RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

Daniel Isaac Narrias Villar

Search for Dark Matter produced in association with a Higgs boson decaying to $b\bar{b}$ using 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector

Dissertation

KIRCHHOFF-INSTITUT FÜR PHYSIK

Dissertation

submitted to the

Combined Faculties of the Natural Sciences and Mathematics

of the Ruperto-Carola-University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Daniel Isaac Narrias Villar

born in Antofagasta

Oral examination on February 6, 2018

Search for Dark Matter produced in association with a Higgs boson decaying to $b\bar{b}$ using 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector

Referees: Priv.-Doz. Dr. Oleg Brandt
Prof. Dr. André Schöning

Abstract

Cosmological and astrophysical observations indicate the existence of Dark Matter. Many beyond Standard Model theories predict associated production of Dark Matter particles with a Higgs boson at energies that can be probed at collider experiments. The signature of the presented analysis consists of a $b\bar{b}$ pair from a Higgs boson decay recoiling against missing transverse momentum from Dark Matter particles, $h(\rightarrow b\bar{b}) + E_T^{miss}$. The presented $h(\rightarrow b\bar{b}) + E_T^{miss}$ search is performed using proton-proton collisions at a centre-of-mass energy of 13 TeV with an integrated luminosity of 36.1 fb^{-1} recorded with the ATLAS detector at the Large Hadron Collider. Several new refined event selections were developed, which increase considerably the experimental sensitivity of the search to beyond Standard Model theories resulting in a $h(\rightarrow b\bar{b}) + E_T^{miss}$ signature. Discovery significance along with its gains from the new refined event selections are presented. The observed data are found to be in agreement with the Standard Model predictions. Upper limits on the production cross-section of $h + E_T^{miss}$ events times the $h \rightarrow b\bar{b}$ branching ratio in a two-Higgs-doublet model with an extra Z' boson scenario are presented at 95% confidence level. The improvement of the limits from the new refined event selections is quantified. A new topological algorithm is suggested for future analyses and its associated increase in experimental sensitivity is also presented.

Zusammenfassung

Kosmologische und astrophysikalische Beobachtungen weisen auf die Existenz Dunkler Materie hin. Viele Theorien jenseits des Standard Modells sagen eine assoziierte Produktion von Dunkle Materie Teilchen mit einem Higgs-Boson bei Energien voraus, die bei Beschleuniger-Experimenten untersucht werden können. Die Signatur der präsentierten Analyse besteht aus einem $b\bar{b}-$Paar als Zerfallprodukt eines Higgs-Bosons zusammen mit fehlendem transversalen Impuls verursacht durch Dunklen Materie: $h(\rightarrow b\bar{b}) + E_T^{miss}$. Die vorgestellte $h(\rightarrow b\bar{b}) + E_T^{miss}-$Suche wird auf einem Datensatz von Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 13 TeV mit einer integrierten Luminosität von 36.1 fb^{-1} durchgeführt. Der Datensatz wurde mit dem ATLAS-Detektor am Large Hadron Collider aufgezeichnet. Für die Analyse wurden mehrere neue Techniken zur Verbesserung der Ereignisselektion entwickelt, welche die experimentelle Sensitivität der Analyse wesentlich erhöhen. Diese neue verfeinerte Ereignisselektion und daraus resultierende Verbesserungen für das Signifikanzniveau werden vorgestellt. Die beobachteten Ergebnisse stimmen mit den Vorhersagen des Standardmodells überein. Für das Zwei-Higgs-Dublett-Modell mit einem zusätzlichen $Z'-$Boson werden obere Ausschlussgrenzen für den Wirkungsquerschnitt von $h(\rightarrow b\bar{b}) + E_T^{miss}-$Ereignissen multipliziert mit dem $h \rightarrow b\bar{b}-$Verzweigungsverhältnis bei einem Konfidenzniveau von 95% CL präsentiert. Die Verbesserung der Ausschlussgrenzen durch die neue verfeinerte Ereignisauswahl wird quantifiziert. Für zukünftige Analysen wird ein neuer topologischer Algorithmus vorgeschlagen und der damit verbundene Anstieg der experimentellen Sensitivität wird präsentiert.

Contents

1 Introduction

Our current knowledge of particle interactions at their most fundamental level relies on the Standard Model of Particle Physics (SM). The SM has successfully described most of the experimental data collected so far, establishing a cornerstone with the discovery of the Higgs boson h in 2012 by the ATLAS and CMS collaborations [1, 2] at the Large Hadron Collider (LHC). However, the SM does not describe all the observed physics phenomena. A fundamental open question is the particle nature of Dark Matter (DM) [3], for which the SM does not provide a viable candidate. Compelling *beyond* Standard Model (BSM) theories predict thermally created DM particles with production cross-sections at the electroweak scale, which can be probed at the LHC. Collider experiments are therefore of great interest in the search for BSM stable weakly interacting particles that can serve as DM candidates, accommodating the observed DM relic abundance [4, 5].

Since DM particles produced at colliders escape undetected the detectors, most collider-based searches for DM focus on the signature of missing transverse momentum recoiling against a SM particle X radiated off the initial state, as illustrated in Figure 1.1, top. This search signature is denoted as $X + E_T^{miss}$. The $X + E_T^{miss}$ signature has been searched for by the ATLAS and CMS experiments, where X represents a light quark or gluon [6, 7], a b- or t-quark [8, 9], a photon [10, 11], or a W or Z boson [12, 13, 14]. The initial state radiation of a Higgs boson h is on the other hand Yukawa-suppressed, which is why the $h + E_T^{miss}$ process is a direct probe of hard interactions with DM particles, as illustrated in Figure 1.1, bottom.

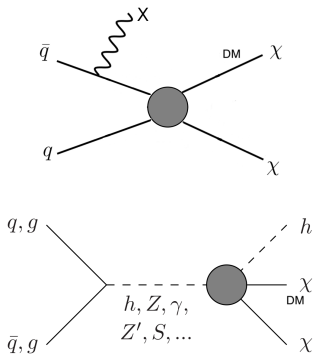

Figure 1.1: Diagrams of production of Dark Matter particles in association with an initial state radiated SM particle X (top) and a Higgs boson h (bottom).

In this thesis, a search for Dark Matter production in association with a Higgs boson decaying to a pair of b-quarks, $h \to b\bar{b}$, in proton-proton collisions at $\sqrt{s} = 13$ TeV using an integrated luminosity of 36.1 fb^{-1} recorded with the ATLAS detector is presented. These results [15], obtained from data taken in Run 2 of the LHC in 2015 and 2016, supersede the $h(\to b\bar{b}) + E_T^{miss}$ search results using 2015 data only, obtained at $\sqrt{s} = 13$ TeV using an integrated luminosity of 3.2 fb^{-1} [16]. The $h(\to b\bar{b}) + E_T^{miss}$ search substantially increases its sensitivity compared to previous results, thanks to the increase in integrated luminosity and to new refined analysis techniques, which correspond to the main contribution of the author to the presented analysis. As a result of this contribution, the reducible $t\bar{t}$ background process was considerably reduced, losing its status of dominant background. At the end of this chapter, the total contributions

of the author are briefly mentioned.

The structure of this thesis is as follows. The theoretical framework, namely the SM and simplified BSM theories used in the presented work, and observational evidence for DM, are introduced in Chapter 2. The LHC and the ATLAS detector are described in Chapter 3. Details about the Monte Carlo (MC) simulations and datasets used in the analysis are given in Chapter 4. The reconstruction of the physics objects in the analysis is discussed in Chapter 5. The event selection and the signal region (SR) of the analysis are introduced in Chapter 6. Details about the new refined event selections, including truth studies, are given in Chapter 7. The control regions (CRs) of the analysis are discussed in Chapter 8. The statistical formalism used for the interpretation of the analysis is introduced in Chapter 9. Systematic uncertainties along with the statistical fit model are introduced in Chapter 10. Postfit distributions, discovery significance and production cross-section limits on $Z' - 2HDM$ models are presented in Chapter 11 for the published $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis and for the analysis without *optimised event selections* using the same integrated luminosity of 36.1 fb^{-1}, in order to assess the impact of the optimised event selections on the sensitivity of the search. A new topological algorithm, the *Likelihood Ordering*, is presented in Chapter 12. Finally, a summary is presented in Chapter 13.

Author's contributions

The author's contributions to the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis are the design, implementation and performance evaluation of several new refined event selections. The author studied their impact on the background processes as well as on diverse BSM signal theories in the SR and CRs. He also assessed their discovery significance gain and determined the improvement on exclusion limits on production cross-section of $Z' - 2HDM$ models, which is the BSM theory used for interpretation of this analysis. These *optimised event selections* are the *b-jet veto*; the *tau veto*, which is based on a boosted-decision-tree tau identification provided by the ATLAS collaboration and on a new custom-built very loose tau identification to identify additional taus; the new H_T ratio event selection, which places a condition on the minimal H_T[1] *fraction* carried by the $h \rightarrow b\bar{b}$ decay; and the ΔR_{bb} event selection, which exploits the angular separation of the $b\bar{b}$ pair from the $h \rightarrow b\bar{b}$ decay. The author performed truth studies on background process composition to assess the flavour components on which the optimised event selections act. The author implemented the track-related tau uncertainties in the analysis framework. The author studied the impact of tau uncertainties on the $t\bar{t}$ normalisation, in order to determine how important tau uncertainties are in the statistical model and whether or not they are negligible. The author developed a new topological algorithm, the *Likelihood Ordering*, to further reduce $t\bar{t}$; its associated discovery significance gains are presented.

Prior to the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis, the author estimated the modelling uncertainties on the track-assisted jet mass from different fragmentation modelling in MC generators, by comparing the jet mass scale between the MC generators PYTHIA8 and HERWIG++ as a function of η, p_T and m/p_T of large-R jets. These results are now included in the stan-

[1] Sum of scalar p_T of jets in the event.

dard jet mass observable used in ATLAS and published in an ATLAS CONF note [17]. In his service task for ATLAS, the author worked on the simulation and implementation of the Parametrised Pile-Up Correction [18], which is used in the ATLAS Level-1 Calorimeter Trigger MC simulation and describes the dynamic pile-up correction made by the new Multi-Chip-Module commissioned for Run 2.

The analysis and work presented in this thesis rely in several places on the work made by other members of the ATLAS collaboration. For example, the reconstruction, identification and calibration of physics objects in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis rely on ATLAS performance studies and ATLAS measurements as well as on related ATLAS software. The unanalised MC samples used in the analysis were produced centrally by an ATLAS analysis framework group. The author did not perform trigger studies nor derived trigger scale factors for the analysis. Studies and the development of the statistical physics model used in the analysis were a common effort of many members in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis group without direct contribution of the author. By using the samples centrally provided by the analysis framework group, the author independently performed the analysis producing analised inputs for statistical interpretation of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis and performed then the statistical interpretation, validating his results against the results put forward by the $h(\rightarrow b\bar{b}) + E_T^{miss}$ group in Reference [15], in which the author is a member and co-author. The author produced all results and plots shown in this thesis, unless stated otherwise and/or cited to the proper sources.

2 Theoretical Framework

The Standard Model of Particle Physics (SM) describes the elementary particles as well as the fundamental forces in nature that govern their interactions. It was formulated in the 1960s and 1970s. During the following decades, several experiments were performed to test the SM, which showed a remarkable agreement with data. Its predictions have been confirmed by experimental discoveries such as, for example, the observation of the gluon in 1979 at DESY [19], of the W and Z bosons in 1983 at CERN [20, 21, 22], of the top quark in 1995 at Fermilab [23, 24], and most recently of the Higgs boson in 2012 at CERN [1, 2]. It is however believed that the SM is a low-energy effective approximation of a more complete theory, since it does not provide, for example, neither a mechanism for the matter-antimatter asymmetry in the Universe nor a viable candidate for Dark Matter.

This chapter proceeds as follows: the Standard Model is briefly introduced in Section 2.1. Next, the observational evidence for Dark Matter is introduced in Section 2.2. Finally, simplified Beyond Standard Model Theories, which predict the production of DM particles in association with a Higgs boson, are introduced in Section 2.3.

2.1 Standard Model of Particle Physics

The Standard Model of Particle Physics is a relativistic quantum field theory that describes the elementary particles of matter and their interactions through the electromagnetic, weak and strong forces from the symmetry group $\mathrm{SU(3)}_C \times \mathrm{SU(2)}_L \times \mathrm{U(1)}_Y$ [26]. The elementary particles in the SM are classified in fermions, gauge bosons and the Higgs bosons, having a spin of 1/2, 1 and 0, respectively. An overview of the particle content in the SM is shown in Figure 2.1. The fermions can be further separated into quarks, which interact through the strong force, and leptons, which do not. Leptons form three families of left-handed doublets $\mathrm{SU(2)}_L$, each consisting of a negatively-electrically-charged lepton and a neutral neutrino, and right-handed singlets only with charged leptons. The charged leptons are the electron (e^-), muon (μ^-) and tau lepton (τ^-), and their corresponding neutrinos are the electron- (ν_e), muon- (ν_μ) and tau-neutrino (ν_τ), respectively. Even though neutrinos have been proven to be massive through neutrino-oscillation experiments, they are assumed as massless in the SM.

The quarks form also three families of left-handed $\mathrm{SU(2)}_L$ doublets and right-handed singlets, each consisting of an *up-type* and a *down-type* quark. The up (u), charm (c) and top (t) quarks are up-type quarks and carry an electric charge of $+2/3$ while the down (d), strange (s) and bottom (b) quarks are down-type quarks and carry an electric charge of $-1/3$. Quarks interact through the weak and strong force; the strong charge is the colour charge and takes *red*, *green* and *blue* values. Only colour-neutral objects have been observed. Quarks bound into colour singlets forming hadrons, commonly in the form of

three-quarks bound states called baryons or quark/anti-quark bound states called mesons. Each fermion has a corresponding antiparticle, which has opposite electric charge, parity and colour but the same mass.

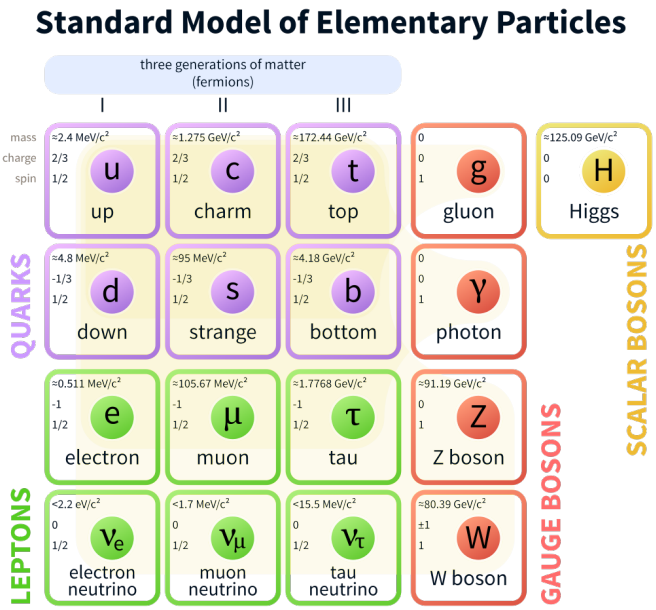

Figure 2.1: Particle content of the Standard Model. Their associated antiparticles not explicitly included. From Reference [25].

The forces in the SM are mediated by gauge bosons. The gauge bosons are a result of the local invariance of the SM Lagrangian density $\mathscr{L}$ under gauge transformations of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ group [26]. The electromagnetic and weak interactions are described by the unified electroweak theory with the symmetry group $SU(2)_L \times U(1)_Y$ [27]. The weak interaction is mediated by three gauge bosons, $W^{\pm}$ and Z, which become massive through the mechanism of *spontaneous symmetry breaking* (SSB) of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ group, the Higgs mechanism [27]. The electromagnetic interaction is mediated by a single massless gauge boson, the photon γ, which is a result of the SM Lagrangian's local invariance under the group $U(1)_Q$ after SSB. The strong interaction is described by *Quantum Chromodynamics* (QCD) with the symmetry group $SU(3)_C$. It is mediated by eight massless gluons, carrying colour and anticolour charge. $SU(2)_L$ and $SU(3)_C$ gauge groups are non-Abelian, which allows self-interactions of the gauge bosons, whereas $U(1)_Q$ gauge group is Abelian, which implies that photons cannot self-interact.

In the following, the Electroweak Theory, the Higgs Mechanism and the the Quantum Chromodynamics Theory are briefly introduced in Subsections 2.1.1, 2.1.2 and 2.1.3, respectively, following the discussions in [27].

2.1.1 The Electroweak Theory

The electroweak theory is a joint description of both electromagnetic and weak interactions. It has four massless gauge fields in the unbroken theory, corresponding to the generators of the group. The B_μ vector field is the generator of the $U(1)_Y$ group and couples to the weak hypercharge, defined as

$$Y = 2(Q - I_3), \tag{2.1}$$

where Q is the electric charge and I_3 is the third component of the weak isospin I_i.

The other three, denoted as W_μ^1, W_μ^2 and W_μ^3, are the generators of $SU(2)_L$ and couple to weak isospin. Weak isospin only acts on left-handed fermions and right-handed antifermions: right-handed fermions are singlets and left-handed fermions are described as doublets under $SU(2)_L$. The left-handed lepton doublets and left-handed quarks doublets are represented as

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L. \tag{2.2}$$

Since weak isospin does not couple to the mass eigenstates of the down-type quarks, the primed quarks are linear combinations of the mass eigenstates, according to the Cabibbo-Kobayashi-Maskawa (CKM) matrix [28, 29]. All fermions have right-handed singlets, except for neutrinos.

The electroweak SM Lagrangian written in terms of the fermion fields, the gauge boson fields and a Higgs doublet Φ is given by

$$\begin{aligned} \mathscr{L}_{EW} = & \sum_f \left(\bar{\psi}_L^f \gamma^\mu D_\mu \psi_L^f + \bar{\psi}_R^f \gamma^\mu D_\mu \psi_R^f \right) \\ & - \frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \\ & + \left(D_\mu \Phi \right)^\dagger \left(D^\mu \Phi \right) - \mu^2 \Phi^\dagger \Phi - \lambda \left(\Phi^\dagger \Phi \right)^2. \end{aligned} \tag{2.3}$$

where the covariant derivative D_μ takes the form

$$D_\mu \psi_L = \left(\partial_\mu + i \frac{g}{2} W_\mu^i \sigma^i + i \frac{g'}{2} B_\mu \right) \psi_L, \tag{2.4}$$

$$D_\mu \psi_R = \left(\partial_\mu + i \frac{g'}{2} B_\mu \right) \psi_R, \tag{2.5}$$

σ^i are the Pauli matrices related to the weak isospin as $I_i = \frac{\sigma_i}{2}$; g and g' are the coupling constants of $SU(2)_L$ and $U(1)_Y$, respectively; ψ_L and ψ_R are the left-handed and right-handed spinor components of the Dirac spinor ψ, respectively; and μ^2 and λ are the mass-like coefficient and the Higgs field quartic self-coupling of the Higgs potential, respectively.

The first line of the SM Lagrangian in Equation 2.3 contains the kinetic energy of fermions and their interaction with the gauge boson; the second line contains the kinetic energy and self-couplings of the gauge bosons, where $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ and $W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g \varepsilon_{ijk} W_\mu^j W_\nu^k$; and the third line is the Lagrangian of the Higgs field, which generates the spontaneous symmetry breaking.

Performing a rotation in the (W_μ^3,B_μ) space, the physical neutral gauge fields A_μ of the photon γ and Z_μ of the Z boson are obtained as follows:

$$A_\mu = \sin(\theta_W)W_\mu^3 + \cos(\theta_W)B_\mu, \tag{2.6}$$

$$Z_\mu = \cos(\theta_W)W_\mu^3 - \sin(\theta_W)B_\mu, \tag{2.7}$$

where the angle θ_W is the *weak mixing* or *Weinberg angle*, satisfying $g\sin(\theta_W) = g'\cos(\theta_W)$, which relates the weak isospin and weak hypercharge couplings. By using the gauge fields W_μ^1, W_μ^2, one defines a field W_μ that annihilates W^+ bosons and creates W^- bosons as

$$W_\mu \equiv \frac{W_\mu^1 - iW_\mu^2}{\sqrt{2}}. \tag{2.8}$$

The $\mathscr{L}_{EW}$ Lagrangian is invariant under $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge transformations. In this form, the Lagrangian contains no mass terms for fermions nor gauge boson. One needs a mechanism to provide fermions and the W and Z bosons with mass. Just adding a Dirac mass term, for example, $m\bar{\psi}\psi = m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$ for fermions, would violate the gauge invariance of $\mathscr{L}_{EW}$ under the symmetry group. The Higgs mechanism circumvents this problem through the SSB, giving mass to particles.

2.1.2 Higgs Mechanism of Spontaneous Symmetry Breaking

The Higgs doublet takes the form

$$\Phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix}, \tag{2.9}$$

where $\phi^+(x)$ is a charged complex scalar field, $\phi^0(x)$ is a neutral complex scalar field and x is the position four-vector.

The Higgs potential $V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda\left(\Phi^\dagger\Phi\right)^2$ in Equation 2.3 is bounded from below only if the coefficient λ of the quartic self-coupling of the Higgs field is positive. The mass-like coefficient μ^2 is assumed to be negative in order to realise the spontaneous breaking of the symmetry $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y \to \mathrm{U}(1)_Q$, where $\mathrm{U}(1)_Q$ is the gauge symmetry group of electromagnetic interactions, which remains unbroken, so the photon γ is massless.

Defining $v \equiv \sqrt{-\frac{\mu^2}{\lambda}}$, the Higgs potential takes the form

$$V(\Phi) = \lambda\left(\Phi^\dagger\Phi - \frac{v^2}{2}\right)^2 + \text{constant}, \tag{2.10}$$

so that the potential reaches its minimum for $\Phi^\dagger\Phi = \frac{v^2}{2}$.

At this minimum, the Higgs field has its vacuum state, which is the lowest energy state. The quantised excitations of each field above the vacuum correspond to particle states. This nonzero value in vacuum is called *vacuum expectation value* (VEV). To have an electrically neutral vacuum, one expresses the Higgs fields at the vacuum state as

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{2.11}$$

which breaks the $SU(2)_L \times U(1)_Y$ symmetry. However, using Equation 2.1,

$$Q\langle \Phi \rangle = \left(I_3 + \frac{Y}{2} \right) \langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ v \end{pmatrix} = 0. \tag{2.12}$$

The vacuum is therefore invariant under $U(1)_Q$ gauge transformations of the form $e^{i\theta Q}\langle \Phi \rangle = \langle \Phi \rangle$, which guarantees the existence of a massless gauge boson of the symmetry group $U(1)_Q$: the photon γ.

The spontaneous symmetry breaking of the SM gauge symmetry group is then denoted as

$$SU(3)_C \times SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \times U(1)_Q. \tag{2.13}$$

Using the freedom to choose a gauge, in the unitary gauge, the Higgs doublet in Equation 2.3 reads as

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{2.14}$$

where $H(x)$ is the physical Higgs boson, obtained by excitations of the neutral Higgs field above the vacuum.

In the unitary gauge, the Higgs Lagrangian in Equation 2.3 reads as

$$\begin{aligned} \mathscr{L}_{Higgs} = & \frac{1}{2}(\partial H)^2 - \lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4} H^4 + \frac{g^2 v^2}{4} W_\mu^\dagger W^\mu + \frac{g^2 v^2}{8\cos^2(\theta_W)} Z_\mu Z^\mu \\ & + \frac{g^2 v}{2} W_\mu^\dagger W^\mu H + \frac{g^2 v}{4\cos^2(\theta_W)} Z_\mu Z^\mu H \\ & + \frac{g^2}{4} W_\mu^\dagger W^\mu H^2 + \frac{g^2}{8\cos^2(\theta_W)} Z_\mu Z^\mu H^2. \end{aligned} \tag{2.15}$$

In the first line in Equation 2.15, the first term is the kinetic energy of the Higgs boson; the second term is the mass term for the Higgs boson, from which one obtains

$$m_H = \sqrt{2\lambda v^2} = \sqrt{-2\mu^2}, \tag{2.16}$$

where the value of μ^2 is not predicted by the SM, so it can be determined only from experiments. The last two terms in the first line are the mass terms for the W and Z gauge bosons. Their masses are then given in the SM by

$$m_W = \frac{gv}{2} \qquad m_Z = \frac{gv}{2\cos(\theta_W)}. \tag{2.17}$$

In the SM, fermions also acquire a mass via the Higgs mechanism, through the presence of Higgs-fermion Yukawa coupling terms in the SM Lagrangian. Details about how

fermions acquire mass after the spontaneous symmetry breaking of the Higgs doublet can be found in Reference [27].

2.1.3 Quantum Chromodynamics Theory

Quantum Chromodynamics is the theory describing carrying colour-charge objects, namely quarks and gluons, and follows the symmetry group $SU(3)_C$. It has eight massless gauge fields, called gluon fields G^i_μ, $i = 1,...,8$, which correspond to the generators of the group. The quark fields are represented as three component vectors of spinors for the three colour states *red*, *green* and *blue*: $\Psi^T = (\psi_r, \psi_g, \psi_b)$. Quantum Chromodynamics written in terms of the coloured fermion fields and the gluon gauge boson fields takes the form

$$\begin{aligned}\mathscr{L}_{QCD} &= \sum_q \left(i\bar{\Psi}_q \gamma^\mu D_\mu \Psi_q - m_q \bar{\Psi}_q \Psi_q\right) - \frac{1}{4} G^i_{\mu\nu} G^{\mu\nu}_i \\ &= \sum_q \bar{\psi}_{q,a} \left(i\gamma^\mu \partial_\mu \delta_{a,b} - \frac{g_s}{2}\gamma^\mu \lambda^i_{ab} G^i_\mu - m_q \delta_{a,b}\right) \psi_{q,b} - \frac{1}{4} G^i_{\mu\nu} G^{\mu\nu}_i .\end{aligned} \tag{2.18}$$

In the first line of Equation 2.18, the sum runs over the six different quark flavours q; the repeated index (following the Einstein summation convention) is $i = 1,...,8$ for the eight generators of $SU(3)_C$; $G^i_{\mu\nu}$ is the gluonic field tensor, where $G^i_{\mu\nu} = \partial_\mu G^i_\nu - \partial_\nu G^i_\mu - g_s f_{ijk} G^j_\mu G^k_\nu$; f_{ijk} are the structure constants of $SU(3)_C$; g_s is the QCD coupling constant, which is related to the strong coupling α_s as $\alpha_s = g_s^2/4\pi$. In the second line, the indices a, b are colour indices and take the values $1,2,3$, representing *red*, *green*, *blue*, respectively; λ^i are the 3×3 Gell-Mann matrices.

α_s depends on the interaction energy. Given a reference energy scale μ^2, the strong coupling after renormalisation takes the form

$$\alpha_s(q^2) = \frac{\alpha_s(\mu^2)}{1 + \frac{11N_C - 2N_f}{12\pi}\alpha_s(\mu^2)\ln\left(\frac{q^2}{\mu^2}\right)}, \tag{2.19}$$

where $N_f = 6$ is the number of quark flavours and $N_C = 3$ is the number of colours. $b_0 = \frac{11N_C - 2N_f}{12\pi}$ is positive: α_s decreases as the probed energy q^2 increases. This energy dependence is referred to as *running of* α_s and leads to the *asymptotic freedom* of quarks and gluons in the proton. With increasing energies and decreasing distances, the strong coupling decreases, so the strength of the strong force is reduced, in such a way that quarks and gluons asymptotically act as free particles. At an energy equal to the mass of the Z boson, $q^2 = m_Z^2$, the strong coupling takes the value of $\alpha_s(m_Z^2) = 0.1185 \pm 0.0006$ [30].

Another key feature of QCD is *colour confinement*. Confinement refers to the exclusive experimental observation of colourless objects, i.e. colour singlet objects carrying no colour charge. It can be understood from the self-interacting nature of gluons. Due to the self-attraction of gluons, the energy stored in the coloured field between two coloured

particles increases linearly with the distance. As the particles separate, it becomes energetically favourable to create a new $q\bar{q}$ pair out of the field energy, decreasing the energy of the system. The $q\bar{q}$ creation process stops when reaching the energy scale of hadronisation.

2.2 Dark Matter

The evidence for DM originates from astronomical observations. The first hints were found in the 1930s by studying the velocity dispersion of galaxies in the Coma cluster of galaxies [31]. In the last years, further astronomical observations have added to this evidence and set further constraints (e.g. weak [32] and strong [33] gravitational lensing, and Cosmic Microwave Background, CMB [34]). These observations indicate that DM is neither composed of baryons nor of any other known particle, and that DM amounts to about a quarter of the content of the Universe, being five times more prevalent than baryonic matter: baryonic matter amounts to about 4% of the Universe, while DM to about 25%. Despite the constraints set on DM properties by cosmological data and the advancing exclusion of phase space of BSM theories from collider experiments, the particle content of Dark Matter (DM) remains one of the major open questions of the contemporary cosmology, astrophysics and particle physics.

In the following, the Standard Cosmological model, the relic abundance of DM, and compelling evidence for DM at different astrophysical length scales will be briefly introduced in Subsections 2.2.1, 2.2.2 and 2.2.3, respectively, following the discussions in [35, 36, 37].

2.2.1 Standard Model of Cosmology

The Standard Model of Cosmology is theoretically based on Einstein's general theory of relativity. It postulates that the Universe in its origin was in a highly compressed and hot state and has been and is still expanding to present time. This is known as the *Big Bang Theory* is based on three cornerstones: the expansion of the Universe from the discovery of the Hubble's law; the fossil record of light elements synthesised in the early Universe in the nucleosynthesis; and the existence of a thermal radiation field, the Cosmic Microwave Background.

Einstein's equations of general relativity are given by

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -\frac{8\pi G_N}{c^4} T_{\mu\nu} + \Lambda g_{\mu\nu}, \tag{2.20}$$

where $R_{\mu\nu}$ and R are the Ricci tensor and Ricci scalar, respectively, $g_{\mu\nu}$ is the metric tensor, G_N is the Newton's constant, $T_{\mu\nu}$ is the energy-momentum tensor, and Λ is the cosmological constant. The Einstein's equations related the geometry of the Universe, in the left of the Equation 2.20, with its energy content described by the energy-momentum tensor, in the right of the equation.

The presence of the cosmological constant indicates that the space-time itself is a source of gravitational field even in the absence of matter. The cosmological constant then represents *vacuum energy*, which contributes to the total energy of the Universe.

From CMB observations, where temperature variations are less than 10^{-4} [38], one infers that the Universe is isotropic on large scales, while from galaxy surveys, one concludes that the Universe is homogeneously distributed at scales larger than ~ 100 Mpc. The latter means that spheres with diameters larger than ~ 100 Mpc centred in any place of the Universe should contain the same amount of matter. The isotropy and homogeneity of the Universe allows us to express the line element as

$$ds^2 = -c^2dt^2 + a(t)^2\left(\frac{dr^2}{1-kr^2} + r^2d\Omega^2\right), \tag{2.21}$$

where the constant k can take the values $k = -1, 0 + 1$, for which the spatial terms in Equation 2.20 reduce to a metric of an open, flat or closed Universe, respectively. For a given value of k, the parameter $a(t)$, called *scale factor*, defines a family of similar spaces: the Friedmann-Lemaitre-Robertson-Walker (FLRW) models. The scale factor depends on time and $\dot{a}/a > 0$ at the present time $t = t_0$, because of the expansion of the Universe.

Using this metric in the Einstein equations allows us to solve them, being one of its tensor components

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G_N}{3}\rho_{tot}, \tag{2.22}$$

which is the so-called Friedmann equation. The ρ_{tot} is the total average energy density of the Universe, which can be expressed as $\rho_{tot} = \rho_m + \rho_r + \rho_\Lambda$ if it gets contributions from matter ρ_m, radiation ρ_r and vacuum energy ρ_Λ.

Only relative changes of the scale factor are measurable. One defines then the Hubble parameter as

$$H(t) = \frac{\dot{a}(t)}{a(t)}, \tag{2.23}$$

which governs the local expansion of the Universe according to the Hubble's law. The Hubble parameter at the present time, $H_0 = H(t_0)$, called Hubble constant, is $H_0 = h \cdot 100\text{kms}^{-1}\text{Mpc}^{-1} = 67.8 \pm 0.9\text{kms}^{-1}\text{Mpc}^{-1}$ [39]. This means that the linear length scale of the Universe is presently being stretched by a fraction of approximately $2 \cdot 10^{-18}$ per second.

From the Friedmann equation 2.22, we see that the Universe is flat with $k = 0$ if the total energy density of the Universe is equal to the *critical density*, defined as

$$\rho_c \equiv \frac{3H^2}{8\pi G_N}. \tag{2.24}$$

The abundance of a substance i is commonly expressed in units of ρ_c, as $\Omega_i \equiv \rho_i/\rho_c$, and the abundance of the total energy in the Universe is denoted as Ω. Then, the Friedmann equation 2.22 can be written as

$$\Omega - 1 = \frac{k}{H^2 a^2}, \tag{2.25}$$

with $\Omega = \Omega_m + \Omega_r + \Omega_\Lambda$, where Ω_m, Ω_r and Ω_Λ refers to the abundance of matter, radiation and vacuum energy in the Universe. The value of Ω determines whether the Universe is open ($\Omega < 1$), flat ($\Omega = 1$) or closed ($\Omega > 1$).

2.2.2 Dark Matter Relic Abundance: WIMPs

DM could have been produced as a thermal relic of the Big Bang. In the early Universe, all particles are in thermal equilibrium as long as their interaction rates are larger than the expansion rate of the Universe. Under these conditions, the interaction $XX \leftrightarrow SM\ SM$ is in equilibrium, where *SM* denotes a SM particle, which is massless and in equilibrium with the photon bath. As the Universe cools down, if it reaches temperatures T lower than DM particle's mass m_X, DM becomes Boltzmann suppressed, dropping exponentially as $e^{-m_X/T}$. As the Universe expands, the probability for DM particles of finding one another to annihilate becomes negligible, so they asymptotically stop annihilating and their number asymptotically approaches a constant, which is their *thermal relic density* or *thermal relic abundance*. One says, the particles *freeze-out* or they *decouple*. After freeze-out, DM is no longer in equilibrium and interactions changing the number of DM particles become negligible, but elastic scattering processes mediating energy exchange between DM particle and other particles (e.g. $X\ SM \leftrightarrow X\ SM$) remain possible.

In the non-relativistic limit $T \ll m_X$, the thermal relic number density of DM particles n_X takes the form [35]

$$n_X \simeq \left(\frac{m_X T_f}{2\pi}\right)^{3/2} e^{m_X/T_f} \simeq \frac{T_f^2}{M_{Pl}\langle\sigma_A v\rangle}, \tag{2.26}$$

where σ_A is the total annihilation cross-section, v the velocity, $\langle ... \rangle$ denotes a thermal average, M_{Pl} is the Planck constant, and $T = T_f$ is the temperature at which the DM particles decouple (freeze-out).

The thermal relic abundance Ω_X of DM particles X at the present time is given by

$$\Omega_X h^2 = \frac{m_X n_X}{\rho_c} \simeq \frac{3 \cdot 10^{-27} \mathrm{cm}^3 s^{-1}}{\langle\sigma_A v\rangle}. \tag{2.27}$$

If one assumes a weakly interacting DM particle, on dimensional arguments the cross-section can be written as

$$\langle\sigma_A v\rangle = k \frac{g^4}{16\pi^2 m_X^2} \cdot B, \tag{2.28}$$

where $B = 1$ or $B = v^2$ for S-wave and P-wave annihilation, respectively, g is the weak hypercharge coupling constant and k parametrises deviations from this formula. With this parametrisation, the relic abundance Ω_X is determined as a function of m_X, given that DM interacts weakly.

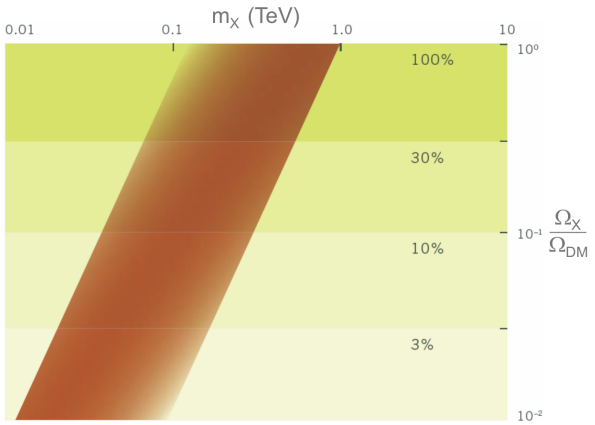

Figure 2.2: In dark orange, the band represents the possible mass values m_X that thermally created DM particles can take, given a relic density fraction Ω_X/Ω_{DM} of the total experimentally observed DM abundance in the Universe, which is $\Omega_{DM} \simeq 0.25$. From Reference [40].

In Figure 2.2, one sees the possible mass values that a weakly interacting and thermally created particle X can take, provided that it contributes in a fraction Ω_X/Ω_{DM} to the total DM density Ω_{DM}. The total DM density is observationally found to be $\Omega_{DM} \simeq 0.25$. If the thermally created particle X contributes to all of the DM density Ω_{DM}, it is predicted to have a mass m_X in the range $m_X \sim 100$ GeV $-$ 1 TeV. A particle that contributes to 10 % of DM (then still playing a role in the structure formation in the Universe) is predicted to have a mass range $m_X \sim 30$ GeV $-$ 300 GeV. This is the so-called *WIMP miracle*: weakly interacting particles with a mass at the weak scale make excellent DM candidates, as they naturally account for the observed relic DM density via thermal freeze-out. These DM candidates are called *WIMPs*.

2.2.3 Observational Evidence for Dark Matter

The observational evidence for DM is overwhelming. In the following, evidence for DM from rotation curves of galaxies, from gravitational lensing and from cosmic microwave background will be briefly summarised.

Rotation Curves of Galaxies

From the Virial theorem, the circular velocities of objects (stars and gas) orbiting the centre of a galaxy are expected to decrease with their distance from the galactic centre. These velocity distributions are called rotation curves of galaxies. In the 1970s, Ford and Rubin [42] observed that the rotation curves of galaxies become flat for large distances from the galactic centre. This means that orbiting gas and dust move at a constant velocity with increasing distances from the galactic centre. In Figure 2.3, the rotation curve of galaxy NGC 6503 is shown, where the flatness of the velocity profile is observed at large galactocentric distances.

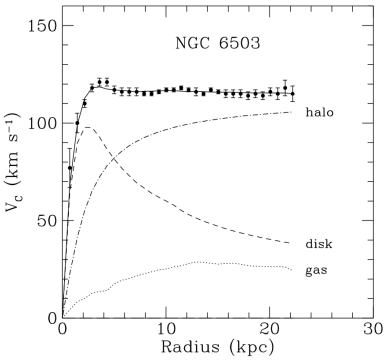

Figure 2.3: Dark-halo fit (solid curve) to the rotation curve of galaxy NGC 6503. The rotation curves of the individual components disk, gas and dark halo shown as a dashed, dotted and dash-dot curve, respectively. [41]

Baryonic matter consisting of gas and stellar objects in the galactic disk could not alone explain the galaxy rotation curves. It was then concluded that galaxies contain matter other than that expected from bright stellar objects in disks and gas. Without additional invisible mass, galaxies would lack of sufficient gravitational energy to keep outer material bound. By considering a dark halo consisting of unknown DM, it is possible to explain the rotation curves, reproducing the flatness of velocity profiles. The massive dark halo hypothesis is in agreement with data as seen in Figure 2.3.

Gravitational Lensing

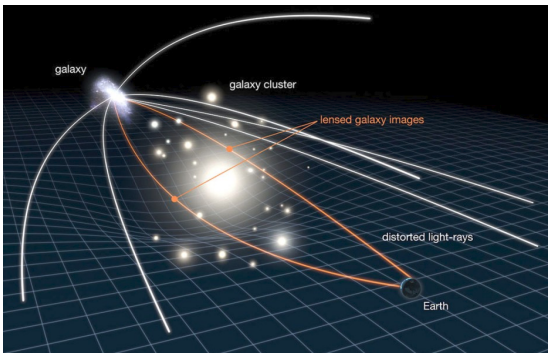

Figure 2.4: Gravitational Lensing.

From Einstein's General Relativity, the gravitational field of a massive object warps spacetime, modifying the geometry. The gravitational field will then cause light rays passing through it to bent and refocus, as illustrated in Figure 2.4. Dense concentrations of mass like the core of a galaxy or a cluster of galaxies can warp spacetime so that light can travel along multiple paths around the lens and still be deflected towards the observer. This is called *strong gravitational lensing*.

Strong gravitational lensing can appear as a luminous ring, *Einstein ring*, if a distant light source is directly behind a circular lens, so that light can travel around any side of it. If the source is slightly offset or the shape of the lens is complex, the source can appear in multiples locations, as seen in Figure 2.5. Each of these multiples images can be magnified or demagnified depending on the focussing of the light path. These effects are rare, which calls for other techniques based on gravitational lensing to study DM.

A foreground mass can be detected by studying the alignment of background luminous sources (e.g. galaxies) around the lensing mass. Gravitational lensing distorts the images of galaxies, by magnifying them and by stretching (*shearing*) them tangentially around the foreground mass. Luminous sources along adjacent lines of sight are coherently sheared by a similar amount, while their intrinsic shapes (for which one corrects by averaging over galaxies) are uncorrelated. This is called *weak gravitational lensing*. The shearing signal (shearing from a foreground mass in sight area) is statistically measured from shapes of

adjacent galaxies. The observable shear field on the sky can be converted into a map of the projected mass distribution (weak lensing mass contours).

Figure 2.6: Composed image showing the galaxy cluster *1E 0657-56*: most of matter (in blue) corresponds to DM, which is clearly separated from the baryonic matter: hot gas (in pink) and galaxies (in orange and white). From Reference [44].

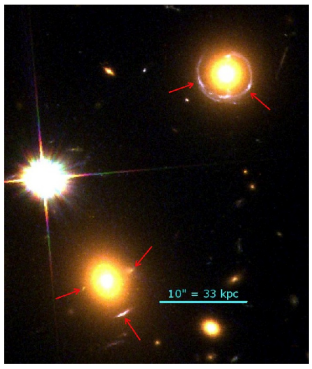

Figure 2.5: Two cases of strong lensing in the merging cluster of galaxies system A520. Strong lensing features marked by red arrows. [43]

Comparison of the matter distribution mapped using weak gravitational lensing with the distribution of baryonic matter using optical and X-ray data can reveal directly the existence of DM, its mass distribution and its interplay with the stellar and gas components. A prominent example is the *bullet cluster*, the galaxy cluster *1E 0657-56*, as seen in Figure 2.6. This cluster was formed as a result of the collision of two large clusters of galaxies. Baryonic matter in the image corresponds to galaxies, detected in optical spectra (in orange and white), and to hot gas, detected in X-rays (shown as two pink clumps). The blue clumps show where most of the mass in the clusters is found, using weak gravitational lensing to map the mass distribution. Most of the matter is clearly separated from the baryonic matter, giving direct evidence for DM. This result cannot be explained by modifying the laws of gravity.

Cosmic Microwave Background

At large temperatures in the early Universe, photons were constantly interacting with free electrons, so they could not travel freely long distances. As the Universe cooled down,

it became energetically favorable for electrons to combine with protons, forming neutral hydrogen atoms, resulting in the decoupling of matter and radiation, which is called *recombination*. Photons were then able to propagate freely and the Universe became *transparent*. The surface of last scattering of photons travelling freely after decoupling is called Cosmic Microwave Background, as seen in Figure 2.7. It follows the radiation of a black-body at $T = 2.72$ K.

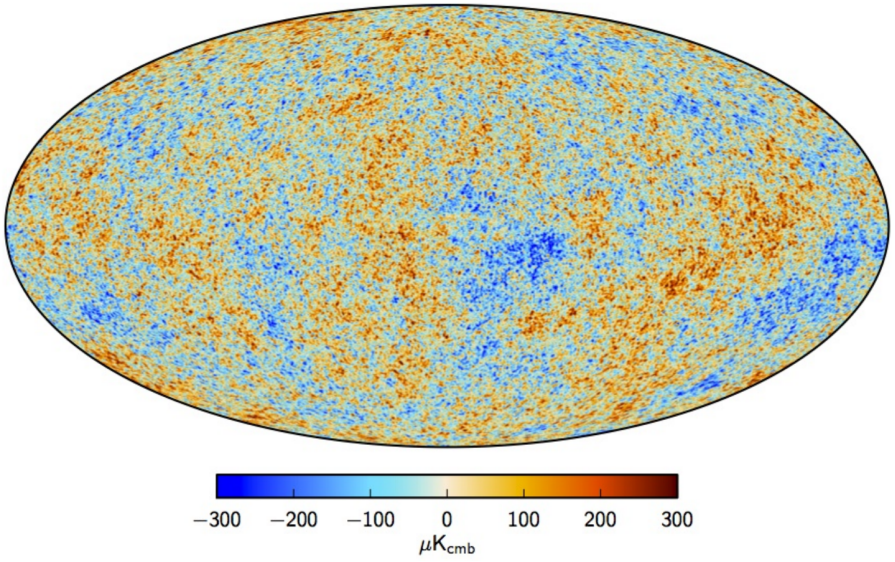

Figure 2.7: CMB intensity map from joint analysis of Planck Collaboration and WAMP. From Reference [38].

The standard theory of structure formation postulates that structure in the Universe, such as galaxies, clusters, voids and filaments formed from small initial fluctuations by gravitational instability. Since baryonic matter and radiation were coupled before recombination, these initial fluctuations must also be present in the CMB. When the Universe became matter-dominated, the small fluctuations in matter started to grow linearly in $a(t)$. Knowing that $\Delta T/T \lesssim 10^{-4}$ in CMB and that baryonic matter can only start clustering after recombination, there is simply not enough time for such small fluctuations in the CMB to explain the highly non-linear structures observed today just from baryonic matter. Pressureless matter fluctuations only can start growing when the Universe is matter-dominated, which motivates the existence of DM: the fluctuations of DM can start growing before recombination as DM does not interact with photons.

The study of CMB anisotropies allows for testing of cosmological models and puts constraints on cosmological parameters. The temperature anisotropies observed in CMB can be expanded using spherical harmonics. If the temperature fluctuations are assumed to be Gaussian, one can reduce the information in CMB maps into a temperature power spectrum. Fitting the power spectrum, one can retrieve constraints on cosmological parameters, as the peaks and their positions in the power spectrum depend on them. Some cosmological parameters as measured by the Planck Collaboration [39] at 95% CL are currently

$$w_b = 0.02229^{+0.00029}_{-0.00027}, \qquad \Omega_m = 0.308 \pm 0.012, \tag{2.29}$$

where $w_b = h^2\Omega_b$ with $\Omega_b = 0.04848^{+0.00143}_{-0.00141}$ the baryonic matter density and Ω_m is the density of the total matter in the Universe. The DM density Ω_{DM} in the Universe is

$$\Omega_{DM} = 0.2588 \pm 0.0072, \qquad [39], \text{Table 7.} \tag{2.30}$$

2.3 Simplified Beyond Standard Model Theories with Particle Dark Matter

In *simplified models* of BSM physics, DM couples to the SM through renormalisable interactions, by keeping the ultraviolet (UV) particles (produced at very high energies) as degrees of freedom in the theory, as opposed to the effective field theory (EFT) approach, where non-renormalisable operators are introduced without specifying the underlying UV physics. The EFT approach is not used throughout the thesis, only renormalisable simplified models are considered. In the following, three simplified models are introduced: $Z' - 2HDM$ model [45], Z' vector mediator model and S scalar mediator model [46] in Subsections 2.3.1, 2.3.2 and 2.3.3, respectively.

Throughout this section, one assumes that the DM particle is a singlet under $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$.

2.3.1 Z'-2HDM Models

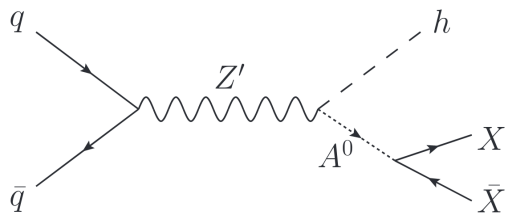

Figure 2.8: Z' couples to a two-Higgs doublet, where A^0 decays essentially to a pair of DM particles X [45].

This simplified model consists of a new Z' gauge boson, acting as a mediator, and of a two-Higgs doublet extension to the SM, with $Z' \to hA^0$ as seen in Figure 2.8, where A^0 is a heavy pseudoscalar, with a large branching ratio to DM particles. The gauge symmetry of the SM is extended by $\mathrm{U}(1)^{'}$, with a gauge boson Z'. This sector contains a SM singlet scalar that leads to spontaneous symmetry breaking of $\mathrm{U}(1)^{'}$, with a Z' mass at a scale above electroweak symmetry breaking. For the Higgs sectors, one assumes a Type 2 two-Higgs doublet model where the Higgs doublet Φ_u couples to up-type quarks while the other Higgs doublet Φ_d couples to down-quarks and leptons, as follows:

$$-\mathscr{L} \supset y_u Q\tilde{\Phi}_u \bar{u} + y_d Q\Phi_d \bar{d} + y_e L\Phi_d \bar{e} + h.c. \tag{2.31}$$

The Higgs doublets Φ_u, Φ_d have hypercharge $Y = 1/2$. In the present case, only u_R and Φ_u have charges under U(1)$^{'}$, with $Q^{'} = 1/2$ for both of them, whereas $Q^{'} = 0$ for Φ_d, Q_L and d_R.

After electroweak symmetry breaking, the Higgs doublets acquire v_u and v_d VEV and can be expressed in unitary gauge as

$$\begin{aligned}\Phi_u &= \frac{1}{\sqrt{2}}\begin{pmatrix} -\sin(\beta)H^+ \\ v_d - \sin(\alpha)h + \cos(\alpha)H - i\sin(\beta)A^0 \end{pmatrix} \\ \Phi_d &= \frac{1}{\sqrt{2}}\begin{pmatrix} \cos(\beta)H^+ \\ v_u + \cos(\alpha)h + \sin(\alpha)H + i\cos(\beta)A^0 \end{pmatrix},\end{aligned} \tag{2.32}$$

where h, H are neutral CP-even scalars, with h corresponding to the SM Higgs observed at $m_h \sim 125$ GeV, and A^0 is a neutral CP-odd scalar; the scalars H, A^0, $H^{\pm}$ are assumed to have mass around or larger than 300 GeV, according to $b \to s\gamma$ constraints [47]; $\tan(\beta) \equiv v_u/v_d$, and α is the mixing angle that diagonalises the $h-H$ mass matrix. From Higgs couplings constraints, Type 2 2HDM is constrained around the alignment limit, namely $\beta \to \alpha + \pi/2$, $\alpha \in (-\pi/2, 0)$. In this limit the h field has SM-like couplings to fermions and gauge bosons. From top Yukawa coupling constraints, $0.3 \lesssim \tan(\beta)$. Therefore, one chooses $\alpha = \beta - \pi/2$ and $0.3 \lesssim \tan(\beta)$.

The Higgs VEVs lead to $Z - Z'$ mass mixing. After diagonalising the gauge boson mass matrix, one considers a small mixing parameter ε, so that the decay width of $Z' \to hA^0$ can be expressed to leading order in ε as

$$\Gamma_{Z' \to hA^0} = (g_Z \cos(\alpha)\cos(\beta))^2 \frac{|p|}{24\pi}\frac{|p|^2}{m_{Z'}^2}, \tag{2.33}$$

where the centre of mass momentum for the decay products is $|p| = \frac{1}{2m_{Z'}^2}\lambda^{1/2}(m_{Z'}^2, m_h^2, m_{A^0}^2)$, with λ the Kaellen triangle function.

One assumes that the pseudoscalar A^0 of the theory possesses a large coupling to DM particles, with branching ratio of order one. DM particles can be fermion particles or scalar particles. For more details, see References [45, 48]. The DM particles in $Z' - 2HDM$ samples throughout the thesis are Dirac fermions.

2.3.2 Z' Vector Mediator Models

A Z' vector mediator is proposed in many new physics scenarios as a result of a new U(1)$^{'}$ symmetry [49], for example in extended gauge theories. Let us consider in the following only an extended gauge model. A gauge extension of the SM assumes that the baryon number B is gauged, with Z' being the gauge boson of U(1)$_B$. If the DM particle χ carries baryon number B_χ, the Lagrangian containing the $Z'-$quark-DM couplings takes the form

$$\mathscr{L}_{Z'-q-\chi} = g_q \bar{q}\gamma^\mu q Z'_\mu + \begin{cases} i g_\chi \chi^\dagger \overleftrightarrow{\partial^\mu} \chi Z'_\mu + g_\chi^2 |\chi|^2 Z'_\mu Z'_\mu & \text{scalar} \\ g_\chi \bar{\chi}\gamma^\mu \chi Z'_\mu & \text{fermion} \end{cases}, \tag{2.34}$$

where the first line and the second line correspond to a fermionic and to a scalar DM particle χ, respectively; g_q and g_χ correspond to the coupling of Z' to quarks and to DM, respectively. Assuming a leptophobic Z' model, these couplings are related to the U(1)$'$ gauge coupling g_B as $g_q = g_B/3$ and $g_\chi = B_\chi g_B$.

The minimal mechanism to generate the Z' mass is to introduce a *baryonic Higgs* scalar, resulting in the spontaneous symmetry breaking of U(1)$'$. After spontaneous symmetry breaking, a physical baryonic Higgs particle h_B remains, which couples to the Z' according to

$$\mathscr{L}_{HiggsB} \supset \frac{m_{Z'}^2}{2}\left(1+\frac{h_B}{v_B}\right)^2 Z'_\mu Z'^\mu, \tag{2.35}$$

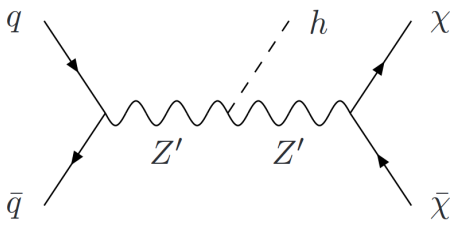

Figure 2.9: LO diagram of s-channel production of a pair of DM particles χ, including Z' boson mediation [46].

where v_B is the baryonic Higgs VEV. The mixing of h_B with the SM Higgs boson results in the coupling $hZ'Z'$ of the form

$$\mathscr{L} \supset -g_{hZ'Z'} h Z'_\mu Z'^\mu, \tag{2.36}$$

where $g_{hZ'Z'} = m_{Z'}^2 \sin(\theta)/v_B$ with θ being the $h - h_B$ mixing angle. The combination of Equations 2.34 and 2.36 results in $h(\to b\bar{b}) + E_T^{miss}$ signals at the LHC, as shown in Figure 2.9.

Unlike $Z' - 2HDM$ models, Z' vector mediator models are kinematically suppressed since Higgs boson radiation off a s-channel propagator leads the Z' vector mediator to being off-shell after the radiation.

2.3.3 Scalar Mediator Models

A new portal into the dark sector may be new scalar particles [50]. A possible realisation is to introduce a real scalar singlet S, with no gauge quantum numbers, which does not couple directly to quarks and leptons but rather through its mixing with the SM Higgs field. Due to gauge invariance, the scalar couples to the SM only through the Higgs field [51], with relevant terms in the scalar potential of the form

$$\begin{aligned} \mathscr{L}_{Higgs-S} &\supset -a|H|^2 S - b|H|^2 S^2 - \lambda |H|^4 \\ &\to -\frac{a}{2}(h+v)^2 S - \frac{b}{2}(h+v)^2 S^2 - \frac{\lambda}{4}(h+v)^4, \end{aligned} \tag{2.37}$$

where a, b are new physics couplings and λ the usual Higgs quartic coupling. The scalar S couples to DM particles χ as $\mathscr{L} \supset -y_\chi \bar{\chi}\chi S$.

The second line in Equation 2.37 takes this form after spontaneous symmetry breaking, leading to mixing terms in the $h-S$ mass matrix. The $h-S$ mass matrix can be diagonalised with a field rotation of the form: $h \to \cos(\theta)h + \sin(\theta)S$, $S \to \cos(\theta)S - \sin(\theta)h$, where the mixing angle θ is defined as $\sin(2\theta) = 2av/(m_S^2 - m_h^2)$. The mixing angle is constraint by current Higgs data, consistent with $\cos(\theta) = 1$ within $\mathcal{O}(10\%)$ uncertainties, so $\sin(\theta) \lesssim 0.4$ [52, 53, 54].

After the field rotation of the two scalar system $h-S$, the quark and DM Yukawa terms take the form

$$\begin{aligned}\mathscr{L} \supset & - y_\chi \bar{\chi}\chi \left[\cos(\theta)S - \sin(\theta)h\right] \\ & - \frac{m_q}{v}\bar{q}q\left[\cos(\theta)h + \sin(\theta)S\right].\end{aligned} \tag{2.38}$$

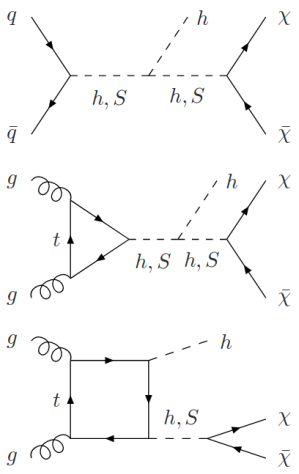

Figure 2.10: LO diagrams of collider production of a pair of DM particles χ, including S scalar mediation [46].

The combination of terms in the Equations 2.37 and 2.38 results in $h(\to b\bar{b}) + E_T^{miss}$ signals as shown in Figure 2.10, where the first two processes depend on h^2S and hS^2 cubic terms in the mediators.

Unlike $Z' - 2HDM$ models, scalar mediator models are kinematically suppressed since Higgs boson radiation off a s-channel propagator leads the scalar mediator to being off-shell after the radiation.

2.3.4 Cross-Section Scaling of Z'-2HDM Models

The phase space of model parameters to study can be greatly reduced if one identifies the parameters whose variation affects neither the kinematics nor the acceptance of the model, as opposed to the parameters on which the kinematics depend. These parameters can be then fixed without loss of generality, since the physics at any phase space parameter point can be obtained by scaling the production cross-section. Therefore, one scans only on the kinematics-dependent parameters, fixing the parameters that affect only the cross-section.

Additionally to the parameter constraints mentioned in Subsection 2.3.1 on Z'-2HDM models, one assumes $m_X < m_{A^0}/2$, so that A^0 is produced on-shell. $\tan(\beta)$ does not influence the kinematics, affecting only the cross-section. Under all these assumptions, the only parameters on which the kinematics depend are the mass of Z', $m_{Z'}$, and the mass of the pseudoscalar Higgs A^0, m_{A^0}.

Using Equation 2.33, the cross-section scaling formula takes the form

$$\sigma_{Z' \to hA^0} = \sigma_0 \cdot \left(\frac{g_Z \cos(\alpha)\cos(\beta)}{g_{Z,0}\cos(\alpha_0)\cos(\beta_0)}\right)^2. \tag{2.39}$$

3 The ATLAS Experiment at the Large Hadron Collider

The $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis was performed using proton-proton collision data recorded by the ATLAS[1] detector at the Large Hadron Collider (LHC) in CERN[2], Geneva. This chapter provides a brief overview on the LHC in Section 3.1, following Reference [55], and on the ATLAS experiment in Section 3.2, following Reference [56].

3.1 The Large Hadron Collider

The Large Hadron Collider is currently the world's largest and most powerful particle accelerator, as it has reached a centre-of-mass energy of $\sqrt{s} = 13$ TeV for proton-proton collisions. The LHC is built in the tunnel of the former LEP[3] experiment. The tunnel has a circumference of 27 km, lying between 45 m and 170 m underground. The LHC is designed to accelerate protons and heavy ions (e.g. lead ions), which are brought into collision at four locations corresponding to four experiments: ATLAS, CMS[4], LHCb[5] and ALICE[6]. ATLAS and CMS are multi-purpose detectors designed to perform precision measurements of the Standard Model (SM) and to search for new physics beyond the Standard Model (BSM) using proton-proton collision data. LHCb is a forward detector that specialises on measuring rare decays of charm and bottom hadrons and studying CP violation using proton-proton collision data. ALICE is a detector that specialises on studying the formation of quark-gluon plasma at extreme energy densities and temperatures in heavy hadron collisions. Proton beams are accelerated by the LHC at a design centre-of-mass energy of $\sqrt{s} = 14$ TeV reaching a design instantaneous luminosity of 10^{34} cm^{-2}s^{-1}, while heavy ion beams are accelerated at a design energy of 2.8 TeV per lead-ion nucleon reaching a design instantaneous luminosity of 10^{27} cm^{-2}s^{-1}.

The protons are brought to the final collision energy by an acceleration chain, reaching intermediate nominal energies at consecutive acceleration stages, as illustrated in Figure 3.1. These consecutive acceleration stages correspond to the linear accelerator LINAC2, the Booster, the Proton Synchrotron (PS) and the Super Proton Synchrotron, before the protons are injected into the LHC to be accelerated to a final nominal centre-of-mass energy of 14 TeV. The protons reach at each mentioned acceleration stage an

[1] ATLAS: *A Toroidal LHC Apparatus*.
[2] CERN: *Conseil Européen pour la Recherche Nucléaire*.
[3] LEP: *Large Electron Positron collider*.
[4] CMS: *Compact Muon Solenoid*.
[5] LHCb: *Large Hadron Collider beaty*
[6] ALICE: *A Large Ion Collider Experiment*.

intermediate nominal energy of 50 MeV, 1.4 GeV, 26 GeV and 450 GeV, respectively.

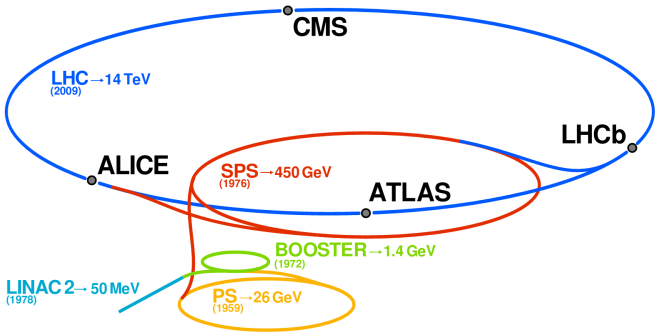

Figure 3.1: LHC, its four main experiments (ATLAS, CMS, LHCb, ALICE) and the intermediate accelerators (LINAC2, Booster, PS, SPS) of the acceleration chain. The nominal energies reached at each acceleration stage and the year of first operation are also shown. Figure from Reference [57].

Two beam pipes in the LHC host the counter-rotating proton beams, which rotate through the LHC ring that consists of eight straight sections connected by eight curved sections. In a long straight section, proton beams are accelerated by eight superconducting radiofrequency cavities at a voltage oscillation frequency of 400 MHz, each cavity providing 2 MeV per proton [58]. The curved sections are equipped with superconducting dipole magnets that provide a magnetic field of 8.3 T to keep the proton beams in the ring-trajectory. At the remaining sections, superconducting quadrupole and octopole magnets focus the proton beams in vertical and horizontal directions, reaching the strongest focussing at the four interaction points. The superconducting magnets are operated at a temperature of 1.9 K, using super-fluid helium for cooling. Each proton beam is structured into 3500 bunches, of which 2808 bunches are filled with about 10^{11} protons. Each bunch is separated at a nominal time of 25 ns from the following one, resulting in a nominal bunch-crossing rate of 40 MHz. At each bunch-crossing, several proton-proton collisions can take place.

The production rate N of a final state X is given by

$$N = L_{inst} \cdot \sigma_{pp \to X} \tag{3.1}$$

where the cross-section $\sigma_{pp \to X}$ quantifies the probability of the process $pp \to X$ to occur in a proton-proton collision and the instantaneous luminosity L_{inst} quantifies how many proton-proton interactions occur per area and time. The luminosity integrated over a specific period of time is referred to as integrated luminosity.

In this thesis, proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV are analysed using an integrated luminosity of 36.1 fb^{-1} recorded with the ATLAS detector in 2015 and 2016.

3.2 The ATLAS Experiment

The ATLAS experiment is a multi-purpose experiment with a wide physics program, covering precision measurements of SM processes, the discovery of the Higgs boson achieved in 2012 by the ATLAS and CMS collaborations [1, 2] and the search for various BSM processes such as new heavy gauge boson production, production of supersymmetry particles, particle dark matter production, extra dimensions, etc. The design of the ATLAS detector relies on the following requirements:

- A precise fine-granularity tracking system with high track momentum resolution and vertex reconstruction capability.
- Fine-granularity calorimeters with large coverage for energy measurements and particle identification.
- A muon system for efficient reconstruction and identification of muons.
- Highly efficient and fast trigger system and data acquisition system to reduce event rates for permanent data storage.

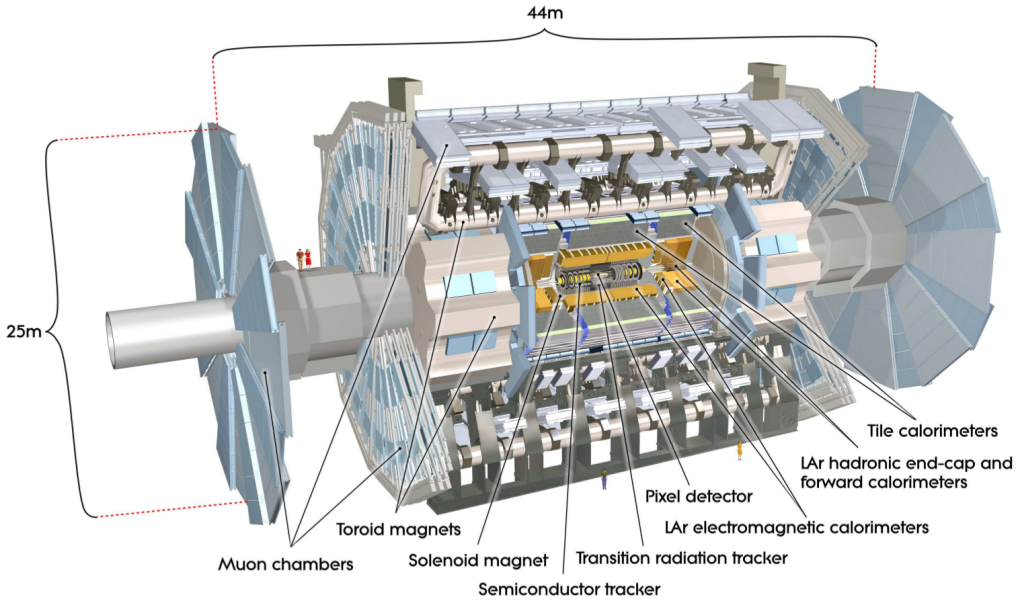

Figure 3.2: Overview of the ATLAS detector. Figure from Reference [56].

Following these requirements, the ATLAS detector and its subsystems were built until its installation in 2008. Figure 3.2 shows an overview of the ATLAS detector. The ATLAS detector consists of three major subsystems, arranged concentrically around the beam axis: the *inner detector*, the *electromagnetic* and *hadronic calorimeters*, and the *muon spectrometers*. For the functionality of the inner detector and the muon spectrometers, a magnetic system is needed in order to bend the trajectories of charged particles,

allowing momentum measurements. A superconducting solenoid magnet between the inner detector and calorimeters provides a 2 T magnetic field for the inner detector. A barrel toroid magnet and two endcap toroid magnets placed symmetrically around the beam axis provide 0.5 T and 1 T for the muon spectrometers, respectively.

The reference coordinate of the ATLAS detector is a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector. The z-axis is defined along the beam pipe. The x-axis points from the IP toward the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r,ϕ) are used in the transverse plane, with ϕ being the azimuthal angle around the z-axis. The polar angle θ is defined from the beam line. The pseudorapidity is defined in terms of θ as $\eta = -\ln\tan(\theta/2)$. The distance between two objects in the $\eta - \phi$ space is $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. The transverse momentum is defined by $p_T = p\ \sin(\theta)$ in the transverse plane.

In the following, the inner detector, the calorimeters, the muon spectrometer and the trigger system are briefly summarised in Subsections 3.2.1, 3.2.2, 3.2.3 and 3.2.4, respectively.

3.2.1 Inner Detector

The inner detector (ID) is designed to precisely measure tracks of charged particles within $|\eta| < 2.5$. As it is immersed in a 2 T solenoidal magnetic field, the trajectories are curved, allowing to measure the momentum of charged particles. By extrapolating the tracks, the inner detector reconstructs the primary vertex and also secondary vertices, which originate from the decay of long-lived particles, such as B-mesons. The inner detector provides electron identification for electrons with 0.5 GeV $< p_T <$ 150 GeV within $|\eta| < 2.0$. The inner detector consists of three subsystems, as illustrated in Figure 3.3: the *silicon pixel detector*, the *silicon microstrip tracker* and the *transition radiation tracker*. The total length of the inner detector is about 6.2 m and its diameter is about 2.1 m.

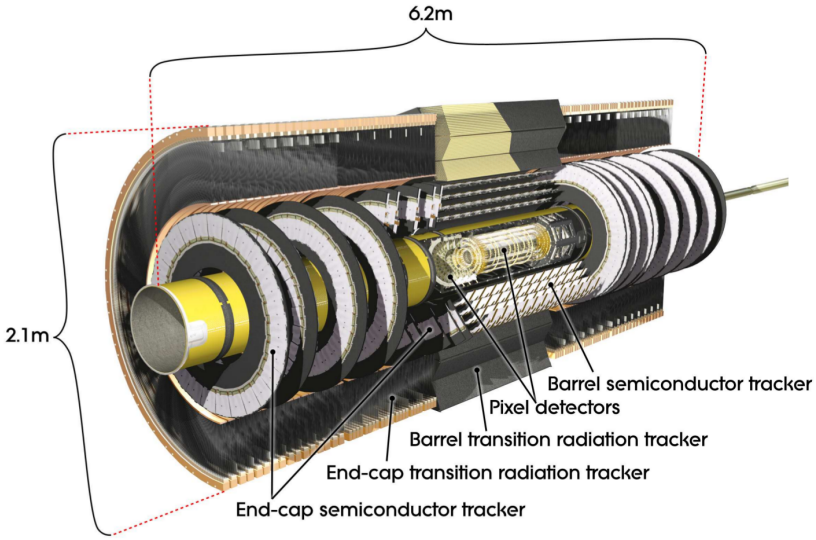

Figure 3.3: Overview of the ATLAS inner detector. Figure from Reference [56].

The silicon pixel detector is the innermost subsystem and is structured in three cylindrical layers in the barrel and three discs as end-caps on either side. It comprises about 80.4 million pixels with a minimum pixel size of 50×400 μm^2, which have an intrinsic accuracy of 10 μm in the $R-\phi$ transverse plane and of 115 μm in the $z-$direction ($R-$direction) for the barrel (for the discs). About 80.4 million channels are necessary for readout.

The silicon microstrip detector (or semiconductor tracker, SCT) is the intermediate subsystem and is structured in four stereo layers of silicon microstrips in the barrel and nine discs as end-caps on either side. Each layer consists of two layers of silicon microstrips with a mean pitch width[7] of about 80 μm. The two layers are arranged back-to-back with an angle of 40 mrad between them, allowing the determination of the coordinates ϕ and z (ϕ and R) in the barrel (end-caps). The SCT provides an intrinsic accuracy of 17 μm in $R-\phi$ and of 580 μm in the $z-$direction ($R-$direction) for the barrel (for the discs). About 6.3 million channels are necessary for readout.

The transition radiation tracker (TRT) is the outermost subsystem and provides measurements within $|\eta| < 2.0$. It consists of drift tubes with a diameter of 4 mm and a length of 144 cm in the barrel and of 37 cm in the end-caps. In the barrel, the drift tubes are arranged in parallel to the beam axis while in the end-caps they point radially to it. The tubes are filled with a gas mixture of 70% xenon, 27% carbon dioxide and 3% oxygen. The produced charges are collected on a gold-plated tungsten wire acting as anode. The space between drift tubes is filled with polypropylene material, which causes particles to emit low energy transition photons, detected then by the drift tubes. Since the transition radiation is proportional to $1/m$, where m is the mass of the particle, light particles such as electrons produce more transition radiation than charged hadrons, allowing electrons to be separately identified.

The momentum resolution of the inner detector is measured to be $\frac{\sigma_{p_T}}{p_T} = (4.83 \pm 0.16) \times 10^{-4}\text{GeV}^{-1} \times p_T$ [59].

3.2.2 Calorimeters

The calorimeters are designed to measure the energy of traversing particles and the position of the resulting particle showers within $|\eta| < 4.9$. Additionally, the calorimeters determine the missing transverse energy E_T^{miss}, originating from undetected particles escaping the detector, and use particle showering information for particle identification. A good spacial resolution is achieved by segmenting the calorimeters in many small cells, whose deposited energy can be individually read out. The calorimeters have a large radial depth to contain the full particle showers for precise energy measurements, providing an almost full coverage of the solid angle. The calorimeters are structured around the beam axis in calorimeter subsystems, as illustrated in Figure 3.4. The innermost subsystem is the *electromagnetic calorimeter* (EMCal), optimised to measure electromagnetic showers induced by electrons and photons. The outermost calorimeter is the *hadronic calorimeter* (HADCal), optimised to measure hadronic showers induced by hadronic particles. In

[7] *Pitch* is the distance from strip to strip.

addition, a *forward calorimeter* (FCal) is used for the forward region.

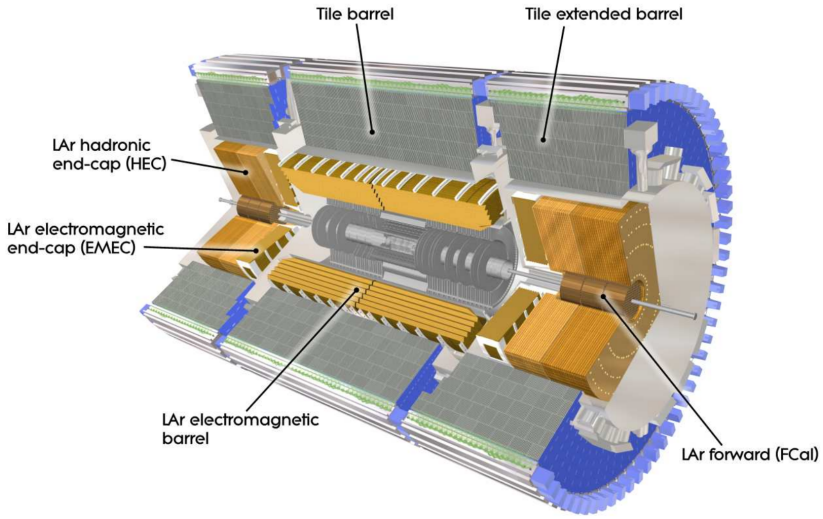

Figure 3.4: Overview of the ATLAS calorimeter system. Figure from Reference [56].

All ATLAS calorimeters are sampling calorimeters, consisting of alternating layers of active and absorber materials. Incident particles interact with the absorber material, inducing a cascade of particles that results in particle showers. The deposited energy of the cascade particles is measured in the active material. The EMCal uses liquid argon (LAr) as active material and lead as absorber in the barrel within $|\eta| < 1.475$ and in the endcaps within $1.375 < |\eta| < 3.2$. The barrel HADCal uses polystyrene-based scintillators (scintillating tiles) as active material and steel as absorber within $|\eta| < 1.7$. The endcap HADCals use LAr as active material and copper as absorber within $1.5 < |\eta| < 3.2$. The electromagnetic (hadronic) FCal uses LAr as active material and copper (tungsten) as absorber within $3.1 < |\eta| < 4.9$.

Electromagnetic Calorimeter

The barrel EMCal consists of two half half-barrels for positive and negative $z-$direction each. The thickness in the barrel depends on $|\eta|$, ranging from 22 to 33 radiation lengths[8] X_0. The absorber lead plates have an accordion shape and the active LAr medium fills the space between plates. The barrel EMCal is longitudinally divided into three layers with different granularity, as illustrated in Figure 3.5, left. The inner layer consists of strip cells with a size of 0.003×0.1 in $\Delta\eta \times \Delta\phi$, whose fine granularity in η allows discrimination between photon and neutral pions decaying to a pair of photons. The inner layer has a thickness of 4.3 X_0. The middle layer is designed to absorb most of deposited energy of traversing particles, so its thickness is 16 X_0. It consists of square cells with a size of 0.025×0.0245 in $\Delta\eta \times \Delta\phi$. The outer layer consists of coarser cells with a size of 0.05×0.0245 in $\Delta\eta \times \Delta\phi$ and has a thickness of 2 X_0.

[8] The radiation length X_0 is defined as the average distance after which the energy of a traversing particle is reduced by a factor of $1/e$ due to electromagnetic interactions with the detector.

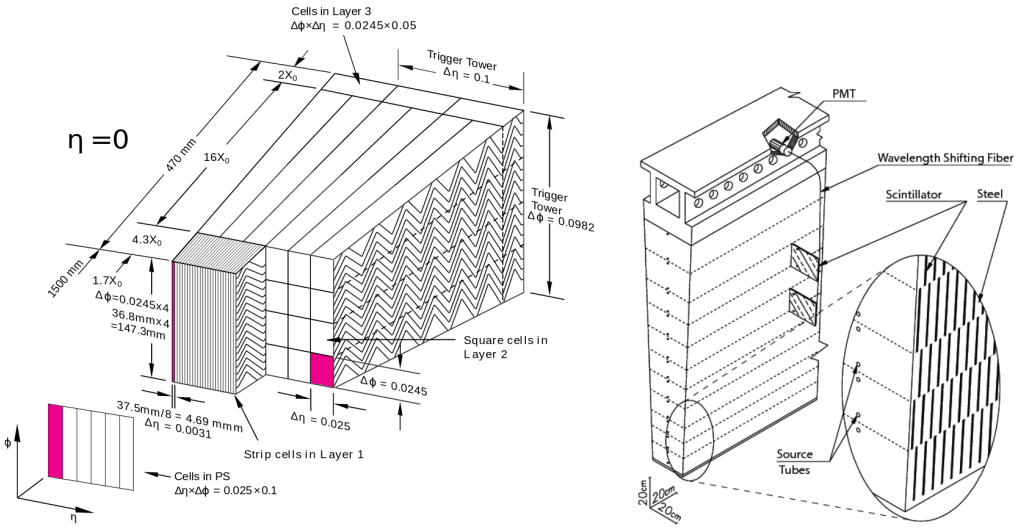

Figure 3.5: Left: Structure of a module in the electromagnetic barrel calorimeter. Right: Structure of a tile module in the hadronic calorimeter. Figure from Reference [56].

The endcap EMCal consists of two co-axial wheels around the beam line, covering $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$ each. Each wheel is divided into eight modules with cell granularity up to 0.1×0.1 in $\Delta\eta \times \Delta\phi$. In front of the EMCals in the barrel and in the endcaps, a LAr presampler layer is installed to correct for energy loss in the upstream material (e.g. ID, solenoid magnet, etc).

The design energy resolution of the EMCal is $\frac{\sigma_E}{E[\text{GeV}]} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$ [56].

Hadronic Calorimeter

The barrel HADCal is divided into a central barrel within $|\eta| < 1.0$ and two extended barrels on each side within $0.8 < |\eta| < 1.7$, which in turn are segmented into three concentric layers. The three concentric layers have interaction lengths of 1.5 λ[9], 4.1 λ and 1.8 λ in the central barrel, and of 1.5 λ, 2.6 λ and 3.3 λ in the extended barrel, respectively. Each barrel calorimeter subsystem consists of 64 modules, of which the structure of a representative module with its optical readout is illustrated in Figure 3.5, right. The scintillation light produced by traversing particles is read out by wavelength-shifting fibres connected to photomultiplier tubes (PMTs). An approximate granularity of 0.1×0.1 in $\Delta\eta \times \Delta\phi$ is achieved by grouping the readout fibres of multiple tiles on the same PMT.

The endcap HADCal consists of two adjacent wheels (front- and rear-wheel) on each side. The granularity is 0.1×0.1 in $\Delta\eta \times \Delta\phi$ within $1.5 < |\eta| < 2.5$ and 0.2×0.2 in $\Delta\eta \times \Delta\phi$ within $2.5 < |\eta| < 3.2$.

The design energy resolution of the HADCal is $\frac{\sigma_E}{E[\text{GeV}]} = \frac{50\%}{\sqrt{E}} \oplus 3\%$ [56].

[9]The interaction length λ is defined as the average distance after which the energy of a traversing particle is reduced by a factor of $1/e$ due to electromagnetic *and* strong interactions with the material detector.

Forward Calorimeter

The FCals are arranged between the beam pipe and the endcaps, consisting of the electromagnetic FCal layer and two hadronic FCal layers on each side. The FCals have a thickness of about 10 λ. The design energy resolution is $\frac{\sigma_E}{E[\text{GeV}]} = \frac{100\%}{\sqrt{E}} \oplus 10\%$ [56].

3.2.3 Muon Spectrometers

The outermost subdetector at ATLAS is the muon spectrometer, which measures the trajectories and momentum of charged particles within $|\eta| < 2.7$. It consists of four tracking detectors: *monitored drift tube chambers* (MDT), *cathode strip chambers* (CSC), *resistive plate chambers* (RPC) and *thin gap chambers* (TGC). In the barrel, the muon chambers are placed on and between the toroid coils in three cylindrical layers around the beam axis, while in the endcaps the chambers are installed in front and behind the endcap toroids in three wheels perpendicular to the beam. High precision tracking is provided by the MDTs within $|\eta| < 2.7$, except in the innermost endcap layer where MDT coverage is limited to $|\eta| < 2.0$ and the CSCs are used for this purpose within $2.0 < |\eta| < 2.7$. Fast tracking for muon triggering and supplementary tracking information is provided in the barrel by RPCs within $|\eta| < 1.05$ while in the endcaps TGCs are used for this purpose within $1.05 < |\eta| < 2.4$. In Figure 3.6, an overview of the ATLAS muon system is shown.

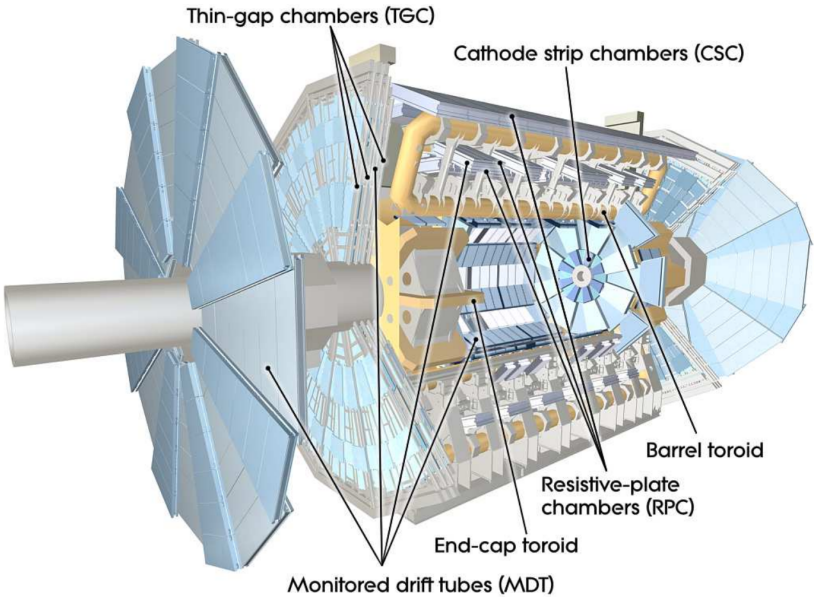

Figure 3.6: Overview of the ATLAS muon system. Figure from Reference [56].

MDT chambers consist of three to eight layers of drift tubes. Each tube is operated with a gas mixture of gas-carbon dioxide at a pressure of 3 bar and has a tungsten-rhenium anode wire in the centre. For a MDT chamber, the position resolution in the plane perpendicular to the tube is 35 μm in $z-$direction, with a drift time less than 700 ns. CSC chambers are multiwire proportional chambers also filled with an argon-carbon dioxide

gas mixture, with wires oriented in the radial direction. For a CSC chamber, the position resolution is 40 μm in $R-$direction (orthogonal to the magnetic field) and 5 mm in $\phi-$direction (parallel to it), with a drift time less than 40 ns and an intrinsic timing resolution of 7 ns.

A RPC chambers consists of two parallel electrode-plates, separated by a gap of 2 mm with an electric field of 4.9 kV/mm. A RPC chamber is operated in avalanche mode, with a position resolution of 10 mm in $z-$direction and of 10 mm in $\phi-$direction and an intrinsic timing resolution of 1.5 ns. The TGC chambers are multiwire proportional chambers filled with a highly quenching gas mixture of carbone dioxide and n-pentane, operated in a quasi-saturated mode due to having the wire-to-wire distance larger than the the wire-to-cathode distance. For a TGC chamber, the position resolution is between 2 mm and 6 mm in $R-$direction and between 3 mm and 7 mm in $\phi-$direction, with an intrinsic timing resolution of 4 ns.

The design momentum resolution of the muon system is $\frac{\sigma_{p_T}}{p_T[\text{GeV}]} = 10\%$ at $p_T = 1$ TeV.

3.2.4 Trigger System

The high event rate at the LHC of 40 MHz and the event size of several MB make it impossible for current computing and storage technology to process and store all events. Additionally, less than one percent of all physics events at the LHC are truly interesting physics processes. Therefore, a *trigger system* is fundamental to reduce the event rates for mass storage and to select only possibly interesting physics events for posterior analysis.

The trigger system in ATLAS is structured in two consecutive levels of event rate reduction: the *level 1 trigger system* (L1 trigger) and the *high level trigger* (HLT). The L1 trigger is based on custom-made hardware, which receives coarse-granularity information from the calorimeters and muon system at 40 MHz. After reconstructing high p_T electrons, photons, muons, jets and E_T^{miss} at the trigger level (trigger objects), the central trigger processor (CTP) takes a decision within 2.5 μs. If an event is accepted by the CTP, a L1 accept (L1A) signal is sent to all ATLAS subdetectors, which initiates the full granularity readout of the subdetectors data in regions of interest (RoI) identified at the L1 trigger. The nominal L1A rate output is 100 KHz, corresponding to a rate reduction factor of 400. The HLT is based on software algorithms, which analyse the full granularity data of the RoIs. If an event is selected by the HLT, it is fully reconstructed and transmitted to mass storage for physics analysis. The nominal HLT rate output is 1 KHz, corresponding to a total rate reduction factor of 40000.

After commissioning in 2017, a new *topological trigger* (L1Topo) system was integrated in the L1 trigger, performing algorithms based on topological or kinematic selections on trigger objects. It was not used in the 2015 and 2016 dataset used in this analysis. A new *fast tracker* (FTK) system will provide global ID track reconstruction at the L1 trigger rate. At the time of writing, the FTK is still in process of commissioning [60].

As the operating instantaneous luminosity increases, the event rate of an individual trigger is kept to design values by increasing the trigger threshold or by applying a *prescale factor*. A *prescaled* trigger with prescale P only selects one in P events passing the trigger requirement.

4 Data, Simulated Monte Carlo Samples and Trigger

Details about the production of data and Monte Carlo samples used in this analysis will be briefly presented in Sections 4.1 and 4.2, respectively. All samples are processed with the CxAOD framework version 26-01, using software release 20.7. Finally, details about the triggers used in the different data-taking periods will be briefly given in Section 4.3.

4.1 Data

This analysis uses proton-proton collision data recorded by the ATLAS detector and collected in 2015 and 2016 at a centre-of-mass energy of 13 TeV, with a bunch crossing of 25 ns. Events are recorded in stable beam conditions at the LHC. The LHC delivered 4.2 fb^{-1} and 38.5 fb^{-1} integrated luminosity in 2015 and 2016, respectively, of which the ATLAS detector recorded 3.9 fb^{-1} and 35.6 fb^{-1} during the respective year. This corresponds to a data taking efficiency of $\sim 93\%$ and $\sim 92\%$ in 2015 and 2016, respectively. The slight data recording inefficiency comes from data acquisition system inefficiencies, detector dead time and ramping up of the ATLAS detector until being fully operational after the LHC declares stable beam conditions.

Occasionally, subdetectors or detector components of the ATLAS detector may run in degraded conditions, resulting in corrupted or incomplete events, lacking of information from the entire ATLAS detector. This may happen if subsystems must be reset (e.g. readout electronics). In this analysis, all detector components, such as calorimeters, tracking detectors, trigger system and muon spectrometers, are required to be fully operational. Only data passing quality criteria, so-called *Good-Run Lists*, are considered. Luminosity summary plots for proton-proton collisions data taking during 2015 and 2016 are presented in Figure 4.1 as a function of time.

The data Good-Run Lists used in this analysis are

- **2015 Dataset:** DATA15_13TEV.PERIODALLYEAR_DETSTATUS-V79-REPRO20-02_DQDEFECTS-00-02-02_PHYS_STANDARDGRL_ALL_GOOD_25NS.XML
- **2016 Dataset:** DATA16_13TEV.PERIODALLYEAR_DETSTATUS-V83-PRO20-15_-DQDEFECTS-00-02-04_PHYS_STANDARDGRL_ALL_GOOD_25NS.XML

The good-run lists dataset amounts in total to an integrated luminosity of 36.1 fb^{-1}, of which 3.2 fb^{-1} were collected in 2015 and 32.9 fb^{-1} in 2016. The mean number of proton-proton interactions per bunch crossing is 14 in the 2015 dataset and 25 in the

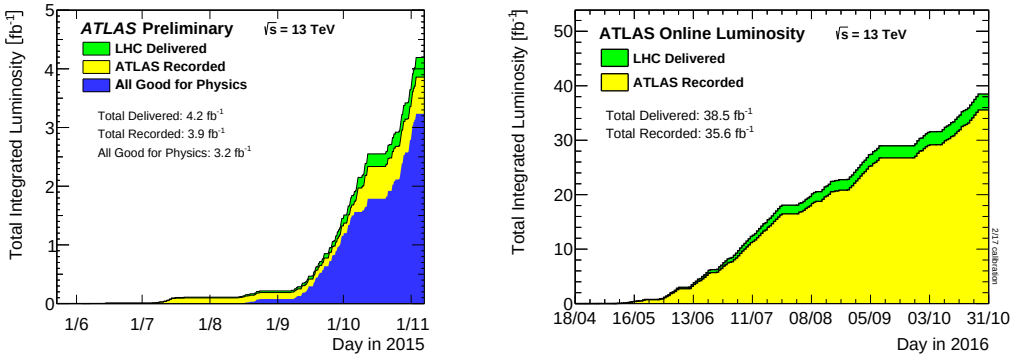

Figure 4.1: Integrated luminosity as a function of time, delivered to ATLAS (green), recorded by ATLAS (yellow) and classified as good quality data (blue) during stable beams for proton-proton collisions at 13 TeV in 2015 (left) and 2016 (right). From Reference [61].

2016 dataset. *Pile up* refers to low-momentum-exchange proton-proton interactions taking place in a given bunch crossing in presence of a hard proton-proton interaction and contributing to the baseline signal in subsequent bunch crossings.

4.2 Monte Carlo Simulated Samples

Simulated MC samples are produced for signal and background processes. The full GEANT 4-based ATLAS simulation [62] is used for all produced MC samples. Information on the MC generators, parton distributions functions (PDFs) and the production cross-sections used for all MC samples are summarised in Table 4.1.

In the following, a brief description of signal MC samples and background MC samples will be presented in Subsections 4.2.1 and 4.2.2, respectively.

Process	**Generator**	**PDFs**	σ_{norm} [pb]
mono-h signals	MADGRAPH + PYTHIA 8	NNPDF30_lo_as_0130	-
$t\bar{t}$	POWHEG + PYTHIA 6 (NLO)	CT10	831.76
Single top			
t-channel	POWHEG + PYTHIA 6 (NLO)	CT10	136.02+80.95
s-channel	POWHEG + PYTHIA 6 (NLO)	CT10	6.35+3.97
Wt-channel	POWHEG + PYTHIA 6 (NLO)	CT10	71.7
V**+jets**			
$W \to \ell\nu$+jets	SHERPA 2.2.1 (NNLO)	NNPDF3.0NNLO	20080
$Z \to \ell\ell, \nu\nu$+jets	SHERPA 2.2.1 (NNLO)	NNPDF3.0NNLO	1906 ($66 < m_{ll} < 116$ GeV)
Diboson			
$Wl\nu Wqq$	SHERPA 2.1 (NLO)	CT10	22.65
$Wl\nu Zqq$	SHERPA 2.1 (NLO)	CT10	10.47
$WqqZll$	SHERPA 2.1 (NLO)	CT10	3.12
$Wqq Z\nu\nu$	SHERPA 2.1 (NLO)	CT10	6.17
$Zqq Z\ell\ell$	SHERPA 2.1 (NLO)	CT10	2.15
$Zqq Z\nu\nu$	SHERPA 2.1 (NLO)	CT10	4.22
SM Vh			
$qq \to Vh(\to bb)$	PYTHIA 8 (LO)	NNPDF2.3LO	0.7639
$gg \to Vh(\to bb)$	POWHEG + PYTHIA 8 (LO)	CT10	0.1057

Table 4.1: List of MC generators, parton distribution functions (PDFs) [63, 64], and the production cross-section (σ_{norm}) per process used for signal ($Z' - 2HDM$, Z'-mediator and S-mediator models) and background processes.

4.2.1 Signal Monte Carlo Samples

Simulated MC samples for the mono-h signal models are generated with the MADGRAPH [65] generator at leading order (LO) in QCD, using the NNPDF30LO [63, 64] PDF set. PYTHIA 8 [66] generator is used with the A14 tune [67] and NNPDF2.3LO PDF set to simulate parton showering and hadronisation. This applies for the simplified models $Z' - 2HDM$, Z' vector mediator and S scalar mediator models.

In this analysis, only $Z' - 2HDM$ models are used for interpretation of statistical results as benchmark signals. Following the notation used in Subsection 2.3.1 and as explained in Subsection 2.3.4, the fixed parameters used in this model are $\tan(\beta) = 1$, $g_Z = 0.8$, $m_X = 100$ GeV and $m_H = m_{H^\pm} = 300$ GeV. The mass of the Z' gauge boson $m_{Z'}$ and of the pseudoscalar Higgs m_A are free parameters, affecting the kinematics. They are scanned in order to probe the regions of phase space where the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis is expected to be sensitive to mono-h signals. $m_{Z'}$ is scanned from 200 GeV to 3 TeV in 200 GeV steps, while m_A is scanned from 200 GeV to 800 GeV in 100 GeV steps.

Z' vector mediator and S scalar mediator models are used in the performance evaluation of the new optimised event selections. Following the notation used in Subsection 2.3.2, the fixed parameters used in the Z' vector mediator model are $\sin(\theta) = 0.3$, $g_\chi = 1$, $g_q = 1/3$ and $g_{hZ'Z'} = m_{Z'}$. The Z' gauge boson mass $m_{Z'}$ and the DM particle mass m_χ are the free parameters, affecting the kinematics, so one scans on these parameters to probe the regions of phase space to which the mono-h analysis is sensitive. Following the notation used in Subsection 2.3.3, the fixed parameters used in the S scalar mediator model are $b = 3$, $y_\chi = 1$ and $\sin(\theta) = 0.3$. The S scalar mass m_S and the DM particle mass m_χ are the free parameters, affecting the kinematics, so one scans on these parameters.

4.2.2 Background Monte Carlo Samples

The MC generators, PDFs and normalisation cross-sections used for the background processes are summarised in Table 4.1. The main background processes in the SR of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis are the production of $V + jets$ (where $V = W$ or $= Z$) and of $t\bar{t}$ events. Top-pair production $t\bar{t}$ is generated with POWHEG, using PYTHIA 6 [68] to simulate parton showering and hadronisation. It is then normalised using cross-sections calculated to NNLO plus next-to-next-to-leading logarithms (NNLL) introducing soft gluon radiation. The $V + jets$ samples are generated with massive b- and c-quarks using SHERPA 2.2.1 [69] generator and the NNPDF3.0NNLO [64] PDF set, and then normalised using cross-sections calculated at next-to-next-to-leading order (NNLO) in QCD. They are generated with filters to separate samples with final states containing hard light, charm and bottom quarks. The single top-quark samples are generated with the POWHEG generator, and normalised to cross-sections calculated at NLO. The diboson (WW, WZ and ZZ) samples are generated with SHERPA 2.1 interfaced to CT10 PDFs at NLO and normalised to cross-sections calculated at NLO. SM Vh samples are generated with PYTHIA 6.

In-time and *out-of-time* pile up are simulated using PYTHIA 8 with A2 tune [70], interfaced to MSTW2008LO [71] PDF set. These simulated pile-up samples, so-called *minimum bias events*, are overlaid with all simulated MC samples. The distribution of the average number of interactions per bunch crossing of simulated minimum bias events is reweighted to reproduce the observed distribution in data.

4.3 Trigger

Events in the 0 and 1 lepton regions are triggered by E_T^{miss} triggers. Events in the 2 lepton region are triggered by single lepton triggers. The trigger items used are the lowest unprescaled triggers in the trigger menu for Run 2 luminosity.

The full list of triggers used in the $h(\to b\bar{b}) + E_T^{miss}$ analysis can be found in Table 4.2.

4.3.1 E_T^{miss} Triggers in the 0 Lepton Region

Different unprescaled E_T^{miss} triggers are used in data (and in MC) according to the increasing instantaneous luminosity in 2015 and 2016, in order to adjust to the nominal events rates necessary for mass storage. The E_T^{miss} triggers in the 0 lepton region as function of the data-taking periods are

- HLT_XE70, L1_XE50, for 2015 data (3.2 fb^{-1}).
- HLT_XE90_MHT_L1XE50, for 2016 data (periods A-D3, 6.1 fb^{-1}).
- HLT_XE100_MHT_L1XE50 OR HLT_XE110_MHT_L1XE50, for 2016 data (periods D4-E3, 3.9 fb^{-1}).
- HLT_XE110_MHT_L1XE50, for 2016 data (periods F1-, 23.2 fb^{-1}).

The lower E_T^{miss} requirement in the event selection is $E_T^{miss} > 150$ GeV, see Chapter 6, at which the E_T^{miss} triggers are not fully efficient. To include consistently the phase space below fully efficiency of E_T^{miss} triggers, it is necessary to correct for mismatching of E_T^{miss} trigger efficiency turn-on in MC with that in data. One uses custom E_T^{miss} *trigger scale factors* to bring the MC E_T^{miss} trigger efficiency turn-on curves to match those in data.

The derivation of E_T^{miss} trigger scale factors used in the $h(\to b\bar{b}) + E_T^{miss}$ analysis are documented in [72, 73].

4.3.2 Data-driven E_T^{miss} Trigger Scale Factors

The E_T^{miss} trigger scale factors are derived in a data-driven way, by comparing the E_T^{miss} trigger efficiency turn-on curves of MC samples with those of data in the 1 lepton region. The 1 lepton region is orthogonal to the 0 lepton region where the scale factors will be applied.

The E_T^{miss} as seen by the HLT trigger does not contain the four-momentum from muon objects [60]. Therefore, the offline E_T^{miss} four-momentum *plus* the muon four-momentum of events in the 1 lepton region (containing a single muon lepton and no other leptons) is kinematically similar to the HLT E_T^{miss}. One denotes it as $E_{T,noMU}^{miss}$. The offline $E_{T,noMU}^{miss}$ in the 1 lepton region in turn kinematically mimics the E_T^{miss} in the 0 lepton region.

The E_T^{miss} trigger efficiency $\varepsilon_{E_T^{miss}}$ is defined as

$$\varepsilon_{E_T^{miss}} = \frac{N^{\underline{o}} \text{ Events passing event selection AND } E_T^{miss} \text{ trigger}}{N^{\underline{o}} \text{ Events passing event selection}}, \tag{4.1}$$

Period	0 lepton	1 lepton	2 lepton
2015	HLT_xe70	HLT_xe70	HLT_e24_lhmedium_L1EM18VH (MC) **OR** HLT_e24_lhmedium_L1EM20VH (data) **OR** HLT_e60_lhmedium **OR** HLT_e120_lhloose **OR** HLT_mu20_iloose_L1MU15 **OR** HLT_mu50
2016 (A)	HLT_xe90_mht_L1XE50	HLT_xe90_mht_L1XE50	HLT_e24_lhtight_nod0_ivarloose **OR** HLT_e60_lhmedium_nod0 **OR** HLT_e60_medium **OR** HLT_e300_etcut **OR** HLT_e140_lhloose_nod0 **OR** HLT_mu24_iloose_L1MU15 (MC) **OR** HLT_mu24_iloose (data) **OR** HLT_mu40
2016 (B-D3)	HLT_xe90_mht_L1XE50	HLT_xe90_mht_L1XE50	HLT_e24_lhtight_nod0_ivarloose **OR** HLT_mu24_ivarmedium **OR** HLT_mu50
2016 (D4-E3)	HLT_xe100_mht_L1XE50 **OR** HLT_xe110_mht_L1XE50	HLT_xe100_mht_L1XE50 **OR** HLT_xe110_mht_L1XE50	HLT_e26_lhtight_nod0_ivarloose **OR** HLT_mu24_ivarmedium
2016 (F1)	HLT_xe110_mht_L1XE50	HLT_xe110_mht_L1XE50	HLT_mu26_ivarmedium
2016 (F2-)	HLT_xe110_mht_L1XE50	HLT_xe110_mht_L1XE50	HLT_e26_lhtight_nod0_ivarloose **OR** HLT_e60_lhmedium_nod0 **OR** HLT_e60_medium **OR** HLT_e300_etcut **OR** HLT_e140_lhloose_nod0 **OR** HLT_mu26_ivarmedium **OR** HLT_mu50

Table 4.2: Summary table of triggers used in 2015 and 2016 data.

where the event selection is as described in Chapter 6, except for the lower E_T^{miss} requirement and the triggers used, which in this case are the lowest unprescaled single muon triggers. $\varepsilon_{E_T^{miss}}$ is a function of $E_{T,noMU}^{miss}$ for each lowest unprescaled muon trigger in the corresponding data-taking period.

The correction scale factors (SFs) are defined as the ratio of E_T^{miss} trigger efficiencies for data over those for MC, as follows:

$$\mathrm{SF} = \frac{\varepsilon_{E_{T,noMU}^{miss}}(\mathrm{Data})}{\varepsilon_{E_{T,noMU}^{miss}}(\mathrm{MC})}. \tag{4.2}$$

To calculate the data-driven corrections for the MC E_T^{miss} trigger efficiency turn-ons as a function of $E_{T,noMU}^{miss}$, the SFs are fitted for each E_T^{miss} trigger with the following functional form

$$f\left(E_{T,noMU}^{miss}\right) = 0.5 \cdot \left[1 + Erf\left(\frac{E_{T,noMU}^{miss} - p_0}{\sqrt{2}p_1}\right)\right], \tag{4.3}$$

in the fit range 120 GeV $< E_{T,noMU}^{miss} <$ 300 GeV.

MC samples are divided in periods according to the luminosities of data periods with corresponding lowest unprescaled E_T^{miss} triggers.

Plots of the SFs for each E_T^{miss} trigger can be found in Reference [73], slide 5.

5 Object Reconstruction

Objects (referring to e.g. electrons, muons, jets, etc) must be reconstructed and identified in order for them to be used in the analysis. The identification criteria are designed to maximise the acceptance of objects as expected in the signal topology while reducing the contamination of those from background processes such as multijet production. In the following, the reconstruction and identification criteria of various objects as used in the analysis are briefly introduced. This is based on studies done by several combined performance groups at the ATLAS collaboration and, more specifically for our analysis, by the SM $Vh(bb)$ analysis group [74].

5.1 Jets

Quarks and gluons fragment and hadronise forming a directional collimated spray of hadrons, called *jets*. Jets are formed using jet algorithms that cluster constituents into jets. The constituents can be for example topological calorimeter cell deposit clusters, called *topoclusters*, in the calorimeters or tracks in the ID. According to the jet algorithm used, its clustering parameters and the constituents considered, various jet collections can be formed targeting different topologies and kinematics. In the following, the jet collections used in the $h(\to b\bar{b}) + E_T^{miss}$ analysis will be briefly introduced.

5.1.1 Small-Radius Jets

Small-radius jets (or small-R jets) are used in the analysis to reconstruct the $h \to b\bar{b}$ decay products as two separated jets in presence of intermediate E_T^{miss} ($E_T^{miss} < 500$ GeV), and to reconstruct the hadronic activity in background processes and signal processes.

Jet Reconstruction

The constituents of small-R jets are topoclusters calibrated using the electromagnetic calorimeter scale [75]. Topoclusters are used as inputs to the anti-k_t algorithm [76], as implemented in FastJet [77], with a radius parameter of $R = 0.4$ to reconstruct jet objects. The anti-k_t algorithm is collinear safe (i.e. the collinear splitting of a particle leads to the same jet) and infrared safe (i.e. the emission of soft gluons leads to the same jet), and is used as the standard jet algorithm at ATLAS. The four-momentum of a jet is defined as the sum of the four-momenta of all constituents that have been clustered into the jet.

Small-R jets are reconstructed in $|\eta| < 4.5$ with $p_T > 7$ GeV as provided by ATLAS. Jets within $|\eta| < 2.5$ are called *central jets* and used to reconstruct the $h \to b\bar{b}$ candidate.

The remaining jets are referred to as *forward jets*. In the $h(\to b\bar{b}) + E_T^{miss}$ analysis, central jets must have $p_T > 20$ GeV while forward jets $p_T > 30$ GeV.

Jet Energy Calibration

The standard ATLAS jet energy scale (JES) calibration is applied to all small-R jets, correcting their jet energy scale to that of stable truth jets at the particle-level [78]. The jet energy scale calibration chain includes various correction steps, listed in the following.

- The *origin correction* changes only the direction of the jet, in order for it to point to the hard-scatter PV.
- The *pile-up correction* is based both on substracting the pile-up p_T density integrated over the jet area and on correcting residual pile-up using MC simulations.
- The *absolute MC-based calibration* corrects the jet energy response to the particle-level energy scale.
- The *global sequential calibration* corrects sequentially the dependence of the jet response on five observables while conserving the overall energy scale at the EM+JES, as described in [78].
- And finally a *residual in-situ calibration* applied only to jets in data for correcting differences in jet response between data and MC simulation, quantified by balancing the p_T of a jet against other well-measured reference object, including photons, Z bosons and calibrated jets. All small-R jets are then calibrated to the hadronic scale as described in [79].

Additionally to the stardard ATLAS jet energy calibration, improvements on the jet energy response to heavy flavour jets involved in the reconstruction of the $h \to b\bar{b}$ decay are applied in the analysis, as described below in Subsection 5.1.4.

B-Tagging

Jets containing b-hadrons are of special importance for our analysis, whose signature is $h(\to b\bar{b}) + E_T^{miss}$. The default b-tagging algorithm at ATLAS [80, 81] is a boosted decision tree (BDT) algorithm, which improves the discrimination between different jet flavours by using three b-tagging algorithms as inputs: *IP3D* that exploits impact parameters information, *SV1* that exploits secondary vertex information, and *JetFitter* that exploits decay chain multi-vertex information.

Hadrons containing a b-quark have a long lifetime ($c\tau \sim 450\ \mu$m). B-hadrons are then characterised by at least a vertex displaced from the hard-scatter primary vertex (PV), at which the b-hadrons decay, giving rise to displacement of tracks relative to the PV. Tracks from b-hadron decay products therefore tend to have large transverse impact parameter[1],

[1] The transverse impact parameter d_0 is the distance of closest approach of the track to the primary vertex in the transverse plane.

d_0, and large longitudinal impact parameter[2], z_0, unlike tracks from the PV. The IP3D algorithm is based on a log-likelihood ratio of d_0/σ_{d_0} and $z_0\sin(\theta)/\sigma_{z_0\sin(\theta)}$ probability distributions for b- and light-flavour jet hypotheses to separate jets containing b-hadrons from light-flavour-originated jets, as described in Reference [81].

For the secondary vertex finding algorithm, SV1, an inclusive displaced secondary vertex is explicitly reconstructed within the jet, corresponding to the location of the b-hadron decay. All pair of tracks displaced[3] from the PV and associated with the jet, are tested for the two-track vertex hypothesis (see Subsection 3.3 in Reference [80]). Two-track vertex candidates are removed if compatible with originating from K^0_S, Λ^0 decays, photon conversions, hadronic interactions in the detector material or random combination of tracks from the PV. All remaining two-track vertices are then fitted into a single inclusive secondary vertex using a Kalman-based χ^2 fit [82] to reconstruct the vertex of the b-hadron decay. Similarly to IP3D, the SV1 algorithm is based on a log-likelihood ratio formalism, as described in Reference [80].

The JetFitter algorithm is a decay chain multi-vertex algorithm that exploits the topological structure of weak b- and c-hadron decays within the jet to reconstruct the full b-hadron decay chain. It assumes that the PV, b- and c-hadron decay vertices lie on the same common line corresponding to the b-hadron flight path. This common line along with the position of the b- and c-hadron vertices are found with a Kalman filter [80]. The JetFitter algorithm is based on an artificial neural network, as described in Reference [80].

The b-tagging algorithm, called *MV2c10* [81], is trained with b-jets as signals and light-flavour jets ($\sim 93\%$) and c-jets ($\sim 7\%$) as background in $t\bar{t}$ events, using anti-k_T calorimeter jets, with $R = 0.4$. The signal efficiency working point used for b-tagging in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis is 70%. Jets that are b-tagged are referred to as *b-jets*.

5.1.2 Large-Radius Jets

When a massive particle such as a top quark, a W/Z gauge boson or a Higgs boson is produced with significant Lorentz boost and decays into quarks, it may be reconstructed as a single jet containing the entire hadronic decay. Large-radius jets (or large-R jets) are used in the analysis for events where the E_T^{miss} is very large ($E_T^{miss} > 500$ GeV) such that the $h \rightarrow b\bar{b}$ is very boosted, reconstructing the b-flavour decay products as a single large-R jet.

Jet Reconstruction

The constituents of large-R jets are calorimeter topoclusters calibrated to the hadronic scale using the local cell signal weighting (LCW) method [83]. The LCW topoclusters are clustered into large-R jets using the FastJet [77] implementation of the anti-k_t algorithm [76] with a radius parameter of $R = 1.0$.

[2]The longitudinal impact parameter z_0 is the distance of closest approach of the track to the primary vertex in the beam-axis.

[3]A track is considered *displaced* if $d_{3D}/\sigma_{d_{3D}} > 2$, where d_{3D} is the (three dimensional) distance of closest approach of the track to the PV and $\sigma_{d_{3D}}$ is its uncertainty.

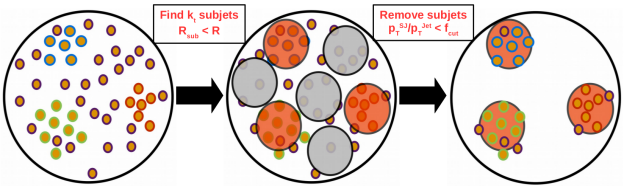

Figure 5.1: Illustration of the trimming algorithm from Reference [84], slide 8. In ATLAS, trimming is used with $R_{sub} = 0.2$ and $f_{cut} = 0.05$.

Using a radius parameter of $R = 1.0$ makes large-R jets more sensitive to contamination from *initial state radiation* (ISR), underlying event and pile-up, which degrades the jet reconstruction, for example, degrading the jet mass resolution or shifting the jet mass scale. Since sources of radiation in the event other than the hard scatter are likely to be much softer, jet reconstruction can be improved by removing soft radiation. Large-R jets at ATLAS are groomed using the *trimming* algorithm [85]. The constituents of the reconstructed large-R jet are *re-clustered* into subjets using the k_t algorithm with radius parameter of $R_{sub} = 0.2$. If a subjet has a p_T lower than 5% of the p_T of the ungroomed parent large-R jet, the subjet with its constituents is discarded, resulting into the final *trimmed* large-R jet. Trimming is illustrated in Figure 5.1.

The jet four-momentum is calculated as the sum of the four-momenta of the constituents. Large-R jets in the analysis have $p_T > 200$ GeV and $|\eta| < 2.0$.

Jet Mass Calibration

The standard mass of large-R jets in ATLAS is a weighted linear combination of two jet mass observables. The *calorimeter-based jet mass*, m^{calo}, for a large-R jet J with topocluster constituents having energy E_i and momentum $\vec{p}_i$ ($|\vec{p}_i| = E_i$) is defined as

$$m^{calo} = \sqrt{\left(\sum_{i \in J} E_i\right)^2 - \left(\sum_{i \in J} \vec{p}_i\right)^2} \tag{5.1}$$

similarly as done for small-R jets.

The angular separation of decay products of a boosted massive particle scales as $1/p_T$. If the particle has a sufficiently high Lorentz boost, the angular separation of the decay products can be comparable with the calorimeter granularity, starting to be limited by the angular resolution of the calorimeter. Since ID tracks have a better angular resolution than the calorimeters, the *track-assisted jet mass*, m^{TA}, can improve performance at high p_T values, using tracking and calorimeter information, as follows:

$$m^{TA} = \frac{p_T^{calo}}{p_T^{track}} \cdot m^{track} \tag{5.2}$$

where p_T^{calo} stands for the p_T of the large-R jet, p_T^{track} for the p_T of the four-momentum sum of tracks associated to the large-R jet, and m^{track} for the invariant mass of track four-momentum sum. The p_T^{calo}/p_T^{track} ratio corrects for the neutral component missed by the tracker.

The jet mass scale (JMS) calibrations corrects on average the reconstructed jet mass to the particle-level jet mass using multijets MC samples after applying the jet energy

calibration, as described in [17]. Large-R calorimeter jets are geometrically matched to *particle-level* truth jets if $\Delta R < 0.6$, where truth jets are reconstructed from detector-stable simulated particles ($c\tau > 10$ mm, with except for muons and neutrinos) as inputs using the anti-k_T algorithm with radius parameter of $R = 1.0$. For each matched large-R calorimeter and truth jet, the jet mass response distribution is given by the $\mathscr{R}_m = m^{reco}/m^{truth}$ ratio[4]. For each $(p_T^{truth}, |\eta_{det}|, m^{truth})$[5] bin, the average jet mass response $\langle \mathscr{R}_m \rangle$ is calculated and after a numerical inversion technique, calibration factors are derived as a function of $(p_T^{reco}, |\eta_{det}|, m^{reco})$[6]. Upon applying the calibration factors, the jet mass m^{calo} and m^{TA} are said to be calibrated if the average response fulfills $\langle \mathscr{R}_m \rangle = 1$.

The performance of the jet mass is determined by the jet mass resolution (JMR), which is associated with the width of the jet mass response distribution, defined as the half of the 68% interquantile range *IQnR* divided by the median[7]. In the ideal case of a Gaussian distribution, IQnR coincides with the standard deviation. Since the m^{calo} and m^{TA} are only slightly correlated[8], it is possible to reduce the response resolution by combining information from both jet mass observables. The *combined mass* of large-R jets is then defined as

$$m^{comb} = \left(\frac{\sigma_{calo}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}} \right) \cdot m^{calo} + \left(\frac{\sigma_{TA}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}} \right) \cdot m^{TA} \tag{5.3}$$

where σ_{calo} and σ_{TA} are the calorimeter-based jet mass and the track-assisted jet mass resolution function, respectively. By construction, the m^{comb} is calibrated if m^{calo} and m_{TA} are calibrated.

In Figure 5.2, the jet mass resolution curves of the m^{calo}, m^{TA} and m^{comb} are shown as a function of the jet p_T, for jets from W or Z bosons produced in $WZ \to qqqq$ processes samples, with matching to particle-level W and Z bosons. One observes that the m^{TA} outperforms the m^{calo} for $p_T \gtrsim 1$ TeV whereas m^{calo} has a better resolution for $p_T \lesssim 1$ TeV. Due to the calorimeter granularity, the angular spread of the hadronic decay products of W/Z bosons (angular separation scales as m/p_T of massive decaying particle) starts to hit the calorimeter angular resolution for $p_T > 1$ TeV, which allows the resolution of m^{TA} to be better than that of m^{calo}. On the other hand, the charged-to-neutral fluctuations dominate the resolution of m^{TA}, which is worse than that of m^{calo} for $p_T < 1$ TeV. The m^{comb} outperforms both the m^{TA} and m^{comb}, having a better JMR over the entire p_T range.

Jet Mass Scale Systematic Uncertainties

The jet mass scale of m^{calo} is probed in data with the ratio $r_{track}^{m} = m^{calo}/m^{track}$ in an inclusive selection of high p_T multijet events [87], as the average of r_{track}^{m} is approximately

[4] m^{reco} and m^{truth} are the mass of the large-R calorimeter and truth jet, respectively.

[5] $|\eta_{det}|$ is the pseudorapidity of the jet based on the detector geometry.

[6] m^{reco} can be either m^{calo} or m^{TA} depending on the mass being calibrated.

[7] IQnR is defined as $q_{84\%} - q_{16\%}$, where $q_{84\%}$ and $q_{16\%}$ are the 84$^{\text{th}}$ and 16$^{\text{th}}$ percentiles of the response distribution, respectively.

[8] Within $|\eta| < 2.0$, the correlation is 0.22 and 0.10 for $p_T > 250$ GeV and $p_T > 1$ TeV, respectively, for W/Z-jets in $Z'/W' \to WZ \to qqqq$ MC samples, as described in Reference [17].

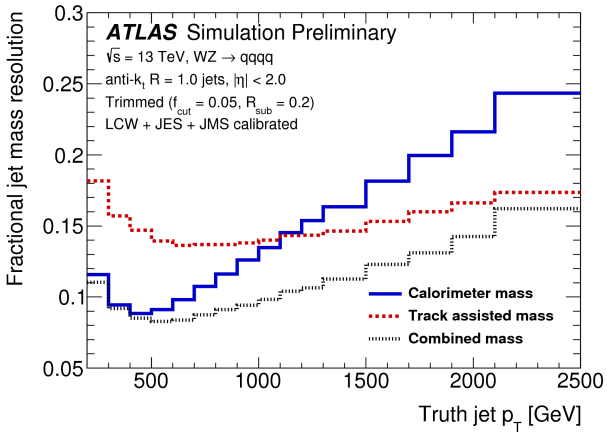

Figure 5.2: The jet mass resolution of the m^{calo}, m^{TA} and m^{comb} as a function of truth jet p_T for W/Z-jets produced in $WZ \to qqqq$ samples, with matching to particle-level W and Z bosons. The resolution is defined as $\frac{1}{2}(\text{IQnR})/median$. The resolutions of m^{calo}, m^{TA} used as input to the m^{comb} are determined using jets in multijet samples, with truth-jet matching. The m^{comb} resolution outperforms that of m^{calo} and m^{TA} over the entire p_T range. From Reference [86].

proportional to the the jet mass scale. In this case, $1 - \langle r^m_{track}\rangle_{Data}/\langle r^m_{track}\rangle_{MC}$ is a measure of the JMS uncertainty of m^{calo}.

In the case of m^{TA}, the jet mass scale (and resolution) uncertainties can be estimated by propagating the detector-based uncertainties on track reconstruction (such as track reconstruction inefficiency and fake track reconstruction) and calorimeter-jet p_T uncertainties (calculated using the r_{track} method: $r^{p_T}_{track} = p^{calo}_T / p^{track}_T$ as in [17]) according to its definition in Equation 5.2. The modelling of fragmentation of hadrons introduces an additional uncertainty on jet mass scale of m^{TA}. The component of charged hadrons at detector level, which are reconstructed in the tracker, depends on the fragmentation modelling at particle level. Charge-to-neutral modelling differences introduce then uncertainties on the jet mass scale of m^{TA}. The author evaluated this uncertainty by comparing the m^{TA} jet mass scale behaviour for two different MC generators based on different fragmentation modelling, namely PYTHIA8 [66] and HERWIG++ [88]. The jet mass scale is parametrised as a function of p_T, $|\eta|$ and m/p_T.

The fragmentation modelling uncertainty is defined as the absolute difference of the jet mass scale of m^{TA} of events generated with PYTHIA8 and HERWIG++. In Figure 5.3, left (from the author), the signed fragmentation modelling uncertainty on the jet mass scale of m^{TA} is shown for dijet MC events as a function of p_T within $|\eta| < 0.4$ and three slices in m/p_T. In this phase space, the fragmentation modelling uncertainty is within 1.5%. The fragmentation modelling uncertainty is within 5% considering the whole phase space. Combining all sources of systematic uncertainties and statistical uncertainty, the total uncertainty on the jet mass scale of m^{TA} is shown in Figure 5.3, right, as a function of p_T, within $|\eta| < 2.0$ and for $m/p_T = 0.1$. It is seen that the JMS uncertainty of m^{TA} is lower than that of m^{calo} (small-dashed curve) for $p_T \lesssim 1$ TeV.

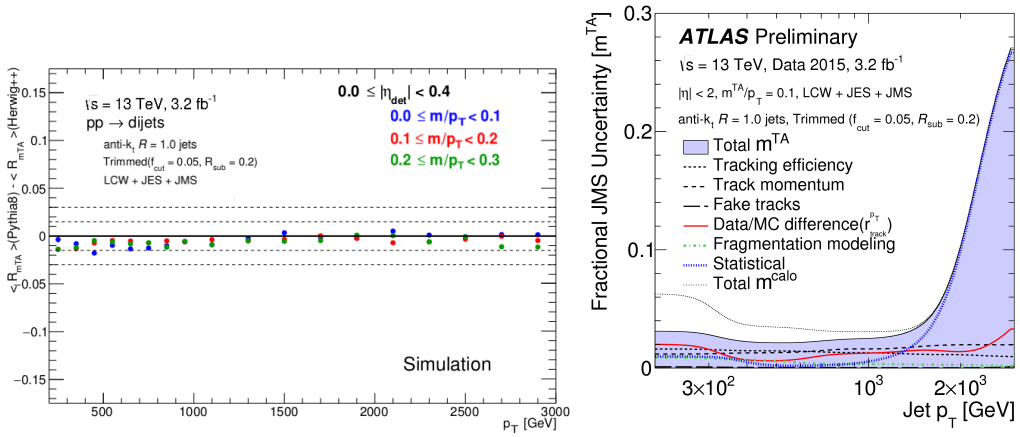

Figure 5.3: Left: Signed fragmentation modelling uncertainty on the jet mass scale of m^{TA} as a function of p_T. Within the shown phase space, the uncertainty is less than 1.5%. Right: Total uncertainty on the jet mass scale of m^{TA} as a function of p_T, within $|\eta| < 2.0$, $m/p_T = 0.1$. The uncertainty is rather stable and less than $\sim 3\%$ up to $p_T = 1$ TeV, after which it starts increasing. The JMS uncertainty of m^{TA} is lower than that of m^{calo} (small-dashed curve) for $p_T \lesssim 1$ TeV. Left Figure from the author. Right Figure from Reference [17].

5.1.3 Track Jets

Track jets are used in the analysis to identify heavy flavour hadrons in large-R jets. Track jets are reconstructed using ID tracks as constituent inputs to the anti-k_T algorithm with a radius parameter of $R = 0.2$ [89]. In boosted topologies, using tracking information (and $R = 0.2$) allows to better reconstruct the b-hadrons from $h \to b\bar{b}$ decay as separated jets, which otherwise would be merging in the calorimeter due to its poorer angular resolution.

The ID tracks constituents of track jets are required to fulfill $p_T > 0.4$ GeV, $|\eta| < 2.5$, having at least 7 hits in the silicon detectors, having no more than one hit in the pixel detector that is shared by multiple tracks, having no more than one missing expected hit in the pixel detector and no more than two missing expected hits in the silicon detectors, and finally that either the track is a constituent of the hard scatter primary vertex or $|z_0 \cdot \sin(\theta)| < 3$ mm.

Track jets are associated to large-R jets using *ghost association* [90]. *Ghosts* are the four-momenta of track jets in the event, with p_T set to an infinitesimal small value, so that ghosts essentially retain the direction of the track jets. All jet constituents in the event, including the ghosts, are reclustered again using the anti-k_T algorithm with $R = 1.0$. The resulting jets are identical to the ungroomed large-R jets in absence of the ghosts, with the addition of the ghosts contained in the large-R jets as constituents. Then track jets are said to be *ghost-associated* to a large-R jet if the track jet is contained within the catchment area as constituent of the ungroomed jet after reclustering.

The b-tagging of track jets in large-R jets is based on MV2c10, using the same training for b-tagging of small-R jets, as introduced in 5.1.1. Only the first two leading track jets ghost-associated to the large-R jet are used to assess the flavour content of the large-R jet. Tracks jets in the analysis must have $p_T > 10$ GeV, $|\eta| < 2.5$ and at least two tracks.

5.1.4 Muon-in-Jet Correction

About 20% of b-hadron decay chains produce a muon and muon neutrino, as quoted in Reference [74], line 667. Muons are minimum ionising particles, so they traverse the calorimeters with a small energy loss, without depositing their energy in the calorimeters. It is then necessary to correct for the four-momentum carried away by the muon when reconstructing b-jets. The four-momentum of a jet with a muon in it is corrected first by removing the four-momentum of the energy loss deposited in the calorimeters by the muon and then adding the four-momentum of the reconstructed muon, described in Subsection 5.2.2. This is called *muon-in-jet* correction. It improves the energy response, the scale and the resolution of the mass of the Higgs candidate, with respect to the truth Higgs mass.

Muons used in the muon-in-jet correction must fulfill the medium ID requirement and $p_T > 4$ GeV. For small-R jets, the effect of semi-muonic decays of b-hadrons is corrected using the closest reconstructed muon within $\Delta R < 0.4$ of the reconstructed b-jet. For large-R jets, the closest reconstructed muon within $\Delta R < 0.2$ of a b-tagged ghost-associated track jet is used for the muon-in-jet correction. Only the leading and subleading ghost-associated track jets, if b-tagged, are considered, so that the four-momentum of up to two reconstructed muons can be added to the large-R jet four-momentum in the muon-in-jet correction.

5.2 Leptons

Lepton objects have two main applications in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis. The first is to veto background events containing leptons for construction of the signal region. The second is to select events containing leptons (electron or muon) used in the control regions to constrain background processes. In the following, electrons, muons and taus will be briefly introduced in Subsections 5.2.1, 5.2.2 and 5.2.3, respectively.

5.2.1 Electron Leptons

As result of the interaction of electrons with the ATLAS detector material, the production of electrons in the detector gives rise to tracks in the inner detector and energy deposits in the electromagnetic (EM) calorimeter. By using different techniques and algorithms applied on the detector information, one can reconstruct, identify and isolate electron objects. This is described below as used in the analysis, following [91].

Electron Reconstruction

Electron candidates are reconstructed only in the central region $|\eta| < 2.47$. The electron candidate reconstruction is based on the geometrical *matching* [92] in $\eta \times \phi$ of a *track candidate* with a *seed-cluster candidate*, defined as follows:

- **Seed-cluster candidate:** In the EM calorimeter, cells are clustered within fixed-size rectangles forming a seed-cluster candidate, as described in [93]. The efficiency of this cluster search varies from 95% at $E_T = 7$ GeV to more than 99% above $E_T = 15$ GeV.

- **Track candidate:** The track candidate reconstruction consists of pattern recognition and track fit, in a first step fitting with the ATLAS Global χ^2 Track Fitter [94] and in a second step refitting with an optimised Gaussian Sum Filter [95], which accounts for non-linear bremsstrahlung effects.

For the four-momentum, ϕ and η of electrons with respect to the beam-line are obtained from the corresponding values of the best track matched to the original seed cluster, while the energy is given by the final calibrated cluster [96].

Background from conversions and secondary particles is reduced upon requiring two conditions: $d_0/\sigma_{d_0} < 5$ and $|\Delta z_0 \cdot \sin(\theta)| < 0.5$ mm, increasing compatibility between matched track and electron cluster.

Electron Identification

A likelihood-based (LH) method is the identification (ID) algorithm used to identify electrons, separating them from *background-like* electron candidates such as hadronic jets or converted photons. The LH ID method is based on discriminating properties of the electron cluster and matched track, such as calorimeter shower shapes, bremsstrahlung effects for distinguishing signal from background, track properties, track-cluster matching related quantities, etc [91]. The optimisation of the ID algorithm is performed by using electron candidates from MC simulations samples of $Z \to ee$ and dijet events, along with $J/\psi \to ee$ and minimum bias events at low E_T.

The identification operating points defined for the electron ID are the *Loose*, *Medium* and *Tight* operating points, in which ordering the background rejection (signal acceptance) consecutively increases (decreases). The operating points are subset of one another: e.g. electrons selected by *Medium* are also selected by *Loose*.

The performance of the LH electron ID algorithm is shown in Figure 5.4 for the three mentioned operating points. The efficiencies for electron candidates at the three operating points increases with E_T.

Electron Isolation

After identification, the *isolation* allows to further disentangle prompt electrons from those originating e.g. from converted photons or light hadrons misidentified as electrons.

- **Calorimeter isolation energy, $E_T^{\text{cone0.2}}(\eta, E_T)$:** sum of positive E_T of topological clusters [97] at the EM scale, within a cone of $\Delta R = 0.2$ around the electron cluster, excluding the region $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ centred at the cluster barycentre.

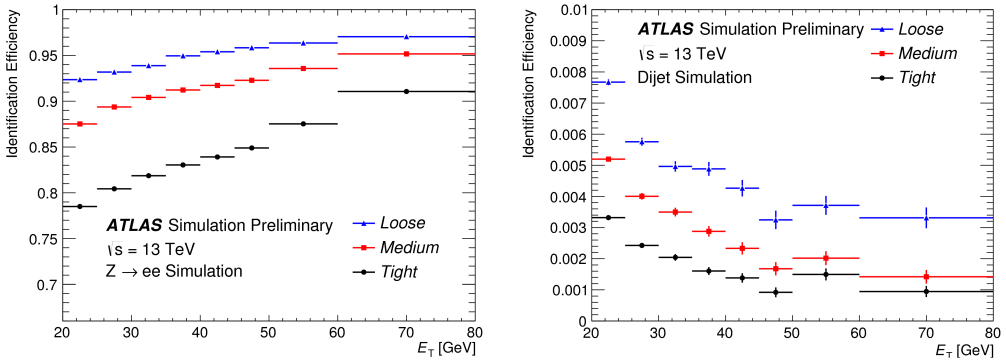

Figure 5.4: In the left, the electron ID efficiency from $Z \to ee$ MC samples, and in the right, the fake electron ID efficiency of hadrons from dijet MC samples, as measured with respect to reconstructed electrons. The candidates in $Z \to ee$ samples are matched to true electrons. From Reference [91].

- **Track isolation momentum, $p_T^{\text{varcone0.2}}(\eta, E_T)$:** p_T sum of all tracks within a cone of $\Delta R = \min(0.2, 10\ \text{GeV}/E_T)$ around the electron track, originating from the primary vertex and satisfying quality requirements, excluding the electron track itself and tracks from converted photons.

In the $h(\to b\bar{b}) + E_T^{miss}$ analysis, only *LooseTrackOnly* and *FixedCutHighPtCaloOnly* are used as isolation criteria. *LooseTrackOnly* targets a constant isolation efficiency of 99% varying $p_T^{\text{varcone0.2}}$ only. *FixedCutHighPtCaloOnly* uses a constant upper threshold on $E_T^{\text{cone0.2}}$ as $E_T^{\text{cone0.2}} < 3.5$ GeV. More details in [91].

Electrons in the $h(\to b\bar{b}) + E_T^{miss}$ analysis

In the $h(\to b\bar{b}) + E_T^{miss}$ analysis, by combining identification and isolation definitions described above, one defines three electron candidates: the *V-loose* electron, the *Z-signal* electron and the *W-signal* electron candidate, whose definition criteria are summarised in Table 5.1.

Electron Candidate	p_{T} [GeV]	Identification	Isolation
V-loose	> 7 GeV	Loose LH	LooseTrackOnly
Z-signal	> 27 GeV	Loose LH	LooseTrackOnly
W-signal	> 27 GeV	Tight LH	FixedCutHighPtCaloOnly

Table 5.1: The electron objects used in the $h(\to b\bar{b}) + E_T^{miss}$ analysis are summarised. All electron candidate definitions require $|\eta| < 2.47$, $d_0/\sigma_{d_0} < 5$ and $|\Delta z_0 \cdot \sin(\theta)| < 0.5$ mm.

5.2.2 Muon Leptons

Muons interact with the detector material of the inner detector (ID), calorimeters and muon spectrometers (MS), so that information from ID and MS, supplemented by information from the calorimeters, is used to reconstruct, identify and isolate muon objects. This is described below as used in the analysis, following [98].

Muon Reconstruction

The muon reconstruction is in a first step performed independently in the ID and MS. Most muons with $p_T \gtrsim 6$ GeV produce a full track in the MS, so a muon is in a second step reconstructed globally combining tracks in the ID and MS. However, other reconstruction algorithms need to be exploited in parallel for example in order to extend the reconstruction toward lower momenta or compensate edge effects in the geometrical acceptance between ID and MS. The parallel reconstruction algorithms are as follows:

- **Combined (CB) muon:** the hits of independently reconstructed muon tracks in the ID and MS are globally refitted to form a combined muon track.
- **Segment-tagged (ST) muon:** a track in the ID is defined as a muon candidate if associated with at least one local track segment in the MDT or CSC chambers, which accounts for muons not fully crossing the MS.
- **Calorimeter-tagged (CT) muon:** a track in the ID is defined as a muon candidate if matched to a calorimeter energy deposit compatible with a minimum ionising particle. It is optimised for $15 < p_T < 100$ GeV and $|\eta| < 0.1$, recovering acceptance lost in the MS due to cabling and services to calorimeters and the ID.
- **MS Stand-alone (SA) muons:** a track in the MS is defined as a muon candidate if compatible with originating from the interaction point and traversing at least two layers of MS chambers (or three layers if in the forward region), which extends the acceptance for muon reconstruction into $2.5 < |\eta| < 2.7$ (outside the ID coverage).

Overlap of muon reconstruction definitions can occur. When two muons share the same ID track, the CB muons are prioritised, then ST muons and finally CT muons.

Muon Identification

The muon identification is a cut-based selection of reconstructed muons, along with requirements on track quality, which allows suppressing *background-like* muons from pion and kaon decays, while selecting *signal-like* muons from W, Z boson decays.

Four muon identification selection criteria are defined, namely *Loose*, *Medium*, *Tight* and *High-p_T* [98], of which only *Loose* and *Medium* muon IDs are used in the analysis.

- **Medium:** Default ATLAS muon ID, for which only CB and SA tracks are considered. CB tracks are required to have more than 3 hits in at least two MDT

layers, while SA tracks are required to have at least three MDT/CSC layers and $2.5 < |\eta| < 2.7$. Additionally, CB and SA tracks are required *p/q significance* less than 7 for compatibility between ID and MS p_T measurements, as described in [98]. In Figure 5.5, the *Medium* ID efficiency is shown as a function of p_T.

- **Loose:** CB and SA muons satisfying the *Medium* selection pass the *Loose* selection. CT and ST muons are required to have $|\eta| < 0.1$. For loose muons within $|\eta| < 2.5$, $\sim 97.5\%$, $\sim 1.5\%$ and $\sim 1\%$ correspond to CB, CT and ST candidates, respectively.

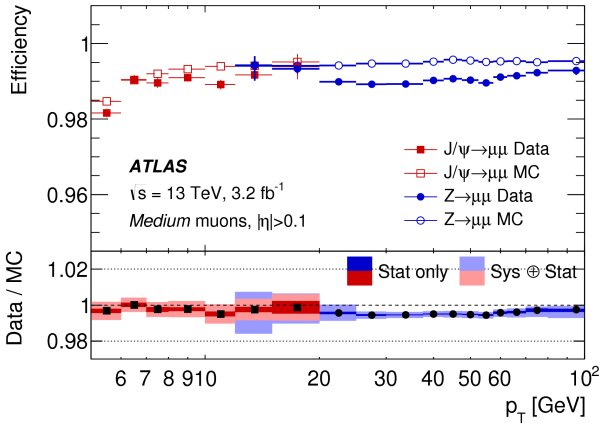

Figure 5.5: Identification efficiency for the medium muon criterion as a function of muon p_T in $0.1 < |\eta| < 2.5$, as obtained with $Z \to \mu\mu$ and $J/\psi \to \mu\mu$ MC events. From Reference [98].

Muon Isolation

The isolation variables used for muons are the same than those introduced for electrons. In the $h(\to b\bar{b}) + E_T^{miss}$ analysis, only *LooseTrackOnly* (see Subsection 5.2.1) and *FixedCutHighPtTrackOnly* are used for muons. *FixedCutHighPtTrackOnly* requires $p_T^{\text{varcone0.2}} < 1.25$ GeV only.

Muons in the $h(\to b\bar{b}) + E_T^{miss}$ analysis

Similarly to electron objects, in the $h(\to b\bar{b}) + E_T^{miss}$ analysis, by combining identification and isolation definitions described above, one defines three muon candidates: the *V-loose* muon, the *Z-signal* muon and the *W-signal* muon candidate, which is summarised in Table 5.2.

5.2.3 Tau Leptons

Tau leptons can decay either leptonically ($\sim 35\%$) or hadronically ($\sim 65\%$), which takes place typically before reaching the ATLAS detector. Since taus decaying to leptons (electrons or muons) cannot be distinguished from prompt leptons, only hadronic tau decays

Muon Candidate	p_{T} [GeV]	$\lvert\eta\rvert$	Identification	Isolation
V-loose	> 7 GeV	< 2.7	Loose ID	LooseTrackOnly
Z-signal	> 25 GeV	< 2.5	Loose ID	LooseTrackOnly
W-signal	> 25 GeV	< 2.5	Medium ID	FixedCutHighPtTrackOnly

Table 5.2: The muon objects used in the $h(\rightarrow b\bar{b})+E_T^{miss}$ analysis are summarised. All muon candidate definitions require $d_0/\sigma_{d_0} < 3$ and $|\Delta z_0 \cdot \sin(\theta)| < 0.5$ mm [98].

are considered for reconstructing tau objects. The tau decay products interact with the ID, as they contain either one (in $\sim 72\%$ of the cases) or three ($\sim 22\%$) charged pions, and with the calorimeters, so that information from both subdetectors is used in the reconstruction and identification. Neutrinos produced in the decays escape the detector, so only the visible part of the hadronic tau decay is reconstructed.

In the $h(\rightarrow b\bar{b})+E_T^{miss}$ analysis, tau objects are only used to veto background processes, unlike electrons and muons that are also used to define control regions. The tau veto is based on two tau object definitions, one is provided by the ATLAS combined performance tau group [99, 100], referred to as *standard tau* throughout the thesis, and another one is custom-built aiming at vetoing taus not identified as *standard* taus, referred to as *extended tau* throughout the thesis. In the following, the *standard* tau and the *extended* tau are introduced.

Tau Reconstruction

Jets as described in Subsection 5.1.1 are used as seeds of the tau reconstruction algorithm, which are additionally required to fulfill $p_T > 20$ GeV and $|\eta| < 2.5$. Tau candidates in the transition region between the barrel and forward calorimeters, $1.37 < |\eta| < 1.52$, are vetoed. The track reconstruction of taus follows dedicated algorithms [99, 100].

The chosen primary vertex of the event does not always correspond to the vertex at which the tau lepton was produced, referred to as tau vertex (TV). The TV is determined as the primary vertex with the largest fraction of p_T sum of tau candidate tracks in $\Delta R < 0.2$ around the jet seed direction. Using the TV improves the reconstruction efficiency at low p_T in a high pile-up environment, compared to using the standard primary vertex.

The tracks of tau candidates are required to have $p_T > 1$ GeV, at least two associated hits in the pixel detector (including the IBL), at least seven hits in the pixel and SCT detectors, $|d_0| < 1.0$ mm[9], and $|\Delta z_0 \cdot \sin(\theta)| < 1.5$ mm[10]. Tracks in $\Delta R < 0.2$ define the *core*[11] define the *core* while tracks in $0.2 < \Delta R < 0.4$ define the *isolation* region. Tau objects are required to have either one or three tracks in the core.

The tau reconstruction efficiency is dominated by the tau tracks and TV reconstruction,

[9] d_0 is in this case the distance of closest approach of the track to the TV in the transverse plane.
[10] z_0 is in this case the distance of closest approach of the track to the TV in the beam-axis.
[11] The direction reference for ΔR is calculated using the TV.

since finding a jet-seed for a tau candidate with $p_T > 20$ GeV within the pseudorapidity acceptance is almost fully efficient. For the tau four-momentum, the (η, ϕ) coordinates are calculated using the vectorial sum of topoclusters within $\Delta R < 0.2$ of the seed jet barycentre, using the TV as the origin while the E_T is provided by the tau-specific calibrated energy from the seed jet [99].

Tau Identification

The tau ID algorithm is designed to reject backgrounds from quark- and gluon-initiated jets. The tau ID algorithm is based on Boosted Decision Tree (BDT) methods and follows the approach described in [100, 101]. The BDT for tau candidates associated with one or three tracks are trained separately using MC samples of $Z/\gamma^* \to \tau\tau$ for signal and of dijet events for background. The input variables are based on tracks and topoclusters found in the core and isolations regions around the tau candidate direction, as described in [99].

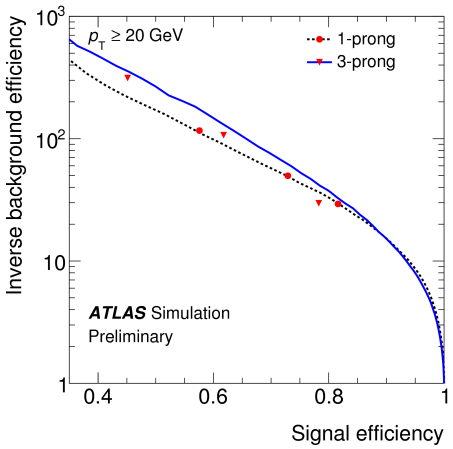

Figure 5.6: Inverse of mistagging efficiency of QCD jets as function of ID efficiency of tau candidates, considering 1-prong and 3-prong tau candidates. Each curve is obtained for constant BDT threshold requirements. The efficiency working points, shown as red points, do not fall on the curve because they are derived for constant signal efficiency as function of p_T. From Reference [100].

The ID efficiency is defined as the fraction of 1-prong (3-prong) hadronic tau decays that are reconstructed as 1-track (3-track) tau candidates (*reconstruction*) passing a given BDT selection criterion (*identification*). Three working points, labelled *Loose*, *Medium* and *Tight*, are provided, corresponding to different tau ID efficiency values. For each working point, the selection on the BDT score is determined as a function of tau p_T and varies, in order to achieve a constant tau signal efficiency.

The target signal efficiencies are 0.6, 0.55 and 0.45 for the 1-track *Loose*, *Medium* and *Tight* working points, respectively, while they are 0.5, 0.4 and 0.3 for the corresponding 3-track working points. The performance *receiver operating characteristic* (ROC) curve is shown for the efficiency working points in Figure 5.6. The efficiency working points do not fall on the curve because the ROC curve is derived for constant BDT threshold requirements and not for constant signal efficiency as function of p_T.

Custom-Built Tau

The aim of building an additional tau object definition is to construct a tau candidate that allows identifying hadronic taus, that are reconstructed as jets but missed by the loose BDT ID. It acts as an effective *very loose* tau identification.

The reconstruction algorithm of these custom-built tau candidates, referred to as *extended tau*, corresponds to that of jets described in Subsection 5.1.1. The tau ID in this case, referred to as *extended tau ID*, is a cut-based selection, consisting of two requirements:

- $1 \leq N_{track} \leq 4$: The first selection requires compatibility of the track multiplicity in the tau candidate with the charged pion multiplicity in hadronic tau decays. The tracks considered are tracks in the core, $\Delta R < 0.2$, associated with the primary vertex and have $p_T > 1$ GeV. Unlike *standard* taus, *extended* tau candidates are not required to have exactly either 1 or 3 tracks in order to account for underestimation of track multiplicity due to tracking inefficiency and for overestimation of track multiplicity due to photon conversion tracks that pass the standard track selection criteria.
- $\Delta\phi(\vec{E}_T^{miss}, \vec{\tau}) \leq \pi/8$: The *extended* tau ID targets hadronic taus that are decay products of leptonically decaying W bosons, $W \to \tau\nu_\tau$. Important background processes in the signal region, such as $t\bar{t}$, single top and $W + jets$, contain a W boson as a final state particle, decaying leptonically to a hadronic tau. From the topology of the decay $W \to \tau\nu_\tau$, the tau candidate is expected to be close to E_T^{miss} in $\Delta\phi$. The actual requirement was chosen after trying different upper values and evaluating the trade-off of significance gain of a tau veto and the signal loss of BSM models (see Subsection 7.2.2).

Taus in the $h(\to b\bar{b}) + E_T^{miss}$ analysis

As introduced above, in the $h(\to b\bar{b}) + E_T^{miss}$ analysis, two tau objects are used: the *standard tau*, which is a loose BDT tau candidate, and the *extended tau*, which is a custom-built tau candidate acting as an effective very loose tau. These tau objects are only used for vetoing background processes. A summary of the tau objects can be found in Table 5.3.

Tau Candidate	p_T [GeV]	$\lvert\eta\rvert$	Identification
Standard tau	> 20 GeV	< 2.5	Loose BDT ID
Extended tau	> 20 GeV	< 2.5	$1 \leq N_{track} \leq 4$, $\Delta\phi(\vec{E}_T^{miss}, \vec{\tau}) \leq \pi/8$

Table 5.3: The tau objects used in the $h(\to b\bar{b}) + E_T^{miss}$ analysis are summarised. *Standard* tau: additionally, it must have either 1 or 3 tracks in core, and is vetoed in $1.37 < |\eta| < 1.52$. *Extended* tau: additionally, it must fulfill the same conditions as required for jets in the $h(\to b\bar{b}) + E_T^{miss}$ analysis, please see 5.1.1.

5.3 Missing Transverse Momentum

The missing transverse momentum, E_T^{miss}, plays a crucial role in the analysis, as it is a proxy of the energy carried away by the dark matter particles escaping the detector. The principle for calculating E_T^{miss} is that the momentum in the plane transverse to the beam axis must be conserved, which means that the sum of the p_T of the collision products must be zero. Any imbalance in the momentum conservation is assigned as E_T^{miss}. Mismeasurements of reconstructed objects or production of particles escaping the acceptance of the detector can also contribute to the momentum imbalance. Dedicated algorithms are then necessary to reconstruct the E_T^{miss}. In the following, the reconstruction of E_T^{miss} as used in the analysis will be briefly introduced, following [102].

In the so-called TST E_T^{miss} scheme [102], the missing transverse energy vector $\vec{E}_T^{miss}$ in the transverse plane is defined as the negative vectorial sum in the transverse plane of all fully calibrated objects in the event, called *hard term*, plus tracks associated to the hard scatter primary vertex, called *track-based soft term*. The objects contributing to the hard term are calibrated and corrected for pile-up. In the $h(\to b\bar{b}) + E_T^{miss}$ analysis, the objects contributing to the hard term are electrons selected as V-loose, small-R jets and muons selected as V-loose, as described in Subsections 5.2.1, 5.1.1 and 5.2.2, respectively. The calorimeter signals are associated with these reconstructed objects following this order in the E_T^{miss} algorithm. The objects contributing to the soft term are ID tracks matched to the hard scatter primary vertex (then almost independent of pile-up), which must not be associated with any reconstructed object considered in the hard term. ID tracks are required to have $p_T > 0.4$ GeV, $|\eta| < 2.5$ and standard ATLAS reconstruction quality requirements, see [103]. The soft term aims at recovering residual contributions from the hard interaction, which are not considered in the analysis objects. ID tracks are required to have $p_T > 0.5$ GeV and $|\eta| < 2.5$.

The final E_T^{miss} used in the event selections and its azimuthal angle are calculated as:

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \tag{5.4}$$

$$\phi^{miss} = \arctan\left(\frac{E_y^{miss}}{E_x^{miss}}\right) \tag{5.5}$$

A track-based E_T^{miss}, referred to as $\vec{p}_T^{miss}$, is introduced to reduce beam-induced and non-collision background events. $\vec{p}_T^{miss}$ is defined as the negative vectorial sum of the transverse momenta of ID tracks matched to the hard scatter primary vertex, fulfilling $p_T > 0.5$ GeV and $|\eta| < 2.5$.

5.4 Overlap Removal

All energy depositions in the calorimeters are considered in the reconstruction jet algorithm. Particles depositing energy in the calorimeters, such as electrons, can be then reconstructed as jets, independently of whether they are identified by other reconstruction

algorithms. If a same particle is reconstructed as different analysis objects, an *overlap removal* procedure is applied to avoid double counting of objects in an event, removing any ambiguity of the reconstructed objects. The overlap removal is applied consecutively as follows, where objects are considered at a given step only if they were not removed at the previous step:

- *Electron/muon*: If a CB muon shares an ID track with an electron, the electron is removed and the muon is kept. If a CT muon shares an ID track with an electron, the muon is removed and the electron is kept.
- *Electron/jet*: Jets within $\Delta R < 0.2$ of an electron are removed, keeping the electron. For any surviving jet, electrons within $\min\Delta R(0.4, 0.04 + 10 \text{ GeV}/p_T^{\text{electron}})$ of the jet are removed.
- *Muon/jet*: Jets within $\Delta R < 0.2$ of a muon are removed if the jet has less than three associated ID tracks matched to the hard primary vertex with $p_T > 0.5$ GeV or both following conditions are met: $p_T^{\text{muon}}/p_T^{\text{jet}} > 0.5$ and $p_T^{\text{muon}}/\sum_{i\in\text{jet}} p_T^{\text{track}_i} > 0.7$, where $\sum_{i\in\text{jet}} p_T^{\text{track}_i}$ is the p_T sum of tracks in the jet matched to the primary vertex with $p_T > 0.5$ GeV. For any surviving jet, muons within $\min\Delta R(0.4, 0.04 + 10 \text{ GeV}/p_T^{\text{electron}})$ of the jet are removed.

Objects, reconstructed as indicated in this Chapter, are used in the following in the $h(\to b\bar{b}) + E_T^{miss}$ analysis. A summary of these objects can be found in Table 5.4

Object	Kinematics	Selection	Additional
Small-R Jets *central*	$p_T > 20$ GeV $\|\eta\| < 2.5$	anti-k_T R =0.4 EM-Topoclusters	b-Tagging: MV2c10 at 70% W.P.
Small-R Jets *forward*	$p_T > 30$ GeV $2.5 \leq \|\eta\| < 4.5$	anti-k_T $R = 0.4$ EM-Topoclusters	
Large-R Jets	$p_T > 200$ GeV $\|\eta\| < 2.0$	anti-k_T $R = 1.0$ LCW-Topoclusters Trimmed ($R_{\text{sub}} = 0.2$, $f_{\text{cut}} = 0.05$)	
Track Jets	$p_T > 10$ GeV $\|\eta\| < 2.5$	anti-k_T $R = 0.2$ ID Tracks $N_{\text{track}} \geq 2$	Ghost-associated $p_T^{\text{track}} > 0.4$ GeV,$\|\eta^{\text{track}}\| < 2.5$ Track quality requirements (Subsec. 5.1.3) b-Tagging: MV2c10 at 70% W.P.
Electrons	*V-loose*: $p_T > 7$ GeV, $\|\eta\| < 2.47$ *Z-signal*: $p_T > 27$ GeV, $\|\eta\| < 2.47$ *W-signal*: $p_T > 27$ GeV, $\|\eta\| < 2.47$	Loose LH Loose LH Tight LH	LooseTrackOnly Isolation LooseTrackOnly Isolation FixedCutHighPtCaloOnly Isolation For all electrons: $d_0/\sigma_{d_0} < 5$, $\|\Delta z_0 \cdot \sin(\theta)\| < 0.5$ mm
Muons	*V-loose*: $p_T > 7$ GeV, $\|\eta\| < 2.7$ *Z-signal*: $p_T > 25$ GeV, $\|\eta\| < 2.5$ *W-signal*: $p_T > 25$ GeV, $\|\eta\| < 2.5$	Loose ID Loose ID Medium ID	LooseTrackOnly Isolation LooseTrackOnly Isolation FixedCutHighPtTrackOnly Isolation For all muons: $d_0/\sigma_{d_0} < 3$, $\|\Delta z_0 \cdot \sin(\theta)\| < 0.5$ mm
Taus	*Standard* tau: $p_T > 20$ GeV, $\|\eta\| = [0, 1.37] \cap [1.52, 2.5]$ *Extended* tau: $p_T > 20$ GeV, $\|\eta\| < 2.5$	Loose BDT ID $1 \leq N_{\text{track}} \leq 4$, $\Delta\phi(\vec{\tau}, \vec{E}_T^{miss}) \leq \pi/8$	$N_{\text{track}} = 1$ or 3 with ΔR(track,tau) < 0.2 $p_T^{\text{track}} > 1$ GeV, track quality requirements ID tracks matched to PV, $p_T^{track} > 1$ GeV Requirements on small-R jet seed.
E_T^{miss}		E_T^{miss}: TST E_T^{miss} algorithm $\vec{p}_T^{miss} = -\sum \vec{p}_T^{\text{track}}$	Hard term: all reconstructed objects. Soft term: ID tracks matched to PV, not associated with any object. $p_T^{\text{track}} > 0.5$ GeV,$\|\eta\|^{\text{track}} < 2.5$, matched to PV.

Table 5.4: Summary of objects used in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis.

6 Event Selection

In order to search for BSM theories giving rise to the signature $h(\rightarrow b\bar{b}) + E_T^{miss}$, one must select a dataset enriched with signal-like events while suppressing background-like events, potentially originating from background processes (see Section 6.3). This can be achieved by exploiting the kinematical and topological features of signal and background events, as well as of objects expected to be found either in signal or background events.

6.1 Signal Region Event Selection

The signal region (SR) is designed to target a signal-enriched phase space, considering the $h \rightarrow b\bar{b}$ decay topology, the E_T^{miss} distribution from dark matter particles and the hadronic activity in a signal event, while addressing features that may allow background processes to mimic the signal or that are exclusive of background topologies. Various signal models lead to the production of dark matter particles carrying varying momentum, giving rise to varying E_T^{miss} in an event. The dark matter particles are produced in association with a Higgs boson, such that the boost of the Higgs decay depends on the momentum carried by the dark matter particles. For moderate boosts, the $h \rightarrow b\bar{b}$ decay can be reconstructed as two small-R jets, while for high boosts, the $h \rightarrow b\bar{b}$ decay is more efficiently reconstructed as a single large-R jet, due to the merging of b-jets in the calorimeters. The region of the SR with moderate E_T^{miss} is called *resolved regime* while the region of the SR with large E_T^{miss} (and Higgs Lorentz boost) is called *merged regime*.

In the event selection, only objects selected as described in Chapter 5 are used. The SR is a 0 lepton region, containing no leptons, which is split into eight orthogonal categories based on the b-tag multiplicity (two categories) and on the E_T^{miss} present in the event (four categories), as will be introduced in Subsections 6.1.1 and 6.1.2.

The event selections defining the SR has a set of selections common to both resolved 6.1.1 and merged 6.1.2 regimes. These common selections are as follows:

- **Good-run Lists:** Only good-run lists datasets in 2015 and 2016 are used in the analysis, as indicated in Section 4.1.
- **Event Cleaning:** Veto data events with corrupted or incomplete events due to failing in various subsystems (e.g. tile calorimeter, semiconductor tracker, etc).
- **Vertex:** Require at least one reconstructed vertex with at least 2 associated tracks.
- **Jet Cleaning:** Veto events with jets failing the loose working point of the jet cleaning algorithm [104]. This ensures a good measurement of E_T^{miss}.
- E_T^{miss} **trigger:** Use of E_T^{miss} trigger items as indicated in Subsection 4.3.1.

- **Lepton Veto:** Veto events with V-loose electrons or with V-loose muons.
- E_T^{miss} **lower cut:** Only events with $E_T^{miss} > 150$ GeV are accepted.
- p_T^{miss} **lower cut:** Only events with $p_T^{miss} > 30$ GeV are accepted in the 1 tag subregion. It targets vetoing non-collision background events.
- **Alignment of E_T^{miss},p_T^{miss}:** Only events with $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{miss}) < \pi/2$ are accepted. In events with real unobserved particles, giving rise to real missing transverse momentum, the E_T^{miss} and p_T^{miss} (see Section 5.3) are expected to be aligned. This cut targets reducing multijet events, because in case of a dijet event with a mismeasured jet, the p_T^{miss} will be small and have no preferred direction.
- **Mismeasured jet:** Veto events with $\min\left[\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{j_{1,2,3}})\right] < \pi/9$, where the $\vec{p}_T^{j_{1,2,3}}$ refers to the $\vec{p}_T$ of the first, second or third sorted jet, with a sorting prioritising central jets ordered in p_T and then forward jets ordered in p_T. If a hard small-R jet in dijet events is mismeasured overestimating its p_T, the E_T^{miss} resulting from this measurement azimuthally points in the direction of the non-mismeasured hard jet (which is the opposite direction of the mismeasured jet). On the other hand, if a jet in dijet events is mismeasured underestimating its p_T, the resulting E_T^{miss} azimuthally points in the direction of the mismeasured jet. This requirement leads to vetoing events, in which any of the three leading jets is angularly close to the E_T^{miss}, reducing the multijet background process. This and the previous event selections are referred to as *anti-QCD* event selections.

Events are split into the resolved and merged regimes according to the E_T^{miss} in the event. Events with $E_T^{miss} < 500$ GeV are classified in the resolved regime, whereas events with $E_T^{miss} > 500$ GeV are classified in the merged regime. Both regimes are described in the following.

6.1.1 Resolved Regime

The resolved regime is defined as the signal phase space with moderate E_T^{miss}, $E_T^{miss} <$ 500 GeV, in the 0 lepton region. In presence of moderate E_T^{miss}, the hadronic decay products of the $h \to b\bar{b}$ decay are reconstructed as two small-R jets and b-tagging is used to assess the flavour. The jets are first sorted by central b-tagged jets and then central non-b-tagged jets, with each subset being ordered in decreasing p_T. The first and second jets of this ordering are assigned to the $h \to b\bar{b}$ decay, and the four-momentum sum of both jets is referred to as *Higgs, h, candidate* in the resolved regime. Events are categorised based on the b-tag multiplicity and the E_T^{miss} in the event. If an event has only one b-jet, the event is sorted into the 1 tag subregion, whereas if it has two or more b-jets, it is sorted into the 2 tag subregion. Events without b-jets are discarded. In the resolved regime, events are split into three E_T^{miss} subregions: 150 GeV $< E_T^{miss} <$ 200 GeV, 200 GeV $< E_T^{miss} <$ 350 GeV and 350 GeV $< E_T^{miss} <$ 500 GeV.

A number of event selections in the resolved regime are inherited from the $h(\to b\bar{b}) + E_T^{miss}$ analysis using 3.2 fb^{-1} in 2015 [16] as follows:

- **Kinematics regime:** Only events with $E_T^{miss} < 500$ GeV are accepted. This ensures, the $h \to b\bar{b}$ decay is reconstructed as two separated small-R jets.
- **Minimum jet multiplicity:** Only events with $N_{\text{jet}^{\text{central}}} \geq 2$ are accepted. From the $h \to b\bar{b}$ decay, events must have at least two central jets.
- **Minimum jet p_T in $h \to b\bar{b}$ decay:** Only events with $p_T^{\text{j1 OR 2}} > 45$ GeV are accepted, where $p_T^{\text{j1 OR 2}}$ refers to the p_T of either the leading or subleading jet. If any of these jets has a p_T larger than 45 GeV, the event is accepted.
- **Trigger efficiency mismodelling removal:** Only events with $\sum_{i=1}^{2(3)} p_T^{j_i} > 120(150)$ GeV are accepted, for events with up to two (more than two) central jets. Jets are sorted in decreasing p_T. This selection was designed to remove a region of phase space that is mismodelled in simulation, because of a dependence of the E_T^{miss} trigger efficiency on the jet activity.
- **Boost of Higgs boson:** Only events with $\Delta\phi(\vec{p}_T^{\,j_1}, \vec{p}_T^{\,j_2}) < 7\pi/9$ are accepted. From the boost of the $h \to b\bar{b}$ decay in $h(\to b\bar{b}) + E_T^{miss}$, the resulting b-jets are azimuthally close to each other. This requirement also reduces multijet events, as the pair of hard jets in dijet events are expected to be produced in average back-to-back.
- $h(\to b\bar{b}) + E_T^{miss}$ **topology:** Only events with $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{\,h}) > 6\pi/9$ are accepted, where $\vec{p}_T^{\,h}$ refers to the h candidate $\vec{p}_T$. In the case of a signal event with signature $h(\to b\bar{b}) + E_T^{miss}$, the E_T^{miss} and the $h \to b\bar{b}$ decay are in average back-to-back.

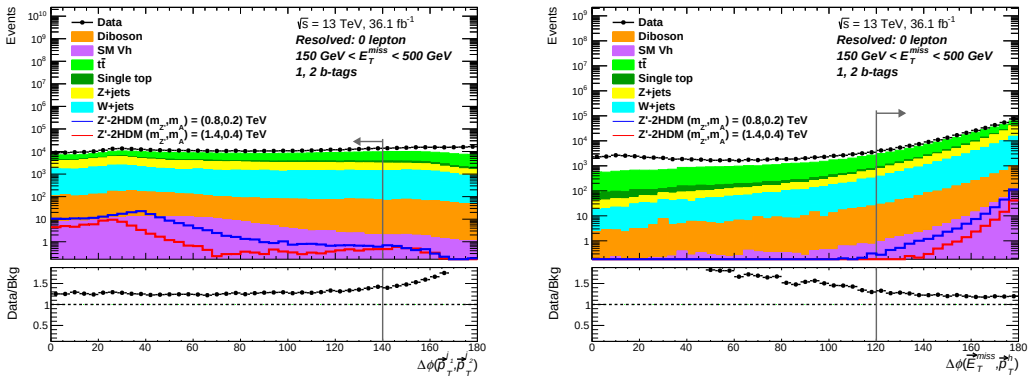

Figure 6.1: Prefit distributions of $\Delta\phi(\vec{p}_T^{\,j_1}, \vec{p}_T^{\,j_2})$ (left) and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{\,h})$ (right) in the 0 lepton region, resolved regime, 1 tag and 2 tag. Two representative $Z' - 2HDM$ models, with cross-section normalised to 10 fb and $(m_{Z'}, m_A) = (1.4\text{ TeV}, 0.4\text{ TeV}), (0.8\text{ TeV}, 0.2\text{ TeV})$, are included as a red and blue line, respectively, along with their cut values shown as a grey line. No background uncertainties are shown.

In Figure 6.1, the prefit distributions of $\Delta\phi(\vec{p}_T^{\,j_1}, \vec{p}_T^{\,j_2})$ and $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{\,h})$ are shown in logarithmic scale upon applying the event selections mentioned above. The data exceeds the background due to the absence of the multijet background, which plays a more sizeable role at this stage of the event selection. A dedicated data-driven multijet estimation is performed to assess the multijet contribution to the background in the SR, as explained in Subsection 8.3. The event selections remove regions having negligible sensitivity to the representative signals. The regions removed are also those where the disagreement between data and MC background is more prominent.

Optimised Event Selections in the Resolved Regime

Previous $h(\to b\bar{b}) + E_T^{miss}$ analyses, e.g. [16], have used event selections similar to those described so far, increasing the sensitivity of the search mainly by using more integrated luminosity. A set of new event selections were developed for the current $h(\to b\bar{b}) + E_T^{miss}$ publication, in order to reduce the $t\bar{t}$ background while increasing the sensitivity of the search. $t\bar{t}$ was the dominant background process in the $h(\to b\bar{b}) + E_T^{miss}$ analysis using 3.2 fb^{-1}, amounting *up to 80% of the total background* in the SR, resolved regime, 2 tag, as shown in Table 2, Reference [16]. These new event selections, which will be described in more detail in Chapter 7, are as follows:

- **Veto of additional b-jets:** Veto events containing more than two b-jets. Signal events contain two b-jets at most from the $h \to b\bar{b}$ decay, so any additional heavy flavour activity may be a sign for background. This reduces $t\bar{t}$ and single top events, where a third reconstructed b-jet can originate from mistagging of c-quark-initiated jets (as of now called *c-jets*) as b-jets in a W hadronic decay. This will be explained in Subsection 7.1.1.

- **Veto of *standard* taus:** Veto events containing any *standard* tau lepton, see Subsection 5.2.3. Signal events are expected to contain no tau leptons in the final state particles, whereas background processes with a $W \to \tau\nu_\tau$ can be then rejected with a tau veto.

- **Veto of *extended* taus:** Veto events containing any *extended* tau lepton, see Subsection 5.2.3. *Extended* tau objects are hadronic tau leptons which were not identified by the loose ID criteria, but still identified by the new *extended* ID criteria for reconstruction of hadronic taus.

- **H_T ratio:** Most of the hadronic activity in a signal event is expected from the $h \to b\bar{b}$ decay, which leads to requiring a certain fraction of the H_T to come from the h candidate, where the H_T is the scalar p_T sum of all jets present in the event in the resolved regime. Only events in which the scalar p_T sum of jets other than the h candidate and, if present, the leading additional jet is smaller than $0.37 \cdot H_T$ are accepted.

- ΔR_{bb}**:** Only events with $\Delta R(\vec{p}^{j_1}, \vec{p}^{j_2}) < 1.8$ are accepted. The jets from the $h \to b\bar{b}$ decay are expected to be geometrically close, whereas b-jets in a $t\bar{t}$ event have in average larger angular separation.

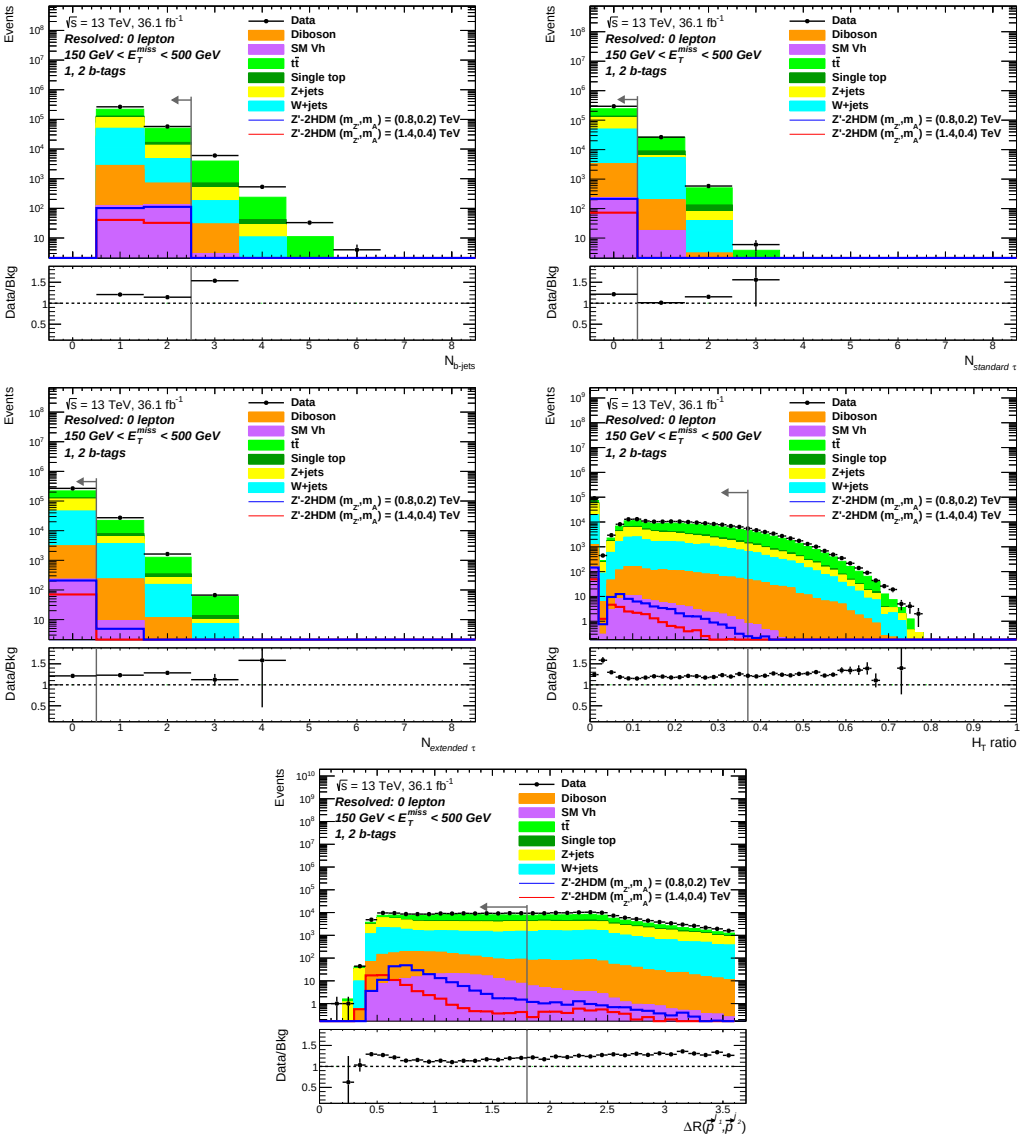

Figure 6.2: Prefit distributions of optimised event selections upon applying consecutively each selection, in the 0 lepton region, resolved regime, 1 tag and 2 tag. The event selection distributions are (as shown from left to right, from top to bottom): multiplicity of b-jets, multiplicity of *standard* taus, multiplicity of *extended* taus, H_T ratio ($\sum_{i=4}^{\infty} p_T^{j_i}/H_T$, where $H_T = \sum_{i=1}^{\infty} p_T^{j_i}$) and $\Delta R(\vec{p}^{j_1}, \vec{p}^{j_2})$. Two representative $Z' - 2HDM$ models, with cross-section normalised to 10 fb and $(m_{Z'}, m_A) = (1.4\text{ TeV}, 0.4\text{ TeV}), (0.8\text{ TeV}, 0.2\text{ TeV})$, are included as a red and blue line, respectively, along with their cut values shown as a grey line. No background uncertanties are shown.

In Figure 6.2, the prefit distributions of the optimised event selections are shown in logarithmic scale upon applying consecutively each selection in the order given above.

The data exceeds the background due to the absence of the multijet background. After including a data-driven multijet estimate, the SM background is brought into agreement with data through the statistics fit model, on account of production rate adjustment of non-multijet background processes and shape adjustment of the E_T^{miss} probability density distribution, as seen later on in Chapter 11. As shown in Figure 6.2 for two representative signal models, the optimised event selections remove regions with negligible sensitivity to the signal.

6.1.2 Merged Regime

The merged regime is defined as the signal phase space with high E_T^{miss}, $E_T^{miss} > 500$ GeV, in the 0 lepton region. At first order, the p_T of the Higgs boson h is equal to the E_T^{miss} present in the event, as the dark matter particles giving rise to the E_T^{miss} are produced recoiling against the h boson. Since the b-quarks from the $h \to b\bar{b}$ decay are angularly separated as $\sim 2m^h/p_T^h$, where m^h and p_T^h are the mass and the p_T of the h boson, for high E_T^{miss}, the angular separation of the b-quarks decreases, until the $h \to b\bar{b}$ decay products can no longer be reconstructed as two separated small-R jets. In this case, for high Lorentz boost of the h boson, the $h \to b\bar{b}$ decay products are reconstructed as a single large-R jet. The leading large-R jet, denoted as J, is referred to as the *Higgs candidate* in the merged regime. The two leading-p_T track jets ghost-associated to the h candidate are used to assess the flavour content of the h candidate. If both of them are b-tagged, the event is sorted into the 2 tag subregion, whereas if only one of them is b-tagged, it is sorted into the 1 tag subregion. If none of the two leading-p_T ghost-associated track jets are b-tagged, the event is discarded. An object Y is said to be *non-associated* if $\Delta R(\vec{p}^h, \vec{p}^Y) > 1.0$.

In addition to the event selections common to both regimes, see Section 6.1, the merged regime has the following event selections:

- **Kinematics regime:** Only events with 500 GeV $< E_T^{miss}$ are accepted. This ensures, the $h \to b\bar{b}$ decay products are reconstructed as a single large-R jet.
- **Minimum large-R jet multiplicity:** Only events with $N_{\text{large-}R\text{ jet}} \geq 1$, having at least one ghost-associated track jet, are accepted.

These event selections are inherited from the $h(\to b\bar{b}) + E_T^{miss}$ analysis using 3.2 fb^{-1} in 2015 [16].

Optimised Event Selections in the Merged Regime

The $t\bar{t}$ background as dominant background amounts *up to 40% of the total background* in the SR, merged regime, SR, in the $h(\to b\bar{b}) + E_T^{miss}$ analysis using 3.2 fb^{-1} [16]. The corresponding optimised event selections in the merged regime aiming at reducing $t\bar{t}$ are as follows:

- **Veto of non-associated b-jets:** Veto events containing any non-associated b-jet. The $h \to b\bar{b}$ decay products are expected to be contained in the large-R jet, so no

b-jet is expected outside the *h* candidate in signal events. In a $t\bar{t}$ event, one of the top quarks is reconstructed as the *h* candidate, while the other one decays giving rise to a non-associated *b*-jet, which is used to veto the event.

- **Veto of non-associated *standard* taus:** Veto events containing any non-associated *standard* tau lepton, see Subsection 5.2.3.

- **Veto of non-associated *extended* taus:** Veto events containing any *extended* tau lepton, see Subsection 5.2.3.

- H_T **ratio:** Only events in which the scalar sum of p_T of non-associated small-*R* jets is smaller than $0.57 \cdot H_T$ are accepted, where H_T is the scalar p_T sum of all non-associated small-*R* jets plus the p_T of the leading large-*R* jet. As in the resolved regime, this condition follows the idea that most of the hadronic activity in a signal event is expected from the $h \to b\bar{b}$ decay.

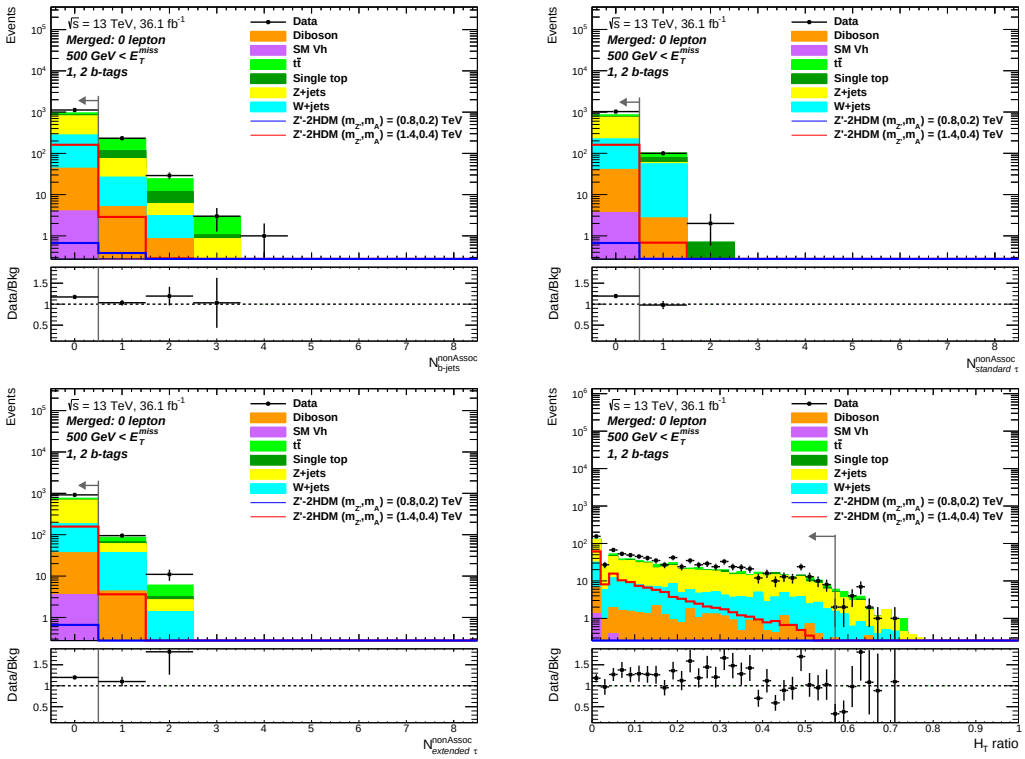

Figure 6.3: Prefit distributions of the optimised event selections upon applying consecutively each selection, in the 0 lepton region, merged regime, 1 tag and 2 tag. The event selection distributions are (as shown from left to right, from top to bottom): multiplicity of non-associated *b*-jets, multiplicity of non-associated *standard* taus, multiplicity of non-associated *extended* taus, H_T ratio ($\sum_{i=1}^{\infty} p_T^{j_i^*}/H_T$, where $H_T = p_T^J + \sum_{i=1}^{\infty} p_T^{j_i^*}$, $p_T^{j_i^*}$ is the p_T of non-associated small-*R* jets and p_T^J is the *h* candidate p_T). Two representative $Z' - 2HDM$ models, with cross-section normalised to 10 fb and $(m_{Z'}, m_A) = (1.4\text{ TeV}, 0.4\text{ TeV}), (0.8\text{ TeV}, 0.2\text{ TeV})$, are included as a red and blue line, respectively. No background uncertainties are shown.

In Figure 6.3, the prefit distributions of the optimised event selections are shown in logarithmic scale upon applying consecutively each selection in the order given above. Similarly to the resolved regime, the SM background is brought into agreement with data through the statistics fit model. The optimised event selections remove regions having negligible sensitivity to the representative signals, as seen in Figure 6.3.

6.1.3 Mass Discriminant

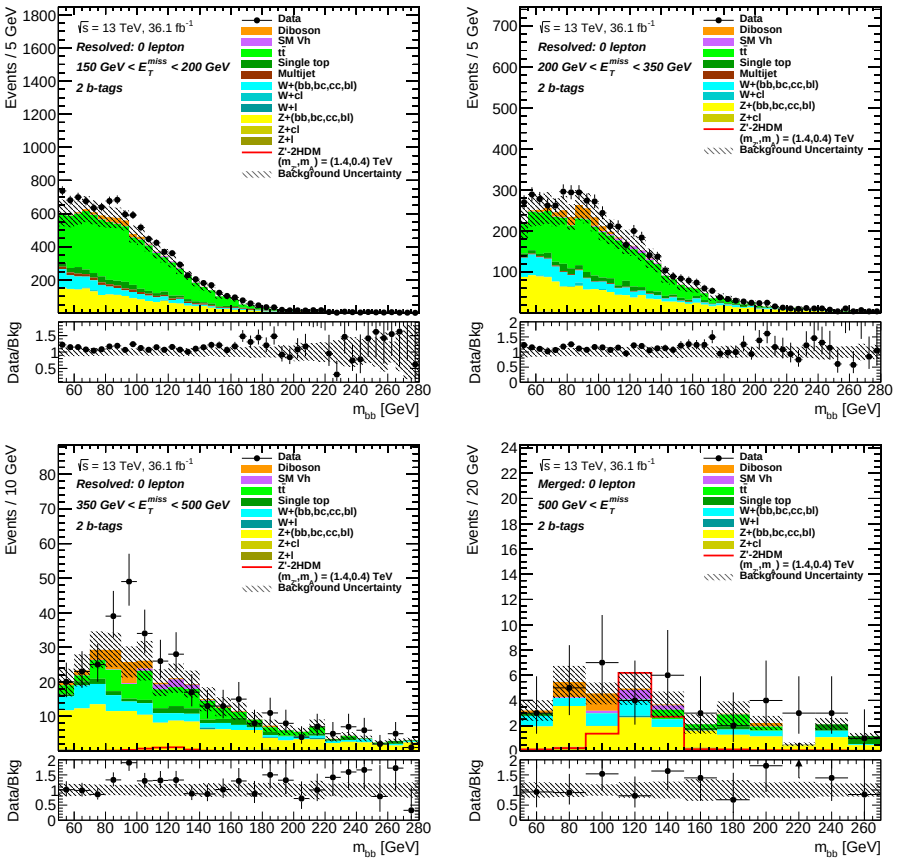

Figure 6.4: Prefit distribution of the h candidate mass after full event selection in the 0 lepton region, 2 tag, for the four E_T^{miss} categories: 150 GeV $< E_T^{miss} <$ 200 GeV, 200 GeV $< E_T^{miss} <$ 350 GeV and 350 GeV $< E_T^{miss} <$ 500 GeV of the resolved regime, and merged regime. A representative $Z' - 2HDM$ model, with cross-section normalised to 10 fb and $(m_{Z'}, m_A) = (1.4\text{ TeV}, 0.4\text{ TeV})$ is shown as a red curve. The total background uncertainty is shown as a hatched band. The labelling of $V + jets$ is defined in Subsection 6.3.

The primordial discriminant of the search in the SR is the invariant mass of the h candidate[1]. In the analysis, one searches for an excess over the SM background prediction in the invariant mass spectrum of the h candidate. In Figure 6.4, the prefit distribution of the h candidate mass is shown after full event selection in the 0 lepton region, 2 tag, for

[1]The E_T^{miss} distribution and b-tag multiplicity give discriminant sensitivity to signal as well.

the four E_T^{miss} categories. The data-driven multijet estimate is only considered in the first two E_T^{miss} categories, as it is negligible for $E_T^{miss} > 350$ GeV. Data and SM background are consistent within uncertainties.

6.2 Event Selection Signal Efficiency

The $Z' - 2HDM$ is the BSM signal model used for the statistical interpretation of the analysis. In Figure 6.5 (Figure 6.6), the cutflows of $Z' - 2HDM$ models as a function of the event selections are shown in the 0 lepton region, resolved (merged) regime, inclusive in b-tag. After the lepton veto, the events are split into resolved and merged regime, and then the event selections are *consecutively* applied *in the order given above*. The event selection efficiency is calculated with respect to the yields before the E_T^{miss} triggers, referred to as *preselection*. Three representative $Z' - 2HDM$ models are considered with $(m_{Z'}, m_A) = (0.8\text{ TeV}, 0.2\text{ TeV})$, $(1.4\text{ TeV}, 0.4\text{ TeV})$ and $(2.2\text{ TeV}, 0.4\text{ TeV})$. The E_T^{miss} spectra of $Z' - 2HDM$ $(0.8\text{ TeV}, 0.2\text{ TeV})$ and $Z' - 2HDM$ $(2.2\text{ TeV}, 0.4\text{ TeV})$ are driven by the resolved and merged regime, respectively, while $Z' - 2HDM$ $(1.4\text{ TeV}, 0.4\text{ TeV})$'s E_T^{miss} spectrum populates both regimes. In the merged regime, $Z' - 2HDM$ $(2.2\text{ TeV}, 0.4\text{ TeV})$ has a higher efficiency in 1 tag than in 2 tag. The reason is that in very boosted topologies, the track jets ghost-associated to the h candidate start merging, so the double b-tagging efficiency of the track jets decreases, being more likely to classify signal events in 1 tag.

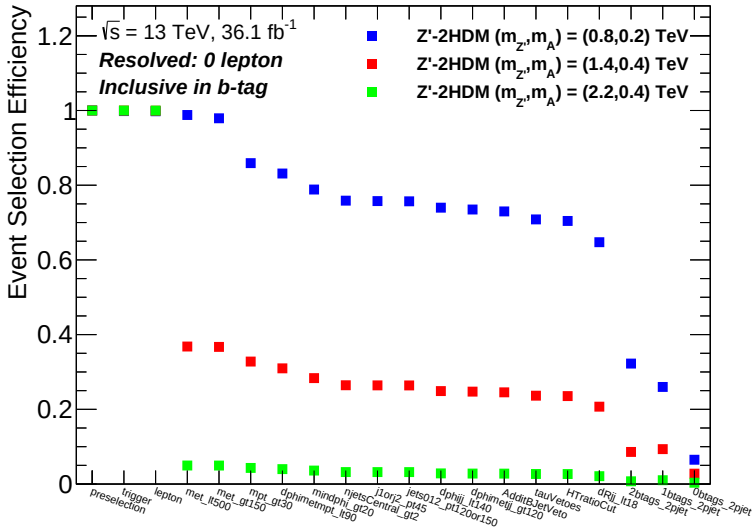

Figure 6.5: Cutflow of representative $Z' - 2HDM$ models as a function of the event selections in the 0 lepton, resolved regime, inclusive in b-tag. The $Z' - 2HDM$ models are $(m_{Z'}, m_A) = (0.8\text{ TeV}, 0.2\text{ TeV})$, $(1.4\text{ TeV}, 0.4\text{ TeV})$ and $(2.2\text{ TeV}, 0.4\text{ TeV})$ shown as blue, red and green marks, respectively.

In Figure 6.7, the $Acceptance \times Efficiency$ is shown for all generated $Z' - 2HDM$ models. The efficiency considered is the efficiency to reconstruct an event and pass all

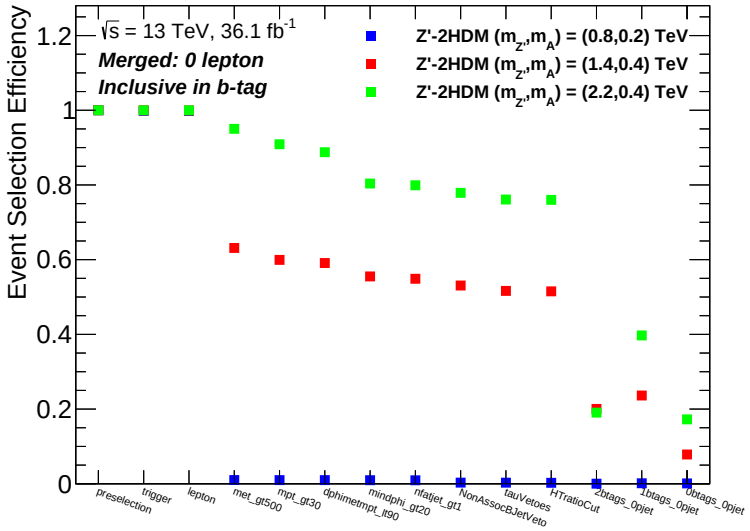

Figure 6.6: Cutflow of representative $Z' - 2HDM$ models as a function of the event selections in the 0 lepton, merged regime, inclusive in b-tag. The $Z' - 2HDM$ models are $(m_{Z'}, m_A) = (0.8\,\text{TeV}, 0.2\,\text{TeV})$, $(1.4\,\text{TeV}, 0.4\,\text{TeV})$ and $(2.2\,\text{TeV}, 0.4\,\text{TeV})$ shown as blue, red and green marks, respectively.

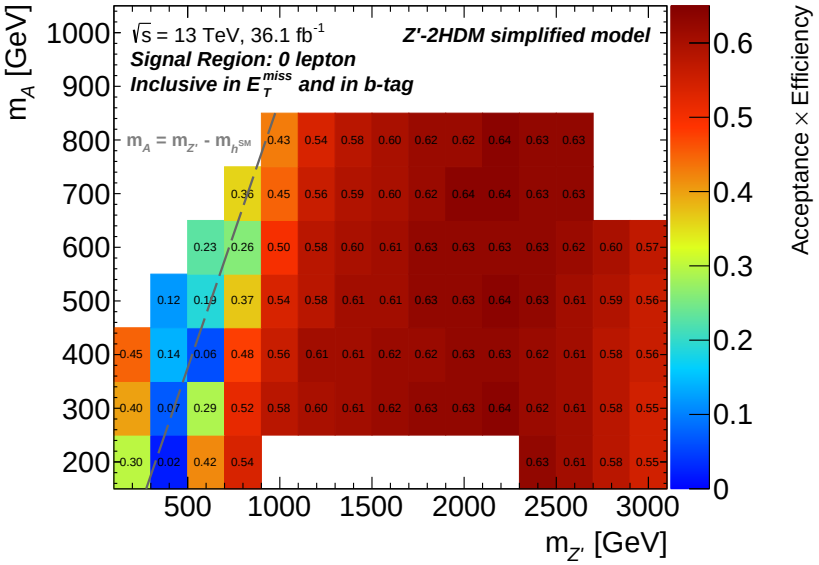

Figure 6.7: $Acceptance \times Efficiency$ for all generated $Z' - 2HDM$ models in the 0 lepton, inclusive in b-tag and in E_T^{miss}. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass.

event selections, inclusive in b-tag and in E_T^{miss}. The $Acceptance \times Efficiency$ is calculated as the ratio of the number of successfully reconstructed events passing all event selections (in which the kinematic acceptance is included) to the total number of generated events. The total number of generated events is calculated as $\sigma \times \mathscr{L}$, where σ is the cross-section normalised to 10 fb for all signal models, and $\mathscr{L} = 36.1$ fb^{-1} is the

integrated luminosity. In the on-shell region, the efficiency increases as $m_{Z'}$ increases at a fixed m_A, and increases as m_A decreases at a fixed $m_{Z'}$, since more E_T^{miss} is present in the event. For very boosted topologies (over $m_{Z'} \sim 2.2$ TeV), the efficiency decreases as $m_{Z'}$ increases at a fixed m_A, due to the merging of the track jets used to assess the flavour content of the h candidate.

6.3 Background Processes

The background processes present in the SR are as follows (the data-driven multijet estimate will be introduced in Subsection 8.3 of Chapter 8):

Top-pair production

Top-pair production, or $t\bar{t}$, background is a main background process in the SR. In Figure 6.8, left, the tree-level Feynman diagrams of $t\bar{t}$ production at the LHC are shown. The decay modes of $t\bar{t}$ can be classified based on the decay modes of the W bosons in the event. If both W bosons decay leptonically, the $t\bar{t}$ event is called *dileptonic*; if only one W decays leptonically, it is called *semileptonic*, as seen schematically in Figure 6.8, right; and if both W bosons decay hadronically, it is called *hadronic*. Hadronic $t\bar{t}$ events are successfully excluded from the SR, as they have no real E_T^{miss} and any jet mismeasurement leading to fake E_T^{miss} is strongly reduced by the anti-QCD selections. The most efficient way for $t\bar{t}$ to be selected into the SR is through a W boson decaying as $W \to \tau\nu_\tau$, where the tau decays hadronically. For such $t\bar{t}$ event in the SR, in addition to the two real b-jets, there is real E_T^{miss} in the event and the hadronic tau is reconstructed as a jet, mimicking a QCD-originated jet. Therefore, semileptonic $t\bar{t}$, with the presence of a hadronic tau, is the dominant component of the $t\bar{t}$ background process in the SR. Dileptonic $t\bar{t}$ contributes less considerably to the SR, with absence of di-electron and di-muon events, which are strongly reduced by the lepton veto.

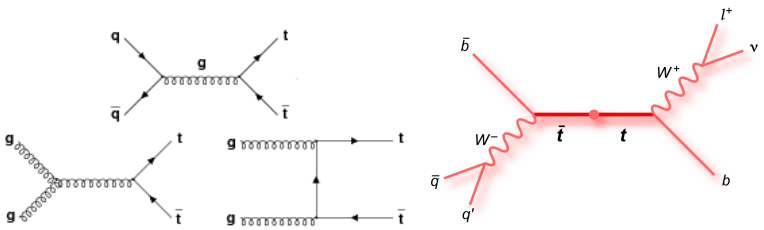

Figure 6.8: Left: Tree-level Feynman diagrams of top-pair production as produced at the LHC. For an $t\bar{t}$ event to be selected into the SR, a W boson decays as $W \to \tau\nu_\tau$, giving rise to real E_T^{miss}. $t\bar{t}$ events have two real hard b-quarks. Right: Scheme of a semileptonic $t\bar{t}$ event.

In Figure 6.9, the truth leptonic flavour distribution of final state particles normalised to unit area is shown for background processes in the SR, resolved regime, 1 tag (left) and 2 tag (right) subregions after full event selection. Events in the sidebands of the h candidate mass spectrum are excluded, considering only events within the h candidate

mass window $(70\,\text{GeV}, 150\,\text{GeV}$. About 70% (70%) of $t\bar{t}$ events in 2 tag (1 tag) contain a hadronic tau from the $W \to \tau\nu_\tau$ decay, giving rise to real E_T^{miss}. The hadronic tau mimics a QCD-originated jet, as it was identified neither by the BDT tau ID nor by the *extended* tau ID. By exploiting the topology of semileptonic $t\bar{t}$, a new algorithm was developed to further reduce $t\bar{t}$ in the SR: the *Likelihood Ordering*, as introduced in Chapter 12.

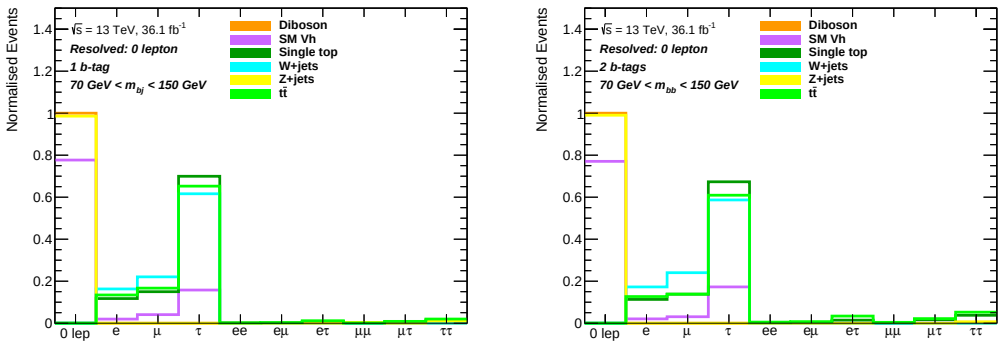

Figure 6.9: Truth leptonic flavour distribution of final state particles normalised to unit are for background processes in the SR, resolved regime, 1 tag (left) and 2 tag (right) within h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$, excluding the sidebands in the h candidate mass spectrum.

Z+jets and W+jets Production

$Z + jets$ is the dominant background process in the SR. In Figure 6.10, left, the tree-level Feynman diagrams of $Z + jets$, where the Z boson is associated with b-jets, are shown. As seen in Figure 6.9, $Z + jets$ events have no lepton (electron, muon or tau) as final state particle, so the decay of the Z boson is $Z \to \nu\nu$ for all $Z + jets$ events in the SR. $Z \to \nu\nu + jets$, with real E_T^{miss}, is a semi-irreducible[2] background in the SR, which is selected into the 2 tag subregion if the Z boson is produced in association with b-jets.

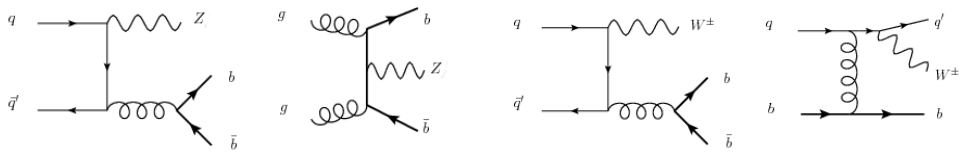

Figure 6.10: Left: Tree-level Feynman diagrams of $Z + jets$, where Z boson is associated with b-jets. $Z \to \nu\nu + jets$ is an irreducible background in the SR. Right: Tree-level Feynman diagrams of $W + jets$, where W boson is associated with b-jets. For $W \to \tau\nu_\tau$, where the tau decays hadronically, the $W + jets$ event has real E_T^{miss}, being able to be selected into the SR.

In Figure 6.10, right, the tree-level Feynman diagrams of $W + jets$, where the W boson is associated with b-jets, are shown. As in the case of $t\bar{t}$, the ideal decay mode of W boson for the $W + jets$ event to be selected into the SR is $W \to \tau\nu_\tau$, where the tau decays

[2] $Z(\to \nu\nu) + b\bar{b}$ has the same signature than $h(\to b\bar{b}) + E_T^{miss}$ but it is not resonant in the h candidate mass spectrum.

hadronically. In this case, there is real E_T^{miss}, b-jets and a hadronic tau mimicking a QCD-originated jet. About 60% of $W + jets$ events in the SR, resolved regime, have a hadronic tau as final state particle, as seen in Figure 6.9.

Throughout the thesis, where appropriate, $V + jets$ events, with V representing a W or Z boson, are assigned a flavour labelling based on the geometrical matching between reconstructed jets and truth jets, where the flavour content of the latter is known at truth level. Only two jets are used for this labelling, corresponding to the leading b-jets. In case of less than 2 b-jets, the leading non-b-tagged jet is considered. *Heavy flavour* refers to $V + jets$ events having two reconstructed jets associated with either two truth b-jets, a truth b-jet and a truth c-jet, two truth c-jets, or a truth b-jet and a truth light-jet (originating from a u-/d-/s-quark). Heavy flavour $V + jets$ events are denoted as $V + (bb, bc, cc, bl)$ or $V + HF$.

Single Top Production

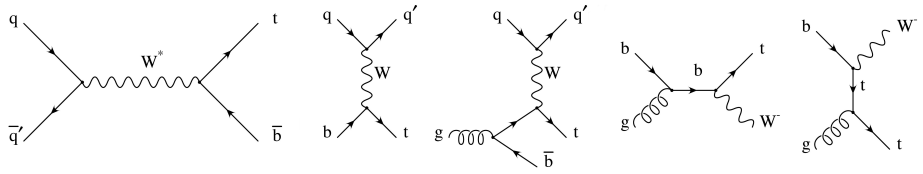

Figure 6.11: Tree-level Feynman diagrams of single top production. For the event to be selected into the SR, a W boson present in the event decays as $W \to \tau\nu_\tau$, giving rise to real E_T^{miss}.

In Figure 6.11, the tree-level Feynman diagrams of *single top* production are shown. The top quark decays as $t \to Wb$. As in the case of $t\bar{t}$ and $W + jets$, for $W \to \tau\nu_\tau$, where the tau decays hadronically, there is real E_T^{miss} in the event and no lepton to be vetoed, as required in the SR. In 2 tag (1 tag), about 75% (70%) of single top events have a hadronic tau, originating from the $W \to \tau\nu_\tau$ decay, see Figure 6.9. For some diagrams, single top is produced in association with a b-jet, being selected into the 2 tag subregion if successfully b-tagged. For some diagrams, the top quark is produced in association with a W boson. If one of the two W bosons decays hadronically, giving rise to a c-jet, the c-jet may be mistagged as a b-jet (8% mistagging efficiency for MV2c10 b-tagging at 70% operating point [105], Table 2), being selected into the 2 tag subregion, or into the 1 tag subregion if not mistagged.

SM Vh and Diboson Production

The SM Vh and Diboson production are background processes that are resonant in the h candidate mass spectrum. In Figure 6.12, left, the tree-level Feynman diagram of SM Vh production is shown, where V stands for a W or Z boson. SM Zh is an irreducible background process, as the Higgs decays as $h \to b\bar{b}$ and the Z boson as $Z \to \nu\nu$, giving rise to real E_T^{miss}. About 80% of SM Vh events in the SR, resolved regime, correspond to this process. The resonance is exactly at the Higgs mass in the h candidate mass spectrum,

overlapping with the signal. SM Wh is more likely selected into the SR, if the W boson decays as $W \to \tau\nu_\tau$, where the tau decays hadronically mimicking a QCD-originated jet. In this case, there is real E_T^{miss} in the event, and the Higgs decays also as $h \to b\bar{b}$ for the event to be selected into the SR. About 17% of SM Vh events in the SR, resolved regime, correspond to SM Wh, with the W boson decaying as $W \to \tau\nu_\tau$, see Figure 6.9.

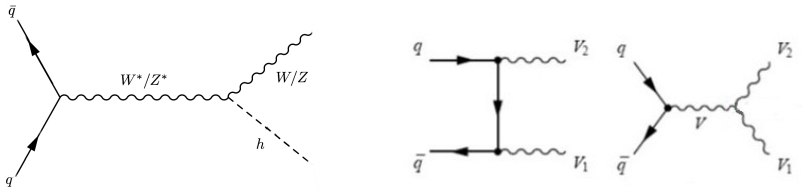

Figure 6.12: Left: Tree-level Feynman diagram of SM Vh production. For the event to be selected into the SR, the Higgs decays as $h \to b\bar{b}$ and the vector boson either as $Z \to \nu\nu$ or $W \to \tau\nu_\tau$ to give rise to real E_T^{miss}. Right: Tree-level Feynman diagram of diboson production. The final state particles can be WW, WZ and ZZ. For the event to be selected into the SR, a vector boson decays either as $Z \to \nu\nu$ or $W \to \tau\nu_\tau$ to give rise to real E_T^{miss}, while the other vector boson decays hadronically.

In Figure 6.12, right, the tree-level Feynman diagrams of diboson production are shown, where V_1 and V_2 stands for a W or Z boson in the diagram. The s-channel includes a triple gauge boson coupling of the form $WW\gamma$ or WWZ, so that the boson mediator V stands in this case for a W, Z or γ boson. The produced particles can be WW, WZ and ZZ. In order for the event to be selected into the SR, it needs to have real E_T^{miss}. Most diboson events favoured by the event selection of the SR correspond to the ZV production, where the Z boson decays as $Z \to \nu\nu$, giving rise to real E_T^{miss}, and the V vector boson (either Z or W) decays hadronically, as deduced from the fractional leptonic flavour distribution of final state particles in Figure 6.9. The diboson production is resonant at the vector boson mass V in the h candidate mass spectrum.

7 Optimised Event Selections

As shown in Table 2 of Reference [16], the $t\bar{t}$ background process dominates the total predicted Standard Model background in the SR, 2 tag and 1 tag for the previous published $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis iteration at 13 TeV using 3.2 fb^{-1} [16]. $t\bar{t}$ represents about 80% and 40% of the total background in the SR, 2 tag, resolved and merged regime, respectively. Since $t\bar{t}$ is a reducible main background process, it is crucial to refine the event selection by developing requirements targeting $t\bar{t}$ and other reducible background processes in order to improve the sensitivity of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ search to new physics.

As introduced in Subsections 6.1.1 and 6.1.2 for the resolved and merged regimes, the refined new event selections are the veto of additional b-jets, the veto of hadronic tau leptons, the H_T ratio cut and the ΔR_{bb} cut. These optimised event selections are described in Section 7.1 (veto of additional b-jets), Section 7.2 (veto of *standard* and *extended* taus), Section 7.3 (H_T ratio cut) and Section 7.4 (ΔR_{bb} cut). The definition of the optimised event selections are summarised in Table 7.1.

	Resolved regime	**Merged regime**
Veto of additional b-jets	< 3 b-jets	no non-associated b-jet
Tau veto	no *standard* tau	no non-associated *standard* tau
***Extended* tau veto**	no *extended* tau	no non-associated *extended* tau
H_T ratio cut	$\sum_{i=4}^{\infty} p_T^{j_i} \Big/ \sum_{i=1}^{\infty} p_T^{j_i} < 0.37$	$\sum_{i=1}^{\infty} p_T^{j_i^*} \Big/ \left(p_T^J + \sum_{i=1}^{\infty} p_T^{j_i^*}\right) < 0.57$
ΔR_{bb} cut	$\Delta R(\vec{p}^{\,j_1}, \vec{p}^{\,j_2}) < 1.8$	—

Table 7.1: Summary of optimised event selections in the resolved and merged regimes, as applied in the SR and CRs. p_T^J stands for the p_T of the leading large-R jet, which is the h candidate in the merged regime. $\vec{p}^{\,j_i^*}$ stands for the p_T of non-associated small-R jets in the merged regime, where *non-associated* means that $\Delta R(\vec{p}_T^{\,J}, \vec{p}^{\,j_i^*}) > 1.0$. The veto of additional b-jets, tau veto, *extended* tau veto, H_T ratio cut and ΔR_{bb} cut are explained in Sections 7.1, 7.2, 7.2.2, 7.3 and 7.4, respectively.

The figure of merit used to evaluate the performance of the optimised event selections is the *Asimov median significance* [106]

$$Z_0 = \sqrt{2\left[(s+b)ln(1+s/b) - s\right]}, \tag{7.1}$$

of a counting experiment with a test statistics under the background-only hypothesis, where the mean background yields b and the signal yields s are known with negligible uncertainty. An event selection improves the sensitivity of the search if the median significance increases upon applying the event selection. The evaluation of the optimised event selections with a full statistical fit model are presented in Chapter 11.

The *significance gain* ΔZ_0 of a given event selection X with cut value X_0 is defined as

$$\Delta Z_0[X \leq X_0] = \left(\frac{Z_0[X \leq X_0] - Z_0}{Z_0^{\text{baseline}}} \right), \tag{7.2}$$

where $Z_0[X \leq X_0]$ is the median significance with the requirement $X \leq X_0$, Z_0 is the median significance without applying the requirement on X and Z_0^{baseline} is the median significance of the so-called *baseline event selection*, which is defined as the event selection without applying any of the optimised event selections.

Throughout the chapter, the *fractional rejection* of a process (signal or background) is defined as the ratio of the number of rejected events by a given optimised event selection to the event yield at the baseline event selection. Only events within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$ are considered for the optimisation of the new refined event selections throughout the chapter, as this window is enriched with signal events. Representative $Z' - 2HDM$, Z'- and S-mediator signal models are used throughout the chapter to optimise the new refined event selections. After choosing an optimised cut value X_0, the fractional rejection by a given optimised event selection is evaluated for the $Z' - 2HDM$ grid, since the $Z' - 2HDM$ models are used for the interpretation of the analysis.

7.1 Veto of Additional B-jets

The *veto of additional b-jets* is motivated by the signal signature $h(\rightarrow b\bar{b}) + E_T^{miss}$, in which at most two hard b-jets in the event are expected (exactly two b-jets if both b-quark-initiated jets from the $h \rightarrow b\bar{b}$ decay are successfully b-tagged). The presence of any additional b-jet not associated to the $h \rightarrow b\bar{b}$ decay would indicate that the event is likely a background event. In the following, the vetoes of additional b-jets in the resolved and merged regimes are presented.

7.1.1 Veto of Additional B-jets in the Resolved Regime

From the $h \rightarrow b\bar{b}$ decay, one expects at most two b-jets in a signal event. Initial state radiation (ISR) jets are dominantly gluon-initiated and u-/d-quark-initiated, so the presence of a third b-jet in a signal event is very unlikely. This motivates the veto of additional b-jets, that is to say, veto of events having more than two b-jets in the resolved regime.

$t\bar{t}$ events also have two b-quarks as hard final state particles, so at first glance it is not evident how a third b-tagged jet is to be found in a $t\bar{t}$ event, which would allow to veto the event. In order to elucidate this, let us consider the truth flavour distribution of the

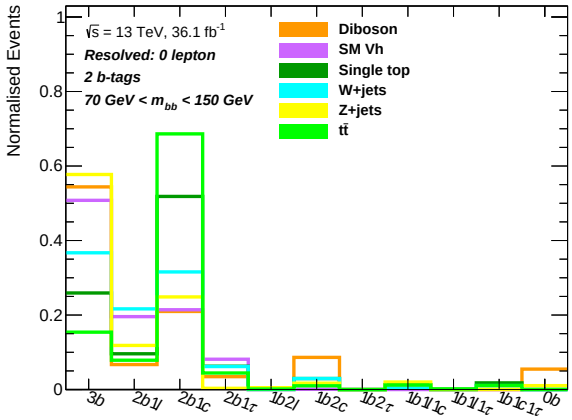

Figure 7.1: Truth flavour distribution of the three leading b-tagged small-R jets in events having more than two b-tagged small-R jets normalised to unit area in the SR, 2 tag, resolved regime within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. About 70% (50%) of the $t\bar{t}$ (single top) events contain two real b-jets and a real c-jet.

three leading b-tagged small-R jets in events having more than two b-tagged small-R jets. In Figure 7.1, this truth flavour distribution is shown normalised to unit area in the SR, 2 tag, resolved regime. About 70% (50%) of $t\bar{t}$ (single top) events having more than two b-tagged small-R jets contain two real b-jets and a real c-jet. For MV2c10 b-tagging at 70% operating point, the mistagging efficiency of c-quark-initiated jets as b-jets is about 8% [105], Table 2. This means that the veto of $t\bar{t}$ (and single top) events having additional b-jets relies on the mistagging of a c-quark-initiated jet as a b-jet from the $W \to c\bar{s}$ decay.

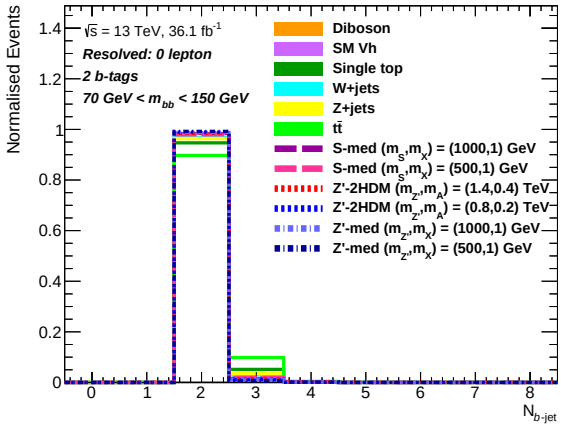

Figure 7.2: Distribution of b-jet multiplicity $N_{b\text{-jet}}$ normalised to unit area in the SR, 2 tag, resolved regime, within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator signal models (in dashed lines) populate negligibly $N_{b\text{-jet}} \geq 3$ bins. The veto of additional b-jets rejects about 10% (5%) of $t\bar{t}$ (single top) in the SR, 2 tag, resolved regime.

In Figure 7.2, the distribution of b-jet multiplicity normalised to unit area is shown in the SR, 2 tag, resolved regime. The representative signal models populate as expected

the two b-jet multiplicity, with negligible presence in higher b-jet multiplicities. On the other hand, about 10% (5%) of $t\bar{t}$ (single top) populate the three b-jet multiplicity, which correspond to the fraction of $t\bar{t}$ (single top) events rejected by the veto of additional b-jets.

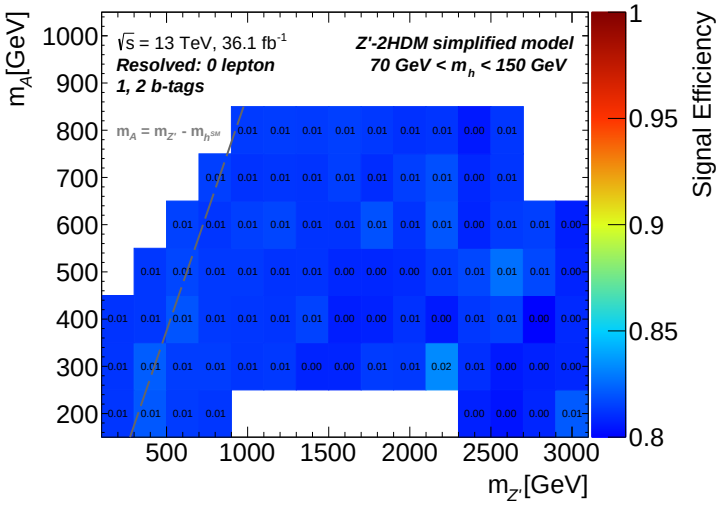

Figure 7.3: Fractional signal rejection from the veto of additional b-jets in the SR, resolved regime for the $Z^{'}-2HDM$ grid within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey dashed line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The fractional rejection is a percent for nearly all signal models.

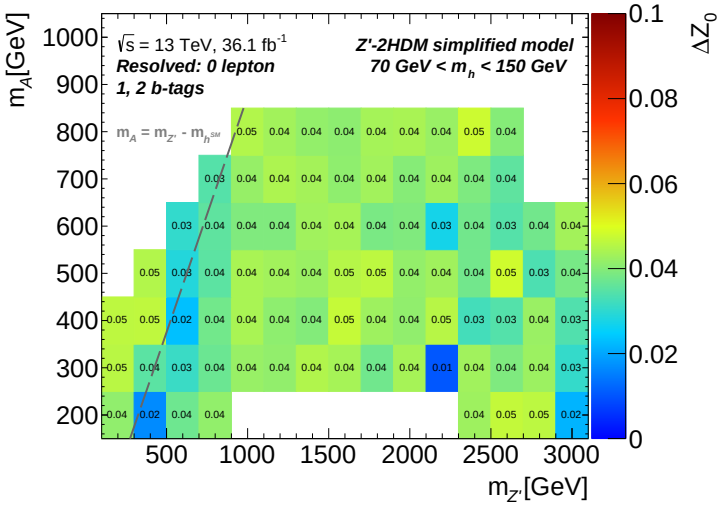

Figure 7.4: Significance gain ΔZ_0 from the veto of additional b-jets in the SR, resolved regime for the $Z^{'}-2HDM$ grid within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey dashed line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The significance gain is $3-4\%$ for nearly all signal models in this region.

In Figure 7.3, the fractional signal rejection from the veto of additional b-jets in the SR,

resolved regime is shown for the $Z^{'} - 2HDM$ mass point grid. The signal rejection from the veto of additional b-jets is a percent for nearly all $Z^{'} - 2HDM$ models in this region.

In Figure 7.4, the significance gain ΔZ_0 from the veto of additional b-jets in the SR, resolved regime is shown for the $Z^{'} - 2HDM$ mass point grid. The significance gain is $3-4\%$ for nearly all mass point signal models in this region.

7.1.2 Veto of Non-Associated B-jets in the Merged Regime

In the merged regime, the $h \to b\bar{b}$ decay is reconstructed as a large-R jet, containing the $b\bar{b}$ decay products. Therefore, in a signal event with signature $h(\to b\bar{b}) + E_T^{miss}$, no b-jet is expected outside the h candidate in the merged regime. This motivates the *veto of non-associated b-jets*.

In the merged regime, the h candidate is the leading large-R jet, which contains the hadronic decay products of a top-quark in $t\bar{t}$ events. Since the top-quarks in a $t\bar{t}$ event are in average produced back-to-back, a real b-jet is expected to be found, which is well separated from the leading large-R jet. In Figure 7.5, the truth flavour distribution of the leading non-associated b-tagged small-R jet is shown normalised to unit area in the SR, merged regime. In about 100% of the $t\bar{t}$ (and single top) events having a non-associated b-jet, the leading non-associated b-jet is a real b-quark-initiated jet. The veto of non-associated b-jets relies therefore on the veto of a real b-quark initiated jet in $t\bar{t}$ events.

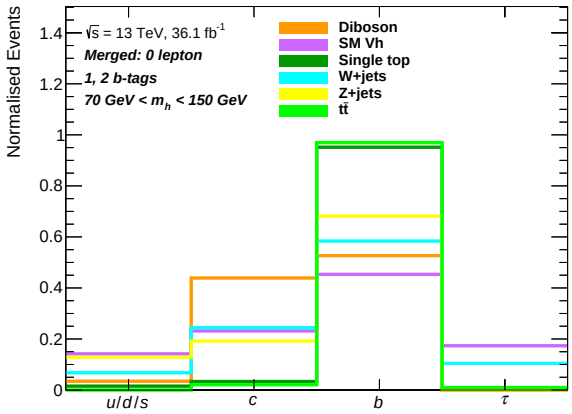

Figure 7.5: Truth flavour distribution of the leading non-associated b-tagged small-R jet normalised to unit area in the SR, merged regime, within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. In about 100% of $t\bar{t}$ events having a non-associated b-jet, the non-associated b-jet is a real b-quark-initiated jet.

Since the h boson is radiated off a s-channel propagator in $Z^{'}$- and S-mediator models, Section 2.3, the mediator becomes off-shell before decaying into a pair of DM particles. In this case, the E_T^{miss} distribution is not so sensitive to the mediator mass, compared to the $Z^{'} - 2HDM$ model whose Jacobian peak in the E_T^{miss} distribution noticeably shifts to higher E_T^{miss} values as the mediator mass $m_{Z'}$ increases. $Z^{'}$- and S-mediator models have generally a sharply falling E_T^{miss} distribution, see Reference [107], pages 6-8. This is especially truth for S-mediator models, for which the E_T^{miss} distribution slightly shifts to

higher E_T^{miss} values as m_S increases. The sensitivity of the analysis to S-mediator models is then driven by the resolved regime. The S-mediator model therefore plays a minor role in the performance evaluation of the optimised event selections in the merged regime.

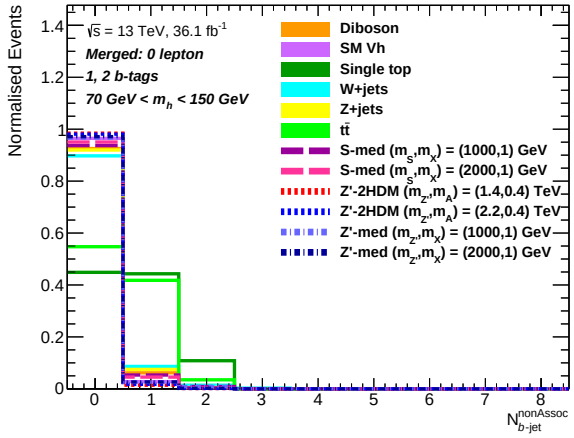

Figure 7.6: Distribution of multiplicity of non-associated b-jets $N_{b\text{-jet}}^{\text{nonAssoc}}$ normalised to unit area in the SR, merged regime within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. Th representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator signal models populate dominantly $N_{b\text{-jet}}^{\text{nonAssoc}} = 0$. About 45% (55%) of $t\bar{t}$ (single top) events populate b-jet multiplicities higher than zero.

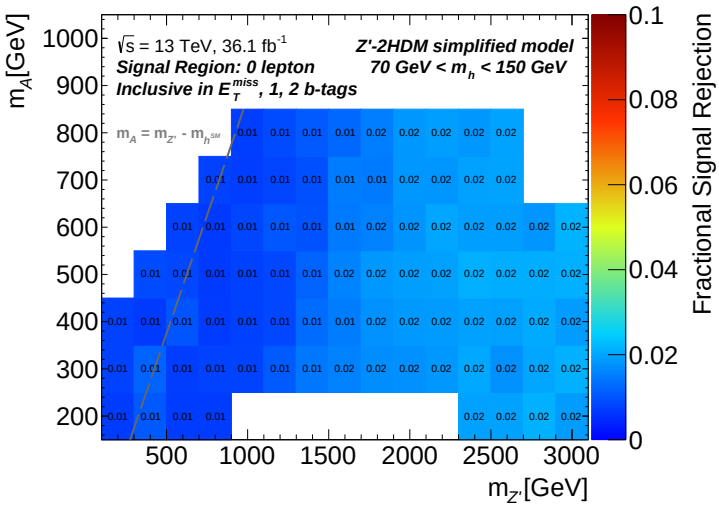

Figure 7.7: Fractional signal rejection from the veto of additional b-jets in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The signal rejection is 1% for $Z^{'} - 2HDM$ models with $m_{Z'} \leq 1.6$ TeV and 2% for $Z^{'} - 2HDM$ models having boosted topologies with $m_{Z'} \geq 1.8$ TeV.

In Figure 7.6, the distribution of multiplicity of non-associated b-jets normalised to unit area is shown in the SR, merged regime. The representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator signal models populate mainly the zero non-associated b-jet multiplicity, with

negligible presence in higher b-jet multiplicities. On the other hand, about 45% (55%) of $t\bar{t}$ (single top) events populate b-jet multiplicities higher than zero.

In Figure 7.7, the fractional signal rejection from the veto of additional b-jets in the SR, inclusive in E_T^{miss} is shown for the $Z^{'} - 2HDM$ grid. The signal rejection from the veto of additional b-jets is 1% for $Z^{'} - 2HDM$ models with $m_{Z'} \leq 1.6$ TeV while it is 2% for $Z^{'} - 2HDM$ models having boosted topologies with $m_{Z'} \geq 1.8$ TeV.

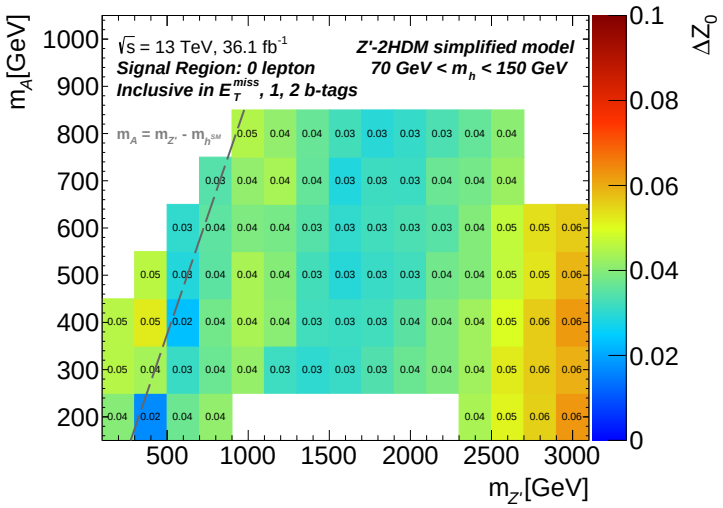

Figure 7.8: Significance gain ΔZ_0 from the veto of additional b-jets in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window (70 GeV, 150 GeV). The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey dashed line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The significance gain ranges between 3% (for most models with $m_{Z'} \leq 2.$ TeV) and 6% (for models with very boosted topologies).

In Figure 7.8, the significance gain ΔZ_0 from the veto of additional b-jets in the SR, inclusive in E_T^{miss} is shown for the $Z^{'} - 2HDM$ grid. The significance gain is about $3-4$% for nearly all signal models with $m_{Z'} \leq 2.$ TeV, reaching values up to 6% for models with very boosted topologies.

7.2 Tau Veto

The *tau veto* is motivated by the signal signature $h(\rightarrow b\bar{b}) + E_T^{miss}$, in which no tau is expected as final state particle. The presence of any hadronic tau[1] would indicate that the event is likely a background event.

In Figure 7.9, the leptonic truth flavour composition of background processes normalised to unit area is shown after the veto of additional b-jets in the SR, resolved (top) and merged (bottom) regime. In the resolved regime, $t\bar{t}$, $W + jets$ and single top background processes consist of events with at least a truth hadronic tau about 75%, 67% and

[1]Taus decaying leptonically are indistinguishable from prompt electrons and muons. Events with leptonically decaying taus would mainly enter the CRs.

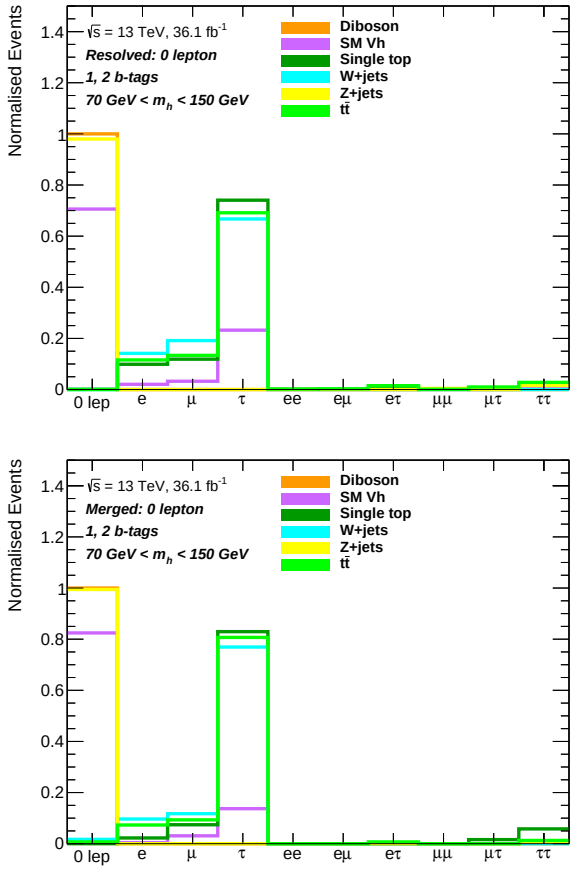

Figure 7.9: Leptonic truth flavour distribution normalised to unit area after the veto of additional b-jets for background processes in the SR, resolved (top) and merged (bottom) regime within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. $t\bar{t}$, $W + jets$ and single top background processes are mainly composed of events with a hadronic tau from the $W \to \tau\nu_\tau$ decay.

75% of the cases, respectively. In the merged regime, $t\bar{t}$, $W + jets$ and single top background processes consist of events with at least a truth hadronic tau about 80%, 76% and 90% of the cases, respectively. The leptonic content in the final state particles of these background processes originates from the leptonic decay of a W boson, which gives rise to real E_T^{miss} and a lepton. Background events with a $W \to \tau\nu_\tau$ decay are favoured in the SR because it gives rise to real E_T^{miss} and a hadronic tau, which is reconstructed as a small-R jet and not restrained by the stringent lepton vetoes of muons and electrons. A tau veto aims to reduce background processes that have events with a hadronic tau from the $W \to \tau\nu_\tau$ decay.

In the following, the tau veto based on the veto of events with a *standard* tau and a *extended* tau will be presented in Subsections 7.2.1 and 7.2.2, respectively. For the definition of these tau objects, please see Subsection 5.2.3.

7.2.1 Tau Veto with Standard Taus

In Figure 7.10, the *standard* tau multiplicity distribution normalised to unit area is shown in the SR, resolved (top) and merged (bottom) regime. Representative signal models have no *standard* taus. On the other hand, in the resolved regime, $t\bar{t}$, $W + jets$ and single top background processes populate higher *standard* tau multiplicities in about 15%, 10% and 18%, respectively. In the merged regime, $t\bar{t}$, $W + jets$ and single top populate higher non-associated *standard* tau multiplicities in about 18%, 22% and 37%, respectively. These percentages represent the fraction of background events per process rejected by the tau veto in the respective regions.

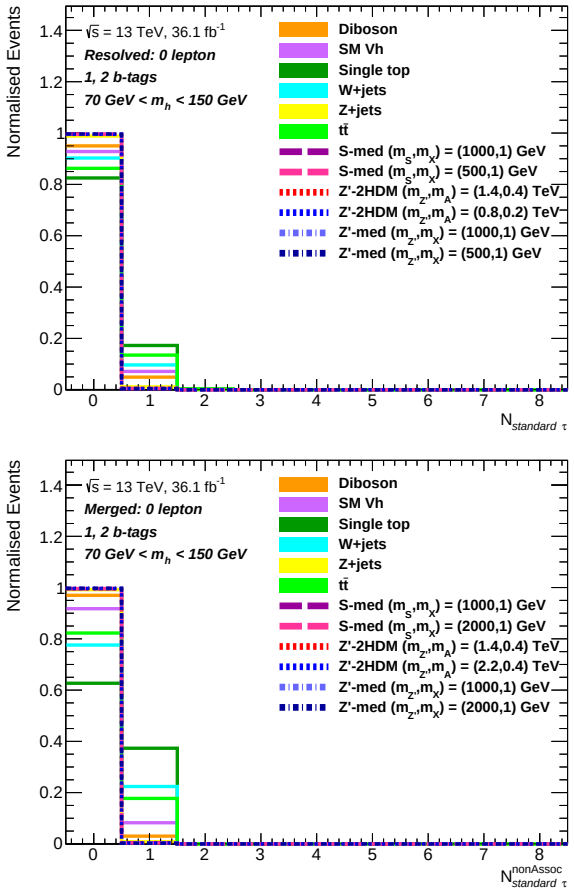

Figure 7.10: *Standard* tau multiplicity distribution normalised to unit area in the SR, resolved (top) and merged (bottom) regime within h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. Representative signal models (in dashed lines) have no *standard* taus. The tau veto mainly reduces $t\bar{t}$, $W + jets$ and single top background processes.

In Figure 7.11, the fractional signal rejection from the tau veto in the SR, inclusive in E_T^{miss} is shown for the $Z^{'} - 2HDM$ grid. The signal rejection is 0% for nearly all $Z^{'} - 2HDM$ models in the SR.

In Figure 7.12, the significance gain ΔZ_0 from the tau veto in the SR, inclusive in E_T^{miss} is shown for the $Z^{'} - 2HDM$ grid. For most signal models in the resolved regime, on-

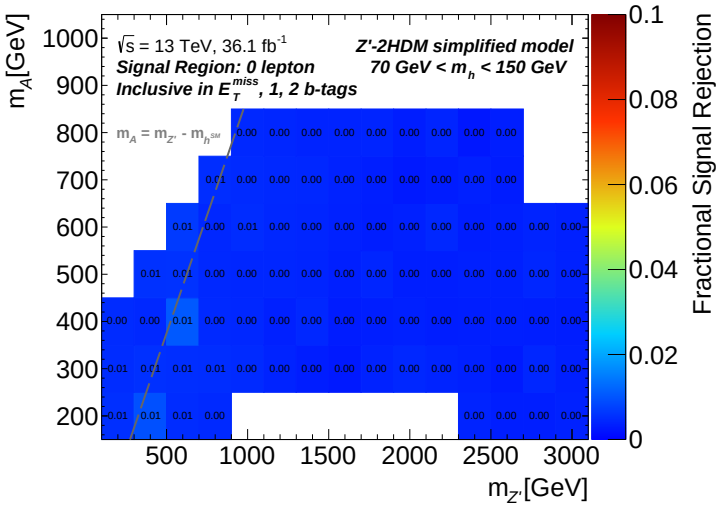

Figure 7.11: Fractional signal rejection from the tau veto in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The tau veto has no effect on the yields of nearly all signal models.

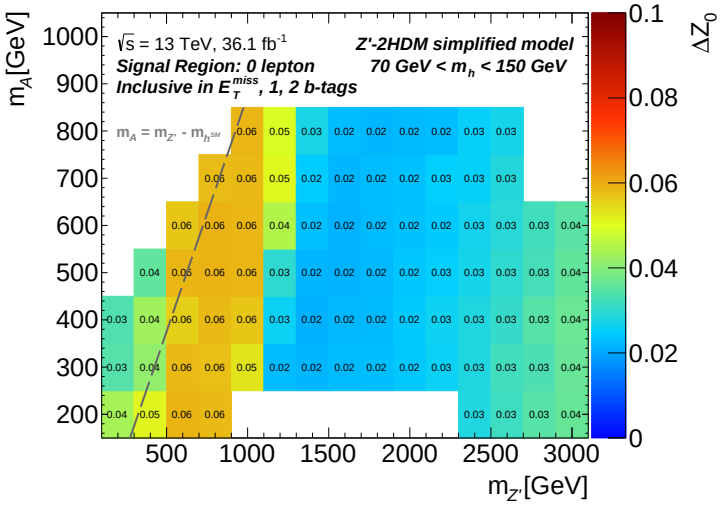

Figure 7.12: Significance gain ΔZ_0 from the tau veto in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The significance gain ranges between 2% and 6%.

shell region, the significance gain is about 6%. The significance gain is $2-3\%$ for signal models in the merged regime with 1.4 TeV $\leq m_{Z'} \leq$ 2.6 TeV, while it is about 4% for signal models with very boosted topologies at large $m_{Z'}$ values.

7.2.2 Tau Veto with Extended Taus

In Figure 7.13, the leptonic truth flavour distribution normalised to unit area after the additional b-jets and tau vetoes is shown for background processes in the SR, resolved regime. Most $t\bar{t}$, $W + jets$ and single top events still contain a hadronic tau after the tau veto. That is why a very loose custom-built tau identification was designed by the author to identify the remaining hadronic taus that were not identified by the standard identification algorithm. The *extended* tau definition is explained in Subsection 5.2.3.

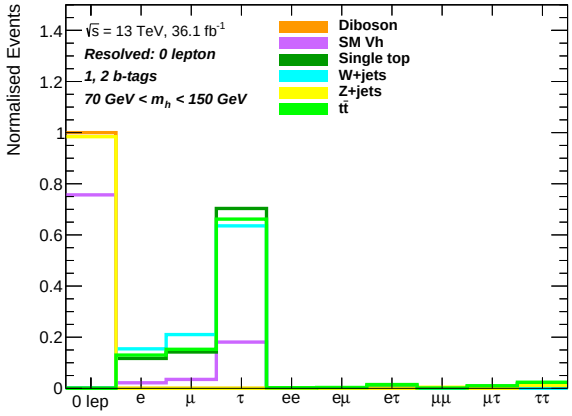

Figure 7.13: Leptonic truth flavour distribution normalised to unit area after the additional b-jets and tau vetoes for background processes in the SR, resolved regime within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. After the tau veto, $t\bar{t}$, $W + jets$ and single top background processes are still mainly composed of events with a hadronic tau.

In Figure 7.14, the *extended* tau multiplicity distribution normalised to unit area is shown in the SR, resolved (top) and merged (bottom) regime. Representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator signal models (in dashed lines) have essentially no *extended* tau. On the other hand, in the resolved regime, $t\bar{t}$, $W + jets$ and single top populate higher *extended* tau multiplicities in about 13%, 7% and 12%, respectively. In the merged regime, $t\bar{t}$, $W + jets$ and single top populate higher non-associated *extended* tau multiplicities in about 20%, 15% and 27%, respectively. These percentages correspond to the fractional rejection per process.

In Figure 7.15, the fractional signal rejection from the *extended* tau veto in the SR, inclusive in E_T^{miss} is shown for the $Z^{'} - 2HDM$ grid. The signal rejection from the *extended* tau veto is 2% for nearly all $Z^{'} - 2HDM$ models.

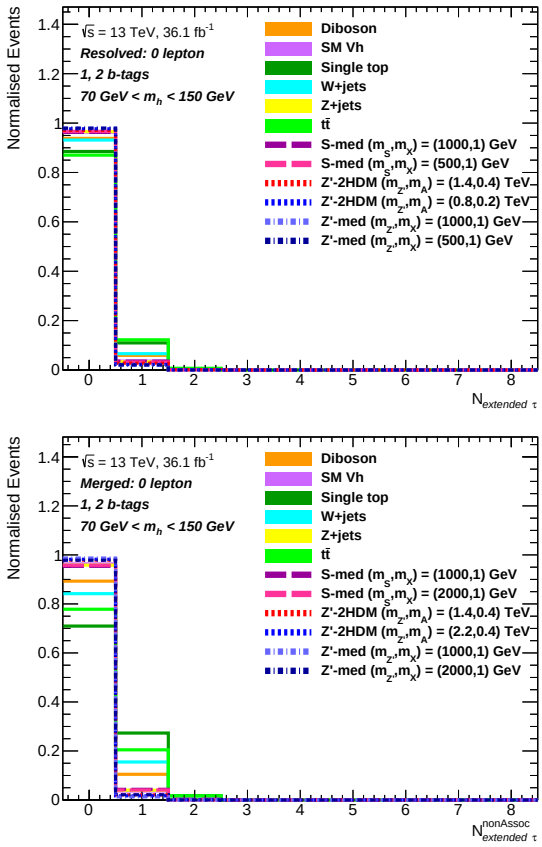

Figure 7.14: *Extended* tau multiplicity distribution normalised to unit area in the SR, resolved (top) and merged (bottom) regime within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. Representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator signal models (in dashed lines) have essentially no *extended* tau. The *extended* tau veto mainly reduces $t\bar{t}$, $W + jets$ and single top background processes.

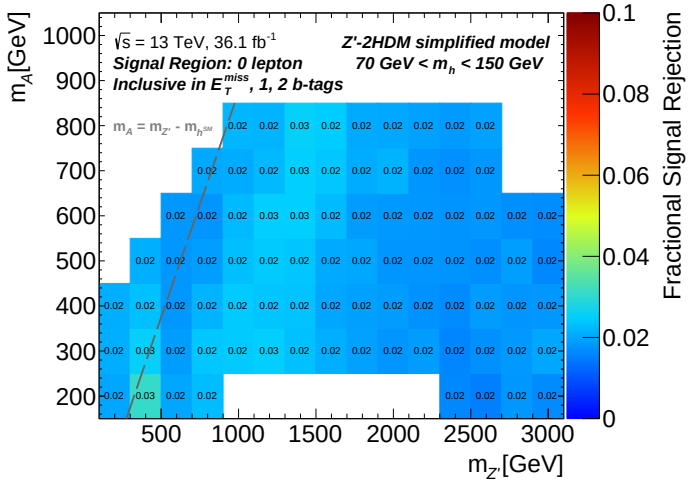

Figure 7.15: Fractional signal rejection from the *extended* tau veto in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. The *extended* tau veto reduces the signal yields by about 2% for nearly all models.

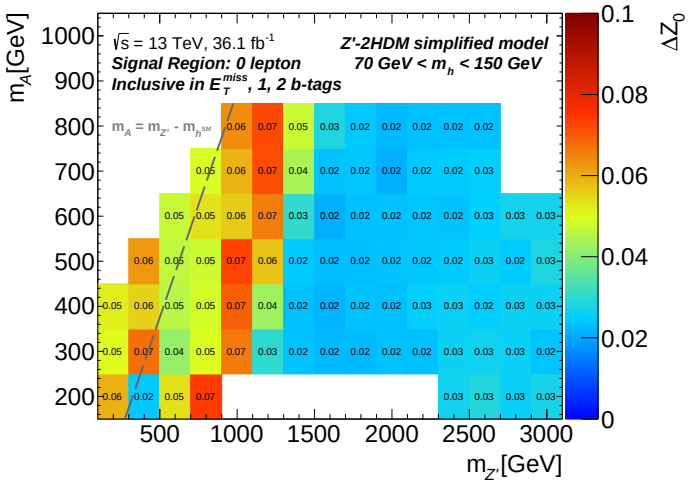

Figure 7.16: Significance gain ΔZ_0 from the *extended* tau veto in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window $(70\text{ GeV}, 150\text{ GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The significance gain from *extended* tau veto ranges between 2% and 7%.

In Figure 7.16, the significance gain ΔZ_0 from the *extended* tau veto in the SR, inclusive in E_T^{miss} is shown for the $Z^{'} - 2HDM$ grid. The significance gain is about $4-7\%$ for signal models to which the sensitivity is driven by the resolved regime, on-shell region, while it is about $2-3\%$ for signal models to which the sensitivity is driven by the merged regime.

7.3 H_T Ratio Cut

In a signal event with signature $h(\rightarrow b\bar{b}) + E_T^{miss}$, the hard final state particles are the pair of b-quarks from the $h \rightarrow b\bar{b}$ decay, so most of the hadronic activity in a signal event is expected from the $h \rightarrow b\bar{b}$ decay. On the other hand, in a $t\bar{t}$ event, the hard final state particles are a pair of b-quarks, a pair of quarks from a $t \rightarrow Wb \rightarrow qq^{'}b$ decay and often a hadronic tau and a neutrino from a $t \rightarrow Wb \rightarrow \ell\nu_\ell b$ decay, see Figure 6.9. The h candidate is most likely reconstructed from the decay products of the $h \rightarrow b\bar{b}$ decay in a signal event while it is most likely reconstructed from the pair of b-quarks (the hadronic decay products of a top quark) in a $t\bar{t}$ event in the resolved (merged) regime. Therefore, while most of the hadronic activity in a signal event is carried by the h candidate, in a $t\bar{t}$ event a considerable fraction of it may be carried by hadronic particles not forming the h candidate. This means that $\frac{H_T(\text{non}-h\text{ candidate})}{H_T} \ll 1$ in a signal event whereas $\frac{H_T(\text{non}-h\text{ candidate})}{H_T} < 1$ in a $t\bar{t}$ event, where H_T is the scalar p_T sum of all jets present in the event and $H_T(\text{non}-h\text{ candidate})$ is the scalar p_T sum of the jets not forming the h candidate.

7.3.1 H_T Ratio Cut in the Resolved Regime

In the resolved regime, the H_T ratio is defined as the ratio of the scalar p_T sum of small-R jets other than the h candidate and, if present, the leading additional small-R jet to the scalar p_T sum of all small-R jets in the event, as seen in Equation 7.3

$$H_T \text{ ratio} \equiv \sum_{i=4}^{\infty} p_T^{j_i} \Big/ \sum_{i=1}^{\infty} p_T^{j_i} \,, \tag{7.3}$$

where $p_T^{j_i}$ is the p_T of the i-th small-R jet in the event in the resolved regime.

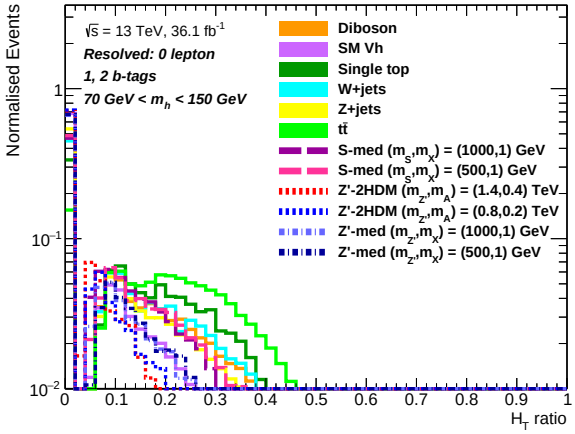

Figure 7.17: H_T ratio distribution normalised to unit area in the SR, resolved regime within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. Representative $Z^{'} - 2HDM$, $Z^{'}$- and S- mediator signal models (in dashed lines) mainly populate small H_T ratio value bins, whereas $t\bar{t}$, $W + jets$ and single top background processes accumulate at larger H_T ratio values.

In Figure 7.17, the H_T ratio distribution normalised to unit area is shown in the SR, resolved regime. Representative $Z^{'} - 2HDM$ and $Z^{'}$-mediator signal models mainly populate small H_T ratio value bins. Representative S-mediator models have a harder H_T ratio distribution, since the hadronic activity is higher for S-mediator processes due to having gluons as initial states[2], see Subsection 2.3.3. Unlike signal models, $t\bar{t}$, $W + jets$ and single top accumulate at larger H_T ratio values, having no presence of events at very low H_T ratio values.

In Figure 7.18, the fractional rejection (top) and the significance gain (bottom) as a function of the H_T ratio upper cut are shown in the SR, resolved regime for representative $Z^{'} - 2HDM$, $Z^{'}$- and S- mediator signal models and the total background. A more stringent H_T ratio cut at lower H_T ratio values increases the significance gain as well as the fractional rejection. A H_T ratio cut of 0.37 was chosen in the analysis to increase the significance gain while having negligible signal rejection. The fractional background rejection at this cut value is about 8%.

[2]Due to the colour factor for gluons, $C_A = 3$, compared to that for quarks, $C_F = 4/3$, a gluon tends to emit more soft- and collinear-radiation than a quark.

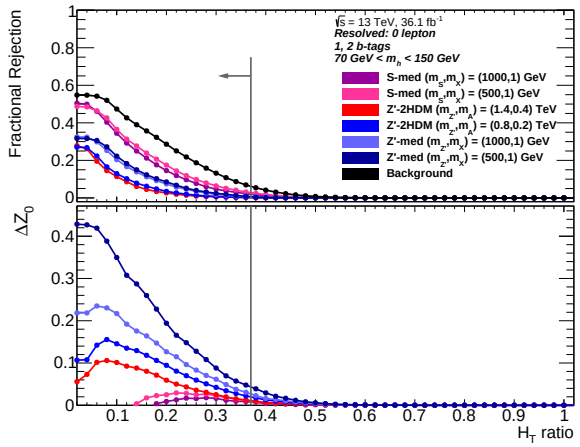

Figure 7.18: Fractional rejection (top) and significance gain ΔZ_0 (bottom) as a function of the H_T ratio upper cut in the SR, resolved regime for representative $Z^{'}-2HDM$, $Z^{'}$- and S-mediator signal models and the total background within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The H_T ratio cut of 0.37 shown as a grey line. At this cut value, the signal rejection is negligible. The fractional background rejection is about 8%.

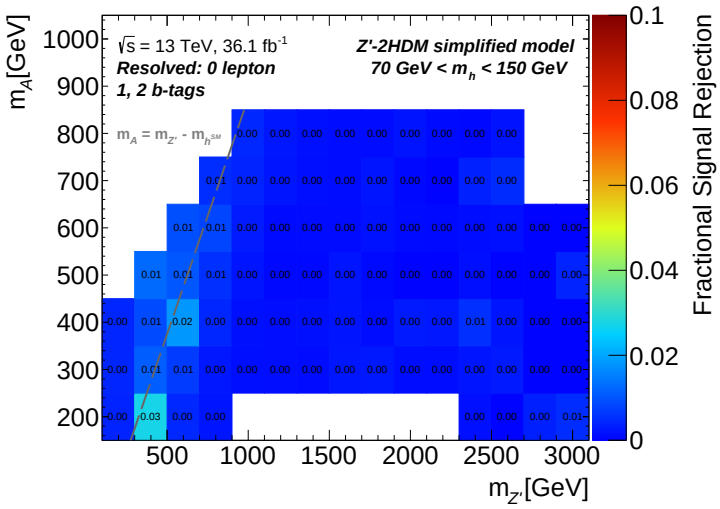

Figure 7.19: Fractional signal rejection from the H_T ratio cut in the SR, resolved regime for the $Z^{'}-2HDM$ grid within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The chosen H_T ratio cut at 0.37 has no effect on the yields of nearly all signal models.

In Figure 7.19, the fractional signal rejection from the H_T ratio cut at 0.37 in the SR, resolved regime is shown for the $Z^{'}-2HDM$ grid. The signal rejection with respect to the former baseline event selection is 0% for nearly all $Z^{'}-2HDM$ models.

In Figure 7.20, the significance gain ΔZ_0 from the H_T ratio cut in the SR, resolved regime is shown for the $Z^{'}-2HDM$ grid within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The significance gain reaches values up to 5% in the resolved regime.

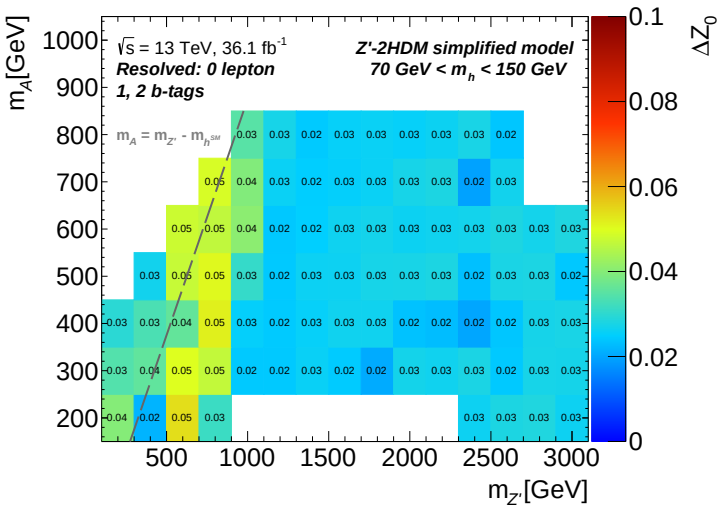

Figure 7.20: Significance gain ΔZ_0 from the H_T ratio cut in the SR, resolved regime for the $Z' - 2HDM$ grid within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The significance gain is up to 5%.

7.3.2 H_T Ratio Cut in the Merged Regime

In the merged regime, the H_T ratio is defined as the ratio of the scalar p_T sum of all non-associated small-R jets to the sum of the scalar p_T of the h candidate (leading large-R jet in the event) *and* the scalar p_T sum of all non-associated small-R jets in the event, as seen in Equation 7.4

$$H_T \text{ ratio} \equiv \sum_{i=1}^{\infty} p_T^{j_i^*} \Bigg/ \left(p_T^J + \sum_{i=1}^{\infty} p_T^{j_i^*} \right), \tag{7.4}$$

where $p_T^{j_i^*}$ stands for the p_T of non-associated small-R jets and p_T^J for the p_T of the h candidate.

In the merged regime, at high E_T^{miss}, the ISR is expected to increase (especially for non-resonant signal models such as Z'- and S-mediator models), which shifts the H_T ratio distribution to higher values. In Figure 7.21, the H_T ratio distribution normalised to unit area is shown in the SR, merged regime. As in the resolved regime, representative S-mediator models have more hadronic activity, reaching larger H_T ratio values than $Z' - 2HDM$ and Z'-mediator models. Given the H_T ratio distribution in the merged regime, a H_T ratio cut does not increase the sensitivity to S-mediator models while the increase in sensitivity to $Z' - 2HDM$ and Z'-mediator models is expected to be small.

In Figure 7.22, the fractional rejection (top) and the significance gain ΔZ_0 (bottom) as a function of the H_T ratio upper cut are shown in the SR, merged regime for representative $Z' - 2HDM$, Z'- and S- mediator signal models and the total background. A H_T ratio cut of 0.57 negligibly reduces the signal yields, except for S-mediator models whose fractional rejection is comparable to that of background. For larger mediator masses, the H_T ratio

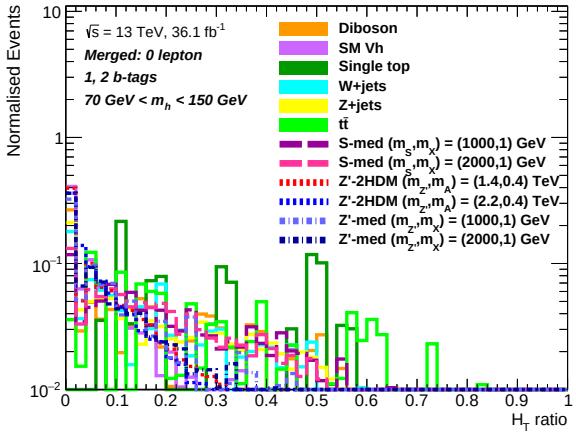

Figure 7.21: H_T ratio distribution normalised to unit area in the SR, merged regime within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. Representative $Z^{'} - 2HDM$, $Z^{'}$- and S- mediator signal models (in dashed lines) populate smaller H_T ratio value bins than $t\bar{t}$.

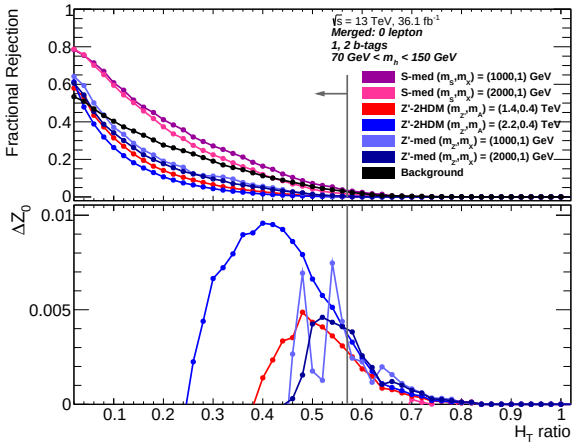

Figure 7.22: Fractional rejection (top) and significance gain ΔZ_0 (bottom) as a function of the H_T ratio upper cut in the SR, merged regime for representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator signal models and the total background within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The H_T ratio cut of 0.57 shown as a grey line. At this cut value, the signal yields rejection is negligible and the significance gain less than a percent in the merged regime.

distribution shifts to smaller H_T ratio values as the momentum carried by the produced h increases more than the ISR activity. The chosen H_T ratio cut of 0.57 increases the significance in less than a percent in the merged regime.

In Figure 7.23, the fractional signal rejection from the H_T ratio cut is shown in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid. The signal rejection with respect to the former baseline event selection is 0% for nearly all $Z^{'} - 2HDM$ models.

In Figure 7.24, the significance gain ΔZ_0 from the H_T ratio cut in the SR, inclusive in E_T^{miss} is shown for the $Z^{'} - 2HDM$ grid. For signal models, to which the sensitivity is driven by the resolved regime, the significance gain reaches values up to 5%, as expected from Figure 7.20. For signal models, to which the sensitivity is driven by the merged

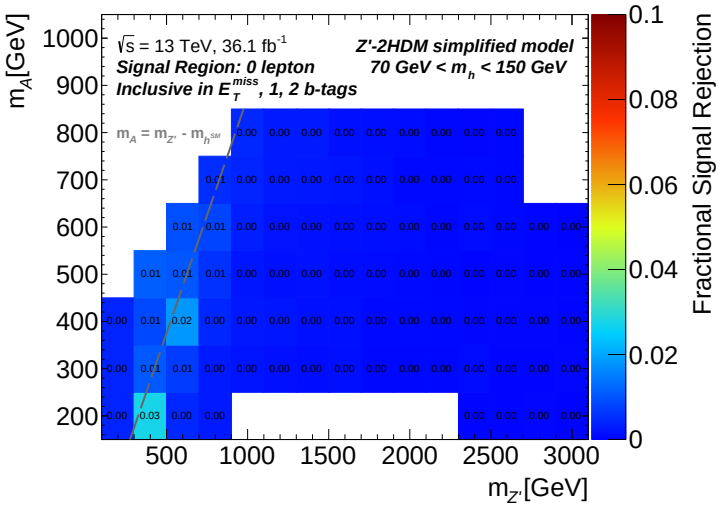

Figure 7.23: Fractional signal rejection from the H_T ratio cut in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The H_T ratio cut has no effect on the yields of nearly all signal models.

regime, the significance gain in the SR ranges between $1 - 2\%$.

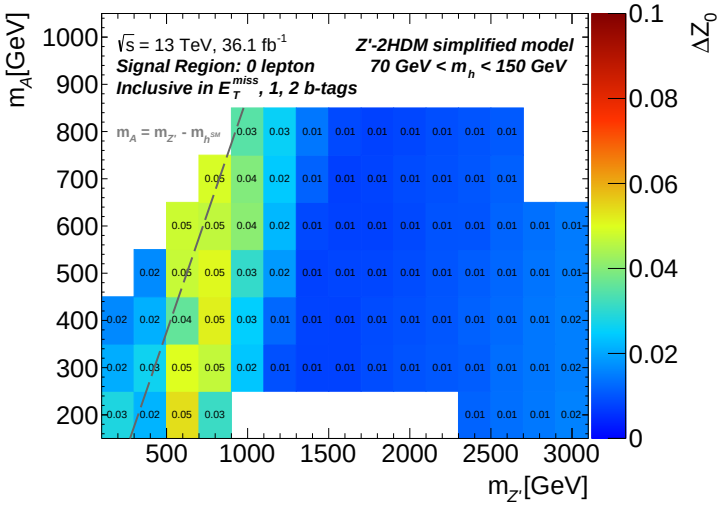

Figure 7.24: Significance gain ΔZ_0 from the H_T ratio cut in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The significance gain ranges between 1% and 5%.

7.4 ΔR_{bb} Cut

In a signal event with signature $h(\rightarrow b\bar{b}) + E_T^{miss}$, the angular separation of the hadronic decay products from the $h \rightarrow b\bar{b}$ decay follows the formula $\Delta R(\vec{p}^{\,b}, \vec{p}^{\,\bar{b}}) = 2p_T^h/m_h$, where p_T^h and m_h stand for the transverse momentum and mass of the h boson, respectively. The b-quark-initiated decay products are collimated due to the boost of the h boson. On the other hand, in a $t\bar{t}$ event, the b-jets forming the h candidate originate from the decays of top quarks produced in average back-to-back, so the angular separation of this pair of b-jets is in average larger than that of the pair of b-jets originating from the decay of the same heavy particle h. The ΔR_{bb} cut is only applied in the resolved regime, since the reconstruction of the $h \rightarrow b\bar{b}$ decay as a large-R jet of radius $R = 1.0$ introduces in the merged regime an effective requirement on the geometrical separation of the b-quark-initiated decay products.

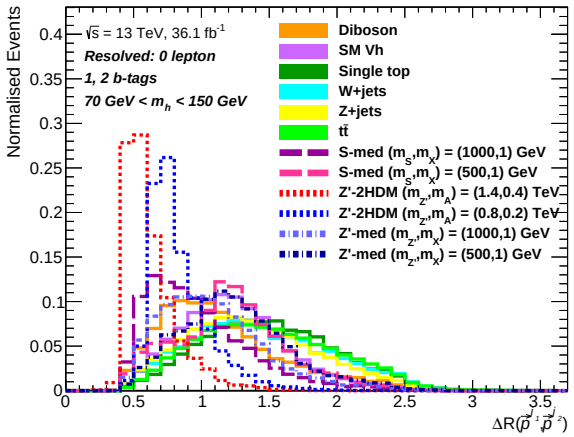

Figure 7.25: $\Delta R(\vec{p}^{\,j_1}, \vec{p}^{\,j_2})$ distribution normalised to unit area in the SR, resolved regime within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. Representative $Z^{'} - 2HDM$, $Z^{'}$- and S- mediator signal models (in dashed lines) populate small angular separation value bins whereas the $\Delta R(\vec{p}^{\,j_1}, \vec{p}^{\,j_2})$ distribution of $t\bar{t}$ tends to be shifted to larger angular separation values.

In Figure 7.25, the $\Delta R(\vec{p}^{\,j_1}, \vec{p}^{\,j_2})$ distribution normalised to unit area is shown in the SR, resolved regime. Representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator signal models (in dashed lines) populate small angular separation value bins as expected. The resonant $Z^{'} - 2HDM$ models accumulate dominantly at angular separations smaller than 1, whereas the $\Delta R(\vec{p}^{\,j_1}, \vec{p}^{\,j_2})$ distributions of non-resonant signal models tend to be shifted to larger angular separation values. Background processes, especially $t\bar{t}$ and single top, accumulate at larger angular separation values than signal models.

In Figure 7.26, the fractional rejection (top) and the significance gain ΔZ_0 (bottom) as a function of the $\Delta R(\vec{p}^{\,j_1}, \vec{p}^{\,j_2})$ upper cut are shown in the SR, resolved regime for representative $Z^{'} - 2HDM$, $Z^{'}$- and S- mediator signal models and the total background. A $\Delta R(\vec{p}^{\,j_1}, \vec{p}^{\,j_2})$ cut of 1.8 was chosen in the analysis to increase the significance gain while having negligible signal rejection. The fractional background rejection at this cut value is less than 10% for the representative signal models. A $\Delta R(\vec{p}^{\,j_1}, \vec{p}^{\,j_2})$ upper cut above 1.8

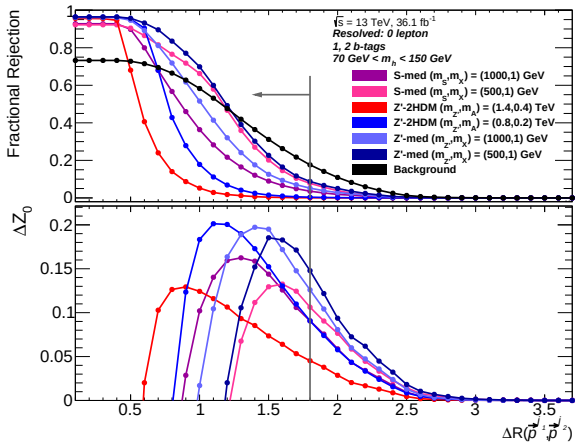

Figure 7.26: Fractional rejection (top) and significance gain ΔZ_0 (bottom) as a function of the $\Delta R(\vec{p}^{j_1}, \vec{p}^{j_2})$ upper cut in the SR, resolved regime for representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator signal models and the total background within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. The $\Delta R(\vec{p}^{j_1}, \vec{p}^{j_2})$ cut of 1.8 shown as a grey line. The fractional signal rejection is less than 10% while the fractional background rejection is about 20%, resulting in a significance gain for the representative signal models.

increases the significance gain for all representative signal models.

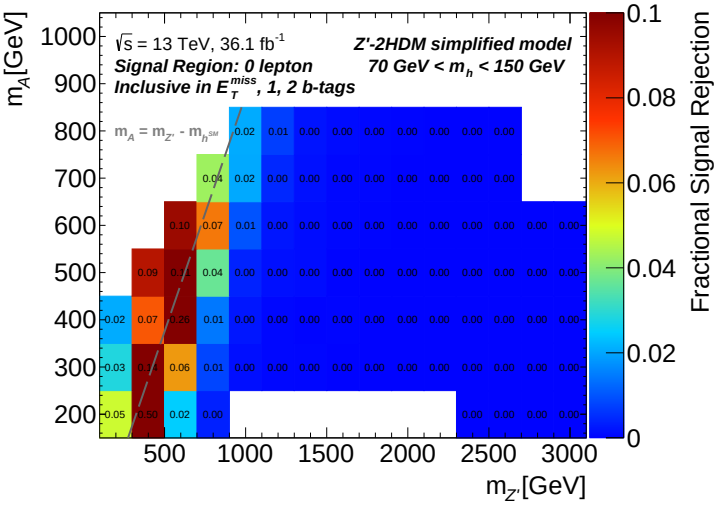

Figure 7.27: Fractional signal rejection from the ΔR_{bb} cut in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z^{'}} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. In the on-shell region, the signal rejection is negligible for $Z^{'} - 2HDM$ models, to which the $h(\to b\bar{b}) + E_T^{miss}$ search is sensitive, reaching up to 6%.

In Figure 7.27, the fractional signal rejection from the ΔR_{bb} cut in the SR, inclusive in E_T^{miss} is shown for the $Z^{'} - 2HDM$ grid. The fractional signal rejection is zero for nearly all $Z^{'} - 2HDM$ models in the on-shell region, reaching up to 6% for some mod-

els to which the sensitivity is driven by the resolved regime. From Figure 6.7 and Figure 7.27, one infers that the $Acceptance \times Efficiency$ of the mass points $(m_{Z'}, m_A) = (400\,\text{GeV}, 200\,\text{GeV})$ and $(600\,\text{GeV}, 400\,\text{GeV})$ before the ΔR_{bb} cut is small, having values of about 0.04 and 0.08, respectively. As seen in Chapter 11, the $h(\rightarrow b\bar{b}) + E_T^{miss}$ search is not sensitive to these signal models.

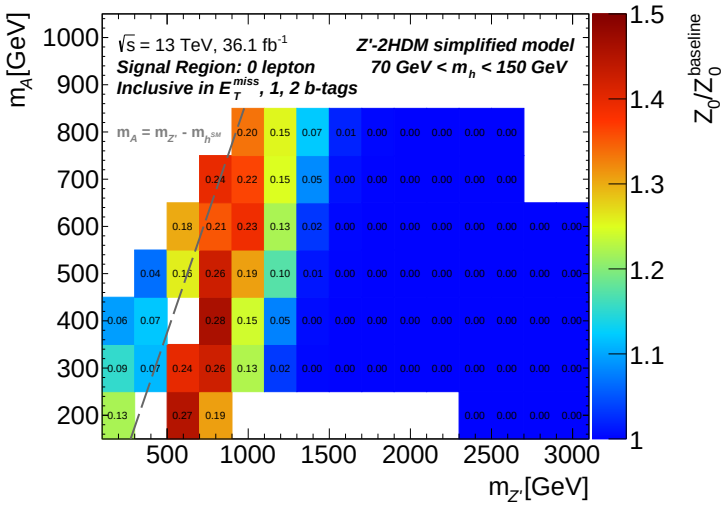

Figure 7.28: Significance gain ΔZ_0 from the ΔR_{bb} cut in the SR, inclusive in E_T^{miss} for the $Z^{'} - 2HDM$ grid within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The significance gain reaches values up to 28%.

In Figure 7.28, the significance gain ΔZ_0 from the ΔR_{bb} cut in the SR, inclusive in E_T^{miss} is shown for the $Z^{'} - 2HDM$ grid. Since the ΔR_{bb} cut is applied only in the resolved regime, the ΔR_{bb} cut increases the sensitivity only to signal models populating the resolved regime. The significance gain reaches values up to 28%.

7.5 Efficiency of Optimised Event Selections

In Figure 7.29, the event selection efficiency for the $t\bar{t}$ background process (top) and total background (down) are shown in the SR upon consecutively applying the optimised event selections. The $t\bar{t}$ background process is considerably reduced by the optimised event selections by about 50% and 70% in the resolved and merged regime, respectively. The total background is reduced by about 45% and by 35% in the resolved and merged regime, respectively, through the optimised event selections.

In Figure 7.30, the signal efficiency in the SR is shown for the $Z^{'} - 2HDM$ mass point grid. The signal efficiency is here defined as the ratio of the signal yields after optimised event selections to those before optimised event selections. The signal efficiency for $Z^{'} - 2HDM$ models is above $\sim 90\%$, reaching values up to 96% for most signal models with boosted topologies at large $m_{Z'}$ values. The only exceptions are mass point models with

small E_T^{miss} available near the transition line at $m_A = m_{Z'} - 125$ GeV, such as $(m_{Z'}, m_A) = (400, 200)$ GeV, $(600, 400)$ GeV. The $h(\to b\bar{b}) + E_T^{miss}$ analysis is not sensitive to these two models, as seen in Chapter 11 and in Figure 6.7.

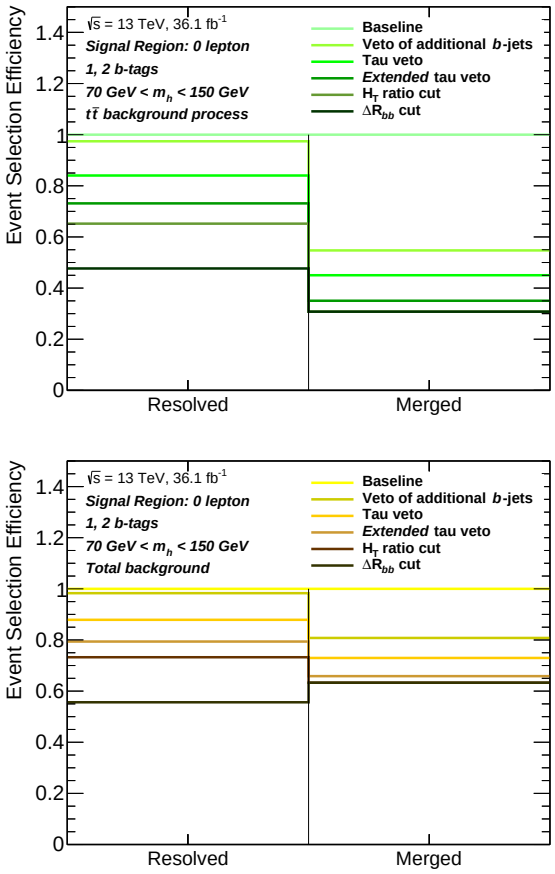

Figure 7.29: Event selection efficiency for the $t\bar{t}$ background process (top) and total background (down) in the SR, upon consecutively applying the optimised event selections (each shown in degrading colours) within the h candidate mass window $(70\text{ GeV}, 150\text{ GeV})$. $t\bar{t}$ (total background) is reduced by about 50% (45%) and 70% (35%) in the resolved and merged regime, respectively through the optimised event selections.

The significance gain from the optimised event selections is presented with a full statistical fit model in Chapter 11.

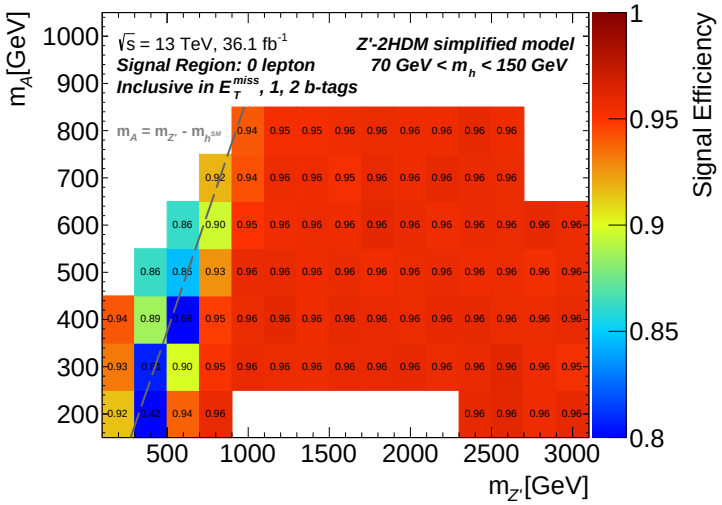

Figure 7.30: Signal efficiency for the $Z^{'} - 2HDM$ mass point grid in the SR within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The signal efficiency from the optimised event selections is above $\sim 90\%$, reaching values up to 96% for most signal models with boosted topologies at large $m_{Z'}$ values.

8 Background Estimation

Control regions (CRs) in the analysis are designed to constrain the main background processes in the signal region (SR). The main background processes are $W + jets$, $Z + jets$ and $t\bar{t}$, which account for more than 90% of the total background in the SR of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis. The CRs in the analysis are the *1 Lepton Control Region*, containing a single muon, and the *2 Lepton Control Region*, containing either two electrons or two muons.

Using simultaneously information from the CRs and the SR, the normalisations of the main background processes are determined from the data by allowing them to be free parameters in the final binned profile likelihood fit, as described in Section 10.3. The 1 lepton CR is used to constrain $t\bar{t}$ and $W + jets$ background processes while the 2 lepton CR is used to constrain $Z + jets$ background process.

The events are categorised based on the lepton multiplicity to define the SR and CRs. Exactly in the same way as done in the SR, the events in the CRs are categorised based on the b-jet multiplicity and on the p_T^V, which represents the E_T^{miss} in the SR and a E_T^{miss}-like variable in the CRs to be defined in the following Sections. Therefore, events in each CR are also categorised into 1 tag and 2 tag, and into four p_T^V bins, namely 150 GeV $< p_T^V <$ 200 GeV, 200 GeV $< p_T^V <$ 350 GeV and 350 GeV $< p_T^V <$ 500 GeV in the resolved regime and $p_T^V >$ 500 GeV in the merged regime.

8.1 1 Lepton Control Region

The 1 lepton CR is designed to constrain the normalisation of heavy flavour $W + jets$ and $t\bar{t}$ background processes. The event selection in this CR is identical to that in the SR as described in Chapter 6, except for the lepton veto: only events with exactly one W-signal muon and no V-loose electron are selected, whose defitions can be found for the muon and electron in Subsections 5.2.2 and 5.2.1, respectively. Having a similar event selection to that in the SR ensures that the events selected in this CR are kinematically and topologically similar to those in the SR. From now on, the 1 lepton CR will be referred to as 1μ CR.

In order for the events in the 1μ CR to be kinematically and topologically similar to those contaminating the SR, the definition of the E_T^{miss} used in the event selection in the 1μ CR must be modified. In the SR, no muon is considered in the reconstruction of E_T^{miss}, while in the 1μ CR, the selected muon is considered. In order for the E_T^{miss} in 1μ CR to kinematically mimic the E_T^{miss} in the SR, one considers the E_T^{miss} four-momentum *plus* the muon four-momentum as the missing transverse energy in the 1μ CR, which is denoted as $E_{T,noMU}^{miss}$. Thus, $p_T^V = E_{T,noMU}^{miss}$. The p_T^{miss} is accordingly modified by adding the muon four-momentum.

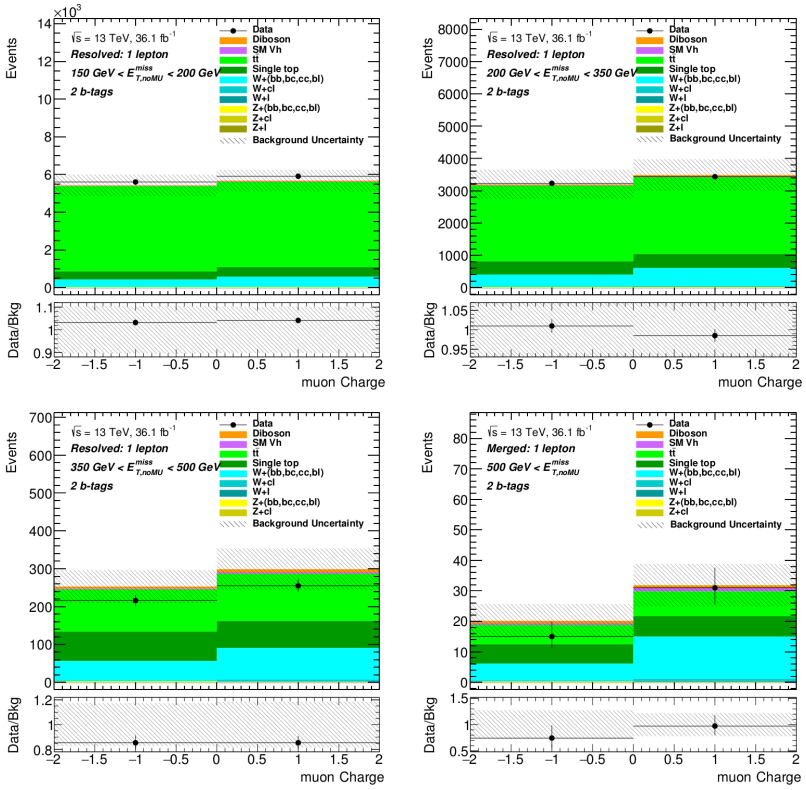

Figure 8.1: Prefit distributions of the muon charge in the 1μ CR, 2 tag, shown for the four subregions in $E^{miss}_{T,noMU}$, within h candidate mass window $(50\,\text{GeV}, 280\,\text{GeV})$. The background uncertainty band includes statistical and systematic uncertainties. The muon charge is asymmetric in $W+jets$, whereas it is symmetric in $t\bar{t}$, as expected. The disagreement between data and background is due to normalisation mismodelling of $t\bar{t}$ and $W+jets$, which is corrected for by the statistical fit model.

In the 1μ CR, the electric charge of the muon is used as discriminant in the statistical fit model to separate $t\bar{t}$ from $W+jets$. For proton-proton collisions, a prevalence of muons with positive electric charge is expected in $W+jets$, whereas $t\bar{t}$ is expected to provide equal number of muons with negative and positive electric charges, see [108]. In Figure 8.1, the prefit distributions of the muon charge in the 1μ CR, 2 tag, are shown for the four subregions in $E^{miss}_{T,noMU}$, within the h candidate mass window $(50\,\text{GeV}, 280\,\text{GeV})$. The background uncertainty includes statistical and systematic uncertainties. The muon charge is asymmetric in $W+jets$, whereas it is symmetric in $t\bar{t}$, as expected in proton-proton collisions. The disagreement between data and background is due to normalisation mismodelling of $t\bar{t}$ and heavy flavour $W+jets$, which is corrected for by the statistical fit model, as shown in Chapter 11.

8.2 2 Lepton Control Region

The 2 lepton CR is designed to constrain the normalisation of heavy flavour $Z+jets$. The $Z+jets$ background process, with $Z \to \nu\nu$, is the dominant background in the SR.

$Z(\to \ell\ell) + jets$ is kinematically very similar to $Z(\to \nu\nu) + jets$ as the p_T of the Z boson is independent of its decay products. This allows constraining the normalisation of $Z + jets$ with a simultaneous profile likelihood fit considering the SR and 2 lepton CR. From now on, the 2 lepton CR will be referred to as 2ℓ CR.

The event preselection is identical to that in the SR, as listed in Section 6.1. The event selection in the 2ℓ CR is as follows:

- **Single lepton triggers:** Events are triggered by the lowest unprescaled single electron or single muon triggers, as listed in Table 4.2.
- **Leptons:** Events must have either exactly two electrons or two muons. In case of two electrons, both of them must be Z-signal electrons, with one of them having a looser p_T cut, namely $p_T > 10$ GeV. The di-electron invariant mass m_{ee} must be within $(83\text{ GeV}, 99\text{ GeV})$. In case of two muons, both of them must be Z-signal muons, with one of them having a looser p_T cut, namely $p_T > 10$ GeV. The di-muon invariant mass $m_{\mu\mu}$ must be within $(71\text{ GeV}, 106\text{ GeV})$ and both muons must have opposite electric charges. The requirement on the dilepton invariant mass suppresses non-resonant background processes, such as $t\bar{t}$, single top and multijet.
- E_T^{miss} **significance:** The E_T^{miss} significance is given by the ratio of the E_T^{miss} to the square root of the scalar sum of p_T of both leptons and all small-R jets in the event. Only events with $E_T^{miss}/\sqrt{H_T^\star} < 3.5\ \sqrt{\text{GeV}}$ are accepted, where $H_T^\star = p_T^{\ell\ell} + \sum_{i=0}^{\infty} p_T^{\text{jet}_i}$ with $p_T^{\text{jet}_i}$ denoting the p_T of a small-R jet. This selection reduces $t\bar{t}$ keeping $Z(\to \ell\ell) + jets$, as the E_T^{miss} originates from finite detector resolution in the case of $Z + jets$ whereas it originates mainly from neutrinos in the case of $t\bar{t}$.

In order for the events in the 2ℓ CR, in majority corresponding to $Z + jets$, to be kinematically and topologically similar to those $Z(\to \nu\nu) + jets$ events contaminating the SR, the transverse momentum of the dilepton system, $p_T^{\ell\ell}$, from the decay $Z \to \ell\ell$, plays the role of the E_T^{miss} in the 2ℓ CR. Thus, $p_T^V = p_T^{\ell\ell}$ in the 2ℓ CR. All remaining event selections, as listed in Chapter 6, are identical between the SR and the 2ℓ CR, except for the p_T^{miss} lower cut and the anti-QCD cuts, which are not applied in the 2ℓ CR.

In the 2ℓ CR, no shape information of the h candidate mass is considered, using only the event yields as discriminant in the statistical profile likelihood fit. In Figure 8.2, the prefit event yields in the 2ℓ CR, 2 tag, are shown for the four subregions in p_T^V within the h candidate mass window $(50\text{ GeV}, 280\text{ GeV})$. The background uncertainty includes statistical and systematic uncertainties. The $Z + jets$ background process is by far the main background process in the 2ℓ CR. The disagreement between data and background is due to normalisation mismodelling of heavy flavour $Z + jets$, which is corrected for by the statistical fit model, as shown in Chapter 11.

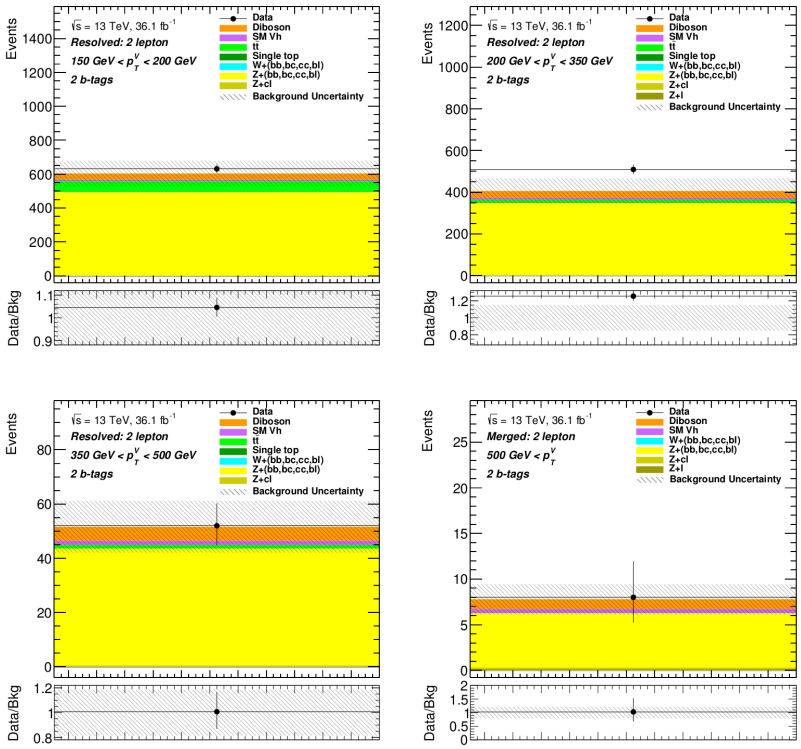

Figure 8.2: Prefit event yields in the 2ℓ CR, 2 tag, for the four subregions in p_T^V ($p_T^{\ell\ell}$), within h candidate mass window $(50\ \text{GeV}, 280\ \text{GeV})$. The background uncertainty includes statistical and systematic uncertainties. The disagreement between data and background is due to normalisation mismodelling of heavy flavour $Z+jets$.

8.3 Data-driven Multijet Estimation

The multijet background process in the SR is poorly modeled by MC simulation due to lack of high statistics. Therefore, the multijet background in the SR is estimated through a dedicated data-driven method to estimate correctly its template from data. For this, a control region enriched with multijet events, called multijet CR, is constructed essentially by inverting the requirement $\min\left[\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{j_{1,2,3}})\right] > \pi/9$ to $\min\left[\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{j_{1,2,3}})\right] < \pi/9$ in the event selection, see Section 6.1. A multijet shape template from the multijet CR is estimated and transferred to the SR through a dedicated multijet fit. This data-driven method consists of two steps:

Multijet Normalisation

The multijet CR is identical to the SR defined in Chapter 6, except for the mentioned inversion of the $\min\left[\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{j_{1,2,3}})\right]$ cut and the relaxation of the event selection. The relaxation of the event selection consists of dropping event cuts with impact on the multijet background process in the SR. These event cuts are $\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{miss}) < \pi/2$,

$\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{h}) > 6\pi/9$ and $\Delta R(\vec{p}^{j_1}, \vec{p}^{j_2}) < 1.8$. The dropping of these event cuts allows the multijet CR to be enriched with multijet events and have sufficient statistics for the derivation of the multijet normalisation to transfer the multijet template to the SR. In the derivation of the multijet normalisation, the data is blinded, so no events with h candidate mass within $(70\text{ GeV}, 140\text{ GeV})$ are considered. This prevents any possible signal contamination, which could get fitted away in the fitting procedure. At this stage, the SR used in the multijet fitting procedure is identical to the multijet CR, except for the non-inversion of the $\min\left[\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{j_{1,2,3}})\right]$ cut and is referred to as relaxed SR.

The multijet background process in the SR mainly originates from the leptonic decay of heavy flavour hadrons within jets, which give rise to real E_T^{miss} from a neutrino and to a non-isolated lepton. Therefore, the discriminant of the multijet fit is chosen to be the multiplicity of jets containing muons. The non-multijet background processes are estimated from MC simulation. The multijet template is defined as the data minus all non-multijet background processes. Under the assumption that the shape of the multijet template is the same in the multijet CR and in the relaxed SR, the multijet template derived in the CR for the jet with muons multiplicity distribution is used in the relaxed SR for the fitting procedure.

The parameter of interest of the multijet fit is the multijet normalisation. The multijet normalisation is allowed to be a free parameter in the binned profile likelihood fit used in the multijet fitting procedure. Non-multijet background processes have normalisations constrained to theory predictions within uncertainties. The uncertainties are assumed to be 6% for $t\bar{t}$, 20% for $Z + jets$, 20% for $W + jets$, 5% for single top, 10% for diboson and 20% for SM Vh background processes. The luminosity normalisation is an overall normalisation parameter of all non-multijet background processes, which is allow to vary in the multijet fit within an uncertainty of 3.4%.

The multijet fit to data in the relaxed SR, considering the multiplicity of jet with muons as discriminant, is performed only for 150 GeV $< E_T^{miss} <$ 200 GeV and 200 GeV $< E_T^{miss} <$ 350 GeV subregions, as the multijet background is negligible for larger values of E_T^{miss}. In Figure 8.3, the postfit distributions of the multiplicity of jet with muons is shown for 1 tag and 2 tag subregions and the E_T^{miss} subregions just mentioned. The multijet normalisations in the 1 tag, 150 GeV $< E_T^{miss} <$ 200 GeV and 200 GeV $< E_T^{miss} <$ 350 GeV subregions are 0.14 ± 0.01 and 0.05 ± 0.01, respectively, while the multijet normalisations in the 2 tag, 150 GeV $< E_T^{miss} <$ 200 GeV and 200 GeV $< E_T^{miss} <$ 350 GeV subregions are $0.12 \pm 0..01$ and 0.06 ± 0.01, respectively.

Multijet Transfer

After obtaining the multijet normalisations from the multijet fit, the final multijet template is derived. At this stage, the multijet CR is defined using the unblinded event selection in the SR by inverting the $\min\left[\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{j_{1,2,3}})\right]$ cut to $\min\left[\Delta\phi(\vec{E}_T^{miss}, \vec{p}_T^{j_{1,2,3}})\right] < \pi/9$. The multijet template is derived for the h candidate mass distribution in the multijet CR as the data minus all non-multijet background processes. The multijet template of the h candidate mass distribution is transferred from the multijet CR to the SR by scaling it

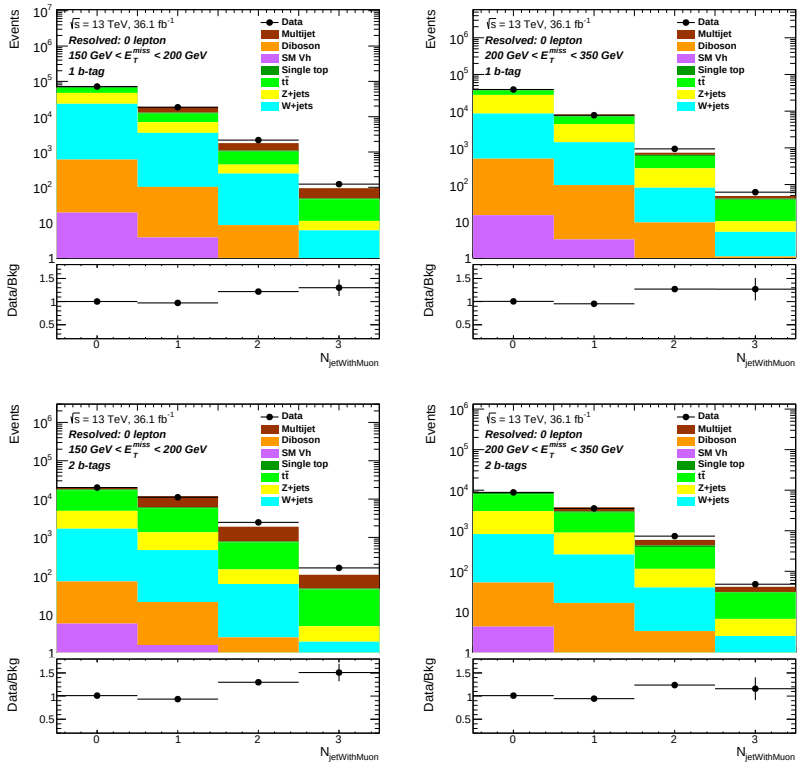

Figure 8.3: Postfit distributions of the multiplicity of jet with muons in the relaxed resolved SR, 1 tag (top) and 2 tag (bottom), 150 GeV $< E_T^{miss} <$ 200 GeV (left) and 200 GeV $< E_T^{miss} <$ 350 GeV (right) subregions.

by the previously derived multijet normalisations. The multijet background process as derived through the data-driven method is included in the total background estimation of the prefit h candidate mass spectrum shown in Figure 6.4 for the SR, 2 tag subregion.

The data-driven multijet estimate is also independently derived for the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis without optimised event selections using the same data-driven method.

9 Statistical Formalism

A fundamental quest of High Energy Physics is the search for *new physics* processes that have been predicted but not yet experimentally seen. One constructs a physics model describing known processes, referred to as *background*, and potential non-SM processes, referred to as *signal*, in order to statistically test it against data. Under a given hypothesis, the statistical analysis of the model against data allows us to make a statement either of discovery or of exclusion of new physics at a certain confidence level.

In the following, the statistical formalism used in the $h(\rightarrow b\bar{b})+E_T^{miss}$ analysis to search for new physics and exclude BSM signal models in the context of a frequentist statistical test is outlined, following References [106, 109, 110].

9.1 Binned Profile Likelihood

Let us consider an experiment where one measures an observable (e.g. the h candidate mass) for each selected event so that the data can be represented as a binned histogram in designed regions (e.g. in the SR). The expected event yield in the i-th bin can be expressed as

$$n_i^{\text{exp}} = \mu s_i + b_i, \tag{9.1}$$

where b_i denotes the expected background yield, s_i the expected signal yield in the i-th bin and μ denotes the *signal strength* defined as the ratio of the signal cross-section to the reference signal cross-section (10 fb in this analysis). The signal rate or signal cross-section is proportional to the signal strength. $\mu = 0$ corresponds to the background-only hypothesis (or null hypothesis), describing only known processes, and $\mu = 1$ to the nominal signal hypothesis, describing background and the sought after signal.

The probability, or *binned likelihood* L, of observing n^{obs} events is given by

$$L(n^{\text{obs}}|\mu) = \prod_{i=1}^{N} \frac{(\mu s_i + b_i)^{n_i^{\text{obs}}}}{n_i^{\text{obs}}!} e^{-(\mu s_i + b_i)}, \tag{9.2}$$

where n_i^{obs} stands for the number of observed events in the i-th bin with $n^{\text{obs}} = \sum_{i=1}^{N} n_i^{\text{obs}}$ and N is the number of bins in the observable histogram.

In a measurement, the signal and background yields are only known within uncertainties. The uncertainties must be incorporated into the likelihood function, which is done in the form of *nuisance parameters* $\vec{\theta} = (\theta_i)$. The values of the nuisance parameters are fitted to data by including constraint terms in the likelihood, representing each uncertainty source. Each nuisance parameter θ_i is constrained either by a Gaussian distribution or a log-normal distribution, with mean θ_i^0 and width σ_{θ_i} experimentally determined through

dedicated *auxiliary measurements* (e.g. as done for the jet energy scale at the ATLAS collaboration). Log-normal distributions are used for normalisation uncertainties. The systematic uncertainties used in the statistical model are described in Chapter 10. The signal and background yields depend on the nuisance parameters and are fitted to data in the likelihood function.

The *binned profile likelihood* function, which incorporates nuisance parameters from uncertainty sources, is given by

$$L(n^{\text{obs}}|\mu,\vec{\theta}) = \prod_{i=1}^{N} \frac{(\mu s_i(\vec{\theta}) + b_i(\vec{\theta}))^{n_i^{\text{obs}}}}{n_i^{\text{obs}}!} e^{-(\mu s_i(\vec{\theta})+b_i(\vec{\theta}))} \cdot \prod_{\theta_j,\theta_k\in\vec{\theta}} \text{Gauss}\left(\theta_j|\theta_j^0,\sigma_{\theta_j}\right) \text{Lognorm}\left(\theta_k|\theta_k^0,\sigma_{\theta_k}\right). \tag{9.3}$$

9.2 Test Statistic

For our purposes, a *test statistic* is a function of the data used to distinguish between the hypothesis that the observed data consists of background only (null hypothesis) and the hypothesis that the data consists of signal and background (alternative hypothesis).

In order to test a hypothesized value of μ, one defines the *profile likelihood ratio* as

$$\lambda(\mu) = \frac{L(\mu,\hat{\hat{\vec{\theta}}}(\mu))}{L(\hat{\mu},\hat{\vec{\theta}})}, \tag{9.4}$$

where $\hat{\hat{\vec{\theta}}}(\mu)$ stands for the *conditional* maximum-likelihood (ML) estimator of $\vec{\theta}$, which is the value of $\vec{\theta}$ that maximises the likelihood function L given a specific value of μ. $\hat{\mu}$ and $\hat{\vec{\theta}}$ stand for the ML estimators of μ and $\vec{\theta}$, respectively, which (unconditionally) maximise L.

The profile likelihood ratio λ takes values $0 \leq \lambda \leq 1$, with λ near 1 implying good agreement between the data and the hypothesized value of μ. Using $\lambda(\mu)$, one defines the test statistic under the background-only hypothesis, $\mu = 0$, as

$$q_0 = \begin{cases} -2\ln\lambda(0) & , \hat{\mu} \geq 0, \\ 0 & , \hat{\mu} < 0, \end{cases} \tag{9.5}$$

considering new physics that can only lead to an increase in the number of events observed; that is why, $q_0 = 0$ for $\hat{\mu} < 0$.

One defines a test statistic under the signal *plus* background hypothesis ($s+b$ hypothesis), for establishing upper limits on the signal strength μ, as

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & , \hat{\mu} \leq \mu, \\ 0 & , \hat{\mu} > \mu, \end{cases} \tag{9.6}$$

where $\hat{\mu} > \mu$ is not taken as part of the rejection region of the test.

The probability density function (PDF) of the test statistic q_μ under the $s+b$ hypothesis is denoted by $f(q|s+b) = f(q_\mu|\mu s(\vec{\theta}) + b(\vec{\theta}))$ while the PDF of q_0 under the background-only hypothesis is denoted by $f(q|b) = f(q_0|b(\vec{\theta}))$.

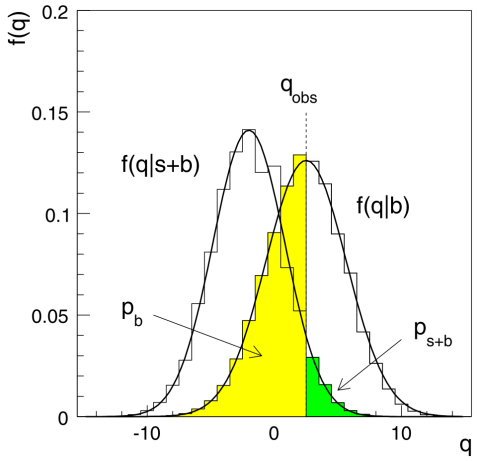

Figure 9.1: $f(q|s+b)$ and $f(q|b)$ distributions under the hypotheses of $\mu = 1$ and $\mu = 0$, respectively. Figure from Reference [106].

At a given observed q^{obs}, which is the value of the test statistic observed from the data, one quantifies the level of disagreement between the data and the hypothesis with the *p-value*. A smaller p-value corresponds to increasing incompatibility between the data and the hypothesis. Under the $s+b$ hypothesis, the p-value, p_{s+b}, is defined as the probability of obtaining a value of the test statistic q_μ that is equal or less likely than the observed test statistic q^{obs},

$$p_{s+b} = \int_{q^{\text{obs}}}^{\infty} f(q|s+b)dq. \tag{9.7}$$

Similarly, under the background-only hypothesis, the p-value, p_b, is defined as the probability of obtaining a value of the test statistic q_0 that is equal or less likely than the observed test statistic q^{obs},

$$p_b = \int_{-\infty}^{q^{\text{obs}}} f(q|b)dq. \tag{9.8}$$

The p_{s+b} and p_b values under the $s+b$ and background-only hypotheses, respectively, are illustrated in Figure 9.1.

The *significance* Z is defined such that a Gaussian distributed variable found Z standard deviations above its mean has an upper-tail probability equal to p, where p corresponds to the p-value. That is,

$$Z = \Phi^{-1}(1-p), \tag{9.9}$$

where Φ^{-1} is the inverse of the cumulative distribution of the Normal distribution. In the High Energy Physics community, the rejection of the background-only hypothesis with a significance of at least $Z_{p_b} = 5$ is regarded as an appropriate level to constitute discovery. This corresponds to a p_b less than 2.87×10^{-7}.

9.3 CL_s Limit Setting

Based on the definition of p_{s+b}, which quantifies the disagreement between the data and the $s+b$ hypothesis, one may exclude a signal model at a confidence level of $1-\alpha =$

95% if $p_{s+b} < \alpha = 0.05$. However, this exclusion criterion can lead to excluding signal models to which the search has no or little sensitivity. If a search has little sensitivity to a specific signal model, the number of expected signal yield is small with respect to the expected background yield. In this case, the distributions of the test statistic for the $s+b$ hypothesis, $f(q|s+b)$, and the background-only hypothesis, $f(q|b)$ (see Figure 9.1), are close together, having a large overlap. If the observed number of events has a sufficiently large downward fluctuation below the expected background, p_{s+b} decreases, so that the search may exclude a signal model to which it has no or low sensitivity. To avoid this, the CL_s probability is defined as

$$\mathrm{CL}_s \equiv \frac{p_{s+b}}{1-p_b}. \tag{9.10}$$

The $s+b$ hypothesis is rejected at 95% confidence level if $\mathrm{CL}_s < 0.05$.

In the scenario of a downward fluctuation, if a search is sensitive to a signal model, the two distributions $f(q|s+b)$ and $f(q|b)$ are well separated, $1-p_b$ is small but p_{s+b} is much more smaller, so $CL_s \ll 1$. On the other hand, if a search has little sensitivity to a signal model, the two distributions are then close together, $1-p_b$ and p_{s+b} are of the same order, so CL_s increases reaching values of order 1, which avoids exclusion of the signal model. As $\mathrm{CL}_s > p_{s+b}$, a subset of the signal models excluded by $p_{a+b} < \alpha$ are excluded, so limits from the CL_s method get weaker: CL_s is a conservative limit setting.

The *confidence interval* at confidence level $\mathrm{CL} = 1-\alpha$ for the signal strength μ of the signal process consists of the set of μ values that are not excluded, for which $\mathrm{CL}_s(\mu) \geq \alpha$. The endpoints of the confidence interval are found by solving (numerically) the equation $\mathrm{CL}_s(\mu) = \alpha$. The upper limit on μ, which can be translated into an upper limit on a cross-section, is the largest μ for which $\mathrm{CL}_s(\mu) \geq \alpha$.

As deduced from the Wald's [111] and Wilks's [112] Theorems, the test statistics q_μ and q_0 follow a χ^2-distribution for one degree of freedom in the large sample limit, which is referred to as *asymptotic formula*. Using the asymptotic formulae, the *Asimov dataset* is defined as the dataset for which the Asimov observed quantities are equal to their expectation values [106], which is equivalent to a dataset in which all statistical fluctuations are suppressed[1]. The Asimov dataset is a representative dataset that replaces the large number of pseudo datasets generated with MC simulations to estimate the test statistic distributions, which decreases the computational demand significantly. For more details, see Reference [106].

[1] This is equivalent to the Asimov ML parameter estimators being equal to the true parameter values.

10 Systematic Uncertainties

Systematic uncertainties are uncertainties in the knowledge of background and signal processes that manifest themselves as uncertainties in the yields and probability density (referred to as *shape*) of the h candidate mass observable for signal and background. Systematic uncertainties arise from sources related to the reconstruction, identification and calibration of objects in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ search, referred to as *experimental systematic uncertainties*, as well as from sources related to theoretical predictions and modelling of background and signal, referred to as *theoretical systematic uncertainties*.

10.1 Experimental Systematic Uncertainties

The experimental systematic uncertainties applied in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ search are summarised in Table 10.1, in which a short description and the alias name of the uncertainty in the analysis are given. The integrated luminosity corresponds to 36.1 fb^{-1} with an uncertainty of 3.4%.

Systematic uncertainty	Description
Data	
Luminosity	Uncertainty on the integrated luminosity
Electrons	
EL_EFF_Trigger_TOTAL_1NPCOR_PLUS_UNCOR	Trigger efficiency uncertainty
EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	Reconstruction efficiency uncertainty
EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR	Identification efficiency uncertainty
EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR	Isolation efficiency uncertainty
EG_SCALE_ALL	Energy scale uncertainty
EG_RESOLUTION_ALL	Energy resolution uncertainty
Muons	
MUON_EFF_TrigSystUncertainty MUON_EFF_TrigStatUncertainty	Trigger efficiency uncertainties
MUON_EFF_STAT MUON_EFF_SYS	Reconstruction and identification efficiency uncertainty for $p_T > 15$ GeV
MUON_EFF_STAT_LOWPT MUON_EFF_SYS _LOWPT	Reconstruction and identification efficiency uncertainty for $p_T < 15$ GeV
MUON_ISO_STAT MUON_ISO_SYS	Isolation efficiency uncertainty
MUON_TTVA_STAT MUON_TTVA_SYS	Track-to-vertex association efficiency uncertainty
MUONS_SCALE	Momentum scale uncertainty
MUONS_ID	Momentum resolution uncertainty from inner detector
MUONS_MS	Momentum resolution uncertainty from muon system
Small-R Jets	
JET_GroupedNP	Energy scale uncertainty parametrised in 3 components
JET_SR1_JET_EtaIntercalibration_NonClosure	Non-closure in the jet response
JET_SR1_JER_SINGLE_NP	Energy resolution uncertainty

FT_EFF_EIGEN_B FT_EFF_EIGEN_C FT_EFF_EIGEN_L	b-tagging efficiency uncertainties for small-R jets
FT_EFF_EIGEN_extrapolation	b-tagging efficiency uncertainty on the extrapolation to high p_T jets
FT_EFF_EIGEN_extrapolation_from_charm	b-tagging efficiency uncertainty on τ-jets
Large-R Jets	
JET_Comb_Baseline_Kin JET_Comb_Modelling_Kin JET_Comb_TotalStat_Kin JET_Comb_Tracking_Kin	Energy scale uncertainties (jet p_T and jet mass scales fully correlated)
FATJET_JER	Energy resolution uncertainty
FATJET_JMS	Mass resolution uncertainty
Track-Jets	
FT_EFF_EIGEN_B FT_EFF_EIGEN_C FT_EFF_EIGEN_L	b-tagging efficiency uncertainties for track jets
FT_EFF_EIGEN_extrapolation	b-tagging efficiency uncertainty on the extrapolation to high p_T jets
FT_EFF_EIGEN_extrapolation_from_charm	b-tagging efficiency uncertainty on τ-jets
E_T^{miss} Trigger	
METTrigStat METTrigSyst	Trigger efficiency uncertainty

Table 10.1: Summary of the experimental systematic uncertainties considered in the analysis. Small uncertainties from jet vertex tagger efficiency and soft term E_T^{miss} are not available in some samples of the MC production, therefore they are not included in the analysis.

10.1.1 E_T^{miss} Trigger Uncertainties

The uncertainties on the trigger correction scale factors (SFs) in MC have a component related to statistical uncertainty, estimated as the $+\sigma$ fit uncertainty band as shown in Reference [113], slide 5, and a component related to systematic uncertainty. For the systematic uncertainty, three sources for variations are considered: the *flavour composition*, the *fit range* of the SF fit and the *background composition*. The flavour composition approach assesses the variation in the SFs by considering data and MC background either inclusive in b-tag or only for 1 tag and 2 tag together since the difference in the calorimeter response to light-quark-initiated and b-quark-initiated jets may affect the E_T^{miss} calculation. The fit range approach assesses the variation in the SFs by changing the low limit of the fit range from 120 GeV to 100 GeV. The background composition approach assesses the variation in the SFs by considering the MC background as consisting of either only $W + jets$ events or only $t\bar{t}$ events, which addresses the change in the E_T^{miss} trigger efficiency depending on the event topology. The variations are calculated as the absolute difference between the SF fits in each case (e.g. for the flavour composition, difference between the SF fit inclusive in b-tag and the SF fit only for 1 tag and 2 tag).

The trigger systematic uncertainty is chosen as the variation with the largest SF fit absolute difference. The largest variation originates from the flavour composition [73]. The trigger systematic uncertainty is shown in Reference [113], slide 6.

In Figure 10.1, the E_T^{miss} trigger SF fit as a function of $E_{T,noMU}^{miss}$ (denoted in the Figure as $\not{E}_T$) is shown for the nominal $h(\to b\bar{b}) + E_T^{miss}$ analysis inclusive in b-tag, for the flavour composition variation considering 1 tag and 2 tag together, and for the $h(\to b\bar{b}) + E_T^{miss}$

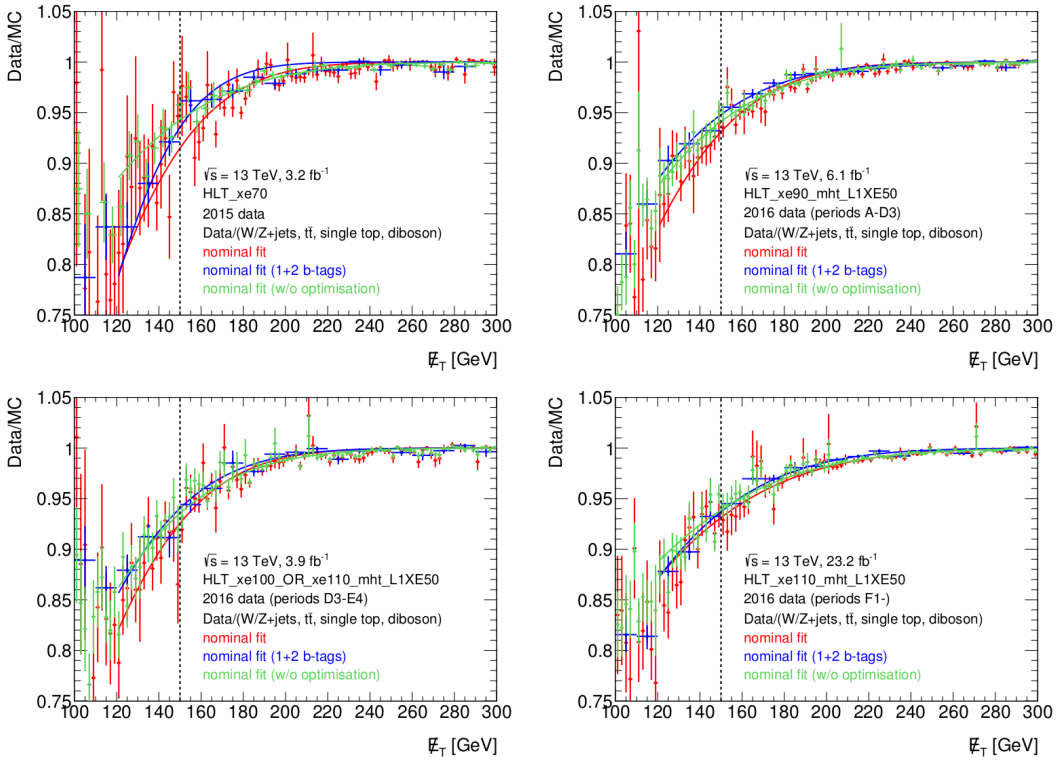

Figure 10.1: E_T^{miss} trigger SF fit as a function of $E_{T,noMU}^{miss}$ (denoted as $\not\!\!E_T$) in red for the nominal $h(\to b\bar{b}) + E_T^{miss}$ analysis inclusive in b-tag, in blue for the flavour composition variation considering 1 tag and 2 tag together, and in green for the $h(\to b\bar{b}) + E_T^{miss}$ analysis inclusive in b-tag without considering the optimised event selections. The E_T^{miss} triggers are listed in Subsection 4.3.1. This figure was provided by S. Suchek, a collaborator of the $h(\to b\bar{b}) + E_T^{miss}$ analysis group.

analysis inclusive in b-tag without considering the optimised event selections described in Chapter 7. The E_T^{miss} triggers are listed in Subsection 4.3.1. The E_T^{miss} trigger uncertainty is symmetrised, so that the systematic uncertainty corresponds to an *up* and *down variation* around the nominal SF fit. The nominal SF fit of the $h(\to b\bar{b}) + E_T^{miss}$ analysis without optimised event selections falls within the E_T^{miss} trigger systematic uncertainty, in which case the same nominal SFs of the $h(\to b\bar{b}) + E_T^{miss}$ analysis can be used for it. Similarly, the same E_T^{miss} trigger systematic uncertainty for the $h(\to b\bar{b}) + E_T^{miss}$ search is used for the $h(\to b\bar{b}) + E_T^{miss}$ search without optimised event selections.

10.1.2 Small-R Jet Uncertainties

For small-R jets, uncertainties from several sources are considered in the analysis. Uncertainties on the jet energy calibration are included [114]. The nominal full set of 88 nuisance parameters for jet energy calibration originate from in-situ calibration, eta intercalibration, calibration for high-p_T jets and pile-up. The $h(\to b\bar{b}) + E_T^{miss}$ analysis uses a strongly reduced set of uncertainties to represent the full set. The reason for this is that the JES uncertainty is subdominant compared to other sources of uncertainties in the analysis,

as will be shown in Chapter 11. Four JES uncertainty components are used, which are a non-closure uncertainty on eta intercalibration and three nuisance parameters combining the remaining nuisance parameters [114]. A single nuisance parameter with only an up variation is used for the jet energy resolution (JER) uncertainty [115]. The analysis tools, JES and JER, and uncertainties on JES and on JER [116] are provided by the *JetEtMiss* combined performance group at the ATLAS collaboration.

Uncertainties on the flavour tagging efficiency SFs are also included in the analysis. The b-tagging efficiency uncertainties are derived independently for small-R jets in the resolved regime and ghost-associated track jets in the merged regime, considering nuisance parameters for light-quark-, c-quark- and b-quark-initiated jets [117, 80, 81]. The analysis tools, b-tagging SFs and uncertainties are provided by the *Flavour Tagging* combined performance group at the ATLAS collaboration.

The jet vertex tagger [118] (JVT) uncertainty has a negligible impact on nominal yields, as shown in Reference [119]. The JVT calibration was not included in some samples of the MC production used for the $h(\rightarrow b\bar{b}) + E_T^{miss}$ publication, so neither JVT SFs nor uncertainties on JVT efficiency are applied in the analysis. It is planned to include them in the next analysis iteration.

10.1.3 Large-R Jet Uncertainties

Uncertainties on the jet p_T and jet mass scale of large-R jets are included in the analysis. The uncertainty sources are the difference between data and MC simulation referred to as *baseline*, the fragmentation *modelling* as explained in Subsection 5.1.2 (calculated by the author for the track-assisted jet mass), the *tracking* reconstruction efficiency, fake rate and q/p_T bias, and finally the statistical uncertainty on the measurement, which are incorporated in the statistical model as four nuisance parameters. The jet p_T and jet mass scale uncertainties are treated as being fully correlated [120]. Uncertainties on the jet p_T resolution and on the jet mass resolution are also included [120]. The analysis tools and uncertainties are provided by the *JetEtMiss* combined performance group at the ATLAS collaboration.

10.1.4 Electron Uncertainties

Similarly to the treatment of E_T^{miss} trigger efficiencies, correction SFs for electron efficiencies in MC simulations are used in the search, which are computed as $SF = \varepsilon(\text{Data})/\varepsilon(\text{MC})$, where ε stands for efficiency. The correction SFs and their uncertainties are determined for reconstruction, identification, isolation and trigger (in the 2ℓ-CR) efficiencies in bins of transverse momentum and rapidity from data and MC using $Z \rightarrow ee$, $W \rightarrow e\nu$ and $J/\Psi \rightarrow ee$ simulated samples [121]. Uncertainties on the electron energy scale and resolution are included too [122]. The analysis tools, correction SFs and uncertainties are provided by the *E/Gamma* combined performance group at the ATLAS collaboration.

10.1.5 Muon Uncertainties

As for electrons, correction SFs for muon efficiencies are also used in the search. The correction SFs and their uncertainties are determined for reconstruction plus identification, isolation, trigger (in the 2ℓ-CR) and track-to-vertex association efficiencies using $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ simulated samples [123]. Uncertainties on the muon energy scale and resolution are included too [123]. The analysis tools, correction SFs and uncertainties are provided by the *Muon* combined performance group at the ATLAS collaboration.

10.1.6 Tau Uncertainties

In the analysis, taus are only used for the tau veto, as described in Section 7.2. Variations on the tau energy scale (TES) or on the tau identification can result in a variation on the event yields as events containing a tau or not would be either selected or discarded, respectively. A tau veto then requires considering systematic uncertainties on the tau identification and TES. For *standard* taus, several sources of variations on the TES are considered [124], while for *extended* taus, several tracking systematic uncertainties are considered [125], as the cut-based *extended* ID depends on the number of ID tracks in the core of a *extended* tau candidate. The tracking uncertainties for taus were implemented in the analysis framework by the author.

In order to assess the impact of tau uncertainties on the search results, the author studied variations from tau uncertainties on the $t\bar{t}$ yield in the SR, since $t\bar{t}$ is a main background process considerably affected by the tau veto. The considered tau uncertainties are listed in Table 10.2. It includes uncertainties on the TES from the detector effects, in-situ calibration and modelling, as well as uncertainties on the tracking from track reconstruction efficiency and fake track reconstruction.

Systematic uncertainty	Short description
Taus	
TAUS_TRUEHADTAU_SME_TES_DETECTOR	Tau energy scale uncertainty from detector effects
TAUS_TRUEHADTAU_SME_TES_INSITU	Tau energy scale uncertainty from in-situ calibration
TAUS_TRUEHADTAU_SME_TES_MODEL	Tau energy scale uncertainty from modelling
TRK_EFF_LOOSE_TIDE	Track efficiency reconstruction uncertainty
TRK_FAKE_RATE_LOOSE	Fake track reconstruction uncertainty

Table 10.2: Summary of the experimental tau systematic uncertainties considered for the tau veto.

Standard taus have a p_T requirement of $p_T > 20$ GeV in the analysis framework. Thus, the low p_T requirement must be shifted to a higher p_T value in order to see the effect of TES up variations. Upon applying the tau uncertainties, the largest mean p_T deviation from nominal of standard taus is found to be 1.2 GeV for $t\bar{t}$ events in the SR, inclusive in E_T^{miss} and in b-tag multiplicity, before applying the tau veto. In Figure 10.2, the $t\bar{t}$ yield variation as a function of the tau uncertainty producing the variation is shown in the SR, inclusive in E_T^{miss} and in b-tag. The largest deviation from the nominal $t\bar{t}$ yield upon applying tau uncertainties is less than 0.5%. The tau uncertainties thus have a negligible

impact on the $t\bar{t}$ yield, so one concludes that the tau uncertainties can be neglected from the statistical fit model.

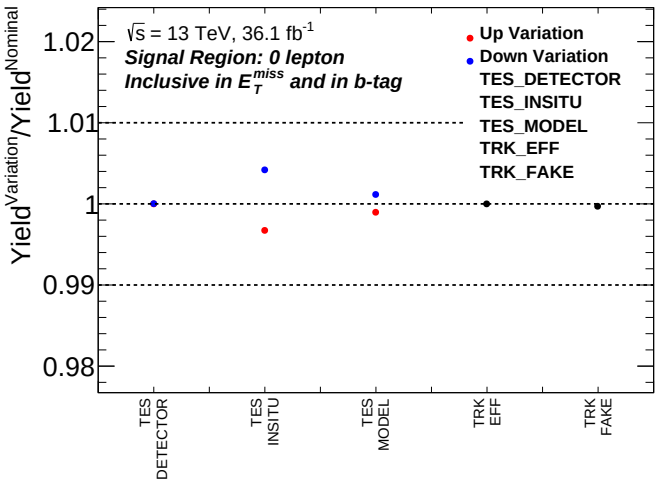

Figure 10.2: $t\bar{t}$ yield variation as a function of the tau uncertainty in the SR, inclusive in E_T^{miss} and in b-tag multiplicity. The largest deviation from the nominal $t\bar{t}$ yield is less than 0.5%.

10.1.7 E_T^{miss} Uncertainties

The uncertainty on the hard term E_T^{miss} (see Section 5.3) results from the propagation of the uncertainties of reconstructed objects used to form the hard term E_T^{miss}, so no explicit nuisance parameter is included in the analysis for the hard term E_T^{miss}. Soft term E_T^{miss} uncertainties are negligible, as shown in Reference [119]. Uncertainties on the soft term E_T^{miss} [126] were not included in some MC samples used for the $h(\rightarrow b\bar{b}) + E_T^{miss}$ publication, so they are not applied in the analysis. It is planned to include them in the next analysis iteration.

10.2 Theoretical Systematic Uncertainties

10.2.1 Background Modelling Uncertainties

The background estimation used in the statistical model depends on the MC simulation modelling (matrix element calculation, parton shower, production of radiation, fragmentation and hadronisation modelling, etc) of the chosen MC generator. An assignment of dedicated modelling uncertainties is performed for the main background processes in the SR, namely $t\bar{t}$, $Z + jets$ and $W + jets$, following methods developed in the SM $Vh(bb)$ analysis, documented in Reference [127]. These are uncertainties on the shape of the p_T spectrum of the V boson for $V + jets$ and of the leptonically decaying W boson for $t\bar{t}$ event, which is related to the E_T^{miss} distribution in the SR, as well as uncertainties on the shape of the h candidate mass spectrum.

The uncertainty estimation as developed by the SM $Vh(bb)$ analysis is as follows:

- $t\bar{t}$ **and** $W + jets$ **samples:** The uncertainties on the shape of the $p_T(\ell, E_T^{miss}) = p_T^V$ and the h candidate mass distributions are estimated in the 1ℓ CR by comparing the p_T^V and the h candidate mass distributions normalised to unit area of different MC generators as described in Reference [128]. The largest variation from the nominal shape is fitted with a functional form and then symmetrised (up and down variation) to define the modelling uncertainty.

- $Z + jets$ **sample:** The uncertainties on the shape of the $p_T(\ell, \ell) = p_T^V$ and the h candidate mass distributions are estimated in the 2ℓ CR, which is enriched with heavy flavour $Z + jets$ events at high purity. As a consequence of the high purity of heavy $Z + jets$ events, the modelling uncertainty is defined as the difference of the shapes of data and MC templates normalised to unit area, which is parametrised with an analytical function.

In Table 10.3, the above mentioned modelling uncertainties on the shape of p_T^V and h candidate mass distributions are summarised for $t\bar{t}$, $Z + jets$ and $W + jets$ samples.

Systematic uncertainty	Description
	$t\bar{t}$
SysTTbarMBB	The top modelling uncertainty on m_{jj}
SysTTbarPTV	The top modelling uncertainty on p_T^V
	Z + jets
SysZMbb	The Z modelling uncertainty on m_{jj}.
SysZPtV	The Z modelling uncertainty on p_T^V
	W + jets
SysWMbb	The W modelling uncertainty on m_{jj}
SysWPtV	The W modelling uncertainty on p_T^V

Table 10.3: Summary of the shape modelling uncertainties for $t\bar{t}$, $Z + jets$ and $W + jets$ samples.

Systematic uncertainties on the normalisation parameters of the background processes other than $t\bar{t}$, $Z + jets$ and $W + jets$ are implemented with a prior from theory predictions. These uncertainties as used in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis are as follows:

- $V + ll$: 10%, where l stands for *light flavour* and V for Z, W bosons.

- $V + cl$: 30%.

- Single top production: 4.6%, 4.4% and 6.2% for s-, t- and Wt-channel, respectively.

- Diboson production: 20%, 26% and 25% for ZZ, WZ and WW diboson production, respectively.

- SM Vh production: 50% uncertainty on Zh *plus* Wh normalisation.

Here, the notation for labelling of $Z+jets$, $W+jets$ follows the explanation in Subsection 6.1.3.

10.2.2 Signal Acceptance Uncertainties

Uncertainties on the acceptance of the $Z^{'}-2HDM$ model, which is the signal model used for analysis interpretation, are included in the statistical model. These uncertainties originate from the signal production modelling. They are evaluated by considering the following variations on modelling parameters used in the signal MADGRAPH + PYTHIA 8 samples:

- **Renormalisation and factorisation scales:** The default renormalisation and factorisation scales used in MADGRAPH [65] when generating signal events are dynamically set to $M_T^2+P_T^2$, where M_T and P_T stand for the transverse mass and transverse momentum of the four-momentum sum of the final state particles. The variation on the scales corresponds to coherently changing them by a factor of 2 and $1/2$.

- **Tune variations:** Uncertainty sources on the final state radiation, initial state radiation and multi-parton interactions are parametrised as a subset of tune variations providing maximal variation coverage for underlying event effects, jet structure effects and additional jet production.

- **Parton showering and hadronisation PDFs:** The variation is evalued by replacing the nominal NNPDF2.3LO PDF [64] with MSTW2008lo68cl PDF [129] and CTEQ6L1 PDF [130] and taking the largest deviation from the nominal PDF in signal acceptance.

The mentioned variations on the signal acceptance are evaluated at particle level mirroring the event selection of the $h(\rightarrow b\bar{b})+E_T^{miss}$ analysis when analysing the signal generated samples with a given variation. The signal acceptance variation from each uncertainty source is calculated for several mass points of the $Z^{'}-2HDM$ model at each of the four E_T^{miss} regions considered in the analysis. The variations are symmetrised. The implementation of these uncertainties was not performed by the author.

10.3 Statistical Fit Model

As follows from Chapters 6 and 8, the events in the $h(\rightarrow b\bar{b})+E_T^{miss}$ analysis are classified in 24 orthogonal channels according to the lepton multiplicity (the SR, 0 lepton region, and the control regions, 1μ CR and 2ℓ CR), to the b-tag multiplicity (either 1 b-tag or 2 b-tag) and to the p_T^V split into 4 categories (three subregions in the resolved regime and the fourth one being the merged regime, where $p_T^V = E_T^{miss}, E_{T,noMU}^{miss}, p_T^{\ell\ell}$ in the SR, 1μ CR and 2ℓ CR, respectively). The event categories are summarised in Table 10.4.

	Signal Region	$t\bar{t}$ **and** W**+jets Control Region**	Z**+jets Control Region**
Lepton number	0 lepton	1 muon	2 leptons
b-tag		1 tag and 2 tag	
p_T^V		$(150,200)$, $(200,350)$, $(350,500)$ and $(500,+\infty)$ GeV	

Table 10.4: Event categories in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis.

The variable used to discriminate between signal and background in the SR is the h candidate mass, the muon electric charge is used as fitting variable in the 1μ CR[1] and the event yield in the 2ℓ CR. In the SR, the bin width used for the h candidate mass spectrum is 5 GeV in the first and second E_T^{miss} categories, 10 GeV in the third E_T^{miss} category in the resolved regime and 20 GeV in the merged regime, using a larger bin width as the event statistics decreases. The course binning used in the 1μ CR and 2ℓ CR increase the robustness of the fitting procedure without decreasing the sensitivity of the search, as the CRs serve the constraining of background normalisation. Only events within the h candidate mass window $(50\text{ GeV}, 280\text{ GeV})$ are considered for the fit.

In the combined profile likelihood fit, four free normalisation parameters are considered, for which the rate is completely determined from data without using prior information. The first one is the *signal strength* μ, which is the parameter of interest of the statistical model. The statistical fit model considers three additional free normalisation parameters of the main background processes, namely $t\bar{t}$, heavy flavour $Z + jets$ denoted as $Z + HF$ and heavy flavour $W + jets$ denoted as $W + HF$[2], which are fitted simultaneously in the SR and CRs.

In addition to these free normalisation parameters, the statistical fit model contains nuisance parameters that are constrained by prior knowledge. These nuisance parameters correspond to experimental systematic uncertainties, background modelling and signal modelling uncertainties. The normalisation parameters of the remaining background processes other than $t\bar{t}$, $Z + HF$ and $W + HF$ follow a log-normal distribution, whereas non-normalisation nuisance parameters follow a Gaussian distribution. These nuisance parameters are individually correlated across all channels.

Following the studies in References [131, 132], partial decorrelation between analysis categories and/or background components is introduced for some normalisation nuisance parameters as follows:

- $Z + HF$ in 0 lepton SR *versus* $Z + HF$ in 2ℓ CR: 20%, correlated across b-tag channels.
- $W + HF$ in 0 lepton SR *versus* $W + HF$ in 1μ CR: 20%, correlated across b-tag channels.

[1] The muon electric charge is used as fitting variable in the 1μ CR because it is asymmetric for $W + jets$ while symmetric for $t\bar{t}$, allowing seperation of both background processes.

[2] Heavy flavour $V + jets$ refers to $V + (bb, bc, cc, bl)$, as introduced in Subsection 6.1.3.

- Ratios $W+bl/W+HF$, $W+bc/W+HF$ and $W+cc/W+HF$: 20%, correlated across channels.
- Ratios $Z+bl/Z+HF$, $Z+bc/Z+HF$ and $Z+cc/Z+HF$: 20%, treated separately in 0 lepton SR and 2ℓ CR, correlated across b-tag channels.
- Multijet: 100%, normalisations decorrelated per channel (where data-driven multijet estimation is included).

The overall $Z+HF$ normalisation is decorrelated between the 0 lepton SR and the 2ℓ CR within an uncertainty of 20%. This means that the $Z+HF$ normalisation parameter is fitted to data in the 2ℓ channel while the normalisation in the 0 lepton channel is controlled by the overall $Z+HF$ normalisation (simultaneously fitted to data in the 0 lepton channel) times an additional nuisance parameter that is simultaneously constrained within an uncertainty of 20% following a Gaussian distribution. The additional scaling nuisance parameter represents the relative acceptance variation of $Z+HF$ in the 0 lepton SR with respect to the 2ℓ CR [131]. The same treatment of the $W+HF$ normalisation is performed between 0 lepton SR and 1μ CR [131]. The flavour composition of the heavy $V+jets$ background is decorrelated, considering separately relative acceptance variations of $V+bl$, $V+bc$ and $V+cc$ with respect to $V+HF$ within an uncertainty of 20%, with V denoting the Z, W bosons. The flavour composition decorrelation for $W+HF$ is considered inclusively across channels, whereas the flavour composition decorrelation for $Z+HF$ is treated individually in the 0 lepton SR and 2ℓ CR, motivated by particle-level studies as documented in References [131, 132]. The decorrelation uncertainties used in the $h(\rightarrow b\bar{b})+E_T^{miss}$ analysis are similar to those quoted in References [131, 132], and it was checked that it barely affects the results.

The statistical fit model of the $h(\rightarrow b\bar{b})+E_T^{miss}$ analysis is a result of common efforts of many members in the mono-h analysis group without direct contribution of the author.

11 Results

The main goal for the development of new refined event selections is to reduce the $t\bar{t}$ background process, which was the dominant background process in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis using 3.2 fb^{-1}, to increase the experimental sensitivity of the search. As seen in Figure 7.29, top, $t\bar{t}$ is reduced by $\sim 50\%$ and $\sim 70\%$ by the new refined event selections in the SR, resolved and merged regime, respectively. In order to assess the increase of search sensitivity from the inclusion of the new refined event selections alone, the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis is performed a second time by the author, keeping the same analysis strategy while dropping the new refined event selections. The $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis results are then compared to those of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis without new refined event selections.

The $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis results are statistically interpreted using two hypotheses: the background-only hypothesis and the signal *plus* background hypothesis. In the following, the analysis results under the background-only hypothesis are presented in Section 11.1 using an integrated luminosity of 36.1 fb^{-1} of proton-proton collisions at 13 TeV. Finally, the analysis results under the signal *plus* background hypothesis are presented in Section 11.2 using the same dataset.

11.1 Search for New Physics

The search for new physics is statistically performed under the background-only hypothesis, which is tested using the test statistic q_0 defined in Equation 9.5. In the following, the fit results obtained using this hypothesis are presented.

11.1.1 Fitting of the Statistical Model to Data

Nuisance Parameter Constraints

Upon fitting the binned profile likelihood statistical model to data, the favoured fitted nuisance parameter values can differ from the *prior value* of the parameter, which is provided either from theoretical predictions or auxiliary measurements. In this case, one says that the nuisance parameter is *pulled*. The *pull* of a nuisance parameter is defined as $(\theta_{\text{fit}} - \theta_0)/\Delta\theta$, where $\theta_{\text{fit}} = \hat{\theta}$ is the fitted nuisance parameter value maximising the profile likelihood function, θ_0 is the prior value of the parameter and $\Delta\theta$ is the prior uncertainty on the nuisance parameter.

In Figure 11.1, the nuisance parameter pulls with their constraints are shown upon fitting to the Asimov data [106] for the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis. As expected from the definition of the Asimov data, all parameter estimators are equal to the expected values,

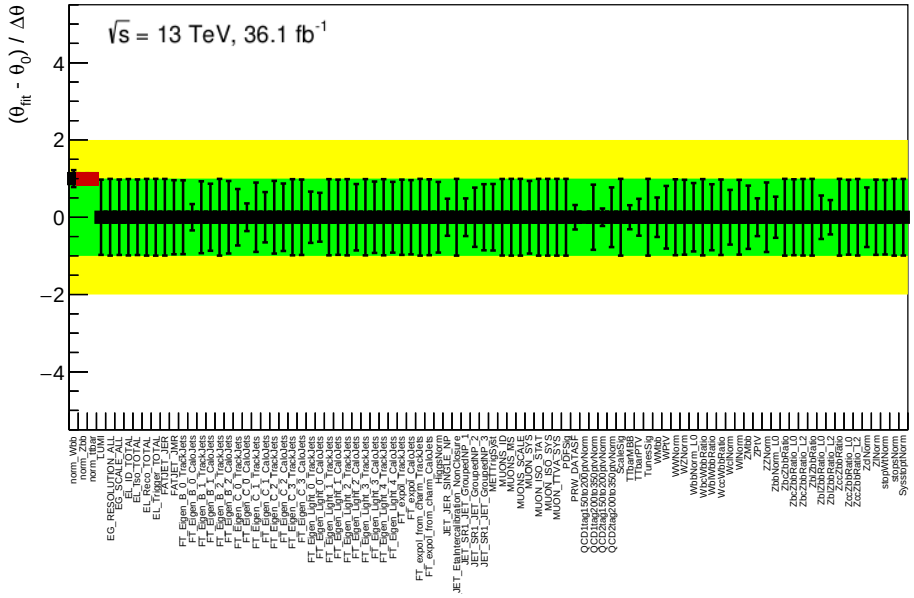

Figure 11.1: The nuisance parameter pulls upon fitting to the Asimov data for the $h(\to b\bar{b}) + E_T^{miss}$ analysis.

so the pulls are centred at zero. The same pull behaviour is found for the $h(\to b\bar{b}) + E_T^{miss}$ analysis without optimised event selections upon fitting to the Asimov data, as shown in Appendix A.1.

In Figure 11.2, the nuisance parameter pulls with their constraints are shown upon fitting to the observed data for the $h(\to b\bar{b}) + E_T^{miss}$ analysis. The statistical model fit behaves well as most pulls are centred around zero for the $h(\to b\bar{b}) + E_T^{miss}$ analysis. The postfit values of the free normalisation nuisance parameters are 1.07 ± 0.20 for the heavy flavour $W + jets$, $W + HF$, 1.12 ± 0.07 for the heavy flavour $Z + jets$, $Z + HF$, and 0.99 ± 0.03 for the $t\bar{t}$ for the $h(\to b\bar{b}) + E_T^{miss}$ analysis. An almost identical behaviour of the nuisance parameter pulls is found for the $h(\to b\bar{b}) + E_T^{miss}$ analysis without optimised event selections upon fitting to the observed data, as shown in Appendix A.1.

The correlation of nuisance parameters is presented in Appendix A.2.

The only nuisance parameters that are pulled by more than a σ, but still within less than about 1.5σ, are TTbarPTV, ZbbNorm_L0 and ZblZbbRatio_L0 nuisance parameters, corresponding to the uncertainty on the shape of the p_T^V distribution of the $V = W(\ell, E_T^{miss})$ boson for $t\bar{t}$ events, to the relative uncertainty of $Z + HF$ between 0 lepton SR and 2ℓ CR, and to the relative flavour component uncertainty of $Z + bl$ to $Z + HF$ in the 0 lepton SR, respectively. The pulling of TTbarPTV is related to the modelling of the E_T^{miss} spectrum for $t\bar{t}$. It was found in the $h(\to b\bar{b}) + E_T^{miss}$ analysis that the POWHEG+PYTHIA 6 $t\bar{t}$ sample used in the analysis provides a E_T^{miss} spectrum too soft to describe the data, which in principle should be covered by TTbarPTV uncertainties. Comparing the E_T^{miss} distribution shape for $t\bar{t}$ from various MC generators to that from the nominal POWHEG+PYTHIA 6 generator, some shape deviations are however not contained within TTbarPTV uncertain-

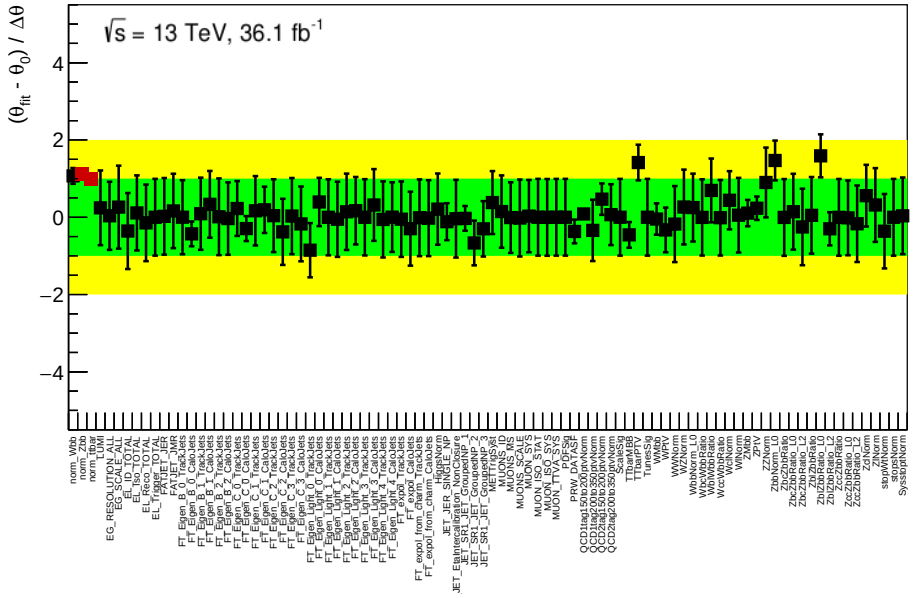

Figure 11.2: The nuisance parameter pulls upon fitting to the observed data for the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis.

ties for the selection used in this analysis [133]. The TTbarPTV uncertainty derived by the SM Vh analysis [131] is underestimated for the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis, explaining the tension in the TTbarPTV pull. Similarly, the ZbbNorm_L0 uncertainty is underestimated [131, 132]. The relative flavour component uncertainty of $Z + bl$ to $Z + HF$ is considered inclusively across b-tag categories. However, the $Z + bl$ acceptance is different between 1 b-tag and 2 b-tag with respect to $Z + HF$, as the $Z + bl$ component dominates the flavour composition of $Z + HF$ in the 1 b-tag while it is subdominant in the 2 b-tag. The analysis is sensitive to the flavour through the b-tag categories: a decorrelation between 1 b-tag and 2 b-tag for $Z + bl/Z + HF$ will resolve the tension in the ZblZbbRatio_L0 pull.

Ranking of Uncertainties

The impact of an uncertainty source is evaluated as the fractional uncertainty on the postfit signal strength $\sigma_{\hat{\mu}}/\hat{\mu}$, which is equivalent to a fractional uncertainty on the signal yield after the fit to data. The statistical model is fitted to the Asimov data using a $Z' - 2HDM$ mass point model with its cross-section from theory prediction. The postfit fractional uncertainty on the signal strength $\hat{\mu}$ from different uncertainty sources is evaluated for three representative $Z' - 2HDM$ mass point models mostly populating low E_T^{miss}, medium E_T^{miss} and large E_T^{miss} values: $Z' - 2HDM$ with $(m_A, m_{Z'}) = (0.6\text{ TeV}, 0.3\text{ TeV})$, $(1.4\text{ TeV}, 0.6\text{ TeV})$ and $(2.6\text{ TeV}, 0.3\text{ TeV})$, respectively.

The total uncertainty on the postfit signal yield is derived from the fit including all uncertainty sources. The statistical uncertainty is estimated by excluding all systematic

uncertainty sources. In order to estimate the impact of a given systematic uncertainty θ_i on the postfit signal yield, the fit is repeated excluding the systematic uncertainty (or family of systematic uncertainties) in question, resulting in a reduced postfit signal yield uncertainty $\sigma_{no\ \theta_i}$. The post signal yield uncertainty is then estimated as $\sigma_{\theta_i} \equiv \sqrt{\sigma^2_{Total} - \sigma^2_{no\ \theta_i}}$, where σ_{Total} is the total uncertainty on the postfit signal yield.

Source of uncertainty	Impact [%]		
	(a)	(b)	(c)
V+jets modelling	5.0	5.7	8.2
$t\bar{t}$, single-t modelling	3.2	3.0	3.9
SM $Vh(b\bar{b})$ norm.	2.2	6.9	6.9
Signal modelling	3.9	2.9	2.1
MC statistics	4.9	11	22
Luminosity	3.2	4.5	5.4
b-tagging, track-jets	1.4	11	17
b-tagging, small-R jets	5.0	3.4	4.7
Calibration small-R jets	1.7	3.8	2.1
Calibration large-R jets	< 0.1	1.2	4.7
Total syst. uncertainty	10	21	36
Statistical uncertainty	6	38	62
Total uncertainty	12	43	71

Table 11.1: Fractional uncertainty on the postfit signal yield upon fitting to the Asimov data using three representative $Z' - 2HDM$ mass point models (a) with $(m_{Z'}, m_A) = (0.6\ \text{TeV}, 0.3\ \text{TeV})$, (b) with $(m_{Z'}, m_A) = (1.4\ \text{TeV}, 0.6\ \text{TeV})$, and (c) with $(m_{Z'}, m_A) = (0.6\ \text{TeV}, 0.3\ \text{TeV})$. Table from Reference [15], in which the author is co-author.

In Table 11.1, the dominant sources of uncertainty in the $h(\rightarrow b\bar{b}) + E_T^{miss}$ search are presented, quoting the fractional uncertainty on the postfit signal yield. The quoted uncertainty groups multiple sources of uncertainty, e.g. the fractional uncertainty on the postfit signal yield from *Calibration small-R jets* groups the jet energy scale and resolution uncertainties as well as the jet mass scale and resolution uncertainties. The dominant theoretical systematic uncertainty arises from the modelling of $V + jets$ background processes. The dominant experimental systematic uncertainties originate from the MC statistics for background processes, the integrated luminosity and b-tagging. The impact of the luminosity uncertainty (that does not affect backgrounds with free normalisations) varies due to the changing background composition with increasing E_T^{miss}. The $h(\rightarrow b\bar{b}) + E_T^{miss}$ is statistically limited for 300 GeV $\lesssim E_T^{miss}$. The method for the uncertainty ranking was not implemented by the author.

11.1.2 Postfit distributions

In Figure 11.3, the postfit invariant mass spectrum of the h candidate is shown in the SR, 2 tag, for the four E_T^{miss} categories upon fitting to the observed data. The SM predictions are found to be in agreement with the observed data[1]. After applying optimised event selections, the irreducible $Z+jets$ is the new dominant background process in the SR for $E_T^{miss} > 200$ GeV.

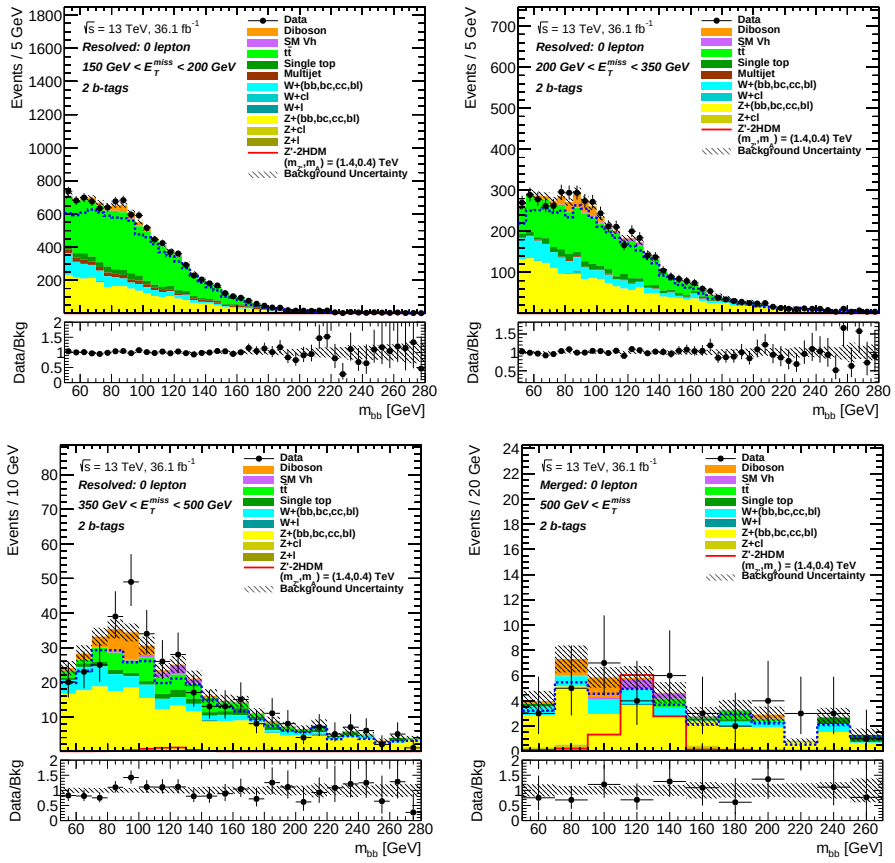

Figure 11.3: Postfit distribution of the h candidate invariant mass in the SR, 2 tag, for the four E_T^{miss} categories. The SM background expectation before the fit to data shown as a dashed blue line. A representative $Z'-2HDM$ model, with cross-section normalised to 10 fb and $(m_{Z'}, m_A) = (1.4\text{ TeV}, 0.4\text{ TeV})$, shown as a red solid line. The total background uncertainty displayed as a hatched band, considering correlations between uncertainties.

In Figure 11.4, the postfit muon charge distributions in the 1μ CR, 2 tag, for the four $E_{T,noMU}^{miss}$ categories are shown in the four upper panels within h candidate mass window $(50\text{ GeV}, 280\text{ GeV})$. The muon charge distribution is symmetric in $t\bar{t}$, from which equal number of muons with negative and positive electric charges are expected, whereas the muon charge distribution is asymmetric in $W+jets$, from which a predominance of muons with positive electric charge is expected in proton-proton collisions [108].

[1]Except for a data point around $m_{bb} = 95$ GeV in the E_T^{miss} bin of 350 GeV $< E_T^{miss} <$ 500 GeV.

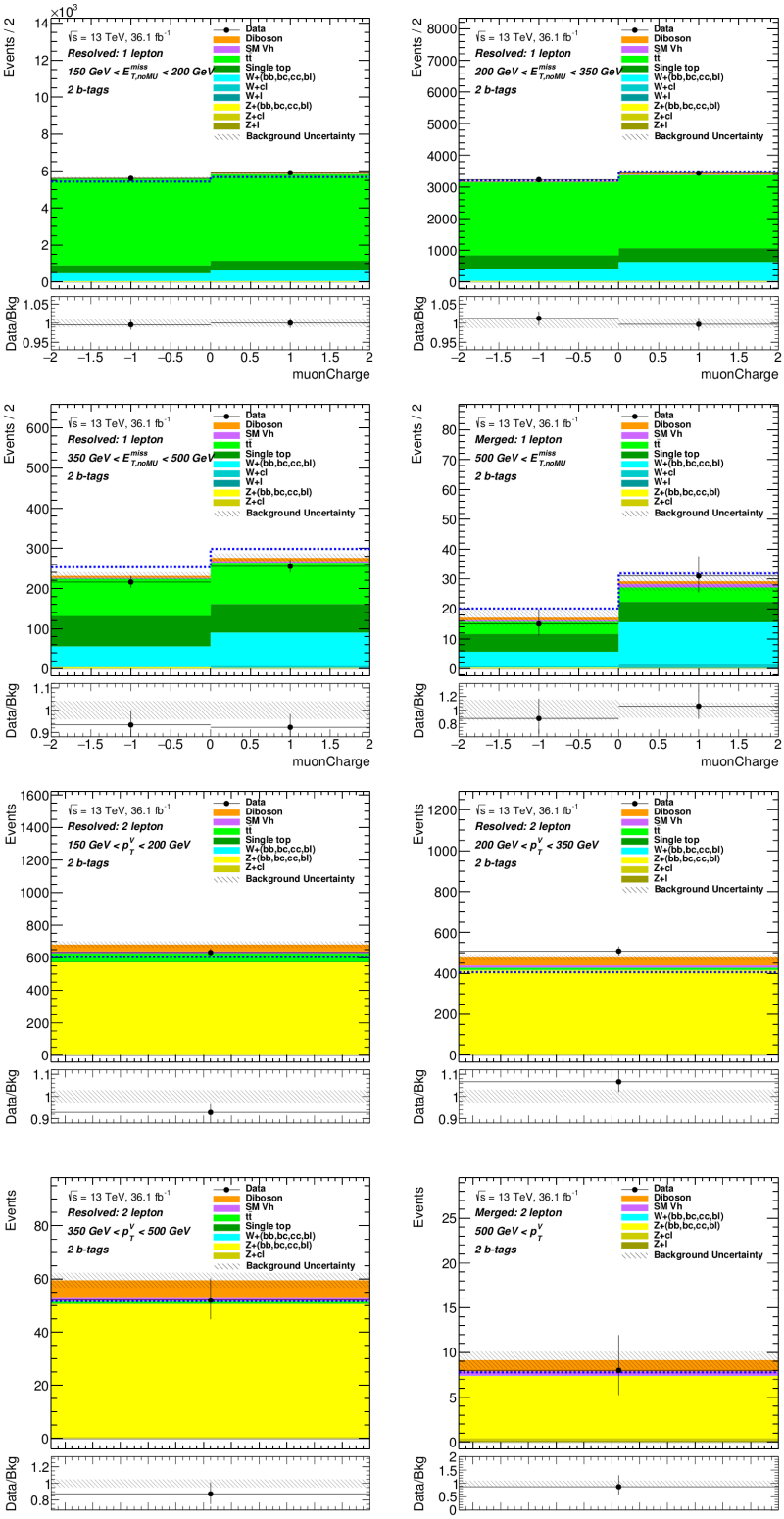

Figure 11.4: Postfit muon charge distributions in the 1μ CR, 2 tag (four upper panels) and postfit event yields in the 2ℓ CR, 2 tag (four lower panels) for the four $E^{miss}_{T,noMU}$ and $p^{\ell\ell}_T$ (p^V_T) categories, respectively, are shown within h candidate mass window $(50\ \text{GeV}, 280\ \text{GeV})$. The SM background expectation before the fit to data shown as a dashed blue line.

In Figure 11.4, the postfit event yields in the 2ℓ CR, 2 tag, for the four p_T^V ($p_T^{\ell\ell}$) categories are shown in the four lower panels within the h candidate mass window $(50\,\text{GeV}, 280\,\text{GeV})$.

11.1.3 Experimental Sensitivity of the Search

The experimental sensitivity of a search is quantified by the median discovery significance for a given signal model, which is approximated using the Asimov dataset. The median discovery significance estimated with the Asimov dataset is referred to as *expected significance*.

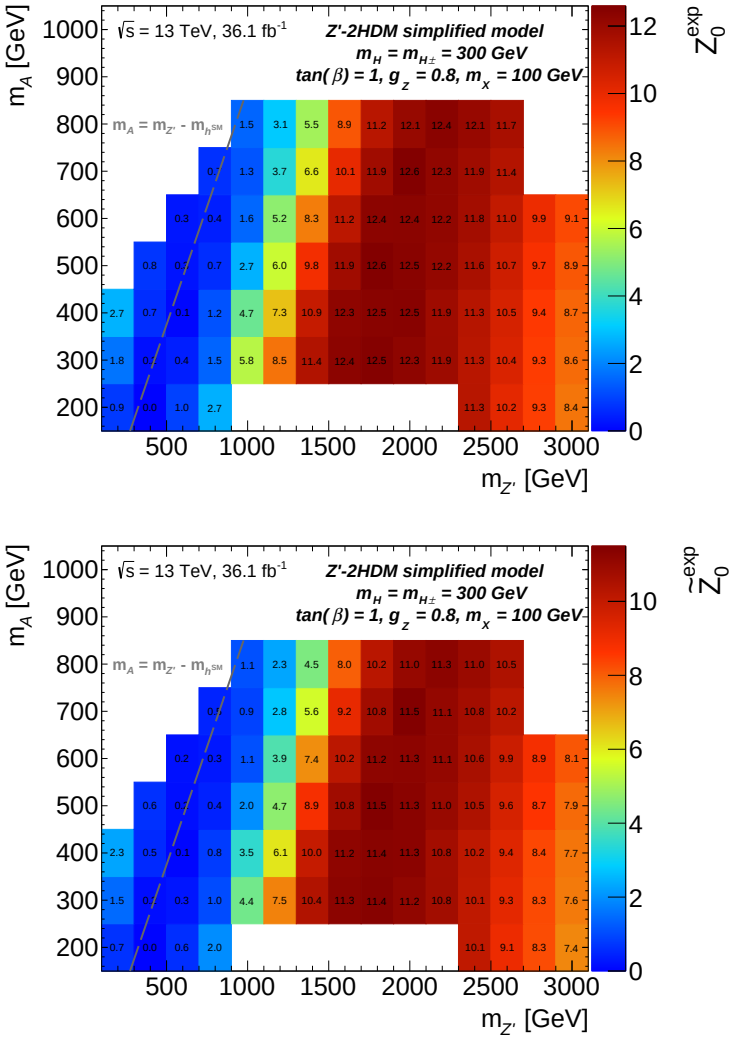

Figure 11.5: Expected significance of the $h(\to b\bar{b}) + E_T^{miss}$ analysis, Z_0^{exp}, (top) and of the $h(\to b\bar{b}) + E_T^{miss}$ analysis without optimised selections, $\tilde{Z}_0^{exp}$, (bottom) are shown for the $Z' - 2HDM$ mass point model grid. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass. The expected significance increases as the signal E_T^{miss} spectrum gets harder.

In Figure 11.5, the expected significance of the $h(\to b\bar{b}) + E_T^{miss}$ analysis (top) and of the $h(\to b\bar{b}) + E_T^{miss}$ analysis without optimised selections (bottom) are shown for the $Z' - 2HDM$ mass point model grid. The E_T^{miss} produced in a signal event scales with $m_{Z'} - m_A$: the signal E_T^{miss} spectrum gets harder (the Jacobian peak shifts to higher E_T^{miss} values) as $m_{Z'}$ increases for a given m_A and as m_A decreases for a given $m_{Z'}$ in the on-shell region [107]. The expected significance increases as the signal E_T^{miss} spectrum gets harder[2]. The $h(\to b\bar{b}) + E_T^{miss}$ analysis is therefore more sensitive to $Z' - 2HDM$ models produced with large E_T^{miss}. For signal models with $m_{Z'} \gtrsim 2.2$ TeV having very boosted topologies, the expected significance decreases as $m_{Z'}$ increases at fixed m_A, due to the increasing merging of the track jets used to tag the flavour content of the h candidate. A quick comparison between the two panels shows that the expected significance is larger for the $h(\to b\bar{b}) + E_T^{miss}$ analysis including the optimised event selections.

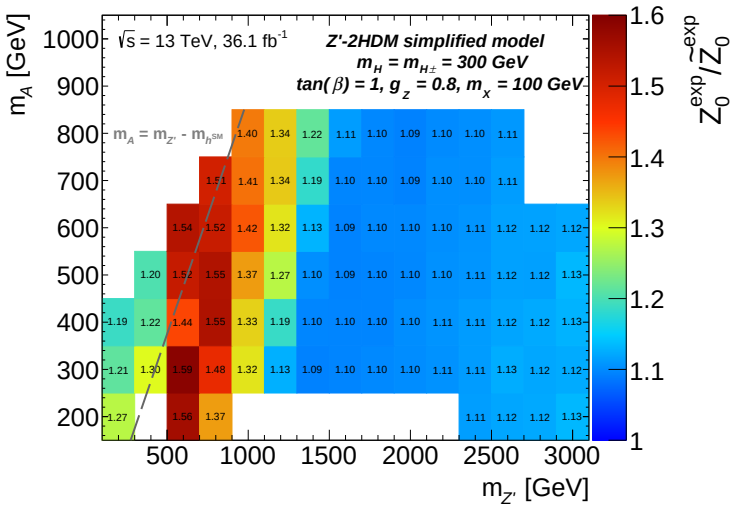

Figure 11.6: Ratio of the expected significance of the $h(\to b\bar{b}) + E_T^{miss}$ analysis to that of the $h(\to b\bar{b}) + E_T^{miss}$ analysis without optimised event selections shown for the $Z' - 2HDM$ mass point model grid. The grey dashed line is the transition line between the on-shell and off-shell regions at $m_A = m_{Z'} - m_{hSM}$, with $m_{hSM} = 125$ GeV. The expected significance of the search increases between about 10% and 60% from inclusion of the optimised event selections.

In Figure 11.6, the ratio of the expected significance of the $h(\to b\bar{b}) + E_T^{miss}$ analysis to that of the $h(\to b\bar{b}) + E_T^{miss}$ analysis without optimised event selections is shown for the $Z' - 2HDM$ mass point model grid[3], which quantifies the gain in experimental sensitivity of the search from inclusion of the optimised event selections. The significance gain takes

[2]The production cross-section for $Z' - 2HDM$ models is normalised to 10 fb in the analysis, in which scenario the expected significance is calculated. Normalising the cross-section allows to decouple signal production rate from topology and kinematics features of the signal models when calculating the sensitivity. The actual production cross-section depends on $\sqrt{s}$, here $\sqrt{s} = 13$ TeV. The sensitivity to signal models with hard E_T^{miss} spectrum is influenced by the interplay between the decreasing signal production cross-section and the decreasing background yield as the signal E_T^{miss} spectrum shifts to larger values.

[3]No significance ratio entry is included at $(m_{Z'}, m_A) = (400, 200)$ GeV because the expected significances are null for this $Z' - 2HDM$ mass point.

values between about 30% and 60% for $Z' - 2HDM$ mass points with most E_T^{miss} spectrum in the resolved regime in the on-shell region. This means that by including the optimised event selections, the $h(\rightarrow b\bar{b}) + E_T^{miss}$ search is experimentally between 30% and 60% more sensitive to $Z' - 2HDM$ models in the resolved regime, increasing considerably the chances of discovery if these models exist in Nature. The significance gain takes values between about 10% and 13% for $Z' - 2HDM$ mass points with most E_T^{miss} spectrum in the merged regime in the on-shell region. The significance gain in the resolved regime is larger than that in the merged regime, because the ΔR_{bb} is applied as a new improvement only in the resolved regime and because the chosen H_T ratio cut is more stringent in the resolved regime than in the merged regime. The expected significance of $Z' - 2HDM$ mass points in the off-shell region increases between about 20% and 55%.

Based on these results, it is concluded that the optimised event selections not only considerably decrease the $t\bar{t}$ background process in the SR, but also considerably increase the experimental sensitivity of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis.

In Figure 11.7, the observed significance of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis obtained upon fitting to the observed data is shown for the $Z' - 2HDM$ mass point model grid. No evidence for new physics in the form of $Z' - 2HDM$ models is found.

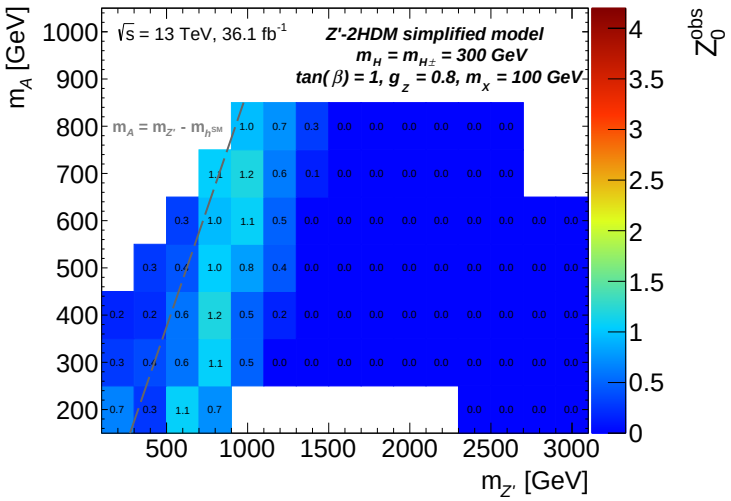

Figure 11.7: Observed significance of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis for the $Z' - 2HDM$ mass point model grid. The grey dashed line is the transition line between the on-shell and off-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$, with $m_{h^{SM}} = 125$ GeV. No evidence for new physics in the form of $Z' - 2HDM$ models is observed.

11.2 Limits on New Physics

Since no evidence for $Z' - 2HDM$ models was experimentally found, exclusion limits for $Z' - 2HDM$ models are set. The exclusion of $Z' - 2HDM$ models is statistically performed using the CL_s limit setting (explained in Subsection 9.3). The signal *plus* background hypothesis for $Z' - 2HDM$ models is excluded at 95% confidence level (CL) imposing $CL_s < 0.05$.

Using the CL_s limit setting, experimental upper limits at 95% CL on the production cross-section of $h + DM$[4] events, σ_{h+DM}, times the $h \to b\bar{b}$ branching ratio, $\mathscr{B}(h \to b\bar{b})$[5], are derived for $Z' - 2HDM$ models as a function of $(m_{Z'}, m_A)$. In the following, *cross-section* is used as shortening for *cross-section times* $h \to b\bar{b}$ *branching ratio*. Any mass point $(m_{Z'_0}, m_{A_0})$ is excluded at 95% CL if the theoretical cross-section of the $Z' - 2HDM$ model with mass parameters $(m_{Z'_0}, m_{A_0})$ is larger than the cross-section upper limit. In practice, the cross-section upper limit as a function of $(m_{Z'}, m_A)$ is interpolated from a set of mass points, whose grid is indicated in Subsection 4.2.1. The interpolation method is chosen to be a linear interpolation of the logarithmic cross-section, which was found to provide a well-behaving interpolation in the $(m_{Z'}, m_A)$ plane [134].

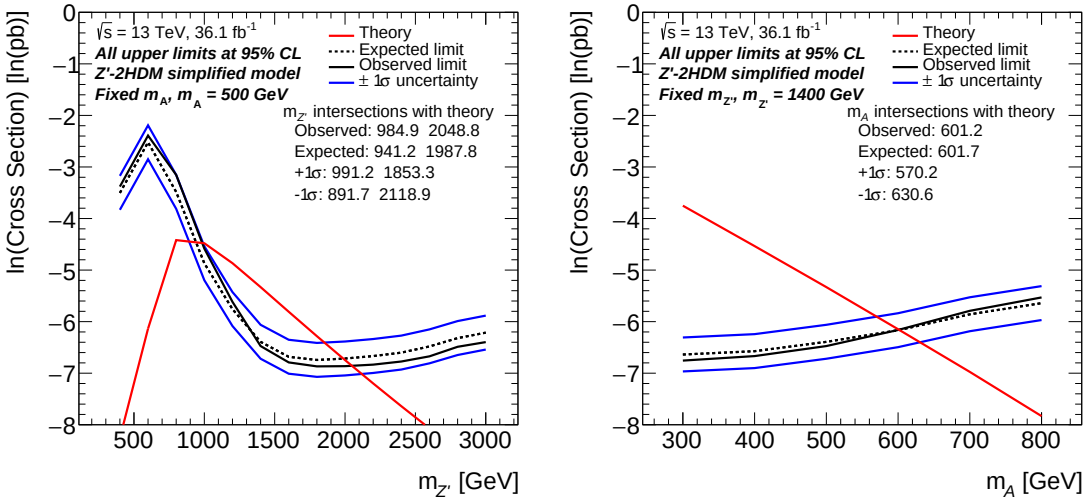

Figure 11.8: Linear interpolation of the upper limits on the logarithmic cross-section as a function of $m_{Z'}$ (m_A) for fixed $m_{A_0} = 500$ GeV ($m_{Z'_0} = 1400$ GeV) in the left (right). Expected and observed limits are shown as a dashed and solid black line, respectively, while theory cross-sections and $\pm 1\sigma$ expected upper limits are shown as a red and blue solid line, respectively. A mass point model with $(m_{Z'}, m_A)$ is excluded at 95% CL if the theoretical cross-section is larger than the cross-section upper limit.

At a fixed m_{A_0} ($m_{Z'_0}$) parameter, the upper limits of all mass point models in the grid with varying $m_{Z'}$ (m_A) parameter are linearly interpolated, as illustrated in Figure 11.8, where the linear interpolation of upper limits on the logarithmic cross-section is shown as

[4] $h + DM$ stands for events with a h boson and dark matter particles in the final state.
[5] The $h \to b\bar{b}$ branching ratio is taken as $\mathscr{B}(h \to b\bar{b}) = 0.571$.

a function of $m_{Z'}$ (m_A) at fixed $m_{A_0} = 500$ GeV ($m_{Z'_0} = 1400$ GeV) in the left (right). Let us consider the left panel in Figure 11.8. At fixed $m_{A_0} = 500$ GeV, the mass point models with $m_{Z'}$ values, for which the linear interpolation curve of the theoretical production cross-section (here referred to as *theoretical cross-section curve*) is larger than the linear interpolation curve of of the upper limits on the observed production cross-section (here referred to as *observed limit curve*), are experimentally excluded at 95% CL. The excluded mass point models with $(m_{Z'}, m_A)$ are those for which $m_A = 500$ GeV and $m_{Z'_1} < m_{Z'} < m_{Z'_2}$, where $m_{Z'_1}$ and $m_{Z'_2}$ are the intersection points between the observed limit curve and the theoretical curve. By repeating this procedure for the mass point grid, contours of upper limits on $\sigma_{h+DM} \times \mathscr{B}(h \to b\bar{b})$ in the $(m_{Z'}, m_A)$ plane are constructed from the intersection points for $Z' - 2HDM$ models, as shown in Figures 11.9, 11.10.

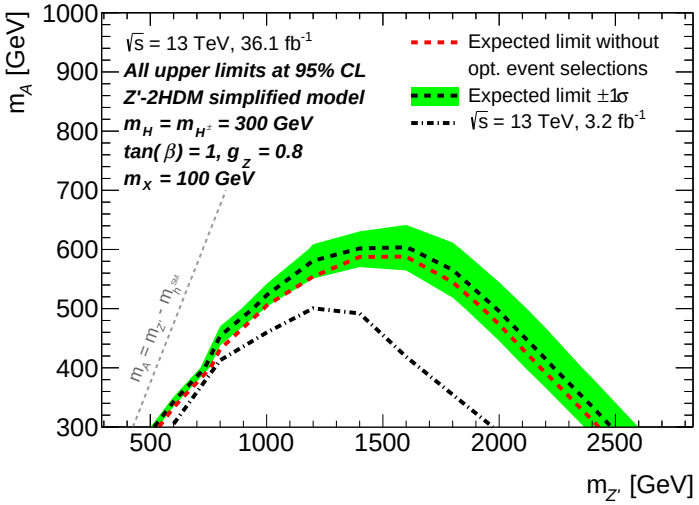

Figure 11.9: Exclusion contours of expected upper limits on $\sigma_{h+DM} \times \mathscr{B}(h \to b\bar{b})$ of $Z' - 2HDM$ models shown in the $(m_{Z'}, m_A)$ plane as a black (red) dashed contour for the $h(\to b\bar{b}) + E_T^{miss}$ analysis including (without) optimised event selections. Uncertainty bands on the expected upper limits shown in green. The expected upper limits are improved more than an uncertainty band up to $m_{Z'} \sim 1.4$ TeV and up to $m_A \sim 0.55$ TeV by the optimised event selections.

In Figure 11.9, the exclusion contours of expected upper limits on $\sigma_{h+DM} \times \mathscr{B}(h \to b\bar{b})$ of $Z' - 2HDM$ models are shown as a function of $(m_{Z'}, m_A)$. The expected upper limits for the $h(\to b\bar{b}) + E_T^{miss}$ analysis and for the $h(\to b\bar{b}) + E_T^{miss}$ analysis without optimised event selections are included to evaluate the improvement on the upper limits from the optimised event selections. The expected upper limits are improved in more than an uncertainty band by the optimised event selections for $m_{Z'}$ up to $m_{Z'} \sim 1.4$ TeV and m_A up to ~ 0.55 TeV.

In Figure 11.10, exclusion contours of the expected and observed upper limits on $\sigma_{h+DM} \times \mathscr{B}(h \to b\bar{b})$ of $Z' - 2HDM$ models are shown as a function of $(m_{Z'}, m_A)$ for the $h(\to b\bar{b}) + E_T^{miss}$ analysis. $Z' - 2HDM$ models with mass parameters up to $m_Z \sim 2.6$ TeV and $m_A \sim 0.6$ TeV are excluded at 95% CL. These conclusions are identical to those put forward by the $h(\to b\bar{b}) + E_T^{miss}$ analysis group in the publication [15], of which the author

is co-author.

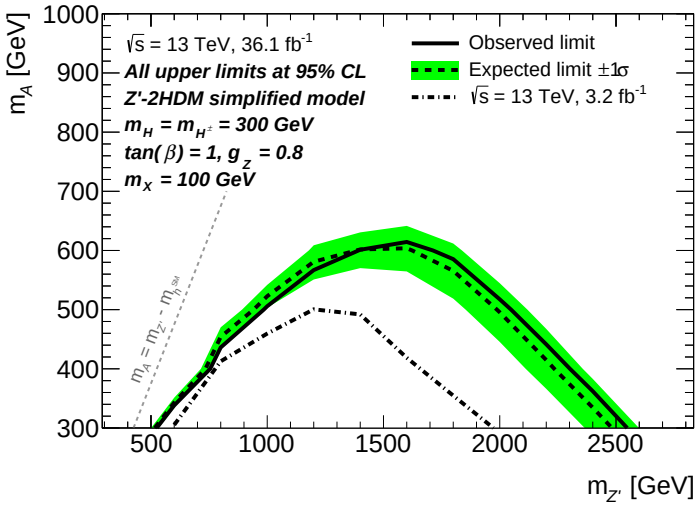

Figure 11.10: Exclusion contours of the observed (expected) upper limits on $\sigma_{h+DM} \times \mathscr{B}(h \to b\bar{b})$ of $Z' - 2HDM$ models shown in the $(m_{Z'}, m_A)$ plane as a black solid (dashed) contour for the $h(\to b\bar{b}) + E_T^{miss}$ analysis. Uncertainty bands on the expected upper limits shown in green. $Z' - 2HDM$ models with $m_{Z'}$ up to ~ 2.6 TeV and m_A up to ~ 0.6 TeV are excluded at 95% CL.

12 The Likelihood Ordering Algorithm

As outlined in Section 7.5, the optimised event selections considerably reduce the $t\bar{t}$ background process by $\sim 50\%$ and $\sim 70\%$ in the SR, resolved and merged regime, respectively. However, $t\bar{t}$ remains a main background process in the two lower E_T^{miss} categories in the SR, resolved regime, as seen in Figure 11.3. A new algorithm, the *Likelihood Ordering*, is designed as an event selection to reduce $t\bar{t}$ in the SR while further increasing the experimental sensitivity of the search.

12.1 Reconstruction of Semileptonic $t\bar{t}$

As seen in Figure 6.9, $\sim 90\%$ ($\sim 95\%$) of $t\bar{t}$ in the SR, 2 tag (1 tag), correspond to semileptonic $t\bar{t}$, of which $\sim 70\%$ ($\sim 70\%$) correspond to semileptonic $t\bar{t}$ with a hadronic tau as final state particle. A scheme of a semileptonic $t\bar{t}$ event is shown with the final state particles in Figure 6.8, right. From now on, the W boson decaying hadronically and leptonically are referred to as *hadronic W* or W_{had}, and *leptonic W* or W_{lep}, respectively. Similarly, a top quark is referred to as *hadronic top* or t_{had}, (*leptonic top* or t_{lep}) if decaying into a *hadronic W* (*leptonic W*) plus a b-quark. Each quark hadronises being reconstructed as a small-R jet in the resolved regime. The hadronic tau is reconstructed as a small-R jet too, mimicking a QCD-originated jet. The Likelihood Ordering algorithm aims at finding the correct *jet-parton assignment* assuming semileptonic $t\bar{t}$ event in order to reconstruct the *hadronic top* mass, $m_{t_{had}}$, and the *hadronic W* mass, $m_{W_{had}}$, which gives a handle to reject semileptonic $t\bar{t}$ events in the SR.

The assignment of a jet to a (parton) final state particle in an event leads to a large number of possible combinations, with only a correct jet-parton assignment in a $t\bar{t}$ event. The combinatorics increase with the number of jets in the event. Considering the two b-quarks, the two quarks from the $W \to q'\bar{q}$ decay and additionally the hadronic tau from the $W \to \tau\nu_\tau$ decay, the number of jet-parton assignments for reconstructing $m_{t_{had}}$, $m_{W_{had}}$ and $m_{t_{lep}}$ at Born level is given by

$$2 \cdot \binom{N_{jet}}{2} \binom{N_{jet}-2}{2} \binom{N_{jet}-4}{1}. \tag{12.1}$$

where N_{jet} stands for the number of central jets in the event, with $N_{jet} \geq 5$. The 2-element combination $\binom{N_{jet}}{2}$ corresponds to the b-quarks; the additional 2 factor takes into account that a b-quark decays either from t_{had} or t_{lep}. From the resting jets, $\binom{N_{jet}-2}{2}$ is a 2-element combination for the quarks from the $W \to q'\bar{q}$ decay, which are indistin-

guishable for the mass $m_{W_{had}}$. The remaining 1-element combination corresponds to the hadronic tau from the $W \to \tau\nu_\tau$ decay.

For $N_{jet} = 5$ the number of possible jet-parton assignments is 60 and for $N_{jet} = 6$ the number of possible jet-parton assignments is 360, of which only one corresponds to the correct jet-parton assignment. The number of possible combinations is decreased under the assumption that a b-tagged jet always results of the hadronisation of a real b-quark. In Figure 12.1, right (left), the truth flavour distribution of the two b-jets (of the b-jet) in the event normalised to unit area is shown in the SR, 2 tag (1 tag), resolved regime. For about 75% (95%) of $t\bar{t}$ events, the two b-jets (the single b-tagged jet) in the event originate from the hadronisation of a truth b-quark. The assignment of b-tagged jets to b-quarks is assumed in the following, reducing the combinatorics. For example, in 2 tag, for $N_{jet} = 5$ the number of possible jet-parton assignments is then 6 and for $N_{jet} = 6$ the number of possible jet-parton assignments is 24.

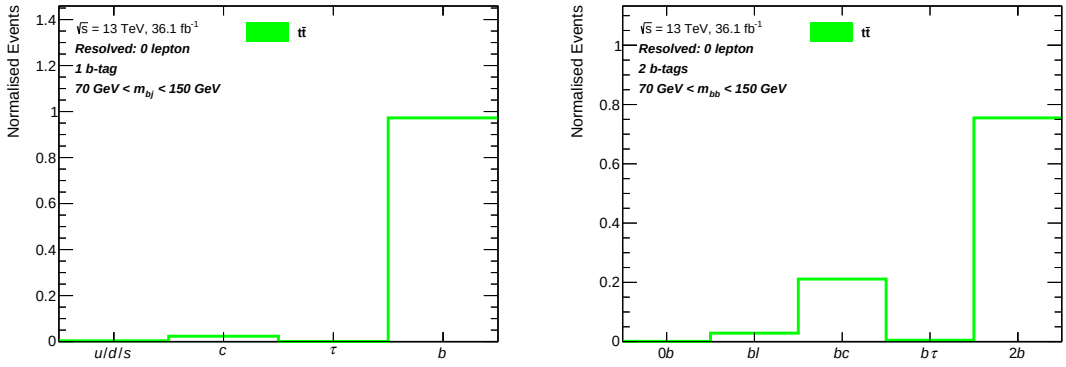

Figure 12.1: Left: B-jet truth flavour distribution normalised to unit area in the the SR, 1 tag, resolved regime. Right: Truth flavour distribution of two b-jets normalised to unit area in the SR, 2 tag, resolved regime, within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$ for $t\bar{t}$ events. For about 75% (95%) of $t\bar{t}$ events in 2 tag (1 tag), SR, the b-tagged jets originate from a b-quark.

For a given jet-parton assignment, the invariant mass of the W bosons and top quarks can be reconstructed using the four-momenta of jets assigned to the final state particles, as follows:

- ***hadronic*** W **mass:** $m^2_{W_{had}} = (p_{q'} + p_{\bar{q}})^2$, where $p_{q'}$ and $p_{\bar{q}}$ denote the four-momentum of the jet assigned to the q' and to the $\bar{q}$ partons, respectively, from the $W \to q'\bar{q}$ decay.

- ***hadronic top*** **mass:** $m^2_{t_{had}} = (p_{q'} + p_{\bar{q}} + p_{b_{had}})^2$, where $p_{b_{had}}$ denotes the four-momentum of the b-jet assigned to the b-quark from the hadronic top decay.

- ***leptonic*** W **mass:** The z-component of the missing momentum, assumed to be the neutrino momentum, is unknown. $m^2_{W_{lep}} = (p_\tau + p_{\nu_\tau})^2 \equiv m^2_W$, where p_τ and p_{ν_τ} denote the four-momenta of the tau and neutrino, respectively, from the $W \to \tau\nu_\tau$ decay, with $p_{\nu_\tau,x} = E^{miss}_x$, $p_{\nu_\tau,y} = E^{miss}_y$, is a quadratic equation in $p_{\nu_\tau,z}$ used for

estimating the mass of the *leptonic top*. m_W is the mass of the W boson from direct measurements [135]. For more details on $p_{\nu_\tau,z}$, see Appendix B.1.

- ***leptonic top* mass:** $m^2_{t_{lep}} = (p_\tau + p_{\nu_\tau} + p_{b_{lep}})^2$, where $p_{b_{lep}}$ denotes the four-momentum of the b-jet assigned to the b-quark from the leptonic top decay.

Using topological information, an *ordering rule* is introduced to rank the jet-parton assignments. This is done based on a *Likelihood* function, after which the algorithm is named. The likelihood for a given jet-parton assignment to correspond to a semileptonic $t\bar{t}$ event is given by

$$\begin{aligned} L = & f_{\mathrm{BW}}(m_{W_{had}}; m_W, \Gamma_W) \cdot f_{\mathrm{BW}}(m_{t_{had}}; m_t, \Gamma_t) \cdot f_{\mathrm{BW}}(m_{W_{lep}}; m_W, \Gamma_W) \cdot f_{\mathrm{BW}}(m_{t_{lep}}; m_t, \Gamma_t) \\ & \cdot \mathrm{Gauss}(m_{t_{had}} - m_{W_{had}}; \Delta m_0, \sigma_{\Delta_{had}}) \, \mathrm{Gauss}(m_{t_{lep}} - m_{W_{lep}}; \Delta m_0, \sigma_{\Delta_{lep}}). \end{aligned} \tag{12.2}$$

The terms of the likelihood function are as follows. The invariant mass m of an unstable massive particle (resonance) follows a relativistic Breit-Wigner distribution f_{BW} of the form

$$f_{\mathrm{BW}}(m; m_0, \Gamma_0) \equiv \frac{1}{(m^2 - m_0^2)^2 + m_0^2 \Gamma_0^2}, \tag{12.3}$$

where m_0 stands for the rest mass and Γ_0 stands for the resonance width of the particle. Therefore, $m_{t_{had}}$, $m_{t_{lep}}$, $m_{W_{had}}$ and $m_{W_{lep}}$ are distributed as f_{BW}. Two constraints on the difference of the top quark mass and the W boson mass are introduced following a Gaussian distribution of mean $\Delta m_0 = m_t - m_W$ and standard deviation σ_Δ, where m_t and m_W stand for the mass of the top quark and the W boson, respectively, and σ_Δ is an empirical value. The mass values are taken as $m_t = 172.5$ GeV [136] and $m_W = 80.4$ GeV [135]. The decay width values are taken as $\Gamma_t = 1.5$ GeV [136] for the top quark and $\Gamma_W = 2.1$ GeV [135] for the W boson. Finally, $\sigma_{\Delta_{had}} = 9$ GeV and $\sigma_{\Delta_{lep}} = 18$ GeV.

The *Likelihood Ordering rule* states that the jet-parton assignment of all possible combinations that maximises the likelihood function L is chosen as the semileptonic $t\bar{t}$ candidate. From the chosen jet-parton assignment, one defines the *hadronic top* and *hadronic W candidates* in the event, used in the following for discriminating between $t\bar{t}$ and signal events. The topological information from the *leptonic top* decay is used to assess the compatibility of the event topology with semileptonic $t\bar{t}$ production. For events with $N_{jet} = 4$, no topological information from the *leptonic top* decay is used in the likelihood function. The likelihood function for events with $N_{jet} = 4$ is found in Appendix B.2. Likelihood Ordering makes sense only for events having at least four central small-R jets, $N_{jet} \geq 4$.

In Figure 12.2 (Figure 12.3), the distribution of the *hadronic W* (*hadronic top*) candidate mass normalised to unit area is shown in the SR, resolved regime, 1 tag in the left panel and 2 tag in the right panel. The $t\bar{t}$ distribution peaks at m_W (m_t) and *rapidly* decreases as $|m_{W_{had}} - m_W|$ ($|m_{t_{had}} - m_t|$) increases, reconstructing correctly the decaying particle mass. The representative signal model distributions are wide and do not have a pronounced peak compared to $t\bar{t}$.

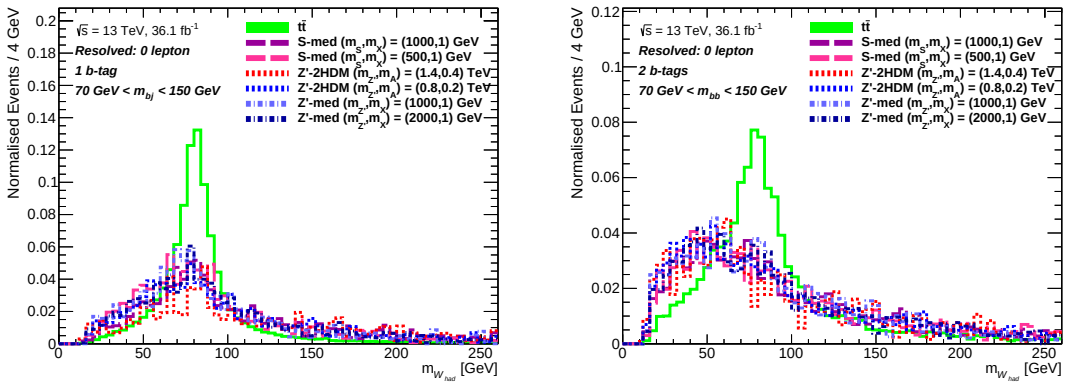

Figure 12.2: *Hadronic W* candidate mass distribution normalised to unit area in the SR, 1 tag (left) and 2 tag (right), resolved regime, within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. Unlike the representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator models, the $t\bar{t}$ distribution peaks at m_W and rapidly decreases as $|m_{W_{had}} - m_W|$ increases.

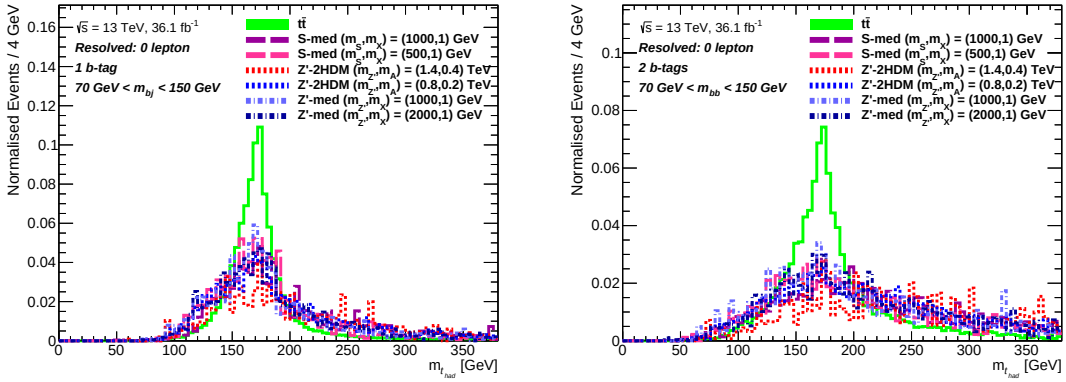

Figure 12.3: *Hadronic top* candidate mass distribution normalised to unit area in the SR, 1 tag (left) and 2 tag (right), resolved regime, within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. Unlike the representative $Z^{'} - 2HDM$, $Z^{'}$- and S-mediator models, the $t\bar{t}$ distribution peaks at m_t and rapidly decreases as $|m_{t_{had}} - m_t|$ increases.

12.2 Elliptic Cut

The mass of the *hadronic W* candidate and of the *hadronic top* candidate are used to discriminate between $t\bar{t}$ and signal events. In Figure 12.4, the distribution of $t\bar{t}$ events as a function of $(m_{t_{had}}, m_{W_{had}})$ normalised to unit area is shown in the SR, 1 tag (left) and 2 tag (right), resolved regime. The event density peaks around (m_t, m_W) as expected. It rapidly decreases for mass points with increasing distance to (m_t, m_W), having approximate elliptic contour curves of equal event density. Therefore, an elliptic cut on $(m_{t_{had}}, m_{W_{had}})$ can significantly reject the $t\bar{t}$ background process. Such elliptic cut has a minor impact on the signal efficiency, because the event density in the $(m_{t_{had}}, m_{W_{had}})$ plane for signal models does not have a pronounced peak around (m_t, m_W) and is wide and signal events accumulate at jet multiplicities lower than four, as shown in Appendix B.3.

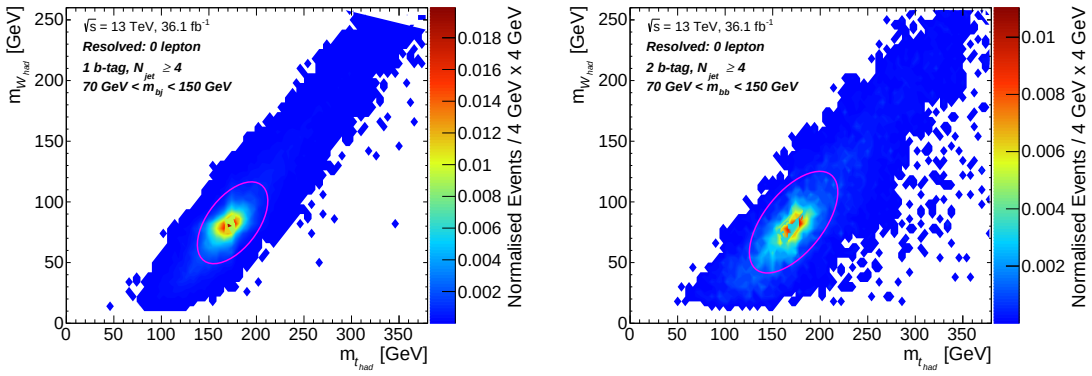

Figure 12.4: Distribution of $t\bar{t}$ events as a function of $(m_{t_{had}}, m_{W_{had}})$ normalised to unit area in the SR, 1 tag (left) and 2 tag (right), resolved regime within the h candidate mass window $(70\,\text{GeV}, 150\,\text{GeV})$. Two ellipses in pink are included.

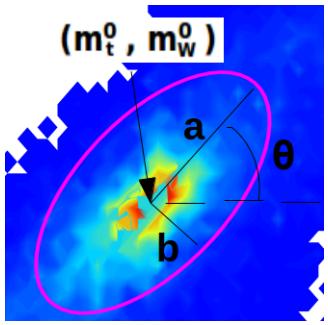

Figure 12.5: Five ellipse parameters to be optimised.

An ellipse is defined by five parameters which can be optimised to define the actual elliptic cut. As illustrated in Figure 12.5, the parameters are the centre of the ellipse (m_t^0, m_W^0), its rotation angle θ, and its semi-major and semi-minor axis lengths a and b, respectively. The final parameter values are iteratively optimised, with each iteration consisting of three steps. The chosen values at each step serve as the initial values for the optimisation of the following step. The average considered in the ellipse optimisation, denoted as $\langle ... \rangle$, is over the representative signal models considered in Figures 12.2, 12.3, to which much of the sensitivity lies in the resolved regime, where the Likelihood Ordering is applied. The steps at each iteration are as follows:

- Step 1: It is a scan on θ, with the centre and semi-major and semi-minor length parameters being fixed. The figure of merit is $\langle \varepsilon_s \rangle / \varepsilon_{t\bar{t}}$, which is maximised by the chosen θ value in this step, where ε_s and $\varepsilon_{t\bar{t}}$ stand for the signal and $t\bar{t}$ efficiency, respectively.
- Step 2: It is a scan on the centre (m_t^0, m_W^0), with the other parameters being fixed. The figure of merit is the averaged significance gain $\langle \Delta Z_0 \rangle$, where Z_0 is defined in Equation 7.1. The chosen ellipse centre maximises the averaged significance gain under the chosen constraint of $1 - \langle \varepsilon_s \rangle \leq 8\%$ (12%)[1] in 2 tag (1 tag).
- Step 3: It is a scan on the semi-axis lengths a, b, with the other parameters being fixed. The figure of merit is the averaged significance gain $\langle \Delta Z_0 \rangle$ again, which is maximised by the chosen semi-axis lengths under the chosen constraint of $1 - \langle \varepsilon_s \rangle \leq 8\%$ (12%) in 2 tag (1 tag).

[1] The choice of values is an arbitrary choice, for which the $t\bar{t}$ rejection in 1 tag and 2 tag is similar.

The optimised ellipse parameter values are: $(m_t^0, m_W^0) = (172\ \text{GeV}, 83\ \text{GeV})$, $a = 56$ GeV, $b = 28$ GeV and $\theta = 40^\circ$ in 2 tag, while $(m_t^0, m_W^0) = (175\ \text{GeV}, 83\ \text{GeV})$, $a = 44$ GeV, $b = 24$ GeV and $\theta = 40^\circ$ in 1 tag. These ellipses are shown in Figure 12.4. All events within the ellipse are rejected. The optimisation scans at the last iteration are given in Appendix B.4.

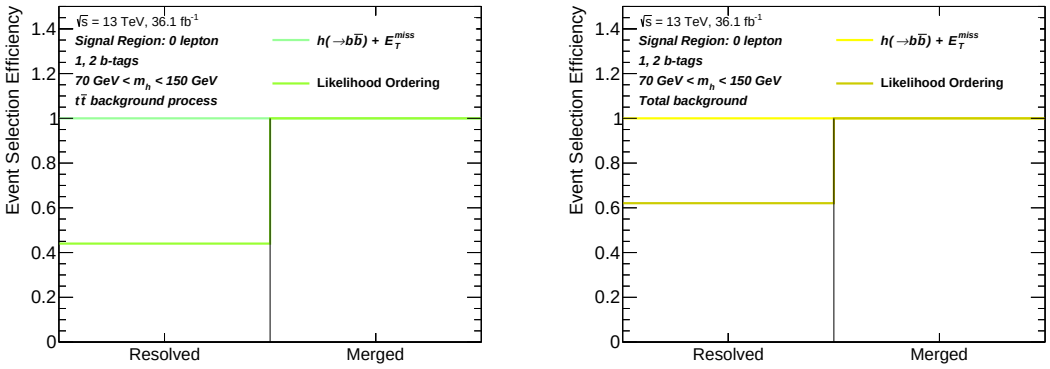

Figure 12.6: Likelihood Ordering event selection efficiency for the $t\bar{t}$ background process (left) and the total background (right) in the SR within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$.

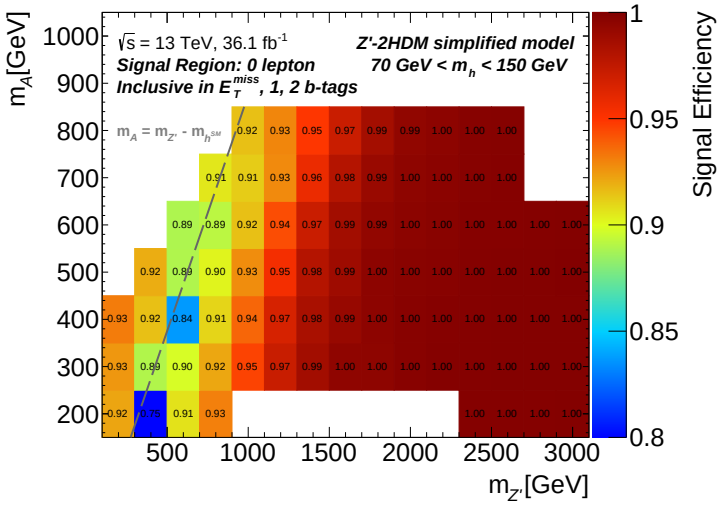

Figure 12.7: Likelihood Ordering event selection efficiency for $Z' - 2HDM$ signal models in the SR, inclusive in E_T^{miss} within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass.

In Figure 12.6, the efficiency of the Likelihood Ordering event selection is shown for the $t\bar{t}$ background process (left) and for the total background (right) in the SR. The $t\bar{t}$ background process is reduced by about 55% while the total background is reduced by about 40% in the resolved regime.

In Figure 12.7, the signal efficiency for $Z' - 2HDM$ models is shown in the SR, inclusive in E_T^{miss}. Signal models, where the sensitivity is driven by the merged regime, have about 100% efficiency. Most signal models, whose sensitivity is dominated by the resolved regime, have efficiency above 90%. The experimental significance gain of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis after including the Likelihood Ordering event selection is shown in Section 12.4.

12.3 Validation of the Likelihood Ordering Algorithm in Data

In the following, the behaviour of the Likelihood Ordering algorithm in data is examined by comparing data to MC background in a region enriched with $t\bar{t}$ events, referred to as $t\bar{t}$ *Control Region* (CR). The $t\bar{t}$ CR is defined based on the 1μ CR by additionally inverting the ΔR_{bb} event selection to $\Delta R(\vec{p}^{\,j_1}, \vec{p}^{\,j_2}) \geq 1.8$ and requiring that events have at least four central jets, $N_{jets} \geq 4$. The $t\bar{t}$ background process amounts to about 92% (75%) of the total background at prefit level in the $t\bar{t}$ CR, 2 tag (1 tag), resolved regime.

The approach of the Likelihood Ordering algorithm is the same in the 1μ CR and SR, except for the combinatorics considered. In the 1μ CR, the tau-jet assignment is not part of the combinatorics, as the muon is well defined, playing the role of the tau in the SR, $W \rightarrow \mu\nu_\mu$.

In Figure 12.8 (Figure 12.9), the *hadronic W* (*hadronic top*) candidate mass distribution is shown in the $t\bar{t}$ CR, resolved regime, 1 tag in left panel and 2 tag in right panel. At prefit level the agreement of data to MC background is good. The Likelihood Ordering algorithm reconstructs well $m_{W_{had}}$ ($m_{t_{had}}$) in data as well as it does in MC.

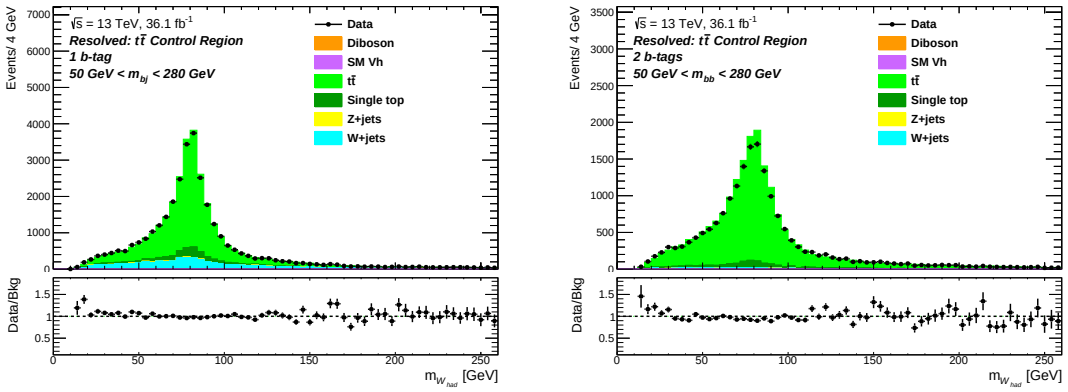

Figure 12.8: *Hadronic W* candidate mass distribution in the $t\bar{t}$ CR, 1 tag (left) and 2 tag (right), resolved regime, within the h candidate mass window $(50\,\text{GeV}, 280\,\text{GeV})$. No systematic uncertainty band is shown.

The performance of the Likelihood Ordering algorithm in data is well modelled in MC, validating the use of this algorithm.

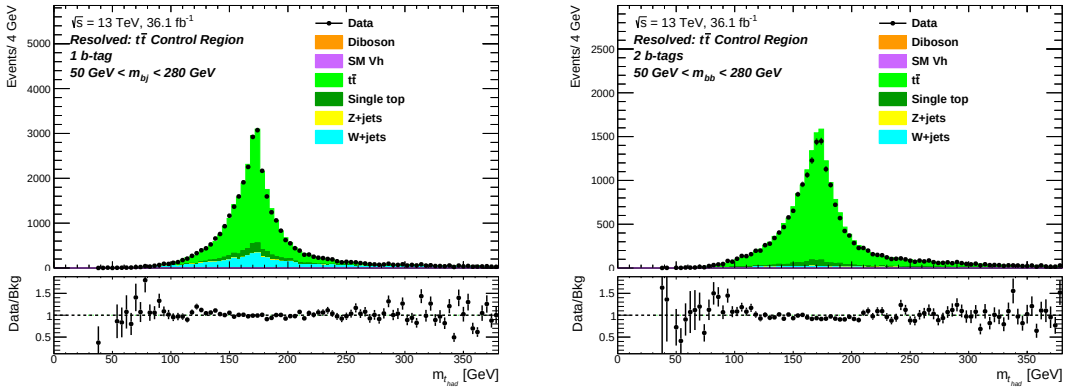

Figure 12.9: *Hadronic top* candidate mass distribution in the $t\bar{t}$ CR, 1 tag (left) and 2 tag (right), resolved regime, within the h candidate mass window $(50\,\text{GeV}, 280\,\text{GeV})$. No systematic uncertainty band is shown.

12.4 Experimental Sensitivity Gain

The Likelihood Ordering algorithm is included in all the channels of the analysis, performing then the entire analysis again to examine the increase in the experimental sensitivity. The approach of the Likelihood Ordering is the same in the SR, 1μ CR and 2ℓ CR, up to a few differences. As mentioned above, no tau-jet assignment enters the possible combinations in the 1μ CR. In the 2ℓ CR, the same possible combinations are considered as in the SR and the z-component of $\vec{E}^{miss}$ is known a priori from $\vec{E}^{miss} \equiv \vec{p}^{\ell\ell}$.

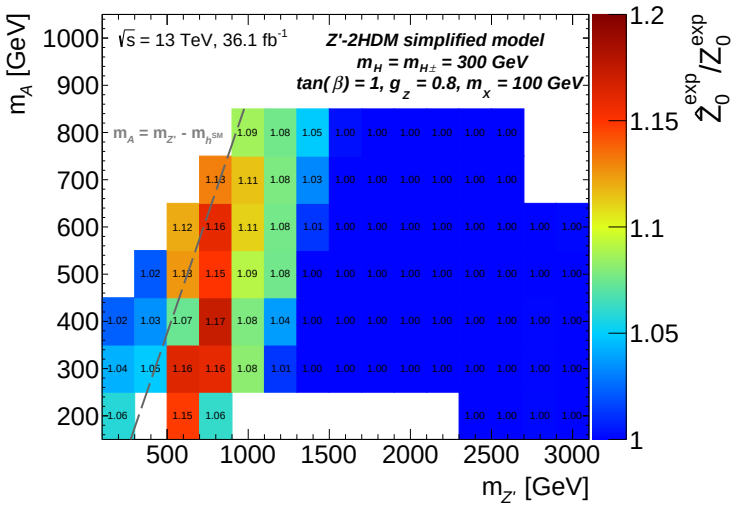

Figure 12.10: Ratio of the expected significance of the $h(\to b\bar{b}) + E_T^{miss}$ analysis including the Likelihood Ordering event selection to that of the $h(\to b\bar{b}) + E_T^{miss}$ analysis, $\hat{Z}_0^{exp}/Z_0^{exp}$, for the $Z' - 2HDM$ model. The transition between off-shell and on-shell regions at $m_A = m_{Z'} - m_{h^{SM}}$ shown as a grey line, where $m_{h^{SM}} = 125$ GeV is the SM-like Higgs mass.

In Figure 12.10, the ratio of the expected significance of the $h(\to b\bar{b}) + E_T^{miss}$ analysis including the Likelihood Ordering event selection to the expected significance of the $h(\to$

$b\bar{b}) + E_T^{miss}$ analysis is shown for the $Z' - 2HDM$ model[2]. The Likelihood Ordering event selection increases the discovery expected significance up to 17% for $Z' - 2HDM$ models, where the sensitivity is driven by the resolved regime. No significance gain is observed for signal models with sensitivity driven by the merged regime, as expected.

Based on these results, it is concluded that the Likelihood Ordering algorithm is a new approach that considerably decreases the semileptonic $t\bar{t}$ background process and increases the experimental sensitivity of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis. This is a promising algorithm for future $h(\rightarrow b\bar{b}) + E_T^{miss}$ analyses and any analysis dealing with semileptonic $t\bar{t}$.

[2]No significance ratio entry is included at $(m_{Z'}, m_A) = (400, 200)$ GeV because the expected significances are null for this $Z' - 2HDM$ mass point.

13 Summary

Many beyond Standard Model theories predict associated production of Dark Matter particles with a Higgs boson at energies that can be probed at collider experiments. Previous to the present analysis, such processes were searched for in final states with missing transverse momentum and a Higgs boson decaying to a $b\bar{b}$ pair with the ATLAS detector using 20.3 fb^{-1} and 3.2 fb^{-1} of proton-proton collisions recorded at a centre-of-mass energy of 8 TeV [137] in 2012 and 13 TeV [16] in 2015, respectively. In these searches, the reducible $t\bar{t}$ background is the dominant background process. In the latter search, the $t\bar{t}$ background represents about 75% of the predicted Standard Model background in the signal region, which restricts the experimental sensitivity of the search.

In this thesis, a search for Dark Matter, DM, produced in association with a Higgs boson, h, decaying to a $b\bar{b}$ pair is performed using proton-proton collisions at a centre-of-mass energy of 13 TeV with an integrated luminosity of 36.1 fb^{-1} recorded with the ATLAS detector [15]. New refined event selections were designed and implemented aiming at the reduction of the $t\bar{t}$ background process and the increase in the experimental sensitivity of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ search. Such refined event selections are the *veto of additional b-jets*; the *veto of standard taus* based on a boosted-decision-tree tau identification, and the *veto of extended taus* based on a new custom-built tau identification; the new H_T *ratio* requirement on the minimal H_T[1] *fraction* carried by the $h \rightarrow b\bar{b}$ decay; and the ΔR_{bb} requirement, which exploits the angular separation of the $b\bar{b}$ pair from the $h \rightarrow b\bar{b}$ decay. These *optimised event selections* reduce the $t\bar{t}$ background by about 50% and 70% in the signal region, resolved and merged regimes, respectively, so that the irreducible $Z + jets$ background becomes the dominant background process. The experimental sensitivity of the search to the $Z' - 2HDM$ model is substantially increased up to about 60% by the optimised event selections.

The observed data are found to be in agreement with the Standard Model predictions. Upper limits on the production cross-section of $h + DM$ events, σ_{h+DM}, times the $h \rightarrow b\bar{b}$ branching ratio, $\mathscr{B}(h \rightarrow b\bar{b})$[2], are derived at 95% confidence level for the $Z' - 2HDM$ model. In the $Z' - 2HDM$ scenario with fixed parameters $\tan(\beta) = 1$, $g_{Z'} = 0.8$, $m_\chi =$ 100 GeV and $m_H = m_{H^\pm} = 300$ GeV, $Z' - 2HDM$ models are excluded at 95% confidence level for $m_{Z'}$ up to 2.6 TeV and m_A up to 0.6 TeV, substantially extending previous limits[3]. The upper limits on $\sigma_{h+DM} \times \mathscr{B}(h \rightarrow b\bar{b})$ in the $Z' - 2HDM$ scenario are improved by more than an uncertainty band on the expected limits through the optimised event selections for $m_{Z'}$ up to ~ 1.3 TeV and m_A up to ~ 0.55 TeV.

After applying the optimised event selections, the background from $t\bar{t}$ production is

[1] H_T is the sum of scalar p_T of jets in the event.

[2] The $h \rightarrow b\bar{b}$ branching ratio is taken as $\mathscr{B}(h \rightarrow b\bar{b}) = 0.571$.

[3] Upper limits are derived for other fixed parameter values by scaling the presented cross-section limits.

still comparable to the $Z+jets$ background process in the resolved regime for events with $E_T^{miss} \leq 350$ GeV. A new algorithm, called *Likelihood Ordering*, is designed and tested in the analysis, reducing the $t\bar{t}$ background by 55% in the signal region, resolved regime. The experimental sensitivity of the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis to the $Z' - 2HDM$ model is further increased up to 17% through the Likelihood Ordering algorithm.

A Postfit Results

A.1 Additional Nuisance Parameter Constraints

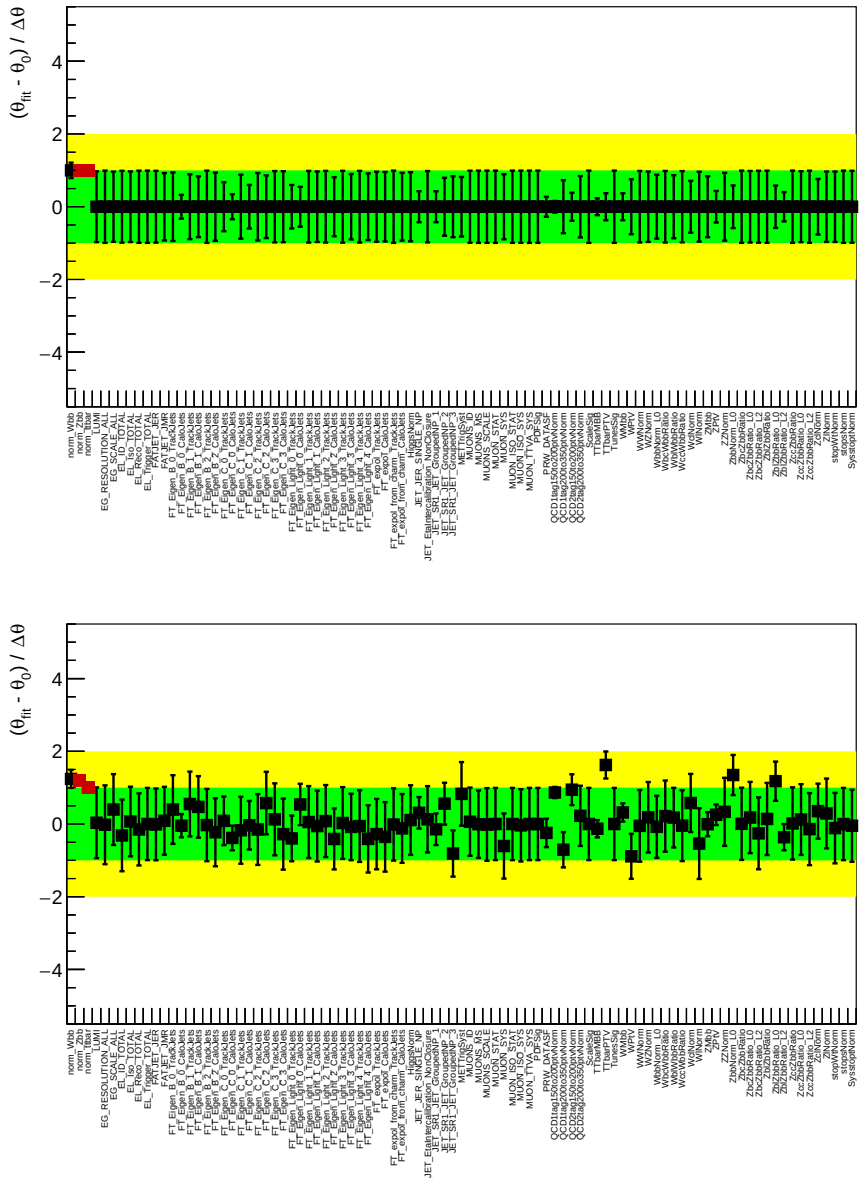

Figure A.1: The nuisance parameter pulls upon fitting to the Asimov data (top) and to the observed data (bottom) for the $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis without optimised event selections.

In Figure A.1, the nuisance parameter (NP) pulls upon fitting to the Asimov data (top) and to the observed data (bottom) for the $h(\to b\bar{b}) + E_T^{miss}$ analysis without optimised event selections. The nomenclature of the NPs is found in Chapter 10. It is observed that the NP pull behaviour is very similar to that for the $h(\to b\bar{b}) + E_T^{miss}$ search, as seen in Figures 11.1, 11.2.

A.2 Correlation of Nuisance Parameters

In Figure A.2, the correlation of NPs is shown for the statistical fit to the observed data under the background-only hypothesis. The nomenclature of the NPs is found in Chapter 10.

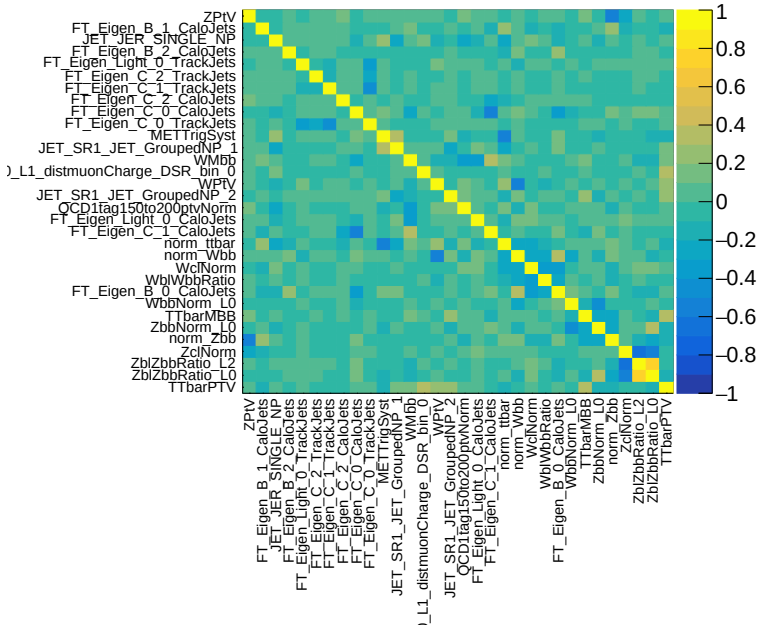

Figure A.2: Nuisance parameter correlations for the statistical fit to the observed data under the background-only hypothesis.

A.3 Postfit Distributions

In Figure A.3, the postfit h candidate invariant mass distribution is shown in the SR, 1 tag, for the four E_T^{miss} categories. The prefit SM background expectation is shown as a dashed blue line. A representative $Z' - 2HDM$ model, with cross-section normalised to 10 fb and parameters $(m_{Z'}, m_A) = (1.4\text{ TeV}, 0.4\text{ TeV})$ is shown as a red curve. The agreement of data to background is good.

In Figure A.4, the postfit distribution of the muon charge is shown in the 1μ CR, 1 tag, for the four subregions in $E_{T,noMU}^{miss}$, within h candidate mass window $(50\text{ GeV}, 280\text{ GeV})$.

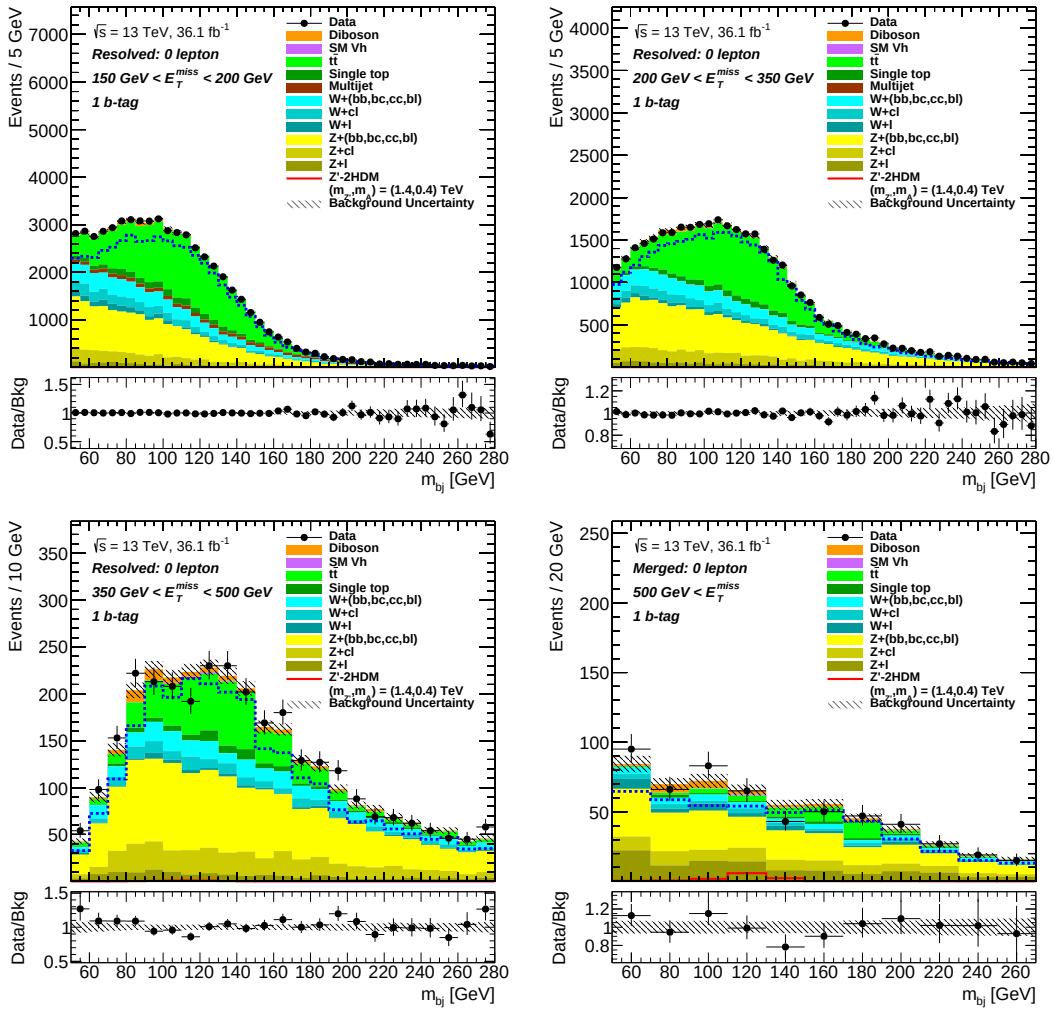

Figure A.3: Postfit distribution of the h candidate invariant mass in the SR, 1 tag, for the four E_T^{miss} categories. The SM background expectation before the fit (prefit distribution) is shown as a dashed blue line. A representative $Z' - 2HDM$ model, with cross-section normalised to 10 fb and parameters $(m_{Z'}, m_A) = (1.4\text{ TeV}, 0.4\text{ TeV})$ is shown as a red curve.

The agreement of data to background is good.

In Figure A.5, the postfit event yields is shown in the 2ℓ CR, 1 tag, for the four subregions in p_T^V ($p_T^{\ell\ell}$), within h candidate mass window $(50\text{ GeV}, 280\text{ GeV})$. The agreement of data to background is good.

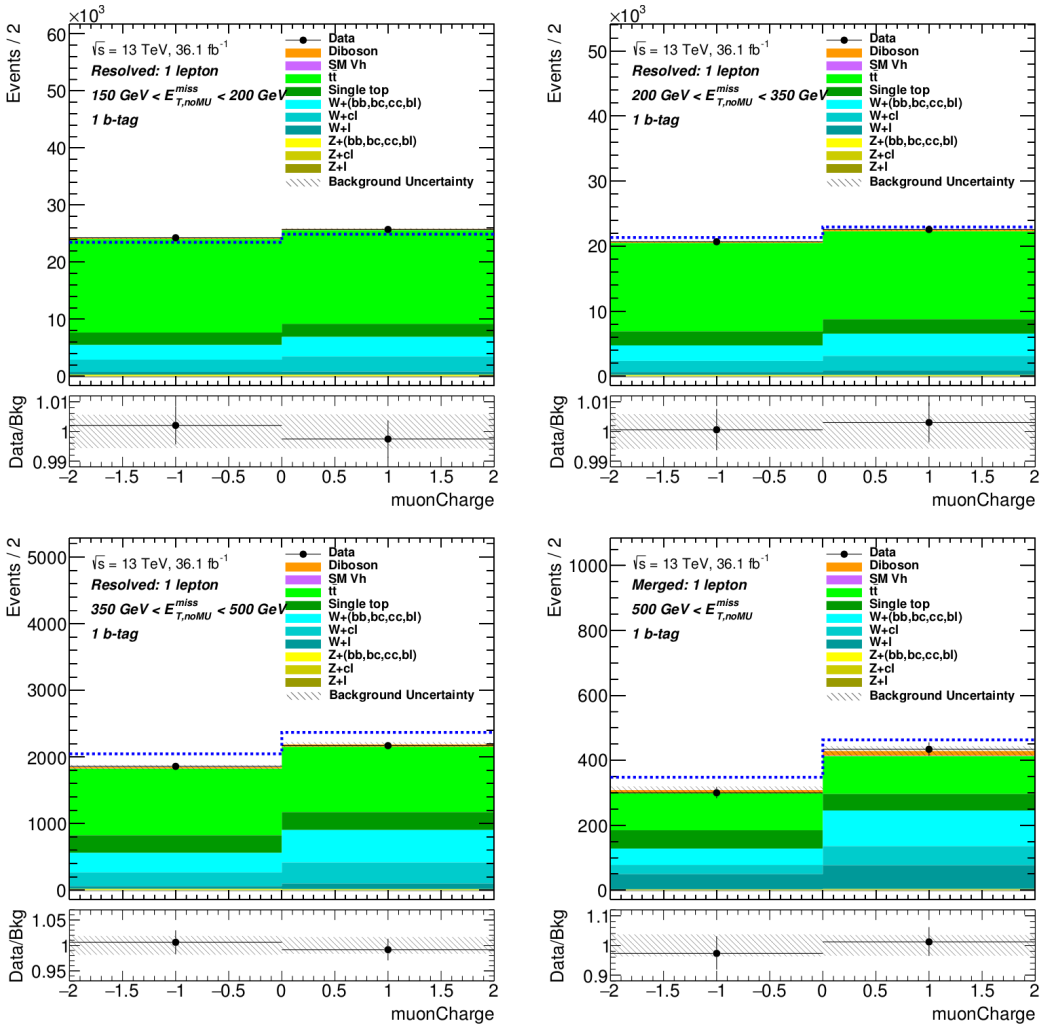

Figure A.4: Postfit distributions of the muon charge in the 1μ CR, 1 tag, for the four subregions in $E^{miss}_{T,noMU}$, within h candidate mass window $(50\,\text{GeV}, 280\,\text{GeV})$.

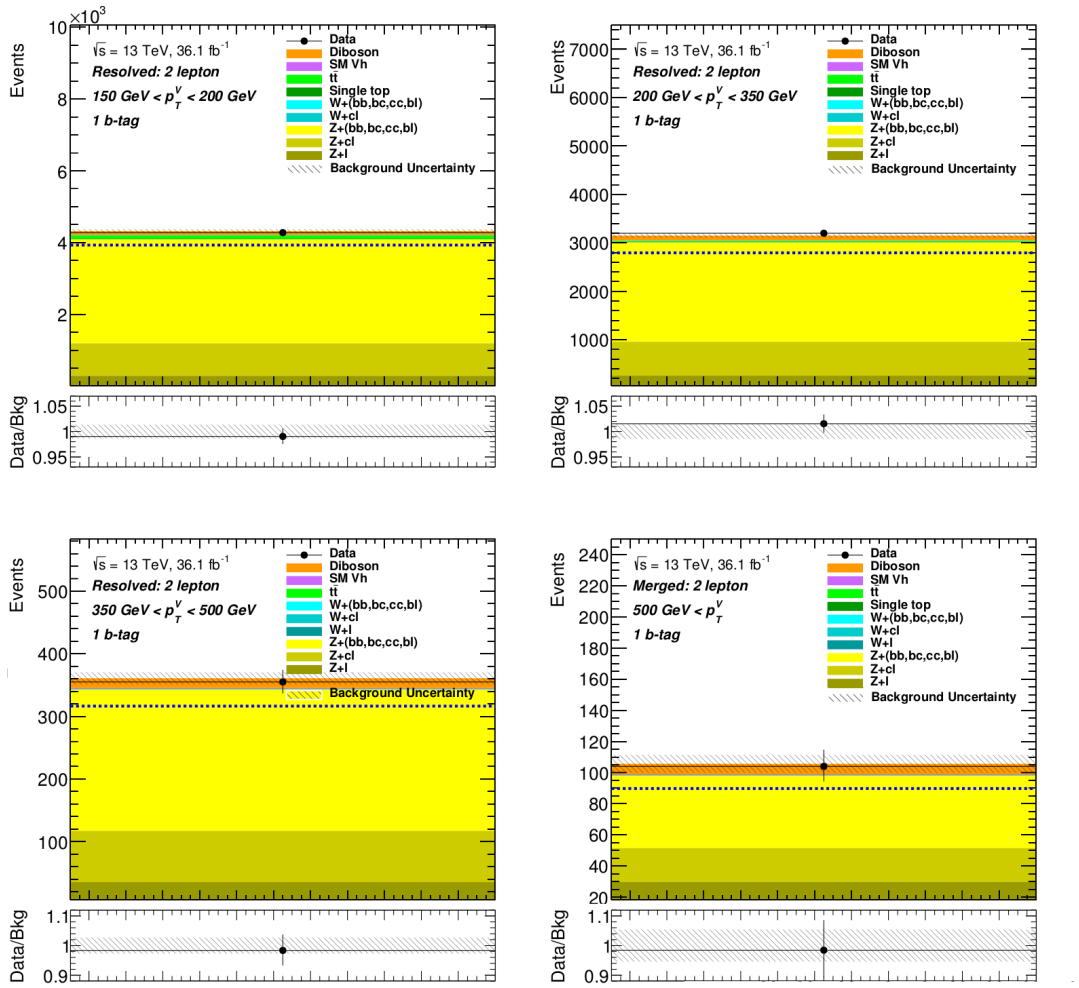

Figure A.5: Postfit event yields in the 2ℓ CR, 1 tag, for the four subregions in p_T^V ($p_T^{\ell\ell}$), within h candidate mass window $(50\,\text{GeV}, 280\,\text{GeV})$.

B Likelihood Ordering Algorithm

B.1 Reconstruction of the z-Component of the Neutrino Momentum

The z-component of the neutrino momentum is used for the four-momenta of the *leptonic top* and *leptonic W* candidates in order to use a constraint on the mass difference between both candidates and include the mass decay Breit-Wigner distribution for the *leptonic top* in the likelihood function in Equation 12.2. One derives the z-component of the neutrino from the decay $W \to \ell\nu$, where the W boson corresponds to the so-called *leptonic W* and ℓ corresponds to a tau in the SR or to a muon in the 1μ CR. Thus,

$$m_W^2 = (p_\ell + p_\nu)^2 = 2(E_\ell E_\nu - \vec{p}_\ell \cdot \vec{p}_\nu), \tag{B.1}$$

assuming the leptons to be massless. Using the identity $p_{\nu,x}^2 + p_{\nu,y}^2 + p_{\nu,z}^2 = |\vec{p}_\nu|^2 = E_\nu$ and Equation B.1, one obtains

$$\begin{aligned} p_{\nu,x}^2 + p_{\nu,y}^2 + p_{\nu,z}^2 =& \frac{1}{E_\ell^2}\left(p_{\nu,x}p_{\ell,x} + p_{\nu,y}p_{\ell,y} + p_{\nu,z}p_{\ell,z} + \frac{m_W^2}{2}\right)^2 \\ =& \frac{1}{E_\ell^2}\left(A^2 + 2Ap_{\nu,z}p_{\ell,z} + p_{\nu,z}^2 p_{\ell,z}^2\right), \end{aligned} \tag{B.2}$$

where $A \equiv p_{\nu,x}p_{\ell,x} + p_{\nu,y}p_{\ell,y} + m_W^2/2$. Equation B.2 is a quadratic equation in $p_{\nu,z}$. Using $p_{\ell,t}^2 \equiv E_\ell^2 - p_{\ell,z}^2$, one solves for $p_{\nu,z}$ as

$$p_{\nu,z} = \frac{2Ap_{\ell,z} \pm \sqrt{4A^2p_{\ell,z}^2 - 4p_{\ell,t}^2[(p_{\nu,x}^2 + p_{\nu,y}^2)E_\ell^2 - a^2]}}{2p_{\ell,t}^2}. \tag{B.3}$$

If the square root in Equation B.3 is negative, $p_{\nu,z} = Ap_{\ell,z}/p_{\ell,t}^2$ is used as proxy for the z-component of the neutrino momentum. If the square root is positive, both $p_{\nu,z}$ solutions are used as proxy for the z-component of the neutrino momentum, considering therefore additional combinatorics. The chosen jet-parton assignment corresponds to the one maximising the likelihood function, as already stated.

B.2 Likelihood Function for $N_{jet} = 4$

In case of events with only four central jets, the likelihood function used to rank the jet-parton assignments takes a different form than that in Equation 12.2. With only four

central jets, no information on the *leptonic top* decay nor on the *leptonic W* decay is used. The combinatorics consider only the *hadronic top* and *hadronic W* candidates.

For an event having four central jets, the likelihood for a given jet-parton assignment is given by

$$L = f_{\mathrm{BW}}(m_{W_{had}}; m_W, \Gamma_W) f_{\mathrm{BW}}(m_{t_{had}}; m_t, \Gamma_t) \,\mathrm{Gauss}(m_{t_{had}} - m_{W_{had}}; \Delta m_0, \sigma_{\Delta_{had}}), \qquad \text{(B.4)}$$

where the parameter values are the same than those used in the likelihood function for events with more than four central jets, as given in Chapter 12.

B.3 $(m_{t_{had}}, m_{W_{had}})$ Distribution for Signal Models

The Likelihood Ordering event selection has a minor impact on the signal yields, as it hardly decreases the signal efficiency, as seen in Figure 12.7 for $Z' - 2HDM$ models. One reason for that is that the central jet multiplicity distribution for signal models peak at jet multiplicities lower than four. As the Likelihood Ordering algorithm only is applied in events having at least four central jets, all signal events having less than four central jets are accepted, so the elliptic cut can by definition only act on a subset of signal events, as seen in Figure B.1, which is the central jet multiplicity distribution for representative signal models in the SR, 1 tag (left) and 2 tag (right), resolved regime.

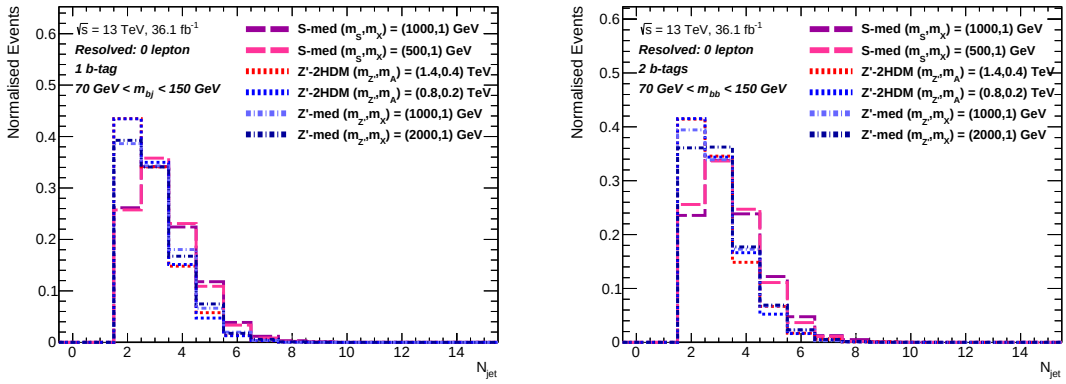

Figure B.1: Central jet multiplicity distribution for representative $Z' - 2HDM$, Z'- and S-mediator signal models in the SR, 1 tag (left) and 2 tag (right), resolved regime within the h candidate mass window $(70\,\mathrm{GeV}, 150\,\mathrm{GeV})$.

Another reason is that the mass observables constructed with the Likehood Ordering algorithm, namely $m_{W_{had}}$ and $m_{t_{had}}$, have a distribution that is wide and has no pronounced peaks for signal models, so the ellipse in $(m_{t_{had}}, m_{W_{had}})$ used for the event selection does not include a significant fraction of events. In Figure B.2 and Figure B.3, the distribution of representative signal model events as a function of $(m_{t_{had}}, m_{W_{had}})$ normalised to unit area are shown in the SR, resolved regime, 2 tag and 1 tag, respectively. The representative signal models include $Z' - 2HDM$ models, Z'- and S-mediator models.

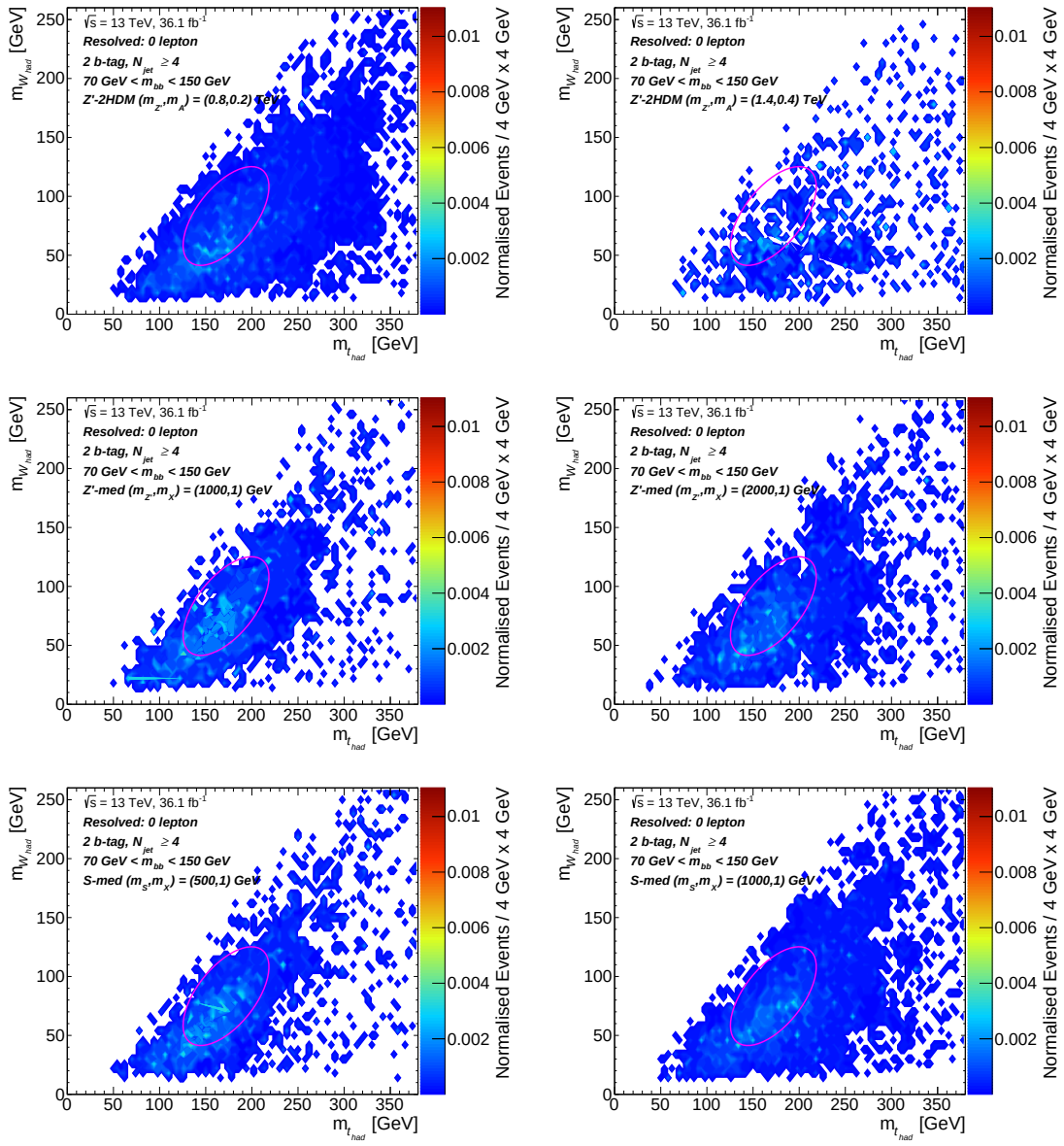

Figure B.2: Distribution of representative signal model events as a function of $(m_{t_{had}}, m_{W_{had}})$ in the SR, resolved regime, 2 tag within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The representative signal models include $Z' - 2HDM$ models, Z'- and S-mediator models.

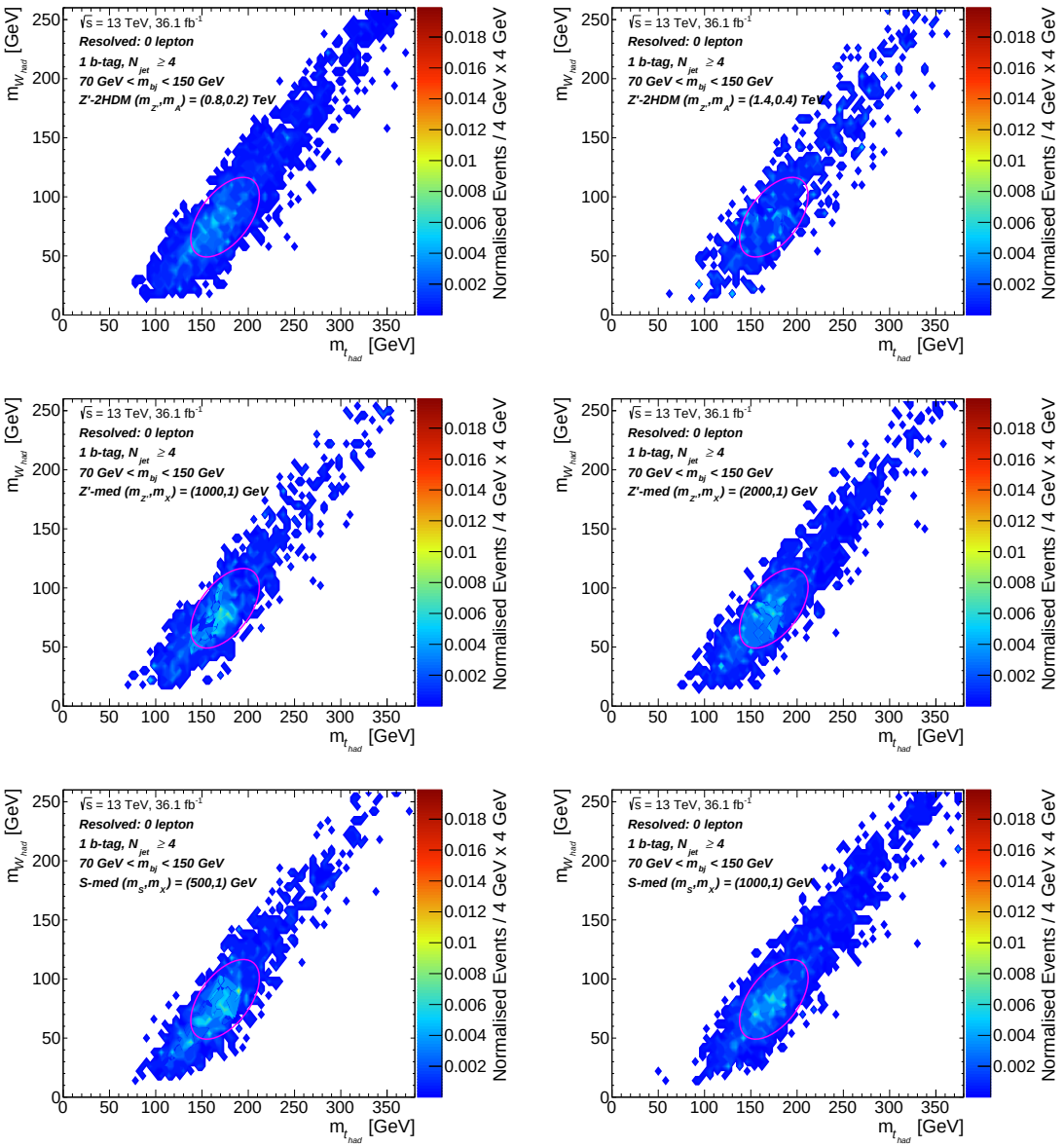

Figure B.3: Distribution of representative signal model events as a function of $(m_{t_{had}}, m_{W_{had}})$ in the SR, resolved regime, 1 tag within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. The representative signal models include $Z' - 2HDM$ models, Z'- and S-mediator models.

B.4 Optimisation Scans

In this Appendix Section, the optimisation scans on the five parameters of the ellipse used for the Likelihood Ordering event selection are shown for the last optimisation iteration, as described in Section 12.2.

In Figure B.4, the optimisation scan on the rotation angle θ, having all other parameters fixed, is shown in the SR, 1 tag (left) and 2 tag (right), resolved regime, within the h candidate mass window $(70\ \text{GeV}, 150\ \text{GeV})$. Representative $Z' - 2HDM$, Z'- and S-mediator signal models are included as well as the $t\bar{t}$ background process. The upper and lower plots correspond to the efficiency and to the ratio of the signal efficiency to the $t\bar{t}$ efficiency, respectively, as a function of the rotation angle θ. The optimised rotation angle θ is chosen to maximise $\langle\varepsilon_s\rangle/\varepsilon_{t\bar{t}}$, where ε_s and $\varepsilon_{t\bar{t}}$ correspond to the signal and $t\bar{t}$ efficiency, respectively, and the average is over the representative signal models. The optimised angle is $\theta = 40°$.

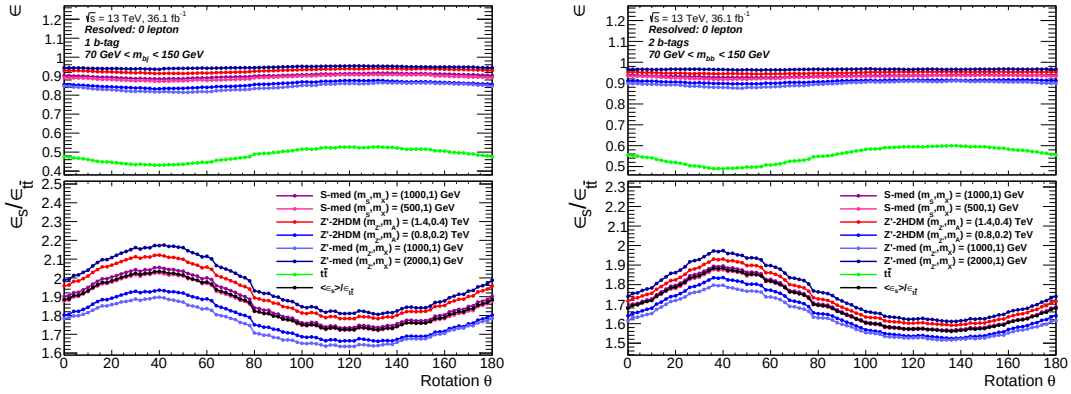

Figure B.4: Efficiency (top) and ratio of signal efficiency to $t\bar{t}$ efficiency (bottom) as a function of the rotation angle θ in the SR, 1 tag (left) and 2 tag (right), resolved regime for representative signal models and the $t\bar{t}$ background process. The optimised θ maximises $\langle\varepsilon_s\rangle/\varepsilon_{t\bar{t}}$, corresponding to $\theta = 40°$.

The following average is over the representative signal models mentioned above. In the following, only the scans for the SR, 2 tag are shown within the h candidate mass $(70\ \text{GeV}, 150\ \text{GeV})$.

The scan on the ellipse center (m_t^0, m_W^0), having all other parameters fixed, is chosen by maximising the averaged significance gain $\langle\hat{Z}_0/Z_0\rangle$ under the constraint $\langle\varepsilon_s\rangle \geq 0.92$. In Figure B.5, the scan on (m_t^0, m_W^0) is shown in the SR, 2 tag, resolved regime, where the translation along the major and minor axes is with respect to $(172\ \text{GeV}, 83\ \text{GeV})$. The chosen translation point is $(0\ \text{GeV}, 0\ \text{GeV})$, so the optimised ellipse center is $(172\ \text{GeV}, 83\ \text{GeV})$.

The scan on the semiaxis lengths a and b, having all other parameters fixed, is chosen by maximising the averaged significance gain $\langle\hat{Z}_0/Z_0\rangle$ under the constraint $\langle\varepsilon_s\rangle \geq 0.92$. In Figure B.6, the scan on (a, b) is shown in the SR, 2 tag, resolved regime, with a the semi-major axis length and b the semi-minor axis length. The chosen semiaxis lengths are $(a, b) = (56\ \text{GeV}, 28\ \text{GeV})$.

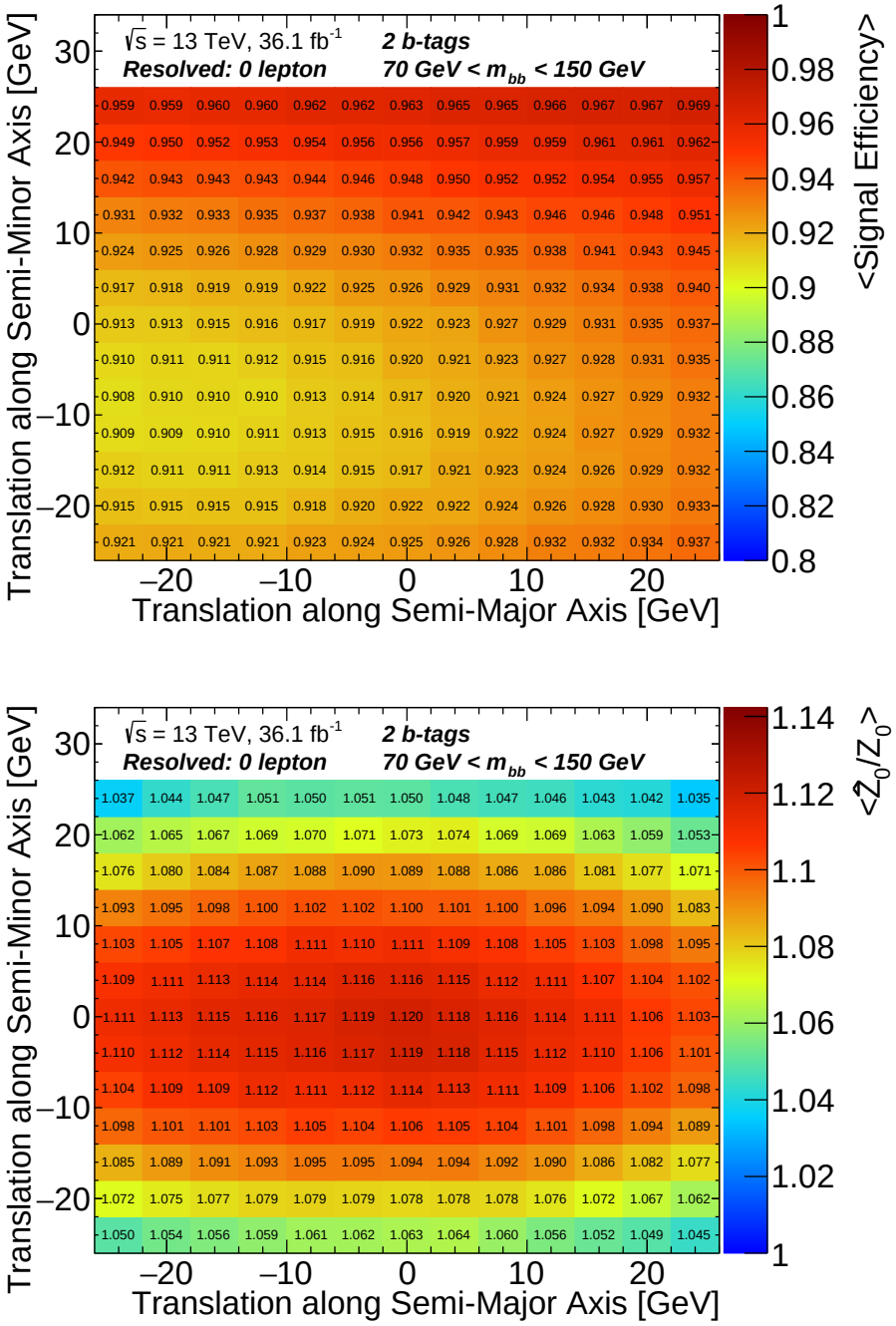

Figure B.5: Scan on the ellipse center (m_t^0, m_W^0), where the translations along the major and minor axes are with respect to $(172\,\text{GeV}, 83\,\text{GeV})$. The averaged signal efficiency (top) and the averaged significance gain (bottom) as a function of the mentioned translations.

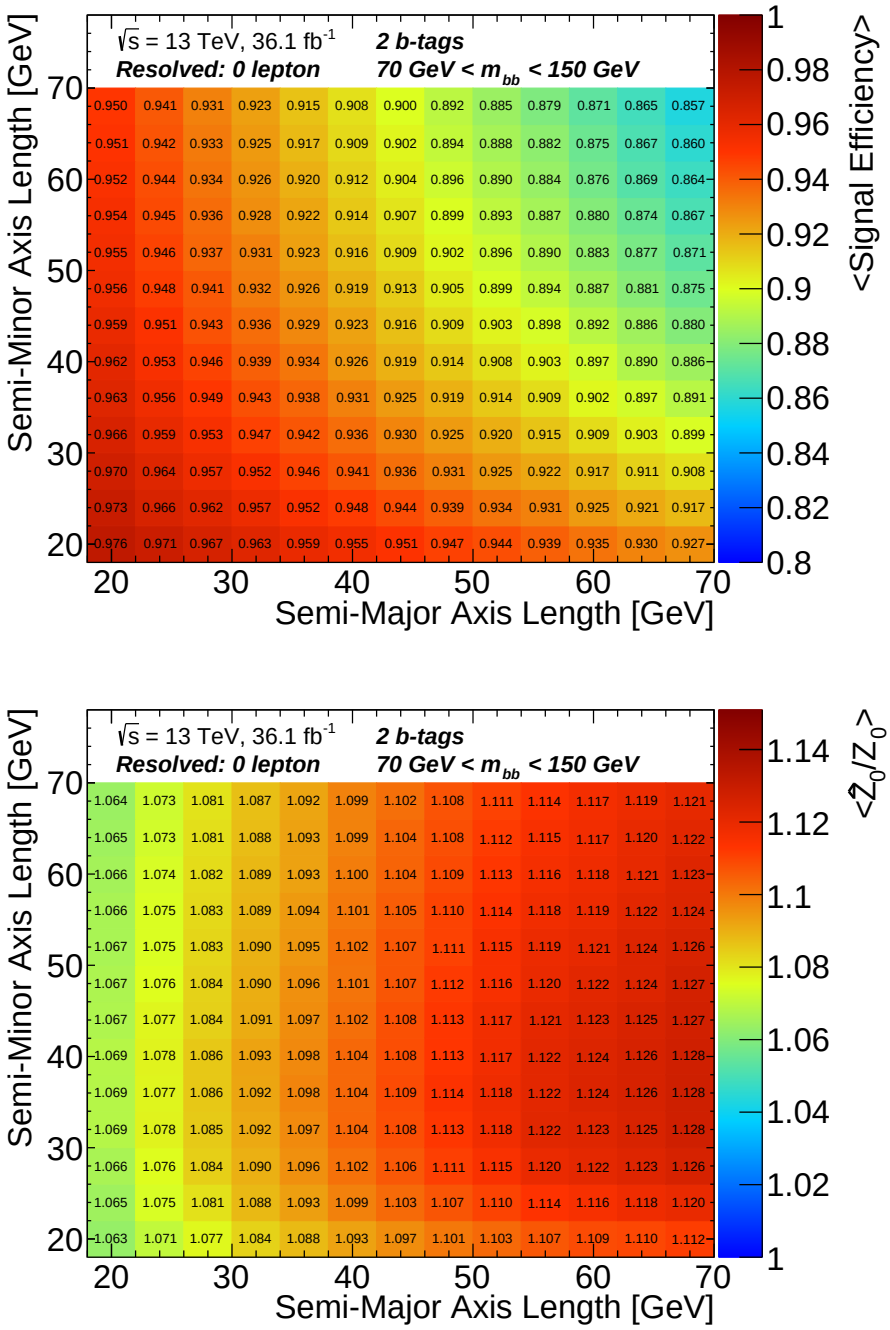

Figure B.6: Scan on semiaxis lengths. The averaged signal efficiency (top) and the averaged significance gain (bottom) as a function of (a, b), where a is the semi-major axis length and b the semi-minor axis length.

C 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**:1–29, 2012.

[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**:30, 2012.

[3] G. Bertone, D. Hooper, and J. Silk. *Particle dark matter: Evidence, candidates and constraints*. *Phys. Rept.*, **405**:279–390, 2005.

[4] R. J. Scherrer and M. S. Turner. *On the Relic, Cosmic Abundance of Stable Weakly Interacting Massive Particles*. *Phys. Rev. D*, **33**:1585, 1986.

[5] P. Gondolo and G. Gelmini. *Cosmic abundances of stable particles: Improved analysis*. *Nucl. Phys. B*, **360**:145–179, 1991.

[6] ATLAS Collaboration. *Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector*, ATLAS-CONF-2017-060. URL: `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2017-060/`, 2017.

[7] CMS Collaboration. *Search for dark matter, extra dimensions, and unparticles in monojet events in proton–proton collisions at* $\sqrt{s} = 8$ *TeV*. *Eur. Phys. J. C*, **75**:235, 2015.

[8] ATLAS Collaboration. *Search for dark matter in events with heavy quarks and missing transverse momentum in pp collisions with the ATLAS detector*. *Eur. Phys. J. C*, **75**:92, 2015.

[9] CMS Collaboration. *Search for Monotop Signatures in Proton-Proton Collisions at* $\sqrt{s} = 8$ *TeV*. *Phys. Rev. Lett.*, **114**:101801, 2015.

[10] ATLAS Collaboration. *Search for dark matter at* $\sqrt{s} = 13$ *TeV in final states containing an energetic photon an large missing transverse momentum with the ATLAS detector*. *Eur. Phys. J. C*, **77**:393, 2017.

[11] CMS Collaboration. *Search for new physics in the monophoton final state in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, CMS-EXO-16-039. URL: `http://cms-results.web.cern.ch/cms-results/public-results/publications/EXO-16-039/`, 2017.

[12] ATLAS Collaboration. *Search for dark matter in events with a Z boson and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. *Phys. Rev. D*, **90**:012004, 2014.

[13] ATLAS Collaboration. *Search for dark matter in events with a hadronically decaying W or Z boson and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. *Phys. Rev. Lett.*, **112**:041802, 2014.

[14] CMS Collaboration. *Search for dark matter and unparticles in events with a Z boson and missing transverse momentum in proton-proton collosions at $\sqrt{s} = 13$ TeV*. *JHEP*, **03**:061, 2017.

[15] ATLAS Collaboration. *Search for dark matter produced in association with a Higgs boson decaying to $b\bar{b}$ using 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. *Phys. Rev. Lett.*, **119**:181804, 2017.

[16] ATLAS Collaboration. *Search for dark matter in association with a Higgs boson decaying to b-quarks in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. *Phys. Lett. B*, **765**:11, 2017.

[17] ATLAS Collaboration. *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, ATLAS-CONF-2016-035. URL: `https://cds.cern.ch/record/2200211`, 2016.

[18] ATLAS Collaboration. *Simulation and parametrisation of dynamic pile-up corrections in the ATLAS Level-1 Calorimeter Trigger*, ATL-COM-DAQ-2015-120. URL: `https://cds.cern.ch/record/2038692`, 2015.

[19] PETRA Collaboration. *Discovery of Three-Jet Events and a Test of Quantum Chromodynamics at PETRA*. *Phys. Rev. Lett.*, **43**:830–833, 1979.

[20] UA1 Collaboration. *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*. *Phys. Lett. B*, **122**:103–116, 1983.

[21] UA2 Collaboration. *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN $p\bar{p}$ Collider*. *Phys. Lett. B*, **122**:476–485, 1983.

[22] UA1 Collaboration. *Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider*. *Phys. Lett. B*, **126**:398–410, 1983.

[23] CDF Collaboration. *Observation of top quark production in $\bar{p}p$ collisions*. *Phys. Rev. Lett.*, **74**:2626–2631, 1995.

[24] D0 Collaboration. *Observation of the top quark*. *Phys. Rev. Lett.*, **74**:2632–2637, 1995.

[25] Particle Data Group, PBS NOVA, Fermilab. *Standard Model particle content*. URL: `https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`, 2008.

[26] ATLAS Collaboration. *The Standard Model*, arXiv:9812242. URL: `https://arxiv.org/abs/hep-ph/9812242`, 1998.

[27] C. Giunti and C.W. Kim. *Fundamentals of Neutrino Physics and Astrophysics*. *Oxford University Press*, 2007.

[28] N. Cabibbo. *Unitary Symmetry and Leptonic Decays*. *Phys. Rev. Lett.*, **10**:531–533, 1963.

[29] M. Kobayashi and T. Maskawa. *CP Violation in the Renormalizable Theory of Weak Interation*. *Prog.Theory.Phys.*, **49**:652–657, 1973.

[30] K. A. Olive et al. *Review of Particle Physics*. *Chin. Phys. C*, **38**:09001, 2014.

[31] F. Zwicky. *Spectral displacement of extra galactic nebulae*. *Helv.Phys.*, **6**:110–127, 1933.

[32] A. Refregier. *Weak Gravitational Lensing by Large-Scale Structure*. *Ann.Rev.Astron.Astrophys.*, **41**:645–668, 2003.

[33] J. A. Tyson, G. P. Kochanski, and I. P. Dell'Antonio. *Detailed Mass Map of CL0024+1654 from Strong Lensing*. *Astrophys. J.*, **498**:L107, 1998.

[34] E. Komatsu et al. . *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*. *Astrophys. J.*, **192**:18–65, 2011.

[35] J. L. Feng. *Dark Matter Candidates from Particle Physics and Methods of Detection*. *Ann.Rev.Astr.Astrophys.*, **48**:495–545, 2010.

[36] G. Bertone, D. Hooper and K. Silk. *Particle Dark Matter: Evidence, Candidates and Constraints*. *Phys.Rep.*, **405**:279–390, 2005.

[37] L. Bergstroem. *Non-baryonic Dark Matter: Observational Evidence and Detection Methods*. *Rep.Prog.Phys.*, **63**:793–843, 2000.

[38] Planck Collaboration. *Planck 2015 results. I. Overview of products and scientific results. Astr.Astrophy.*, **594**:A1, 2016.

[39] Planck Collaboration. *Planck 2015 results. XIII. Cosmological parameters. Astr.Astrophy.*, **594**:A13, 2016.

[40] HEPAP LHC/ILC Subpanel . *Discovering the Quantum Universe*. URL: `http://www.linearcollider.org/`, 2006.

[41] K.G. Begeman, A.H. Broeils and R.H. Sanders. *Extended rotation curves of spiral galaxies: dark haloes and modified dynamics*. *Mon.Not.R.Astr.Soc.*, **249**:523–537, 1991.

[42] V.C. Rubin and W.K. Ford. *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*. *Astrophys. J.*, **159**:379, 1970.

[43] D. Clowe et al. *On dark peaks and missing mass: a weak lensing mass reconstruction of the merging cluster system Abell 520. Astrophys. J.*, **758**:128–144, 2012.

[44] X-ray: NASA/CXC/CfA/M.Markevitch et al.; Optical: NASA/STScI; Magellan/U.Arizona/D.Clowe et al.; Lensing Map: NASA/STScI; ESO WFI; Magellan/U.Arizona/D.Clowe et al. *Bullet Cluster: Galaxy Cluster 1E 0657-56*. URL: `http://chandra.harvard.edu/photo/2006/1e0657/`, 2006.

[45] A. Berlin, T. Lin and L.-T. Wang. *Mono-Higgs Detection of Dark Matter at the LHC. JHEP*, **1406**:078, 2014.

[46] L. Carpenter et al. *Mono-Higgs-boson: A new collider probe of dark matter. Phys. Rev. D*, **89**:075017, 2014.

[47] G.C. Branco et al. *Theory and phenomenology of two-Higgs-doublet models*. *Phys. Rept.*, **516**:1–102, 2012.

[48] A. Berlin, S. Gori, T. Lin and L.-T. Wang. *Pseudoscalar Portal Dark Matter*. *Phys. Rev. D*, **92**:015005, 2015.

[49] P. Langacker. *The physics of heavy $Z^{'}$ gauge bosons. Rev. Mod. Phys.*, **81**:1199, 2009.

[50] J. March-Russell, S. M. West, D. Cumberbatch and D. Hooper. *Heavy dark matter through the Higgs portal. JHEP*, **07**:058, 2008.

[51] D. O'Connell, M.J. Ramsey-Musolf and M.B. Wise. *Minimal extension of the standard model scalar sector. Phys. Rev. D*, **75**:037701, 2007.

[52] G. Belanger et al. *Status of invisible Higgs decays. Phys.Lett.B*, **723**:340–347, 2013.

[53] A. Falkowski, F. Riva and A. Urbano. *Higgs at last. JHEP*, **11**:111, 2013.

[54] J. Ellis, T. You. *Updated global analysis of Higgs couplings. JHEP*, **06**:103, 2013.

[55] P. Bryant and L. Evans. *LHC Machine*. *JINST*, **3**:S08001, 2008.

[56] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider*. *JINST*, **3**:S08003, 2008.

[57] Felix Müller. *Jet production measurements at the ATLAS experiment*, Heidelberg University. URL: https://www.kip.uni-heidelberg.de/Veroeffentlichungen/download.php/5442/temp/2966.pdf, 2014.

[58] X. C. Vidal and R. C. Manzano. *Taking a closer look at the LHC*. URL: https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.rf_cavities, 2017.

[59] ATLAS Collaboration. *The ATLAS Inner Detector commissioning and calibrations*. *Eur. Phys. J. C*, **70**:787, 2010.

[60] ATLAS Collaboration. *Performance of the ATLAS tirgger system in 2015*. *Eur. Phys. J. C*, **77**:317, 2017.

[61] ATLAS Collaboration. *Luminosity Public Results in Run 2* . URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2, 2014.

[62] ATLAS Collaboration. *The ATLAS Simulation Infrastructure*. *Eur. Phys. J. C*, **70**:823, 2010.

[63] H.-L. Lai et al. *New parton distributions for collider physics*. *Phys. Rev. D*, **82**:074024, 2010.

[64] R.D. Ball et al. *Parton distributions with LHC data*. *Nucl. Phys. B*, **867**:244, 2013.

[65] J. Alwall et al. *MadGraph/MadEvent v4: The New Web Generation*. *JHEP*, **09**:028, 2007.

[66] T. Sjostrand, S. Mrenna and P. Skands. *A brief introduction to PYTHIA 8.1*. *Comput.Phys.Commun.*, **178**:852, 2008.

[67] ATLAS Collaboration. *ATLAS Run 1 Pythia8 tunes*, ATL-PHYS-PUB-2014-021. URL: https://cds.cern.ch/record/1966419, 2014.

[68] T. Sjostrand, S. Mrenna and P. Skands. *PYTHIA 6.4 Physics and Manual*. *JHEP*, **0605**:026, 2006.

[69] T. Gleisberg et al. *Event generation with SHERPA 1.1*. *JHEP*, **0902**:007, 2009.

[70] ATLAS Collaboration. *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003. URL: https://cds.cern.ch/record/1474107, 2012.

[71] G. Watt and R.S. Thorne. *Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs*. *JHEP*, **08**:052, 2012.

[72] S. Suchek, O. Brandt and S. Meehan. HLT_XEX *triggers in resolved regime*, $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis group. URL: https://indico.cern.ch/event/538483/contributions/2219261/attachments/1298921/1938583/monohbb.pdf, 2016.

[73] S. Suchek, O. Brandt and S. Meehan. *MET trigger scale factors for 2015/2016: implementation and validation*, $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis group. URL: https://indico.cern.ch/event/538483/contributions/2430502/attachments/1394748/2125800/ssuchek_monoHbb_12012017_v01.pdf, 2017.

[74] A. Buzatu and W. Wang. *Object selections for SM Higgs boson produced in association with a vector boson in which H→bb and V decays leptonically with Run-2 data: Object support note for VH(bb) 2015+2016 dataset publication*, ATL-COM-PHYS-2016-1674. URL: https://cds.cern.ch/record/2233686, 2016.

[75] W. Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002. URL: https://cds.cern.ch/record/1099735, 2008.

[76] M. Cacciari, G.P. Salam and G. Soyez. *The anti-k_t jet clustering algorithm*. *JHEP*, **0804**:063, 2008.

[77] M. Cacciari, G.P. Salam and G. Soyez. *FastJet User Manual*. *Eur. Phys. J. C*, **72**:1896, 2012.

[78] ATLAS Collaboration. *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, CERN-EP-2017-038. URL: http://inspirehep.net/record/1519834, 2017.

[79] ATLAS Collaboration. *Jet energy scale measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*. *Eur. Phys. J. C*, **75**:17, 2015.

[80] ATLAS Collaboration. *Performance of b-jet identification in the ATLAS experiment*. *JINST*, **11**:04008, 2016.

[81] ATLAS Collaboration. *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012. URL: https://cds.cern.ch/record/2160731, 2016.

[82] G. piacquadia and C. Weiser. *A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS*. *J. Phys.*, **119**:032032, 2009.

[83] ATLAS Collaboration. *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*. *Eur. Phys. J. C*, **77**:490, 2017.

[84] Joe Taenzer. *Large-R jet reconstruction and calibration at 13 TeV with ATLAS*. URL: https://indico.cern.ch/event/579660/contributions/2496135/attachments/1495415/2326534/BOOST17_largeR_reco_calib_ATLAS.pdf, 2017.

[85] D. Krohn, J. Thaler and L.-T. Wang. *Jet Trimming*. *JHEP*, **1002**:084, 2010.

[86] ATLAS Collaboration. *ATLAS Combined Mass resolution and W mass reconstruction performance*, JETM-2017-002. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2017-002/, 2017.

[87] ATLAS Collaboration. *Performance of jet substructure techniques for large-R jets in proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*. *JHEP*, **1309**:076, 2013.

[88] M. Baehr et al. *Herwig++ Physics and Manual*. *Eur. Phys. J. C*, **58**:639–707, 2008.

[89] ATLAS Collaboration. *Flavour Tagging with Track-Jets in Boosted Topologies with the ATLAS Detector*, ATL-PHYS-PUB-2014-013. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-013/, 2014.

[90] M. Cacciari, G.P. Salam and G. Soyez. *The Catchment Area of Jets*. *JHEP*, **0804**:005, 2008.

[91] ATLAS Collaboration. *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024. URL: http://cds.cern.ch/record/2157687, 2016.

[92] ATLAS Collaboration. *Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data*. *Eur. Phys. J. C*, **77**:195, 2017.

[93] W. Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*, ATLAS-LARG-PUB-2008-002. URL: https://cds.cern.ch/record/1099735, 2008.

[94] T. Cornelossen et al. *The global χ^2 track fitter in ATLAS*. *J.Phys.Conf.Ser.*, **119**:032013, 2008.

[95] ATLAS Collaboration. *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047. URL: http://cds.cern.ch/record/1449796, 2011.

[96] ATLAS Collaboration. *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*. *Eur. Phys. J. C*, **74**:3071, 2014.

[97] ATLAS Collaboration. *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*. *Eur. Phys. J. C*, **77**:490, 2017.

[98] ATLAS Collaboration. *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}$=13 TeV*. *Eur. Phys. J. C*, **76**:292, 2016.

[99] ATLAS Collaboration. *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s}$=13 TeV*, ATLAS-CONF-2017-029. URL: https://cds.cern.ch/record/2261772, 2017.

[100] ATLAS Collaboration. *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045. URL: https://cds.cern.ch/record/2064383, 2015.

[101] ATLAS Collaboration. *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s}$=8 TeV*. *Eur. Phys. J. C*, **75**:303, 2015.

[102] ATLAS Collaboration. *Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton-proton collisions at $\sqrt{s} =$ 13 TeV*, ATL-PHYS-PUB-2015-027. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-027/, 2015.

[103] ATLAS Collaboration. *Expected performance of missing transverse momentum reconstruction for the ATLAS detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-023. URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-023/, 2015.

[104] ATLAS collaboration. *Jet Cleaning*. URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets2015, 2017.

[105] ATLAS Collaboration. *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATLAS-PHYS-PUB-2016-012. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-012/, 2016.

[106] G. Cowan, K. Cranmer, E. Gross and O. Vittels. *Asymptotic formulae for likelihood-based tests of new physics*. *Eur. Phys. J. C*, **71**:1554, 2011, [Erratum: Eur. Phys. J. C 73,2501(2013)].

[107] D. M. Portillo, P. Francavilla and S. de Cecco. *Preliminary study: simplified models in the search for dark matter*, $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis group. URL: https://indico.cern.ch/event/443590/contributions/1098919/attachments/1193046/1732211/mono-Hmeeting_23-11-2015.pdf, 2015.

[108] C.-H. Kom and W. J. Stirling. *Charge asymmetry in W + jets production at the LHC*. *Eur. Phys. J. C*, **69**:67, 2010.

[109] J. Beringer et al. (Particle Data Group). *Review of Particle Physics*. *Phys. Rev. D*, **86**:010001, 2012.

[110] ATLAS Statistics Forum. *The CL_s method: information for conference speakers*. URL: https://twiki.cern.ch/twiki/pub/AtlasProtected/StatisticsTools/CLsInfo.pdf, 2016.

[111] A. Wald. *Tests of statistical hypotheses concerning several parameters when the number of observations is large*. *Trans. Am. Math. Soc.*, **54(3)**:426–482, 1943.

[112] S.S. Wilks. *The large-sample distribution of the likelihood ratio for testing composite hypotheses*. *Ann. Math. Stat.*, **9**:60–62, 1938.

[113] S. Suchek, O. Brandt and S. Meehan. *MET triggers scale factors for 2015/2016: brand new Trigger Tool arrived!*, $h(\rightarrow b\bar{b}) + E_T^{miss}$ analysis group. URL: https://indico.cern.ch/event/538483/contributions/2444435/attachments/1398696/2133402/ssuchek_monoHbb_19012017.pdf, 2017.

[114] ATLAS collaboration. *Uncertainty release for analyses using 2015 and 2016 data in 20.7 (MC15c) releases*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetUncertainties20152016Data20p7, 2016.

[115] ATLAS collaboration. *Jet energy resolution recommendations*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetResolution2015Prerecom, 2016.

[116] ATLAS collaboration. *JetEtMiss recommendations for MC15 (release 20.7)*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetEtmissRecommendations2016, 2016.

[117] ATLAS collaboration. *Flavour tagging recommendations in release 20.7 (February 2017)* . URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTagCalib2015, 2017.

[118] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018. URL: https://cds.cern.ch/record/1700870, 2014.

[119] ATLAS collaboration. *Systematics study: impact of MET soft term and JVT efficiency*. URL: https://indico.cern.ch/event/538483/contributions/2527699/attachments/1432840/2201659/talk.pdf, 2016.

[120] ATLAS collaboration. *Recommendations for* $2015+2016$ *data analysis aiming for 2017 winter conferences*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetUncertainties2016Moriond2017, 2017.

[121] ATLAS collaboration. *Electron Efficiencies for Run2 (scale factors and uncertainties)*. URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ElectronEfficiencyRun2, 2016.

[122] ATLAS collaboration. *Electron and Photon Calibration for Run2 (energy scale and resolution)*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaCalibrationRun2, 2016.

[123] ATLAS collaboration. *Muon combined performance analysis guidelines*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesMC15, 2016.

[124] ATLAS collaboration. *Tau systematic recommendations for Moriond 2017*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TauRecommendationsMoriond2017, 2016.

[125] ATLAS collaboration. *Tracking systematic recommendations for Moriond 2017*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TrackingCPMoriond2017, 2016.

[126] ATLAS collaboration. *Usage of E_T^{miss} in analyses: rebuilding and systematics*. URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/METUtilities, 2017.

[127] ATLAS collaboration. *Signal and Background Modelling Studias for the Standard Model Vh, $h \to b\bar{b}$ and Related Searches: Modelling support note for $Vh(bb)$ 2015 + 2016 dataset publication* . URL: https://cds.cern.ch/record/2235887, 2017.

[128] ATLAS collaboration. *Studies on top-quark Monte Carlo modelling for Top2016*. URL: https://cds.cern.ch/record/2216168, 2017.

[129] A.D. Martin et al. . *Parton distributions for the LHC*. *Eur. Phys. J. C*, **63**:189, 2009.

[130] J. Pumplin et al. . *New generation of parton distributions with uncertainties from global QCD analysis*. *JHEP*, **0207**:012, 2002.

[131] ATLAS collaboration. *Theoretical studies for the Standard Model Vh, $h \to b\bar{b}$ and related searches*. URL: https://cds.cern.ch/record/2151835, 2016.

[132] ATLAS collaboration. *Theoretical studies for the 2HDM $A \to Zh$, $h \to b\bar{b}$ search and the SM Vh, $h \to b\bar{b}$*. URL: https://cds.cern.ch/record/2111370, 2016.

[133] O. Brandt et al. *Mono-H Modelling and Fit Studies*, $h(\to b\bar{b}) + E_T^{miss}$ analysis group. URL: https://indico.cern.ch/event/538483/contributions/2483456/attachments/1415492/2167157/rroehrig_ttbarMod.pdf, 2017.

[134] D. Guest. *Interpolating exclusion on the $h(\to b\bar{b}) + E_T^{miss}$ grid*, $h(\to b\bar{b}) + E_T^{miss}$ analysis group. URL: https://indico.cern.ch/event/538483/contributions/2548655/attachments/1441663/2219793/dguest-interpolation.pdf, 2017.

[135] Particle Data Group. *The W Boson*. URL: http://pdg.lbl.gov/2017/listings/rpp2017-list-w-boson.pdf, 2017.

[136] Particle Data Group. *The Top Quark*. URL: http://pdg.lbl.gov/2017/listings/rpp2017-list-t-quark.pdf, 2017.

[137] ATLAS Collaboration. *Search for dark matter produced in association with a Higgs boson decaying to two bottom quarks in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*. *Phys. Rev. D*, **93**:072007, 2016.

Acknowledgements

Primero que nada, me gustaría agredecer a una persona sin quien nada de esta tesis y nada del arduo trabajo en ATLAS durante estos años habría sido posible: mi esposa. Mi Amor, mi Armella-Lucia, nadie más que tú sabe bien lo que esta tesis significa, sólo tú conoces todo el esfuerzo, tiempo y sacrificios que han permitido llegar exitosamente hasta este punto, sorteando injusticias y dificultades mucho más allá de y mucho más complicadas que lo académico. Agradezco infinitamente tu amor incondicional, tu apoyo incondicional, tu palabra constante, cada detalle del diario vivir que hicieron posible este camino. Esta tesis no es sólo mía, es tuya, que sin ti, nada de ella sería. Agradezco a mi familia, a mis padres Franco y Rosa y a mi hermana Francesca, que siempre estuvieron allí a mi lado a pesar de la distancia, brindando consejos, escuchando, alegrándose conmigo, enojándose conmigo, apenándose conmigo, luchando conmigo. Gracias infinitas por todo, mi familia, mis amados, a ustedes debo más que un logro académico, gran parte de lo que soy. A ustedes todos, mis Ángeles, les agradezco su guía constante, nuevas visiones y perspectivas, la salida correcta, su amor incondicional. Este final es un comienzo.

I would like to thank my supervisor Priv.-Doz. Dr. Oleg Brandt for the opportunity to pursue doctoral studies at Kirchhoff-Institute for Physics. It was an exciting period of challenges, researching and learning. I am grateful for the many fruitful Physics discussions, for transmiting much of his knowdledge in High Energy Physics to us and for always managing to allocate time when it was needed regardless of the many projects going on. Thank you for following the analysis, working hard side by side in the ATLAS mono-Higgs group and pushing forward the publication of our analysis. I would like to also thank my co-supervisors Prof. Dr. André Schöning and Prof. Dr. Hans-Christian Schultz-Coulon for overseeing the doctoral process and giving valuable and fair input and advice during the Ph.D. time. Special thanks to Prof. Dr. André Schöning for kindly agreeing to referee this thesis.

Many thanks as well to Santiago Casas, Merve Sahinsoy, Stanislav Suchek, Jennifer Thompson, Thomas Spieker, Claire Antel, Miguel Campos and Rainer Stamen for taking time to proofread my thesis.

I would like to thank all my collegues and friends in the ATLAS group of our institute. Getting to know you all and even becoming friends with some of you gave a lot of sense to my time at the institute. It was very motivating to share Physics discussions with you and learn from each other. In the middle of big challenges, it was refreshing to stop by, have a chat or just laugh about any topic. It was a gift to have all these sincere moments with you whatever the circumstances. Thank you all, thank you KIP for these four years that went by in the blink of an eye, I will always remember all I learnt, which was much more than Physics.