat{\phi}^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \hat{\phi}_1 + i\hat{\phi}_2 \\ \hat{\phi}_3 + i\hat{\phi}_4 \end{pmatrix} \tag{2.17}$$

The Lagrangian for the Higgs field is

$$\hat{\mathcal{L}}_H = \left(\hat{D}_\mu \hat{\phi}\right)^\dagger \left(\hat{D}^\mu \hat{\phi}\right) - \hat{V}(\hat{\phi}^\dagger \hat{\phi}), \tag{2.18}$$

where $\hat{V}(\hat{\phi}^\dagger \hat{\phi})$ is the Higgs potential and the gauge covariant derivative is given by

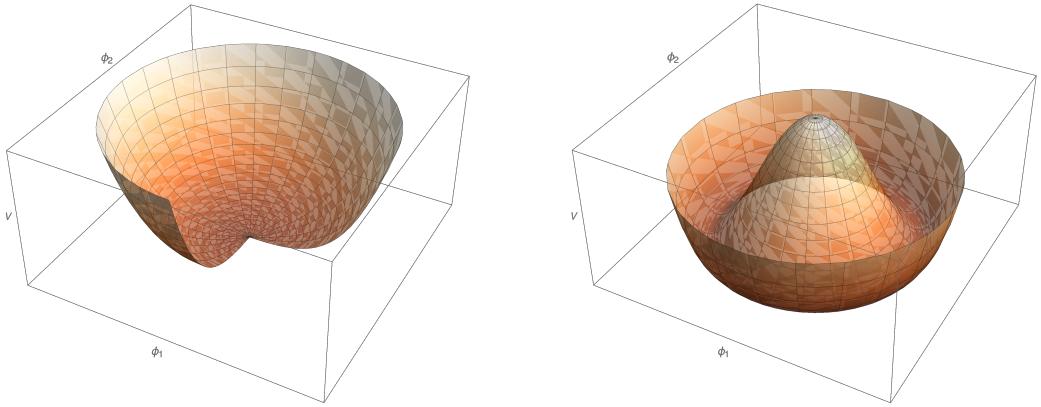
$$\hat{D}_\mu \hat{\phi} = \left(\partial_\mu + \frac{ig}{2} \vec{\tau} \cdot \hat{\vec{W}}_\mu + \frac{ig'}{2} \hat{B}_\mu\right) \hat{\phi}, \tag{2.19}$$

when acting on the $\hat{\phi}$ field.

The Lagrangian $\hat{\mathcal{L}}_H$ is invariant under $SU(2)_L \otimes U(1)_Y$ transformations. The potential is written in the most general and renormalizable form consistent with the $SU(2)_L \otimes U(1)_Y$ gauge invariance

$$\hat{V}(\hat{\phi}^\dagger \hat{\phi}) = \mu^2 \hat{\phi}^\dagger \hat{\phi} + \lambda \left(\hat{\phi}^\dagger \hat{\phi}\right)^2. \tag{2.20}$$

The first term of equation (2.20) is a mass term and the second is a self-interaction term. We can require λ to be positive to ensure that $\hat{V}(\hat{\phi}^\dagger \hat{\phi})$ is bounded from below. $\hat{V}(\hat{\phi}^\dagger \hat{\phi})$ can take two different forms according to the value of μ^2. If



(a) **(b)**

Figure 2.2: The shape of the Higgs potential for $\mu^2 > 0$ (a) and $\mu^2 < 0$ (b). In the first case, the minimum is unique and lies at the center of the potential, whereas in the second case there is a set of minima which lie on a circle. By choosing a particular point in the circle of minima and setting this to be the ground state, the symmetry is spontaneously broken.

μ^2 is positive ($\mu^2 > 0$), the potential $\hat{V}(\hat{\phi}^\dagger\hat{\phi})$ simply describes a scalar field with mass μ. In that case the ground state or the vacuum state (configuration in which the energy of the field is minimized) is unique and corresponds to $\langle 0|\,\hat{\phi}\,|0\rangle = 0$, as shown in Figure 2.2a. However, if μ^2 is negative ($\mu^2 < 0$), the minimum of the potential occurs for

$$\langle 0|\,\hat{\phi}^\dagger\hat{\phi}\,|0\rangle = \frac{1}{2}\langle 0|\left(\hat{\phi}_1^2 + \hat{\phi}_2^2 + \hat{\phi}_3^2 + \hat{\phi}_4^2\right)|0\rangle = -\frac{\mu^2}{2\lambda}. \tag{2.21}$$

As shown in Figure 2.2b, the ground state is degenerate, there is an infinite set of minimum values for the potential. These values are found to lie on a circle and when one of these values is chosen as the ground state, the symmetry is broken. This vacuum does not hold the same symmetry of the initial Lagrangian. Whenever this occurs we say that the symmetry was spontaneously broken. For example, if we choose $\langle 0|\,\hat{\phi}_i\,|0\rangle = 0$ for $i = 1, 2, 4$ and $\langle 0|\,\hat{\phi}_3\,|0\rangle = \sqrt{-\mu^2/\lambda}$, the vacuum expectation value will be

$$\langle 0|\,\hat{\phi}\,|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix} \quad \text{with} \quad \langle 0|\,\hat{\phi}_3\,|0\rangle \equiv v = \sqrt{\frac{-\mu^2}{\lambda}}. \tag{2.22}$$

To determine the excitations (i.e. the particle spectrum), it is necessary to

expand the fields around their value at the minimum. Oscillations about equation (2.22) can be parametrized by

$$\hat{\phi}(x) = \frac{1}{\sqrt{2}} \exp\left(-i\vec{\tau}\cdot\hat{\vec{\theta}}(x)/v\right)\begin{pmatrix}0\\ v+\hat{H}(x)\end{pmatrix}, \tag{2.23}$$

where $\tau_i (i = 1,2,3)$ are the Pauli matrices.

The three $\hat{\theta}(x)$ can be removed by an appropriate gauge transformation $\hat{\phi} \rightarrow \hat{\phi}' = \exp\left(i\vec{\tau}\cdot\hat{\vec{\theta}}(x)/v\right)\hat{\phi}$. The field $\hat{\phi}$ will be rotated into the form

$$\hat{\phi}'(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v+\hat{H}(x)\end{pmatrix}. \tag{2.24}$$

Thus, of the four scalar fields, the only one that does not vanish after the rotation is the $\hat{H}(x)$ field, the three $\hat{\theta}(x)$ fields were removed. The Hermitian scalar field $\hat{H}(x)$ is understood as the physical Higgs boson. The gauge choice of equation (2.24) is called unitary gauge.

The $\hat{\theta}(x)$ fields are Goldstone bosons. It can be shown, by the Goldstone's theorem [19], that a massless scalar spinless particle, known as Goldstone boson, appears when a continuous symmetry is spontaneously broken. For a local gauge symmetry the Goldstone bosons disappear from the spectrum and the massless gauge boson associated with the broken symmetry will acquire mass and a third polarization state (longitudinal polarization), this is the Higgs mechanism.

Each broken symmetry is associated with one massless Goldstone boson. The massless gauge fields that correspond to spontaneously broken non-Abelian symmetries will acquire mass, whereas those gauge fields that correspond to respected symmetries will not. If $SU(2)_L \otimes U(1)_Y$ symmetry were completely broken we would expect four Goldstone modes. However, the vacuum choice (eq. 2.22) carries no electric charge, $Q = (I_3 + {}^{Y}/_{2})$ is conserved and the vacuum will be invariant under $U(1)_Q$ transformations. So, since the $U(1)_Q$ symmetry of electromagnetism is conserved, the photon will remain massless and we expected only three of the four degrees of freedom from the original Higgs doublet to be Goldstone bosons. These three degrees of freedom will become the longitudinal polarization state of the three $W+$, $W-$ and Z bosons and make them massive (A massless particle has only two transverse polarization degrees of freedom). The fourth degree of freedom will become a scalar particle, the Higgs boson.

Introducing equation (2.24) into equation (2.18) and defining the $\hat{W}^{\pm}_{\mu}$, $\hat{Z}_{\mu}$ and $\hat{A}_{\mu}$ fields as

$$\hat{W}^{\pm}_{\mu} = \frac{1}{\sqrt{2}}\left(\hat{W}^{1}_{\mu} \mp i\hat{W}^{2}_{\mu}\right) \tag{2.25}$$

$$\hat{Z}_{\mu} = \cos\theta_W \hat{W}^{3}_{\mu} - \sin\theta_W \hat{B}_{\mu} \tag{2.26}$$

$$\hat{A}_{\mu} = \sin\theta_W \hat{W}^{3}_{\mu} + \cos\theta_W \hat{B}_{\mu}, \tag{2.27}$$

with θ_W the weak mixing angle, defined by

$$\sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}} \quad \cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}}, \tag{2.28}$$

the full Lagrangian for the Electroweak Symmetry Breaking becomes:

$$\hat{\mathcal{L}}_H = \frac{1}{2}\partial_\mu \hat{H}\partial^\mu \hat{H} + \left[m_W^2 \hat{W}^{+\mu}\hat{W}^{-}_{\mu} + \frac{m_Z^2}{2}\hat{Z}^\mu \hat{Z}_\mu\right]\left(1+\frac{\hat{H}}{v}\right)^2 - \hat{V}(\hat{H}), \tag{2.29}$$

with the potential given by

$$\hat{V}(\hat{H}) = \frac{1}{2}m_H^2\hat{H}^2 + \sqrt{\frac{\lambda}{2}}m_H\hat{H}^3 + \frac{\lambda}{4}\hat{H}^4 - \frac{m_H^4}{16\lambda} \quad \text{with} \quad m_H = \sqrt{2\mu^2}. \tag{2.30}$$

As expected, the second term of equation (2.29) now includes the mass terms for the $W^{\pm}$ and Z fields, while the photon remains massless. The (tree-level) masses are given by

$$m_W = \frac{gv}{2}; \quad m_Z = \frac{\sqrt{g^2+g'^2}v}{2}; \quad m_A = 0 \tag{2.31}$$

implying the relation

$$\frac{m_W}{m_Z} = \cos\theta_W \tag{2.32}$$

Equation (2.29) also contains the interaction terms between the vector bosons and the Higgs (i.e. ZZH, W^+W^-H, $ZZHH$ and W^+W^-HH). The first term of equation (2.30) is a (tree-level) mass term for the Higgs boson, which is given by $m_H = \sqrt{2\mu^2} = \sqrt{2\lambda}v$. The value of m_H cannot be calculated by the theory, since it depends on the coupling λ which is a free parameter. The potential also includes terms for the Higgs self-interactions between three and four Higgs bosons. The couplings are represented in Figure 2.3.

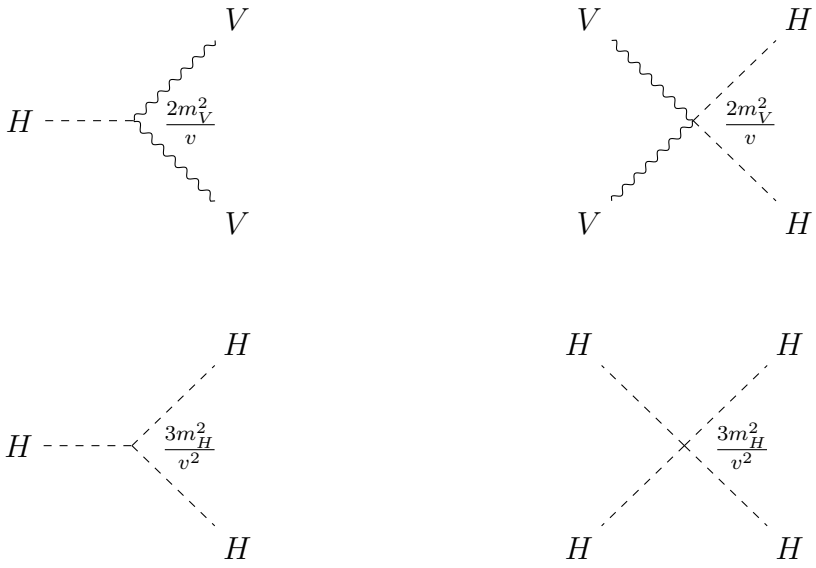


Figure 2.3: Interactions between the Higgs boson and weak bosons $V \equiv W^{\pm}, Z$ and Higgs self-interactions, the couplings are represented in each vertex.

2.4 Fermion Masses

It is possible to introduce in the Lagrangian an SU(2)-invariant interaction between the fermions and the Higgs doublet. These couplings are called Yukawa couplings and, after the spontaneous breaking of symmetry, these couplings will generate the mass terms for the fermions. The Yukawa Lagrangian can be expressed in the from

$$\hat{\mathcal{L}}_Y = -y_e \bar{l}_{mL} \hat{\phi} e_{mR} - y_d \bar{q}_{mL} \hat{\phi} d_{mR} - y_u \bar{q}_{mL} \hat{\phi}_C u_{mR} + h.c., \tag{2.33}$$

where $h.c.$ means Hermitian conjugate of the preceding terms, $\hat{\phi}_C = i\tau_2 \hat{\phi}^{\dagger}$ is the charge-conjugate of $\hat{\phi}$ and the couplings y_f are arbitrary.

Substituting equation (2.24) in the Yukawa Lagrangian gives

$$\hat{\mathcal{L}}_Y = \frac{-1}{\sqrt{2}} (v + H) \left(y_e \bar{e}_{mL} e_{mR} + y_d \bar{d}_{mL} d_{mR} + y_u \bar{u}_{mL} u_{mR} \right) + h.c. \tag{2.34}$$

$$= -m_e \bar{e}_m e_m - m_d \bar{d}_m d_m - m_u \bar{u}_m u_m - \frac{m_e}{v} \bar{e}_m e_m H - \frac{m_d}{v} \bar{d}_m d_m H - \frac{m_u}{v} \bar{u}_m u_m H, \tag{2.35}$$

where the second line was obtained using $\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L = \bar{\psi}\psi$ and defining $m_e = y_e v/\sqrt{2}$, $m_d = y_d v/\sqrt{2}$ and $m_u = y_u v/\sqrt{2}$. The fermions, except for neutrinos, have

acquired masses by coupling with the Higgs boson after spontaneous symmetry breaking. Yukawa terms cannot be written for the neutrinos since the SM contains no ν_R state, leaving neutrinos massless. However neutrinos masses can still be included in the SM with Majorana terms.

The last tree terms of equation (2.35) describe the interactions, represented in Figure 2.4, between the fermions and the Higgs boson. The coupling is proportional to the fermion masses, which have to be measured, and the Higgs couples most strongly to the heaviest fermion.

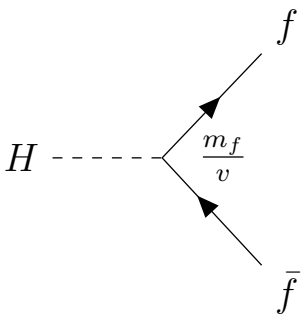


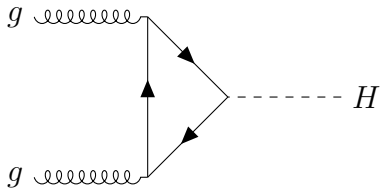
Figure 2.4: Higgs boson coupling to massive fermions.

2.5 Higgs Boson Production at the LHC

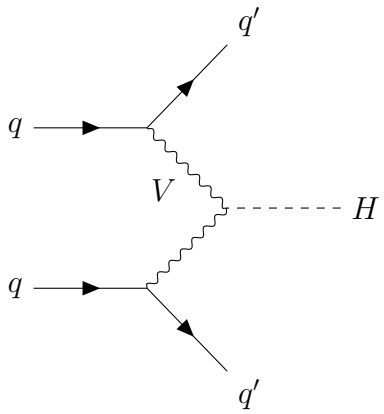
As seen in the previous sections, the Higgs couplings to the bosons and fermions are proportional to the corresponding particle mass, which means that the Higgs production and decay modes are dominated by processes involving couplings between the Higgs and heavy particles, such as the third generation of fermions an the weak vector bosons. The Higgs does not couple to massless gauge bosons (photons and gluons). Their interactions cannot occur at tree-level and need to be mediated by a loop of quarks (mainly top and bottom quarks, due to their large masses) or $W^{\pm}$ bosons (only in the photon case).

The dominant processes for producing Higgs bosons in pp collisions at the LHC, in order of decreasing cross section, are: gluon-gluon fusion (ggF), vector boson fusion (VBF), vector boson-associated production or "Higgsstrahlung" (VH), and Higgs associated production with $t\bar{t}$ and $b\bar{b}$ (qqH). Figure 2.5 shows examples of Feynman diagrams for those production modes. The theoretical cross section for each production mode is given in Table 2.2 and plotted as a function of the center-of-mass energy $\sqrt{s}$ for $m_H = 125$ GeV in Figure 2.6.

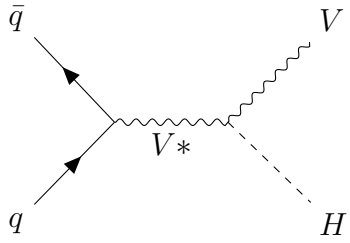
The ggF process is the dominant production mode for the Higgs boson at the LHC, contributing about 87% of the total Higgs cross section. This significant contribution is owing to the large presence of gluons in proton collisions.



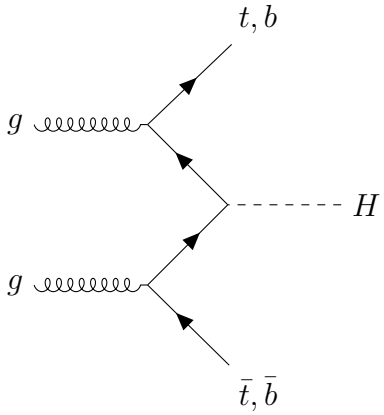
(a) Gluon fusion ($pp \to H$).



(b) Vector boson fusion ($pp \to qqH$).



(c) Vector boson associated production ($pp \to VH$).



(d) Heavy quarks associated production ($pp \to t\bar{t}$ ($b\bar{b}$)H).

Figure 2.5: Feynman diagrams of the dominant production modes for the Higgs boson at the LHC.

The second most probable production mode is VBF, in which the Higgs is produced from $W^{\pm}$ or Z bosons that originate from two initial state quarks. These quarks will scatter through to the final state and form two jets. This channel contributes about 7% of the total Higgs cross section.

The VH process occurs when one quark and one antiquark interact to from an off-shell vector boson that radiates a Higgs boson and one on-shell vector boson.

This final state vector boson will further decay into leptons or hadrons. The contribution of this process is smaller due to the fact that the initial state antiquark does not come from the valence of the colliding protons but from the sea.

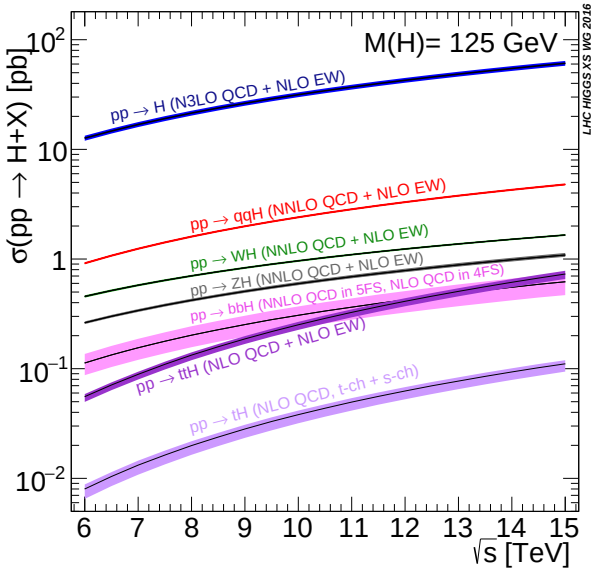


Figure 2.6: Higgs boson production cross sections as a function of the center-of-mass-energies. The theoretical uncertainties are indicated as bands [20].

Table 2.2: SM Higgs boson production cross sections for $m_H = 125.09$ GeV in $p\bar{p}$ collisions at $\sqrt{s} = 13$ and 14 TeV [20]. The fraction f of the total cross section is shown together with the expected number of events N, considering an integrated luminosity of 147 and 350 fb^{-1} for the Run 2 and 3, respectively.

	$\sqrt{s} = 13$ TeV			$\sqrt{s} = 14$ TeV		
	σ (pb)	f (%)	N	σ (pb)	f (%)	N
ggF	$48.51^{+7.8\%}_{-9.9\%}$	87.137	$7.13 \cdot 10^6$	$54.60^{+7.8\%}_{-9.9\%}$	87.091	$1.91 \cdot 10^7$
VBF	$3.922^{+2.3\%}_{-2.0\%}$	7.045	$5.77 \cdot 10^5$	$4.442^{+2.3\%}_{-2.0\%}$	7.085	$1.55 \cdot 10^6$
WH	$1.370^{+2.4\%}_{-2.6\%}$	2.461	$2.01 \cdot 10^5$	$1.510^{+2.3\%}_{-2.6\%}$	2.409	$5.28 \cdot 10^5$
ZH	$0.882^{+5.4\%}_{-4.6\%}$	1.584	$1.30 \cdot 10^5$	$0.983^{+5.4\%}_{-4.8\%}$	1.568	$3.44 \cdot 10^5$
ttH	$0.498^{+9.4\%}_{-12.8\%}$	0.895	$7.32 \cdot 10^4$	$0.603^{+9.6\%}_{-12.8\%}$	0.962	$2.11 \cdot 10^5$
bbH	$0.486^{+20.1\%}_{-23.9\%}$	0.873	$7.14 \cdot 10^4$	$0.552^{+20.1\%}_{-24.1\%}$	0.88	$1.93 \cdot 10^5$
tH	$0.0028^{+4.6\%}_{-4.0\%}$	0.005	$4.12 \cdot 10^2$	$0.0032^{+4.5\%}_{-3.9\%}$	0.005	$1.12 \cdot 10^3$

The qqH production is the mechanism with the smallest contribution, where one Higgs an two final state heavy quarks are produced from a pair of gluons or a pair of quark-antiquark. The $t\bar{t}H$ channel is a extremely important process that gives direct access to the Higgs Yukawa coupling to top quarks and can be used to probe the CP nature of the Higgs boson. The importance of the $t\bar{t}H$ channel will be further discussed in section 4.1.

2.6 Higgs Boson Decay Modes

The Higgs will predominantly decay into the heaviest pairs of weak vector bosons or fermions ($H \to V\bar{V}, f\bar{f}$) if the decay mode is kinematically allowed ($m_H > 2m_{V,f}$). As a result, since $m_H < 2m_t$, an on-shell Higgs cannot decay into pairs of top quarks despite its large Yukawa coupling. Therefore, the decay into pairs of bottom quarks is the one with the highest branching ratio. In the case of a Higgs decaying into a pair of $W^{\pm}$ or Z bosons, one of the bosons needs to be off-shell for the decay to be kinematically allowed. The Higgs can also decay into pairs of massless gauge bosons through loops of massive particles. The dominant decay modes are $H \to b\bar{b}$ and $H \to WW^*$, followed by $H \to gg$, $H \to \tau^+\tau^-$, $H \to c\bar{c}$, $H \to ZZ^*$, $H \to \gamma\gamma$, $H \to Z\gamma$ and $H \to \mu^+\mu^-$. The branching ratios of the different Higgs decay channels are plotted as a function of the Higgs mass in Figure 2.7. Table 2.3 shows the decay branching ratios for a Higgs boson with a mass of 125.09 GeV.

The sensitivity of a search channel does not depend only on the corresponding branching ratios. The production cross section, the reconstructed mass resolution, the selection efficiency and the level of background in the final state also need to be considered. For a Higgs boson with mass $m_H = 125$ GeV, the decay channels $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ provide the best resolution of the reconstructed mass as the detected photons and leptons can be very precisely measured. Although the $H \to b\bar{b}$ and $H \to WW^*$ channels have large branching ratio, the reconstructed mass resolution is lower. The $H \to b\bar{b}$ channel is not easily distinguished from its large QCD background, the b-jets originated from this channel are identified with considerably worse resolution than the one achieved with leptons and photons. The $H \to WW^*$ channel contains either jets or neutrinos in the final state. The latter cannot be directly detected and must be reconstructed from the missing energy. Hence, in spite of the small branching ratio, channels including photons and leptons may still be preferred.

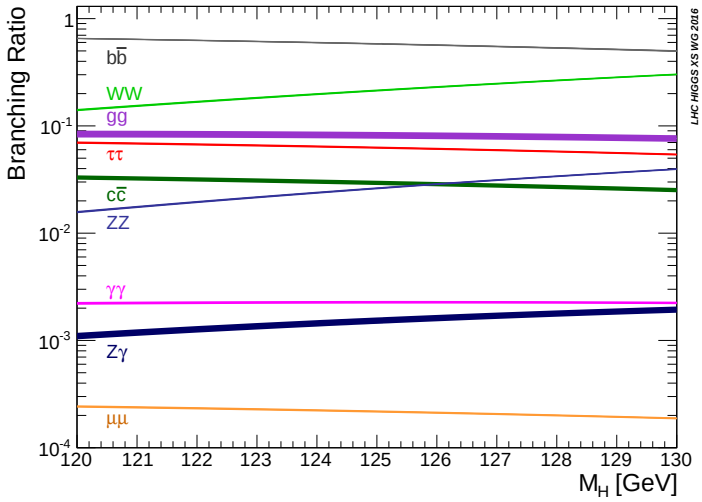


Figure 2.7: Higgs boson branching ratios and their uncertainties for the mass range around 125 GeV [20].

Table 2.3: Higgs boson branching ratios and Higgs total width together with their total relative uncertainties for a SM Higgs boson with $m_H = 125.09$ GeV [20].

Decay channel	Branching ratio (%)	
$H \to b\bar{b}$	58.09	$^{+2.14\%}_{-2.18\%}$
$H \to WW^*$	21.52	$^{+2.64\%}_{-2.6\%}$
$H \to gg$	8.18	$^{+8.21\%}_{-8.15\%}$
$H \to \tau^+\tau^-$	6.256	$^{+2.77\%}_{-2.74\%}$
$H \to c\bar{c}$	2.884	$^{+7.75\%}_{-3.41\%}$
$H \to ZZ^*$	2.641	$^{+2.62\%}_{-2.59\%}$
$H \to \gamma\gamma$	0.227	$^{+3.27\%}_{-3.33\%}$
$H \to Z\gamma$	0.1541	$^{+7.26\%}_{-7.37\%}$
$H \to \mu^+\mu^-$	0.02171	$^{+2.8\%}_{-2.86\%}$
$\Gamma_H = 4.100 \cdot 10^{-3}$	$^{+2.32\%}_{-2.31\%}$ GeV	

2.7 Higgs Boson Discovery and Properties

The discovery of a particle with a mass around 125 GeV and consistent with the Higgs boson was announced by both the ATLAS [1] and CMS [2] collaborations on the fourth of July 2012. The Higgs boson was a remarkably difficult particle to be observed and the search had been ongoing for almost half a century. This is due

to the fact that its mass depends on the coefficient λ of the Higgs self-coupling, as seen in section 2.3. The numerical value of λ is not given by the theory and it cannot be obtained by any other observable. Experiments could not look for the Higgs boson at a specific mass and had to be designed for any value of m_H.

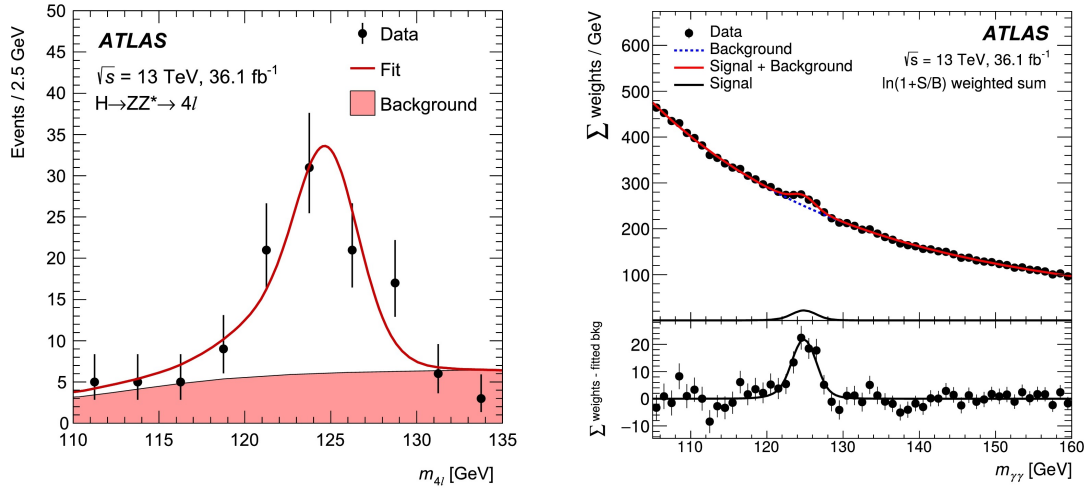


Figure 2.8: Invariant mass distribution of the four leptons (a) and diphoton system (b) overlaid with the results of the fits and the background in each case [21].

After the discovery, the primary goal of the Higgs studies has been to obtain precise measurements of the different Higgs parameters and properties, such as measurements of its mass, decay width, coupling constants, spin and parity quantum numbers in order to verify if the discovered Higgs boson is consistent with the SM predictions. A combined measurement [22] of the Higgs boson mass was performed by ATLAS and CMS with information from both the $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ channels. The measurement is derived from a combined fit to the four-lepton and diphoton invariant mass spectra. Figure 2.8 shows examples of these distributions. The various results obtained by ATLAS and CMS can be combined to produce a more precise value. The latest combined measured mass of the Higgs boson performed using Run 1 data at $\sqrt{s} = 7$ and 8 TeV was found to be

$$m_H = 125.09 \pm 0.21 \text{ (stat)} \pm 0.11 \text{ (syst) GeV}. \tag{2.36}$$

Once the mass is know, the production cross sections and decay branching ratios of the Higgs boson predicted by the SM can be precisely calculated. The signal strength parameter μ is commonly used to characterize the Higgs boson yields and

compare the predicted values for production and decay to the experimental results. The signal strength can be defined as

$$\mu = \mu_i \times \mu^f = \frac{\sigma_i \times BR^f}{(\sigma_i \times BR^f)_{SM}}, \tag{2.37}$$

where σ_i is the cross section of the production mode i and BR^f is the branching ratio of the decay channel f. The values of σ_i and BR^f cannot be measured separately since each detected process involves a production and a decay, and only the product of μ_i and μ^f can be extracted experimentally. The most recent measurement of the global Higgs signal strength obtained by the ATLAS collaboration is [23]

$$\mu = 1.13 \pm 0.05\ (\text{stat}) \pm 0.05\ (\text{exp})^{+0.05}_{-0.04}\ (\text{sig th}) \pm 0.03\ (\text{bkg th}) \tag{2.38}$$

where the total uncertainty is decomposed into components for statistical uncertainties, experimental systematic uncertainties, and theory uncertainties on signal and background modelling. The measurement is in agreement with the SM predictions.

In order to obtain values of the Higgs couplings and search for possible deviations from the SM predictions, the expected SM Higgs cross sections and partial decay widths are scaled by coupling modifiers k_j as [24]

$$(\sigma \cdot BR)\,(i \to H \to f) = \sigma_{SM}\,(i \to H) \cdot BR_{SM}\,(H \to f) \cdot \frac{k_i^2 k_f^2}{k_H^2} \tag{2.39}$$

where $k_j = 1$ denoting SM predictions, with $j = i,f$ for the initial and final states of the process, respectively. k_H is one scale factor for the Higgs boson total width. With this parameterization, known as the kappa framework, the expected coupling strengths can be tested through fits to the experimental data, combining data from several different channels.

As seen in section 2.3 and 2.4, the SM predicts a linear relation between fermion couplings to the Higgs boson and fermion masses, and a linear relation between vector boson couplings to the Higgs boson and the square of vector boson masses. The reduced couplings used in the parameterization are defined as $y_f = k_f m_f/v$ for fermions and $y_V = \sqrt{k_V} m_V/v$ for vector bosons. Figure 2.9 shows a summary of the best-fit results for the Higgs boson couplings obtained by ATLAS and CMS, displaying a trend consistent with the SM predictions.

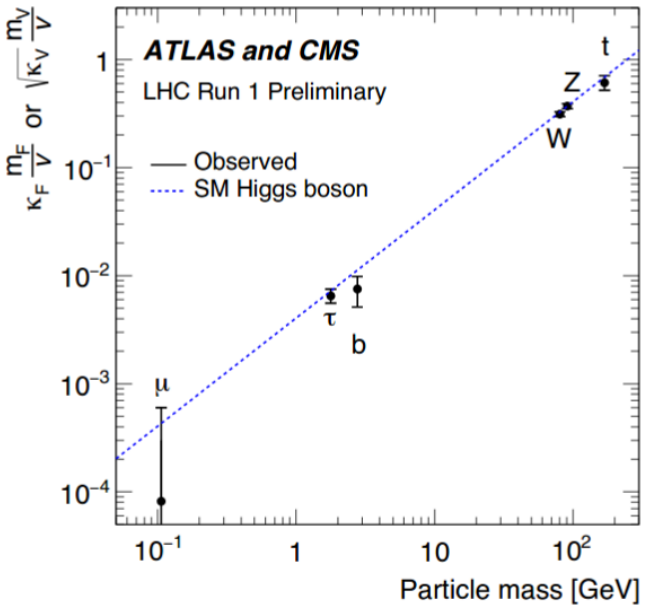


Figure 2.9: Higgs boson reduced coupling strength modifiers to the different fermions and vector bosons measured by ATLAS and CMS using Run 1 data, as a function of the particle mass [25].

It is also essential to probe the Higgs boson spin (J), charge conjugation (C) and parity (P) quantum numbers. In the SM, the Higgs is a scalar and CP-even particle ($J^{PC} = 0^{++}$), which means that it is invariant under CP-transformations. The ATLAS and CMS collaborations have searched for CP-odd contributions in the Higgs boson couplings to gauge bosons, and so far all results are in agreement with the CP-even SM prediction [26–29]. However, these channels probe only the bosonic couplings. The measurement of the Higgs boson couplings to fermions may provide information that the bosonic channels do not, since these couplings are sensible to CP-odd contributions at tree-level, whereas the bosonic ones are not. Exploring the Yukawa couplings structure is essential to fully understand the observed Higgs boson. Such structure could indicate a CP violation in the Higgs sector, which could potentially help in the explanation of baryogenesis. The importance of this measurement is further addressed in Chapter 4.

3

The LHC and the ATLAS Detector

In order to test the theory and its predictions, sophisticated particle accelerators are required to provide high energy collisions at high rates, as well as multilayer detectors that can efficiently identify and reconstruct the different types of particles originated from the collisions. In this chapter, a brief introduction of the Large Hadron Collider (LHC) is given, together with its foreseen High-Luminosity upgrade. In the second part of this chapter, the ATLAS detector and all of its components is described, as well as the trigger and data acquisition systems.

3.1 The Large Hadron Collider

The Large Hadron Collider [30], located at CERN (European Organization for Nuclear Research), is the world's most powerful superconducting hadron accelerator and collider. It lies between 45 and 170 m under the France-Switzerland border near Geneva, in a 26.7 km tunnel that first hosted the Large Electron-Positron Collider (LEP). The LHC was built to explore new energy frontiers and explain some of the most fundamental questions of physics.

The LHC was designed to collide proton beams with a center-of-mass energy of 14 TeV and luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. It was also designed to collide heavy ions, in particular, lead (Pb) nuclei, with an energy of 2.8 TeV per nucleon and a peak luminosity of 10^{27} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. To achieve such high energies, the particles need to pass through a series of systems that successively increase their energy. The protons obtained by the ionization of hydrogen atoms are injected into the linear particle accelerator LINAC 2, where they are accelerated to an energy of 50 MeV. Then, the protons are accelerated by the Proton Synchrotron Booster (PSB) up to 1.4 GeV, by the Proton Synchrotron (PS) up to 25 GeV, and by the Super Proton Synchrotron (SPS) up to 450 GeV. Finally, the proton beam is split

and injected into the LHC. The beams are introduced into different LHC beam pipes in opposite directions, where the beams are further accelerated and reach their final energy.

Inside the LHC, the particles circulate in bunches of approximately 10^{11} protons, with a bunch spacing of 25 ns. Then the bunches collide at four crossing points, around which the detectors are positioned. Each detector is dedicated to certain kinds of study: ATLAS (A Toroidal LHC ApparatuS) [31] and CMS (Compact Muon Solenoid) [32] are general-purpose detectors, ALICE (A Large IonCollider Experiment) [33] is designed to study the quark-gluon plasma in Pb-Pb and p-Pb collisions, and LHCb (LHC-beauty) [34] is dedicated to the study of bottom quark physics.

One important concept of particle accelerators is the notion of Luminosity, which measures the capacity of a particle collider to produce the demanded number of interactions. It is used to quantify the performance of the collider. For a given process i, the number of events N_i can be obtained by the product of the production cross section σ_i of this process and the time integral over the instantaneous luminosity $\mathcal{L}(t)$:

$$N_i = \sigma_i \int \mathcal{L}(t)\ dt. \tag{3.1}$$

For bunched beams the instantaneous luminosity can be expressed by

$$\mathcal{L} = \frac{N_p^2 n_b f_{rev}}{4\pi\sigma_x\sigma_y} F, \tag{3.2}$$

where N_p is the number of protons contained in each bunch, n_b is the number of bunches injected at the LHC per revolution, f_{rev} is the beam revolution frequency ($\approx$ 11 kHz), σ_x and σ_y characterize the transverse dimensions of the beam and F is the geometric luminosity reduction factor, which is a small correction factor to account for collisions with a transverse offset or crossing angle between beams at the interaction point.

Table 3.1 summarizes the main operation parameters of the LHC for each data taking period until the end of Run 2. Figure 3.1 shows the total integrated Luminosity delivered by the LHC and recorded by ATLAS during the Run 2. The detectors start working only after the beams circulating in the beam pipe are stable so the total luminosity recorded by the detectors will be different from the total delivered by the LHC.

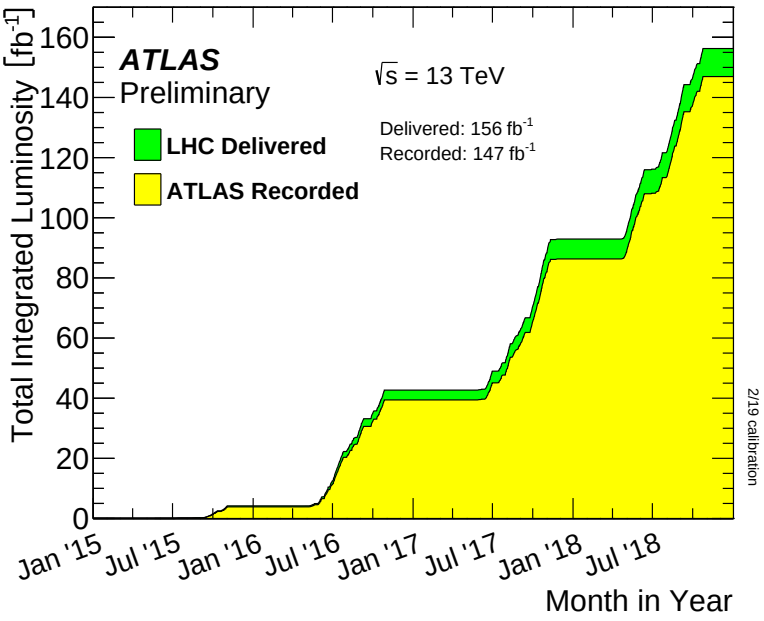


Figure 3.1: Cumulative luminosity delivered to ATLAS (green) and recorded by ATLAS (yellow) during stable beams for *pp* collisions at 13 TeV center-of-mass energy in LHC Run 2 [35].

Table 3.1: Typical LHC running parameters for *pp* collisions during operation in Run 1 (2010 - 2012) and Run 2 (2015 - 2018), shown together with the design parameters [35–39]. In 2017, the LHC was run in two configurations: standard, with bunch spacing of 25 ns, and '8b4e', denoting a pattern of eight bunches separated by 25 ns followed by a four bunch-slot gap. Values are presented for both configurations.

	Run 1			Run 2				
Parameter	2010	2011	2012	2015	2016	2017	2018	Design
Beam energy (TeV)	3.5	3.5	4.0	6.5	6.5	6.5	6.5	7.0
Max number of bunch pairs	368	1380	1380	2232	2208	2544/1909	2544	2808
Bunch population (10^{11} protons)	1.0	1.3	1.5	1.1	1.1	1.1/1.2	1.1	1.15
Bunch spacing (ns)	150	50	50	25	25	25/8b4e	25	25
Max peak luminosity (10^{34} cm^{-2}s^{-1})	0.021	0.35	0.77	0.5	1.3	1.6	1.9	1.0
Total integrated luminosity ($1fb^{-1}$)	0.048	5.5	22.8	4.0	38.5	50.2	63.4	-
Number of Interactions per Crossing		9.1	20.7	13.4	25.1	37.8	36.1	24

By increasing the luminosity, the number of collisions per bunch crossing will increase and, with it, the probability of observing a rare process. However, due to the large bunch density, multiple interactions can occur simultaneously. The overlap of the electronic signals from these additional interactions with the primary event is known as pile-up, it increases with the luminosity and presents a significant challenge to physics analyses. Figure 3.2 shows the distributions of pile-up for the data collected by the ATLAS experiment during the LHC Run 2.

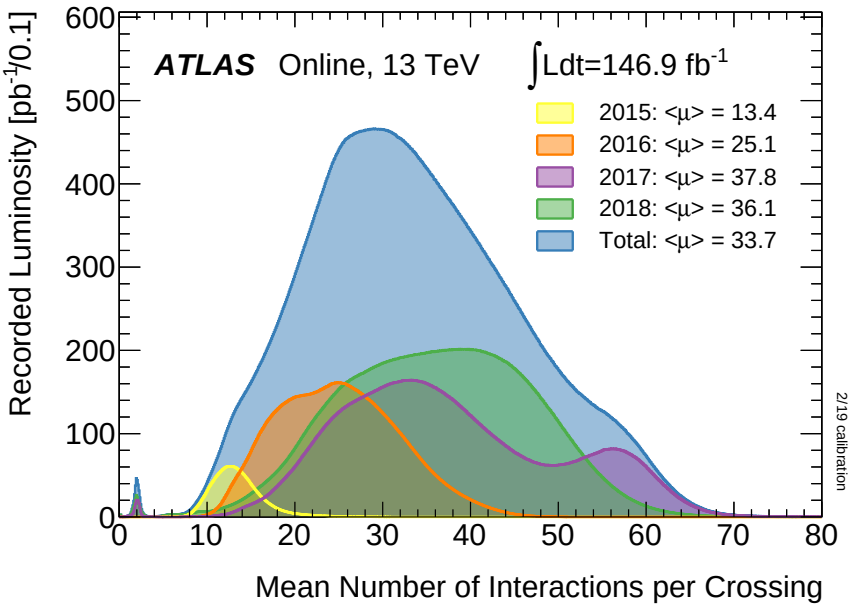


Figure 3.2: Luminosity-weighted distribution of the mean number of interactions per crossing for the 13 TeV data from 2015 - 2018 collected by the ATLAS experiment. The time averaged pile-up, $\langle\mu\rangle$, per year is shown [35].

3.1.1 The High-Luminosity Large Hadron Collider

The High-Luminosity Large Hadron Collider (HL-LHC) [41–43] project is an upgrade to the LHC that proposes a boost in the performance of the machine to increase its potential for physics discoveries. The project is planned to start in 2027 and aims to increase the instantaneous luminosity by at least a factor of five beyond its initial design value. This luminosity will allow for the HL-LHC to collect a dataset about one order of magnitude larger than what would be expected by the end of 2023 with the LHC baseline program. The timeline for LHC program including the HL-LHC upgrade is represented in Figure 3.3.

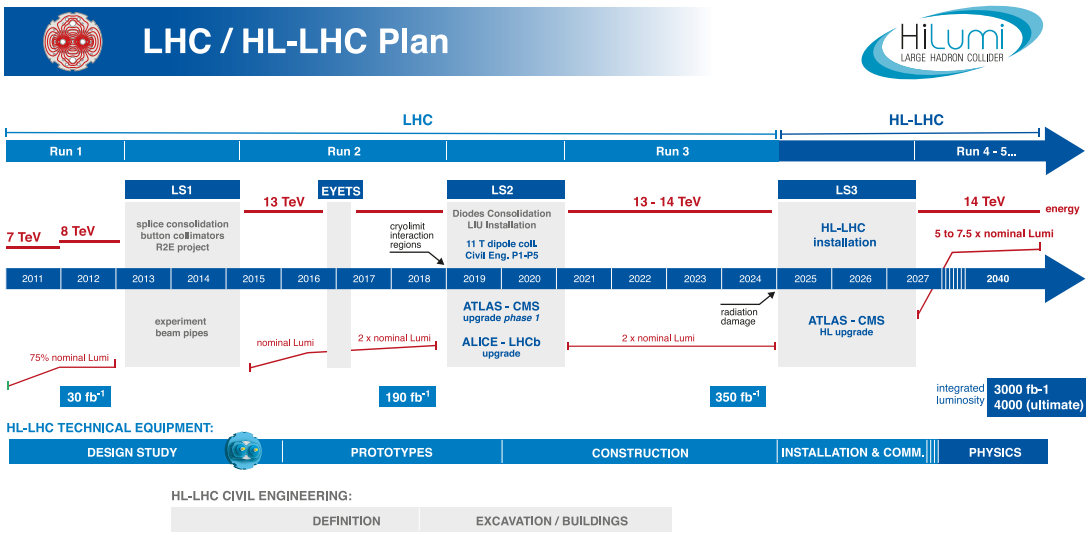


Figure 3.3: LHC plan for the next decade and beyond showing the energy of the collisions and the luminosity. After the third long shutdown (LS3) the machine will be in the High Luminosity configuration [41].

At the start of the operation of the HL-LHC, the ATLAS experiment will have collected a dataset corresponding to an integrated luminosity of roughly 300 fb^{-1}. The HL-LHC program is expected to provide ten times more integrated luminosity, 3000 fb^{-1}, during an operating period of approximately 10 years. An instantaneous luminosity of 7.5×10^{34} cm^{-2}s^{-1} and a performance of more than 300 fb^{-1} per year is required in order to accomplish this goal. The collisions will be held at a center-of-mass energy of $\sqrt{s} = 14$ TeV, in a pile-up environment of $140 - 200$ simultaneous events and bunch spacing of 25 ns.

An increase in luminosity means that there will be more collisions per bunch crossing and therefore more data will be collected. This will lead to a considerable improvement in the experimental precision of the SM measurements. Besides that, it will provide access to rare processes not previously seen and extend the precision of important measurements that may hint for the existence of physics beyond the SM. For example, the measurements of the coupling of the Higgs boson to SM particles is expected to reach a level where the effects of new physics could possibly be seen. Furthermore, the HL-LHC will be able to probe the Higgs pair production, which gives access to the Higgs trilinear self-coupling, providing constraints on the Higgs potential and improving our understanding of the EWSB mechanism.

With the increase in the integrated luminosity and presumed potential gains in technique, several studies were performed [44–46] to estimate the future performance of the HL-LHC, as well as to predict the expected systematic uncertainties of physics results. General recommendations were defined for the HL-LHC uncertainties. Theoretical uncertainties are considered to be reduced by a factor of two of the current values, due to higher-order calculation and reduced PDF uncertainties [47]. Intrinsic statistical uncertainty in the measurement is divided by a factor of $\sqrt{L/L_{ref}}$, where L is the projection integrated luminosity and L_{ref} the luminosity of the reference Run-2 analysis. The same factor is also usually applied for experimental systematic uncertainties but, since these systematic uncertainties will never be negligible, a limit for this factor is set to 50% of the currently value.

3.2 The ATLAS Detector

The ATLAS detector is one of two general-purpose detectors at the LHC. It is located at interaction Point 1 in a cavern 100 m below ground. Being a cylindrical shaped detector with dimensions of 44 m in length and 25 m in height, it is the most voluminous detector at the LHC. It consists of distinct subsystems wrapped concentrically in layers around the interaction point, in the center of the detector, to measure the trajectory, momentum, and energy of particles, allowing the identification of each particle. The charged particles trajectories are bent by a large magnet system, in order to obtain an accurate measurement of the momentum of these particles.

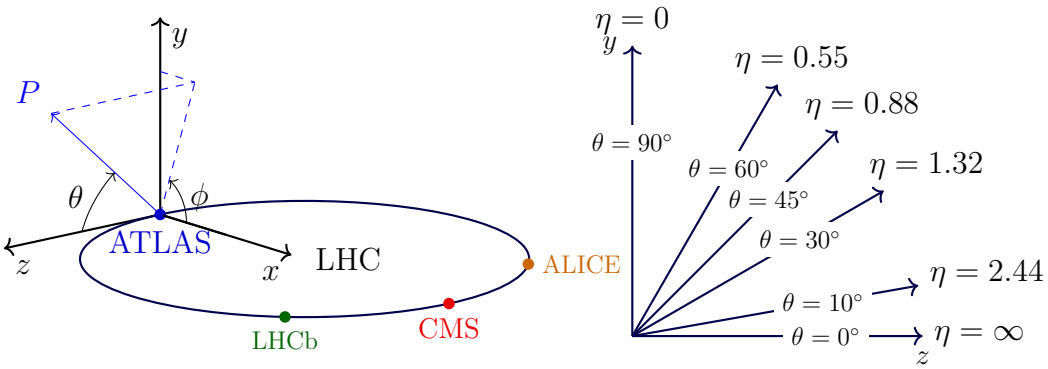


Figure 3.4: Schematic representation of the coordinate system used by the ATLAS collaboration (left) and values of pseudorapidity (η) shown on a polar plot. (right). Images taken and adapted from [48].

ATLAS uses a right-handed coordinate system, with its origin being set at the nominal interaction point in the center of the detector. The z-axis is determined by the direction of the beam line, while the x-y plane is set transverse to it. The x-axis pointing towards the center of the LHC ring and the y-axis pointing upwards. Cylindrical coordinates can be used in the transverse plane, with ϕ ($-\pi < \phi < \pi$) being the azimuthal angle around the z-axis and r the radius pointing outside the detector. The pseudorapidity η is preferred over the polar angle θ ($0 < \theta < \pi$) to describe the angle of a particle relative to the beam direction since rapidity differences ($\Delta\eta$) are Lorentz invariant under boosts along the z-axis. The pseudorapidity is defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{3.3}$$

The angular distance between particles in the pseudorapidity-azimuthal plane can be defined as

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{3.4}$$

Figure 3.4 shows a representation of the ATLAS right-handed coordinate system as well as a representation of η for different values of the polar angle θ.

Table 3.2: Performance goals of each component of the ATLAS detector [31]. The units for p_T and E are in GeV. The $\oplus$ symbol represents the sum in quadrature.

Detector component	Required resolution	η coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	± 2.5	
EM calorimetry	$\sigma_E/E = {}^{10\%}/_{\sqrt{E}} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic calorimetry (jets)			
barrel and end-cap	$\sigma_E/E = {}^{50\%}/_{\sqrt{E}} \oplus 3\%$	± 3.2	± 3.2
forward	$\sigma_E/E = {}^{100\%}/_{\sqrt{E}} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV	± 2.7	$3.1 < \lvert\eta\rvert < 4.9$

The major components of the ATLAS detector are the Inner Detector (ID), the electromagnetic and hadronic calorimeters (ECAL and HCAL), the Muon Spectrometer (MS) and the Magnet System, which are illustrated in Figure 3.5 and will be detailed in the following sections. Table 3.2 summarizes the general performance goals of each component of the ATLAS detector.

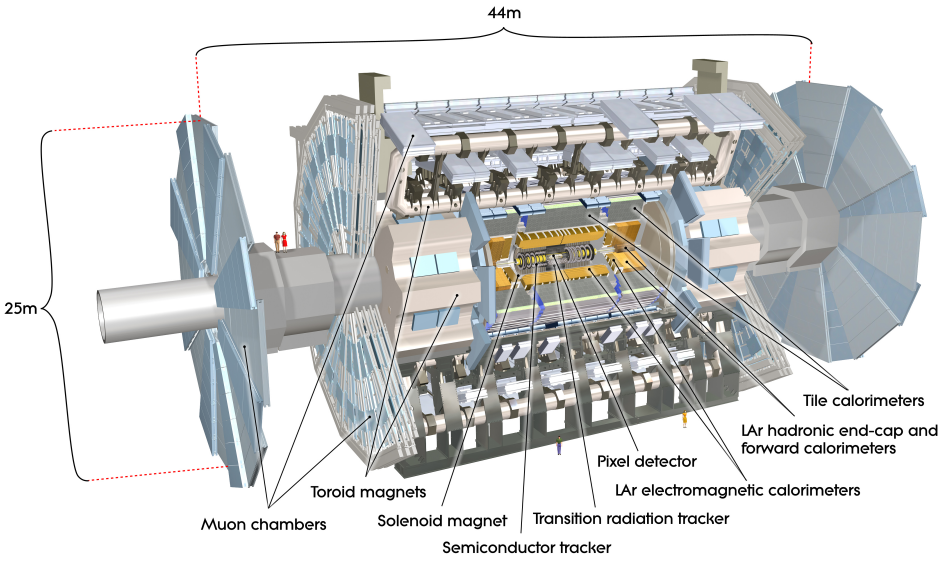


Figure 3.5: Cutaway diagram of the ATLAS detector, illustrating the Inner Detector (pixel detector, solenoid magnet and transition radiation tracker), the Liquid Argon Calorimeters, the Tile Calorimeter, the Muon Detectors, and the toroid and solenoid magnets [49].

3.2.1 The Inner Detector

The Inner Detector (ID) [50–53] is the tracking system of the ATLAS detector, it is the closest component to the beam pipe and is responsible for the reconstruction of charged particles trajectories and for the determination of their momenta. The ID provides excellent coverage up to $|\eta| < 2.5$ and full coverage in ϕ, with a transverse momentum resolution of $^{\sigma_{p_T}}/_{p_T} = 0.05\% p_T \oplus 1\%$ in the plane perpendicular to the beam axis and transverse impact parameter resolution of 10 μm for high momentum particles in the central η region.

As the charged particles pass through the ID, they ionize the medium of the multiple layers of the detector. The trajectories of these particles can be reconstructed from the individual electronic signals left in the different layers. The layout of the ID is illustrated in Figure 3.6. It consists of three different sub-components: the Pixel Detector [54], the Semiconductor Tracker (SCT) [55] and the Transition Radiation Tracker (TRT) [56]. Each sub-component is divided into two regions, a barrel region, in which the components are arranged in concentric cylinders around the beam direction, and end-cap regions, in which the components are located on disks perpendicular to the beam axis. The three sub-layers are

immersed in a magnetic field of 2 T parallel to the beam axis. The magnetic field, which is generated by a superconducting solenoid (Section 3.2.4) that extends over a length of 5.3 m and a diameter of 2.5 m, induces the deflection of the charged particles and allows the measurement of their momenta.

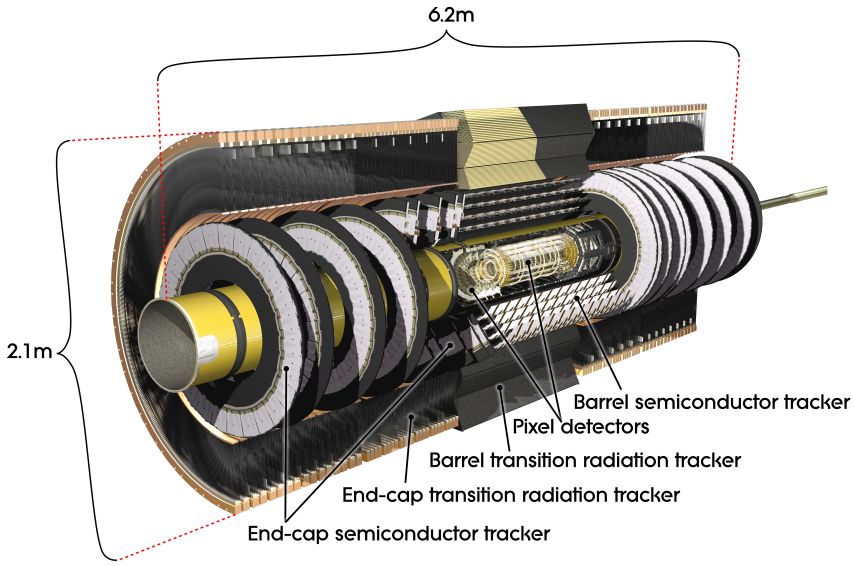


Figure 3.6: Cutaway diagram of the ATLAS inner detector, illustrating the Pixel detectors and the barrel and end-cap regions of the Semiconductor Tracker and of the Transition Radiation Tracker [57].

The pixel detector is the innermost layer of the ID and, therefore, receives the largest flow of particles, which requires the greatest granularity. The pixel tracker consists of 1744 silicon pixel modules arranged in three concentric barrel layers and two end-caps of three disks each. It provides at least three measurement points for particles emanating from the collision point. Each module contains 47232 pixels of size 50×400 μm^2 and a spatial hit resolution of 14 μm in the (r,ϕ) plane and 115 μm along z. In May 2014 the Insertable B-layer (IBL) [58] was installed to add an additional fourth pixel layer at a smaller radius, between the beam-pipe and the pixel detector, as shown in Figure 3.7. This upgrade was intended to enhance the track and vertex reconstruction and thus allow for a better identification of jets originating from b-quarks.

The SCT is positioned outside the Pixel detector and consists of 4088 modules of stereo silicon-strip sensors arranged in four concentric barrels layers and two end-caps of nine disks each. The sensors are composed of two layers of 6 cm long silicon strips. The layers of strips are arranged at a small angle of 40 mrad, allowing the

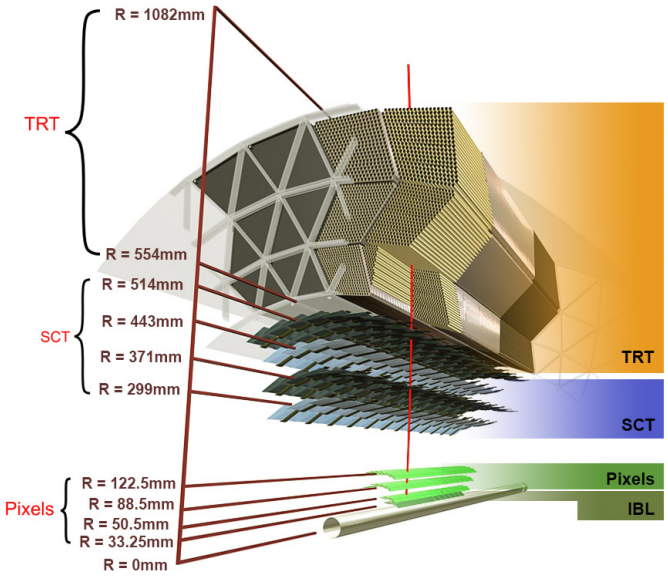


Figure 3.7: Detailed layout of the Inner Detector, including the Pixel Detector, the Semiconductor Tracker (SCT), the Transition Radiation Tracker (TRT) and the new Insertable B-Layer (IBL) [58].

measurement of both r and ϕ. Then, every barrel or disk will produce two strip measurements which are combined to build space-points. The SCT will provide eight precision measurement (four space-points) for each track.

The TRT is the outermost sub-detector layer of the ID. It consists of 298304 proportional drift tubes, referred to as straws, with a diameter of 4 mm. The straws are parallel to the beam axis in the barrel region and are radially arranged in wheel-like structures in the end-cap regions. The tubes are filled with a gas mixture based on Xe (70%), CO_2 (27%) and O_2 (3%). Whenever a charged particle passes through the drift tubes, it ionizes the gas mixture inside the tubes. The produced ions move in the electric field and generate a detectable signal proportional to the energy left by the particle. This measurement allows the distinction between different types of particles, such as electrons from heavier charged hadrons. The TRT tubes provide about 36 hits per track.

3.2.2 The Electromagnetic and Hadronic Calorimeters

Most of the particles that leave the ID (except muons and neutrinos) are stopped by the calorimeter system. This system is designed to provide precise energy deposition measurements with a good position accuracy. The electromagnetic

calorimeter (ECAL) covers the region $|\eta| < 3.2$ and provides measurements of the energy of electrons and photons, whereas the hadronic calorimeter (HCAL) covers the region $|\eta| < 4.9$ and provides measurements of the energy of hadronic showers (jets). The ATLAS calorimeter system is illustrated in Figure 3.8.

Both ECAL and HCAL are sampling calorimeters. Sampling calorimeters are made of alternating layers of an absorber material, which reduces the particle energy and induces showering, and an active material that provides the signal. The electromagnetic calorimeter is composed of layers of lead absorbers interleaved with electrodes immersed in Liquid Argon (LAr) active material. It is divided into a barrel region ($|\eta| < 1.475$) and two end-cap components ($1.375 < |\eta| < 3.2$). The barrel region consists of two identical half-barrels, separated by a small gap at the center, and each end-cap calorimeter is divided into two coaxial wheels.

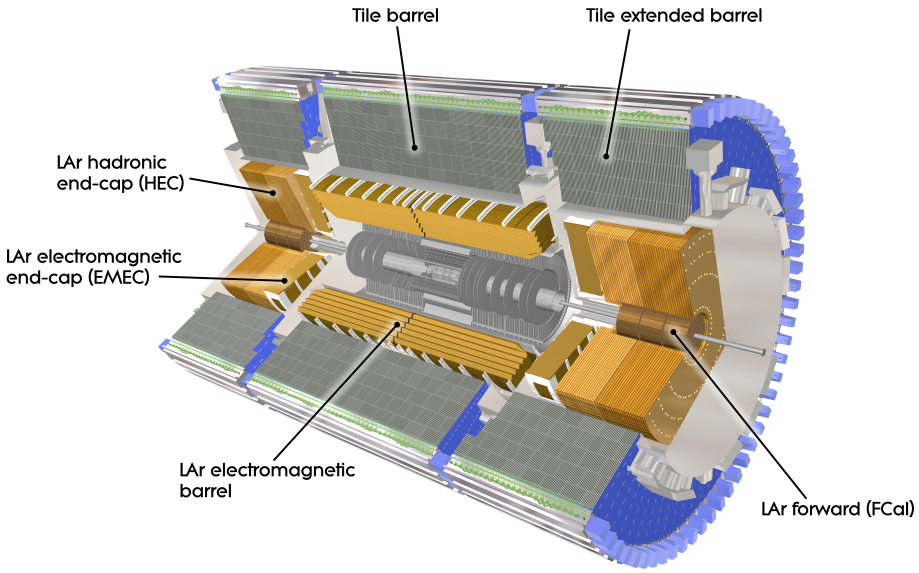


Figure 3.8: Cutaway diagram of the ATLAS calorimeter system, illustrating the barrel and end-cap regions of the Electromagnetic and Hadronic Calorimeters [59].

The HCAL is composed by the Tile calorimeter (TileCAL), the LAr hadronic end-cap calorimeter (HEC) and LAr forward calorimeter (FCAL). The TileCAL is located directly outside the ECAL system and provides a coverage range of $|\eta| < 1.7$. It uses steel as the absorber and scintillating tiles as the active material. The HEC extends the end-cap coverage out to $|\eta| < 3.2$ and consists of two independent wheels in each end-cap, placed directly behind the ECAL end-cap. The wheels are

composed of parallel copper absorber plates interleaved with 8.5 mm LAr gaps as the active medium. The FCAL is composed of three modules in each end-cap: The innermost layer uses copper as an absorber and is optimized for electromagnetic detection, the other two layers use tungsten as an absorber and are optimized for hadronic showers detection. The FCAL cover the range of $3.1 < |\eta| < 4.9$.

3.2.3 Muon Spectrometer

The Muon Spectrometer (MS) surrounds the calorimeter system and is responsible for the identification and measurement of momenta of muons, which are the only charged particles expected to leave the calorimeters. The MS covers the pseudorapidity range of $|\eta| < 2.7$. The tracking is based on the deflection of muons due to the presence of a magnetic field generated by toroidal magnets, which consist of a barrel toroid magnet of 1 T in the $|\eta| < 1.4$ region and end-cap toroids magnet of 0.5 T in the region $1.6 < |\eta| < 2.7$. In between $1.4 < |\eta| < 1.6$, the magnetic field is a combination of the fields produced by both toroids.

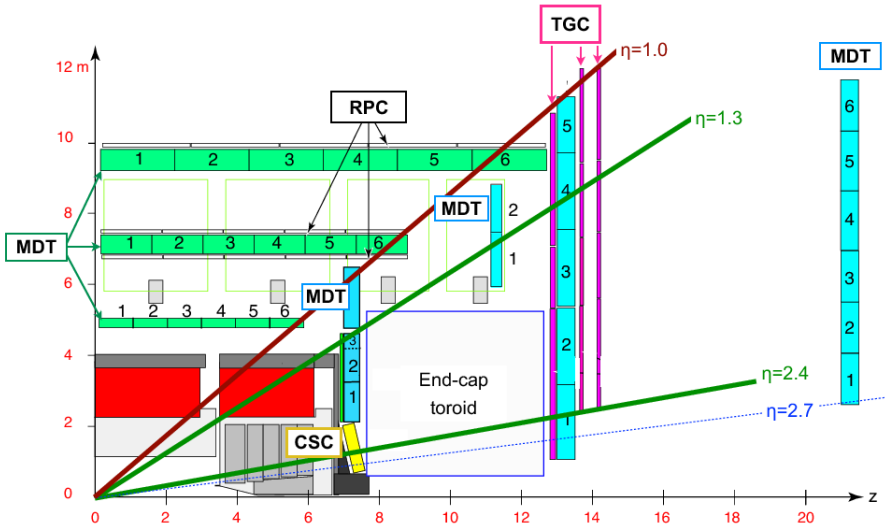


Figure 3.9: Schematic diagram showing a quarter-section of the ATLAS muon system in a plane containing the beam axis [60].

The MS is composed of four types of muon chambers: the Monitored Drift Tubes (MDT), the Cathode Strip Chambers (CSC), the Thin Gap Chambers (TGC) and the Resistive Plate Chambers (RPC). The layout of the Muon Spectrometer is shown in Figure 3.9. The MDT provide precise measurements of the

momentum in the pseudorapidity range $|\eta| < 2.7$, except in the innermost end-cap layer where their coverage is limited to $|\eta| < 2$. In this region the CSC are used instead, due to their higher rate capability and time resolution. The two fast trigger chambers TGC and RPC are used to complement the tracking system, the RPC in the barrel region ($|\eta| < 1.05$) and the TGC in the end-cap ($1.05 < |\eta| < 2.4$). These trigger chambers are capable of generating track information within a few tens of nanoseconds after the passage of the particle, as well as bunch-crossing identification.

3.2.4 Magnet System

The ATLAS magnet system is composed of one central solenoid [61], which provides the Inner Detector with a longitudinal magnetic field, and a toroidal magnet system [62, 63] that generates the field for the Muon Spectrometer. Figure 3.10 shows the layout of the ATLAS magnet system.

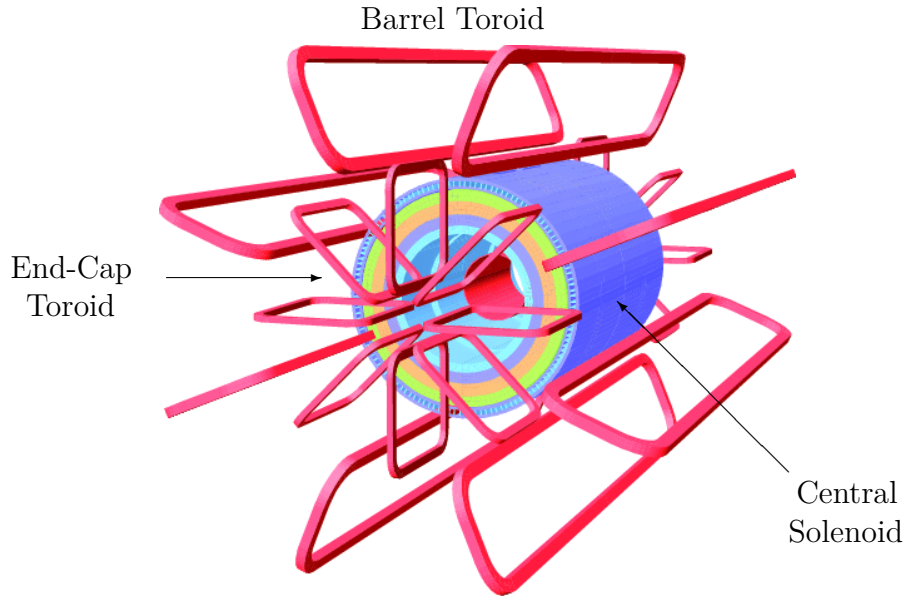


Figure 3.10: Schematic view of the layout of the ATLAS magnet system, illustrating the barrel toroid, the two end-cap toroids and the central solenoid. The central solenoid lies inside the calorimeter volume [31].

The central solenoid is a superconducting magnet positioned between the ID and the ECAL with 5.8 m in length with inner diameter of 2.46 m and outer diam-

eter of 2.56 m. The solenoid produces an axial magnetic field of 2 T in the center of the ID and was carefully optimized to keep the material as transparent as possible to traversing particles. The evenness and strength of this magnetic field induce the bending of highly energetic particles, allowing their moments to be measured with great precision. In addition, particles with small values of momentum (below hundreds of MeV) will loop repeatedly in the field and are unlikely to be measured.

The toroidal magnet system is located outside the HCAL and within the Muon Spectrometer. The toroidal system achieves a complete angular coverage with three very large air-cored toroids, a barrel toroid magnet and two end-cap toroids. This system provides a field of 1 T, in the barrel region, and 0.5 T, in the end-cap region, for the Muon Spectrometer. Each toroid is composed of eight air-core superconducting coils. The overall size of the barrel toroid system is 25.3 m in length, with inner diameter of 9.4 m and outer diameter of 20.1 m. While the end-cap toroids are 5 m deep, with inner diameter of 1.6 m and outer diameter of 10.7 m.

3.2.5 The Trigger and Data Acquisition System

Due to the large Luminosity and bunch-crossing frequency, the LHC produces roughly 10^9 proton collisions per second and, consequently, produces an enormous amount of data that cannot be stored, thus selective triggering is needed. The Trigger and Data Acquisition system (TDAQ) [64–66] performs a run-time event selection that effectively rejects the less interesting events and significantly reduces this large amount of data to a more manageable one. The TDAQ for the Run 2, shown in Figure 3.11, includes a hardware-based first-level trigger (L1) and a software-based high-level trigger (HLT).

The L1 process information from the calorimeters and the muon spectrometer to perform the initial event selection. It aims to identify signatures from high transverse momentum muons, electrons, photons, jets, and τ leptons decaying hadronically. It also searches for events with large total transverse energy and large missing transverse energy. The trigger identifies detector coordinates in η and ϕ where these objects were detected as regions of interest (RoI). These coordinates, the type of event identified and the energy is stored in the RoI data that will be subsequently used by the HLT. The L1 trigger decides whether to store the event in less than 2.5 μs after the bunch crossing. This reduces the approximately 40 MHz input event rate to approximately 100 kHz.

The data from the RoIs received from L1 as well as the data from the full event can be used in the HLT where it is processed through sophisticated reconstruction and selection algorithms using high granularity calorimeter information, tracking information from the ID, and precision measurements from the muon spectrometer. The threshold cuts and the particle identification are improved due to better information on energy deposition and track reconstruction in the inner detector. The event rate is reduced from the L1 input of 100 kHz to approximately 1 kHz. The events selected by the HLT are stored for further analysis.

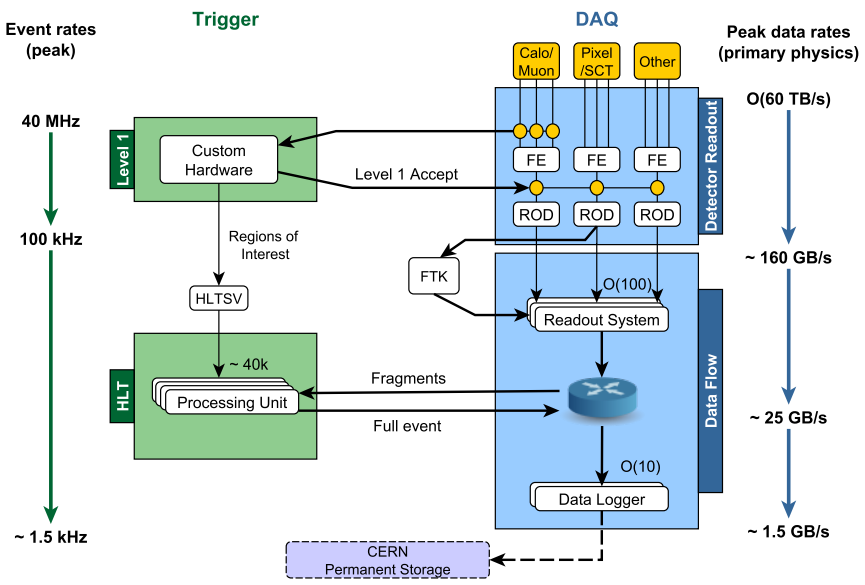


Figure 3.11: Diagram of the ATLAS Trigger and Data Acquisition system in Run 2 showing expected peak rates and bandwidths through each component [67].

4

State of the Art

This chapter presents the state of the art of $t\bar{t}H$ searches and studies of the CP properties of the top quark Yukawa interaction. The importance of obtaining a precise measurement of the top quark Yukawa coupling through $t\bar{t}H$ production is addressed in Section 4.1, together with a description of this channel. The Lagrangian considered in this analysis for the top Yukawa coupling that accounts for a possible CP-odd component is detailed in Section 4.2. Then, Section 4.2.1 describes several CP sensitive variables that were used during the training of multivariate techniques and in the fitting. Finally, Section 4.2.2 gives an overview of the experimental status of the $t\bar{t}H$ production and CP measurements.

4.1 The $t\bar{t}H$ production

As seen in the Section 2.7, it is necessary to investigate the structure and the CP properties of the top quark Yukawa coupling. The measurement of a CP-odd component in one of the Higgs boson couplings would constitute an important discovery of new physics. Furthermore, it could potentially help in the explanation of the observed asymmetry between matter and antimatter in the Universe, which requires new sources of CP violation in addition to the ones predicted by the SM.

The top quark has the strongest Yukawa coupling[1] to the Higgs boson due to its large mass. Deviations from the SM would be more evident in this coupling, which makes it a strong candidate for this measurement. On-shell top quarks are too massive to be produced in any Higgs boson decay. Instead, the coupling y_t can be directly[2] obtained through $t\bar{t}H$ and tH ($tHjb$ and WtH) processes. The

[1] From equation (2.35) we have $y_t = \sqrt{2}m_t/v \approx 1$. This value may indicate a possible special role of the top quark in the EWSB mechanism.

[2] The top quark lifetime is small ($\approx 0.5 \times 10^{-24}$ s [6]), it decays before occurring hadronization

ATLAS and CMS Collaborations reported in 2018 the first observation of the $t\bar{t}H$ production [3, 4]. Figure 4.1 shows examples of tree-level diagrams for this process.

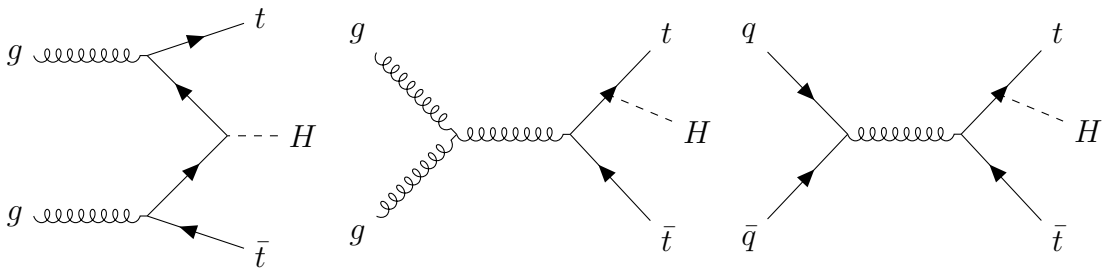


Figure 4.1: Tree-level Feynman diagrams for the $t\bar{t}H$ production at the LHC.

The top quarks decay mainly into a W boson and a bottom quark, with a branching fraction of $(95.7 \pm 3.4)\%$ [6]. Then the bottom quark hadronizes and the W boson can decay hadronically, into a quark-antiquark pair, or leptonically, into a charged lepton and the corresponding neutrino. Table 4.1 shows the branching fraction of the different top quark decay modes.

Table 4.1: Top quark branching ratios [6].

Decay channel ($t \to Wb$)	Branching ratio (%)
$t \to e\nu_e b$	13.3 ± 0.6
$t \to \mu\nu_\mu b$	13.4 ± 0.6
$t \to \tau\nu_\tau b$	7.1 ± 0.6
$t \to q\bar{q}b$	66.5 ± 1.4

Three different channels are expected for the $t\bar{t}H$ process, according to the possible decay modes of the W boson originated from the decay of the two top quarks:

- The all-hadronic channel, where both W bosons decay hadronically. Although this channel present the highest branching ratio (about 46%), it is extremely challenging to distinguish it from the large multijet background, in view of the large number of jets in its final state.

and transfers its spin information directly to its decay products.

- The dilepton channel, where both W bosons decay leptonically. This channel has the lowest branching ratio (about 9%) and it is less precisely reconstructed because of the two neutrinos in the final state. However, it is easily distinguished from the background due to the presence of two isolated leptons.
- The single lepton or l+jets channel, where one W boson decays leptonically and the other decays hadronically. This channel contains one lepton (and only one neutrino) in its final state and, thus, it presents a clean signature with a higher branching ratio of about 30% (not including the τ leptons, which further decay into hadrons or leptons and are treated separately).

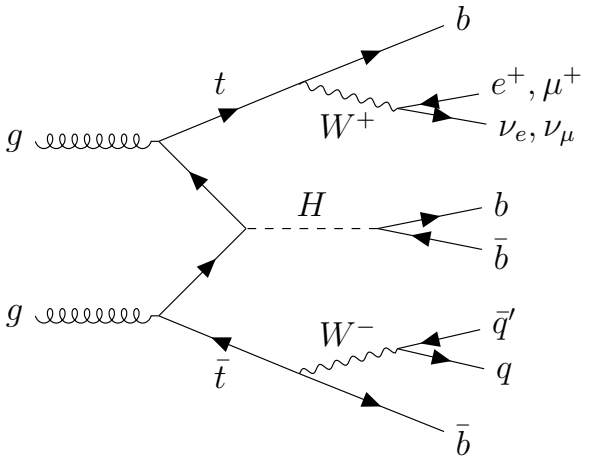


Figure 4.2: Tree-level Feynman diagrams for the $t\bar{t}H(H \to b\bar{b})$ channel, including the subsequent decays of the top quarks.

At the LHC, the $t\bar{t}H$ production accounts for only 1% of the total Higgs boson production cross-section (Figure 2.6). In order to compensate this small cross-section, analyses may take advantage of the large branching fraction of the $H \to b\bar{b}$ decay channel (Figure 2.7). However, this channel also suffers from the presence of an overwhelming background of $t\bar{t}$+jets events and a $t\bar{t}b\bar{b}$ irreducible [3] component, which is badly modelled and introduces challenges to the analysis. The $t\bar{t}H(H \to b\bar{b})$ channel is represented in Figure 4.2 and is the channel used in this thesis.

[3] The irreducible background has the same final state particles as the ones we are searching for. The more stringent requirements are applied to eliminate this background, more signal events are lost. This creates a trade off between background rejection and signal acceptance.

4.2 Study of the $t\bar{t}H$ CP Properties

Deviations to the top Yukawa coupling from the SM may be described as [20]

$$\hat{\mathcal{L}}_{htt} = -k_t \frac{y_t}{\sqrt{2}} \hat{\bar{\psi}}_t \hat{\psi}_t \hat{h}, \tag{4.1}$$

where k_t parameterizes deviations from the SM Yukawa coupling to the top quark from equation (2.35). A value of $k_t = 1$ corresponds to the SM case, and $k_t \neq 1$ would have a direct impact on the $t\bar{t}H$ cross-section. The current measurements are consistent with the $k_t = 1$ case within an uncertainty of approximately 20% [3, 4]. However, the effects of physics BSM could also arise from a different top Yukawa coupling structure and not only by deviations in the total $t\bar{t}H$ cross-section. This would originate significant differences also on the predicted $t\bar{t}H$ kinematics.

The existence of CP-odd Higgs boson is predicted by certain BSM models, such as the two-Higgs doublet model and the MSSM. If multiple Higgs doublets are considered, the resulting Yukawa coupling will contain two components, one CP-even and one CP-odd [68]. The pure CP-odd scenario was already excluded at 99.98% confidence level [26, 27]. However, couplings with a CP-odd component are still permitted by experimental data [69].

The most general Yukawa Lagrangian that accounts for a possible CP-odd $t\bar{t}H$ contribution can be written as [70]:

$$\hat{\mathcal{L}}'_{htt} = -k'_t \frac{y_t}{\sqrt{2}} \hat{\bar{\psi}}_t \left(\cos\alpha + i\gamma_5 \sin\alpha\right) \hat{\psi}_t \hat{h}, \tag{4.2}$$

where $y_t = \sqrt{2}m_t/v$ is the SM Higgs Yukawa coupling, α represents the CP mixing angle that regulates the relative phase of the scalar and pseudoscalar components, and k'_t is the generalized coupling modifier. The SM is obtained when $|\cos\alpha| = 1$ and $k'_t = 1$, in that case h is the SM (CP-even) Higgs boson, which is usually represented by H ($J^{CP} = 0^{++}$). The quantity $\bar{\psi}_t \gamma_5 \psi_t$ is a pseudoscalar, it transforms as a Lorentz scalar as well as $\bar{\psi}_t \psi_t$, but it is odd under parity transformations. The pure pseudoscalar scenario occurs when $\cos\alpha = 0$, in that case h is usually represented by A ($J^{CP} = 0^{+-}$).

The Lagrangian can also be written as

$$\hat{\mathcal{L}}'_{htt} = -\frac{y_t}{\sqrt{2}} \hat{\bar{\psi}}_t \left(k_t + i\gamma_5 \tilde{k}_t\right) \hat{\psi}_t \hat{h} \quad \text{with} \quad \begin{cases} k_t = k'_t \cos\alpha \\ \tilde{k}_t = k'_t \sin\alpha \end{cases} \tag{4.3}$$

where the pure CP-even interaction is given by $k_t = 1$ and $\tilde{k_t} = 0$, while the pure CP-odd interaction is given by $k_t = 0$ and $\tilde{k_t} = 1$.

4.2.1 Phenomenological Studies

Several variables, that are sensitive to the scalar and pseudoscalar components of the fermions Yukawa couplings, have been extensively explored by recent phenomenological papers [71–74] in order to achieve a better discrimination between the different CP scenarios, and also to reduce the contribution from the dominant irreducible background ($t\bar{t}b\bar{b}$). These observables exhibit significant differences between the samples of pure CP-evens, pure CP-odd and the background. The variables that show the greatest discriminating power can be used in the training of a multivariate algorithm that may be used in the fit (Section 6.4).

The first set of observables used in this work includes angular variables that are defined as a product of trigonometric functions of two distinct three-dimensional angles calculated in different reference frames. Figure 4.3 shows examples of these angular variables. The notation θ_y^x represents the angle between the direction of the y system, in the rest frame of x, and the direction of the x system, in the rest frame of its parent system. In the phenomenological studies, these observables were chosen based on the best forward-backward asymmetry, as well as the set of variables that presented the lowest correlation between them. In this thesis, the choice of variables was based on the classification efficiency of the multivariate algorithm built with some of these variables. The procedure used in this thesis to classify the CP structure of events is further detailed in Section 6.4 and 6.5.

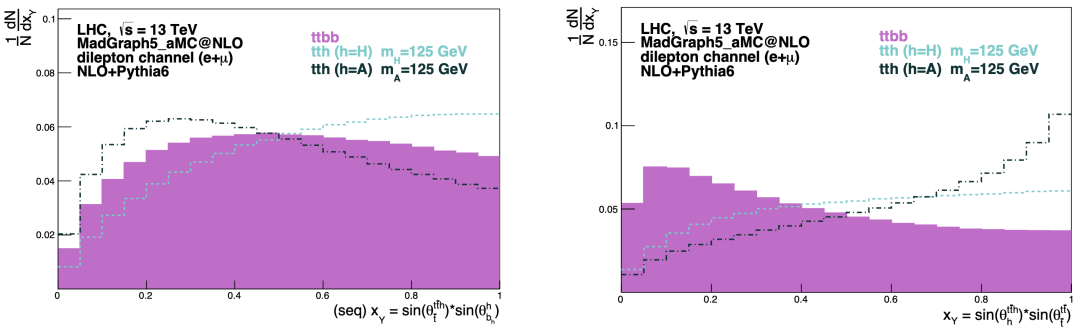


Figure 4.3: Distributions of angular variables for the dilepton channel after event selection and full kinematic reconstruction. The distributions are shown for pure CP-even $t\bar{t}H$ and pure CP-odd scenario and for the $t\bar{t}b\bar{b}$ background [72, 73].

In addition to the angular variables, another set of observables, defined in terms of transverse and longitudinal projections of the particle momenta with respect to the beam axis, was also used:

$$b_2^f(i,j) = \frac{(\vec{p}_i^{\,f} \times \hat{k}_z) \cdot (\vec{p}_j^{\,f} \times \hat{k}_z)}{|\vec{p}_i^{\,f}||\vec{p}_j^{\,f}|} \tag{4.4}$$

$$b_4^f(i,j) = \frac{(\vec{p}_{i,z}^{\,f}) \cdot (\vec{p}_{j,z}^{\,f})}{|\vec{p}_i^{\,f}||\vec{p}_j^{\,f}|} \tag{4.5}$$

where $\vec{p}_i^{\,f}$ is the total momentum of the particle i and $\vec{p}_{i,z}^{\,f}$ its z-component, with $i,j = \{t, \bar{t}, h\}$ and f being the reference frame where the momentum is measured. $\hat{k}_z$ corresponds to the beam line, which defines the z-direction. Figure 4.4 shows the distribution of the b_2^f for $i = t$ and $j = \bar{t}$ calculated at the $t\bar{t}h$ reference frame and b_4^f calculated at the laboratory frame.

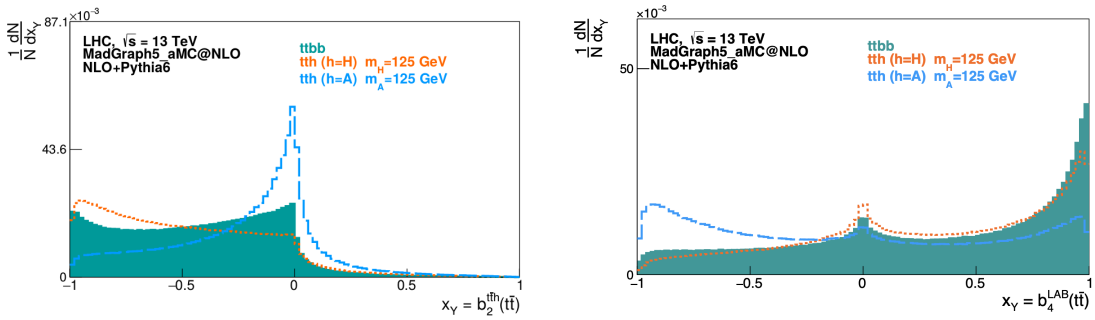


Figure 4.4: Normalized $b_2^{t\bar{t}h}(t\bar{t})$ (left) and $b_4^{LAB}(t\bar{t})$ (right) distributions for pure CP-even $t\bar{t}H$ and pure CP-odd scenario and for the $t\bar{t}b\bar{b}$ background, at NLO including shower effects [74].

4.2.2 Observation of $t\bar{t}H$ Production and CP Measurements

The observation of the $t\bar{t}H$ production by the ATLAS and the CMS collaborations [3, 4] was announced in 2018 and proves the tree-level coupling of the Higgs boson to the top quarks. The ATLAS measurement was obtained through the combination of $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ decay channels with $H \to b\bar{b}$ and multilepton searches from Refs. [75, 76]. The analysis used a dataset corresponding to an integrated luminosity of 79.8fb^{-1} at $\sqrt{s} = 13$ TeV collected in 2017 for the $\gamma\gamma$ and ZZ^* decay modes, while for the other channels a 36.1fb^{-1} dataset collected in 2015 and 2016 at $\sqrt{s} = 13$ TeV was used.

The events in the single-lepton and dilepton channels were classified into non-overlapping regions based on the total number of jets, as well as the number of b-tagged jets at different b-tagging efficiency. A combined fit was performed to the different regions of the individual analyses using the profile likelihood method[4]. All Higgs boson production processes besides $t\bar{t}H$ production were considered as background and their cross sections were fixed to the SM expectations. The respective cross-section uncertainties were treated as systematic uncertainties. Figure 4.5 shows the $t\bar{t}H$ signal strength measured in the analysis.

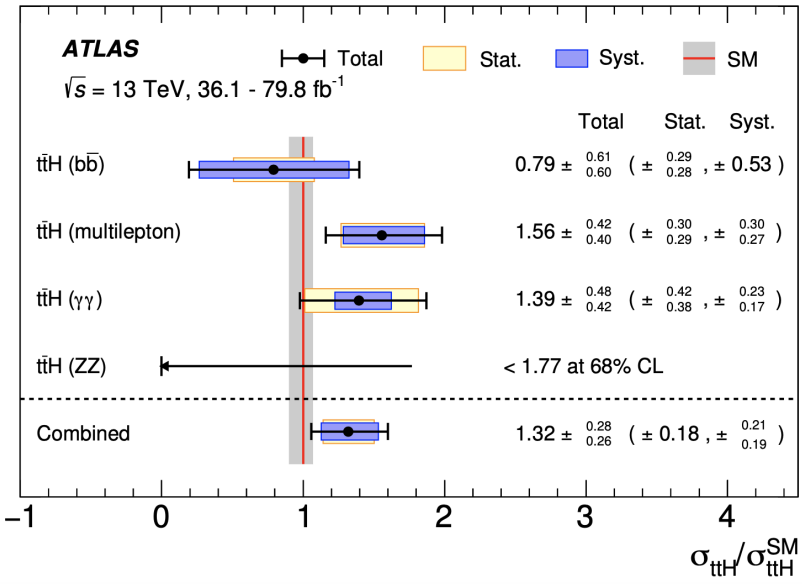


Figure 4.5: Measurements of the $t\bar{t}H$ production cross section divided by the SM prediction. the statistical and systematic uncertainties, as well as the total uncertainty are represented. The SM cross-section prediction is indicated by a red vertical line, together with a grey band that represents its uncertainties [3].

In 2020, both the CMS and the ATLAS collaboration reported the first measurement sensitive to the CP structure of the top Yukawa interaction in the $t\bar{t}H(H \to \gamma\gamma)$ channel [77, 78]. The $t\bar{t}H(H \to \gamma\gamma)$ ATLAS analysis was performed using 139 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data from 2015 to 2018 recorded with the ATLAS detector. This analysis was based on two different independent boosted decision trees (BDT) algorithmin order to improve the analysis sensitivity. One BDT is used to separate events into signal ($t\bar{t}H$ and tH events[5]) and into

[4]For a more detailed discussion on the profile likelihood method see chapter A

[5]tH production is also included as signal since it is particularly sensitive to variations from SM due to destructive interference between CP-even and CP-odd tH diagrams. The tH cross

background, and the other to discriminate between the two different CP scenarios (CP-odd and CP-even).

The events are classified into different categories, according to the output of the background rejection BDT. The number of categories and their boundaries are optimized based on the discriminating potential between the CP-even and CP-odd scenarios. In total, 12 categories were used in the hadronic region and 8 in the leptonic region (where at least one top quark decays leptonically). Figure 4.6 shows the distribution of data in the hadronic channel in CP BDT as function of background rejection BDT. The blue lines corresponds to pure SM signal, and the red lines corresponds to pure CP-odd signal. The projections of the two BDTs are also shown in the figure.

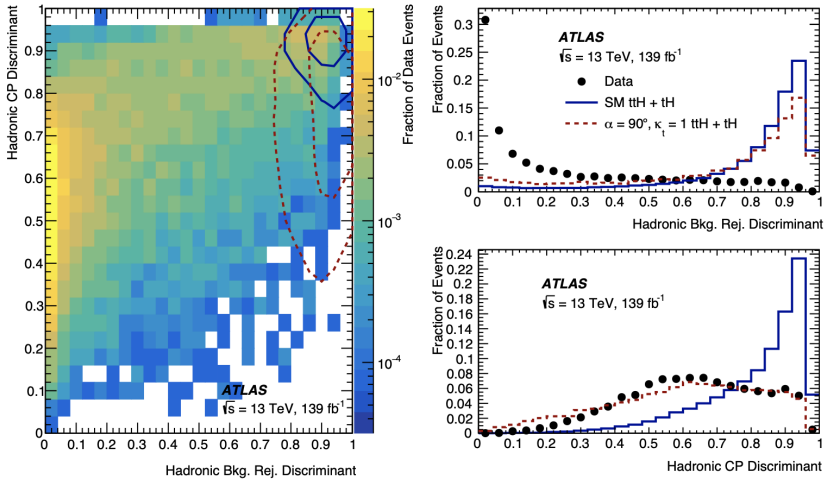


Figure 4.6: Two-dimensional BDT distribution in the selected data events from the Hadronic region showing the Background Rejection BDT and CP BDT (left). The inner (outer) lines capture 25% (50%) of the $t\bar{t}H$ and tH signal events for CP-even (blue) and CP-odd (red) hypotheses. Together with the projections onto the background rejection and CP BDT axes (right) [78].

A fit is performed to the diphoton invariant mass $m_{\gamma\gamma}$ distribution in all the categories simultaneously. The $t\bar{t}H$ production is observed with a significance of 5.2σ, assuming a CP-even coupling. The tH process is not observed with a limit of 12 times the SM expectation at 95% confidence level. The analysis exclude values of α greater than 43° at 95% confidence level and a pure CP-odd coupling ($\alpha = 90$) at 3.9σ. Figure 4.7 shows the observed 2D contour of $k'_t \cos\alpha$ and $k'_t \sin\alpha$.

section strongly varies with α [79]. It is also present in the analysis after event selection and gives sensitivity to the sign of the CP mixing angle (Section 7.1).

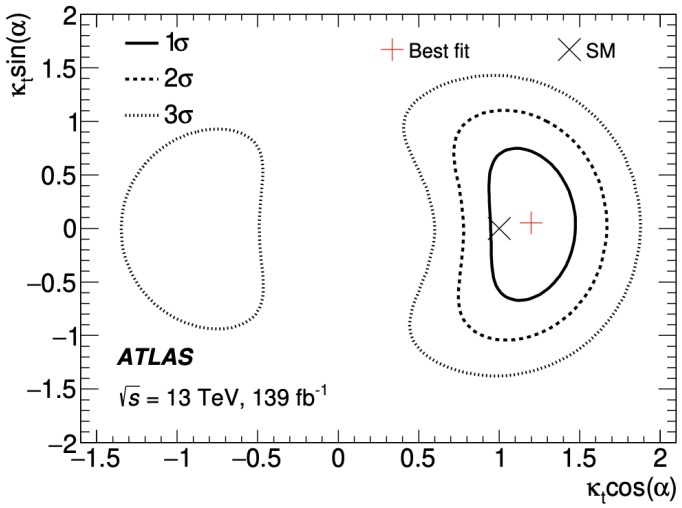


Figure 4.7: Two-dimensional likelihood contours for $k'_t \cos\alpha$ and $k'_t \sin\alpha$ with ggF and $H \to \gamma\gamma$ constrained by the Higgs boson coupling combination [78].

In summary, all the measurements are in agreement with the SM and the analysis constrain the pure CP-odd coupling. However, since the Higgs-photon coupling is loop-induced, the $H \to \gamma\gamma$ could be affected by BSM effects, and a search in the $H \to b\bar{b}$ decay channel is still necessary.

The first analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons [80] was also published in 2020 by the CMS collaboration. The study uses the full pp dataset collected at $\sqrt{s} = 13$ TeV by the CMS detector in 2016, 2017 and 2018, corresponding to an integrated luminosity of 137fb^{-1}. The analysis selects events of ggF, VBF and VH production where the Higgs decays into two τ leptons, and one τ decays to a muon and the other hadronically or when both τ leptons decay hadronically. Several multivariate analysis techniques, that exploits angular correlations between the decay planes of the τ leptons, are used to discriminate between the different decay modes and separate the signal from background events. The analysis obtained a value of $4 \pm 17°$ for the CP mixing angle at 68% confidence level and $4 \pm 36°$ at 95% confidence level. The pure CP-odd Higgs boson was exclude with 3.2σ. Figure 4.8 shows the obtained contours on the τ Yukawa couplings.

In summary, all of these searches are consistent with the SM and severely constrain the allowed parameter space for new physics. However, it is important to study all Yukawa couplings since any deviation from a purely scalar interaction

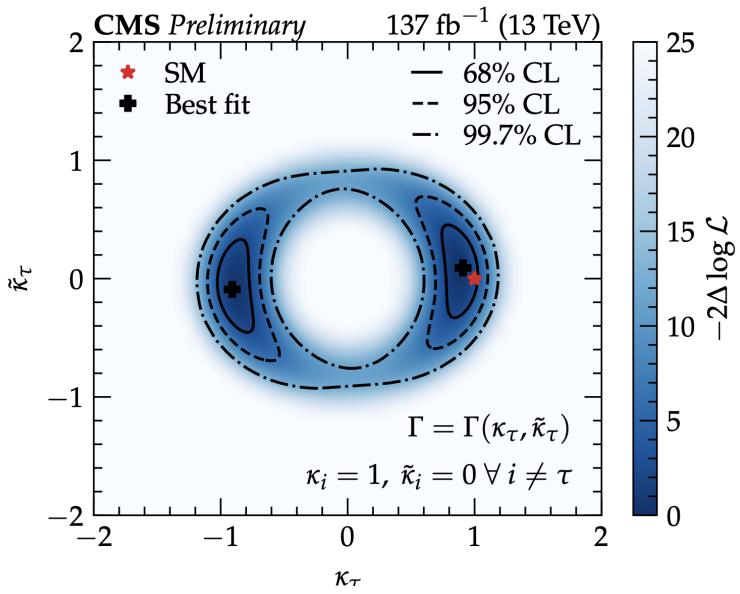


Figure 4.8: Two-dimensional likelihood contours in the $(k_\tau, \tilde{k}_\tau)$ plane [80], where k_τ and $\tilde{k}_\tau$ are the reduced CP-even and CP-odd τ Yukawa couplings, respectively.

would be a direct indication of physic BSM. Besides, CP violation in the Higgs couplings may be small, as expected in several BSM models, and more precise measurements are required. These measurements are expected to be achieved with the increase of the LHC luminosity and during the high luminosity LHC (HL-LHC) plan (Section 3.1.1) or with future colliders. It would be interesting to predict when such level of precision will be achieved. In this work projections of the $t\bar{t}H(H \rightarrow b\bar{b})$ analysis for higher values of luminosity are also presented in Section 7.2 and the expected limits to the CP mixing angle for the HL-LHC are obtained.

5

Event Generation and Object Reconstruction

In the first section of this chapter, the necessary techniques and concepts behind the generation of Monte Carlo (MC) events and detector simulation are briefly described. The MC event generators, as well as the data and MC samples used in this analysis are presented. In the second section, the reconstruction of final state objects is described in detail. Finally, in the third section, a study of the validation of the fast detector simulation Atlfast-II [81] is presented, using variables that exploit jet substructure information with $t\bar{t}H$ and $t\bar{t}$ samples in the boosted regime.

5.1 Monte Carlo Event Generation

The simulation of proton collisions plays an essential role in almost all experimental analyses and planning for future experiments. The simulated events are obtained by tools called Monte Carlo event generators that use numerical MC techniques to provide a detailed simulation of particle collisions. The collision is described by a few main steps that take place at different energy scales: from very short distance scales, up to the scale of hadron formation and decay. These can be treated separately. At short distances, the constituent partons[1] of the incoming protons interact to produce a few energetic outgoing partons, leptons or gauge bosons. The matrix elements (ME) of these hard processes are calculated with perturbative QCD. At larger distances, the confinement of partons into hadrons dominates. These soft processes cannot be calculated perturbatively but have to be phenomenologically modelled.

[1] In the parton model, the hadrons are considered as a composition of point-like particles called partons. These are quarks, antiquarks and gluons.

As represented in Figure 5.1, the basic steps involved in the proton collisions processes that need to be simulated are: the hard process, the parton shower, the hadronization and the subsequent decay of hadrons.

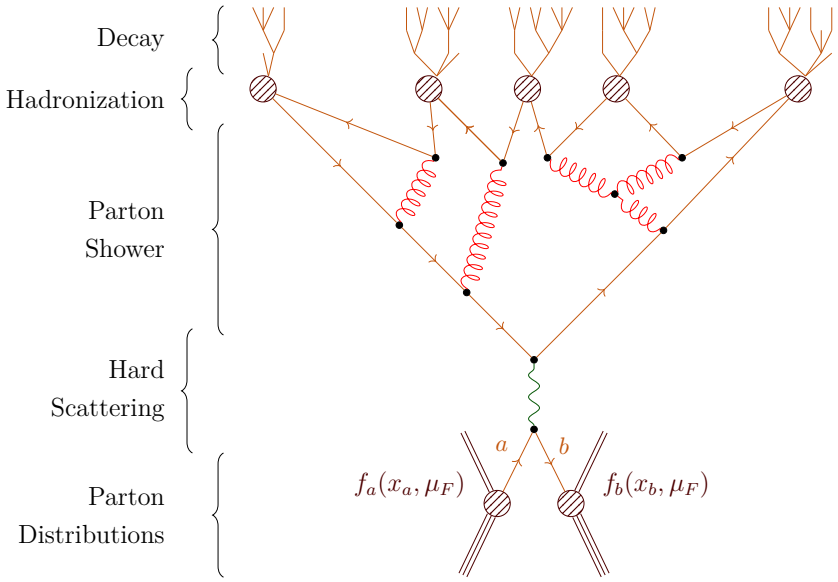


Figure 5.1: Diagram showing the structure of a proton-proton collision event generation. The steps in the event generation chain are indicated in the figure.

For each collision between bunches of hadrons, only a small number of events contain a high momentum-transfer process of interest. Since it is not feasible to simulate all possible collisions, only the hard process is simulated, which is the process with the highest momentum transfer in the event. The hard process can be described by perturbation theory, as partons are asymptotically free at high energies. The cross sections for a hard scattering process, such as $pp \to n$, at hadron colliders can be computed in collinear factorization through [82]

$$\sigma = \sum_{ab} \int dx_a dx_b f_a(x_a, \mu_F) f_b(x_b, \mu_F) d\hat{\sigma}_{ab\to n}(\mu_F, \mu_R) \tag{5.1}$$

$$= \sum_{ab} \int dx_a dx_b d\Phi_n f_a(x_a, \mu_F) f_b(x_b, \mu_F) \times \frac{1}{2\hat{s}} |\mathcal{M}_{ab\to n}|^2(\Phi_n; \mu_F, \mu_R), \tag{5.2}$$

where $f_a(x_a, \mu_F)$ and $f_b(x_b, \mu_F)$ are the parton distribution functions (PDFs), which are the probability densities associated to a parton of type (a) and (b) carrying a given fraction x of the respective initial proton momentum. $\hat{\sigma}_{ab\to n}$ is the cross section for the partonic process, that depends on the μ_F factorization scale and on the μ_R renormalization scale. $d\Phi_n$ denotes the differential phase

space element over the n final-state particles, $1/2\hat{s}$ represents the parton flux and $|\mathcal{M}_{ab\to n}|^2$ is the ME squared of the partonic process.

Scattered color charges emit QCD radiation in the form of gluons. The initial and final state parton involved in the hard process are colored particles and will therefore radiate gluons. The gluons themselves will radiate others gluons and also produce pairs of quark–antiquark. This will lead to showers of outgoing partons called parton showers. As the parton shower evolves downwards in momentum scale it reaches a point where perturbation theory breaks down and it is necessary to switch to hadronization models. At low energy scales, the uncolored partons start to bind together and form colorless hadrons. These models describe the confinement of partons into hadrons. Many of the hadrons are produced in excited states during hadronization and need to be further decayed into stable states. In addition to the hard process, more than one pair of partons from the initial protons may interact with each other. These secondary interactions form the underlying event in each proton collision and contaminate the hard process that was already simulated. The underlying event is also simulated using phenomenological models, whose parameters are tuned using experimental data.

In this analysis, MadGraph5_aMC@NLO [83], PowhegBox [84–86], OpenLoops [87, 88] and Sherpa [89] were used for the generation of events. To preserve all spin correlations, the MadSpin package [90, 91] was used for decays at LO of the events produced with MadGraph5_aMC@NLO. Together with Pythia8 [92–94] or Herwig7 [95, 96], for the simulation of parton shower and hadronization. The following section explains which generators were used for each samples.

After the simulation of parton showers, hadronization and decay of most of the unstable particles, a further simulation can be performed to account for the response of a specific detector and obtain a more realistic description of what would be observed in the LHC. For physics analyses that require a high level of accuracy, a complex detector simulation software, such as Geant4 [97], is used to fully simulate the interactions of particles with the detector matter content, the electronic response of each detector component, and the algorithms used to identify and reconstruct the different physics objects. In this analysis, both the full ATLAS detector simulation [98] based on Geant4 and a faster simulation [81] were used. The fast simulation replaces the full Geant4 simulation by parameterizations of the ECAL and HCAL showers. Both simulations were found to give similar results for the variables used in this analysis.

5.1.1 Data and Monte-Carlo samples

The proton–proton collision data used in this analysis were recorded by the ATLAS detector from 2015 to 2018 at a center-of-mass energy of $\sqrt{s} = 13$ TeV, and corresponds to a total integrated luminosity of 139.0 fb^{-1} for the full run-2 dataset.

The generated MC samples are divided into signal and background. For the background samples, several processes were produced. The $t\bar{t}$+jets, composed by $t\bar{t} + \geq 1b$, $t\bar{t} + \geq 1c$, and $t\bar{t}$+light jets, are the most relevant background processes for this analysis, especially the large irreducible $t\bar{t}b\bar{b}$ production that is associated to a large theory uncertainty on its modelling. The $t\bar{t}b\bar{b}$ samples were modelled at NLO in QCD with PowhegBoxRes and OpenLoops, using Pythia8 for the PS and hadronization. These samples were used as nominal $t\bar{t} + \geq 1b$ prediction. The $t\bar{t} + \geq 1c$ and $t\bar{t}$+light samples were generated at NLO with PowhegBox v2 and Pythia8.230 for PS and hadronization.

In addition, other important backgrounds were also produced. The single top and tW production were generated using PowhegBox v2 at NLO. The PS and hadronization was modelled with Pythia8.230. The $t\bar{t}Z$ and $t\bar{t}W$ processes were produced with MadGraph5_aMC@NLO v2.3.3 at NLO and Pythia8.210 for showering and hadronization. The production of four tops, tZq and tWZ were also produced with MadGraph5_aMC@NLO v2.3.3, but PS and hadronization were modelled using Pythia8.230 for the $t\bar{t}t\bar{t}$ process and Pythia8.212 for tZq and tWZ. Finally, Z+jets, W+jets and diboson production were generated with Sherpa v2.2.1 and Sherpa PS. The diboson production was modelled at NLO accuracy, while the Z+jets and W+jets processes were produced at NLO for up to 2 jets, and LO for up to 4 jets.

Table 5.1 contains a list of the background samples used in the analysis and the MC generators used in the production of each sample.

For the signal samples, the $t\bar{t}H$ samples were generated at NLO in QCD using the MadGraph5_aMC@NLO generator and parton showers were modelled with Pythia8. Three samples were generated using the Lagrangian described by equation (4.3), that allows for a CP-odd term in the top–Higgs coupling. The first for the SM ($\alpha = 0$), the second corresponding to pure CP-odd interaction ($\alpha = 90°$) and the other for the maximal mixing between CP-odd and CP-even ($\alpha = 45°$). States with different CP composition were generated by reweighting the

Table 5.1: Background samples generated in the analysis, the generators used to simulate the process considered.

Sample	QCD order	ME Generator	PS Generator
$t\bar{t}b\bar{b}$	NLO	PowhegBoxRes and OpenLoops	Pythia8
$t\bar{t}+\geq 1c$ and $t\bar{t}$+light	NLO	PowhegBox v2	Pythia8.230
Single-top and tW	NLO	PowhegBox v2	Pythia8.230
$t\bar{t}V$ ($V = Z$,W)	NLO	MadGraph5_aMC@NLO v2.3.3	Pythia8.210
$t\bar{t}t\bar{t}$	NLO	MadGraph5_aMC@NLO v2.3.3	Pythia8.230
tZq	LO	MadGraph5_aMC@NLO v2.3.3	Pythia8.212
tWZ	NLO	MadGraph5_aMC@NLO v2.3.3	Pythia8.212
V+jets ($V = Z$,W)	NLO (up to $2j$) + LO (up to $4j$)	Sherpa v2.2.1	Sherpa PS
Diboson	NLO	Sherpa v2.2.1	Sherpa PS

pure CP-even and CP-odd signals as described in Section 6.6. Then, the sample with maximal mixing was used for validating that procedure. The samples were generated using the Higgs Characterization [99] model implemented in Feynrules [100, 101], that accounts for the Yukawa coupling presented in Section 4.2.

The tWH and $tHjb$ processes were also included as signal samples as discussed in Section 4.2.2. These were also generated using the MadGraph5_aMC@NLO at NLO in QCD and showered with Pythia8. For both processes, multiple samples were used for different values of α and k_t': α from 0 to 90° in steps of 15° with $k_t = 1$, $\alpha = 0$ with $k_t = (-1, 0.5, 2)$ and, finally, $\alpha = 45°$ with $k_t = 2$.

5.2 Object Reconstruction

After the detection of the signals produced by the interactions of the particles and the detector components explained in Chapter 3.2, the events need to be reconstructed into physical objects (such as electrons, muons or jets) so the momentum, the vertex and tracks of each particle, as well as the physical processes, can be determined. Both the data and the MC samples need to be reconstructed for the analysis. The main physics objects that are present in this work are electrons, muons, jets, and b-jets, which are the possible final state objects of the $t\bar{t}H(H \to b\bar{b})$ channel.

5.2.1 Leptons

The reconstruction of electrons is performed by associating the clusters of energy detected in the ECAL with reconstructed tracks in the ID [102]. The elec-

tron candidates, originated from the decays of the top quarks, are required to have $p_T > 10$ GeV with $|\eta| < 2.47$ and be outside the calorimeter barrel–endcap transition region ($1.37 < |\eta| < 1.52$). Then, an identification algorithm is applied to determine whether the electron candidates are signal-like or background-like (fake electrons) objects, such as jets or photons mimicking the electron signature. This algorithm is based on a likelihood discriminant that combines variables related to the electron cluster and track measurements. Three identification working points (WP) are typically provided: *Loose*, *Medium*, and *Tight*, with increasing background rejection power. In this analysis, electrons must satisfy the *Medium* identification criterion.

The reconstruction of muons is achieved by matching track segments or complete tracks in the Muon Spectrometer (MS) with tracks reconstructed in the Inner Detector (ID) [103]. The hits from both the ID and MS sub-detectors are used in a global refit to obtain the combined track. Muons are usually first reconstructed in the MS, then extrapolated inward and matched to an ID track. In this work, muons are required to have $p_T > 10$ GeV and $|\eta| < 2.5$.

Table 5.2: Reconstruction requirements for leptons used in this analysis. ${}^{IP_{r\phi}}/_{\sigma_{IP_{r\phi}}}$ is the transverse impact parameter significance, where $IP_{r\phi}$ is the distance of closest approach of the particle track to the beam axis. $|IP_z|$ is the longitudinal impact parameter, which is the value of z of the point on the track that determines $IP_{r\phi}$.

	Electrons	Muons	τ_{Had}
p_T	> 10 GeV	> 10 GeV	> 25 GeV
$\lvert\eta\rvert$	< 2.47 and not in $[1.37, 1.52]$	< 2.5	< 2.5
$\frac{IP_{r\phi}}{\sigma_{IP_{r\phi}}}$	< 5	< 3	-
$\lvert IP_z\rvert$	< 0.5 mm	< 0.5 mm	< 1.5 mm

In addition to the electrons and muons, hadronically decaying tau leptons (τ_{Had}) [104] are reconstructed from jets formed using the anti-k_t algorithm [105] with a distance parameter $R = 0.4$ and three-dimensional clusters of calorimeter cells called topo-clusters [106]. Then, a multivariate discriminant is used to distinguish the τ_{Had} candidates from jets, based on the track collimation, further jet substructure, and kinematic information. Three working points corresponding to distinct tau identification efficiencies are defined. In this analysis, the τ_{Had} candidates must satisfy $p_T > 25$ GeV, $|\eta| < 2.5$ and pass the Medium identification working point.

The requirements used in the reconstruction of leptons in this analysis are represented in Table 5.2 .

5.2.2 Small-R Jets

Due to the QCD color confinement, the final state partons produced in the collisions will hadronize and form collimated sprays of hadrons called jets. In this analysis, the jets were reconstructed using the anti-k_t algorithm implemented in the FastJet package [107] using topo-clusters in the calorimeter as inputs. The anti-k_t algorithm involves a symmetric distance d_{ij} between pairs of objects i and j given by

$$d_{ij} = min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta R_{ij}^2}{R^2}, \tag{5.3}$$

where $p_{T,i}$ is the transverse momentum of the i-th object, ΔR_{ij} is the angular distance (Section 3.2) between the i-th and j-th object, and R is the radius parameter that governs the size of the cone. R was set to 0.4 and jets were required to have $p_T > 25$ GeV and $|\eta| < 2.5$. In addition, a jet vertex tagger (JVT) [108] algorithm was used to identify jets consistent with the primary vertex and reduce the effect of pileup.

5.2.3 Large-R Jets

The angular separation between two particles decaying from a symmetric two-body decay of a massive particle can be expressed by

$$\Delta R \approx \frac{2m}{p_T} \tag{5.4}$$

where m and p_t are the mass and transverse momentum of the parent particle. When the original particle has a high transverse momentum (boosted regime), the angular separation between its decay products will be small, and they can be contained in a single larger-radius jet referred as large-R jet. The large-R jet is used to reconstruct all the decay products of the original particle as a single object/jet. This approach is more efficient than reconstructing all of the products as individual jets. Several observables, named substructure variables, can be constructed in order to characterize the internal structure of the large-R jet and to identify the original particle that produced the jet.

Usually, only a few choices of R are used for all analyses since every jet

configuration (algorithm, radius and grooming parameters[2]) must be calibrated to correct the reconstructed jet four-momentum and account for features of the detector, the jet reconstruction algorithm and pile-up. One solution to this is to build the large jets out of the constituents of small-R (R = 0.4) jets instead of topo-clusters, using a jet reclustering technique [109]. This allows for a direct propagation of the calibrations and systematic uncertainties associated with the small-R jets, eliminating the need for any further calibration or uncertainty. Figure 5.2 compares a normal large-R jet with a reclustering one.

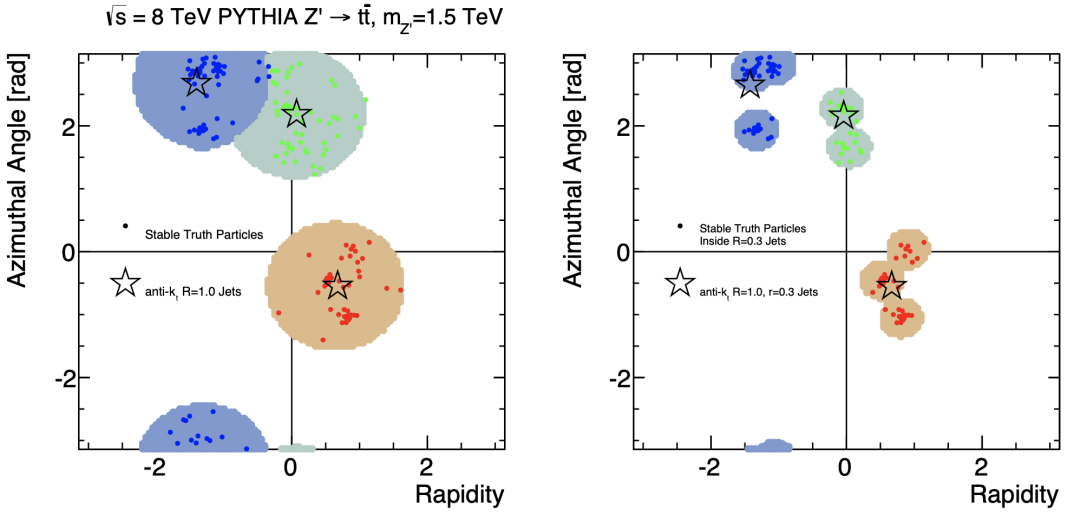


Figure 5.2: Example event which has been clustered using the anti-k_T $R_{large} = 1.0$ (left) and with anti-k_T $R_{large} = 1.0$ reclustered $R_{small} = 0.3$ anti-k_T jets (right) [109].

In this analysis, the anti-k_T $R = 1.0$ reclustered jets took as input the anti-k_T small-R jets with $R = 0.4$. These reclustered large-R jets were required to have a reconstructed invariant mass higher than 50 GeV, $p_T > 200$ GeV and at least two small-R jets inside. The resulting jets were, then, used to identify top quarks and Higgs bosons in signal events in the boosted regime and decaying into collimated hadronic final states.

[2]Due to the large area of large-R jets, these jets suffer from strong contamination of soft particles coming from pile-up and underling events. Grooming algorithms are used to eliminate these effects.

5.2.4 b-tagged Jets

Jets that originate from the hadronization of a b quark, referred to as b-jets, can be differentiated from the other jets by exploiting the special properties of b-hadrons, such as the long lifetime that originates a displaced secondary vertex. These jets are reconstructed via b-tagging algorithms that rely on multivariate techniques. In this analysis, the MV2c10 tagger was used and trained on simulated $t\bar{t}$ events to discriminate b-jets from light jets and c-labelled jets [110]. The jets were b-tagged at 60% and 70% efficiency WP (that corresponds to an average b-tagging efficiency of 60% and 70%, respectively) for the dilepton and single lepton resolved channels, and at the 85% efficiency WP for the single lepton boosted channel.

5.2.5 Overlap Removal

An overlap removal procedure is applied to the reconstructed leptons and small-R jets to prevent multiple objects being reconstructed from the same detector signals. The nearest jet within $\Delta R_y = \sqrt{(\Delta y)^2 + (\Delta\phi)^2} = 0.2$ of a selected electron is removed to avoid double-counting of electron energy deposits in the jet reconstruction. If the closest jet that survives the selection is within $\Delta R_y = 0.4$ of the electron, the electron is discarded.

5.2.6 Missing Transverse Energy

Since the protons collide along the beam axis, the total p_T of the particles after the collision must be zero to conserve energy-momentum. The neutrinos do not interact with any component of the detector and leave an imbalance in the measured transverse momentum. In the SM, this missing transverse energy (MTE) may be attributed to the neutrinos, and it is calculated as the magnitude of the negative vector sum of the p_T of all selected reconstructed objects [111].

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\mu} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,jets} + E_{x(y)}^{miss,soft}, \qquad (5.5)$$

where $E_{x(y)}^{miss}$ is the x (or y) component of the MTE. Each term corresponds to the negative sum of the energy of the objects e, μ, τ, γ, and jets. projected in the x (or y) axis. The last term, which is called soft term, accounts for objects not associated with any object passing the selection cuts. These can be ID tracks or calorimeter signals.

The magnitude of the MTE and the azimuthal angle ϕ^{miss} can be calculated by

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \tag{5.6}$$

$$\phi^{miss} = \arctan\left(\frac{E_y^{miss}}{E_x^{miss}}\right). \tag{5.7}$$

Due to conservation of the 4-momentum in the W boson decay, the neutrino reconstruction is dictated by the relation

$$m_W^2 = (p_{lep} + p_\nu)^2, \tag{5.8}$$

where, m_W is the W boson mass, p_{lep} is the lepton 4-momentum and p_ν is the neutrino 4-momentum, considering that the neutrino has zero mass and that the missing transverse momentum components are the transverse momentum components of the neutrino. In case two solutions for p_ν are found, the invariant mass of the reconstructed leptonic top candidate (discussed in section 5.2.7) is computed for both solutions, the p_ν value which gives the closest invariant mass to the top mass is selected.

5.2.7 Higgs and Top Candidates

In the resolved regime (Section 6.2), a boosted decision tree referred to as reconstruction BDT is used to match reconstructed jets to the decay products of top quarks and Higgs bosons. The W boson, top quark and Higgs boson candidates are built from combinations of jets and leptons. Simulated events are used to iterate over all allowed combinations. The reconstruction BDT is trained to distinguish between correct and incorrect jet assignments. In the training, invariant masses and additional kinematic variables that exploit the event topology are used as input.

In the single lepton resolved channel, the leptonically decaying W boson candidate is reconstructed from the lepton and neutrino four-momenta. The latter is constructed form the MTE as described in the previous section. The hadronically decaying W boson is build from a pair of jets. The top quark candidates are formed from one W boson candidate and one jet. The Higgs boson is reconstructed from the two b-jets. In the dilepton resolved channel, the top quarks are reconstructed from one b-jet and one lepton, and the neutrinos are not considered

in the reconstruction.

In the boosted regime, in order to achieve a better reconstruction matching of Higgs and top final objects, a multi-class deep neural network (DNN) was trained with $t\bar{t}H$ samples, using variables related to the reclustered large-R jets and its constituents, to identify the parent particle that generated the large-R jet. The DNN classify the jets into three categories: Higgs-boson, top quark and QCD jets. The network outputs probabilities for each of the three possible labels: P(H) for Higgs, P(T) top quarks and P(Q) for QCD jets. Figure 5.3 shows the probabilities for the Higgs and top categories. The observed efficiency for the DNN identification is 76% for the Higgs bosons and 67% for the top quarks.

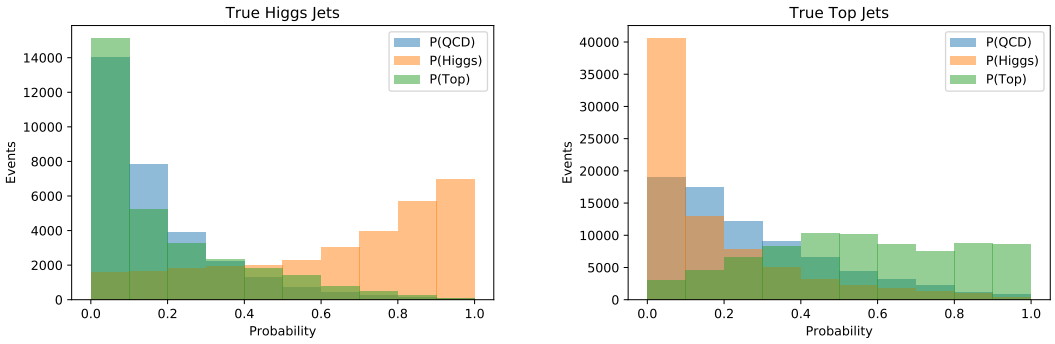


Figure 5.3: Output of the DNN for the Higgs and top categories, where P(Higgs) + P(Top) + P(QCD) = 1. A high discrimination power is shown.

The boosted Higgs bosons are reconstructed from large-R jets containing at least two constituent jets, among which exactly two are b-tagged at the 85% WP. The Higgs candidates must satisfy $p_T > 300$ GeV and $m_H^{rec} \in [100 - 140)$ GeV and $P(H) > 0.6$. If more than one boosted Higgs candidate is found, the candidate with the constituent jets invariant mass closest to the Higgs mass is selected.

If additional large-R jets are found with $p_T > 300$ GeV and $P(T) > 0.3$, they are considered as potentially coming from one top quark that decays hadronically and are used for the reconstruction of a hadronic top quark candidate. If more than one boosted top quark candidate is found, the one with the mass closest to the top quark mass is selected.

Following the identification of the boosted Higgs and the boosted hadronic top candidates, a small-R jet is needed for the reconstruction of the leptonic top candidate. The small-R jet, the reconstructed lepton and neutrino are used for the reconstruction of the leptonic top candidate, which is required to have an invariant

mass of $m_{t_{lep}}^{rec} \in [130-200)$ GeV. In addition, the jet cannot be overlapping the two jets utilized in the reconstruction of the Higgs and hadronic top candidates. In case more than one small-R jet fulfill these requirements, the one with the greatest transverse momentum is chosen. If no jet satisfying these requirements is found, the leptonic top is defined only by the reconstructed lepton and neutrino.

If no hadronic top candidate is found, the invariant mass of the reconstructed hadronic top and leptonic top are evaluated for all combinations of small-R jets not overlapping with the Higgs candidate. Both masses are required to be inside the interval $[70-195)$ GeV. The combination with minimum value of the sum $|m_{t_{had}}^{rec} - 172.5| + |m_{t_{lep}}^{rec} - 172.5|$ is considered. If no combination is found, the hadronic top is reconstructed from the three jets with the highest transverse momentum and which are not overlapping with the Higgs candidate, whereas the leptonic top is reconstructed only from the lepton and the neutrino.

5.3 Validating Substructure Variables with Atlfast-II

In a high-energy physics analysis, a realistic simulation of the detector response, with an accurate description of the interactions between the particles and the detector components, is essential. However, due to the large number of events demanded by the rapidly increasing LHC luminosity, a full detector simulation with small statistical uncertainty on MC samples can be challenging and extremely time-consuming. In some cases, the use of a simplified and fast approach may be a satisfactory alternative. A study [112], that might benefit future analysis, was performed in this thesis to compare the fast detector simulation Atlfast-II [113] with the ATLAS full simulation, and verify the modelling of substructure variables of large-R jets using $t\bar{t}H$ and $t\bar{t}$ MC samples.

Most of the CPU time required for fully simulating each event is spent in the calorimeter systems. The Atlfast-II simulation applies the FastCaloSim package [81, 114], which reduces the calorimeter simulation time to a few seconds per event, using a parametrized simulation of the particle energy response and the energy distribution in the ATLAS calorimeter. For the inner detector and the muon system, Atlfast-II uses the Geant4 simulation.

The studied jet substructure variables include the n-Subjettiness (τ_1^{wta}, τ_2^{wta}, τ_3^{wta}) [115, 116], ratios of n-Subjettiness (τ_{32}^{wta}, τ_{21}^{wta}), energy correlation functions (ECF_1, ECF_2, ECF_3) [117], energy correlations ratios ($D2$, $C2$, $E3$) [118], the

K_t distance between subjets during merging ($Split_{12}$, $Split_{23}$) [119] and others. In addition to these variables, several kinematic observables were also studied, such as the transverse momentum (p_{T}), the pseudorapidity (η), the azimuthal angle (ϕ), the energy (e) and the invariant mass (m) of the leading and subleading large-R jets.

Considerable differences were found in the shapes of the distributions of these observables. These differences are larger for the substructure variables than for the kinematic variables. Figure 5.4 and 5.5 show examples of the mismodelling of some of these observables in $t\bar{t}H$ and $t\bar{t}$ events. Although the mismodelling appears to be smaller for the samples of $t\bar{t}$, it is still significant. A full simulation may be necessary for a proper modelling of $t\bar{t}H$ and $t\bar{t}$ events in the boosted regime.

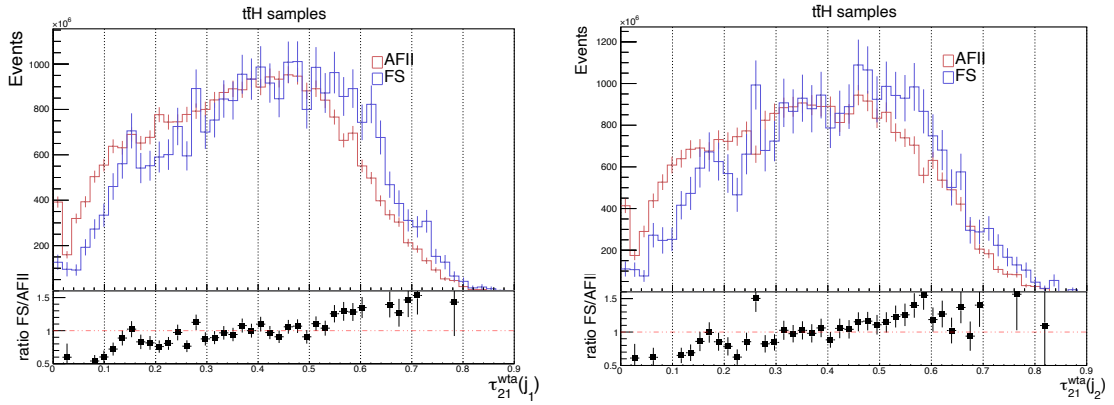


Figure 5.4: τ_{21}^{wta} distributions for the leading (left) and subleading (right) large-R jets using $t\bar{t}H$ samples generated with AF-II and FS detector simulations [112].

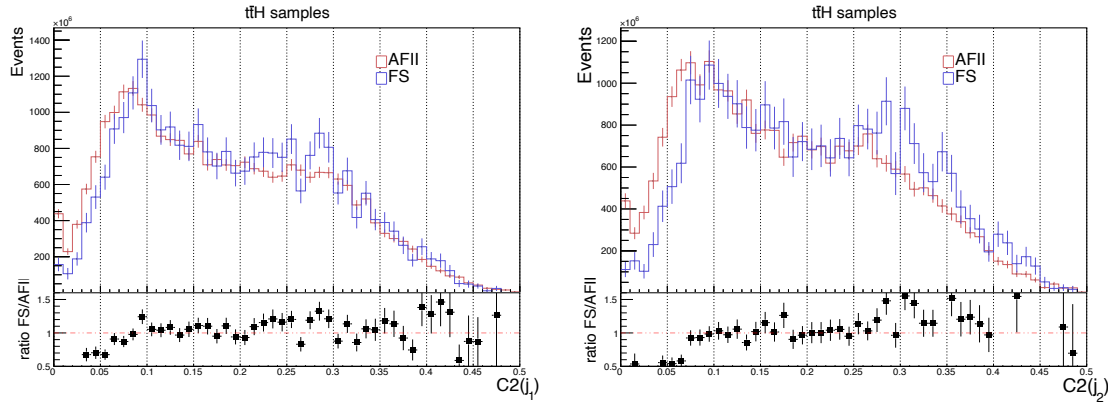


Figure 5.5: $C2$ distributions for the leading (left) and subleading (right) large-R jets using $t\bar{t}H$ samples generated with AF-II and FS detector simulations [112].

6

Analysis Strategy

In the long run, this analysis is intended to measure the CP properties of the top-Higgs coupling, relying on the differential cross-sections with respect to sensitive observables as well as the $t\bar{t}H$ inclusive cross-section. The first step in this process is to demonstrate sensitivity to reject the pure CP-odd hypothesis. Both the signal strength and the signal shape in CP-discriminating distributions are constrained simultaneously. Events are classified according to the number of jets and b-tagged jets at specific working points (WPs). Several multivariate techniques were used to improve the analysis sensitivity. One boosted decision tree (BDT) was trained to separate signal from background processes and a second BDT to combine several CP-sensitive variables into a single classifier. The analysis is performed using the signal ($t\bar{t}H$ and tH) and background MC samples described in Section 5.1.1.

In the first section of this chapter, the trigger and selection applied to the events are presented, followed by a detailed description of the classification of events into signal and control regions in Sections 6.2 and 6.3, where the classification BDT is described. Section 6.4 presents the CP classifier. Section 6.5 describes the variables used in the fit in each region of the analysis. Section 6.6 explains the parameterization of the $t\bar{t}H$ and tH yields as a function of α and k_t'. The systematic uncertainties are presented in section 6.7 and the fit is described in Section 6.8. Finally, Section 6.9 compares two different analysis strategies and explains why the simplified strategy was chosen.

As explained in Section 4.2, this analysis is based on two parameters of interest (POI), the CP mixing angle α and the coupling modifier k_t'. A detailed description of all the statistical analysis techniques applied in this work to measure these two parameters is presented in Appendix A.

6.1 Event Selection

Before the analysis selection, events were recorded using unprescaled single-lepton triggers with different p_T thresholds. This means that only events with at least one electron or one muon are selected. All the trigger menus and the p_T thresholds used are summarized in Table 6.1, for each of the four years of data-taking. The recorded events were required to fire the low lepton p_T threshold triggers (20, 24 and 26 GeV) with additional isolation requirements or higher threshold triggers (50, 60, 120 and 140 GeV) with a less restrictive identification requirement (Section 5.2.1).

Table 6.1: Single-lepton trigger menus used in the single lepton and dilepton channels in 2015 and 2016 to 2018, where p_T^e and p_T^μ represent the momentum threshold for the electron and the muon, respectively.

Years	Single-lepton triggers	Threshold (GeV)
2015	HLT_e24_lhmedium_L1EM20VH	$p_T^e = 24$
	HLT_e60_lhmedium	$p_T^e = 60$
	HLT_e120_lhloose	$p_T^e = 120$
	HLT_mu20_iloose_L1MU15	$p_T^\mu = 20$
	HLT_mu50	$p_T^\mu = 50$
2016–2018	HLT_e26_lhtight_nod0_ivarloose	$p_T^e = 26$
	HLT_e60_lhmedium_nod0	$p_T^e = 60$
	HLT_e140_lhloose_nod0	$p_T^e = 140$
	HLT_mu26_ivarmedium	$p_T^\mu = 26$
	HLT_mu50	$p_T^\mu = 50$

Then, events were divided according to the number of leptons accepted by the trigger algorithm, into the single lepton (exactly one lepton) and dilepton channels (exactly two leptons). In both channels, at least one reconstructed lepton, with $p_t > 27$ GeV, is required to match a lepton reconstructed by the trigger system, within the angular distance $\Delta R > 0.15$ and with the same flavor.

In the dilepton channel, the two leptons must have opposite electric charge, and if the leptons are either two electrons or two muons, their invariant mass is required to be greater than 15 GeV (to reject low-energy hadronic resonances, such as the Υ meson) and outside of the Z-boson mass range $83 - 99$ GeV. In addition, at least three jets are required, at least three of which must be b-tagged using the 70% efficiency WP (Section 5.2.4).

In the single lepton channel, at least four jets and one Higgs candidate are required for the boosted category or at least five jets for the resolved category. In

the boosted category, the Higgs candidate must have $p_T > 300$ GeV and a DNN score of $P(H) > 0.6$ (Section 5.2.7). While, in the resolved category, four of the five jets must be b-tagged using the 70% efficiency WP, and only events which are not in the boosted category are selected.

6.2 Signal and Control Regions

The selected resolved events, as described in the previous section, were classified into regions, depending on the number of leptons, the number of jets, the number of b-tagged jets using the 60% and 70% WPs. The signal-enriched regions are denoted as *inclusive signal regions*, while the regions with lower signal purity are denoted as control regions. One signal region ($\mathrm{SR}_{\mathrm{inc}}^{\geq 4j,\geq 4b}$) and three control regions ($\mathrm{CR}_{\mathrm{high}}^{\geq 4j,3b}$, $\mathrm{CR}_{\mathrm{low}}^{\geq 4j,3b}$, $\mathrm{CR}_{\mathrm{high}}^{\geq 3j,3b}$) were defined for the dilepton channel. For the single lepton channel, one signal region ($\mathrm{SR}_{\mathrm{inc}}^{\geq 6j,\geq 4b}$) and two control regions ($\mathrm{CR}_{\mathrm{high}}^{\geq 5j,\geq 4b}$, $\mathrm{CR}_{\mathrm{low}}^{\geq 5j,\geq 4b}$) were defined, as well as one single *boosted signal region* ($\mathrm{SR}_{\mathrm{boosted}}$) for events in the boosted category. Table 6.2 summarizes the various regions in which the events were divided.

Table 6.2: Summary of the event classification for each region of the analysis.

Region	leptons	jets	b-tag		Higgs candidates
			70%	60%	
$\mathrm{SR}_{\mathrm{inc}}^{\geq 4j,\geq 4b}$	$= 2$	≥ 4	≥ 4	-	-
$\mathrm{CR}_{\mathrm{high}}^{\geq 4j,3b}$			$= 3$	$= 3$	-
$\mathrm{CR}_{\mathrm{low}}^{\geq 4j,3b}$				< 3	-
$\mathrm{CR}_{\mathrm{high}}^{\geq 3j,3b}$		$= 3$		$= 3$	-
$\mathrm{SR}_{\mathrm{inc}}^{\geq 6j,\geq 4b}$	$= 1$	≥ 6	≥ 4	-	-
$\mathrm{CR}_{\mathrm{high}}^{\geq 5j,\geq 4b}$		$= 5$		≥ 4	-
$\mathrm{CR}_{\mathrm{low}}^{\geq 5j,\geq 4b}$				< 4	-
$\mathrm{SR}_{\mathrm{boosted}}$		≥ 4	-	-	≥ 1

Regions with three b-tagged jets at the 70% WP in the dilepton lepton channel and regions with five jets in the single lepton channel determine the control regions, which are referred to as *high* or *low* control regions according to the signal to background ratio. The control regions were used in a combined fit with the signal regions to provide constraints on systematic uncertainties and to better separate signal from background.

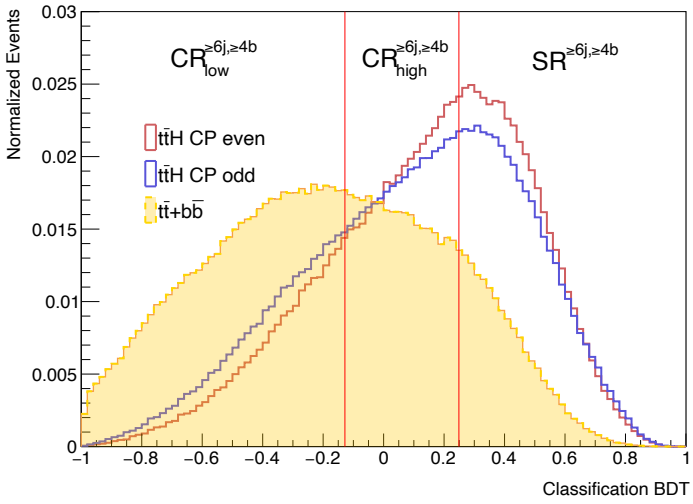


Figure 6.1: Classification BDT output for samples of CP-even and CP-odd $t\bar{t}H$ production as well as for the main background in the $\mathrm{SR}_{\mathrm{inc}}^{\geq 6j,\geq 4b}$. The vertical lines represent the boundaries of the regions resulting from the division based on the signal to background ratio.

6.3 Signal Classification

The regions with events containing at least four jets in the dilepton channel or at least six jets in the l+jets channel, and four b-tagged jets at the 70% WP were referred to as *inclusive signal regions* since these regions exhibit a higher percentage of $t\bar{t}H$ when compared to the control regions. However, these regions still have a considerable percentage of background, especially $t\bar{t}b\bar{b}$ events. In view of this, a BDT is trained to distinguish between the CP-even[1] $t\bar{t}H$ signal from the background and further split these regions according to its output. This BDT is referred to as *classification* BDT. The resulting regions, which are summarized in Table 6.3, are denoted with a subscript *high* or *low* depending on the signal-to-background ratio. Figure 6.1 shows the BDT output used for splitting the $\mathrm{SR}_{\mathrm{inc}}^{\geq 6j,\geq 4b}$ region.

Some input variables used in the training of the CP-BDT (discussed in the next section) require the reconstruction of the top quark pair. In the dilepton channel, some events will fail to be reconstructed due to the presence of two neutrinos in the final state. These events will be treated separately, and are classified in the $\mathrm{CR}_{\text{no-reco}}^{\geq 4j,\geq 4b}$ region.

[1]The dependence on the CP mixing angle is not significant and the background is rejected well, as seen from Figure 6.1.

Table 6.3: Summary of the inclusive regions separated into high and low signal and control regions after the classification of events based on the output of the classification BDT. The BDT selection range for each region is indicated in the table. The boosted inclusive region is not divided into other regions, but a selection is applied.

Inclusive region	Resulting regions	Classification BDT selection
$\mathrm{SR}_{\mathrm{inc}}^{\geq 4j,\geq 4b}$	$\mathrm{CR}_{\mathrm{no\text{-}reco}}^{\geq 4j,\geq 4b}$	-
	$\mathrm{CR}_{\mathrm{low}}^{\geq 4j,\geq 4b}$	BDT$\in [-1, -0.0862]$
	$\mathrm{SR}_{\mathrm{low}}^{\geq 4j,\geq 4b}$	BDT$\in [-0.0862, 0.1862]$
	$\mathrm{SR}_{\mathrm{high}}^{\geq 4j,\geq 4b}$	BDT$\in [0.1862, 1]$
$\mathrm{SR}_{\mathrm{inc}}^{\geq 6j,\geq 4b}$	$\mathrm{CR}_{\mathrm{low}}^{\geq 6j,\geq 4b}$	BDT$\in [-1, -0.128]$
	$\mathrm{CR}_{\mathrm{high}}^{\geq 6j,\geq 4b}$	BDT$\in [-0.128, 0.249]$
	$\mathrm{SR}^{\geq 6j,\geq 4b}$	BDT$\in [0.249, 1]$
$\mathrm{SR}_{\mathrm{boosted}}$	$\mathrm{SR}_{\mathrm{boosted}}$	BDT$\in [0.05, 1]$

In addition, the boosted region is also considered as an inclusive signal region, and a selection is applied to the output of the BDT to reject most of the $t\bar{t}$+light background, while keeping the sensitivity to signal. Figure 6.2 and 6.3 show the expected composition of the different background events and the signal to background ratios for each region of the l+jets and dilepton channel, respectively, estimated from simulated events.

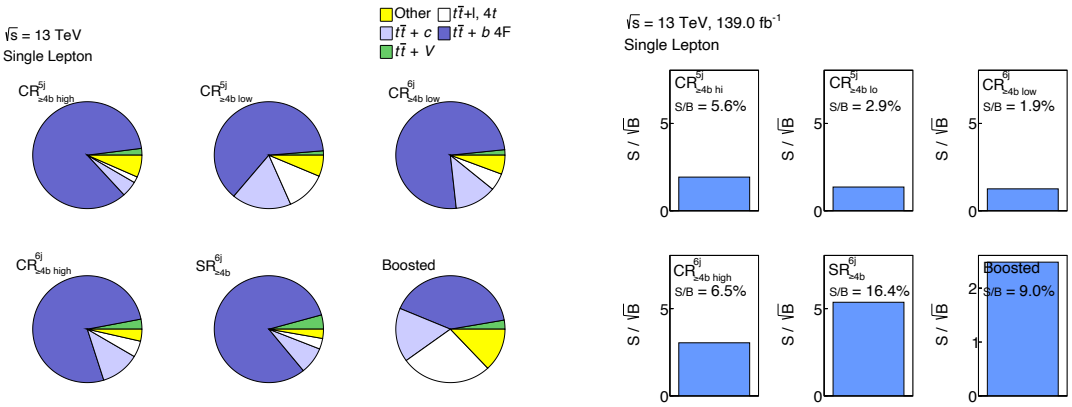


Figure 6.2: Fraction of the different backgrounds (left) and signal purity (right) for each region in the l+jets channel.

The toolkit for multivariate analysis (TMVA) [120] is used to train the classification BDT on statistically independent MC samples, using variables that exploit the differences between the kinematics of signal and background events, as well as

$b-$tagging variables. In addition, in order to achieve a better signal purity, the outputs of intermediate multivariate analysis (MVA) techniques are also used in the training, such as the output of the DNN training (Section 5.2.7) and techniques that explore the signal event topology and the properties of the possible jet-parton combinations when reconstructing the Higgs-boson and top quark candidates. In the dilepton channel, the BDT is trained against $t\bar{t} + \geq 1b$ backgrounds. In the l+jets resolved channels, the BDT is trained against all $t\bar{t}$ backgrounds. In the l+jets boosted channel, the BDT is trained against all the backgrounds.

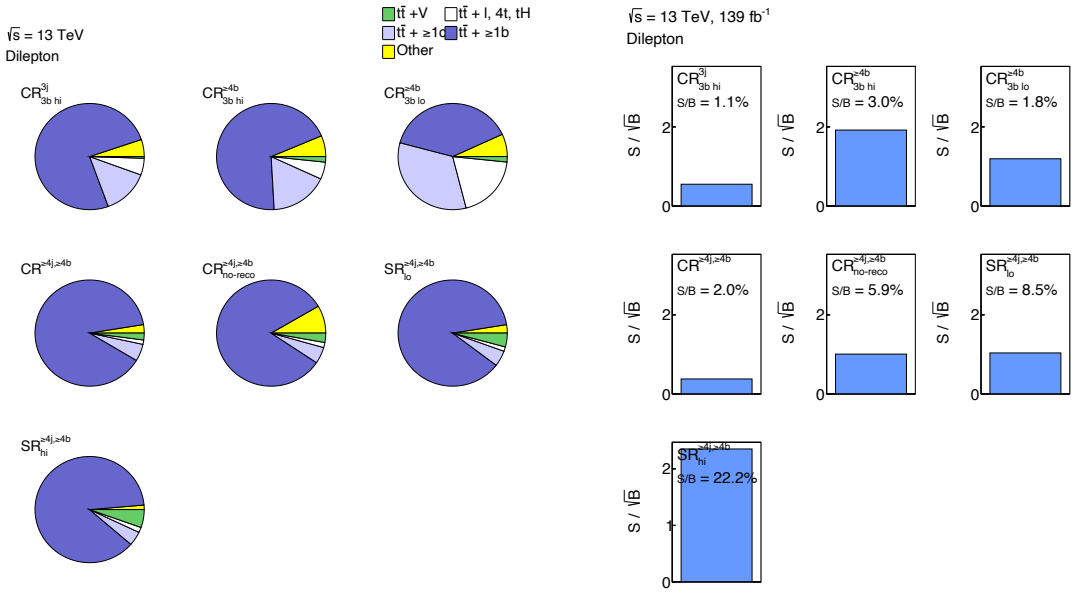


Figure 6.3: Fraction of the different backgrounds (left) and signal purity (right) for each region in the dilepton channel.

6.4 Classification of the CP Structure

After the classification of events according to the signal to background ratio, the successfully reconstructed $t\bar{t}H$ events in $\mathrm{SR}_{\mathrm{inc}}^{\geq 4j,\geq 4b}$ regions of the dilepton channel are used to train a second BDT, referred to as CP-BDT. This BDT is trained with TMVA to separate CP-even-like and CP-odd-like samples. In the training of the CP-BDT, CP-sensitive variables presented in Section 4.2.1 where used as input variables, including the lab-frame observable $b_4(t, \bar{t})$ and angular variables calculated in boosted reference frames. These variables were selected to optimize the discrimination between the pure CP-even and pure CP-odd samples,

and were ranked according to the effectiveness of their selection. A list of all these observables is presented in Table 6.4. Figure 6.4 shows the distributions of the two highest-ranked input variables and Figure 6.5 shows the distribution of the CP-BDT output.

CP-BDT Input Variables
$b_4(t,\bar{t})$
$\sin(\theta^{t\bar{t}H}_{t_A})\sin(\theta^{t_A}_{b_{2,H}})^{t_A-H}$
$\sin(\theta^{H}_{b_{t_A}})^{t_A-H}\sin(\theta^{H}_{W_{t_A}})^{t_A-H}$
$\sin(\theta^{t\bar{t}H}_{t_B})\sin(\theta^{t_B}_{b_{1,H}})^{t_B-H}$
$\sin(\theta^{t\bar{t}H}_{t_B})\cos(\theta^{t_B}_{b_{2,H}})^{t_B-H}$
$\sin(\theta^{t\bar{t}H}_{t_B})\sin(\theta^{t_A}_{b_{2,H}})^{t_A-H}$
$\sin(\theta^{t\bar{t}H}_{t_B})\sin(\theta^{t_A}_{b_{1,H}})^{t_A-H}$
$\sin(\theta^{t\bar{t}H}_{t_B})\sin(\theta^{H}_{b_{t_A}})^{t_A-H}$

Table 6.4: Input variables used in the training of the CP-BDT. The variables are arranged according to their relative importance in distinguishing between the two CP scenarios, starting with $b_4(t,\bar{t})$ as the most important. θ^x_y represents the angle between y and x systems. y is calculated in the x frame and x in the frame of its parent particles. A (and B) represents the closest (farthest) object from the Higgs boson, while 1 (2) indicates the (sub-)leading object. A superscript indicates the frame in which a variable is calculated.

The $b_4(t,\bar{t})$ observable, as defined in equation (4.5), is sensitive to angular differences between the top and anti-top distributions in the final state [121]. In CP-odd $t\bar{t}H$ production, the top quarks tend to be produced closer to the beam pipe and in opposite directions more often than in CP-even $t\bar{t}H$ production. For this reason, the $b_4(t,\bar{t})$ is the highest-ranked CP-sensitive variable in the CP-BDT for the dilepton channel. The angles calculated in boosted reference frames are also important since they probe the helicity information and spin correlations [72, 73]. These variables are constructed using the helicity formalism [122] and, by conservation of angular momentum, the top-quark spin is highly correlated to the angular distributions of its decays.

The CP-BDT is only used in $\mathrm{SR}^{\geq 4j,\geq 4b}_{\mathrm{inc}}$ regions of the dilepton channel. For the other regions, different CP discriminant variables are used as input to the fit. The $b_2^{t\bar{t}H}(t,\bar{t})$ variable, as defined in equation (4.4), is used as discriminant in the

l+jets channel. It exploits the smaller azimuthal separation between top quarks in CP-odd $t\bar{t}H$ production. To enhance its discriminating power, the $b_2^{t\bar{t}H}(t,\bar{t})$ observable is computed in the $t\bar{t}H$ rest frame [74, 121]. In the next section, all of the variables used in the fit are presented for each region.

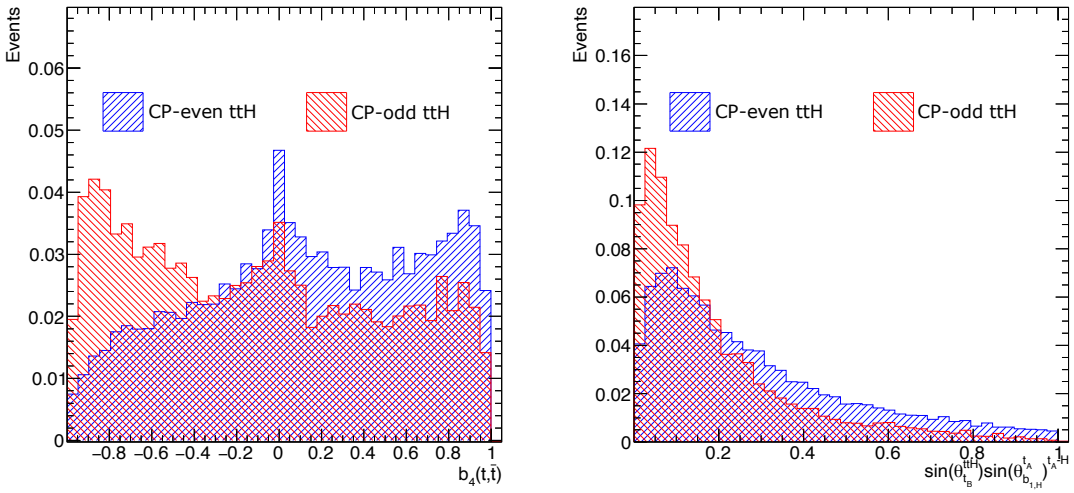


Figure 6.4: Distributions of $b_4(t,\bar{t})$ (left) and $\sin(\theta_{t_A}^{t\bar{t}H})\sin(\theta_{b_{2,H}}^{t_A})^{t_A-H}$ (right) for samples of CP-even $t\bar{t}H$ (blue) and CP-odd $t\bar{t}H$ (red) production.

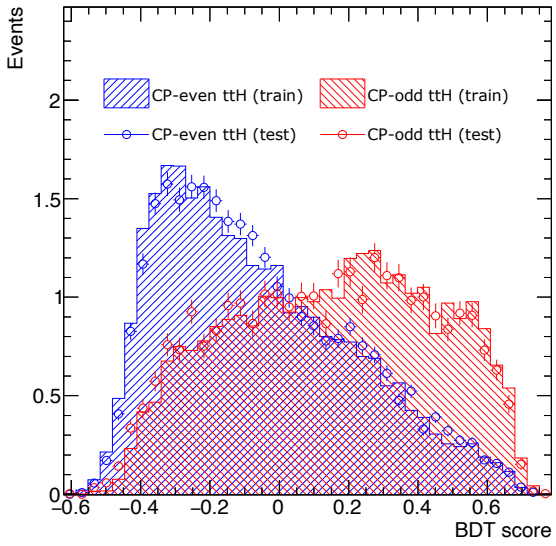


Figure 6.5: CP BDT output for samples of CP-even $t\bar{t}H$ (blue) and CP-odd $t\bar{t}H$ (red) production. The shaded areas represent the training samples, while the dots represent the test samples.

6.5 Fitted Variables

In order to increase the sensitivity of the analysis to the CP-odd $t\bar{t}H$ production and measure the CP mixing angle, CP discriminant variables are used as input to a profile likelihood fit as described in Section 6.8. The fit is performed to the signal and control regions, simultaneously, to constrain the backgrounds and obtain the coupling modifier k_t' and the CP mixing angle α. In control regions, it can be more interesting to use well-modelled variables that are sensitive to the presence of signal, instead of CP-sensitive variables, since these regions do not contain a high percentage of signal. All the fitted variables are presented in Table 6.5 for each region.

Table 6.5: Summary of all the fitted variables used in each regions of the analysis for the single lepton and dilepton channels.

Single Lepton Channel		Dilepton Channel	
Region	Fitted Variable	Region	Fitted Variable
$\mathrm{CR}^{5j,\geq 4b}_{\mathrm{low}}$ $\mathrm{CR}^{5j,\geq 4b}_{\mathrm{high}}$	$\Delta R^{avg}_{b\bar{b}}$	$\mathrm{CR}^{3j,3b}_{\mathrm{high}}$ $\mathrm{CR}^{\geq 4j,3b}_{\mathrm{low}}$ $\mathrm{CR}^{\geq 4j,3b}_{\mathrm{high}}$	Yield
$\mathrm{CR}^{\geq 6j,\geq 4b}_{\mathrm{low}}$ $\mathrm{CR}^{\geq 6j,\geq 4b}_{\mathrm{high}}$ $\mathrm{SR}^{\geq 6j,\geq 4b}$	$b_2^{t\bar{t}H}(t,\bar{t})$	$\mathrm{CR}^{\geq 4j,\geq 4b}$ $\mathrm{SR}^{\geq 4j,\geq 4b}_{\mathrm{low}}$ $\mathrm{SR}^{\geq 4j,\geq 4b}_{\mathrm{high}}$	CP-BDT
$\mathrm{SR}_{\mathrm{boosted}}$	Class BDT	$\mathrm{CR}^{\geq 4j,\geq 4b}_{\text{no-reco}}$	$\Delta\eta_{ll}$

The output of the CP-BDT is used in the dilepton regions resulting from splitting of $\mathrm{SR}^{\geq 4j,\geq 4b}_{\mathrm{inc}}$ regions, except for the $\mathrm{CR}^{\geq 4j,\geq 4b}_{\text{no-reco}}$ region. In this *no-reco* region, the difference between the pseudorapidity of the two leptons ($\Delta\eta_{ll}$) is used instead. $\Delta\eta_{ll}$ is significantly larger in CP-odd $t\bar{t}H$ production [123], and was shown to be the highest-ranked CP-sensitive variable that does not require top-quark pair reconstruction. A better significance is achieved when events that can not be reconstructed are excluded from the CP BDT and are fitted separately to the $\Delta\eta_{ll}$. Additionally, the CP BDT can be trained with only fully reconstructed events. For this reason, $\Delta\eta_{ll}$ is a preferred choice for $\mathrm{CR}^{\geq 4j,\geq 4b}_{\text{no-reco}}$.

In the single lepton $\mathrm{SR}^{\geq 6j,\geq 4b}_{\mathrm{inc}}$ regions, the $b_2^{t\bar{t}H}(t,\bar{t})$ variable is used in the fit. The original strategy was to fit the CP-BDT in these regions of the single lepton

channel as well, and not only for the dilepton channel. However, several studies were carried out with the intention to simplify the analysis. It was shown that fitting the $b_2^{t\bar{t}H}(t,\bar{t})$ in the single lepton channel gives almost the same sensitivity as using the CP-BDT in this channel. Section 6.9 further details why fitting the $b_2^{t\bar{t}H}(t,\bar{t})$ observable might be a better option for the single lepton channel.

In the $\mathrm{CR}_{\mathrm{high}}^{3j,3b}$, $\mathrm{CR}_{\mathrm{low}}^{\geq 4j,3b}$ and $\mathrm{CR}_{\mathrm{high}}^{\geq 4j,3b}$ regions of the dilepton channel, only the event yield is fitted, since the high statistics of these regions contribute to strongly constrain the normalization of $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ backgrounds. On the other hand, in the l+jets $\mathrm{CR}_{\mathrm{low}}^{5j,\geq 4b}$ and $\mathrm{CR}_{\mathrm{high}}^{5j,\geq 4b}$ regions, the fitted variable is the average angular distance $\Delta R_{b\bar{b}}^{avg}$ between any pair of $b-$jets within the first four $b-$jets. This variable is used to constrain the shape of the background.

Finally, the output of the classification BDT is used in the l+jets boosted channel as input to the fit. The amount of signal in the boosted region is expected to be enough to add sensitivity to the CP-odd coupling, due to the significantly larger fraction of high$-p_T$ Higgs bosons expected in CP-odd $t\bar{t}H$ production, as shown in Figure 6.6. For this reason, the classification BDT is chosen instead of a CP discriminant in the $\mathrm{SR}_{\mathrm{boosted}}$ region.

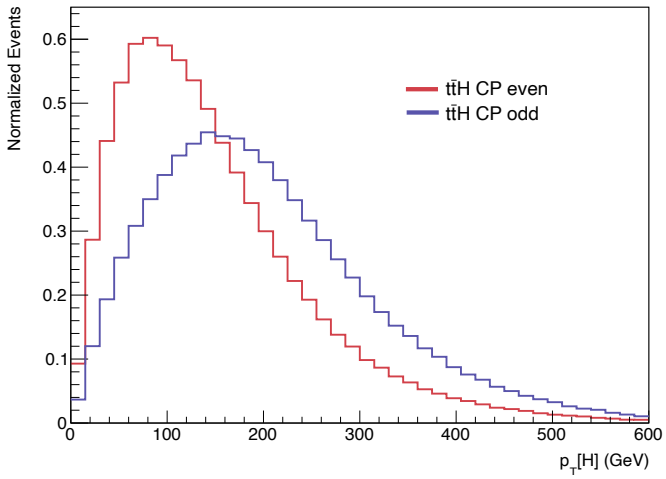


Figure 6.6: Higgs boson transverse momentum distributions at parton-level for the pure CP-even and pure CP-odd $t\bar{t}H$ production.

Figure 6.7 shows the two-dimensional distributions of the CP-BDT output as a function of the classification BDT output for the reconstructed events in the dilepton $\mathrm{SR}_{\mathrm{inc}}^{\geq 4j,\geq 4b}$ regions. The events are plotted separately for CP-even and CP-odd $t\bar{t}H$ production, as well as for the $t\bar{t}$+jets background. Similarly, Figure

6.8 shows the distributions of $b_2^{t\bar{t}H}(t,\bar{t})$ as a function of the classification BDT output for the l+jets $\mathrm{SR}_{\mathrm{inc}}^{\geq 6j,\geq 4b}$ regions. The region's boundaries are represented in the figure as vertical lines and the binning of each region as horizontal lines. In both cases, significant differences between the background, CP-odd signal, and CP-even signal can be observed. Several bins are sensitive to the presence of a CP-odd signal.

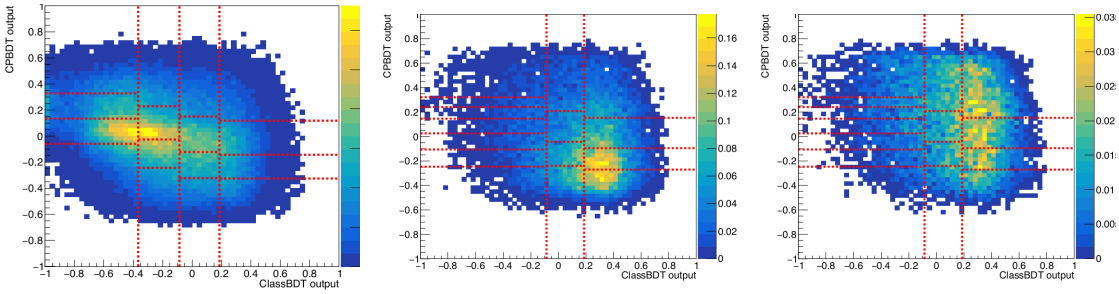


Figure 6.7: Two-dimensional distributions of the classification and CP BDTs, for $t\bar{t}$+jets (left), CP-even $t\bar{t}H$ (middle) and CP-odd $t\bar{t}H$ (right) in the fully reconstructed dilepton $\mathrm{SR}_{\mathrm{inc}}^{\geq 4j,\geq 4b}$ regions. The vertical dashed lines represent the regions boundaries, and the horizontal lines represent the bin edges in the CP BDT.

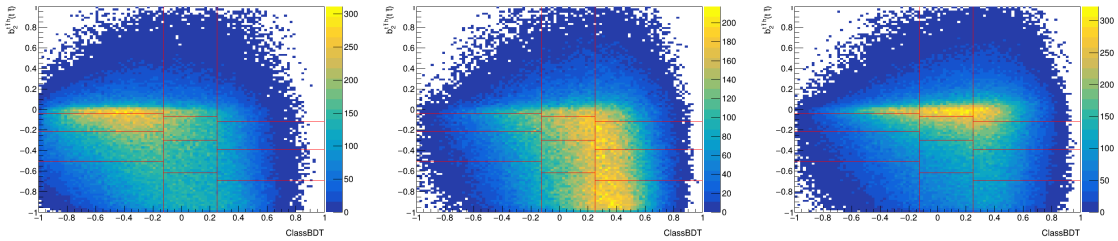


Figure 6.8: Two-dimensional distributions of the classification BDT and $b_2^{t\bar{t}H}(t,\bar{t})$, for $t\bar{t}+\geq 1b$ (left), CP-even $t\bar{t}H$ (middle) and CP-odd $t\bar{t}H$ (right) in the l+jets $\mathrm{SR}_{\mathrm{inc}}^{\geq 6j,\geq 4b}$ regions. The vertical dashed lines represent the regions boundaries, and the horizontal lines represent the bin edges in $b_2^{t\bar{t}H}(t,\bar{t})$.

6.6 Signal Yields Parameterization

Since the signal samples of Section 5.1.1 are available only for a few points in the (α, k_t') parameter space, the signal yields need to be parameterised as a function of the mixing angle α and the coupling strength k_t' to estimate these two parameters and avoid having to simulate a significant large amount of events. The expected yield from $t\bar{t}H$ process in a certain analysis bin in a given region is

parameterised as a weighted sum of CP-even and CP-odd samples:

$$y_{t\bar{t}H} = y_{\text{even}}(k'_t \cos\alpha)^2 + y_{\text{odd}}(k'_t \sin\alpha)^2, \tag{6.1}$$

where y_{even} and y_{odd} are the expected yields from the $t\bar{t}H$ samples for the pure CP-even and pure CP-odd scenario, respectively. Any term that accounts for a possible interference between CP-even and CP-odd contributions is neglected.

The parameterization of equation (6.1) was validated at truth-level in both channels using several different variables, and was shown to be a good approximation. The validation was performed by comparing the results of this parameterization to those obtained from the sample generated in the maximal mixing scenario (Section 5.1.1). Figure 6.9 shows examples of variables used in this validation.

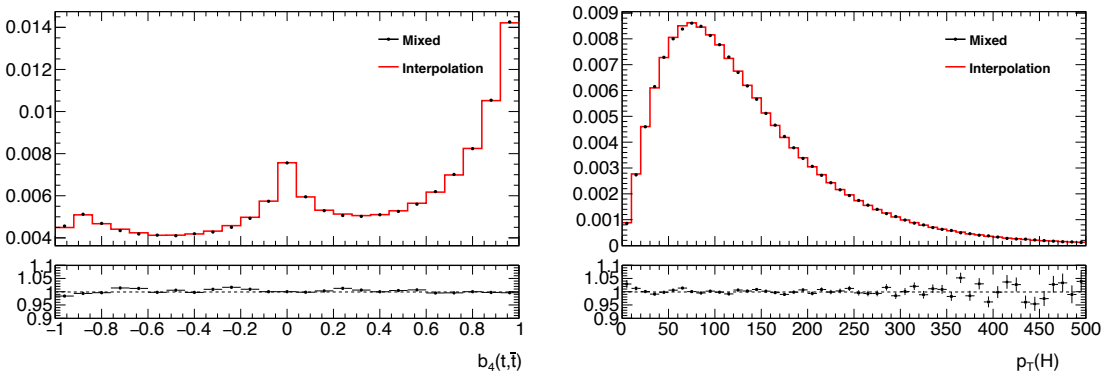


Figure 6.9: Comparison of $b_4(t,\bar{t})$ (left) and Higgs transverse momentum (right) distributions for the maximal mixing scenario ($\alpha = 45°$, $k'_t = 1$) using the parameterization from equation (6.1) (labelled "Interpolation") and using the samples generated for this scenario (labelled "Mixed") in the l+ jets channel.

For the tWH and $tHjb$ processes, the parameterization can be expressed as

$$\begin{aligned} y_{tH}/y_{SM} =& A(k'_t \cos\alpha)^2 + B(k'_t \sin\alpha)^2 + C(k'_t \cos\alpha) + \\ &+ D(k'_t \sin\alpha) + E(k'^2_t \cos\alpha \sin\alpha) + F, \end{aligned} \tag{6.2}$$

where y_{tH} is the expected yield for either tWH or $tHjb$ processes separately. In this case, the effects of destructive interference between the distinct amplitudes proportional to the t-H and H-W couplings in the tWH process need to be considered [124]. This interference results in a large dependence of the $tHjb$ and tWH cross-sections on the coupling parameters. The third and fourth terms account for the destructive interference between the t-H coupling and W-H coupling for

the CP-even and CP-odd scenarios, respectively. The E term accounts for the interference between CP-even and CP-odd t-H couplings. The F term represents the pure W-H contribution. The leading order Feynman diagrams for $tHjb$ and tWH production and the t-H and W-H couplings are represented in Figure 6.10.

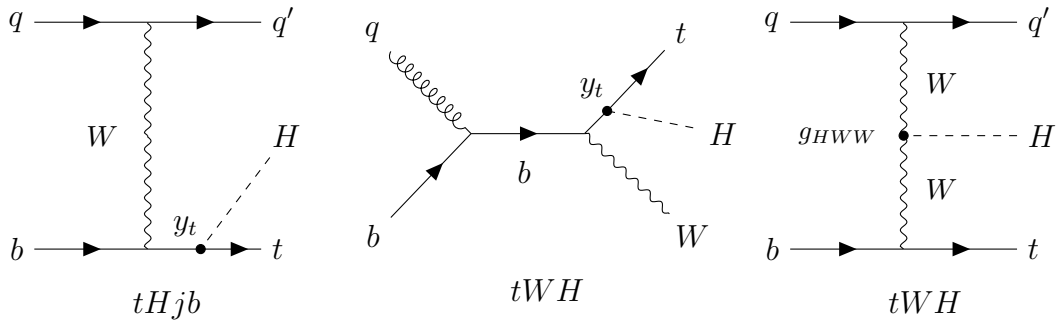


Figure 6.10: Leading order Feynman diagrams for $tHjb$ and tWH production, where the Higgs boson couples either to the top quark (left and center) or the W boson (right). The couplings are represented in the figure.

The coefficients for each bin are determined using a fit to the generated $tHjb$ and tWH signal MC samples (described in Section 5.1.1) at different values of the mixing angle α and coupling modifier k'_t. The best-fit values and uncertainties for the coefficients of equation (6.2) are represented in Figure 6.11 for the single lepton channel. The parameterizations of both samples were in good agreement with the yields obtained directly from the MC samples. An example of the comparison between the tH MC yields and the fit in one of the bins of the single lepton boosted region is shown in Figure 6.12.

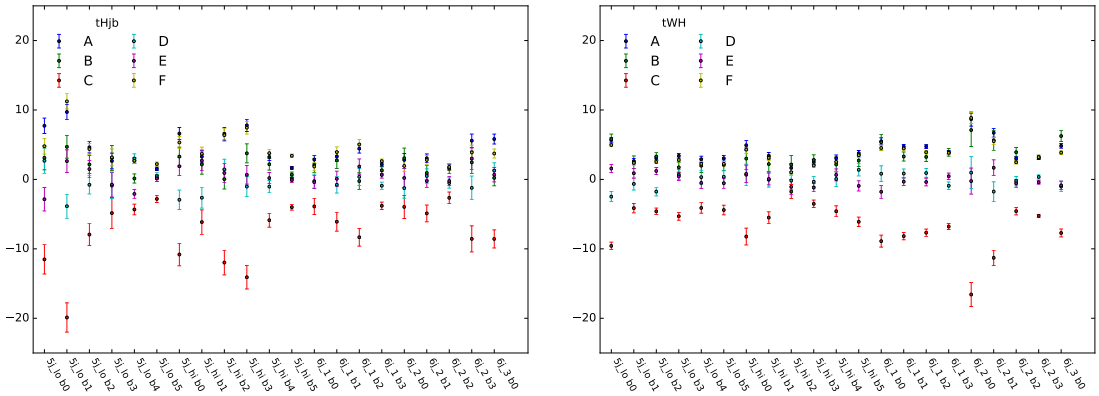


Figure 6.11: Best-fit values and uncertainties for the coefficients of the $tHjb$ (left) and tWH (right) parameterization in the resolved regions of the l+jets channel.

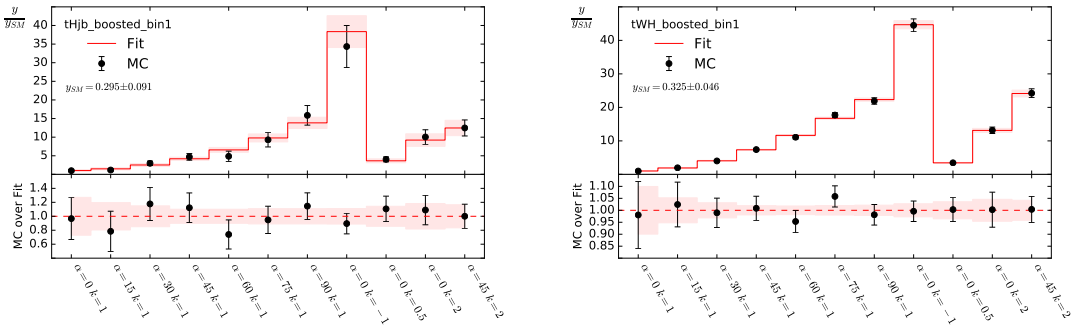


Figure 6.12: Parameterization of the $tHjb$ (left) and tWH (right) samples in the bin 1 of the l+jets boosted region for different values of α and k_t'. The MC yields are shown in black and the parameterization in red. The uncertainty in the tH samples is propagated to the fitted coefficients, then propagates to the parameterised yields.

6.7 Systematic Uncertainties

Different sources of systematic uncertainty affect the analysis sensitivity: experimental uncertainties, such as uncertainties related to the luminosity, pile-up modelling and the identification and reconstruction of the physics objects, and theory uncertainties related with the modelling of MC signal and background processes. As described in Appendix A, several nuisance parameters can be assigned to each source of systematic uncertainty for a more accurate treatment. The various uncertainties may affect only the normalization of signal and background samples, or also the shapes of the final discriminant distributions. All systematic uncertainties included in the analysis are listed in Table 6.6, together with the number of independent components of each systematic uncertainty and the type (if it affects the normalization only or also the shapes) of the systematic.

The experimental uncertainties are related to the luminosity, pileup modelling and reconstructed objects. An uncertainty of 1.7% is assigned to the value of the luminosity. Uncertainties associated with reconstructed leptons arise from the trigger, reconstruction, identification, isolation and resolution of the measurement of the lepton momentum. Uncertainties associated with jets arise from the efficiency of pileup rejection (jet vertex tagger), as well as jet energy scale and resolution. The uncertainties associated with leptons and jet energy scales are propagated to the missing transverse energy, and additional uncertainties in the scale and resolution of soft terms are considered. Finally, uncertainties associated

Table 6.6: Systematic uncertainties considered in the analysis. The type of the systematic and the number of components are indicated in the table. "N" represents that the uncertainty is taken as normalization-only for all processes and channels affected, whereas "SN" represents that the uncertainty is taken on both the shapes and the normalization.

Systematic uncertainty	Type	Components
Experimental Uncertainties		
Luminosity	N	1
Pileup modelling	SN	1
Physics Objects		
Electrons	SN	7
Muons	SN	15
Jet energy scale	SN	31
Jet energy resolution	SN	9
Jet vertex tagger	SN	1
$E_{\mathrm{T}}^{\mathrm{miss}}$	SN	3
***b*-tagging**		
Efficiency	SN	45
Mis-tag rate (c)	SN	20
Mis-tag rate (light)	SN	20
Signal and Background Modelling		
Signal		
H branching fractions	N	3
$t\bar{t}H$ cross-section	N	2
$t\bar{t}H$ modelling	SN	4
tH modelling	SN	4
$t\bar{t}$ Background		
$t\bar{t}$ cross-section	N	1
$t\bar{t}+\geq 1c$ normalization	N	1
$t\bar{t}+\geq 1b$ normalization	N (free floating)	1
$t\bar{t}+$ light modelling	SN	4
$t\bar{t}+\geq 1c$ modelling	SN	4
$t\bar{t}+\geq 1b$ modelling	SN	6
Other Backgrounds		
$t\bar{t}W$ cross-section	N	2
$t\bar{t}Z$ cross-section	N	2
$t\bar{t}W$ modelling	SN	1
$t\bar{t}Z$ modelling	SN	1
Single top cross-section	N	3
Single top modelling	SN	7
W+jets normalization	N	3
Z+jets normalization	N	3
Diboson normalization	N	1
$4t$ cross-section	N	1
tZq and tWZ cross-sections	N	3

with b-tagging arise from b-tagging efficiency of jets and mis-tagging efficiency of c-jets and light-jets.

The signal and background modelling uncertainties are theoretical uncertainties and arise from theoretical branching fractions and cross-sections of simulated MC samples, as well as from other aspects of the modelling to which the analysis is sensitive that are estimated by comparing predictions of different MC generators. While cross-section and branching fraction uncertainties affect only the normalization of distribution, the latter can also affect the shapes. These uncertainties can arise from the choice of NLO MC generator used for the simulation of the events, initial and final state QCD radiation, underlying event, parton distribution functions, as well as from the choice of parton shower and hadronization models.

In Section 7.1.1 the effect of the different sources of systematic uncertainties is assessed. Uncertainties on the background modelling have the strongest impact on the analysis presented in this thesis.

6.8 Measurement of the CP Mixing Angle

To extract the value of the CP mixing angle α, the distributions of the discriminant variables (summarized in Table 6.5) from each region of the analysis are combined into a profile likelihood fit[2]. A binned likelihood function $L(\alpha, k_t', \vec{\theta})$, that depends on the CP mixing angle α and the coupling modifier k_t', as well as on the set of nuisance parameters $\vec{\theta}$ that encode the effects of systematic uncertainties, is constructed as a product of Poisson probability terms over all bins of the analysis.

A scan is performed over all the allowed parameter space and the log-likelihood function is minimized. The measured best-fit values correspond to the minimum of the log-likelihood. Figure 6.13 shows an example of a scan performed to the log-likelihood function.

In order to define the confidence intervals and hypothesis exclusion a test statistic $q_{\alpha,k_t'}$ is built as:

$$q_{\alpha,k_t'} = -2\ln\frac{L(\alpha, k_t', \hat{\hat{\vec{\theta}}}(\alpha, k_t'))}{L(\hat{\alpha}, \hat{k}_t', \hat{\hat{\vec{\theta}}})}. \tag{6.3}$$

[2]For a more detailed discussion on statistical analysis techniques and the maximum likelihood estimator see Appendix A.

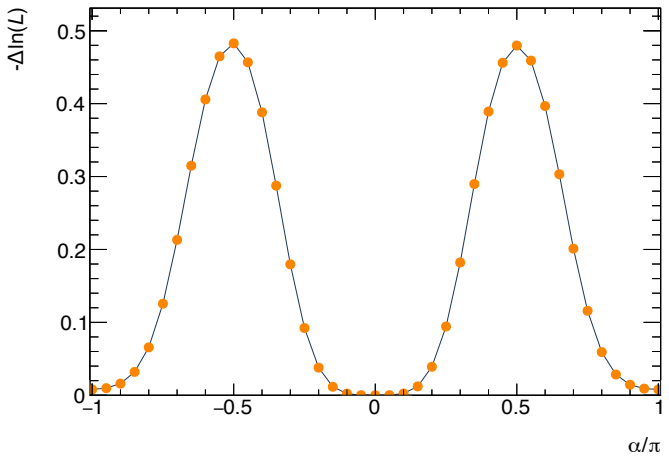


Figure 6.13: Example of one-dimensional log-likelihood scan on α/π for the single lepton channel.

where, $\hat{\alpha}, \hat{k}'_t$ and $\hat{\vec{\theta}}$ are the measured best-fit values, while $\hat{\hat{\vec{\theta}}}(\alpha, k'_t)$ is the values of θ that minimize the log-likelihood for a specific value of α and k'_t. As explained in Appendix A, q_{α,k'_t} can be asymptotically approximated to a χ^2 distribution. A certain point in the (α, k'_t) plane can be excluded with a confidence level equal to the cumulative χ^2 distribution up to q_{α,k'_t}, and the exclusion contours can be built.

The number of expected events in a certain bin depends on α, k'_t and the nuisance parameters. The profile likelihood fit is performed simultaneously to all signal and control regions as the expectations for signal and background are adjusted according to the corresponding systematic uncertainties, and their fitted values correspond to the ones that best fit the data. This procedure takes advantage of the highly populated background-dominated regions included in the fit to constrain the nuisance parameters corresponding to each systematic uncertainty, resulting in a simultaneous estimate of the POI and the nuisance parameters.

The $t\bar{t} + \geq 1b$ normalization factor is included in the likelihood function as a free parameter. All the other nuisance parameters are implemented with Gaussian or log-normal priors.

6.9 Analysis Simplification Strategy

In the first implementation of the analysis strategy, referred to as *original strategy*, the CP-BDT was also trained in the single lepton channel and fitted in the $\mathrm{SR}_{\mathrm{inc}}^{\geq 6j,\geq 4b}$ regions of this channel. However, several simplified strategies were

tested, and the CP-BDT was replaced with the highest-ranked variable used in the training of the dilepton channel, which was the $b_2^{t\bar{t}H}(t,\bar{t})$ observable. This is the current strategy, which was used during the rest of this work and is summarized in Table 6.5, and will be referred to as b_2 *simplification* in this section.

The Asimov fit presented in Section 7.1, which uses a dataset constructed from the SM expected values of signal and background in each bin, was performed for the current $b2$ strategy and the original strategy. Figure 6.14 shows the one-dimensional log-likelihood scan on α/π for the single lepton channel using both strategies. It can be seen that fitting the $b_2^{t\bar{t}H}(t,\bar{t})$ in the $\mathrm{SR}_{\mathrm{inc}}^{\geq 6j,\geq 4b}$ single lepton regions gives almost the same sensitivity as using the CP-BDT in this channel. A loss of only 0.01σ is observed. In view of this, the $b_2^{t\bar{t}H}(t,\bar{t})$ observable was preferred in this channel to reduce one layer of complexity of the analysis.

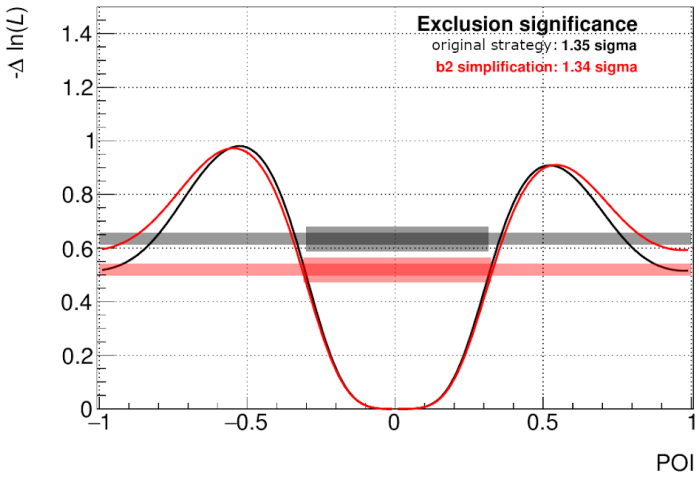


Figure 6.14: One-dimensional log-likelihood scan on α/π for the single lepton channel using pure CP-odd Asimov data for the baseline strategy and the current $b_2^{t\bar{t}H}(t,\bar{t})$ strategy. Note that the expected exclusion significance presented this figure is different from the value obtained in Chapter 6.2. This is due to the fact that one more nuisance parameter was introduced in the fit after this simplification study.

In addition, a better background modelling is achieved, since most of the important systematic uncertainties (Figure 7.4) are more constrained after the simplification. Figure 6.15 shows the best-fit values of the background modeling nuisance parameters obtained after the background only fits (discussed in Section 7.1.2) to data for the two strategies. It can be seen that some of the nuisance parameters have their pulls and associated uncertainties reduced. A larger pull

is observed for the $t\bar{t} + \geq 1b$ 4F radiation (initial state radiation). However, the $t\bar{t} + \geq 1b$ 4F NLO generator, parton shower and 4FS vs 5FS are significantly more constrained. Overall, fitting the $b_2^{t\bar{t}H}(t,\bar{t})$ in the single lepton channel seems to be a good alternative to the CP-BDT without a significant loss of sensitivity.

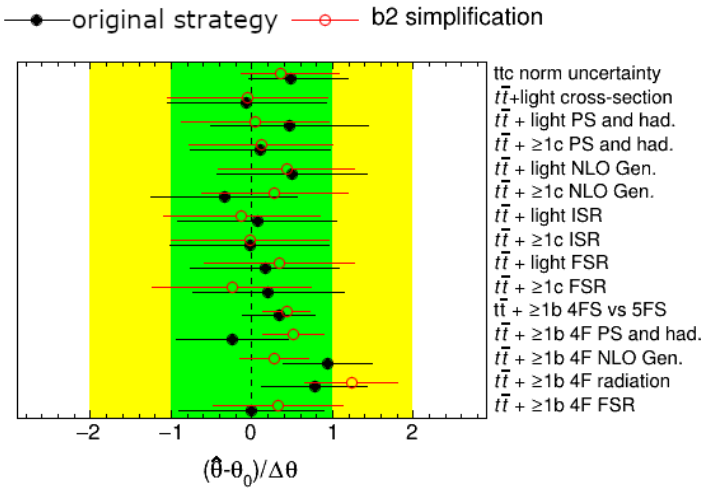


Figure 6.15: Comparison of background modelling nuisance parameters from background only fits to the single lepton channel between the baseline strategy and the current $b_2^{t\bar{t}H}(t,\bar{t})$ strategy. The green (yellow) area represent the $\pm 1(2)\sigma$ intervals on the post-fit systematic uncertainty. The points and horizontal bars represent the pulls and constraints in units of standard deviation, respectively.

7

Results

The main results of this work are described in this chapter. The expected limit obtained for the baseline analysis of the CP mixing angle and the coupling modifier are presented in Section 7.1. The combined CP-odd exclusion significance and the exclusion contours are also given. The different sources of systematic uncertainties are ranked according to their impact on the CP mixing angle in Section 7.1.1 and the efficacy of the background model is assessed in Section 7.1.2. In the second part of this chapter, the projections of the search from the LHC Run 2 to the HL-LHC are explored. Several luminosities are considered for different scenarios of systematic uncertainties. The pure CP-odd exclusion significance is taken as benchmark of the analysis sensitivity and is represented as a function of the luminosity for each of these scenarios in Section 7.2.

7.1 Expected Limit for the CP Mixing Angle

In order to estimate the expected analysis sensitivity and background constraining power, the Asimov dataset was used. This dataset is constructed from the exact expected values of signal and background, as predicted by the nominal signal and background MC samples. It is used to validate the fit model and the performance of the analysis before fitting the real data. Ideally, the best-fit values obtained for the POI when using an Asimov dataset are, by definition, their nominal values.

For a pure CP-even Asimov dataset, the values obtained in a fit combining both the single lepton and the dilepton channels for the mixing angle α and coupling modifier k_t' are

$$\alpha/\pi = 0.00^{+0.29}_{-0.27} \quad \text{and} \quad k_t' = 1.00^{+0.19}_{-0.22}, \tag{7.1}$$

which corresponds to an expected uncertainty interval on α of $[-0.27\pi, 0.29\pi]$ (or $[-48.5°, 52.2°]$). An expected pure CP-odd exclusion significance of 1.60σ is obtained for the combined fit. For the fit performed in the single lepton and the dilepton channels separately, pure CP-odd exclusion significances of 1.29σ and 0.98σ were obtained, respectively. Figure 7.1 shows the one-dimensional log-likelihood scan of α/π for both channels and their combination.

The expected two-dimensional CP-odd exclusion contours using the $(k_t, \tilde{k}_t)$ parameterisation from equation (4.3) is shown in Figure 7.3. This figure shows the pure CP-odd scenario outside the 1σ contour but inside the 2σ contour, and the inverted coupling scenario inside the 1σ contour. Note that the figure shows the best-fit point placed at (0.98,0.20) instead of (1.0,0.0) as expected from an Asimov fit. This is due to a numerical error during the likelihood scan in this parameterization, the scan does not have enough granularity to find the exact minimum.

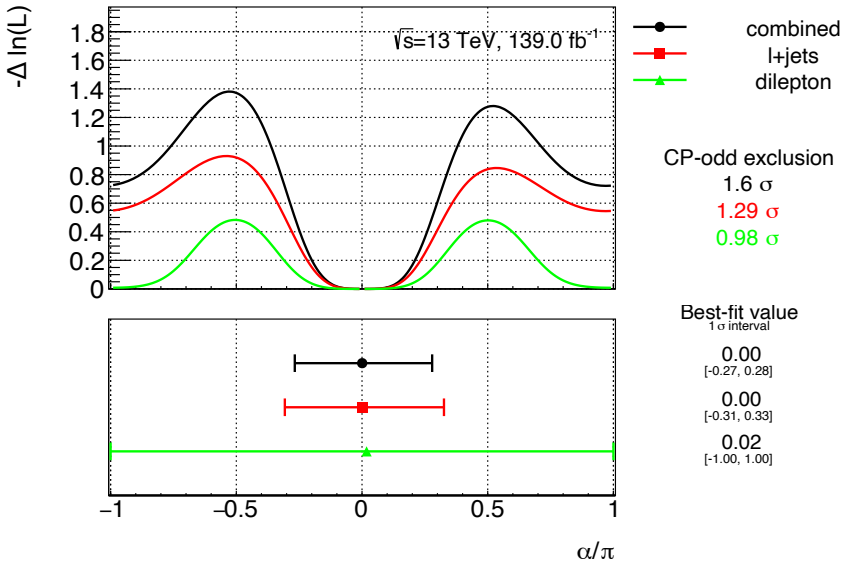


Figure 7.1: One-dimensional log-likelihood scan for the l+jets, dilepton and combined channels using pure CP-even Asimov data. The best-fit values for α and its uncertainty is represented for each channel, together with the expected CP-odd exclusions.

Similarly, the pure CP-odd case has also been studied. For this Asimov dataset, the combined best-fit values for α and k_t' are

$$\alpha/\pi = 0.50^{+0.25}_{+0.84} \cup^{-0.84}_{-0.24} \quad \text{and} \quad k_t' = 1.00^{+0.22}_{-0.29}, \tag{7.2}$$

which corresponds to an expected exclusion significance of 1.88σ for a pure CP-even signal. For the fit performed in the single lepton and the dilepton channels separately, CP-even exclusion significances of 1.68σ and 0.70σ were obtained, respectively. The one-dimensional log-likelihood scans for all the channels and the combination are shown in Figure 7.2. Figure 7.3 also shows the expected two-dimensional CP-even exclusion contours using the $(k_t, \tilde{k}_t)$ parameterisation.

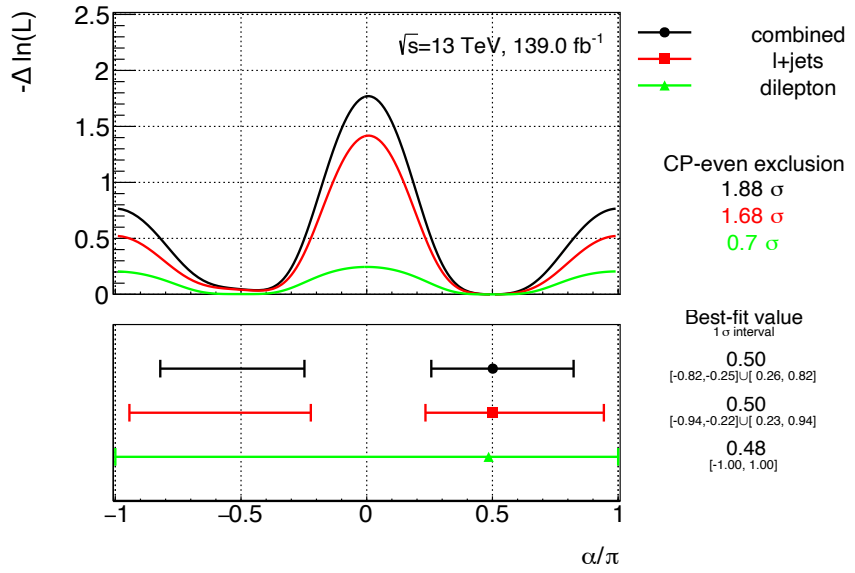


Figure 7.2: One-dimensional log-likelihood scan for the l+jets, dilepton and combined channels using pure CP-odd Asimov data. The best-fit values for α and its uncertainty is represented for each channel, together with the expected CP-even exclusions.

It can be seen that the log-likelihood scans are asymmetric. For example, the $\alpha/\pi = 0.5$ minimum is preferred over $\alpha/\pi = -0.5$ in the scan of Figure 7.2. This asymmetry is due to the presence of the tH signal. As seen from equation (6.2), the tH yields and, consequently, the analysis are sensitive to the sign of the CP mixing angle α.

The ATLAS analysis based on the $t\bar{t}H$ ($H \to \gamma\gamma$) channel, described in Section 4.2.2, excluded the pure CP-odd scenario at 3.9σ with stronger constraints to the allowed parameter space. As expected, the study presented in this work, using the $H \to b\bar{b}$ decay channel, seems to be less sensitive to the t-H coupling parameters, since this channel has several experimental challenges that the $H \to \gamma\gamma$ channel does not. For example, the presence of an overwhelming multijet background, while the $H \to \gamma\gamma$ has a unique and well defined final state with two

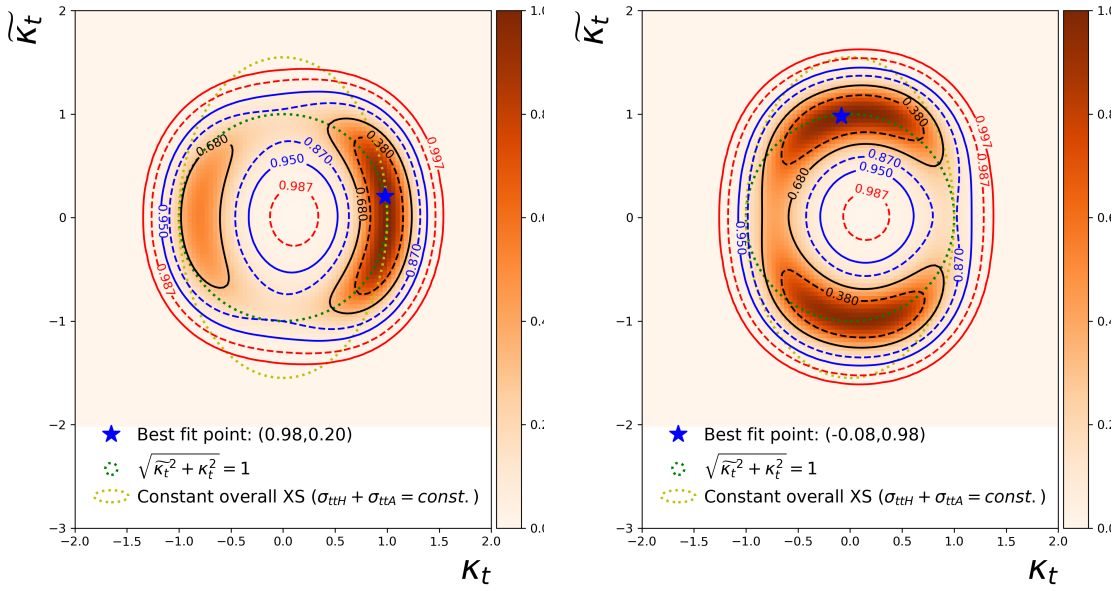


Figure 7.3: Expected exclusion contour in the $(k_t, \tilde{k}_t)$ plane for the combined, dilepton and l+jets channels using pure CP-even (left) and CP-odd (right) Asimov data.

photons. Nevertheless, searching for a CP-odd component in the $H \to b\bar{b}$ decay channel is still necessary, since the $H \to \gamma\gamma$ is loop-induced and may not be model independent (i.e. can be affected by effects of BSM physic), while the $H \to b\bar{b}$ channel provides a direct measurement.

7.1.1 Impact of systematic uncertainties

To assess the impact of the many sources of systematic uncertainties, several one-dimensional log-likelihood scans are performed across the allowed range of α using Asimov data (described in Section 7.1), where each scan is performed for each nuisance parameter, by fixing the corresponding parameter to its pre-fit value and allowing all other nuisance parameters to vary. The nominal model is compared with the model where each systematic uncertainty was removed from the fit, and the impact on the uncertainty of α is given by the square root of the quadratic difference between the 65% CL intervals of each case.

Figure 7.4 shows the impact of the 6 highest-ranked systematic uncertainties using the combined fit (dilepton and single lepton channels). The uncertainties related to the $t\bar{t}$ background (especially $t\bar{t}$+$\geq$1b) modelling have the largest impact on α.

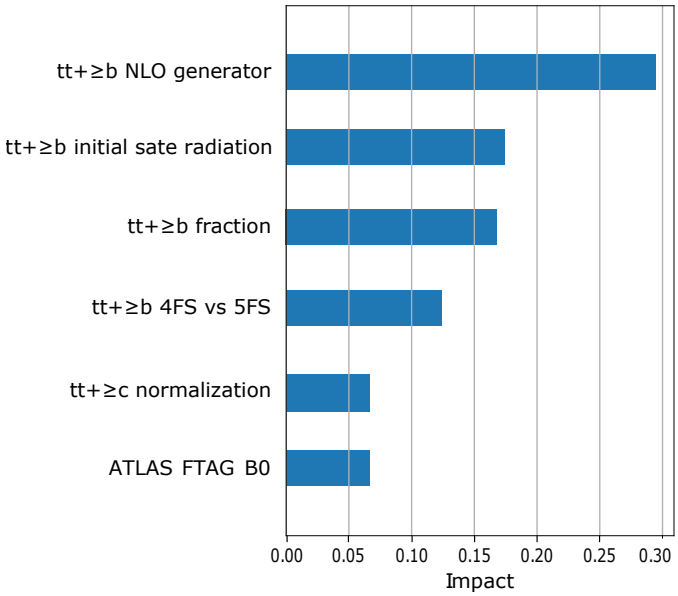


Figure 7.4: Impact of the highest ranked systematics for the combined fit. The values on the horizontal axis give the impact on α/π.

The highest-ranked uncertainty is due to the choice of $t\bar{t} + \geq 1b$ NLO MC generator, this uncertainty is calculated by comparing samples generated with MadGraph5_aMC@NLO+Pythia8 (Section 5.1.1) with the nominal samples generated with PowhegBox+Pythia8.

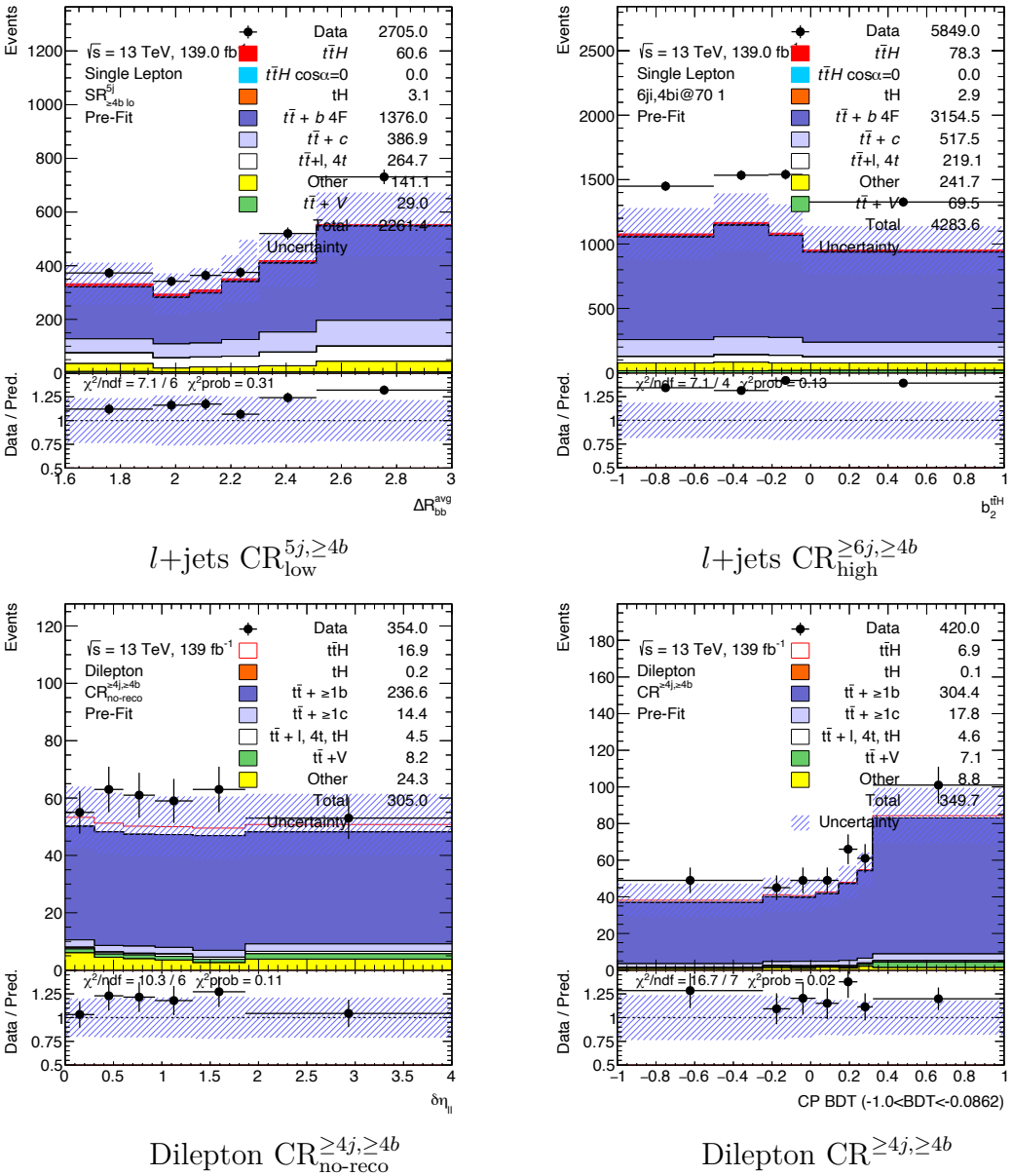
The $t\bar{t}+\geq 1b$ initial state radiation and $t\bar{t}+\geq 1b$ fraction also presented a large impact on α. The former is obtained by varying several settings on the nominal samples, such as the strong-coupling constant and the amount of QCD radiation. The latter is assigned to the relative importance of the subcomponents ($t\bar{t} + 1b$, $t\bar{t} + 1B$ and $t\bar{t} + \geq 2b$) included in the $t\bar{t} + \geq 1b$ background, and it is obtained by comparing the fractions of each subcategory in samples of PowhegBox+Pythia8 and PowhegBox+Herwig7.

The $t\bar{t} + \geq 1b$ 4FS vs 5FS (4-flavor versus 5-flavor schemes) systematic uncertainty is also highly-ranked. This uncertainty arises from a comparison between background samples of $t\bar{t}+\geq 1b$ generated with PowhegBox+Pythia8 in the 4-flavor scheme with those generated in the 5-flavor scheme.

The $t\bar{t} + \geq 1c$ normalization factor and the *ATLAS FTAG B0*, which is an experimental uncertainty related to b-tagging efficiency, presented a noticeable, but smaller, impact on α as well.

7.1.2 Background-Only Fit

A comparison between distributions of data and MC has been performed in bins where the signal to background ratio is small ($\leq 7.7\%$ of the total expected yield) to assess the accuracy of the background model. This comparison has been carried out before and after a fit performed to data using only background MC

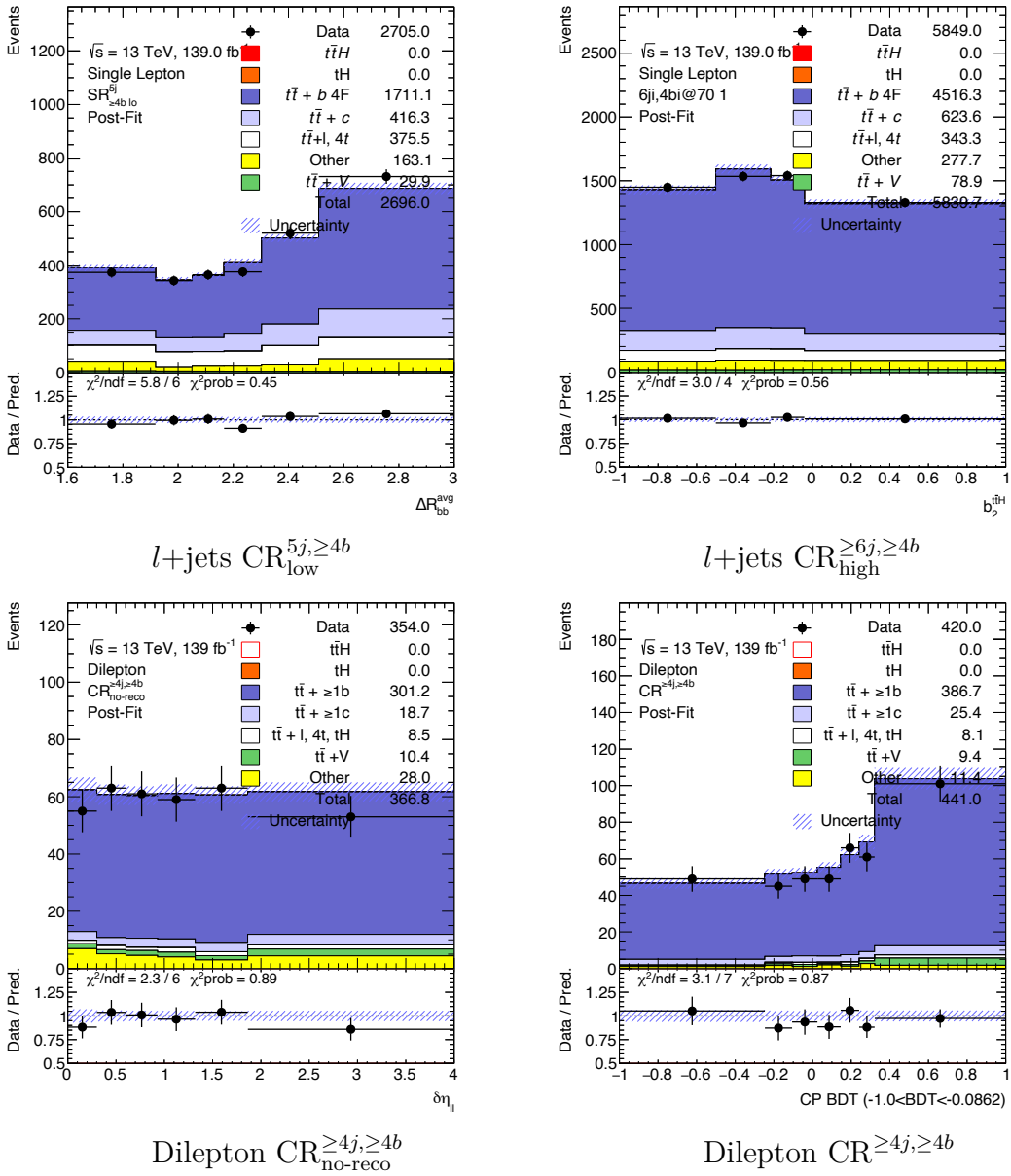


l+jets $\mathrm{CR}^{5j,\geq 4b}_{\mathrm{low}}$ l+jets $\mathrm{CR}^{\geq 6j,\geq 4b}_{\mathrm{high}}$

Dilepton $\mathrm{CR}^{\geq 4j,\geq 4b}_{\mathrm{no\text{-}reco}}$ Dilepton $\mathrm{CR}^{\geq 4j,\geq 4b}$

Figure 7.5: Pre-fit distributions for the $\mathrm{CR}^{5j,\geq 4b}_{\mathrm{low}}$, $\mathrm{CR}^{\geq 6j,\geq 4b}_{\mathrm{high}}$, $\mathrm{CR}^{\geq 4j,\geq 4b}_{\mathrm{no\text{-}reco}}$, $\mathrm{CR}^{\geq 4j,\geq 4b}$ regions before background only fit.

(no signal is considered). The modelling of all the distributions that are used to perform the fit, as well as the distributions of the input variables to the CP BDT, has been inspected. Figure 7.5 and 7.6 shows the pre- and post-fit agreement between data and MC distributions of some variables that are used in the fit in the dilepton and single lepton channels. The distributions show a reasonable agreement in these regions, which is reflected by relatively high (> 0.5) p-values.



l+jets $\mathrm{CR}^{5j,\geq 4b}_{\mathrm{low}}$ l+jets $\mathrm{CR}^{\geq 6j,\geq 4b}_{\mathrm{high}}$

Dilepton $\mathrm{CR}^{\geq 4j,\geq 4b}_{\text{no-reco}}$ Dilepton $\mathrm{CR}^{\geq 4j,\geq 4b}$

Figure 7.6: Post-fit distributions for the $\mathrm{CR}^{5j,\geq 4b}_{\mathrm{low}}$, $\mathrm{CR}^{\geq 6j,\geq 4b}_{\mathrm{high}}$, $\mathrm{CR}^{\geq 4j,\geq 4b}_{\text{no-reco}}$, $\mathrm{CR}^{\geq 4j,\geq 4b}$ regions after background only fit.

The best-fit values and associated uncertainties for the nuisance parameters associated to the background modelling are presented in Figure 7.7. The pulls are obtained after the background only fits to data for the single lepton, dilepton and combined channels. The pulls are calculated from the difference between the best-fit value ($\hat{\theta}$) and the nominal value (θ_0) of the nuisance parameters, divided by their pre-fit uncertainty ($\Delta\theta$).

Overall, deviations of the fitted nuisance parameters from their nominal values are small and compatible between the two channels. In the combined fit, the pulls are smaller and the nuisance parameters are slightly more constrained. In addition, the dilepton channel compensates for the large nuisance parameter pull in the $t\bar{t}+\geq 1b$ 4F initial state radiation, which is observed in single lepton channel, and that was discussed in Section 6.9. This pull is smaller in the combination and it is still the only pull larger than one standard deviation.

It is also interesting to note that the dilepton channel is particularly important to constrain the $t\bar{t}+\geq 1c$ background. This is reflected in a significant decrease of the pull for the $t\bar{t}+\geq 1c$ normalization uncertainty in the combination.

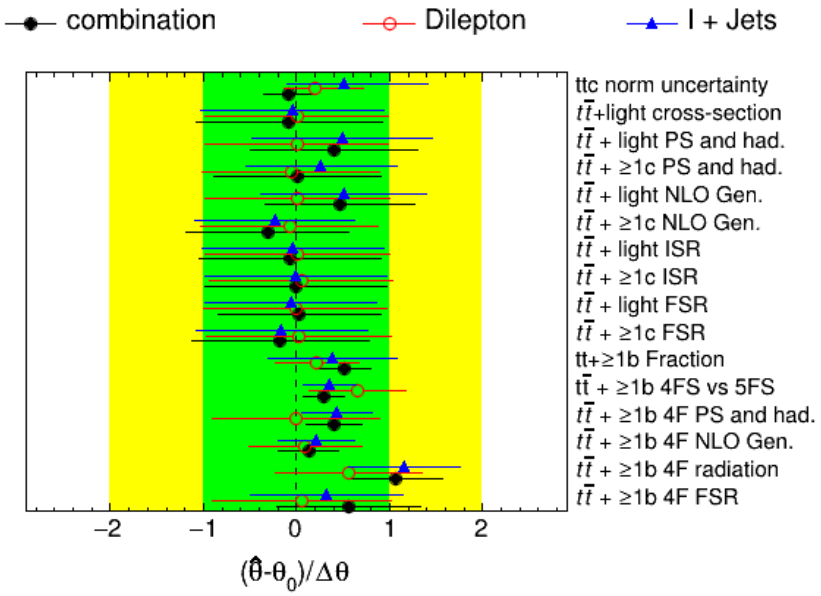


Figure 7.7: Background modelling nuisance parameters from background only fits to the l+jets, dilepton and combined channels. The green (yellow) area represent the $\pm 1(2)\sigma$ intervals on the post-fit systematic uncertainty. The points and horizontal bars represent the pulls and constraints in units of standard deviation, respectively.

7.2 HL-LHC Sensitivity to CP Mixing Angle

In order to obtain the expected sensitivity of the analysis for higher values of luminosity and predict when the CP-odd component will be excluded with a significance of 3σ, Asimov fits are performed using Asimov data produced for different values of luminosity (139, 300, 1000, 2000 and 3000 fb^{-1}). The expected pure CP-odd exclusion significance is obtained for each of these values, serving as a benchmark of the expected analysis sensitivity.

Due to technical reasons only the single lepton channel is used in the fits for this study. However, the dilepton channel is essentially used to constrain the $t\bar{t}+ \geq 1c$ background and the significance is not significantly improved when both channels are combined, as seem from Figure 7.1. The effects of different systematics uncertainties on the future expected sensitivity can still be assessed using only the single lepton channel, which dominates the sensitivity.

A decrease in the systematic uncertainties due to detector improvements and higher data statistics is expected to be achieved with the HL-LHC. In view of this, several scenarios were considered for the evolution of systematic uncertainties:

- **Statistical Only Scenario:** This scenario is included for comparison and only considers statistical uncertainties.
- **Run 2 Scenario:** This scenario assumes that the systematic uncertainties will remain unchanged with respect to the LHC Run 2 period.
- **50% Systematic Uncertainties Scenario:** This scenario assumes that both theory and experimental systematic uncertainties will decrease over time. All the systematics included in the current Run-2 analysis are reduced by a factor two with respect to their relative uncertainty values. Except for the luminosity uncertainty, which is reduced to 40% of its original value. This scenario is similar to the YR18 scenario referred in [44].
- **No Instrumental Uncertainties Scenario:** In this scenario all of the experimental uncertainties, including luminosity uncertainty, are removed from the fit to assess the effect of these uncertainties in the significance.
- **No Theory Uncertainties Scenario:** This scenario neglects all of the theoretical uncertainties related with the simulation and modelling of processes. The theoretical cross section uncertainties are still included in this scenario.
- **No Cross Section Uncertainties Scenario:** In this scenario all of the theoret-

ical cross section uncertainties of processes are removed to assess the effect of these uncertainties in the significance.

Table 7.1: Considered uncertainty scenarios and the changes applied to each type of systematic uncertainty in each scenario.

Source Uncert.	Stat. Only	Run 2 Syst.	50% Syst.	No Inst.	No Theory	No Cross Sec.
Instrumental	None	Unchanged	0.5	None	Unchanged	Unchanged
Int. Luminosity	None	Unchanged	0.4	None	Unchanged	Unchanged
Theory (shapes)	None	Unchanged	0.5	Unchanged	None	Unchanged
Cross Section	None	Unchanged	0.5	Unchanged	Unchanged	None

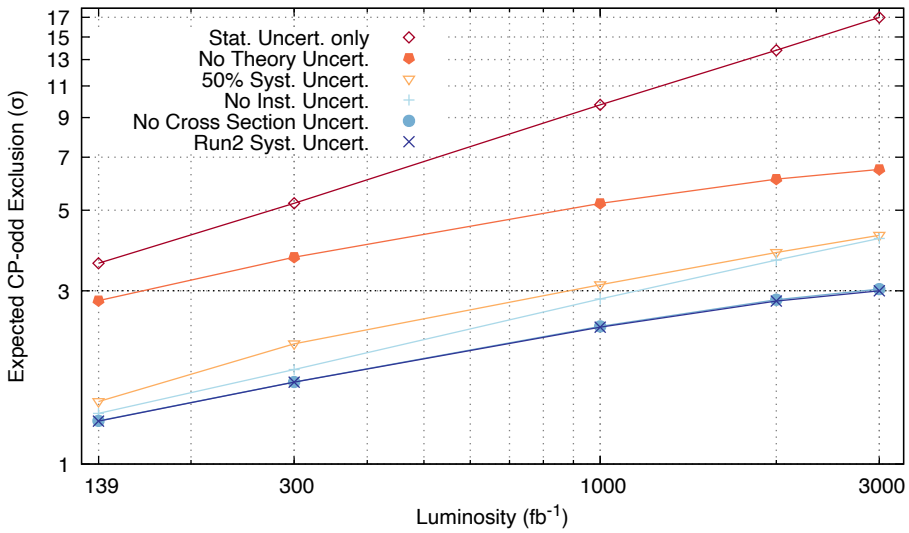


Figure 7.8: Expected CP-odd exclusion in standard deviations σ as a function of the LHC luminosity. Several different scenarios for the systematic uncertainties are displayed.

A list containing the changes applied to each type of systematic uncertainty for each scenario is presented in Table 7.1. The expected CP-odd exclusion is represented as a function of the LHC luminosity for each scenario in Figure 7.8.

The scenario with only statistical uncertainties exhibits a significant increase in the expected exclusion significance. In this scenario, an exclusion greater than 3σ would be achieved for the current LHC luminosity. If there were no systematic uncertainties the significance would evolve with the square root of the luminosity, while in the other scenarios it would evolve logarithmically. This reveals that the significance is being strongly limited by systematic uncertainties.

For the Run 2 scenario, the 3σ exclusion significance is expected to be obtained for a luminosity of approximately 3000fb^{-1}. This value is expected to be reached only at the end of the HL-LHC plan (Section 3.1.1) approximately in year 2035.

In the optimistic 50% Syst. scenario, where the uncertainties are divided by a factor of two, an expected exclusion significance greater than 3σ is obtained for a luminosity of about 900fb^{-1}, which is expected to be reached approximately in year 2027. For 3000fb^{-1}, an exclusion of 4.27σ is expected.

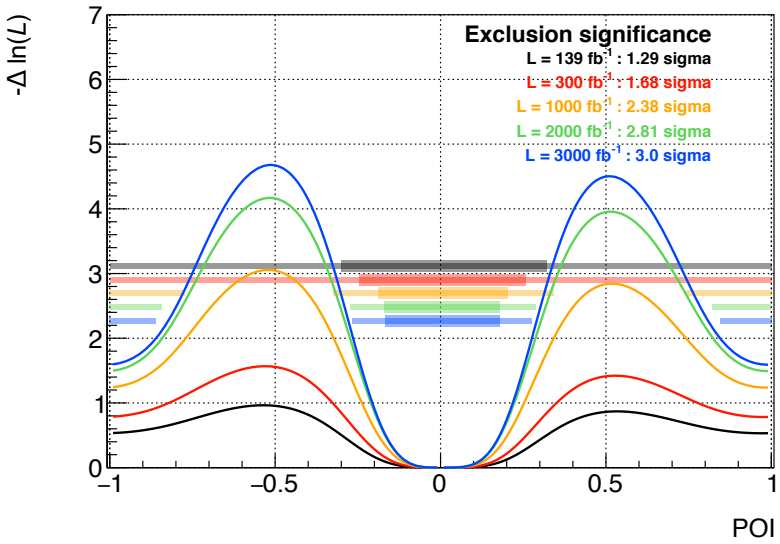
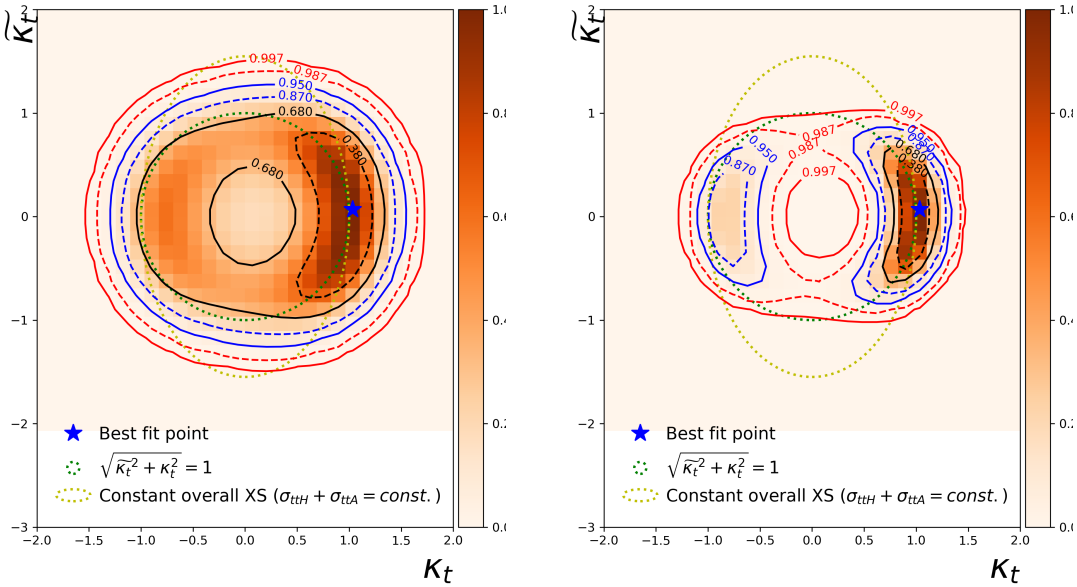


Figure 7.9: One-dimensional log-likelihood scan for the l+jets channel using pure CP-even Asimov data for different values of luminosity. The expected CP-odd exclusions are shown for each luminosity. The scans were obtained using the Run 2 systematic uncertainty scenario.

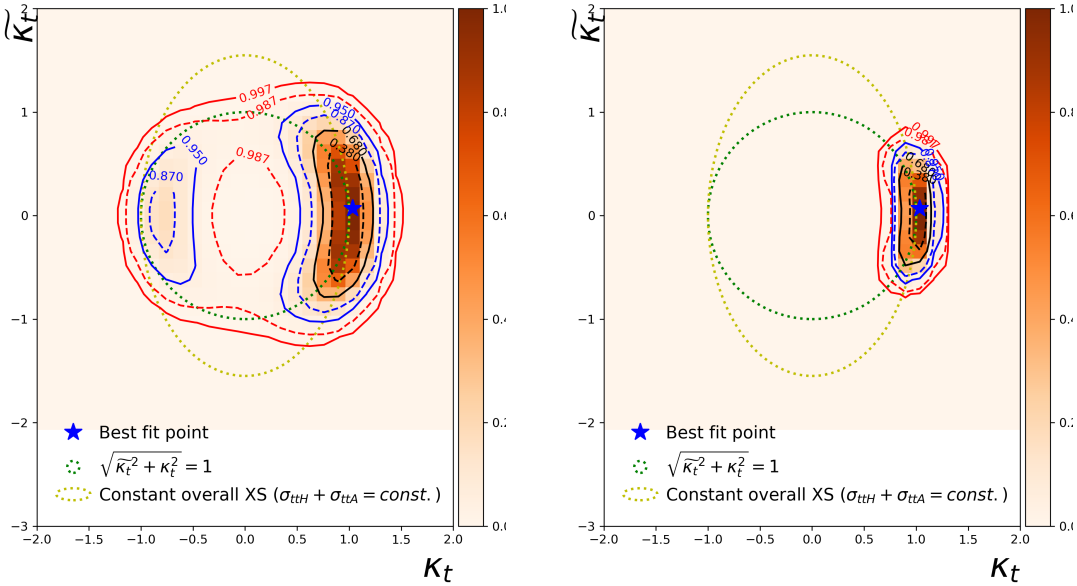
A large increase in the significance is obtained for the case were uncertainties on the modelling of processes are neglected (represented in red). Although this scenario is not fully realistic, it reveals that improvements in theoretical calculations at higher orders of accuracy are demanded. Besides, when no cross section uncertainty is considered, only a small difference is achieved when comparing to the Run 2 scenario. This means that the uncertainties on the shapes of the distributions have a more significant impact on the expected significance than the theoretical cross sections. The impact of the instrumental uncertainties on the significance also appears to be far less extreme than the impact from the uncertainties on the modelling.



Run 2 Uncert.
$L = 139\text{fb}^{-1}$

Run 2 Uncert.
$L = 3000\text{fb}^{-1}$

Figure 7.10: Expected exclusion contour for the l+jets channel using pure CP-even Asimov data for a luminosity of 139 (left) and 3000 fb^{-1} (right). These contours were obtained using the Run 2 systematic uncertainty scenario.



No Theory Uncert.
$L = 139\text{fb}^{-1}$

No Theory Uncert.
$L = 3000\text{fb}^{-1}$

Figure 7.11: Expected exclusion contour for the l+jets channel using pure CP-even Asimov data for a luminosity of 139 (left) and 3000 fb^{-1} (right). These contours were obtained using the scenario where the theoretical uncertainties on the shapes of distributions are neglected.

The log-likelihood scan on α and the expected CP-odd exclusion obtained with the Run 2 systematic uncertainty scenario are represented in Figure 7.9 for each luminosity. Figure 7.10 shows the expected exclusion contours in the $(k_t, \tilde{k}_t)$ plane using this pessimistic scenario for the current luminosity and for 3000 fb^{-1}. For 3000 fb^{-1}, the pure CP-odd scenario is outside the 3σ contour, while the inverted coupling scenario is outside the 1σ contour and inside the 2σ contour.

Figure 7.11 shows the expected contours for the current luminosity and for 3000 fb^{-1} using the scenario where the theoretical uncertainties on the shapes of distributions are removed. For 3000 fb^{-1}, the pure CP-odd scenario and the inverted coupling scenario are outside the 3σ contour, and the coupling parameters are severely more constrained when compared to the Run 2 scenario.

8

Conclusion

This thesis describes the search for CP-odd $t\bar{t}H$ production, in the $H \to b\bar{b}$ decay channel. The analysis uses a dataset of proton-proton collisions corresponding to an integrated luminosity of 139 fb^{-1} at a center-of-mass energy of $\sqrt{s} = 13$ TeV collected during the LHC Run 2 by the ATLAS detector. Events in the dilepton and single lepton channels are classified into control and inclusive signal regions according to the final state particles of each event, such as the number of jets and the number of b-tagged jets. A boosted decision tree (BDT) is used to discriminate between signal and background events in the inclusive regions and further divide these regions. A profile likelihood fit is performed simultaneously to all analysis regions to determine the CP mixing angle and the coupling modifier, as the expectations for signal and background are adjusted according to the corresponding systematic uncertainties. Several CP-discriminant variables are used in the fit, including the CP-BDT which combines several CP-sensitive variables into a single classifier.

The analysis relies on CP-sensitive observables to discriminate between different CP scenarios and measure the CP mixing angle and the coupling modifier of the top Yukawa coupling. In addition to these CP-discriminant distributions, other variables are fitted in the control regions of the analysis, in order to constrain the normalization and shapes of the background. Despite the fact that these are not sensitive to the CP of the t-H coupling, they help to improve the sensitivity and robustness of the analysis.

For a luminosity of 139 fb^{-1}, the pure CP-odd coupling is expected to be excluded at 1.60σ when both channels are combined in a fit to an Asimov dataset that follows the Standard Model expectations. When the fit is performed only in the single lepton channel, which dominates the analysis sensitivity, an expected CP-odd exclusion of 1.29σ is obtained. While with the dilepton channel, which is

particularly important for constraining the $t\bar{t}+\geq 1c$ background, an exclusion significance of 0.98σ is achieved. The expected two-dimensional confidence intervals on the $(k_t, \tilde{k}_t)$ plane are derived. The inverted coupling scenario (i.e. $k'_t = -1$) is expected to be inside the 1σ contour.

The impact of the many sources of systematic uncertainties in the analysis sensitivity was investigated. It was shown that the uncertainties with the greatest impact on the CP mixing angle are the ones related with the modelling of the $t\bar{t}+\geq 1b$ background. In addition, a background-only fit was performed to the regions with small signal to background ratio in order to assess the accuracy of the background model. A fairly reasonable agreement was observed with high p-values.

Furthermore, the current analysis strategy was compared with an older version, in which the CP-BDT was also being fitted in the single lepton channel. However, it was shown that fitting a CP-discriminant variable referred to as $b_2^{t\bar{t}H}(t,\bar{t})$ is almost as good as fitting the CP-BDT in this channel. The analysis sensitivity decreases by just 0.01σ when applying this simplification. In addition, a better background modelling is achieved when the $b_2^{t\bar{t}H}(t,\bar{t})$ observable is used.

This work also explores projections of this search from the LHC Run 2 to the HL-LHC. The projections are based on the single lepton channel of the analysis. Integrated luminosities of up to 3000 fb^{-1} are considered in the projections, with different scenarios for the expected systematic uncertainties. The pure CP-odd top Yukawa coupling is expected to be excluded with a significance greater than 3σ after 3000 fb^{-1} in the Run 2 uncertainties scenario. In the 50% systematics scenario, this same exclusion is predicted to be achieved before 1000 fb^{-1}. In this scenario, the expected sensitivity of the search will be improved by almost a factor of 3 by the end of the HL-LHC program. However, when only statistical uncertainties are considered, the Run 2 luminosity is enough for an exclusion of 3.57σ. Clearly, the systematic uncertainties are significantly limiting the precision of the measurement.

In addition, three more scenarios were also considered by removing specific types of systematic uncertainties to asses its effects on the exclusion significance. The uncertainties on the modelling of processes, which is dominated by uncertainties on the modelling of background, presented the greatest impact on the expected significance. In summary, the sensitivity to the signal is strongly limited by the mismodelling of background, and a better understanding of the theory, with

higher orders of accuracy calculations, especially for the background events, would strongly benefit the analysis.

During this work, the author participated and actively contributed to several stages of an ATLAS analysis, which is expected to be published soon. Among these contributions, we highlight the tH parameterization as a function of the mixing angle α and the coupling strength k_t' from Section 6.6 and the modelling of background presented in Section 7.1.2, as well as fit studies and analysis simplification strategies, such as the $b_2^{t\bar{t}H}$ alternative to the CP-BDT as CP-discriminant variable in the dilepton channel, mentioned in Section 6.5. The author also presented the projections of this search for the HL-LHC in Section 7.2 and a comparison between distributions of large-R jet substructure variables using Atlfast-II and full detector simulation in Section 5.3.

Appendices

A

Statistical Analysis Techniques

In this analysis, two parameters of interest (POI) are being considered, the CP-mixing angle α and the coupling modifier k'_t. The POI are the unknown parameters of the theoretical model that predicts the distribution of the observables, and the objective of the analyses is to measure these parameters. In view of this, a text-statistic that considers the two POI is required. The following sections cover all the techniques applied for the measurement these parameters, starting with an explanation of the maximum likelihood estimator, followed by a description on how the uncertainties on the parameters of interest and the significance are measured. Finally the test statistic used for measuring the CP-mixing angle α and the one used for simultaneously measuring α and the coupling modifier k'_t are discussed.

A.1 The Maximum Likelihood Estimator

For a sample of N measurements of variables $\vec{x} = x_1, ..., x_l$, with distributions that take into account both intrinsic physics randomness (theory) and detector effects described according to some unknown parameters, the probability density functions f that describe the data are expressed as a likelihood function:

$$L(\vec{x}; \vec{\mu}, \vec{\theta}) = \prod_{i=1}^{N} f(x_1^i, ..., x_l^i; \mu_1, ..., \mu_k, \theta_1, ..., \theta_m). \tag{A.1}$$

The parameters can be divided into POI, represented as $\vec{\mu}$, and nuisance parameters, represented as $\vec{\theta}$. The nuisance parameters are the additional parameters that need to be taken into account when evaluating the POI. The values of the POI depend on the values of the nuisance parameters, and these parameters arise, for example, due to the finite resolution of the detectors, miscalibrations, presence of background, etc.

The parameters for which the theoretical probability density functions best fits

the experimental data sample corresponds to the set of parameters that maximizes the likelihood function. The maximum likelihood method consists in determining the parameter set that corresponds to the maximum value of the likelihood function. This is equivalent to minimizing the $-\log L(\vec{x};\vec{\mu},\vec{\theta})$, which may be more convenient since the product of all the terms is transformed into a sum of logarithms. Usually, a scan around the minimum value is performed to determine the uncertainty contour in the fitted parameters.

This analysis is divided into several regions (Section 6.2), each with multiple bins of observables. The number of events in each bin follows a Poisson distribution and the expected number of events in each bin depends on the unknown parameters of the theory that one wants to estimate. A binned likelihood function is built as a product of Poisson probability terms over all the bins considered in the analysis:

$$L(\vec{n};\vec{\mu},\vec{\theta}) = \prod_{i=1}^{n_{bins}} \frac{e^{-\nu_i(\mu_1,...,\mu_k,\theta_1,...,\theta_m)}\nu_i(\mu_1,...,\mu_k,\theta_1,...,\theta_m)^{n_i}}{n_i!}, \tag{A.2}$$

where, n_{bins} is the number of bins, n_i is the number of entries for each bin i and ν_i is the expected number of entries, that depends on the parameters $\vec{\mu}$ and $\vec{\theta}$. This approach is usually preferred when dealing with an extensive number of measurements N and when the implementation of an unbinned likelihood would demand intensive computing power.

In addition, a constraint term for the nuisance parameters is introduced in the likelihood function:

$$L(\vec{n};\vec{\mu},\vec{\theta}) = \prod_{i=1}^{n_{bins}} \text{Poiss}(n_i;\nu_i(\mu_1,...,\mu_k,\theta_1,...,\theta_m)) \prod_{j=1}^{m} G(\theta_j^0;\theta_j), \tag{A.3}$$

where m is the number of nuisance parameters and θ_j^0 is the nominal value of each nuisance parameter, which is distributed according to one PDF $G(\theta_j^0;\theta_j)$, usually a Gaussian or log-normal distribution. The nominal value is usually an estimation of the true value of the nuisance parameter, and its fitted value corresponds to the one that best fits the data. Background-dominated regions, referred to as control regions, are used to constrain the systematic uncertainties on the measurement.

A.2 The Profile Likelihood Ratio

The profile likelihood method is commonly used in LHC analyses for the treatment of nuisance parameters. Is is based on a likelihood ratio that is used to construct a test statistic and test a certain POI against its alternatives over the parameter space:

$$\lambda(\vec{\mu}) = \frac{L(\vec{x}; \vec{\mu}, \hat{\hat{\vec{\theta}}}(\vec{\mu}))}{L(\vec{x}; \hat{\vec{\mu}}, \hat{\vec{\theta}})}, \tag{A.4}$$

where, in the denominator, $\hat{\vec{\mu}}$ and $\hat{\vec{\theta}}$ are the best fit values of $\vec{\mu}$ and θ, which are fitted simultaneously. Whereas, in the numerator, $\hat{\hat{\vec{\theta}}}(\vec{\mu})$ corresponds to the best fit value of $\vec{\theta}$ obtained for a fixed alternative value of $\vec{\mu}$.

The following test statistic can be build for the purpose of establishing a limit on the parameters $\vec{\mu}$.

$$q_\mu = -2 \log \lambda(\vec{\mu}), \tag{A.5}$$

which, by construction, has a minimum at $q_\mu(\hat{\vec{\mu}}) = 0$. A scan around the minimum can be performed to determine the uncertainty in the value of $\hat{\vec{\mu}}$, given by the intersection with $-\log \lambda(\vec{\mu}) = 0.5$ that corresponds to a significance of 1σ.

The p-value can be calculated to quantify the level of agreement between a hypothesized $\vec{\mu}$ and the data as

$$p_\mu = \int_{q_\mu, obs}^{\infty} f(q_\mu|\vec{\mu}) dq_\mu, \tag{A.6}$$

where $f(q_\mu|\vec{\mu})$ is the PDF of q_μ assuming that $\vec{\mu}$ is true. In the limit of large samples and considering that the Wilks' theorem [125] holds (i.e. ensuring that the true values of the evaluated parameters lie within the permitted parameter space), the test statistic q_μ will be asymptotically distributed as $f(q_\mu|\vec{\mu}) \approx \chi^2_k$ with degrees of freedom equal to the number of components in $\vec{\mu} = (\mu_1, ..., \mu_k)$. The hypothesized $\vec{\mu}$ can be excluded at a certain confidence level (CL) if the p-value is found below a specified threshold. The confidence interval is build for a certain CL with the values of $\vec{\mu}$ that cannot be excluded for that value of CL. In this limit, the exclusion significance of $\vec{\mu}$ can be approximated to $Z_\mu \approx \sqrt{q_\mu} = \sqrt{-2 \log \lambda(\vec{\mu})}$ [126].

For the implementation of the statistical analysis, the TRexFitter [127] framework was used. All parameters were allowed to vary, with the constraints that $-\pi \leq \alpha \leq \pi$ and $0 \leq k'_t$, and the log-likelihood function was minimized to obtain

the best-fit values of the CP-mixing angle α and the coupling modifier k_t.

In order to obtain the confidence intervals and hypothesis exclusion, a test statistic q_α was constructed as

$$q_\alpha = -2 \ln \frac{L(\alpha, \hat{\hat{k}}'_t(\alpha), \hat{\hat{\vec{\theta}}}(\alpha))}{L(\hat{\alpha}, \hat{k}'_t, \hat{\vec{\theta}})}, \tag{A.7}$$

where, $\hat{\alpha}, \hat{k}'_t$ and $\hat{\vec{\theta}}$ are the measured best-fit values, while $\hat{\hat{k}}'_t(\alpha)$ and $\hat{\hat{\vec{\theta}}}(\alpha)$ are the values of k'_t and θ that minimize the log-likelihood for a specific value of α. One-dimensional scans are performed along the allowed range of α.

A second test statistic q_{α,k'_t} was build to define the confidence region in the (α, k'_t) plane. q_{α,k'_t} was defined as

$$q_{\alpha,k'_t} = -2 \ln \frac{L(\alpha, k'_t, \hat{\hat{\vec{\theta}}}(\alpha, k'_t))}{L(\hat{\alpha}, \hat{k}'_t, \hat{\vec{\theta}})}. \tag{A.8}$$

In this case, two-dimensional scans of q_{α,k'_t} are performed along α and k'_t and the confidence region was represented in the $(k_t, \tilde{k}_t)$ parameterization (see equation 4.3).

Bibliography

[1] ATLAS Collaboration, *"Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,"* Phys. Lett. B **716** (2012), 10.1016/j.physletb.2012.08.020.

[2] CMS Collaboration, *"Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,"* Phys. Lett. B **716** (2012), 10.1016/j.physletb.2012.08.021.

[3] ATLAS Collaboration, *"Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector,"* Phys. Lett. B **784**, 173 – 191 (2018).

[4] CMS Collaboration, *"Observation of $t\bar{t}H$ Production,"* Phys. Rev. Lett. **120**, 231801 (2018).

[5] C. Burgard, "TikZ diagram of the Standard Model of physics." `http://www.texample.net/tikz/examples/model-physics/`.

[6] Particle Data Group, *"Review of Particle Physics,"* Phys. Rev. D **98** (2018), 10.1103/PhysRevD.98.030001.

[7] F. Englert and R. Brout, *"Broken Symmetry and the Mass of Gauge Vector Mesons,"* Phys. Rev. Lett. **13**, 321–323 (1964).

[8] P. Higgs, *"Broken Symmetries and the Masses of Gauge Bosons,"* Phys. Rev. Lett. **13**, 508–509 (1964).

[9] P. Higgs, *"Broken symmetries, massless particles and gauge fields,"* Phys. Lett. **12**, 132–133 (1964).

[10] G. Guralnik, C. Hagen, and T. Kibble, *"Global Conservation Laws and Massless Particles,"* Phys. Rev. Lett. **13**, 585–587 (1964).

[11] P. Higgs, *"Spontaneous Symmetry Breakdown without Massless Bosons,"* Phys. Rev. **145**, 1156–1163 (1966).

[12] T. Kibble, *"Symmetry Breaking in Non-Abelian Gauge Theories,"* Phys. Rev. **155**, 1554–1561 (1967).

[13] E. Noether, *"Invariant variation problems,"* Transport Theory and Stat. Phys. **1**, 186–207 (1971).

[14] F. Halzen and A. Martin, *Quarks and leptons: an introductory course in modern particle physics* (Wiley, New York, NY, 1984).

[15] I. Aitchison and A. Hey, *Gauge theories in particle physics: a practical introduction; 3rd ed.*, Graduate student series in physics (IOP, Bristol, 2004).

[16] D. Griffiths, *Introduction to elementary particles; 2nd rev. version*, Physics textbook (Wiley, New York, NY, 2008).

[17] G. Kane, *Modern elementary particle physics: explaining and extending the standard model; 2nd ed.* (Cambridge University Press, Cambridge, 2017).

[18] P. Langacker, *The standard model and beyond; 2nd ed.*, High energy physics, cosmology and gravitation (CRC Press, 2017).

[19] J. Goldstone, A. Salam, and S. Weinberg, *"Broken Symmetries,"* Phys. Rev. **127**, 965–970 (1962).

[20] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector," (2017), arXiv:1610.07922 [hep-ph] .

[21] ATLAS Collaboration, *"Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4l$ and $H \to \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector,"* Phys. Lett. B **784**, 345 – 366 (2018).

[22] ATLAS and CMS Collaborations, *"Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments,"* Phys. Rev. Lett. **114**, 191803 (2015), arXiv:1503.07589 [hep-ex] .

[23] ATLAS Collaboration, *Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2018-031 (CERN, Geneva, 2018).

[24] LHC Higgs Cross Section Working Group, "LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle," (2012), arXiv:1209.0040 [hep-ph] .

[25] ATLAS and CMS Collaborations, *"Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV,"* JHEP **08**, 045 (2016), arXiv:1606.02266 [hep-ex] .

[26] CMS Collaboration, *"Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV,"* Phys. Rev. D **92**, 012004 (2015).

[27] CMS Collaboration, *"Combined search for anomalous pseudoscalar HVV couplings in VH(H $\to bb^-$) production and H $\to$ VV decay,"* Phys. Lett. B **759**, 672 – 696 (2016).

[28] ATLAS Collaboration, *"Evidence for the spin-0 nature of the Higgs boson using ATLAS data,"* Phys. Lett. B **726**, 120–144 (2013), arXiv:1307.1432 [hep-ex] .

[29] ATLAS Collaboration, *"Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector,"* Eur. Phys. J. C **75**, 476 (2015), [Erratum: Eur.Phys.J.C 76, 152 (2016)], arXiv:1506.05669 [hep-ex] .

[30] L. Evans and P. Bryant, *"LHC Machine,"* JINST **3**, S08001 (2008).

[31] ATLAS Collaboration, *"The ATLAS Experiment at the CERN Large Hadron Collider,"* JINST **3**, S08003 (2008).

[32] CMS Collaboration, *"The CMS experiment at the CERN LHC,"* JINST **3**, S08004 (2008).

[33] ALICE Collaboration, *"The ALICE experiment at the CERN LHC,"* JINST **3**, S08002 (2008).

[34] LHCb Collaboration, *"The LHCb Detector at the LHC,"* JINST **3**, S08005 (2008).

[35] ATLAS Collaboration, "ATLAS Luminosity Public Results for Run 2," https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Luminosity_summary_plots_for_201.

[36] ATLAS Collaboration, "ATLAS Luminosity Public Results for Run 1," https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults#2010_pp_Collisions.

[37] R. Alemany-Fernandez *et al.*, *"Operation and Configuration of the LHC in Run 1,"* (2013).

[38] R. Bruce *et al.*, *"LHC Run 2: Results and challenges," Proceedings, 57th ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams (HB2016): Malmö, Sweden, July 3-8, 2016*, , MOAM5P50 (2016).

[39] ATLAS Collaboration, *"Luminosity determination in pp collisions at* $\sqrt{s} =$ 13 *TeV using the ATLAS detector at the LHC,"* (2019).

[40] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1*, Tech. Rep. CERN-LHCC-99-014 (CERN, Geneva, 1999).

[41] G. Apollinari *et al.*, *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*, CERN Yellow Reports: Monographs (CERN, Geneva, 2015).

[42] ATLAS Collaboration, *ATLAS Phase-II Upgrade Scoping Document*, Tech. Rep. CERN-LHCC-2015-020. LHCC-G-166 (CERN, Geneva, 2015).

[43] D. Contardo *et al.*, *Technical Proposal for the Phase-II Upgrade of the CMS Detector*, Tech. Rep. CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02 (CERN, Geneva, 2015).

[44] P. Azzi *et al.*, "Report from Working Group 1: Standard Model Physics at the HL-LHC and HE-LHC," in *Report on the Physics at the HL-LHC,and Perspectives for the HE-LHC*, Vol. 7 (2019) pp. 1–220, arXiv:1902.04070 [hep-ph] .

[45] CMS Collaboration (CMS Collaboration), *Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC*, Tech. Rep. CMS-NOTE-2018-006. CERN-CMS-NOTE-2018-006 (CERN, Geneva, 2018).

[46] ATLAS Collaboration, *Expected performance of the ATLAS detector at the High-Luminosity LHC*, Tech. Rep. ATL-PHYS-PUB-2019-005 (CERN, Geneva, 2019).

[47] R. Abdul Khalek *et al.*, *"Towards Ultimate Parton Distributions at the High-Luminosity LHC,"* Eur. Phys. J. C **78**, 962 (2018), arXiv:1810.03639 [hep-ph] .

[48] "How to draw diagrams in LaTeX with TikZ," https://wiki.physik.uzh.ch/cms/latex:tikz.

[49] J. Pequenao, "Computer generated image of the whole ATLAS detector," https://cds.cern.ch/record/1095924, Mar, 2008.

[50] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*, Tech. Rep. CERN-LHCC-97-016 (CERN, Geneva, 1997).

[51] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 2*, Tech. Rep. CERN-LHCC-97-017 (CERN, Geneva, 1997).

[52] ATLAS Collaboration, *"The ATLAS Inner Detector commissioning and calibration,"* Eur. Phys. J. **C70**, 787–821 (2010), arXiv:1004.5293 [physics.ins-det] .

[53] *Alignment of the ATLAS Inner Detector and its Performance in 2012*, Tech. Rep. ATLAS-CONF-2014-047 (CERN, Geneva, 2014).

[54] ATLAS Collaboration, *"ATLAS pixel detector electronics and sensors,"* JINST **3**, P07007–P07007 (2008).

[55] ATLAS Collaboration, *"Operation and performance of the ATLAS semiconductor tracker,"* JINST **9**, P08009 (2014), arXiv:1404.7473 [hep-ex] .

[56] ATLAS TRT Collaboration, in *Astroparticle, particle and space physics, detectors and medical physics applications. Proceedings, 8th Conference, ICATPP 2003, Como, Italy, October 6-10, 2003* (2003) pp. 497–501, arXiv:hep-ex/0311058 [hep-ex] .

[57] J. Pequenao, "Computer generated image of the ATLAS inner detector," https://cds.cern.ch/record/1095926, Mar, 2008.

[58] ATLAS Collaboration, *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*, Tech. Rep. ATL-PHYS-PROC-2016-104 (CERN, Geneva, 2016) 15 pages, EPS-HEP 2015 Proceedings.

[59] J. Pequenao, "Computer Generated image of the ATLAS calorimeter," https://cds.cern.ch/record/1095927, Mar, 2008.

[60] ATLAS Collaboration, *"Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s}$ = 8 TeV,"* Eur. Phys. J. **C75**, 120 (2015), arXiv:1408.3179 [hep-ex] .

[61] A. Yamamoto *et al.*, *"The atlas central solenoid,"* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **584**, 53 – 74 (2008).

[62] ATLAS Collaboration, *ATLAS barrel toroid: Technical Design Report*, Tech. Rep. (Geneva, 1997).

[63] ATLAS Collaboration, *ATLAS end-cap toroids: Technical Design Report*, Tech. Rep. (Geneva, 1997) electronic version not available.

[64] ATLAS TDAQ Collaboration, *"The ATLAS data acquisition and high level trigger system,"* JINST **11**, P06008–P06008 (2016).

[65] ATLAS Collaboration, *"Performance of the ATLAS Trigger System in 2015,"* Eur. Phys. J. **C77**, 317 (2017), arXiv:1611.09661 [hep-ex] .

[66] ATLAS Collaboration, *"The run-2 ATLAS trigger system,"* J. Phys. Conf. Ser. **762**, 012003 (2016).

[67] ATLAS Collaboration, "ATLAS experiment - Public results," `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ`, 2017.

[68] D. Fontes *et al.*, *"Large pseudoscalar Yukawa couplings in the complex 2HDM,"* JHEP **2015**, 60 (2015).

[69] H. Bahl *et al.*, *"Indirect $\mathcal{CP}$ probes of the Higgs-top-quark interaction: current LHC constraints and future opportunities,"* (2020), arXiv:2007.08542 [hep-ph] .

[70] P. Artoisenet *et al.*, *"A framework for Higgs characterisation,"* JHEP **11**, 043 (2013), arXiv:1306.6464 [hep-ph] .

[71] S. Amor dos Santos *et al.*, *"Angular distributions in $t\bar{t}H(H \to b\bar{b})$ reconstructed events at the LHC,"* Phys. Rev. D **92**, 034021 (2015), arXiv:1503.07787 [hep-ph] .

[72] S. Amor Dos Santos *et al.*, *"Probing the CP nature of the Higgs coupling in $t\bar{t}h$ events at the LHC,"* Phys. Rev. D **96**, 013004 (2017), arXiv:1704.03565 [hep-ph] .

[73] D. Azevedo, A. Onofre, F. Filthaut, and R. Gonçalo, *"CP tests of Higgs couplings in $t\bar{t}h$ semileptonic events at the LHC,"* Phys. Rev. D **98**, 033004 (2018), arXiv:1711.05292 [hep-ph] .

[74] A. Ferroglia, M. Fiolhais, E. Gouveia, and A. Onofre, *"Role of the $t\bar{t}h$ rest frame in direct top-quark Yukawa coupling measurements,"* Phys. Rev. D **100**, 075034 (2019), arXiv:1909.00490 [hep-ph] .

[75] ATLAS Collaboration, *"Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector,"* Phys. Rev. D **97** (2018), 10.1103/physrevd.97.072003.

[76] ATLAS Collaboration, *"Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector,"* Phys. Rev. D **97** (2018), 10.1103/physrevd.97.072016.

[77] CMS Collaboration, *"Measurements of $\mathrm{t\bar{t}}H$ production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel,"* (2020), arXiv:2003.10866 [hep-ex] .

[78] ATLAS Collaboration, *"Study of the CP properties of the interaction of the Higgs boson with top quarks using top quark associated production of the Higgs boson and its decay into two photons with the ATLAS detector at the LHC,"* (2020), arXiv:2004.04545 [hep-ex] .

[79] J. Ellis, D. Hwang, K. Sakurai, and M. Takeuchi, *"Disentangling Higgs-Top Couplings in Associated Production,"* JHEP **04**, 004 (2014), arXiv:1312.5736 [hep-ph] .

[80] *Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, Tech. Rep. CMS-PAS-HIG-20-006 (CERN, Geneva, 2020).

[81] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, Tech. Rep. ATL-PHYS-PUB-2010-013 (CERN, Geneva, 2010).

[82] A. Buckley *et al*, *"General-purpose event generators for LHC physics,"* Phys. Rep. **504**, 145–233 (2011).

[83] J. Alwall *et al.*, *"The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,"* JHEP **07**, 079 (2014), arXiv:1405.0301 [hep-ph] .

[84] S. Frixione, P. Nason, and C. Oleari, *"Matching NLO QCD computations with parton shower simulations: the POWHEG method,"* JHEP **2007**, 070–070 (2007).

[85] S. Alioli *et al.*, *"A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,"* JHEP **2010** (2010), 10.1007/jhep06(2010)043.

[86] H.B. Hartanto, B. Jäger, L. Reina, and D. Wackeroth, *"Higgs boson production in association with top quarks in the POWHEG BOX,"* Phys. Rev. D **91** (2015), 10.1103/physrevd.91.094003.

[87] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini, *"Scattering Amplitudes with Open Loops,"* Phys. Rev. Lett. **108**, 111601 (2012), arXiv:1111.5206 [hep-ph] .

[88] Ansgar Denner, Stefan Dittmaier, and Lars Hofer, *"Collier: A fortran-based complex one-loop library in extended regularizations,"* Computer Physics Communications **212**, 220 – 238 (2017).

[89] E. Bothmann *et al.*, *"Event generation with Sherpa 2.2,"* SciPost Phys. **7** (2019), 10.21468/scipostphys.7.3.034.

[90] P. Artoisenet *et al.*, *"Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations,"* JHEP **2013** (2013), 10.1007/jhep03(2013)015.

[91] P. Artoisenet *et al.*, *"Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations,"* JHEP **2013** (2013), 10.1007/jhep03(2013)015.

[92] T. Sjostrand, S. Mrenna, and P. Skands, *"PYTHIA 6.4 Physics and Manual,"* JHEP **05**, 026 (2006), arXiv:hep-ph/0603175 [hep-ph] .

[93] T. Sjostrand, S. Mrenna, and P. Skands, *"A Brief Introduction to PYTHIA8.1,"* Comput. Phys. Commun. **178**, 852–867 (2008), arXiv:0710.3820 [hep-ph] .

[94] T. Sjöstrand *et al.*, *"An Introduction to PYTHIA 8.2,"* Comput. Phys. Commun. **191**, 159–177 (2015), arXiv:1410.3012 [hep-ph] .

[95] M. Bähr *et al.*, *"Herwig++ physics and manual,"* EPJ C **58**, 639–707 (2008).

[96] J. Bellm *et al.*, *"Herwig 7.0/Herwig++ 3.0 release note,"* EPJ C **76** (2016), 10.1140/epjc/s10052-016-4018-8.

[97] J. Allison *et al*, *"Geant4 developments and applications,"* IEEE Transactions on Nuclear Science **53**, 270–278 (2006).

[98] ATLAS Collaboration, *"The ATLAS Simulation Infrastructure,"* EPJ C **70**, 823–874 (2010).

[99] P. Artoisenet *et al.*, *"A framework for Higgs characterisation,"* JHEP **11**, 043 (2013), arXiv:1306.6464 [hep-ph] .

[100] A. Alloul *et al.*, *"FeynRules 2.0 — A complete toolbox for tree-level phenomenology,"* Comput. Phys. Commun. **185**, 2250–2300 (2014).

[101] C. Degrande *et al.*, *"UFO – The Universal FeynRules Output,"* Comput. Phys. Commun. **183**, 1201–1214 (2012).

[102] *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, Tech. Rep. ATLAS-CONF-2016-024 (CERN, Geneva, 2016).

[103] ATLAS Collaboration, *"Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13$ *TeV,"* EPJ C **76** (2016), 10.1140/epjc/s10052-016-4120-y.

[104] *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, Tech. Rep. ATL-PHYS-PUB-2015-045 (CERN, Geneva, 2015).

[105] M. Cacciari, G. Salam, and G. Soyez, *"The anti-Kt jet clustering algorithm,"* JHEP **2008**, 063–063 (2008).

[106] ATLAS Collaboration, *"Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1,"* EPJ C **77** (2017), 10.1140/epjc/s10052-017-5004-5.

[107] M. Cacciari, G. Salam, and G. Soyez, *"FastJet user manual,"* EPJ C **72** (2012), 10.1140/epjc/s10052-012-1896-2.

[108] ATLAS Collaboration, *"Performance of pile-up mitigation techniques for jets in pp collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector,"* EPJ C **76** (2016), 10.1140/epjc/s10052-016-4395-z.

[109] B. Nachman *et al.*, *"Jets from jets: re-clustering as a tool for large radius jet reconstruction and grooming at the LHC,"* JHEP **2015** (2015), 10.1007/jhep02(2015)075.

[110] M. Aaboud *et al.*, *"Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV,"* JHEP **2018** (2018), 10.1007/jhep08(2018)089.

[111] *Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at $\sqrt{s}$= 13 TeV*, Tech. Rep. ATL-PHYS-PUB-2015-027 (CERN, Geneva, 2015).

[112] L. Coelho, "Study of the difference between substructure variables of large-R jets using AFII and FS simulations," https://cds.cern.ch/record/2740890.

[113] W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, Tech. Rep. ATL-SOFT-PROC-2012-065 (CERN, Geneva, 2012).

[114] ATLAS Collaboration, *"The ATLAS calorimeter simulation FastCaloSim,"* J. Phys. Conf. Ser. **331**, 032053 (2011).

[115] Jesse Thaler and Ken Van Tilburg, *"Identifying Boosted Objects with N-subjettiness,"* JHEP **03**, 015 (2011), arXiv:1011.2268 [hep-ph] .

[116] Daniele Bertolini, Tucker Chan, and Jesse Thaler, *"Jet Observables Without Jet Algorithms,"* JHEP **04**, 013 (2014), arXiv:1310.7584 [hep-ph] .

[117] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler, *"Energy Correlation Functions for Jet Substructure,"* JHEP **06**, 108 (2013), arXiv:1305.0007 [hep-ph] .

[118] Andrew J. Larkoski, Ian Moult, and Duff Neill, *"Power Counting to Better Jet Observables,"* JHEP **12**, 009 (2014), arXiv:1409.6298 [hep-ph] .

[119] Simone Marzani, Gregory Soyez, and Michael Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, Vol. 958 (Springer, 2019) arXiv:1901.10342 [hep-ph] .

[120] A. Hoecker *et al.*, *"TMVA - Toolkit for Multivariate Data Analysis,"* (2007), arXiv:physics/0703039 [physics.data-an] .

[121] J. Gunion and X. He, *"Determining the CP nature of a neutral Higgs boson at the LHC,"* Phys. Rev. Lett. **76**, 4468–4471 (1996), arXiv:hep-ph/9602226 .

[122] J. Richman, *An experimenter's guide to the helicity formalism*, Tech. Rep. CALT-68-1148 (Calif. Inst. Technol., Pasadena, CA, 1984).

[123] F. Demartin *et al.*, *"Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction,"* EPJ C **74**, 3065 (2014), arXiv:1407.5089 [hep-ph] .

[124] J. Ellis, D. Hwang, K. Sakurai, and M. Takeuchi, *"Disentangling Higgs-Top Couplings in Associated Production,"* JHEP **04**, 004 (2014), arXiv:1312.5736 [hep-ph] .

[125] S. Wilks, *"The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses,"* Annals Math. Statist. **9**, 60–62 (1938).

[126] G. Cowan *et al.*, *"Asymptotic formulae for likelihood-based tests of new physics,"* Eur. Phys. J. C **71**, 1554 (2011), [Erratum: Eur.Phys.J.C 73, 2501 (2013)], arXiv:1007.1727 [physics.data-an] .

[127] "TRexFitter framework twiki page," `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TtHFitter`.