VNIVERSITAT DE VALÈNCIA

Dark matter search in the top-quark sector with the ATLAS detector at the LHC

Florencia L. Castillo

Directores:

Dr. Jose Enrique García Navarro

Dr. Carlos Escobar Ibáñez

Doctorat en Física

Falcutat de Física

Departament de Física Atòmica, Molecular i Nuclear

September 2020

*El **Dr. Jose Enrique García Navarro**, científico titular del Consejo Superior de Investigaciones Científicas (CSIC) y el **Dr. Carlos Escobar Ibáñez**, Contratado investigador distinguido (CIDEGENT) del Consejo Superior de Investigaciones Científicas (CSIC)*

Certifican:

Que la presente memoria, "Dark matter search in the top-quark sector with the ATLAS detector at the LHC" ha sido realizada bajo su dirección en el Departament de Física Atómica, Molècular i Nuclear de la Universitat de València por ***Florencia L. Castillo*** y constituye su tesis para optar al título de Doctora en Física por la Universitat de València.

Y para que así conste, en cumplimiento de la legislacíon vigente, firman el presente certificado.

Valencia, a día 23 de julio de 2020:

Dr. Jose Enrique García Navarro

Dr. Carlos Escobar Ibáñez

Con el visto bueno del tutor, el ***Dr. Juan Zúñiga Román***

JUAN|ZUÑIGA|ROMAN

Firmado digitalmente por JUAN|ZUÑIGA|ROMAN
Fecha: 2020.07.20 14:17:55 +02'00'

“Tantas veces me mataron
Tantas veces me morí
Sin embargo estoy aquí resucitando
Gracias doy a la desgracia y a la mano con puñal
Porque me mató tan mal
Y seguí cantando”

- María Elena Walsh.

ABSTRACT

Astronomical and cosmological observations support the existence of invisible matter that can only be detected through its gravitational effects, thus making it very difficult to study. This component, called dark matter, makes up about 26.8% of the known universe. Experiments at the LHC, located at CERN, search for new particles to be dark matter candidates. A dark matter production can consist of an excess of events with a single final-state object X recoiling against large amount of missing momentum of energy called mono-X signal.

The studies presented in this thesis are focused on the mono-X signature, with X being a top quark, named mono-top. The topology is studied, where the W boson from the associated top-quark decays into a lepton (electron or a muon) and a neutrino. Firstly, a sensitivity search of dark matter production in an extension of the Standard Model featuring a two-Higgs-doublet model and an additional pseudo-scalar is presented. The pseudo-scalar is the mediator which decays to the dark matter particles. This analysis uses all the data collected by the ATLAS experiment at the LHC during Run 2 (2015–2018) corresponding to an integrated luminosity of 139 fb^{-1}, at a centre-of-mass energy of 13 TeV. A multivariate analysis based on a boosted decision tree is performed in order to enhance the discrimination of signal events from the main background. The results are expressed as 95% confidence level limits on the parameters of the signal models considered. No significant excess is found with respect to Standard Model predictions. The region below of $\mathrm{m_{H^{\pm}}}$ = 800 GeV (1100 GeV) with $\tan\beta = 0.3$ is excluded with 95% confidence level of the observed (expected) limit.

Furthermore, the prospect of the potential discovery of the non-resonant production of an exotic state decaying into a pair of dark matter candidates in association with a right-handed top quark in the context of an effective dark matter flavour-changing neutral interaction, at the HL-LHC is presented as well. The HL-LHC project is expected to operate at a centre-of-mass energy of 14 TeV aiming to provide a total integrated luminosity of 3000 fb^{-1}. The number of signal and background events

is estimated from simulated particle-level truth information after applying smearing functions to mimic an upgraded ATLAS detector response in the HL-LHC environment. The expected exclusion limit (discovery reach) at 95% confidence level for the mass of the exotic state is calculated to be 4.6 TeV (4.0 TeV), using a multivariate analysis based on a boosted decision tree.

CONTENTS

To my family, for their unconditional love.

CHAPTER 1

The Standard Model of particle physics

"If we believe in nothing, if nothing has any meaning and if we can affirm no values whatsoever, then everything is possible and nothing has any importance."
- Albert Camus, The Rebel

The Standard Model (SM) of particle physics is the name given in the 1970s to the theory of fundamental particles and their interactions. It incorporates special relativity and quantum mechanics, and therefore it is based on quantum field theory. The SM describes the interactions between elementary fermions through the electroweak (EW) and strong forces under the symmetry group $\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ [1–4] [1]. An overview of their constituents is described in Section 1.1.
The EW theory [2] describes both the electromagnetic and the weak forces by extending the quantum electrodynamics (QED) theory, which is explained in Section 1.2. The quantum cromodymamics (QCD) theory is a gauge theory characterising the strong interaction. At the end of the last century, the discovery of unstable hadrons [5–7] was a fundamental key for the development of this theory. An explanation of this force is summarised in Section 1.3. Section 1.4 describes the Brout–Englert–Higgs (BEH) mechanism [8–10] which introduces the spontaneous symmetry breaking (SSB) to explain the mass of bosons and fermions. Finally, Section 1.5 shows a list of known properties of Nature that can not yet be explained by the SM.

1.1 The particles of the Standard Model

The SM is a quantum field theory where particles of matter and antimatter are defined as local excitation of fields permeating space. The interactions are described

[1] Where the subscripts C,L, and Y indicate colour, weak isospin and hypercharge, respectively.

by a gauge theory, where the associated Lagrangian does not change, i.e. it is invariant under local transformations. Those invariancies give rise to symmetries corresponding with conserved physical parameters [11]. The matrices representing these invariant transformations are the generators of a symmetry group. The SM is a gauge theory, based on the symmetry group $\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$.

The $\mathrm{SU(3)_C}$ group has eight generators, which are spin-1 particles called gluons (g), responsible for the strong interaction. Then, the $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ part is the EW side, with four generators. Corresponding to three gauge bosons (particles with integer spin) W and Z for the weak interaction, and the massless photon (γ) for the electromagnetic force. Fermions (particles with half-integer spin) are divided into two groups, quarks and leptons. Both are organised in three generations [12]. The first generation of quarks and leptons forms common matter, up (u) and down (d) quarks, the electron (e), and the electron neutrino (ν_e). The second generation contains the strange (s) and charm (c) quarks, the muon (μ), and the muon neutrino (ν_μ). The third generation includes the top (t) and bottom (b) quarks, the tau (τ) lepton, and the tau neutrino (ν_τ). Through higher generations, particles have larger mass than the previous generation, with the possible exception of the neutrinos (whose small but non-zero masses have not been accurately determined).

The last piece of the SM was recently (in 2012) discovered with the verification of the BEH mechanism by the A Toroidal LHC Apparatus (ATLAS) [13] and Compact Muon Solenoid (CMS) [14] Collaborations at Large Hadron Collider (LHC) in Conseil Européen pour la Recherche Nucléaire (CERN) complex [15, 16]. These two experiments successfully detected the quantised excitation of the BEH field, the so-called Higgs boson with a mass around 125 GeV. This discovery confirmed the Higgs mechanism, which is discussed further in Section 1.4. Properties of boson and fermion particles [17] are summarised in Table 1.1.

1.2 The electroweak interaction

The EW theory describes both the electromagnetic force (i.e QED) and the weak force. Superficially, these forces appear quite different. The weak force acts only across distances smaller than the atomic nucleus, while the electromagnetic force can

Particle	*Symbol*	*Mass* [GeV]	*Charge* [e]	*Strong*	*Weak*	*Electromagnetic*
Up quark	u	0.0022	$2/3$	✓	✓	✓
Down quark	d	0.0047	$-1/3$	✓	✓	✓
Charm quark	c	1.275	$2/3$	✓	✓	✓
Strange quark	s	0.95	$-1/3$	✓	✓	✓
Top quark	t	173.3	$2/3$	✓	✓	✓
Bottom quark	b	4.18	$-1/3$	✓	✓	✓
Electron	e	0.00051	-1	✗	✓	✓
Electron neutrino	ν_e	$< 2 \times 10^{-6}$	0	✗	✓	✗
Muon	μ	0.106	-1	✗	✓	✓
Muon neutrino	ν_μ	$< 2 \times 10^{-6}$	0	✗	✓	✗
Tau	τ	0.106	-1	✗	✓	✓
Tau neutrino	ν_τ	$< 2 \times 10^{-6}$	0	✗	✓	✗
Photon	γ	0	0	✗	✗	✓
Z	Z	91.19	0	✗	✓	✗
W	W	80.38	± 1	✗	✓	✓
Gluon	g	0	0	✓	✗	✗
Higgs	H	125.18	0	-	-	-

TABLE 1.1: Summary of the properties of SM particles [17].

extend for great distances (as observed in the light of stars reaching across entire galaxies), weakening only with the square of the distance. But, above of certain energy (on the order of 246 GeV) they merge into a single force, i.e. these two forces are different facets of a single and more fundamental, the EW force.

The unification of electromagnetic and weak interactions was proposed by Glashow, Salam and Weinberg [1–3]. They gave a picture of the EW theory as a chiral theory, i.e. a theory that distinguishes between left- and right-chiral fermion fields[2]. Using symmetry arguments, the model for the combined EW interactions should contain a doublet for the left-handed particles and the electromagnetic interactions. Therefore it is characterised by the following symmetry group:

$$\mathrm{SU(2)_L \times U(1)_Y},$$

[2]An object or a system is chiral if it is distinguishable from its mirror image. Human hands are an example of chiral objets.

where $U(1)_Y$ is the symmetry group of the electromagnetic interaction, where Y is the weak hypercharge[3]. The weak interaction is represented with the $SU(2)_L$ symmetry group, where L indicates that gauge fields of $SU(2)_L$ are only coupled to left-handed fermions. The fermions are divided into left and right handed helicity eigenstates where the left ones transform as doublets, and the right ones as singlets. Singlets do not interact weakly[4]. The weak hypercharge Y is defined by Gell–Mann–Nihijima formula [18]

$$Q = I_3 + \frac{Y}{2},$$

where Q is the electric charge and I_3 is the third component of the weak isospin. The weak hypercharge in the SM is a conserved quantum number. The three generators of the $SU(2)_L$ symmetry group are the three gauge boson fields $W_\mu^{\alpha=1,2,3}$ with a coupling strength g_W. One vector boson field B_μ is associated to gauge $U(1)_Y$ symmetry with a coupling strength g_Y. The covariant derivative of the EW interaction is:

$$D_\mu = \partial_\mu - \frac{1}{2} g_W \sigma_\alpha W_\mu^\alpha - \frac{1}{2} i g_Y Y B_\mu^\alpha, \tag{1.1}$$

where σ_α ($\alpha \in \{1,2,3\}$) are the Pauli matrices, i.e the generators of $SU(2)_L$. The complete EW Lagrangian is

$$\mathscr{L}_{EW} = \sum \bar{f} \gamma^\mu D_\mu f - \frac{1}{4} F_{\mu\nu}^\alpha F_\alpha^{\mu\nu}, \tag{1.2}$$

where γ^μ are the Dirac matrices. This Lagrangian describes the interaction of fermions f with EW gauge bosons. The EW field strength tensor $F_{\mu\nu}^\alpha$ is defined as

$$F_{\mu\nu}^\alpha = \partial_\mu W_\mu^\alpha - \partial_\nu W_\mu^\alpha + g_W f_{abc} W_\mu^b W_\nu^c + \partial_\mu B_\nu^\alpha - \partial_\nu B_\mu^\alpha, \tag{1.3}$$

where the term $F_{\mu\nu}^\alpha F_\alpha^{\mu\nu}$ in $\mathscr{L}_{EW}$ (Equation 1.2) is commonly known as the kinematic term $\mathscr{L}_{kin}$.

[3]The weak hypercharge is a quantum number relating the electric charge and the third component of weak isospin.

[4]Neutrinos are an exception since they only interact weakly. Consequently, they do not have a right-handed component in the SM.

The constructed Lagrangian successfully describes electromagnetic and weak interaction. Unfortunately, it contains no mass terms and therefore predicts massless particles. Experimental observations indicate the opposite, since the W and Z bosons are massive and only the γ remains massless [19, 20]. Adding mass terms for EW gauge fields to the Lagrangian defined in Equation 1.2 would break the gauge invariance. This inconsistency between theory and observation is solved by the introduction of the SSB, which is explained in Section 1.4.

1.3 The strong interaction

The strong interaction, as its name suggests, is the strongest force currently known in Nature. QCD is the theory framework that describes the strong interaction, where quarks and gluons interact under the $\mathrm{SU}(3)_\mathrm{C}$ symmetry group, with C being the colour-charge. The quarks are the only fermions carrying some type of colour charge as triplets inside the $\mathrm{SU}(3)_\mathrm{C}$ symmetry group. Leptons are colourless, thus they are singlets under $\mathrm{SU}(3)_\mathrm{C}$. QCD has three different kinds of charges, which are commonly named: red, green and blue, known as RBG-colour scheme. The QCD covariant derivative can be written as

$$D_\mu = \partial_\mu - \frac{1}{2} i g_s \lambda_\alpha G_\mu^\alpha,$$

where g_s is the strong coupling constant, λ_α ($\alpha \in 1, \ldots, 8$) are the Gell-Mann matrices, i.e. the generators of $\mathrm{SU}(3)_\mathrm{C}$, and G_μ^α are the eight gauge bosons force carriers, named gluons. One important feature of the strong force is the asymptotic freedom and confinement of QCD state, where the strength of the strong coupling g_s (usually referred to as α_S) becomes weak at small distances (high energy scales) and strong at a large distance (low energy scales), leading to the confinement of quarks and gluons within composite hadrons.

This theory can be expressed similarly to QED, therefore the kinematic term of the Lagrangian of QCD can be written as

$$\mathscr{L}_{\mathrm{QCD,kin}} = -\frac{1}{4} G_{\mu\nu}^\alpha G_\alpha^{\mu\nu},$$

where $G^{\alpha}_{\mu\nu}$ is the strong strength tensor, i.e.

$$G^{\alpha}_{\mu\nu} = \partial_{\mu} G^{\alpha}_{\mu} + g_s f_{\alpha bc} G^{b}_{\mu} G^{c}_{\nu}. \tag{1.4}$$

The field-strength tensor for QCD (Equation 1.4) has a crucial difference with QED (Equation 1.3): the self-interaction of gauge bosons is possible in QCD as opposed to QED, where photons couple only to charged fermions[5].

1.4 Electroweak symmetry breaking

The BEH mechanism [8–10] uses the SSB to allow particles to acquire mass. An additional scalar field is introduced into the Lagrangian to break the $SU(2)_L \times U(1)_Y$ symmetry, and to explain the experimental evidence of massive W and Z EW gauge bosons[6]. The SSB ensures that γ remains massless i.e. the Lagrangian is invariant under the electromagnetic symmetry group $U(1)_{EM}$. The scalar field with such property is called Higgs field (ϕ), and can be expressed as,

$$\phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.5}$$

Equation 1.5 carries four degrees of freedom, the four $\phi_{i=1,2,3,4}$ real scalar fields. The additional term in the Lagrangian due to the Higgs field is,

$$\mathscr{L}_{\text{Higgs}} = \left(D_{\mu}\phi\right)^{\dagger} \left(D^{\mu}\phi\right) - \left(\mu^2 \phi^{\dagger}\phi + \lambda \left(\phi^{\dagger}\phi\right)^2\right),$$

where the superscript † refers to the transpose conjugate of a matrix, D_{μ} is the covariant derivative of the EW interaction shown in Equation 1.1. The second term is known as Higgs potential ($V(\phi)$). The shape of the potential V depends on the real parameters μ^2 and λ. The parameter λ must be above zero for the theory to remain stable, while μ^2 can be negative or positive. Picking $\mu^2 > 0$ produces a trivial potential with only one minimum located at $\phi^{+} = \phi^{0} = 0$, which does not allow for

[5] Unlike the photon, which is electrically neutral, the gluon has a colour charge.

[6] The fermions (quarks and leptons) acquire a mass proportional to the vacuum expectation value of the Higgs, described by the Yukawa interaction.

symmetry breaking. However, setting $\mu^2 < 0$ the potential is symmetric, with a local maximum at the origin and a set of local minima ring,

$$\min\{V(|\phi|)\} = \sqrt{-\frac{\mu^2}{2\lambda}},$$

the Higgs field has a non-zero value in this minima ring, which is the vacuum expectation value (VEV), the VEV is represented by ν, obtaining,

$$\langle 0|\,\phi\,|0\rangle = \sqrt{-\frac{\mu^2}{2\lambda}} \equiv \frac{\nu}{2} \neq 0.$$

The minimum of the Higgs potential can be chosen that the charged term of the Higgs doublet is zero while the neutral has the VEV (Equation 1.5),

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu \end{pmatrix}.$$

The SSB leaves the $U(1)_Y$ symmetry of the electromagnetic group unaltered. Parametrising a small excitation of the scalar doublet around the selected minimum in unitarity gauge results to

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu + H \end{pmatrix}, \tag{1.6}$$

where the H is the named Higgs boson and has a mass of $m_H = \sqrt{-\mu^2}$. After the SSB, the $W^{\pm}$, Z^0 and the photon are described by linear combination of W_μ and B_μ, such that $W^{\pm}$ and Z^0 acquire mass.

The γ and Z boson fields are obtained by adding the Equation 1.6 into the kinematic term (Equation 1.4). After the diagonalisation of the resulting matrix, the fields are:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} := \begin{pmatrix} \cos(\theta_W) & \sin(\theta_W) \\ -\sin(\theta_W) & \cos(\theta_W) \end{pmatrix} \begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix}$$

with

$$\cos(\theta_W) := \frac{g_W}{\sqrt{g_W^2 + g_Y^2}},$$

where θ_W is the Weinberg angle [2]. The charged vector boson, and its complex conjugate are defined as $W_\mu^\pm := \frac{1}{\sqrt{2}}(W_\mu^1) \mp iW_\mu^2$ and the massive bosons are:

$$m_W = \frac{g_W v}{2},$$

$$m_Z = \frac{g_W v}{2\cos(\theta_W)}.$$

Similarly, the Higgs field allows fermions to obtain their masses through Yukawa interactions:

$$\mathscr{L}_{\text{Yukawa}} = -y_\ell^{ij} \ell_{Li}^\dagger e_R \phi - y_q^{ij} q_{Li}^\dagger u_{Rj} i\sigma_2 \phi^* - y_q^{ij} q_{Li}^\dagger d_R \phi + \text{h.c.} \tag{1.7}$$

The matrices y^{ij} describe the called Yukawa couplings between the single Higgs doublet ϕ and the fermions, σ_2 is the second Pauli generator[7] and i ($\in \{1,2,3\}$) represents the three fermion generators. The terms in the Yukawa Lagrangian (Equation 1.7) allow all fermions (except for neutrinos) to obtain mass thought an interaction with the Higss boson.

The Yukawa coupling matrices are non-diagonal, therefore the mass of fermions are a mixture of the EW eigenstates. The Cabibbo–Kobayashi–Maskawa (CKM) matrix [21, 22] describes the mixing between the three different families of quarks.

To summarise, the SM has 19 free parameters (with the massless left-handed neutrinos), all of them have been experimentally measured. These include the masses of the elementary particles, six quarks, three leptons, the Higgs boson mass, the Higgs self-coupling strength λ, the gauge couplings g_s, g_W and g_Y of the three symmetries $SU(3)_C$, $SU(2)_L$ and $U(1)_Y$, four parameters of the CKM mixing angles and one Charge conjugation-Parity (CP) violation phase.

[7] The Pauli matrices are a set of three 2×2 complex matrices which are hermitian and unitary.

1.5 Limitations of the Standard Model

Although the excellent success of the SM, there are some aspects of Nature that can not be explained by this theory, some of them are:

- Astronomical and cosmological observations support the existence of invisible matter [23–25] that has only been detected through its gravitational effects. Dark matter (DM) makes up about 26.8% of the energy of the known universe [26]. This topic is widely explained in the next chapter.

- The observation of the accelerated expansion of the universe [27] can be described with a new component called dark energy, an enigmatic phenomenon that acts in opposition to gravity. This new energy source makes up to about 68.3% of the energy in the known universe [26]. Currently there is no clue about its nature.

- Observed matter–antimatter asymmetry in the early universe after the Big Bang is not explained by the SM. The existence of an imaginary phase in the CKM matrix allows a small CP violation, but this does not cover the asymmetry observed [28].

- The strong-CP problem [29] relates to the question of why there is not CP violation in the strong interaction. While the weak interaction can have small CP violation, the strong force naturally allows CP violation, but has not been observed. Indeed, it is experimentally constrained to be very small by measurements on neutron dipole moment [30]. The dynamics that would cancel the neutron electric dipole moment also produces a new particle: the axion [31], which solves the strong-CP problem.

- The SM predicts massless neutrinos. Observations have revealed that neutrinos have non-zero masses [32]. Although, the introduction of massive neutrinos seems to be a minimal extension of the SM, the nature of the mass of the neutrino particle is still unknown.

- One of the open outstanding questions is why the SM contains three generations of matter particles. In physical world, it seems that only the first-generation particles play crucial roles, the two remaining seem "unnecessary" [33]. One possible answer could be found with the extra-dimensional models [34, 35] or discrete symmetries [36–38], since such theories require three generations as a consistency condition. However, they still are hypothetical.

- The fermion masses vary over 11 orders of magnitude (from neutrinos to the top quark). The SM does not explain about the origin of these differences between fermion masses [39].

- The discovery of the Higgs boson confirmed the SSB. But the SM requires an unnatural fine-tuning of its parameters to produce a Higgs boson mass around 100 GeV [15, 16]. This value is much smaller than the Plank scale[8] (10^{19} GeV) [40], so there is a huge gap between fundamental scales. Since the Higgs boson couples to all massive particles, the quantum corrections (Δm_H^2) to a fermion (f) mass are proportional to Λ, the cut-off energy scale,

$$\Delta m_H^2 = -N_f \frac{\lambda_f^2}{8\pi^2}[\Lambda^2 + ...],$$

  where λ_f is the coupling to the Higgs with fermion and N_f is the number of fermions. Since the quantum corrections are proportional to Λ^2, where the cut-off value is the Plank scale, in order to get a Higgs mass around 100 GeV, it is necessary to include counteracting terms around 10^{-30} GeV, which is extremely unnatural. Consequently, the universe is considerably fine-tuned, or there is a new process which cancels the divergences and sets the Higgs boson mass, requiring new physics between the EW scale (100 GeV) and the Planck scale. This theoretical problem is known as the hierarchy problem [41]. In order to solve this problem various theories introduce additional particles and

[8] In particle physics and physical cosmology, the Planck scale is an energy scale around 10^{19} GeV (the Planck energy, corresponding to the mass–energy equivalence of the Planck mass, 2.17645×10^{-8} kg) at which quantum effects of gravity become strong.

symmetries. Supersymmetry (SUSY) theory is the most popular of the extension of the SM which solves the hierarchy problem by the introduction of the supersymmetry partners. This cancels completely the corrections avoiding the fine-tuning [42].

- Last but not least is the absence of gravity in the SM, which does not have a quantum description.

Finally, there is a model in which, at high energies, the three gauge interactions of the SM comprising all fundamental forces except gravity are merged into a single force, the so-called Grand Unified Theory (GUT) [43]. Experiments confirmed that at high energy the electromagnetic interaction and weak interaction unify into a single EW interaction. GUT models predict that at even higher energy, around the 10^{16} GeV scale [44], the strong interaction and the EW interaction would unify into a single interaction. Although this unified force has not been directly observed, the various GUT models theorise its existence. Some of the problems of the SM discussed above, such us the existence of three quark-lepton families or the hierarchy problem, could naturally be solved by extended GUTs.

CHAPTER 2

Dark Matter

"And I say to any man or woman. Let your soul stand cool and composed before a million universes."

- Walt Whitman

The existence of DM has been discussed for more than a century [45]. The history of the DM problem can be traced back to the 1930s, but it was not until the early 1970s that the issue of "missing matter" was widely recognised as problematic [46]. The evidence is based on a set of observations, including anisotropies of the cosmic microwave background (CMB), baryonic acoustic observation, type Ia supernovae, gravitational lensing of galaxy clusters and rotation curves of supernovae [26, 47].

Section 2.1 describes the most important observational evidences of DM. The SM of particle physics contains no suitable particle(s) to explain these observations. Proposed candidates for DM are introduced in Section 2.2. One of the preferred candidates are the Weakly Interacting Massive Particles (WIMPs), with a mass around the EW scale. The WIMPs fit the model of a relic DM particle from the early universe, when all particles were in a state of thermal equilibrium [48–50]. This procedure is explained in Section 2.3. Section 2.4 shows the different approaches to search DM. Section 2.5 presents the theoretical highlighted developments in collider DM searches. Section 2.6 describes two different approaches of DM production that have been studied in this thesis. Further models used in DM searches are presented in Section 2.7. Finally, Section 2.8 provides a status of the constraints to DM models in searches at colliders.

2.1 Observational evidence of dark matter

First evidence that suggested the existence of DM came from rotational curves of spiral galaxies. The velocity of the rotation is typically obtained from 21 cm hydrogen

spectral line, and this has been measured for different types of galaxies and galaxies [51]. On the other hand, one can compute the velocity of an galactic object ($v(r)$) using Newtonian dynamics,

$$v(r) = \sqrt{\frac{G_N M(r)}{r}} \quad M(r) = 4\pi \int \rho(r) r^2 dr, \tag{2.1}$$

where r is the distance of visible objects to the galactic centre, $M(r)$ is the mass of the galaxy inside a radius r, $\rho(r)$ is the matter density and G_N is the gravitational constant. In distribution of luminous matter, the circular velocity is expected to fall proportional to $1/\sqrt{r}$ at large distances beyond the galactic disk. However, the measurements show that the speed does not depend of the distance from the galactic centre [52, 53]. Observations propose a $v(r)$ almost constant, thus, Equation 2.1 implies the existence of a spherical halo $M(r) \propto r$, or $\rho(r) \propto r^{-2}$ composed by matter that does not emit light, pointing to the existence of DM.

Besides rotational curves, other effects can not easily be explained without DM. Since the light emitted by distant galaxies travels through space, the path, the size and the cross-section of a light bundle may be affected by all the matter between the light source and the observer. This matter causes distortion of the travel path of the light, producing the so-called, gravitational lensing. Gravitational lensing can be strong [54], resulting in the formation of multiple images of the light-emitting galaxy, or can be weak [55] creating a distortion, stretchering or magnifying the image observed of galaxies. Although difficult to measure for an individual galaxy, galaxies clustered close together will exhibit similar lensing patterns. Both lensing effects can only be produced by gravitational interaction, reaffirming the existence of DM.

The first evidence of DM at extragalactic scale was observed in 1933, when F. Zwicky investigated the Coma cluster [23]. The mass of a galaxy in a cluster can be measured looking at the dispersion velocity of galaxies which compose the cluster. For the Coma cluster, the mass-to-light ratio as a bound system exceeds by an order of magnitude what can be reasonably ascribed to the known stellar populations. Therefore, it was concluded that there was some type of non-luminous matter. This technique was applied to many other clusters [56], with similar results.

Regardless of all observations mentioned above, these do not give a quantification of the amount of DM in the universe. Fortunately, the CMB provides this measurement. Currently the CMB has a nearly uniform temperature around 2.7 K [17]. However, the microwave sky shows that the temperature of the CMB is not exactly the same in all directions, it is anisotropic[1]. The angular power spectrum of the anisotropy of the CMB contains information about the formation of the universe and its content. Recent measurements by the Plank Collaboration [26] determined the density of DM in our universe to be $\Omega_{\rm DM}^{\rm obs}{\rm h}^2 = 0.1187 \pm 0.0017$, while the baryonic matter density is $\Omega_{\rm b}^{\rm obs}{\rm h}^2 = 0.02214 \pm 0.00024$. The baryonic matter contributes only a 26% of the matter in the universe.

Since then, various attempts have been made to modify Newtonian dynamics to accommodate the anomalies observed. Several modifications of General Relativity manage to explain the rotation curves of the galaxies without the need of DM [57–59]. But other observations consistent with the existence of DM can not be explained within these theories [60].

The discovery of gravitational waves in 2016 [61] constrained even more these modified-gravity theories [62–65]. The speed of gravitational waves was measured being the same of the speed of the light within 10^{-15} [66], which is not consistent with the modified-gravity theories prediction. Consequently, nowadays the existence of a new kind of particle as source of DM is the preferred theoretical model.

Although most of their properties are unknown, some characteristics can be extracted from observations. In particular DM can be classified as "cold", "warm", or "hot" according to its velocity (more precisely, its free streaming length). These categories, that refer to velocity, indicate how far a corresponding objects moved due to random motions in the early universe before they slowed down due to cosmic expansion. Current models favour the cold DM scenario, which are mostly in agreement with observations of cosmological large-scale structure.

[1] There are small fluctuations in the temperature across the sky at the level about of 10^{-5}.

Properties of DM candidates

The astrophysical and cosmological observations that support the existence of DM, also provide information about its characteristics. These are the general assumptions about the DM candidates:

- The DM candidates do not interact with light, do not emit, reflect or absorb light of any frequency. This means that DM candidates must have a small enough electromagnetic coupling [67].
- The DM has attractive gravitational interactions and it is either stable or has a lifetime much larger than the lifetime of the universe ($t_0 = 13.08 \pm 0.04$ Gyr) [68].
- The DM has been mostly assumed to be collision-less, indeed the upper limit on DM self-interactions is very large[2] ($\sigma_{self}/m < 1$ cm^2/g) [70].
- The current DM candidates are "cold", which means non-relativistic at the time of freeze-out [71].
- They are non-baryonic [49].

2.2 Dark matter candidates

Several theories provide DM candidates, each with its own particular way of interacting. However, some theories are more likely to be successful than others. One of the first proposed candidates was the Massive Astrophysical Compact Halo objects (MACHOs). These objects include black holes, neutron stars, and brown and white dwarfs, which are composed of ordinary matter, but they emit very little light [72]. Primordial black holes belong to the class of MACHOs objects. They are naturally good DM candidates: they are (nearly) collision-less and stable (if sufficiently massive), they have non-relativistic velocities, and they were created very early in the

[2]The best limits on self-interactions are derived from the Bullet cluster [69], where the DM of the two colliding galaxies had to pass through each other.

universe. Nevertheless, tight limits on their abundance have been set up from various astrophysical and cosmological observations [73]. Latest studies show that it is unlikely that enough of these dark bodies make up the vast amount of DM in the universe [74]. Therefore, the non-baryonic proposal is the favorite to continue to the hunting.

Neutrinos were also good DM candidates, since a beyond of SM physics theoretical model is needed to explain their masses (mentioned in Section 1.5). But, with the non-zero measurement of neutrinos, the expected relic density can not explain the amount of DM in the universe [49]. Other well motivated aspirants to be DM are sterile neutrinos as massive right-handed neutrinos [75]. As their name suggests, sterile neutrinos would not interact with other matter through the weak force, but they would interact gravitationally, making it very difficult to detect them. Nowadays there is no evidence of their existence [76].

Additional well motivated DM candidates are the axions and axion-like particles. This kind of new DM particle named QCD axion [31] is predicted to be light and very weakly coupled. Additionally, axions were introduced to solve the "strong CP problem" (see Section 1.5). At the present time, there are strong constraints on axions and axion-like models, and still nothing points to their existence [77].

Anyhow, the type of DM candidates that have attracted most attention are the WIMPs, they appear for a long time as perfect DM particles, having a mass at the weak-interaction scale, approximately between 10 GeV and 1 TeV. Their existence could produce naturally the relic abundance [49] and conceivably mitigate the egregious hierarchy problem [78]. A brief introduction is shown in Section 2.3.

Since there is no information on the mass of DM particle, the mass range is really large and many theories provide candidates. There are DM candidates as in the Little Higgs models [79, 80], WIMPzilla particles [81], superfluid dark matter [82], extra dimension models [45]. These different mechanics of DM production can cover different mass range values, making them complementary between each other. This is illustrated in Figure 2.1. However, so far the theoretical framework under the DM phenomena, it remains unknown.

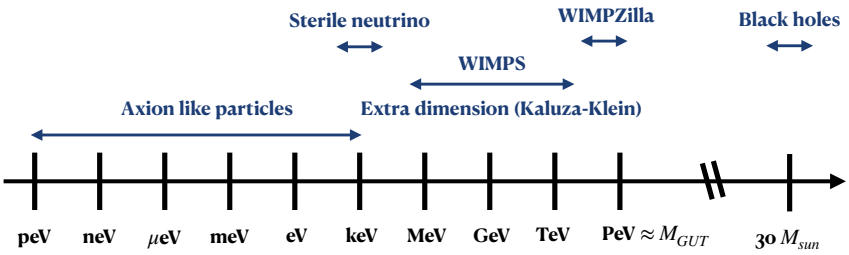


FIGURE 2.1: Representative mass range of the DM candidates. In the figure $\mathrm{M_{GUT}}$ represents the scale where the GUT can be possible.

2.3 The WIMP as DM candidates

The relic abundance of DM points to DM candidates (if a particle) being WIMPs. At an early state of the universe, after the inflation stage, all particles (SM and DM particles) were in a thermal equilibrium "soup". The DM particles suffered self-annihilation within this particle soup. It has been measured that the universe expands, and the rate of this process decreases due to drop in particle density. The density [83] is given by Boltzmann expression,

$$\frac{\mathrm{dn}}{\mathrm{dt}} + 3\mathrm{Hn} = -\langle\sigma\mathrm{v}\rangle(\mathrm{n}^2 - \mathrm{n}_{\mathrm{eq}}^2), \tag{2.2}$$

where in this formula H is the so-called Hubble parameter[3]. The σ is the particle interaction cross-section, v is the velocity and $\mathrm{n_{eq}}$ is the equilibrium number density. At some stage, when the universe was sufficiently large, particles in the soup stopped annihilating and the relic abundance remains, usually known as "freeze-out". This procedure starts when the universe has expanded sufficiently such that the particle density is too low for self-annihilation, that means $\mathrm{H} \sim \langle\sigma\mathrm{v}\rangle\mathrm{n}$. Knowing the relic abundance today allows to estimate σ, although this does not give information on the

[3]The Hubble parameter H is calculated from the scale factor of the universe a (dimension-less), where $\mathrm{H} = \dot{\mathrm{a}}/\mathrm{a}$.

WIMP mass directly. It is defined the subscript f as a freeze-out point and χ denoting the WIMP particle, then the solution of Boltzmann (Equation 2.2), can be written as,

$$n_f \sim (m_\chi T_f)^{3/2} e^{-m_\chi/T_f} \sim \frac{T_f^2}{M_{Pl}\langle\sigma v\rangle n},$$

where m_χ is the WIMP mass, T_f is the temperature and $M_{Pl} = \sqrt{\hbar c/G}$ is the Planck mass. The factor $x_f = -m_\chi/T_f$ is found to be ~ 20 for WIMP masses [84]. The relic density, then becomes,

$$\Omega_\chi = \frac{m_\chi n_0}{\rho_c} = \frac{m_\chi T_0^3}{\rho_c}\frac{n_0}{T_0^3} \sim \frac{m_\chi T_0^3}{\rho_c}\frac{n_f}{T_f^3} \sim \frac{x_f T_0^3}{\rho_c M_{Pl}}\frac{1}{\langle\sigma v\rangle} \tag{2.3}$$

where $\rho_c = \frac{3H^3}{8\pi G}$ is the critical density of the universe (flat space-time) and T_0 is the current age of the universe. This solution is only valid if $m_\chi > T_f$, therefore the particle has to be non-relativistic at the freeze-out. This approximation shows the dependence of the relic density with the annihilation rate at the freeze-out. In fact, the larger (smaller) is the annihilation, the smaller (larger) is the density of DM remaining in the universe today. Figure 2.2 shows this dependence [85]. The most convincing motivation for the WIMPs is the "WIMP miracle", in which a particle with a m_χ on the weak scale ($\mathcal{O}$(GeV) to $\mathcal{O}$(TeV)) with a weak coupling to the SM naturally matches the observed relic density measured in CMB experiments ($\Omega_{DM}^{obs} h^2 \approx 0.12$) [50].

2.4 Searches for dark matter

There are three methods to search for DM candidates, the detection of a WIMP-nucleus interaction by the measurement of a nuclear recoil, the detection of its annihilation or decay products and the production of DM at colliders. Figure 2.3 shows a schematic diagram of these methods [86].

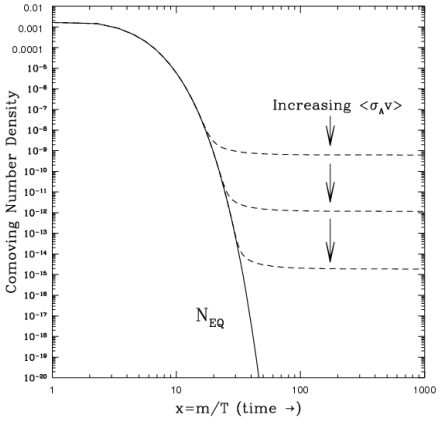


FIGURE 2.2: The number density (with respect to the expansion of the universe) as a function of the DM mass divided by the temperature. The black solid line indicates DM density if equilibrium is maintained. The dashed line specifies when the freeze-out mechanism occurs. Increasing the thermally averaged annihilation rate times velocity $\langle \sigma v \rangle$ leads to higher freeze-out temperature and a lower remaining number density, which is the DM relic abundance.

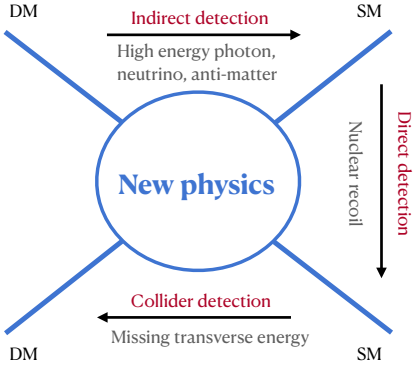


FIGURE 2.3: Schematic diagram to show the relation among the direct detection, indirect detection and collider detection of DM. The arrows indicate the direction of reaction.

2.4.1 Direct detection

Direct searches are based on studying nuclear recoils looking for DM-nucleus interactions. The choice of the target element (i.e. composition) influences the sensitivity of the experiment. There are three methods which can be used: the scintillation, ionisation and heat; producing photons, charge particles and phonons[4], respectively. Some experiments use just one, others use two methods together, since this can offer advantages in background rejection. Prediction of the cross-section for the nucleon-DM interaction are divided into spin-dependent and spin-independent. The spin dependent models correspond to coupling of the type axial-vector while the spin independent are scalar or vector couplings [87]. This division can be studied, since the amplitude entering in the scattering cross-section is the sum of the amplitudes among interactions with all the nucleons (in a detector), separately. If the interactions are spin-independent, the amplitude is proportional to the number of nucleons. However, when a spin-dependent interaction occurs, the amplitude changes sign if the spin of the scattered nucleon is reversed. Since the nucleons in a nucleus have alternating spins, the contribution to the total amplitude from different nucleons will cancel out, leaving a total which depends on the total spin[5].

The most relevant characteristic of direct detection experiment is not the measurement of the recoil energy itself but how the energy is deposited. For instance, the Cryogenic Dark Matter Search (SuperCDMS) experiment [90] separates signal from background using ionisation yields within a germanium and silicon crystal substrates. While the Cryogenic Rare Event Search with Superconducting Thermometers (CRESST) [91] uses scintillating calcium-tungstate crystals. The noble gases xenon (Xe) and argon (Ar) are popular targets, since they are powerful scintillators[6]. Experiments like XENON1T [92] and Large Underground Xenon (LUX) [93] use Xe,

[4]In physics, a phonon is a collective excitation in a periodic, elastic arrangement of atoms or molecules in condensed matter, specifically in solids and some liquids.

[5]This total spin can be the same order obtained from a single nucleus (unpaired nucleon) or can be almost zero. The total cross-section is proportional to the square of the amplitude, so this can be much larger for spin-independent interactions of WIMPS with heavy nuclei. Therefore, direct detection experiments are much less sensitive to spin-dependent interactions than spin-independent (see Figure 2.11) [88, 89].

[6]A scintillator is a material which produces light when particles interact with it.

while Darkside2 [94] uses Ar. On the other hand, the Superheated Drop Detector (SDD) experiment [95] works as a classic bubble chamber. The PICO experiment [96] is an example of a DM particle detector using this technique.

There are important basic assumptions built in the calculations used in direct detection of DM [97]:

- There is a smooth halo of DM particles in our galaxy described by the Maxwell velocity distribution [98].
- Local DM density is assumed [99–101].
- The nucleus is considered as a hard sphere characterised by the Helm form factor [102].

In addition, one characteristic of direct searches is that WIMP-nucleon scattering cross-section can be measured as a function of the WIMP mass.

2.4.2 Indirect detection

Indirect detection experiments look for the annihilation or decay products of DM candidates. To enhance detection reach, usually searches focus on regions where the DM is expected to be denser, such as the galactic centre or dwarf galaxies. The basis of most indirect searches is to look for an excess of either γ-rays or positrons coming from the following annihilation,

$$\chi + \chi \rightarrow \gamma + \gamma,$$

$$\chi + \chi \rightarrow e^{+} + e^{-}.$$

An excess search requires a good understanding of the γ-ray and positron spectra that arrive to the Earth. Neutrinos, like γ-ray, preserve spectral information and point back to the source, making them useful astroparticles for indirect searches. On the other hand, searches for products of DM annihilation and decay in charged cosmic-ray fluxes can be highly sensitive, especially due to low backgrounds for antimatter produced by standard astrophysical processes. In such cases, the measurement of an

excess of anti-proton cosmic rays could provide a unique and sensitive signature of DM annihilation or decay [103].

Different types of detectors are used: air-shower arrays, Cherenkov telescopes, neutrino telescopes or particle detectors in balloons or satellites. There are two sources of possible DM signatures where only neutrino telescopes can search: annihilations in the centre of the Sun and Earth. Only neutrinos will escape unaltered the dense interiors of these objects. Neutrino telescopes have placed limits on DM annihilation cross-sections and spin-dependent WIMP-nucleon scattering. The IceCube [104] and ANTARES [105] experiments, both consisting of arrays of vertical strings of photomultiplier tubes located under the Antarctic ice or the Mediterranean sea, respectively, look for Cherenkov light produced by muons. They have performed searches for neutrinos produced in DM annihilation, that are competitive with direct searches [106].

Things are a bit different in searches for DM candidates at the galactic centre, halo, or other galaxies, since the neutrino, γ or cosmic-ray spectra depend on the thermally averaged product of the DM self-annihilation cross-section times the DM velocity, $\langle \sigma v \rangle$, and the halo profile chosen [87].

Notable results from high energy cosmic rays were found by the Alpha Magnetic Spectrometer (AMS) experiment, which measured an increase in positron fraction measurement up to about 200 GeV [107] and an excess in anti-proton to proton ratio [108]. To accommodate these results, the mass of the DM particle had to be within 0.5 and 1 TeV [109, 110].

Direct and indirect DM searches are affected by large experimental uncertainties in the initial state. These are often of astrophysical nature, such as velocity and density distributions of the DM in the universe, reaching equal sensitivity [111]. A summary of the exclusion contours for (vector and axial-vector) DM-mediator mass using direct and indirect searches is given in Section 2.8.

2.4.3 Detection at colliders

Searches for WIMP DM candidates have been performed at several high-energy colliders, such as the Large Electron Positron Collider (LEP) at CERN [112, 113] and

the Tevatron at Fermilab [114]. Currently the LHC is operating at CERN, which provides the greatest sensitivity and allows access to the highest energy scales for new phenomena including DM [115]. ATLAS and CMS experiments at CERN are multipurpose experiments with similar physics programme: the measurement of Higgs boson properties, precise measurements of the SM and searches for physics beyond the SM (BSM), including the nature of DM. More detailed descriptions of the LHC accelerator and ATLAS experiment are given in the next chapter.

Direct production of DM is an extremely rare processes at colliders. It can not be directly measured, hence a mediator particle is needed, which connects the DM sector with SM particles (e.g. a quark, a photon, a vector boson, etc.). The experimental signature for DM searches, includes an imbalance in the transverse[7] momentum ($\mathrm{p_T}$) measured in the detector, since the DM candidate escapes without leaving any signal in the detector. This final state is motivated by the assumption that proton–proton (pp) collisions produce two WIMPs, which remain undetected. Since the DM is not detected they are often searched in association of a single object in the detector, which happens in many models. These signatures are called mono-X signatures [115], with X being a visible SM particle that recoils against missing transverse momentum originated from undetected DM particle. Seccion 2.5 describes the models of WIMP DM that are investigated at hadron colliders.

2.5 Phenomenology for hadron collider searches

Many BSM theories require new particles at the EW scale which are either DM candidates themselves or might couple to DM candidates. This section shows some frameworks implemented to look for DM production at colliders. One of them is the Effective Field Theory (EFT), which allows to look for new physics without precise knowledge of the full theory behind the process. Other frameworks, such as simplified models, are constructed with a minimal an extension to the SM, adding only a couple of new hypothetical BSM particles to the SM. In addition, simplified models

[7]The term transverse refers to the transverse plane to the direction of the beam pipe. The ATLAS coordinate system is detailed in Section 3.2.1.

retain contact with ultra-violet (UV) completions[8], which reduce to the simplified models in some particular low-energy limit.

2.5.1 Effective field theories

The EFT operators can be applied to a vast number of models, and they provide meaningful predictions. However, there are several constraints on the validity of EFTs. An important caveat is that since EFTs are non-renormalisable field theories, they will break down at larger energies, leading to non-physical results. This occurs typically when the energy scales probed are of the order of the mass of the mediator particle. At this point, integrals involved in the computation of the observable quantities diverge [116, 117]. The advantage of the EFT approach is that each operator and energy scale describe a range of processes, depending on the direction of the arrow of time in Figure 2.4. As shown in this figure, DM annihilation, scattering, and production can all be described by the same operator. This process shows a fermionic DM particle χ and SM quark q that are coupled via a scalar interaction, and the strength of the interaction is given by an energy scale $\mathrm{M_{eft}}$.

The EFT framework (for DM searches) in the LHC can describe the effects of heavy particles (or "mediators") in the low-energy theory where these particles have been integrated out. But the LHC accelerator delivers scattering events at energies so high, that they may directly produce the mediator itself, and in this case the EFT description fails. The EFT analyses remain a useful tool for LHC searches, but this simple fact calls for a careful and consistent use of the EFT. Where the EFT description has to be implemented in the range of its validity [118].

2.5.2 Simplified models

All processes involving effective operators can be replaced by so-called "simplified models". These models are designed such they are applicable to the full range of energies reached by the LHC [116, 117, 119, 120]. These models can be described by a set of parameters: the coupling to the SM ($\mathrm{g_q}$) particles between DM sector ($\mathrm{g_\chi}$),

[8] A more general high energy theory which is well-defined at arbitrarily high energies.

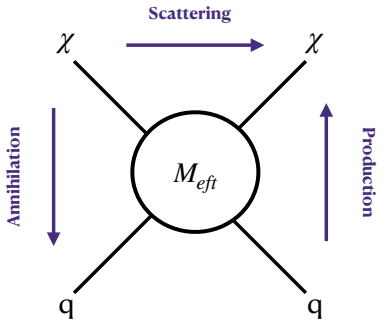


FIGURE 2.4: Representative EFT interaction between DM sector and the SM.

the mass of the mediator (m_{med}), the mass of the DM candidate (m_χ) and the width of the mediator (Γ_{med})[9]. Figure 2.5 shows simplified Feynman diagrams for the t- and s-channel production within simplified models [111].

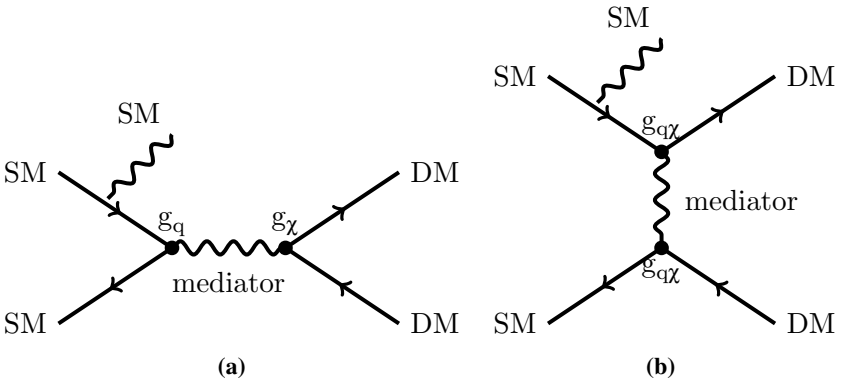


FIGURE 2.5: Simplified DM model for the **(a)** *s*-channel production and the **(b)** *t*-channel production [121].

In the *s*-channel DM production, the mediator is usually a boson, a vector or axial-vector for spin 1, or scalar- and pseudo-scalar for spin 0 interactions. If the

[9]This value of the width of the mediator assumes the minimum decay width, which implies that the width of the mediator is not a single parameter, but it is obtained from the other four parameters to allow only decays to the minimal set of particles specified in a given model.

mediator has spin 1, then it can be equally capable of coupling to all quarks. If its spin is 0, then the coupling is proportional to the Yukawa coupling [122]. For the LHC searches, the selection of the coupling values was recommended by the LHC Dark Matter Working Group (DMWG) [122] in 2015. They were chosen to be g_q = 1 for spin 1 mediators, g_q = 0.25 to spin 0 mediators, and g_χ coupling (between the DM and the mediator) was elected to be 1.

If the mediator couples only to a quark and a DM particle, a coloured exchanging particle is required, and it is mediated via a *t*-channel process as shown in Figure 2.5(**b**). These couplings can be universal with regard to the colour and generations, but can violate universality with preferential couplings to the first two generations [111, 123]. Another possibility is that the DM carries a flavour index, and the strongest couplings might occur to third generation particles (top-flavoured DM).

2.5.3 Two-Higgs-doublet models

The simplest possible extensions of the SM is the two-Higgs-doublet model (2HDM) [124]. The 2HDM is one of the simplest low-energy effective Higgs models and could be the result of various UV-complete BSM theories (like the Minimal Supersymmetric Standard Model (MSSM)). One motivation of 2HDM comes from axion models, where the strong CP problem can be sorted out by imposing a global $\mathrm{U}(1)$ Peccei-Quinn symmetry that allows to rotate away the CP violating phase allowed in the kinetic Lagrangian of QCD. If one sticks to the SM matter content, the introduction of a second Higgs doublet is required in order to solve the strong CP problem [125].

The 2HDM is generated by adding to the complex scalar doublet, Φ_1 of the SM, another doublet, Φ_2, which alters the dynamics of EW symmetry-breaking. The model has 8 degrees of freedom, three of them give rise to the masses for the EW gauge bosons, and the remaining five manifest themselves as physical states. There are different types of 2HDM theories depending on which type of fermions couples to which doublet Φ [126].

In the Type I 2HDM models all quarks and charged leptons are coupled only to second (Φ_2) scalar doublet (fermiophobic) [126]. The type II 2HDM is one the most

studied, since the couplings of the MSSM are a subset of the couplings of this type [127, 128], where up-type quarks couple to Φ_2 and down-type quarks couple to Φ_1.

Such models can be described in terms of the following physical parameters [129]: mass eigenstates for two CP-even neutral Higgs bosons (m_h, m_H, where $m_h < m_H$), one CP-odd neutral Higgs boson (m_A) as well as a pair of charged ones ($m_{H^\pm}$), the ratio of Higgs doublets ($\tan\beta$) VEV, the mixing angle of the two doublet (α), the mixing angle which diagonalises the mass-squared matrices and the quadrilinear couplings to the scalar doublets with the a boson ($\lambda_{P1}, \lambda_{P2}$ and λ_3).

A special case occurs in the the alignment limit i.e. $\cos(\beta-\alpha) \to 0$, where α represents the mixing angle between the two CP-even weak spin-0 eigenstates. In the alignment limit, the lighter CP even Higgs boson h has couplings exactly like the SM-Higgs boson. In another hand, when $\sin(\beta-\alpha) \to 0$, the heavier CP even boson, i.e. H is SM-like, leaving h to be lighter than the discovered Higgs boson.

2.6 Dark matter search in the top-quark sector

This thesis focuses on mono-top signature models (i.e. mono-X with X being a top quark). Two different approaches for mono-top production are studied. The first one is given for the EFT description (Section 2.6.1), another under the 2HDM framework (Section 2.6.2). In the following sections a picture of these mono-top productions are described.

2.6.1 Non resonant and resonant DM production models in the LHC

The non-resonant mono-top signal is produced via a flavour-changing neutral interaction as shown in Figure 2.6, where a right top quark, a light-flavour up-type quark and an exotic massive vector-like particle V can be parameterised through a general Lagrangian [130, 131]:

$$\mathscr{L}_{\text{int}} = a V_\mu \bar{u}\gamma^\mu P_R t + g_\chi V_\mu \bar{\chi}\gamma^\mu \chi + \text{h.c.},$$

the V mediator is coupled to a pair of DM particles (represented by Dirac fermions $\chi\bar{\chi}$) whose strength can be controlled through the parameter g_χ and where P_R represents the right-handed chirality projector. The parameter a stands for the coupling constant between the massive invisible vector-like boson V, and the t- and u-quarks, and γ^μ are the Dirac matrices. A detailed description of further assumptions presented in these benchmark models can be found in Refs. [130, 132].

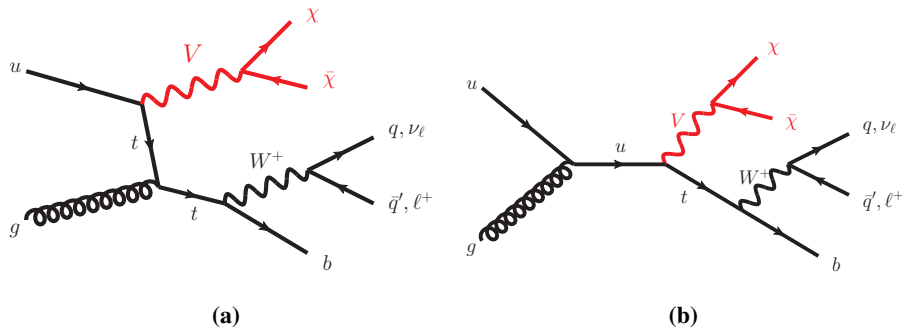


FIGURE 2.6: Representative leading-order Feynman diagrams corresponding to the mono-top signals searched for non-resonant **(a)** *t*- and **(b)** *s*-channel DM production in association with a top-quark.

Another possibility is the resonant case, corresponding to the production of a coloured charge-2/3 scalar (ϕ) decaying into a top quark and a spin-1/2 DM particle (χ) [133]. This process, represented in Figure 2.7, is described by the following Lagrangian [130, 131]:

$$\mathscr{L}_{\text{int}} = \lambda\phi\bar{d}^{c}P_{R}s + y\phi\bar{\chi}P_{R}t + \text{h.c.},$$

where the parameters λ and y represent the couplings of the charged scalar to the d- and s-quarks and to the top quark and the DM particle χ, respectively.

The mono-top signature can be divided in two topologies depending on the decay of the W boson associated to the top-quark decays: if the W decays into a lepton and a neutrino, it is called leptonic channel, or when the W boson decays into hadrons receiving the name of hadronic channel. In this thesis, the topology studied is the leptonic channel.

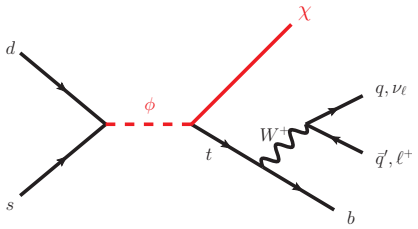


FIGURE 2.7: Resonant production of a coloured scalar ϕ that decays into a DM particle and a top-quark

In Ref. [134] it was observed that the contribution to the search of the resonant DM model, the leptonic channel is negligible compared with the hadronic channel. Therefore, this work only considers the non-resonant DM signal model (analysis developed in Chapter 7).

2.6.2 2HDM type II plus a pseudo-scalar mediator

In the analysis presented in Chapter 6 in this thesis, the model considered is type II 2HDM with the additional pseudo-scalar mediator to DM (2HDM+a) [135]. This category of models represents one of the simplest UV complete and renormalisable theories with a spin 0 mediator [114, 122, 135–141].

The alignment ($\cos(\beta-\alpha) = 0$) and decoupling limit is assumed. Such that the lightest CP-even state of the Higgs-sector, h, can be identified as the Higgs boson. And the EW VEV is set to 246 GeV, i.e. the same value as in the SM.

The additional pseudo-scalar mediator of the model, a, couples the DM to the SM particles and mixes with the CP-odd pseudo-scalar A. The masses of the heavy CP-even Higgs boson, H, and charged bosons, $\mathrm{H}^{\pm}$, are set equal to the mass of the heavy CP-odd partner $\mathrm{m_A}$, and $\lambda_{\mathrm{P1,2,3}} = 3$. This mass splitting constraints and quadrilinear coupling values are imposed by EW precision measurements [135, 142].

In addition, unitary couplings ($\mathrm{y}_\chi = 1$) between the a and the DM particle χ is considered. The DM particle mass is set to 10 GeV.

The mono-top production mode is covered for three possible channels, the *t*-channel, *s*-channel and tW generations. The mono-top signal cross-section is enhanced by the on-shell production[10] of the $\mathrm{H}^{\pm}$ boson. The resonant enhancement is larger for the tW production, which cross-section dominates the mono-top production (especially for lower values of $\mathrm{m}_{\mathrm{H}^{\pm}}$). These processes are shown in Figure 2.8.

For the *t*-channel processes, the main contribution comes from the two middle diagrams shown in Figures 2.8(**c**)-(**d**). These diagrams interfere destructively[11], and the amount of interference decreases with increasing $\mathrm{H}^{\pm}$ mass, increasing the cross-section at high masses of $\mathrm{H}^{\pm}$. The two dominant *s*-channel production modes are shown in Figure 2.8(**e**)-(**f**). These last two processes are studied in this thesis in Chapter 6 and in Appendix B.

[10] In quantum field theory, virtual particles are named off-shell, since they do not satisfy the energy–momentum relation; real exchange particles do satisfy this relation and are termed on-shell.

[11] The observed destructive interference ensures perturbative unitarity of the process.

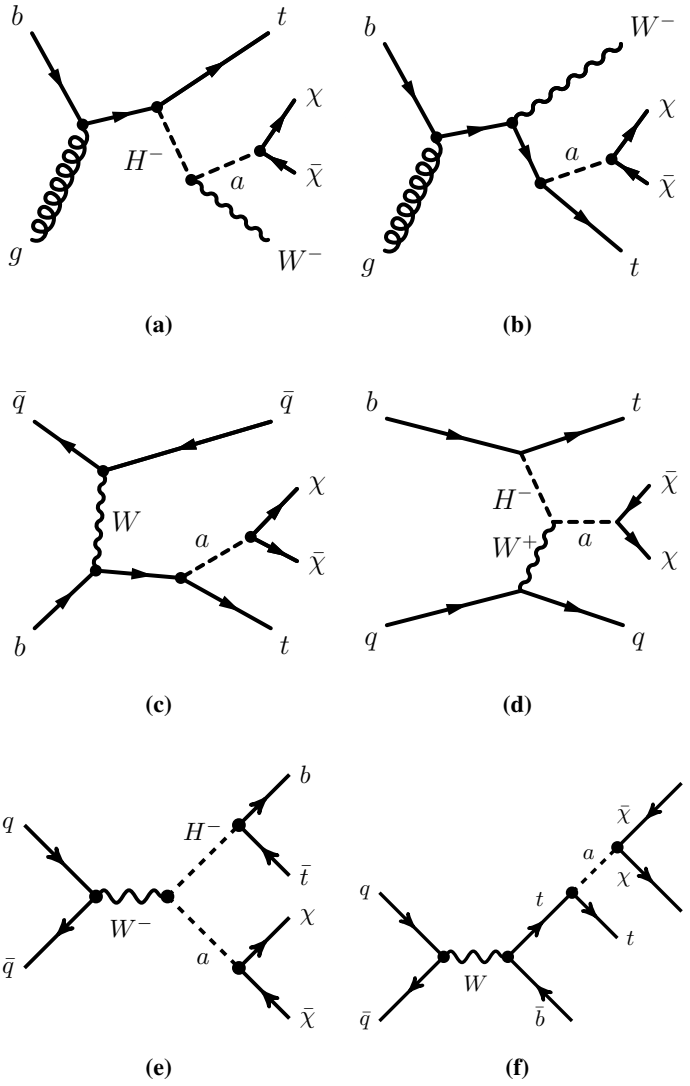


FIGURE 2.8: Schematic representation of the dominant production and decay modes for the 2HDM+a model: tW: **(a)**-**(b)**, *t*-channel: **(c)**-**(d)**, and *s*-channel: **(e)**-**(f)** in single-top-quark final states.

2.7 Other searches

Additionally, a known SM particle, like Higgs boson, could couple directly to DM particles, these models have been also studied, often but not exclusively, associated to the Higgs boson [143–145]. In these models the particle DM candidate interacts only through their couplings with the Higgs sector of the theory, the so-called Higgs-portal models. The Higgs-portal interaction can be elementary, compatible with a dark sector that consists of the DM candidate only [144, 145]. If the DM candidate has a mass equal or less than half of the Higgs boson mass, then the Higss boson mediator can decay into pairs of DM particles, opening the chance to study this signature [146, 147]. Also, if the DM mass is higher than the Higgs boson mass, its production becomes off-shell and searches in this region are challenging [148]. Searches at colliders have looked for a mediator [149, 150], which might also decays into SM particles, usually mostly to quarks and gluons [151–153].

The lack of evidence in DM collider searches motivates the reorientation to the study of models where non-prompt signature can be possible and to investigate the impact on DM searches [154–156]. A field gaining attraction is the search for long-lived particles in the context of DM. Good progress has been achieved to introduce a new production mode [157, 158], dark sectors, non-E_T^{miss}signatures [159, 160] and dark-photon searches [161, 162].

2.8 Summary search results

So far, none of the searches for DM candidates described in this chapter have shown any sign of a DM signal. Results are then interpreted as exclusion limits (i.e. upper limits at a 95% confidence level) on the different signal models [163]. Constraints on selected mediator-based DM models using the ATLAS detector during the years 2015-2016 are summarised in Ref. [164].

Figure 2.9 shows the exclusion contours in the DM-mediator mass plane with two representative scenarios: a leptophobic (when quark couplings of the vector mediator are much larger than the lepton couplings) vector mediator scenario and a leptophilic (with non-vanishing coupling to leptons) vector mediator scenario. The same studies,

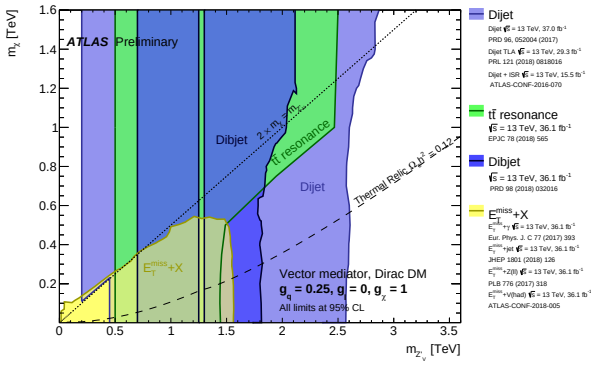
but for axial-vector mediator case, are provided in Figure 2.10. For each scenario, thermal relic curve[12] indicates combinations of DM candidate(s) and mediator mass that are consistent with a DM density. Models of DM production with vector or axial-vector mediators are mainly constrained by dijet resonance searches and associate production of DM and objects from initial state radiation (ISR). The very strong limits obtained from the dijet resonance searches are due to the particular choice of the universal quark couplings used as benchmark.

Comparison between results from searches for DM at colliders with indirect and direct detection experiments can be computed by transforming the vector and axial-vector mediator limits on the spin-independent DM-proton and spin-dependent DM-nucleon scattering cross-section as a function of the DM mass, respectively. Figure 2.11 presents a comparison of the limits obtained by the ATLAS detector and direct detection experiments. As can be seen, the exclusion contours for vector and axial-vector neutral mediator models are complementary to the ones obtained by ATLAS and direct detection. The ATLAS searches have a better sensitivity for low values of the DM mass, since direct experiments lose sensitivity[13]. It is important to notice, that the results of the interpretation in both the spin-dependent and spin-independent cross-sections have the dependence on the mediator mass and the couplings selection, thus the comparison is only valid for the conditions of the specific model and couplings.

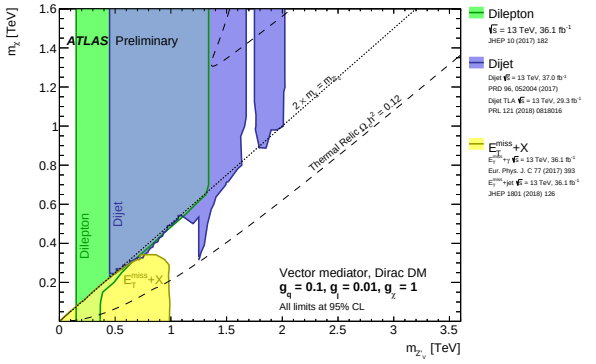
For the 2HDM+a model, the constraints from searches in ATLAS are presented for the first time in Ref. [164]. As visible in the results presented in Figure 2.12, the exclusion sensitivity is vastly dominated by the $\mathrm{h}(\mathrm{b\bar{b}})$+$\mathrm{E_T^{miss}}$ and $\mathrm{Z}(\ell\ell)$+$\mathrm{E_T^{miss}}$ results. It is important to investigate the sensitivity in terms of $\tan(\beta)$, the ratio of Higgs doublets VEV. Figure 2.12 shows the contour limit on this parameter. Although the exclusion obtained is dominated also in this case by the $\mathrm{h}(\mathrm{b\bar{b}})$+$\mathrm{E_T^{miss}}$ and

[12]Thermal and non-thermal relics have a different relationship between their relic abundance and their properties such as mass and couplings, so the distinction is especially important for DM detection efforts. For example, the WIMP class of particles can be defined as those particles which are created thermally, which we call only as relic density. For instance, DM axions come mostly from non-thermal processes.

[13]The nuclear recoils produced by low DM masses are currently challenging to detect as the recoil energies are below current detector thresholds [165].

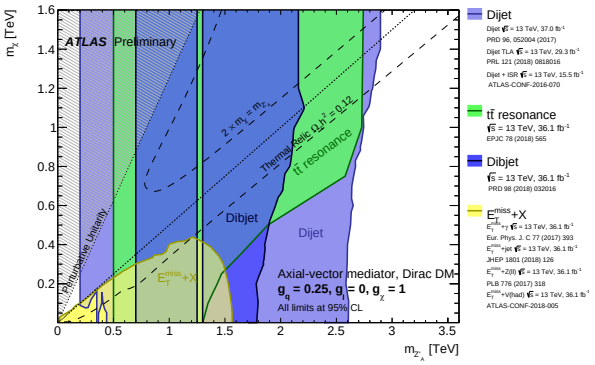


(a)

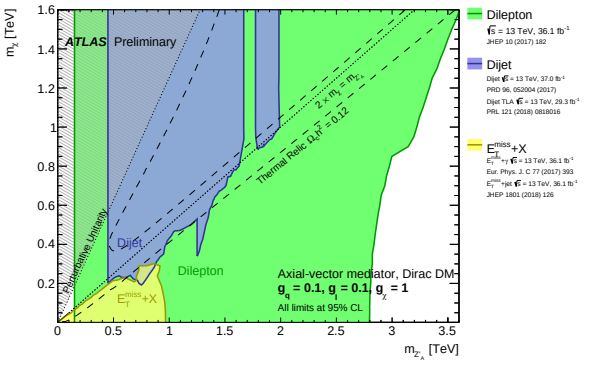


(b)

FIGURE 2.9: Regions in a DM mass-mediator mass plane excluded at 95% CL by the dijet, dilepton and mono-X searches within ATLAS experiment, for vector mediator simplified models with two values of couplings, leptophobic scenario **(a)** and leptophilic scenario **(b)**. Exclusions limits were obtained for DM coupling g_χ=1, a quark coupling g_q universal to all flavors, finally the lepton coupling g_ℓ, noted the plots. The dashed curve of indicates the thermal relic, i.e. the combinations of DM and mediator mass that are consistent with a DM density measure. A dotted curve indicates the kinematic threshold where the mediator can decay on-shell into DM. Excluded regions that are in tension with the perturbative unitary considerations are indicated by shading in the upper left corner [164].



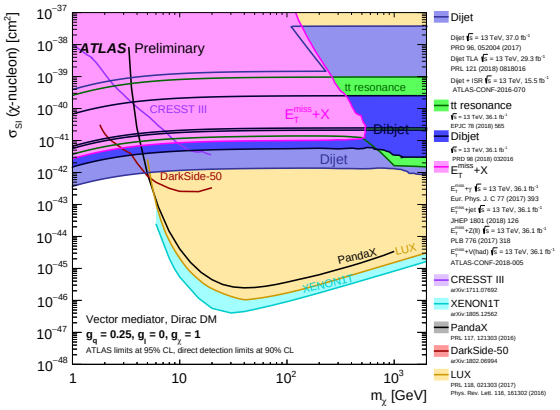
(a)



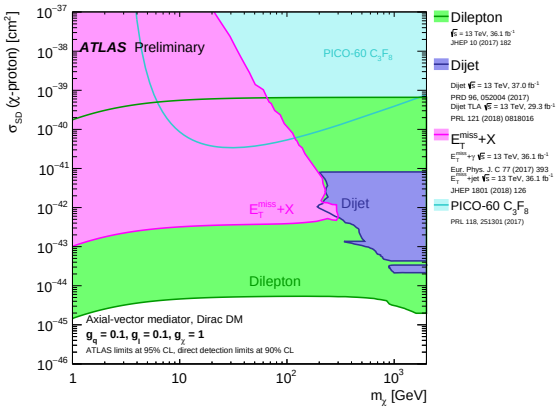
(b)

FIGURE 2.10: Regions in a DM mass-mediator mass plane excluded at 95% CL by the dijet, dilepton and mono-X searches, within ATLAS experiment, for axial-vector mediator simplified models with two values of couplings, leptophobic scenario **(a)** and leptophilic scenario **(b)**. Exclusions limits were obtained for DM coupling g_χ=1, a quark coupling g_q universal to all flavors, finally the lepton coupling g_ℓ, noted the plots. The dashed curve of indicates the thermal relic, i.e. the combinations of DM and mediator mass that are consistent with a DM density measure. A dotted curve indicates the kinematic threshold where the mediator can decay on-shell into DM. Excluded regions that are in tension with the perturbative unitary considerations are indicated by shading in the upper left corner [164].

$Z(\ell\ell)$+E_T^{miss} results, two additional signatures $t\bar{t}$+E_T^{miss} and $t\bar{t}t\bar{t}$ signatures can contribute at $\tan(\beta) \sim 0.5$.

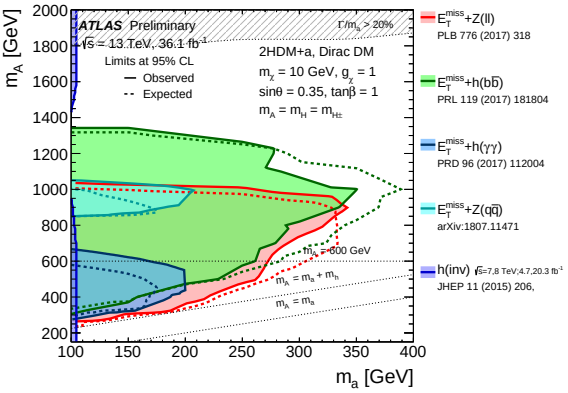


(a)

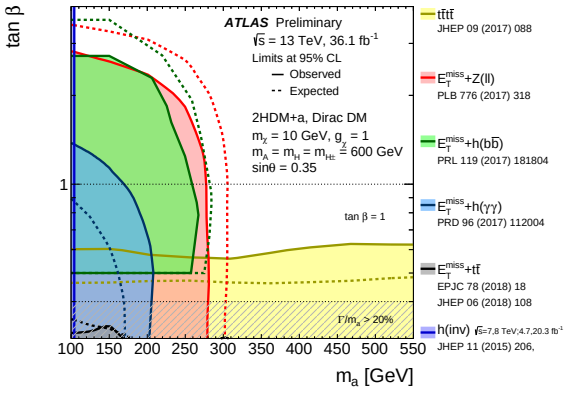


(b)

FIGURE 2.11: Comparison of the inferred limits to the constraints from direct detection experiments ((**a**): CRESST III, DarkSide-50, XENON1T, LUX and PandaX, (**b**): PICO-60 C_3 F_8) on the spin-dependent WIMP-proton scattering cross-section in the context of the vector leptophobic model (**a**). and on the spin-independent WIMP-nucleon scattering cross-section in the context of the axial-vector leptophilic model (**b**). ATLAS limits are shown at 95% CL and direct detection limits at 90% CL. ATLAS searches and direct detection experiments exclude the shaded areas. Exclusions beyond the canvas are not implied in the ATLAS results. The dijet and mono-X exclusion regions represent the union of exclusions from all analyses of that type [164].



(a)



(b)

FIGURE 2.12: Regions in a (m_a,m_A) **(a)** and (m_a,tan(β)) **(b)** planes excluded from data at 95% CL by mono-X and four top analyses, within ATLAS experiment. The dashed grey regions at the top of **(a)** and the bottom of **(b)** indicate the region where the width of any of the Higgs bosons exceeds 20% of its mass [164].

CHAPTER 3

The LHC and the ATLAS detector

"Somewhere, something incredible is waiting to be known."
- Carl Sagan

The analysis presented in this thesis is performed using *pp* collision data recorded by the ATLAS detector, one of the four main detectors located at the LHC ring. Section 3.1 gives an overview of the LHC accelerator complex with a brief summary of the experiments. Section 3.2 presents a description of the ATLAS detector, their components and its purpose during the recording of the huge number of particles produced in *pp* collisions. Section 3.3 describes the main upgrades of the ATLAS detector for the High-Luminosity LHC (HL-LHC) stage.

3.1 The Large Hadron Collider

The LHC is a two-ring-superconducting-hadron accelerator [166] located at the CERN complex. The LHC is the largest and the highest energy particle accelerator in the world. It is located in a tunnel with a circumference of 26.7 km where the Large Electron-Positron collider (LEP) [167] was previously housed. The tunnel is just outside Geneva, on the border between France and Switzerland. The LHC is composed of superconducting magnets and radiofrequency cavities that bend and accelerate, respectively, the beams of protons (or lead ions) to 99.9% the speed of light. Protons are accelerated up to an energy of 6.5 TeV[1].

The LHC injection chain makes use of a staged accelerator systems [166]. The protons for the beams are extracted from a tank of ionised hydrogen gas and linearly accelerated to momenta of 50 MeV by the Linear Accelerator-2 (LINAC-2)[2]

[1] During the Run 2, period among 2015 to 2018 data-taking.
[2] Replaced by LINAC-4 after Run 2.

[168, 169]. Afterwards, the protons are injected into the first circular accelerator, the so-called Proton Synchrotron Booster (BOOSTER) with 630 m of circumference, where beams reach an energy of 1.4 GeV. Then, these protons go into the second circular accelerator, the Proton Synchrotron (PS), rising the beams up to 25 GeV. They continue their way into the third circular accelerator, the Super Proton Synchrotron (SPS) with 7 km of circumference, where they increase their energy up to 450 GeV. Finally, the beams enter into the LHC by two different 2 km-long tunnels. At this last stage, the LHC uses its radiofrequency cavities to accelerate the proton beams to up 6.5 TeV. Figure 3.1 shows a schematic of the CERN accelerator complex [170].

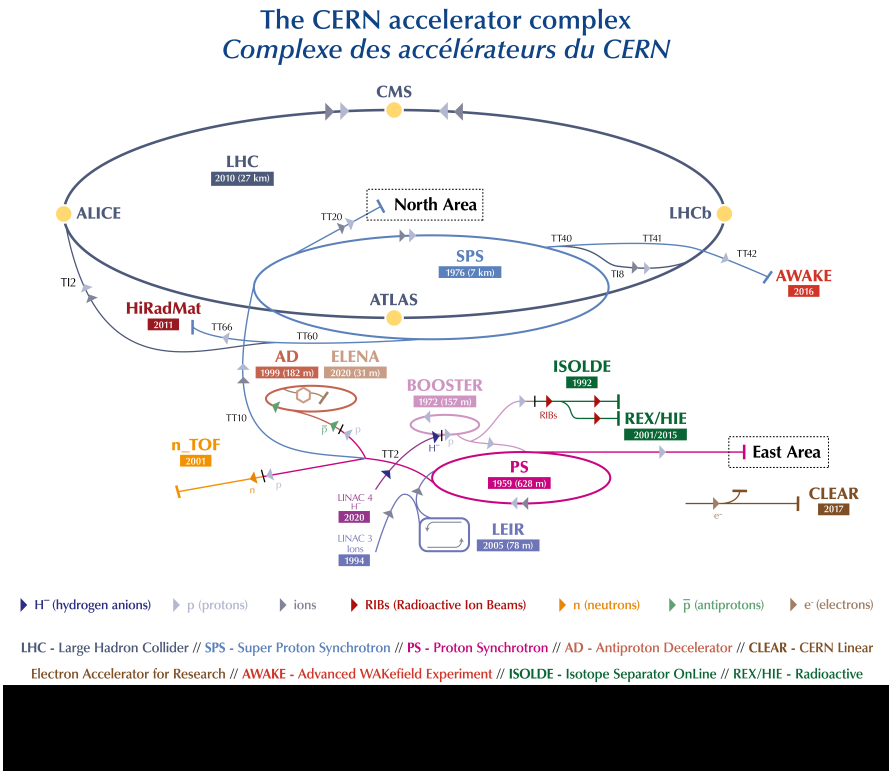


FIGURE 3.1: A schematic of the CERN accelerator complex.

Protons collide in seven points, and in these points the different experiment are located. There are four main experiments on the LHC ring: ATLAS, CMS, Large

Hadron Collider beauty (LHCb) [171], and A Large Ion Collider Experiment (ALICE) [171]. The ATLAS and CMS experiments are multipurpose detectors designed to have sensitivity to a wide variety of final-state objects and topologies, to study the SM EW interactions, the Higgs boson production, soft QCD and possible signatures of BSM physics. The LHCb experiment is an asymmetric detector constructed to study the decay of B-hadrons and the matter-antimatter composition of the universe. The ALICE experiment studies the primordial soup of quark-gluon plasma in the early universe.

Besides the four big LHC experiments, there are smaller experiments located next to the main interaction points. They are:

- Large Hadron Collider forward (LHCf) is a special-purpose experiment for astroparticle (cosmic ray) physics [172]. It consists of two detectors placed at zero degree collision angle $\pm$ 140 m from ATLAS interaction point to measure very forward photons and neutral pions, as source to simulate cosmic rays in laboratory conditions.

- Total Elastic and diffractive cross-section Measurement (TOTEM) [173] which aim is the measurement of the total pp cross-section with a luminosity-independent method and the study of elastic and diffractive scattering at the LHC. It is a forward detector located near CMS.

- Monopole and Exotics Detector at the LHC (MoEDAL) [174] is located inside the LHCb cavern and it specialises in the search for magnetic monopoles or dyons[3] and other highly ionising stable massive particles and pseudo-stable massive particles.

As mentioned above, the LHC accelerator provides high-energy pp collisions and the different detectors record the result of these collisions. Anyhow, in order to deeply analyse these recorded pp collisions and constrain our knowledge of the SM and its phenomenology, these collisions have to be understood from the theoretical point of view. Thus, a summary of the phenomenology of pp collisions in the LHC is

[3] In physics, a dyon is a hypothetical particle with both electric and magnetic charges.

given in Section 3.1.1. Since the deep inelastic scattering experiments at the Stanford Linear Accelerator Center (SLAC) in 1968 [175], it is well known that the proton is not a point-like particle and it is indeed formed by quarks, antiquarks and gluons. In Section 3.1.2, the glorious complexity of the proton structure is discussed and different concepts such as the production cross-section, parton distribution functions as well as the underlying event are introduced. Finally, in Section 3.1.3 and 3.1.4, the terms luminosity and pile-up are described, respectively.

3.1.1 Phenomenology of *pp* collisions

The total *pp* cross-section at $\sqrt{s}$ = 13 TeV is approximately 110 mb [176], which can be divided in luminosity-independent elastic (31 mb) and inelastic (79 mb) cross-section. Elastic scattering of the protons and diffractive events are observed by the detectors. However, only the inelastic scattering generates particles at sufficient high angles (with respect to the beam axis) such that these can be seen in the detectors.

Scattering processes at hadron colliders such as the LHC are classified as either soft or hard. QCD is the underlying theory for all such processes, but the approach (and the level of understanding) is very different for the two cases. An important property of strong interactions is asymptotic freedom, since the interaction strength decreases with energy, as already mentioned in Section 1.3. This means that in high energy collisions, quarks and gluons may be treated effectively as free particles, allowing for perturbative QCD calculations. On the contrary, at low energies quarks and gluons interact strongly, forming hadrons such as the proton. A fundamental parameter of QCD is α_S (Equation 1.4) which indicates the effective strength of the strong interaction in the process and that depends on the energy scale Q, typically the momentum transfer in the given process. Most of the interactions are soft, i.e. long distance collisions between the two protons, in which the transverse momentum exchanged between both is so small that most of the energy escapes along the beam pipe. The rates and properties of these soft interactions are dominated by non-perturbative QCD effects. The most interesting events are hard interactions, in

which the inelastic scattering between the two protons can be treated as direct parton–parton[4] scattering. In this case, the rates and event properties can be predicted with precision using perturbation theory.

3.1.2 The proton structure

Protons are a bound state of partons (quarks and gluons), its constituent quarks (valence quarks) are *uud*. In addition to these quarks, the proton has also the so-called sea quarks, coming from gluon splitting into quark and antiquark pairs. The momentum of the proton is shared among all these constituents and its structure depends on the energy scale at which it is probed. For instance, at lower energies ($Q \sim 1$ GeV), the momentum of the proton is mainly distributed among the three valence quarks. When energy increases ($1 < Q \leq 100$ GeV), the emission of gluons is more probable and these gluons carry some of the initial momentum of the quarks. The proton dynamics can be understood in terms of the fraction of the proton momentum distributed among its constituent partons. Formally, it is expressed as a probability distribution, called a parton distribution function (PDF). The interacting partons carry only a fraction x_1 and x_2 of this momentum; therefore, the centre-of-mass energy of the interaction ($\hat{s}$) is only a fraction of the total energy (s): $\hat{s} = x_1 x_2 s$.

3.1.2.1 The cross-section for processes in hadron-hadron collisions

The space-time picture of a *pp* collision is shown in Figure 3.2, where can be separated the long (i.e. non-perturbative QCD) and short (i.e. hard parton-level interactions, perturbative QCD) distance physics processes. The factorisation theorem, first proposed by Drell and Yan [178] obtains (for hadronic collisions) the cross-section of a hard scattering process. It is divided in a partonic cross-section (short distance), which is process-dependent and calculable in pertubative QCD, and a universal part corresponding to the distribution of partons inside the colliding hadrons (long distance), provided by the PDFs.

[4] In particle physics, the parton model is a model of hadrons, such as protons and neutrons, introduced by R. Feynman in one of his (informal) talks [177], for interpreting the cascades of radiation produced by QCD processes and interactions in high-energy collisions.

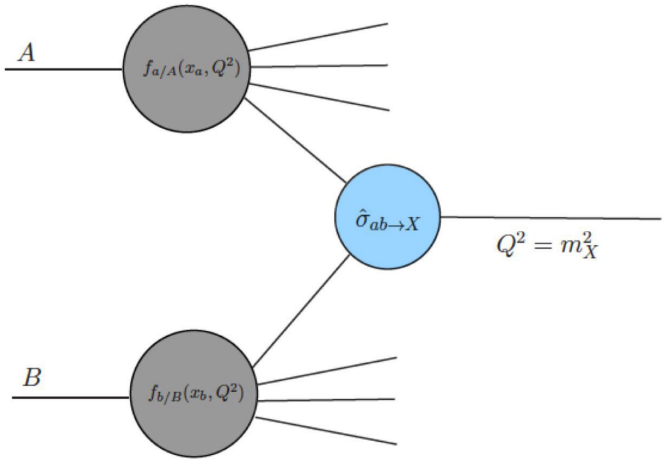


FIGURE 3.2: Schematic diagram for particle production in *pp* collisions.

As mentioned before, the PDFs give the probability that the parton (a) carries a fraction of the momentum of the hadron $\mathrm{x_a}$:

$$\mathrm{f_a(x_a, \mu_F^2)},$$

and depends on the factorisation scale μ_F^2 which separates long and short distance physics, i.e. above that scale one can rely on perturbative calculations.

The total cross-sections for an X production in *pp* collisions, $\sigma_{\mathrm{pp\to X}}$, can be calculated from the convolution of the cross-section of the incoming partons a and b, $\hat{\sigma}_{\mathrm{ab\to X}}$, and the PDFs for the partons a and b inside the two protons, $\mathrm{f_a}$ and $\mathrm{f_b}$:

$$\sigma_{\mathrm{pp\to X}} = \sum_{\mathrm{ab}} \int \mathrm{dx_a dx_b f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) d\hat{\sigma}_{ab\to X}(x_a, x_b, \mu_F^2)},$$

where,

$$\hat{\sigma}_{\mathrm{ab\to X}} = [\sigma_{\mathrm{LO}}(\mathrm{x_a, x_b}, \mu_F^2) + \alpha_s(\mu_R)\sigma_{\mathrm{NLO}}(\mathrm{x_a, x_b}, \mu_F^2) + \alpha_s(\mu_R)^2\sigma_{\mathrm{NNLO}}(\mathrm{x_a, x_b}, \mu_F^2) + \ldots]_{\mathrm{ab\to X}}.$$

The partonic cross-section $\hat{\sigma}_{ab\to X}$ can be calculated in perturbative QCD and written in terms of leading order (LO), next-to-leading order (NLO), next-to-next-to-leading order (NNLO) and so on processes. Also, two scales are used in the definition of $\hat{\sigma}_{ab\to X}$, the already mentioned factorisation scale μ_F and the renormalisation scale μ_R. It is important to recall that higher-order terms cause logarithmic divergences due to soft and collinear gluon emissions. The renormalisation factor appears to replace the divergent integrals such that large logarithms vanish. Actually the physical cross-section is independent of μ_F and μ_R. But, the truncation of the infinite perturbative series at finite order typically results in a non-negligible scale dependence. To obtain a faithful prediction, higher order corrections are computed until this dependence can be reduced. Usually, both scales are assumed to be equal and are chosen at the order of magnitude of momentum scales of the hard process.

3.1.2.2 Parton distribution functions

The PDFs describe the parton content of the hadrons, and give the probability density for a parton a in the hadron h with a certain longitudinal momentum fraction x_a at a scale μ_F^2. This probability is parametrised as a function of the momentum fraction x_a of a parton in a proton. The PDFs are not possible to extract from analotical calculations, given the nonperturbative nature of QCD. But, can be extracted from global fits to data, mainly deep inelastic scattering (DIS) data, since they are very extensive and precise. Fixed-target lepton-nucleon scattering experiments at SLAC, Fermilab, CERN and from the electron-proton HERA collider at Deutsches Elektronen Synchrotron (DESY) provide the backbone of parton distribution analysis. The lepton-nucleon data include electron, muon and neutrino DIS measurements on hydrogen, deuterium and nuclear targets. The additional physical processes which are used in the fits are:

- The single jet inclusive production in nucleon-nucleon interactions, selecting jets with large transverse energy; this quantity is dependent on the gluon distribution.

- Dilepton production in the virtual photon Drell-Yan process $pN \rightarrow \mu^+\mu^- + X$, which is a probe of the sea quark distribution.
- EW Z and W boson production $pp^- \rightarrow W^+(W^-) + X$ at the Tevatron collider which is sensitive to the up and down quark and antiquark distributions.

3.1.2.3 The Underlying event

The underlying event (UE) [179] is all what is seen in a hadron collider event which is not coming from the primary hard scattering (high energy, high momentum impact) process. After the hard parton–parton interaction, a coloured/beam remnant that did not participate at the hard process itself is left over. So, multiple parton interactions may occur within a collision. Figure 3.3 shows the UE, which is the collection of all the soft processes that go along with the high-p_T interaction of interest, except the hard process itself. It consists of the initial and final state radiation (ISR and FSR) plus beam–beam remnants.

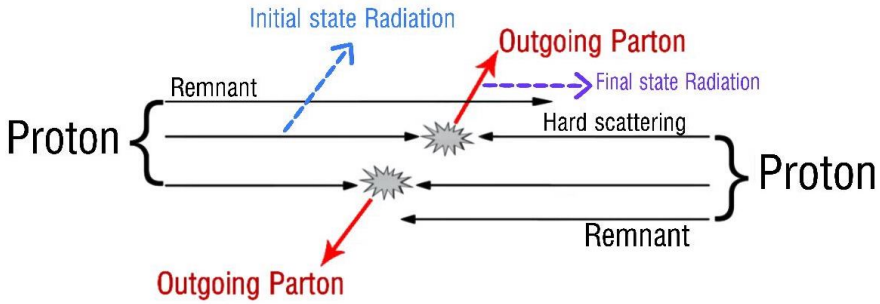


FIGURE 3.3: An illustrative event in hadronic collisions: hard parton–parton scattering. The resulting event contains particles that originate from the two outgoing partons and from the ISR and FSR and particles that come from the break-up of the protons (beam-beam remnants). The UE is everything except the two outgoing hard scattered "jets", i.e. consists of the beam–beam remnants plus ISR and FSR processes.

The modelling of these soft interactions is important because they may impact other high-p_T measurements and also affect the detector resolution. However, these

low-p_T scattering processes are difficult to predict using perturbative methods only. For this reason, predictions of the models are compared to LHC data.

3.1.3 Luminosity

The particle beams usually come in bunches (see Figure 3.4), depending on the number of protons in each bunch, and how the beams collide, the number of interactions inside the detectors can be very different. The instantaneous luminosity ($\mathscr{L}$) is a measurement of the number of (inelastic) collisions that can be produced per second and cm^2. It is defined as follows:

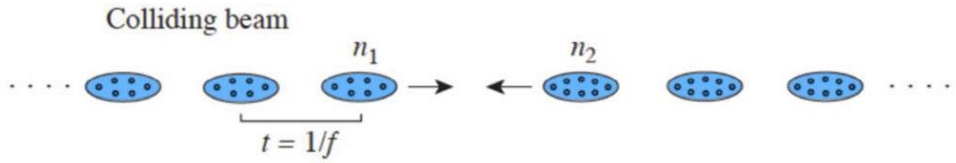


FIGURE 3.4: Colliding beams with a bunch crossing frequency f.

$$\mathscr{L} = \frac{N_1 N_2 f_{rev} k \gamma}{4\pi\varepsilon_n \beta^\star} F, \tag{3.1}$$

where N_1 and N_2 represent the number of protons for each colliding bunch, f_{rev} stands for the revolution frequency of the bunch, the symbol k represents the number of bunches per beam, γ is the Lorentz factor, ε_n is the emittance of the beams, which is determined by the beam injectors, $\beta^\star$ is the beta function[5], and finally F denotes the geometric factor due to the crossing angle of the beams [180]. For the LHC, the nominal values are N_1, N_2 equal to 10^{11} protons per bunch, $k = 2{,}808$ bunches and revolution frequency (f_{rev}) of 40 MHz [166].

The LHC is designed for a value of the centre-of-mass energy of *pp* collisions of $\sqrt{s}$ = 14 TeV at an instantaneous luminosity of $1 \times 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$, but the nice performance of the LHC, allowed the luminosity to reach up to $2 \times 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ for

[5]Or amount of "squeeze" of the beam.

the first time in October 2017, during Run 2 [181]. The integration of Equation 3.1 over time T gives the total amount of luminosity acquired. i.e.:

$$\mathscr{L}_{\text{total}} = \int_0^{\text{T}} \mathscr{L}\, \text{dt}.$$

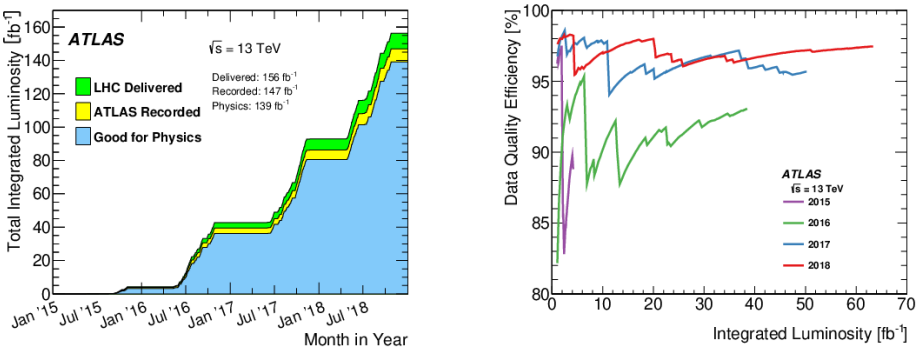
The integrated luminosity of the data delivered to ATLAS by the LHC between the time at which the LHC declares "stable beams"[6] and the time when sensitive subsystems are switched off to allow a beam dump or beam studies is typically defined as the "delivered" luminosity. The subset of these data that are recorded by ATLAS is defined as the "recorded" luminosity. Combining the availability of all subsystems, ATLAS achieved an average data-taking efficiency of 95.6% during Run 2. The ATLAS experiment collected 3.2 fb^{-1}, 32.9 fb^{-1}, 44.3 fb^{-1} and 58.5 fb^{-1}, from 2015 to 2018. For the whole Run 2 at $\sqrt{\text{s}}$ = 13 TeV, this corresponds to 139 fb^{-1} of data, certified as being good for physics analysis. The luminosity collected and the data taking efficiency of the ATLAS detector during Run 2 campaigns [181] are both shown in Figure 3.5.

3.1.4 The Pile-up effect

Pile-up is a challenge matter among detectors and for the acquisition and analysis of the data. In particle physics, pile-up is called to the situation where the detector is being affected by several events at the same time. Due to the large number of protons in colliding bunches at the LHC, it is not unlikely that multiple independent hard interactions occur during one bunch crossing, called in time pile-up. This is equivalent to the mean number of interactions (μ) and it increases with increasing instantaneous luminosity:

$$\mu = \frac{\text{N}_1\text{N}_2}{4\pi\sigma_{\text{x}}\sigma_{\text{y}}}\sigma_{\text{inel}}, \tag{3.2}$$

[6] Data-taking for physics begins as soon as possible after the LHC declares "stable beams", a condition indicating that stable particle collisions have been achieved.



(a) **(b)**

FIGURE 3.5: The cumulative integrated luminosity **(a)** delivered to and recorded by ATLAS between 2015 and 2018 during stable beam *pp* collision data-taking at $\sqrt{s}$ = 13 TeV is shown. The cumulative data quality efficiency versus total integrated luminosity **(b)** delivered to the ATLAS experiment between 2015 and 2018 is also shown.

where N_1 and N_2 represent the number of protons for each colliding bunch, σ_x and σ_y are the Gaussian widths in the horizontal and vertical plane of the bunch respectively and σ_{inel} the *pp* inelastic cross section. The mean number of interactions per crossing corresponds the mean of the Poisson distribution on the number of interactions per crossing calculated for each bunch. In addition, the spacing between the bunches is shorter than the response time of the detectors, so additional collisions from different bunches are recorded simultaneously, referred to as out of time pile-up.

3.2 The ATLAS detector

The ATLAS detector is a multi-purpose particle detector installed at one of the four collision points along the LHC. The measurements and searches performed by ATLAS consider a large variety of final-state particles and topologies. The particles have to be identified and reconstructed and therefore, the detector must have excellent efficiency, granularity, acceptance and resolution. The ATLAS detector is composed of several layers, each of them designed to measure signals left by specific types of

particles. It is disposed in a cylindrical manner, symmetric around the beam axis and the collision point. All these layers build up an accurate and elaborate picture of the collisions that are taking place, identifying the different types of the particles in the event, their trajectories and their electric charges.

Figure 3.6 shows the layout of the ATLAS detector [182]. The particles are produced at the centre and pass through several detector subsystems, which are described in the following sections. The dimensions of the ATLAS detector are 46 m in length, 25 m in height and a weight of about 7000 tonnes.

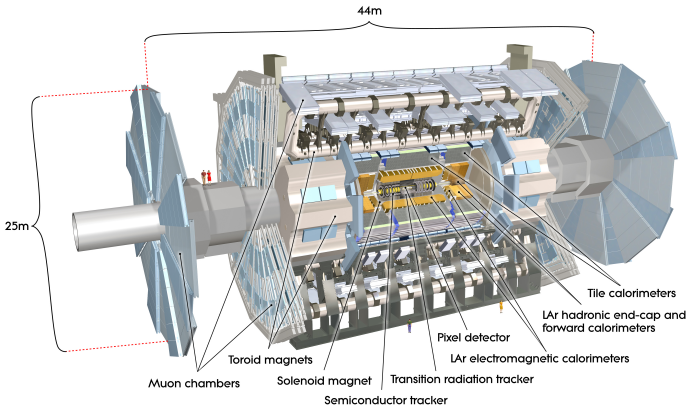


FIGURE 3.6: Layout and dimensions of ATLAS.

The ATLAS detector was designed to record *pp* collision data at high energy and luminosity. The detector has several requirements to accomplish:

- High accuracy measurements of the energy of photons, charged leptons, jets and missing transverse energy.
- High momentum resolution and reconstruction efficiency of charged-particle tracks through the detector.
- Identification of particles within a wide angular coverage with a high granularity in the detector.

- An efficient trigger system to be able to quickly select event data of interest for physics analyses from the collisions.

3.2.1 Coordinate system

The detector is cylindrically symmetric around the beam axis and is forward-backward symmetric with respect to the collision point. Consequently, the ATLAS detector employs a right-handed coordinate system with its origin at the nominal interaction point and the z-axis along the beam pipe. The x-y plane is defined orthogonally to the beam, where the positive x-direction points toward the centre of the LHC and the y-axis points upwards. In the cylindrical coordinates, the azimuthal angle in the transverse plane is named ϕ and it is the angle around the z-axis, where $\phi = 0$ corresponds to pointing along the +x-axis. The polar angle taken from beamline is called θ. These coordinates allow to define the transverse plane, i.e. the x-y plane. The transverse momentum of a particle of four-momentum, $p^{\mu} = (E/c, \vec{p})$, is $p_T = \sqrt{p_x^2 + p_y^2}$. For convenience the ATLAS detector uses instead of the angle θ, the pseudorapidity (η) or the rapidity (y). For a particle with a four-momentum p^{μ} and an angle θ, η can be written as:

$$\eta = -\ln\left(\tan\left(\theta/2\right)\right).$$

This quantity is Lorentz invariant for boosts along the beam pipe, and it is preferred instead of the angle θ. The pseudorapidity η can be expressed in function of the particle momentum as well,

$$\eta = \frac{1}{2}\ln\left(\frac{p + p_z}{p - p_z}\right),$$

where $|\vec{p}| = p$, and p_z is the longitudinal momentum of the particle. The rapidity y is defined as,

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right),$$

with E being the energy the particle. In the ultra-relativistic limit, $\mathrm{p} \gg \mathrm{m}$, this is approximated the η. The utility of the η compared to the y in high energy physics is that the former induces geometrical quantities without knowing any particle property. Angular separations along the η-ϕ plane are measured by the quantity $\Delta\mathrm{R}$, where two objects are separated by $\Delta\phi$ and $\Delta\eta$ (or Δy),

$$\Delta\mathrm{R} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2},$$

$$\Delta\mathrm{R}^y = \sqrt{(\Delta y)^2 + (\Delta\phi)^2}.$$

3.2.2 Inner detector

The Inner Detector (ID) [183, 184], the subdetector closest to the interaction point, composes the tracking system of the ATLAS detector. The primary aim of the ID is to precisely measure trajectories, also often referred to as "tracks", of charged particles in the region closest to the beam pipe. The ID is immersed in a 2 T magnetic field[7] produced by a superconducting solenoid and it detects charged particles with $|\eta| < 2.5$. The ID has 3 subsystems, the Pixel detector (including the Insertable B-layer (IBL) [185]), the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The parts closest to the beam pipe have finer resolution. Figure 3.7 shows the ID subsystems. Table 3.1 summarises the characteristic of each subdetector.

The IBL is the innermost layer of the Pixel detectors. It is located at a radius of 3.3 cm from the beam axis. This detector is built to cope with high radiation and high occupancy and to reconstruct charged-particle tracks efficiently and precisely as well as for the identification of heavy-flavour jets [185, 187]. It contains 14 staves which cover the region $|\eta| < 3.0$ with over 12 million silicon pixels with a typical size of 50 µm in r-ϕ and 250 µm in z each. The staves distribution ensures a full hermetic coverage in ϕ , with an overlapping angle between the staves of 1.82 rad. Each stave is instrumented with 12 two-chip silicon planar modules, covering the

[7] The magnitude of the Earth's magnetic field, at its surface, ranges from 25 to 65 µT. The ATLAS magnetic system is about 40.000 times larger.

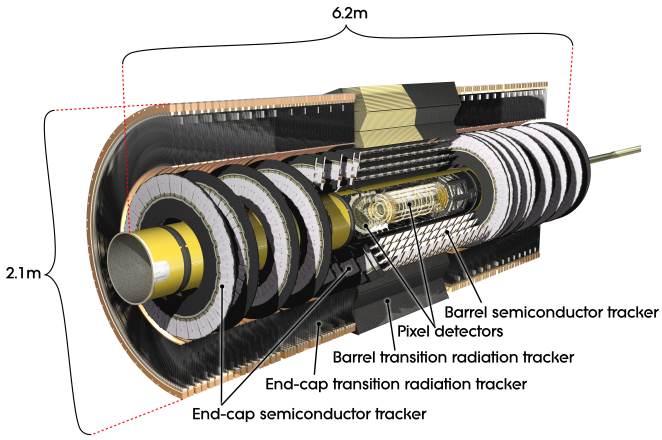


FIGURE 3.7: The ATLAS Inner Detector [186].

region of $|\eta| < 2.7$, and 8 single chip modules with silicon 3D sensors, four at each end of the stave. The expected hit resolution is 8 μm in r-ϕ and 40 μm in z [188].

Subdetector	Element size [μm]	Intrinsic resolution [μm]
IBL	50×250	8×40
Pixel	50×400	10×115
SCT	80	17×580
TRT	4000	130

TABLE 3.1: Summary of the main characteristics of the ID subdetectors. The intrinsic resolution of the IBL, the Pixel and the SCT is reported along r-ϕ and z, while for TRT is only along r-ϕ plane.

The Pixel [189] and SCT [190] detectors employ silicon pixels and strips, respectively. These detectors record hits corresponding to the particles crossing the sensors. Charged particles crossing them excite the electrons in the conduction band of the silicon. The barrel region of the Pixel detector contains three layers of silicon sensors at radii of 5.05, 8.85, and 12.25 cm. The forward regions of the detector consist of three disks per side. The pixels have a size of 50×400 μm^2, what translates in a resolution of the Pixel detector of 10 μm in the transverse direction, and 115 μm in the longitudinal direction. The closeness of the Pixel detector to the interaction point

together with its high resolution enables it to reconstruct the interaction vertices very precisely.

The SCT detector offers additional precision tracking. It has a similar geometry to the Pixel detector, consisting of 4088 silicon strip modules, arranged in four concentric barrels (with 2112 modules) and two endcaps of nine disks each (with 988 modules per endcap). The SCT detector is located in a radial region of 299–560 mm. The intrinsic resolution of the SCT modules is 17 µm in r-ϕ plane and 580 µm in z. The resolution of the SCT detector is lower than in the Pixel detector, as it is located further from the interaction point, and its spacial resolution is obtained by combining information from the two sides of the modules. The barrel layers consist of silicon detector units of 780 readout strips each with a pitch of 80 µm, while the endcap disk detector units are similar but with a tapered geometry.

Both the Pixel and SCT silicon modules are cooled to $\sim -10^{\circ}$C to reduce noise from thermal excitations [190].

The outermost layer of the ID is the TRT [191]. It employs 4 mm diameter drift tubes (polyimide straws) providing tracking coverage and electron identification information up to $|\eta| < 2.0$. They are nominally filled with a 70%-27%-3% gas mixture of xenon-carbon dioxide-oxygen gas, embedded with polypropylene fibres which detect the transition radiation generated by charged particles as they move between the materials. Particles going through the detector ionise gas molecules and induce movement of charges towards electrodes. The subsystems provide a resolution of 130 µm and a discrimination power between electrons and charged hadrons. The barrel of the TRT stretches from 563 to 1066 mm in z with respect to the interaction point, while the endcap covers 848 to 2710 mm in r. Additionally to the particle identification, the TRT improves momentum resolution as track curvature can be constrained over a larger path length.

3.2.3 Calorimeters

The calorimeters are situated outside the solenoidal magnet that surrounds the ID. They consist of two main types of calorimeters, one for electromagnetic energy measurements and one for hadronic energy measurements. They are composed of multiple layers of dense materials in order to absorb the particles flowing out from the interaction point. The incoming particles create showers[8] which are initiated as these particles pass through layers of absorber material inducing an output signal proportional to the input energy of the shower. The calorimeter system has a pseudorapidity coverage of $|\eta| < 4.9$. Figure 3.8 shows a cut-away view of the calorimeter systems.

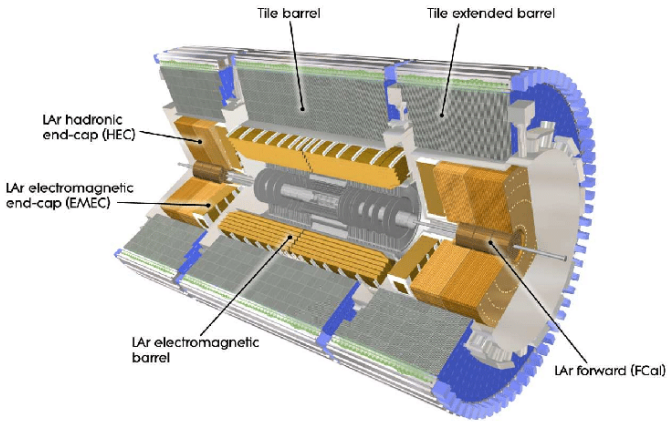


FIGURE 3.8: Cut-away view of the ATLAS Calorimeters [192].

The electromagnetic barrel and two endcaps (EMEC) calorimeters [191] are positioned closest to the beam pipe. They use liquid Argon (LAr) as the active medium to detect the showers and to measure the energy of electrons and photons within $|\eta| < 3.2$. The barrel covers $|\eta| < 1.475$ and it consists of three layers with lead (Pb) absorbers arranged in an accordion-like folding pattern which allows to have a full azimuth in acceptance. This layer has a significant role to identify collimated photons coming from neutral pion decays. When a high energy electron (or positron)

[8] A shower is a cascade of secondary particles produced as the result of a high-energy particle interacting with dense matter.

crosses the material, which is used as an absorber due to its heavy nucleus, it emits photons through bremsstrahlung processes. These photons convert into e^+e^- pairs, creating further interaction with the absorber material (a shower). If the energy of these particles is less than the critical energy required to create additional particles (which occurs at around 10 MeV), showering process stops. The LAr was selected due to its radiation hardness and linear response with energy. The EMECs contain four wheels, that are each divided into eight wedges shaped modules and cover the range $1.375 < |\eta| < 3.2$ with a three-layer Pb-LAr.

The hadronic calorimeters are designed to measure and absorb the energy of outgoing hadrons. They consist on the Tile calorimeter (TileCal) [193] which is made of three layers and two wheels in the endcap, with a coverage of $|\eta| < 4.9$. The TileCal is made from alternating layers of plate absorbers, and plastic scintillator tiles, as the active material in the barrel region, $|\eta| < 1.5$. Copper (Cu) with LAr are the active materials in the endcap where $1.5 < |\eta| < 3.2$, then Cu and Tungsten (W) absorbers with LAr in the forward region, $3.2 < |\eta| < 4.9$. The forward region is essential for measuring forward jets and to compute the total transverse momentum of all objects produced from a hard scatter event.

To measure the energy deposited by a particle passing through the detector, the active material converts the energy of the incoming hadrons into electromagnetic radiation, which is converted into electrical signals by photomultiplier tubes. Hadronic showers differ from the electromagnetic ones since the former ones decay through the strong interaction, and produce showers of mostly composed of pions, around a third of these pions decay electromagnetically. The hadronic showers are typically much wider than electromagnetic showers.

3.2.4 Muon Spectrometer

Since muons behave like minimum ionising particles, they typically deposit only a few GeV of energy in the calorimeters before escaping from them, and produce tracks in the ID. The outermost layers of the ATLAS detector are dedicated to detect and measure the energy of muons together with the ID. The Muon Spectrometer (MS) covers the region of $|\eta| < 2.7$. It is split between a barrel region extending to $|\eta| <$

1.05 and the endcap magnets ($1.4 < |\eta| < 2.7$). The MS detector is composed by a tracking and a triggering systems. The system is subject to a magnetic field produced by a toroid which bends the muons allowing to calculate the transverse momentum of the muon with 3% precision. Figure 3.9 shows the MS system.

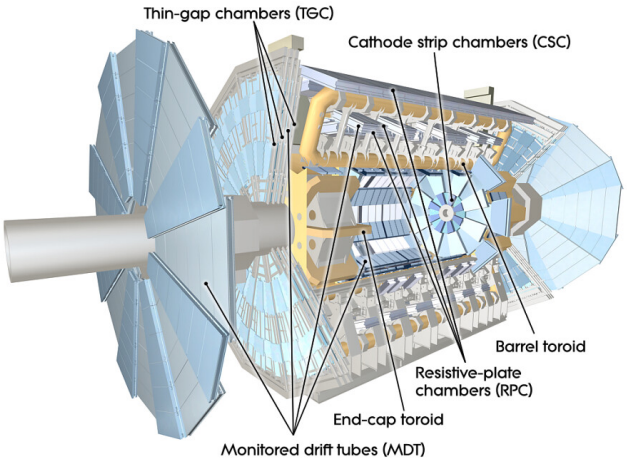


FIGURE 3.9: Cut-away view of the muon detection systems within the ATLAS detector [194].

The precision tracking chambers are mostly monitoring drift tubes (MDTs) with coverage of $|\eta| < 2.7$. The MDT chambers are arranged in three layers along the trajectory of the track, which allow the determination of the momentum from the sagitta[9] of the curvature of the track in the magnetic field. In the barrel part of the detector the three layers form coaxial cylinders. In the endcap part they form large circular disks centered at the beam axis. These are made of 30 mm tubes filled with Ar (93 %) and CO_2 (7 %). Cathode strip chambers (CSCs) are used to handle higher muon fluxes in the forward regions, but neither of these are adequated for triggering purposes due to long drift times in both components. To this end, the trigger chambers are a combination of resistive-plate chambers (RPCs) at $|\eta| < 1.05$ and thin-gap

[9] In geometry, the sagitta of a circular arc is the distance from the center of the arc to the center of its base.

chambers (TGCs) in the forward region at $1.05 < |\eta| < 2.4$, and they are similar to the CSCs, but they have smaller time resolutions.

3.2.5 Trigger and data acquisition

The LHC accelerator collides bunches of protons every 25 ns (40 MHz). However, it is not possible to process and store the large amount of information produced at this rate, so it is only possible to record a small fraction of those events. Additionally, the vast majority of collisions produce events that are of no interest to the physics being investigated. For each bunch crossing, the trigger system [195] uses a combination of fast electronics and algorithms to make real-time decision to filter interesting events for offline analysis. The trigger system is composed by three levels called the level-1 (L1), level-2 (L2) and event filter (EF). The L2 and EF are collectively referred to as the high-level trigger (HLT). The L1 is applied at hardware level and can reduce the total rate by a factor of 200. It checks the high momentum objects or missing energy based on low granularity inputs from all calorimeters and the dedicated muon chambers. The decision is reached within 2.5 μs of the bunch crossing and triggers the information from the front-end boards to be sent from the detector. The L1 trigger system manages the dead-time to limit the rate of events which are interesting. This system tries to avoid overlapping readouts or overflowing in data buffers. Events which "survive" the L1 trigger, are buffered in the Read-Out System (ROS) where will be processed by the HLT system.

The HLT is a software level trigger that does finer tracking in regions of interest to reconstruct physics quantities, taking the events in the ROS and reducing the incoming event rate from about 100 kHz to roughly 1.5 kHz. These data are then sent to LHC Computing Grid. This is a distributed computing facility, with servers located around the world, that processes collision data and it is also used for analyses by ATLAS Collaboration members. An overview of the trigger and Data Acquisition systems (DAQ) system is shown in Figure 3.10.

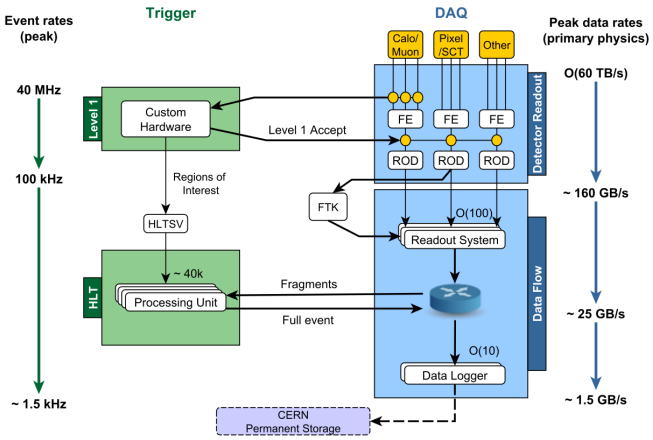


FIGURE 3.10: The ATLAS DAQ System in LHC Run 2. Events passing the L1 hardware trigger (top left) are passed to the HLT (bottom left) via the farm supervisor node (HLTSV), now including assembly of Regions of Interest. Simultaneously to this, event data from the detector front-end electronics systems are sent to the ROS via optical links from the Readout Drivers (RODs) in response to a L1 trigger accept signal. These data are then buffered in the ROS and made available for sampling by algorithms running in the HLT. Once the HLT accepts an event it is sent to permanent storage via the Data Logger [195].

3.3 Upgraded ATLAS detector at the HL-LHC

The HL-LHC is currently expected to begin its operations at the second half of 2027, with a nominal instantaneous luminosity of 7.5×10^{34} cm^{-2} s^{-1} at $\sqrt{s} = 14$ TeV. The expected average number of inelastic *pp* collisions per bunch crossing (pile-up) is as high as 200. This will be significantly higher than the average pile-up of 50 archived during 2018 data-taking at 2.1×10^{34} cm^{-2} s^{-1} (more details are given in Section 4.5), so pile-up is one of the biggest concerns at the HL-LHC. This programme aims to provide a total integrated luminosity of 3000 fb^{-1} by 2037. Upgrades of the ATLAS detector will be necessary to maintain its performance in this expected higher luminosity environment and to mitigate the impact of radiation damage and detector ageing. A brief description of the upgrades outlined in the ATLAS detector at the

HL-LHC are given in the following:

Inner Tracker

The ATLAS Inner Tracker [196] will be completely replaced in order to provide excellent tracking to face the high pile-up environment expected at the HL-LHC. The new silicon-only design (referred to as "ITk") will achieve improved momentum resolution for reconstructed tracks and extend the $|\eta|$ coverage from $|\eta| < 2.5$ to $|\eta| < 4.0$ with a lower material budget than in Run 2. The performance of the ITk will be as good, and in most cases better, than the existing ID in an environment with significantly higher overlapping events. The resulting ITk layout is showed in Figure 3.11. A silicon pixel detector composed of 5 barrel layers will be placed closest to the beamline. A silicon strip detector with 4 barrel layers will extend tracking out to higher radii. A series of rings will extend coverage to the forward region [197, 198].

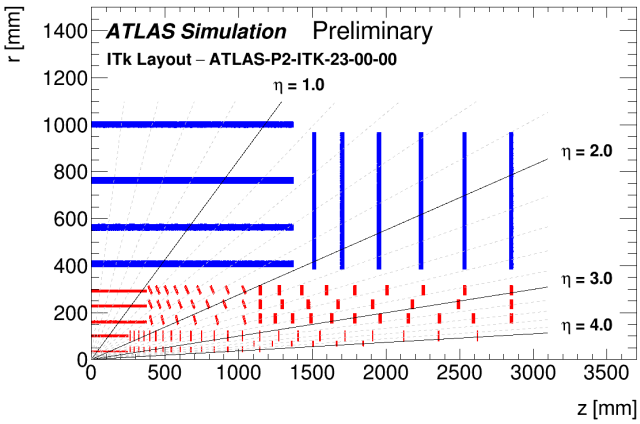


FIGURE 3.11: The schematic depiction of the ITk Layout, including the updated layout for the Pixel detector. Only one quadrant and only active detector elements are shown. The active elements of the strip detector are shown in blue, and those of the pixel detector are shown in red. The horizontal axis is along the beam line with zero being the interaction point. The vertical axis is the radius measured from the interaction region [196].

Calorimeters

The ATLAS LAr Calorimeter will have entirely new frontend and readout electronics optimised to withstand radiation conditions. The electronics architecture is designed to output full-granularity digitised signals at 40 MHz. These upgrades will combat the HL-LHC conditions with active pile-up correction techniques using nearby bunch crossings to maintain an excellent energy resolution over a wide dynamic range [199]. The ATLAS TileCal will use new frontend and readout electronics, power supplies, and optical link interface boards to withstand increased radiation conditions [200].

High-Granularity Timing Detector

The ATLAS High-Granularity Timing Detector (HGTD), which will precisely measure the timings of charged particles, will be installed covering $2.4 < |\eta| < 4.0$ in front of the LAr calorimeter to reduce background from pile-up jets, as the increased pile-up expected in high-luminosity running will require additional mitigation strategies. A timing resolution of 30 ps for minimum-ionising particles is expected [201].

Muon Spectrometer

A large fraction of the ATLAS MS frontend and on- and off-detector readout and trigger electronics will be replaced to enable higher trigger rates and longer latencies. Additional muon chambers will be installed to maintain muon identification and reconstruction performance, increase trigger acceptance, and suppress the rate of random coincidences. The possibility to extend the muon acceptance to $|\eta| < 4$ is still under study (high-η tagger), although most performance results presented to date for HL- LHC studies do not yet take possible improvements from this extension into account in their projections [202].

Trigger and Data Acquisition

The trigger and data acquisition systems will be improved to preserve high signal acceptance in the high-rate and high-occupancy HL-LHC environment [203]. The

improvements will include higher bandwidth readout using high granularity measurements and tracking information earlier in the trigger. The hardware-based first-level[10] trigger acceptance rate is planned to be 400-1000 kHz, while the software-based HLT acceptance rate will be 10 kHz, i.e. an increase of about a factor 10 compared to the HLT at the current ATLAS detector. The b-jet identification efficiency and light-flavour-quark rejection of the projected ATLAS detector at the HL-LHC is expected to be similar to that of the current detector while the c-jet rejection is expected to be about a factor of two lower than that of the Run 2 detector [204].

[10]The expression "first-level trigger" is used here to refer generically to the earliest trigger level, which is the L1 trigger in the current ATLAS system and will be the Level-0 trigger in the Phase-II upgrade.

CHAPTER 4

Data and simulated events

All models are wrong, but some are useful.
- George Box

As already mentioned in the previous chapter, over the course of Run 2, the ATLAS detector recorded an unprecedented amount of LHC *pp* collision data at $\sqrt{s}$ = 13 TeV. These data are exploited to test the SM as well as also other BSM models and theories, to discover new particles, and to explore nature at the smallest scales with the most powerful accelerator in the world.

Unfortunately, understanding the final states of such high-energy particle collisions is an extremely challenging theoretical problem. Typically hundreds of particles are produced, and in most processes of interest their momenta range over many orders of magnitude. All the particle species of SM, and maybe some beyond, are involved. A very convenient way of obtaining theoretical predictions is to use Monte Carlo (MC) event generators. These are essential components of almost all experimental analyses and they are widely used by theorists and experiments to make predictions, comparisons to data and even preparations for future experiments. Basically, the simulation of events of a given physics process starts with the calculation of its parton-level objects at a fixed order in perturbation theory, and follows with the simulation of the parton cascades and non-perturbative effects, such as the hadronisation and multiple parton scatterings. Finally they must be integrated over a final-state phase space of huge and variable dimension in order to obtain predictions of experimental observables. Putting all these elements together, one has MC event generators capable of simulating a wide range of the most interesting SM processes that are expected at the LHC or BSM processes that may occur, which can be used for several distinct purposes in particle physics experiments.

A *pp* collision event[1] can be a real event, i.e. an actually recorded event, or an event that is simulated by a MC event generator. Anyhow, independently of the nature of the events, these are processed by a common ATLAS software, called ATHENA [205]. If real data are processed, the detector response is evaluated through different algorithms (pattern recognition, track fitting, energy measurement, etc) in order to reconstruct final-state objects. If simulation data are processed, MC generated particles are transported through the detector material and magnetic field, and the response of the detector, including its electronics, is simulated. Once this is done, exactly the same algorithms mentioned for data are used in order to reconstruct the simulated final-state objects.

This section is organised as follows. Section 4.1. describes the data recorded by the ATLAS detector from 2015 to 2018. Then, the basis of MC simulations is discussed in Section 4.2. Several MC event generators are employed to produce signal and background processes, and their features are described in Section 4.3. Section 4.4 is devoted to the ATLAS simulation and reconstruction framework. Section 4.5 and Section 4.6 describe these simulations for the Run 2 analysis and for the HL-LHC analysis, respectively. Finally, Section 4.7 is dedicated to discuss how the choice of the MC generator and its set of tunes[2] can affect the kinematic properties and the reconstruction efficiency of one of the main backgrounds of the two analyses presented in this thesis, the top-quark pair ($\mathrm{t\bar{t}}$) production. The implementation of higher-order QCD corrections in $\mathrm{t\bar{t}}$ is important for the understanding and subtraction of backgrounds in searches for new physics along many analyses [206]. To study $\mathrm{t\bar{t}}$ production, MC simulation samples are employed and evaluated. These processes can produce a b-quark for each top-quark decay, and a c-quark from W boson decays. This section shows the studies of heavy flavour hadron production and decay properties, among different MC generators.

[1] Indeed, an event can be a *pp* collision, or a single particle traversing the detector, e.g. a cosmic muon or a testbeam event.

[2] Accordingly each model presents several free parameters which must be optimised to produce a reasonable description of measured observables. This optimisation process is known as tuning, and the resulting parameter sets are referred to as MC generator tunes.

4.1 Data event samples

The analysed data in the analysis presented in Chapter 6 are part of the 25 ns *pp* collisions delivered by the LHC from 2015 to 2018 at $\sqrt{s} = 13$ TeV and collected by the ATLAS detector. Events are selected from a common data stream using unprescaled single-lepton triggers and unprescaled E_T^{miss} triggers, which are discussed in Section 5.7 and further described in Refs. [207–209].

The registered data are filtered at the luminosity block[3] level using good-runs lists (GRLs) which required the LHC stable beams flag to be true and the data quality (DQ) flags for all detectors and trigger sub-components to be green. The amount of data used by the Run 2 analysis corresponds to an integrated luminosity of 138.96 fb^{-1}. The total uncertainties on the integrated luminosities for each individual year of data-taking range from 2.0 to 2.4 %, and are partially correlated between years [211]. The total uncertainty on the combined 2015–2018 integrated luminosity is 1.7%. These uncertainties are derived from the calibration of the luminosity scale using x-y beam-separation scans, following a methodology similar to that detailed in Ref. [212], and using the LUCID-2 detector for the baseline luminosity measurements [213]. The partial and total integrated luminosities together with their uncertainties and number of events are given in Table 4.1.

Year	Number of events	Integrated luminosity [pb^{-1}]
2015	220.58M	3219.56 $\pm$ 2.1%
2016	1057.84M	32988.1 $\pm$ 2.2%
2017	1340.80M	44307.4 $\pm$ 2.4%
2018	1716.77M	58450.1 $\pm$ 2.0%
2015–2018	4335.99M	138965.16 $\pm$ 1.7%

TABLE 4.1: Integrated luminosity per year with their relative uncertainties. Additionally, the number of events per year are shown.

[3]In the ATLAS experiment, a luminosity block is a time interval of data recording over which the experimental conditions are assumed to be constant. In particular, it is assumed that the instantaneous luminosity is constant over the duration of the luminosity block. Luminosity blocks are set by the ATLAS central trigger processor and their duration is usually around one minute. If the data-recording configuration changes, e.g. if a trigger prescale is adjusted, a new luminosity block is started [210].

Since the colliding bunches contain a huge number of protons ($> 10^{11}$), for each bunch crossing there are many *pp* interactions. As discussed in Section 3.1.4, the number of interactions per bunch crossing is called pile-up. The pile-up can induce degradation of object resolutions, the indefiniteness of the primary vertex (PV) location, and incorrect or missed reconstructions of particle track.

The pile-up is implemented in simulations by overlaying a varying number of simulated minimum-bias interactions on top of the hard scatter of the event. The average number of interactions per bunch crossing (i.e. the μ profile, defined in Equation 3.2) in the ATLAS detector for the luminosity delivered during Run 2 [181] is shown in Figure 4.1

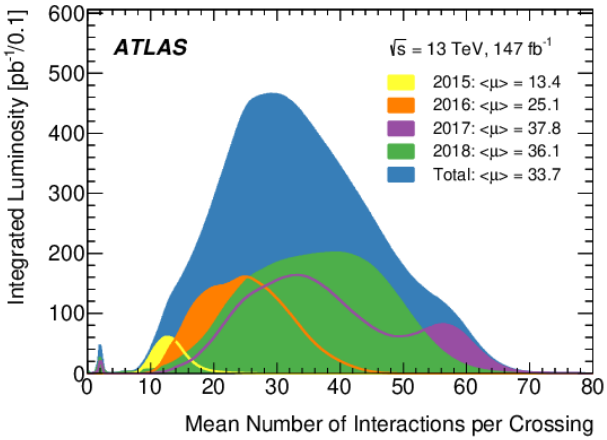


FIGURE 4.1: Luminosity-weighted distribution of the mean number of interactions per bunch crossing, μ, for the full Run 2 *pp* collision dataset at $\sqrt{s}$ = 13 TeV. The μ corresponds to the mean of the Poisson distribution of the number of interactions per crossing (see Equation 3.2) calculated for each proton bunch. It is calculated from the instantaneous per bunch luminosity. All data recorded by ATLAS during stable beams are shown, including machine commissioning periods, special runs for detector calibration, LHC fills with a low number of circulating bunches or bunch spacing greater than 25 ns. The integrated luminosity and the mean μ value for each year are given.

4.2 Monte Carlo simulation

A MC event generator is a program that simulates high-energy collisions, provides a list of final-state particles and their momenta, and thus gives a realistic estimate of signals and backgrounds that describe fundamental physics [214]. This simulation starts with the calculation of the parton level cross-section $\hat{\sigma}$ at a fixed order in perturbation theory (discussed in Section 3.1.2.1), and follows with the simulation of the parton cascades and non-perturbative effects, such as the hadronisation and multiple parton scatterings. In this section, a general overview of these techniques is given. The structure of a *pp* collision is shown in Figure 4.2, where the partons, UE, parton shower (PS), hadronisation and final-state hadrons are depicted.

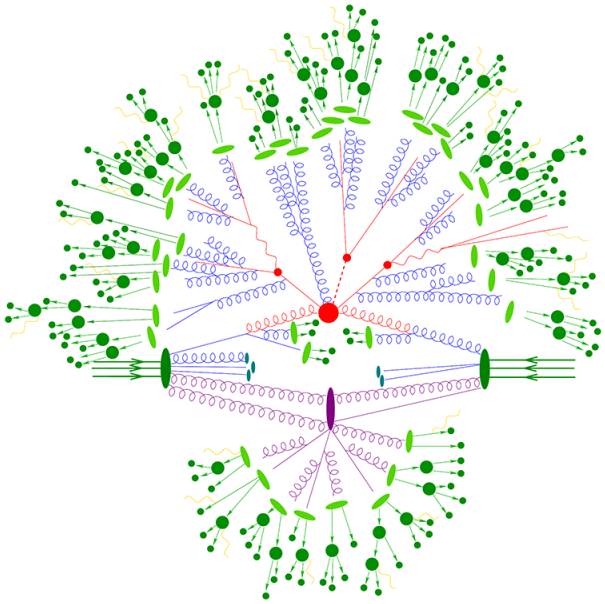


FIGURE 4.2: A sketch of the structure of a *pp* collision, including the initial-state partons (blue), the UE (violet), the hard scattering process and subsequent PS (red), the hadronisation (light green) and the final-state hadrons (green) [215].

4.2.1 Parton shower simulation

The primary process normally is a hard process where particles interact at high momentum, up to the predictable scale of hadron formation and decays. The partons (quarks and gluons) incoming or outgoing of the hard subprocess are, afterwards, part of the PS, which is the next step of the simulation. These particles involve large momentum transfers, particles with electric and colour charges can emit QED (i.e. photons) and QCD (i.e. gluons) radiation, respectively. The PS algorithm mimics the remaining terms of the perturbative expansion in α_S by emitting gluons which will eventually split into more partons. Consider the following example, where the PS splitting probability at a time t can be given by a function $f(t)$. The splitting can only happen at the instant t, so the splitting probability $P(t)$ satisfies:

$$P(t) = \frac{d}{dt}\Big(1 - N(t)\Big) = f(t)N(t). \tag{4.1}$$

Consequently, if the process starts at $t = 0$ with $N(0) = 1$, the solution of Equation 4.1 is,

$$P(t) = f(t)\,\exp\Big[-\int_0^t f(s)ds\Big],$$

therefore, the so-called Sudakov ($\Delta(t_2, t_1)$) factor, which provides the simplest example of factorisation theorems, can be written as,

$$\Delta(t_2, t_1) = 1 - \int_{t_1}^{t_2} P(s)ds = 1 + \exp\left(-\int_0^t f(s)ds\right)\Big|_{t=t_1}^{t_2} \leq 1.$$

The PS can be angle- or p_T-ordered, depending on the chosen scale t. For the first case, the angle of emission with respect to the incoming parton is decreased in each step, while in the second case, the ordering variable in time is the parton transverse momentum. The successive branching is stopped when a certain cut-off scale is reached. For the ISR, the scale variation is reversed (angles and momenta are increased), while for FSR, the scales are forced to decrease in each step.

4.2.2 Hadronisation simulation

After showering the interaction scale falls and the strong interaction causes the confinement of the coloured partons into colourless hadrons (primary hadrons), which may decay into secondary hadrons (eventually). The hadronisation process is not easy to compute with non-perturbative techniques so the MC event generators use phenomenological models to describe it. This process is based on the parton-hadron duality hypothesis, which establishes that, as the hadronisation is a long-distance process involving only small momentum transfers, the flows of momentum and quantum numbers at the hadron level must follow those for the parton level. This implies that partons are recombined with other partons to form hadrons if they are close in phase space. As perturbation theory works well down to low scales $Q \sim 1$ GeV, the assumption is made that α_s can be defined non-perturbatively for arbitrary low scales Q. These assumptions are supported by heavy quark spectra and event shapes data, but they do not provide a model for the way hadronisation actually happens.

The mechanism by which hadrons are formed from partons is simulated using two models, namely the Lund string model [216, 217] and the cluster model [218], which are schematically described in Figure 4.3.

The main characteristics of these algorithms are as follows:

- The Lund string model describes the colour dynamics between quarks in terms of strings, assuming a linear confinement potential. When the quarks separate in the phase space, this confinement potential increases the energy of the string up to the mass threshold of a new $q\bar{q}$ pair. In this instance, the string is broken and the new $q\bar{q}$ pair gives raise to the formation of hadrons. When a gluon splits perturbatively, an additional string segment is created, whereas the remaining gluons at the end of the PS lead to kinks in the string segment that connects them.

- The cluster model is based on the preconfinement property of the shower, by which neighbouring colour connected partons have an asymptotic mass distribution that falls steeply at high masses and is asymptotically independent of Q^2 and universal. The method starts with a non-perturbative splitting $g \to q\bar{q}$,

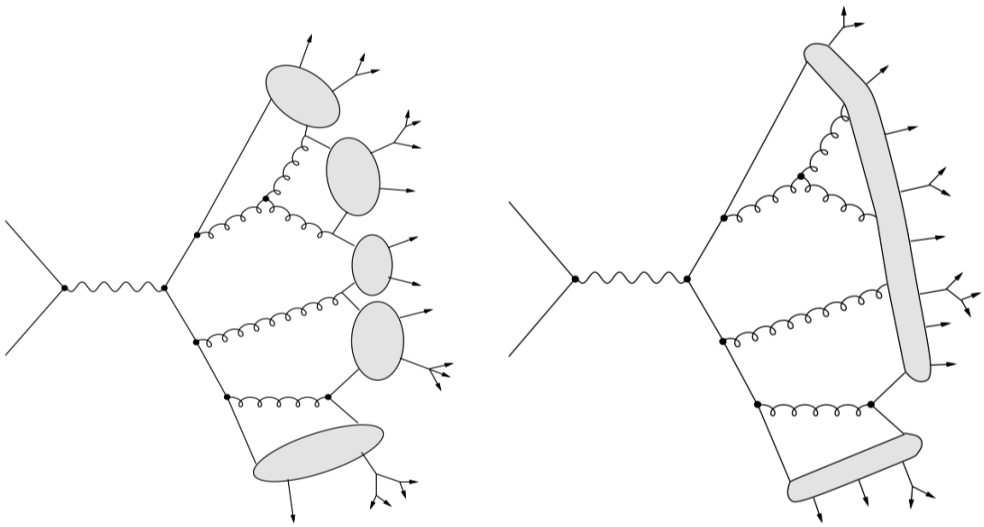


FIGURE 4.3: Schemes for the cluster **(a)** and the Lund string **(b)** hadronisation models [219].

and follows with the association of $q\bar{q}$ pairs into colour singlet combinations, which are assumed to form clusters. These clusters decay into pairs of hadrons following an isotropic pattern.

4.2.3 Underlying event simulation

In hadron collider events that contain a hard subprocess, there is extra hadron production that cannot be ascribed to showering from the coloured partons participating in the subprocess. The UE is believed to arise from collisions between those partons in the incoming hadrons that do not directly participate in the hard subprocess (as described in Section 3.1.2.3). The UE is illustrated the in sketch shown in Figure 4.2. The parameters involved in the simulation model of the UE have to be tuned using experimental data.

4.2.4 Hadron decays simulation

Finally, the last step of event generation is the decay of unstable hadrons[4]. As, the decay process simulation has to be consistent, the sum of branching fractions for all decays must be the unity. The experimental data indicate that a large fraction of observed final-state particles come from the decays of excited hadronic states, so most of the states listed in the Review of Particle Physics [17] need to be included, together with their decay modes.

4.2.5 Pile-up simulation

The effect of multiple interactions in the same and neighbouring bunch crossings (pile-up) is modelled by overlaying over the original hard-scattering event. These simulated minimum-bias events are generated with PYTHIA 8 (v8.186) [220] using the NNPDF2.3 LO set of PDFs [221] and the third ATLAS set of tuned parameters for minimum-bias events (A3 tune) [222]. The MC events are weighted to reproduce the $\langle\mu\rangle$ observed in the data. In this procedure, the $\langle\mu\rangle$ value from the data is divided by a factor of 1.03 ± 0.07, a rescaling which makes the number of reconstructed PV agree better between data and simulation and reproduces the visible cross-section of inelastic *pp* collisions as measured in the data [223].

4.3 Monte Carlo event generators

The general features described above are used by several MC event generator programs to describe the experimental data. A description is given for the most commonly used, which are by the way the ones used in the analyses carried out in this thesis. The MC generators implemented in this thesis are:

- The NLO perturbative QCD computations as well as PS programs are fundamental tools for the present-days particle physics phenomenology. The demand for better and better predictions from high energy experiments calls for improving the precision of existing PS generators, including NLO corrections [224].

[4]Typical hadron-collider definition of stable particles is defined when their lifetime obey that $c\tau > 10$ mm.

The POWHEG-BOX program [225–228] is an example of the NLO generators. Its formalism for every event is built by producing the $2 \to 2$ or $2 \to 3$ hard scattering matrix element[5] (ME). The renormalisation and factorisation scales μ_R and μ_F are set to be equal to the transverse momentum of the hard partons (p_T^{Born}). The cut on the minimum transverse momentum of the generated hard partons may affect the value of the cross-section due to the low-p_T divergence of the $2 \to 2$ cross-section. POWHEG-BOX is matched either to PYTHIA [229] or HERWIG [230] for the PS step of the event generation.

- MADGRAPH is another ME event generator [231]. Any $2 \to 1$, $2 \to 2$ and $2 \to 3$ processes can be implemented. Given the process, MADGRAPH automatically creates the amplitudes for all the relevant subprocesses and produces the mappings for the integration over the phase space. Once the events have been generated the event information (i.e. particle identifications, momenta, spin, colour connections) it may be passed directly to a shower MC program. The available models are SM with or without Higgs boson, full and simplified 2HDM and MSSM. MADGRAPH can simulate processes to LO accuracy for any user-defined Lagrangian, and to NLO accuracy in the case of QCD corrections to SM processes.

- SHERPA [232] MC event generator provides a complete hadronic final state in simulations of high-energy particle collisions. It is built with interactions of the SM and new physics processes. The events can be constructed with $2 \to 2$ processes. The emission of additional QCD partons of initial and final states is employed by the PS model based on phenomenological cluster-hadronisation, while a multiple-interaction model is used for the UE. SHERPA includes approximately 200 decay tables that contain 2500 decay channels, where the majority of the decaying particles are hadrons.

- PYTHIA 8 shower generator [229] is the most widely used and established MC event generator. It was written in C++ code unlike its predecessor PYTHIA 6

[5] The matrix element method is used by MC event generators to store the direct connection between the underlying theory of particle physics processes and physical observables measured by the particle detector.

that used Fortran. It is a LO event generator implementing the calculations of $2 \rightarrow 1$ and $2 \rightarrow 2$ MEs. It is matched to a p_T-ordered PS accounting for the ISR and FSR. Thus, the relevant scale is $Q^2 = -p_T^2$, which increases its value as the shower evolves. The fragmentation is based on the Lund string model [216, 217]. The UE is applied using a multiple-interaction model [233].

- HERWIG [230, 234] MC generator implements a wide variety of QCD processes. It was developed in Fortran but newer versions, HERWIG++ and HERWIG 7, are released in C++. The default PS is angular-ordered. This means that, for a branching $a \rightarrow bc$ the relevant scale is given by $Q^2 = 2E_a^2(1-\cos\theta_{bc})$, where E_a is the energy of the particle a and θ_{bc} is the angle between the branching products b and c. The scale Q^2 increases with the evolution of the shower, thus the angle between the products of sucessive branchings decreases as the shower evolves. HERWIG 7 makes use of the cluster model for hadronisation, and a eikonal multiple-interaction model for UE [218]. Also, HERWIG 7 improves perturbative input NLO QCD calculation for all SM processes.

4.4 Simulation and reconstruction software in the ATLAS experiment

After the event generation step, the trigger and detector simulation is performed with a dedicated ATLAS software infrastructure. The ATLAS software is implemented in the ATHENA framework [205], which is derived from the GAUDI framework developed for the LHCb experiment [235]. It ensures that the requested algorithms are run in the correct order, and it offers common services like message logging, access to data on disk, and filling of histograms and ntuples. The output of an algorithm is written to a common place in memory, called the "transient event store", from where the next algorithm can retrieve the output and process it further. The output can also be written to disk, using the POOL persistency scheme. The framework ensures that the data is read from disk (if it is not already present in the event store) if an algorithm requests it. In this way it is possible to run a chain of algorithms in one job, or to split it up into several jobs that read their input from disk.

The ATHENA framework is used to either generate simulated events or to reconstruct both simulated and real events. A schematic summary is shown in Figure 4.4.

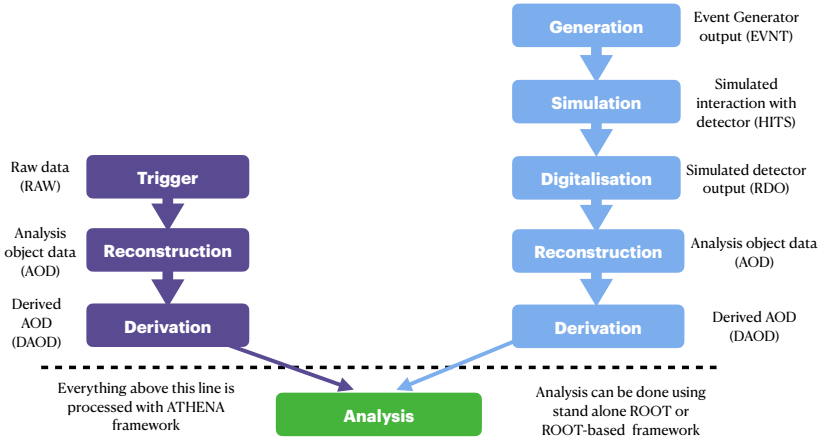


FIGURE 4.4: Schematic of the simulation and reconstruction procedure for data and MC events in ATLAS experiment by ATHENA framework.

The steps to simulate and reconstruct events are schematically:

- Event generation: The first step in the simulation chain is the simulation of the *pp* collision itself. As mentioned before, several programs are available to perform this task. These programs calculate the particles that are produced in the collision, with their four momenta. Same of these MC generators can be run within ATHENA framework.

- Detector simulation: For MC events, the passage of the generated particles through the detector is simulated. This task is performed by the GEANT4 toolkit [236], which is controlled by the G4ATLAS application. GEANT4 transports the particles through the magnetic field and simulates the interactions with the detector material. Examples of such interactions are: multiple scattering, energy loss, and photon conversions. Also the decay of unstable particles is simulated.

- Detector response: At this stage, the response of the detector including electronics is simulated (also called digitisation). In the simulation, a volume can be declared to be "sensitive", e.g. a silicon sensor. If a particle hits such a volume, a "hit" is written out. The digitisation consists of simulating the response of the detector to the energy deposits in these hits. Both the response of the detector and of the electronics are simulated.

- Reconstruction: Starting from the simulated data produced or from real data recorder by the detector, various algorithms are run to reconstruct the event. This includes algorithms that perform pattern recognition, track fitting, vertex determination, energy measurement, etc. When simulated data is considered, the reconstructed objects are matched to the simulated input ("truth"), such that the quality of the reconstruction can be studied. For data, as first step, it is collected in Raw data (RAW), the RAW are read in and converted to digits. These digits are equivalent to the output of the detector response simulation. This happens for events which pass the ATLAS trigger system (which is also properly emulated for MC simulated data). Then, the output is written to AOD ("analysis object data") and Derived AOD (DAOD), for data as well as for simulation. The AOD file contains only the information that is relevant for physics analysis. But, AODs are too big to be analysed directly so the container is reduced according to the needs of the physics groups, producing DAODs which are much smaller.

- Physics analysis: The reconstructed particles can be used to study a particular physics process. At this step data and simulation can be analysed either using the ATHENA framework or the ROOT package [237]. The latter is the standard CERN tool for analysing, histogramming and storing data and MC simulated data.

A major ingredient for the event simulation is the detector description. This part of the software has the information about the geometry and material of the detector, such as material properties, and the position and orientation of detector elements. All the simulation and reconstruction steps in ATLAS get their detector information from

a common source, called GeoModel [238]. This commonality ensures that each step uses exactly the same information, thus preventing inconsistencies.

Another important database for the reconstruction is the conditions database. This database holds the alignment constants, which are used by the detector description to correct the positions of the sensors. It also holds the calibration constants. The calibration constants also comprise the list of detector channels that are dead or noisy. Since the performance of the detector may change over time (for example, readout boards may be added or removed), the constants in the conditions database are grouped in blocks, so-called intervals of validity (IOVs). The reconstruction checks which IOV corresponds to the dataset that is reconstructed, and retrieves the constants that belong to this IOV.

4.5 2HDM+a analysis: Signal and background simulated samples

4.5.1 Simulated signal event samples

The MEs of the signal processes for the 2HDM+a model (described in Section 2.6.2) are generated at LO [231] using the MADGRAPH (v2.6.2) generator interfaced to PYTHIA 8 (v8.212) with the A14 tune for the modelling of PS, hadronisation and the description of the UE. It makes use of the 5 flavour scheme (5FS)[6], with the NNPDF30NLO PDF set. The signal event samples are generated with the requirement that one lepton[7] (e or μ) with $p_T > 20$ GeV and $|\eta| < 2.8$ and $E_T^{miss} > 60$ GeV are present (the definition of lepton and E_T^{miss} is done in Chapter 5). A scan is simulated, spanning a wide range of model parameters, which is described as:

- $m_A = m_H = m_{H^\pm}$,
- $m_a = 250$ GeV,
- $\sin\theta = 0.7$,

[6] Its refers to a consistently massless treatment of the b-quark, which can therefore be found in both initial and final states, while the 4 flavor treats the b-quarks as massive and allows them to be in the final state only [239].

[7] This primary selection has an efficiency on the event selection about a 25% among signals.

- m_χ = 10 GeV,

where m_A, m_H, $m_{H^\pm}$ are the CP-odd neutral Higgs boson, CP-even neutral Higgs boson and a pair of charged ones, respectively. The $\tan\beta$ is the ratio of Higgs doublets VEV. The neutral sector mixing term $\sin\theta$ where θ represents the mixing angle of the two CP-odd weak spin-0 eigenstates, A and a. The m_χ is the mass assumed to the DM particle candidate. The signals elected covers the values of $\tan\beta$: [0.3,..., 20] and $m_{H^\pm}$: [300,...,1750] GeV.

4.5.2 Simulated background event samples

For the background samples, several MC event generators are combined with PS and hadronisation programs. The most important backgrounds for the single top-quark DM process are $t\bar{t}$ and W+jets production. The single top-quark *tW*, *t*-channel and *s*-channel productions also contribute to the backgrounds. Less significant backgrounds are Z+jets and diboson (WW, ZZ and WZ) processes, as well as $t\bar{t}$ and single top-quark productions associated with vector or Higgs bosons ($t\bar{t}Z$, $t\bar{t}W$, $t\bar{t}H$, tZq and tWZ).

Top-quark pair process

Top-quark pair production constitutes one of the most important background processes to this single top-quark analysis. Di-leptonic $t\bar{t}$ events may mimic the single top-quark *t*-channel production final-state signature of two jets, one high-p_T lepton and large amount of missing transverse momentum, if one of the leptons is not identified.

The production of $t\bar{t}$ events is modelled using the POWHEG-BOX (v2) [225–228, 240–242] program, which simulates the MEs at NLO in the strong coupling constant α_S with the NNPDF3.0 NLO [231] PDF set and the h_{damp} parameter[8] set to 1.5 m_t [243]. The functional form of the μ_R and μ_F scale is set to the default scale $\sqrt{m_t^2 + p_T^2}$. Top quarks are decayed at LO using MADSPIN to preserve all spin correlations. The events are interfaced with PYTHIA 8 (v8.230) [229] for the PS and

[8]The h_{damp} parameter controls the p_T of the first additional emission beyond the LO Feynman diagram in the PS and therefore regulates the high-p_T emission against which the $t\bar{t}$ system recoils.

hadronisation, using the A14 tune and the NNPDF2.3 LO PDF set. The decays of bottom and charm hadrons are simulated using the EvtGen (v1.6.0) program (more details are given in Section 4.7.1).

The NLO $t\bar{t}$ inclusive production cross-section is corrected to the theory prediction at NNLO in QCD including the resummation of next-to-next-to-leading logarithmic (NNLL) soft-gluon terms calculated using Top++2.0 [244–249]. For *pp* collisions at $\sqrt{s}$ = 13 TeV, this cross-section corresponds to 832 ± 51 pb using a top-quark mass of m_t = 172.5 GeV.

Single top-quark processes

Although having a smaller contribution, the single top-quark *tW*, *s*-channel and *t*-channel associated productions are considered.

The single top-quark *tW* associated production and the *s*-channel process are modelled using the Powheg-Box (v2) generator which provides MEs at NLO in the strong coupling constant α_S in the 5FS with the NNPDF3.0 NLO PDF set. In both cases, the functional form of the μ_R and μ_F scale is set to the default scale $\sqrt{m_t^2 + p_T^2}$. In the *Wt* associated production, the diagram removal (DR) scheme [250] is employed to handle the interference with $t\bar{t}$ production [243]. Top quarks are decayed at LO using MadSpin to preserve all spin correlations. The events are interfaced with Pythia 8 (v8.230) using the A14 tune and the NNPDF2.3 LO PDF set. The decays of bottom and charm hadrons are simulated using the EvtGen (v1.6.0) program.

The inclusive cross-section of the *tW* associated production is corrected to the theory prediction calculated at NLO in QCD with NNLL soft gluon corrections [251, 252]. For *pp* collisions at $\sqrt{s}$ = 13 TeV, this cross-section corresponds to 71.7 ± 3.8 pb, using a top-quark mass of m_t = 172.5 GeV. The uncertainty on the cross-section due to PDF is calculated using the MSTW2008 90% CL NNLO PDF set, and is added in quadrature to the scale uncertainty.

For the *s*-channel process, the inclusive cross-section is corrected to the theory prediction calculated at NLO in QCD with HatHor (v2.1). For *pp* collisions at $\sqrt{s}$ = 13 TeV, this cross-section corresponds to $6.35^{+0.23}_{-0.20}$ pb and $3.97^{+0.19}_{-0.17}$ pb

for top-quark and top-antiquark production, respectively, using a top-quark mass of m_t = 172.5 GeV.

The *t*-channel single top-quark events are produced using the POWHEG-BOX (v2) generator which provides MEs at NLO in the strong coupling constant α_S in the 4FS with the NNPDF3.0 NLO nf4 [231] PDF set. The functional form of the μ_R and μ_F scale is set to $\sqrt{m_b^2 + p_{T,b}^2}$ following the recommendation of Ref. [240], where m_b and $p_{T,b}$ are the mass and $p_T$of the b-quark from the initial gluon splitting, so-called "second b-quark". The factorisation scale is set to μ_F^2 = - p_W^2 for the spectator quark and μ_F^2 = $p_{\bar{b}}^2$ + m_b^2 for the gluon, where p_W and $p_{\bar{b}}$ are the three-momenta of the exchanged W boson and of the b-antiquark originating from the gluon splitting, respectively. This sample is normalised to the predicted single top-quark *t*-channel production cross-section, which was calculated at NLO in QCD with HATHOR (v2.1) [251, 252]. For *pp* collisions at $\sqrt{s}$ = 13 TeV , this cross-section corresponds to $54.9^{+2.3}_{-1.9}$ pb and $29.7^{+1.7}_{-1.5}$ pb for top quark and anti-quark production, respectively, using a top-quark mass of m_t = 172.5 GeV.

Single-boson processes

The QCD V+jets production is simulated with the SHERPA (v2.2.1) [232] ME+PS generator. Additional hard parton emissions [253] are matched to a PS based on Catani-Seymour dipoles [254]. The NNPDF3.0 NNLO set [231] of PDFs as well as the dedicated set of tuned PS parameters developed by the SHERPA authors for this version are used.

The ME+PS matching [255] is employed for different jet multiplicities which are then merged into an inclusive sample using using an improved CKKW matching procedure [256, 257] which is extended to NLO accuracy using the MEPS@NLO prescription [258]. These particular simulations are NLO accurate for up to two additional partons and LO accurate for up to four additional partons. The virtual QCD correction for MEs at NLO accuracy are provided by the OPENLOOPS library [259, 260].

Each final state (i.e. W+jets and Z+jets) is sliced in the variable max($H_T, p_T(V)$) at the parton level [261], where the boson transverse momentum, $p_T(V)$, is worked

out from the two parton-level leptons and the parton-level H_T is given by the scalar p_T sum of all parton-level jets with $p_T > 20$ GeV. Here, parton-level jets are constructed from all remaining ME-level partons (excluding the two leptons) using the anti-k_t algorithm with a jet-radius parameter 0.4. For $\max(H_T, p_T(V)) < 500$ GeV, the slices are also filtered according to their b-hadron and c-hadron content at the particle level.

Other minor processes

Other minor backgrounds are: $t\bar{t}Z$, $t\bar{t}W$, $t\bar{t}H$, tZq, tWZ, WW, WZ, ZZ processes.

The MADGRAPH5_aMC@NLO (v2.3.3) generator is used to modelled $t\bar{t}Z$ and $t\bar{t}W$ event productions, which provides NLO ME in $\alpha_S$with the NNPDF3.0 NLO PDF set. The form of μ_R and μ_F is set to the default scale $0.5\times\sum_i\sqrt{m_i^2+p_{T,i}^2}$, where the sum runs over all the particles generated from the ME calculation. The events are interfaced with PYTHIA 8 (v8.210) for the PS and hadronisation, using the A14 tune and the NNPDF2.3 LO PDF set. The cross-sections are calculated at NLO QCD and NLO EW accuracies using MADGRAPH5_aMC@NLO and reported in Ref. [262]. The $t\bar{t}Z$ cross-section is further supplemented with an off-shell (down to 5 GeV) correction. The predicted values are $0.88^{+0.09}_{-0.11}$ pb and $0.60^{+0.08}_{-0.07}$ pb for $t\bar{t}Z$ and $t\bar{t}W$ respectively, where the uncertainties are from variations of μ_R and μ_F scales as well as $\alpha_S$variations.

The production of $t\bar{t}H$ events is modelled using the POWHEG-BOX generator at NLO with the NNPDF3.0 NLO PDF set. The events are interfaced with PYTHIA 8 (v8.230) using the A14 tune and the NNPDF2.3 LO PDF set. The cross-section is calculated at NLO QCD and NLO EW accuracy using MADGRAPH5_aMC@NLO and reported in Ref. [262]. The predicted values are 507^{+35}_{-50} fb, where the uncertainties are from variations of μ_R and μ_F scales as well as α_S variations.

The tZq event sample is simulated using the MADGRAPH5_aMC@NLO (v2.3.3) generator at NLO with the NNPDF3.0 NLO PDF set. The events are interfaced with PYTHIA 8 (v8.230) using the A14 tune and the NNPDF2.3 LO PDF set. Off-resonant events away from the Z-boson mass peak are included. The 4FS is used where all the

quark masses are set to zero, except for the top and bottom quarks. The functional form of the μ_R and μ_F scale is set to the $4\sqrt{m_b^2 + p_{T,b}^2}$, where the b-quark is the external one produced from gluon splitting in the event. The cross-section, calculated at NLO using MADGRAPH5_aMC@NLO (v2.3.3) with the NNPDF3.0 NLO PDF set, is 800 fb, with an uncertainty of $^{+6.1}_{-7.4}\%$. The uncertainty is computed by varying the μ_R and μ_F scales by a factor of two and by a factor of 0.5.

The tWZ sample is simulated using the MADGRAPH5_aMC@NLO (v2.3.3) generator at NLO with the NNPDF3.0 NLO PDF set. The events are interfaced with PYTHIA 8 (v8.212) using the A14 tune and the NNPDF2.3 LO PDF set. In this simulated event samples, top quark decays inclusively while Z boson decays to a pair of leptons. The 5FS is implemented. The the μ_R and μ_F scales are set to the top-quark mass. The DR1 scheme is employed to handle the interference between tWZ and $t\bar{t}Z$, and is applied to the tWZ event sample.

In all these samples, top quarks and W/Z are decayed at LO using MADSPIN to preserve all spin correlations. The decays of bottom and charm hadrons are simulated using the EVTGEN (v1.2.0) program.

Fully leptonically and semileptonically decaying diboson samples are simulated with the SHERPA (v2.2) ME+PS generator as for V+jets production. The ME+PS matching is the same as for the single-boson processes. These particular simulations are NLO accurate for up to one additional parton and LO accurate for up to three additional parton emissions using factorised on-shell decays. The virtual QCD correction for MEs at NLO accuracy are provided by the OPENLOOPS library. The calculation is performed in the G_μ scheme using, ensuring an optimal description of pure EW interactions at the EW scale.

Physics process		Generator	PDF Set	Parton shower	UE Tune	Cross-section normalisation
Signal		MADGRAPH (v2.6.2)	NNPDF 3.0 [231]	PYTHIA 8	A14 [263]	LO [231]
Top pair ($t\bar{t}$)		Powheg-Box (v2) [227]	NNPDF 3.0	PYTHIA 8	A14	NNLO+NNLL [249]
Single-top	t-channel	Powheg-Box (v2)	NNPDF 3.0	PYTHIA 8	A14	NNLO+NNLL [264]
	s- and Wt-channel	Powheg-Box (v2)	NNPDF 3.0	PYTHIA 8	A14	NNLO+NNLL [265, 266]
V+jets ($V = W/Z$)		SHERPA (v2.2.1) [232]	NNPDF 3.0	SHERPA	Default	NNLO [267]
Diboson		SHERPA (v2.2.1 – 2.2.2)	NNPDF 3.0	SHERPA	Default	NLO
$t\bar{t}+V$, $V = W, Z, h$		MADGRAPH5_aMC@NLO (v2.3.3)	NNPDF 3.0	PYTHIA 8	A14	NLO [268]

TABLE 4.2: List of generators used for the different processes. Information is given about the underlying-event tunes, the PDF sets and the perturbative QCD highest-order accuracy (LO, NLO, NNLO, and NNLL) used for the normalisation of the different samples.

4.6 Non-resonant model analysis for HL-LHC: Signal and background simulation samples

Only MC simulation event samples are used for the analysis developed for the HL-LHC project, which expects to come into operation at the end of 2027. Samples of events generated using MC simulations are produced using different event generators interfaced to various PS and hadronisation generators.

4.6.1 Simulated signal event samples

The signal samples are generated following the same procedure of Ref. [134]. For the ME calculations, samples of signal events (see Section 2.6.1) generated using the non-resonant model are produced using the MADGRAPH5_aMC@NLO (v2.3.3) [268] generator at LO using the NNPDF3.0 LO [231] PDF set. The PS, hadronisation and the UE are handled by the PYTHIA 8.30 event generator [229] with the A14 [263] set of tuned parameters, using the NNPDF2.3 LO PDF set [221]. TheEVTGEN (v1.6.0) program [269] is used to describe the properties of the bottom and charmed hadron decays. All these MC simulation samples are generated for a range of the mediator masses between m_V = 1.0 and 5.0 TeV, in steps of 0.5 TeV. The benchmark DM particle masses are assumed to be m_χ = 1 GeV (larger masses, up to around 100 GeV, can be considered since kinematic distributions predicted by the model do not change

as shown in Ref. [270]). The values of the coupling parameter a is set to 0.5 and g_χ is set to 1.0.

4.6.2 Simulated background event samples

Samples of simulated events for background processes include production of single-top quarks, $t\bar{t}$, W/Z boson in association with jets, vector-boson pairs, associated production of a $t\bar{t}$ pair and a W/Z boson and single-top quark in association with a Z boson.

The $t\bar{t}$ production and EW production of single-top quarks in the *s*-channel, associated *Wt* and *t*-channel are produced using the NLO, implemented with POWHEG-BOX (v2) generator [225–227]. In the $t\bar{t}$ event generation the resummation damping factor h_{damp} is set to 1.5 m_t. The *t*-channel event generation the 4FS is used, treating the b-quark as massive. For $t\bar{t}$ and *s*-channel the NNPDF3.0 NLO PDF set is used in the ME generation, while NNPDF3.04f NLO PDF set is used for the *t*-channel, and CT10 [271] PDF set for the associated *Wt* process. All these simulation samples except the latter are interfaced to PYTHIA 8 for the PS, fragmentation and the UE simulation, using the A14 set of tuned parameters and the NNPDF2.3 LO PDF set. The associated *Wt* production sample is interfaced to PYTHIA 6 [272], using the CT10 PDF set and the corresponding Perugia 2012 tuneable parameters [273].

The MADGRAPH5_aMC@NLO generator at LO using the NNPDF3.0 NLO PDF set is implemented for the W boson production in association with jets. These W+jets event samples are simulated for up to one additional parton at NLO and up to two additional partons at LO. The Z+jets production is produced using the POWHEG-BOX (v1) generator at NLO in QCD with the CT10 PDF set and the AZNLO [274] set of tuned parameters of the UE are used. The final-state photon radiation is modelled by the PHOTOS [275] MC simulation. Both productions are interfaced with PYTHIA 8 generator for the PS, fragmentation and UE, using the CT10 PDF set in the case of W+jets and CTEQ6L1 [276] PDF set in the case of Z+jets.

Samples of vector-boson pairs events (WW, ZZ, WZ), containing up to three additional partons where at least one of the bosons decays leptonically, are produced using the SHERPA generator [232] with the NNPDF3.0 NNLO PDF set.

The associated productions of a $t\bar{t}$ pair and either a W or Z boson ($t\bar{t}W$, $t\bar{t}Z$) are generated using MADGRAPH5_aMC@NLO at NLO using the NNPDF3.0 NLO [231] PDF set. The generated events are then processed with PYTHIA 8 to perform the fragmentation and hadronisation, and to generate the UE, using the NNPDF2.3 LO PDF set and the A14 set of tuned parameters.

Samples of single-top quark production in association with a Z boson events (tZq) are generated at LO in QCD using MADGRAPH5_aMC@NLO in the 4FS with the CTEQ6L1 LO PDF set. The Z boson is simulated to be on-shell and off-shell Z/γ^* contributions and their interference are not taken into account. The PS, hadronisation and the UE are generated by PYTHIA 8 with the A14 set of tuned parameters using the NNPDF2.3 LO PDF set.

In all background samples where PYTHIA 6 or PYTHIA 8 are used, the EVTGEN (v1.2.0) program was also used to model bottom and charmed hadron decays.

The MC programs implemented to simulate signal and background (SM) processes are listed in Table 4.3.

Physics process		Generator	PDF Set	Parton shower	UE Tune	Cross-section normalisation
Signal		MADGRAPH5_aMC@NLO (v2.3.3)	NNPDF 3.0 [231]	PYTHIA 8	A14 [263]	LO [231]
Top pair ($t\bar{t}$)		Powheg-Box (v2) [227]	NNPDF 3.0	PYTHIA 8	A14	NNLO [267]
Single-top	t-channel	Powheg-Box (v2)	NNPDF 3.0	PYTHIA 8	A14	NLO [271]
	s-channel	Powheg-Box (v2)	NNPDF 3.0	PYTHIA 8	A14	NLO
	Wt-channel	Powheg-Box (v2)	NNPDF 3.0	PYTHIA 6	A14	NLO
W+jets		MADGRAPH5_aMC@NLO (v2.3.2)	NNPDF3.0	PYTHIA 8	A14	NLO
Z+jets		Powheg-Box (v1)	NNPDF3.0	PYTHIA 8	AZNLO	NLO
Diboson		SHERPA (v2.2.1)	NNPDF3.0	SHERPA	Default	NLO
$t\bar{t}+V$, $V=W,Z,h$		MADGRAPH5_aMC@NLO (v2.3.2)	NNPDF2.3	PYTHIA 8	A14	NLO
tZq		MADGRAPH5_aMC@NLO (v2.3.2)	NNPDF2.3	PYTHIA 8	A14	LO

TABLE 4.3: List of generators used for the different processes. Information is given about the underlying-event tunes, the PDF sets and the perturbative QCD highest-order accuracy (LO, NLO, NNLO, and NNLL) used for the normalisation of the different samples.

4.7 Monte Carlo dependence of heavy-flavour production fractions and decays

To explore new physics scenarios the MC simulation predictive performance needs to be evaluated. Particularly, the precise description of heavy-flavour (bottom and charm) hadrons at the LHC. The good knowledge of the heavy-flavour properties allows to have accuracy on measurements. It can expose physics BSM, constrain SUSY and test QCD. In particular, the heavy-flavour description plays a key role in the identification of $\mathrm{t\bar{t}}$ processes. Several studies [134, 243, 277, 278] were presented with modern MC event generators, in order to study their prediction goodness. The study of the $\mathrm{t\bar{t}}$ production is particularly relevant for the analyses presented in this thesis, since the $\mathrm{t\bar{t}}$ process is the major background contributor. The choice of the MC generator can affect the kinematic properties of the simulated $\mathrm{t\bar{t}}$ events and the reconstruction efficiency. The implementation of higher-order QCD corrections is important for the understanding and subtraction of this background [206]. To develop the studies, MC simulation samples of $\mathrm{t\bar{t}}$ production are generated, where the heavy-flavour hadron are produced from the b-quark for each top-quark decay, and a c-quark from W boson decays. In particular, one relevant study is to check the effect of MC parameters (beyond default parameters) and look for better or optimal values. The MC event generators selected are SHERPA (v2.2.1), PYTHIA 8 (v8.210) and HERWIG 7 (v7.0.1). The latter two are PS generators, which the partonic hard process MEs are calculated using POWHEG implementation, needed to provide NLO corrections.

The following subsections show detailed studies of heavy flavour hadron production and decay properties for different MC generators using *pp* and $\mathrm{e^{-}e^{+}}$ collisions.

4.7.1 Monte Carlo simulation samples of heavy-flavour production

In the following sections there are different versions of MC event generators, they are:

- The version 2.2.1 of SHERPA calculates the NLO ME processes by the OPENLOOPS library [259] combining the COLLIER library [260] to provide the loop integrals for NLO generation. This representation includes up to one additional

parton at NLO accuracy and four additional partons at LO accuracy. Samples set of PDF is given by NNPDF 3.0 NLO description [231]. Two samples of SHERPA are used, the nominal with the Heavy Baryon Enhancement parameter (HBE) set up to four, and the varied with HBE = 1. The HBE value affects the b-quark and c-quark hadronisation procedure [279]. Their production fractions is an important element to reproduce in MC simulation samples, in order to identify jets from b-quarks and c-quarks.

- The interplay between POWHEG and PYTHIA 8 can be optimised by choosing the best value of the h_{damp} parameter [277] (i.e. resummation damping factor) whose value affects the ME/PS matching in POWHEG and regulates the high-p_T radiation. The default value for PYTHIA 6 interfaced to POWHEG is $h_{damp} = m_t$, using a top-quark mass of m_t = 172.5 GeV. While the setup of POWHEG interfaced to PYTHIA 8 is $h_{damp} = 1.5\ m_t$ using the A14 tune [263]. This choice is compared with unfolded data at the centre-of-mass energies 7, 8, and 13 TeV in Ref. [243], giving a good description of the data. PYTHIA version 8.210 interfaced to POWHEG generator is used in the studies shown in the next sections. It is implemented with the A14 set of tunable parameter as well, at least one lepton filter and $h_{damp} = 1.5\ m_t$. The renormalisation, factorisation scale and the PDF (NNPDF 3.0 NNLO) choice are the same as for the SHERPA sample.

- POWHEG interfaced to HERWIG version 7.0.4 uses an h_{damp} parameter set to $1.5 m_t$. Moreover, HERWIG 7 uses the H7UE set of tunable parameters [234]. The used PDF set is the same as in the PYTHIA 8 and SHERPA samples, i.e. NNPDF 3.0 NLO.

- Finally, the EVTGEN (v1.2.0) [269] program provides decay models of heavy-flavour hadrons, especially for B and D mesons. It has detailed models for semileptonic decays, CP-violating decays and produces correct results for the angular distributions in sequential decays. Some simulation samples used in the following sections are implemented EVTGEN package.

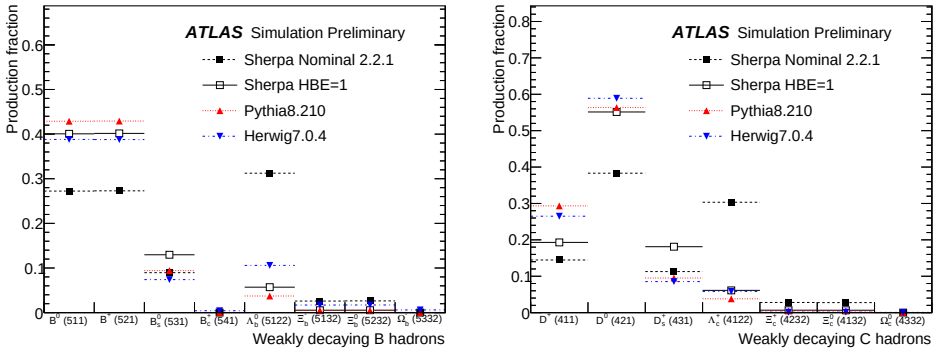
4.7.2 Heavy-flavour hadron production fraction studies

Having a correct modelling of the heavy-flavour production fractions is an important element to reproduce in MC simulation samples the efficiencies of identifying jets from b- and c-quarks (what is known as tagging). Charm tagging is especially sensitive to these fractions, due to the large differences in lifetimes and semileptonic branching fractions among the charm hadron species. This section describes different MC simulations to study their prediction of the production fraction of heavy hadrons. Previous studies are shown in Ref. [278], presenting that the relative rates of hadron production depend on the hadronisation model and the values of tunable parameters of MC generators. In these MC simulation samples the condition to the jets are $p_T > 25$ zGeV and $p_T < 2000$ GeV.

The studies of the heavy-flavour production fractions in PYTHIA 8, HERWIG 7 and SHERPA 2.2.1 are shown here. For the latter, the effect of varying the HBE is studied by comparing the default (HBE = 4) with a lower HBE factor (HBE = 1). Top-quark pair production samples are used as benchmark processes. Charm hadron production fractions include only those hadrons produced directly (not from bottom hadrons decays). Figure 4.5 shows the heavy-flavour production fractions for bottom and charm hadrons. The SHERPA varied event sample (HBE = 1) has a better agreement with PYTHIA 8 and HERWIG 7 than the nominal sample (HBE = 4). Fractions obtained from the generators are compared with experimental data in Table 4.4 and Table 4.5, showing that SHERPA varied event sample is in good agreement with the World Average as well. In fact, SHERPA nominal event sample produces bottom (B^+, B^0, B_s^0) and charm (D^+, D^0, D_s^0) fractions significantly lower.

Species	SHERPA Nominal	SHERPA (HBE = 1)	PYTHIA 8	HERWIG 7	World Average [280]
B^+	27.3	40.1	42.9	38.8	40.4 ± 0.6
B^0	27.2	40.1	42.9	38.7	40.4 ± 0.6
B_s^0	9.0	13.0	9.4	7.4	10.3 ± 0.5
Baryons	36.5	6.8	4.8	15.1	8.8 ± 1.2

TABLE 4.4: Percentage probability that a b-quark decays to a B-hadron of a given species for SHERPA (v2.2.1), PYTHIA 8 (v8.210) and HERWIG 7 (v7.0.1) event samples and for the World Average.



(a) **(b)**

FIGURE 4.5: Relative rates for production of different heavy-flavour species: B-hadrons **(a)** and D-hadrons **(b)** [279].

Species	SHERPA Nominal	SHERPA (HBE = 1)	PYTHIA 8	HERWIG 7	World Average [281]
D^+	14.5	19.3	29.3	26.5	22.56 ± 0.77
D^0	38.5	55.1	56.4	58.9	56.43 ± 1.51
D_s^0	11.3	18.1	9.5	8.5	7.97 ± 0.45
Baryons	35.9	7.5	4.8	6.1	10.8 ± 0.91

TABLE 4.5: Percentage probability that a c-quark decays to a D-hadron of a given species for the SHERPA (v2.2.1), PYTHIA 8 (v8.210) and HERWIG 7 (v7.0.1) event samples and for the World Average.

Production fractions vary about 10% among generators [9]. The production fraction of D_s^0 is overestimated for all generators. In case of D^+ (B_s^0) SHERPA calculates a fraction smaller (higher) than the World Average. Also HERWIG 7 underestimates the production fraction of B_s^0.

Results show that, the HBE (= 4) factor set to the default SHERPA (v2.2.1) is inconsistent with measured flavour production fractions and should be reduced to a value of 1.

[9] Comparison only done with respect the SHERPA varied sample (HBE = 1).

4.7.3 Heavy-flavour fragmentation studies

The fragmentation of a b- or c-quark into a heavy-flavor hadron is modelled by the generators using a non-perturbative fragmentation function $\mathrm{D}_{\mathrm{X}}^{h}(z)$ that describes the probability that a heavy quark of species *h* hadronises into a *X* hadron carrying, fraction *z* of the quark's momentum.

The fragmentation function exhibits scaling violations and has to be estimated with the corresponding scale μ. These fragmentation functions are assumed to be universal, aside from the scale dependence, which can be calculated perturbatively.

A possible approach of the fragmentation of b- and c-quarks can be calculated with phenomenological models of hadronisation using the non-perturbative fragmentation function $\mathrm{D}_{\mathrm{X}}^{\mathrm{h}}(z,\mu^2)$. The modelling of the fragmentation functions differs among the generators, but in all cases the parameters of the models are tuned using data from $\mathrm{e}^{+}\mathrm{e}^{-}$ collisions [278].

To assess the performance of the three generators, the following strategy is used:

- For PYTHIA 8 two different α_{S} values are used in order to study the change of bottom fragmentation function. One is the standard value, used in most MC programs ($\alpha_{\mathrm{S}} = 0.137$) and the other value ($\alpha_{\mathrm{S}} = 0.127$) is the ATLAS default value. Furthermore, PYTHIA 8 is interplaced with EVTGEN to provide a better description of weak decays of heavy flavour hadrons.

- The Herwig 7 event generator's Matchbox module allows the full implementation of LO/NLO simulations, adding a subtractive (MC@NLO-like)[10] or multiplicative (POWHEG-like) matching scheme for NLO calculations. In these studies subtractive option is employed. The subtractive process employs the ME corrections using adaptive methods, and is able, for the case of the angular-ordered shower (PS default implementation), to add truncated showers on top of it to fully account for large-angle, soft emissions.

- The SHERPA Nominal setup MC generator is implemented (i.e. HBE = 4).

[10] The MC@NLO formalism allows to incorporate NLO QCD MEs consistently into a PS framework.

Figure 4.6 shows the b-quark fragmentation function at $\sqrt{s} = 91.2$ GeV for e^+e^- collision. The fragmentation function is E/E_{beam} where E is the energy of the B-hadron and the *y*-axis shows,

$$F(x) \equiv \frac{1}{N_B}\frac{dN_B}{dx}.$$

where N_B in the number of weakly decaying B-hadrons. Figure 4.6 shows different MC event generators and experimental measurements from DELPHI [282] and SLD [283] detectors. It is observed that the best agreement is given by the PYTHIA 8 generator compared to DELPHI and SLD data. The different values of α_S do not show a significant effect on the fragmentation function. SHERPA (HERWIG 7) has a higher (smaller) peak compared with experimental data. Table 4.6 presents the mean value of the b-quark fragmentation function for the different MC event generators and the World Averange values. The mean value of the fragmentation function for PYTHIA 8 and SHERPA is more similar to the SLD value and HERWIG 7 is approaching the result from DELPHI. These studies show differences in the mean values of the b-fragmentation at $\sqrt{s} = 91.2$ GeV of about 4.5 %.

	E_B/E_{beam}
PYTHIA 8	0.7281±0.0002
PYTHIA 8 ATLAS	0.7290±0.0002
SHERPA	0.7110±0.0005
HERWIG 7	0.6846±0.0002
Delphi Data	0.699 ± 0.011
SLD Data	0.709 ± 0.005

TABLE 4.6: Mean values of fragmentation function computed by PYTHIA 8 (v8.210), HERWIG 7 (v7.0.4) and SHERPA (v2.2.1) at $\sqrt{s} = 91.2$ GeV and experimental data from DELPHI and SLD.

In e^+e^- collisions, the fragmentation function can be measured relative to the beam energy, since the momentum transfer of the hard scatter is fixed by the energy of the initial beams (aside from initial-state photon radiation). Nonetheless, fragmentation in hadron collisions must be parameterised with respect to the output of a

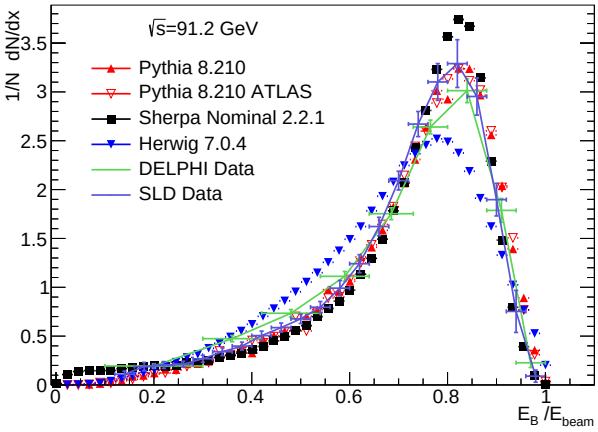


FIGURE 4.6: The b-quark fragmentation function computed by PYTHIA 8 (v8.210), HERWIG 7 (v7.0.4) and SHERPA (v2.2.1) for centre-of-mass of energy $\sqrt{s} = 91.2$ GeV.

jet-finding algorithm. By convention, the p_T of the jet containing the heavy-flavour hadron is used to approximate the unknown p_T of the heavy flavor parton. So, the fragmentation function $f(z)$ can be written as,

$$f(z) \equiv \frac{\vec{p}_{\text{hadron}} \cdot \vec{p}_{\text{jet}}}{p^2_{\text{jet}}},$$

where $\vec{p}_{\text{hadron}}$ and $\vec{p}_{\text{jet}}$ stand for the vector momentum of heavy-flavour hadron and the jet (with a magnitude value of p_{jet}).

Figure 4.7 shows the b-quark fragmentation function computed by PYTHIA 8, HERWIG 7 and two versions of SHERPA, the nominal with HBE = 4 and the varied with HBE = 1 (same samples at Section 4.7.2). These MC event samples are built using the anti-k_t algorithm (R = 0.4), applied to particle-level truth objects[11]. This study shows that the SHERPA varied setup (HBE = 1) produces a b-fragmentation

[11] The reconstruction of the jet with anti-k_t algorithm behaves like a cone algorithm based on cluster jet algorithms; R is the radius parameter of the algorithm [284]. And truth particle jets are made of stable particles (excepts muons and neutrinos).

function more compatible with the PYTHIA 8 and HERWIG 7 generators, than with the nominal.

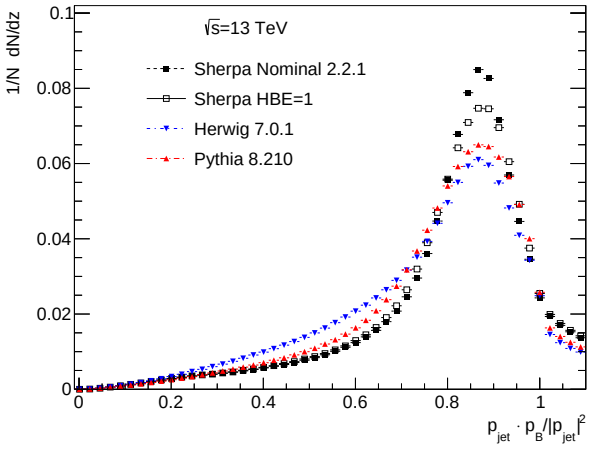


FIGURE 4.7: The b-quark fragmentation function computed by PYTHIA 8 (v8.210), HERWIG 7 (v7.0.1) and SHERPA (v2.2.1) nominal setup (HBE = 4) and SHERPA (v2.2.1) varied setup (HBE = 1) at $\sqrt{s}$ = 13 TeV.

4.7.4 Heavy flavour hadron decay studies

Heavy-flavour decay properties are studied for different MC event generators. All with weakly decaying B- and D-hadrons MC event generators studied are: PYTHIA 8 (v8.210), PYTHIA 6 (v6.428), HERWIG++ (v7.0.1), HERWIG 7 (v7.0.1) and SHERPA (v2.2.1), where the MC events are selected from $t\bar{t}$ events and their decays are then analysed. The differences among MC event generators can affect the b-tagging[12] efficiency and the mistag rate for D-hadrons passing the b-tagger requirement in simulated data. An alternative approach is to re-decay all heavy flavor hadrons using a unified decay description. Results are presented here for each generator for both with and without EVTGEN[13]. In some cases, the data in these standard tables differ from

[12]The b-tagging method is used for jet flavor tagging identification of jets originating from bottom quarks.

[13]With the exception of SHERPA generator, which EVTGEN generator is not interfaced.

the values reported by the PDG. These differences are highlighted in the figures that follow.

Figures 4.8-4.9 compare the lifetimes of four weakly decaying B- (B^0, B^+, B_S^0 and Λ_b) and D-hadrons (D^0, D^+, D_S and Λ_c) respectively. It is shown that using EVTGEN the agreement of the lifetimes with respect the experimental averages in the PDG is improved. It is also observed that for B^+ and Λ_c hadrons, the SHERPA generator shows a good agreement with experimental data. However the rest of B- and D-hadrons exhibit discrepancies between SHERPA and the World Average listed by the PDG (and the other generators). On the other hand, as seen in Ref. [278], the B_S^0 still shows discrepancies with respect the PDG values, even if the MC event generators are using the EVTGEN package.

The heavy-flavour model implemented by SHERPA has known issues and limitations. These kind of studies suggest that the SHERPA MC generator needs to be interfaced with EVTGEN to eliminate discrepancies. This implementation is already used within the ATLAS Collaboration. The improvements are prepared to be available in the SHERPA (v2.2.5) official software.

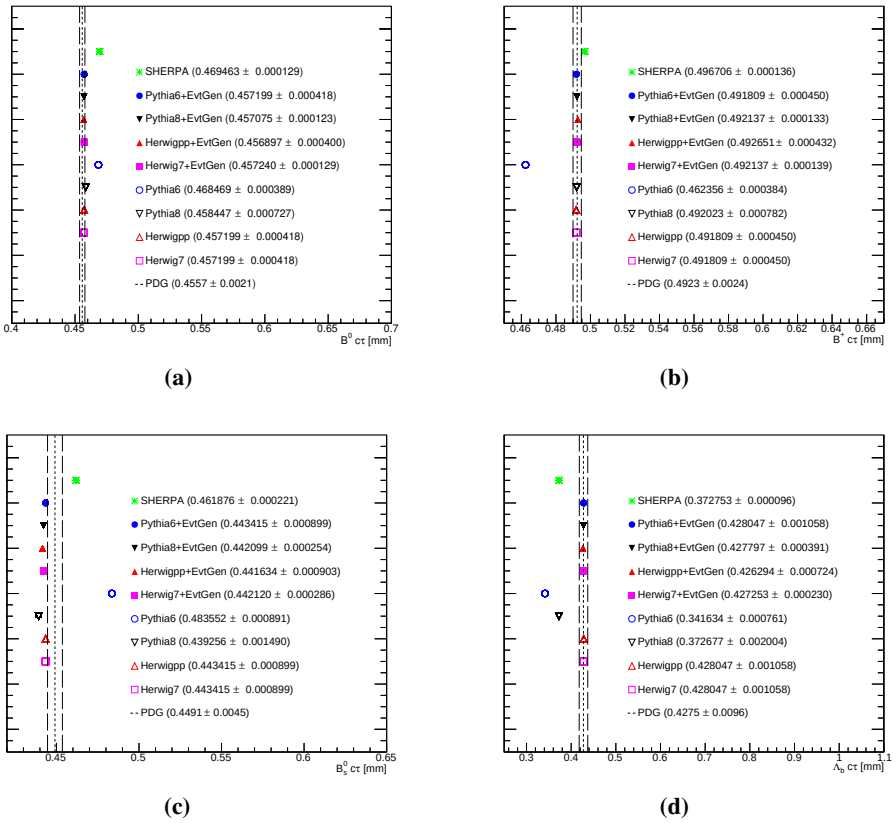


FIGURE 4.8: Comparison of lifetimes of weakly decaying bottom hadrons B^0 **(a)** , B^+ **(b)**, B_S^0 **(c)** and Λ_b **(d)** among five generators: SHERPA (v2.2.1), PYTHIA 6 (v6.428), PYTHIA 8 (v8.210), HERWIG++ (v7.0.1) and HERWIG 7 (v7.0.1) all with and without EVTGEN (with the exception of SHERPA).

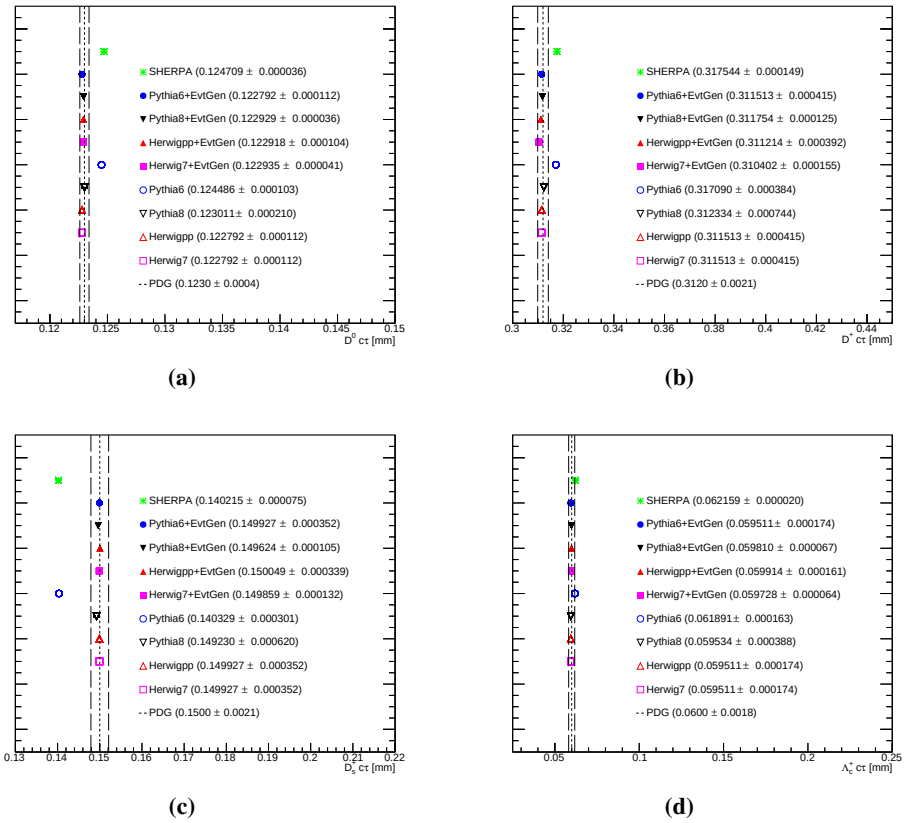


FIGURE 4.9: Comparison of lifetimes of weakly decaying charm hadrons D^0 **(a)**, D^+ **(b)**, D_S **(c)** and Λ_c **(d)** among five generator: SHERPA (v2.2.1), PYTHIA 6 (v6.428), PYTHIA 8 (v8.210), HERWIG++ (v7.0.1) and HERWIG 7 (v7.0.1) all with and without EVTGEN (with the exception of SHERPA).

CHAPTER 5

Object definitions and event reconstruction

"We are not lost. We're locationally challenged."
- John M. Ford.

Once data are selected by the trigger system (see Section 3.2.5), the information from all sub-detectors is gathered to build the "raw" event as already mention in Chapter 4. Raw data need to be converted from electronic pulses in the detector to physical objects. Such objects include photons, electrons, muons, jets (from quarks and gluons) and missing momentum. These objects form the basic elements of any analysis. This process is named reconstruction. This chapter give an overview of the objects used in the analyses presented in this thesis. A visualisation of how ATLAS detects particles is shown in Figure 5.1. This simplified illustration shows the main energy deposit of particles across the ATLAS detector [285]. For HL-LHC analysis particle-level objects instead of fully reconstructed objects are used. Section 5.8 gives an overview of the definitions.

5.1 Tracks and vertices

The trajectories of charged particles (tracks) are reconstructed mainly with the information from the ID. A charged particle passing though the ID will interact with the active sensors and produce a series of pulses in the sub-detectors system. These pulses are transformed into three-dimensional space coordinates, as space-points (see the coordinates definition in Chapter 3). Since the ID is submerged in a solenoidal magnetic field, charged-particles traces are curved trajectories. This allows to reconstruct the momentum and charge. It is important to recall that when particles interact with the material detector, multiple effects can occur, and these can induce deviations and/or energy loss. Effects such as ionisation, bremsstrahlung, Compton

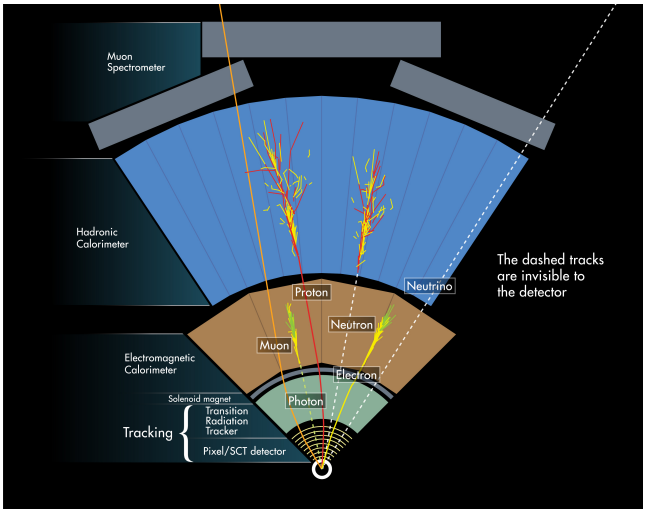


FIGURE 5.1: Diagram illustrating how different particles interactions with the layers of the ATLAS detector.

scattering and hadronic interactions with atomic nuclei induce the particles to lose energy. These effects have to be considered in the simulation and/or reconstruction processes as mentioned in Chapter 4.

The track reconstruction is based on two complementary algorithms, one initiates from the centre of the ID and works outwards (inside-out), the other starts in the TRT and works inwards (outside-in). The inside-out procedure starts with information provided by track seeds, formed from sets of few space-points using Pixel and SCT. The PV is also reconstructed with this inside-out procedure. Then these tracks are extended outwards to the TRT [286]. The inside-out method accounts for the majority of tracks reconstructed in ATLAS but it is complemented by an outside-in approach that starts with the TRT hits and moves inwards [287]. Track reconstruction based on the Kalman filter [288] is used to improve the performance. This algorithm progressively constrains material scattering points and applies Gaussian sum filters [289] for the description of energy loss due to bremsstrahlung. The inwards algorithm only considers tracks with $p_T > 400$ MeV to reduce the algorithm processing time.

Identification of the PV is needed in order to properly identify the hard scatter

processes. Thus, the vertices are defined by associating track origins with individual points and are optimised with a χ^2 function. The dominant source of fake vertices is the fictitious track association and inefficiency is guided by merged and split vertices[1]. The PV of an event is determined to have the largest Σp_T^2 of all associated tracks [290]. These algorithms apply cluster position measurements with precise knowledge of module alignments. These are regulary calibrated in reliance on track χ^2 optimisation as well as cosmic-ray data [291].

5.2 Electrons and photons

The reconstruction of electron and photon objects is performed by identifying deposits of energy in the electromagnetic calorimeter, and for electrons a particle track recorded in the ID [292, 293]. The clusters are required to be $|\eta_{\text{cluster}}| < 2.47$, excluding the transition region between the barrel and endcap EM calorimeters, $1.37 < |\eta_{\text{cluster}}| < 1.52$, which contains a large amount of inactive material and therefore limited instrumentation.

The signal from each EM cell is processed in intervals of 25 ns. The space is distributed into a grid of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ rad towers which combine the calorimeter layers. Cluster seeds by individual towers with energy above 2.5 GeV are searched for within the EM calorimeter middle layer [294]. Clusters are labelled as electron candidates if they are matched with a primary ID track or unconverted photon candidates if no suitable track exists[2]. Then, the clusters are constructed with 3×7 (5×5) central-layer cells in the barrel (endcap) for electron and converted photon, and 3×5 (5×5) cells in the barrel (endcap) for unconverted photon.

Since the photon is reasonably easy to identify, is almost background free. Electron identification, on the other hand, is more difficult due to other background, especially jets. Jets are always non-negligible background, since they have large cross-section.

[1] Split vertices result in a deterioration of the vertex track multiplicity and invariant mass. In order to mitigate splitting, nearby vertices are allowed to be merged into a single vertex.

[2] Converted photons are those produced when photon interacts with the detector material to create e^+e^- pairs.

To determine whether the reconstructed electron candidates are prompt electrons[3], electron candidates are required to have a transverse momentum of $p_T > 7$ GeV. Further requirements on the EM shower shape, calorimeter energy to track momentum ratio, and other discriminating variables are combined into a likelihood-based object quality requirement, optimised for strong background rejection. There are three levels of identification in this likelihood-based method, labelled "loose", "medium" and "tight" [295, 296]. The electrons selected with "tight" are a subset of the electrons identified as "medium", which they are a subset of "loose" electrons. The "loose" electron have a larger identification efficiency but also much larger background than tight.

Additionally, an isolation criteria is applied. Isolation is meant to reject candidates coming from other sources than prompt boson decays (hadrons faking an electron signature, heavy-flavour decays or photon conversions). The isolation criteria are implemented by cuts on the energy deposited in the calorimeter cells inside a cone of radius $\Delta R = 0.2$ around the electron and on the transverse momentum sum of the tracks in a cone of size 0.3. The efficiency of these criteria depends on the electron p_T, with better efficiencies to higher values of p_T. For instance, electron identification with isolation cut (including tight) applied is is 90% efficient for electrons with $p_T = 25$ GeV and 99% efficient for electrons with $p_T = 60$ GeV.

Electron tracks are also required to be consistent with originating at the beam line. The following requirements are applied: the significance of the transverse impact parameter d_0^{BL} must satisfy $|d_0^{BL}/\sigma(d_0^{BL})| < 5$ and the longitudinal impact parameter z_0^{BL} relative to the PV has to obey $|\Delta z_0^{BL} \sin\theta| < 0.5$ mm, where BL stands for beam line[4].

Finally, once the electrons and photons are reconstructed and identified, their energies are calibrated. The electron cluster energy and track momentum measurements are combined, improving the resolution at low energies [297, 298]. For the

[3]Prompt electrons (or muons) are the ones involved in the hard-scatter process, while, non-prompt electrons (or muons) are the ones from weak decay products of b-jets and c-jets.

[4]The transverse and longitudinal impact parameters d_0 and z_0 are defined to the closest point in a trajectory associated to a PV in the transverse plane and z-direction, respectively.

Run 2 analysis presented in Chapter 6, the electron candidates have to satisfy the condition to be identified as "tight" with $p_T > 30$ GeV and $|\eta| < 2.47$.

5.3 Muons

Muons candidates are reconstructed with the information from the ID and the MS, and occasionally from the electromagnetic calorimeter if the MS is not available [299–301]. Muon track candidates in the MS are combined to create local track segments, which are later combined to form the final MS track. There are four different types of muon definitions depending on how tracks are formed and which sub-detectors are used.

- The stand-alone (SA) muons, reconstructed in a range $2.5 < |\eta| < 2.7$ where the ID does not cover, they only have a track in the MS. The parameters of the muon track at the interaction point are determined by extrapolating the track back to the point of closest approach to the beam pipe.

- The combined (CB) muons are defined with tracks reconstructed independently of the ID and the MS. Then, these both tracks are combined. It is only valid for muons within the ID acceptance (i.e. $|\eta| < 2.5$).

- Segment tagged (ST) muons are reconstructed by starting from a track in the ID. If the extrapolated track can be matched to at least one local segment, the track is considered corresponding to a muon. This allows the possibility to increase the acceptance for muons which crossed only one MS chamber layer.

- Calorimeter-tagged (CM) muons are recognised by a track from the ID associated to an energy deposit in the calorimeter compatible minimum ionising particle. These type of reconstructed muons recover acceptance in the region where the MS is only partially instrumented.

Among the types described above the CB muons are the ones with the highest purity. Muons are classified into "loose", "medium" and "tight" categories, similarly

to electrons. Track isolation is implemented to reduce impurity of events where a muon is created from a quark decay.

Similarly to electrons, the significance of the transverse impact parameter d_0^{BL} must satisfy $|d_0^{BL}/\sigma(d_0^{BL})| < 5$ and the longitudinal impact parameter z_0^{BL} relative to the PV has to obey $|\Delta z_0^{BL} \sin\theta| < 0.5$ mm. In the Run 2 analysis described in this work, the muons must have $p_T > 30$ GeV and $|\eta| < 2.5$, and satisfy the "medium" identification criteria.

5.4 Jets

Jets are conical showers of hadrons that come from an initial particle production. They are composed of neutral and charged particles and deposit most of their energy in the hadronic calorimeter with extensive showers[5]. The jets are then reconstructed using the anti-k_t algorithm [284]. This algorithm iterates over a set of topo-cluster[6] objects to evaluate the separation d_{ij} for all pairs of objects detected i, j and the distance measure d_i^B of each object i to the beamline:

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

$$d_i^B = \frac{1}{k_{t,i}^2},$$

where $k_{t,i}$ is the transverse momentum of the *ith* object and R is the distance parameter, which defines a cone of the size of the jets. Usually the cone size is selected to be 0.4 or 0.6, though the most standard used in ATLAS is 0.4.

The process continues until it finds $d_i^B < d_{ij}$, then the jet is reconstructed and removed from the list of entities. The procedure is repeated until all entities are recognised as jets or merged into jets.

[5] Contrary to electrons and photons, where large percentage of their energy is stored in a narrow electromagnetic calorimeter shower.

[6] In ATLAS, the jet reconstructions are typically locally calibrated topo-clusters built from noise-suppressed topological groupings of calorimeter cell.

If around a cone of radius 2R of a high-p_T (hard) object there is not another hard object, then the low-p_T (soft) objects surrounding in a cone of radius R will be merged with it. If two hard objects are identified around R and 2R of distance, then their relative momenta will define the position of the boundary between them and determines the soft objects located to the nearest hard one. This creates jets with conical edges, making their energy calibration easier [284].

The average energy contribution from pile-up is subtracted according to the jet area and the jets are calibrated as described in Ref. [302]. To further reduce the effect of pile-up interactions, the jets with $|\eta| < 2.8$ and $p_T < 120$ GeV are required to satisfy the "medium" working point of the jet vertex tagger (JVT), a tagging algorithm that identifies jets originating from the PV using track information [303, 304]. In this thesis, the baseline jets are selected in the region $|\eta| < 4.5$ and having $p_T > 20$ GeV. Forward jets are required to have $p_T < 50$ GeV and $|\eta| > 2.5$.

Jets originating from the hadronisation process of b-quarks are identified with multivariate discriminants based on track and secondary vertex observables [305–307]. This is known as b-tagging. There are several efficiency levels implemented in the b-tagging algorithm, the one used in this thesis for the Run analysis is the 77% efficiency as implemented in MV2c10 tagging algorithm. This b-flavour jet efficiency has a c-jet rejection rate ~ 6, i.e. 1/6 of all c-jets are wrongly reconstructed as b-jets. The light jet rejection is estimated to be 134. Since b-taggers need high-p_T tracks ($p_T > 20$ GeV) and as the use of information from the ID, b-tagged jets can be only reconstructed in the region $|\eta| < 2.5$.

5.5 Overlap removal

The reconstruction of the same energy deposits as multiple objects may happen and it is resolved using the following overlap removal criteria. When two electrons share a track, the electron with lower-p_T is removed. An electron sharing a track with a muon is removed to avoid cases where a muon mimics an electron through radiation of a hard photon. A CM muon is removed if it shares a track with an electron in order to minimise electron fakes. Jets overlapping with selected electron candidates within an η–ϕ cone of size $\Delta R = \min(0.4, 0.04 + 10\ \text{GeV}/p_T)$ are removed from the

event to reduce the proportion of electrons being reconstructed as jets. Any electron found close to non-pile-up jet within a cone of radius $\Delta R = 0.4$ is also removed, in order to reduce backgrounds from non-prompt, non-isolated electrons coming from heavy-flavour hadron decays. Any non-b-jet with less than three tracks originating from the PV are removed if found within a cone of radius $\Delta R < 0.2$ from a muon or if it has a muon ID track ghost-associated to it. This is to reduce fake jets from muons depositing energy in the calorimeters. Finally, similarly to electrons, muons with a distance $\Delta R = \min(0.4, 0.04 + 10\ \text{GeV}/p_T)$ from any of the surviving jets are removed to avoid contamination of non-prompt muons from heavy-flavour hadron decays. All these criteria is summarised in Table 5.1.

Reject	Against	Criteria
electron	electron	shared track, $p_{T,1} < p_{T,2}$
muon	electron	is calo-muon and shared ID track
electron	muon	shared ID track
jet	electron	not a b-jet and $\Delta R < 0.2$
electron	jet	$\Delta R < \min(0.4, 0.04 + 10\ \text{GeV}/p_T^{\text{electron}})$
jet	muon	not a b-jet and NumTrack < 3 and (ghost-associated or $\Delta R < 0.2$)
muon	jet	$\Delta R < \min(0.4, 0.04 + 10\ \text{GeV}/p_T^{\text{muon}})$

TABLE 5.1: Summary of the overlap removal criteria.

5.6 Missing transverse momentum

The transverse momentum of an event is defined as the vector sum of photons, electrons, muons and jets, known as hard physics object; and soft term objects, which are calculated from tracks emerging from the hard-scatter vertex but not associated a hard physics object [308]. The sum of the vector transverse momenta of all the products of a collision should be zero due to the conservation of momentum in the transverse plane. The missing transverse momentum (with a magnitude E_T^{miss}) is defined as the negative sum of all detected particles and represents the sum all non detected objects: long-lived final states such as neutrinos or DM candidates which

do not interact with the ATLAS detector as well as out-of-acceptance objects such as those at high η. Each contribution from these objects is added in a separated term and summed in E_T^{miss} variable as shown in the following equations:

$$\vec{p} = \sum_{\gamma} \vec{p}^{\gamma} + \sum_{e} \vec{p}^{e} + \sum_{\mu} \vec{p}^{\mu} + \sum_{jet} \vec{p}^{jet} + \sum_{soft} \vec{p}^{soft},$$

$$E_{x,y}^{miss} = -p_{x,y},$$

$$E_T^{miss} = \sqrt{p_x^2 + p_y^2}.$$

The uncertainties involved in the E_T^{miss} variable not only contain the uncertainties of the reconstructed objects but those related to soft terms. Their influence contains the modelling of the UE and its impact on the p_T scale and resolution of unclustered energy [308]. In Run 2 analysis presented in this text, the minimal requirement in E_T^{miss} is 200 GeV.

5.7 Trigger

The ATLAS event trigger system is composed by the hardware-based L1 trigger and the software-based HLT (described in Section 3.2.5). The trigger selection is applied in simulation as well, and correction factors are included to reproduce their performance in data. Two main types of triggers are used in the Run 2 analysis, a lepton-based and E_T^{miss}-based triggers.

To select leptons there are some requirements that the event has to obey. For instance, the electron triggers select a calorimeter cluster matched to a track. To do this identification, the electrons must satisfy a criteria based on a multivariate technique using a likelihood discriminant. In 2015 data-taking, the electrons were triggered by requiring at L1 a transverse energy deposit E_T^{miss} above of 20 GeV, a reduced calorimeter granularity is used at this step. Then, for HLT stage, the full granularity of the calorimeter and the tracking information are available. The reconstructed calorimeter cluster is matched to a track at this point. The trigger electron object is

then required to be isolated with medium identification and to have E_T^{miss} > 24 GeV. In 2016–2018, electrons have to satisfy a tight identification at HLT together with an isolation criteria and have the E_T^{miss} > 26 GeV. The E_T^{miss} threshold varies slightly as a function of η to compensate for passive material in front of the calorimeter [207]. Two additional trigger requirements are used to avoid efficiency losses due to identification and isolation at high p_T. Those requirements include a selection of loose electrons with E_T^{miss} > 120 GeV in 2015 and E_T^{miss} > 140 GeV in 2016–2018.

Muons are triggered by tracks with a match between MS and ID. In 2015, muons had to satisfy a medim isolation requirement and p_T > 20 GeV. At a transverse momentum above 50 GeV this trigger is complemented by a trigger which does not require isolation, such the some inefficiency is recovered for high transverse momentum region [207]. In 2016–2018, the isolation criterion was tightened and the threshold increased to p_T > 26 GeV. As for electron triggers, during all these years, in order to avoid efficiency losses due to isolation at high p_T, another muon trigger without any isolation requirement was used, selecting loose muons with p_T > 50 GeV.

In addition to lepton trigger, E_T^{miss} triggers were also used. This trigger is calculated based on the calorimeter cells. The E_T^{miss} triggers are obtained by summing over all jets reconstructed using the anti-k_t algorithm with a radius of 0.4, after a Local Cluster Weight (LCW) hadronic-scale calibration.

For Run 2 analysis, events with E_T^{miss} < 250 GeV pass either lepton triggers or E_T^{miss} triggers. Only E_T^{miss} trigger is used when events have E_T^{miss} > 250 GeV. The systematic uncertainties that are ingrained in the object reconstruction for the Run 2 analysis are detailed in Chapter 6.

5.8 HL-LHC particle-level object definitions

The HL-LHC is a major upgrade of the LHC which will enable the LHC to collect up to about 3000 fb^{-1} of data (total expected for $\sim$ 2037).

The detector response of the upgraded ATLAS detector is emulated applying smearing functions to simulated final-state particles at truth level[7]. Thus particle-level definitions are used for electrons, muons, jets and E_T^{miss}, which are the final-state

[7] A truth level simulation is called when MC does not contain experimental effects.

objects used by HL-LHC analysis. The smearing functions applied on the pile-up scenario emulate the reconstruction and the energy calibration expected for HL-LHC environment. These functions are described in Ref. [309].

Particle-level charged leptons are defined as electrons and muons that originate from a W boson decay, including those emerging from a subsequent tau decay. The leptons from hadron decays either directly or via a tau decay are rejected. The selected leptons (electrons and muons) are dressed[8] with photons within a cone of size $\Delta R = 0.1$, which implies that their final four-momenta are the vector sum of the dressing photons and their original lepton four-momenta.

In order to simulate the electron/muon track match requirement (i.e. the overlap removal between electrons and muons), events are rejected if a matching in ϕ and θ of 0.005 is found between these two particle-level objects. Identification efficiencies [311] are applied to the lepton candidates emulate detector response. These have their energy, p_T and η smeared according to the detector resolution. The contribution due to pile-up is taken into account before applying detector resolution effects.

Particle-level neutrinos are required, similarly to electrons and muons, not to originate from a hadron or quark decay. The E_T^{miss} is calculated from the negative vector sum of true final-state particles within the detector acceptance.

Particle-level jets are reconstructed using the anti-k_t algorithm [284, 304] implemented in the Fastjet [304] algorithm, with a radius parameter of 0.4, similarly to the reconstructed jets used in the Run 2 analysis. All stable particles are used to reconstruct the jets, except the selected leptons, neutrinos and the photons associated with these leptons.

The b-tagging identification is applied for jets within $|\eta| < 2.5$ and a parametrised efficiency tagging is used, function of the true flavour of the jet, p_T and η. Since the b-tagging is particularly sensitive to the contamination of pile-up tracks, tracks with large impact parameters are considered. Therefore tracks from nearby pile-up are likely to be selected in order to mitigate effects from pile-up. These efficiencies are

[8] To define dressed leptons, a cone or a jet algorithm can be used to cluster all photons around the direction of a lepton after partial QED radiation recovery [310].

evaluated considering the latest layout of the ITk detector [198] using the MV2 b-tagging algorithm [204, 306, 307] at the 70% working point.

The criterion to avoid double counting of electrons and jets is similar to Run 2 analysis. As HL-LHC analysis is performed using only simulated processes, a fraction of the particle-level jets are removed, according to the expected mis-identification rate shown in Ref. [312]. Energy, $\mathrm{p_T}$ and η of remaining jets are smeared according to the detector resolution. Pile-up jets are rejected using tracking information. The expected systematic uncertainties for the HL-LHC analysis are detailed in the Chapter 7.

CHAPTER 6

2HDM+a mono-top search using Run 2 dataset

"We live on an island surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance."
- John Archibald Wheeler.

This chapter describes the search for the single top-quark production associated with a DM candidate, within the 2HDM+a model, described in Chapter 2. The analysis makes use of data collected with the ATLAS detector at $\sqrt{s} = 13$ TeV during LHC Run 2 (2015–2018) corresponding to an integrated luminosity of 139 fb^{-1}.

This chapter is organised as follows: it starts with an introduction to the basic characteristics of the selected signature in Section 6.1. The event selection is discussed in Section 6.2. Section 6.3 gives a detailed analysis strategy, introducing the multivariate analysis definition and evaluation. Background estimation is presented in Section 6.4. Section 6.5 shows the optimised signal region defined to perform this analysis. The source of systematic uncertainties considered are listed and explained in Section 6.6. To extract the main results of this analysis an hypothesis test is performed. The procedure can be seen in Section 6.7. To finalise this chapter, Section 6.8 discusses the main results.

6.1 Experimental signature

The tj signature (also named tj channel) in the 2HDM+a model is defined containing the single top-quark production associated with DM candidates and a hadronic jet. The origin of this process in the framework of a 2HDM+a model is described in Chapter 2, and the relevant diagrams are shown in Figure 6.1. This signal signature

originates from the interplay between the charged Higgs boson ($H^{\pm}$), the W boson, the top quark and the pseudo-scalar a (i.e. DM mediator). Two processes allow this production, the *t*-channel (Figures 6.1(**a**)-(**b**)) and the *s*-channel (Figures 6.1(**c**)-(**d**)). Both Figures 6.1(**b**)-(**c**) involve the $H^{\pm}$ boson, so the cross-section of these diagrams will depend on the $m_{H^{\pm}}$. Figures 6.2(**a**)-(**b**) shows the cross-section dependency of the channels as a function of $\tan\beta$.

The *s*-channel production has in overall very small cross-section, but its contribution can dominate the parameter space where $m_{H^{\pm}} - m_a < m_W$[1]. Looking at the dependence of the cross-section for both production times the *s*- and *t*-channel in Figure 6.2, it can be seen that *t*-channel has a large dependency on $\tan\beta$ with a minimum at $\tan\beta = 5$.

Figures 6.2(**c**)-(**d**) present the *t*-, *s*-channel, and tj cross-section behaviour for different values of $m_{H^{\pm}}$. The shape (kink) at masses below 500 GeV for the *t*-channel is caused by the interference between the different diagrams. Figure 6.1(**b**) has a minimum value around 400 GeV. It can be noticed that contrary to *s*-channel, *t*-channel increases for larger values of $m_{H^{\pm}}$, this is true independently of the β value.

For signals with $m_a = 250$ GeV and $\tan\beta = 0.3$ ($\tan\beta = 0.5$), the dominant process contributing for the tj cross-section is the *t*-channel with a relative contribution ranging from 76% (51%) for a $m_{H^{\pm}} = 300$ GeV to a 99% (99%) when $m_{H^{\pm}} = 1750$ GeV. However, the *s*-channel production is larger for tj signals when $m_{H^{\pm}} = 300$ GeV and $1 \leq \tan\beta \leq 20$, contributing from 80% when $\tan\beta = 1$ to 56% for $\tan\beta = 20$.

Sensitivity studies for both production channels are performed, applying individual optimisation, and show that *s*-channel production is not sensitivity enough. Dedicated studies are presented in Appendix B. Thus, the analysis described in this chapter will concentrate only on the *t*-channel signature. This signature will be referred to as either tj or *t*-channel in this chapter.

[1] This statement is only valid for Feynman diagrams shown in Figures 6.1(**b**)-(**c**)

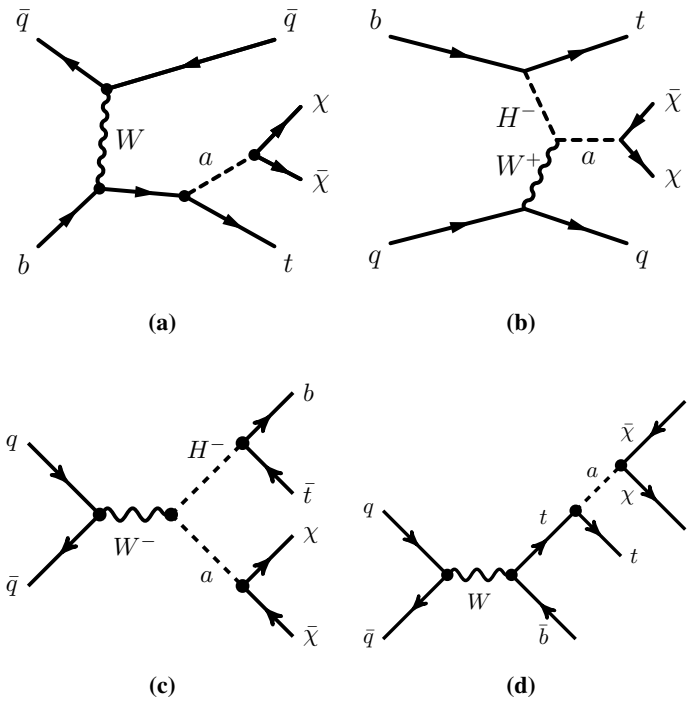


FIGURE 6.1: Feynman diagrams for the dominant production and decay modes for the 2HDM+a model: t-channel: **(a)**-**(b)**, and s-channel: **(c)**-**(d)** in single top-quark final states.

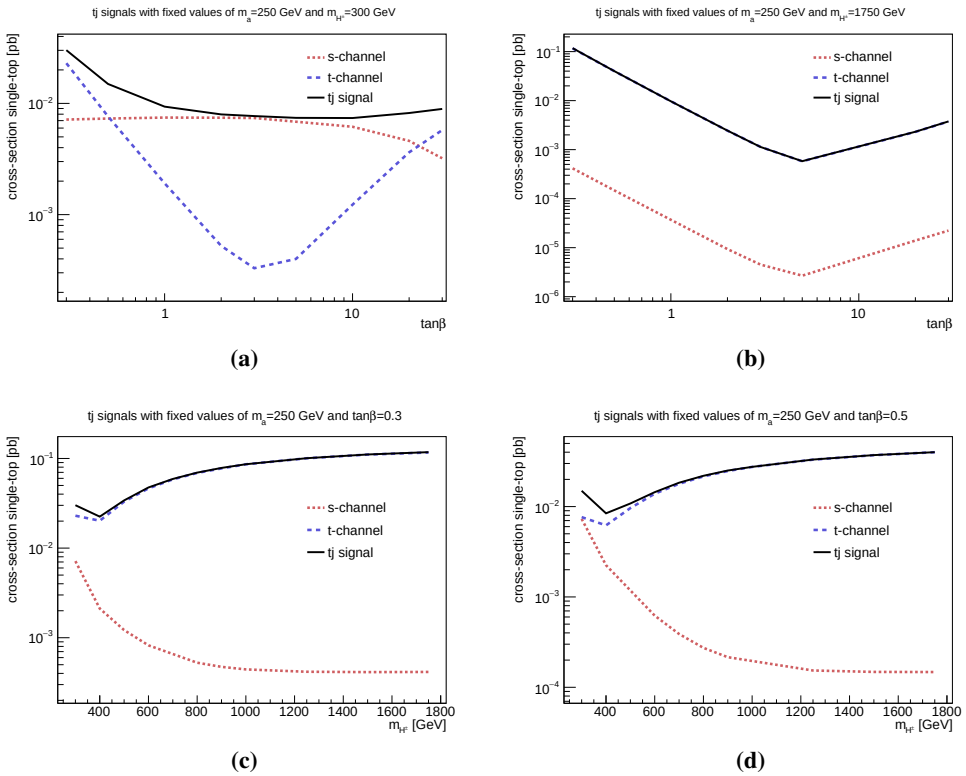


FIGURE 6.2: Production of the decay modes for the 2HDM+a model: *t*-channel (blue dashed line) and *s*-channel (red dotted line) in the tj (black line) in single top-quark final states. The parameter space is covered by the free-parameters proposed in this work, $\tan\beta$ and $m_{H^{\pm}}$.

6.2 Pre-selection

The signal has the signature of one lepton (e or μ) from the top-quark decay chain, a b-tagged jet, E_T^{miss} from the decay of the W boson, the DM particles, and additional jets that can be present.

As an initial step, events having at least one electron or muon with $p_T > 30$ GeV and $|\eta| < 2.5$ are required. Events are selected with at least one jet but not more than four with a p_T cut of 30 GeV, one or two of which must be b-tagged. The fourth jet of the event, if present, is required to have a p_T value less than 50 GeV[2]. The p_T threshold of the highest b-tagged jet (b-jet 1) is required to have $p_T > 50$ GeV, if a second one exits it must have at least $p_T > 30$ GeV. Additionally, the requirement $|\Delta\phi_{min}| > 0.5$ rad is applied, with $|\Delta\phi_{min}|$ being the smallest azimuthal angular distance between the E_T^{miss} and the four leading signal jets. This requirement helps to discriminate against events with mismeasured E_T^{miss}.

In order to maximise the sensitivity of this study, and in addition to the previous requirements, further discrimination is obtained by applying additional selection criteria according to the kinematic properties of the signal.

The E_T^{miss} is required to be higher than 200 GeV, which allows to reduce further background processes, where lower E_T^{miss} is expected. Another important variable is $m_T(\ell, E_T^{miss})$, the transverse mass of the lepton–E_T^{miss} system defined as follows:

$$m_T(\ell, E_T^{miss}) = \sqrt{2 p_T(\ell) E_T^{miss} \left(1 - \cos\Delta\phi(\ell, E_T^{miss})\right)}, \tag{6.1}$$

where $p_T(\ell)$ denotes the magnitude of the lepton transverse momentum and $\Delta\phi(\ell, E_T^{miss})$ is the azimuthal difference between the lepton momentum and the E_T^{miss} directions. This variable is requested to be larger than 60 GeV. In background events the spectrum of this quantity falls rapidly for values higher than the W-boson mass. A summary of the pre-selection criteria is presented in Table 6.1. This set of cuts defines the pre-selection region. The main background for this pre-selection region is the $t\bar{t}$ with a contribution of 61.8%, followed for W+jets with 26.1%.

[2]It is designed to remove background from the $t\bar{t}$ processes, without a sizeable reduction on signal.

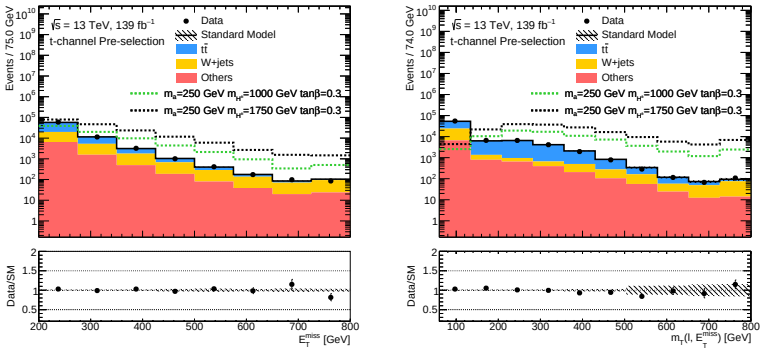
Variable	Requirement
number of jets, $\mathcal{N}_{\text{jets}}$	$1 \leq \text{N} \leq 4$
number of b-jets, $\mathcal{N}_{\text{b–jets}}$	$1 \leq \text{N} \leq 2$
number of leptons, $\mathcal{N}_{\ell}$	1
$\text{p}_{\text{T}}^{\text{jet 1, 2, 3, 4}}$ [GeV]	> 30
$\text{p}_{\text{T}}^{\text{jet 4}}$ [GeV]	< 50
$\text{p}_{\text{T}}^{\text{b-jet 1}}$ [GeV]	> 50
$\text{p}_{\text{T}}^{\text{b-jet 2}}$ [GeV]	> 30
$\text{p}_{\text{T}}(\ell)$ [GeV]	> 30
$\lvert\Delta\phi_{\text{min}}\rvert$ [rad]	> 0.5
$\text{m}_{\text{T}}(\ell, \text{E}_{\text{T}}^{\text{miss}})$ [GeV]	> 60
$\text{E}_{\text{T}}^{\text{miss}}$ [GeV]	> 200

TABLE 6.1: Summary of the pre-selection region requirements for the tj channel.

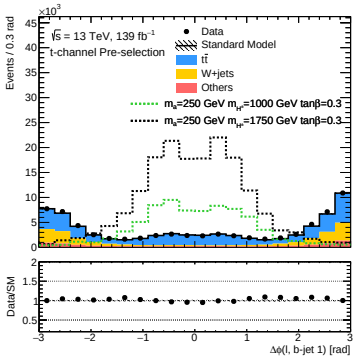
Further variables are defined in order to understand the signal behaviour. Table 6.2 shows the full list of variables. These variables are used to discriminate between the differences of kinematic properties for background and signal. The mass related variables, such as $\text{m}(\text{b-jet 1}, \ell)$ and $\text{m}(\text{b-jet 1}, \ell, \text{E}_{\text{T}}^{\text{miss}})$ are calculated from the four-momentum sum of the components shown in the parentheses. For the $\eta_{(\ell,\text{b-jet 1})}$ variable, η is calculated from the four-momentum sum between lepton and b-jet 1. The main variables of this analysis are shown in Figures 6.3-6.4, the remaining are presented in the Figures 6.5-6.6. All SM backgrounds are stacked on top of each other and scaled to their respective cross sections. The data is shown in black and a potential signal distribution is scaled to same number of events as background.

Variable name	**Description**
Kinematic variables	
$\mathrm{E_T^{miss}}$	Magnitude of the missing transverse momentum
$\mathrm{m_T}(\ell, \mathrm{E_T^{miss}})$	Transverse mass of lepton–$\mathrm{E_T^{miss}}$ system
$\mathrm{p_T}(\text{jet 1})$	Transverse momentum of jet 1
$\mathrm{m}(\text{b-jet 1}, \ell)$	Mass of the b-jet 1 and lepton
$\mathrm{m}(\text{b-jet 1}, \ell, \mathrm{E_T^{miss}})$	Mass of the b-jet 1, lepton and $\mathrm{E_T^{miss}}$-system
$\eta^{\text{jet 1}}$	η of jet 1
$\eta_{(\ell,\text{b-jet 1})}$	η of the lepton and b-jet 1
Azimuthal differences	
$\Delta\phi(\ell, \text{b-jet 1})$	$\Delta\phi$ between the lepton and b-jet 1
$\Delta\phi(\ell, \mathrm{E_T^{miss}})$	$\Delta\phi$ between the lepton and $\mathrm{E_T^{miss}}$
Angular distance differences	
$\Delta\mathrm{R}(\ell, \text{b-jet 1})$	$\Delta\mathrm{R}$ between the lepton and b-jet 1

TABLE 6.2: List of all the discriminating variables and their definitions used in the *t*-channel analysis.

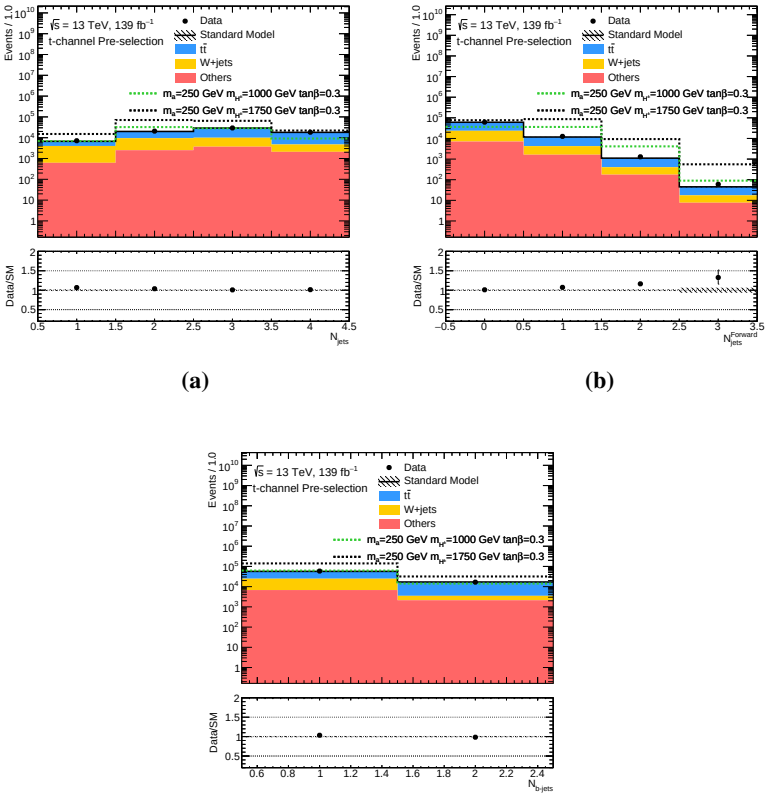


(a) **(b)**



(c)

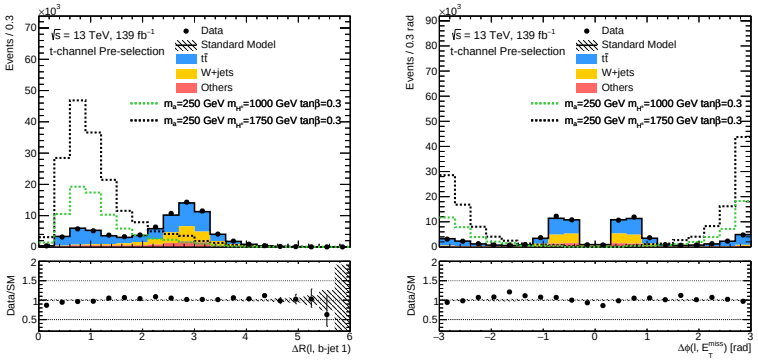
FIGURE 6.3: Distributions at pre-selection-level of *t*-channel analysis for the $\mathrm{E_T^{miss}}$ **(a)**, $\mathrm{m_T}(\ell, \mathrm{E_T^{miss}})$ **(b)** and $\Delta\phi(\ell, \mathrm{b\text{-jet}\ 1})$ **(c)**. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $\mathrm{t\bar{t}V}$, Z+jets, $\mathrm{t\bar{t}h}$ and tWZ) is the combination of all the backgrounds not including $\mathrm{t\bar{t}}$ and W+jets. The SM backgrounds and the signals correspond to the simulation predictions normalised to the theoretical predictions. The last bin contains the overflow events.



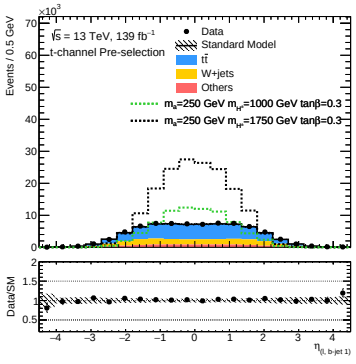
(a) **(b)**

(c)

FIGURE 6.4: Distributions at pre-selection-level of t-channel analysis for the number of jets ($\mathcal{N}_{\text{jets}}$) **(a)**, number of forward jets ($\mathcal{N}_{\text{jets}}^{\text{Forward}}$) **(b)** and number of b-jets ($\mathcal{N}_{\text{b-jets}}$) **(c)**. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. The SM backgrounds and the signals correspond to the simulation predictions normalised to the theoretical predictions.

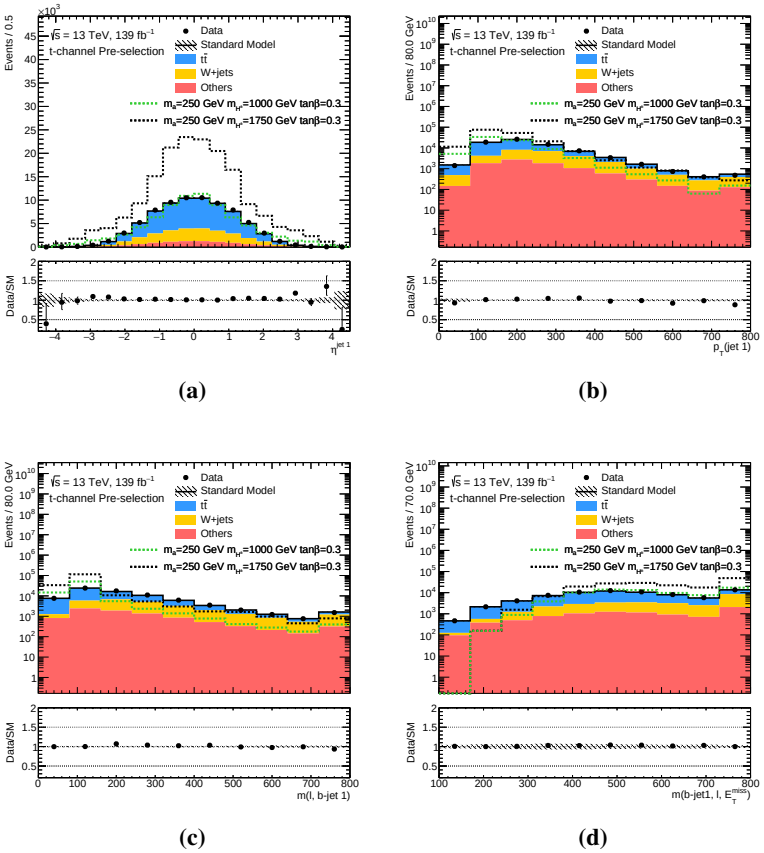


(a) **(b)**



(c)

FIGURE 6.5: Distributions at pre-selection-level of *t*-channel analysis for the $\Delta R(\ell, \text{b-jet 1})$ **(a)**, $\Delta\phi(\ell, E_T^{\text{miss}})$ **(b)** and $\eta_{(\ell,\text{b-jet 1})}$ **(c)**. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. The SM backgrounds and the signals correspond to the simulation predictions normalised to the theoretical predictions.



(a) **(b)**

(c) **(d)**

FIGURE 6.6: Distributions at pre-selection-level of t-channel analysis for the $\eta^{\text{jet 1}}$ **(a)**, $\mathrm{p_T(jet\ 1)}$ **(b)**, $\mathrm{m(b\text{-}jet\ 1, \ell)}$ **(c)** and $\mathrm{m(b\text{-}jet\ 1, \ell, E_T^{miss})}$ **(d)**. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $\mathrm{t\bar{t}V}$, Z+jets, $\mathrm{t\bar{t}h}$ and tWZ) is the combination of all the backgrounds not including $\mathrm{t\bar{t}}$ and W+jets. The SM backgrounds and the signals correspond to the simulation predictions normalised to the theoretical predictions. The last bin contains the overflow events.

6.3 Boosted Decision Tree

Due to the low cross-section of the *t*-channel production a cut-based analysis is not sufficient to achieve a good discrimination of the signal. An approach based on multivariate analysis techniques is applied to enhance signal-background separation. A Boosted Decision Tree (BDT) [313] classifier is trained specifically to separate the *t*-channel signal from all backgrounds, using events that pass the pre-selection criteria listed in Table 6.1. A dedicated Appendix A highlights the procedure and its characteristics. The backgrounds are trained against seven different signal points, selected from the region with largest *t*-channel signal production:

- m_a = 250 GeV
  - $m_{H^\pm}$ = 1000 GeV, 1250 GeV both with $\tan\beta = 1$,
  - $m_{H^\pm}$ = 1250 GeV, 1500 GeV, 1750 GeV; all with $\tan\beta = 0.5$,
  - $m_{H^\pm}$ = 1500 GeV, 1750 GeV both with $\tan\beta = 0.3$.

To train the BDT, all signal event points together with background events are considered as the full set of events, using the k-fold cross validation method [314]. This strategy is based on dividing this set of events in 80% (for training) and 20% (for testing). These events are randomly split based on their event number. A cross-validation is performed splitting the events into five sets. For each of the BDTs, training is done on 80 % of events and tested in the remaining 20%. Notice that each of the testing samples is statistically independent from the others, as well as independent from the events used in the corresponding training. Collision data are not involved, neither in the training nor in the test part. Figure 6.7 shows an illustration of the BDT testing-training procedure.

Each of the resulting BDTs is evaluated on events that were never used for training that BDT, which allows the use of the full statisical power of the sample. The variables entering on the BDT training are:

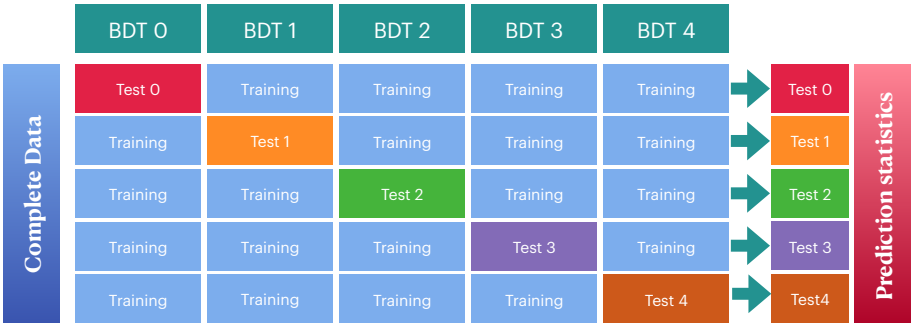


FIGURE 6.7: Representative scheme of the BDT strategy used involving 80% training and 20% testing. The events are randomly split based on their event number. This procedure is applied to obtain BDTs 0 to 4.

- $m(\text{b-jet } 1, \ell)$,
- $\Delta\phi(\ell, \text{b-jet } 1)$,
- $m_T(\ell, E_T^{miss})$,
- $m(\text{b-jet } 1, \ell, E_T^{miss})$,
- $\Delta\phi(\ell, E_T^{miss})$,
- $p_T^{\text{jet } 1}$,
- $\eta_{(\ell,\text{b-jet } 1)}$,
- $\Delta R(\ell, \text{b-jet } 1)$,
- $\eta^{\text{jet } 1}$.

The distributions of the variables entering in the training procedure are shown in Figure 6.8, for both signal and background, these distributions are normalised to one. Their definitions are shown in Table 6.2.

The choice of the parameters of the BDT are studied during an optimisation phase: the maximum depth of each tree, the minimum percentage of events at each node (minimum node size), the number of trees (i.e. number of iterations), the parameter which controls the gradient descent of the boost (learning rate β, usually named shrinkage) and the granularity number of cuts (number of cuts). The parameters which control the behaviour of the BDT can be tuned to obtain the best performance. The selected BDT parameters are listed in Table 6.3. The purpose of each value is detailed in Appendix A.

Figure 6.9 shows the BDTs distribution for training and testing event samples. These distributions offer a check of overtraining. Overtraining is a situation where in machine learning model can predict traininig data with high accuracy but can not

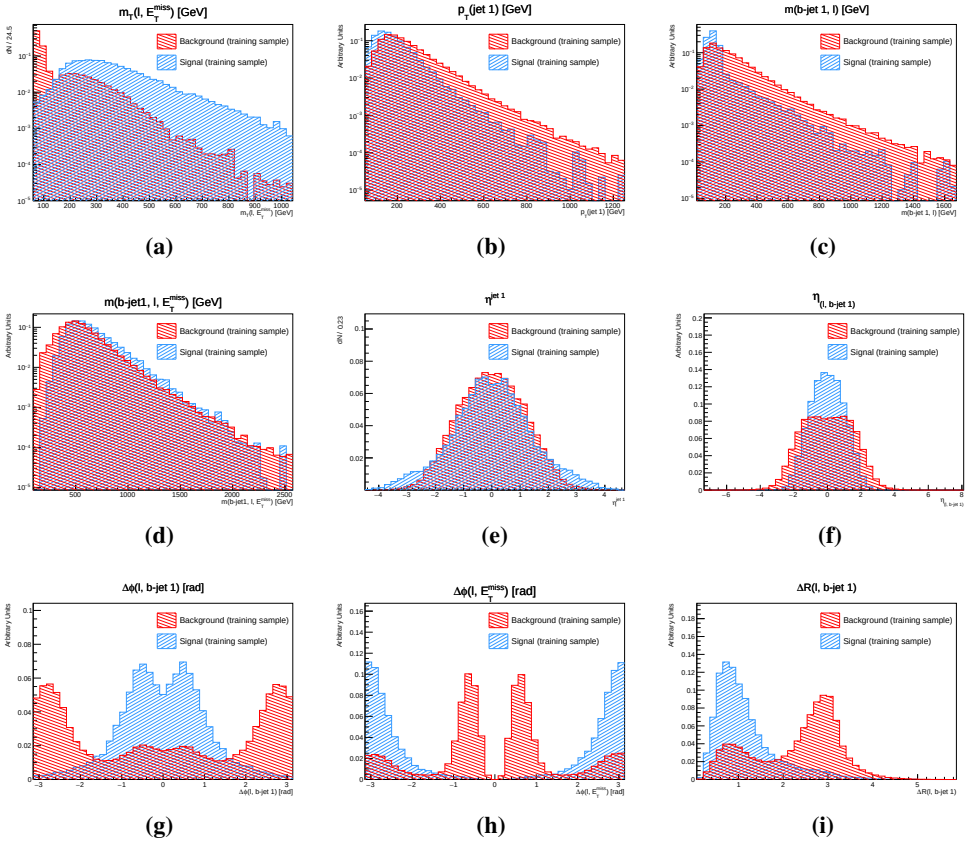


FIGURE 6.8: Distributions of signal and background processes among variables used by the BDT in the training. There are shown the $m_T(\ell, E_T^{miss})$ **(a)**, $p_T(\text{jet } 1)$ **(b)**, $m(\text{b-jet } 1, \ell)$ **(c)**, $m(\text{b-jet } 1, \ell, E_T^{miss})$ **(d)**, $\eta^{\text{jet } 1}$ **(e)**, $\eta_{(\ell,\text{b-jet } 1)}$ **(f)**, $\Delta\phi(\ell, \text{b-jet } 1)$ **(g)**, $\Delta\phi(\ell, E_T^{miss})$ **(h)** and $\Delta R(\ell, \text{b-jet } 1)$ **(i)**.

generalise to new data, performing poorly on these. The ratio in the bottom panel of each plot shows that there is no overtraining observed, since performance for training and testing samples is similar.

The correlation coefficients between pairs of input variables for the *t*-channel analysis are shown in Figure 6.10. Appendix A details the linear correlation coefficients between variables and its interpretation. It shows the correlation among the different

Parameter	Setting
BoostType	GradBoost
Number of trees	1500
Minimum node size	2.5%
Maximum depth of tree	5
Number of cuts	20
Learning rate β	0.20

TABLE 6.3: BDT parameter settings optimised for the search discussed in this analysis.

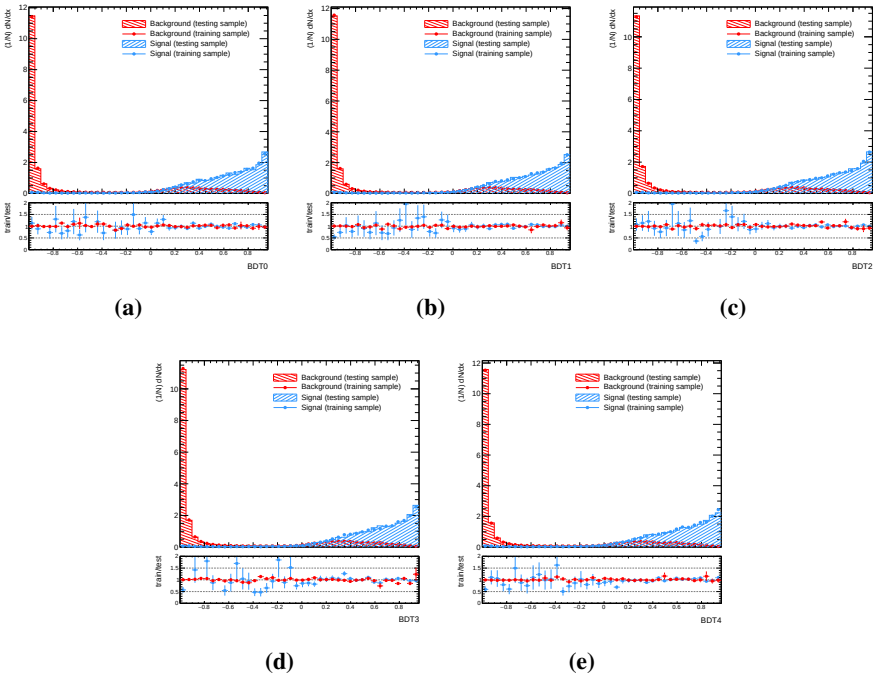
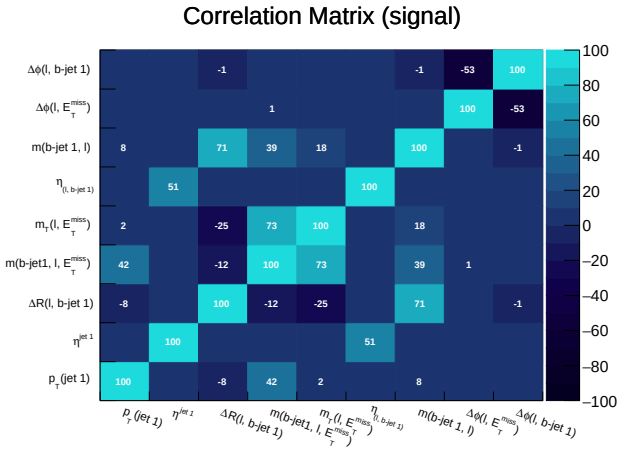


FIGURE 6.9: Overtraining test for the five BDTs. Is is shown the distributions of the BDT scores obtained for background and signal events after the training phase (dotted markers) and testing phase (filled area).

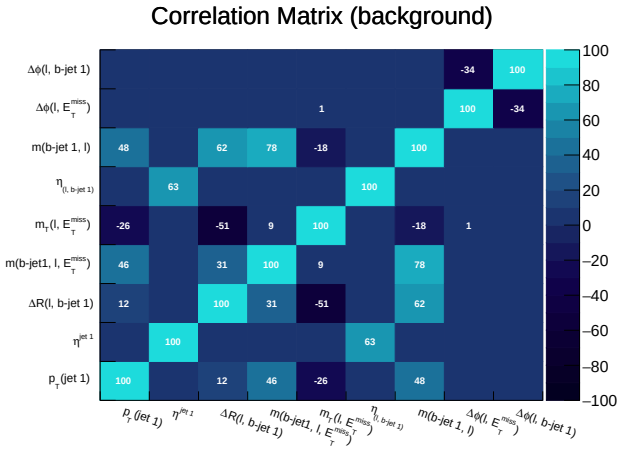
training variables (training sample) for signal and background processes. For signal, a moderate correlation between $\eta_{(\ell,\text{b-jet 1})}$ and $\eta^{\text{jet 1}}$ variables (linear coefficient = 51) is seen, since the signal is characterised by having one b-jet, and most of the time the leading jet and the b-jet are the same object (since the b-jet is quite boosted). The $\Delta R(\ell, \text{b-jet 1})$ and $m(\text{b-jet 1}, \ell)$ variables (linear coefficient = 71), denoting a natural correlation given by their definition (since they are both calculated by the same objects). Indeed this moderate correlation shows the behaviour of the *t*-channel signal, the selected lepton and the b-jet are originated from the one boosted top quark. The same idea is valid for the $m_T(\ell, E_T^{miss})$ and $m(\text{b-jet 1}, \ell, E_T^{miss})$ variables (linear coefficient = 73). Finally the $\Delta\phi(\ell, \text{b-jet 1})$and $\Delta\phi(\ell, E_T^{miss})$ variables (linear coefficient = -53), the correlation is affected for the signal topology, the b-jet and the E_T^{miss} are final states of the signal production. It is worth to mention the correlation for the background between the $\Delta R(\ell, \text{b-jet 1})$ and $m_T(\ell, E_T^{miss})$ variables (linear coefficient = -51). This effect is observed since the main background, the $t\bar{t}$ production, has a large contribution when $\Delta R(\ell, \text{b-jet 1}) \sim 3$, and at low $m_T(\ell, E_T^{miss})$. This occurs, since for events the leading jet does not originate from the same top quark as the lepton (since the other top quarks gives 3 jets, and the leptonic topology only yields 1 jet). However if the leading jet is the b-jet from the top quark, $\Delta R(\ell, \text{b-jet 1})$ most probable around 1, this produces a small bump at that value which can be seen in Figure 6.5a. No strong correlation are observed.

Finally, Figure 6.11 shows the signal efficiency versus background rejection (i.e ROC curve) of the final BDT. A high value of the integral of the curve corresponds to large discrimination power of the BDT, so values close to 1 are desired. The achieved ROC-curve integral is 0.932.

Figure 6.12 shows the result of the evaluation of the set of 5 BDTs, to MC simulation and data. The agreement between data and MC simulation is excellent, showing no pathological behaviour introduced by the BDT.



(a)



(b)

FIGURE 6.10: Correlation coefficients for the variables used in the BDT training for both signal **(a)** and background **(b)** processes.

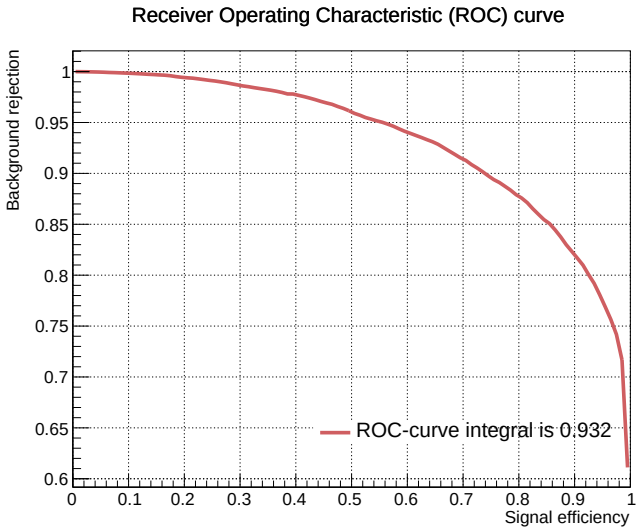


FIGURE 6.11: ROC-curve, i.e. signal efficiency versus background rejection.

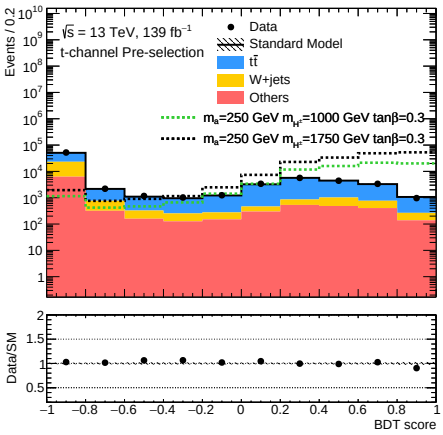


FIGURE 6.12: The BDT score distribution at pre-selection-level (region used for training, it is evaluated the set of 5 BDTs together) for the *t*-channel analysis. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. The SM backgrounds and the signals correspond to the simulation predictions normalised to the theoretical predictions.

6.4 Background estimation

The dominant background processes are estimated using a data-driven technique based on control regions (CRs). These CRs are defined to be orthogonal to the signal region (SR) and they are added in a simultaneous fit with the SR in order to constrain the backgrounds normalisation. The CRs for two of the largest backgrounds, $\mathrm{t\bar{t}}$ and W+jets are defined in this analysis. Definition of the CRs is designed to keep signal contamination in these regions small as possible.

The normalisation factors[3] for the background processes are calculated from a simultaneous likelihood fit (explained in Section 6.7). The normalisation factors derived in the fit are validated in the so-called validation region (VRs), containing negligible signal contamination as well. These regions, designed to be orthogonal to both CRs and SR, are used to cross-check the extrapolation of the normalisation factors (calculated from the CRs).

The dominant background for *t*-channel signal comes from $\mathrm{t\bar{t}}$ production as shown in Section 6.1. The $\mathrm{t\bar{t}}$ production constitutes the most important background processes to single top-quark events. Di-leptonic $\mathrm{t\bar{t}}$ events may mimic the single top-quark *t*-channel production final-state signature of two jets, one high-$\mathrm{p_T}$ lepton and large amount of missing transverse momentum, if one of the leptons is not identified. Control and validation regions for the $\mathrm{t\bar{t}}$ process are defined, referred to as *CR(*$\mathrm{t\bar{t}}$*)* and *VR(*$\mathrm{t\bar{t}}$*)*, respectively. These regions are built using the pre-selection definition, but with alternative requirements on the $\Delta\phi(\ell, \mathrm{b\text{-jet}\ 1})$, $\mathscr{N}_{\mathrm{b-jets}}$ and the BDT score. These requirements are meant to give orthogonality with SR and allow to define a pure enough region enhanced in $\mathrm{t\bar{t}}$ processes. The orthogonality with the SR ensured in the *CR(*$\mathrm{t\bar{t}}$*)* by requiring events to have $|\Delta\phi(\ell, \mathrm{b\text{-jet}\ 1})| > 1.5$ rad and in the *VR(*$\mathrm{t\bar{t}}$*)* with the selection of events with a BDT score less than 0.5. The *VR(*$\mathrm{t\bar{t}}$*)* such that the goodness of the normalisation factors can be checked closed to the SR.

A minor but not negligible contribution of background comes from W+jets events. The contribution of this background comes from misidentification of a quark jet as an electron or muon. This background is further suppressed by the use of b-jet tagging,

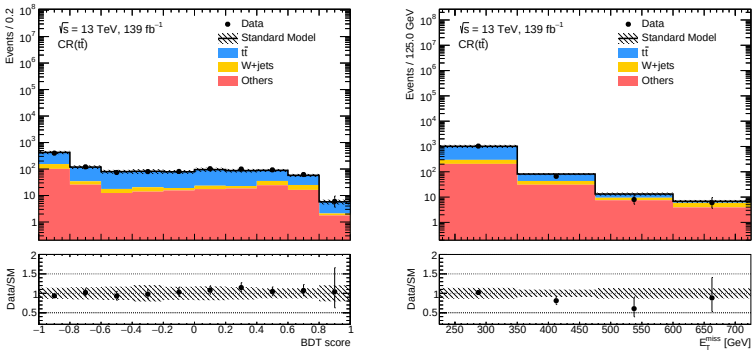
[3]The normalisation factor for the $\mathrm{t\bar{t}}$ and W+jets backgrounds are sets of free floating parameters in the fit and they are adjusted to match the observed event yield in the associated CR.

but a misidentification can be introduced (due to the fake rate of the b-tagging), and contributes to the SR. Dedicated regions are defined to estimate this background, referred to as *CR*(W+jets) and *VR*(W+jets). These regions are orthogonal but kinematically close to the SR, chosen to be enriched in W+jets events. Events are selected to satisfy the pre-selection, but changing the requirements on $m_T(\ell, E_T^{miss})$, $\mathcal{N}_{jets}$, $\mathcal{N}_{b-jets}$ and $\Delta\phi(\ell, \text{b-jet 1})$. For *CR*(W+jets) and *VR*(W+jets) the variable which it to be mutually exclusive with respect to the SR is the $m_T(\ell, E_T^{miss})$. The events contained in *CR*(W+jets) and *VR*(W+jets) are required to have a value of $m_T(\ell, E_T^{miss})$ smaller than 100 GeV, in opposite way of SR where the events require to have a $m_T(\ell, E_T^{miss})$ > 100 GeV. A summary of the background estimate strategy is presented in Table 6.4.

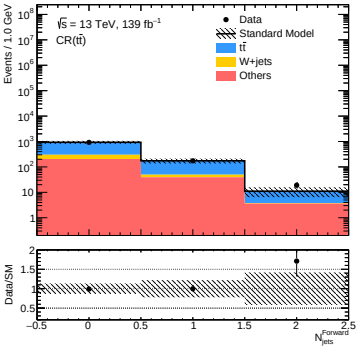
	t-channel
CR W+jets	$\lvert\Delta\phi(\ell, \text{b-jet 1})\rvert > 1.5$ rad
	1 b-jet, 1-2 jets
	$60\text{ GeV} < m_T(\ell, E_T^{miss}) < 100\text{ GeV}$
CR $t\bar{t}$	$\lvert\Delta\phi(\ell, \text{b-jet 1})\rvert > 1.5$ rad
	2 b-jets
	$m_T(\ell, E_T^{miss}) > 100\text{ GeV}$
VR W+jets	$\lvert\Delta\phi(\ell, \text{b-jet 1})\rvert < 1.5$ rad
	$60\text{ GeV} < m_T(\ell, E_T^{miss}) < 100\text{ GeV}$
	1 b-jet
VR $t\bar{t}$	$\lvert\Delta\phi(\ell, \text{b-jet 1})\rvert < 1.5$ rad
	$m_T(\ell, E_T^{miss}) > 100\text{ GeV}$
	BDT score < 0.5

TABLE 6.4: Schematic summary of the analysis strategy for backgrounds estimate. Only the requirements that differ from the SR (or from the CR in the case of VRs) are indicated. The requirements are schematically represented in terms of relevant physics parameter.

Figures 6.13-6.14 show various distributions for the CRs, while Figures 6.15-6.16 contain the relevant distributions for the VRs. The *CR*($t\bar{t}$) and *VR*($t\bar{t}$) have a purity of $t\bar{t}$ about 68.8% and 84.4%, respectively. While *CR*(W+jets) and *VR*(W+jets) have about 62.2% and 44.5% of W+jets purity, respectively.

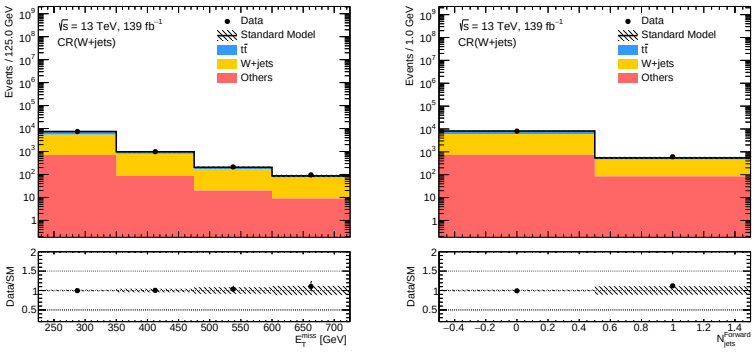


(a) **(b)**



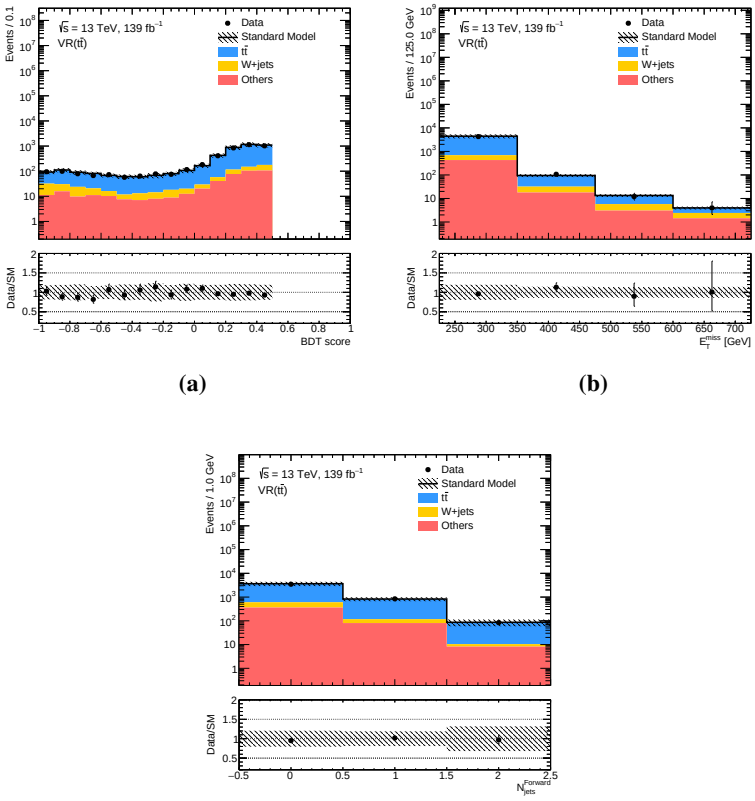
(c)

FIGURE 6.13: Distributions in the $t\bar{t}$ CR for the *t*-channel analysis. The variables presented in *CR*($t\bar{t}$) are: the BDT score **(a)**, E_T^{miss} **(b)** and the number of forward jets ($\mathcal{N}_{jets}^{Forward}$) **(c)**. Uncertainty band includes statistical, experimental and theoretical systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental and theoretical systematics used are detailed in Section 6.6. The normalisation factors obtained from the background-only fit (Table 6.8) are applied to the $t\bar{t}$ and W+jets MC processes.



(a) **(b)**

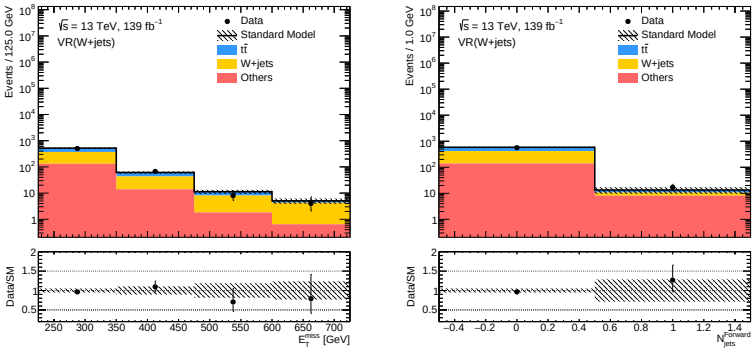
FIGURE 6.14: Distributions in the W+jets CR for the *t*-channel analysis. The variables presented in *CR*(W+jets) are: the E_T^{miss} **(a)** and the number of forward jets ($\mathcal{N}_{jets}^{Forward}$) **(b)**. Uncertainty band includes statistical, experimental and theoretical systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental and theoretical systematics used are detailed in Section 6.6. The normalisation factors obtained from the background-only fit (Table 6.8) are applied to the $t\bar{t}$ and W+jets MC processes.



(a) **(b)**

(c)

FIGURE 6.15: Distributions in the $\mathrm{t\bar{t}}$ VR for the *t*-channel analysis. The variables presented in *VR*($\mathrm{t\bar{t}}$) are: the BDT score **(a)**, $\mathrm{E_T^{miss}}$ **(b)** and the number of forward jets ($\mathcal{N}_{\mathrm{jets}}^{\mathrm{Forward}}$) **(c)**. Uncertainty band includes statistical, experimental and theoretical systematics. Background type "Others" (single top-quark, Diboson, $\mathrm{t\bar{t}V}$, Z+jets, $\mathrm{t\bar{t}h}$ and tWZ) is the combination of all the backgrounds not including $\mathrm{t\bar{t}}$ and W+jets. Experimental and theoretical systematics used are detailed in Section 6.6. The normalisation factors obtained from the background-only fit (Table 6.8) are applied to the $\mathrm{t\bar{t}}$ and W+jets MC processes.



(a) **(b)**

FIGURE 6.16: Distributions in the $t\bar{t}$ and W+jets VRs for the *t*-channel analysis. The variables presented in *VR*(W+jets) are: the E_T^{miss} **(a)** and the number of forward jets ($\mathcal{N}_{jets}^{Forward}$) **(b)**. Uncertainty band includes statistical, experimental and theoretical systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental and theoretical systematics used are detailed in Section 6.6. The normalisation factors obtained from the background-only fit (Table 6.8) are applied to the $t\bar{t}$ and W+jets MC processes.

6.5 Signal region

The *t*-channel final state presents one b-jet from the decay of the top quark, an isolated lepton from the decay of the W boson and large amount of E_T^{miss} due to the neutrino from the W decay and from the undetected DM candidate particles.

In addition to the pre-selection requirements, further discrimination between the *t*-channel signal and background events is achieved by applying additional criteria. For instance, the signal is enriched in regions with lower $\Delta\phi(\ell, \text{b-jet 1})$ since the lepton and the (leading) b-jet are originated from the decay of a top quark, which is also produced quite boosted. Therefore, a selection imposing the rejection of events with $|\Delta\phi(\ell, \text{b-jet 1})| < 1.2$ rad is used. For higher values of E_T^{miss} and $m_T(\ell, E_T^{miss})$ the background decreases significantly, as seen in Figure 6.3(**a**)-(**b**). The lower threshold values on E_T^{miss} and $m_T(\ell, E_T^{miss})$ are increased to 225 GeV and 100 GeV, respectively. For *t*-channel signal production, the extra jet produced is mainly forward ($|\eta^{jet}| > 2.5$). The requirement of having such a jet allows to enhance the sensitivity to the *t*-channel signal. Backgrounds have a larger contribution of events in the $\mathcal{N}_{jets}^{Forward} = 0$. The requirement of $\mathcal{N}_{jets}^{Forward} \geq 1$ rejects more background than signal. Thus, at least a forward jet is required in order to increase signal significance as shown in Figure 6.17(**c**).

The final SR includes the BDT requirement. The BDT score larger than 0.6 is selected to maximise signal acceptance. The acceptance times detector efficiency after applying all selection criteria for the *t*-channel production of the signal models is between 0.37% (0.36%) and 0.73% (0.67%) for m_a = 250 GeV, $\tan\beta$ = 0.3 (0.5) and $m_{H^\pm} \in$ [500,1750] GeV. The final discriminant is implemented in four bins on the BDT score: [0.6, 0.75], [0.75, 0.85], [0.85, 0.9] and [0.9-1.0], referred to as bins [0-3]. Implementation of these bins in the statistical interpretation has been found to improve exclusion reach (see Figure 6.18(**b**)).

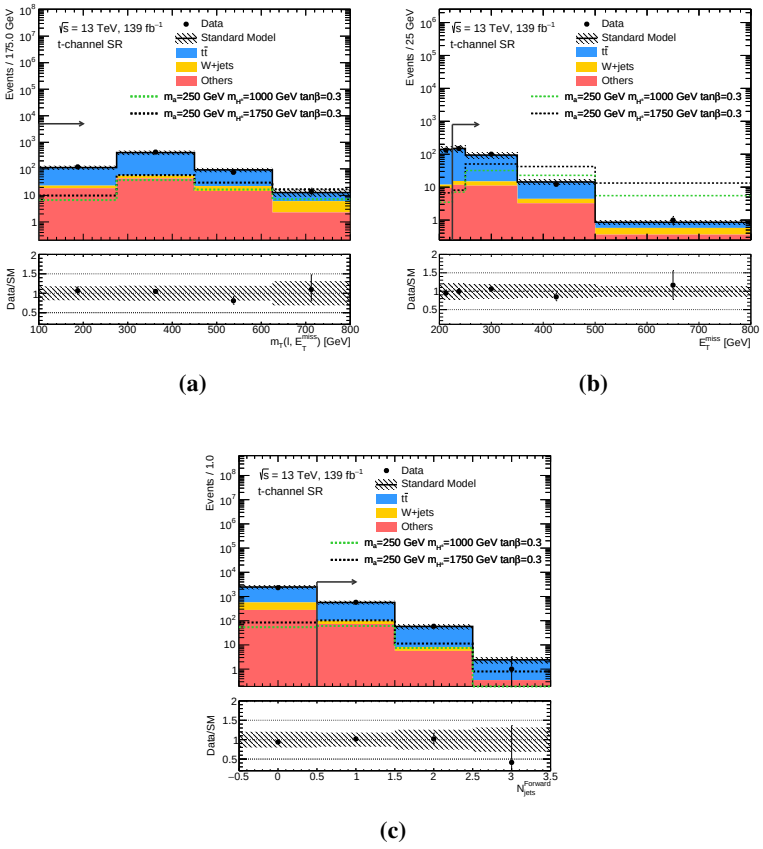
As it has been pointed earlier, the dominant background in the *t*-channel SR originates from $t\bar{t}$ production (contributes with a 83.1% of the total background). Single top quark processes comprise the second largest background in the SR (8.8%)[4].

[4] As it is difficult to obtain a pure single top quark CR, some checks were done in VRs, with medium level purity in single top-quark production, showing good agreement between data and MC.

The third contributor of backgrounds comes from W+jets events (contributes with a 4.2%). Figures 6.17-6.18 present the distributions of important kinematic variables in the SR. A summary of the general analysis strategy is presented in Table 6.5. The Figure 6.19 shows an schematic diagram of the main regions of the analysis presented.

	t-channel
Signal topology	1 ℓ, 1-4 jets, 1-2 b-jets, at least one forward jet $E_T^{miss} > 225$ GeV
Reducible bkg suppression	$m_T(\ell, E_T^{miss}) > 100$ GeV
Main kinematic selections	$\lvert\Delta\phi(\ell, \text{b-jet 1})\rvert < 1.2$ rad, BDT score
Final discriminant	BDT-binned fit
Backgrounds estimated in CRs	W+jets, $t\bar{t}$

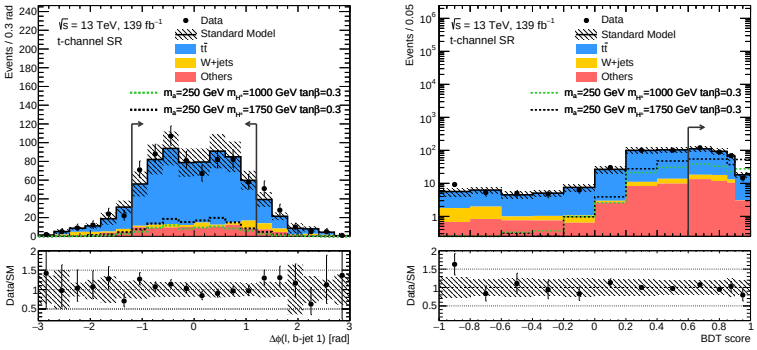
TABLE 6.5: Schematic summary of the analysis strategy. Only the requirements that differ from the pre-selection are indicated. The requirements are schematically represented in terms of relevant physics parameter. The exact value of the requirements have been optimised.



(a) **(b)**

(c)

FIGURE 6.17: Distributions in SR of *t*-channel analysis, there are shown the $\mathrm{m_T}(\ell, \mathrm{E_T^{miss}})$ **(a)**, $\mathrm{E_T^{miss}}$ **(b)** and the number of forward jets ($\mathscr{N}_{\mathrm{jets}}^{\mathrm{Forward}}$) **(c)**. Uncertainty band includes statistical and experimental systematics. Background type "Others" (single top-quark, Diboson, $\mathrm{t\bar{t}V}$, Z+jets, $\mathrm{t\bar{t}h}$ and tWZ) is the combination of all the backgrounds not including $\mathrm{t\bar{t}}$ and W+jets. Experimental systematics used are detailed in Section 6.6. The normalisation factors obtained from the background-only fit (Table 6.8) are applied to the $\mathrm{t\bar{t}}$ and W+jets MC processes.



(a) **(b)**

FIGURE 6.18: Distributions in SR of *t*-channel analysis, there are shown $\Delta\phi(\ell, \text{b-jet 1})$ **(a)** and the BDT score **(b)**. Uncertainty band includes statistical and experimental systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental systematics used are detailed in Section 6.6. The normalisation factors obtained from the background-only fit (Table 6.8) are applied to the $t\bar{t}$ and W+jets MC processes.

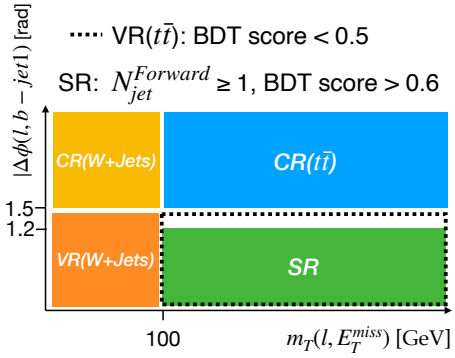


FIGURE 6.19: Sketch depicting the control, signal and validation regions in the $\Delta\phi(\ell, \text{b-jet 1})$ and $m_T(\ell, E_T^{miss})$ plane. At the top, there are shown the extra requirements for *VR*($t\bar{t}$) and SR definition.

6.6 Sources of systematic uncertainties

All signal and background processes have systematic uncertainties some of which are process-dependent (e.g. generator specific) and others generic (e.g. jet energy scale). These systematic uncertainties have either an impact on the rate of the individual contributions or on the shape of their associated kinematic and angular distributions. The precise estimation of the impact of these uncertainties on the normalised event yields in all SRs and CRs as well as on the shape of the discriminating variables is crucial for achieve a good estimation of the physical quantities. Table 6.6 shows an enumeration of the sources of systematic uncertainty evaluated in this analysis.

The systematic uncertainties are included as nuisance parameters (NPs) in the fit, as discussed in Section 6.7.

6.6.1 Statistical uncertainty

Statistical fluctuations in the MC simulated event samples contribute to the overall systematic uncertainty. These uncertainties arise from the statistics of background MC simulation on one hand, and from the statistics of signal MC simulation on the other.

6.6.2 Experimental uncertainties

Event-level uncertainties

The event (process-independent) uncertainties affecting the overall normalisation of the processes relate to both the luminosity and pile-up measurements. The relative uncertainty on the full Run 2 dataset luminosity is 1.7% as shown in Section 4.1. This uncertainty is derived from the calibration of the luminosity scale using x-y beam-separation scans, following a methodology similar to that detailed in Ref. [212], and using the LUCID-2 detector for the baseline luminosity measurements [213]. The luminosity uncertainty is applied to each MC simulated process in order to scale them to match the expected number of events at the given luminosity for each year.

Statistical uncertainties			
MC simulation and data			
Experimental uncertainties			
Event-level			
Luminosity measurement			
Pile-up modelling			
Object reconstruction			
Jets:	*Flavour tagging:*	*Leptons:*	E_T^{miss}*:*
Jet Energy Scale (JES)	b-tagging	Identification Efficiency	Soft-term Resolution
Jet Energy Resolution (JER)	Mis-tagging Efficiency	Reconstruction Efficiency	Soft-term Scale
Jet Vertex Tagger (JVT)		Isolation Efficiency	
		Trigger Efficiency	
Theoretical uncertainties			
Background			
Source of Uncertainty:		*Background Processes:*	
Hard-Scatter Generation (ME)		$t\bar{t}$ and single top-quark	
Fragmentation (PS)		$t\bar{t}$ and single top-quark	
Additional Radiation (ISR and FSR variation)		$t\bar{t}$, single top-quark and $t\bar{t}V$	
Interference ($t\bar{t}$-tW)		$t\bar{t}$ and tW	
Factorisation and renormalisation scales (μ_R and μ_F) variations		V+jets and Diboson	
CKKW/QSF and α_s variations		V+jets	
Normalisation factors		$t\bar{t}$ and W+jets	
Signal			
Fragmentation (PS)			
Factorisation and renormalisation scales (μ_R and μ_F) variations			
Cross-section uncertainty			

TABLE 6.6: Summary of the sources of systematic uncertainties. The sources are divided depending of their nature: statistical, experimental and theoretical.

All MC simulated event samples are reweighted to match the observed distribution of the average number of interactions per bunch-crossing in data [315] as discussed in Section 4.2.5. To account for the difference in pile-up distributions between data and MC simulations, an uncertainty related to the MC scale factors is applied. Up and down variations of these uncertainties related to the pile-up scale factors are applied.

Jet uncertainties

The systematic uncertainties on reconstructed jet objects are related to the jet energy scale (JES), jet energy resolution (JER) and jet-vertex-tagger (JVT). The effects of the JES and JER uncertainties are among, the dominant sources of uncertainty on this analysis. The specific implementation of these uncertainties for this analysis are:

- Jet energy scale: The JES uncertainty is estimated based on MC studies and in-situ measurements by comparing the response of the calorimeter to the jet energy at generator level and its uncertainty depending on $\mathrm{p_T}$ and η, including the flavour composition of jets and pile-up effects [316]. To determine the jet energy scale (JES) and its associated uncertainty, information from test-beam data, LHC collision data and simulation are used. Data taken at $\sqrt{s}$ = 13 TeV is used to calibrate the residual uncertainty on the JES [302]. Events with a vector boson and additional jets are used to calibrate jets in the central region. Di-jet events are exploited to calibrate forward jets against the jets in the central region of the detector. This results in a set of 7 NPs (from nearly 100 NPs), each with an up/down variation, which can have different jet $\mathrm{p_T}$ and η dependencies. These are: the three eta intercalibration non-closure uncertainties as single stand-alone NPs, three additional NPs which are combinations of all the remaining parameters and a flavour response term (dominated by the gluon response).

- Jet energy resolution: The JER uncertainty is derived as one-side variation by comparing data to MC simulation via the transverse momentum balance between a jet and a reference object such as a photon, Z boson, or multi-jet system in data [317]. The JER can be parametrised with a stochastic term, a noise term and a constant term. The uncertainty is extracted from di-jet events by measuring the width of the di-jet asymmetry distribution across $\mathrm{p_T}$ and η. Measurements using zero-bias[5] data with random cones are used to constrain

[5]Zero bias data are composed by data collected with random triggers in coincidence with colliding bunches.

the noise term. The performed fit gives about 117 NPs. An eigenvector decomposition is used which gives a smaller set with 7 NPs. First, nominal smearing is applied on jets. If the JER in MC simulation is found to be smaller than in data, the MC simulation is smeared to match the average resolution in data. If the JER is lower in data, nothing is done. The uncertainties of the JER are then propagated by smearing the jets in the MC simulation with a Gaussian with $\sigma^2_{\text{smear}} = (\sigma_{\text{nominal}} - \sigma_{\text{NP}})^2 - \sigma^2_{\text{nominal}}$ where σ_{nominal} is the nominal JER after the previous smearing and σ_{NP} is the 1σ variation of the JER uncertainty component. Finally, if the JER in data is lower than in the MC simulation, the difference is taken as an additional uncertainty. Therefore, in total 8 NPs are used for the JER uncertainty: one data/MC simulation comparison term, and seven modelling/theory components.

- Jet vertex tagger efficiency: Uncertainties on the JVT are provided as a two sided variation covering the differences in JVT efficiency measured in data and MC simulation, based on scale factors derived in Z+jets events [318]. These scale factors, used to account for differences between the JVT efficiency in simulation and data, are derived using $Z(\rightarrow \mu^+\mu^-)$ + jets events [319]. A control region is used to estimate the pile-up contamination in the signal region. The contribution from hard-scatter jets in the control region is subtracted and a conservative uncertainty of 30% is used to cover a potential mis-modelling. Another systematic uncertainty is derived by comparing different MC generators for the Z+jets simulation [319]. The statistical uncertainty of the derived scale factors is also taken into account.

Flavour tagging uncertainties

There are uncertainties in the jet flavor tagging efficiencies (b-jets), as well as in the measured mis-tagging efficiencies associated with c- and light jets. They are a mixture of statistical, experimental, and modelling uncertainties incurred during the flavour tagging calibration procedures [306].

The b-tagging algorithm needs to be calibrated in order to have a match between the performance in simulation and in data. Therefore, corrective scale factors are

derived from data [320]. The b-tag and c-tag efficiencies and the mis-tag rate for light-flavour jets are measured and scale factors are calculated as the ratio of the efficiencies (or mis-tag rates) in data and simulation. In general, the scale factors depend on jet p_T and η. Uncertainties are propagated into the analysis via 1 NP for b-jets, 1 NP for c-jets, 1 NP for light-flavour jets, and 2 NPs for extrapolation of the b-tagging weights and the c-jet mistag rate to high-p_T regimes.

Lepton uncertainties

The uncertainties on the measurement of charged leptons are related to the electron and muon reconstruction, identification, trigger, and isolation efficiencies in a similar procedure to the flavour tagging. Systematic variations incurred in the associated scale-factor measurements are applied in this analysis. Additional uncertainties related to the lepton kinematics due to the resolution and scale of the electron (muon) energy (momentum) measurement are needed. The muon momentum measurement uncertainties are derived for both the ID and MS measurement of the combined muons used in the analyses. The specific implementation of this analysis are:

- Charged lepton reconstruction, identification, isolation and trigger: For charged leptons the reconstruction, identification, isolation and trigger performances differ between data and MC. To correct for these differences, scale factors are applied. They are estimated with the tag-and-probe method with electrons and muons from Z boson, W boson and J/ψ decays using methods similar to those from Refs. [292, 301]. The uncertainties are evaluated by varying up and down by 1σ the predicted event yields and re-applying the event selection to the signal and backgrounds.

- Charged lepton momentum scale and resolution: The precision of the charged lepton momentum scale and resolution may be different between the simulated events and the observed data. The simulation is checked with reconstructed distributions of $Z \to \ell^+\ell^-$ and $J/\psi \to \ell^+\ell^-$ masses using methods similar to the ones used in Ref. [293, 301]. In the case of electrons, also events with $W \to e\nu$ are used. For muons, corrections to momentum scale and resolution

are applied only to the simulation. Uncertainties on momentum scale and resolution of muons originating from the ID and the MS are considered and varied separately. The impact on the measurement of the electron (muon) energy (momentum) and resolution uncertainties are evaluated by scaling or smearing up and down the charged lepton transverse energy/momentum by 1σ and re-applying the object and event selections to the simulated event samples.

Missing transverse momentum

Uncertainties of the soft-track component are derived from the level of agreement between data and MC simulation of the $\mathrm{p_T}$ balance between the hard and soft $\mathrm{E_T^{miss}}$ components (i.e. $\mathrm{p_T^{hard}} = -\mathrm{p_T^{soft}}$, the hard component refers to all objects do not identified as soft term, see Section 5.6.) Three different uncertainties are considered: an offset along the $\mathrm{p_T}$ (hard) axis (1 NP: up and down variations), as well as the smearing resolution along and perpendicular to the $\mathrm{p_T}$ (hard) axis (2 NP: one-sided variations). These effects are estimated by varying the scales and resolutions up and down by 1σ before re-doing the selection of the simulated event samples.

6.6.3 Theoretical uncertainties

Theoretical uncertainties for the backgrounds include the effects of different MC generators on the hard scattering process and the use of different software for showering to cover effects of the modelling of fragmentation and hadronisation. The effect of varying scale choices such as the renomalisation and factorisation scales, initial- and final-state radiation parameters, along with choices of scales used for PS matching are included in the fit procedure as additional NPs in the fit.

MC generator, PS and interference modelling

The uncertainties due to the choice of the MC event generator, PS and hadronisation models and scales are evaluated for the *t*-channel signal and for the top-quark backgrounds by considering alternative generators or varied parameters of the baseline event generators. These uncertainties are estimated by varying one by one the different processes (in an uncorrelated way).

- ME modelling ($t\bar{t}$ and single top-quark backgrounds): To assess the uncertainty due to the choice of the matching scheme in the single top ME generation, the nominal MC generator POWHEG-BOX is compared to aMC@NLO. This uncertainty is not considered for the $t\bar{t}$ process. The PS and hadronisation are simulated with PYTHIA 8 in both cases.

- PS and hadronisation model ($t\bar{t}$ and single top-quark backgrounds): To describe the impact of the uncertainty coming from the chosen PS and hadronisation models, the nominal simulated sample which uses PYTHIA 8 is compared to another sample using HERWIG 7. POWHEG-BOX is used as ME generator for both simulation samples.

- ISR/FSR ($t\bar{t}$ single top-quark backgrounds and $t\bar{t}V$): To estimate the uncertainty originating from ISR modelling, alternative weights are used in the ME and in the PS within the baseline POWHEG-BOX+ PYTHIA 8 sample. To simulate higher parton radiation, the factorisation and renormalisation scales are varied by a factor of 0.5 in the ME while using the VAR3C up variation from the A14 tune [263]. For lower parton radiation, μ_R and μ_F are varied by a factor of 2.0 while using the VAR3C down variation in the PS. The impact of FSR is evaluated using PS weights which vary the renormalisation scale for QCD emission in the FSR by a factor of 0.5 and 2.0, respectively.

- Interference ($t\bar{t}$-tW): To evaluate the uncertainty due to the interference between the tW associated and $t\bar{t}$ productions at NLO, the baseline setup (using the DR scheme) is compared to a simulated event sample using the diagram subtraction (DS) scheme to handle the interference [250]. In the DR approach, all NLO diagrams that overlap with the doubly resonant $t\bar{t}$ contributions are removed from the calculation of the tW amplitude. This approach accounts for the interference term, but it is not gauge invariant. In the DS approach, a subtraction term is built into the amplitude to cancel out the $t\bar{t}$ component close to the top-quark resonance while respecting gauge invariance.

V+jets variations

SHERPA (v2.2) samples have associated ME matching, renormalisation and factorisation scales uncertainties, which are implemented in the generator and affects V+jets backgrounds. The comparison is made from the nominal SHERPA configuration there are four parameters that can be varied to investigate uncertainties on the modelling of the W/Z+jets process:

- Matrix element matching (CKKW): This uncertainty varies the scale taken for the calculation for the overlap between jets from the ME and the PS. The nominal value for this parameter is 20 GeV. The up variation increases this to 30 GeV, while the down variation decreases the nominal value to 15 GeV.
- Renormalisation scale: Varies the scale for the running strong coupling constant for the underlying hard process, μ_{R} is varied from 2 and 0.5 with respect to the nominal.
- Factorisation scale: Varies the scale used for the PDFs, μ_{F} is varied from 2 and 0.5 with respect to the nominal.

The PDF uncertainties are evaluated on the V+jets backgrounds using the central values of the variation PDF sets that come as internal weights in these samples. These are found to be < 1% and are sub-dominant when compared to many of the modelling uncertainties that are evaluated, thus the PDF uncertainties of the backgrounds are not included.

Normalisation factors

The event yields associated with the simulated top-quark pair events, single top-quark, W/Z+jets and diboson processes are estimated using the selection acceptances and the theoretically predicted cross-sections. The main backgrounds ($\mathrm{t\bar{t}}$ and W+jets) normalisation factors are extracted from the data-driven fits as described in Section 6.7.1. The theoretical relative uncertainties of the cross-sections are considered to be negligibly compared to the systematic uncertainty for the main backgrounds. Since actually they are absorbed by the normalisation factors. For minor backgrounds like $\mathrm{t\bar{t}V}$, diboson and Z+jets, the cross-section uncertainty is negligible compared with the ISR, FSR, μ_{F}, μ_{R} and CKKW/QSF variations associated.

Signal theoretical uncertainties

Uncertainties for the μ_R and μ_F are evaluated using the 3-point variation scheme wherein μ_R and μ_F are scaled simultaneously by a factor 2 and then a factor 1/2 to give the up and down variations respectively. For the PS and hadronisation uncertainties, only the eigentune variations of the A14 tune are considered[6]. For the scale variation samples, effects on the cross-section are de-coupled from kinematic effects by scaling to the nominal cross-section. These variations are added in quadrature to the total theory uncertainty.

6.7 Statistical interpretation

Analyses presented in this thesis are known as "counting experiments", meaning that only the number of events, from data and the predicted background and signal, are used as input. These numbers are calculated from all relevant regions in the analysis: the CRs and the SR, described in previous sections. To obtain the expected number of event counting in the main regions (*r*):

$$N_r^{exp}(\mu_{sig}, \mu_{bkg}, \theta) = \mu_{sig} N_{r,sig}^{exp}(\theta) + \sum_{b \in bkg} \mu_{bkg} N_{r,bkg}^{exp}(\theta),$$

where $N_{r,sig}^{exp}$ and $N_{r,bkg}^{exp}$ represent the expected signal and background yields across SR (runs over all bins implemented in SR) and CRs, defined for this analysis. The μ_{sig} is a scaling parameter applied to the signal to test the sensitivity of the search, the so-called signal-strength. The null hypothesis (background-only hypothesis) corresponds to μ_{sig} = 0 while the test hypothesis (signal-plus-background hypothesis) corresponds to μ_{sig} = 1. And the μ_{bkg} are the 2 normalisation factors associated with the main background processes, i.e. $t\bar{t}$ and $W + jets$. The θ are a set of fit NPs, which represent the different systematic uncertainties discussed in Section 6.7 [322].

A likelihood is constructed as the product of Poisson distributions used to perform the hypothesis test. This likelihood is calculated using the HistFitter framework

[6]This approach is adopted by recent A14 tune, which provide "eigentune" method. In eigentune construction, there are an infinity of ways in which to make model variations "cover" all data [321].

which is based on the HistFactory implementation [323]. The definition of the likelihood,

$$L_0(\mu_{sig}, \mu_{bkg}, \theta) = \prod_{r \, in \, regions} \frac{N_r(\mu_{sig}, \mu_{bkg}, \theta)^{N_r^{obs}}}{N_r^{obs}!} exp[N_r^{exp}(\mu_{sig}, \mu_{bkg}, \theta)] \quad (6.2)$$

where N_r^{obs} is the observed data yield (collected by the ATLAS detector over the whole Run 2) in a region *r*.

On the other hand, the systematic uncertainties related with the measurements of the N_r^{exp} when the NP have a value of $\theta = 0$ represent[7] to the central (nominal) value of the set of parameters associated with the uncertainty. Therefore, the values $\theta_i = \pm 1$ give shifts in the parameter values by their $\pm 1\sigma$ variation. The systematic uncertainties described in Section 6.6 are commonly extracted with only the $\pm 1\sigma$ variation with respect their nominal value. Given this, Equation 6.2 is transformed to:

$$L(\mu_{sig}, \mu_{bkg}, \theta) = L_0(\mu_{sig}, \mu_{bkg}, \theta) \prod_{i=1}^{n} \frac{1}{2\pi} exp(-\frac{\theta_i^2}{2}). \quad (6.3)$$

The best prediction for the parameters μ_{sig}, μ_{bkg}, θ are calculated with the fit to the observed data, N^{obs}, via the maximisation of the likelihood function given by Equation 6.3. The values of all parameters prior to (after) the minimisation procedure are referred to as the pre-fit (post-fit) values. The pre-fit values for the μ_{bkg} parameters are set to 1 (not applying normalisation factor) and the θ_i are set to 0, corresponding to the central values of the Gaussian-shifted parameters associated with the systematic uncertainties.

A useful statistic test to set upper limits in the signal strength (when a signal hypothesis is considered) is the profile likelihood ratio,

$$q_{\mu_{sig}} = -2 \ln \frac{L(\mu_{sig}, \hat{\hat{\mu}}_{bkg}, \hat{\hat{\theta}})}{L(\hat{\mu}_{sig}, \hat{\mu}_{bkg}, \hat{\theta})} \quad (6.4)$$

[7] The $\theta = 0$ means all of the θ_i equal to zero.

where $\hat{\hat{\theta}}$ denotes the value of θ that maximises the likelihood for the specified μ (conditional maximum-likelihood estimator of θ) and the denominator is the maximised (unconditional) likelihood function ($\hat{\mu}$ and $\hat{\theta}$ are their maximum-likelihood estimators).

Since a statistic test is specified, and its expected distribution under a given hypothesis is obtained, a p-value can be defined in order to compute the probability that the observed data originates from the considered hypothesis (value of μ_{sig}).

$$p_{(\mu=\mu_{sig})} = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|\mu)dq_\mu;$$

where $f(q_\mu|\mu)$ is the probability density function of q_μ assuming the hypothesis μ. Usually, to estimate the density function of a statistic test, numerical methods through MC simulations are needed.

In a general way, it becomes reasonable to say that the supposed hypothesis is inconsistent with reality, therefore the hypothesis described by the particular value of μ_{sig} is meant to be excluded if the p-value is below to certain threshold. Under this thinking, it is assumed if the p-value is found below $\alpha = 0.05$, then the value is declared to be excluded at a confidence level (CL) of 95%.

To claim a discovery, the null hypothesis ($\mu_{sig} = 0$) has to be rejected. The thresholds at which new physics can be said to have been observed and discovered are much more stringent than that used for the exclusion of a specified hypothesis. Incompatibilities with the null-hypothesis at the level of[8] $p_0 = 1.3 \times 10^{-3}$ (3 σ) and $p_0 = 2.9 \times 10^{-7}$ (5 σ) are required to affirm the state that observation and discovery, respectively, of new phenomena has happened.

At LHC experiments the statement obtained by a signal hypothesis is a key value, since the specific signal model can be fundamentally considered no longer important to be searched for. The mere p-value, extracted from the observed data is subject to statistical fluctuations and it can conduce to unphysical exclusions when a downward fluctuation in the observed number of events arises. This fact can lead to a fake exclusion of a broad region of new physics that would perhaps no longer be looked into

[8]The p_0 is the p-value value when $\mu_{sig} = 0$, means when the hypothesis is compatible with the background-only hypothesis.

future analyses or experiments. To avoid this giving false results, the LHC experiments calculate the CL_s values instead of the p-values [163, 324]. The CL_s can be obtained as:

$$CL_s = \frac{p_{\mu_{sig}}}{1 - p_0},$$

where the quantities $p_{\mu_{sig}}$ and p_0 quantify the consistency between the signal-plus-background and background-only hypotheses, respectively. Downward fluctuations in data, as those described above, will lead to larger values of p_0 (meant to have good description of SM processes and containing considerable amount of statistics). Thus leading to larger values of CL_s that avoid fake exclusion.

A given signal hypothesis with $\mu_{sig} = 1$ is considered excluded at 95% CL when $CL_s \leq 0.05$. The CL_s metric is also used to calculate the *upper limits*. A 95% CL *upper limit* on a given signal hypothesis specified is the largest value of μ_{sig} satisfying $CL_s \geq 0.05$. The interpretation corresponds to the extraction of the largest possible signal cross-section that is unable to be excluded (and therefore smaller values of μ_{sig}). Since this values is smaller than the signal cross-section, its existence is still consistent with the observed data and can not be excluded.

6.7.1 Background-only fit

To evaluate the impact of the CRs on the background estimation (extrapolation of normalisation values μ_{bkg}) in the SR, the so-called background-only fit is performed. The background-only fit is a profile-likelihood fit (Equation 6.3), using only CRs in the likelihood. The outcome of running a background-only fit to data in the CRs is shown in Table 6.7, which presents the MC predicted yields for the background processes before (pre-fit) and after (post-fit) the background-only fit, and for the observed data, among the CRs and VRs defined. The post-fit yields in the CRs are expected to agree with the observed data counts, since the latter are used as constraints in the fit model. The agreement observed between the post-fit MC yields and the observed data in the VRs shows that the extrapolation, at least in terms of the corrected normalisation at the MC simulation, is performing well.

Process	*CR*($t\bar{t}$)	*CR*(W+jets)	*VR*($t\bar{t}$)	*VR*(W+jets)
Observed Data	1117	8711	4398	582
Post-fit Total SM	1117 ± 30	8711 ± 90	4558 ± 720	599 ± 29
Post-fit $t\bar{t}$	760 ± 137	2235 ± 676	3819 ± 837	171 ± 50
Post-fit W+jets	107 ± 28	5647 ± 657	281 ± 49	280 ± 40
Post-fit Others	250 ± 120	830 ± 95	457 ± 112	148 ± 10
Pre-fi total SM	1108 ± 186	8245 ± 212	4532 ± 574	576 ± 28
Pre-fit $t\bar{t}$	760 ± 118	2232 ± 124	3817 ± 553	170 ± 11
Pre-fit W+jets	98 ± 18	5183 ± 99	256 ± 28	257 ± 18
Pre-fit Others	250 ± 120	830 ± 96	458 ± 113	148 ± 10

TABLE 6.7: Results of the background only fit for *t*-channel CRs and VRs (statistical, experimental and theoretical uncertainties are considered). Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ)

The extracted normalisation values for $t\bar{t}$ and W+jets backgrounds (free-floating scale factors) from a background-only fit setup are presented in Table 6.8. The free-floating scale factors obtained from $t\bar{t}$ and W+jets background are compatible with the unity, as expected.

$\mu_{t\bar{t}}$	1.00±0.27
$\mu_{\mathrm{W+jets}}$	1.10±0.13

TABLE 6.8: Normalisation correction factors for the $t\bar{t}$ ($\mu_{t\bar{t}}$) and W+jets ($\mu_{\mathrm{W+jets}}$) processes acquired from the background-only fit to the CRs. The uncertainties on the quoted numbers are due to the statistical, experimental and theoretical systematic uncertainties implemented in the fit.

6.8 Results

A profile likelihood fit is calculated, following the calculations described in the previous section. All the CRs and the binned[9] SR on the BDT score: [0.6, 0.75], [0.75, 0.85], [0.85, 0.9] and [0.9-1.0], are included. The event yields of the SR (including all uncertainties described in Table 6.6) and the observed data therein are shown in Table 6.9. There is no significant deviation between the observed data counts in the SR as compared to the post-fit prediction of the total SM background, within the statistical and systematic uncertainties. Figure 6.20 shows the post-fit MC prediction compared to the observed data of the selected binning in BDT score in the SR implemented in the fit studies.

The magnitude of the contributions arising from experimental, theoretical modelling, background normalisation and statistical uncertainties are summarised in Figure 6.21 as relative uncertainty on the total background yield for each SR in the analysis.

The dominant experimental systematics are the jet related uncertainties (JES and JER). It is expected in single top-quark analyses that the JES and JER calibration procedures play an important role. These are applied after jet reconstruction, are meant to ensure the correct measurement of the average jet energy across the whole detector (inter-calibration), and are designed to be independent of the pile-up conditions. The theoretical ones, the most relevant are the hard-scatter generation (from the ME) of $t\bar{t}$ and single top-quark productions, the fragmentation (from the PS) and the additional radiation (from ISR), these last two coming from $t\bar{t}$ generation. The theory based uncertainties, related to the simulation of top-quark pair (or single-top quark) signal events, describing the hard scattering environment, PS and ISR/FSR are expected to be dominant in this kind of analysis (because it is selected events containing such process). Since the models need to be tuned to reference data in order to get accurate predictions for phase-space regions opening up at the LHC (as discussed in Chapter 4), these uncertainties are difficult to reduce. Additionally, signal theoretical uncertainty has about a 15% of contribution among signals. The magnitude

[9] A binned fit refers to take several SRs defined by different intervals (which can be selected after a certain value) in one variable.

of the contributions arising from detector, theoretical and statistical uncertainties are summarised in Figure 6.21 as relative uncertainty on the total background yield.

Figure 6.22 shows the main correlation of the NPs entering in the fit for one signal, the sample m_a = 250 GeV, $m_{H^\pm}$ = 1750 GeV and $\tan\beta$ = 0.3. The high correlation between the $\mu_{t\bar{t}}$ with the JES and JER uncertainty, it is excepted, since they are the main experimental uncertainties contributors, the factor $\mu_{t\bar{t}}$ constrains their variations. It can be observed a low correlation between the ME-$t\bar{t}$ with the JES and JER uncertainty. Figure 6.23 shows the influence of the main systematic uncertainties (taking account the correlation presented in Figure 6.22) on the fitted value of the signal strength parameter μ_{sig} of the signal m_a = 250 GeV, $m_{H^\pm}$ = 1750 GeV and $\tan\beta$ = 0.3. The modelling related to the $t\bar{t}$ uncertainties (PS and ISR) are the most dominant, followed for the PS of the single top quark. To the experimental related uncertainties, as expected the JES and JER are the main contributors of their group.

Process	SR bin 0	SR bin 1	SR bin 2	SR bin 3
Observed Data	338	189	64	44
Post-fit Total SM	335 ± 72	187 ± 39	67 ± 18	36 ± 7
Post-fit $t\bar{t}$	280 ± 73	151 ± 40	53 ± 16	30 ± 7
Post-fit W+jets	14 ± 5	12 ± 9	$2.7^{+7}_{-2.7}$	0.00 ± 0.00
Post-fit Others	40 ± 13	23 ± 8	10 ± 3	6 ± 3
Total SM	333 ± 47	186 ± 31	67 ± 11	36 ± 5
Pre-fit $t\bar{t}$	280 ± 43	151 ± 23	54 ± 9	30 ± 4
Pre-fit W+jets	13 ± 4	11 ± 8	$2.5^{+7}_{-2.5}$	0.00 ± 0.00
Pre-fit Others	40 ± 13	23 ± 8	10 ± 3	6 ± 3

TABLE 6.9: Results of the background only fit for *t*-channel SRs. This fit includes the statistical, experimental and theoretical uncertainties.

Since there is no significant deviation between the prediction and observed data, hypothesis tests are calculated in order to assess which regions of the 2HDM+a parameter space can be excluded (signal hypothesis is rejected). The procedure explained in Section 6.7 is implemented. CL_s is evaluated for each point in $\tan\beta$, $m_{H^\pm}$ plane to determine it is excluded at 95% CL. These hypothesis tests are performed

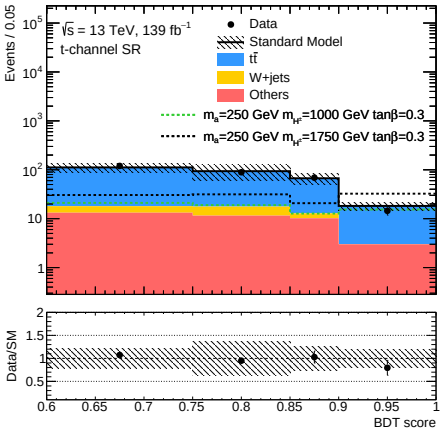


FIGURE 6.20: The distributions of the BDT score at SR definition of the *t*-channel analysis. Uncertainty band includes statistical, experimental and theoretical systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental and theoretical systematics used are detailed in Section 6.6. The normalisation factors obtained from the background-only fit (Table 6.8) are applied to the $t\bar{t}$ and W+jets MC processes.

over the *t*-channel signals, in the SR and the CRs (*CR*($t\bar{t}$) and *CR*(W+jets)) with a profile likelihood along with the observed data. The resulting hypothesis test is presented in Figure 6.24. Upper limit for expected (dashed line) and observed (continue line) are shown in this figure. The excluded signals have a 95% CL limit σ/σ_{BSM} below or equal to 1. Therefore the region below of $m_{H^{\pm}}$ = 800 GeV (1100 GeV) with $\tan\beta = 0.3$ is excluded with 95% CL of the observed (expected) limit.

The search for the single top-quark production associated to a DM generation in the context of 2HDM+a mediator, is studied by two different signatures. One is the presented in this thesis, the tj production. This channel as it is discussed at the beginning of this chapter is dominated for the *t*-channel production. The second is the tW signature (tW associated with DM production), composed by the one lepton (tW_{1L}) and two leptons (tW_{2L}) channels. Figure 6.25 shows a comparison across the mono-top signatures. A statistical combination of the tW_{1L} and tW_{2L} channels

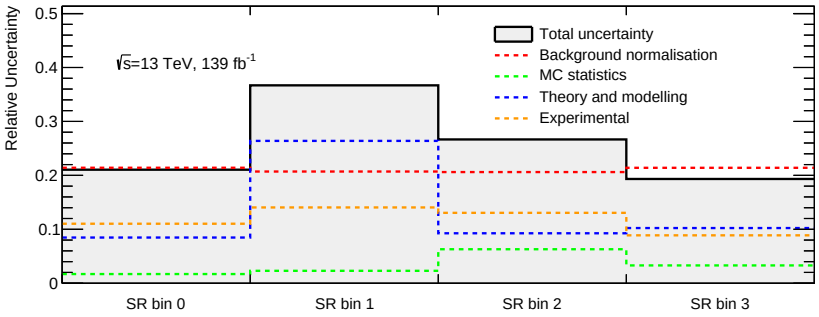


FIGURE 6.21: Relative uncertainties for the total background yield in each SR for the three analysis channels, including the contribution from the different sources of uncertainty. The experimental category contains all detector-related systematic uncertainties and is dominated by jet energy scale and resolution. Individual uncertainties can be correlated, and do not necessarily add up quadratically to the total background uncertainty.

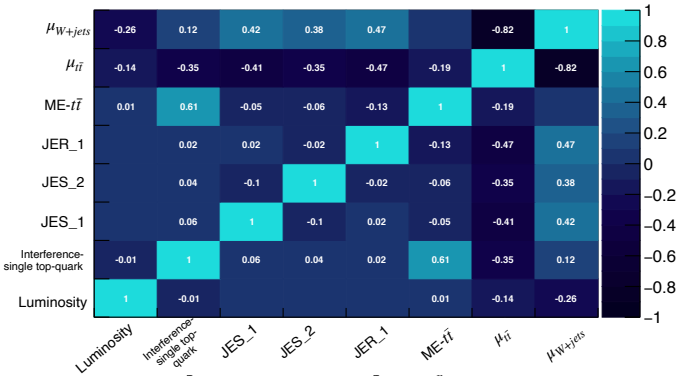


FIGURE 6.22: Correlation matrix of NPs for t-channel analysis with data where the absolute value of the correlation is $\geq 25\%$. Evaluated for the signal sample m_a = 250 GeV, $m_{H^\pm}$ = 1750 GeV and $\tan\beta = 0.3$. The suffix is the decomposition of the respective uncertainty in subgroups.

results is performed to maximise the sensitivity to tW+DM model. The sensitivity

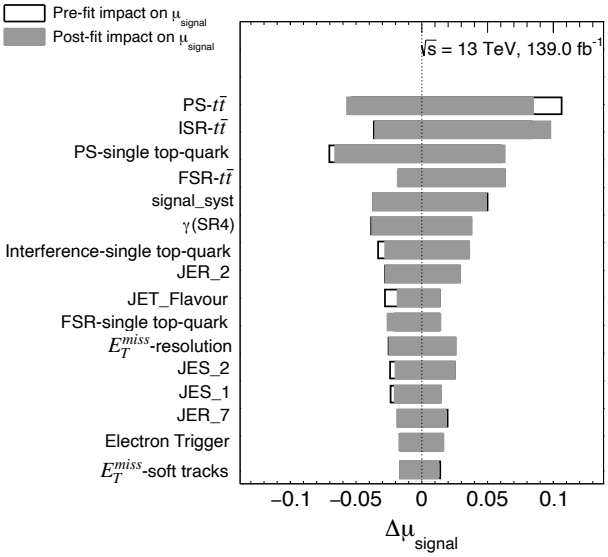
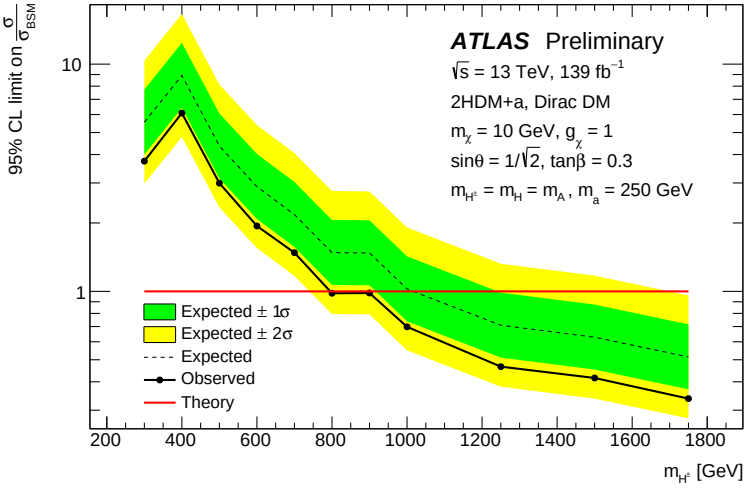
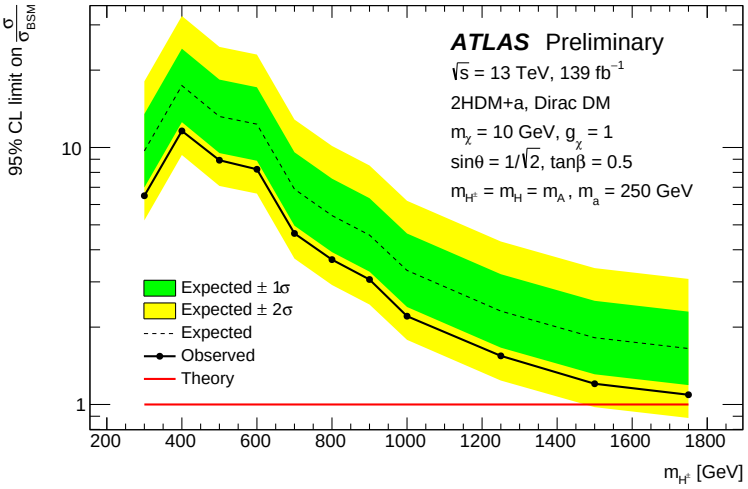


FIGURE 6.23: Ordered list of the 16 NPs with the highest impact on the signal strength μ_{sig}. Evaluated for the signal sample $\mathrm{m_a}$ = 250 GeV, $\mathrm{m_{H^{\pm}}}$ = 1750 GeV and $\mathrm{tan}\beta$ = 0.3. The fitted value of the signal strength parameter μ (top axis values), before the systematics are fitted to the data (empty rectangles) and after (full rectangles). For each NP the shift in the signal strength is obtained for a fit with the NP fixed to its pre-fit or post-fit ($\Delta\mu_{\mathrm{signal}}$) values, respectively. The suffix is the decomposition of the respective uncertainty in subgroups.

of the tj channel is smaller compared to the tW channels due to the smaller cross-section of the DM *t*-channel production process. But at higher $\mathrm{m_{H^{\pm}}}$, i.e. $\mathrm{m_{H^{\pm}}}$ > 1750 GeV the *t*-channel production can become dominant (compared with the tW signature).



(a)



(b)

FIGURE 6.24: The solid (dashed) line shows observed (expected) 95% CLs upper limits for signals with $\tan\beta = 0.3$ **(a)** and $\tan\beta = 0.5$ **(b)**, using fit to four bins in the BDT score. Experimental and theoretical systematic uncertainties, as described in Section 6.6, are applied to background and signal simulated samples and illustrated by the ± 1 standard-deviation (± 2 standard-deviation) green (yellow) [325].

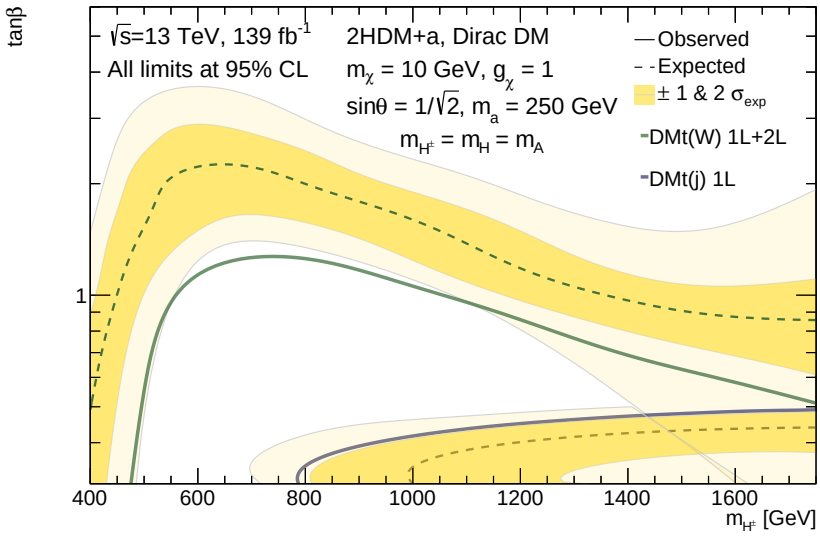


FIGURE 6.25: The expected (dashed line) and observed (solid line) exclusion contours as a function of $(\tan\beta, m_{H^{\pm}})$ for tj+DM (blue line at the bottom of the plot) and tW+DM (solid green-line), below each line is excluded by the analysis. Experimental and theoretical systematic uncertainties, as described in Section 6.6, are applied to background and signal simulated samples and illustrated by the ± 1 standard-deviation and ± 2 standard-deviation yellow bands.

CHAPTER 7

DM mono-top search prospects at the HL-LHC

"If you would be a real seeker after truth, it is necessary that at least once in your life you doubt, as far as possible, all things."

- René Descartes

This chapter describes a search for DM at the future HL-LHC using the ATLAS detector at $\sqrt{s}$ = 14 TeV with 3000 fb^{-1} of assumed integrated luminosity, targeting events with one top-quark and a large amount of E_T^{miss} [326, 327]. This mono-top production was previously searched for by the CDF collaboration at $\sqrt{s}$ = 1.96 TeV using data from the Tevatron with an integrated luminosity of 7.7 fb^{-1} [328], and also by the ATLAS and CMS collaborations using Run 1 data at $\sqrt{s}$ = 8 TeV [329, 330] and part of the Run 2 data at $\sqrt{s}$ = 13 TeV corresponding to an integrated luminosity of 36 fb^{-1} [134, 331]. Different models predict this signature in a resonant and non-resonant production (discussed in Chapter 2). This thesis focuses on the non-resonant production.

This chapter is organised as follows. Section 7.1 describes the event selection of the signal. Section 7.2 shows an optimisation study to perform a cut-based signal analysis. Then, Section 7.3 gives an introduction to the multivariate analysis and the criteria applied for this analysis method. Section 7.4 describes the statistical test to extract the CLs to calculate the upper limits of the signal productions. Finally, Section 7.5 presents a result comparison between the cut-based and the BDT-analysis. An extrapolation to expected results assuming Run 3 data collection (300 fb^{-1}) is also shown.

7.1 Event selection and analysis strategy

The experimental signature of the non-resonant[1] signal events is associated to the top-quark decays. The final state of this channel consists of one lepton (e or μ) from the top-quark decay chain, missing transverse momentum from the decay of the W boson, the DM particles productions, and a b-jet (as can be noticed in Figure 7.1). In Chapter 2 the mechanism of production for this signal is already described. Sensitivity to this signal production is studied expected full luminosity at the HL-LHC (3000 fb^{-1}) and compared to existing limit for Run 2 data (36 fb^{-1}).

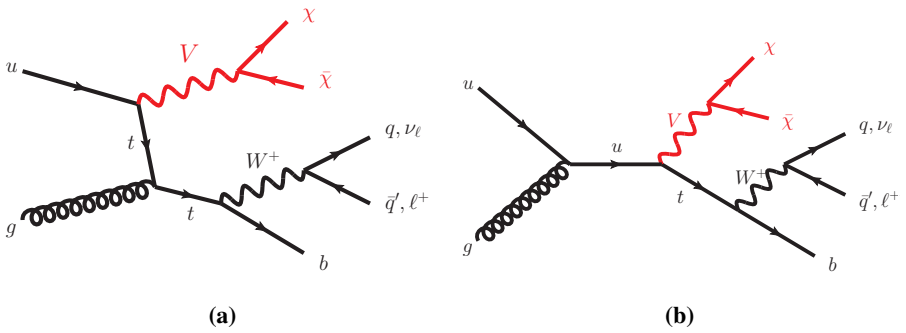


FIGURE 7.1: Representing LO Feynman diagrams corresponding to the mono-top signals searched for non-resonant **(a)** t- and **(b)** *s*-channel DM production in association with a top quark.

The signal event candidates are selected by requiring exactly one lepton (electron or muon) with $p_T > 30$ GeV, exactly one jet with $p_T > 30$ GeV identified as a b-jet and $E_T^{miss} > 100$ GeV. Since the considered signal process favours final states with positive leptons, events with negative lepton charge are rejected[2]. These criteria defines the base selection.

[1] As was mentioned in Chapter 2, previous studies show that the contribution to the search of the resonant DM model, the leptonic channel is negligible compared with the hadronic channel. So, this chapter is only focused in the non-resonant signal production.

[2] The positive leptons from the top quark are favoured in the non-resonant production, since the up-type quark initiated production of top quark is preferred with respect to the top antiquark production, due the PDF structure of the proton.

In addition to the base selection, further discrimination between the signal events and background events is achieved by imposing additional requirements. The transverse mass of the lepton–E_T^{miss} system, $m_T(\ell, E_T^{miss})$, is required to be larger than 100 GeV in order to reduce the background contribution. In background events (mainly W+jets and $t\bar{t}$) the spectrum of this quantity decreases rapidly for values higher than the W-boson mass. For signal events instead, the spectrum has a tail at higher mass values, as seen already in previous searches performed by ATLAS at $\sqrt{s} = 8$ TeV [329] and at $\sqrt{s} = 13$ TeV [134].

It is expected that the lepton and the b-jet are close to each other, since both originate from the decay of a top quark, which is also produced quite boosted. Therefore, events are required to have an azimuthal difference between the lepton momentum and the b-jet momentum directions ($|\Delta\phi(\ell, \text{b-jet})|$) of less than 2.0 rad, this helps to reject mainly W+jets and $t\bar{t}$ backgrounds. This behaviour is presented as well in the ΔR between the lepton and the b-jet. Lower values of $\Delta R(\ell, \text{b-jet})$ distribution can reduce the W+jets and $t\bar{t}$ backgrounds. Table 7.1 shows a summary of the criteria which defines the pre-selection region. Figure 7.2 shows the expected number of jets on the pre-selection region. Figure 7.3 shows the distributions of $\Delta\phi(\ell, \text{b-jet})$, $\Delta R(\ell, \text{b-jet})$, $m_T(\ell, E_T^{miss})$ and E_T^{miss} for this region.

Variable	Requirement
number of leptons, $\mathcal{N}_\ell$	1
$p_T(\ell)$ [GeV]	> 30
Lepton charge sign	+ 1
$p_T(\text{b-jet})$ [GeV]	> 30
E_T^{miss} [GeV]	> 100
number of jets, $\mathcal{N}_{jets}$	< 10
number of b-jets, $\mathcal{N}_{b-jet}$	1
$m_T(\ell, E_T^{miss})$ [GeV]	> 100
$\lvert\Delta\phi(\ell, \text{b-jet})\rvert$ [rad]	< 2.0

TABLE 7.1: Overview of the selection criteria used to define the pre-selection region.

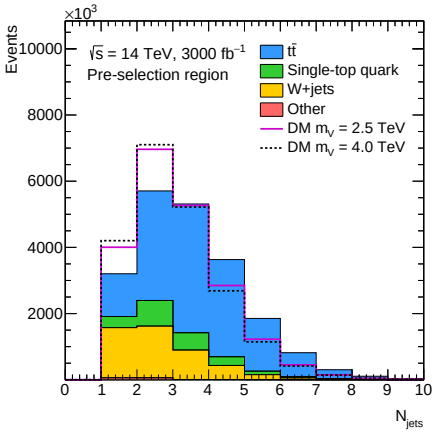


FIGURE 7.2: Distributions of the number of jets ($\mathcal{N}_{\mathrm{jets}}$). The stack distribution shows the background prediction which includes $\mathrm{t\bar{t}}$, single-top quark, W+jets and Other (i.e. Z+jets, dibosons, $\mathrm{t\bar{t}W/Z}$ and tZq). Solid and dashed lines represent the mono-top signal corresponding to a mediator mass of 2.5 and 4.0 TeV, respectively. The background event samples are normalised to their theoretical predictions and the signal event samples are normalised to the number of background events.

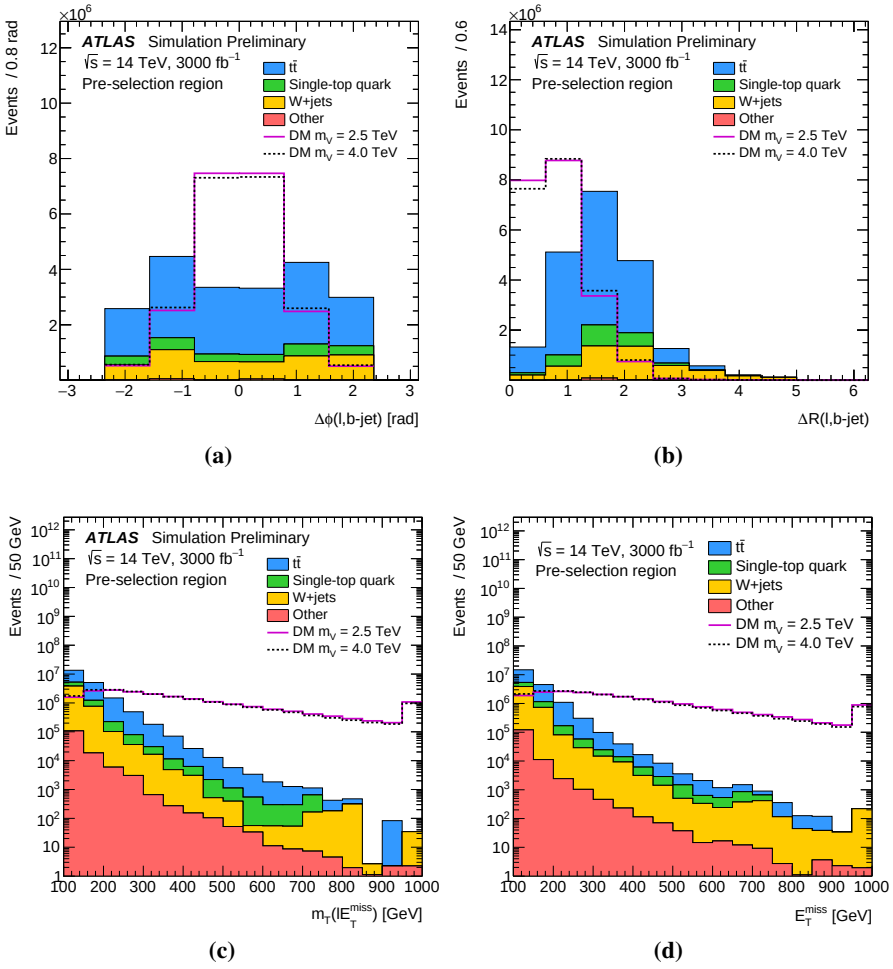
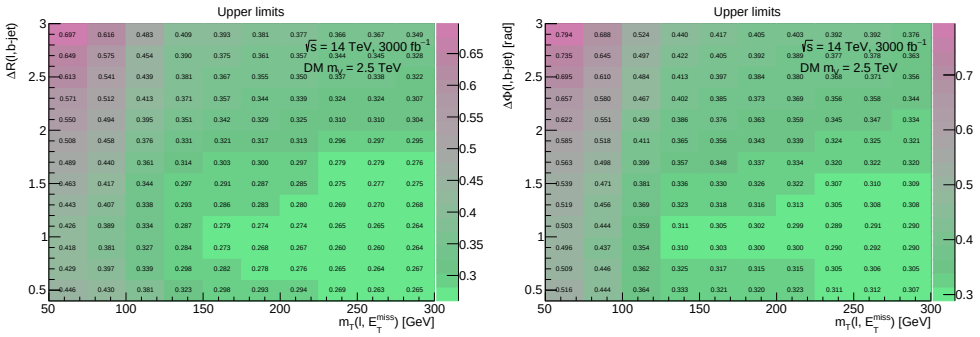


FIGURE 7.3: Distributions of $\Delta\phi$ **(a)** between the lepton and the b-jet, ΔR **(b)** between the lepton and the b-jet, transverse mass of the lepton–E_T^{miss} system **(c)**, and E_T^{miss} **(d)**. The stack distribution shows the background prediction which includes $t\bar{t}$, single-top quark, W+jets and Other (i.e. Z+jets, dibosons, $t\bar{t}W/Z$ and tZq). Solid and dashed lines represent the mono-top signal corresponding to a mediator mass of 2.5 and 4.0 TeV, respectively. The background event samples are normalised to their theoretical predictions and the signal event samples are normalised to the number of background events [327].

The following sections develop two different approaches to perform the analysis, one based on simple selection criteria (cut-based), and another using a multivariate technique. These studies are optimised to define the best SR of the mono-top signal generation (under their respective philosophic implementation).

7.2 Cut-based analysis

Events used in this study are selected with the pre-selection criteria together with additional requirements in three variables properly optimised. The optimisation is done by varying systematically the thresholds of $\Delta\phi(\ell, \text{b-jet})$, $\Delta R(\ell, \text{b-jet})$ and $m_T(\ell, E_T^{miss})$. The tested selection on $m_T(\ell, E_T^{miss})$ ranges between > 50 GeV and > 300 GeV in steps of 25 GeV. Selections on the angular variables range from < 0.5 to < 2.9, in steps of 0.2. The figure of merit used in this process is the excluded upper limit obtained from the likelihood fit (procedure explained in Section 7.4), without considering the systematics uncertainties. The E_T^{miss} is used as discriminant variable in the fit. The results of the optimisation are shown in Figure 7.4.



(a) **(b)**

FIGURE 7.4: Expected excluded signal strength as a function of the value of the lower $m_T(\ell, E_T^{miss})$ threshold and of the higher $\Delta R(\ell, \text{b-jet})$ **(a)** and $\Delta\phi(\ell, \text{b-jet})$ **(b)** threshold for a mediator mass m_V = 2.5 TeV.

The fitting procedure is the same as the one described in Section 7.5. The benchmark signal used for the study correspond with the mediator mass of m_V = 2.5 TeV. Optimal set of requirements is found to be the pre-selection criteria with $\Delta R(\ell, \text{b-jet}) < 1.2$ and $m_T(\ell, E_T^{miss}) > 225$ GeV and with no requirements on $\Delta\phi(\ell, \text{b-jet})$. Additionally, a requirement on $E_T^{miss} > 150$ GeV is applied to further reduce the background. These criteria define the selection region of the cut-based analysis, and they are summarised in Table 7.2.

Criteria	Requirement
Pre−selection	applied
$\Delta R(\ell, \text{b-jet})$	< 1.2
$m_T(\ell, E_T^{miss})$ [GeV]	> 225
E_T^{miss} [GeV]	> 150

TABLE 7.2: Overview of the selection criteria of the cut-based analysis.

7.3 BDT-based analysis

Alternative to the cut-based, a gradient BDT algorithm provided by the Toolkit for Multivariate Analysis [313] is used to enhance signal selection. All the details of the procedure and algorithm characteristics can be found in Appendix A.

The BDT is trained to discriminate signals from the dominant $t\bar{t}$ background. To train the BDT, since no significant difference is observed for the different mediator mass values, the simulation event sample with m_V = 2.5 TeV is used. Half of the events of both signal and background event samples are selected randomly and used to train the BDT. The other half is used to probe the BDT behaviour in order to avoid overtraining. The pre-selection region (see Table 7.1) is applied to the simulated event samples to perform the training procedure.

Five basic parameters of the BDT are studied during the optimisation: the maximum depth of each tree, the minimum percentage of events at each node (minimum

node size), the granularity number of cuts (number of cuts), the number of trees (i.e. number of iterations) and the parameter which controls the gradient descent of the boost (learning rate β, usually named shrinkage). As already mentioned in Chapter 6, these parameters control the behaviour of the BDT can be tuned to obtain the maximum performance. The chosen final configuration is shown in Table 7.3. The purpose of each value is detailed in Appendix A.

Parameter	Setting
BoostType	GradBoost
Number of trees	1000
Minimum node size	2.5%
Maximum depth of tree	2
Number of cuts	30
Learning rate β	0.20

TABLE 7.3: BDT parameter settings optimised for the search discussed in this analysis.

The variables entering the BDT are selected from a pool of fundamental quantities, such as $\mathrm{p_T}$ of jets and b-jets, and angular distances. The variables selected are the ones showing "a priori" the best discriminating power. In particular, $\Delta\phi(\ell, \mathrm{b\text{-}jet})$ and $\mathrm{m_T}(\ell, \mathrm{E_T^{miss}})$ are found to be the most effective variables. A full list and description of the variables used in the BDT training is given in Table 7.4, and their input (training) distributions are presented in Figures 7.5 and 7.6.

Variable name	Description
Kinematic variables	
$\mathrm{E_T^{miss}}$	Magnitude of the missing transverse momentum
$\mathrm{p_T}(\mathrm{b\text{-}jet})$	Transverse momentum of the b-jet
$\mathrm{p_T}(\mathrm{jet\ 1})$	Transverse momentum of the leading jet
$\mathrm{p_T}(\ell)$	Transverse momentum of the lepton
$\mathrm{m_T}(\ell, \mathrm{E_T^{miss}})$	Transverse mass of lepton–$\mathrm{E_T^{miss}}$ system
Azimuthal differences	
$\Delta\phi(\ell, \mathrm{jet\ 1})$	$\Delta\phi$ between the lepton and the leading jet
$\Delta\phi(\ell, \mathrm{b\text{-}jet})$	$\Delta\phi$ between the lepton and the b-jet
$\Delta\phi(\ell, \mathrm{E_T^{miss}})$	$\Delta\phi$ between the lepton and $\mathrm{E_T^{miss}}$
Angular distance differences	
$\Delta\mathrm{R}(\ell, \mathrm{jet\ 1})$	$\Delta\mathrm{R}$ between the lepton and the leading jet
$\Delta\mathrm{R}(\ell, \mathrm{b\text{-}jet})$	$\Delta\mathrm{R}$ between the lepton and b-jet
Masses	
Leading-jet mass	Mass of the leading jet

TABLE 7.4: List of variables entering the BDT and their definitions.

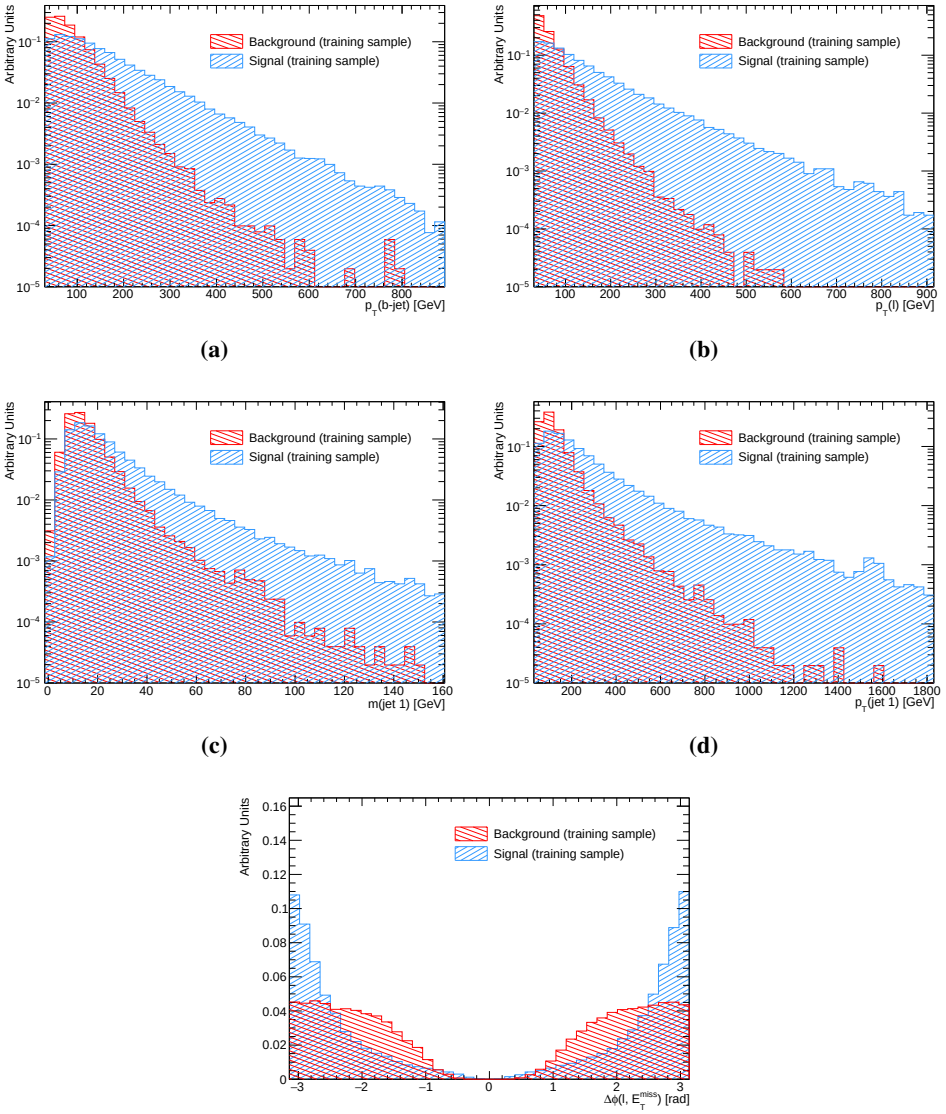


FIGURE 7.5: Distribution of signal and background for the different variables entering the BDT training. Signal corresponds with a mediator mass m_V = 2.5 TeV and $t\bar{t}$ background. There are presented the p_T(b-jet) **(a)**, $p_T(\ell)$ **(b)**, Leading-jet mass **(c)** p_T(jet 1) **(d)** and $\Delta\phi(\ell, E_T^{miss})$ **(e)**.

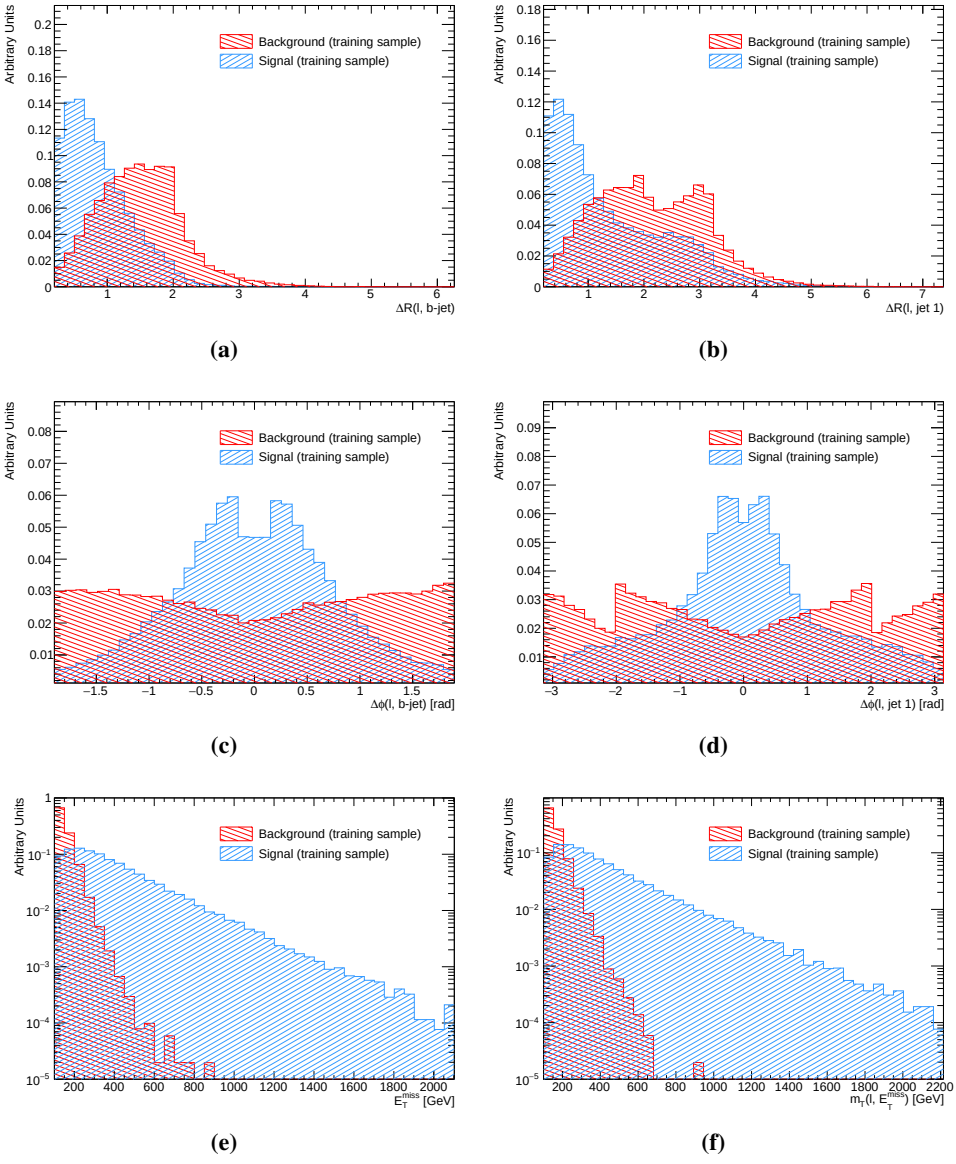


FIGURE 7.6: Distribution of signal and background for the different variables entering the BDT training. Signal corresponds with a mediator mass m_V = 2.5 TeV and $t\bar{t}$ background. There are presented the $\Delta R(\ell, \text{b-jet})$ **(a)**, $\Delta R(\ell, \text{jet } 1)$ **(b)**, $\Delta\phi(\ell, \text{b-jet})$**(c)**, $\Delta\phi(\ell, \text{jet } 1)$ **(d)**, E_T^{miss} **(e)** and $m_T(\ell, E_T^{miss})$ **(f)**.

Figure 7.7 depicts the distributions of BDT scores obtained for background and signal events after the training phase (dotted markers) and testing phase (filled area). In these distributions overtraining can be detected by comparing the score distributions for the training and testing samples for both signal and background. Overtraining would show as difference between testing and training distributions, no such effect is observed in this case. To show better that there are no differences between training and testing is added the ratio between them, which shows that there is no overtraining (since their ratio is compatible with one, including statistical variations at 20%).

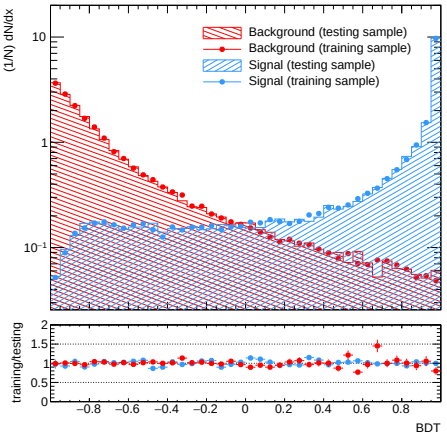


FIGURE 7.7: Comparison of the BDT scores obtained for signal m_V = 2.5 TeV and $t\bar{t}$ background in the BDT training and testing phase, respectively. Both trainings are found to be not overtrained.

Figure 7.8 shows the signal efficiency versus background rejection (ROC curve) of the final BDT. A high value of the integral of the curve corresponds to a large discrimination power of the BDT, so values close to 1 are desired. In this analysis the achieved ROC-curve integral is 0.95.

It is important to check correlation between variables. Figure 7.9 shows the matrices indicating the correlation among the different training variables (training sample) for signal and background samples. In Appendix A is explained the linear correlation coefficients and their utility. For the signal is seen a moderate correlation between

E_T^{miss} and $m_T(\ell, E_T^{miss})$ (linear coefficient = 61), while for the background the correlation is much lower. This is due to the fact that the signal is characterised by having a bigger contribution on the E_T^{miss} variable at high values than the background. Then E_T^{miss} with p_T(jet 1) has a large correlation for the signal (linear coefficient = 78). This correlation appears since the E_T^{miss} is calculated from the visible objects (like jets). A moderate correlation is shown for the signal and not for background samples, since p_T(jet 1) discribution is "softer" for the background. Also, m(jet 1) with p_T(jet 1) has a large correlation for the signal (linear coefficient = 72), since is the same object. For background, only a moderate correlation between $m(jet1)$ with p_T(jet 1) (linear coefficient = 71) is shown. Strong correlations are not observed.

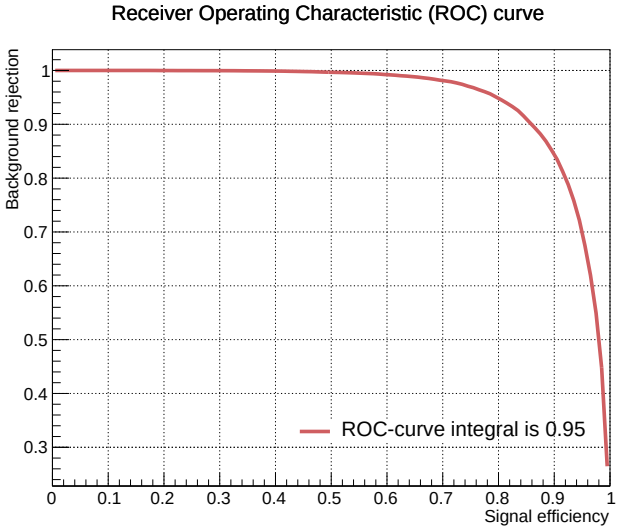
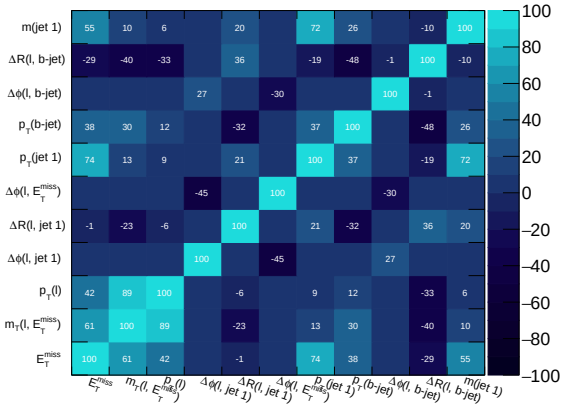
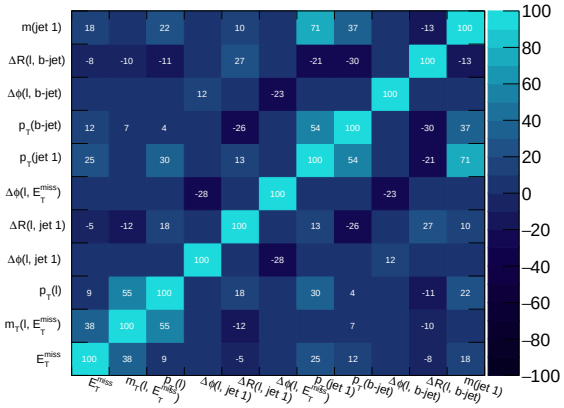


FIGURE 7.8: ROC-curve, i.e. signal efficiency versus background rejection. The ROC-curve is obtained using signal m_V = 2.5 TeV.



(a)



(b)

FIGURE 7.9: Correlation matrix for $m_V = 2.5$ TeV signal **(a)** and for $t\bar{t}$ background **(b)**.

The distribution of the BDT score in pre-selection region is presented in Figure 7.10. Signal is clearly picked to high values of BDT score. The SR is defined by requiring events with BDT score > 0.9 and E_T^{miss} > 150 GeV. This value is chosen to maximise the significance while leaving sufficient statistics. Table 7.5 shows the SR

selection criteria for the BDT-based analysis.

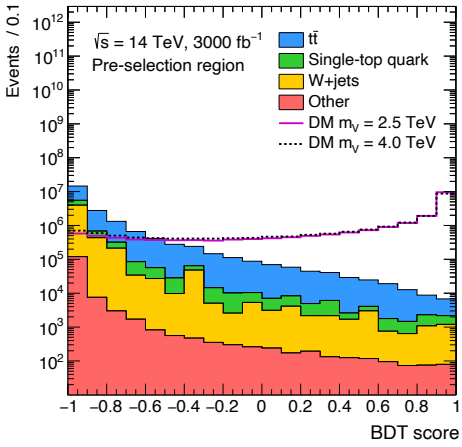


FIGURE 7.10: Response of the BDT algorithm for events in the pre-selection region. The stack distribution shows the background prediction which includes $t\bar{t}$, single-top quark, W+jets and Other (i.e. Z+jets, dibosons, $t\bar{t}W/Z$ and tZq). Solid and dashed lines represent the mono-top signal corresponding to a mediator mass of 2.5 and 4.0 TeV, respectively. The background event samples are normalised to their theoretical predictions and the signal event samples are normalised to the number of background events.

Criteria	Requirement
$\mathrm{Pre-selection}$	applied
$\mathrm{BDT\,score}$	> 0.9
E_T^{miss} [GeV]	> 150

TABLE 7.5: Overview of the SR criteria for the BDT-based analysis, pre-selection correspond to criterion defined in Table 7.1.

7.4 Likelihood fit

Hypothesis testing is performed using a frequentist approach implemented in the TRexFitter package (within the RooStats framework [332]), which uses the asymptotic approximation described in Ref. [333].

For the statistical analysis all backgrounds except the $t\bar{t}$ production are merged in a non-$t\bar{t}$ background to avoid problems of poor statistics in the SR. This allows to use a binned likelihood fit. The shape of the E_T^{miss} distribution is used in the statistical analysis, since a study of different variables shows this variable being the most sensitive for the presence of new physics. The binning of this distribution is optimised for the sensitivity of the analysis in the SR while ensuring the stability of the fit. This results in a non-equidistant binning which exhibits wider bins in regions with a large signal contribution, while preserving a sufficiently large number of background events in each bin: [150-200], [200-250], [250-300], [300-450], [450-700], [700-1000] GeV.

The statistical interpretation used is detailed in Section 6.7. It is noted here that the quantities described by μ_{bkg} in Equation 6.3 correspond to normalisation correction factors for a given SM background process. In this analysis these factors (applied to the main background processes) are not defined, since there is not data to extract them. Therefore, μ_{bkg} are not allowed to vary during the fit, and they are fixed to unity.

7.5 Results

The BDT-based approach is selected given the significantly better results obtained compared to the cut-based analysis, as it will be shown later in this section. Thus, unless explicitly stated, the content on this section refers to the BDT-analysis.

The signals analysed are those with mediator mass between m_V = 1.0 and 5.0 TeV, in steps of 0.5 TeV for the non-resonant model assuming m_χ = 1 GeV, a = 0.5 and g_χ = 1. The result does not include MC statistical uncertainties, since for the HL-LHC stage, they are expected to be negligible. But it incorporates the effects of systematic uncertainties. The uncertainties due to the limited size of the simulated samples are

taken into account in each bin of the fitted distributions as NPs. The theoretical modelling of signal and background has the largest prior, 15%. The second largest source of uncertainty is the one relative to the E_T^{miss}reconstruction, with 6% prior. The JES and JER contribute with a total of 5%. The uncertainty on the requirements for pile-up jets rejection is 5%. The ones on lepton identification and b-tagging efficiencies are 1.2% and 2.5%, respectively. The uncertainty on the expected luminosity is also taken into account, with a 1% effect. A summary of the systematic uncertainties and their effect estimation is shown in Table 7.6. These uncertainties are estimated from previous analysis at $\sqrt{s}$ = 13 TeV corresponding to an integrated luminosity of 36 fb^{-1} of data presented by the ATLAS Collaboration [134], but some reductions are implemented for the HL-LHC stage, since it is expected that uncertainties will be better understood at that time. The b-tagging efficiencies, E_T^{miss} reconstruction, the JES and JER are reduced to the half. The lepton identification efficiency, the pile-up and the modelling systematics uncertainties are expected to be the same.

Systematic effects	
Theoretical uncertainties	
Background	15%
Signal	15%
Experimental uncertainties	
E_T^{miss} soft-term resolution	6%
JES and JER	5%
Pile-up modelling	5%
b-tagging efficiency	2.5%
Lepton identification efficiency	1.2%
Luminosity measurement	1%

TABLE 7.6: Summary of the sources of systematic uncertainties and their estimated effects, extrapolated from the same analysis developed by ATLAS Collaboration at $\sqrt{s}$ = 13 TeV corresponding to an integrated luminosity of 36 fb^{-1} of data. Accordingly reduced to the expected HL-LHC performance.

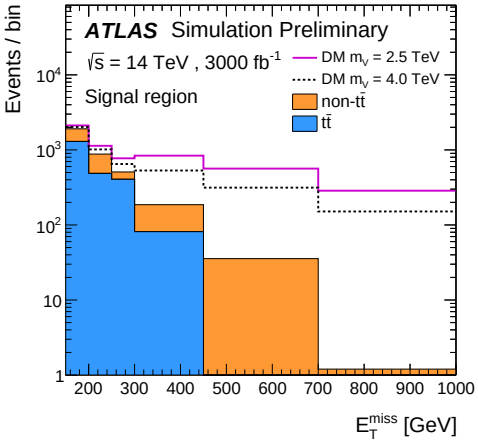


FIGURE 7.11: Expected post-fit E_T^{miss} distribution in the SR. The stack distribution shows the $t\bar{t}$ and non-$t\bar{t}$ background predictions. Solid and dashed lines represent the mono-top signal corresponding to a mediator mass of 2.5 and 4.0 TeV, respectively. The signal event samples are normalised to the number of background events. The binning is the same as the optimised, non-equidistant binning used in the fit. Last bin includes overflow events [326, 327].

Figure 7.12 shows the expected 95% CL upper limits as a function of the mediator mass for the non-resonant model assuming m_χ = 1 GeV, a = 0.5 and g_χ = 1. The expected mass limit at 95% CL is 4.6 TeV while the discovery reach (based on 5σ significance, it is the upper limit at which a discovery is rejected) is 4.0 TeV. The effect of possible improvements in the systematic uncertainties at the HL-LHC project is estimated by reducing them by a half. This estimation has the effect of increasing the exclusion limit (discovery reach) by 80 (50) GeV.

Comparing with Run 2 analysis corresponding to an integrated luminosity of 36 fb^{-1} of data [134], the expected mass limit at 95% CL is improved by 2 TeV. This result was obtained with the combination of the leptonic and hadronic decays of the W associated to the top-quark. The Run 2 analysis with only the leptonic channel reaches an expected limit of 1.3 TeV, which should be compare with the estimated 4.6 TeV for the HL-LHC.

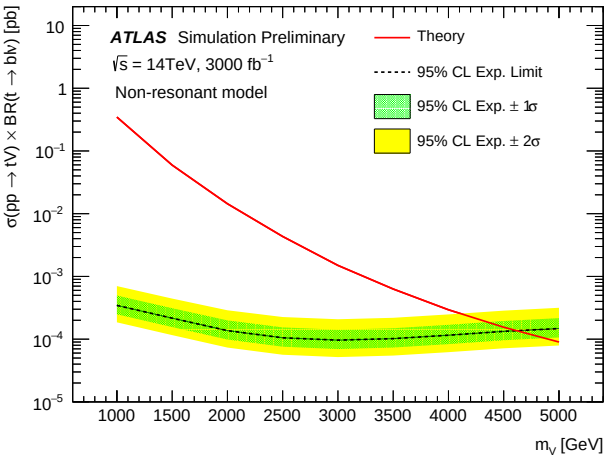


FIGURE 7.12: Expected 95% CLs upper limits on the signal cross-section as a function of the mass of the mediator for the non-resonant model assuming m_χ = 1 GeV, a = 0.5 and g_χ = 1 using a BDT analysis. The MC statistical uncertainty is not considered but the full set of systematics, extrapolated from a $\sqrt{s}$ = 13 TeV analysis is considered [326, 327].

The expectations for the equivalent of Run 3 integrated luminosity (300 fb^{-1}) are also tested, obtaining an exclusion limit (discovery reach) of 3.7 TeV (3.2 TeV).

On the other hand, the expected mass limit at 95% CL obtained with the cut-based analysis, assuming an integrated luminosity of 3000 fb^{-1} and including the same systematic uncertainties (presented in Figure 7.6), is 3.2 TeV. As anticipated at the beginning of the section, this limit is significantly lower than what is obtained with the BDT-based analysis.

CHAPTER 8

Conclusion

"Never confuse education with intelligence,
you can have a PhD and still be an idiot."
- Richard P. Feynman.

Astronomical and cosmological observations support the existence of invisible matter that has only been detected through its gravitational effects, thus making it very difficult to study. This component, called DM, makes up about 26.8% of the energy of the known universe.

Several theories provide DM candidates though an interesting possibility is that DM has a mass of order 100 GeV and it interacts weakly with matter (i.e. as WIMP). Since the start of the LHC data taking there has been a continuous evolution in the design of particle physics models incorporating this new kind of particle. The present thesis is based on one of the latest evolution of these models, featuring both DM and extended Higgs sector, implemented by an extension of the SM featuring a two-Higgs-doublet model and an additional pseudo-scalar (known as 2HDM+a). In this framework, the pseudo-scalar mediator (a) and the CP-odd pseudo-scalar (A) decay into DM particles. Among the possible final states through which DM production would manifest itself it chooses states with a single top quark in the final state (mono-top signature). This signature is particularly interesting from the theoretical point of view, as it implies the coupling of DM to third generation quarks. Besides this, it is particularly promising from the experimental point of view, as the final state including the leptonic decay of the W boson from the single top quark has only moderate SM backgrounds, and it presents several kinematic handles which allow the separation of the signal from the background. This thesis explores the DM production in the tj signature. Two processes allow this production *t*-channel and *s*-channel though it is dominated by the *t*-channel production, so only this process is studied in detailed.

Due to the small production cross-section of the tj signature a cut-based analysis is not sufficient to achieve a good discrimination of the signal against SM background processes. Therefore, an approach based on multivariate analysis techniques is applied to enhance the signal to background separation. In particular, a BDT classifier is trained specifically to separate the tj signal from all backgrounds.

Hypothesis tests are calculated in order to assess which regions of the 2HDM+a parameter space can be excluded (signal hypothesis is rejected). The $\mathrm{CL_s}$ construction is evaluated to determine whether a given point in the (tanβ, $\mathrm{m_{H^{\pm}}}$) plane can be excluded at 95% CL. These hypothesis tests are performed over the tj signals, in the SR and the CRs (*CR(*$\mathrm{t\bar{t}}$*)* and *CR*(W+jets)) with a profile likelihood along with the observed data. The resulting hypothesis tests are presented in Figure 6.24. This figure shows the upper limits for signals with tanβ = 0.3 and tanβ = 0.5 of *t*-channel analysis. The excluded signals have a 95% CL limit on $\sigma/\sigma_{\mathrm{BSM}}$ below or equal to 1. Therefore the region below $\mathrm{m_{H^{\pm}}}$ = 800 GeV (1100 GeV) with tanβ = 0.3 is excluded with 95% CL of the observed (expected) limit. While among tanβ = 0.5 signals none of the points are excluded. The most dominant uncertainties are related from the $\mathrm{t\bar{t}}$ modelling uncertainties.

The search for the single top-quark production associated to DM particles in the context of 2HDM+a mediator, is studied by two different signatures. One is the presented in this thesis, i.e. the tj production (as mentioned it is dominated for the *t*-channel production). The second is the tW signature, covering for the one lepton ($\mathrm{tW_{1L}}$) and two leptons ($\mathrm{tW_{2L}}$) generations. Figure 6.25 shows a comparison across the mono-top signatures. A statistical combination of the results obtained for the $\mathrm{tW_{1L}}$ and $\mathrm{tW_{2L}}$ channels is performed to maximise the sensitivity to tW+DM model. The sensitivity of the tj channel is smaller compared to the tW channel due to its smaller cross-section. Anyhow at higher $\mathrm{m_{H^{\pm}}}$, i.e. $\mathrm{m_{H^{\pm}}}$ > 1750 GeV, the *t*-channel production can become dominant.

As mentioned above, understanding the nature of DM is one of the main objectives of the High Energy Physics community. Therefore, a prospective study is also presented, developed as one of the inputs for the CERN report with the European Strategy for the particle physics about the expected potential and performance of the

HL-LHC. The potential discovery of the non-resonant production of DM in association with a right-handed top quark in the context of an effective DM flavour-changing neutral interaction is also presented. The HL-LHC project is expected to operate at $\sqrt{s}$ = 14 TeV aiming to provide a total integrated luminosity of 3000 fb^{-1}. In the work presented in this thesis, the number of signal and background events is estimated from simulated particle-level truth information after applying smearing functions to mimic an upgraded ATLAS detector response in the HL-LHC environment. A BDT algorithm is used to enhance signal selection, discriminating the signals from the dominant $t\bar{t}$ background.

Hypothesis tests are calculated in order to access the value of the mediator mass for the non-resonant mono-top model to be excluded. The Figure 7.12 shows the expected 95% CL upper limits as a function of the mediator mass for the non-resonant model assuming m_χ = 1 GeV, a = 0.5 and g_χ = 1. The result does not include MC statistical uncertainties, but incorporates the full systematic effects of systematic uncertainties. These uncertainties are estimated from previous analysis using Run 2 data at $\sqrt{s}$ = 13 TeV presented by the ATLAS Collaboration [134] (considering some expected reductions for the HL-LHC stage). The theoretical modelling of signal and background has the largest prior. The expected mass limit at 95% CL is 4.6 TeV while the discovery reach (based on 5σ significance) is 4.0 TeV. A summary of the expected mass limits obtained among different scenarios are given in Table 8.1. In this table can be seen that only the leptonic production at HL-LHC environment duplicates the expected limit obtained in Run 2 stage, in which one combines the leptonic and hadronic channels.

Experiment environment	Analysis technique	Expected mass limit at 95% CL [TeV]
HL-LHC, $\sqrt{s}$ = 14TeV, 3000 fb^{-1} (leptonic)	multivariate technique	4.6
HL-LHC, $\sqrt{s}$ = 14TeV, 3000 fb^{-1} (leptonic)	cut-based	3.2
Run 3, $\sqrt{s}$ = 14 TeV, 300 fb^{-1} (leptonic)	multivariate technique	3.7
Run 2, $\sqrt{s}$ = 13 TeV, 36 fb^{-1} (leptonic + hadronic)	cut-based	2.6
Run 2, $\sqrt{s}$ = 13 TeV, 36 fb^{-1} (leptonic)	cut-based	1.3

TABLE 8.1: Summary of the expected limits of the non-resonant mono-top production among different LHC stage and final decays.

Capítulo 9

Resumen

¡Seamos realistas, pidamos lo imposible!
-Ernesto Che Guevara

El modelo estándar de la física de partículas (usualmente conocido como "Standard Model") es una teoría desarrollada en la década de los 70 [1–4]. Es una teoría de campo cuántico donde las partículas de materia y antimateria se definen como excitaciones locales de campos que permean el espacio. Basada en las ideas de la unificación y simetrías [11], describe las partículas elementales como entes irreducibles cuya cinemática está regida por las cuatro interacciones fundamentales conocidas (exceptuando la gravedad, cuya principal teoría, la relatividad general, no encaja con la formulación matemática del mundo cuántico). Hasta la fecha, casi todas las pruebas experimentales descritas por el modelo estándar están de acuerdo con sus predicciones. Sin embargo, estas no alcanzan a ser una teoría completa de las interacciones fundamentales, debido a varias cuestiones sin resolver.

Una de las cuestiones que el modelo estándar deja fuera de su alcance, es la explicación del 26.8% del universo concido [26]. Varias observaciones astronómicas y cosmológicas respaldan la existencia de materia invisible (usualmente llamada materia oscura, "dark matter" en inglés), que solo puede detectarse a través de sus efectos gravitacionales [23–25]. Debido a la falta de una partícula que se comporta como un candidato a materia oscura, varias modelos fueron propuestos para extender el modelo estándar.

Uno de los candidatos a materia oscura son los WIMPs (sigla en inglés de weakly interacting massive particles; en castellano "partículas masivas débilmente interactuantes"). Estas partículas, además de interactuar por medio de la gravedad con la materia visible, también lo hacen a través de interacciones nuevas (no asociadas al modelo estándar) que se estiman tienen una magnitud similar a las interacciones asociadas a

la fuerza nuclear débil. No se pueden ver directamente, ya que no interactúan y no emiten radiación electromagnética. Dentro del "paradigma de las WIMP", la materia oscura se produjo en el origen del universo e inicialmente se encontraba en contacto térmico con el modelo estándar. Sin embargo, al expandirse y enfriarse el universo, estas partículas salieron del equilibrio térmico lo cual produjo un "congelamiento" en el tiempo de la densidad de materia oscura. Su existencia explicaria naturalmente las mediciones de la densidad de materia en el universo temprano (relic density) [48, 50].

Un enfoque para la detección de WIMPs en la naturaleza es la posibildad de crearla en el laboratorio. Los detectores dentro del Gran Colisionador de Hadrones (LHC) pueden ser capaz de detectar WIMPs, debido a que interacciona con la materia ordinaria, puede ser detectada indirectamente por (grandes cantidades de) la energía faltante o el impulso que escapan a los detectores del LHC [115].

Esta tesis esta principalmente enfocada a la búsqueda de materia oscura producida en el LHC, usando el detector ATLAS (A Toroidal LHC Apparatus).

9.1 Producción de materia oscura

La producción de materia oscura en colisionadores generalmente implica una partícula mediadora, que conecta su producción, con partículas del modelo estándard (por ejemplo, un jet[1], un fotón, un bosón, etc.). Esta estrategia experimental de búsqueda de materia oscuras incluye la medición de la energía transversal faltante (E_T^{miss}), ya que la partícula desconocida creada candidato a materia oscura escapa sin ninguna pista dentro del detector. Este estado final está motivado por la suposición de que las colisiones *pp* producen dos WIMPs, que se crean sin ser detectados, y la presencia de un solo objeto visible dentro del detector, estos canales a menudo se llaman mono-X [115] .

[1] Palabra inglesa que significa "chorro"

9.1.1 Modelo de dos dobletes de Higgs cargados

Una de las extensiones más simples posibles del modelo estándar, es el modelo de dos dobletes de Higgs cargados (2HDM) [124].

Hay diferentes tipos de los modelos de 2HDM, lo mas conocidos son:

- El tipo I 2HDM que se obtiene si todos los quarks y leptones cargados se acoplan solo al segundo doblete escalar (Φ_2) [126].
- El tipo II 2HDM, uno de los más estudiados, dado que los acoplamientos del modelo estándar supersimétrico mínimo (MSSM) son un subconjunto de los acoplamientos de este modelo [127, 128]. Donde los quarks de tipo up se acoplan a Φ_2 y los quarks de tipo down se acoplan a Φ_1.

En esta tesis se considera un sector extendido del tipo II 2HDM junto con un mediador pseudoescalar adicional para la produción de materia oscura (2HDM+a) [135]. Esta categoría de modelos representa una de las teorías ultravioleta (UV) completas y renovables más simples con un mediador de espín 0 [114, 122, 136–141].

Dentro de la producción de un top en solitario como estado final, se puede encontrar la produción tj (un top mas un jet) o tW (un top mas un bosón W). La sección transversal del top en solitario con materia oscura en estado final se ve reforzada por la producción on-shell[2] de los bosones $\mathrm{H}^{\pm}$, que puede contribuir a la restricción de los parámetros libres en los modelos 2HDM+a. La produción tj contiene las producciones de materia oscura por los canales- t y s. Estos procesos se muestran en la Figura 9.1.

Los procesos del canal-t son las contribuciones dominantes dentro de las señales de tj. Provienen de los dos diagramas intermedios que se muestran en la Figura 9.1, estos diagramas interfieren destructivamente[3], y este efecto disminuye con el aumento de la masa de los $\mathrm{H}^{\pm}$, donde el canal-t puede alcanzar más sensibilidad.

[2]En la teoría de campo cuántico, las partículas virtuales se nombran fuera de off-shell, ya que no satisface la relación energía-momento; las partículas de intercambio real que si satisfacen esta relación y se denominan en on-shell.

[3]La interferencia destructiva observada garantiza la unitaridad del proceso.

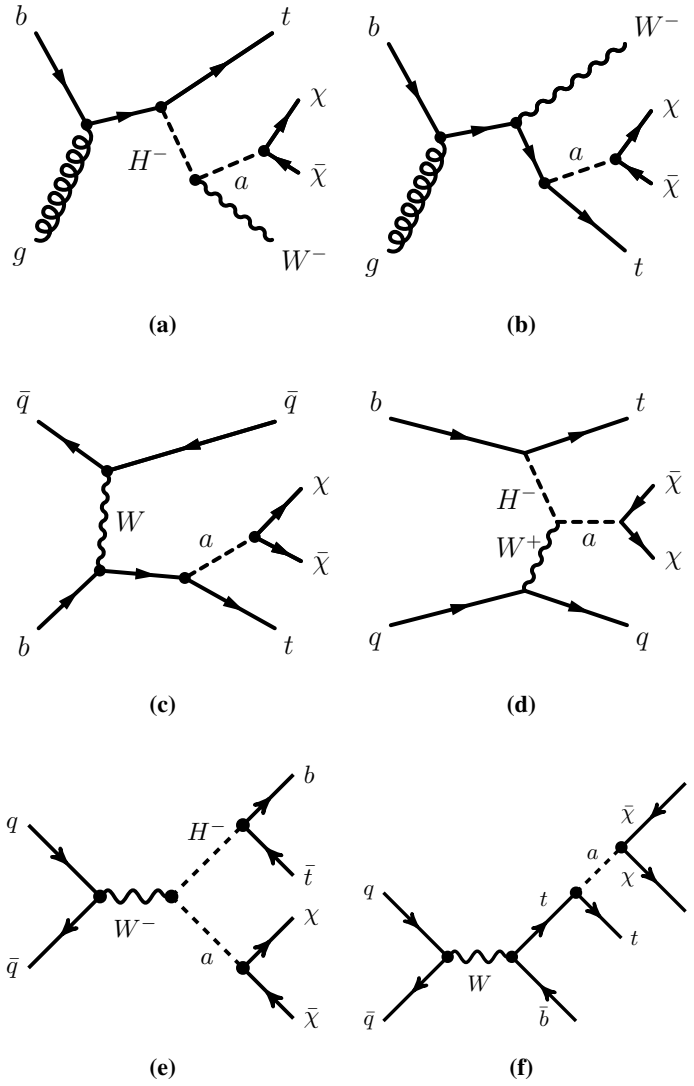


FIGURA 9.1: Representación esquemática de los modos de producción y decaimiento dominantes para el modelo 2HDM+a: tW : **(a)**-**(b)**, canal-t: **(c)**-**(d)** y canal-s: **(e)**-**(f)** en estados finales de quark de un top.

9.1.2 Producción mono-top no resonante

Esta tesis se centra en modelos de señal mono-top, por lo que se ha estudiado la sensibilidad de la producción no resonante de un estado exótico V, el cual se descompone en un par de candidatos a materia oscura en asociación con un quark top [131]. La producción mono-top no resonante es a través de una interacción neutral que cambia el sabor del quark involucrado, como se muestra en la Figura 9.2. Donde un quark top derecho (t), un quark up de sabor ligero (u) y un mediador vectorial masivo V interacionan.

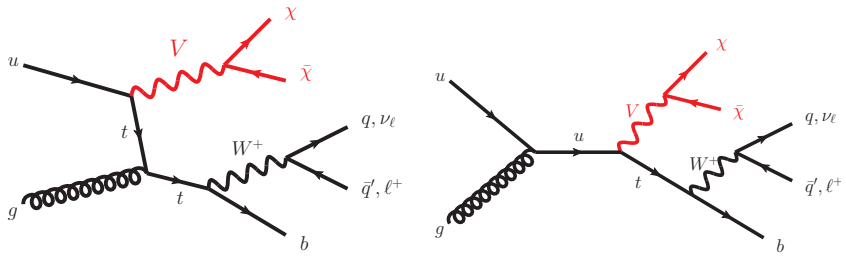


FIGURA 9.2: Los diagramas de Feynman representativos correspondientes a las señales mono-top no resonante. La generación canal-t **(a)** y la canal-s **(b)** de materia oscura asociada a la producción de un quark top.

9.2 ATLAS y el LHC

El LHC es el mayor y más potente acelerador de partículas construido. Éste puede acelerar protones, con sus 27 km de circunferencia, hasta una energía nominal de centro de masas de 14 TeV. Durante el Run 1 (2010-2012), el LHC operó a 7 y 8 TeV, mientras produjo colisiones a 13 TeV en el Run 2 (2015-2018).

Estas altas energías, son posibles gracias a campos magnéticos enormes. El LHC esta diseñado con dipolos magnéticos superconductores que llegan a generar un campo magnético de 8 T. Los cuales para que actúen como superconductores, los dipolos son enfriados a una temperatura de 1.9 K utilizando helio líquido superfluido.

Las cavidades de radio-frecuencia se usan para acelerar los protones, cuadrupolos magnéticos para enfocar los haces y otros sistemas complejos de imanes completan el gran LHC.

El LHC es capaz de proporcionar un gran número de colisiones por segundo, lo que se denomina luminosidad. La luminosidad nominal del LHC es de L = $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, aunque esa cantidad ya ha sido sobrepasada durante el Run 2 (en Octubre, en el año 2017).

En el LHC se sitúan cuatro grandes detectores, cada uno en uno de los puntos de colisión de los haces: ATLAS, CMS, LHCb y ALICE. Los estudios presentados en esta tesis se engloba dentro del experimento ATLAS. Este detector es considerado multipropósito, diseñado para estudiar un amplio espectro de procesos: desde medidas de precisión del modelo estándar, a la búsqueda de nuevas partículas. El detector ATLAS se compone de diferentes subdetectores, especializado cada uno en la detección y medida de diferentes partículas.

La Figura 9.3 muestra el diseño del detector ATLAS. Las partículas se producen en el centro y pasan a través de varios subsistemas de detección. Las dimensiones de ATLAS son 46 m de longitud, 25 m de altura y un peso de aproximadamente 7000 toneladas.

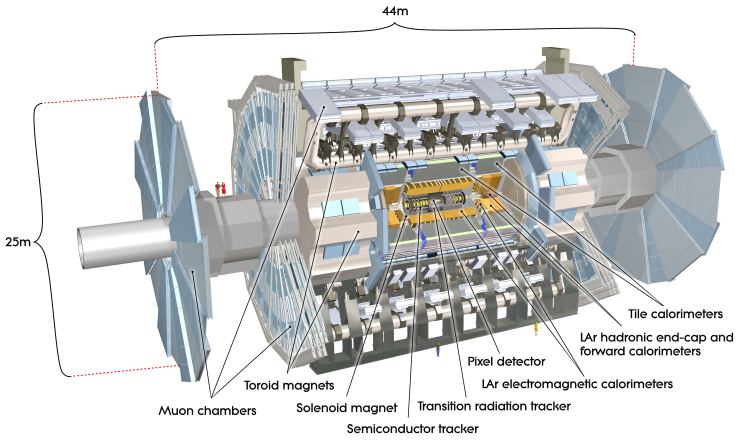


FIGURA 9.3: Diseño y dimensiones del detector ATLAS [182].

De fuera hacia adentro el detector ATLAS esta formado por los siguientes subsistemas

- Las cámaras de muones: estos detectores se especializan en la detección y a la medida de la trayectoria de muones. Cuenta con un sistema de imanes, proporcionando un intenso campo magnético toroidal, el cual curva los muones para determinar su momento.
- Los calorímetros: detectores diseñados para parar las partículas que lo atraviesan y medir su depósito de energía. El detector ATLAS está formado por dos tipos distintos de calorímetros, el calorímetro electromagnético, optimizado para medir la energía de partículas que interactúan electromagnéticamente, como electrones y fotones, y el calorímetro hadrónico, diseñado para medir la energía depositada por la produción de jets. Los jets son conos estrechos de hadrones y otras partículas producidas a partir de la hadronización de un quark o de un gluón.
- El detector interno: detector especializado para la medida y reconstrucción de las trayectorias de las partículas cargadas que lo atraviesan, también conocidas como trazas. Se encuentra inmerso en un sistema de imanes. Un solenoide lo rodea y proporciona un campo magnético uniforme de 2 T en el interior del mismo. El campo generado, curva la trayectoria de las partículas cargadas que lo atraviesan, la dirección de dicha curva, revela la carga de la partícula, y su radio de curvatura, su momento.

9.2.1 Detector ATLAS para el HL-LHC

Actualmente se espera que el LHC de alta luminosidad (HL-LHC) comience sus operaciones en la segunda mitad de 2027, con una luminosidad instantánea nominal de 7.5×10^{34} cm^{-2} s^{-1} a $\sqrt{s} = 14$ TeV de energía de centro de masa. Éste programa tiene como objetivo proporcionar una luminosidad integrada total de 3000 fb^{-1} para 2037.

Para esto, será necesario actualizar el detector ATLAS para mantener su rendimiento en el entorno de mayor luminosidad, y para mitigar el impacto del daño por radiación

y el envejecimiento del detector. La identificación de trazas se reemplazará por completo para el HL-LHC, utilizando un detector de silicio[4] completo. Diseñado con mayor granularidad, mayor lectura ancho de banda y presupuesto de material reducido [197, 198]. Entre otras cosas se mejoraran, las cámaras muónicas [202] para extender la cobertura de identificación del muon. Las mejoras también incluirán lecturas de mayor ancho de banda.

9.3 Reconstrucción de objetos

La información que recopila todos los subsistemas del detector ATLAS se deben combinar para transformar las señales en objetos físicos reconstruidos como electrones, muones, jets, energía transversal faltante y demás. Lo objetos que utilizaremos en los análisis presentes en las próximas secciones son:

- Los electrones se reconstruyen combinando un depósito de energía en el calorímetro electromagnético con una traza en el detector interno.
- Los muones se reconstruyen combinando segmentos de trazas reconstruidas al detector interno y en la cámara de muones. Se les aplica un criterio similar al aplicado a los electrones para asegurar la calidad de los muones reconstruidos.
- Los jets se reconstruyen a partir de agregados topológicos en el calorímetro hadronico. Los jets originados por quarks de tipo b se identifican mediante diferentes algoritmos que explotan las propiedades de estas partículas.
- La energía transversal faltante $\mathrm{E_T^{miss}}$ puede calcularse debido a la conservación de momento en el plano transverso al haz. La suma vectorial del momento transverso de todos los productos de una colisión debe ser cero. Un desequilibrio en el momento se denomina momento transversal faltante $\mathrm{E_T^{miss}}$, indicando la presencia de partículas que no han interaccionado con el detector ATLAS.

[4] Detector que se encuentra dento del detector interno, en la parte mas cercana al punto de colisión.

9.4 Búsqueda de materia oscura en el Run 2

Esta sección contiene la señal tj del modelo 2HDM+a. El análisis utiliza los datos recopilados por el experimento ATLAS a $\sqrt{\mathrm{s}}$ = 13 TeV, durante el Run 2 y corresponden a una luminosidad integrada de 139 fb^{-1}.

9.4.1 Características experimentales básicas

El canal tj contiene las producciones del canal-t y el canal-s para la generación de candidatos a materia oscura (ver Figura 9.1). Estos dos procesos se diferencian por los estados finales de cada uno (dado el b-jet extra producido por el canal-s) y por la sección transversal de su generación.

Se ha estudiado que las señales con valores de $\mathrm{tan}\beta$ = 0.3 ($\mathrm{tan}\beta$ = 0.5), el proceso dominante que contribuye para la sección transversal tj es el canal-t con un porcentaje de 76% (51%) para un $\mathrm{m_{H^{\pm}}}$ = 300 GeV a un 99% (99%) cuando $\mathrm{m_{H^{\pm}}}$ = 1750 GeV.

Los estudios que se han realizado, mostraron que no hay sensibilidad a la producción del canal-s. De ahora en adelante solo se muestran los estudios dedicados a la generación del canal-t de materia oscura.

9.4.2 Selección de la señal

En un colisionador de protones como el LHC se producen gran variedad de sucesos, entre los que hay que seleccionar la señal de interés. En este análisis se busca la desintegración leptónica de un quark top producido en solitario a través del canal-t y gran faltante de energía, $\mathrm{E_T^{miss}}$. Entonces, como primera estancia se requiere eventos que contengan una producción de un leptón (electrón o muón) y un jet identificado como b-jet.

Además de estos requisitos básicos, se aplica un corte en $\mathrm{E_T^{miss}}$ mayor a 200 GeV, el cual ayuda a reducir la contribución de todos los fondos. Luego, la masa transversal del sistema lepton-$\mathrm{E_T^{miss}}$ se define de la siguiente manera:

$$\mathrm{m_T}(\ell, \mathrm{E_T^{miss}}) = \sqrt{2\,\mathrm{p_T}(\ell)\,\mathrm{E_T^{miss}}\,\left(1 - \mathrm{cos}\Delta\phi(\ell, \mathrm{E_T^{miss}})\right)}, \tag{9.1}$$

donde $p_T(\ell)$ es la magnitud del momento transversal del leptón y $\Delta\phi(\ell, E_T^{miss})$ es la diferencia azimutal entre el momento del leptón la dirección de E_T^{miss}. Se requiere que esta variable sea mayor que 60 GeV para reducir la contribución de fondo. En eventos de fondo el espectro de esta cantidad disminuye rápidamente para valores superiores a la masa del bosón W. Un resumen de los criterios de preselección se presenta en la Tabla 9.1. El fondo principal en esta región es $t\bar{t}$ con 61.8 %, seguido de los W+jets con 26.1 %.

Variable	tj
$\mathscr{N}_{jets}$	$1 \leq N \leq 4$
$\mathscr{N}_{b-jets}$	$1 \leq N \leq 2$
$\mathscr{N}_{\ell}$	1
$p_T^{jet1,2,3}$ [GeV]	> 30
p_T^{jet4} [GeV]	< 50
p_T^{b-jet1} [GeV]	> 50
p_T^{b-jet2} [GeV]	> 30
$p_T(\ell)$ [GeV]	> 30
$\lvert\Delta\phi_{min}\rvert$	> 0.5
$m_T(\ell, E_T^{miss})$ [GeV]	> 60
E_T^{miss} [GeV]	> 200

TABLA 9.1: Resumen de la selección de la región de preselección para el canal-t.

Debido a la baja sección transversal de la señal tj[5] la sensibilidad de un análisis basado en cortes no es suficiente para lograr una buena discriminación de la señal. Se aplica, entonces, un enfoque basado en técnicas de análisis multivariado para mejorar la separación de la señal del fondo. Un clasificador llamado Boosted Decision Trees (BDT), entrenado específicamente para separar la señal de todos los fondos, utiliza eventos que pasan los criterios de preselección mostrados en la Tabla 9.1. La variable

[5]El canal-t y el canal-s tienen una sección transversal menor en comparación con la producción tW de esta señal, ver Ref. [129].

resultante que se obtiene al aplicar este algoritmo se usa para definir la región de señal (SR) del canal-t.

Además de los requisitos de preselección, se logra una mayor discriminación entre la señal del canal-t y los eventos de fondo mediante la aplicación de criterios adicionales. La presencia de señal se enriquece en las regiones con menor valor de $\Delta\phi(\ell, \text{b-jet 1})$ y un valor superior de $m_T(\ell, E_T^{miss})$, mientras que el fondo disminuye significativamente. Por lo tanto, se utiliza una selección que impone que los eventos cumplan la condición de tener un valor de $\Delta\phi(\ell, \text{b-jet 1})$ menor que 1.2. Los valores de $m_T(\ell, E_T^{miss})$ y E_T^{miss} se incrementan a 100 GeV y 225 GeV respectivamente. Se agrega una condición extra en los forward jets ($\mathcal{N}_{jets}^{Forward}$) para mejorar la sensibilidad a la señal canal-t[6]. Finalmente, se impone la condición de que la distribución obtenida por el clasificador BDT sea mayor a 0.6. El fondo dominante en la SR canal-t es por la generación del $t\bar{t}$ (el cual contribuye con un 83.1 % del fondo total). Otra contribución no despreciable, proviene de los eventos de W+jets (contribuye con un 4.2 %), y se puede notar en las Figuras 9.4. Dedo este hecho se han definido regiones para testear la correcta normalización de estos fondos en la SR. Un resumen de la estrategia global del análisis se muestra en la Tabla 9.2.

	canal-t
Topología de la señal	1 ℓ, 1-4 jets, 1-2 b-jets, al menos un forward jet $E_T^{miss} > 200$ GeV $m_T(\ell, E_T^{miss}) > 60$ GeV
Selecciones cinemáticas	$\Delta\phi(\ell, \text{b-jet 1}) < 1.2$, BDT
Principal variable discriminante	ajuste en la BDT
Estimación de fondos	W+jets, $t\bar{t}$

TABLA 9.2: Resumen esquemático de la estrategia global del análisis. Solo se indican la condiciones que difieren del SR. Las condiciones son representadas esquemáticamente en términos de parámetros físicos relevantes.

[6]Las señales originadas por la producción canal-t están caracterizada por tener eventos con forward jets, los cuales tienen un ángulo de menor apertura (comparado con los b-jets) con respecto a la línea de colisión.

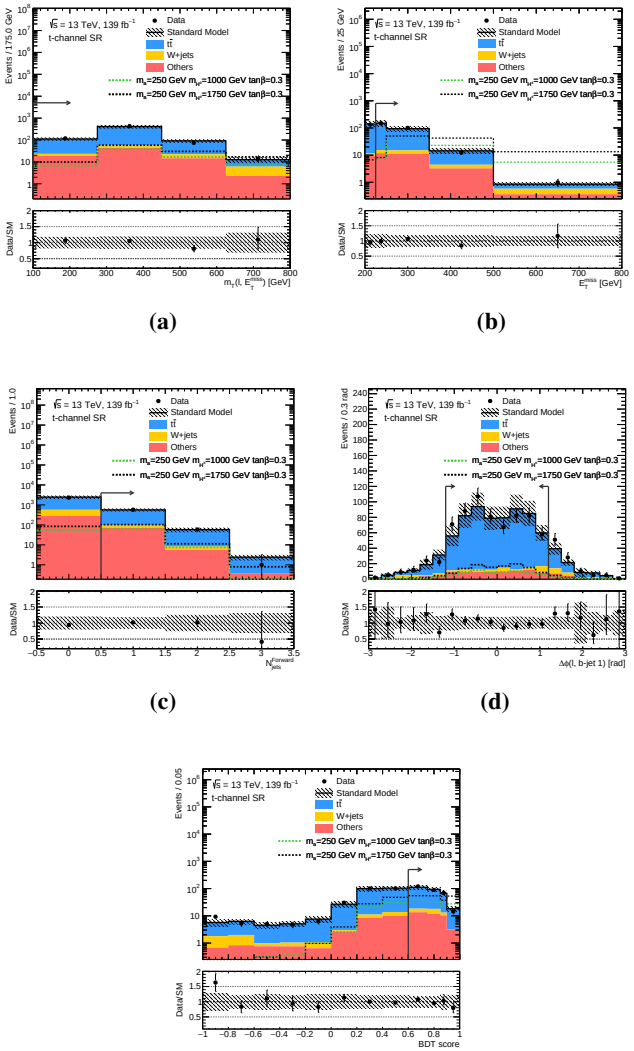


FIGURA 9.4: Distribuciones a nivel de preselección para el análisis canal-t . Las incertidumbres incluyen sistemáticos estadísticos y experimentales. El tipo de fondo llamador "Otros" (single top-quark, Diboson, $t\bar{t}V$, Z+jets , $t\bar{t}h$ y tWZ) es la combinación de todos los fondos sin incluir $t\bar{t}$ y W+jets. Los factores de normalización obtenidos del ajuste se aplican a los procesos $t\bar{t}$ y W+jets.

9.4.3 Estimación de fondos

Los procesos de fondo dominantes se estiman utilizando una técnica basada en datos extraidos por las definición de regiones de control (CR). Estas CR se definen como ortogonales a SR y son agregadas en un ajuste simultáneo con la SR para limitar la normalización de los fondos.

Los factores de normalización derivados por el ajuste usando las CR, se validan en las regiones de validación (VR). Estas regiones, son ortogonales tanto para CR como para SR, están destinadas para validar cualquier extrapolación introducida por los CR.

En la Tabla 9.3 se muestran un resumen de la estrategia de la estimación de fondos.

	canal-t
CR W+jets	$\Delta\phi(\ell, \text{b-jet } 1) > 1.5$, 1 b-jet, 1-2 jets $m_T(\ell, E_T^{miss}) > 60$ GeV
CR $t\bar{t}$	$\Delta\phi(\ell, \text{b-jet } 1) > 1.5$ 2 b-jets $m_T(\ell, E_T^{miss}) > 100$ GeV
VR W+jets	$\Delta\phi(\ell, \text{b-jet } 1) < 1.5$ $m_T(\ell, E_T^{miss}) > 60$ GeV 1 b-jet
VR $t\bar{t}$	$\Delta\phi(\ell, \text{b-jet } 1) < 1.5$ $m_T(\ell, E_T^{miss}) > 100$ GeV BDT < 0.5

TABLA 9.3: Resumen esquemático de la estrategia de análisis para la estimación de fondos. Solo se indican los requisitos que difieren del SR (o del CR en el caso de los VR). Las condiciones se representan esquemáticamente en términos de las variables físicas más relevantes.

9.4.4 Incertidumbres sistemáticas

La simulación y la reconstrucción de todos los procesos físicos utilizados para los procesos del modelo estándar y la posible producción de partículas mareria oscura están asociados con incertidumbres sistemáticas, que son de base experimental y teórica. Las incertidumbres que más afectan a éste análisis son las provenientes de la reconstrucción de los jets, y aquellas que afentan al modelado del decaimiento del quark top.

9.4.5 Resultados

Para estimar la posible presencia de señal de materia oscura se hace un ajuste de máxima verosimilitud a los datos de la distribución de BDT en la SR bajo la hipótesis de la presencia única de fondo. En el ajuste también se tiene en cuenta el número de eventos en las CR. El resultado del ajuste para un modelo representativo se muestra en la Figura 9.5, donde se ve que los datos son compatibles con una hipótesis de sólo fondo.

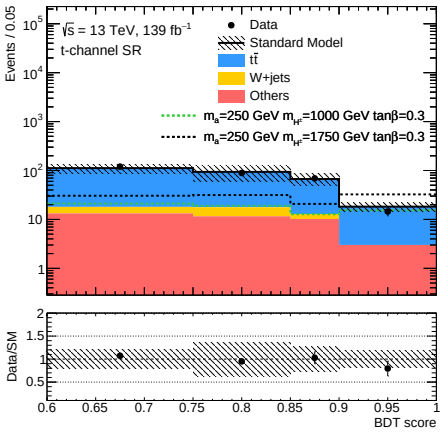


FIGURA 9.5: Distribucion de la BDT usada para extraer los límites superiores. Las incertidumbres incluyen sistemáticos estadísticos, experimentales y teóricas.

Dado que no se observa un exceso con respecto a los datos, los resultados se han utilizado para establecer un límite superior (de existencia) en la masa de los Higgs cargados para los valores de $\tan\beta$ = 0.3 y 0.5 (Figura 9.6). Las señales excluidas tienen un límite del intervalo de confianza (CL) del 95 % $\sigma/\sigma_{\mathrm{BSM}}$ por debajo o igual a 1. Por lo tanto, la región inferior de $m_{H^{\pm}}$ = 800 GeV (1100 GeV) con $\tan\beta$ = 0.3 se excluye con 95 % CL del límite observado (esperado).

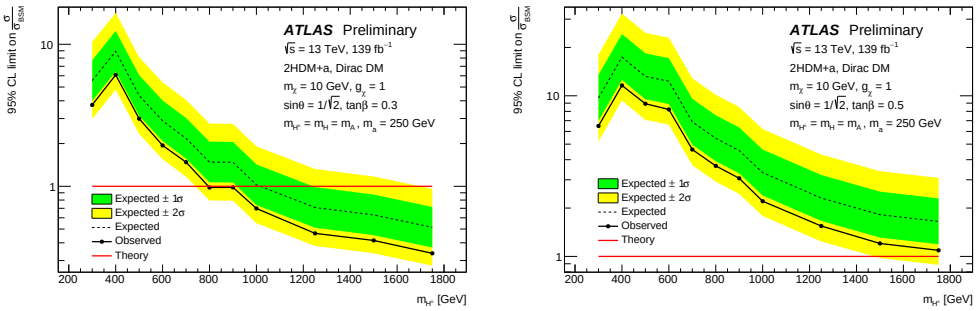


FIGURA 9.6: La línea continua (discontinua) muestra los límites superiores observados (esperados) de 95 % CLs para señales con $\tan\beta$ = 0.3 **(a)** y $\tan\beta$ = 0.5 **(b)**, ajustando en la distribución de la BDT. Las incertidumbres sistemáticas incluyen las de origen estadísticas, experimentales y teóricas.

9.5 Búsqueda de materia oscura para HL-LHC

Esta sección describe la búsqueda de materia oscura realizada por el entorno HL-LHC utilizando el experimento ATLAS a 14 TeV con una luminosidad integrada de 3000 fb^{-1}, seleccionando eventos con un quark top y una gran cantidad de energía transversal faltante $E_{\mathrm{T}}^{\mathrm{miss}}$ [326]. Esta producción ha sido previamente buscada por la colaboración CDF con datos a 1.96 TeV en el acelerador Tevatron [328], también por las colaboraciones ATLAS y CMS durante el Run 1 con 8 TeV de energía [329, 330]. Y en el Run 2 con 13 TeV correspondientes a una luminosidad integrada de 36 fb^{-1} [134, 331].

9.5.1 Selección de la señal

La señal que se busca en este caso es la producción no resonante de un proceso con estado final dado por un quark top, y gran faltante de energía $\mathrm{E_T^{miss}}$ (ver Figura 9.2). En este caso también estudiaremos el caso leptónico, es decir, que el bosón W que se produce del decaimiento del quark top, sean cosiderados solos los eventos donde el W decae a un electrón o un muón.

Dado que el proceso de señal considerado favorece los estados finales con leptones positivos, los eventos con carga negativa de leptones son rechazados. Mejor separación entre los eventos de señal y los eventos de fondo se logra aplicando criterios en la variable $\mathrm{m_T}(\ell, \mathrm{E_T^{miss}})$, la cual debe ser mayor que 100 GeV para reducir la contribución de fondos.

En los eventos de señal, el leptón y el b-jet generados están cerca uno del otro, dado a la descomposición de un quark top. Por lo tanto, se requiere que los eventos tengan una diferencia azimutal entre el momento del leptón y el momento del b-jet $\Delta\phi(\ell, \mathrm{b\text{-jet}\ 1})$, de menos que 2.0, que desfavorece los fondos W+jets y dibosones. La Tabla 9.4 muestra un resumen de los criterios anteriores que define la región de preselección.

Variable	Condición
$\mathscr{N}_\ell$	1
$\mathrm{p_T}(\ell)$ [GeV]	> 30
Signo de la carga de leptón	> 0
$\mathrm{p_T}(\mathrm{b\text{-jet}\ 1})$ [GeV]	> 30
$\mathrm{E_T^{miss}}$ [GeV]	> 100
$\mathscr{N}_{\mathrm{b-jet}}$	1
$\mathrm{m_T}(\ell, \mathrm{E_T^{miss}})$ [GeV]	> 100
$\lvert\Delta\phi(\ell, \mathrm{b\text{-jet}\ 1})\rvert$	< 2.0

TABLA 9.4: Descripción general de los criterios de selección utilizados para definir el región de preselección.

Se hace un análisis basado en técnicas de multivariado. Como en el analisis anterior, se define una BDT. La variable BDT está entrenada para discriminar las señales con respecto al fondo dominante $t\bar{t}$. Solo los eventos con valor de BDT > 0.9 y E_T^{miss} > 150 GeV ingresan en la definición de la SR.

9.5.2 Resultados

Para el análisis estadístico todos los fondos excepto los de producción de $t\bar{t}$ se fusionan en un fondo llamado non-$t\bar{t}$ para evitar problemas de falta de estadística en la SR. La forma de la distribución de la E_T^{miss} se utiliza la extracción del resultado final. Dado que se espera que sea la variable más sensible a la presencia de nueva física. Las incetidumbres que tienen mayor impacto en el ajuste final son el modelado teórico de la señal y del fondo. La segunda mayor fuente de incertidumbre es la reconstrucción de la E_T^{miss} y de los jets. La Figura 9.7 muestra la distribución de la E_T^{miss} en la SR.

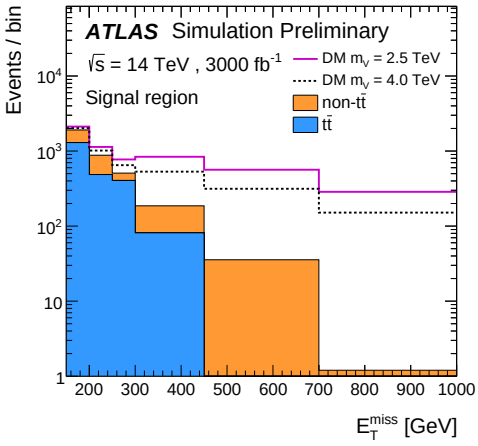


FIGURA 9.7: Distribución esperada de la E_T^{miss} en la SR. La distribución muestra las predicciones de fondo $t\bar{t}$y non-$t\bar{t}$. Las líneas continuas y discontinuas representan la señal correspondiente a una masa mediadora de 2.5 y 4.0 TeV, respectivamente. Las muestras de eventos de señal se normalizan al número de eventos de fondo.

La Figura 9.8 muestra los límites superiores esperados en función de la masa del mediador para el modelo no resonante, donde el límite superior es 4.6 TeV.

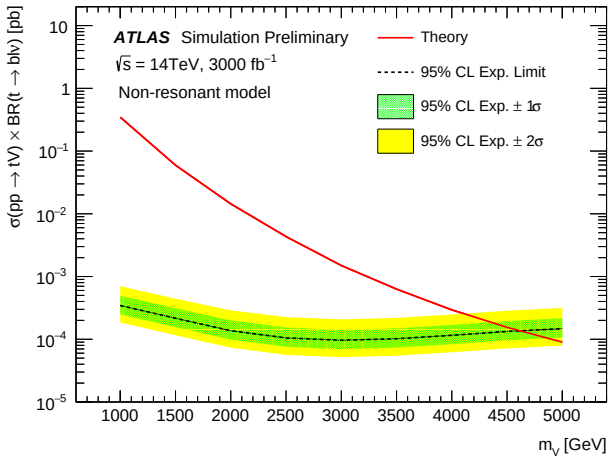


FIGURA 9.8: Límites superiores esperados en la sección transversal de la señal en función de la masa del mediador para el modelo no resonante suponiendo usando un análisis BDT. El error estadístico no se considera en las incertidumbres, sino el conjunto estimado de sistemáticos experimentales y teóricos.

APPENDIX A

The Boosted Decision Tree method

In high energy particle physics, the general technique for identifying relevant events among a high background environment involves applying a set of requirements such that non-relevant events are discarded. Selection criteria are based on variables which show differences of their distribution between the signals and backgrounds. This set of selection is chosen to maximise the signal events (our searches) while minimising the number of background events (faking final-state objects which look similar to the ones from the signal). These criteria are generally referred event selection cuts. However, with the cut-based technique some true signal candidates would be bound to fail the series of final-state object and event cuts, while some background candidates would pass. This results both in a loss in signal events and in an impure final-state event sample.

A decision tree applies a multiple-cut technique in a sophisticated way, to improve the selection of the signal candidates which otherwise would be lost and further remove background that would normally pass. These criteria are based on nodes that classify an event to be signal or background, respectively, for a set of discriminating variables, previously selected. It is a way of organising and choosing the selection cuts applied to a candidate depending on whether it passed or failed the previous cuts. A node is the decision point in the tree in which a variable and cut value are provided and the candidate is determined to either pass or fail it. The pass or failure determines which node the candidate will encounter next.

Decision tree training uses a set of known signal and background training events, each with weight w_i, to build a tree structure of cuts node by node. The classification at each node is performed on discriminating variables to find a cut that best separates signal and background events and if is determined by the so-called Gini index [313],

$$\mathrm{Gini} = \mathrm{p} \cdot (1 - \mathrm{p}),$$

where p is the purity,

$$\mathrm{p} = \frac{\sum_{\mathrm{s}} \mathrm{w_s}}{\sum_{\mathrm{s}} \mathrm{w_s} + \sum_{\mathrm{b}} \mathrm{w_b}},$$

and $\mathrm{s(b)}$ is the weighted total number of signal (background) events which landed on the node during training. Then, the Gini index for a node is 0 (i.e minimal) when the purity is either 1 or 0 (i.e. pure signal or pure background) and it is maximised when the purity is 0.5 (i.e. maximally mixed sample). The number of total trees to create (i.e. number of iterations) can be elected by the numbers of trees parameter. The training procedure then chooses the variable and cut value, that maximises the increase in the impurity function between an initial node and the sum of the daughter nodes weighted with their corresponding fraction of events (with a granularity set by the parameter number of cuts). To avoid decision trees memorising the properties of single events, the depth of the decision tree is limited to a maximum number and the presence of a minimum amount of events at each node is required. This property is usually called as the minimum node size.

However, a single decision tree would not result in any improvement on the signal and background classification since it is nothing else than a simple cut-based analysis. But the boosting process enters here in order to significantly increase the performance of this single tree [334]. In general, the boosting process uses the training results of the first tree to increase the weights of candidates that were misclassified. A new tree is then trained using these weights,

$$\mathrm{w} = \exp(\alpha) = (1 - \delta_{\mathrm{miscl}}/\delta_{\mathrm{miscl}})^{\beta},$$

where δ_{miscl} is the misclassification error and β is the learning rate assigned to all misclassified events of the decision tree and subsequently, all events are reweighted accordingly to preserve the total normalisation. Using this boosting method, many trees are then trained with new weights calculated after each retraining. In the studies presented in this thesis, boosting is applied according to the (discrete) Adaptive Boost

algorithm (AdaBoost) [335], where the classification algorithm is called Boosted Decision Tree (BDT). A Gradient Boost (GradientBoost) method is implemented, adding the functionality of an additive training tree classification method where trees are build in series (iteratively) and compared to each other based on a mathematically derived score of splits. These splits are implemented by random sub-samples of the training events for growing the trees. As the AdaBoost, the GradientBoost works best on weak classifiers, meaning small individual decision trees with a depth of often just 2 to 4 (maximum depth of tree)[1].

The final classification of an event after a successful training phase is achieved by the so-called BDT score [336],

$$\mathrm{BDT\,score} = \sum_{i=0}^{N_{Trees}} \frac{\alpha_i q_i}{\alpha_i},$$

where q_i is indicating whether the event is classified as signal or background by the *i-th* tree,

$$q_i = \begin{cases} +1 & \text{if the i-th tree classifies the event as signal,} \\ -1 & \text{if the i-th tree classifies the event as background.} \end{cases}$$

The overtraining can be detected by comparing the score distributions for the training and testing samples for both signal and background events. Another important parameter to know the goodness of the evaluation of a multivariate classification techniques is the Receiver Operating Characteristic (ROC) curve [337] which shows the background rejection against the signal efficiency. The ROC curve is created by plotting the true positive rate (signal efficiency) against the false positive rate (background rejection) at various threshold settings. An indicator of this test is the area under the ROC curve, which value represents the expected performance of an event classification algorithm (i.e. the larger the area the better the performance) [338].

The linear correlation coefficient (sometimes called Pearson's Correlation Coefficient [339]), commonly denoted as r, is a measure of the strength of the linear relationship between two variables. The value of r has the following properties:

[1] Given such small trees, they are much less prone to overtraining compared to simple decision trees.

- r value between -100 and +100[2].
- The further an r value is from zero, the stronger the relationship between the two variables.
- The sign of r indicates the nature of the relationship: A positive r indicates a positive relationship, while a negative r indicates a negative relationship.

Generally speaking, one may think of the values of r in the following manner:

- If |r| is between 85 and 100, there is a strong correlation.
- If |r| is between 50 and 85, there is a moderate correlation.
- If |r| is between 10 and 50, there is a weak correlation.
- If |r| is less than 10, there is no apparent correlation.

This coefficient is important since variables with strong correlations should not be used.

[2]Usually r goes from -1 until 1, but it is following the TMVA output [313], it is a percentage.

APPENDIX B

2HDM+a *s*-channel search using Run 2 dataset

This appendix studied the *s*-channel signal production for tj 2HDM+a generation described in Chapter 6. The *t*-channel production of this searches has an overall higher generation, as presented in Section 6.1. A detailed study to optimised *s*-channel searches is presented here. To achieve good pre-selection for *s*-channel signal, a criteria to select events containing at least 2 b-jets is required (since it is a characteristics of *s*-channel signal. See Figure 6.1(**c**)-(**d**)). This appendix is organised as follows: firstly it is defined the pre-selection in Section B.1. A multivariate method is applied, and the details are listed in Section B.2. The Section B.3 presents the background estimation of the *s*-channel analysis. The definition to optimise the signal event selection is described in Section B.4. The statistical interpretation is mentioned in Section B.5. The Section B.6 presents the main results of this appendix.

B.1 Pre-selection

The variables used to define to pre-selection region of the *s*-channel analysis are the same to *t*-channel analysis, but requiring the events containing 2 and 3 b-jets (the *t*-channel uses 1 or 2 b-jets). Table B.1 shows the pre-selection criteria for this analysis. The main background for this pre-selection is the $\mathrm{t\bar{t}}$ processes with 76.9%, followed by single top quark with 10.7%. Further variables are defined in order to understand the signal behaviour. Table B.2 shows a summary of list of selected variables. In opposite way to *t*-channel, the *s*-channel variables exploit the two b-jets selection. The mass related variables, such as $\mathrm{m(b\text{-jet}\,1, \ell)}$, $\mathrm{m(b\text{-jet}\,2, \ell)}$, $\mathrm{m(b\text{-jet}\,1, b\text{-jet}\,2)}$ and

$\mathrm{m}(\text{b-jet } 1, \ell, \mathrm{E_T^{miss}})$ are calculated from the four-momentum sum of the components shown in the parentheses.

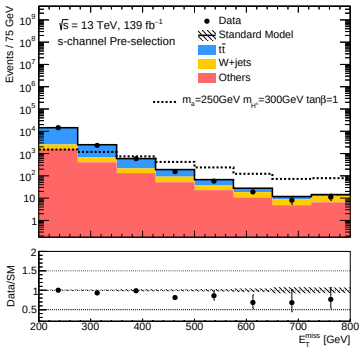
The main distributions of these variables are presented in Figures B.1-B.2, the remaining in Figures B.3-B.4.

Variable	*s*-channel
number of jets, $\mathscr{N}_{\mathrm{jets}}$	$2 \leq \mathrm{N} \leq 4$
number of b-jets, $\mathscr{N}_{\mathrm{b-jets}}$	$2 \leq \mathrm{N} \leq 3$
number of leptons, $\mathscr{N}_{\ell}$	1
$\mathrm{p_T^{jet\ 1,\ 2,\ 3,\ 4}}$ [GeV]	> 30
$\mathrm{p_T^{jet\ 4}}$ [GeV]	< 50
$\mathrm{p_T^{b\text{-}jet\ 1}}$ [GeV]	> 50
$\mathrm{p_T^{b\text{-}jet\ 2}}$ [GeV]	> 30
$\mathrm{p_T}(\ell)$ [GeV]	> 30
$\lvert\Delta\phi_{\mathrm{min}}\rvert$ [rad]	> 0.5
$\mathrm{m_T}(\ell, \mathrm{E_T^{miss}})$ [GeV]	60
$\mathrm{E_T^{miss}}$ [GeV]	200

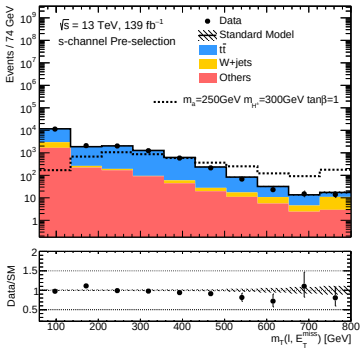
TABLE B.1: Summary of the pre-selection requirements for the *s*-channel analysis.

Variable name	**Description**
Kinematic variables	
$\mathrm{E_T^{miss}}$	Magnitude of the missing transverse momentum
$\mathrm{m_T}(\ell, \mathrm{E_T^{miss}})$	Transverse mass of lepton–$\mathrm{E_T^{miss}}$ system
$\mathrm{m}(\text{b-jet 1}, \ell)$	Mass of b-jet 1 and lepton
$\mathrm{m}(\text{b-jet 2}, \ell)$	Mass of b-jet 1 and lepton
$\mathrm{m}(\text{b-jet 1}, \mathrm{b-jet2})$	Mass of b-jet 1 and b-jet 2
$\mathrm{m}(\text{b-jet 1}, \ell, \mathrm{E_T^{miss}})$	Mass of b-jet 1, lepton and $\mathrm{E_T^{miss}}$-system
Azimuthal differences	
$\Delta\phi(\ell, \text{b-jet 1})$	$\Delta\phi$ between the lepton and b-jet 1
$\Delta\phi(\ell, \mathrm{E_T^{miss}})$	$\Delta\phi$ between the lepton and $\mathrm{E_T^{miss}}$
Angular distance differences	
$\Delta\mathrm{R}(\ell, \text{b-jet 1})$	$\Delta\mathrm{R}$ between the lepton and b-jet 1
$\Delta\mathrm{R}(\text{jet 1}, \text{jet 2})$	$\Delta\mathrm{R}$ between the the jet and jet 2

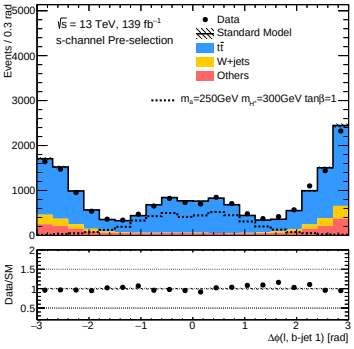
TABLE B.2: List of all the discriminating variables and their definitions used in the *s*-channel analysis.



(a)

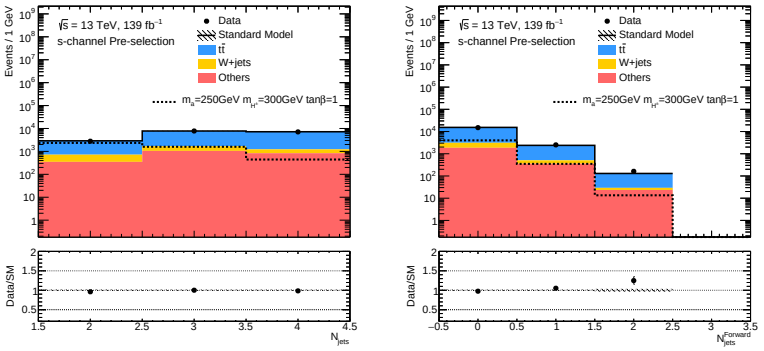


(b)

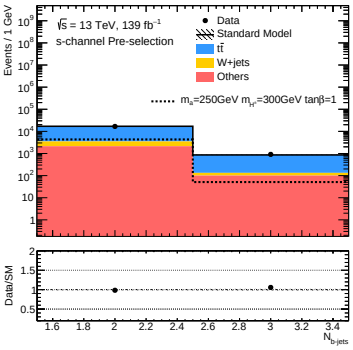


(c)

FIGURE B.1: Distributions at pre-selection-level of the *s*-channel analysis for the E_T^{miss} **(a)**, $m_T(\ell, E_T^{miss})$ **(b)** and $\Delta\phi(\ell, \text{b-jet 1})$ **(c)**. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. The SM backgrounds correspond to the simulation predictions normalised to the theoretical predictions. Signal is scaled ten times its cross-section. The last bin contains the overflow events.

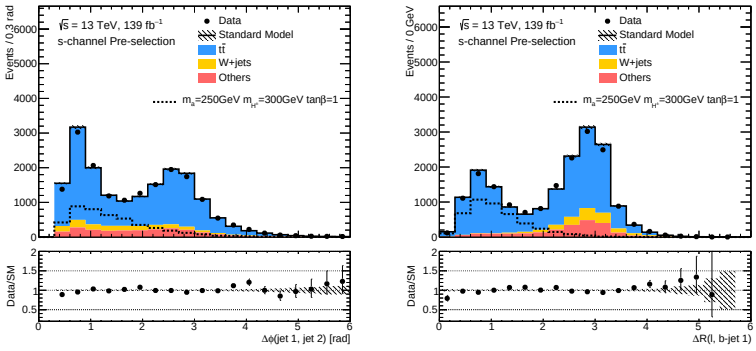


(a) **(b)**

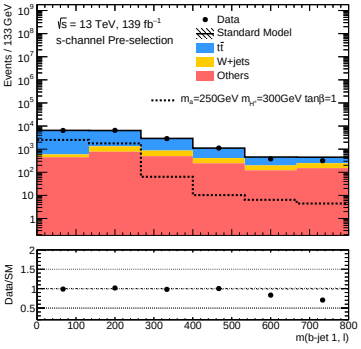


(c)

FIGURE B.2: Distributions at pre-selection-level of the *s*-channel analysis for the number of jets ($\mathscr{N}_{\text{jets}}$) **(a)**, number of forward jets ($\mathscr{N}_{\text{jets}}^{\text{Forward}}$) **(b)** and number of b-jets ($\mathscr{N}_{\text{b-jets}}$) **(c)**. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. The SM backgrounds correspond to the simulation predictions normalised to the theoretical predictions. Signal is scaled ten times its cross-section.

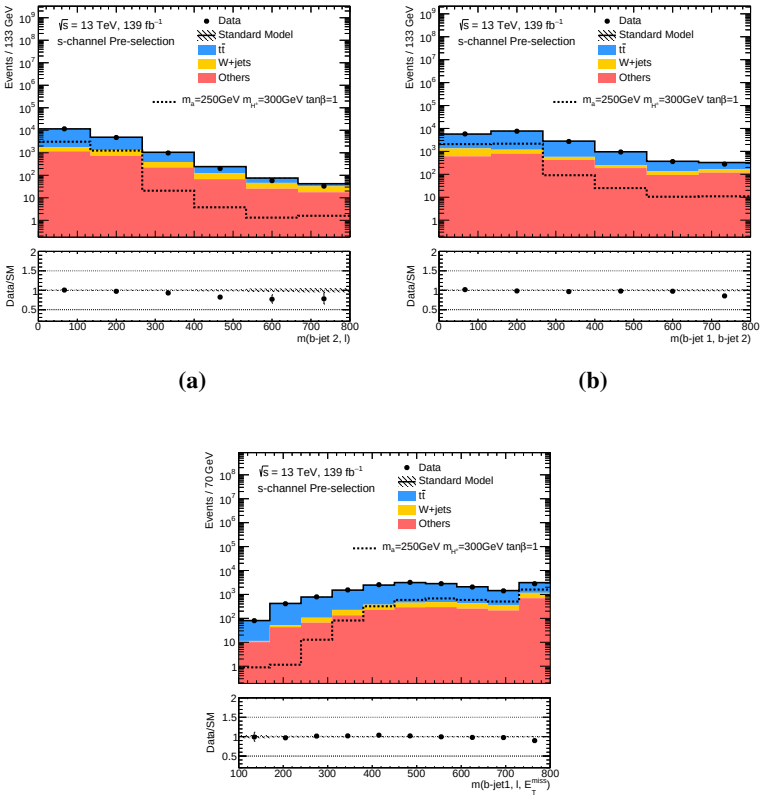


(a) **(b)**



(c)

FIGURE B.3: Distributions at pre-selection-level of the *s*-channel analysis for $\Delta\mathrm{R}(\text{jet 1}, \text{jet 2})$ **(a)**, $\Delta\mathrm{R}(\ell, \text{b-jet 1})$ **(b)** and $\mathrm{m}(\text{b-jet 1}, \ell)$ **(c)**. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $\mathrm{t\bar{t}V}$, Z+jets, $\mathrm{t\bar{t}h}$ and tWZ) is the combination of all the backgrounds not including $\mathrm{t\bar{t}}$ and W+jets. The SM backgrounds correspond to the simulation predictions normalised to the theoretical predictions. Signal is scaled ten times its cross-section. The last bin contains the overflow events.



(a) **(b)**

(c)

FIGURE B.4: Distributions at pre-selection-level of the s-channel analysis for the $\mathrm{m(b\text{-jet}\,2, \ell)}$ **(a)**, $\mathrm{m(b\text{-jet}\,1, b-jet2)}$ **(b)** and $\mathrm{m(b\text{-jet}\,1, \ell, E_T^{miss})}$ **(c)**. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $\mathrm{t\bar{t}V}$, Z+jets, $\mathrm{t\bar{t}h}$ and tWZ) is the combination of all the backgrounds not including $\mathrm{t\bar{t}}$ and W+jets. The SM backgrounds correspond to the simulation predictions normalised to the theoretical predictions. Signal is scaled ten times its cross-section. The last bin contains the overflow events.

B.2 Boosted Decision Tree

To performed the BDT-based analysis, the same idea presented in Section 6.3 is followed, but the elected set of signal are meant to have good contribution of *s*-channel production. The samples are ten, and they are:

- $m_a = 250$ GeV, $m_{H^\pm} = 300$ GeV, 400 GeV both with $\tan\beta = 1, 2, 3, 5, 10$.

The strategy is detailed in Section 6.3, based on dividing this set of events in 80% (for training) and 20% (for testing). The variables entering into the BDT training are selected to exploit the topology fact of the *s*-channel having two b-jets productions at its final state. This is opposite to the *t*-channel, which its characterisation is to have one b-jet at final state. The variables used for the training in the *s*-channel analysis are:

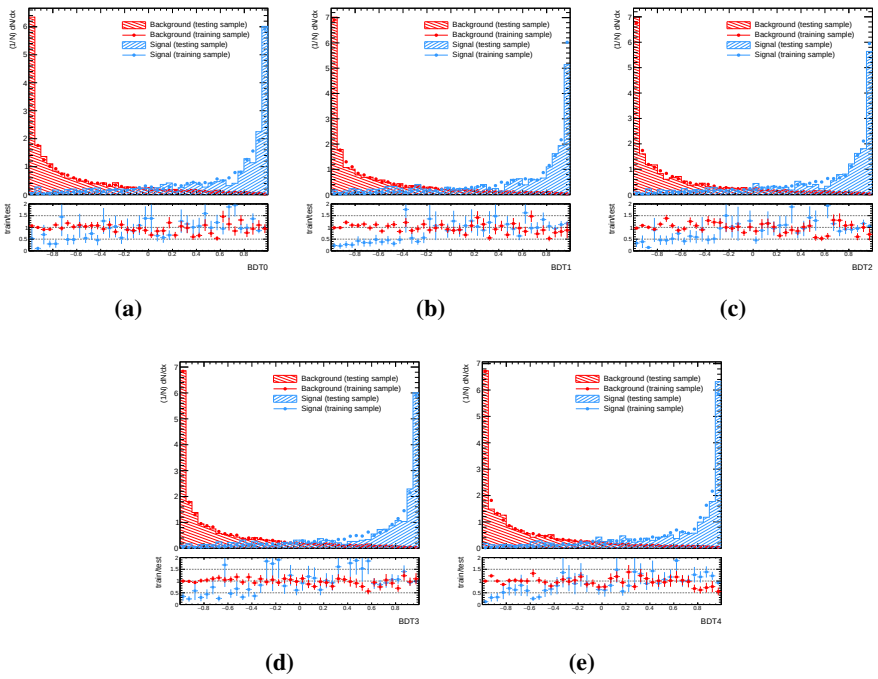
- $m(\text{b-jet } 1, \ell)$,
- $m(\text{b-jet } 2, \ell)$,
- $m(\text{b-jet } 1, \text{b-jet } 2)$,
- $m(\text{b-jet } 1, \ell, E_T^{miss})$,
- $\Delta\phi(\ell, \text{b-jet } 1)$,
- $\Delta\phi(\ell, E_T^{miss})$,
- $\Delta R(\ell, \text{b-jet } 1)$,
- $\Delta R(\text{jet } 1, \text{jet } 2)$,
- $m_T(\ell, E_T^{miss})$.

The parameters used to define the BDT are presented in Table 6.3, i.e. same configuration as the *t*-channel analysis. Figure B.5 shows the BDT distributions for training and testing samples, showing no noticeable differences between the 5 BDT obtained. The statistical fluctuation observed in these BDT score distributions are due the lower number of events (compared with *t*-channel analysis) having more than 2 or 3 b-jets (instead 1 or 2). The event selection for the *s*-channel analysis reduces the statistics about 80 %, if it is compared with the *t*-channel analysis. Figure B.6 shows the distributions of signal and background for all variables used by the BDT in the training.

The correlation coefficients between pairs of input variables are shown in Figure B.7, strong correlation is not observed. The moderate correlation seen among *s*-channel BDT definition are the same found it for the *t*-channel analysis. The origins

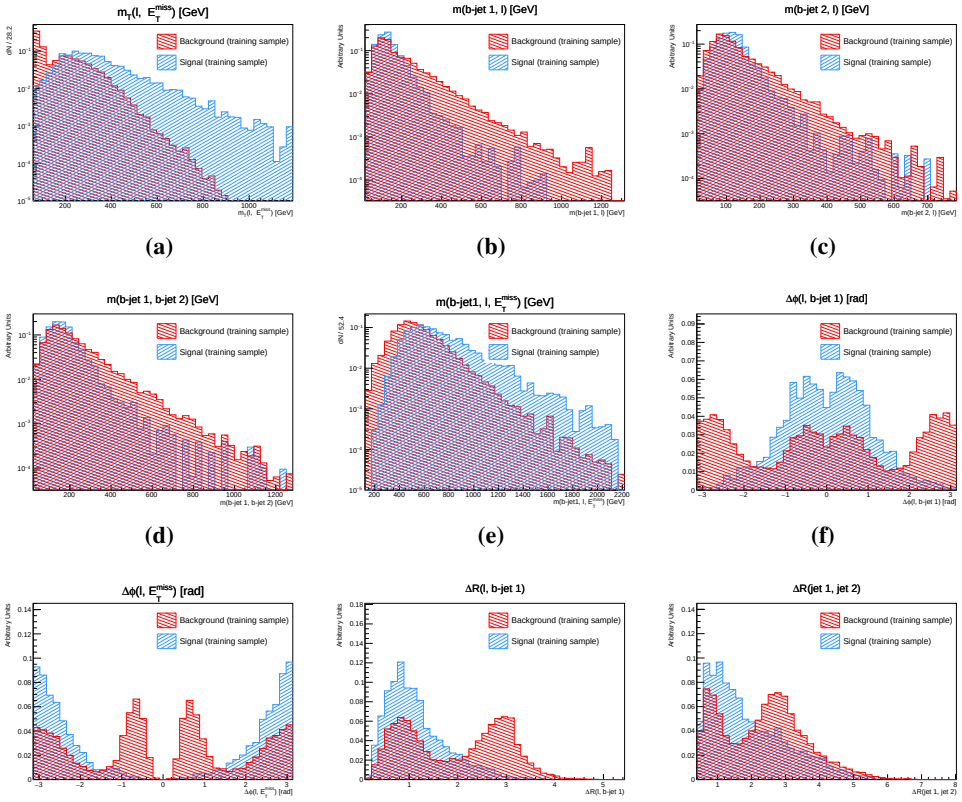
of this correlation are explained in Section 6.3. Figure B.8 presents the signal efficiency versus background rejection (i.e. ROC curve) of the final BDT score, where good performance is seen. The value of the integral of the curve is the same calculated for the t-channel analysis, it is 0.932.

As a result, the 5 BDT scores applied to data and all MC simulation samples is presented in Figure B.9, good agreement between data and MC is observed.



(a) **(b)** **(c)** **(d)** **(e)**

FIGURE B.5: Overtraining test for the five BDTs.

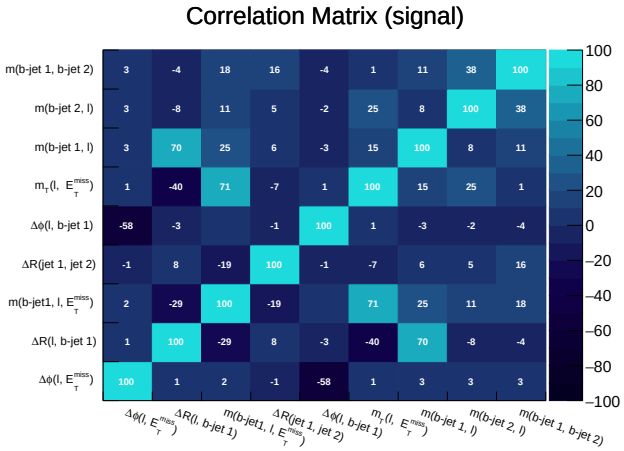


(a) **(b)** **(c)**

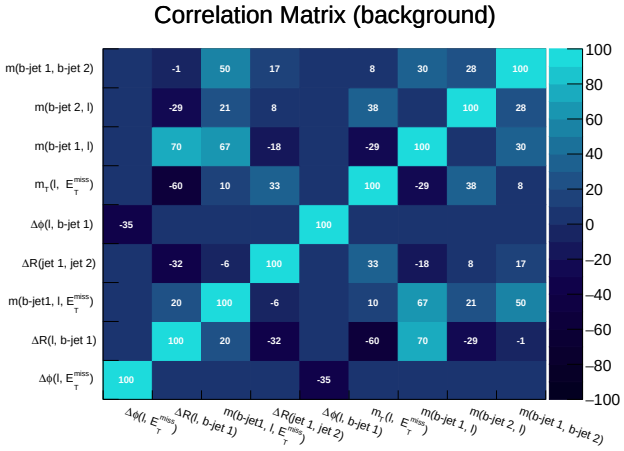
(d) **(e)** **(f)**

(g) **(h)** **(i)**

FIGURE B.6: Distribution of signal and background for all variables used by the BDT in the training. There are shown the $m_T(\ell, E_T^{miss})$ **(a)**, $m(\text{b-jet } 1, \ell)$ **(b)**, $m(\text{b-jet } 2, \ell)$ **(c)**, $m(\text{b-jet } 1, \text{b-jet } 2)$ **(d)**, $m(\text{b-jet } 1, \ell, E_T^{miss})$ **(e)**, $\Delta\phi(\ell, \text{b-jet } 1)$ **(f)**, $\Delta\phi(\ell, E_T^{miss})$ **(g)**, $\Delta R(\ell, \text{b-jet } 1)$ **(h)** and $\Delta R(\text{jet } 1, \text{jet } 2)$ **(i)**.



(a)



(b)

FIGURE B.7: Correlation coefficients for the signal **(a)** and background **(b)** processes.

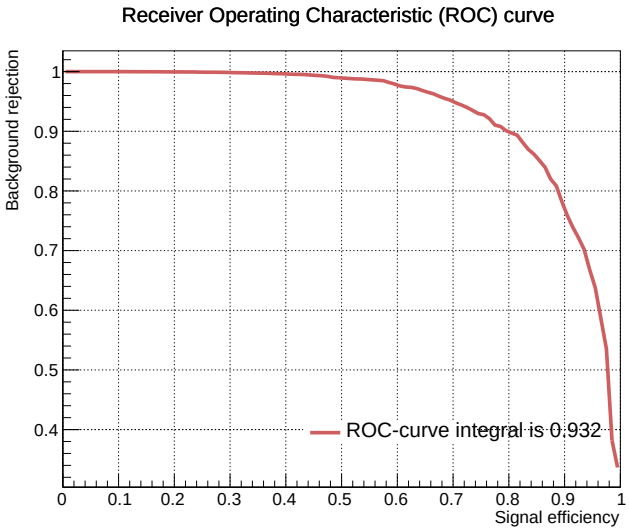


FIGURE B.8: ROC-curve, i.e. signal efficiency versus background rejection.

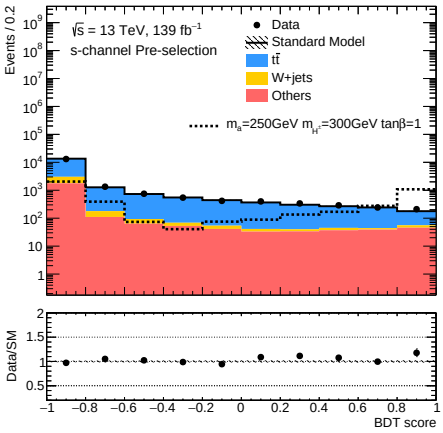


FIGURE B.9: The BDT score distribution at pre-selection-level (region used for training, it is evaluated the set of 5 BDTs together) for the *s*-channel analysis. Uncertainty band includes statistical uncertainties. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. The SM backgrounds correspond to the simulation predictions normalised to the theoretical predictions. Signal is scaled ten times its cross-section.

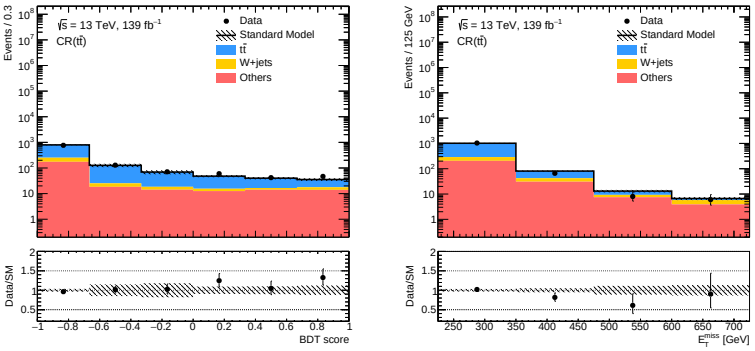
B.3 Background estimation

The same idea as the one presented in Section 6.4 is followed in the *s*-channel analysis. A data driven technique, based on CR is used. Anyhow, it is only defined for $t\bar{t}$ this time, as it is the main background contributor. As well for the VR, only it is defined for $t\bar{t}$. The definition of the CR for $t\bar{t}$ is the same applied for *t*-channel analysis, while VR $t\bar{t}$ differs with the number of b-jets (since the pre-selection has this difference) from *t*-channel. A summary of the background estimate strategy is seen in Table B.3.

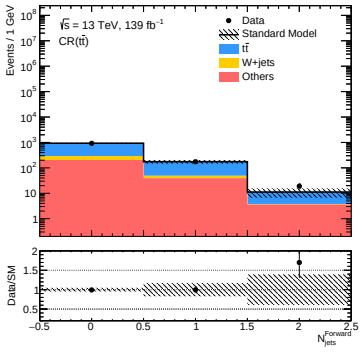
	s-channel
CR $t\bar{t}$	$\lvert\Delta\phi(\ell, \text{b-jet 1})\rvert > 1.5$ rad
	2 b-jets
VR $t\bar{t}$	$\lvert\Delta\phi(\ell, \text{b-jet 1})\rvert < 1.5$ rad
	$m_T(\ell, E_T^{miss}) > 100$ GeV
	BDT score < 0.5

TABLE B.3: Schematic summary of the analysis strategy for the main background estimate. Only the requirements that differ from the SR (or from the CR in the case of VR) are indicated. The requirements are schematically represented in terms of relevant physics parameter. The exact value of the requirement has been optimised and is described in the text of the dedicated section.

Figure B.10 shows various distributions of the $t\bar{t}$ CR, while Figure B.11 contains the $t\bar{t}$ VR studies. The $t\bar{t}$ CR and $t\bar{t}$ VR have a purity of $t\bar{t}$ about a 68.8% and 92.2%, respectively.

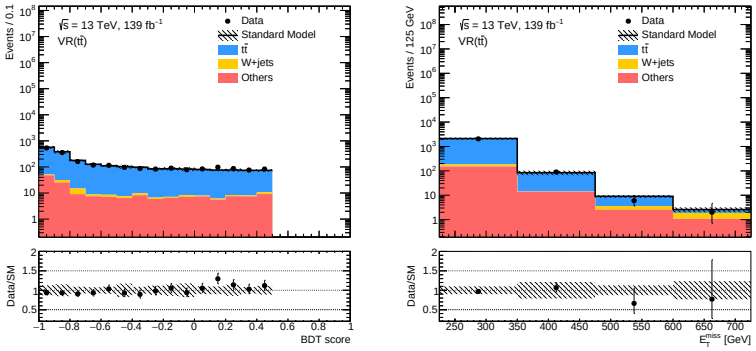


(a) **(b)**

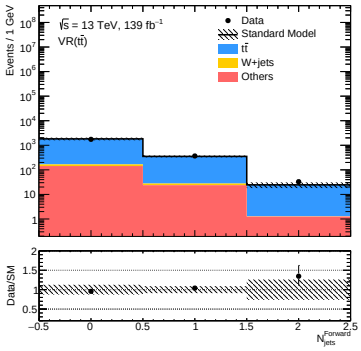


(c)

FIGURE B.10: Distributions in the $t\bar{t}$ CR for the *s*-channel analysis. The variables presented in *CR*($t\bar{t}$) are: the BDT score **(a)**, E_T^{miss} **(b)** and the number of forward jets ($\mathcal{N}_{jets}^{Forward}$) **(c)**. Uncertainty band includes statistical and experimental systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, $Z+jets$, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental systematics used are detailed in Section 6.6. The normalisation factor obtained from the background-only fit (Table B.5) is applied to the $t\bar{t}$ MC process.



(a) **(b)**

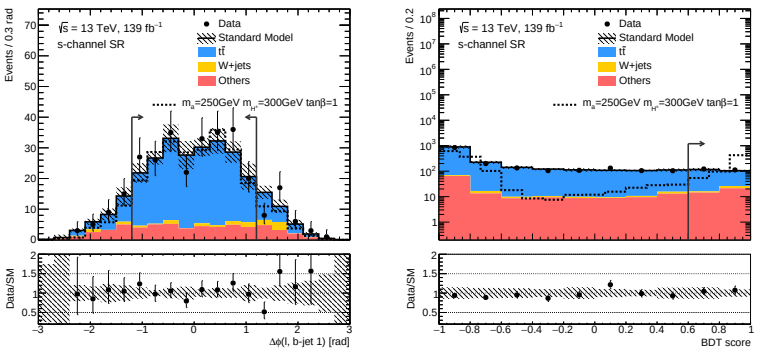


(c)

FIGURE B.11: Distributions in the $t\bar{t}$ VR for the *s*-channel analysis. The variables presented in *VR($t\bar{t}$)* are: the BDT score **(a)**, E_T^{miss} **(b)** and the number of forward jets ($\mathcal{N}_{jets}^{Forward}$) **(c)**. Uncertainty band includes statistical and experimental systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, $Z+jets$, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental systematics used are detailed in Section 6.6. The normalisation factor obtained from the background-only fit (Table B.5) is applied to the $t\bar{t}$ MC process.

B.4 Signal region

The *s*-channel final state contains no forward jets, therefore in order to improve its selection, events not containing forward jets are required in the SR. Following the same idea of the *t*-channel a selection imposing the rejection of events with $|\Delta\phi(\ell, \text{b-jet 1})| < 1.2$ rad is used. The lower threshold values on $m_T(\ell, E_T^{miss})$ and E_T^{miss} are increased to 100 GeV and 225 GeV, respectively. Further requirement on the BDT score to be larger than 0.6 is imposed. Figures B.12-B.13 show the distribution of the main variables in SR.



(a) **(b)**

FIGURE B.12: Distributions in SR for the *s*-channel analysis, there are shown $\Delta\phi(\ell, \text{b-jet 1})$ **(a)** and the BDT score **(b)**. Uncertainties include statistical and experimental systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, $Z+\text{jets}$, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental systematics used are detailed in Section 6.6. Signal is scaled five times its cross-section. The normalisation factor obtained from the background-only fit (Table B.5) is applied to the $t\bar{t}$ MC process.

The dominant background in the *s*-channel SR originates from $t\bar{t}$ production (contributes with a 76% of the total background[1]). A dedicated region, described

[1] Followed for single top-quark production, and 11.4%, the W+jets production with 5.6%.

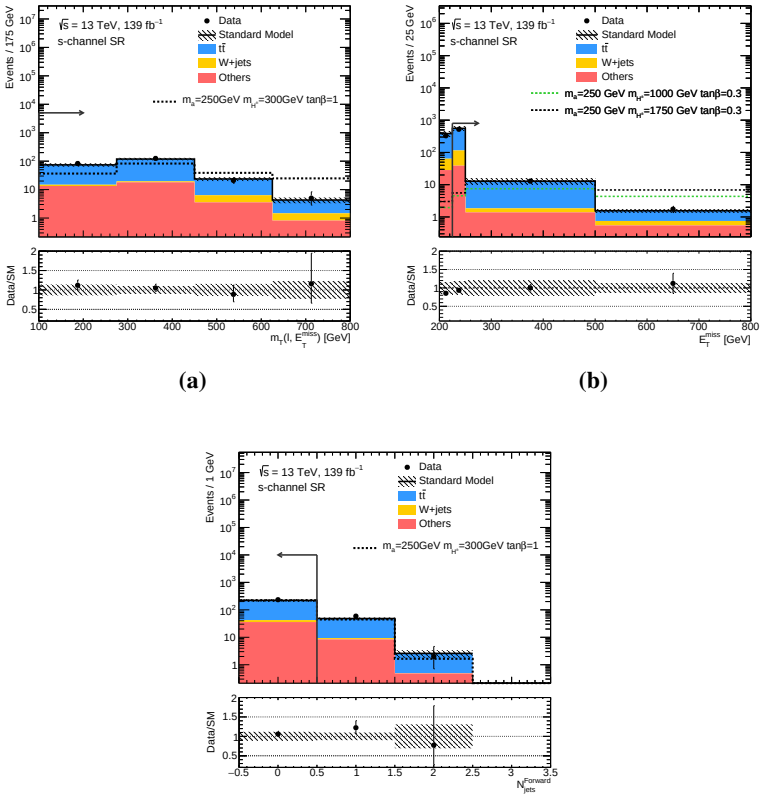


FIGURE B.13: Distributions in SR for the *s*-channel analysis, there are shown the $m_T(\ell, E_T^{miss})$ **(a)**, E_T^{miss} **(b)** and the number of forward jets ($\mathcal{N}_{jets}^{Forward}$) **(c)**. Uncertainty band includes statistical and experimental systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, $Z + jets$, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental systematics used are detailed in Section 6.6. Signal is scaled five times its cross-section. The normalisation factor obtained from the background-only fit (Table B.5) is applied to the $t\bar{t}$ MC process. For $m_T(\ell, E_T^{miss})$ and E_T^{miss} the last bin contains the overflow events.

in the previous section is defined to extract the normalisation factor of the $t\bar{t}$ contribution.

B.5 Statistical interpretation

A simultaneous fit is performed to the SR and $t\bar{t}$ CR, to constrain the main background sources and to extract exclusion limits, as discussed in Section 6.7. The hypothesis test is applied using all experimental uncertainties introduced in Section 6.6.

B.5.1 Background-only fit

The background-only fit is a profile-likelihood fit (see Equation 6.3), where only the region contributing to the likelihood is the $t\bar{t}$ CR. The outcome of running a background-only fit to data in the $t\bar{t}$ CR and $t\bar{t}$ VR is shown in Table B.4. This table presents the MC predicted yields for the background processes both before and after the background-only fit, together with the observed data. The statistical and experimental systematic uncertainties are included in the fit. The agreement observed between the post-fit MC prediction and the observed data in the $t\bar{t}$ VR shows that the extrapolation, at least in terms of the corrected $t\bar{t}$ background normalisation, is well performed.

Process	$t\bar{t}$ CR	$t\bar{t}$ VR
Observed Data	1117	2145
Post-fit Total SM	1117 ± 33	2200 ± 225
Post-fit $t\bar{t}$	771 ± 53	2005 ± 229
Post-fit W+jets	98 ± 15	28 ± 4
Post-fit Others	247 ± 27	167 ± 9
Total SM	1106 ± 46	2172 ± 113
Pre-fit $t\bar{t}$	760 ± 15	1976 ± 115
Pre-fit W+jets	98 ± 15	28 ± 4
Pre-fit Others	247 ± 27	168 ± 9

TABLE B.4: Results of the background only fit for s-channel $t\bar{t}$ CR and VR. Statistical and experimental uncertainties are considered.

The extracted normalisation value for $t\bar{t}$ background (free-floating scale factor) from a the background-only fit is presented in Table B.5.

$\mu_{t\bar{t}}$	1.0 ± 0.1

TABLE B.5: Normalisation correction factor for the $t\bar{t}$ ($\mu_{t\bar{t}}$) process acquired from the background-only fit to the CR. The uncertainties on the quoted numbers are due to the statistical and experimental systematic uncertainties implemented in the fit.

B.6 Results

A profile likelihood fit is calculated. The $t\bar{t}$ CR and the binned SR on the BDT score: [0.6, 0.75], [0.75, 0.85], [0.85, 0.9] and [0.9-1.0], are included. The event yields of the SR region and the observed data are shown in Table B.6. The statistical and experimental systematic uncertainties are included in the fit (experimental uncertainties are listed in Table 6.6). It can be observed that there is no significant deviation between the data counts in the SR as compared to the post-fit prediction of the total SM background. Figure B.14 shows the post-fit MC prediction compared to the observed data of the selected binning in the BDT score in the SR.

The effect of the uncertainties in the SR are summarised in Table B.7. The main dominant systematics in the experimental ones are the JES and JER, same behaviour shown for the *t*-channel analysis. The Figure B.15 shows the main correlation of the NPs entry in the fit, in the exclusion limit of the signal sample: m_a =250 GeV, $m_{H^\pm}$ = 300 GeV and $\tan\beta$ = 1, only it is observed one correlation with a value higher to 25%, the corresponding to the normalisation factor and the NP associated with the luminosity (-47%, it is 30% higher to the *t*-channel analysis, given the lower statistics of the *s*-channel). The Figure B.16 shows the influence of the main systematic uncertainties on the fitted value of the signal strength parameter μ_{sig} of the signal m_a = 250 GeV, $m_{H^\pm}$ = 300 GeV and $\tan\beta$ = 1. As expected the higher contributions comes from the systematic related to the jets.

As there is no significant deviation between the prediction and observed data, hypothesis tests are calculated in order to find which regions of the 2HDM+a parameter

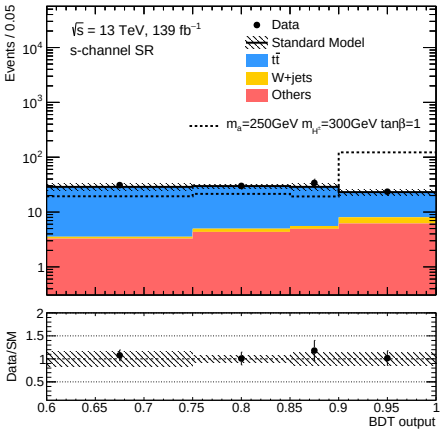


FIGURE B.14: Distribution of the BDT score in SR for *s*-channel analysis. Uncertainty band includes statistical and experimental systematics. Background type "Others" (single top-quark, Diboson, $t\bar{t}V$, Z+jets, $t\bar{t}h$ and tWZ) is the combination of all the backgrounds not including $t\bar{t}$ and W+jets. Experimental systematics used are detailed in Section 6.6. Signal is scaled five times its cross-section. The normalisation factor obtained from the background-only fit (Table B.5) is applied to the $t\bar{t}$ MC process.

Process	SR bin 0	SR bin 1	SR bin 2	SR bin 3
Observed Data	93	60	34	47
Post-fit Total SM	86 ± 13	59 ± 5	28 ± 4	46 ± 6
Post-fit $t\bar{t}$	75 ± 14	50 ± 5	23 ± 4	30 ± 4
Post-fit W+jets	0.80 ± 0.75	1 ± 0.7	0.6 ± 0.4	4 ± 3
Post-fit Others	10 ± 1	9 ± 1	5 ± 1	12 ± 1
Total SM	85 ± 9	58 ± 4	28 ± 3	45 ± 5
Pre-fit $t\bar{t}$	75 ± 9	49 ± 3	23 ± 3	30 ± 3
Pre-fit W+jets	0.80 ± 0.7	1 ± 1	0.6 ± 0.4	4 ± 2
Pre-fit Others	10 ± 1	9 ± 1	5 ± 1	12 ± 1

TABLE B.6: Results of the background only fit for *s*-channel SR. This fit includes the statistical and experimental uncertainties.

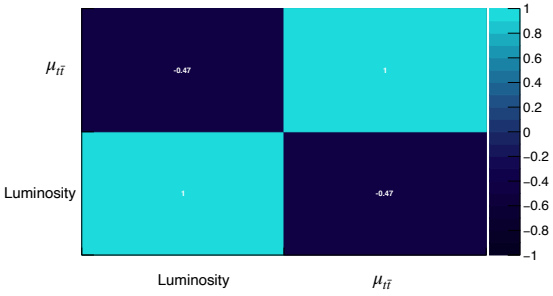


FIGURE B.15: Correlation matrix of NPs for *s*-channel analysis with data where the absolute value of the correlation is $\geq 25\%$. Evaluated for the signal sample m_a = 250 GeV, $m_{H^\pm}$ = 300 GeV and $\tan\beta$ = 1.

space can be excluded. The same procedure is implemented as in Chapter 6, calculating the CL_s values to evaluate whether a given point in the in the ($\tan\beta$, $m_{H^\pm}$) plane can be excluded at 95% CL. These hypothesis tests are performed over the tj signals, in the *s*-channel SR and the $t\bar{t}$ CR with the profile likelihood along with the observed data. The resulting hypothesis test is presented in Figure B.17, where any

Uncertainty in the binned *s*-channel SR				
Uncertainty of region	SR bin 0	SR bin 1	SR bin 2	SR bin 3
Total background expectation	86	60	29	46
Total background uncertainty	± 13	± 5	± 4	± 6
Systematic, experimental	± 9	± 3	± 3	± 5
Statistical, MC samples	± 9	± 7	± 5	± 7
Statistical, $\mu_{t\bar{t}}$ scale-factor	± 6	± 4	± 2	± 2

TABLE B.7: Dominant systematic uncertainties in the background estimates in the *s*-channel SR, expressed in terms of number of events. Individual uncertainties can be correlated, and do not necessarily add up quadratically to the total background uncertainty.

sensitivity to this production is observed. Therefore, this is the main reason behind the tj analysis does not consider an exclusive detailed search of the signals produced for *s*-channel generation in the 2HDM+a model.

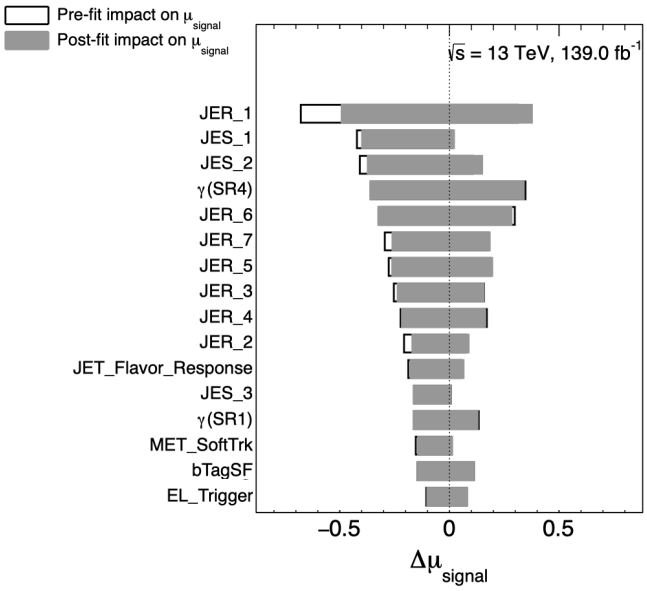


FIGURE B.16: Ordered list of the 16 NPs with the highest impact on the signal strength μ_{sig}. Evaluated for the signal sample m_a = 250 GeV, $m_{H^{\pm}}$ = 300 GeV and $\tan\beta$ = 1. The fitted value of the signal strength parameter μ (top axis values), before the systematics are fitted to the data (empty rectangles) and after (full rectangles). For each NP the shift in the signal strength is obtained for a fit with the NP fixed to its pre-fit or post-fit ($\Delta\mu_{\mathrm{signal}}$) values, respectively. The suffix is the decomposition of the respective uncertainty in subgroups.

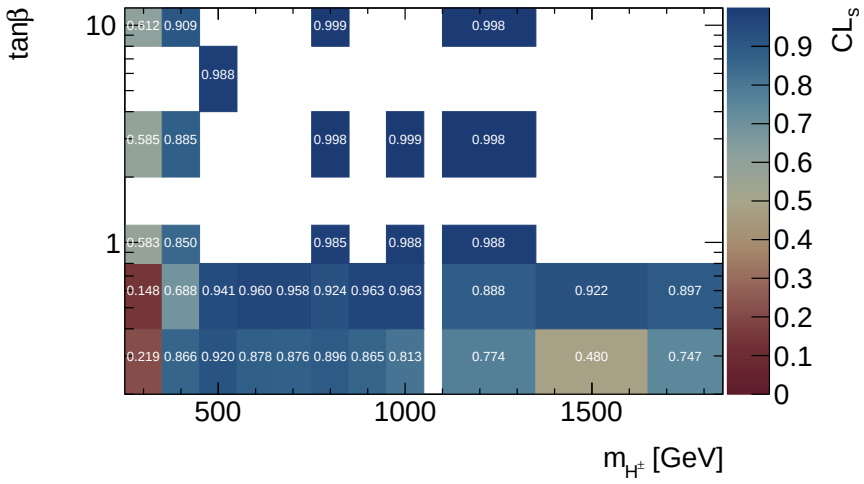


FIGURE B.17: The 95% CLs upper limits (expected) among tj signals for *s*-channel analysis. The fit uses the four bins in the BDT score. Statistical and experimental systematic uncertainties are included in the fit. Signals excluded show a value of $CL_s < 0.05$. It concludes no sensitivity for *s*-channel production along tj signals.

ACKNOWLEDGEMENTS

La parte más difícil de escribir ha llegado, agradecer a todas las personas que me han ayudado e incentivado durante estos últimos años, no parece una tarea fácil. Para empezar este trabajo ha sido posible gracias a Jose, el cual me ha sabido guiar y enseñar durante el transcurso del doctorado, siempre cuidando de mí. También, Carlos ha sido otra persona la cual hace que esta tesis sea cual es. Estoy profundamente agradecida por su ayuda en la creación de este trabajo, sin duda no sería lo mismo sin sus consejos y guía.

I am really grateful to all people that I have the lucky to work with them. Offering me to learn the collaborative way to work. This would not have been possible without the many people who have helped me along the way. Their number is many, too far to count or list here. I hope they know who they are, though, and without them I would not have succeeded to do this work.

Quiero hacer una mención especial a Josep Navarro, realmente este trabajo no hubiera terminado en el tiempo que lo hizo sin su ayuda en los últimos meses de escritura, como se lo he dicho varias veces, me ha salvado. También quiero agradecer a Oscar Estrada, el cual me ha ayudado apenas al llegar a Valencia, y uno de mis amigos formados en esta etapa.

Mudarme tan lejos de mi familia y amigos fue muy difícil, muchas veces la distancia me ha pasado factura. Durante ese tiempo no hubiese sido posible para mí superarlo sin tener amigos en los que apoyarme. Por eso quiero agradecer en especial a tres personas que han estado para mí durante este tiempo: Julián Alcaide, Jesús Guerrero y Martín Perelló. Sin ustedes estos años la cuesta arriba hubiese sido más empinada. Aunque no fueron los únicos he decir, de verdad no puedo estar más que feliz con el ambiente que encontré en Valencia. Los buenos ratos con el "Escuadrón IFIC", las cervezas, las risas, las anécdotas compartidas con muchas personas, las cuales espero que se sientan debidamente mencionadas por estas líneas.

Finalmente quiero agradecer a mi familia por el apoyo estos años y siempre. A mi madre, que siempre está para hablar, sin importar la hora que sea. Cuando solo veo el suelo, sos quien levanta mi cabeza, sin vos realmente no puedo. A mi padre, por

su amor y su confianza incondicional en mí. A mi hermana por cuidar de los viejos, y por sus locuras, las cuales siempre te arranca una sonrisa. A mi abuela, que es lo parecido al amor que conozco y uno de los pilares en lo que tengo para apoyarme, siempre alentándome. A mi tía, por quererme como si fuera una hija más para ella. A mi prima, por su apoyo y aliento durante estos años.

Qué más decir, realmente han sido años fantásticos, jamás me imaginé que podría llegar a vivir lo que viví durante ellos, muchas gracias a todas y todos por los tiempos compartidos! Solamente resta decir una última cosa para terminar esta tesis: *Fi de trajecte*.

BIBLIOGRAPHY

[1] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22**.4 (1961) 579–588.

[2] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[3] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367–377.

[4] M. Srednicki, *Quantum Field Theory*, Cambridge Univ. Press (2007).

[5] G. D. Rochester and C. C. Butler, *Evidence for the Existence of New Unstable Elementary Particles*, Nature **160** (1947) 855–857.

[6] C. F. Powell and G. P. S. Occhialini, *Nuclear physics in photographs: tracks of charged particles in photographic emulsions*, Clarendon Press (1947).

[7] O. Chamberlain et al., *Observation of Antiprotons*, Phys. Rev. **100**.3 (1955) 947–950.

[8] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**.9 (1964) 321–323.

[9] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**.16 (1964) 508–509.

[10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**.20 (1964) 585–587.

[11] E. Noether, *Invariant variation problems*, Transport Theory and Statistical Physics **1**.3 (1971) 186–207.

[12] A. Pich, *The Standard Model of Electroweak Interactions.* (2008), arXiv: 0705.4264 [hep-ph].

[13] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**.08 (2008) S08003.

[14] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3**.08 (2008) S08004.

[15] 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. **B716** (2012) 1–29, arXiv: 1207.7214 [hep-ex].

[16] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. **B716** (2012) 30–61, arXiv: 1207.7235 [hep-ex].

[17] M. Tanabashi et al., *Review Particle Data Group*, Phys. Rev **D98**.3 (2018) 030001.

[18] T. Nakano and K. Nishijima, *Charge Independence for V-particles**, Progress of Theoretical Physics **10**.5 (1953) 581–582.

[19] G. Arnison et al., *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV*, Phys. Lett. **B122** (1983) 103–116.

[20] M. Banner et al., *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, Phys. Lett. **B122** (1983) 476–485.

[21] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10**.12 (1963) 531–533.

[22] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[23] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, Helvetica Physica Acta **6** (1933) 110–127.

[24] K. G. Begeman, A. H. Broeils, and R. H. Sanders, *Extended rotation curves of spiral galaxies: dark haloes and modified dynamics*, Monthly Notices of the Royal Astronomical Society **249**.3 (1991) 523.

[25] G. Bertone and T. M. P. Tait, *A new era in the search for dark matter*, Nature **562**.7725 (2018) 51–56, arXiv: 1810.01668 [astro-ph.CO].

[26] Planck Collaboration, *Planck 2015 results - XIII. Cosmological parameters*, Astron. Astrophys. **594** (2016) A13.

[27] A. G. Riess, *The Case for an Accelerating Universe from Supernovae*, Publications of the Astronomical Society of the Pacific **112**.776 (2000) 1284–1299.

[28] A. D. Sakharov, *Violation of CP in variance,Casymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34**.5 (1991) 392–393.

[29] J. E. Kim and G. Carosi, *Axions and the strong* CP *problem*, Rev. Mod. Phys. **82**.1 (2010) 557–601.

[30] P. G. Harris et al., *New Experimental Limit on the Electric Dipole Moment of the Neutron*, Phys. Rev. Lett. **82**.5 (1999) 904–907.

[31] R. D. Peccei and H. R. Quinn, CP *Conservation in the Presence of Pseudoparticles*, Phys. Rev. Lett. **38**.25 (1977) 1440–1443.

[32] P. F. de Salas et al., *Status of neutrino oscillations 2018: 3σ hint for normal mass ordering and improved CP sensitivity*, Phys. Lett. **B782** (2018) 633–640, arXiv: 1708.01186 [hep-ph].

[33] M. Ibe, A. Kusenko, and T. T. Yanagida, *Why three generations?*, Phys. Lett. **B758** (2016) 365 –369.

[34] E. Witten, *An SU(2) anomaly*, Phys. Lett. **B117**.5 (1982) 324 –328.

[35] B. A. Dobrescu and E. Poppitz, *Number of Fermion Generations Derived from Anomaly Cancellation*, Phys. Rev. Lett. **87** (3 2001) 031801.

[36] I. Hinchliffe and T. Kaeding, *B- and L-violating couplings in the minimal supersymmetric standard model*, Phys. Rev. **D47**.1 (1993) 279–284.

[37] J. L. Evans et al., *Nonanomalous Discrete* R *Symmetry Decrees Three Generations*, Phys. Rev. Lett. **109** (18 2012) 181801.

[38] R. N. Mohapatra and M. Ratz, *Gauged discrete symmetries and proton stability*, Phys. Rev. **D76**.9 (2007) 095003.

[39] E. Sheldon, *Colliders and Neutrinos: The Window into Physics beyond the Standard Model*, Contemporary Physics **50**.6 (2009) 671–673.

[40] G't. Hooft et al., *Recent Developments in Gauge Theories.*, NATO Sci. Ser. **B59** (1980) 1–438.

[41] H.-C. Cheng and I. Low, *TeV symmetry and the little hierarchy problem*, JHEP **09** (2003) 051, arXiv: hep-ph/0308199.

[42] S. P. Martin, *A Supersymmetry primer*, Adv. Ser. Direct. High Energy Phys. (1998) 1–98, arXiv: hep-ph/9709356.

[43] W. De Boer, *Grand unified theories and supersymmetry in particle physics and cosmology*, Progress in Particle and Nuclear Physics **33** (1994) 201–301.

[44] H. D. Politzer, *Asymptotic freedom: An approach to strong interactions*, Physics Reports **14**.4 (1974) 129 –180.

[45] G. Bertone and D. Hooper, *History of dark matter*, Rev. Mod. Phys. **90**.4 (2018) 045002, arXiv: 1605.04909 [astro-ph.CO].

[46] J. de Swart, G. Bertone, and J. van Dongen, *How Dark Matter Came to Matter*, Nature Astron. **1** (2017) 0059, arXiv: 1703.00013 [astro-ph.CO].

[47] J. Silk et al., *Particle Dark Matter: Observations, Models and Searches*, Cambridge Univ. Pres. (2010).

[48] M. Srednicki, R. Watkins, and K. A. Olive, *Calculations of Relic Densities in the Early Universe*, Nucl. Phys. **B310**.3-4 (1988) 693–713.

[49] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, Phys. Rept. **405** (2005) 279–390, arXiv: hep-ph/0404175.

[50] J. L. Feng, *Dark Matter Candidates from Particle Physics and Methods of Detection*, Ann. Rev. Astron. Astrophys. **48** (2010) 495–545, arXiv: 1003.0904 [astro-ph.CO].

[51] A. Bosma, *21-cm line studies of spiral galaxies. 2. The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types.*, Astron. J. **86** (1981) 1825.

[52] V. C. Rubin and W. K. Jr. Ford, *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, Astrophys. J. **159** (1970) 379–403.

[53] V. C. Rubin, W. K. Jr. Ford, and Norbert T., *Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa through Sc*, Astrophys. J. Lett. **225** (1978) L107–L111.

[54] A. N. Taylor et al., *Gravitational Lens Magnification and the Mass of Abell 1689*, Astrophys. J. **501**.2 (1998) 539–553.

[55] H. Hoekstra, H.K.C Yee, and M. D. Gladders, *Current status of weak gravitational lensing*, New Astronomy Reviews **46**.12 (2002) 767 –781.

[56] M. Azzaro, F. Prada, and C. M. Gutierrez, *Motion properties of satellites around external spiral galaxies*, 2004, arXiv: astro-ph/0310487.

[57] M. Milgrom, *A Modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis*, Astrophys. J. **270** (1983) 365–370.

[58] J. W. Moffat, *Scalar-tensor-vector gravity theory*, JCAP **0603** (2006) 004, arXiv: gr-qc/0506021.

[59] E. P. Verlinde, *Emergent Gravity and the Dark Universe*, SciPost Phys. **2**.3 (2017) 016.

[60] G. Bertone and T. Tait, *A new era in the search for dark matter*, Nature **562** (2018) 51–56.

[61] LIGO Scientific Collaboration and Virgo Collaboration, *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett. **116**.6 (2016) 061102, arXiv: 1602.03837 [gr-qc].

[62] S. Boran et al., *GW170817 Falsifies Dark Matter Emulators*, Phys. Rev. **D97**.4 (2018) 041501, arXiv: 1710.06168 [astro-ph.HE].

[63] J. Sakstein and B. Jain, *Implications of the Neutron Star Merger GW170817 for Cosmological Scalar-Tensor Theories*, Phys. Rev. Lett. **119**.25 (2017) 251303.

[64] H. Wang et al., *The GW170817/GRB 170817A/AT 2017gfo Association: Some Implications for Physics and Astrophysics*, Astrophys. J. **851**.1 (2017) L18.

[65] J. D. Bekenstein, *Relativistic gravitation theory for the modified Newtonian dynamics paradigm*, Phys. Rev. **D70**.8 (2004) 083509.

[66] B. P. Abbott et al., *GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral*, Phys. Rev. Lett. **119**.16 (2017) 161101, arXiv: 1710.05832 [gr-qc].

[67] J. M. Overduin and P. S. Wesson, *Dark matter and background light*, Phys. Rept. **402** (2004) 267–406, arXiv: astro-ph/0407207.

[68] S. Palomares-Ruiz, *Model-Independent Bound on the Dark Matter Lifetime*, Phys. Lett. **B665** (2008) 50–53, arXiv: 0712.1937 [astro-ph].

[69] D. Clowe et al., *A direct empirical proof of the existence of dark matter*, Astrophys. J. **648** (2006) L109–L113, arXiv: astro-ph/0608407.

[70] G. B. Gelmini, *The Hunt for Dark Matter*, arXiv: 1502.01320 [hep-ph].

[71] E. W. Kolb and M. S. Turner, *The Early Universe*, Front. Phys. **69** (1990) 1–547.

[72] MACHO Collaboration, *The MACHO project dark matter search*, ASP Conf. Ser. **88** (1996) 95, arXiv: astro-ph/9510104.

[73] B. Carr, F. Kuhnel, and M. Sandstad, *Primordial Black Holes as Dark Matter*, Phys. Rev. **D94**.8 (2016) 083504, arXiv: 1607.06077 [astro-ph.CO].

[74] T. D. Brandt, *Constraints on MACHO Dark Matter from Compact Stellar Systems in Ultra-Faint Dwarf Galaxies*, The Astrophysical Journal **824**.2 (2016) L31.

[75] S. K. Kang, *Roles of sterile neutrinos in particle physics and cosmology*, Int. J. Mod. Phys. **A34**.10 (2019) 1930005, arXiv: 1904.07108 [hep-ph].

[76] M. Drewes et al., *A White Paper on keV Sterile Neutrino Dark Matter*, JCAP **1701** (2017) 025, arXiv: 1602.04816 [hep-ph].

[77] N. Du et al., *Search for Invisible Axion Dark Matter with the Axion Dark Matter Experiment*, Phys. Rev. Lett. **120** (15 2018) 151301.

[78] A. de Gouvea, D. Hernandez, and T. M. P. Tait, *Criteria for Natural Hierarchies*, Phys. Rev. **D89**.11 (2014) 115005, arXiv: 1402.2658 [hep-ph].

[79] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, *Electroweak symmetry breaking from dimensional deconstruction*, Phys. Lett. **B513** (2001) 232–240, arXiv: hep-ph/0105239.

[80] N. Arkani-Hamed et al., *The Minimal Moose for a Little Higgs*, JHEP **0208**.08 (2002) 021.

[81] E. W. Kolb, D. J. H. Chung, and A. Riotto, *WIMPzillas!*, AIP Conf. Proc. **484**.1 (1999) 91–105, arXiv: hep-ph/9810361.

[82] L. Berezhiani and J. Khoury, *Theory of dark matter superfluidity*, Phys. Rev. **D92**.10 (2015) 103510.

[83] O. F. Piattella, UNITEXT for Physics, Cham: Springer, 2018, arXiv: 1803.00070 [astro-ph.CO].

[84] M. Drees, F. Hajkarim, and E. R. Schmitz, *The Effects of QCD Equation of State on the Relic Density of WIMP Dark Matter*, JCAP **06** (2015) 025, arXiv: 1503.03513 [hep-ph].

[85] D. Hooper, *Theoretical Advanced Study Institute in Elementary Particle Physics: The Dawn of the LHC Era* (2010) 709–764, arXiv: 0901.4090 [hep-ph].

[86] X.-J. Bi, P.-F. Yin, and Q. Yuan, *Status of Dark Matter Detection*, Front. Phys.(Beijing) **8** (2013) 794–827, arXiv: 1409.4590 [hep-ph].

[87] C. P. de los Heros, *Status of direct and indirect dark matter searches*, 2020, arXiv: 2001.06193 [astro-ph.HE].

[88] J. Li and A. W. Thomas, *Bottom quark contribution to spin-dependent dark matter detection*, Nuclear Physics **B906** (2016) 60 –76.

[89] D. G. Cerdeno and A. M Green, *Direct detection of WIMPs*, Cambridge University Press (2010), arXiv: 1002.1912 [astro-ph.CO].

[90] R. Agnese et al., *New Results from the Search for Low-Mass Weakly Interacting Massive Particles with the CDMS Low Ionization Threshold Experiment*, Phys. Rev. Lett. **116**.7 (2016) 071301.

[91] CRESST Collaboration, *Description of CRESST-II data*, 2017, arXiv: 1701.08157 [physics.ins-det].

[92] E. Aprile et al., *Physics reach of the XENON1T dark matter experiment.*, JCAP **2016**.04 (2016) 027–027.

[93] D. S. Akerib et al., *Improved Limits on Scattering of Weakly Interacting Massive Particles from Reanalysis of 2013 LUX Data*, Phys. Rev. Lett. **116**.16 (2016) 161301, arXiv: 1512.03506 [astro-ph.CO].

[94] P. Agnes et al., *Results From the First Use of Low Radioactivity Argon in a Dark Matter Search*, Phys. Rev. **D93**.8 (2016) 081101, arXiv: 1510.00702 [astro-ph.CO].

[95] R. E. Apfel, *The superheated drop detector*, Nucl. Instrum. Meth. **162**.1 (1979) 603–608.

[96] PICO Collaboration, *Dark matter search results from the complete exposure of the PICO-60* C_3F_8 *bubble chamber*, Phys. Rev. **D100**.2 (2019) 022001.

[97] G. Arcadi et al., *The waning of the WIMP? A review of models, searches, and constraints*, Eur. Phys. J. **C78**.3 (2018) 203.

[98] J. D. Sloane et al., *Assessing Astrophysical Uncertainties in Direct Detection with Galaxy Simulations*, Astrophys. J. **831** (2016) 93, arXiv: 1601.05402 [astro-ph.GA].

[99] M. Pato et al., *Systematic uncertainties in the determination of the local dark matter density*, Phys. Rev. **D82**.2 (2010).

[100] C. Moni Bidin et al., *On the local dark matter density*, Astronomy and Astrophysics **573** (2014) A91.

[101] H. Silverwood et al., *A non-parametric method for measuring the local dark matter density*, Mon. Not. Roy. Astron. Soc. **459**.4 (2016) 4191–4208, arXiv: 1507.08581 [astro-ph.GA].

[102] C. Kelso et al., *The impact of baryons on the direct detection of dark matter*, JCAP **1608** (2016) 071, arXiv: 1601.04725 [astro-ph.GA].

[103] J. M. Gaskins, *A review of indirect searches for particle dark matter*, Contemp. Phys. **57**.4 (2016) 496–525, arXiv: 1604.00014 [astro-ph.HE].

[104] IceCube Collaboration, *The IceCube data acquisition system: Signal capture, digitization, and timestamping*, Nucl. Instrum. Meth. **A601**.3 (2009) 294 – 316.

[105] E. V. Korolkova, *The Status of the ANTARES experiment*, Nucl. Phys. B Proc. Suppl. **136** (2004) 69–76, ed. by C. Aramo, A. Insolia, and C. Tuve, arXiv: 0408239.

[106] M. G. Aartsen et al., *Search for annihilating dark matter in the Sun with 3 years of IceCube data*, Eur. Phys. J. **C77**.3 (2017) 146, arXiv: 1612.05949 [astro-ph.HE].

[107] AMS Collaboration, *High Statistics Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station*, Phys. Rev. Lett. **113**.12 (2014) 121101.

[108] AMS Collaboration, *Precision Measurement of the Proton Flux in Primary Cosmic Rays from Rigidity 1 GV to 1.8 TV with the Alpha Magnetic Spectrometer on the International Space Station*, Phys. Rev. Lett. **114**.17 (2015) 171103.

[109] M. Boudaud et al., *A new look at the cosmic ray positron fraction*, Astron. Astrophys. **575** (2015) A67, arXiv: 1410.3799 [astro-ph.HE].

[110] M. Ibe et al., *Wino dark matter in light of the AMS-02 2015 data*, Phys. Rev. **D91**.11 (2015) 111701.

[111] B. Penning, *The pursuit of dark matter at colliders—an overview*, J. Phys. G. Nucl. Partic. **45**.6 (2018) 063001.

[112] J. R. Ellis et al., *Dark matter in the light of LEP*, Phys. Lett. **B245** (1990) 251–257.

[113] P. J. Fox et al., *LEP shines light on dark matter*, Phys. Rev. **D84**.1 (2011).

[114] Y. Bai, P. J. Fox, and R. Harnik, *The Tevatron at the Frontier of Dark Matter Direct Detection*, JHEP **12** (2010) 048, arXiv: 1005.3797 [hep-ph].

[115] O. Buchmueller, C. Doglioni, and L. T. Wang, *Search for dark matter at colliders*, Nature Phys. **13**.3 (2017) 217–223, arXiv: 1912.12739 [hep-ex].

[116] N. F. Bell et al., *Dark matter at the LHC: Effective field theories and gauge invariance*, Phys. Rev. **D92** (5 2015) 053008.

[117] Y. Bai and T. M.P. Tait, *Searches with mono-leptons*, Phys. Lett. **B723**.4 (2013) 384 –387.

[118] A. De Simone and T. Jacques, *Simplified models vs. effective field theory approaches in dark matter searches*, Eur. Phys. J. **C76**.7 (2016) 367, arXiv: 1603.08002 [hep-ph].

[119] N. Bell et al., *Unitarisation of EFT Amplitudes for Dark Matter Searches at the LHC*, JHEP **08** (2016) 125, arXiv: 1606.02722 [hep-ph].

[120] J. Abdallah et al., *Simplified Models for Dark Matter and Missing Energy Searches at the LHC*, 2014, arXiv: 1409.2893 [hep-ph].

[121] A. Boveia and C. Doglioni, *Dark Matter Searches at Colliders*, Ann. Rev. Nucl. Part. Sci. **68** (2018) 429–459, arXiv: 1810.12238 [hep-ex].

[122] J. Abdallah et al., *Simplified Models for Dark Matter Searches at the LHC*, Phys. Dark Univ. **9-10** (2015) 8–23, arXiv: 1506.03116 [hep-ph].

[123] P. Agrawal et al., *Flavored dark matter, and its implications for direct detection and colliders*, Phys. Rev. **D86**.5 (2012) 055002.

[124] T. D. Lee, *A Theory of Spontaneous* T *Violation*, Phys. Rev. **D8**.4 (1973) 1226–1239.

[125] A. Celis, J. Fuentes-Martín, and H. Serôdio, *Effective Aligned 2HDM with a DFSZ-like invisible axion*, Phys. Lett. **B737** (2014) 185–190, arXiv: 1407.0971 [hep-ph].

[126] R. Enberg et al., *Signatures of the Type-I 2HDM at the LHC*, PoS **CORFU2018** (2018) 013, arXiv: 1812.08623 [hep-ph].

[127] J. F. Gunion and H. E. Haber, *The CP conserving two Higgs doublet model: The Approach to the decoupling limit*, Phys. Rev. **D67** (2003) 075019, arXiv: hep-ph/0207010.

[128] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) 1–102, arXiv: 1106.0034 [hep-ph].

[129] P. Pani and G. Polesello, *Dark matter production in association with a single top-quark at the LHC in a two-Higgs-doublet model with a pseudoscalar mediator*, Phys. Dark Univ. **21** (2018) 8–15, arXiv: 1712.03874 [hep-ph].

[130] D. Abercrombie et al., *Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum*, Phys. Dark Univ. **27** (2020) 100371, arXiv: 1507.00966 [hep-ex].

[131] J. Andrea, B. Fuks, and F. Maltoni, *Monotops at the LHC*, Phys. Rev. **D84** (2011) 074025, arXiv: 1106.6199 [hep-ph].

[132] I. Boucheneb et al., *Revisiting monotop production at the LHC*, JHEP **01** (2015) 017, arXiv: 1407.7529 [hep-ph].

[133] J. Wang et al., *Search for the signal of monotop production at the early LHC*, Phys. Rev. D **86** (2012) 034008, arXiv: 1109.5963 [hep-ph].

[134] ATLAS Collaboration, *Search for large missing transverse momentum in association with one top-quark in proton-proton collisions at* $\sqrt{s}$ *= 13 TeV with the ATLAS detector*, JHEP **05** (2019) 041, arXiv: 1812.09743 [hep-ex].

[135] LHC Dark Matter Working Group, *LHC Dark Matter Working Group: Next-generation spin-0 dark matter models*, Phys. Dark Univ. **27** (2020) 100351, arXiv: 1810.09420 [hep-ex].

[136] S. Cassel, D. M. Ghilencea, and G. G. Ross, *Electroweak and Dark Matter Constraints on a Z-prime in Models with a Hidden Valley*, Nucl. Phys. **B827** (2010) 256–280, arXiv: 0903.1118 [hep-ph].

[137] P. J. Fox and E. Poppitz, *Leptophilic Dark Matter*, Phys. Rev. **D79** (2009) 083528, arXiv: 0811.0399 [hep-ph].

[138] O. Buchmueller et al., *Characterising dark matter searches at colliders and direct detection experiments: Vector mediators*, JHEP **01** (2015) 037, arXiv: 1407.8257 [hep-ph].

[139] G. Busoni et al., *On the Validity of the Effective Field Theory for Dark Matter Searches at the LHC*, Phys. Lett. **B728** (2014) 412–421, arXiv: 1307.2253 [hep-ph].

[140] G. Busoni et al., *On the Validity of the effective field theory for dark matter searches at the LHC, Part II: Complete analysis for the* s*-channel*, JCAP **1406** (2014) 060, arXiv: 1402.1275 [hep-ph].

[141] G. Busoni et al., *On the validity of the effective field theory for dark matter searches at the LHC Part III: analysis for the* t*-channel*, JCAP **1409** (2014) 022, arXiv: 1405.3101 [hep-ph].

[142] N. Chen et al., *Type-II 2HDM under the Precision Measurements at the Z-pole and a Higgs Factory*, JHEP **03** (2019) 023, arXiv: 1808.02037 [hep-ph].

[143] G. Arcadi, Y. Mambrini, and F. Richard, *Z-portal dark matter*, JCAP **2015**.03 (2015) 018.

[144] S. Westhoff, *Probing Higgs-Portal Dark Matter with Weakly Interacting Mediators*, 2016, arXiv: 1605.05193 [hep-ph].

[145] H. Han and S. Zheng, *Higgs-portal scalar dark matter: Scattering cross section and observable limits*, Nucl. Phys. **914** (2017) 248–256.

[146] V. Khachatryan et al., *Searches for invisible decays of the Higgs boson in pp collisions at* $\sqrt{s}$ *= 7, 8, and 13 TeV*, JHEP **02** (2017) 135, arXiv: 1610.09218 [hep-ex].

[147] ATLAS collaboration, *Search for an invisibly decaying Higgs boson or dark matter candidates produced in association with a* Z *boson in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. **B776** (2018) 318–337, arXiv: 1708.09624 [hep-ex].

[148] N. Craig et al., *The Higgs Portal Above Threshold*, JHEP **02** (2016) 127, arXiv: 1412.0258 [hep-ph].

[149] J. Alitti et al., *A Search for new intermediate vector mesons and excited quarks decaying to two jets at the CERN* $\bar{\mathrm{p}}\mathrm{p}$ *collider*, Nucl. Phys. **B400** (1993) 3–24.

[150] R. M. Harris and K. Kousouris, *Searches for Dijet Resonances at Hadron Colliders*, IJMP **A26** (2011) 5005–5055, arXiv: 1110.5302 [hep-ex].

[151] M. Fairbairn et al., *Constraints on Z' models from LHC dijet searches and implications for dark matter*, JHEP **09** (2016) 018, arXiv: 1605.07940 [hep-ph].

[152] M. Fairbairn and J. Heal, *Complementarity of dark matter searches at resonance*, Phys. Rev. **D90**.11 (2014) 115019, arXiv: 1406.3288 [hep-ph].

[153] ATLAS collaboration, *Search for new phenomena in dijet events using 37 fb*$^{-1}$ *of* pp *collision data collected at* $\sqrt{\mathrm{s}}$ *=13 TeV with the ATLAS detector*, Phys. Rev. **D96**.5 (2017) 052004, arXiv: 1703.09127 [hep-ex].

[154] V. V. Khoze, A. D. Plascencia, and K. Sakurai, *Simplified models of dark matter with a long-lived co-annihilation partner*, JHEP **06** (2017) 041, arXiv: 1702.00750 [hep-ph].

[155] S. Banerjee et al., *Novel signature for long-lived particles at the LHC*, Phys. Rev. **D98**.11 (2018) 115026, arXiv: 1706.07407 [hep-ph].

[156] O. Buchmueller et al., *Simplified Models for Displaced Dark Matter Signatures*, JHEP **09** (2017) 076, arXiv: 1704.06515 [hep-ph].

[157] M. Papucci, A. Vichi, and K. M. Zurek, *Monojet versus the rest of the world I: t-channel models*, JHEP **11** (2014) 024, arXiv: 1402.2285 [hep-ph].

[158] H. An and H. Wang L.-T.and Zhang, *Dark matter with* t*-channel mediator: a simple step beyond contact interaction*, Phys. Rev. **D89**.11 (2014) 115014, arXiv: 1308.0592 [hep-ph].

[159] T. Cohen et al., *LHC Searches for Dark Sector Showers*, JHEP **11** (2017) 196, arXiv: 1707.05326 [hep-ph].

[160] T. Cohen, M. Lisanti, and H. K. Lou, *Semivisible Jets: Dark Matter Undercover at the LHC*, Phys. Rev. Lett. **115**.17 (2015) 171804, arXiv: 1503.00009 [hep-ph].

[161] S. Biswas et al., *Dark-Photon searches via Higgs-boson production at the LHC*, Phys. Rev. **D93**.9 (2016) 093011, arXiv: 1603.01377 [hep-ph].

[162] J. D. Wells, *How to Find a Hidden World at the Large Hadron Collider*, 2008, arXiv: 0803.1243 [hep-ph].

[163] A. L. Read, *Presentation of search results: the CLs technique*, J. Phys. G. Nucl. Partic. **28**.10 (2002) 2693–2704.

[164] ATLAS collaboration, *Constraints on mediator-based dark matter and scalar dark energy models using* $\sqrt{s} = 13$ *TeV* pp *collision data collected by the ATLAS detector*, JHEP **05** (2019) 142, arXiv: 1903.01400 [hep-ex].

[165] T. Emken et al., *Direct Detection of Strongly Interacting Sub-GeV Dark Matter via Electron Recoils*, JCAP **09** (2019) 070, arXiv: 1905.06348 [hep-ph].

[166] L. Evans and P. Bryant, *LHC Machine*, JINST **3**.08 (2008) S08001.

[167] LEP Collaboration, *LEP design report*, CERN-LEP-84-01, 1984.

[168] L. Arnaudon et al., *Linac 4 technical design report*, CERN-AB-2006-084, 2006.

[169] A. Lombardi, *Commissioning of CERN Linac 4*, CERN-ACC-2018-196 (2018) TH1P01.

[170] E. Mobs, *The CERN accelerator complex*, CERN-GRAPHICS-2019-002, 2019.

[171] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3**.08 (2008) S08005.

[172] LHCf Collaboration et al., *The LHCf detector at the CERN Large Hadron Collider*, JINST **3**.08 (2008) S08006.

[173] TOTEM Collaboration et al., *The TOTEM Experiment at the CERN Large Hadron Collider*, JINST **3**.08 (2008) S08007.

[174] J. Pinfold et al., *Technical Design Report of the MoEDAL Experiment*, CERN-LHCC-2009-006. MoEDAL-TDR-001, 2009.

[175] M. Riordan, *The Discovery of quarks*, Science **256** (1992) 1287–1293.

[176] TOTEM Collaboration, *First measurement of elastic, inelastic and total cross-section at $\sqrt{s} = 13$ TeV by TOTEM and overview of cross-section data at LHC energies*, Eur. Phys. J. C **79**.2 (2019) 103, arXiv: 1712.06153 [hep-ex].

[177] R. P. Feynman, *The behavior of hadron collisions at extreme energies*, Conf. Proc. C **690905** (1969) 237–258.

[178] J. M. Campbell, J.W. Huston, and W.J. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, Rept. Prog. Phys. **70** (2007) 89, arXiv: hep-ph/0611148.

[179] R. Field, *The Underlying Event in Hadronic Collisions*, Annual Review of Nuclear and Particle Science **62**.1 (2012) 453–483.

[180] P. Grafström and W. Kozanecki, *Luminosity determination at proton colliders*, Prog. Part. Nucl. Phys. **81** (2015) 97–148.

[181] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15**.04 (2020) P04003, arXiv: 1911.04632 [physics.ins-det].

[182] J. Pequenao, *Computer generated image of the whole ATLAS detector*, CERN-GE-0803012, 2008.

[183] S. Haywood et al., *ATLAS inner detector: Technical Design Report, 2*, CERN-LHCC-97-017, 1997.

[184] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*, CERN-LHCC-97-016, 1997.

[185] M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013. ATLAS-TDR-19, 2010.

[186] J. Pequenao, *Computer generated image of the ATLAS inner detector*, CERN-GE-0803014, 2008.

[187] A. La Rosa, *The ATLAS Insertable B-Layer: from construction to operation*, JINST **11**.12 (2016) C12036, arXiv: 1610.01994 [physics.ins-det].

[188] ATLAS Collaboration, *Study of the material of the ATLAS inner detector for Run 2 of the LHC*, JINST **12** (2017) P12009, arXiv: 1707.02826 [hep-ex].

[189] N. Wermes and G. Hallewel, *ATLAS pixel detector: Technical Design Report*, CERN-LHCC-98-013, 1998.

[190] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, JINST **9**.08 (2014) P08009.

[191] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*, CERN-LHCC-96-041, 1996.

[192] M. Marjanovic, *ATLAS Tile calorimeter calibration and monitoring systems*, IEEE Trans. Nucl. Sci. **66**.7 (2019) 1228–1235, arXiv: 1806.09156 [p.i-det].

[193] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*, CERN-LHCC-96-042, 1996.

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

[195] J. G. P. Vazquez, *The ATLAS Data Acquisition System in LHC Run 2*, Journal of Physics: Conference Series **898** (2017) 032017.

[196] ATLAS Collaboration, *Expected Tracking Performance of the ATLAS Inner Tracker at the HL-LHC*, ATL-PHYS-PUB-2019-014, 2019.

[197] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*, CERN-LHCC-2017-021. ATLAS-TDR-030, 2017.

[198] ATLAS Collaboration, *Technical Design Report for the ATLAS Inner Tracker Strip Detector*, CERN-LHCC-2017-005. ATLAS-TDR-025, 2017.

[199] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS LAr Calorimeter*, CERN-LHCC-2017-018. ATLAS-TDR-027, 2017.

[200] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter*, CERN-LHCC-2017-019. ATLAS-TDR-028, 2017.

[201] S.M. Mazza, *A High-Granularity Timing Detector (HGTD) for the Phase-II upgrade of the ATLAS detector*, JINST **14**.10 (2019) C10028.

[202] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*, CERN-LHCC-2017-017. ATLAS-TDR-026, 2017.

[203] ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*, CERN-LHCC-2017-020. ATLAS-TDR-029, 2017.

[204] ATLAS Collaboration, *Expected performance of the ATLAS detector at the High-Luminosity LHC*, ATL-PHYS-PUB-2019-005, 2019.

[205] ATLAS Collaboration, *ATLAS Computing: technical design report*, Technical Design Report ATLAS, 2005.

[206] G. Zanderighi, *Quantum Chromodynamics and Collider Processes*, 2013, URL: https://cds.cern.ch/record/2009137.

[207] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. **C77** (2017) 317, arXiv: 1611.09661 [hep-ex].

[208] ATLAS Collaboration, *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, ATL-DAQ-PUB-2016-001, 2016.

[209] ATLAS Collaboration, *Trigger Menu in 2016*, ATL-DAQ-PUB-2017-001, 2017.

[210] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76**.12 (2016) 653, arXiv: 1608.03953 [hep-ex].

[211] ATLAS Collaboration, *Luminosity determination in* pp *collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019.

[212] ATLAS Collaboration, *Luminosity determination in* pp *collisions at* $\sqrt{s} = 8\,TeV$ *using the ATLAS detector at the LHC*, Eur. Phys. J. **C76** (2016) 653, arXiv: 1608.03953 [hep-ex].

[213] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13**.07 (2018) P07017.

[214] B. Webber, *Parton shower Monte Carlo event generators*, Scholarpedia **6**.12 (2011) 10662.

[215] J. M. Katzy, *QCD Monte-Carlo model tunes for the LHC*, Progress in Particle and Nuclear Physics **73** (2013) 141 –187.

[216] B. Andersson et al., *Parton fragmentation and string dynamics*, Physics Reports **97**.2 (1983) 31 –145.

[217] T.Sjöstrand, *The Lund Monte Carlo for jet fragmentation and e+e- physics - jetset version 6.2*, Compt. Phys. Commun. **39**.3 (1986) 347 –407.

[218] I. Borozan and M.H. Seymour, *An Eikonal model for multiparticle production in hadron hadron interactions*, JHEP **09** (2002) 015, arXiv: hep-ph/0207283.

[219] B. R. Webber, *Fragmentation and hadronization*, eConf **C990809** (2000) 577–606, arXiv: hep-ph/9912292.

[220] T. Sjostrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, arXiv: 0710.3820 [hep-ph].

[221] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[222] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016.

[223] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at* $\sqrt{s} = 13\,TeV$ *with the ATLAS Detector at the LHC*, Phys. Rev. Lett. **117** (2016) 182002, arXiv: 1606.02625 [hep-ex].

[224] S. Alioli, *NLO and Parton Showers: The POWHEG-BOX*, 2010, arXiv: 1009.2348 [hep-ph].

[225] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146.

[226] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph].

[227] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph].

[228] S. Frixione, P. Nason, and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph].

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

[230] M. Bahr et al., *Herwig++ physics and manual*, Eur. Phys. J. **C58** (2008) 639, arXiv: 0803.0883 [hep-ph].

[231] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph].

[232] T. Gleisberg, S. Hoeche, M. Schoenherr, F. Siegert, J. Winter, *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph].

[233] T. Sjöstrand, M. van Zijl, *A multiple-interaction model for the event structure in hadron collisions*, Eur. Phys. J. **C36**.7 (1987).

[234] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. **C76**.4 (2016) 196, arXiv: 1512.01178 [hep-ph].

[235] G. Barrand et al., *GAUDI — A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140**.1 (2001) 45 –55.

[236] S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[237] R. Brun, F. Rademakers, and S. Panacek, *ROOT, an object oriented data analysis framework*, CERN School of Computing (CSC 2000), 2000 11–42.

[238] J. Boudreau and V. Tsulaia, *The GeoModel Toolkit for Detector Description*, Proceedings of Computing in High Energy Physics and Nuclear Physics, 2005.

[239] F. Krauss and D. Napoletano, *Towards a fully massive five-flavor scheme*, Phys. Rev. **D98**.9 (2018) 096002, arXiv: 1712.06832 [hep-ph].

[240] R. Frederix, E. Re, and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, JHEP **09** (2012) 130, arXiv: 1207.5391 [hep-ph].

[241] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. **C71** (2011) 1547, arXiv: 1009.2450 [hep-ph].

[242] S. Alioli et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09** (2009) 111, arXiv: 0907.4076 [hep-ph].

[243] ATLAS Collaboration, *Studies on top-quark Montecarlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016.

[244] M. Beneke et al., *Hadronic top-quark pair production with NNLL threshold resummation*, Nucl. Phys. **B855** (2012) 695–741, arXiv: 1109.1536 [hep-ph].

[245] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612–622, arXiv: 1111.5869 [hep-ph].

[246] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to* $\mathrm{q\bar{q}} \to \mathrm{t\bar{t}} + \mathrm{X}$, Phys. Rev. Lett. **109** (2012) 132001, arXiv: 1204.5201 [hep-ph].

[247] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080, arXiv: 1210.6832 [hep-ph].

[248] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through* $O(\alpha_S^4)$, Phys. Rev. Lett. **110** (2013) 252004, arXiv: 1303.6254 [hep-ph].

[249] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph].

[250] S. Frixione et al., *Single-top hadroproduction in association with a* W *boson*, JHEP **07** (2008) 029, arXiv: 0805.3067 [hep-ph].

[251] M. Aliev et al., *HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR*, Comput. Phys. Commun. **182** (2011) 1034–1046, arXiv: 1007.1327 [hep-ph].

[252] P. Kant et al., *HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, Comput. Phys. Commun. **191** (2015) 74–89, arXiv: 1406.4403 [hep-ph].

[253] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph].

[254] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph].

[255] S. Höche et al., *A critical appraisal of NLO+PS matching methods*, JHEP **09** (2012) 049, arXiv: 1111.1220 [hep-ph].

[256] S. Catani et al., *QCD matrix elements + parton showers*, JHEP **11** (2001) 063, arXiv: hep-ph/0109231.

[257] S. Höche et al., *QCD matrix elements and truncated showers*, JHEP **05** (2009) 053, arXiv: 0903.1219 [hep-ph].

[258] S. Höche et al., *QCD matrix elements + parton showers. The NLO case*, JHEP **04** (2013) 027, arXiv: 1207.5030 [hep-ph].

[259] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph].

[260] A. Denner, S. Dittmaier, and L. Hofer, *Collier: a fortran-based Complex One-Loop LIbrary in Extended Regularizations*, Comput. Phys. Commun. **212** (2017) 220–238, arXiv: 1604.06792 [hep-ph].

[261] ATLAS Collaboration, *ATLAS simulation of boson plus jets processes in Run 2*, ATL-PHYS-PUB-2017-006, 2017.

[262] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, CERN Yellow Reports: Monographs **2** (2017), arXiv: 1610.07922 [hep-ph].

[263] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014.

[264] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. **D83** (2011) 091503, arXiv: 1103.2792 [hep-ph].

[265] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, Phys. Rev. **D82** (2010) 054018, arXiv: 1005.4451 [hep-ph].

[266] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. **D81** (2010) 054028, arXiv: 1001.5034 [hep-ph].

[267] S. Catani et al., *Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO*, Phys. Rev. Lett. **103** (2009) 082001, arXiv: 0903.2120 [hep-ph].

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

[269] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152–155.

[270] CMS Collaboration, *Search for dark matter in events with energetic, hadronically decaying top quarks and missing transverse momentum at $\sqrt{s}$ = 13 TeV*, JHEP **06** (2018) 027, arXiv: 1801.08427 [hep-ex].

[271] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, arXiv: 1007.2241 [hep-ph].

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

[273] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Phys. Rev. **D82** (2010) 074018, arXiv: 1005.3457 [hep-ph].

[274] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: 1406.3660 [hep-ex].

[275] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and Physics Documentation*, Comput. Phys. Commun. **199** (2016) 86–101, arXiv: 1011.0937 [hep-ph].

[276] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, arXiv: hep-ph/0201195.

[277] ATLAS Collaboration, *Simulation of top quark production for the ATLAS experiment at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2016-004, 2016.

[278] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high pT jets*, ATL-PHYS-PUB-2014-008, 2014.

[279] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling with Sherpa and MG5aMCatNLO*, ATL-PHYS-PUB-2017-007, 2017.

[280] Y. Amhis et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016*, Eur. Phys. J. **C77**.12 (2017) 895, arXiv: 1612.07233 [hep-ex].

[281] E. Lohrmann, *A Summary of Charm Hadron Production Fractions* (2011), arXiv: 1112.3757 [hep-ex].

[282] DELPHI Collaboration, *A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole*, Eur. Phys. J. **C71**.2 (2011) 1–29.

[283] SLD Collaboration, *Measurement of the b quark fragmentation function in Z0 decays*, Phys. Rev. **D65** (2002) 092006, arXiv: hep-ex/0202031, (Erratum: Phys. Rev. **D66**, 079905 (2002)).

[284] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph].

[285] J. Pequenao and P. Schaffner, *How ATLAS detects particles: diagram of particle paths in the detector*, CERN-EX-1301009, 2013.

[286] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, Eur. Phys. J. **C77**.10 (2017) 673, arXiv: 1704.07983 [hep-ex].

[287] T. Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*, Journal of Physics: Conference Series **119**.3 (2008) 032014.

[288] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262**.2 (1987) 444 –450.

[289] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, 2012.

[290] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. **C77**.5 (2017) 332, arXiv: 1611.10235 [physics.ins-det].

[291] J. Jiménez Peña, *Alignment of the ATLAS Inner Detector Upgraded for the LHC Run II*, J. Phys. Conf. Ser. **664**.7 (2015) 072025.

[292] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024, 2016.

[293] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at* $\sqrt{s} = 13$ TeV, ATL-PHYS-PUB-2016-015, 2016.

[294] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, 2008.

[295] ATLAS Collaboration, *Electron identification measurements in ATLAS using* $\sqrt{s}$ *= 13 TeV data with 50 ns bunch spacing*, ATL-PHYS-PUB-2015-041, 2015.

[296] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at* $\sqrt{s}$ *= 13 TeV*, Eur. Phys. J. **C79**.8 (2019) 639, arXiv: 1902.04655 [p.i-det].

[297] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. **C74**.10 (2014) 3071, arXiv: 1407.5063 [hep-ex].

[298] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data*, JINST **14**.03 (2019) P03017, arXiv: 1812.03848 [hep-ex].

[299] ATLAS Collaboration, *Muon reconstruction efficiency and momentum resolution of the ATLAS experiment in proton-proton collisions at* $\sqrt{s}$ *= 7 TeV in 2010*, Eur. Phys. J. **C74**.9 (2014) 3034, arXiv: 1404.4562 [hep-ex].

[300] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, Eur. Phys. J. **C74**.11 (2014) 3130, arXiv: 1407.3935 [hep-ex].

[301] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s}$ *=13 TeV*, Eur. Phys. J. **C76**.5 (2016) 292, arXiv: 1603.05598 [hep-ex].

[302] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. **D96** (2017) 072002, arXiv: 1703.09665 [hep-ex].

[303] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in* pp *collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Eur. Phys. J. **C76** (2016) 581, arXiv: 1510.03823 [hep-ex].

[304] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. **C72** (2012) 1896, arXiv: 1111.6097 [hep-ph].

[305] ATLAS Collaboration, *Expected performance of the ATLAS* b*-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, 2015.

[306] ATLAS Collaboration, *Optimisation of the ATLAS* b*-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016.

[307] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using* $t\bar{t}$ *events at* $\sqrt{s} = 13$ *TeV*, JHEP **08** (2018) 089, arXiv: 1805.01845 [hep-ex].

[308] ATLAS Collaboration, *Performance of missing transverse momentum reconstructionfor the ATLAS detector in the first proton-proton collisions at at* $\sqrt{s}$= *13 TeV*, ATL-PHYS-PUB-2015-027, 2015.

[309] ATLAS Collaboration, *Performance assumptions based on full simulation for an upgraded ATLAS detector at a High-Luminosity LHC*, ATL-PHYS-PUB-2013-009, 2013.

[310] ATLAS Collaboration, *Proposal for truth particle observable definitions in physics measurements*, ATL-PHYS-PUB-2015-013, 2015.

[311] ATLAS Collaboration, *ATLAS Phase-II Upgrade Scoping Document*, CERN-LHCC-2015-020. LHCC-G-166, 2015.

[312] ATLAS Collaboration, *Expected performance for an upgraded ATLAS detector at High-Luminosity LHC*, ATL-PHYS-PUB-2016-026, 2016.

[313] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide.* CERN-OPEN-2007-007, 2007, arXiv: physics/0703039.

[314] M. Stone, *Cross-Validatory Choice and Assessment of Statistical Predictions*, Journal of the Royal Statistical Society: Series B (Methodological) **36**.2 (1974) 111–133.

[315] Z. Marshall, *Simulation of Pile-up in the ATLAS Experiment*, J. Phys. Conf. Ser. **513** (2014) 022024.

[316] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. **C75** (2015) 17, arXiv: 1406.0076 [hep-ex].

[317] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-015, 2015.

[318] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083, 2013.

[319] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014.

[320] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex].

[321] A. Buckley and H. Schulz, *Tuning of MC generator MPI models*, 2019 281–301, arXiv: 1806.11182 [hep-ph].

[322] G. Ranucci, *The profile likelihood ratio and the look elsewhere effect in high energy physics*, Nucl. Instrum. Meth. **A661**.1 (2012) 77–85.

[323] ATLAS Collaboration, *Reproducing searches for new physics with the ATLAS experiment through publication of full statistical likelihoods*, ATL-PHYS-PUB-2019-029, 2019.

[324] A. L. Read, *Modified frequentist analysis of search results (the CLs method)*, CERN-OPEN-2000-205, 2000.

[325] ATLAS Collaboration, *Search for dark matter associated production with a single top quark in $\sqrt{s} = 13$ TeV (pp) collisions with the ATLAS detector*, ATLAS-CONF-2020-034, 2020.

[326] X. C. Vidal et al., CERN Yellow Reports: Monographs **7** (2019) 585–865, arXiv: 1812.07831 [hep-ph].

[327] ATLAS Collaboration, *Prospects for a search of invisible particles produced in association with single-top quarks with the ATLAS detector at the HL-LHC*, ATL-PHYS-PUB-2018-024, 2018.

[328] T. Aaltonen et al., *Search for a dark matter candidate produced in association with a single top quark in* $\mathrm{p\bar{p}}$ *collisions at* $\sqrt{\mathrm{s}} = 1.96$ *TeV*, Phys. Rev. Lett. **108** (2012) 201802, arXiv: 1202.5653 [hep-ex].

[329] ATLAS Collaboration, *Search for invisible particles produced in association with single-top-quarks in proton–proton collisions at* $\sqrt{\mathrm{s}} = 8$ *TeV with the ATLAS detector*, Eur. Phys. J. **C75** (2015) 79, arXiv: 1410.5404 [hep-ex].

[330] V. Khachatryan et al., *Search for Monotop Signatures in Proton-Proton Collisions at* $\sqrt{\mathrm{s}} = 8$ *TeV*, Phys. Rev. Lett. **114**.10 (2015) 101801, arXiv: 1410.1149 [hep-ex].

[331] A. M. Sirunyan et al., *Search for dark matter in events with energetic, hadronically decaying top quarks and missing transverse momentum at* $\sqrt{\mathrm{s}} = 13$ *TeV*, JHEP **06** (2018) 027, arXiv: 1801.08427 [hep-ex].

[332] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, 2003, arXiv: physics/0306116.

[333] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, arXiv: 1007.1727 [physics.data-an], (Erratum: Eur. Phys. J. **C73** (2013) 2501).

[334] R. E. Schapire, *The strength of weak learnability*, Machine Learning **5**.2 (1990) 197–227.

[335] Y. Coadou, *Boosted Decision Trees and Applications*, EPJ Web Conf. **55** (2013) 02004.

[336] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, J. Comput. Syst. Sci. **55**.1 (1997) 119 –139.

[337] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters **27**.8 (2006) 861 –874.

[338] J.A. Hanley and B. Mcneil, *The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve*, Radiology **143** (1982) 29–36.

[339] Stephen M. Stigler, *Francis Galton's Account of the Invention of Correlation*, Statist. Sci. **4**.2 (1989) 73–79.