UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVII CICLO DEL DOTTORATO DI RICERCA IN FISICA

Top Quark Physics at ATLAS: From Cross-Section Measurements to Search for New Particles

Settore scientifico-disciplinare: FIS/01 FISICA SPERIMENTALE

DOTTORANDO / A
LAURA PINTUCCI

COORDINATORE
PROF. FRANCESCO LONGO

SUPERVISORE DI TESI
DR. MICHELE PINAMONTI

CO-SUPERVISORE DI TESI
DR. GIANCARLO PANIZZO

ANNO ACCADEMICO 2023/2024

Abstract

This thesis presents experimental studies of the top quark within the ATLAS experiment at the Large Hadron Collider (LHC). The top quark, with its mass of 172.5 GeV, is the heaviest known elementary particle and possesses unique properties: it has the strongest coupling to the Higgs boson among all SM particles and decays before hadronization, allowing its study as an almost free quark.

Two complementary physics analyses are presented in this thesis. First, the measurement of the single top quark production cross section in the t-channel at a centre-of-mass energy of $\sqrt{s} = 5.02$ TeV, using pp collision data corresponding to an integrated luminosity of 255 pb^{-1} is detailed. After event selections a Boosted Decision Tree is employed for signal-to-background discrimination, and a Profile Likelihood fit is used to extract the total t-channel cross-section, the individual top- and antitop-quark production cross-sections, and their ratio. The t-channel single top quark process is observed for the first time at $\sqrt{s} = 5.02$ TeV with a significance of 6.1σ. The inclusive cross-section for the t-channel single-top production is measured to be $\sigma(tq + \bar{t}q) = 27.1^{+4.4}_{-4.1}$ (stat.) $^{+4.4}_{-3.7}$ (syst.) pb, while the individual cross-sections are measured to be $\sigma(tq) = 19.8^{+3.9}_{-3.1}$ (stat.) $^{+2.9}_{-2.2}$ (syst.) pb and $\sigma(\bar{t}q) = 7.3^{+3.2}_{-2.1}$ (stat.) $^{+2.8}_{-1.5}$ (syst.) pb respectively. The ratio between top- and antitop-quark production is found to be $R_t = 2.73^{+1.43}_{-0.82}$ (stat.) $^{+1.01}_{-0.29}$ (syst.). All measurements are in good agreement with the Standard Model (SM) predictions.

Second, a search for new heavy resonances decaying to top-antitop quark pairs is conducted using the full *Run 2* dataset, consisting of 140 fb^{-1} if pp collisions at $\sqrt{s} = 13$ TeV. The analysis targets three specific Beyond the Standard Model scenarios: a leptophobic Z' boson from Top-colour-assisted technicolour, and two variants of Randall-Sundrum models predicting Kaluza-Klein excitations of the graviton and gluon. The search is performed in the semileptonic decay channel, targeting both resolved and boosted topologies, implementing a specialized overlap-removal technique to enhance the sensitivity at high invariant masses. No significant deviation from the SM prediction is observed. Exclusion limits are set on the production cross-section times branching ratio for the considered signal models, obtaining a higher sensitivity in the extracted limits with respect to the previous ATLAS and CMS searches in the same decay channel.

The thesis also includes technical contributions to the ATLAS collaboration combined performance effort, focusing on electron identification in environments with nearby hadronic jets. This study examines the electron identification process and its efficiencies in challenging scenarios where electron showers overlap with jet constituents. This work contributes to the development of current and especially future physics analyses in the ATLAS collaboration.

Contents

Introduction

The Standard Model (SM) of particle physics is the most successful framework for describing the fundamental constituents of matter and their interactions. Particle colliders over the past few decades, in particular since the beginning of the Large Hadron Collider (LHC) programme, have been able to provide numerous experimental evidence of its validity. Despite its predictive power, several theoretical and experimental observations suggest it may be an incomplete description of Nature, and possible new solutions are being searched outside of this model, through the investigation of Beyond the Standard Model (BSM) physics. The top quark, as the heaviest known elementary particle with a mass of 172.5 GeV [1], plays a unique role in this context.

The top quark has some unique characteristics that make it particularly interesting to study: thanks to its large mass, it has the strongest coupling to the Higgs boson field among all fermions, and it decays before the hadronization process begins, with its width of $\Gamma_t = 1.42^{0.19}_{-0.15}$ GeV [1]. It is therefore the only quark that can be experimentally studied almost as a *free* particle. Its large mass, close to the electroweak symmetry breaking scale, makes it particularly sensitive to potential new physics phenomena and provides a sensitive probe for BSM theories. In particular, new physics could introduce additional top-quark production or decay mechanisms. Various new physics models predicting different kinds of new particles decaying to $t\bar{t}$ have been proposed over the years, with some of them still keeping their relevance.

This thesis presents two complementary experimental investigations involving top quarks conducted within the ATLAS collaboration [2] at the Large Hadron Collider (LHC) [3]: one is a SM measurement of single top quark production in the t-channel at a centre-of-mass energy of $\sqrt{s} = 5.02$ TeV with a dataset of pp collisions collected by the ATLAS experiment corresponding to an integrated luminosity $L = 255$ pb^{-1}, while the other is a search for new physics, specifically looking for possible new heavy resonances decaying into top-antitop-quark pairs using the full dataset collected between 2015 and 2018 at $\sqrt{s} = 13$ TeV with an integrated luminosity of $L = 140$ fb^{-1}.

Single top quark production, first observed at the Tevatron $p\bar{p}$ collider in 2009 [4, 5], provides a unique window into electroweak interactions and allows direct measurement of the CKM matrix element $|V_{tb}|$. The t-channel production mechanism is particularly interesting as it probes the b-quark content of the proton. The measurement discussed in this thesis at $\sqrt{s} = 5.02$ TeV represents the first observation of this process at this specific energy, despite the limited integrated luminosity of $L = 255$ pb^{-1}. Events with one electron or muon in the final state and exactly two hadronic jets, with exactly one containing a b-hadron decay are selected. To enhance the signal-over-background ratio the so-called spectator jets (the one not containing the b-hadron decay) are selected in the pseudo-rapidity range of $1.5 < |\eta| < 4.0$. Further selections on the separation among final state products and on the reconstructed W-boson and t-quark candidates are also used to suppress background contributions. A boosted Decision Tree (BDT) is applied to better separate signal from background events, and its output is used to perform a Profile Likelihood Fit. The total t-channel

cross-section, the top-quark and antitop-quark t-channel cross-sections and their ratio are extracted. The results are compared with values predicted by different Parton Distribution Functions sets and the results have been recently published [6].

The search for heavy resonances decaying to $t\bar{t}$ pairs is one of the fundamental ways to discover new physics. Among the large set of BSM theories predicting new heavy particles that couple preferentially to top quarks, in this thesis, three models are specifically tested belonging to two macro classes: top-colour-assisted technicolour and Randall-Sundrum (RS) theories of warped extra dimensions. To the first class belongs a specific model predicting a Leptophobic Z' spin-1 colour singlet new boson [7]. Concerning the second class (RS), two models have been tested: a model predicting a Kaluza-Klein (KK) excitation of the graviton corresponding to a spin-2 colour singlet [8], while in the other case a model that creates KK excitations of the gluon, which predicts a spin-1 colour octet [9].

This search for heavy spin-1 and -2 resonances decaying into a top-antitop quark pair is performed in the semileptonic decay channel of $t\bar{t}$. The full *Run-2* dataset of 140 fb^{-1} of pp collision data is used, extending the sensitivity compared to previous ATLAS results in the same decay channel [10]. Final states are selected with exactly one electron or muon, large missing transverse energy and at least one hadronic jet containing a b-hadron decay. The analysis considers both resolved and boosted topologies, to optimize the selections and reconstruction techniques for the distinct experimental signatures that arise when the top quarks are produced with different kinematic properties. A special overlap-removal technique is used for the selection of reconstructed physics objects that overlap with each other. This technique is used here to increase the statistics of reconstructed events where the top and antitop quarks are highly boosted, enhancing the sensitivity of the search at high invariant masses of the reconstructed $t\bar{t}$ system. A profile likelihood fit is performed in the hypothesis of SM only and no significant deviations are observed. Exclusion limits are then set on the production cross-section times branching ratio for the BSM models considered.

The work discussed in this thesis also contributes to technical aspects related to performance studies of reconstructed object identification in the ATLAS detector. In Appendix A, a comprehensive study of electron identification efficiencies in environments with nearby jets, including the challenging regime where electron showers overlap with jet constituents is given. This contribution has been performed for current and especially future development of physics analyses in the ATLAS collaboration.

The structure of the thesis is the following: Chapter 1 establishes the theoretical framework of the SM with particular attention to the physics of the top quark. The experimental setup of the ATLAS experiment at the LHC is discussed in Chapter 2, while Chapter 3 details the analysis techniques which have been used in the two physics analyses presented in this work, including physics object reconstruction methodologies, Monte Carlo generators and data-driven process estimation. Chapters 4 and 5 present the two main analyses: the single top t-channel cross-section measurement at $\sqrt{s} = 5.02$ TeV is discussed in Chapter 4, and Chapter 5 describes the search for $t\bar{t}$ resonances at $\sqrt{s} = 13$ TeV.

Chapter 1

Standard Model and Top Quark Physics

1.1 The Standard Model

The Standard Model (SM) of particle physics defines the fundamental particles and their interactions within a gauge quantum structure. Originally formulated in the 1960s by Sheldon Glashow, Steven Weinberg, and Abdus Salam, the model has evolved significantly over the decades. The SM treats particles as field excitations that interact through three primary forces: electromagnetic, weak, and strong forces, complemented by the Higgs mechanism, which acts similarly to a fourth force. An enormous number of measurements have validated this theory, which is tested daily at the Large Hadron Collider (LHC)[3].

This section introduces the SM, focusing on features relevant to this thesis and highlighting the crucial role of the top quark.

1.1.1 Particles and their Interactions

The SM is a quantum field theory, based on a set of fields corresponding to the known fermions and on the gauge symmetries $SU(3)_c \times SU(2)_L \times U(1)_Y$. Fundamental or elementary particles, i.e. those with no internal structure (listed in Figure 1.1), are classified into two main types based on their spin properties: fermions, possessing half-integer spins and following Fermi-Dirac statistics rules, and bosons, which have integer spins and obey to Bose-Einstein statistics.

Fermions are categorized into two primary groups according to the forces they interact with. Quarks undergo interactions mediated by strong, electromagnetic, and weak forces. In contrast, leptons do not interact via the strong force. There are six distinct quark flavours organized into three generations with increasing mass: (u,d), (c,s), and (t,b). Each quark is paired with an anti-quark, with the same mass but opposite quantum numbers. Each generation consists of two quarks: one carrying a $+2/3$ e charge (*up-type*) and the other a $-1/3e$ charge. Additionally, quarks are assigned one of three colour quantum numbers (red, blue, or green) and they form hadrons (composite particles) to maintain colour neutrality. Hadrons are grouped into two types: mesons, consisting of a quark and an anti-quark, and baryons, made up of three quarks. Similarly, leptons are structured into three generations, each including a negatively charged particle (e, μ, τ) and the corresponding neutrino $(\nu_e, \nu_{mu}, \nu_{tau})$, a neutrally charged and light particle that interacts solely through the weak force. As for quarks, each lepton also has its antiparticle. Fermions are categorized into two primary groups according to the forces they interact with.

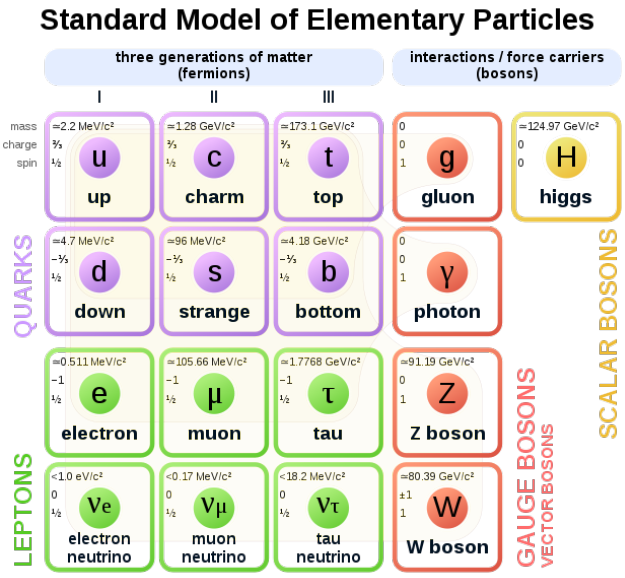


Figure 1.1: Diagram showing fundamental particles of the SM.

Quarks undergo interactions mediated by strong, electromagnetic, and weak forces. In contrast, leptons do not interact via the strong force. There are six distinct quark flavours organized into three generations with increasing mass: (u,d), (c,s), and (t,b). Each quark is paired with an anti-quark, with the same mass but opposite quantum numbers. Each generation consists of two quarks: one carrying a $+2/3$ e charge (*up-type*) and the other a -1/3e charge (*down-type*). Additionally, quarks are assigned one of three colour quantum numbers (red, blue, or green) and they form hadrons (composite particles) to maintain colour neutrality. Hadrons are grouped into two types: mesons, consisting of a quark and an anti-quark, and baryons, made up of three quarks. Similarly, leptons are structured into three generations, each including a negatively charged particle (e, μ, τ) and the corresponding neutrino $(\nu_e, \nu_{mu}, \nu_{tau})$, a neutrally charged and light particle that interacts solely through the weak force. As for quarks, each lepton also has its antiparticle.

Among bosons, there are vector bosons with spin 1 and scalar bosons with spin 0. Vector bosons, also referred to as gauge bosons, act as carriers of the three fundamental forces defined by the SM. The strong force is mediated by eight gluons (g), that have no mass and each one carries a colour and anti-colour charge. The weak force is carried by massive particles: two charged bosons $W^{\pm}$ and one neutral boson Z. The electromagnetic force is transmitted by the mass-less photon (γ). The Higgs boson, the sole elementary scalar boson within the SM, emerges from the spontaneous breaking of the electroweak symmetry through the Higgs mechanism.

1.1.2 The Electroweak Lagrangian and the Higgs Mechanism

The electromagnetic (EM) theory and the weak interaction are combined together in the electroweak (EW) theory [11] as two manifestations of one interaction, which is based on the group $SU(2)_L \times U(1)_Y$. The $SU(2)_L$ is the weak isospin group, a non-abelian group with the corresponding three gauge fields $W^{\mu}_{k=1,2,3}$. The $U(1)_Y$ is the weak hyper-charge group, an abelian group, with the corresponding gauge field B_{μ}. Fermion spinors ψ consist of right-handed (RH) and left-handed (LH) components, defined as:

$$\psi(x)_R = \frac{1}{2}(1+\gamma^5)\ \psi(x) \quad \text{and} \quad \psi(x)_L = \frac{1}{2}(1-\gamma^5)\ \psi(x) \tag{1.1}$$

where γ^5 is proportional to the four Dirac matrices. The weak $(V-A)$ interaction only couples to LH particles and RH antiparticles, therefore parity is not conserved in weak interactions. Fermions can be LH doublets with weak isospin $I = \frac{1}{2}$, or RH singlets with $I = 0$; as an instance the first leptonic generation is

$$\begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}, \quad e_R. \tag{1.2}$$

The EM interaction can be seen as a subgroup of the EW group, thus the EM generator can be written as the linear combination of the EW generator:

$$Q = T_3 + \frac{Y}{2}. \tag{1.3}$$

where T_3 is the third of the three weak isospin group generators $\vec{T} = \frac{1}{2}\vec{\sigma}$ proportional to the Pauli matrices (σ_i); while $\frac{Y}{2}$ is the weak hyper-charge group generator. The $SU(2)_L \times U(1)_Y$ gauge group would imply the existence of four massless gauge bosons: two charged physical gauge bosons $W^{\pm}$, and two neutral ones γ and Z. However, three of these bosons $(W^{\pm}, Z)$ are observed to be massive. To account for these masses, the Higgs mechanism is employed, which initiates EW spontaneous symmetry breaking.

The $SU(2)_L \times U(1)_Y$ local gauge symmetry can undergo the spontaneous symmetry breaking mechanism, proposed by Higgs, Englert and Brout [12, 13] by considering the Higgs Lagrangian:

$$\mathcal{L}_{Higgs} = (D^{\mu}\phi)^{\dagger}(D_{\mu}\phi) - V(\phi) \tag{1.4}$$

with ϕ as a weak isospin doublet of two complex scalar fields, and the covariant derivative $D_\mu = \partial_\mu + ig_W\vec{T}\cdot\vec{W}_\mu + ig'^{Y}/_2 B_\mu$. In the covariant derivative g_W is the coupling strength for the weak force of the $SU(2)$ group, while g' the coupling strength for the electromagnetic force of the $U(1)$ symmetry group. The first term of $\mathcal{L}_{Higgs}$ is the kinematic part, while the second term $V(\phi) = \frac{1}{2}\mu^2\phi^{\dagger}\phi + \frac{1}{4}\lambda(\phi^{\dagger}\phi)^2$ is the potential. The shape of $V(\phi)$ depends on the signs of λ and μ^2 (real constants). The sign of μ^2 can be negative, in which case the potential has a set of degenerate minima (the "vacuum states") $\phi^{\dagger}\phi = {v^2}/_2 = -{\mu^2}/_{2\lambda}$. The local minimum is no longer unique and thus leads to the spontaneous symmetry breaking (SSB). To keep the photon mass equal to zero, the expectation value in the vacuum needs to be non-zero only for the first component of the field (which takes the value $\frac{v}{\sqrt{2}}$, with $v \approx 246$ GeV). Expanding the field around the minimum it can be written as:

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ v + \eta(x) + i\phi_4(x) \end{pmatrix}. \tag{1.5}$$

The field can now be written in the unitary gauge as:

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.6}$$

here $h(x)$ is a scalar field. Writing the field in this way, the kinematic term in Equation 1.4 introduces mass terms for the three gauge bosons:

$$m_{W^{\pm}} = \frac{1}{2}g_W v \quad \text{and} \quad m_Z = \frac{v}{2}\sqrt{g_W^2 + g'^2} \tag{1.7}$$

.

The Yukawa Lagrangian takes into account the interaction of the Higgs field with fermions:

$$-\mathcal{L}_{Yu} = \sum_{ij} \Gamma_u^{ij} \overline{Q}_L^i (i\sigma_2)\phi^* u_R^j + \Gamma_d^{ij} \overline{Q}_L^i \phi d_R^j + \Gamma_e^{ij} \overline{L}_L^i \phi e_R^j + [h.c.] \tag{1.8}$$

here *h.c.* stands for hermitian conjugate of all the previous terms, $\overline{Q}_L^i$ and $\overline{L}_L^i$ are the leptons and quarks LH doublets, u_R^j, d_R^j and e_R^j are the RH leptons and quarks singlets, $\Gamma_{u,d,e}^{ij}$ are the 3×3 complex Yukawa matrices, ϕ is the Higgs doublet. If we now consider a Higgs doublet as in 1.6 in the Yukawa Lagrangian, the SSB mechanism makes the mass terms for fermions appear as:

$$m_f = \frac{y_f v}{\sqrt{2}} \tag{1.9}$$

where y_f are the Yukawa couplings, that can be obtained diagonalizing the Γ matrices. These constants set not only the mass of the fermion but also its coupling to the Higgs boson. The top quark is the fermion with the largest mass as well as the strongest coupling to the Higgs boson, with $y_t \approx 1$.

In the SM neutrinos are assumed to be massless. Anyway it is possible to give them masses with a mechanism similar to that of other fermions. Experiments show that neutrinos should have tiny masses compared to other fermions, suggesting that a different mechanism might be responsible for their masses.

The complex Γ matrices in the Yukawa Lagrangian (see eq. 1.8) do not need to be diagonal, and therefore allow different fermion generations to mix among themselves. The Cabibbo-Kobayashi-Maskawa (CKM) matrix is a 3×3 unitary matrix that describes the mixing between the weak eigenstates of the negatively charged quarks (*d', s', b'*) and the corresponding mass eigenstates (*d, s, b*):

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.10}$$

The CKM matrix coefficients shown in equation 1.10 are the coupling strengths of the W boson with the physical up- and down-type quarks, which create flavour-changing charged currents. The diagonal matrix elements are experimentally found to be close to unity. In particular, the World average value for V_{tb} is $|V_{tb}| = 1.010 \pm 0.027$[1]. As a consequence the top quark almost always couples to the bottom quark, both in production and in decay.

1.1.3 Quantum Chromodynamics

Quantum Chromodynamics (QCD)[14] is a gauge theory based on the $SU(3)_c$ group, with the three colour charges *r, g* and *b*. It regulates the strong interactions between partons (quarks and gluons) via the non-Abelian $SU(3)_c$ group. The interaction mediators are eight massless gluons which each carry a colour and an anti-colour charge. The QCD Lagrangian can be written as composed of three parts:

$$\mathcal{L}_{QCD} = \mathcal{L}_G + \mathcal{L}_q + \mathcal{L}_{int} \tag{1.11}$$

here the first term $\mathcal{L}_G = -\frac{1}{4} F_A^{\alpha\beta} F_{\alpha\beta}^A$ is the kinetic part for the gluon field, with $F_A^{\alpha\beta} = \partial^\alpha A_A^\beta - \partial^\beta A_A^\beta + g_s f_{ABC} A_B^\alpha A_C^\beta$, where A_A^μ are the eight gluons ($A, ..., H = 1, ..., 8$) and f_{ABC} is the structure constant of the $SU(3)_c$ group. The second term of equation

1.11 is the kinetic term for the massive quark fields $\mathcal{L}_q = \overline{\psi}_a(i\gamma_\mu\delta_{ab}\partial^\mu - M_a\delta_{ab})\psi_b$, here ψ_a is the quark field in the fundamental representation ($a, b = 1, 2, 3$). Finally, the last term of equation 1.11 contains the interaction among quarks and gluons $\mathcal{L}_{int} = -\overline{\psi}_a(g_s\gamma^\mu(t^A)_{ab}A^\mu_a)\psi_b$, here $t^A = \frac{1}{2}\lambda^A$ are the group generators proportional to the Gell-Mann matrices.

The strong coupling constant g_s varies with the energy scale of the interaction (Q^2), decreasing at higher momentum transfers and increasing at lower ones. As a consequence of the running of the strong coupling constant, the QCD theory can be divided into two main regimes: the low distance one, where the theory of the strong force can be interpreted as a perturbative theory, referred to as *perturbative QCD*, and the large distance case, where g_s becomes large and its effects cannot be treated in perturbation theory anymore, the *strong QCD* regime. The running of the g_s coupling can also be interpreted through its two main effects: the so-called *asymptotic freedom* allows quarks to act nearly as free particles at small distances. On the other hand, at large distances quarks are bound together in colourless states. This phenomenon is referred to as *colour confinement*: attempting to separate quarks makes a new quark-antiquark pair with opposite colour charges appear from the vacuum, forming new bound states. Quarks cannot be found isolated, while they exist in colourless bound states called *hadrons*, which can be made of a quark and an antiquark (*mesons*) or of three quarks (*baryons*). An important experimental consequence of this *colour confinement* effect is that quarks in high-energy particle interactions are seen as collimated streams of hadrons, referred to as *jets*. The properties of a jet, including its energy and direction, reflect those of the originating quark. The process through which the quark evolves into a jet is called *fragmentation*. It can be divided into a first phase, the *parton shower*, in which QCD effects can be described using perturbative QCD, and a second one, where strong QCD acts, forming new hadrons, the so-called *hadronization*.

High-energy proton collisions can be described through parton interactions weighted by the Parton Distribution Functions (PDFs). The PDF $f_{a/A}(x/Q^2)$ describes the probability density that the parton a in the proton A carries a fraction of the total proton momentum $x = p_a/p_A$. These PDFs depend on the specific hadron (proton, neutron or heavier nuclei), while they are independent of the hard process between hadrons. They can not be calculated analytically but must be extracted from fits on various experimental data sets of different processes like deep inelastic scattering, Drell-Yann events and jet production.

1.2 Top quark physics

The existence of the top quark was suggested after the discovery of the bottom quark b, its weak isospin counterpart in 1977. It was then discovered in 1995 by the CDF and DØ experiments using $p\bar{p}$ collision data at the energy in the centre-of-mass $\sqrt{s} = 1.8$ TeV at the Fermilab Tevatron collider [15, 16]. After that, in 2010 both the ATLAS and CMS collaborations measured the top-quark pair production using pp collision data at $\sqrt{s} = 7$ TeV at the Large Hadron Collider (LHC) [17, 18].

The work of this thesis focuses on the top quark, from different points of view. The interest in studying physics processes related to such particle is inspired by its unique characteristics: it is an up-type quark with an electric charge of $+2/3$; it is the heaviest known fundamental particle, with a mass of 172.57 ± 0.29 GeV [1]; it has the strongest coupling among all SM particles to the Higgs boson, as a consequence of its heavy mass; it decays before the hadronization process begins, thanks to its width $\Gamma_t = 1.42^{0.19}_{-0.15}$ GeV [1], and it is therefore the only quark that can be observed almost

as a free particle. As explained in Section 1.1.2, the top quark almost always decays into a W boson and b-quark. This property can be used as a clear experimental signature to identify top quark events based on its decay products.

The large mass of the top quark significantly influences the electroweak symmetry breaking process, making it a sensitive probe for physics beyond the Standard Model (BSM). For instance, new physics could introduce additional top-quark production or decay mechanisms. Various extensions of the Standard Model propose particles that strongly couple to the top quark, aiming to address unresolved puzzles such as the substantial gap between the scales of electroweak interactions, typically around 100 GeV, and gravitational interactions at the Planck scale, which is approximately 10^{19} GeV.

1.2.1 Top quark production

The top quark can be produced at hadron colliders, such as the LHC, in two main ways: via strong interaction as top-antitop pairs ($t\bar{t}$), and via electroweak interaction as a single top or antitop quark.

Top quark pair ($t\bar{t}$) **production** is the main production mode of top quarks at hadron colliders. At leading order (LO) in perturbation theory there are two distinct sub-processes: quark-antiquark ($q\bar{q}$) annihilation and gluon-gluon (gg) fusion (their respective Feynman diagrams are illustrated in Figure 1.2 and 1.3. Near the kinematic production threshold, $q\bar{q}$ annihilation predominates when the incoming quarks are valence quarks, as in $p\bar{p}$ collisions at the Tevatron. Differently, at higher energies gg fusion becomes the dominant mechanism, both for $p\bar{p}$ and pp collisions, accounting for approximately 80% of the $t\bar{t}$ cross-section at the LHC's center-of-mass energy $\sqrt{s} = 13$ TeV. Next-to-leading order (NLO) [19], and Next-to-next leading order (NNLO) QCD computations [20, 21] incorporate associated quark production, gluon bremsstrahlung, and virtual adjustments to the LO processes. The calculation at NNLO order has been refined with Next-to-next-leading logarithm (NNLL) [22] computations, which systematically add higher order corrections coming from soft gluon radiation. Currently, computations at NNLO precision in perturbative QCD, NNLL soft-gluon resummation mentioned above are implemented in the Top++2.0 programm [23] to calculate the SM $t\bar{t}$ cross-section.

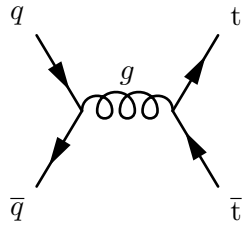


Figure 1.2: LO Feyn. diagram for $t\bar{t}$ via $q\bar{q}$ annihilation

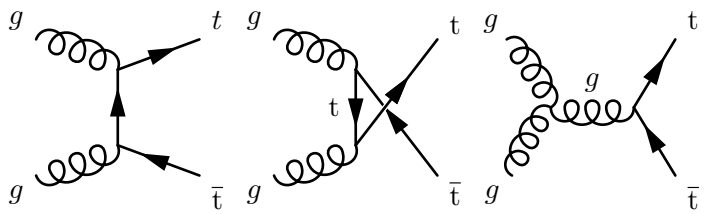


Figure 1.3: LO Feynman diagrams for the top quark pair production via gluon-gluon fusion process

Single top quark production is mediated by the electroweak interaction in the SM, and it involves a Wtb vertex. It has first been observed in 2009 at the Tevatron [4, 5]. Relative to $t\bar{t}$ production, the signal from single top-quark production is quite small and challenging to distinguish from background processes (such as top-quark pair production itself) often resulting in less precise cross-section measurements. Despite being mediated by the weak interaction, the production cross-section of a single top-quark is within an order of magnitude of $t\bar{t}$ production. This relative

abundance comes from the proton composition, richer in bottom quarks and gluons at the lower energy necessary for producing a single top quark ($\approx$ 200 GeV) versus two of them ($\approx$ 400 GeV) [24]. Accurate single top-quark cross-section measurements are sensitive to new physics affecting the production or decay vertices [25]. In the SM the single top-quark production cross-section is directly proportional to the square of the Cabibbo-Kobayashi-Maskawa (CKM) matrix element V_{tb}. At the LHC single top quark can be produced via three main modes:

- ***t*-channel** production, that involves a W boson being exchanged between a light-quark line and a heavy-quark line, and is the primary method for generating single top quarks at the LHC;
- ***s*-channel** production, that is the production and then decay of an off-shell W boson, and it has a minor contribution to single top quark production at the LHC;
- tW **mode** (or W-associated production), is the production in the final state of a top quark (antiquark) and of a W boson.

The main Feynman diagrams for these three production modes are shown in Figure 1.4. Accurate determinations of the cross-sections for the three production modes can

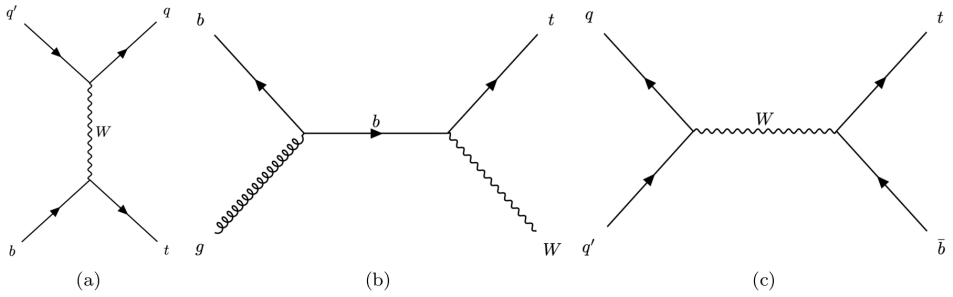


Figure 1.4: Representative diagrams for electroweak single top-quark production in the (a) t-channel, (b) W-associated production (tW) and (c) s-channel.

significantly influence PDF constraints. The three channels complement each other and $t\bar{t}$ production. For instance, the t-channel and tW cross-sections are sensitive to the b-quark PDF and at the same time are anti-correlated to the W- and Z-boson production cross-sections. Meanwhile, the s-channel can be used as a control process, being non-sensitive to the b-quark PDF but correlated to the W/Z production cross-section.

As for $t\bar{t}$ production, NNLO predictions are available for the t-channel single top production cross-section [26] and have been computed with MCFM program [27]. The tW production cross-section is available for the sum of top and anti-top quarks production and was computed at NLO in QCD with the addition of third-order corrections of soft-gluon resummation at NNLL [28]. Predictions at NLO have been calculated for the s-channel production mode [29], including NNLL corrections [30].

1.2.2 Top quark decay

Top quark decay can be categorized based on the final-state particle types. Despite the top quark (antiquark) decaying almost 100% of the time into a W^- (W^+) boson and a b quark (antiquark), the W boson can have different final states after the decay. In 1/3 of the occasions the W is decaying to a charged lepton and a neutrino, while 2/3 of the times it decays to a quark-antiquark pair. Top quark decays with a b

quark, a charged lepton and a neutrino in the final state are called *leptonic* decays, while decays where a b quark plus a quark-antiquark pair are found in the final state are referred to as *hadronic*. When the W boson decays into a charged lepton and neutrino, each of the three generations of leptons has the same probability, while when it decays hadronically the probability of decaying to a certain quark depends on the CKM matrix element values, and therefore decays to b quarks are suppressed as $|V_{cb}| = (41.1 \pm 1.2) \cdot 10^{-3}$.

Top quark pair events are usually divided into three categories based on the decay mode:

- *dileptonic* (or *fully leptonic*) decays, in case both top and anti-top decay leptonically. It has a probability of $\sim 1/9$, and it has in the final state two leptons of opposite charge, a neutrino and an anti-neutrino, a b quark and a $\bar{b}$ anti-quark;

- *semileptonic* decays, that have only one of the two tops decaying leptonically. The probability is of $\sim 4/9$, and the final state is composed of one charged lepton, one neutrino, a quark−anti-quark pair, a b quark and a $\bar{b}$ anti-quark;

- *fully hadronic* decays, in which case both top and anti-top decay to quark-antiquark pairs. This decay mode has a probability of $\sim 4/9$, and two quark−antiquark pairs, a b quark and a $\bar{b}$ antiquark are found in the final state.

1.3 Beyond the Standard Model

Currently, the Standard Model (SM) stands as the most precise framework for describing the experimental phenomenology observed in particle physics. Its predictions have been validated with remarkable precision over recent decades through experiments at particle colliders. Nevertheless, the SM fails to account for certain physical phenomena, underlying its limitations as a comprehensive theory of particle physics. Consequently, ongoing research in particle physics is directed toward looking for evidence of phenomena that lie Beyond the Standard Model (BSM). Here are listed some of the major open items, that the SM is not able to fully explain:

- *Matter-antimatter asymmetry*: despite the SM prediction of an equal amount of matter and antimatter in our Universe, we observe a great unbalance between the two. Present cosmological models suggest that in the early Universe, particles, antiparticles and radiation were in continuous interaction via particle creation and annihilation. As the Universe expanded, a critical temperature was reached and particle creation in thermal equilibrium stopped. According to the SM, if particles were produced in equal quantities of matter and antimatter, the complete annihilation of matter with antimatter would be expected, resulting solely in the release of energy. Numerous theories attempting to explain this phenomenon exist, but the problem still remains unsolved.

- *Dark matter (DM)*: there are several experimental observations, such that galaxy rotation curve [31], gravitational lensing observations [32] and precision measurements of the cosmic microwave background (CMB) anisotropies [33], that indicates the existence of DM. By dark matter we mean a type of matter that interacts gravitationally but not electromagnetically, which is not included in the SM. Dark matter candidates have been theorized, such as Weakly Interacting Massive Particles (WIMPs), supersymmetric particles, and primordial black holes but no experimental evidence has yet been found.

- *Gravity*: this interaction, despite being fundamental in the world we live in, is not part of the SM. Theoretically, gravity should have a corresponding particle carrier, often referred to as the graviton, like the other three fundamental interactions. Currently, the scientific community is intensively pursuing a theory that would merge General Relativity with Quantum Field Theory, some examples are approaches with non-local or loop quantum gravity models [34] and theories like supergravity [35]. There is anyway no model, unifying SM and general relativity that can be tested with current experiments, and the problem stays open.

Another issue which affects every attempt to create new BSM models is the *Hierarchy problem*. Particle masses in the SM depend on the energy scale considered and undergo loop corrections. A lot of fine-tuning is needed to maintain the experimental value of the Higgs boson mass in the SM theory at the energy explored in modern particle experiments. Such high-precision fine-tuning is difficult to justify from a theoretical point of view and it is often described as "unnatural".

At this point in time, it is evident that the SM is a great framework, describing extremely well fundamental particle interactions, and its predictions have been verified by hundreds of measurements performed at the LHC. On the other hand, it is also evident that it cannot be a complete theory that describes the whole fundamental particle world. For this reason, we continue to test SM predictions, as well as to look for new phenomena predicted by BSM, seeking any deviation in data from the SM that would give us a hint to understand new physics. Such a program can be exploited in two main ways:

- **Precision measurement** of SM processes, which test the SM predictions over a vast range of energies and can take into account the possible contribution of new physics effects to those SM processes. The presence of new physics would therefore be seen as significant deviations from the SM predictions in certain sensitive observables.

- **Direct search** for new particles predicted by new physics models in not-yet-explored regions of the parameter phase-space. One of the most common ways to search for new particles is to look for resonances in the invariant mass spectrum of a system of final state particles. As the position of the peak is directly related to the mass of the produced new particle, a localized excess in data over the SM prediction would allow to immediately characterize the signal.

In the next Section, some specific new physics models that can be tested at the LHC through direct searches are discussed. In particular new particles decaying to a top quark pair are discussed, as they are the theoretical background to the search for new physics presented in Chapter 5.

1.3.1 Heavy resonances decaying to a top quark pair

The relevance of the top quark in the search for new physics, and in particular the search of heavy resonances decaying to a top quark pair, was already discussed in Section 1.2. There exist various new physics models predicting different kinds of new particles decaying to $t\bar{t}$, among which a set of them considers s-channel resonances. Three specific new particles that can be created according to these kinds of models are detailed in the following:

- **Spin-1 colour singlet**, electrically neutral vector boson, usually referred to as $\mathbf{Z'}$, predicted by top-colour-assisted Technicolor models [36], as well as other BSM theories. In particular, one of these models, referred to as *Leptophobic Non-Standard Topcolor* Z' (Model IV) in [7], predicts a Z' mainly produced by $q\bar{q}$ annihilation that decays into first and third-generation quarks. In the model, f_1 and f_2 are free parameters related to the coupling of the Z' to the up-type and down-type quark, while the parameter $cot(\theta_H)$ controls the width and the production cross-section. In the analysis detailed in Chapter 5, the parameters are fixed $f_1 = 1$ and $f_2 = 0$ to maximize the branching fraction of the heavy boson to $t\bar{t}$, which reaches 33%. The parameter $cot(\theta_H)$ is tuned for each mass value considered so that the width of the resonance is 1% of its mass. This model has negligible interference with the SM $t\bar{t}$ production process.

- **Spin-2 colour singlet**, predicted in models that consider extra dimensions of space, which cause Kaluza-Klein (KK) excitations of the graviton like the so-called Randall–Sundrum (RS) models [37]. A model often considered is a RS one with a single warped extra dimension that has one warped extra dimension where the SM fields are in the bulk [8]. The RS bulk graviton is usually indicated as G_{KK}. For the dimensionless coupling of the KK gravitons to the SM fields, $8\pi\frac{k}{M_{PI}}$ (here k is the curvature of the warped extra dimension and M_{PI} is the Planck mass) values ≈ 1 are favoured. Choosing the value 1 for this parameter the main production mode is gluon-gluon fusion, and in the mass range between 0.4 TeV and 3 TeV, the width stays between 3 and 6%. The branching ratio to $t\bar{t}$ is around 18% at the mass of 400 GeV and it increases rapidly up to 68% at 1 TeV, where it becomes more stable.

- **Spin-1 colour octet**, predicted in models with extra dimensions of space which creates KK excitations of the gluon [9]. In an RS model with a single warped extra dimension, the KK excitation called g_{KK}, is mainly produced via $q\bar{q}$ annihilation. The couplings of the model can take different values. The values chosen in the study described in Chapter 5 are for the strong coupling of the KK gluon to the SM gluon $g_q = -0.2g_s$ (here g_s is the SM gluon coupling), and the left-handed (LH) coupling to the top quark $g_{L,t} = g_s$, and then optimize the right-handed (RH) one $g_{R,t}$ so that the width of the resonance is $\approx 30\%$.

In figure 1.5 the most relevant s-channel production Feynman diagram at Leading Order (LO) is shown for each of the three new physics particles described.

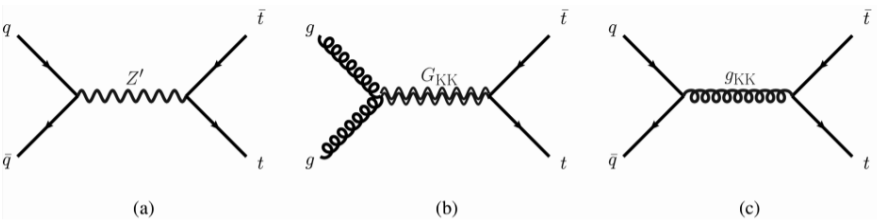


Figure 1.5: Main LO s-channel production diagram at for a Z' boson in (a), a KK graviton in (b) and a KK gluon in (c).

Chapter 2

ATLAS at the Large Hadron Collider

Recent advancements in high-energy physics have driven the need for experiments operating at higher energies and producing larger datasets to achieve greater precision. At the CERN accelerator complex, the Large Hadron Collider (LHC) accelerates proton and heavy-ion beams to relativistic speeds, reaching a centre-of-mass energy of up to $\sqrt{s} = 13.6$ TeV, allowing the study of fundamental constituents of matter at energy scales never investigated before. This chapter introduces the LHC and the ATLAS detector in detail, while a brief introduction to their upgrades is given.

2.1 The Large Hadron Collider (LHC) and High Luminosity LHC

The Large Hadron Collider (LHC), the world's largest particle collider [3], is located in the Geneva region, between France and Switzerland. Its underground tunnel, 100 meters below the surface, spans a circumference of 27.6 km and was previously occupied by the Large Electron-Positron (LEP) Collider. The LHC started its operation in 2010, leading after just two years to the discovery of the Higgs boson [38, 39]. Most of the time the LHC accelerates two counter-rotating proton beams, achieving a peak instantaneous luminosity of a few $\sim 2 \cdot 10^{34}\text{cm}^2\text{s}^{-1}$. It can also accelerate heavy-ion (Pb) beams up to $\sqrt{s} = 5.5$ TeV. The underground tunnel contains a complex system of superconducting magnets cooled with super-fluid helium, generating magnetic fields of 8 T required to bend the beams into their circular shape and to keep them focused. The LHC operates as the last stage of the CERN accelerator complex, accelerating the beams to their highest reachable energies and colliding them in four experimental points, as depicted in Figure 2.1.

Figure 2.1 shows the four main experiments located in the LHC tunnel: ATLAS, CMS, ALICE, and LHCb, situated at the collision points of the beams. ATLAS (A Toroidal LHC ApparatuS) [2] is a multi-purpose detector aimed at exploring a broad range of phenomena in *pp* collisions, to both precisely measure SM processes and to search for BSM physics. The CMS (Compact Muon Solenoid) experiment [41] has a similar purpose as ATLAS, but the detector utilizes different technologies and materials, and is slightly smaller. ALICE (A Large Ion Collider Experiment) [42] primarily focuses on analysing heavy ion collisions, to study strong interactions under high energy density conditions. It also analyzes *pp* collisions to provide baseline data for heavy-ion studies. LHCb (Large Hadron Collider beauty) [43] is dedicated to heavy-flavour physics, particularly studying hadrons containing bottom and charm

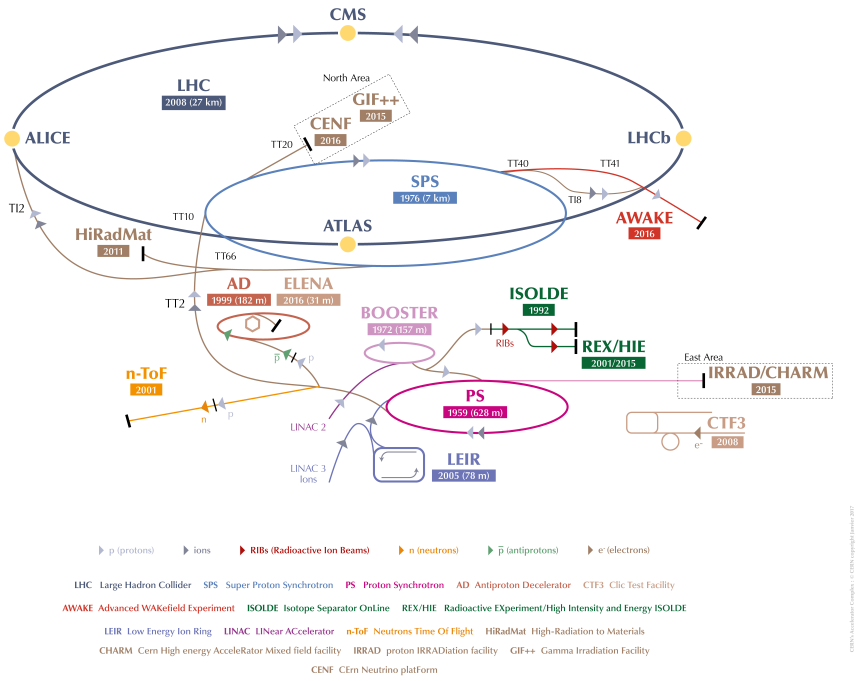


Figure 2.1: Schematic view of the CERN accelerator complex[40].

quarks and probing new potential sources of CP violation.

2.1.1 The LHC operations

The LHC operations are periodically interrupted every few years to perform upgrades to both the collider machine and the experiments. It first collided proton beams between 2009 and 2013 in the so-called *Run 1* at energies in the centre-of-mass of $\sqrt{s} = 7$ and 8 TeV, it then restarted operations in *Run 2* between 2015 and 2018 at $\sqrt{s} = 13$ TeV, and it started the *Run 3* in 2022, which is still ongoing, running at an energy in the centre-of-mass $\sqrt{s} = 13.6$ TeV.

Apart from the collision energy, the so-called luminosity is a key parameter to characterize the performance of a particle collider. A high luminosity is essential to perform precision SM measurements and to search for rare BSM processes. The instantaneous luminosity of a particles collider is defined as $L = {}^{1}/_{\sigma}{}^{dN}/_{dt}$, where σ is the process cross-section and ${}^{dN}/_{dt}$ is the event rate per second. For two colliding beams with a Gaussian profile, this luminosity is expressed by equation 2.1

$$L = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \tag{2.1}$$

here N_1 and N_2 are the particle counts in each bunch, N_b is the number of bunches, f is the collision rate, and σ_x and σ_y are the linear dimensions of the bunches. The bunches have approximately $1.2 - 1.4 \cdot 10^{11}$ protons each and are spaced 25 ns apart, corresponding to a frequency of 40 MHz. The LHC is designed to contain 2808 bunches. The integrated luminosity $\int L$ is defined as the instantaneous luminosity collected over a period of time.

While certainly beneficial for the physics programme of a particle collider, high luminosity levels produce an important experimental complication, commonly re-

ferred to as *pile-up*. This phenomenon is related to the high number of interacting particles in each bunch crossing. This causes multiple inelastic scattering events other than the main one to be studied. In the ATLAS experiment the average pile-up level registered during the so-called *Run 1*, quantified by $\langle\mu\rangle$, defined here as the average number of collisions per bunch crossing, ranged between 9 and 21. During *Run 2* values $13 << \mu >< 38$ have been registered further increasing in the course of *Run 3* until reaching around 60 interactions per bunch crossing with the current LHC conditions.

2.1.2 The High-Luminosity LHC

The LHC is currently the most powerful particle collider in the World and it will remain so for at least another decade. It represents one of the most sophisticated machines ever built by men and holds great scientific potential. This is why already in 2006 CERN priority was to "*fully exploit the physics potential of the LHC*" [44] and it was later confirmed repeatedly in the European Strategy for particle physics. As a consequence in 2010, CERN created the High Luminosity LHC (HL-LHC) project, which was later approved by the CERN Council between 2013 and 2014 [45].

The main goals of the HL-LHC project are to upgrade and optimize the LHC machine so that it will be able to reach a peak luminosity of $5 \cdot 10^{34}$ cm$^{-2}s^{-1}$ and an integrated luminosity of 250 fb^{-1} per year with the goal of collecting a total luminosity of 3000 fb^{-1}. The increase in instantaneous luminosity comes with an increase in pile-up, which is expected to reach $< \mu >\sim 140 - 200$. A schematic timeline of the LHC and HL-LHC upgrade is shown in Figure 2.2. The physics

Figure 2.2: LHC/HL-LHC plan and schedule as in 2022 [46].

program at the LHC, currently running at $\sqrt{s} = 13.6$ TeV, has been giving us a broader picture of physics at the TeV scale, which continues to be updated. Some of the fundamental inputs to better understand the future of the energy-frontier physics are extremely precise SM measurements, in particular in the context of the recently discovered Higgs boson and its self-interaction, and searches for new physics, especially for extremely rare new processes. Such findings can only be fully achieved with the luminosity collected with the HL-LHC program. On the other hand, the numbers associated with the upgraded characteristics of the HL-LHC pose important challenges to all the experiments along the ring. Various upgrades to the four main experiments are foreseen in the next years, especially during the long shutdown of the LHC before the start of the HL-LHC operation. These upgrades are needed

both to have detectors resistant to the new challenging physics environment (i.e. the increased radiation produced by the colliding beams), as well as to fully exploit the physics possibilities of the HL-LHC program.

2.2 The ATLAS Detector and its Upgrade

Out of the four main experiments at the LHC, the ATLAS detector is the biggest one with a height of 25 m and a length of 44 m. It is centred around the beam interaction point with a cylindrical structure that covers almost the full solid angle, as shown in Figure 2.3. ATLAS (like CMS) is a general-purpose detector, that looks at a great

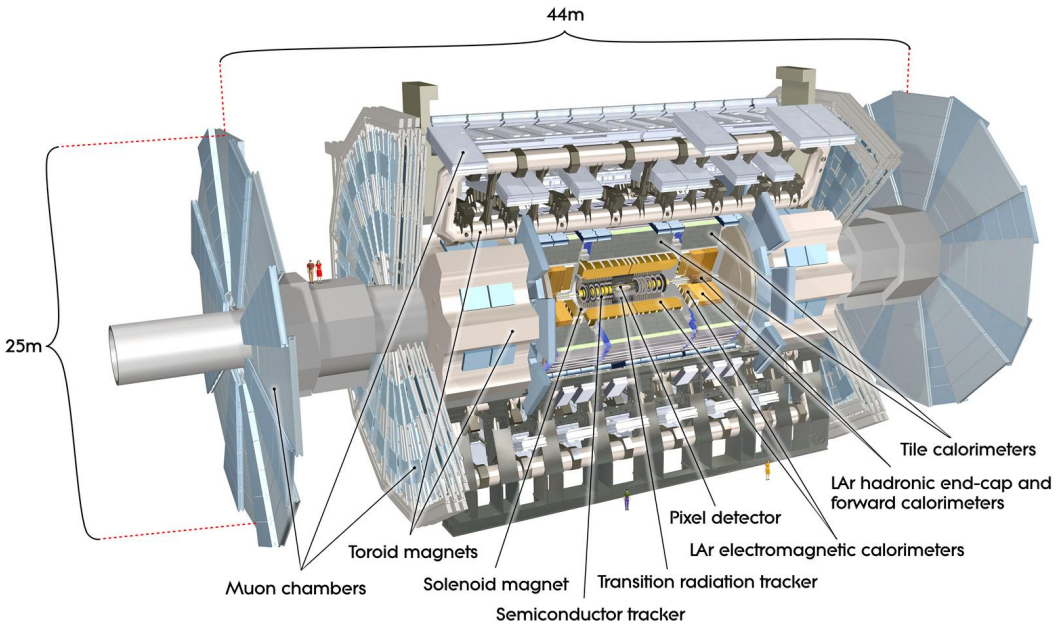


Figure 2.3: The ATLAS detector [2].

variety of SM and BSM processes. In order to study them properly with collision produced by the LHC, several requirements are needed by the experiment: high detector granularity, high-level calorimeters (both electromagnetic and hadronic), precise muon identification, fast electronics, resistance to radiation, as well as an efficient and fast trigger system.

In the ATLAS experiment, a standard coordinate system is defined, which is an orthonormal right-handed one. The origin is at the beam interaction point, the z axis is parallel to the beam direction, the x axis is perpendicular to the beam, pointing towards the centre of the LHC, and the y axis points upwards. The plane $x-y$ plane, transverse to the beam direction, is commonly defined as the *transverse plane*.

Particles measured in the ATLAS detector can be described as a function of a few variables:

- azimuthal angles, defined as the angle in the transverse plane $(x-y)$;
- the polar angle θ measured in the $z-x$ plane;
- the pseudo-rapidity $\eta = -ln(\theta/2)$, which is Lorentz-boost invariant for massless particles;

- the angular distance between two particles can be described as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The ATLAS detector is an ensemble of various subdetector systems, which are all made up of two main sections: a cylindrical one, along the beam axis, and the so-called endcaps, that close the cylinder perpendicularly to the beam axis. Closest to the beam pipe the Inner Detector (ID) is essential for charged particle trajectory reconstruction. Around it, there are two calorimeter systems: the closer to the beam pipe is the electromagnetic (EM) calorimeter, used to reconstruct electrons and photons and to measure their energy, while the outer one is the hadronic (HAD) calorimeter, necessary to stop and measure the energy of hadrons. The outermost subdetector is the Muon Spectrometer (MS), used to precisely identify muons, the only charged particles to reach this layer of the detector. Two magnet systems are part of the ATLAS detector: the smallest one is a solenoid, that is placed between the ID and the EM calorimeter, generating a field of 2 T; the second one is a complex system of 8 magnets, placed outside the MS, creating a toroidal magnetic field.

For this complex machine to work effectively and collect useful data for physics studies, a big team of ATLAS scientists takes care of its operation, together with engineers and technicians. The main part of this work is carried out at *Point 1*, one of the LHC beam interaction points, where the ATLAS experiment is located. A fundamental system, needed to collect meaningful data from the physics research point of view is the ATLAS trigger system. It is needed to drastically reduce the enormous rate of data generated by collisions and only record the *interesting* ones.

In the next sections, the various ATLAS subsystems are presented.

2.2.1 The Inner Detector

The ID is composed of three parts: the Pixel detector, the SemiConductor Tracker (SCT), and the Transition Radiation Tracker (TRT), as shown in Figure 2.4. The total length of the whole ID cylinder is 6.2 m with a radius of 1.15m. It is fundamental for charged particle identification through track reconstruction, and it is designed to have a momentum resolution of $\frac{\sigma(p_T)}{p_T} = 0.05\%\ p_T$. It can measure with high precision the direction, the p_T and the sign of charged particles emerging from the interaction region, allowing the primary and secondary vertex reconstruction. The ID covers the full ϕ angle and the pseudo-rapidity region up to $|\eta| < 2.5$.

Pixel Detector

Placed around the beryllium beam piped, the Pixel detector is made of silicon pixels, most of which are $50 \times 400\ \mu\text{m}^2$ in size and 250 μm thick. At the core of each pixel is a reverse-biased p-n junction connected to a readout channel, capable of detecting charged particles passing through it. The interaction of a charged particle with this sensible region generates multiple electron-hole pairs, which produce an electrical signal as they move toward their respective electrodes. Pixels are grouped together in modules, which are arranged in three layers in the barrel and three layers for each of the two end-caps. The three barrel layers are placed at radii of 50, 89 and 123 nm, while for each side the three end-cap disks are placed between $|z| = 495$ mm and 650 mm away from the centre.

In between *Run 1* and *Run 2*, during what is called the *Long Shutdown 1* (LS1), another layer was added, to face the increase in luminosity. It is called "Insertable B Layer" (IBL) and reaches an intrinsic resolution of 10 μm in the $x - y$ plane and 66.5 μm in the z plane. The integration of this system into ATLAS enhances track and vertex reconstruction efficiencies, in particular boosting the efficiency of b-hadron

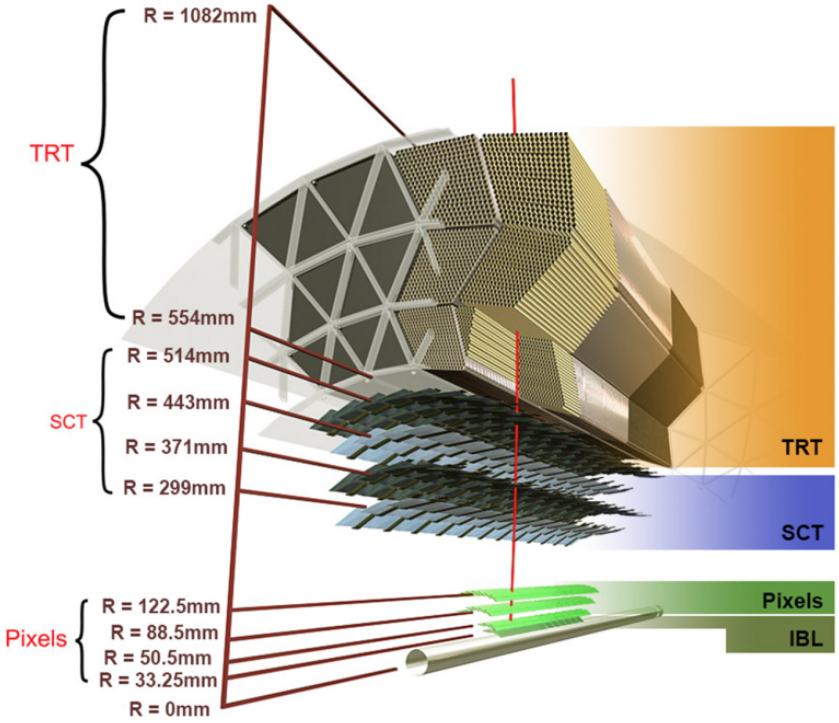


Figure 2.4: Schematic view of the different layers of the Inner Detector.

associated jets. On average, with the inclusion of the IBL, the Pixel detector records four hits per track, both in Monte Carlo (MC) simulation and in observed data [47].

SemiConductor Tracker

This second part of the ID is made of silicon micro-strips with a pitch of 80 μm, grouped in modules. It is made of four layers in the barrel with radii in between $R = 299$ mm and 514 mm, while there are nine disks for each end-cap at a distance from the centre between ~ 0.9 m and ~ 2.7 m. In the barrel, the strips are aligned parallel to the beam axis, with a uniform pitch of 80 μm. In the end-caps, the strips are oriented radially and have a variable pitch. Particles coming from the beam-interaction region create typically four space points in the SCT.

The Pixel detector and the SCT are refrigerated and maintained at a temperature of -7°C. The cooling is necessary to reduce the radiation damage on the silicon.

Transition Radiation Tracker

The TRT is a gaseous detector made of drift tubes alternated with transition radiation material. The straw tubes are 144 cm (37 cm) long in the barrel (end-cap) region, and they have a radius of 2 mm. The gas used is a mixture in *Run 2* of 70% Xe, 27% CO_2 and 3%O_2, and an over-pressure between 5 and 10 mbar. Each tube has a wire (the anode) with a potential difference with respect to the straw shell (the cathode). They collect ions created by particles passing through the gas. The TRT has a spatial resolution of 130 μm for particles with $|\eta| < 2$ and delivers an almost continuous tracking, with approximately 36 hits per track.

This part of the detector also enables particle identification by using tubes interspersed with layers of polypropylene fibres and foils. As a charged particle crosses the

boundary between materials with differing refractive indices, it emits X-ray radiation proportionally to its relativistic factor. With two threshold levels set by the discriminator, the TRT can distinguish electrons from hadrons by comparing the ratio of high threshold hits to total hits. This sub-detector operates at room temperature and is therefore insulated.

2.2.2 The Calorimeters

The ATLAS calorimeter system gives information on the energy and position of particles. It is composed of two types of calorimeters: the Liquid Argon (LAr) calorimeter and the Tile calorimeter. The structure of this system can be divided into three parts: the Electromagnetic Calorimeter (ECAL), the Hadronic Calorimeter (HCAL) and the Forward Calorimeter (FCAL), as shown in Figure 2.5. All calorimeters are sampling ones, with different active and passive materials. They are structured so that particles create *showers* while passing through them: a particle interacting with a layer of passive material creates a cascade of energetic particles that are detected by the following layer, made of active material, and each of these secondary particles creates a new particle cascade when encountering the next passive layer. The calorimeter system covers a range of pseudo-rapidity up to $|\eta| < 4.9$ and contains all electromagnetic and hadronic showers.

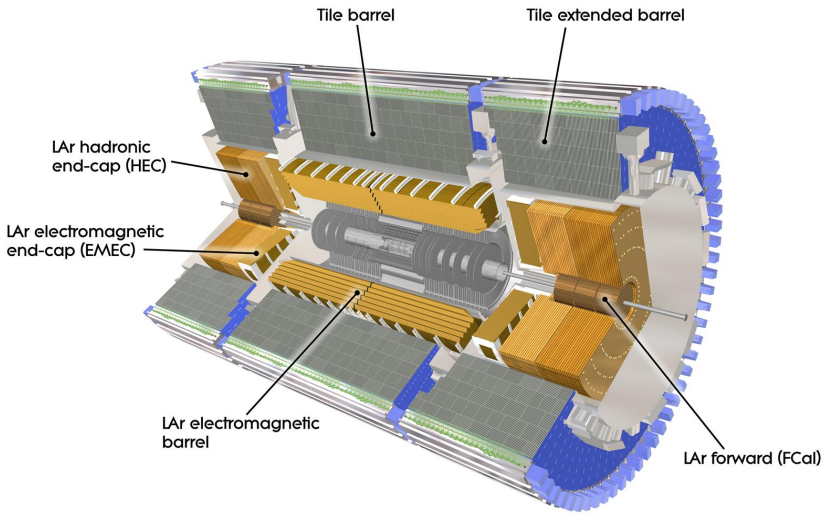


Figure 2.5: Schematic view of the ATLAS Calorimeter System [2].

ECAL

The electromagnetic calorimeter's main purpose is the precision measurements of the energy of electrons and photons, and it is composed of the barrel part and the end-caps. It is designed with an accordion geometry to have full coverage in the azimuthal angle ϕ and a fast extraction of the signal. Liquid Argon (Ar) is the active material used, while Lead is the passive material. Particles with typical LHC energies interact with the ECAL mainly via bremsstrahlung and electron-positron pair production. These kinds of interaction are characterized by the radiation length X_0. Therefore for the ECAL to contain the full EM shower of electrons and photons must have a large radiation length.

The barrel calorimeter is made of two half-barrels which goes to pseudo-rapidity up to $|\eta| < 1.475$. The two halves are separated in the transverse plane by 4 mm and each of them is divided into 16 modules. The thickness of a single module varies from 22 X_0 at $\eta = 0$ to 33 X_0 at $|\eta| = 1.3$. Each end-cap is divided into two wheels, that cover between $1.375 < |\eta| < 3.2$. The inner wheel is thick between 26 X_0 and 36 X_0, while the outer one is between 24 X_0 and 38 X_0. The barrel has an inner diameter of 2.8 m and an outer of 4 m, while the end-caps have an inner radius of 0.3 m and an outer of 2 m. The region between the barrel and the end-caps, with pseudo-rapidity $1.37 < |\eta| < 1.52$, is referred to as EM *crack region* and is characterized by worse performance due to the passive material in front of it.

HCAL

The hadronic calorimeter, together with the forward one, has as its main purpose the reconstruction of hadronic jets, the measurement of their energy, and the determination of the missing transverse energy (E_T^{miss})[1]. The HCAL is divided into the barrel, which is a Tile calorimeter, and the end-caps (Hadronic End-cap Calorimeter (HEC)), which are LAr calorimeters.

The barrel part is split into the central barrel, with $|\eta| < 1$ made of three layers thick $\sim 1.5\ \lambda$, $4.1\ \lambda$ and $1.8\ \lambda$, and two extended barrels that cover $0.8 < |\eta| < 1.7$ with three layers each with a thickness approximately of $1.5\ \lambda$, $2.6\ \lambda$ and $3.3\ \lambda$. The Tile calorimeter uses polystyrene scintillator tiles, with steel as the passive material. Tiles scintillate when crossed by charged particles, producing photons that are then read out by photo-multipliers.

Behind the ECAL end-cap, each HEC has two wheels covering the region $1.5 < |\eta| < 3.2$. The HEC is a LAr calorimeter with copper as a passive material, and it operates similarly to the ECAL LAr calorimeter.

FCAL

The forward calorimeter covers the higher range in pseudo-rapidity, between $3.1 < |\eta| < 4.9$ and it is located in the end-caps. Each end-cap is split into three modules: the first one serves as an electromagnetic calorimeter with the passive material being copper, while the second and third modules have tungsten as passive materials, being optimized to make hadrons shower. The forward calorimeter is particularly close to the beam pipe and receives more radiation, therefore the passive material as well as the geometry are chosen to make it particularly radiation-resistant.

2.2.3 The Muon Spectrometer

The MS is responsible for triggering and measuring muons, by reconstructing their tracks and determining their momentum, thanks to the toroidal magnetic field in which it is immersed, as explained in Section 2.2.4. The p_T of the muons can indeed be reconstructed using the magnetic field information together with three different points of a track. Moreover, in the ATLAS experiment, one can get two independent measurements of the muon transverse momentum using the MS as well as the ID (as in 2.2.1).

The MS is made of four subsystems:

[1]The missing transverse energy is defined in ATLAS as the sum of the energy in the whole detector associated with reconstructed objects in the transverse plane in an event. This definition is based on the idea that the total initial momentum in the transverse plane before the proton colliding is zero and should be conserved after the collision. The E_T^{miss} is often used as an approximated value for the neutrino energy, which cannot be directly detected by the ATLAS detector.

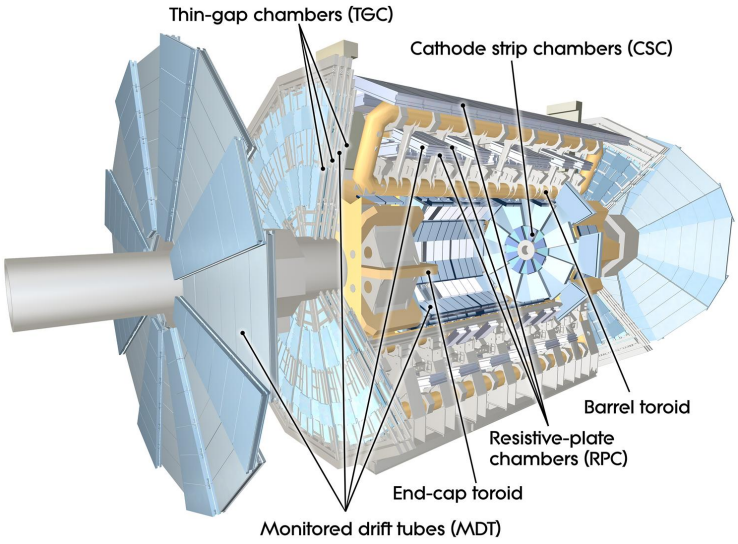


Figure 2.6: Schematic view of the Muon Spectrometer [2].

- **the Monitored Drift Tubes (MDT)**, with layers of drift tubes stacked together to create MDT chambers. The MDT covers the pseudo-rapidity range up to $|\eta| < 2.7$. A muon leaves on average 20 hits in this subsystem. Each tube is made of aluminium and has an inner cable of a gold-plated tungsten-rhenium alloy. The tubes are filled with a mixture of Ar and CO_2;
- **the Cathode Strip Chambers (CSCs)** are multi-wire drift chambers 7 m long, and cover the region between $2. < |\eta| < 2.7$. A muon is expected to leave 4 measurements in the CSC. The anode wires are kept at a potential difference of 1.9 kV and are immersed in a gas mixture with the same gasses of the MDTs but in different percentages;
- **the Resistive Plate Chambers (RPCs)** are made of resistive plates parallel to each other with in between a gas mixture of $C_2H_2F_4$, SF_4 and isobutane. This subsystem is divided into three concentric layers around the beam axis. A muon leaves on average 6 hits in the RPCs. They cover the region up to $|\eta| < 1.05$ and are used to also trigger on muons.
- **the Thin Gap Chambers (TGCs)** is made of two concentric rings that cover the range in pseudo-rapidity between $1.05 < |\eta| < 2.4$, where each muon is expected to leave 9 hits and are used to trigger on the muons. The TGCs are multi-wire proportional chambers with a potential difference between the anode wire and the cathode of 2.9 kV. The chambers are filled with a mixture of CO_2 and n-pentane.

Around the $|\eta| \sim 0$ region there is a small gap in the MS where the services for the ID, solenoid magnet and calorimeters pass through to exit the ATLAS detector.

2.2.4 The Magnet System

The ATLAS Magnet System is composed of two separate subsystems, with the fields generated independently of each other. The two subsystems, shown in Figure 2.7, are the central solenoid and the toroid magnets:

- The central solenoid is embedded between the ID and the calorimeters, and it is a cylinder aligned with the beam axis, It provides an axial magnetic field

of 2 T. At the same time, it is designed to minimize the thickness in terms of radiation lengths in front of the calorimeters.

- The toroid magnets, are divided between the barrel region and the end-caps. The included magnetic fields curve particle trajectories in different regions of the MS: the barrel toroid curves muons in the region $|\eta| < 1.4$, the end-caps toroids in the region $1.6 < |\eta| < 2.7$, and a combination of the two fields for the region in between. The barrel region is 25 m long and produces a magnetic field between 1.5 T×m and 5.5 T×m, while the end-cap toroids are 5 m long and create a field that varies from 1 to 7.5 T×m.

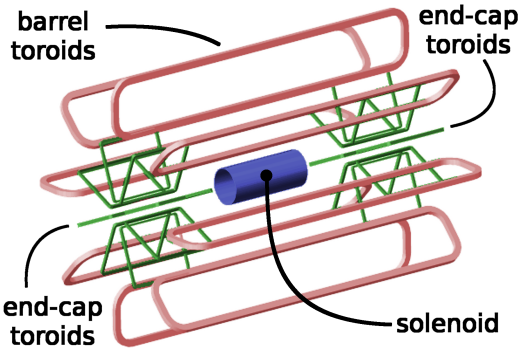


Figure 2.7: Overview of the ATLAS magnet system, composed of the solenoid magnet and the toroid magnets [48].

2.2.5 ATLAS Trigger System

The LHC generates approximately one billion interactions per second, creating a data output of 60 Terabytes per second. Given that only a minimal portion of this data is interesting for physics research, there is a need to use a trigger system to select events. ATLAS implements a dual-tier trigger system that is divided into two parts: first, there is a hardware-based Level 1 (L1) trigger and afterwards, there is a software-based High-Level Trigger (HLT). The general mechanism of the trigger system is schematized in Figure 2.8.

- The L1 trigger processes signals of lower granularity from the calorimeters and muon detectors to identify high-energy objects. It effectively reduces the event rate from a bunch-crossing rate of 40 MHz down to 100 kHz. It collects data on electrons, photons and jets from the calorimeters, such as transverse energies and isolation. For muons, it also considers the coincidence of hits across multiple layers of the Muon Spectrometer (MS). The L1 must deliver its decision to the front-end electronics within 2.5 μs, specifying Regions of Interest (RoI) where potential trigger objects are detected.

- The High-Level Trigger (HLT) is a software trigger. It uses the RoI data produced at the previous level to run a simplified version of the event reconstruction using the complete granularity and precision of the calorimeters and muon chambers. The reconstruction sequence employs various *Processing Units* (PUs), selection applications that take a decision in a few 100 ms. It is designed to reduce the rate of data to 1 kHz with a latency of the order of 10^2 ms.

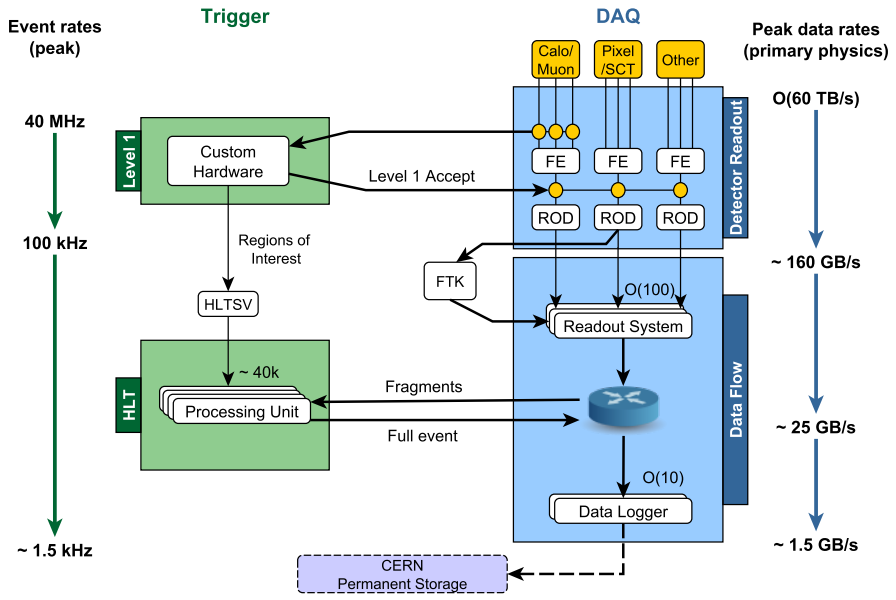


Figure 2.8: Functional diagram of the ATLAS Trigger and Data Acquisition system [49].

Trigger rates can be changed by adjusting thresholds or implementing different selection cuts. Once an event is approved by the L1 trigger, the data are digitized as signals and formatted into raw data before being sent to the Data Acquisition (DAQ) system. Events passing the HLT are archived in permanent storage at the CERN data centre.

2.2.6 Luminosity Detectors

Other than the main structure of the ATLAS detector already described, three smaller systems are situated in the ATLAS forward region [50]. The LUminosity measurement using Cherenkov Integrating Detector (LUCID) [51] and Absolute Luminosity For ATLAS (ALFA) [52] have the goal of measuring precisely the luminosity delivered to the experiment, while the Zero-Degree Calorimeter (ZDC) [53] main purpose is to assess the centrality of heavy-ion collisions. The **LUCID** detector captures inelastic proton-proton scattering in the forward direction. It is located at ± 17 m from the interaction region and is composed of polished aluminium tubes at around 10 cm from the beam pipe, filled with C4 F10 gas. It covers a high pseudo-rapidity range between $5.6 < |\eta| < 6.0$. **ALFA** situated at ± 240 m from the interaction point. It is made of scintillating fibre trackers positioned within the Roman pots, which can approach the beam up to 1 mm. It tries to measure luminosity through elastic scattering at small angles. Finally, the **ZDC** can be found at ± 140 m, featuring layers of quartz rods and tungsten plates to detect neutral particles at $|\eta| < 8.3$ in both proton-proton and heavy-ion collisions.

2.2.7 The Inner Tracker project: ATLAS Upgrade

The ATLAS collaboration is preparing major updates for its experiment in preparation for the HL-LHC upgrade. Both the expected HL-LHC peak luminosity and

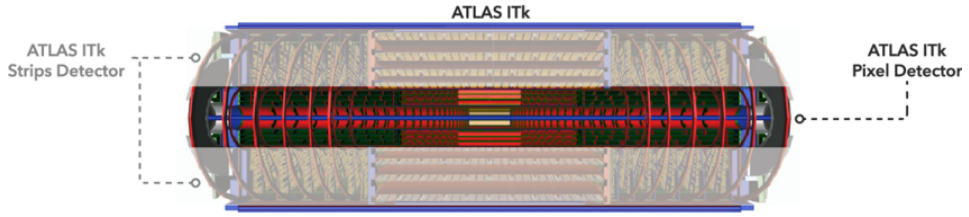


Figure 2.9: Schematic view of the ATLAS ITk Pixel Detector surrounded by the Strip Detector. [56].

pile-up level are indeed expected to be well above those considered for the ATLAS detector design, as already mentioned in 2.1.2. In particular, the radiation levels for the innermost parts of the ID will be 20 times higher than the present ones, up to a non-ionizing radiation fluence of $\phi = 210^{16} n_{eq}$ cm^{-2} over the expected lifetime. The current ID will not be able to handle these conditions, therefore it will be replaced at the end of the current LHC run (*Run 3*) with a new all-silicon detector called Inner Tracker (ITk). This upgraded tracker will be also able to maintain the current tracking performance in the new high-occupancy environment. The ITk will have a new readout scheme which will be part of the upgrade of the trigger system for HL-LHC, providing the possibility to use track-based trigger chains.

The new layout of the ITk detector will be divided into two parts the **Pixel Detector** [54], closer to the beam, will be around twice large and four times longer than the current Pixel detector, and the **Strip Detector** [55] surrounding it, that will cover three times the silicon area of the current detector. The design structure of the ITk system is shown in Figure 2.9. The Pixel detector will be composed of hybrid planar and 3D silicon pixels disposed in five layers, while the Strip detector will be made of silicon micro-strip layers, separated by the Pixel subsystem by a Pixel Support Tube (PST). The two innermost layers of the Pixel Detector will be replaceable and separated by the rest of the ITk through an Inner Support Tube (IST). The total coverage of the ITk will be up to $|\eta| < 4.0$ in pseudo-rapidity.

In order to study and investigate the detailed performance of the modules, various beam tests have been performed before and after irradiation, both at DESY and at CERN. The DESY tests used electrons with an energy of 4 to 4.8 GeV, while those at CERN used 120 GeV pions. At both DESY and CERN beam tests, the EUDET-style telescopes are used, which consist of six *MIMOSA26* pixel sensors. Another pixel layer with a different read-out system was used to achieve better timing for the telescope so that individual tracks could be matched to hits on the strip module to be tested.

Chapter 3

Data Analysis Tools and Techniques

3.1 ATLAS Physics Objects

: To analyze the data collected by the ATLAS detector, particles must be reconstructed and identified from the signatures they leave in various parts of the detector. A complex system of procedures used to associate these signals with specific particles is standard practice within the ATLAS collaboration and is continuously updated and refined. A *physics object* is here defined as a candidate particle that has been reconstructed and identified. The following sections give a brief description of the physics objects used within the ATLAS collaboration and specifically in the physics analyses discussed in this thesis.

3.1.1 Primary Vertex requirement

Charged particle trajectories (tracks) with $p_T > 0.4$ GeV and $|\eta| < 2.5$ are reconstructed using data from the ID. Initially, hits in the Pixel Detector serve as seeds for the track reconstruction algorithm. This information is subsequently integrated with data from the SCT to form track candidates, which are then refined with inputs from the TRT measurements. These tracks are input into the primary vertex identification algorithm, which employs an iterative approach consisting of two main stages: initially, a *finding* algorithm identifies a vertex seed using the information of the track point of closest approach to the beam spot. Afterwards, a *vertex fitting* algorithm uses the seed and available tracks to determine the vertex position and assess each track's compatibility with the interaction coordinates.

Multiple interactions per bunch crossing occurred during *Run 2* and thus numerous vertices are identified per event. The primary vertex for each bunch crossing is designated as the one associated with the group of tracks having the highest squared p_T sum.

3.1.2 Electrons

Electrons are expected to leave hits in the ID and create electromagnetic showers in the EM calorimeter. The electron candidate object reconstruction is initially independently performed for ID tracks (as described in 3.1.1) and electromagnetic clusters, i.e. clusters of energy deposits in the EM calorimeter cells. To identify these electromagnetic clusters and to associate with them the relative four-momentum vector a *TopoCluster* algorithm is used [57]. The reconstructed tracks are then matched to the electromagnetic clusters, creating *Super-clusters*.

Prompt electron candidates are then selected using a likelihood-based identification. Input variables used for the identification process are related to primary electron track, electromagnetic shower shape and track-cluster compatibility. The input variables distributions are determined in separate bins of the electron candidate transverse energy E_T and pseudo-rapidity η. Some of the input distributions are also corrected in MC-simulated samples to match the distributions observed in data. Four main identification Working Points (WPs) are defined: VeryLoose, Loose, Medium, and Tight WP. They are ordered in increasing background rejection and decreasing efficiency [58].

A calibration of the electron energy is necessary to consider effects like energy losses in passive materials, EM shower leakages and fluctuations in the energy deposited in the calorimeters. Such calibration is derived from MC simulation but is then refined with corrections to energy scale and resolution extracted from data using reference SM processes: $Z \to e^+e^-$ and $J/\psi \to e^+e^-$ [59].

To enhance the rejection of backgrounds such as non-prompt electrons and hadrons misidentified as electrons, so-called isolation requirements are usually applied to identified candidates. Variables are constructed that quantify the amount of activity in the vicinity of the candidate object. Two classes of isolation variables are considered: calorimeter-related variables and track-related ones. Selections on these variables have been optimised to identify prompt electrons and related selections are applied.

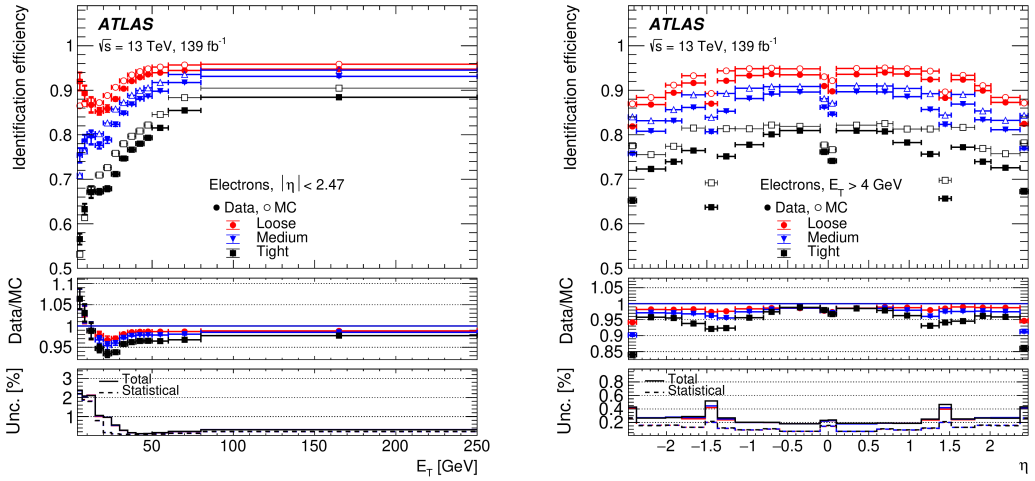
The reconstruction, identification and isolation processes come with related efficiencies, which can differ between data and MC samples. In order to correct for these differences between data and MC, electron Scale Factors (SFs) are applied to MC simulations. Such SFs are defined as the ratio of the efficiency measured in data and the efficiency in the MC simulation $SF = \varepsilon_{DATA}/\varepsilon_{MC}$. The electron efficiency can be factorised as a product of the electron reconstruction, identification, and isolation, as in A.2

$$\varepsilon = \varepsilon_{reco} \cdot \varepsilon_{id} \cdot \varepsilon_{iso} = \frac{N_{reco}}{N_{clus}} \cdot \frac{N_{id}}{N_{reco}} \cdot \frac{N_{iso}}{N_{id}}, \tag{3.1}$$

where N_{clus} is the number of reconstructed electromagnetic (EM) calorimeter clusters, N_{reco} is the number of reconstructed electron candidates, N_{id} is the number of identified electron candidates passing previous selections, N_{iso} is the number of isolated electron candidates passing previous selections. In cases where the electron is used to trigger the acquisition of the event, also the trigger efficiency $\varepsilon_{trigger} = N_{trig}/N_{iso}$ is considered in the previous equation A.2.

and N_{trig} is the number of triggered electron candidates passing previous selections. Figure 3.1, shows the electron identification efficiencies with respect to the electron E_T and η for different identification WPs. To obtain these efficiency measurements in data, a clean and unbiased sample of electrons is needed. The decay of the Z boson into an electron-positron pair can be used. In this case, the so-called *tag-and-probe* method is typically employed: one of the electrons is required to pass strict selection criteria (referred to as *tag* electron), while the other one serves as an unbiased source to measure the efficiency of passing a specific step of the selection (referred to as *probe* electron).

Among physics processes studied in the ATLAS collaboration, it is common to have final states with both a charged lepton (and specifically an electron) and one or more jets. The separation between the electron and the jet can be quantified through the angular distance $\Delta R(\text{e, jet}) = \sqrt{(\Delta\eta(\text{e, jet}))^2 + (\Delta\phi(\text{e, jet}))^2}$. The vicinity in ΔR of a hadronic jet to the electron can modify the efficiency of selection steps such as the identification and isolation of the electron. The dependencies of the electron identification or isolation efficiency can be partially mis-modelled in MC simulation. It can therefore be useful to study electron efficiencies and SFs with



(a) El. identification efficiencies wrt E_T.

(b) El. identification efficiencies wrt to η.

Figure 3.1: Identification efficiencies of electrons as a function of the electron's 3.1(a) transverse momentum and 3.1(b) pseudo-rapidity for the different identification WPs. The top panels show the efficiencies obtained in data and MC simulation. The middle panels show their data to MC ratio that is applied as a correction factor in analyses. The bottom panels show the statistical and the total uncertainties in the data/MC ratio.

respect to the angular distance to the closest jet. Moreover, as discussed also in Section 3.1.5, the decay of boosted particles, such as boosted top quarks decaying leptonically, can create a topology with an electron and a jet close to each other, with a $\Delta R(\text{e, jet}) \leq 0.4$. The electron identification process is however optimized considering only electrons with $\Delta R(\text{e, jet}) > 0.4$, and SFs are computed only for those electrons. Further studies, to exploit information on electrons close to jets are discussed in Appendix A.

Uncertainties on the energy calibration, reconstruction, identification and isolation efficiency are related to statistical uncertainty, background estimation, signal modelling, and different estimation technique comparison as detailed in [59, 60]. Such uncertainties are propagated to the relative SFs.

3.1.3 Muons

Muons are charged leptons with a mass $\sim$ 200 times larger than electrons, as a consequence they interact very little with the bulk of calorimeters in the ATLAS detector. The Muon Spectrometer system (see 2.2.3), located outside of the hadronic calorimeter, is then primarily used to identify and measure muons produced in *pp* and heavy ion collisions. A detailed report on muons treatment in the ATLAS detector can be found in [61].

Tracks from the Inner Detector and MS are first reconstructed independently and then combined to create a candidate muon track across all the layers of the detector. Tracks in the ID are first reconstructed using the Pixel detector and SCT hits and then are extended with TRT information (more details on track reconstruction performances in the ATLAS detector in [62]). Candidate tracks in the MS are constructed starting from track segments, that are made of hit patterns found in individual MS stations. Afterwards track segments are combined into approximated track candidates which are then fitted to create MS tracks. Finally, *combined muon*

candidates are reconstructed by combining the ID tracks with the MS tracks through a global fit procedure. Potentially, muons can be reconstructed with other combinations of detector information as explained in [61], but only combined muons are considered in the work discussed in this thesis.

Following the reconstruction process, muon candidates are identified through quality requirements designed to suppress background contributions originating from non-prompt muons, spurious hits, and signals in the MS caused by jets not fully contained in the hadronic calorimeter. Four main identification WPs are defined: the **Loose** WP, which maximises the reconstruction efficiency; the **Medium** WP, which is the standard one for ATLAS physics analyses has loose selection on the compatibility between ID and MS muon candidate momentum measurements; **Tight** WP, which optimises the muon identification purity; the **High-p_T** WP which tries to optimise the momentum resolution for muons with $p_T > 100$ GeV.

Similarly to electron isolation, discussed in Section 3.1.2, the isolation of a muon candidate is a key element to discriminate prompt muons from background muons, as muons from decay of quarks and misidentified hadrons, which are usually generated close to a jet and are surrounded by high detector activity. As for electrons, track-based and calorimeter-based variables are used to evaluate the isolation of muons. Various isolation WPs are defined for usage in ATLAS physics analyses, but here only those used in the analyses discussed in this thesis are briefly discussed. The t-channel single top production measurement (presented in Chapter 4) uses the **Tight Fixed Radius** WP. Upper cuts are applied on the calorimeter cluster energy within a cone of the size of $R = 0.2$ around the muon track and on the ratio of the summed p_T of additional tracks within a cone of $R = 0.3$, as detailed in [61]. The search **Tight, Track-Only Variable-Radius** WP is used in the search for heavy resonances decaying to top quark pair (see Chapter 5). This WP requires the sum of the track transverse momenta within a variable-radius cone around the muon direction, excluding the muon track, to be less than 6% of the muon p_T. The track isolation cone size is defined as $R = \min(^{10\text{GeV}}/_{p_T^\mu}\ ,\ 0.3)$, thus, the cone radius increases with decreasing muon transverse momentum (p_T^μ) up to a maximum of 0.3.

Reconstruction, identification and isolation efficiencies for muons are computed both in data and MC samples with the *tag-and-probe* method, as for electrons similarly as for electrons [61]. The processes $Z \to \mu^+\mu^-$ and $J/\psi \to \mu^+\mu^-$ are used, with one muon as a reference tag selected with stringent criteria and the other as the probe. Scale factors are then derived as the ratio between efficiencies in data and MC samples and used to correct the MC prediction. Systematic uncertainties on the scale factors take into account uncertainties in the background estimation and variations of selection requirements. The same physics processes are used to correct the muon momentum scale, which can be affected for example by mis-modelling of the detector or magnetic field, and detector misalignment. Alignment corrections are also implemented for both the MS and ID detectors. Some residual charge-dependent biases, particularly on the sagitta[1], can still be found after these calibrations and are corrected minimizing the variance in the di-muon invariant mass distribution. Systematic uncertainties are assigned to these various corrections, accounting for uncertainties in the extrapolations outside the phase space of the $Z \to \mu^+\mu^-$ decays, as well as of potential biases from selection requirements and signal and background modelling [63].

[1]The sagitta is defined as the distance between the midpoint of the *arc* and the midpoint of the related chord.

3.1.4 Jets

The electromagnetic and hadronic calorimeters of the ATLAS detector are built so that particles like photons, electrons and hadrons that pass through them create showers of secondary particles. In the case of hadronic showers, developing mostly in the hadronic calorimeter, these are called *hadronic jets* or simply *jets*.

In ATLAS the anti-k_T algorithm [64] is used to reconstruct jets, as it is insensitive to soft and collinear gluon emissions. This algorithm takes as inputs a set of constituents, that can be reconstructed particles, energy clusters in the calorimeters of ID tracks: these inputs are added to the clustered jet based on two specific distances d_{ij} and d_iB. The distance between two particles i, j is

$$d_{ij} = min(\frac{1}{k_{t,i}^2}, \frac{1}{j_{t,j}^2})\frac{\Delta R(i,j)}{R^2} \tag{3.2}$$

where $k_{t,i}$ is the transverse momentum of particle i, $\Delta R(i,j)$ is the angular distance between the particles and R is the radius parameter of the jet. A measurement of the distance between particle i and the beamline is described as $d_{iB} = 1/k_{t,i}^2$. If the smallest of $d_{i,j}$ and d_{iB} distances is d_{iB}, then the object i is classified as a jet; while if $d_{i,j}$ is the smallest, then objects i and j are combined to form a new object k. The procedure is then repeated until all objects have been considered and clustered. A visual representation of the anti-k_T algorithm is shown in Figure 3.2. Jets used in the analyses discussed in this thesis are defined using *particle flow* objects, which use both track and calorimeter type of information [65].

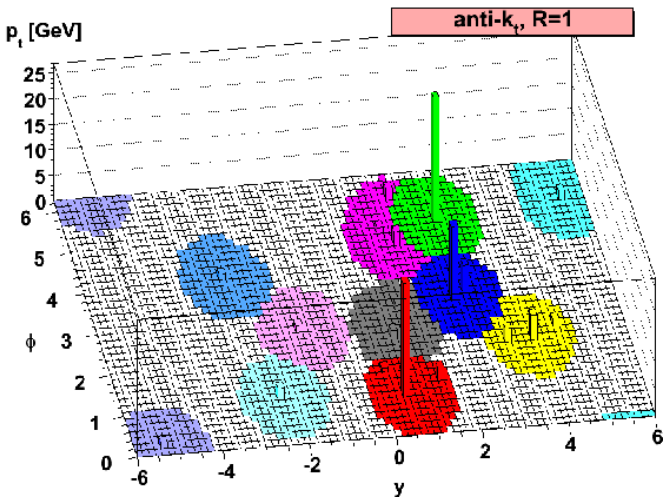


Figure 3.2: Jet clusters created with the anti-k_T algorithm and a radius parameter $R = 1.0$, in a MC sample generated with Herwig [64].

The radius parameter R plays an important role in the jet reconstruction, as it characterizes the size of the cone of the jet itself. Within the ATLAS Collaboration standard jets have a radius parameter of $R = 0.4$. Other kinds of jets can also be used, such as the *large-R jets*, typically with R set to 1.0. Such large-R jets are often used to reconstruct highly boosted heavy particles with collimated decay products.

Jet Energy Corrections

Jets reconstructed in the detector need to undergo an energy calibration process, to correct for energy loss in the passive calorimeter layers and other effects. Such

calibration aims at aligning the measured energy to the particle-level jet energy in MC simulation. These corrections modify the jet p_T, energy and mass [66], and proceed as follows.

To mitigate the surplus energy from additional proton-proton interactions occurring either during the same or adjacent bunch crossings, *pile-up* corrections are applied. The *absolute JES* (Jet Energy Scale) calibration aims to correct the visible jet energy and direction to match those found in MC dijet events. Using the same dijet events, the *GSC* (Global Sequential Calibration) is derived to minimize effects related to detector inefficiencies and physics effects by applying layer-specific and overall jet energy adjustments. These corrections are derived from MC simulation only and are applied both to data and MC samples. Afterwards *in-situ* corrections are applied to deal with any left discrepancy between data and MC. These in-situ corrections are derived using selected data events with well-known reference objects, such as Z bosons and photons. Furthermore, jet energy resolution (JER) is also corrected. A *relative JER* is determined using di-jet events, where one of the jets is used as a probe jet and the other one as the reference tag.

Residual effects of pile-up can simulate jets, particularly particularly low-p_T jets. A likelihood-based Jet Vertex Tagger (JVT) reduces these sort of pile-up effects with a technique that tags pile-up jets using a 2D likelihood estimate [67]. In the analyses discussed in this thesis, a JVT score greater than 0.5 is required for jets with $p_T > 60$ GeV and $|\eta| < 2.4$.

Jet in-situ correction at 5.02 TeV

The jet energy calibrations implemented for data collected by the ATLAS collaboration in *Run 2* at $\sqrt{s} = 13$ TeV are not necessarily applicable to the dataset collected at $\sqrt{s} = 5.02$ TeV, due to the smaller amount of pile-up, lower underlying event activity and the use of lower calorimeter topological cluster noise thresholds both in data taking and in event reconstruction. The compatibility of the jet corrections between these two setups has been already studied in an ATLAS measurement using data collected at $\sqrt{s} = 5.02$ TeV [68], and it was found that an additional JES in-situ correction is needed with respect to those used for the standard $\sqrt{s} = 13$ TeV dataset. This section briefly describes the in-situ correction derived in the context of the t-channel single top cross-section measurement at $\sqrt{s} = 5.02$ TeV described in Chapter 4.

The Z–jet p_T balance technique is used to estimate the effect of the correction on the jet energy scale. This analysis aims to select events with a single jet recoiling against a Z boson decaying to two leptons and compare their transverse momenta. Events are selected if they contain a pair of opposite-sign, same-flavour charged leptons (e^+e^- or $\mu^+\mu^-$) both with $p_T > 20$ GeV and passing identification and selection requirements (as in the t-channel single-top analysis described in 4.2.1) and at least one jet with $p_T > 10$ GeV and $|\eta| > 4.0$ (as described in 4.2.1). Furthermore, the two leptons are required to have an invariant mass in the Z-mass window of $81 < m_{ll} < 101$ GeV. To obtain a back-to-back topology between the Z and the jet, the azimuthal angle needs to be $|\Delta\phi(Z, jet)| > 2.8$. The reference p_T can be defined as $p_T^{ref} = p_T^Z \cdot \cos\Delta\phi(Z, jet)$. Events with a second jet are required to have $p_T^{sub-leadjet} < \max(10\text{ GeV}, 0.1 \cdot p_T^{ref})$. Figure 3.3 shows the transverse plane view of the Z boson and its recoiling jet. The per-event balance quantity is defined as $r = p_T^{jet}/p_T^{ref}$, and its mean value $\langle r \rangle$ as a function of the p_T^{ref} is used to calibrate the jets. The in-situ correction is derived as a scale factor as the ratio between $\langle r \rangle$ in data and MC simulation in bins of the jet p_T and $|\eta|$. Figure 3.4(a) shows the two-dimensional correction map. The uncertainty associated with this in-situ correction

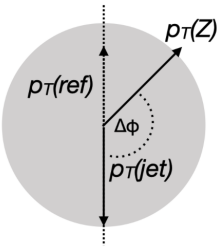
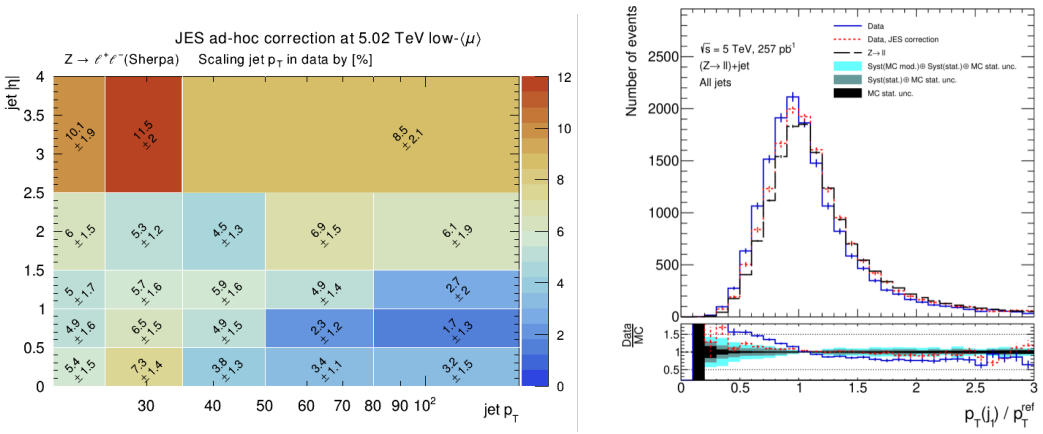


Figure 3.3: Projection of the Z boson p_T along the recoiled jet axis, which is defined as p_T^{ref}.



(a) In-situ JES correction map.

(b) Z-jet balance quantity $\langle r \rangle$.

Figure 3.4: On the left the in-situ correction for the JES map, is shown as a scale factor to be applied on data as a function of the jet p_T and $|\eta|$. The error is related to the statistical uncertainty of the data and MC samples. On the right the Z-jet balance $\langle r \rangle$ distribution for all jets considered in the study. Data are shown before (solid blue line) and after (dashed red line) applying the in-situ correction. The bottom box shows the ratio between the data and the MC sample.

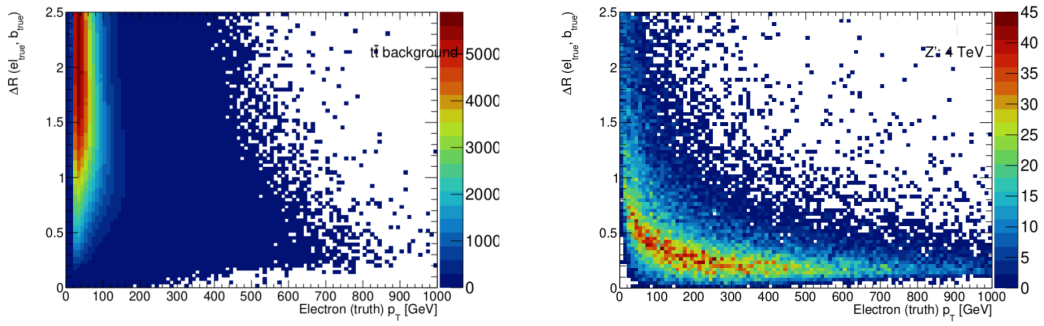
includes a statistical and a systematic component. The latter is evaluated by comparing the correction obtained with a `Sherpa` sample and a `Powheg+Pythia8` sample. Figure 3.4(b) shows the agreement between data and a `Sherpa` MC sample before and after the correction is applied. It can be seen that before applying the correction the agreement is poor, while afterwards, they are compatible within uncertainties.

3.1.5 Overlap Removal

The so-called overlap removal (OR) procedure is applied in physics analyses to avoid double counting of candidate events, energy deposits and reconstructed tracks. It is indeed possible that the electronic signals left in the detector by one particle are used for the reconstruction of multiple physics objects. The following steps define the standard ATLAS OR procedure[2], and are performed in this order:

- any muon found to share a track with an electron is removed;
- any electron found to share a track with a muon is removed;

[2]The steps of the OR procedure here described are valid for analysis not using τ particles and photons, differently some more steps in the OR need to be considered.



(a) SM $t\bar{t}$ production.

(b) $Z' \to t\bar{t}$ production.

Figure 3.5: Angular separation between the electron and the b-jet, $\Delta R(e_{true}, b\text{-jet}_{true})$, originating from a leptonically-decaying top quark, as a function of the true electron p_T for 3.5(a) SM $t\bar{t}$ production and 3.5(b) a 4 TeV mass Z' heavy resonance decaying to $t\bar{t}$, both of them decaying dileptonically.

- any jet found within a ΔR of 0.2 of an electron is removed;
- any electron subsequently found within ΔR of 0.4 of a jet is removed;
- any jet with fewer than three tracks associated to it found within ΔR of 0.2 of a muon is removed;
- any jet with fewer than three tracks associated to it which has a muon inner-detector track ghost-associated[3] to it, is removed;
- any muon subsequently found within ΔR of 0.4 of a jet is removed;
- any τ found within a ΔR of 0.2 of any type of muon with $p_{\mathrm{T}} > 2$ GeV is removed;
- any jet found within a ΔR of 0.2 of a τ is removed;
- any jet found within ΔR of 0.4 of a photon is removed.

The jets referred to in the OR procedure are the *standard jets* (as defined above in 3.1.4), while no OR removal procedure is applied with respect to the *large-R jets*.

Special Overlap Removal for boosted toplogies

By following the standard ATLAS OR procedure described above, there cannot be events with leptons (electrons or muons) within an angular distance of $\Delta R \leq 0.4$ from a jet. Instead, highly boosted leptonically-decaying top quarks typically result in collimated lepton–jet pairs. Figure 3.5 shows the $\Delta R(e, b\text{-jet})$ as a function of the electron p_T comparing the distribution in simulated SM $t\bar{t}$ production versus simulate BSM heavy resonance Z' production decaying to $t\bar{t}$ (see 1.3.1 for details on the BSM model used). It is clear, looking at Figure 3.5 that the number of events with $\Delta R(e, b\text{-jet}) \leq 0.4$ is significantly large for the BSM Z' signal process at high-p_T, while it is quite small for SM $t\bar{t}$ events. Using the standard ATLAS OR in this case, would cause to throw away a large portion of potential signal events.

[3]Ghost-association is a technique used to define the jet area by making use of infinitely soft particles, called "ghosts" [69]

Some alternative OR procedures have been developed to minimize this effect. The *sliding* ΔR OR includes a special treatment of the OR between muons and jets and is widely used in the ATLAS Collaboration, while the *electron-in-jet OR*, which is characterized by a special treatment of the OR between electrons and jets, has been specifically developed for the physics analysis presented in Chapter 5.

A short description of these two special OR techniques is given:

- **Sliding ΔR OR for muons**: a variable-ΔR association is used instead of the fixed-ΔR cone used in the standard ATLAS OR. The ΔR used for the muon–jet OR is determined as: $\Delta R = \min(0.4\ ,\ 0.04 + 10\ \text{GeV}/p_T^{\mu})$. In this way, the size of the cone decreases while the p_T of the muon increases, reducing signal selection inefficiencies due to close-by jets for sufficiently energetic muons.

- **Electron-in-jet OR**: distinguishing an electron candidate object that overlaps with a candidate jet object can be complicated as both these physics objects deposit their energies in the calorimeter, and they leave similar signatures in the ATLAS detector. In general, when an electron candidate object and a jet candidate object are found close in angular distance there can be three main possibilities: the candidate jet is the result of a true electron wrongly reconstructed also as a jet; the reconstructed electron is not a real prompt electron (it could be a non-prompt electron from a c- or b-hadron semi-leptonic decay or a hadron or photon mis-identified as an electron) and should be discarded; both the jet and the electron are two separate prompt particles and their energy deposits overlap. The procedure applied for each electron is:

  1. the electron's four-momentum is subtracted from the jet's four-momentum if the two are within a distance of $\Delta R < 0.2$;
  2. if the jet p_T after this subtraction falls below a threshold (p_T^{max}), the jet is discarded;
  3. if any jets remains within $\Delta R < 0.2$ of the electron, afterwards the electron is removed.

  The threshold p_T^{max} is defined as $^{p'_T}/_{p_T^0+b} < c$ where p_T^0 is the jet p_T before the subtraction (mentioned above at point 1) and p'_T is the jet p_T after the subtraction, while the two constants are chosen to be $b = 20$ GeV and $c = 0.37$. These optimal values of b and c were determined using $Z \to e^+e^-$ simulated events. This technique allows the inclusion of electrons with reconstructed jets within $\Delta R < 0.4$, which are otherwise excluded by ATLAS standard electron efficiency calibrations.

  In the context of this thesis, various studies on electron efficiencies and SFs have been performed, as well as studies to better understand the interplay between electrons and jets in this particular topology. Such studies are detailed in Appendix A.

3.1.6 Missing Transverse Energy

Based on the principle of momentum conservation, the total transverse momentum in each event should be zero. The Missing Transverse Energy (MET) E_T^{miss} can be defined as the imbalance in the total measured transverse energy. It can indicate the presence of weakly interacting particles in the event final state. In the SM these particles are mainly identified as neutrinos, but in general, the presence of MET in

an event could hint at the production of BSM particles. From a mathematical point of view, the MET can be defined in the x and y direction as:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,\text{jets}} + E_{x(y)}^{miss,\mu} + E_{x(y)}^{miss,\text{soft}} \tag{3.3}$$

where each term is the negative vectorial sum of the momenta of the respective particles, while the soft term refers to all the energy deposits in the detector not associated with any reconstructed particle or jet. the transverse component can thus be extracted:

$$\begin{cases} E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \\ \phi^{miss} = \arctan(E_y^{miss}/E_x^{miss}) \end{cases} \tag{3.4}$$

Systematic effects related to the scale and resolution are computed by comparing data and MC samples. Systematic uncertainties originating from the reconstructed particles used to compute the MET are evaluated separately and then propagated to the MET calculation.

3.1.7 Flavour tagging

Hadronic jets can originate from highly energetic gluons or quarks of different flavours. Being able to identify or *tag* the flavour of quark-initiated jets is a key ingredient of a large fraction of measurements and searches at modern particle-collider experiments. Jets containing energetic b-hadrons, hence likely originating from a b-quark, are referred to as b-jets. Their identification, referred to as b-jet-tagging or simply b-tagging, can significantly help to select candidate top-quark production events, as top quarks decay into a b quark and a W boson essentially 100% of the times.

The distinctive features of a b-hadron decay are used to identify $b$jets. In particular, b-hadrons produced in pp collisions at the LHC energies travel at around the speed of light, decaying after $\sim 300\mu$m, hence creating *secondary vertexes* that can be resolved from the primary vertex thanks to the precision of the inner tracking system. Various algorithms have been developed over the years by the ATLAS Collaboration to identify b-jets, among which the DL1r. The DL1r algorithm is based on a feed-forward neural network, taking as input the kinematic variables of the jet and its associated tracks

The DL1r algorithm uses as input kinematic variables, variables related to track transverse and longitudinal impact parameter significance, as well as secondary-vertex reconstruction information. It outputs the probabilities for a jet to be a b-jet (p_b), a c-jet (p_c) and a light-jet (p_{light})[4]. The discriminant D_b is constructed as:

$$D_b = \frac{p_b}{f_c p_c + (1 - f_c) p_{light}} \tag{3.5}$$

where f_c is the fraction of c-jets in the background. Different working points are defined based on the b-jet identification efficiency.

When applying a b-tagging algorithm, scale factors need to be applied to correct for possible differences in efficiencies between data and MC simulation. These scale factors are extracted from $t\bar{t}$ events in the case of b-tagging efficiency and c-jet misidentification rate, and from Z+jets events for measuring the light-jet misidentification rate. Systematic uncertainties in these scale factors are obtained by propagating through the calibration analyses the statistical uncertainties in data and MC samples, JES and JER uncertainties, as well as signal and background modelling uncertainties [70, 71, 72].

[4]Here light jet refers to jets originating from up, down and strange quarks.

3.1.8 Boosted top-tagging

Many searches for physics BSM need to fully exploit processes with boosted massive particles such as W, Z or Higgs bosons or top-quarks. It is important to reconstruct and accurately identify hadronic decay modes of these massive particles to reject background events and to be able to build invariant mass distributions of candidate heavy BSM resonances.

For instance, in the analysis discussed in Chapter 5, the $t\bar{t}$ system is studied in highly boosted topologies. Hadronic decays of such energetic top quarks give raise to *large-R jets*, with distinctive multi-prong structures due to the topology of the chain decay $t \to W \to q\bar{q}'$. Machine-learning techniques can be used to identify such hadronically decaying boosted top quarks. This procedure is referred to as boosted-top-tagging or simply top-tagging. In particular, ATLAS developed a top-tagging algorithm is based on a Deep Neural Network (DNN) [73]. This DNN top-tagger exploits the sub-structure of *large-R jets* to distinguish top-quark jets from jets originating from lighter quarks or gluons. The tagger uses jet-level discriminants such as the jet p_T and mass, as well as information about the dispersion of the jet constituents, including energy correlation functions [74], the so-called N-subjettiness[5] [75], and splitting scales [76].

A DNN score between 0 and 1 is obtained, with top-quark-initiated jets having values close to 1 and light-quark- and gluon-initiated jets having values close to 0. Multiple specific definitions of the DNN top-tagger are available. In the analysis discussed in Chapter 5 the *inclusive* top-tagging criteria is used, optimised for jets containing some but not necessarily all of the decay products of the hadronically decaying top quark [77].

This top-tagging process can have different efficiencies when applied to data and MC samples, and therefore dedicated SFs are derived and applied to data [73].

3.2 Monte Carlo Simulation

In high-energy physics experiments such as ATLAS, theoretical predictions for many observables cannot be directly computed analytically, except in a few straightforward cases. In most cases, when comparing experimental data with Standard Model predictions, MC simulations are indispensable. These simulations employ stochastic methods to perform numerical integration by randomly sampling outcomes from theoretical probability distributions. This section gives an overview of the MC simulation used in the ATLAS experiment that are relevant to the physics analyses discussed in this thesis.

3.2.1 Event simulation

The ATLAS simulation framework [78] is segmented into various parts, the initial step is the generation of events, where a series of particles are produced for subsequent detector simulation. Figure 3.6 shows this initial phase, which can be categorized into four main sub-steps:

- **Hard-scattering**: also known as the matrix element (ME) generation, involves the collision of two partons. It describes short-distance parton-parton

[5]The N-subjettiness is defined as: $\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \cdots, \Delta R_{N,k}\}$, where k runs over the constituent particles in a given jet, $p_{T,k}$ are their transverse momenta, and $\Delta R_{J,k} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is the distance in the rapidity-azimuth plane between a candidate subjet J and a constituent particle k. The normalization factor is $d_0 = \sum_k p_{T,k} R_0$.

interactions that are calculated with perturbative QCD, as well as taking into account the effect of the PDFs.

- **Parton shower (PS)**: refers to the radiation of coloured particles and photons by the scattered partons after the hard scattering process. It also includes the initial state radiation (ISR), and emission of gluons or quarks by the incoming partons before the scattering event.
- **Hadronization:** Following the PS, it takes place at energy scales close to the non-perturbative limit of QCD. It describes the transition of partons into hadrons, which subsequently decay. Two models commonly employed to simulate this process are the Lund string fragmentation model [79] and the cluster fragmentation model [80].
- **Multiple interactions and underlying event:** In addition to the primary parton-parton collision, additional ones can occur simultaneously. This phenomenon is called Multiple Parton Interactions (MPI). The events created by the MPI are called *underlying events* (UE).

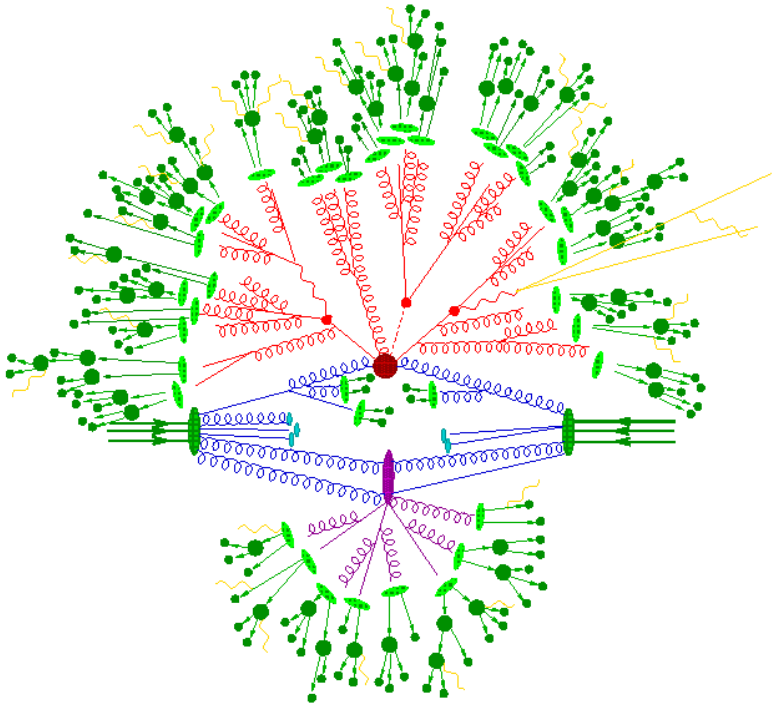


Figure 3.6: Schematic representation of a hard collision event as produced by a MC simulation [81]. The red part represents the hard scattering and the parton shower steps, which then progresses in the hadronization in green. The purple part of the representation shows multi-parton interactions.

3.2.2 Monte Carlo generators

Various Monte Carlo generators are available to the high-energy physics community, and multiple are used by the ATLAS collaboration. In the following, those used in the analyses discussed in this thesis are discussed.

The ME computations is handled by Powheg Box [82], MadGraph5_aMC@NLO [83] and Sherpa [84] at NLO. Events created by these generators are then passed to PS

generators like PYTHIA [85], HERWIG [86] and SHERPA. HERWIG treats hadronization through the cluster fragmentation model, while PYTHIA and SHERPA use the Lund string fragmentation model. The output of the PS and hadronization processes can then be passed to the EVTGEN [87] generator which is specialised in b- and c-hadrons decays.

After events are generated, they undergo a comprehensive simulation of particle interactions with the ATLAS detector. This step is essential to compare the simulated events with experimental data collected by the detector. The Monte Carlo simulation program utilized is GEANT4 [88], a software toolkit that simulates a wide range of physical processes, including electromagnetic, hadronic, and optical phenomena across diverse materials, with energy scales from 250 eV to several TeV. Key components of the GEANT4 toolkit are the simulation of the detector's detailed geometry and materials, the generation of primary particles for each event, the tracking of particle trajectories through materials and electromagnetic fields, and the simulation of the detector's sensitive components response to interactions with these particles.

The comprehensive simulation via GEANT4 of particle interactions within the ATLAS detector is both time-consuming and computationally demanding. When such precision is not essential, and efficiency is prioritized in the production of the final MC samples, a faster simulation alternative is employed. This method utilizes parameterized models of showers within the calorimeters, known as FASTCALOSIM [89].

3.2.3 Uncertainties related to Monte Carlo generators

Monte Carlo generators try to simulate QCD and EW theory for particle collisions at the LHC. The results have shown to be extremely precise. Nevertheless, various approximations and choices of parameter values are done during the simulation. Therefore several systematic uncertainties are associated with the use of MC samples. In this Section, only those relevant for the analyses described in Chapters 4, 5 are briefly described.

As the ME computation is an approximation of perturbative QCD, an uncertainty related to missing higher-order QCD correction in the ME computation is applied. They are estimated by independently varying the re-normalization (μ_R) and factorization (μ_F) scales by multiplicative factors of 2 and 0.5.

In the parton shower, the ISR is simulated, and the uncertainty related to the amount of the ISR can be assessed by varying the corresponding parameter of the A14 PS tune (`Var3c`) [90]. Similarly, the uncertainty related to the Final-State-Radiation (FSR) is evaluated by varying the nominal value of the scale μ_R^{FSR} by factors of 2 and 0.5.

The PS model used by a specific MC generator can be compared with that of a different MC generator to take into account the uncertainty related to the choice of such a model. For instance, when using a MC sample generated with POWHEGBOX+PYTHIA the difference with the same process generated with POWHEGBOX+HERWIG can be taken as a systematic uncertainty.

Matching the ME computation to the PS model in a MC generator can be performed in different ways and usually comes with the choice of a cut value of the transverse momentum of emitted particles. An uncertainty related to this choice, referred to as *matching* uncertainty can be estimated in samples generated with POWHEGBOX+PYTHIA by varying the parameter `pThard` from its standard value of 0 to 1.

In the POWHEG simulated sample the PS is regulated also by the h_{damp} parameter, which controls the amount of radiation (gluons emitted from initial and final state partons) in the parton shower, acting as a damping factor for the high p_T

emissions. The uncertainty related to the choice of the value for this parameter in $t\bar{t}$ simulated samples is usually assessed by comparing the nominal sample with the one produced by choosing $h_{damp} = 3m_{top}$.

PDFs are estimated by fit to experimental data and come with a related uncertainty. It can be estimated in different ways, based also on the specific PDF set used in the analysis. In some cases, it is common to vary internal parameters for the re-weighting of the PDF set which is used as a systematic variation, while in other cases a PDF set is compared with a completely different one.

In the case of the single top tW production mode, a choice on the method to remove the double-counting of diagrams contributing also to the $t\bar{t}$ production is needed. The method used in the nominal samples in this thesis is called *diagram removal*. An uncertainty on this choice can be estimated by comparing the nominal sample with another one which uses a different method called *diagram subtraction* [91].

3.3 Fake and non-prompt lepton estimation with the Matrix Method

Many ATLAS measurements and searches require the presence of one or more leptons in the final state (here we only consider electrons and muons as leptons), as indeed is done in the analyses presented in this thesis. We can define here *real leptons* as electrons or muons produced directly in the hard-scattering process or directly in the decay of a short-lived non-hadronic resonance. Typical ATLAS analyses aim to select *real leptons* in the final state.

In the ATLAS detector lepton candidates are reconstructed based on signals from the inner tracker, calorimeters, and muon spectrometer, as explained in 3.1.2, 3.1.3. After the reconstruction process, identification criteria are applied to select *real leptons* and suppress background candidates, which can be categorized as *non-prompt leptons* from the semileptonic decay of hadrons, or from photon conversions in detector material, and *fake leptons* where the reconstructed object is not a lepton.

Monte Carlo simulators do not usually describe precisely the rates of identification and selection of *non-prompt* and *fake* leptons. These processes can vary significantly based on small details of the simulation and are also part of the non-perturbative regime, which is not reliable in MC simulation. Furthermore, they depend on the simulation of particle interaction with the detector material for *non-prompt* leptons and on low probability events in which a hadronic jet in a multi-jet event gives rise to a *fake lepton*. Such complicated and rare processes would require an excessive amount of computing resources to obtain a precise simulated sample. To estimate these backgrounds *data-driven* approaches are used, in which special techniques are applied to data, using information extrapolated from MC simulation.

The ATLAS collaboration provides standard tools and prescriptions that physics analyses can use to estimate this background, as referenced in [92]. Specifically, a few main methods are available: the *fake-factor* method, which applies a factor F to account for differences observed between simulation and data control samples, and the *matrix method*, which is described in the next section.

3.3.1 The Matrix Method

The estimation is based on the definition of two sets of selections: the *tight* and the *baseline*, which contains as a subset the first one. The first set of criteria should be the same as that used in the physics analysis to select leptons, which are here called

tight leptons. The *baseline* selection instead applies less stringent requirements on the leptons, which translates into a higher rate of *fake* or *non-prompt* leptons. Leptons that satisfy the *baseline* criteria but not the *tight* ones are called *loose leptons*. These two sets of criteria are chosen so that the **real efficiency** (ε_r), defined as the fraction of *real leptons* in the *baseline class* that pass the *tight criteria*, is substantially larger than the **fake efficiency** (ε_f), that is the fraction of *fake/non-prompt* leptons in the *baseline class* that pass the *tight criteria*.

The real efficiency ε_r is usually estimated using the SFs (see previous section 3.1.2), and 3.1.3 that are precisely calibrated by the ATLAS collaboration with the *tag-and-probe* method in $Z \to \ell\ell$ events and can be applied to simulated events using the *tight criteria* relevant for the specific analysis. The fake efficiency ε_f can be estimated with different methods based on the specific analysis, which can be more sensitive to one source of *fake* or *non-prompt* leptons than another one.

Once the real and fake efficiencies are estimated, the number of tight (N_t) and loose leptons (N_l) observed in data can be related to the number of events with *real* and *fake/non-prompt leptons* in the *baseline* sample as:

$$\begin{pmatrix} N^t \\ N^l \end{pmatrix} = \begin{pmatrix} \varepsilon_r & \varepsilon_f \\ 1-\varepsilon_r & 1-\varepsilon_f \end{pmatrix} \begin{pmatrix} N_r^b \\ N_f^b \end{pmatrix}, \tag{3.6}$$

where N_r^b (N_f^b) is the number of *real lepton* (*fake/non-prompt*) in the *baseline* sample. At this point, inverting the matrix in eq. 3.6 one can obtain the number of events with a *fake/non-prompt lepton* in the *tight* sample:

$$N_f^b = \frac{1}{\varepsilon_r - \varepsilon_f}[(\varepsilon_r - 1)N^t + \varepsilon_r N^l] \quad , \quad N_f^t = \varepsilon_f N_f^b. \tag{3.7}$$

Each event in the *baseline lepton* data sample can now be assigned a *matrix method weight* based on eq. 3.7:

$$w = \frac{\varepsilon_f}{\varepsilon_r - \varepsilon_f}(\varepsilon_r - \delta), \tag{3.8}$$

where $\delta = 1$ if the event passes the *tight criteria* and $\delta = 0$ otherwise.

3.4 Statistical Analysis

As in many experimental physics results using pp collision data, the two analyses discussed in this thesis make use of statistical techniques based on profile likelihood fits. In this section, a brief introduction to the main statistical analysis techniques used in Chapters 4,5 is given.

3.4.1 Profile Likelihood Fit

In the analyses discussed in this thesis, binned profile likelihood fits are performed using a software framework based on the `HistFactory` [93] and `RooFit` [94] packages. A binned likelihood (LH) function $L(\mu, \theta)$ is constructed as a product of Poisson probability terms over all bins considered in the measurement. Often, this LH function depends on a single parameter of interest (POI), which is often an overall multiplicative factor to the signal prediction or *signal strength* μ and a set of nuisance parameters (NPs) that encode the effects of systematic uncertainties on the signal and background predictions ($\vec{\theta}$). The likelihood then includes multiplicative factors or *constraint terms* for each of the NPs, representing prior knowledge of the values of these parameters, or auxiliary measurements previously performed to determine them. Each of these constrained NPs is associated with a single independent source

of systematic variation, affecting the signal and/or background predictions in the various bins.

Such binned LH function can then be written as:

$$L(\mu, \vec{\theta}; \vec{n}) = \prod_{i=1}^{N_b} \text{Poiss}(n_i; \mu S_i(\vec{\theta}) + B_i(\vec{\theta}) \prod_{j=1}^{N_p} \text{C}(t_j; \theta_j, \sigma_j) \tag{3.9}$$

The variables involved in Equation 3.9 are defined as:

- N_b: number of bins;
- n_i: number of data events in i-th bin;
- μ: The parameter of interest (POI): here defined as the ratio of the observed cross-section to the expected cross-section: $\mu = \frac{\sigma_{\text{observed}}}{\sigma_{\text{expected}}}$:
- $S_i(B_i)$: number of expected or predicted signal (background) events in the bin i;
- N_p: number of nuisance parameters considered (systematic uncertainties);
- θ_j: the set of nuisance parameters corresponding to systematic uncertainties;
- t_j: the nominal values of the nuisance parameters θ_j;

The function C is typically a Gaussian in the case of NPs associated with systematic uncertainties, or a Gamma function for NPs related to statistical uncertainty.

With such a LH function defined, the optimal value (or *best-fit* value) for the POI can be found by performing an N-dimensional maximization on the LH function (where N is the number of NPs), or equivalently by minimizing the negative-log LH function. In other words, when obtaining the best-fit value for a POI, one is getting as by-products also the best-fit values of all the NPs in the model.

The uncertainties in these best-fit values, for the POI and for each of the NPs, are evaluated as follows. For each considered parameter, denoted as k, the so-called profile likelihood ratio is defined as:

$$\lambda(k) = \frac{L(k, \hat{\hat{\theta}}_k)}{L(\hat{k}, \hat{\theta})},$$

where $L(\hat{k}, \hat{\theta})$ is the likelihood evaluated at its maximum, i.e. $\hat{k}$ and $\hat{\theta}$ are the best-fit values for the parameter k and for all the other parameters, respectively; instead $\hat{\hat{\theta}}_k$) represent the best-fit values of the parameters other than k, obtained from a fit performed while keeping the parameter k fixed. Such likelihood ratio is a function of the sole parameter k, while all the other parameters are *profiled* away. This profile likelihood ratio has the useful property that its natural logarithm, and more specifically the quantity $-2ln\lambda(k)$, is distributed like a χ^2, with one degree of freedom, in the limit of large statistics, usually referred to as *asymptotic regime*. Thanks to this property, the uncertainty on the parameter k can be easily defined as the shift from its best-fit value needed to increase the quantity $-2log\lambda(k)$ by 1: $\hat{\sigma_k} : -2log\lambda(\hat{k} \pm \hat{\sigma_k}) = 1$.[6]

[6]Notice that this formula assumes a perfectly symmetric behaviour of the negative-log-likelihood-ratio with respect to the parameter k. If this symmetry is not realised, different values are found for the positive and negative errors.

3.4.2 Significance of a measurement

In order to claim the observation of a process, a quantification of the *significance* of an excess of measured events is needed. Such significance is defined as the degree of incompatibility of the observed data with the hypothesis of the absence of signal, or the *background-only* hypothesis. In general, the significance can be computed as a transformation of the so-called p_0-value. The p_0-value is defined as the probability, while assuming a certain *null hypothesis* (typically corresponding to the absence of signal process), of getting data of equal or greater incompatibility than the observed one, with the predictions of the null hypothesis. The incompatibility with the observed data is defined through a *test statistics*.

In the case of the profile likelihood formalism, the best test statistics are proven to be the profile likelihood ratio, $\lambda(\mu)$, or one of its transformations (e.g. $-2log\lambda(\mu)$). Here we considered the signal strength μ of the tested POI (as in eq. 3.9). At this point, two hypotheses are tested and compared: $H = H_1$, which refers to the presence of a signal ($\mu > 0$), and the null hypothesis $H = H_0$, i.e. no signal and only background ($\mu = 0$). The profile likelihood ratio is then built from this hypothesis:

$$\lambda(0) = \frac{L(\mu = 0, \hat{\hat{\theta}}_0)}{L(\hat{\mu}, \hat{\theta})},$$

where the numerator corresponds to the hypothesis H_0, as it is a *conditional* maximized LH function for the given value of $\mu = 0$, while the denominator corresponds to H_1, as it is a *unconditional* maximized LH function, where the presence of a signal is allowed. In other words, this test statistic tells how much better the data are fitted by a signal-plus-background model with respect to a background-only model. Finally, the p-value p_0 for the negative log of this ratio is computed, as the probability of getting a worse value of the test statistics than the observed one, usually exploiting the known properties of the profile likelihood ratio, in the asymptotic regime. Finally, the significance of the signal process is computed as $Z_0 = \Phi^{-1}(1-p_0)$, where Φ^{-1} is the quantile (inverse of the cumulative distribution) of the standard Gaussian probability distribution.

3.4.3 Exclusion Limits

In the context of searches for new physics, in the absence of significant excess in the data over the background-only prediction, upper limits at 95% confidence level (CL) are usually derived for the predicted signal process. The upper limits are calculated using the CL_s frequentist formalism, described in [95]. The test statistic used for the limit computation is defined as a profile likelihood ratio:

$$q_{\mu_{\text{signal}}} = -2ln(\frac{L(\mu_{\text{signal}}, \hat{\hat{\theta}}_\mu)}{L(\hat{\mu}_{\text{signal}}, \hat{\theta}_\mu)}) \tag{3.10}$$

where $\hat{\mu}_{\text{signal}}$ and $\hat{\theta}_\mu$ are the best-fit values of the signal strength and NPs, while $\hat{\hat{\theta}}_\mu$ is the best-fit values for the NPs for a given value of μ_{signal}. Also for this exclusion limit computation, the asymptotic behaviour of the negative-log-likelihood ratio is exploited. In addition, so-called *expected* exclusion limits are computed together with the observed ones. These are obtained by applying the same procedure as for the extraction of the observed exclusion limits but using a synthetic dataset built from the background-only model. Usually, such expected limits are reported together with corresponding $\pm1\sigma$ and $\pm2\sigma$ bands, obtained by sampling the synthetic dataset used for the extraction of the expected limits with Poisson fluctuations.

Chapter 4

Cross-section Measurement: single top t-channel

4.1 Introduction

This Chapter presents the first measurement of t-channel single-top-quark production in pp collisions at the centre-of-mass energy of $\sqrt{s} = 5.02$ TeV, based on 255 pb^{-1} of ATLAS data collected during LHC *Run 2*. As mentioned in 1.2.1 top quarks are mainly produced in pairs at the LHC, but are also produced as singly, via EW processes. Among the three main production modes of single top, the t-channel one is the primary at the LHC. Single top t-channel production has been measured for the first time at the Tevatron at $\sqrt{s} = 1.96$ TeV [96, 97] in $p\bar{p}$ collisions, and then at the LHC at 7 and 8 TeV in pp collisions both by the ATLAS and CMS collaborations [98, 99, 100, 101]. Measurements of the t-channel production cross section at $\sqrt{s} = 13$ TeV have been performed by the CMS collaboration with an integrated luminosity of $\int L = 35.9$ fb^{-1} [102], and more recently by the ALTAS collaboration with the full *Run 2* dataset corresponding to $\int L = 139$ fb^{-1} [103].

The result presented in this Chapter is based on data collected during the LHC *Run 2* corresponding to an integrated luminosity of just $\int L = 255$ pb^{-1}, which is three orders of magnitude less than the full *Run 2* dataset. This dataset has a lower centre-of-mass energy with respect to the one at which the rest of *Run 2* data are collected, i.e. $\sqrt{s} = 13$ TeV, as well as a lower instantaneous luminosity and a consequent lower pile-up . These pp collisions were produced at the LHC as a reference run for the subsequent heavy-ion collisions ($PbPb$) at the same energy per nucleon of $\sqrt{s_{NN}} = 5.02$ TeV. The proportionality among the energy and the production cross-section is tested, as well as higher-order corrections from QCD [104] at a different energy than those studied before. The measurement of t-channel single-top production process, and in particular the extraction of the cross-section ratio between top-quark and anti-top-quark production, is a useful input to global PDF fits, as the different centre-of-mass energy can provide different constraints.

At the energy of $\sqrt{s} = 5.02$ TeV the single top t-channel cross-section predicted by the SM is $\sigma^{SM}(tq+t\bar{q}) = 30.3^{+0.5}_{-0.5}$ pb. The observation of this process with a small dataset of $\int L = 255$ pb^{-1} is therefore challenging. The analysis presented in this thesis addresses this challenge through the optimization of event selection to improve the signal-to-background ratio, as well as enhancing the separation between signal and background events using machine learning algorithms. This Chapter shows the details of the observation of t-channel single top production, as well as of the measurement of the total cross-section for this process($\sigma(t+\bar{t})$), of the same cross-section for the top and anti-top separately ($\sigma(t)$ and $\sigma(\bar{t})$) and their ratio R_t. The

results discussed here have been recently published in a paper [6].

4.2 Data and Monte Carlo samples

In the following sections, a short description of the physics object definitions, data and Monte Carlo samples used in the t-channel single top cross-section measurement is given. This section frequently references Chapter 3, where a more detailed description of the general concepts is given.

4.2.1 Physics Object Definition

Electron candidates are reconstructed and identified as explained in section 3.1.2. In this analysis, electrons are required to pass the `Medium` likelihood identification criteria, to have $|\eta_{cluster}| < 2.47$ and not to be in the *crack region*. The isolation criteria used for electrons is the `FixedCutTight` out of those detailed in 3.1.2.

Muon candidates are reconstructed and identified as explained in section 3.1.3. They are required to pass the `Medium` identification criteria, and to have $|\eta| < 2.5$. The isolation criteria used for muons is the `FCTight_FixedRad` out of those detailed in 3.1.3.

Both electrons and muons are matched to the primary vertex using standard impact parameter requirements in ATLAS (see section 3.1.1) by requiring that $|z_0 sin(\theta)| <$ 0.5 mm and $|d_0/\sigma(d_0)| < 5(3)$ for electrons (muons). Calibration and efficiency scale factors are applied to electrons and muons derived for the low pileup $\sqrt{s} = 5.02$ TeV dataset [105, 106].

Jets are reconstructed using the anti-k_T jet algorithm [64] with a radius parameter of $R = 0.4$ following the procedure for the *standard jets* as explained in 3.1.4. They are required to have $p_T > 20$ GeV and $|\eta| < 4.0$. Jet energy calibrations derived for the standard ATLAS *Run 2* measurements with high-pileup at $\sqrt{s} = 13$ TeV need an additional correction due to different pile-up, detector and reconstruction conditions for the dataset at $\sqrt{s} = 5.02$ TeV used in this analysis. This additional jet in-situ correction is derived with a dedicated Z/jet p_T balance study detailed in section 3.1.4.

Out of the jet objects, b-jets are identified with the DL1r algorithm, as detailed in section 3.1.7. The DL1r working point used in this analysis is the tighter one, to increase the signal purity, with a 60% b-tagging efficiency for b-jets in simulated $t\bar{t}$ events. It corresponds to a rejection factor of ~ 1155 against jets originating from light quarks[1], and ~ 29 against jets originating from charm quarks. The b-tagging efficiencies at $\sqrt{s} = 5.02$ TeV were found to be compatible with those at $\sqrt{s} = 13$ TeV within the statistical uncertainty, from a previous ATLAS measurement [68], thus the same is used in this analysis.

To prevent double counting of hits, energy deposits in the detector, reconstructed tracks or clusters, the ATLAS standard overlap removal procedure is used as described in 3.1.5. The calculation of the missing transverse momentum E_T^{miss} also uses the standard ATLAS definition explained in 3.1.6.

4.2.2 Data Samples

The analysis is performed using the dataset collected during November 2017 LHC pp collision run at $\sqrt{s} = 5.02$ TeV and corresponds to a total integrated luminosity

[1]Here the up, down and strange quarks are referred to as light quarks.

of 254.94 pb^{-1} after data quality requirements [107]. The average number of interactions per bunch crossing ranged from approximately 0.5 to 4 in this dataset with an average of $\langle\mu\rangle = 2$ interactions per bunch [108].

4.2.3 Monte Carlo Simulation

The analysis relies on dedicated Monte Carlo (MC) samples generated at $\sqrt{s} =$ 5.02 TeV. These samples are processed through the full ATLAS detector simulation based on GEANT4 [88]. Both the MC samples and data samples are elaborated with the same reconstruction algorithms and analysis chain.

The signal t-channel single-top quark process is simulated using the POWHEG BOX V2 [109] generator with PYTHIA 8.2 program [85] with the NNPDF3.0 NNLO PDF set [110] and the A14 tune [90] for parton showering and hadronization. Two alternative signal samples are generated: the first uses the same MC generators with the difference that the `pTHard` parameter in PYTHIA 8 is changed from zero to one, while the second is generated with POWHEG+HERWIG7.1.6 generator. All the signal samples are normalized to NNLO cross-sections calculated using the MCFM program [111]. The inclusive cross-section is $\sigma^{theory}(t+\bar{t}) = 30.3^{+0.4}_{-0.3}(\text{QCD scales})^{+0.6}_{-0.4}(\text{PDF}+\alpha_s)$, where the quoted uncertainties include those from the choice of renormalization scale μ_R, the factorization scale μ_F, the uncertainty on the PDFs, and the uncertainty in the value of the strong coupling constant α_s.

The $t\bar{t}$, Wt and s-channel background processes are all similarly generated using POWHEG BOX V2 interfaced with PYTHIA 8.2 , using the NNPDF3.0 NNLO PDF set and the A14 tune. An alternative $t\bar{t}$ sample is produced using POWHEG+HERWIG7.1.6 for parton showering. All $t\bar{t}$ samples are normalized to NNLO+NNLL cross-section predictions calculated with the TOP++ (v2.0) program [112].

For the Wt associated production process, the diagram-removal (DR) scheme [113] is used to handle the interference between the $t\bar{t}$ and Wt final states. Two alternative samples are generated: one with the same generators but using the diagram subtraction (DS) scheme [114], and the second one using POWHEG+HERWIG7.1.6 [86]. The Wt samples are normalized to NLO+NNLL cross-section calculations [115]. An alternative sample for the s-channel process is produced with the POWHEG+HERWIG7.1.6 generator. The s-channel samples are normalised to the generator-level NLO cross-section prediction.

The W+jets and Z+jets backgrounds are simulated with SHERPA2.2 [84] using NLO matrix elements with up to two partons, and LO matrix elements for up to four partons, with the NNPDF3.0 NNLO PDF set. These samples are normalized using NNLO cross-section predictions [116]. The diboson production (WW, WZ, ZZ) is generated using SHERPA2.1 with the CT10 PDF set [117].

For all samples, heavy flavour decays are modelled using the EVTGEN-1.6.0 program [87], except for processes modelled using the SHERPA generator. The top-quark mass is set to $m_t = 172.5$ GeV in all relevant samples.

Small corrections derived from data-to-simulation comparisons are applied as scale factors (SFs) to the simulated lepton trigger and reconstruction efficiencies to improve agreement with the observed data response.

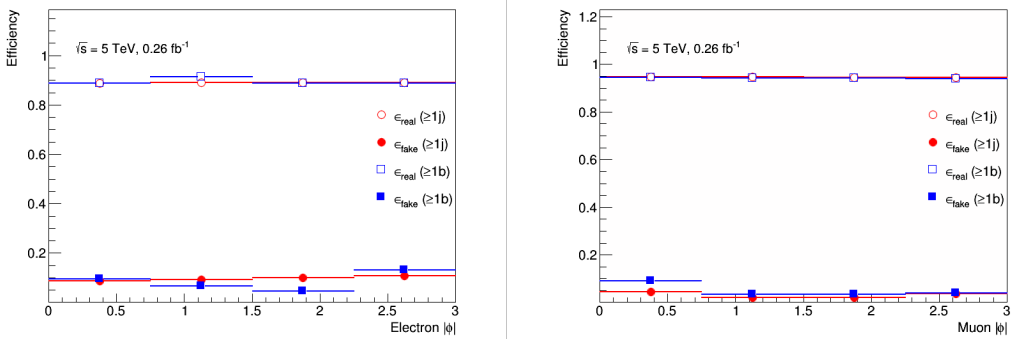
4.2.4 Fake and non-prompt lepton background estimation

In the single top t-channel analysis, as detailed in 4.3.1, at least one lepton is required in the final state. This selection is needed to identify the leptonic decay of the W boson produced by the top quark. Non-prompt leptons, hadrons, and photons can satisfy the lepton selection criteria and therefore create a *non-prompt/fake*

background. This background is not well described by MC simulation and is estimated with a data-driven method: the Matrix Method. A detailed explanation of this background and of the method is given in a previous section 3.3. The tight and loose samples in the Matrix Method for this analysis are selected, after imposing the final kinematic selection as in 4.3.1, based on the lepton tight or loose requirements, defined in 4.2.1.

The real efficiency (ε_r) is measured in MC simulation in a region dominated by Z boson and $t\bar{t}$ production: events are split into two regions if they contain two electrons ($e^{\pm}e^{\mp}$) or two muons ($\mu^{\pm}\mu^{\mp}$) of opposite electric charge, they are required in both cases to have at least one b-tagged jet and an invariant mass of the two leptons of 80 GeV $< m_{\ell\ell} <$ 100 GeV. Fake efficiencies (ε_f) are extracted from MC simulation and not directly from data due to limited statistics, considering events that fail the *real lepton* selections (as defined in 4.3.1). The electron ε_f is computed in events with one electron and one muon with the same electric charge ($e^{\pm}\mu^{\pm}$), while muon ε_f in events with two muons with the same electric charge ($\mu^{\pm}\mu^{\pm}$).

The main source of *fake/non-prompt* events for electrons are coming from light flavour hadrons from hadrons, while for muons it is non-prompt muons from heavy-flavour hadron and τ-lepton decays. The ε_r and ε_f are applied separately for electrons and muons and parameterised in bins of leptons $|\phi|$, as shown in Figure 4.1. The choice of the parametrisation is done after studying basic kinematic distributions such as lepton $|\eta|$ and $|\Delta\phi(\ell, \text{jet})|$.



(a) Electron efficiencies.

(b) Muon efficiencies.

Figure 4.1: The ε_r and ε_f efficiencies for electrons 4.1(a) and muons 4.1(b) as a function of the lepton $|\phi|$.

4.3 Analysis Strategy

The basic structure of the adopted analysis strategy comes from the typical topology of the signal events, as depicted in the representative Feynman diagram shown in Figure 4.2. The final state is characterized by a single charged lepton, and a neutrino from the W boson decay, a b-jet originated from the top-quark decay as the W, and finally, another jet, created by the *spectator quark*, is the result of the diffraction of the incoming quark.

In this context, the t-channel single top production process is defined as the signal, while background processes include top quark pair production, W+jets production, other single top production modes and fake and non-prompt lepton background.

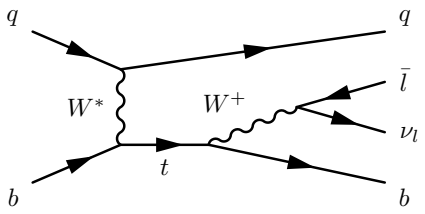


Figure 4.2: Representative tree-level Feynman diagram of the t-channel single top-quark production and leptonic decay of the top quark.

4.3.1 Event Selection

Based on the t-channel production final states, all the candidate events are required to pass a single electron or a single muon trigger. Both the triggers used are fully efficient for leptons with a $p_T > 18$ GeV and therefore only events passing this requirement are further considered [118, 119].

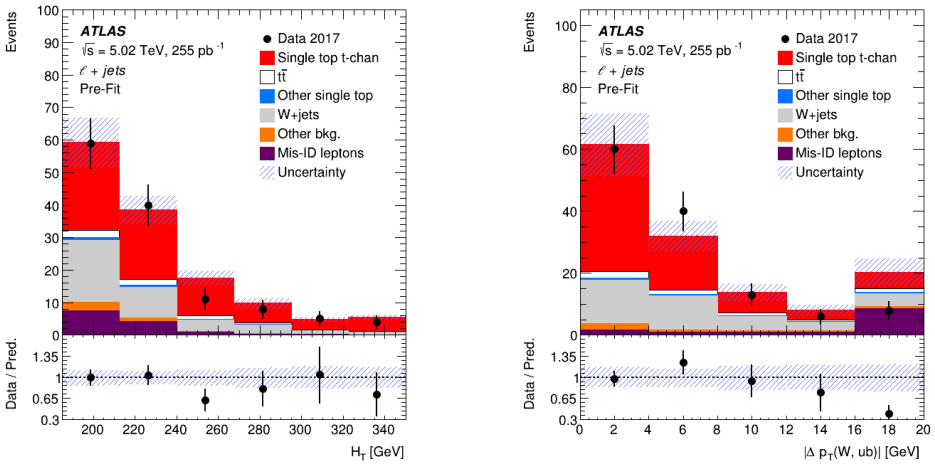
At least one reconstructed vertex with at least two associated tracks that have a $p_T > 0.5$ GeV is required for all events. When multiple vertices are reconstructed, the vertex with the largest sum of the squared p_T is taken [120].

Events are required to have at least one lepton candidate (an electron or muon object as defined in 4.2.1), exactly two reconstructed jets with $p_T > 23$ GeV, and exactly one jet tagged as a b-jet with the `DL1r` 60% WP. The b-jet is required to have $|\eta| < 2.5$, while the other jet is selected if $|\eta| < 4.0$. This selection, for the candidate jet originating from the spectator quark, includes also forward jets to enhance the signal-over-background ratio, as better explained in 4.3.4. Events are also required to have $E_T^{miss} > 15$ GeV.

Additionally, to reduce the contribution from the $t\bar{t}$ background process the candidate b-jet and the spectator jet are required to have $\Delta\eta > 1.5$. To mitigate the effect of multi-jet background the *triangular cut* is applied, defined as $(E_T^{miss} + m_W^T) > 70$ GeV, as well as the selection $m_T^W > 35$ GeV. These two selections aim to reduce the multi-jet background containing a fake or non-prompt lepton, which tends to have low E_T^{miss} and m_T^W with respect to the signal process. Here the transverse W boson mass is defined as $m_T^W = \sqrt{2p_T^{lep}E_T^{miss}(1 - \cos(\Delta\phi(l, \vec{p}_T^{miss})))}$. A further selection requiring $H_T > 185$ GeV is applied, with H_T defined as the sum of the transverse momentum of the lepton, of those of the jets and of E_T^{miss}, to reduce the W+jets background contribution. A cut on the invariant mass of the lepton–b-jet system, $m_{tb} < 165$ GeV, is applied to reject background events from processes not involving a top-quark. Finally, to further characterize the t-channel production process, events are selected if they have $m_W < 102$ GeV and 140 GeV $< m_{top} < 225$ GeV, where m_W and m_{top} are the reconstructed W-boson and top-quark mass values, as described in the next Section 4.3.2. Figure 4.3, shows the data and MC distributions for events passing the event selections here described for the H_T observable and for the absolute value of the p_T difference between the W boson and the up-quark and b-quark system.

4.3.2 Top quark reconstruction

The reconstruction of top quarks from their decay is a recurring problem of top-quark-related analyses. Reconstruction techniques can differ significantly based on the decay channel of the top. In the analysis described in this Chapter the top-quark decays leptonically. The main challenge in this final state is related to the neutrino,



(a) H_T distribution.

(b) $|\Delta p_T(W, ub)|$ distribution.

Figure 4.3: The H_T in 4.3(a) and $|\Delta p_T(W, ub)|$ in 4.3(b) distributions for data (black dots) and MC samples after event selections. The error bars on the dots represent the statistical uncertainty on the data while the blue cross-hatched lines correspond to the total uncertainties on the MC prediction. The lower panels show the ratio between the data and the prediction.

which is not detected by the ATLAS detector.

The first step in this procedure is to reconstruct the W boson from its decay products. Its four-momentum can be obtained as $P_W = P_l + P_\nu$. Here the lepton momentum P_l is known with good accuracy, while the neutrino momentum P_ν can only be estimated from the E_T^{miss} quantity and therefore only the x and y components can be known. Since the neutrino is originated from the decay of the W boson, the pole mass of the latter can be used as a constraint, together with the requirement that the neutrino is massless. In this way the z component of the neutrino four-momentum can be extracted as:

$$P_{\nu,z}^2 - 2 \cdot \frac{\mu \cdot P_{l,z}}{E_l^2 - P_{l,z}^2} \cdot P_{\nu,z} + \frac{E_l^2 \cdot p_T^2 - \mu^2}{E_l^2 - P_{l,z}^2} = 0 \tag{4.1}$$

with

$$\mu = \frac{m_W^2}{2} + \cos\Delta\phi(l, E_T^{miss}) \cdot p_{l,T} \cdot p_{\nu,T} \tag{4.2}$$

where $P_{l,z}$ and E_l are the z-component of the momentum and the energy of the charged lepton respectively, $p_{l,T}$, $p_{\nu,T}$ are the transverse momentum of the charged lepton and neutrino, $\Delta\phi(l, E_T^{\text{miss}})$ is the azimuthal angle difference between the charged lepton momentum and the missing transverse momentum. In general, two solutions $P_{\nu,z}^{A,B}$ of this quadratic equation can be found:

$$P_{\nu,z}^{A,B} = \frac{\mu \cdot P_{l,z}}{P_{l,T}^2} \pm \sqrt{\frac{\mu^2 \cdot P_{l,z}^2}{P_{l,T}^4} - \frac{E_l^2 \cdot p_{\nu,T}^2 - \mu^2}{P_{l,T}^2}} \tag{4.3}$$

In the case of two real roots in the quadratic function, the root which gives a reconstructed top-quark mass closest to the on-shell top-quark mass $m_{top} = 172.5$ GeV is chosen. In the case of an imaginary root, which can occur due to the limited E_T^{miss} resolution, only the real component is considered.

With the z-component of the neutrino's four-momentum, the four-momentum of the W-boson is calculated. The top-quark is then reconstructed from its decay products: the four-momentum of the top-quark is calculated as $P_{top} = P_W + P_b$, where P_b is the 4-momentum of the b-tagged jet which is also measured with good accuracy.

4.3.3 Signal Discrimination

To achieve a high accuracy when measuring the t-channel single top production process it is important to have a good separation between the signal process versus the background ones. In this respect, a machine learning approach is used in this analysis: a Boosted Decision Tree (BDT) is implemented with the `XGBoost` package [121].

A BDT is an algorithm that enhances classification performance by combining multiple decision trees into a single classifier. The algorithm builds these trees sequentially, with each subsequent tree focused on correcting misclassifications from previous trees through the *boosting* process. While individual trees in the ensemble are intentionally kept simple, their combination yields a robust classifier. The boosting process iteratively assigns higher weights to misclassified events, ensuring that the algorithm focuses on challenging cases.

Training step

In the training step, the BDT is provided with simulated samples of signal events and of all the background processes considered in this analysis: $t\bar{t}$, W+jets, tW, s-channel, Z+jets, and diboson production. The multi-jet fake/non-prompt lepton background sample estimated with the Matrix method is also given as input to the BDT training. The truth MC classification information for all events is passed to the BDT so it can learn to recognize signal versus background events.

The BDT is trained on MC datasets without information on the MC event weights, that are used in the statistical analysis described in Section 4.4, and that match the physical yields for each process. This way the number of events used in the training for each background process would not match the correspondent physical yields. Therefore, background samples are scaled so that proportions between the number of events for background processes used in the BDT match the correspondent physical yields.

In total 30 different input features for the BDT have been considered, with information on invariant mass of different combinations of final state particles, reconstructed intermediate particles, angular and transverse momentum variables of the final state objects, difference between kinematic observables of the final state particles and more complicated functions of kinematic variables of the leptons and jets in the final state. The BDT provides a ranking of input variables through `XGBoost` importance metric. The feature importance is calculated using a weight, which represents the relative number of times a variable is used to split the data across all trees. An optimization study on the number of input variables to use has been performed: only the first 5, 9 or 25 highest-ranked input variables are used. The results are evaluated by looking at the output metrics of the BDT such as the Area Under the Curve (AUC)[2] and accuracy[3], as well as the separation S of the BDT output

[2]The AUC represents the probability that a randomly chosen signal event is ranked higher than a randomly chosen background event by the classifier. It is defined as the Area Under the ROC (Receiver Operating Characteristic) curve).

[3]Accuracy is defined as the ratio of correctly classified events (both signal and background) to the total number of events.

variable between the signal sample and the background processes, with computed as $S = 1/2\ {}^{(s-b)^2}/_{(s+b)}$. Taking into account that a smaller number of input features for the BDT corresponds to lower computing resource usage and a reduced risk of overtraining, the optimal number of input features chosen is 9. Table 4.1 lists the 9 variables, with a short description.

Feature	**Definition**
m_{lb}	Invariant mass of the lepton (l) and the tagged jet (b)
$H_{\mathrm{T}}(l, \mathrm{jets}, E_T^{miss})$	Scalar sum of p_T of the charged lepton, the jets, and E_T^{miss}
$\lvert\eta(u)\rvert$	Absolute value of the η of the light quark jet (u)
$\lvert\Delta p_T(W, ub)\rvert$	Absolute value of the scalar difference in p_T between the reconstructed W-boson and the jet pair
$\lvert\Delta\eta(b, l)\rvert$	Absolute value of the difference in η between the charge lepton and the tagged jet (b)
$\lvert\Delta\eta(u, l)\rvert$	Absolute value of the difference in η between the charge lepton and the light-quark jet (u)
$\lvert\eta(l)\rvert$	Absolute value of η of the charged lepton
$m(lvb)$	Invariant mass of the reconstructed top-quark
Transverse sphericity	$\frac{2\lambda_2}{\lambda_1+\lambda_2}$ where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and $\lambda_1, \lambda_2, \lambda_3$ are eigenvalues of $S^{ab} = \frac{\sum_i p_i^a p_i^b}{\sum_i p_i^2}$; $a, b = x, y, z; i =$ jet index.

Table 4.1: The 9 input features used in the training of the Boosted Decision Tree.

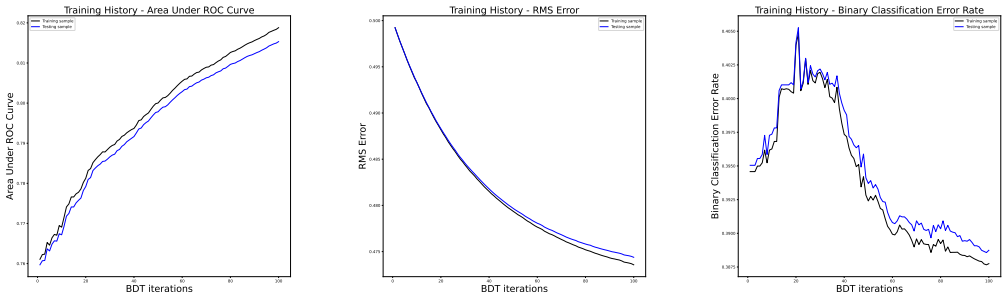
The hyper-parameters of the BDT have also been studied and optimized to obtain a minimal amount of computing resources while not decreasing the BDT output performances. The performance evaluation here is the same as for the number of input features described above. The BDT hyper-parameter chosen values are listed in Table 4.2.

Parameter	**Value**
Number of Trees	100
Maximum Depth	5 nodes
Learning Rate	0.01
Number of Epochs	100
Loss Function	Binary Cross-Entropy

Table 4.2: BDT training hyper-parameters values.

Testing step

A test for over-training is performed: the discriminant constructed from the training (the BDT output variable) is applied to a statistically independent sample that has not been used in the training step. Figure 4.4 shows different metrics evaluated on the training and testing samples. The test step is performed using the *k-fold cross validation* technique. The training set is split into three equal-sized partitions (folds). One of them is designated as the *test sample* and the other two as the *training sample* and the BDT is then trained and tested with this scheme. This procedure is repeated until each fold has been used as the test set exactly once. This ensures an orthogonal test set for each training and data efficiency is gained. The Receiver



(a) Area Under the ROC curve. (b) Root mean square error. (c) Mean square error.

Figure 4.4: Metrics used to evaluate the model performance, in black for the training sample and in blue for the testing sample.

Operating Characteristic (ROC) curve is a graphical representation used to assess the performance of a binary classifier. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The TPR is the ratio of True Positives (TP) to the actual positives (TP + FN), and the FPR is the ratio of False Positives (FP) to the actual negatives (TN + FP). The ROC curve is shown in Figure 4.5 for each of the folds and their average.

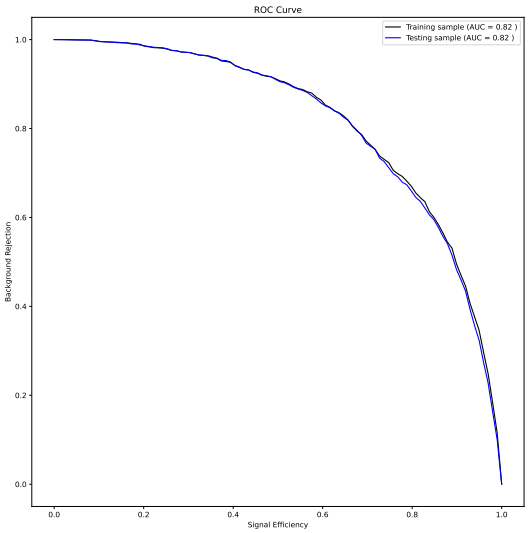
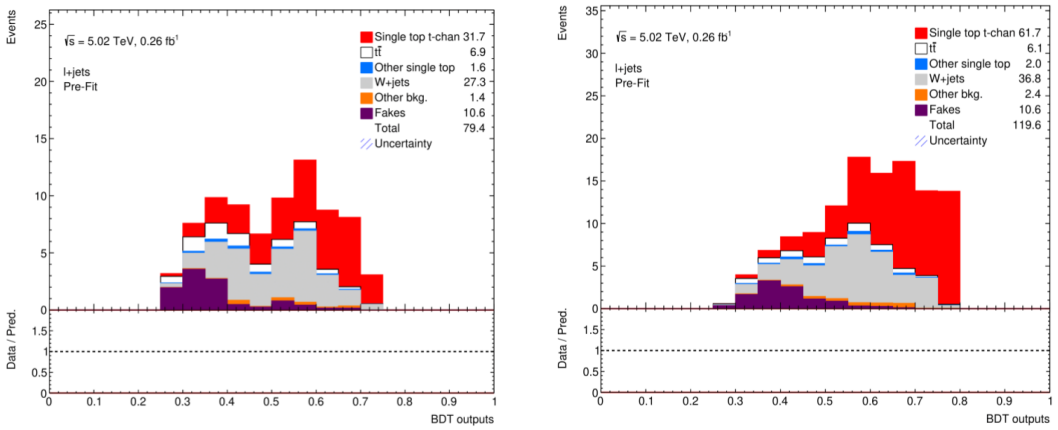


Figure 4.5: Background rejection versus signal efficiency obtained by varying the threshold value on the BDT output variable for the events of the test sample showing the 3 folds for the training region. The average ROC curve value of each of the individual folds is shown as a solid blue line

4.3.4 Forward jet selection optimization

In the ATLAS collaboration, the standard jet physics objects used in physics analyses are required to be reconstructed in the central region of the detector, $|\eta| < 2.5$. Anyway, as it can be seen in the Feynman diagram in Figure 4.2, the jet corresponding to the *spectator quark* tends to be in the forward region ($|\eta| > 2.5$).



(a) Jet $|\eta| < 2.5$.

(b) Jet $|\eta| < 4.0$.

Figure 4.6: BDT output distribution for the t-channel single top production signal and the background processes considered MC samples. On the left 4.6(a) when the jet pseudo-rapidity selection is $|\eta| < 2.5$, while on the right 4.6(b) for the jet pseudo-rapidity selection is $|\eta| < 4$. It is evident how in the latter case the signal events are larger with respect to the background events than in 4.6(a). It can also be seen that the BDT is able to distinguish better the signal process versus background ones in the case of 4.6(b).

To find the optimal value to select the pseudo-rapidity of the candidate jet, various selection values have been considered to optimize the signal over background ratio. Table 4.3 shows the values considered for the pseudo-rapidity jet selection and the corresponding separation value between signal and background computed with respect to the Boosted Decision Tree (BDT) output variable (detailed in Section 4.3.3). It is clear that a higher cut on the absolute value of the pseudo-rapidity

$\|\eta\|_{\max}$	Separation (%)
2.5 (standard)	16.4
3.0	22.5
3.5	22.5
4.0	26.1

Table 4.3: Separation between signal and background processes for the BDT output distribution for different $|\eta|_{\max}$ selections.

that enhances the signal-over-background ratio, also improves the BDT ability to separate signal versus background. As an example in Figure 4.6 the BDT output is shown for two values considered for the spectator jet $|\eta|$.

4.4 Statistical Analysis and Systematics

After the training and testing step of the BDT, the architecture obtained is applied to whole data and MC samples. The BDT output distribution is then used to perform a binned Profile Likelihood Fit as described below in Section 4.4.2, which takes into account the statistical uncertainty, as well as all possible sources of systematic uncertainties, detailed in the next Section (4.4.1).

4.4.1 Systematic Uncertainties

Several sources of systematic uncertainties that can affect the normalization and shape of the signal and background BDT output distributions are considered.

Luminosity and Pileup

The luminosity for the low-pileup pp collision data recorded by ATLAS at $\sqrt{s} = 5.02$ TeV has been calibrated following the methodology discussed in [122], where the uncertainty on the integrated luminosity is computed to be of 1.0%. As the level of pileup in the dataset used in this analysis is low ($\langle\mu\rangle = 2$), no separate pileup re-weighting uncertainty was included.

Reconstructed physics objects

Uncertainties in the reconstruction, identification and calibration of the physics objects are considered. Leptons have uncertainties related to the reconstruction, identification and isolation SFs, as well as uncertainties relative to the electron and muon energy scale and resolution as explained in 3.1.2, 3.1.3.

Uncertainties related to the JES and JER are included as for ATLAS standard analyses with pp collision data at $\sqrt{s} = 13$ TeV (more details in 3.1.4. Furthermore, since a dedicated JES in-situ correction is extracted for the low pileup $\sqrt{s} = 5.02$ TeV data, two additional non-closure systematic uncertainties on the jet energy scale are added. The uncertainty in the efficiency of the JVT selection is also added.

The uncertainties in the b-tagging calibration, determined separately for b-jets, c-jets and light-jets, are applied, to account for differences between data and simulation (more details in 3.1.7). The uncertainty in the E_T^{miss} is considered as explained in 3.1.6.

Uncertainties in signal and background modelling

Various uncertainties in the modelling of the signal process are considered. Here a list of those applied in this analysis is given, while an explanation can be found in Section 3.2.3:

- missing higher-order QCD corrections in the ME computation;
- uncertainty on the amount of ISR, and FSR;
- uncertainties on the choice of parton shower and hadronisation model;
- uncertainty in the matching between ME computation and PS model;
- PDF uncertainty for the PDF set used in this analysis, the `PDF4LHC15`, is evaluated following the recommendation in [123]. The uncertainties due to the PDF choice are only estimated for the t-channel signal and $t\bar{t}$ background.

Uncertainties coming from the background processes modelling are also considered. Uncertainties due to higher-order QCD correction, ISR, FSR, choice of the PS and hadronisation model are considered for the $t\bar{t}$ sample in the same way as for the signal process. Moreover, an uncertainty on the choice of the h_{damp} parameter and on the choice of the ME generator are considered for this background process (as detailed in 3.2.3). Finally, an additional normalization uncertainty of $^{+7.5}_{-7.7}\%$ is applied on the $t\bar{t}$ cross-section [124].

Uncertainties on the multi-jet fake/non-prompt lepton background estimation are considered separately for the e+jets and the μ+jets channels and a conservative

50% normalization uncertainty is used. A shape uncertainty for this background was studied and found to be negligible and therefore not included in the results. However, to account for any possible shape differences in the multi-jet background estimation, the parameters referring to the statistical uncertainty for each bin (here referred to as γ_j) corresponding to this background process are separated from those of the other processes.

A normalization uncertainty of $^{+5.6}_{-5.5}$% is applied to the tW background cross-section, while 9.5% to the s-channel one [125]. The effects of the renormalization scale, factorization scale, ISR, and FSR are also evaluated for these two background processes as for the signal one. The tW simulated sample also takes into account the uncertainty on the diagram-removal technique, as detailed in 3.2.3.

A normalisation uncertainty based on the conservative Berends scaling [126] that is dependant on the number of jets found in a particular event is used to cover the higher order effects to the W+jets cross-section. A 34% normalisation uncertainty is used, which is further split into W+light jets, $W+\geq 1c$-jets, and $W+\geq 1b$-jets contributions. The effects of the renormalization scale and factorization scale are also considered for these processes, taking into account only the shape effect of the variation.

The normalisation uncertainty in the Z+jets and diboson backgrounds is conservatively taken to be 50%, accounting both for uncertainties in the cross-sections and in the acceptance.

4.4.2 Profile Likelihood Fit

The statistical analysis is performed based on a Profile Likelihood Fit to the BDT output distribution. A detailed explanation of the Profile Likelihood Fit formalism is given in Section 3.4.1.

In particular, selected events are split into two different categories, based on the electric charge of the lepton in the final state: the l^++jets and l^-+jets signal regions. This is done to differentiate the contribution of single top-quark versus antitop-quark t-channel production. A potential contamination in the two defined regions, caused by events of the other topology, can be caused by wrongly assigned charge leptons. The contribution of this contamination was checked in MC simulation to be less than $1^o/_{oo}$. The single-top-quark production cross-section (σ_t) signal strength μ_t, and the single-antitop-quark production cross-section ($\sigma_{\bar{t}}$) signal strength $\mu_{\bar{t}}$ are extracted from the Profile Likelihood fit to the BDT response distributions in the l^++jets and l^-+jets signal regions.

The total single-top-quark t-channel production cross-section ($\sigma_{t-chann}$) and the ratio R_t between σ_t and $\sigma_{\bar{t}}$ are extracted from the same fit by parametrizing the individual signal strengths μ_t and $\mu_{\bar{t}}$ as functions of the t-channel production cross section signal strength $\mu_{t-chann}$ and the signal-strength ratio μ_{R_t} as in Equation 4.4:

$$\mu_t = \mu_{t\text{-chan.}}\mu_{R_t}\frac{R_t^{\text{predicted}}+1}{\mu_{R_t}R_t^{\text{predicted}}+1}, \mu_{\bar{t}} = \mu_{t\text{-chan.}}\frac{R_t^{\text{predicted}}+1}{\mu_{R_t}R_t^{\text{predicted}}+1} \tag{4.4}$$

Here $R_t^{\text{predicted}} = 2.03$. The significance on $\sigma_{t-chann}$ is measured according to the procedure defined in Sec. 3.4.2.

All experimental uncertainties described in Section 4.4.1 are included as NPs in the Profile LH fit. Systematic uncertainties with an effect less than 0.01% on the total yields in the signal region are pruned to improve the fit stability.

Shape effects which arise from statistical fluctuations in the systematic uncertainties distributions are mitigated by means of a smoothing procedure.

4.5 Results

The fit to data results in a signal strength value for the t-channel single-top production cross-section of $\mu_{t\text{-chan.}} = 0.89^{+0.14}_{-0.13}(\text{stat.})\ ^{+0.15}_{-0.12}(\text{syst.})$ and for the Ratio between top and antitop quark cross-sections of $\mu_{R_t} = 1.35^{+0.70}_{-0.41}(\text{stat.})\ ^{+0.50}_{-0.14}(\text{syst.})$. These signal strength values correspon to: $\sigma_{t-chann} = 27.1^{+4.4}_{-4.1}(\text{stat})^{+4.4}_{-3.7}(\text{syst})$ pb and $R_t = 2.73^{+1.43}_{-0.82}(\text{stat})^{+1.01}_{-0.29}(\text{syst})$ respectively. The Pearson correlation coefficient between $\sigma_{t-chann}$ and R_t is measured to be -30%. The significance of the signal observation is evaluated using the asymptotic approximation [127]. The background-only hypothesis is rejected with an observed (expected) significance of 6.1 (6.4) standard deviations.

The NPs shifts with respect to the input values given to the Profile LH fit, as well as their constraints, defined as post-fit uncertainties relative to their uncertainties prior to the fit (i.e. to the standard deviation of the associated constraint terms in the likelihood), are shown in Figure 4.7. The best-fit values of all NPs are found to be within 0.2 standard deviations from their input values and all constraints are smaller than 10%.

In Figure 4.8 the fitted NPs corresponding to the statistical uncertainty in the prediction of each bin (γ_j) are shown. It can be seen that the statistical uncertainty for the *fake/non-prompt* multijet background bins, especially for the μ+jets component, is quite large.

To check the fit stability, a profile likelihood scan was performed by scanning over various values of the two POIs and plotting the difference in the post-fit negative-log-likelihood function for the scanned value and the best-fit value of the POI, profiling over all the nuisance parameters at each scanned point. The resulting plot is shown in Fig. 4.9 and the two-dimensional scan is in Figure 4.10. These Likelihood scans highlight that the negative log-likelihood is a smooth function both for the signal-strength parameter related to the t-channel cross-section and for the parameter related to the ratio of the top- and antitop-quark cross-sections. The two curves are not symmetric in general, but zooming in around the minimum it can be verified that both of them are approximately symmetric for small variations.

To look which of the different sources of uncertainties have a higher impact on the final fit results, the so-called *ranking plots* are produced. The impact of each source of systematic uncertainty is obtained by comparing the best-fit values of the POI in the nominal fit to data and in an alternative fit, where the NP associated to the systematic uncertainty source is shifted and fixed to its $\pm 1\sigma$ variation. This procedure is repeated four times: twice to evaluate the impact up and down pre-fit, shifting the NP value to its 1σ variation before the fit, and twice to evaluate the post-fit impact shifting the NP by its post-fit positive 1σ and negative 1σ variation. In the ranking plot NPs are order from the top to the bottom based on their impact on the POI.

Figure 4.11 shows the ranking plots for both $\mu_{t\text{-chan.}}$ and μ_{R_t} after a fit to data. In each plot, the impact of the first 20 systematics is shown in light-blue (blue) for the down (up) variations, with an empty (full) box for the pre-fit (post-fit) impact, which are extracted as explained above. The ranking plots show also the relative shifts and constraints of the associated NPs: the post-fit best value of the fit is shown as a black dot, and its post-fit $\pm 1\sigma$ variation with a black line.

The total uncertainty in $\sigma_{t-chann}$ is dominated by the statistical uncertainty followed by uncertainties in the signal modelling, misidentified leptons modelling, JER, the in-situ correction, JES, and the modelling of $W+ \geq 1b$ jets. The total uncertainty on R_t is dominated by the statistical uncertainty, followed by uncertainties in

the misidentified leptons modelling, JER, the statistical uncertainty in the prediction, the JES the in-situ correction, and JES. The BDT output distribution for the l^++jets and l^-+jets regions after the Profile LH fit are shown in Figure 4.12.

As explained in Section 4.4.2, the single-top-quark and antitop-quark t-channel production cross-sections can be extracted from the fit results, and are measured to be $\sigma_t = 19.8^{+3.9}_{-3.1}(\text{stat})^{+2.9}_{-2.2}(\text{syst})$ pb and $\sigma_{\bar{t}} = 7.3^{+3.2}_{-2.1}(\text{stat})^{+2.8}_{-1.5}(\text{syst})$ pb. The uncertainties are calculated using error propagation techniques and account for the correlation between μ_{R_t} and $\mu_{t-chann}$. The t-channel single-top production cross-section depends on $f_{LV}^2 \cdot |V_{tb}|^2$ (f_{LV} is a left-handed form factor that is unity in the SM, V_{tb} is a component of the CKM matrix). From the fit to data, assuming that the CKM matrix elements $|V_{td}|, |V_{ts}| \ll |V_{tb}|$ and that the Wtb vertex is left-handed, the measured cross-section gives $f_{LV} \cdot |V_{tb}| = 0.94^{+0.11}_{-0.10}$ [6].

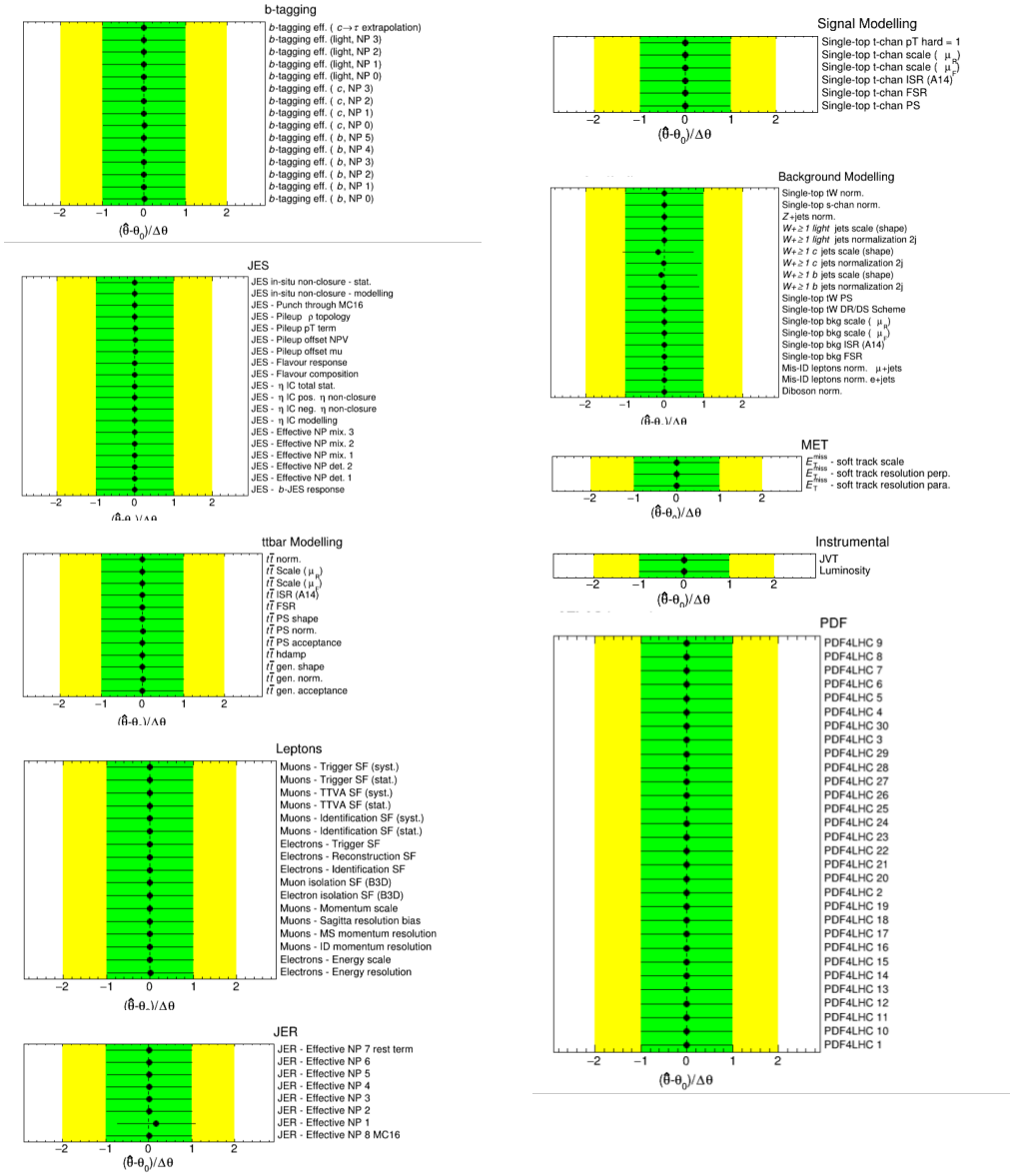


Figure 4.7: Nuisance parameter relative shifts and constraints after a fit to data using the BDT response distributions. NPs are arranged in different groups based on the different systematic categories they belong to. The NPs correspond to systematic uncertainties explained in 4.4.1.

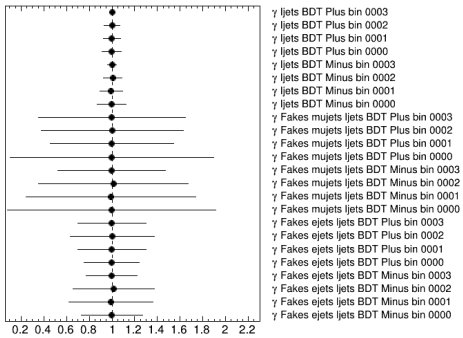


Figure 4.8: Nuisance parameter post-fit values and uncertainties corresponding to the statistical uncertainty (γ_j) after a fit to data using the BDT response distributions.

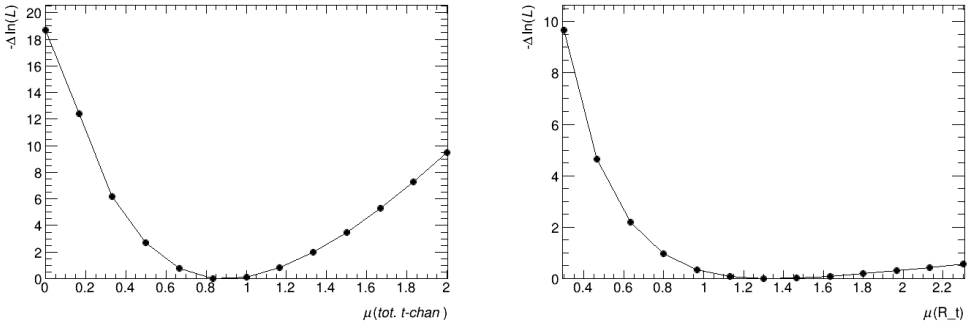


Figure 4.9: Likelihood scan for $\mu_{t\text{-chan.}}$ on the left, and μ_{R_t} on the right in a fit to data using the BDT response.

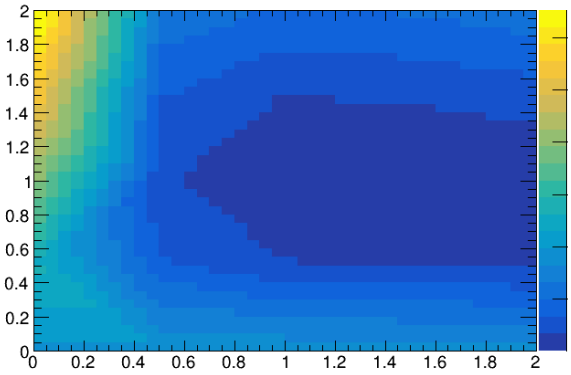


Figure 4.10: Two dimensional likelihood scan for μ_{R_t} and $\mu_{t\text{-chan.}}$ in a fit to data using the BDT response.

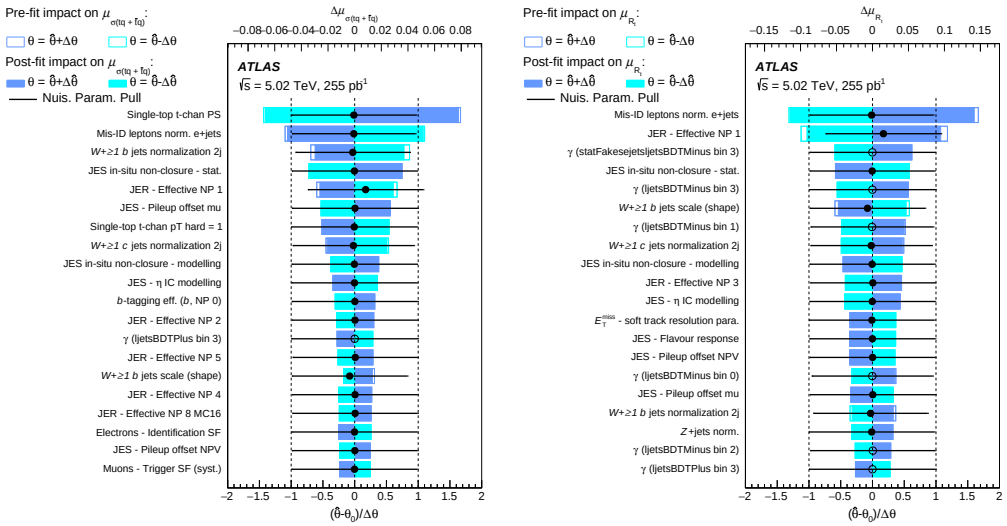


Figure 4.11: Ranking plot showing the impact of the top 20 sources of uncertainties on the signal strength in a fit to data with the BDT response distributions, together with the relative shifts and constraints on the associated nuisance parameters. The impacts are obtained as described in the text.. The ranking plot for $\mu_{t\text{-chan.}}$ is on the left and the plot for μ_{R_t} is on the right.

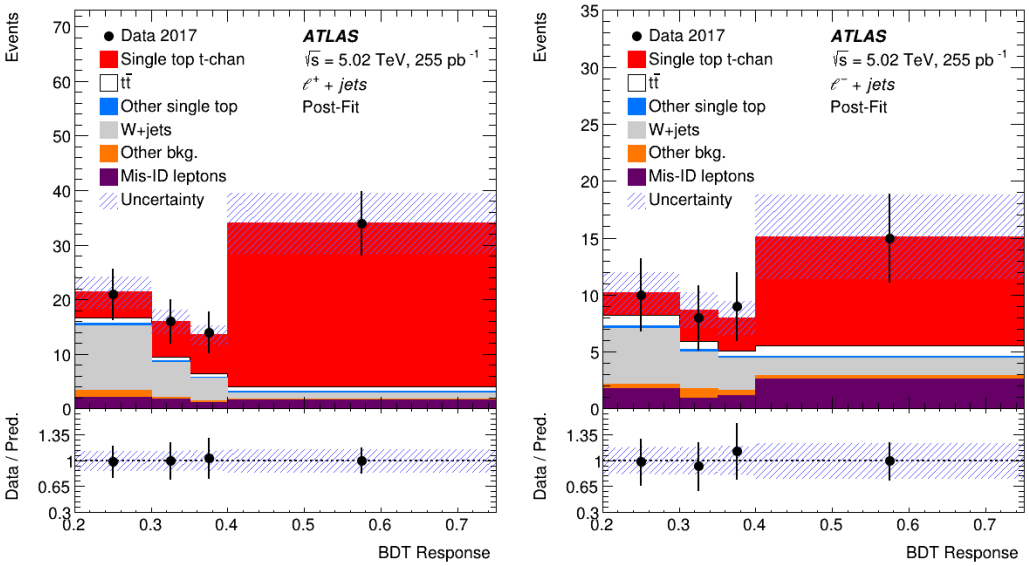


Figure 4.12: The post-fit BDT output distribution in the (left) l^++jets and (right) l^-+jets channels after a fit to data. The bottom panel shows the ratio of data to the total prediction.

4.5.1 Summary plots

Figure 4.13 shows on the left ATLAS measurements of the t-channel single top production process at different centre-of-mass energies, among which that described in this Chapter. The measurements are compared to the NLO prediction from the MCFM MC generator [128, 129]. On the right in Figure 4.13 the measured $\sigma_{t-chann}$ is compared with various PDF predictions at $\sqrt{s} = 5.02$ TeV.

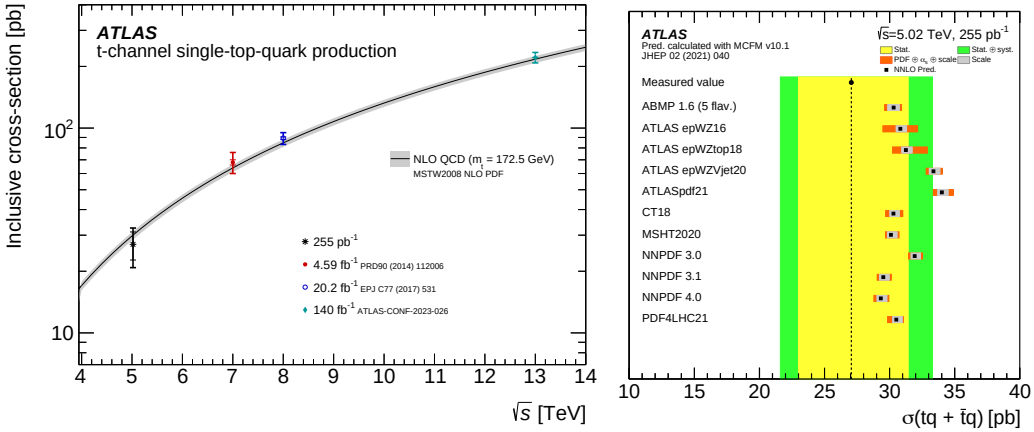


Figure 4.13: ATLAS measurements of the t-channel single top production cross-section at various energies in the centre-of-mass compare to theoretical calculations at NLO in QCD in (a). Comparison of the measured $\sigma_{t-chann}$ shown as a dashed line with its statistical (systematic) uncertainty in the yellow (green) band, compared to predictions based on different PDF sets.

Chapter 5

Search for New Particles: $t\bar{t}$ Resonances

5.1 Introduction

This Chapter presents a search for $t\bar{t}$ resonances with the full *Run 2* dataset corresponding to $\int L = 140\text{fb}^{-1}$ of pp collisions at an energy in the centre-of-mass of $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC.

As already discussed in Sections 1.2 and 1.3, several BSM theories predict heavy new particles coupling strongly to the top quark. Among these new theories, some of them predict additional heavy neutral bosons with spin-1 or spin-2: the Topcolor-Assisted-Technicolor (in particular the *Leptophobic Non-Standard Topcolor* Z' [36]) that include a spin-1 Z' particle; the Randall–Sundrum (RS) models of warped extra dimensions that cause Kaluza-Klein (KK) excitation of the graviton, corresponding to a spin-2 boson G_{KK} [8], or RS models with extra dimensions of space that create KK excitations of the gluon, corresponding to a spin-1 boson g_{KK} [9]. Due to its large coupling to the Higgs field, the new states are expected to have strong couplings to the top quark. These three models have been discussed in detail in Section 1.3.1.

The analysis presented in this Chapter is a search for heavy resonances decaying to $t\bar{t}$, testing the three models discussed above. Resonant production of a top-antitop quark pair, if observed, would provide a clear signature of BSM physics, hence $t\bar{t}$ resonance searches represent an important part of the BSM programme of the LHC experiments.

Searches for heavy resonances decaying to $t\bar{t}$ have been performed with the ATLAS and CMS experiments using pp collisions at $\sqrt{s} = 7$ TeV[130, 131, 132, 133], 8 TeV [134, 135] and 13 TeV [10, 136, 137, 138]. Both ATLAS and CMS collaborations published the results of the searches in the fully hadronic and semileptonic decay channels with pp collision data at $\sqrt{s} = 13$ TeV with $\int L = 35.9$ fb^{-1}. The CMS collaboration also included the dileptonic decay channel. The only search using the full *Run 2* dataset has been published by the ATLAS collaboration with fully hadronic decay channel [136] and provides the best exclusion limits for high-mass (≥ 1.4 TeV) narrow resonances.

The analysis presented in this Chapter is a search for $t\bar{t}$ resonances in the semileptonic decay channel, targeting both narrow and wide resonances with masses between 0.4 and 5 TeV. It has been developed in collaboration with the team working on the corresponding search in the dileptonic decay channel, and a combination of the results for the two searches is foreseen.

5.2 Data and Monte Carlo Samples

In the following sections, a short description of the physics object definitions, data and Monte Carlo samples, and data-driven background estimations used in the $t\bar{t}$ resonance search is given. This Section frequently references a previous Chapter 3, where a more detailed description of the general concepts is given.

5.2.1 Physics Objects

To study the invariant mass spectrum of $t\bar{t}$ pairs in the semileptonic decay channel, it is necessary to define a set of physics objects, that are described in the following.

Electron candidates are reconstructed and identified as detailed in Section 3.1.2. In this analysis they are required to pass the `Tight` likelihood identification criteria, to have $|\eta_{cluster}| < 2.47$ and not to be in the *crack-region*. The isolation WP used for electrons is the `TightTrackOnly_VarRad`, which only relies on track-related information and not on calorimeter-based ones. In addition, this WP considers only tracks in a cone around the electron with decreasing radius for increasing electron p_T, resulting in a fully efficient isolation selection at high p_T, ideal for the selection of highly energetic top quarks decaying leptonically.

Muon candidates are reconstructed and identified as explained in Section 3.1.3, and the `Medium` WP is selected for the identification. They are also required to have $|\eta| < 2.5$. The same isolation WP as for electrons is used for candidate muons.

Both electrons and muons are matched to the primary vertex using standard impact parameter requirements in ATLAS (see section 3.1.1) by requiring that $|z_0 sin(\theta)| <$ 0.5 mm and $|{}^{d_0}/_{\sigma(d_0)}| < 5(3)$ for electrons (muons). Both electrons and muons energy and momentum are calibrated as explained in Sections 3.1.2, 3.1.3.

In this analysis two types of jets are used: the *standard jets*, reconstructed with the anti-k_T algorithm with a radius $R = 0.4$ and required to have $|\eta| < 2.5$; and the *large-R jets*, reconstructed with the anti-k_T algorithm with a radius $R = 1.0$ and required to have $|\eta| < 2.0$. The *large-R* jets used in this analysis are also required to have $p_T > 350$ GeV. Details on jet reconstruction and jet energy calibration can be found in Section 3.1.4.

Standard jets are selected as b-jets based on the DL1r algorithm, as explained in Section 3.1.7. The efficiency WP used in this analysis corresponds to a 77% b-tagging efficiency, which gives a good compromise between signal efficiency and background rejection. *Large-R jets* are identified as originating from a top-quark (top-tagged) with the DNN top tagging algorithm discussed in 3.1.8 using the *inclusive* criteria. The WP that corresponds to an identification efficiency of 80% for top jets, with a rejection factor for light-flavour quark and gluon jets of approximately 10 for top quarks with p_T between 500 and 3000 GeV[73].

A special Overlap removal technique is used in this analysis. The standard OR used in the ATLAS collaboration would discard all events with an angular distance between the candidate lepton and jet less than 0.4, which constitute a large part of signal events where a boosted top-antitop quarks pair is created from the decay of a heavy resonance. Therefore the standard OR procedure is followed for all objects except removal between candidate leptons and candidate jets. The *sliding* ΔR OR for muons and the *electron-in-jet OR* procedures are followed when muons or electrons are found to overlap with jets. A detailed explanation of these special OR procedures and the motivation for their use in analyses like the one discussed here is given in Section 3.1.5. The calculation of the missing transverse energy E_T^{miss} uses the standard ATLAS definition as in Section 3.1.6.

5.2.2 Data Sample

The analysis is performed using the *pp* collision data collected during the full *Run 2* period, corresponding to an integrated luminosity of $\int L = 140$ fb^{-1} after data quality requirements [107] at the centre-of-mass energy of $\sqrt{s} = 13$ TeV. The average number of interactions per bunch crossing (pile-up) corresponds to $\langle\mu\rangle = 33.7$.

5.2.3 Monte Carlo Simulation

The dominant and irreducible background is the SM $t\bar{t}$ production. It is simulated with PowhegBox v2 [139] at NLO accuracy in QCD interfaced with Pythia8 [85], with the NNPDF3.0nlo PDF set [110] and the A14 tune [90]. The h_{damp} parameter is set to $1.5 \cdot m_t$, where the mass of the top quark is assumed $m_t = 172.5$ GeV. An iterative recursive reweighting algorithm, the so-called *NNLO reweighting*, is used to correct the $t\bar{t}$ MC sample to NNLO in QCD processes and NLO in EW interactions. A detailed explanation of this procedure is given in Section 5.2.6. The decays of the bottom and charm hadrons are modelled with EvtGen [87] for this MC sample as well as for others considered in this analysis.

Alternative SM $t\bar{t}$ samples are generated, for the evaluation of modelling systematic uncertainties. The first has the same setup of the nominal sample described above, but with the Herwig7.7 instead of Pythia8, using the H7UE tune [86] and the MMHT2014lo PDF set [140]. All $t\bar{t}$ samples are normalised to the cross-section value computed with the `Top++2.0` [112] programme to NNLO perturbative QCD including NNLL corrections.

Single top production is simulated with PowhegBox v2 for the s-channel and tW mode, while with PowhegBox v1 [141] with the four flavour scheme for NLO ME calculations for the t-channel. For all single top samples, the ME generator is interfaced with Pythia8 with the A14 tune and the NNPDF2.3lo PDF set. MadSpin [142] is used to simulate top-quark decays in the t-channel sample, which preserves all spin correlations. The overlap between tW and $t\bar{t}$ production is treated with the DR scheme [113]. An alternative tW sample is produced using the DS scheme [114], instead of the DR one. The single top samples are all normalized to their theoretical cross-sections calculated at NNLO+NNLL accuracy [143, 144, 145].

The background from $t\bar{t} + V$ (where $V \in W, Z$) production is generated at NLO in QCD with MadGraph5_aMC@NLO [83] and the NNPDF3.0nlo PDF set interfaced with Pythia8 using the A14 tune. The samples are normalised to the theoretical cross-section calculated at NLO in QCD [146]. The $t\bar{t} + H$ production is simulated at NLO accuracy in QCD with PowhegBox v1 interfaced with Pythia8 using the NNPDF3.0nlo PDF set and the A14 tune. The sample is normalised to the theoretical cross-section calculated at NLO in QCD and NLO in EW accuracy [147]. In all $t\bar{t} + V$ and $t\bar{t} + H$ the EvtGen program is used for the decay of bottom and charm hadrons.

The W+jets, Z+jets ad diboson samples are generated with Sherpa [84] using the NNPDF3.0nlo PDF set. The Z+jets sample is normalised to the theoretical cross-section computed at NLO in QCD precision [148]. The W+jets sample normalisation is corrected with data-driven factors, which also correct the relative flavour fractions, as discussed in Section 5.2.5.

All signal processes are simulated at LO in QCD with MadGraph5_aMC@NLO interfaced with Pythia8 with the NNPDF2.3nlo PDF set and the A14 tune. These MC generators allow for the spin information to be correctly propagated to the final-state leptons. The Z' signal samples are generated with masses of the boson between 400 GeV and 5 TeV. The cross-section was normalised to the predictions obtained at

NLO in QCD in [149]. The g_{KK} signal samples are generated between 400 GeV and 5 TeV, the G_{KK} signal samples are generated between 400 GeV and 3 TeV. Both for the g_{KK} and the G_{KK} samples, the value of the cross-sections obtained from the generator is used.

5.2.4 Fake and non-prompt lepton background estimation

The background created from strong production of multijet events where at least one of them is misidentified as a lepton is one of the most challenging to estimate, as already discussed in Section 3.3. A fully data-driven approach is used to estimate this background, using the Matrix Method [92], already described in 3.3.1.

In the fake and non-prompt lepton background estimation for this analysis, the real efficiency ε_r and the fake efficiency ε_f are estimated in data. The sample used to estimate the ε_r uses the same selections described in 5.3.1. Looser selections are instead used to estimate the ε_f to have a sample enriched in multijet events. These samples, referred to as *Control Regions* (CRs), are required to have at least one electron or muon passing the `Loose` identification WP (see Section 3.1.2), while no leptons passing the `Tight` WP should be present. Further selections applied are listed in Table 5.1. A set of events with similar kinematic characteristics to those of events used to perform the search presented in this Chapter, but still enriched in multijet events is used as *Validation Regions* for the fake and non-prompt lepton background estimate. The selections applied to the validation regions are listed in Table 5.1.

	CR e	CR μ	VR 1	VR 2
Lepton	1 Loose e	1 Loose μ	1 Loose e	1 Loose μ
Veto	no Tight μ	no Tight e	no Tight μ	no Tight e
$E_{\mathrm{T}}^{\mathrm{miss}}$	< 20 GeV	< 20 GeV	< 20 GeV	> 20 GeV
$E_{\mathrm{T}}^{\mathrm{miss}} + m_{\mathrm{T}}^{W}$	< 60 GeV	< 60 GeV	> 60 GeV	> 60 GeV
Jets	≥ 4 jets			
b-jets	≥ 1 b-jet			

Table 5.1: Selections applied to the multijet electron and muon control regions and validation regions.

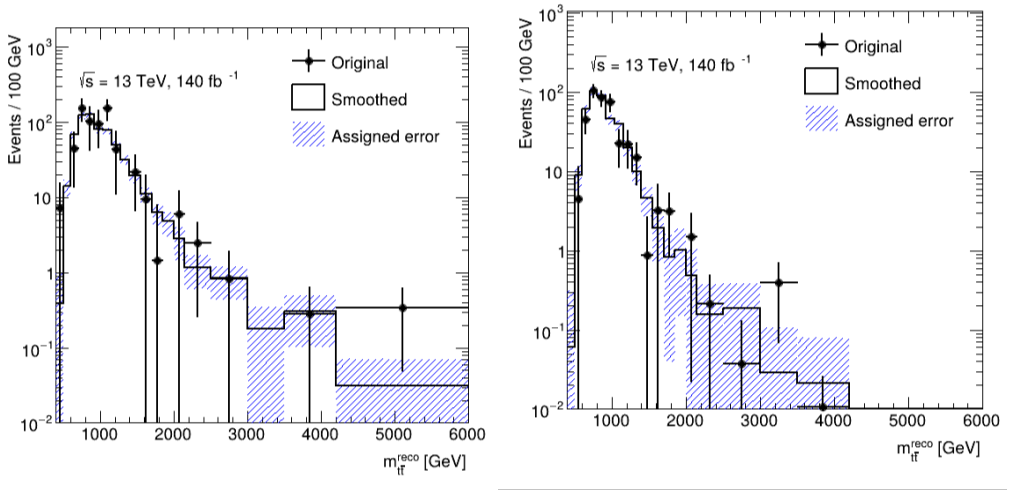
The real and fake rates are parametrised as a function of the lepton p_T and the calorimeter-based isolation variable E_T^{cone20}. In addition, the rates are derived separately for the two cases $\Delta R(\ell, j) < 0.4$ and $\Delta R(\ell, j) \geq 0.4$ (here j is the nearest selected jet). The fake rates for electrons vary from 5% to 86%, with the largest values occurring at high electron p_T, with low nearby calorimeter activity. This trend can be expected based on the special isolation requirements which use a decreasing size of the isolation cone with increasing p_T of the lepton and therefore can lead to looser isolation selection at high p_T. The fake rates for muons vary from 9% to 84%, with the largest values occurring at low muon p_T, with low nearby calorimeter activity, which is typical for soft muons arising from semileptonic decays of hadrons inside jets. Good agreement is found both in CRs and VRs between data and MC prediction for SM background processes.

The low statistics of the multijet background for values of the top-antitop quark reconstructed invariant mass, $m_{t\bar{t}}$, leads to large statistical fluctuations with a non-smooth distribution, especially in the merged topology considered in this analysis (that is described in Section 5.3.1). Therefore a smoothing procedure is applied to the multijet background estimation for events in the merged topology. The smoothing

is derived separately for events with an electron and a muon in the final state. The smoothing procedure is derived with the following steps:

1. fit the bulk of the un-binned data sample with a Landau distribution and the tail with an exponential distribution,
2. build a combined function from the two fitted ones, the two functions end/start at the separation value $m_{t\bar{t}}^{cut}$, where continuity between the two functions is ensured,
3. based on the overall fitted function, new binned data samples are generated, where the content of each bin is derived from the function integral in that bin.

Figure 5.1 shows the original and the smoothed multijet estimations for events in the merged topology.



(a) e+jets final state.

(b) μ+jets final state.

Figure 5.1: Smoothed histograms (purple line) compared to the original estimate (black points) for the multijet background estimation in the merged topology. The shaded area represents the uncertainty on the smoothed histogram.

5.2.5 Data-driven correction to the W+jets background

The normalisation of the W+jets background can be poorly modelled by simulation. Normalisation factors are derived from data to correct the normalisation of this background and are applied as scale factors before the final fit. The *Charge-Asymmetry* (CA) *method* is employed to derive the normalisation factors. It relies on the knowledge that W^+ bosons should be produced at the LHC at a higher rate than W^- bosons due to the asymmetry in the quark content of the proton. The ratio of the production cross-section for W^+ over W^- is known from theory with a higher accuracy than the total W+jets cross-section in events with three or more jets. The charge of the lepton in the final state is used to determine the charge of the W boson that originated it. The total number of W+jets events in data $N_{W^+} + N_{W^-}$ is given by

$$N_{W^+} + N_{W^-} = \frac{r_{MC} + 1}{r_{MC} - 1}(D_{corr+} - D_{corr-}), \tag{5.1}$$

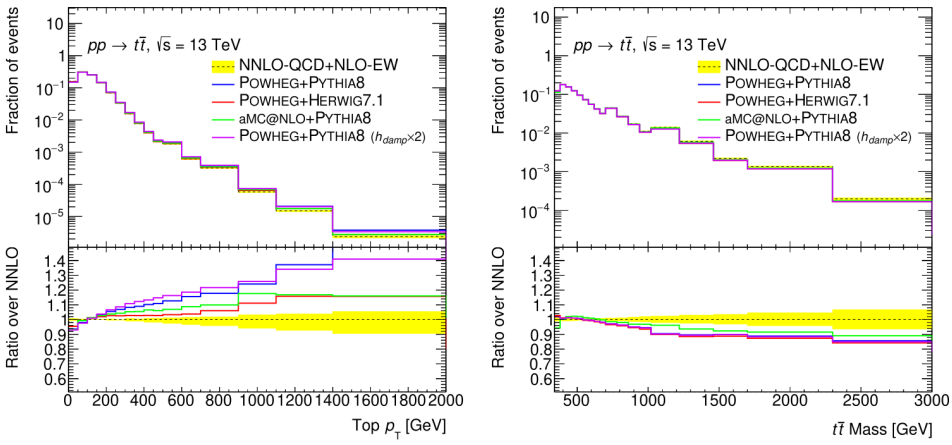
where r_{MC} is the ratio of the number of W+jets events with a positively charged lepton over that with a negatively charged lepton obtained from the MC simulation

and $D_{corr+(-)}$ the number of observed data events with a positively (negatively) charged lepton.

A significant advantage of the CA method is that the flavour-symmetric backgrounds like $t\bar{t}$, multijet production and Z+jets have no charge asymmetry. Hence, the extraction of the normalisation factors via the CA method can be performed in any sample of events using the analysis events selections, in dedicated CRs or from inclusive event samples. The normalisation factors derived from events with the analysis event selections are found to be in agreement within statistical uncertainties with those obtained using events in the CR.

5.2.6 Higher order reweighting

To take advantage of the recent theoretical predictions for $t\bar{t}$ differential cross-section, currently available at NNLO in QCD including NLO EW effects [150], a NNLO reweighting technique is used. The predictions for several kinematic observables from this computation can be significantly different than those obtained from the MC generator, for instance, Figure 5.2 shows a comparison of the top-quark p_T and $m_{t\bar{t}}$ distributions predicted by various MC generators and the higher order calculation in Reference [150]. The differences in the predictions are clearly visible, and they are not covered by the yellow uncertainty band which includes scale and PDF uncertainties on the higher-order prediction(see Section 3.2.3 for an explanation of these systematics). To implement this more accurate theory prediction into the MC



(a) Top-quark p_T distribution.

(b) $m_{t\bar{t}}$ distribution.

Figure 5.2: Comparison of the higher-order analytical predictions in Reference [150] (dashed black line) with various MC generator predictions (solid lines) for the top-quark p_T in 5.2(a) and for the invariant $t\bar{t}$ mass in 5.2(b).

samples, an iterative, recursive reweighting algorithm is used to correct the kinematics of the SM $t\bar{t}$ samples. Each step of the procedure reweights the distribution of a target kinematic variable from the MC generator to that of the calculated higher-order predictions, taking into account any corrections already applied by the previous steps of the algorithm. The reweighting is performed event-by-event for each of the distributions, and applied to the kinematic distributions in the following order: $m_{t\bar{t}}$ first, then the transverse momentum of the top-quark $p_T(t)$ and antitop-quark $p_T(\bar{t})$, and finally again to the $p_T(t)$ and $p_T(\bar{t})$. To avoid introducing artificial asymmetry between the top-quark and antitop-quark kinematic distributions, when applying the procedure to the p_T distributions, two weights are derived for each event, one

considering the top quark p_T and the other for the antitop-quark p_T and their geometrical average is then used. This whole procedure is repeated three times to obtain good agreement between the reweighted MC-generated sample distributions and the theoretical prediction in Reference [150].

5.3 Analysis Strategy

Representative Feynman diagrams for the BSM signals considered in this analysis have already been shown in Figure 1.5, in Section 1.3.1 where these models have been presented. The main irreducible background is the SM $t\bar{t}$ production. A Feynman diagram for $t\bar{t}$ production with semi-leptonic decay is shown in Figure 5.3.

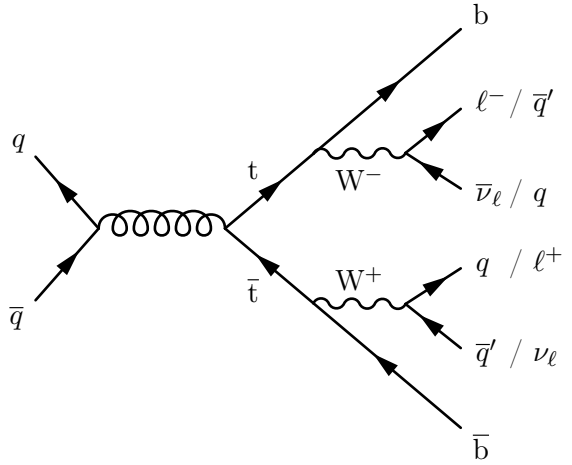


Figure 5.3: SM $t\bar{t}$ production Feynman diagram decaying semileptonically.

The final state considered in this analysis has a lepton and a neutrino from the W boson decaying leptonically, a quark-antiquark pair from the W boson decaying hadronically and a b and a $\bar{b}$ quarks from the top- and antitop-quarks decaying to a $\bar{b}W^+$ and bW^- respectively.

In this search for a heavy resonance decaying to $t\bar{t}$ events a vast range of invariant mass for the heavy resonance is considered, from 400 GeV up to 5 TeV. In the case of a low $t\bar{t}$ invariant mass range, all final state products are expected to be separated in space, while for the highly boosted events, top-quark decay products are expected to be collimated and thus not separated. Therefore two main topologies are considered: a *resolved* topology where all final state objects are separated from each other and a *merged* topology where the hadronic decaying top-quark is reconstructed as a single physics object (a *large-R* jet). This splitting allows us to optimize the reconstruction and identification efficiency in different kinematic regimes.

The main background processes, other than the $t\bar{t}$ production, are the W+jets production, the QCD multijet background, single-top production, Z+jets production, $t\bar{t}+V$ and $t\bar{t}+H$ production and diboson production.

5.3.1 Event Selection

Events are required to have at least one reconstructed vertex with at least two associated tracks with $p_T > 0.5$ GeV. The primary vertex is selected as the one with the largest sum of the squared p_T [120].

Based on the final state considered for this analysis, all candidate events are required to pass a single electron [151] or a single muon trigger [152]. The trigger

efficiency is mostly constant in the transverse momentum for leptons with $p_T >$ 28 GeV and therefore only events passing this requirement are further considered.

All events are required to contain exactly one electron or muon candidate (as defined in Section 5.2.1), and events with a second electron or muon with $p_T > 25$ GeV are vetoed to ensure orthogonality with the corresponding analysis considering the dileptonic decay of the $t\bar{t}$ system. Events are also required to have at least one b-tagged *standard* jet (as defined in Section 4.2.1).

To reduce the multijet background contribution, a minimum amount of missing transverse energy is required: $E_T^{miss} > 20$ GeV. Additionally, to reduce the same background, the *triangular cut* is applied $E_T^{miss} + m_T^W > 60$ GeV, where m_T^W is the W boson transverse mass calculated from the selected lepton and the $\vec{p}_T^{miss}$ as $m_T^W = \sqrt{2p_T^{lep}E_T^{miss}(1 - \cos(\Delta\phi(l, \vec{p}_T^{miss})))}$.

Further event selections are different for the merged and resolved topologies and are described in the following. To obtain orthogonal signal samples for the final fit, only events that do not meet the requirements for the merged topology are considered for the resolved one. This choice is based on the better $m_{t\bar{t}}$ resolution that can be obtained in the merged topology compared to the resolved one (see Section 5.3.2).

Both for the resolved and the merged topologies, events entering the signal regions (SRs) are selected based on the absolute pseudo-rapidity difference between the hadronically decaying reconstructed top-quark and the candidate lepton $\Delta\eta(t^{had}, \ell)$. The selection on this observable is optimised as a function of $m_{t\bar{t}}$: and the chosen values are reported in Table 5.2 for events in the merged region and in Table 5.3 for events in the resolved region.

$m_{t\bar{t}}$ [GeV]	$\Delta\eta(t^{\text{had}}, \ell)$ range
$m_{t\bar{t}} < 640$	< 1.5
$640 \leq m_{t\bar{t}} < 800$	< 2.0
$800 \leq m_{t\bar{t}} < 1280$	< 2.5
$1280 \leq m_{t\bar{t}} < 1700$	< 3.0
$m_{t\bar{t}} \geq 1700$	< 3.5

Table 5.2: Selection values on the $\Delta\eta(t^{had}, \ell)$ observable as a function of the reconstructed $m_{t\bar{t}}$ for the merged region.

$m_{t\bar{t}}$ [GeV]	$\Delta\eta(t^{\text{had}}, \ell)$ range
$m_{t\bar{t}} < 1040$	< 1.0
$1040 \leq m_{t\bar{t}} < 1850$	< 1.5
$1850 \leq m_{t\bar{t}} < 2500$	< 2.0
$2500 \leq m_{t\bar{t}} < 3500$	< 2.5
$m_{t\bar{t}} \geq 3500$	< 3.0

Table 5.3: Selection values on the $\Delta\eta(t^{had}, \ell)$ observable as a function of the reconstructed $m_{t\bar{t}}$ for the resolved region.

Events that do not pass the se $\Delta\eta(t^{had}, \ell)$ selections are used to define **validation regions VRs** (which are then divided in the merged and resolved topology in the same way as for events passing the $\Delta\eta$ selection).

Merged topology selections

Events in the merged category must have at least one *standard* jet within an angular distance from the selected lepton of $\Delta R(\ell, \text{jet}) < 2.0$: such jets are considered to be candidates for the b-jet from the leptonically decaying top-quark. If none of those jets is b-tagged, the jet with the highest p_T is chosen as the candidate b-jet. If more than one jet with $\Delta R < 2.0$ is b-tagged, the one with the highest p_T is chosen as the candidate b-jet.

Exactly one top-tagged *large-R* jet must be present and it needs to have an angular distance with respect to the lepton and the selected b-jet from the leptonically decaying top-quark greater than 1.5. The *large-R* jet is also required to have a

mass greater than 100 GeV, which helps to avoid fake contributions coming from mistakenly identified top-tagged *large-R* jets. Events passing these selections are part of the merged SR or merged VR if they respectively pass or not the $\Delta\eta(t^{had}, \ell)$ in Table 5.2.

Resolved topology selections

As mentioned above, only events failing the merged topology selections are considered for the resolved category. Events are then required to have at least four *standard* jets, and at least one of them b-tagged. The jets are selected if they have a transverse momentum greater than $p_T > 25$ GeV. The $t\bar{t}$ system is reconstructed with a χ^2 algorithm described in Section 5.3.2. Only well-reconstructed events are selected, asking that the χ^2 value obtained from the algorithm is $\log_{10}(\chi^2) < 0.9$. This requirement has a 60% efficiency on SM $t\bar{t}$ events.

To further increase the search sensitivity, events passing the resolved-topology selection are split into two sub-categories based on whether one or both of the reconstructed top quarks have an associated b-jet. The *resolved2b* category contains events with both reconstructed top-quarks have an associated b-jet, while the *resolved1b* category contains events where only one of the two reconstructed top-quarks has an associated b-jet. Events passing the resolved topology selections are therefore split into 4 categories: the *resolved1b* SR (VR) for events in the *resolved1b* category that pass (not pass) the $\Delta\eta$ selection described in Table 5.3, and the *resolved2b* SR (VR) for events in the *resolved2b* category that pass (not pass) the $\Delta\eta$ selection.

Figure 5.4 shows data/MC comparison in the *resolved1b* and *resolved2b* SR regions for the leading jet p_T and for the leading lepton p_T[1]

5.3.2 Invariant mass reconstruction of the $t\bar{t}$ system

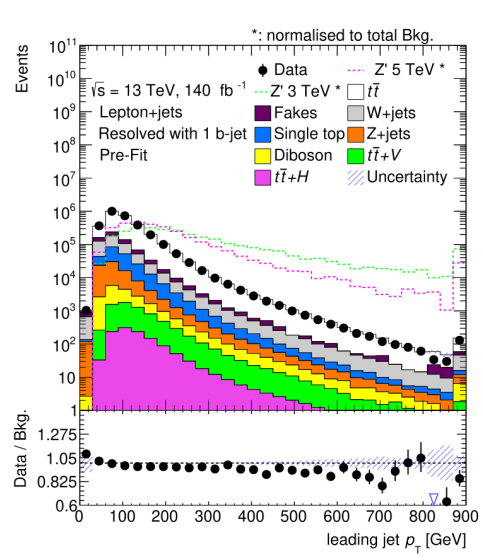
The main observable used in the analysis, on which the presence of new physics signals is tested, is the reconstructed invariant mass of the $t\bar{t}$ system, $m_{t\bar{t}}$. The $t\bar{t}$ system is reconstructed by taking as input the final-state particles measured in the detector, i.e. the physics objects. This reconstruction is performed differently for events in the resolved and merged topologies due to the different objects present in each of the two categories. Still, the neutrino reconstruction is the same in both categories and is described in the next paragraph.

The reconstruction of the $t\bar{t}$ system is complicated by the emission of extra radiation, which can reduce the sensitivity of the analysis, particularly at high resonance masses. The effect of this, at the simulation level, can be seen in Figure 5.5, where the generated Z' mass is compared to the total mass of the $t\bar{t}$ system decay products taken from the generator itself.

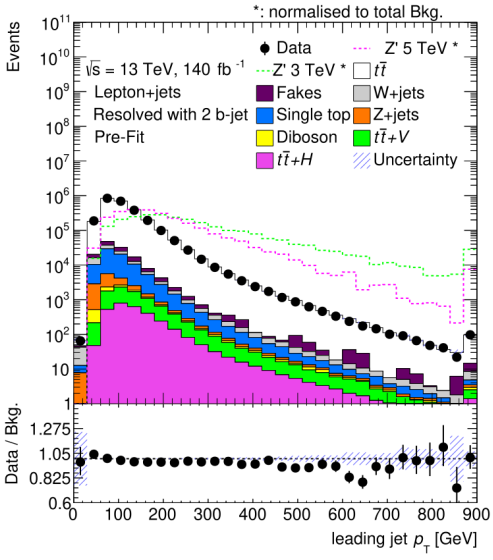
Neutrino reconstruction

The transverse momentum of the neutrino p_T^ν is identified with the E_T^{miss} in the event, assuming that a single neutrino is the only source of missing transverse energy in the event. Instead, the longitudinal component p_z^ν cannot be determined solely from the missing transverse energy in the event and it is here calculated imposing the W boson mass constraint $m_W^2 = (p^\nu + p^\ell)^2$, in a similar way as in the case of the single-top cross-section analysis discussed in Chapter 4, Section 4.3.2. There are three possibilities when solving this quadratic equation:

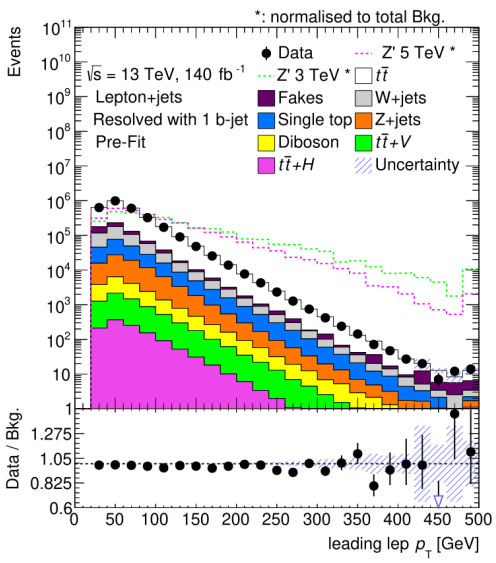
[1]Here leading refers to the object of that type with the highest transverse momentum.



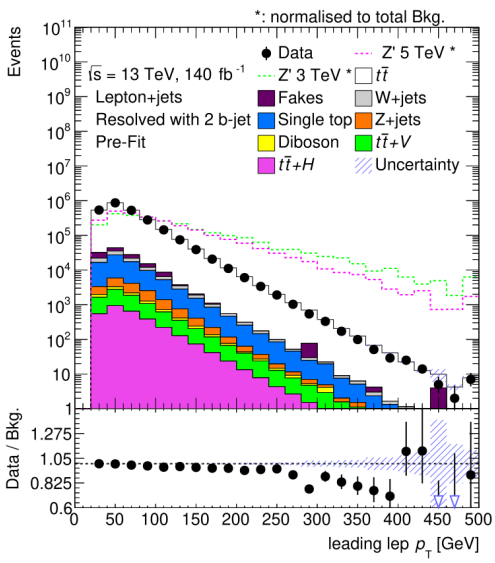
(a) Resolved 1b jet p_T.



(b) Resolved 2b jet p_T



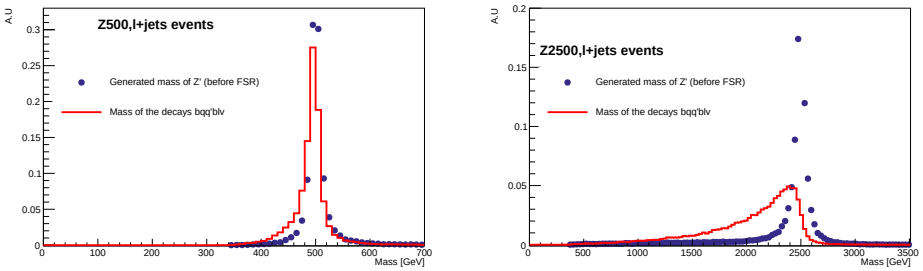
(c) Resolved 1b lep p_T



(d) Resolved 2b lep p_T

Figure 5.4: Pre-fit jet p_T distributions in the *resolved1b* SR category 5.4(a) and *resolved2b* SR category 5.4(b), and lepton p_T distribution in the *resolved1b* SR category 5.4(c) and *resolved2b* SR category 5.4(d) comparing data and simulation. Only statistical uncertainties are included.

- if there is only one real solution, that is taken as the neutrino longitudinal momentum;
- if no real solution is found, the missing transverse energy vector is rescaled and rotated by a minimal amount to obtain exactly one real solution. It is assumed that this effect is caused by a mismeasurement of the E_T^{miss} in the event,
- if two real solutions are found the procedure is different for the resolved and merged topology: in the first case the solution resulting in a smaller χ^2 value is taken, while for events in the merged topology the solution with the smallest



(a) Z' resonance at 500 GeV. (b) Z' resonance at 2.5 TeV.

Figure 5.5: Mass of the generated Z' signal compared to the mass of the $t\bar{t}$ decay products at generator level of the 500 GeV mass resonance in 5.5(a) and 2.5 TeV in 5.5(b).

p_z^ν is chosen.

Event reconstruction in the resolved topology

A χ^2 minimization approach is used to select the four jets originating from the $t\bar{t}$ decay out of all of the selected *standard* jets in the event. The χ^2 function is defined as:

$$\chi^2 = \left[\frac{m_{jj} - m_W}{\sigma_W}\right]^2 + \left[\frac{(m_{jjb} - m_{jj}) - m_{(t_h - W)}}{\sigma_{(t_h - W)}}\right]^2 + \left[\frac{m_{jl\nu} - m_{t_l}}{\sigma_{t_l}}\right]^2 + \left[\frac{(p_{T,jjb} - p_{T,jl\nu}) - (p_{T,t_h} - p_{T,t_l})}{\sigma_{\Delta p_T}}\right]^2 . \quad (5.2)$$

The first term is a constraint requiring the mass of a pair of jets m_{jj} to be close to the W boson mass. The second term constrains the mass of a three-jet system that includes a b-tagged jet m_{jjb} to be close to the mass of the hadronically-decaying top quark (t_h). Since m_{jj} and m_{jjb} are strongly correlated, the mass of the hadronically decaying W boson is first subtracted to decouple this second term from the first one. The third term is used to constrain the mass of the leptonically decaying top quark (t_l), and the final term constrains the transverse momenta of the two top quarks to be similar, as expected for SM $t\bar{t}$ production at LO. All possible jet permutations are considered and the one with the lowest value of χ^2 is chosen in each event.

This estimate is improved using the precise knowledge of the W boson and top-quark masses. First, the momenta of the two jets assigned to the hadronic W boson decay are corrected based on the constraint that their total invariant mass must be close to the W boson mass. Then, the momenta of the third jet assigned to the hadronic top-quark decay is corrected based on the constraint that the total invariant mass of these three jets must be close to the top-quark mass. No scaling is applied to the jet assigned to the leptonically-decaying top quark in order to avoid over-correcting the jet momenta beyond the range compatible with the jet energy resolution and the uncertainty on the jet energy scale, caused by the poorer E_T^{miss} resolution. This correction does not change the χ^2 minimization process and therefore does not affect the jet-parton assignment, but improves the $m_{t\bar{t}}$ resolution.

Event reconstruction in the merged topology

For the merged topology, the hadronically decaying top quark can be straightforwardly identified with the top-tagged selected *large-R* jet. The leptonically decaying top-quark is reconstructed from the selected candidate jet identified with the *b*-jet (as explained in Section 5.3.1), the selected lepton and the reconstructed neutrino.

5.4 Statistical Analysis and Systematic Uncertainties

The reconstructed invariant mass of the $t\bar{t}$ system in the *resolved1b*, *resolved2b* SRs and in the merged SR are used to perform a simultaneous Profile Likelihood Fit. The Statistical analysis setup, as well as the modelling of the background processes, have been verified first in the VRs (defined in Section 5.3.1). Both statistical and systematic uncertainties are taken into account in the fit.

5.4.1 Systematic Uncertainties

Several sources of systematic uncertainties that can affect the normalization and shape of the invariant mass distributions predicted for the various background and signal processes, in the different signal and validation regions are considered and described in the following.

Luminosity and Pileup

The uncertainty on the combined integrated luminosity of the dataset collected between 2015 and 2018 corresponds to 0.83% [122]. The MC samples are corrected to the luminosity profile measured in data with a *pileup weight*, and its relative uncertainty is taken into account.

Reconstructed physics objects

Uncertainties related to the reconstruction, identification and isolation of leptons are considered through uncertainties on the corresponding SFs. In addition, uncertainties arising from the energy/momentum scale and resolution of the leptons are also considered. In the case of electrons within an angular distance of ΔR from the closest jet less than 0.4, an additional uncertainty of 10% is added to the standard identification SF uncertainty as discussed in Section 3.1.5.

Uncertainties related to the JES and JER are included for all jets as explained in Section 3.1.4. Additionally, *large-R* jets also have uncertainties related to the Jet Mass Scale (JMS) and Jet Mass Resolution (JMR).

The uncertainties on the *b*-tagging calibration are determined separately for *b*-, *c*- and light-jets (as explained in Section 3.1.7). An additional uncertainty to take into account of the flavour-tagging SFs extrapolation at high p_T is applied. Events in the merged topology are also affected by an uncertainty related to the top tagging SFs. Finally, the uncertainty on the E_T^{miss} is applied as explained in Section 3.1.6.

Modeling Uncertainties

The following sources of modelling uncertainties are the main ones affecting MC simulated background samples and are considered in this analysis, although some of them may be negligible for certain small background processes. They are listed here, while a more detailed explanation can be found in Section 3.2.3:

- missing higher orders in the perturbative QCD ME computation;

- uncertainty on the amount of ISR and FSR;
- choice of the PDF sets;
- uncertainty on the method for the matching and merging of ME calculation with PS model;
- uncertainty on the choice of the parton shower and hadronisation model,

For the $t\bar{t}$ background process, renormalization and factorization scale uncertainties are estimated separately for (anti)top p_T and $m_{t\bar{t}}$ distributions by varying independently the scales up and down by a factor of two. For this sample, also the uncertainty in the EW contribution is estimated by comparing the nominal spectra to those obtained with the `NNPDF3.0QED` PDF set. Furthermore, two uncertainties are assigned to the iterative reweighting method, which takes into account possible biases due to the choice of the order in which the reweighting is applied and possible differences between the parton-level top-quark definitions in the higher-order calculations and the NLO MC prediction. An additional uncertainty with respect to those defined in Section 3.2.3, referred to as *lineshape uncertainty*, related to the treatment of $t\bar{t}$ spin correlations in different generators, is estimated by comparing the nominal sample with an alternative sample generated with the same setup as the nominal one but using MadSpin to decay the top quarks.

Uncertainties on the cross-section normalization of all background processes are applied, the values can be found in Table 5.4. Uncertainties on the multi-jet fake/non-

Table 5.4: Summary of the normalization uncertainties on the calculated cross-sections for the different background components.

Background process	Up variation [%]	Down variation [%]
$t\bar{t}$	+5.6	-6.1
Single-top: Wt-channel	+5.4	-5.4
Single-top: t-channel	+4.3	-3.7
Single-top: s-channel	+4.4	-4.1
$t\bar{t} + Z$	+10.4	-12.0
$t\bar{t} + W$	+13.3	-12.0
$t\bar{t} + H$	+6.8	-9.8
$W + jets$	+50.0	-50.0
$Z + jets$	+30.0	-30.0
Diboson	+50.0	-50.0
Multijet	+50.0	-50.0

prompt lepton background estimation are considered separately for the e+jets and the μ+jets channels and other than the conservative 50% normalization uncertainty, an additional systematic uncertainty related to the smoothing procedure applied in the boosted topology is used. Normalization uncertainties on the $W + jets$ background process are set to a conservative 50% based on the CA normalization factors derivation method. Only the normalization uncertainty is considered for the Z+jets and diboson processes which are minor backgrounds in this analysis and are set to a conservative 50%. The normalization uncertainties for the other processes are derived from the theory calculations mentioned in Section 5.2.3.

The modelling uncertainties are treated as uncorrelated between all signal and background samples. Additionally, the uncertainties in the SM $t\bar{t}$ background related to the modelling of the PS and hadronisation, the ME computation matching to the PS, and the choice of the h_{damp} parameter are treated as uncorrelated between all 3 signal regions. These uncertainties are obtained by comparing the nominal to an alternative sample in each case (the so-called *two point systematics*), an approach that typically yields conservative uncertainties with non-negligible constraints in the Profile LH fit. These uncertainties are kept de-correlated among signal regions to prevent constraints from propagating across regions, thus minimizing the constraints on these nuisance parameters. Additionally, uncertainties in the SM $t\bar{t}$ background related to the choice of μ_R and μ_F, amount of ISR/FSR cannot be assumed to be fully correlated across different topologies and are therefore de-correlated among resolved and merged topologies.

5.4.2 Profile Likelihood Fit and Exclusion Limit

The agreement between the data and the SM prediction is tested with a Profile LH fit of the expected distributions in $m_{t\bar{t}}$ to the observed distributions, both with a background-only hypothesis and including a signal contribution in the model. When testing the presence of a particular signal, a single Parameter of Interest (POI) is included, as a signal strength parameter μ_{signal} multiplying the signal prediction for the tested signal benchmark, interpreted as the ratio of the observed over the predicted value of the pure-signal process cross section times its branching fraction to $t\bar{t}$. For each tested signal hypothesis, the signal strength is extracted from a simultaneous fit in all the three signal regions considered: the *resolved1b* and *resolved2b* for the resolved topology and the signal region for the merged topology. In addition, another parameter corresponding to the normalization of the SM $t\bar{t}$ background common to all signal regions is left free to float during the fit.

The Profile LH fit has already been described in Section 3.4.1. In the absence of significant excess in the data over the background-only prediction, upper limits at 95% CL_s are derived on the production cross-section times branching ratio to $t\bar{t}$ for the resonances predicted by the models discussed in Section 1.3.1. The upper limits are calculated as explained in Section 3.4.3.

Systematic uncertainties with an effect of less than 0.01% are removed from the model in order to improve the fit stability. Shape effects which arise from statistical fluctuations in the systematic uncertainties distributions are smoothed for most of the NPs distributions.

5.4.3 Background modelling and fit setup validation

The modelling of the SM background as a function of $m_{t\bar{t}}$ is tested in the three Validation Regions, defined in Section 5.3.1. This is particularly important to check the description of data-driven background estimation (such as the *fake/non-prompt* multijet background and the CA normalization factors for the W+jets background), as well as modelling of systematic uncertainties related to background processes, $t\bar{t}$ production being the most relevant one.

Figure 5.6 shows the $m_{t\bar{t}}$ distribution in the three VRs defined (resolved topology with one b-jet, resolved topology with two b-jets, and merged topology) for data and SM background expectation. Good agreement of the data with the SM background expectation within the total uncertainty is observed in VRs.

The $m_{t\bar{t}}$ distribution after the Profile LH fit to the three VRs under the background-only hypothesis ($\mu_{\text{signal}} = 0$) is shown in Figure 5.7, and the expected constraints

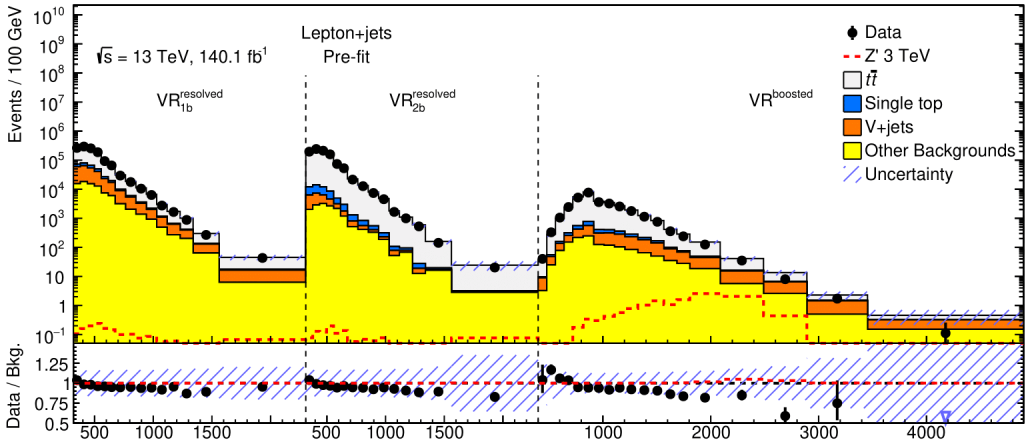


Figure 5.6: Pre-fit distributions of the reconstructed $m_{t\bar{t}}$ for the three VRs. In the lower panels, the ratio of the data and background prediction is shown. The expected relative deviation from the background prediction in the presence of a Z' resonance with a mass of 3 TeV, scaled by a factor of five for better visibility, is also shown in both panels.

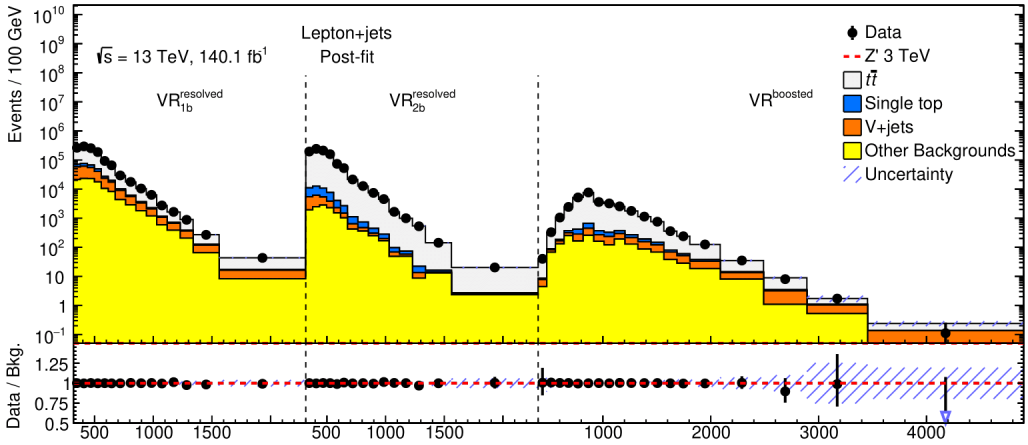


Figure 5.7: Post-fit distributions of the reconstructed $m_{t\bar{t}}$ for the three VRs. In the lower panels, the ratio of the data and background prediction is shown. The expected relative deviation from the background prediction in the presence of a Z' resonance with a mass of 3 TeV, scaled by a factor of five for better visibility, is also shown in both panels.

and variations of the central value of the NPs are shown in Figure 5.8.

Figure 5.8 shows some small shifts of some NPs. The largest ones are observed on JES uncertainty NPs related to pileup, η and energy scale modelling, on the W+jets normalization in the boosted topology, for $t\bar{t}$ FSR in the *resolved1b* NP relative to the shape of this uncertainty, and on the *fake/non-prompt* multijet background estimation.

The upper limit at 95% CL_s has been extracted for all signal sample masses considered in this analysis in the validation region to compare the effect of background estimation techniques, in a region of the phase space where the analysis is supposed

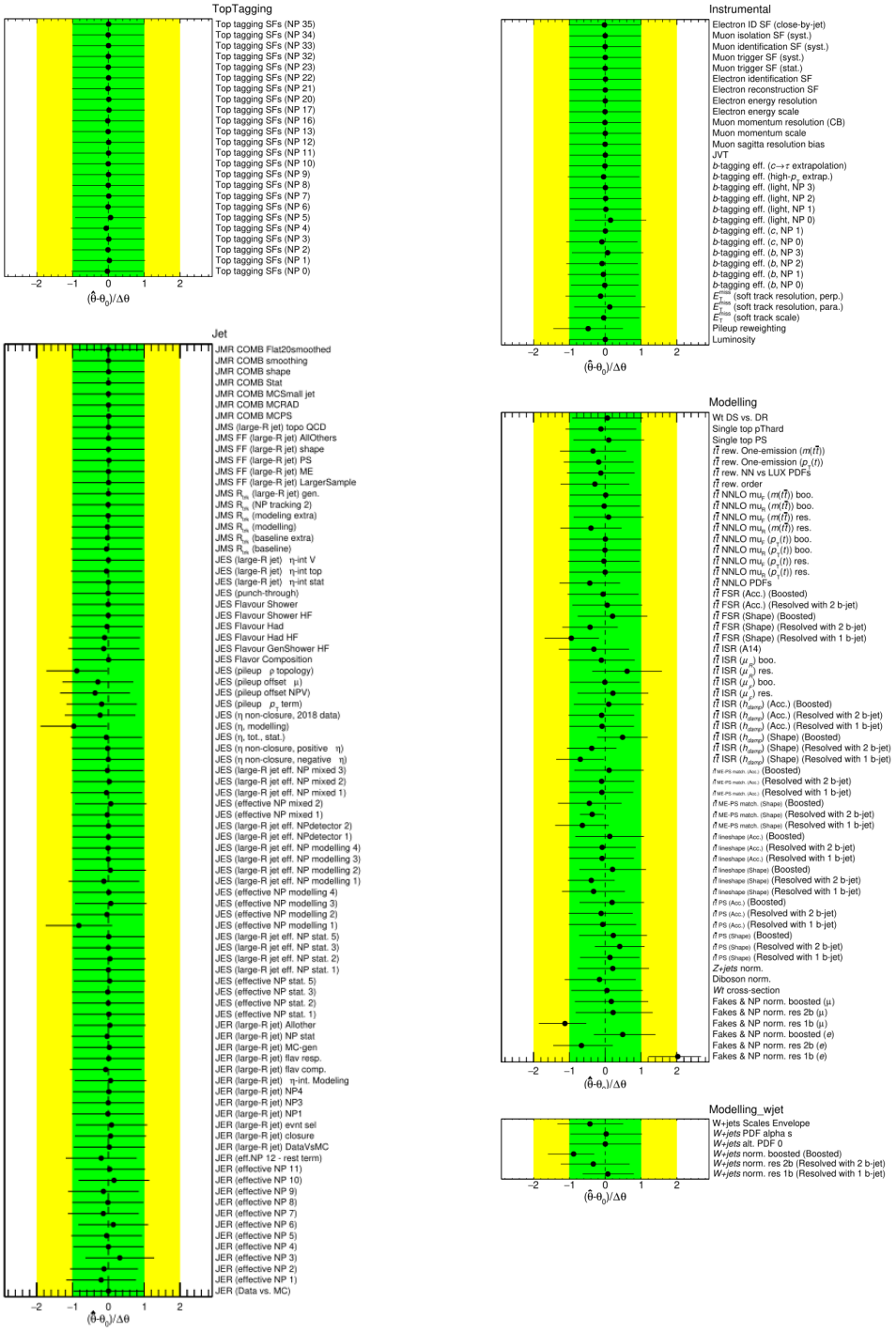


Figure 5.8: Nuisance parameters after a fit to data using the $m_{t\bar{t}}$ distribution in the VRs. NPs are arranged in different groups based on the different systematic categories they belong to. The NPs correspond to systematic uncertainties explained in 5.4.1.

to not be sensitive. The expected upper limit, extracted using an Asimov dataset[2], is shown in Figure 5.9. As a test to check the effect of the smoothing procedure

[2]An Asimov dataset is an artificial dataset constructed by setting all observed quantities to their expected values according to the nominal model used. This approach allows for the evaluation of expected experimental sensitivities without the influence of statistical variations in the actual data [153].

applied in the boosted topology to the *fake/non-prompt* multijet background (described in 5.2.4) the expected limits have been extracted and shown in Figure 5.9 in the case of no smoothing procedure applied to this multijet background. It can be clearly seen that in the VRs the effect of the smoothing is within the 2σ sensitivity band for the 95% CL_s. The plot also shows the observed upper limit on the signal cross-section times branching ratio to $t\bar{t}$ for the Z' signal in the VRs. The values are expected not to be very sensitive as the selection on the $\Delta\eta(t^{had}, \ell)$ reported in Table 5.3 and 5.2 is chosen to enhance the signal over background ratio in the signal regions while decreasing it in the validation regions.

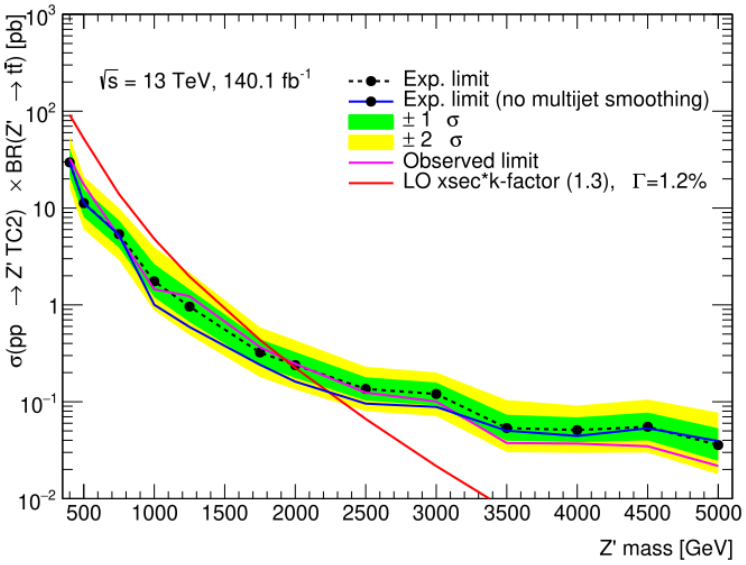
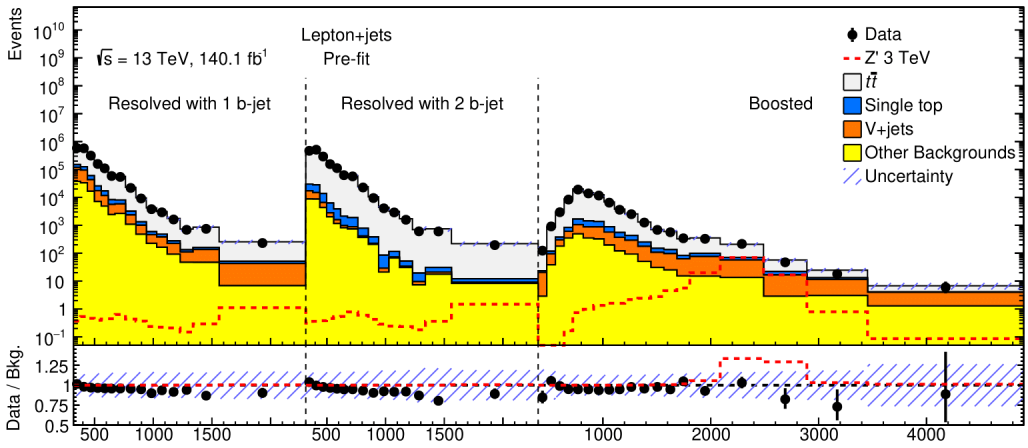


Figure 5.9: Observed and expected upper limits on the signal cross-section times branching ratio to $t\bar{t}$ for the Z' signal at 95% CL, extracted from the validation regions.

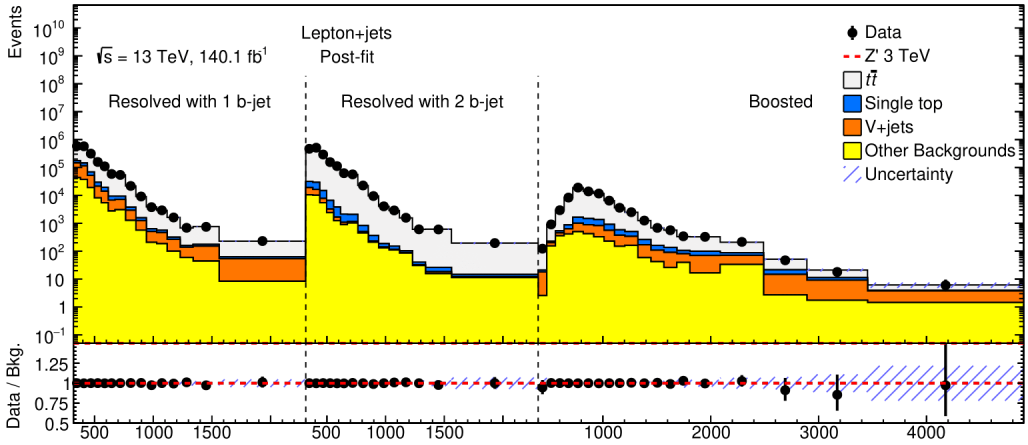
5.5 Results

The agreement between the observed data and the SM prediction in the signal regions is tested with a background-only hypothesis fit in which only the nuisance parameters are allowed to vary ($\mu_{\text{signal}} = 0$ for all the signal benchmarks). The $m_{t\bar{t}}$ distributions in the three signal regions before and after the fit to the full 140 fb^{-1} dataset are shown in Figure 5.10(a), 5.10(b) respectively. The observed post-fit $m_{t\bar{t}}$ distribution is compatible with the post-fit expected spectra within the (constrained) uncertainty bands in all signal regions.

Figure 5.11 shows the constraints and variations of the central value of the NPs after the fit to data. Some shifts of some NPs are observed, all within a 1σ variation. In particular the first NPs related to the JES uncertainty, the W+jets normalization in the *resolved1b* SR NP, the multijet background normalization NPs, the NP related to the $t\bar{t}$ NNLO PDFs and the NP relative to the $t\bar{t}$ FSR in the *resolved1b* SR show the largest shifts. Moreover, some constraints are observed for NPs corresponding to background modelling uncertainties, with the strongest constraints appearing on the NPs related to the ME-PS matching uncertainty in the *resolved1b* and *resolved2b* SRs.



(a) Pre-fit distribution.



(b) Post-fit distribution.

Figure 5.10: Pre-fit 5.10(a) and post-fit 5.10(b) distributions of the reconstructed $m_{t\bar{t}}$ for the three signal regions. The expected relative deviation from the background prediction in the presence of a Z' resonance with a mass of 3 TeV, scaled by a factor of five for better visibility, is also shown in both panels.

A new method to compute the impact of different sources of systematic uncertainty has been recently suggested [154]. In this way, the contribution to the uncertainties in the POIs of each systematic source r can be computed as the covariance between the POIs with respect to the selected systematic uncertainty r. Figures 5.12 show the contribution of the first 20 NPs after a fit to data with a function including background and signal prediction. In particular, Figure 5.12(a) and 5.12(b), show the NPs contribution ordered from the largest from top to bottom for the Z' signal model, corresponding respectively to a mass of 750 GeV and 1 TeV. The plots also show the relative shifts and constraints of the associated NPs: the post-fit best value of the fit is shown as a black dot, and its post-fit $\pm 1\sigma$ variation with a black line. The dominant contributions are related to jet uncertainties, $t\bar{t}$ modelling uncertainties and a couple of γ-parameters, representing the statistical uncertainty for bins in the resolved and boosted SRs. Some of the best-fit values for the NPs are shifted from the pre-fit value, still remaining within the 1σ shift. Figure 5.13 shows

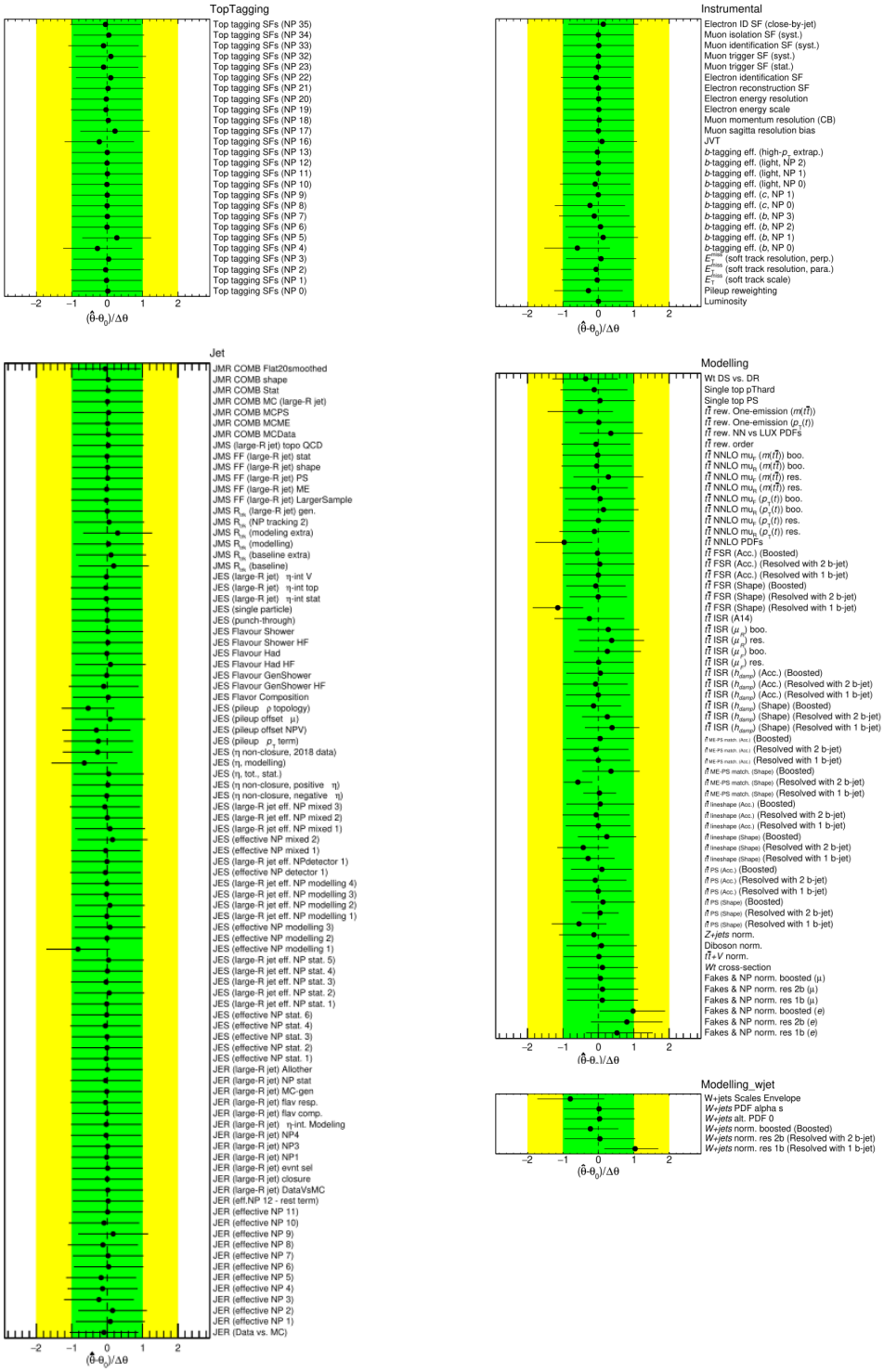
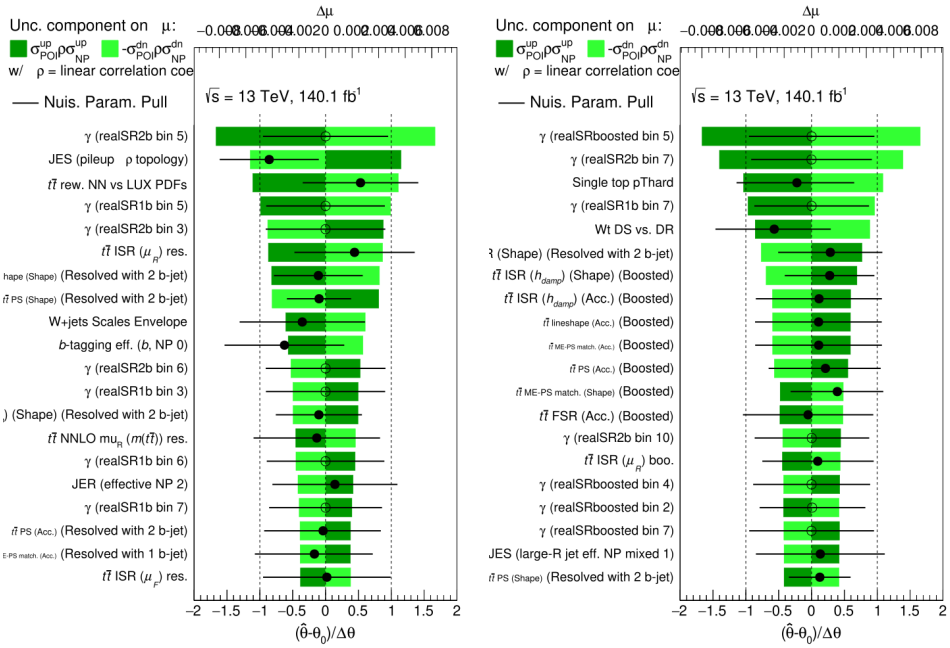


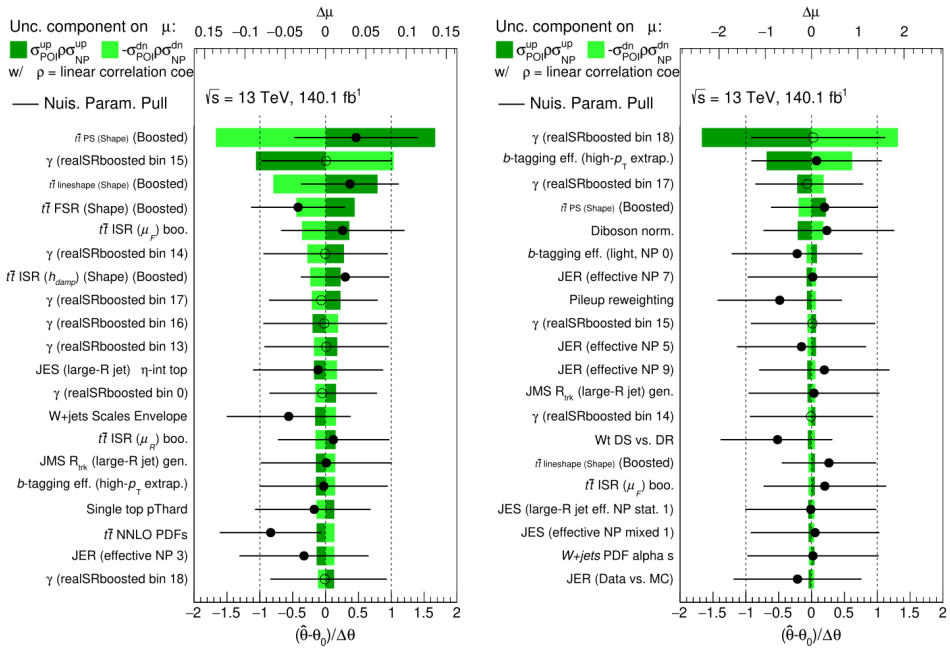
Figure 5.11: Nuisance parameters after a fit to data using the $m_{t\bar{t}}$ distribution in the SRs. NPs are arranged in different groups based on the different systematic categories they belong to. The NPs correspond to systematic uncertainties explained in 5.4.1.

the same ranking plots but for Z' signal with higher masses: 3 TeV and 5 TeV. In Figures 5.13(a), and 5.13(b) the contribution to the POI uncertainty is dominated by γ-parameters relative to the statistical uncertainty of bins of the boosted SR. In particular, they are relative to the highest bins in the $m_{t\bar{t}}$ distribution in the boosted SR used for the Profile LH fit.



(a) Z' 750 GeV (b) Z' 1 TeV

Figure 5.12: Ranking plot showing the contribution of the top 20 sources of uncertainties on the signal strength in a fit to data, together with the relative shifts and constraints on the associated nuisance parameters.

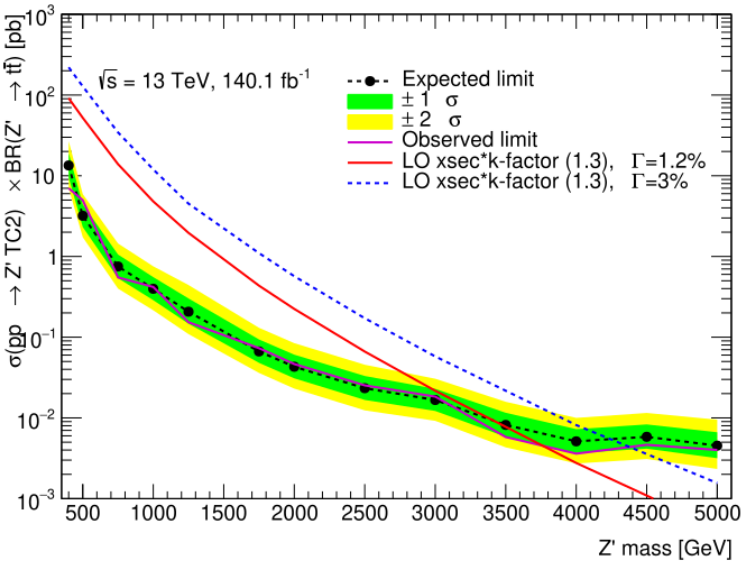


(a) Z' 3 TeV (b) Z' 5 TeV

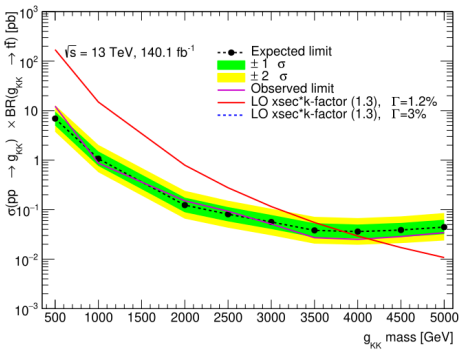
Figure 5.13: Ranking plot showing the contribution of the top 20 sources of uncertainties on the signal strength in a fit to data, together with the relative shifts and constraints on the associated nuisance parameters.

No significant excess is found over the range of signal masses considered, and the largest local significance for the Z' signal corresponds to 1.75σ for a mass of $Z' = 500$ GeV. The upper limit at 95% CL_s has been extracted for all signal sample masses considered in this analysis. The expected upper limit, extracted using an Asimov dataset, and the observed ones using the collected experimental data are shown in Figure 5.14(a) for the Z' signal, in Figure 5.14(b) for the g_{KK} signal and Figure 5.14(c) for the G_{KK} signal. The cross-section limits are extracted for each mass point and are interpolated with straight lines in the regions between the mass points. The observed limit is within the 2σ error band (shown in the Figures in yellow) for all signal masses considered showing no significant deviation from the SM background-only prediction. The considered model can be excluded for the case of the Leptophobic Non-Standard Topcolor Z' up to $m_{Z'} = 3.8$ TeV ($m_{Z'} = 3.5$ TeV expected), for the case of the Kaluza-Klein graviton excitation spin-2 colour singlet up to $m_{G_{KK}} = 1.2$ TeV ($m_{G_{KK}} = 1.2$ TeV expected), and for the Kaluza-Klein gluon excitation spin-1 colour octet up to $m_{g_{KK}} = 4.1$ TeV ($m_{g_{KK}} = 3.85$ TeV).

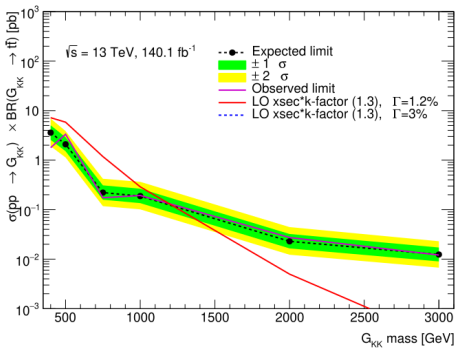
The search provides the tightest upper limits on the production cross-section times branching ratio for low-mass $t\bar{t}$ resonances, excluding, for example, values of the cross-section times branching ratio greater than 7 pb for a Z' with a mass of 450 GeV, and provides the strongest constraints on broad resonances with a relative width of 30%. These exclusion limits show also a clear improvement with respect to the previous ATLAS analysis searching for heavy particles decaying into top-quark pairs in the semileptonic decay channel at $\sqrt{s} = 13$ TeV using an integrated luminosity of $\int L = 36.1$ fb^{-1} [10]. For instance, the excluded mass value for the Leptophobic Non-Standard Topcolor Z' signal was there of $m_{Z'} = 3.2$ TeV ($m_{Z'} = 2.85$ TeV expected), showing that the result discussed in this thesis improves the excluded signal mass value of 0.6 TeV (0.65 TeV) in the observed (expected) excluded signal mass. Similar improvements are found also for the KK gluon and KK graviton signal models.



(a) Z' signal.



(b) g_{KK} signal.



(c) G_{KK} signal.

Figure 5.14: Observed and expected upper limits on the signal cross-section times branching ratio to $t\bar{t}$ for the Z' signal in 5.14(a), for the g_{KK} signal in 5.14(b), and for the G_{KK} in 5.14(c). The theoretical predictions for the production cross-section times branching ratio of the signal process decaying to $t\bar{t}$ at the corresponding masses are also shown.

Conclusions

This thesis presented two complementary physics analyses involving top quarks, performed using data collected by the ATLAS experiment at the Large Hadron Collider, contributing both to precision measurements of Standard Model processes and searches for physics beyond the Standard Model. The candidate was strongly involved in both these physics analyses.

The first analysis achieved the first observation of single top-quark production in the t-channel at the centre-of-mass energy of $\sqrt{s} = 5.02$ TeV, despite the limited integrated luminosity of 255 pb^{-1}. The analysis was performed by selecting events with exactly one charged lepton and two jets, one of which required to be b-tagged. Jets produced in the forward region of the ATLAS detector ($2.5 < |\eta| < 4.0$) are selected to enhance the signal-over-background ratio. A profile likelihood fit to a Boosted Decision Tree discriminant distribution in the ℓ^++jets and ℓ^-+jets channels is performed. The measurement yields a total t-channel single top production cross-section of $\sigma(tq + \bar{t}q) = 27.1^{+4.4}_{-4.1}$ (stat.) $^{+4.4}_{-3.7}$ (syst.) pb, allowing to claim the observation of the process with a significance of 6.1 standard deviations over the background-only hypothesis. The ratio of top to anti-top quark production cross-sections was measured to be $R_t = 2.73^{+1.43}_{-0.82}$ (stat.) $^{+1.01}_{-0.29}$ (syst.), and the value of $f_{LV} \cdot |V_{tb}|$ was extracted to be $0.94^{+0.11}_{-0.10}$ [6]. These results, despite having larger uncertainties than measurements at higher centre-of-mass energies which use significantly larger datasets, provide an independent test of the Standard Model predictions at this new energy and offer new valuable input for constraining parton distribution functions. The candidate played an essential role in the development of the analysis strategy, and its optimisation, as well as in the extraction of the results.

The second analysis consisted in a search for new spin-1 and spin-2 heavy resonances decaying into top-quark pairs, in final states with exactly one lepton, using the full Run 2 dataset, consisting in 140 fb^{-1} of pp data collected at $\sqrt{s} = 13$ TeV. Topologies where the products of hadronic top-quark decays are fully resolved as well as those where they are collimated into larger-radius jets were exploited, to optimise the sensitivity of the search over the full $m_{t\bar{t}}$ range. Ad-hoc *overlap removal* techniques were used in the analysis, allowing to separate charged leptons from hadronic jets while keeping a high selection efficiency at high mass, where leptonic top-quark decay products are mostly collimated. No significant excess over the Standard Model prediction was observed over the whole inspected mass range. Upper limits were set on the production cross-section times branching fraction for three benchmark models. The search provided the tightest upper limits on the production cross-section times branching ratio for low-mass $t\bar{t}$ resonances, excluding, for example, values of the cross-section times branching ratio greater than 7 pb for a Z' with a mass of 450 GeV, and provides the strongest constraints on broad resonances with a relative width of 30%. The presence of a leptophobic Z' boson is excluded up to 3.8 TeV (expected 3.5 TeV), Kaluza-Klein gluon up to 4.1 TeV (expected 3.9 TeV), and Kaluza-Klein gravitons up to 1.2 TeV (expected 1.2 TeV). The excluded values

of the tested signal models substantially extend the sensitivity compared to previous searches in the same decay channel. The candidate contributed in a decisive way to the finalisation of the search, taking care of its re-optimisation and of the statistical analysis for the extraction of the results.

The technical developments presented in the Appendix, regarding electron identification in challenging environments with nearby jets, personally carried out by the candidate, will contribute to the ongoing improvement of physics object reconstruction in ATLAS. These developments are valuable for current physics analyses in ATLAS, and even more so for future searches for new physics. Indeed, higher energy regions and tails of mass distributions will become increasingly populated with data, thanks to the rapidly growing integrated luminosity collected during the ongoing LHC Run-3 data-taking and the foreseen HL-LHC run. In addition, higher instantaneous luminosity values, predicted for the HL-LHC run, will correspond to harsher pile-up conditions, thus increasing the probability of overlap between leptons and jets.

In summary, this thesis presented key measurements and technical developments related to the physics of the top quark from multiple perspectives. It advanced our understanding of Standard Model top-quark production through the t-channel single top-quark production measurement, which observed this process for the first time at the specific energy of $\sqrt{s} = 5.02$ TeV with a small dataset of 255 pb^{-1}. Then it explored the high-energy frontier by searching for new heavy resonances decaying into $t\bar{t}$ pairs with a dataset of 140 fb^{-1}, and setting new constraints on the tested new physics models. Lastly, it also contributed to technical challenges within the ATLAS collaboration, such as the characterization of electron identification in dense hadronic environments.

Appendices

Appendix A

Studies of Electrons Close to Jets

Electrons serve as clean probes for both Standard Model measurements and search for new physics in particle physics analyses. Leptonic final states have a relatively low background compared to hadronic channels, with a clear experimental signature in detectors, and are present in numerous physics processes.

The identification and measurement of electrons can have different levels of complexity depending on their topological configuration. Isolated electrons benefit from both inherent physical advantages and extensive experimental optimization. From a physics perspective, isolated electrons are easier to reconstruct and identify, as nearby particle deposits do not contaminate their detector signatures. Additionally, since most ATLAS physics analyses focus on signatures with isolated electrons, optimized, and extensively validated reconstruction and identification procedures have been developed for isolated electrons. In contrast, the scenario of electrons with small angular separation from jets has received limited dedicated study and is usually excluded from standard physics analyses.

The topological configuration with electrons close to jets is especially relevant in highly boosted scenarios or in processes involving heavy particle decays, where the decay products can be highly collimated. Some important physics processes that are studied with LHC collision data for which electrons close to jets are a crucial tool are:

- Heavy resonance searches, as $Z' \to t\bar{t}$ (for instance the search discussed in Chapter 5) or $W' \to tb$ lead to boosted top quarks. In these cases, the W boson from the top decay can produce an electron that overlaps with the b-jet from the same top decay;

- Vector Boson Fusion (VBF) and Vector Boson Scattering (VBS) processes at high energies, where EW bosons can be produced with a significant boost, leading to collimated decay products;

- Searches for BSM particles with multi-step decay chains, such as super-symmetric scenarios where the new physics particle decays via W bosons and light quarks, potentially resulting in nearby electrons and jets.

As the LHC continues to probe higher energy regimes, such topological configurations become more common due to the boost in final state particles created by heavy particles. The current ATLAS standard practice of excluding electrons with an angular distance $\Delta R(e, \text{jet}) < 0.4$ reduces the signal acceptance and limits our sensitivity to new physics signatures.

The challenges in reconstructing and identifying electrons close to jets arise from several factors: the overlap of calorimeter deposits between the electron and the jet,

the difficulty in correctly reconstructing overlapping tracks and matching them to the correct energy clusters in the calorimeters, and the complex interplay between electromagnetic and hadronic shower developments in the detector. Standard electron identification techniques become significantly less efficient in this scenario. Physics analyses interested in using electrons close to jets need precise measurements of the identification, isolation, reconstruction and trigger electron efficiencies and SFs. Anyway, electron identification efficiencies and relative SFs have not been computed for electrons with $\Delta R(e, \text{jet}) < 0.4$ in the ATLAS collaboration.

In this Appendix the electron identification efficiency as a function of the angular distance ΔR(e,jet) between the electron and the closest jet and the relative SFs are computed in Section A.1. In this context identification efficiencies and SFs are studied also for electrons with $0.2 < \Delta R(e, jet) < 0.4$. In Section A.2 the distributions of variables used for identifying electrons in the ATLAS detector are compared for electrons with $\Delta R \geq 0.4$ and with $0.2 < \Delta R < 0.4$.

A.1 Electron identification efficiencies with respect to the angular distance from a jet

A.1.1 The Tag-and-Probe method

Electron reconstruction, identification and isolation SFs are applied to MC simulations to correct the corresponding efficiencies with respect to data. As briefly mentioned in Section 3.1.2, electron SFs are defined as the ratio of the efficiency in data and the efficiency in the MC simulation A.1:

$$SF = \frac{\varepsilon_{DATA}}{\varepsilon_{MC}}. \tag{A.1}$$

The electron efficiencies can be factorised as a product of the electron reconstruction, identification, isolation, and trigger efficiencies:

$$\varepsilon = \varepsilon_{reco} \cdot \varepsilon_{id} \cdot \varepsilon_{iso} \cdot \varepsilon_{trig} = \frac{N_{reco}}{N_{clus}} \cdot \frac{N_{id}}{N_{reco}} \cdot \frac{N_{iso}}{N_{id}} \cdot \frac{N_{trig}}{N_{iso}}, \tag{A.2}$$

where N_{clus} is the number of reconstructed electromagnetic (EM) calorimeter clusters, N_{reco} is the number of reconstructed electron candidates, N_{id} is the number of identified electron candidates passing previous selections, N_{iso} is the number of isolated electron candidates passing previous selections, and N_{trig} is the number of triggered electron candidates passing previous selections.

To obtain these efficiency measurements in data, a clean and unbiased sample of electrons is needed. The decay of the Z boson into an electron-positron pair is used, in this case, one of the electrons is required to pass strict selection criteria (referred to as *tag*), while the other one serves as an unbiased source to measure the efficiency (referred to as *probe*). The background estimation method considered in this Appendix is the Z-mass method, which uses the invariant mass distribution of the electron-positron pair distribution to subtract the background from data, as defined in [155].

In the ATLAS collaboration, electron efficiencies are computed with the so-called Tag&Probe software [156]. To perform the studies described in this Appendix some developments of the software have been implemented, such as the treatment of jets as physics objects, the implementation of the angular distance between two objects $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, the possibility to choose the variables with respect to which compute the efficiency and SFs, the choice of the closest jet in angular distance

from the electron probe and a modified *basic overlap removal*. As a baseline, the Tag&Probe software does not use any overlap removal technique. Since most electrons are reconstructed both as an electron object and wrongly as a jet object, an overlap removal technique is needed when considering electrons with respect to jets. In this context, the *basic overlap removal* technique implemented in the software the reconstructed jet in the event is rejected if it is found to be within ΔR(e,jet) < 0.2 from any electron in that event.

A.1.2 Datasets and event selections

Electron identification efficiencies shown in this Section have been computed using data collected during *Run 2* in 2018 at an energy in the centre-of-mass of $\sqrt{s} = 13TeV$. Such data are compared with MC simulation of the signal process$Z \to e^+e^-$ process, simulated for the nominal sample with PowhegBox v2 [82] interfaced with Pythia8, while a second one is simulated with the Sherpa generator.

Some minimal event selections have been applied to both the data and MC samples:

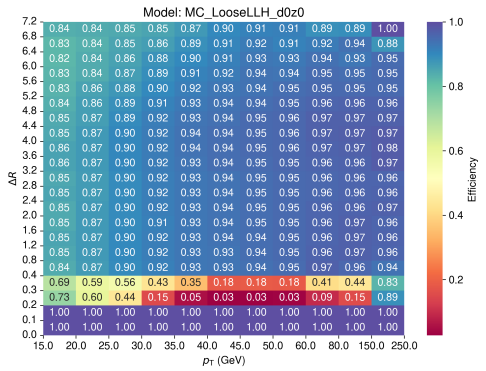
- tag electron: $p_T > 25$GeV, $|\eta| < 2.47$, $|\sigma_{d_0}| < 5.0$, $|\delta_{z_0}|sin(\theta) < 0.5$;
- tag jet: $p_T > 25$GeV, $|\eta| < 2.47$;
- probe electron: $p_T > 15$GeV, $|\eta| < 2.47$;
- at least one tag jet.

A.1.3 Electron identification efficiency and SFs maps with respect to ΔR(e,jet)

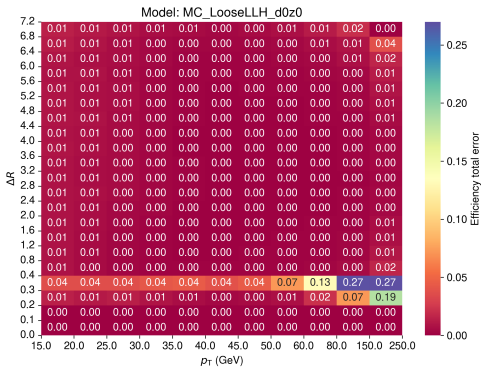
Electron identification efficiency has been computed both for data and MC samples with respect to the transverse momentum p_T and the pseudorapidity η of the electron probe, and the angular distance ΔR(e,jet), for three identification working points (WPs): Loose, Medium and Tight [58], which have been already discussed in Section 3.1.2. Figures A.1 and A.2 show electron identification efficiency map with respect to p_T and ΔR(e,jet). The first two bins in ΔR ($0.0 < \Delta R < 0.2$) do not have any event as expected based on the *basic overlap removal* technique implemented (as explained in A.1.3).

These efficiency maps show that the identification efficiency drops significantly for events with $0.2 < \Delta R(e, jet) < 0.4$. Furthermore, from the error maps, it can be seen that the error on the efficiency is 0.01 for events with $\Delta R > 0.4$, while it is significantly larger for events with $\Delta R < 0.4$. This is mainly caused by the statistical uncertainty, which is much larger for $\Delta R < 0.4$, in particular for high p_T bins.

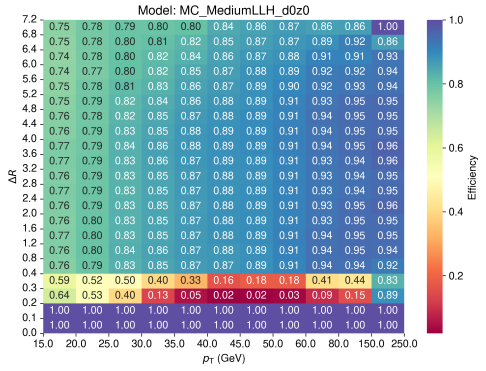
Similar electron identification efficiency maps with respect to the electron probe η and the angular distance ΔR(e,jet) have been also produced and they show similar characteristics to those commented for the previous efficiency maps.



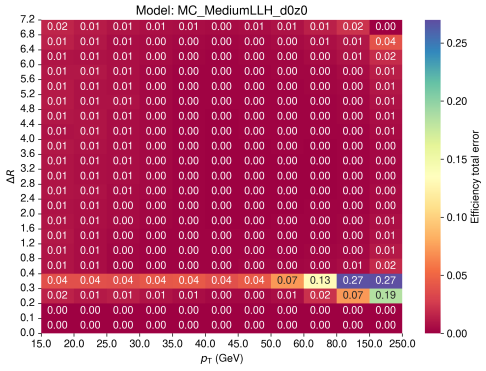
(a) Loose WP efficiency



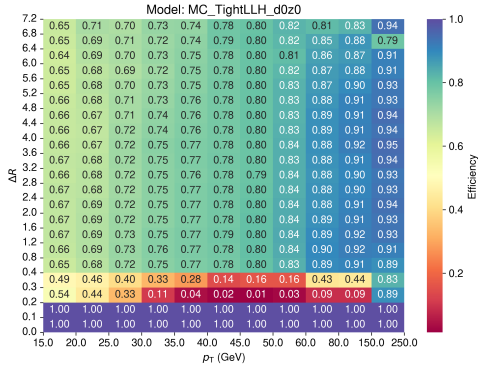
(b) Loose WP error



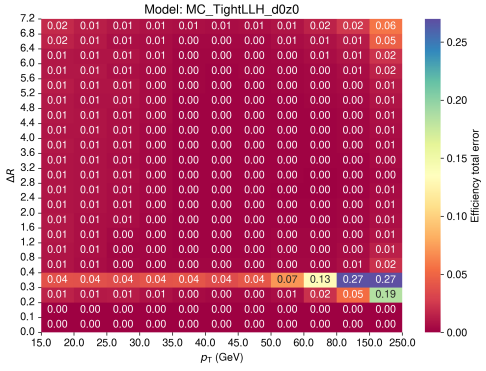
(c) Medium WP efficiency



(d) Medium WP error

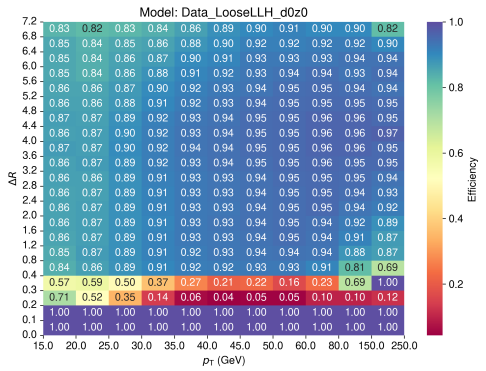


(e) Tight WP efficiency

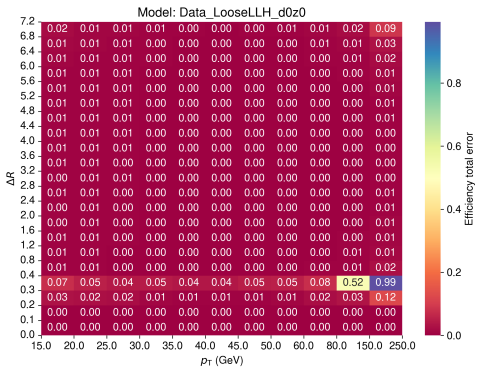


(f) Tight WP error

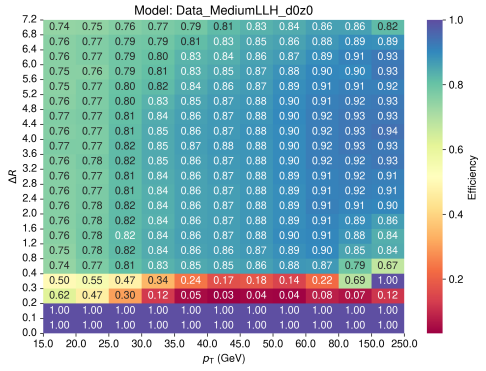
Figure A.1: Electron identification efficiencies for Loose, Medium and Tight identification WP (a, c, e) and their errors (b, d, f) for MC sample with respect to p_T of the electron probe and the angular distance between the electron probe and the tag jet ΔR(e,jet).



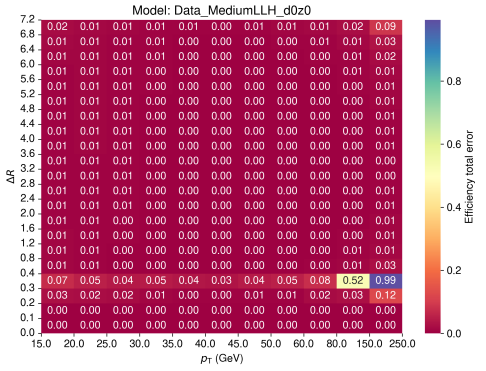
(a) Loose WP efficiency



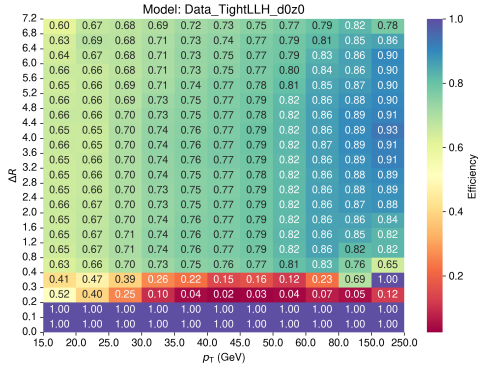
(b) Loose WP error



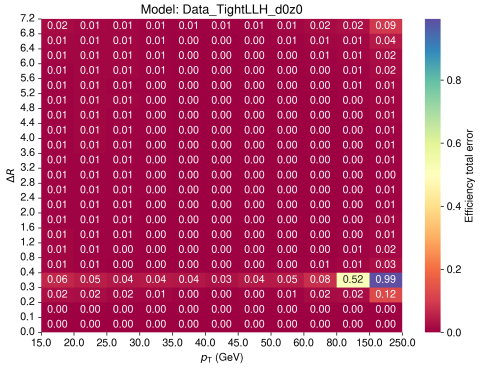
(c) Medium WP efficiency



(d) Medium WP error



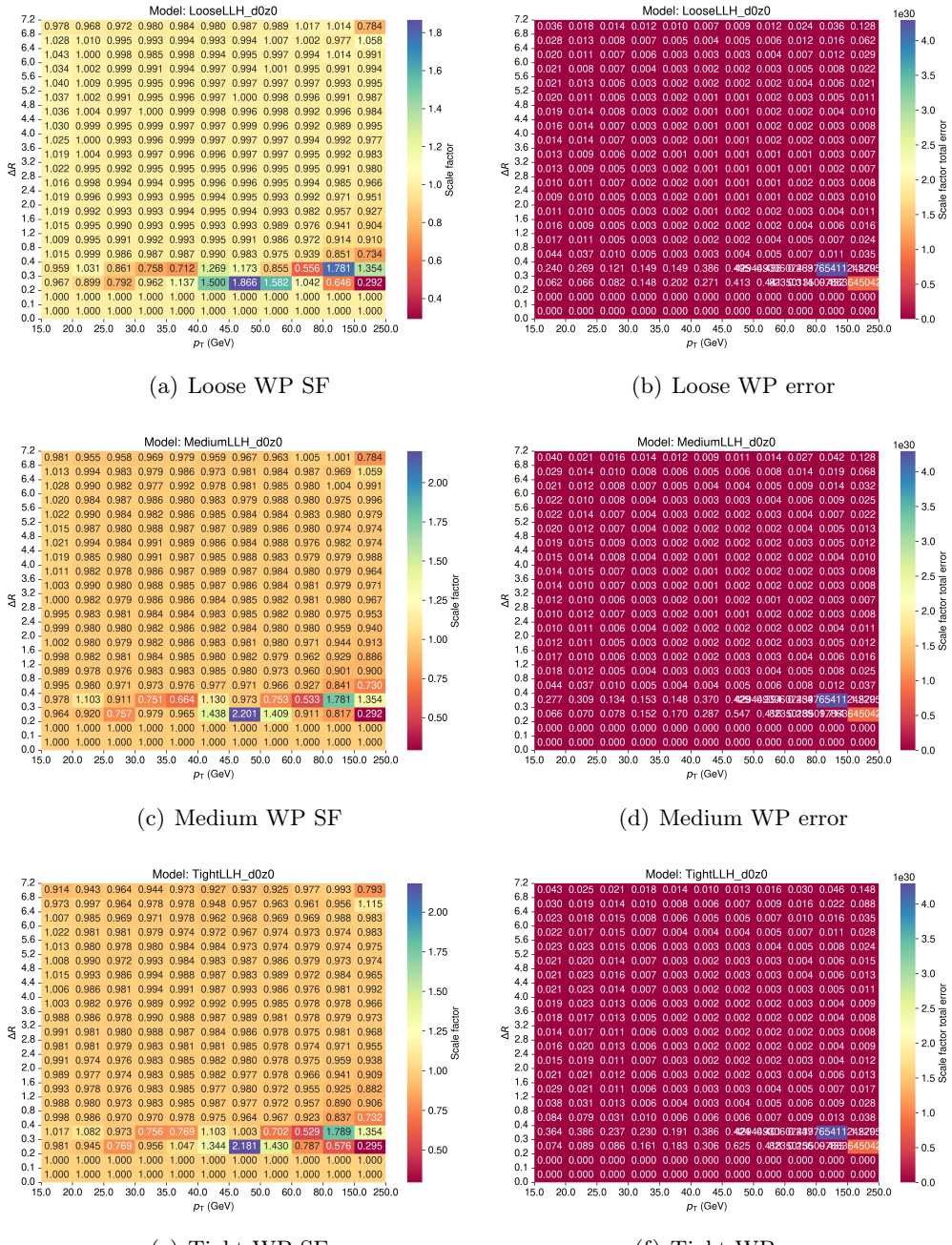
(e) Tight WP efficiency



(f) Tight WP error

Figure A.2: Electron identification efficiencies for Loose, Medium and Tight identification WP (a, c, e) and their errors (b, d, f) for data sample with respect to the electron probe p_T and the angular distance between the electron probe and the tag jet ΔR(e,jet).

Electron identification SFs are computed using the electron identification efficiencies shown in Figure A.1 and A.2. The SFs map and their uncertainties with respect to the electron probe p_T and the angular distance between the electron probe and the tag jet ΔR for the three identification WPs already mentioned are shown in Appendix/figuresA A.3. These maps show that SFs have values close to one for most bins in p_T and ΔR. The SFs values vary more in the two bins for $0.2 < \Delta R < 0.4$, for such bins the uncertainties are bigger, as a consequence of the higher statistical uncertainties for the corresponding efficiencies in data and MC.



(a) Loose WP SF

(b) Loose WP error

(c) Medium WP SF

(d) Medium WP error

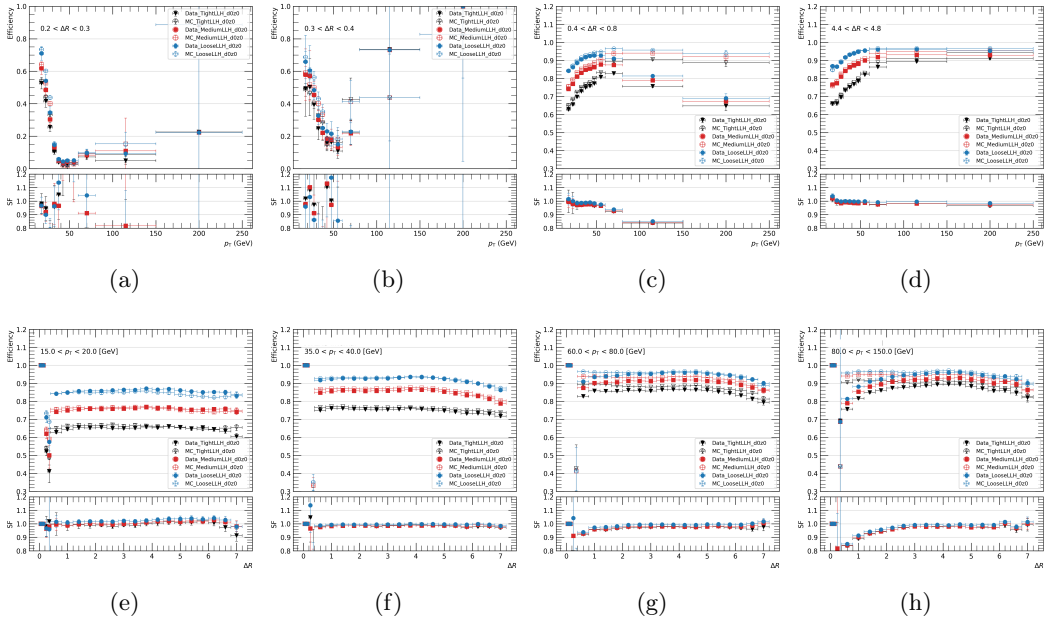
(e) Tight WP SF

(f) Tight WP error

Figure A.3: Electron identification SFs for Loose, Medium and Tight WP (a, c, e) and their errors (b, d, f) with respect to p_T and $\Delta R(e, jet)$.

To look at trends in the efficiencies and SFs with respect to one variable at

a time, plots of the electron identification efficiencies and SFs have been produced with respect to electron probe p_T, for various ΔR bins, as well as with respect to ΔR for p_T bins. Figure A.4 show a few representative plots. Figure A.4(a) and A.4(b) show a different trend for efficiencies than the other efficiency plots with respect to the electron probe p_T for ΔR(e,jet) > 0.4 bins (here only A.4(c) and A.4(d) are show, anyway all other bins have a similar trend). The difference in this efficiency trend for ΔR(e,jet) > 0.2 is investigated in the following, different reasons might be the cause: a possibility is that not all electrons and jets reconstructed objects in events with ΔR(e,jet) < 0.4 are real electron and jets, another possible explanation is that for electrons with ΔR(e,jet) < 0.4 the reconstruction process is not correct as the electron and jet energy clusters and hits overlap with each other. Plots in



(a) (b) (c) (d)

(e) (f) (g) (h)

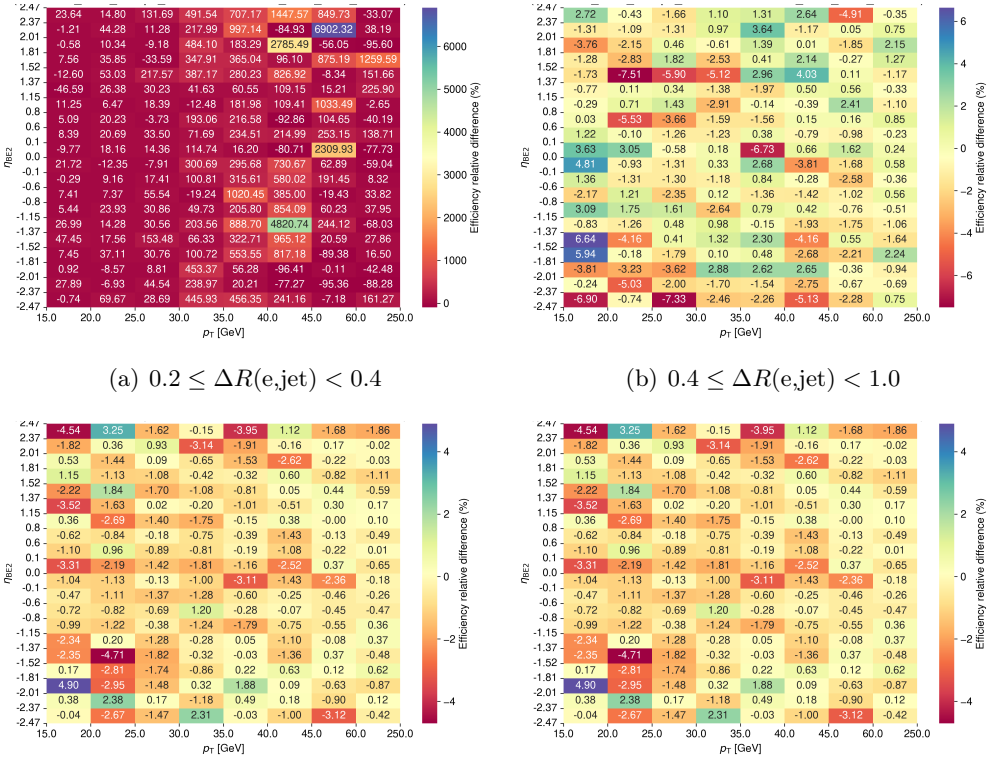
Figure A.4: Electron identification efficiencies with respect to the electron probe p_T for various ΔR(e,jet) bins in A.4(a) to A.4(d), and efficiencies with respect to ΔR(e,jet) for various electron probe p_T bins. The bottom panel shows the correspondent electron identification SF values. Three identification WPs are shown, in black the Tight WP, in red the Medium WP and in blue the Loose WP.

Figures A.4(g) and A.4(h) show that electron identification SFs vary with ΔR(e,jet) for high electron probe p_T bins, effects which are not present in Figures A.4(e) and A.4(f). In particular for electron $p_T > 60$ GeV the SFs reach minimum values 0.8 even for ΔR greater than 0.4.

A.1.4 Different MC generators comparison in electron identification efficiencies

The choice of MC generator can influence the determination of electron identification SFs, particularly for electrons close to jets. Different generators employ different models for parton showering, hadronization, and underlying event simulation, which affect the electron's calorimeter isolation and shower shape variables. The accurate modelling of the jet's internal structure and its interaction with nearby electrons requires precise simulation of both the perturbative QCD processes governing the initial parton shower and the non-perturbative effects controlling hadronization.

To look at this effect on the electron identification efficiencies and SFs shown in the previous Section, efficiencies have been computed for the POWHEG+PYTHIA MC generated sample and for the SHERPA MC generated sample. In this context, the electron identification efficiencies have been computed for four ΔR(e,jet) bins corresponding to: $[0.2 - 0.4]$, $[0.4 - 1.0]$, $[1.0 - 2.0]$ and $[\geq 2.0]$. The identification efficiency and SFs maps are computed with respect to the electron probe p_T and η. The ratio between the efficiencies values obtained with POWHEG+PYTHIA and SHERPA MC generators multiplied by a factor of 100 is shown in Figures A.5. The corresponding SFs map with the ratio between SFs computed with the two MC generators multiplied by a factor of 100 is shown in Figures A.6. The efficiency maps



(a) $0.2 \leq \Delta R$(e,jet) < 0.4

(b) $0.4 \leq \Delta R$(e,jet) < 1.0

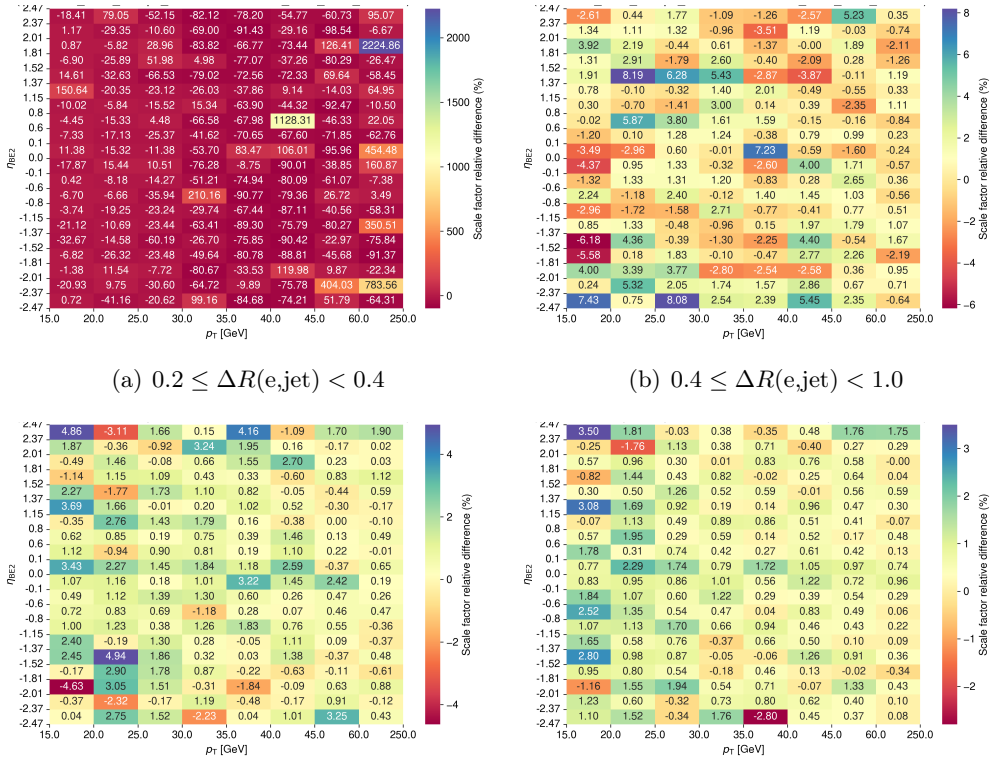
(c) $1.0 \leq \Delta R$(e,jet) < 2.0

(d) ΔR(e,jet) ≥ 2.0

Figure A.5: Electron identification efficiencies difference between the POWHEG+PYTHIA and the SHERPA MC generator for the identification Loose WP with respect to the electron probe p_T and η for four different ΔR(e,jet) intervals. Values of the efficiency in Figure A.5(a) for some bins at high p_T are completely dominated by the statistical error and are therefore not significant.

in Figures A.5(b), A.5(c), A.5(d) show a relative efficiency difference of maximum 7.5% in the so called *crack-region* ($1.37 < |\eta| < 1.52$), and a maximum of 7.3% when excluding the *crack-region*. In general many bins of the relative efficiency difference are over 1%. These values are obtained for what we consider the standard definition of the electron physics objects in the ATLAS collaboration, and it indicates that providing electron identification SFs for different MC generators (differently from what is currently the standard approach) would probably enhance the precision of physics analyses using electrons in the final state. The relative efficiency difference for electrons with $0.2 \leq \Delta R$(e,jet) < 0.4 shown in Figure A.5(a) shows percentage values quite large, and larger that for ΔR(e,jet) ≥ 0.4. As explained above this is

an expected behavior, coming from the important interplay between the hadronic jet parton shower and the hadronization process description of the MC generator models with the electronic shower of the electron probe. It is to be noted here that, as already discussed for the efficiency maps in A.1, values of the efficiency for $0.2 \leq \Delta R(\text{e,jet}) < 0.4$ in bins at high p_T of the electron probe are completely dominated from the statistical error and are therefore not significant. The behaviour of relative



(a) $0.2 \leq \Delta R(\text{e,jet}) < 0.4$

(b) $0.4 \leq \Delta R(\text{e,jet}) < 1.0$

(c) $1.0 \leq \Delta R(\text{e,jet}) < 2.0$

(d) $\Delta R(\text{e,jet}) \geq 2.0$

Figure A.6: Electron identification scale factors difference between the Powheg+Pythia and the Sherpa MC generator for the identification Loose WP with respect to the electron probe p_T and η for four different ΔR(e,jet) intervals. Values of the SFs in Figure A.6(a) for some bins at high p_T are completely dominated by the statistical error and are therefore not significant.

SFs difference between the Powheg+Pythia and the Sherpa MC generator shown in Figure A.6, is very similar to that discussed for the relative efficiency difference in Figure A.5. Variations of up to $\sim 7\%$ are found for the standard electron object definition in the ATLAS collaboration ($\Delta R(\text{e,jet}) \geq 0.4$), while considerably larger differences are seen in the case of $0.2 \leq \Delta R(\text{e,jet}) < 0.4$.

A.2 Study of identification variables for electron close to jets

The electron identification algorithm currently used in the ATLAS collaboration is based on a Likelihood function which uses different observables describing both the track left by the candidate electron in the ID (referred to as track-related variables), and the electromagnetic shower of the candidate electron in the calorimeters (referred to as shower-related variables), more details on the identification process can be found

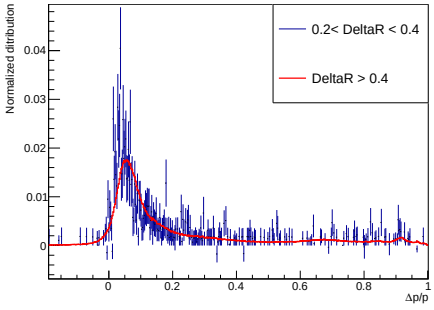
in [58]. The identification algorithm has been optimized in the ATLAS collaboration only considering electrons for which ΔR(e,jet) ≥ 0.4 or for electrons in events with no jets at all. Considering the results shown in Figure A.4(a) and A.4(b), the possibility that track-related and shower-related variables used for the electron identification have considerably different behaviour for electron close to jets ΔR(e,jet) < 0.4 than for the standard electron object definition in ATLAS is studied.

The variables used for the electron Likelihood identification are listed, with a short explanation in Table A.1. To investigate the MC shape of these variables their distributions have been plotted in different bins of the electron p_T and η for electrons with ΔR(e,jet) ≥ 0.4 in blue and for electrons with $0.2 < \Delta R$(e,jet) < 0.4 in red. the binning used is electron $p_T = [15., 20., 30., 40., 50.]$ and electron $|\eta| = [0.00, 0.60, 0.80, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37, 2.47]$.

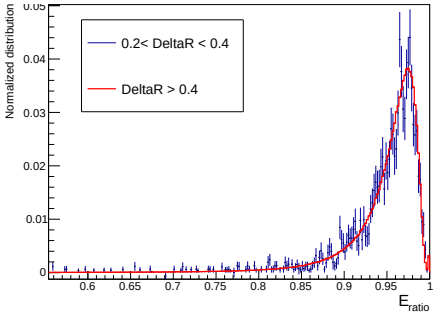
Type	Description	Name
Hadronic leakage	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $\|\eta\| < 0.8$ and $\|\eta\| > 1.37$)	R_{Had1}
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $\|\eta\| > 0.8$ and $\|\eta\| < 1.37$)	R_{Had}
Third layer of EM cal.	Ratio of the energy in the third layer to the total energy	f_3
Middle layer of EM calorimeter	Lateral shower width, $\sqrt{(\Sigma E_i \eta_i^2)/(\Sigma E_i) - ((\Sigma E_i \eta_i)/(\Sigma E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	$W_{\eta 2}$
	Ratio of the energy in 3×3 cells over the energy in 3×7 cells centered at the electron cluster position	R_ϕ
	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centered at the electron cluster position	R_η
Strip layer of EM calorimeter	Shower width, $\sqrt{(\Sigma E_i (i - i_{\mathrm{max}})^2)(\Sigma E_i)}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η, and i_{max} is the index of the highest-energy strip	W_{stot}
	Ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies	E_{ratio}
	Ratio of the energy in the strip layer to the total energy	f_1
Track quality	Number of hits in the B-layer (discriminates against photon conversions)	n_{BLayer}
	Number of hits in the pixel detector	n_{Pixel}
	Number of total hits in the pixel and SCT detectors	n_{Si}
	Transverse impact parameter	d_0
	Significance of transverse impact parameter defined as the ratio of d_0 and its uncertainty	σ_{d_0}
	Momentum lost by the track between the perigee and the last measurement point divided by original momentum	$\Delta p/p$
TRT	Total number of hits in the TRT	n_{TRT}
	Ratio of the number of high-threshold hits to the total number of hits in the TRT	F_{HT}
Track–cluster matching	$\Delta\eta$ between the cluster position in the strip layer and the extrapolated track	$\Delta\eta_1$
	$\Delta\phi$ between the cluster position in the middle layer and the extrapolated track	$\Delta\phi_2$
	Defined as $\Delta\phi_2$, but the track momentum is rescaled to the cluster energy before extrapolating the track to the middle layer of the calorimeter	$\Delta\phi_{\mathrm{Res}}$
	Ratio of the cluster energy to the track momentum	E/p

Table A.1: Electron Likelihood identification input variables [58]

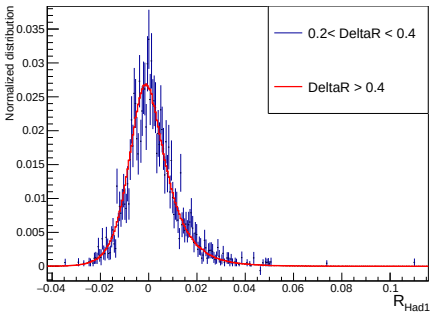
A similar trend for most of the variables used as input for the likelihood identification is found to have a similar shape both for electrons with ΔR(e,jet) ≥ 0.4 and with $0.2 < \Delta R$(e,jet) < 0.4. Some representative plots are shown in Figure A.7. Looking over all distributions, some identification variables are found to have sig-



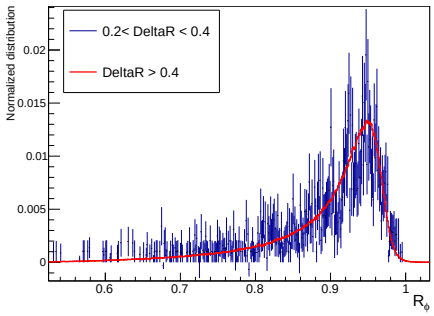
(a) $^{\Delta p}/_{p}$ for $15 < p_T < 30$ GeV and $0.0 < |\eta| < 0.6$



(b) E_{ratio} for $30 < p_T < 50$ GeV and $1.52 < |\eta| < 1.81$



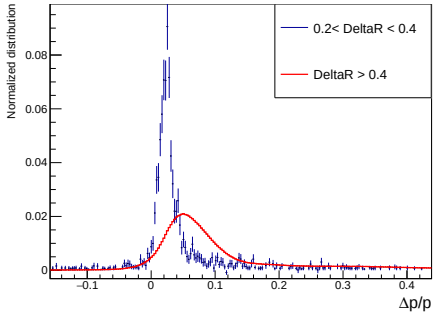
(c) R_{Had1} for $30 < p_T < 50$ GeV and $2.01 < |\eta| < 2.37$



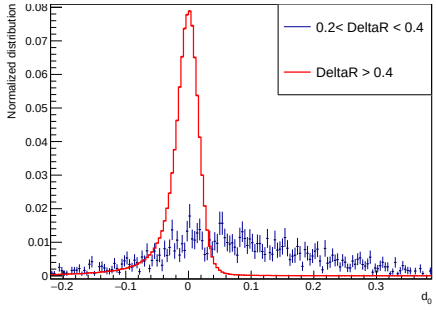
(d) R_{ϕ} for $15 < p_T < 30$ GeV and $0.8 < |\eta| < 1.15$

Figure A.7: Some representative electron varaibles used in the Likelihood electron identification for specific electron p_T and $|\eta|$ intervals. The red lines shows the distribution for electrons with ΔR(e,jet) ≥ 0.4, while the blue dots show the distribution for electrons with $0.2 < \Delta R$(e,jet) < 0.4.

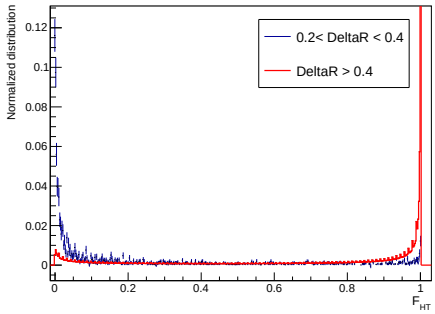
nificantly different shapes for electrons with ΔR(e,jet) ≥ 0.4 and electrons with $0.2 < \Delta R$(e,jet) < 0.4. It is noticed that this happens particularly at higher values of the electron p_T ($p_T \geq 40$ GeV) and higher values of the electron $|\eta|$. It is noticed that the variables more affected by this difference are the so-called track-related variables, such as $^{\Delta p}/_{p}$, n_{TRT}, $\Delta\phi_{Res}$, d_0. Some representative plots are shown in Figure A.8. The difference observed in some of the variables describing the electron track characteristics in the detector when comparing electrons close to jets ($0.2 < \Delta R$(e,jet) < 0.4) to electrons with ΔR(e,jet) ≥ 0.4, might be caused by a non-correct reconstruction of the electron track, caused by various hits left in the ID very close by between the electron and the hadronic jet. This suggests that a further study of the electron reconstruction process for electrons close to jets could lead to a better understanding of this behaviour.



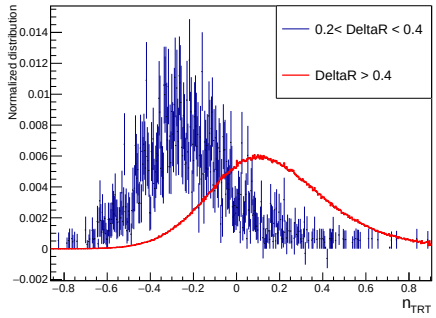
(a) $\Delta p/p$ for $30 < p_T < 50$ GeV and $0.0 < |\eta| < 0.6$



(b) d_0 for $30 < p_T < 50$ GeV and $1.81 < |\eta| < 2.01$



(c) F_{HT} for $30 < p_T < 50$ GeV and $1.52 < |\eta| < 1.81$



(d) TRT_{PID} for $30 < p_T < 50$ GeV and $1.52 < |\eta| < 1.81$

Figure A.8: Some representative electron variables used in the Likelihood electron identification for specific electron p_T and $|\eta|$ intervals. The red line shows the distribution for electrons with ΔR(e,jet) ≥ 0.4, while the blue dots show the distribution for electrons with $0.2 < \Delta R$(e,jet) < 0.4.

Bibliography

[1] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024.

[2] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[3] LHC Machine. *JINST*, 3:S08001, 2008.

[4] T. Aaltonen et al. First Observation of Electroweak Single Top Quark Production. *Phys. Rev. Lett.*, 103:092002, 2009.

[5] V. M. Abazov et al. Observation of Single Top Quark Production. *Phys. Rev. Lett.*, 103:092001, 2009.

[6] ATLAS Collaboration. Measurement of t-channel single-top-quark production in pp collisions at s=5.02 TeV with the ATLAS detector. *Phys. Lett. B*, 854:138726, 2024.

[7] Robert M. Harris, Christopher T. Hill, and Stephen J. Parke. Cross-Section for Topcolor Z'_t Decaying to $t\bar{t}$. 11 1999.

[8] A. Liam Fitzpatrick, Jared Kaplan, Lisa Randall, and Lian-Tao Wang. Searching for the Kaluza-Klein Graviton in Bulk RS Models. *JHEP*, 09:013, 2007.

[9] Ben Lillie, Lisa Randall, and Lian-Tao Wang. The Bulk RS KK-gluon at the LHC. *JHEP*, 09:074, 2007.

[10] ATLAS Cllaboration. Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 78(7):565, 2018.

[11] Abdus Salam and John Clive Ward. Electromagnetic and weak interactions. *Phys. Lett.*, 13:168–171, 1964.

[12] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

[13] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[14] H. Fritzsch, Murray Gell-Mann, and H. Leutwyler. Advantages of the Color Octet Gluon Picture. *Phys. Lett. B*, 47:365–368, 1973.

[15] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.

[16] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[17] ATLAS Collaboration. Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 71:1577, 2011.

[18] CMS Collaboration. First Measurement of the Cross Section for Top-Quark Pair Production in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 695:424–443, 2011.

[19] P. Nason, S. Dawson, and R. Keith Ellis. The Total Cross-Section for the Production of Heavy Quarks in Hadronic Collisions. *Nucl. Phys. B*, 303:607–633, 1988.

[20] Peter Bärnreuther, Michal Czakon, and Alexander Mitov. Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$. *Phys. Rev. Lett.*, 109:132001, 2012.

[21] Michał Czakon, Paul Fiedler, and Alexander Mitov. Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$. *Phys. Rev. Lett.*, 110:252004, 2013.

[22] M. Beneke, P. Falgari, S. Klein, and C. Schwinn. Hadronic top-quark pair production with NNLL threshold resummation. *Nucl. Phys. B*, 855:695–741, 2012.

[23] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185:2930, 2014.

[24] Scott S. D. Willenbrock and Duane A. Dicus. Production of Heavy Quarks from W Gluon Fusion. *Phys. Rev. D*, 34:155, 1986.

[25] J. A. Aguilar-Saavedra. A Minimal set of top anomalous couplings. *Nucl. Phys. B*, 812:181–204, 2009.

[26] Edmond L. Berger, Jun Gao, C. P. Yuan, and Hua Xing Zhu. NNLO QCD Corrections to t-channel Single Top-Quark Production and Decay. *Phys. Rev. D*, 94(7):071501, 2016.

[27] John Campbell, Tobias Neumann, and Zack Sullivan. Single-top-quark production in the t-channel at NNLO. *JHEP*, 02:040, 2021.

[28] Nikolaos Kidonakis and Nodoka Yamanaka. Higher-order corrections for tW production at high-energy hadron colliders. *JHEP*, 05:278, 2021.

[29] Sarah Heim, Qing-Hong Cao, Reinhard Schwienhorst, and C. P. Yuan. Next-to-leading order QCD corrections to s-channel single top quark production and decay at the LHC. *Phys. Rev. D*, 81:034005, 2010.

[30] Nikolaos Kidonakis. NNLL resummation for s-channel single top quark production. *Phys. Rev. D*, 81:054028, 2010.

[31] Edvige Corbelli and Paolo Salucci. The Extended Rotation Curve and the Dark Matter Halo of M33. *Mon. Not. Roy. Astron. Soc.*, 311:441–447, 2000.

[32] Douglas Clowe, Marusa Bradac, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *Astrophys. J. Lett.*, 648:L109–L113, 2006.

[33] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[34] Hanno Sahlmann. Loop Quantum Gravity - A Short Review. In *Foundations of Space and Time: Reflections on Quantum Gravity*, pages 185–210, 1 2010.

[35] P. Van Nieuwenhuizen. Supergravity. *Phys. Rept.*, 68:189–398, 1981.

[36] Christopher T. Hill. Topcolor assisted technicolor. *Phys. Lett. B*, 345:483–489, 1995.

[37] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.

[38] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

[39] CMS Collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[40] `https://cds.cern.ch/record/2197559`. Accessed: 21/09/2021.

[41] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[42] ALICE collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[43] LHCb collaboration. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[44] CERN. Safety Code F, Radiation Protection, CERN. *EDMS 335729*, 2006.

[45] High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1. 4/2017, 2017.

[46] https://hilumilhc.web.cern.ch/article/ls3-schedule-change. Accessed: 03-10-2024.

[47] Track Reconstruction Performance of the ATLAS Inner Detector $\sqrt{s} = 13$ TeV. *https://cds.cern.ch/record/2037683*.

[48] J.Goodson. http://www.jetgoodson.com/images/thesisImages/magnetSystems.png. Accessed: 03-10-2024.

[49] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ. Accessed: 03-10-2024.

[50] ATLAS Cllaboration. Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC. *Eur. Phys. J. C*, 76(12):653, 2016.

[51] A. Sbrizzi. A Cherenkov detector for monitoring ATLAS luminosity. *Nucl. Phys. B Proc. Suppl.*, 215:232–234, 2011.

[52] S. Abdel Khalek et al. The ALFA Roman Pot Detectors of ATLAS. *JINST*, 11(11):P11013, 2016.

[53] Sebastian White. The ATLAS zero degree calorimeter. *Nucl. Instrum. Meth. A*, 617:126–128, 2010.

[54] Technical Design Report for the ATLAS Inner Tracker Pixel Detector. 2017.

[55] Technical Design Report for the ATLAS Inner Tracker Strip Detector. 4 2017.

[56] R. Müller. System tests of the ATLAS ITk planar and 3D pixel modules. *JINST*, 18(03):C03014, 2023.

[57] W. Lampl et al. Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, 2008.

[58] ATLAS Cllaboration. 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. C*, 79(8):639, 2019.

[59] ATLAS Cllaboration. Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data. *JINST*, 14(03):P03017, 2019.

[60] ATLAS Collaboration. Electron and photon efficiencies in LHC Run 2 with the ATLAS experiment. *JHEP*, 05:162, 2024.

[61] ATLAS Collaboration. Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data. *Eur. Phys. J. C*, 74(11):3130, 2014.

[62] M. Aaboud et al. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2. *Eur. Phys. J. C*, 77(10):673, 2017.

[63] ATLAS Collaboration. Studies of the muon momentum calibration and performance of the ATLAS detector with *pp* collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 83(8):686, 2023.

[64] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[65] ATLAS Cllaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C*, 77(7):466, 2017.

[66] ATLAS Collaboration. Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 81(8):689, 2021.

[67] 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. C*, 76(11):581, 2016.

[68] ATLAS Collaboration. Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 5.02$ TeV with the ATLAS detector. *JHEP*, 06:138, 2023.

[69] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Catchment Area of Jets. *JHEP*, 04:005, 2008.

[70] ATLAS Collaboration. ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset. *Eur. Phys. J. C*, 83(7):681, 2023.

[71] ATLAS Collaboration. Measurement of b-tagging efficiency of c-jets in $t\bar{t}$ events using a likelihood approach with the ATLAS detector. ATLAS Conference Note ATLAS-CONF-2018-001, CERN, 2018.

[72] ATLAS Collaboration. Calibration of light-flavour b-jet mistagging rates using ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV . ATLAS Conference Note ATLAS-CONF-2018-006, CERN, 2018.

[73] ATLAS Cllaboration. Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC. *Eur. Phys. J. C*, 79(5):375, 2019.

[74] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. Energy Correlation Functions for Jet Substructure. *JHEP*, 06:108, 2013.

[75] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 03:015, 2011.

[76] Jesse Thaler and Lian-Tao Wang. Strategies to Identify Boosted Tops. *JHEP*, 07:092, 2008.

[77] ATLAS Collaboration. Boosted hadronic vector boson and top quark tagging with ATLAS using Run 2 data. Public Note AATL-PHYS-PUB-2020-017, CERN, 2020.

[78] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823–874, 2010.

[79] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.

[80] B. R. Webber. A QCD Model for Jet Fragmentation Including Soft Gluon Interference. *Nucl. Phys. B*, 238:492–528, 1984.

[81] Enrico Bothmann et al. Event Generation with Sherpa 2.2. *SciPost Phys.*, 7(3):034, 2019.

[82] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[83] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[84] Enrico Bothmann et al. Event Generation with Sherpa 2.2. *SciPost Phys.*, 7(3):034, 2019.

[85] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[86] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016.

[87] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152–155, 2001.

[88] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[89] ATLAS Collaboration. The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim. ATLAS Public Note, ATL-PHYS-PUB-2010-013, CERN, 2010.

[90] ATLAS Collaboration. ATLAS Pythia 8 tunes to 7 TeV data. ATLAS Public Note ATL-PHYS-PUB-2014-021, CERN, 2014.

[91] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J. C*, 71:1547, 2011.

[92] ATLAS Collaboration. Tools for estimating fake/non-prompt lepton backgrounds with the ATLAS detector at the LHC. *JINST*, 18(11):T11004, 2023.

[93] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. 6 2012.

[94] Wouter Verkerke and David P. Kirkby. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003.

[95] Alexander L. Read. Presentation of search results: The CL_s technique. *J. Phys. G*, 28:2693–2704, 2002.

[96] Victor Mukhamedovich Abazov et al. Model-independent measurement of t-channel single top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Lett. B*, 705:313–319, 2011.

[97] T. Aaltonen et al. Observation of Single Top Quark Production and Measurement of $|V_{tb}|$ with CDF. *Phys. Rev. D*, 82:112005, 2010.

[98] ATLAS Collaboration. Comprehensive measurements of t-channel single top-quark production cross sections at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev. D*, 90(11):112006, 2014.

[99] CMS Collaboration. Measurement of the Single-Top-Quark t-Channel Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 12:035, 2012.

[100] ATLAS Cllaboration. Fiducial, total and differential cross-section measurements of t-channel single top-quark production in pp collisions at 8 TeV using data collected by the ATLAS detector. *Eur. Phys. J. C*, 77(8):531, 2017.

[101] CMS Collaboration. Measurement of the t-channel single-top-quark production cross section and of the $| V_{tb} |$ CKM matrix element in pp collisions at $\sqrt{s}$= 8 TeV. *JHEP*, 06:090, 2014.

[102] CMS Collaboration. Measurement of the single top quark and antiquark production cross sections in the t channel and their ratio in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 800:135042, 2020.

[103] ATLAS Collaboration. Measurement of t-channel production of single top quarks and antiquarks in pp collisions at 13 TeV using the full ATLAS Run 2 data sample. *JHEP*, 05:305, 2024.

[104] John M. Campbell, Rikkert Frederix, Fabio Maltoni, and Francesco Tramontano. Next-to-Leading-Order Predictions for t-channel Single-Top Production at Hadron Colliders. *Phys. Rev. Lett.*, 102:182003, 2009.

[105] ATLAS Collaboration. VElectron corrections for low pile-up runs taken in 2017 and 2018. ATLAS Conference Note, ATL-COM-PHYS-2019-077, CERN, 2019.

[106] ATLAS Collaboration. Muon corrections for low pile-up runs taken in 2017 and 2018. ATLAS Conference Note, ATL-COM-PHYS-2019-072, CERN, 2015.

[107] ATLAS Collaboration. ATLAS data quality operations and performance for 2015–2018 data-taking. *JINST*, 15(04):P04003, 2020.

[108] ATLAS Collaboration. Luminosity determination for low-pileup datasets at $\sqrt{s} = 5$ and 13 TeV using the ATLAS detector at the LHC. ATLAS Conference Note ATLAS-CONF-2020-023, CERN, 2020.

[109] Rikkert Frederix, Emanuele Re, and Paolo Torrielli. Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO. *JHEP*, 09:130, 2012.

[110] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244–289, 2013.

[111] John Campbell and Tobias Neumann. Precision Phenomenology with MCFM. *JHEP*, 12:034, 2019.

[112] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185:2930, 2014.

[113] Stefano Frixione, Eric Laenen, Patrick Motylinski, Bryan R. Webber, and Chris D. White. Single-top hadroproduction in association with a W boson. *JHEP*, 07:029, 2008.

[114] Chris D. White, Stefano Frixione, Eric Laenen, and Fabio Maltoni. Isolating Wt production at the LHC. *JHEP*, 11:074, 2009.

[115] Nikolaos Kidonakis and Nodoka Yamanaka. Higher-order corrections for tW production at high-energy hadron colliders. *JHEP*, 05:278, 2021.

[116] Charalampos Anastasiou, Lance J. Dixon, Kirill Melnikov, and Frank Petriello. High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO. *Phys. Rev. D*, 69:094008, 2004.

[117] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

[118] ATLAS Collaboration. Electron corrections for low pile-up runs taken in 2017 and 2018. ATLAS Conference Note, ATL-COM-PHYS-2019-077, CERN, 2019.

[119] ATLAS Collaboration. Muon corrections for low pile-up runs taken in 2017 and 2018. ATLAS Conference Note, ATL-COM-PHYS-2019-072, CERN, 2019.

[120] ATLAS Collaboration. Vertex reconstruction performance of the ATLAS detector at $\sqrt{s} = 13$ TeV. ATLAS Public Note, ATL-PHYS-PUB-2015-026, CERN, 2015.

[121] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. 3 2016.

[122] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. *Eur. Phys. J. C*, 83(10):982, 2023.

[123] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.

[124] J. Ferrando and D. Wendland. Reference $t\bar{t}$ production cross sections for use in ATLAS analyses. ATLAS Conference Note ATL-COM-PHYS-2014-112, CERN, 2021.

[125] Nikolaos Kidonakis. Next-to-next-to-next-to-leading-order soft-gluon corrections in hard-scattering processes near threshold. *Phys. Rev. D*, 73:034001, 2006.

[126] Frits A. Berends, H. Kuijf, B. Tausk, and W. T. Giele. On the production of a W and jets at hadron colliders. *Nucl. Phys. B*, 357:32–64, 1991.

[127] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[128] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann. HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR. *Comput. Phys. Commun.*, 182:1034–1046, 2011.

[129] P. Kant, O. M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz, P. Rieck, and P. Uwer. HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions. *Comput. Phys. Commun.*, 191:74–89, 2015.

[130] ATLAS Collaboration. A search for $t\bar{t}$ resonances in lepton+jets events with highly boosted top quarks collected in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 09:041, 2012.

[131] ATLAS Collaboration. Search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using 4.7 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. D*, 88(1):012004, 2013.

[132] CMS Collaboration. Search for Anomalous $t\bar{t}$ Production in the Highly-Boosted All-Hadronic Final State. *JHEP*, 09:029, 2012. [Erratum: JHEP 03, 132 (2014)].

[133] CMS Collaboration. Search for Resonant $t\bar{t}$ Production in Lepton+Jets Events in pp Collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 12:015, 2012.

[134] ATLAS Collaboration. A search for $t\bar{t}$ resonances using lepton-plus-jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 08:148, 2015.

[135] CMS Collaboration. Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Phys. Rev. D*, 93(1):012001, 2016.

[136] ATLAS Collaboration. Search for $t\bar{t}$ resonances in fully hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 10:061, 2020.

[137] CMS Collaboration. Search for $\mathrm{t\bar{t}}$ resonances in highly boosted lepton+jets and fully hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 07:001, 2017.

[138] CMS Collaboration. Search for resonant $\mathrm{t\bar{t}}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 04:031, 2019.

[139] John M. Campbell, R. Keith Ellis, Paolo Nason, and Emanuele Re. Top-Pair Production and Decay at NLO Matched with Parton Showers. *JHEP*, 04:114, 2015.

[140] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J. C*, 75(5):204, 2015.

[141] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 09:111, 2009. [Erratum: JHEP 02, 011 (2010)].

[142] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015, 2013.

[143] Nikolaos Kidonakis. Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production. *Phys. Rev. D*, 83:091503, 2011.

[144] Nikolaos Kidonakis. NNLL resummation for s-channel single top quark production. *Phys. Rev. D*, 81:054028, 2010.

[145] Nikolaos Kidonakis. Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-. *Phys. Rev. D*, 82:054018, 2010.

[146] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[147] ATLAS Collaboration. https://twiki.cern.ch/twiki/bin/view/ LHCPhysics/CERNYellowReportPageAt13TeVttH_Process. Atlas webpage, CERN, 2020.

[148] Stefano Catani, Leandro Cieri, Giancarlo Ferrera, Daniel de Florian, and Massimiliano Grazzini. Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO. *Phys. Rev. Lett.*, 103:082001, 2009.

[149] Roberto Bonciani, Tomás Jezo, Michael Klasen, Florian Lyonnet, and Ingo Schienbein. Electroweak top-quark pair production at the LHC with Z' bosons to NLO QCD in POWHEG. *JHEP*, 02:141, 2016.

[150] Michal Czakon, David Heymes, Alexander Mitov, Davide Pagani, Ioannis Tsinikos, and Marco Zaro. Top-pair production at the LHC through NNLO QCD and NLO EW. *JHEP*, 10:186, 2017.

[151] ATLAS Collaboration. Performance of electron and photon triggers in ATLAS during LHC Run 2. *Eur. Phys. J. C*, 80(1):47, 2020.

[152] ATLAS Collaboration. Performance of the ATLAS muon triggers in Run 2. *JINST*, 15(09):P09015, 2020.

[153] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[154] Andrés Pinto, Zhibo Wu, Fabrice Balli, Nicolas Berger, Maarten Boonekamp, Émilien Chapon, Tatsuo Kawamoto, and Bogdan Malaescu. Uncertainty components in profile likelihood fits. *Eur. Phys. J. C*, 84(6):593, 2024.

[155] ATLAS Collaboration. Support Note for electron ID: measurement of electron reconstruction and identification efficiencies. ATL-COM-PHYS-2017-1352, 2018.

[156] https://tagandprobe.readthedocs.io/en/stable/overview.html, Visited on the 28/10/2024.