DIPARTIMENTO DI MATEMATICA E FISICA

CORSO DI DOTTORATO DI RICERCA IN FISICA

XXXII CICLO

Measurement of the ratio of top quark branching fractions $\mathscr{B}(t \to Wb)/\mathscr{B}(t \to Wq)$ in proton-proton collisions at 13 TeV using the ATLAS detector

Dottoranda:
Valentina VECCHIO

Docente Guida:
Prof. Giuseppe SALAMANNA

Coordinatore:
Prof. Giuseppe DEGRASSI

A Mariastella e Francesco.

Contents

Introduction

The measurement of the *Cabibbo-Kobayashi-Maskawa* (CKM) [1,2] matrix element $|V_{\mathrm{tb}}|$ is part of the current Large Hadron Collider (LHC) [3] physics program. The Standard Model (SM) gives no prediction on this quantity, therefore it is necessary to measure it.
Under some hypothesis, it is possible to determine $|V_{\mathrm{tb}}|$ from the single top quark production cross section measurement. Both the CDF and D0 collaborations of Tevatron [4] and ATLAS [5] and CMS [6] experiments at LHC performed this measurement. The average of Tevatron and LHC value gives $|V_{tb}| = 1.019 \pm 0.025$ [7].
The module of V_{tb} can be also measured in top pairs ($t\bar{t}$) events from the quantity R_{b} defined as the the ratio between the top quark decay rate into bottom quarks and all the possible top quark decay rates:

$$R_{\mathrm{b}} := \frac{\mathscr{B}(t \to Wb)}{\mathscr{B}(t \to Wq)} = \frac{|V_{\mathrm{tb}}|^2}{\Sigma_q |V_{tq}|^2}. \tag{1}$$

In the SM scenario q stands for *down*, *strange* and *bottom* quarks and in the assumption that only three generations of quarks exist, $R_{\mathrm{b}} = |V_{tb}|^2$. If, however, more than three generation of quarks existed, $|V_{tb}|$ and thus R_{b} could assume lower values than unity.
R_{b} has been measured by the CDF collaboration both in the dilepton channel [8], finding $R_{\mathrm{b}} = 0.87 \pm 0.07$, and in the final state containing one lepton [9], where $R_{\mathrm{b}} = 0.94 \pm 0.09$. Assuming three generations of quarks the following values of $|V_{\mathrm{tb}}|$ have been respectively determined: $|V_{tb}| = 0.93 \pm 0.04$ and $|V_{tb}| = 0.97 \pm 0.05$. The D0 collaboration provides the combined result of the dilepton and one lepton measurements [10], reporting $R_{\mathrm{b}} = 0.90 \pm 0.04$.
The latest measurement of R_{b} has been performed by the CMS collaboration using Run 1 data at a center of mass energy of 8 TeV in the dilepton channel [11]. The result of the CMS measurement is $R_{\mathrm{b}} = 1.014 \pm 0.003(stat.) \pm 0.032(syst.)$ and a lower limit on $|V_{tb}| > 0.975$ at 95% C.L.
This work presents the measurement of R_{b} performed using ATLAS data collected during Run 2 in 2016 and 2017, at a center of mass energy of 13 TeV. A model independent approach is used, where R_{b} is extracted performing a likelihood fit on the distribution of the b-jets multiplicity. The analysis selects events containing one electron, one muon and exactly two hadronic jets. This results in an extremely pure sample of $t\bar{t}$ events with only 8% of contribution from background processes, involving mostly single top, diboson and other SM processes. A kinematical reconstruction of the events is performed to pair the two jets in the events with the lepton which is most likely coming from the decay of the same top quark. The outcome of the kinematical reconstruction allows to further categorise the events into four regions: three of them are enriched in events in which at least one of the jets is not coming from the top decay, while the remaining one is enriched into events in which both the jets are coming from the top decay. The events falling in this latter region are used to extract R_{b} while the three remaining regions are used to correct using data

the predicted fractions of lepton-jet mis-assignment. This measurement relies on an ad-hoc calibration of the b-tagging efficiency, performed using multi-jets events. The baseline calibration of the identification rate of bottom-jets is widely performed using $t\bar{t}$ events, under the assumption of $R_{\mathrm{b}} = 1$. Therefore, using this calibration would introduce a bias on the measurement of interest in this work. For this reason, an alternative calibration is put in place in the context of the R_{b} measurement. It is based on the so-called $p_{\mathrm{T}}^{\mathrm{rel}}$ method, which exploits the kinematical properties of the semi-leptonic decays of B-hadrons to identify B meson decays within a hadronic jet.
This work is organised as follows. Chapter 1 gives an overview of the Standard Model. Chapter 2 gives the state of the art and the strategy for the precision measurement of the top quark decay heavy flavor content. In Chapter 3 the LHC complex and the ATLAS detector are described. In the same Chapter, a dedicated Section 3.3 contains the description of the reconstruction and calibration of the physics objects. In Chapter 4 a detailed description the b-tagging algorithm and its calibration are given. The methodology used to perform the measurement of R_{b} is extensively shown in Chapter 5 where the events selection 5.2, the kinematical reconstruction 5.3, the fitting strategy 5.4 and a detail discussion of all sources of systematic uncertainty considered 5.5 are provided.
The results are presented in Section 5.8 and while Chapter 6 finally draws the conclusions of this work.

Chapter 1

The Standard Model of Particle Physics

The Standard Model of particle physics describes the elementary particles and their interactions. This theory has been developed in the second half of 1900 and it is the most thoroughly tested theories of modern physics. The SM gives highly precise predictions of the existing experimental observations. Moreover, it provided a road map for the discovery of new particles. The recent top quark discovery by the Tevatron experiments CDF and D0 [12,13] and the tau neutrino observation [14] are two examples of the high prediction power of the SM.
The latest success of the SM comes with the observation of the Higgs Boson, predicted in 1964 by P. Higgs and F. Englert [15,16] and later in 1967 by S. Weinberg [17]: this particle represented the missing piece of the SM, as its associated field gives an explanation of the massive nature of the fundamental particles (more details in Section 1.4). In 2012, ATLAS [18] and CMS [19] experiments at the LHC claimed the discovery of a particle compatible with the Higgs Boson, having a mass of roughly 125 GeV/c^2.
Although its robustness, the SM is not able to provide an explanation for a variety of experimental observations such as the neutrinos mass and oscillations, the origin of the dark matter and the asymmetry in the matter and anti-matter budget of the universe. Moreover, this theory is completely decoupled from General Relativity, which is the current most accurate description of the Gravitational Interaction.
These latter arguments are all indications pointing to the existence of physics beyond the SM (BSM). In this context, the precision measurements of the SM particles properties could lead the way to an extension of the SM theory or its complete revolution.
In this thesis, the measurement of the top quark branching fractions ratio is presented. The aim of this work is to probe for any BSM physics through a precise measurement of a SM parameter. This Chapter gives the basis of the theoretical framework, covering a brief overview on the SM 1.1 together with the description of the Electroweak 1.2 and Quantum Chromodynamics 1.3 theories. Section 1.4 briefly summarises the spontaneous symmetry breaking mechanism. These first sections prepare the field for a closer look to the physics of top quark, described in Section 1.5.

1.1 Introduction to the Standard Model

The SM is the theory describing the particles and the way they interact. In this context the particles are classified into four groups: leptons, quarks, gauge bosons and one scalar boson.
Leptons and quarks are fermions of spin 1/2, they are constituent of the matter and

they are divided into three doublets. Concerning the leptons, each doublet have one charge lepton and one neutrino. The location of leptons in doublets is based on the flavour (or generation). The lightest charged lepton is the electron, followed by the muon and finally the tau. Neutrinos are considered to be massless in the SM scenario, but there are experimental evidences of their massive nature [20–22].

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}.$$

Quark are classified as follows: in each doublet there always is a quark-up like, which has $+2/3$ electric charge and a quark-down like which has $-1/3$ electric charge.

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}.$$

According to the Dirac equation all these fermions have an associated anti-particle, having opposite quantum numbers. Leptons and quarks interact via the electromagnetic and weak force, while only quarks feel the strong force.
The gauge bosons are the mediators of the interactions. Gluons are massless bosons, mediators of the strong force. Also the photon happens to be massless: this particle mediates the electromagnetic force. Finally the $W^{\pm}$ and Z massive vector bosons are the mediators of the weak force, responsible of respectively the charged and neutral current.
The SM is a Quantum Field Theory (QFT) based on the gauge symmetry $SU(2)_L \times U(1)_Y \times SU(3)_C$. A brief description of the basic concepts of QFT are given in the following.

1.1.1 Quantum Field Theory: the basis

Let ψ be a physical system whose dynamics is described by a given Lagrangian $\mathscr{L}$ which is invariant under a global symmetry G. Following Noether's theorem, the associated current and charge are conserved. By promoting the global symmetry to a local one, the original free theory transforms into an interacting theory. In order to get the theory invariant under local transformations vector boson fields are introduced, interacting with the ψ field in a gauge invariant manner.
Taking as an example the U(1) symmetries, the following transformation

$$\psi \rightarrow \psi' = e^{i\alpha}\psi, \tag{1.1}$$

rotates the field by a global phase α. An example of this kind of symmetry is the $U(1)_L$, where the conserved current is the leptonic current and the conserved charge is the leptonic number. Promoting the previous transformation to a local transformation means introducing a space dependency in the phase α

$$\psi \rightarrow \psi' = e^{i\alpha(x)}\psi, \tag{1.2}$$

giving birth to a gauge boson field. The U(1) group has one generator and correspondingly one gauge field: in the Quantum Electrodynamics (QED) case this particle is the photon. In the SU(2) case there are three gauge bosons: in the Weak (WK) case these particles are the $W^{\pm}$ and Z bosons. In the SU(3) case eight gauge bosons are present: in the Quantum Chromodynamics (QCD) case these particles are the gluons, responsible of the exchange of the color charge.

Given these premises, the SM is now a well define QFT in which the QED and WK are unified into the symmetry group of the Electroweak (EWK) interactions $SU(2)_L \times U(1)_Y$, while the strong interaction is defined by the $SU(3)_C$ symmetry group.

1.2 Electroweak Theory

The Electroweak theory has been formulated between 1960 and 1970 [23–25]. Here the electromagnetic and weak interactions were unified to describe the leptons interactions with photons and weak mediators. The same is valid for quarks, which also interact via the strong interaction (described in Section 1.3).
Let the $\psi = \psi(x)$ field be the Dirac spinor of a spin $\frac{1}{2}$ fermion, whose dynamics is described by the Dirac Lagrangian

$$\mathscr{L} = \bar{\psi}(i\partial_\mu \gamma^\mu - m)\psi, \tag{1.3}$$

where $\bar{\psi} = \psi^\dagger \gamma^0$ and γ^μ are the Dirac gamma matrices. Applying the same transformation as Equation 1.2 the Lagrangian becomes

$$\mathscr{L} \to \mathscr{L}' = \mathscr{L} - \bar{\psi}\gamma^\mu \partial_\mu \alpha(x)\psi. \tag{1.4}$$

Requiring the conservation of the U(1) symmetry implies the definition of the covariant derivative D_μ, which has to ensure

$$D_\mu \psi(x) \to D'_\mu \psi(x)' = e^{i\alpha(x)} D_\mu \psi(x). \tag{1.5}$$

The previous requirement is satisfied when

$$D_\mu \psi(x) = (\partial_\mu + ieA_\mu)\psi(x). \tag{1.6}$$

A_μ is the gauge field of QED, whose Lagrangian is the following

$$\mathscr{L}_{QED} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi}(i\partial_\mu \gamma^\mu - m)\psi, \tag{1.7}$$

where the first term is the kinetic term. In Abelian gauge group case, as the QED, the $F_{\mu\nu}$ tensor can be defined in terms of D_μ.

$$[D_\mu, D_\nu]\psi = ieF_{\mu\nu}\psi. \tag{1.8}$$

The unified EWK theory is obtained requiring the $SU(2)_L \times U(1)_Y$ symmetry, under the $SU(2)_L$ local gauge isospin transformation

$$\psi \to \psi' = e^{i\alpha(x)\cdot\frac{\sigma}{2}}\psi, \tag{1.9}$$

with σ the vector of Pauli spin matrices (generators of the SU(2) symmetry).
The resulting EWK Lagrangian consists of the kinetic part of both SU(2) and U(1) and the interaction term between matter and fields

$$\mathscr{L}_{EWK} = -\frac{1}{4}\Sigma_{a=1}^{3} W^a_{\mu\nu} W^{\mu\nu a} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{\psi}_L (iD_\mu \gamma^\mu)\psi_L + \bar{\psi}_R (iD_\mu \gamma^\mu)\psi_R, \tag{1.10}$$

where

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g f^{abc} W^b_\mu W^c_\nu, \tag{1.11}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{1.12}$$

B_μ is the U(1) field while the three SU(2) vector fields are W^a_μ. f^{abc} is the Levi-Civita symbol and g the electroweak coupling.
The first and second components of W^a_μ are combined to get $W^\pm_\mu = \frac{1}{2}(W^1_\mu \mp W^2_\mu)$, while the B and W_3 are rotated by the Weinberg angle θ_W to generate the EM vector potential and the neutral weak vector potential Z.
Applying all these definitions, one obtains:

$$A_\mu = B_\mu cos\theta_W + W^3_\mu sin\theta_W, \tag{1.13}$$

$$Z_\mu = -B_\mu sin\theta_W + W^3_\mu cos\theta_W. \tag{1.14}$$

In the EWK theory, the third component of the weak isospin I_3 and the weak hypercharge Y are invariant under gauge transformations. The relation between these two quantities and the electric charge is the following

$$Q = I_3 + \frac{Y}{2} \tag{1.15}$$

1.3 Quantum Chromodynamics Theory

The Quantum Chromodynamics is a quantum field theory that describes the strong interaction via the $SU(3)_C$ symmetry group. Here the color charge C is invariant under gauge transformations. The strong mediators are eight spin-1 and massless gluon field $G^a_\mu (a = 1, ..., 8)$ that interact with quarks q. The resulting Lagrangian is the following

$$\mathscr{L}_{QCD} = -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a + \bar{q}(iD_\mu \gamma^\mu - m)q, \tag{1.16}$$

where the gluon field strength

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + e_s f^{a\beta\gamma} G_{\mu\beta} G_{\nu\gamma}, \tag{1.17}$$

denoting the non-Abelian nature of QCD. The algebra of $SU(3)_C$ is indeed non-commuting: this implies that self-interactions of the gluon fields are permitted by the theory. This structure leads also two properties of the QCD theory: the asymptotic freedom and the confinement. Both of these properties are connected to the behaviour of the strong coupling constant α_S as a function of the transfer momentum p of the interacting particles. In particular, α_S decreases with the increasing of p^2 and asymptotically vanishes for infinite value of p^2. In this regime, QCD can be treated with a perturbative approach. Concerning the confinement, one can take as an example the separation to an infinite distance of two colored particles. The aim of this easy exercise would be to understand if a colored particle can be observed alone. However, it appears that the creation of quark-antiquarks pairs from the vacuum is energetically favored compared to the binding energy, having as a result the impossibility to isolate a colored particle.

1.4 The mass of the particles

The EWK and QCD Lagrangians do not contain terms related to the gauge boson and fermion masses: this is in contrast with several experimental evidences that showed these particles are massive. By spontaneous breaking the symmetry via the

Higgs mechanism, the Yukawa interaction term arises, giving an explanation to the massive nature of the fundamental particles. The idea is to introduce ad additional complex scalar multiplet, having spin-0, of the $SU(2)_L$ gauge group

$$\phi = \begin{pmatrix} \phi+ \\ \phi^0 \end{pmatrix}, \tag{1.18}$$

with electrically charged and neutral components. The corresponding Higgs Lagrangian is

$$\mathscr{L}_{Higgs} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi), \tag{1.19}$$

where the potential $V(\phi^\dagger \phi)$ contains quadratic terms in ϕ so that the theory is renormalizable

$$V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{1.20}$$

Spontaneous symmetry breaking is induced if the minimum of V is obtained for non-vanishing ϕ values. This occurs because $\mu^2 < 0$ and $\lambda > 0$ and the potential minimum is then reached when

$$|\phi_0^2| = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}, \tag{1.21}$$

with v the vacuum expectation value different from 0. Selecting one specific vacuum state breaks the EWK gauge symmetry, preserving the full gauge symmetry of the Lagrangian.
Since the ϕ doublet is $SU(2)_L$ invariant, it is always possible to find a gauge transformation that removes the upper component. Applying this transformation one obtains:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{1.22}$$

Starting from here, a deviation σ around the vacuum expectation value gives:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+\sigma \end{pmatrix}, \tag{1.23}$$

where σ is exactly the scalar field which represents the physical Higgs boson. By adding a term in the EWK Lagrangian that takes into account this new scalar field it is possible to obtain mass terms for fermions and bosons.
The interaction between the Higgs and the fermion fields in the form the Yukawa Lagrangian is:

$$\mathscr{L}_{Yukawa} = \Sigma_{f=l,q} y_f [\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi \psi_L], \tag{1.24}$$

where y_f are the matrices containing the Yukawa coupling constants between the fermions and the Higgs boson. The matrices y_f can be diagonalised in order to get the eigenvalues of the Yukawa couplings using unitary transformations that will redefine the fermion fields. In the leptonic sector this transformation has no effect due to the absence of right-handed neutrinos. However, in the quark sector, the rotation to the mass eigenstate basis provides a mixing among the fermions which is the manifestation of the weak interactions. The mixing among the weak eigenstates of the down-type quarks (d', s', b') and the corresponding mass eigenstates (d, s, b)

is characterised by the known Cabibbo-Kobayashi-Maskawa matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.25}$$

where

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1.26}$$

The off-diagonal elements of the CKM matrix explain that W bosons can couple to two quarks belonging to two different generations. The CKM matrix has four parameters; three mixing angles that control the mixing among each generation pair and one complex phase responsible for CP-violating phenomena. The resulting formula for the mass of the fermions is

$$m_f = y_f \frac{v}{\sqrt{2}}, \tag{1.27}$$

while the coupling to the gauge vector bosons is $2m_V^2/v$. From these two latter results one can conclude that the coupling of the Higgs boson in proportional to the mass of the particle, meaning it will be more favorably produced in association with heavy particles, and will decay more favorably into the heaviest particles that are kinematically allowed.
In this way, the Standard Model scheme is finally completed.

1.5 Top Quark physics

The *top* quark has been discovered in 1995 by the Tevatron experiments CDF and D0 [12, 13]. Since then the most massive particle of the SM of particle physics has been studied in great detail, both at the Tevatron and at the LHC. The world average of the *top* quark mass, coming from the combination of the measurement performed by the Tevatron and LHC experiments [26] is

$$m_{top} = 173.34 \pm 0.27(stat) \pm 0.71(syst)\ GeV/c^2. \tag{1.28}$$

Being its Yukawa coupling $y_top \approx 1$, the *top* quark is the heaviest of the fundamental particles known: thanks to its high mass the *top* quark has a very short lifetime, smaller than the typical hadronisation time. For this reason, contrary to what happens to the lighter quarks, the *top* quark is never confined in bound states (so called *hadrons*). This peculiarity offers the unique opportunity to study the properties of a "bare" quark.
The *top* quark cross section at the LHC is significantly high (roughly $1nb$ over the total $100\mu b$): for this reason the LHC is also known to be a "*top* factory".
Given the big amount of *top* quarks produced at the LHC, precision measurements of their properties can be performed. These measurements are motivated by a variety of reasons: first of all the *top* quark is a background in many searches for physics BSM. Its precise understanding is therefore crucial to claim new physics.
Moreover, thanks to its special properties described above the *top* quark is considered to be a good candidate to probe for physics BSM. In this direction, it is crucial to measure all the production, decay and coupling properties of this particle with the highest achievable precision.

In this Section a description of the *top* quark production and decay at the hadron colliders is given.

1.5.1 Production of Top Quarks

The top quark can be produced at hadron colliders in pairs ($t\bar{t}$) via the strong interaction or singly (*single top*) via the electroweak interaction.

Top pairs

Figure 1.1 shows the Feynman diagrams of $t\bar{t}$ at the Leading Order (LO) of QCD: the resulting cross-section is proportional to α_S^2 and it is strongly dependent on the collision energy reached by the accelerator and on the Parton Distribution Functions (PDFs) of the initial state partons.

Here the physics effects at high energies are factorised with the small energies

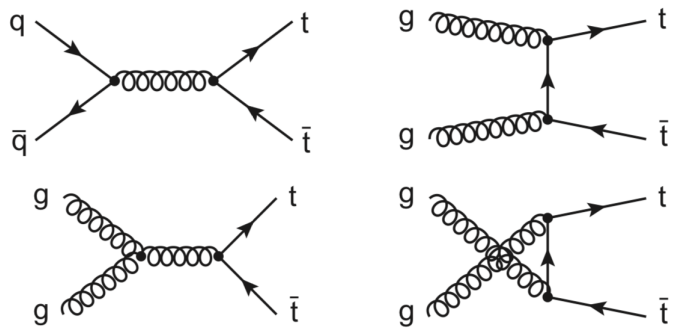


FIGURE 1.1: Feynman Diagrams of the $t\bar{t}$ production at the hadron colliders at the LO of QCD. $q\bar{q}$ annihilation (top left), gluon-gluon (gg) fusion in the s-channel (bottom left), gg fusion in the t-channel (top right), and gg fusion in the u-channel (bottom right) [27]

effects. The total cross section for the production of the $t\bar{t}$ final state via these amplitudes can be expressed as a cross section for the high-energy parton-parton scattering weighted by the PDFs of the participating partons integrated over all the parton momenta and summed over all parton types. The hard scattering cross sections can be computed in perturbative QCD, while the PDFs are usually independently determined from fits to Deep Inelastic Scattering (DIS) cross section measurements performed by various experiments. The resulting factorisation formula is the following

$$\sigma = \Sigma_{j,k} \int_0^1 dx_j dx_k f_j(x_j, \mu_F^2) f_k(x_k, \mu_F^2) \hat{\sigma_{jk}}(x_j x_k s, \mu_F, \alpha_S(\mu_R)). \tag{1.29}$$

The PDFs $f_i(x_i, \mu_R^2)$ are universal functions that describe the probability to find a parton i with a given momentum fraction x_i when the hadron is probed at a momentum transfer of μ_F. The factorisation scale μ_F gives the energy scale that separates physics processes at short and long distances. The PDF absorbs all long-distance effects in the initial state that would lead to infrared and/or collinear divergences. The hard scatter cross-section $\hat{\sigma}$ is a function of the partonic center-of-mass energy squared $\hat{s} = x_i x_k s$ (with s being the pp center-of-mass energy squared), the factorisation scale and the strong coupling constant α_S. The partonic cross-section $\hat{\sigma}$ is computed in

perturbation theory. It follows that both the PDFs and $\hat{\sigma}$ have a residual dependence on the factorisation and renormalisation scales, due to uncalculated high orders.
In proton-proton collisions, the quark anti-quark annihilation can take place between valence quarks or sea quarks and anti-quarks. In Figure 1.2 the $t\bar{t}$ cross section is plotted as function of the center-of-mass energy of the collisions. Here both the measured and predicted values are showed. The center-of-mass energy $\sqrt{s}$ of the

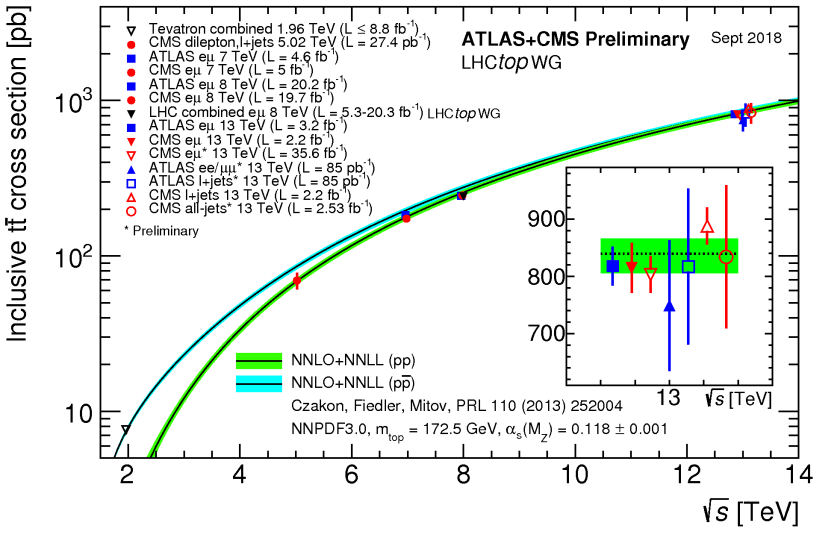


FIGURE 1.2: Summary of LHC and Tevatron measurements of the top-pair production cross-section as a function of the centre-of-mass energy compared to the NNLO QCD calculation complemented with NNLL resummation. The theory band represents uncertainties due to renormalisation and factorisation scale, parton density functions and the strong coupling.

collisions determines at which momentum fraction x the partons in the initial state protons are probed: to produce a $t\bar{t}$ pair at rest, the partonic center-of-mass energy $\hat{s}$ must be equal to at least twice the *top* quark mass. For larger $\sqrt{s}$, smaller x values and larger momentum transfers get relevant, and it becomes increasingly likely to probe a gluon inside the protons. As at the LHC at $\sqrt{s} = 13$ TeV the PDF of soft gluons is the dominant contribution, the $t\bar{t}$ production is dominated by the gluon-gluon fusion, with a relative contribution of approximately 90%.
At the next-to-leading order (NLO) the processes included in the $t\bar{t}$ calculation are a mixture of quark-gluon (qg) or antiquark-gluon ($\bar{q}g$).

Single top

Figure 1.3 shows the LO Feynman diagrams of single top production at the hadron colliders. The production processes are classified by the virtuality of the W boson exchanged in the process: the *t-channel* and *s-channel* production modes involves a virtual W boson, while the *Wt-channel* produces a top quark association to a physical W boson. As shown in Figure 1.4, the most abundant single top production process at the LHC is the *t-channel* production followed by the *Wt-channel* and the *s-channel* production.
Single top production in the *t-channel* is mediated by a space-like virtual W boson. The process can be calculated in a scheme in which the initial state bottom quark

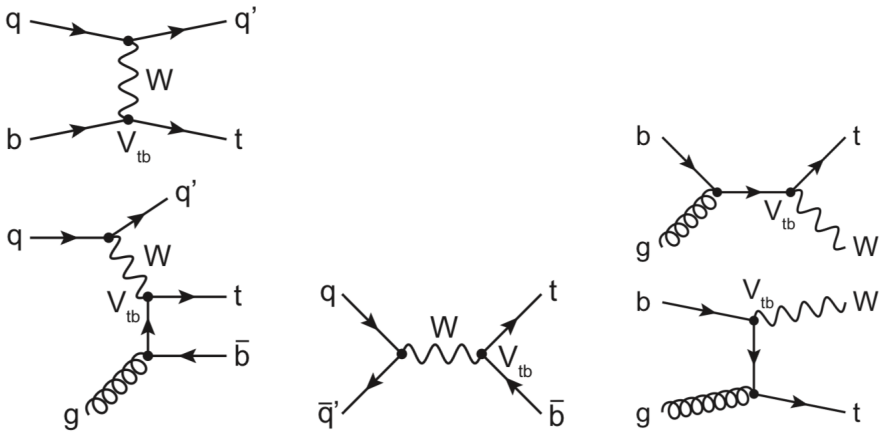


FIGURE 1.3: Feynman Diagrams of the singletop top production at the hadron colliders at the LO of EWK. t-channel production in the five-flavour scheme and four-flavour scheme (left), s-channel production (center), and associated Wt production (right) [27]

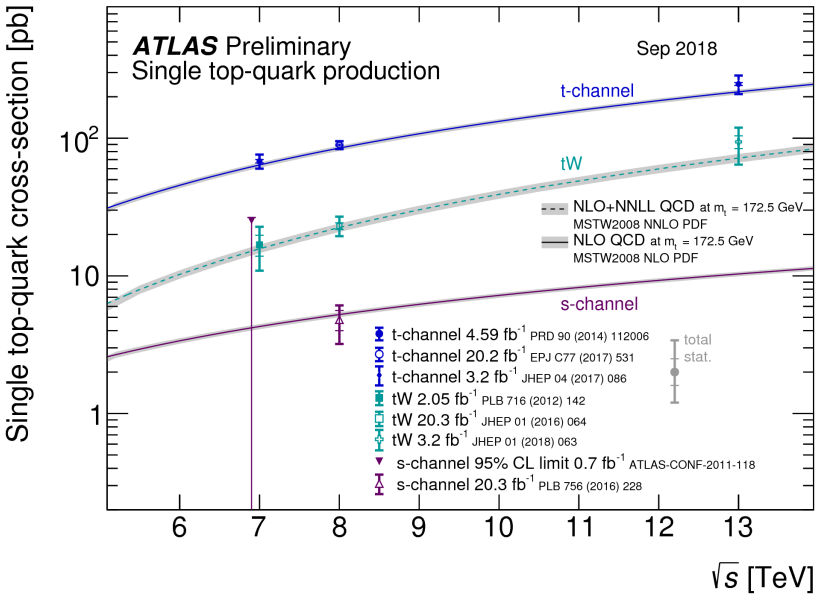


FIGURE 1.4: Summary of ATLAS measurements of the single top production cross-sections in various channels as a function of the center of mass energy compared to theoretical calculations based on NLO QCD and on NLO QCD complemented with NNLL resummation.

originates from flavour excitation in the proton (five-flavour scheme 5FS). Alternatively, the calculation of this process can be performed in the assumption that an initial-state gluon splits into a $b\bar{b}$ pair and one of the bottom quarks interacts with the virtual W boson (four-flavour scheme 4FS), where the spectator bottom quark is typically emitted at rather small angles with respect to the beam-axis.
Theoretically the *t-channel* process is known differentially to NNLO, while the *s-channel* is known to approximate NNLO.
As already explained above, in the *Wt-channel* the W boson is real. The NLO of QCD corrections to *Wt-channel* production share the same final state with $t\bar{t}$ production. This overlap is a problem if one wants to have separate predictions for these two precesses [28, 29]. For this reasons the *Wt-channel* prediction is calculated removing this overlap in two possible ways: either removing all the amplitudes before calculating the cross-section (*diagram removal* DR) or cancelling the overlap after the cross-section calculation via a subtraction term (*diagram subtraction* DS).

1.5.2 Top Quark decay

Once the top quark is produced via one of the processes showed in the previous section, it decays via the EWK interaction. As the top quark decay time is smaller than the hadronisation time, it does not form any bound state.
Being the top a up-type quark it can only decay into a W boson associated to a down-type quark. The probability for each type of down quark to be produced in the top decay is given by the relevant CKM matrix element $|V_{tq}|$, with $q = d, s, b$.

The CKM elements $|V_{td}|$ and $|V_{ts}|$ determination relies on the measurements of $B - \overline{B}$ oscillations mediated by box diagrams with top quark, or loop mediated K and B decays. Theoretical uncertainties due to hadronic effects are reduce by taking the ratios of processes: these ratios are equal in the flavour SU(3) limit to determine $|V_{td}/V_{ts}|$.
The value of the Δm_d is obtained from the measurement of the B^0_d oscillations: several measurements have been performed by LEP, Tevatron, B factories and LHC experiments: their combination yields $\Delta m_d = 0.510 \pm 0.003\ (stat) \pm 0.002\ (syst) ps^{-1}$ [30].
In the B^0_s system, Δm_s is measured by LHCb [31] to be $\Delta m_s = 17.768 \pm 0.023\ (stat) \pm 0.006\ (syst)\ ps^{-1}$.
Assuming $|V_{\mathrm{tb}}| = 1$, and using lattice QCD calculation one obtains:

$$|V_{td}| = (8.4 \pm 0.6) \times 10^{-3}, \quad |V_{ts}| = (40.0 \pm 2.7) \times 10^{-3}. \tag{1.30}$$

The uncertainties are dominated by lattice QCD, therefore the contrain on $|V_{td}/V_{ts}|$ from $\Delta m_d/\Delta m_s$ is more reliable theoretically. These provide a theoretically clean and significantly improved constrain [7]

$$|V_{td}/V_{ts}| = 0.216 \pm 0.001 \pm 0.011. \tag{1.31}$$

The determination of $|V_{\mathrm{tb}}|$ is currently performed throught the measurement of the top quark branching fractions ratio and the single top cross section: a review of these measurements is given in Chapter 2.

The SM predictions for the V_{tq} matrix elements are currently derived from CKM unitary considerations. The values obtained by a recent global CKM fit [7] yields are reported in Table 1.1, where the SM constrains, such as the three generations unitarity, are imposed. From this latter result the top quark is considered to decay only in $t \to Wb$ in all the theoretical calculation.

$\lvert V_{td}\rvert$	$\lvert V_{ts}\rvert$	$\lvert V_{tb}\rvert$
0.00886 ± 0.00032	0.0405 ± 0.001	0.99914 ± 0.00005

TABLE 1.1: V_{tq} matrix elements as estimated by global CKM fit imposing SM constrains [7].

However, if more than three generation of quarks exist, $|V_{tb}|$ can assume lower values: for this reason is important to reach the highest precision on the determination of the CKM matrix elements, which could give informations about possible BSM particles coupling to the top quark.
In many extensions of the SM, the unitarity of the 3×3 CKM matrix can be violated throught the mixing of a fourth generation of quarks with the other three generations, or by non-universality of the quark couplings in electroweak interations. The extensions of the SM proposed in [32] yield to a value of $|V_{\mathrm{tb}}|$ considerably different from one.

The decay signature of the top quark are usually categorised according to the decay mode of the W boson. The W boson can decay into a lepton and its neutrino $W \to l\nu_l$ or into a quark pair $W \to q\bar{q}'$. The top is called *leptonic* in the first case and *hadronic* in the second case.
The $t\bar{t}$ events are therefore classified using the same logic: when both the W have a leptonic decay the event is defined to be *dileptonic*; if one W decays leptonically and the second one hadronically the event is classified as *semileptonic*; finally if both the W bosons decay hadronically the event is called *fully hadronic*.

Chapter 2

Probing the top quark decay: state of the art and strategy

At the LHC it is possible to perform precision measurements in the top quark sector: at a center of mass energy of 13 TeV the cross section of top quark inclusive production is 0.8 nb which represents a big portion of the total $pp \to X \approx 100\ \mu b$ cross section. This large amount of data can be selected requiring lepton triggers: as result background processes with an higher cross section (i.e. multi-jet production) are strongly filtered out. This results in a pure sample of top quark events through which the QCD and EWK theories can be probed.
The measurements of the top quark cross section are performed looking at the different final states and in several kinematical regions: they represent a stringent test on the QCD and on the gluon density function inside the proton.
The EWK theory can be tested through top quark properties measurements such as the polarisation, the decay width, the spin correlation, the charge asymmetry and the couplings.
In this Chapter a review of the determinations of one of the couplings, $|V_{\mathrm{tb}}|$, is given. The state of the art is shown in Section 2.1 while Section 2.2 links the past measurements to the one performed in this work.

2.1 State of the art of $|V_{\mathrm{tb}}|$ measurement

Under some assumptions, the CKM matrix element $|V_{\mathrm{tb}}|$ (defined in the Section 1.4) is extracted in two ways from hadron colliders data:

- from the determination of the single top cross section (Section 2.1.1), assuming $|V_{td}|, |V_{ts}| \ll |V_{tb}|$;
- from the measurement of the top quark decay rate to bottom quarks, R_b, (Section 2.1.2), assuming $|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2 = 1$.

Before the advent of the TeV scale hadron colliders $|V_{\mathrm{tb}}|$ was extracted from a weak constraint from precision EWK data, where top quarks enter in loops. The sensitivity is best in $\Gamma(Z \to b\bar{b})$ and yields $|V_{tb}| = 0.77^{+0.18}_{-0.24}$ combining LEP, SLC, Tevatron and neutrino scattering experiments [33]. This latter measurement does not rely on any assumption on the CKM matrix unitarity.

In the extensions of the SM proposed in [32] the value of $|V_{\mathrm{tb}}|$ can be considerably different from one. One model is provided where the presence of new vector-like top singlet leads to a global rescaling of the V_{tq} matrix elements, leaving R_b unchanged. In the second method illustrated in the article, a complete new fourth generation is

added and the R_b measurement can be used as a direct constrain.

2.1.1 Extraction of $|V_{tb}|$ from single top cross section

The primary method for the extraction of $|V_{tb}|$ is through the measurement of the single top cross sections, which are directly proportional to $|V_{tb}|^2$. The estimate of the coupling at the tWb production vertex, $|f_{LV}V_{tb}|$, is obtained from the measured single top quark cross section $\sigma_{meas.}$ and its corresponding theoretical expectation $\sigma_{theo.}$

$$|f_{LV}V_{tb}| = \sqrt{\frac{\sigma_{meas.}}{\sigma_{theo.}(V_{tb}=1)}}. \tag{2.1}$$

The f_{LV} term is a form factor that parameterizes the possible presence of anomalous left-handed vector couplings. By construction, this from factor is exactly one in the SM scenario, while it can be different from one in models of new physics processes. This estimation assumes that $|V_{td}|, |V_{ts}| \ll |V_{tb}|$ and that the tWb interaction involves a left-handed weak coupling.
Both the CDF and D0 collaborations of Tevatron and ATLAS and CMS experiments at LHC performed this measurement.
The combination of Tevatron experiments [34], CDF and D0, is $|V_{tb}| = 1.02^{+0.06}_{-0.05}$, corresponding to $|V_{\rm tb}| > 0.92$ at the 95% C.L. Here f_{LV} is assumed to be one and the leading contribution to the uncertainty is coming from the normalisation of the background processes.
The combination of the LHC experiment [35] gives $|f_{LV}V_{tb}| = 1.02 \pm 0.04(meas.) \pm 0.02(theo.)$. The leading contributions to the final uncertainty are in order the theoretical modeling, the jets and detector modeling, the uncertainty on the luminosity and the background normalisation.
The summary of the ATLAS and CMS [36] results is reported in Figure 2.1. For each result, the contribution to the total uncertainty originating from the uncertainty on the theoretical prediction for the single top production cross-section is shown along with the uncertainty originating from the experimental measurement of the cross-section.
The average of Tevatron and LHC value [7] is

$$|V_{tb}| = 1.019 \pm 0.025. \tag{2.2}$$

Under the same assumptions of $|V_{td}|, |V_{ts}| \ll |V_{tb}|$, it is also possible to constrain the V_{tq} matrix elements from the measurement of the single top differential cross sections. Limits are extracted and reported in [37], where a global χ^2 fit is performed and the V_{tq} values obtained setting $|V_{td}| = |V_{ts}| = 0$ and $|V_{tb}| = 1$ are

$$|V_{td}| = 0.000^{+0.023}_{-0.000}, \quad |V_{ts}| = 0.000^{+0.041}_{-0.000}, \quad |V_{tb}| = 0.986^{+0.008}_{-0.008}. \tag{2.3}$$

The estimation of the V_{tq} values is also performed considering $|V_{td}|, |V_{ts}|, |V_{tb}|$ as free parameters and then integrating over two of them. This yields to the following result

$$|V_{td}| = 0.000^{+0.038}_{-0.000}, \quad |V_{ts}| = 0.000^{+0.069}_{-0.000}, \quad |V_{tb}| = 0.980^{+0.009}_{-0.012}. \tag{2.4}$$

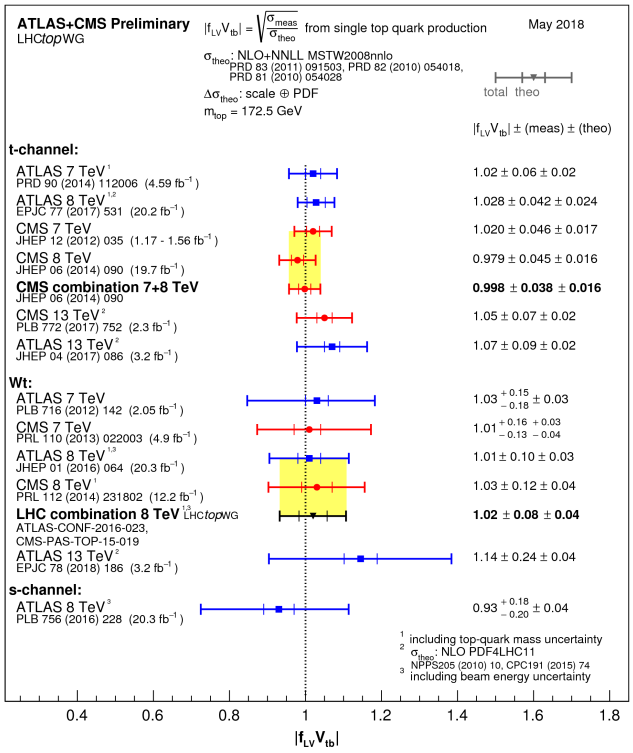


FIGURE 2.1: Summary of the ATLAS and CMS extractions of the CKM matrix element V_{tb} from single top quark measurements.

2.1.2 Measurement of the branching fractions ratio

The top quark decay rate into bottom quarks is defined as:

$$R_b := \frac{\mathcal{B}(t \to Wb)}{\mathcal{B}(t \to Wq)} = \frac{|V_{tb}|^2}{\Sigma_q |V_{tq}|^2}. \tag{2.5}$$

The module of V_{tb} can be extracted from the measurement of R_b in $t\bar{t}$ events, under the assumption of $|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2 = 1$. In the SM scenario q stands for down, strange and bottom quarks and in the assumption that only three generations of quarks exist $R_b = |V_{tb}|^2$.
R_b is usually extracted through a parametric fit on the distribution of number of b-jets generated by the top quark decay, where a b-jets is an hadronic jet initiated by a bottom quark.
R_b has been measured by the CDF collaboration both in the dilepton channel [8], finding $R_b = 0.87 \pm 0.07$, and in the final state containing one lepton [9], where $R_b = 0.94 \pm 0.09$. Assuming three generations of quarks, these translate respectively into the the following values of the CKM matrix element : $|V_{tb}| = 0.93 \pm 0.04$ and $|V_{tb}| = 0.97 \pm 0.05$. Both the CDF measurements are limited by the systematic uncertainty, of which the biggest contribution is coming from the correction to the the b-jets identification efficiency and it is respectively the ≈4-5%and ≈7-8%. In the one lepton case the background normalisation has a significant contribution with ≈ 5%.
The D0 collaboration provides the combined result of the dilepton and one lepton measurements [10], reporting $R_b = 0.90 \pm 0.04$. Also in the D0 case, the uncertainty associated to the b-jets identification efficiency represents the leading contribution to the total uncertainty.

The latest measurement of R_b has been performed by the CMS collaboration using Run 1 data at a center of mass energy of 8 TeV in the dilepton channel [11].
The result of the CMS measurement is

$$R_b = 1.014 \pm 0.032\ (tot.) = 1.014 \pm 0.003\ (stat.) \pm 0.032\ (syst.), \tag{2.6}$$

with a relative uncertainty of 3.2% and a lower limit on $|V_{tb}| > 0.975$ at 95% C.L. The biggest contribution to the uncertainty reported by CMS is coming from the identification of b-jets, which is the 2.4%.

2.2 Strategy for a model independent determination of R_b : past lessons learned and new challenges

The current status of the art described in the last section shows that the estimation of $|V_{tb}|$ is currently limited by the knowledge of the systematic effects. Given the different assumptions made in the $|V_{tb}|$ extraction from the single top cross section and the R_b measurement, it is not trivial to compare the two techinques. However, the measurement of the single top cross section and R_b can be combined to perform and extraction of $|V_{tb}|$ which is model independent. The method and the limits on the CKM matrix elements V_{tq} derived are shown in [38].
The estimation of $|V_{tb}|$ under the hypotesis of $|V_{td}|, |V_{ts}| \ll |V_{tb}|$ relies on the implicit assumpton of $R_b = 1$: it follows that a more precise determination of R_b can help to put even more stringent limits using this kind of combined fits. Moreover, from the experimental point of view, some points in favour of a new R_b measurement arise if compared to the one based on single top events:

- the available data sample is bigger as the $t\bar{t}$ cross section is $\approx$ 2-3 times the single top cross section
- the experimental signature is clearer, resulting in a smaller contribution to the final uncertainty coming from the estimation of the background events which is one of the limiting factors of all the single top measurements
- the access to the tWb vertex is in the decay vertex only; in the single top case $|V_{tb}|$ is estimated from the production vertex under the assumption that $|V_{tb}| = 1$ in the decay vertex. This assumption is implicit in the request of at least on b-jet in the selection of the events
- the leading uncertainty on the R_b estimation is coming from the calibration of the b-jets identification efficiency: in Section 4 it is shown how this uncertainty can be reduced

Moreover, as ATLAS never measured R_b before, having an independent experimental setup cross-checking the previous results can be useful, especially in the context of a future combination.
All these motivations support the model independent approach here presented in which R_b is measured using the partial LHC Run 2 dataset collected by the ATLAS detector.

2.2.1 The methodology

In order to perform the measurement of R_b with ATLAS data, addressing the main issues presented thus far, the following strategy has been developed for this work.

The dilepton final state with one electron, one muon and exactly two jets is chosen. Requiring two leptons of different flavour ensures an high discrimination power against the SM background processes, mostly related to the single top and di-boson production. As in this way the contribution from the backgrounds is minimised, also the uncertainty associated to their determination will have a smaller impact in the final result.
In a so-selected sample of $t\bar{t}$ events R_{b} can be measured counting how many times the top quark decayed into a bottom quark.
When a b-quark is generated at the LHC, it is reconstructed inside the ATLAS detector as an hadronic jet having peculiar properties given mainly by the b-hadrons lifetime and mass. Thanks to these properties a jet containing a b-hadron, hereafter referred to as b-*jet*, can be identified by dedicated algorithms that will be described in Section 4. The performance of the identification of b-jets (b-*tagging*) is quantified by the b-tagging efficiency ε_{b} which is defined as the fraction of b-jets identified by the algorithm, b-*tagged jets*, over all the b-jets.
The b-tagging algorithms can also mis-identify charm and light jet as b-jet: the *mis-tag rate*, η_q, is the fraction of the charm/light jets wrongly b-tagged among all the charm/light jets.
The presence of exactly two jets in the final state implies two $t\bar{t}$ topologies:

- the case of the LO diagram, in which both the jets are coming from the top quark decay;
- the case in which one or both the jets are not produced in the top quark decay (NLO and higher orders).

Stemming from this it is possible to define the jet-to-top assignment fractions $\alpha_i (i = 0, 1, 2)$ as the fractions of events in which 0, 1 or 2 jets coming from the top decay are selected, where $\Sigma_{i=0}^{2} \alpha_i = 1$. It follows that a crucial part of the measurement consists in the precise determination of the α_i fractions as the R_{b} has to be extracted only from the jets coming from the top decay.

The strategy presented in the following is similar to the one presented in Ref. [39] and makes use of all the ingredients described thus far.
The number of events having a k-number of b-tagged jets in two jets events $N_{events}^{ttbar}(k\, b - tags)$ can be parametrised as function of R_{b} , the b-jets identification efficiency, the charm and light mis-identification rates and the jet-to-top assignment fractions.

$$N_{events}^{ttbar}(k\, b - tags) = N_{events}^{ttbar}(2jets) \times P_k(R_b, \varepsilon_b, \eta_q, \alpha_i), \tag{2.7}$$

where the k-index runs from 0 to 2, the maximum number of jets.
The equation encodes the reasoning articulated in the following: the $N_{events}^{ttbar}(k\, b - tags)$ expression is proportional to the number of events having two jets selected and to the P_k probability, which contains the dependency from the listed parameters.
Following such considerations, one can devise the set of equations for events with 0, 1 and 2 b-tagged jets. For the sake of simplicity, only the $N_{events}^{ttbar}(2\, b - tags)$ is shown,

the 0 and 1 b-tags case are reported in the Appendix A

$$
\begin{aligned}
&N_{events}(2\ b-tags) \\
&\quad = N_{events}(2jets) \times P_2(R_b, \varepsilon_b, \eta_q, \alpha_i) && (2.8)\\
&\quad = N_{events}(2jets) \times \{ && (2.9)\\
&\quad \alpha_2 \cdot [R_b^2\varepsilon_b^2 + 2R_b(1-R_b)\varepsilon_b\eta_q + (1-R_b)^2\eta_q^2] + && (2.10)\\
&\quad \alpha_1 \cdot [R_b\varepsilon_b\eta_q + (1-R_b)\eta_q^2] + && (2.11)\\
&\quad \alpha_0 \cdot \eta_q^2 && (2.12)\\
&\quad \} && (2.13)
\end{aligned}
$$

The parameter R_b is extracted by comparing the observed number of b-tagged jets $\hat{N}_{ev}(k\ b-tags)$ to the prediction model presented thus far. These equations are valid assuming that the amount of b-jets generated by other sources than the top quark decays is negligible: this assumption is further investigated in Section 5.4, where a detailed description of the fit is given.

2.2.2 The main challenges

The two main challenges in this method arise from the b-tagging efficiency calibration and the estimation of the jet-to-top assignment fractions.

Calibration of b-tagging efficiency

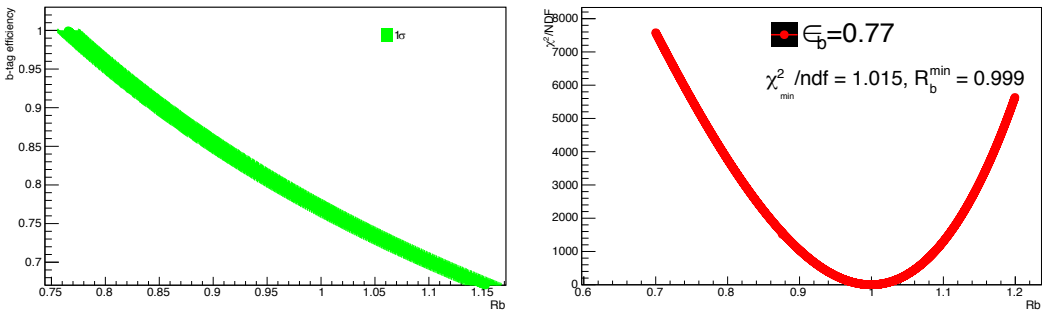
The uncertainty on the calibration of the b-tagging efficiency is the leading contribution from the documented Tevatron and LHC measurements and it is expected to be the same for the estimation presented in this work.
A calibration of the b-tagging efficiency using data events is necessary to correct the performance of the b-tagging algorithms, as their development is based on simulated informations: the aim of the calibration is to take into account any mis-modeling of the Monte Carlo (MC) simulation.
In the assumption that top quarks decay only into bottom quarks ($|V_{tb}|$ =1), $t\bar{t}$ events provide a sample of well determined b-flavour. Nowadays they constitute the standard candle used to calibrate the b-tagging efficiency. However, because of the underlying assumption that R_b =1, the results of a $t\bar{t}$ - based b-tagging calibration [40] cannot be used in this analysis.
To avoid this problem two alternatives are considered:

- Simultaneously extract R_b and ε_b
- Calibrate ε_b with an alternative method

The first option is not suitable: a counting analysis like the present is not able to disentangle the effect of a decreased R_b from that of a smaller ε_b. The two are fully correlated in the b-tagging multiplicities that are fitted, making it impossible to isolate R_b. This is mathematically demonstrated in Equation 2.8, where the terms significantly different from zero have R_b and ε_b appearing in a product. Under the assumption of R_b close to one, as measured by previous experiments, all 1-R_b terms are then close to zero. Moreover, the value of the mis-tage rate η_q is expected to be of the order of 10^{-2-3}.
To illustrate this point, 10k toy experiments are generated. Assuming the model in Eq. 2.7, a χ^2 function is built and minimised

$$\chi^2 = \sum_{i=0}^{10k} \frac{(\widehat{N}_{ev,i}(k\ b-tags)(R_b=1,\varepsilon_b=0.77) - \widehat{N}_{ev}(k\ b-tags)(R_b,\varepsilon_b))^2}{\widehat{N}_{ev}(k\ b-tags)(R_b,\varepsilon_b)}, \quad (2.14)$$

where $(\widehat{N}_{ev,i}(k\ b-tags)(R_b=1,\varepsilon_b=0.77)$ is extracted from a Poisson distribution centered on the number of events observed in the simulation. The value of ε_b = 0.77 is chosen because it is the nominal value of the b-tagging efficiency used in the $t\bar{t}$ simulation. It is found that if the R_b and ε_b are let to float the minimum of χ^2/NDF in the interval of 1σ are distributed in a band of possible solutions, showed in Figure 2.2a. The 1σ band is defined by half of the minimum of χ^2/NDF.
The same exercise is repeated but fixing the b-tagging efficiency to its nominal value $\varepsilon_b = 0.77$. Figure 2.2b shows the χ^2/NDF as function of R_b . The minimum found is a unique value $R_b = 0.999$ which agrees with what has been set in the simulation, in which $R_b = 1$. This latter result further illustrates that the fit needs an external input for the b-tagging efficiency, which must be estimated from an independent sample of events.
Stemming from the above situation, for the R_b measurement it is necessary to perform an external calibration of the b-tagging efficiency. The uncertainty associated to



(A) (B)

FIGURE 2.2: Figure (A) shows the b-tagging efficiency as function of R_b found by the minimum of χ^2/NDF by the 10k toy experiments. Here both ε_b and $R_b$are free parameters in the interval 0.7-1.2. The green band delimits the 1σ interval. Figure (B) shows χ^2/NDF as function of R_b when ε_b is fixed to 0.77

this calibration needs to be as small as possible: the uncertainty on R_b as a function of the uncertainty on ε_b is reported in Figure 2.3. This curve is derived generating

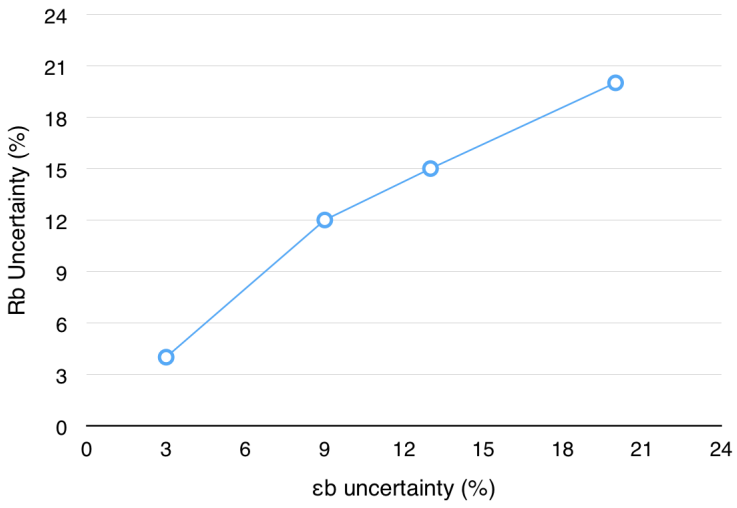


FIGURE 2.3: Percentage uncertainty on R_b as function of the percentage uncertainty on ε_b derived using toy experiments.

toy experiments with an up and down variation of the ε_b of 3%, 9%, 14% and 20% respectively. It shows how the uncertainty on ε_b directly translates to an uncertainty on R_b of the same size.

An alternative calibration method is identified, which was developed in the past and relies on multi-jets events. In this method, events in which hadrons decay semileptonically into muons are used. Exploiting the kinematical correlation between the muon momentum and the mass of its hadron parent, a variable discriminating in the hadron flavour is built. This variable is called p_T^{rel} and it is defined as the muon

transverse momentum in the jet-plus-muon system axis. The fraction of b-jets identified by ATLAS b-tagging algorithms is extracted fitting on $p_{\mathrm{T}}^{\mathrm{rel}}$, giving an estimate on data of the b-tagging efficiency. As reported in the Ref. [41], where the method is described, the uncertainty achieved by the $p_{\mathrm{T}}^{\mathrm{rel}}$ calibration during Run 1 was around 5% of which the 4% coming from the extrapolation of the correction factors from the semileptonic to the inclusive sample of b-hadrons decay.
Section 4.2.5 shows the results of the calibration using $p_{\mathrm{T}}^{\mathrm{rel}}$ on the Run 2 dataset.

Estimation of the jet-to-top assignment fractions

The second main challenge of this work is finding a way to estimate the fraction of $t\bar{t}$ events in which a jet can be assigned to a top decay: the jet-to-top assignment fractions α_i.
The first solution probed makes use of the truth record of the simulation in which b-quarks associated to the top quark decay at the parton level are stored. Here, a spatial matching between these b-quarks and the jets is performed: the final goal is to find the jet initiated by the top quark decay products and thus classify jets into those coming from top-quark and those coming from other sources, such as ISR and FSR activity. In principle, this method should provide the extact origin of the jets. However, the jet matching turnes out to be strongly dependent from the value of the $\Delta R(jet, bottom\ quark)$ [1] cone used.
Table 2.1 reports the fractions of jets from top as function of the matching cone value: the fractions range from 69% of the $\Delta R < 0.1$ case to the 82% of the $\Delta R < 0.6$ case. The reason for the trend observed in the fraction of jets associated to the top decay is

ΔR	Jets from top	Jets not from top
0.1	69%	31%
0.2	77%	23%
0.4	81%	19%
0.6	82%	18%

TABLE 2.1: Fraction of jet associated or not to the bottom quark from the top decay in the $t\bar{t}$ simulation for different values of the matching come going from 0.1 to 0.6.

believed to be the analysis acceptance: jets created in a top quark decay could fail the reconstruction requirements or the selection criteria of this analysis. As a result, the fractions of jets from top increases with the value of the matching cone, without ever reaching 100%, which is the expected value for a simulation having $R_b = 1$. From this result one can conclude that the spatial matching between the truth information of the top quark decay and the reconstructed jet is not well defined in the ATLAS simulation. Therefore, none of the matching cone values can be used to estimate the jets from top fractions as there is no way to determine which of them is giving the right estimate.

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

To overcome this issue an estimation of the jet-to-top assignment fractions is performed using the $t\bar{t}$ simulation exploiting the fact that none of the simulated light jets are coming from the top decay. This method implies the definition of control regions enhanced in $t\bar{t}$ events having at least one jet not from the top decay, which are used to correct the simulated fractions using data events. The procedure is extensively explained in Chapter 5, where the possible biases introduced by the assumptions of this method are investigated (see Section 5.3).

Chapter 3

The Large Hadron Collider and the ATLAS Detector

The data analysed in this thesis have been collected in the 2016 and 2017 by the *A Toroidal LHC ApparatuS* (ATLAS) experiment [5] . The ATLAS detector is placed at one of the interaction points of the proton beam lines of the Large Hadron Collider [3], located in Geneva at European Organisation for Nuclear Research (CERN). In this chapter a description of the collider and the ATLAS experiment are reported.

3.1 The Large Hadron Collider

The LHC is a proton-proton (or Pb-Pb) collider designed for several purposes, mainly to test the validity of the SM and to look for its possible extension.
The LHC is a circular accelerator built in the same tunnel which hosted the LEP, it has a 27 kilometers circumference at about 100 meters beneath the ground.
The Large Hadron Collider is designed to work at the center of mass energy of $\sqrt{s}$ = 14 TeV, 7 TeV for each of the two beams. The accelerator ran at $\sqrt{s}$ = 7 TeV in the 2010 and 2011 and reached $\sqrt{s}$ = 8 TeV in 2012. After a two years stop, in which the machine has been upgraded, it restarted in 2015 delivering to the experiment protons collision at $\sqrt{s}$ = 13 TeV.
The LHC is composed by of a chain of accelerators of which the 27 Km ring is the last step. Figure 3.1 shows the whole accelerator system connected to the LHC.
The protons from hydrogen atoms, stripped of their electrons, are accelerated in a linear accelerator (Linac2) up to 50 MeV. They are then injected in a circular accelerator, the Proton Synchrotron Booster (PSB). Here the beam reaches an energy of about 1.4 GeV and it is injected in a larger accelerator, the Proton Synchrotron (PS), where the beam energy rises to 26 GeV. The protons before being injected in the main ring are accelerated in another synchrotron, the Super Proton Synchrotron (SPS) . In the SPS each beam reaches an energy of 450 GeV. The protons in the SPS are also accumulated in bunches to have an higher interaction probability during the collisions. By project the maximum number of protons for each bunch and the number of bunches for each beam are 115 billion and 2808 respectively. At full operation the interactions between the two beams take place at discrete intervals never shorter than 25 ns, with a maximum collision rate of 40 MHz.
The whole LHC accelerator is composed of several superconducting magnets: 1232 dipoles and 392 quadrupoles. The first ones are part of the system that bends the charged particles path during the not linear parts of the collider, while the quadrupoles are mainly used to focus the beams around the collision areas. The nominal field strength is of 8.33 T and it is generated by an electric current 11.700 A for each dipole.

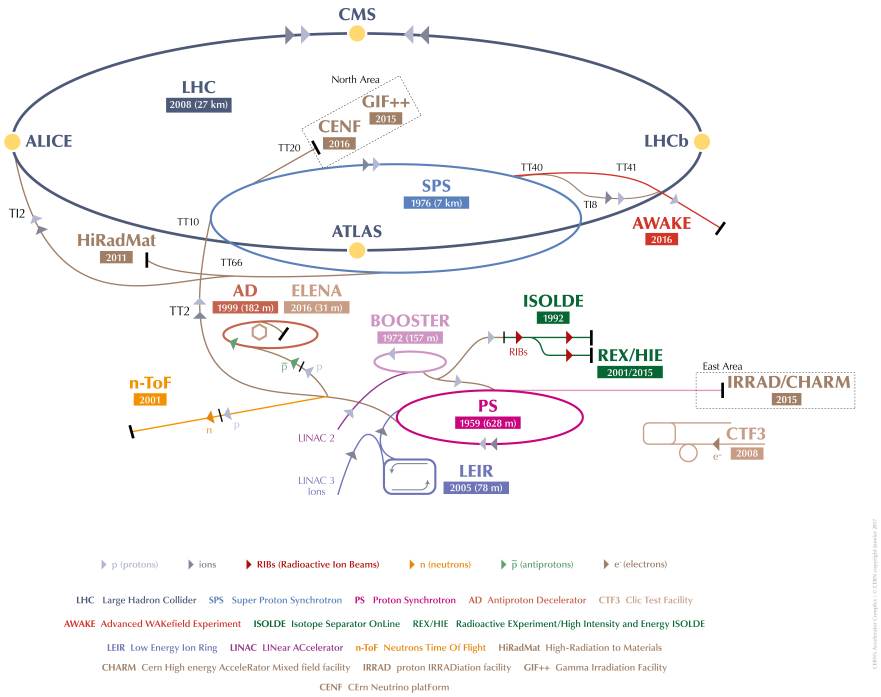


FIGURE 3.1: The LHC accelerator complex shown in a cartoon containing the chain of the smaller machines together with the LHC ring. [42]

3.1.1 Luminosity and Pileup

The concept of *luminosity* is used to quantify the performance of a particle collider and the amount of proton-proton collisions.
The instantaneous luminosity can be expressed as a function of the rate of pp interactions and in terms of the beam parameters as follows:

$$\mathscr{L} = \frac{N_b^2 n_b \nu}{4\pi\sigma_x\sigma_y} F, \tag{3.1}$$

where:

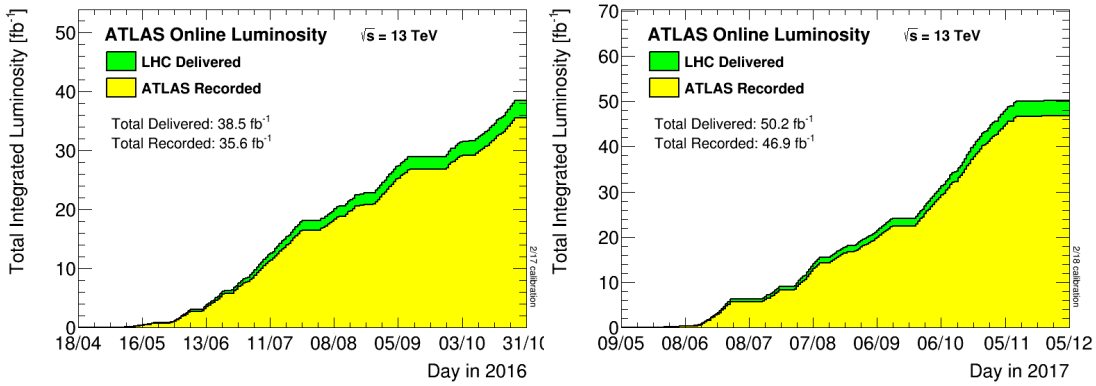
- N_b is the number of protons per bunch
- n_b is the number of bunch per beam
- ν is the revolution frequency
- σ_x and σ_y stand for the horizontal and vertical Gaussian widths of the colliding beams
- F is the geometric luminosity reduction factor, a correction factor accounting for the crossing angle between beams

The different beam parameters of the LHC for the 2016 and 2017 data taking periods are presented in Table 3.1.

Parameter	2016	2017
Beam Energy [TeV]	6.5	6.5
Bunch spacing [ns]	25	25
Max number of bunches (n_b)	2220	2556
Protons per bunch (N_b) [10^{11}]	1.15	1.15
Peak luminosity [$10^{33} cm^{-2} s^{-1}$]	13.8	20.9
Integrated luminosity [fb^{-1}]	38.5	50.2

TABLE 3.1: Values of the beam parameters during Run 2 in 2016 and 2017 [43, 44].

The *integrated luminosity* is obtained by integrating the instantaneous luminosity on the live time of the experiment $L = \int \mathscr{L} dt$.
Figures 3.2a and 3.2b show the integrated luminosity delivered by LHC and recorded by the ATLAS experiment for 2016 and 2017, respectively.



(A) (B)

FIGURE 3.2: Cumulative luminosity as function of time delivered to (green) and recorded by ATLAS (yellow) during stable beams for pp collisions at 13 TeV in 2016 (A) and 2017 (B) [45]

The luminosity recorded by ATLAS is smaller than the delivered as effect of the detector inefficiencies.

The expected number of events N_{exp} for a process having a production cross-section σ is given in terms of the integrated luminosity as

$$N_{exp} = \sigma \int \mathscr{L} dt. \tag{3.2}$$

Given the high density of the beam bunches and the high frequency of collisions, many hadronic interactions may occur simultaneously to the the hard scatter process. This phenomena is called *pileup*: it consists of an higher number of energy deposits in the detector that can be due to:

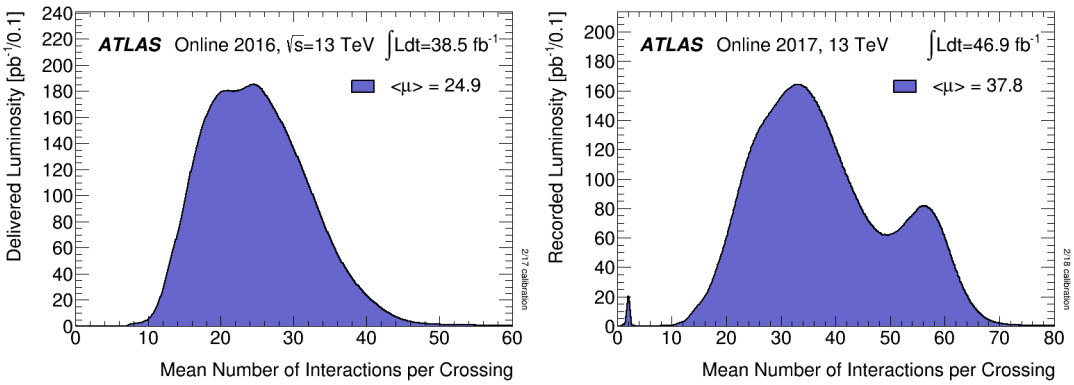
- additional proton-proton interactions in the same bunch crossing, the so-called *in-time pileup*. In this case several interaction points are produced, resulting in an higher number of particles emerging from the collision.
- detector signals reconstructed in an event as an effect of interactions occurred in the previous bunch crossing. This pileup effect is related to the integration

time of some detectors and it is called *out-of-time pileup*. The out-of-time pileup usually affects the signal in the calorimeters.

Increasing the number of protons per bunch crossing or the number of bunches results in higher luminosity but also raises the level of pileup. At higher number of protons corresponds an higher number of interactions in each bunch crossing, resulting in an higher in-time pileup activity. Increasing the number of bunches reduces their spacing, causing an overlap between the different bunch crossings and thus an higher out-of-time pileup activity.
The distribution of the mean number of interactions per crossing is sensitive to all these effects. It corresponds to the mean of a Poisson distribution of the number of interactions per crossing calculated for each bunch. It is calculated as

$$< \mu >:= \frac{L_{bunch} \times \sigma_{inel}}{\nu}, \tag{3.3}$$

where L_{bunch} is the measured instantaneous luminosity per number of colliding bunch pairs, σ_{inel} is the inelastic cross section for pp interactions which is 80 mb for at 13 TeV and ν is the LHC revolution frequency. In 2016, the average number of interactions per bunch is measured to be $< \mu >= 24.9$ which increased to $< \mu >= 37.8$ in the 2017 data taking period. Figures 3.3a and 3.3b show the distribution of $< \mu >$ for the 2016 and 2017 data respectively. The increase of $< \mu >$ is consequence of the increase



(A) (B)

FIGURE 3.3: Mean number of interaction per bunch crossing for 2016 (A) and 2017 (B) collision data. The integrated luminosity reported is relative to all data recorded by ATLAS [45].

in the instantaneous luminosity which reached the peak of $L = 13.8 \times 10^{33} cm^{-2} s^{-1}$ in 2016 and $L = 20.9 \times 10^{33} cm^{-2} s^{-1}$ in 2017.

3.2 The ATLAS Detector

The ATLAS detector, shown in Figure 3.4, is a general purpose detector located at Interaction Point 1 on the LHC ring. It is 25 meters high, its length is 44 meters and its weight is 7000 tons.
The ATLAS detector consists of a series of concentric cylindric sub-detectors, placed around the interaction point where the proton beams of the LHC collide. Starting from the beam line going outwards there are three major components: the Inner Detector (ID), the Calorimeters and the Muon spectrometer (MS).

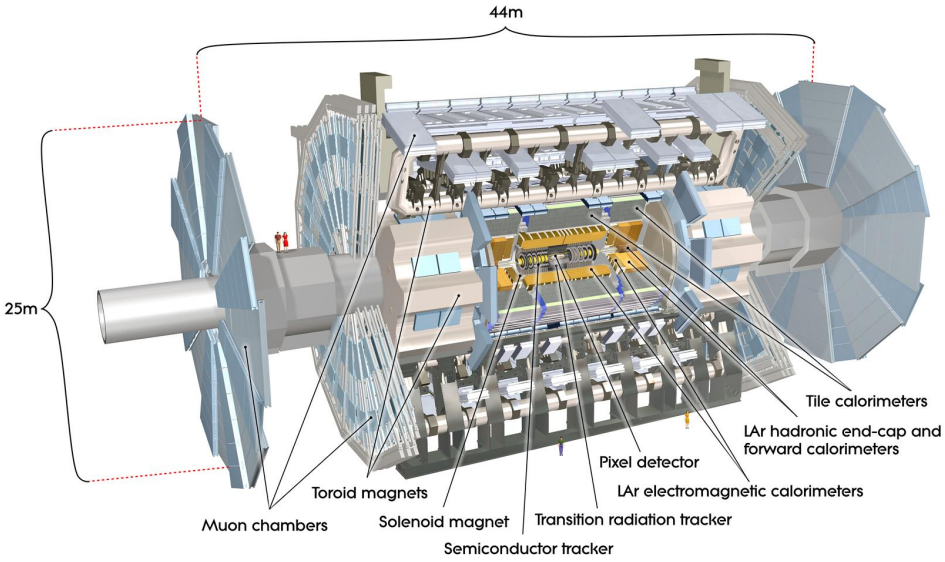


FIGURE 3.4: A computer generate image of the ATLAS detector from the longitudinal view. It shows the ID, the Calorimeters, the MS and the magnetic system [5].

The ID consists of a system of tracking detectors enclosed by a thin solenoid magnet, providing an axial magnetic field of 2T. As a result, inside the ID the trajectory of all the charged particles is bent, allowing the measurement of their momenta. The Calorimetric system measures the energy of the neutral and charged particles escaping the ID. It is divided into an electromagnetic and hadronic calorimeter, measuring the energy of electrons and photons, and hadronic particles, respectively. The MS is implied for the identification of muons and the measurement of their momentum. It operates in presence of a magnetic field generated by air-core toroids magnets.

A two-level trigger system is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to reduce the accepted rate to at most 100kHz. This is followed by a software-based trigger that reduces the accepted event rate to 1kHz on average depending on the data-taking conditions.

ATLAS uses a right-handed coordinate system with its origin at the nominal IP in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. The angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

3.2.1 The Inner Detector

The ID [46], shown in Figure 3.5, is the nearest detector to the collision point. It provides precise tracking of charged particles having $p_T > 0.1GeV$ and $|\eta| < 2.5$. The

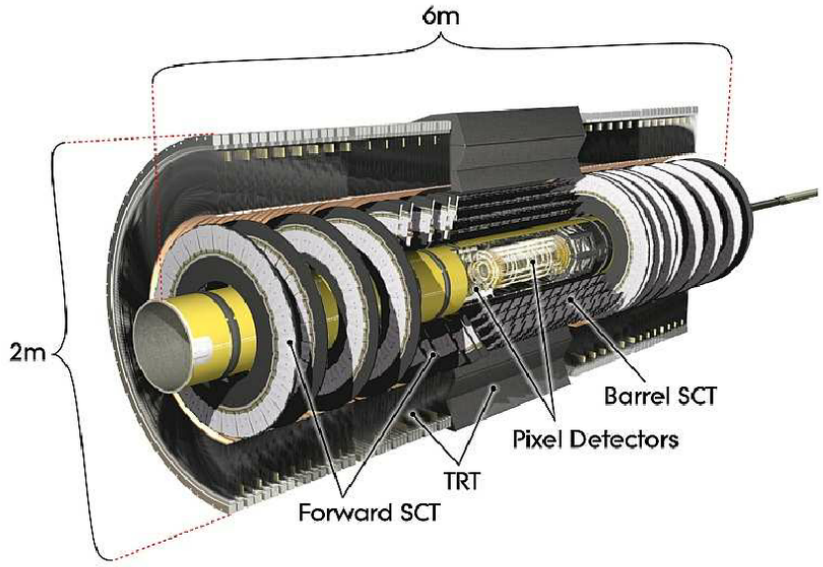


FIGURE 3.5: Inner Detector longitudinal view showing the barrel modules and the forward disks for the Pixel detectors, the SCT and the TRT [46].

ID is immersed in a uniform 2T axial magnetic field generated by the central superconducting solenoid held at 4.5 K by liquid helium in the region of $|\eta| < 1.6$. The magnetic field bends the charged particles, allowing their momenta to be accurately measured using the curvature of their tracks. Each sub-detector is split into cylindrical concentric barrel modules covering the central region and disk-shaped end-cap modules covering the forward regions.
The ID comprises three complementary sub-detectors: the Pixel Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT).

The Pixel Detector and the Insertable B-Layer

The Pixel Detectors [47] consists of 1744 silicon pixel modules arranged in three concentric barrel layers and two endcaps of three disks each. The size of each pixel is 50×400 μm^2, allowing to achieve a spatial resolutions of 10 μm in the $R \times \phi$ plane and 115 μm in z for the barrel and 10 μm in the $R \times \phi$ plane and 115 μm in R for the endcap.
A new component has been added to this configuration during the shutdown between Run 1 and Run 2, the Insertable B-Layer (IBL) [48]. The IBL is currently the innermost layer whose aim is to increase the performances of the tracking at the new highest instantaneous luminosity. Moreover, the additional informations provided by the IBL allow to improve the identification of jets initiated by b-quarks, giving more informations for the reconstruction of the secondary vertex of decay of b-hadrons. Its $R-\phi$ resolution is $< 10\mu m$.

The Semiconductor Tracker

The barrel SCT [49] consists of 4088 modules of silicon-strip detectors arranged in four concentric barrels and two endcaps of nine disks each. Each silicon detector is

$6.36 \times 6.40\ cm^2$ with 768 readout strips each with 80 μm pitch. Each module consists of four detectors. On each side of the module, two detectors are wire-bonded together to form 12.8 cm long strips.
The forward modules are very similar in construction but use tapered strips, with one set aligned radially.
The detector contains 61 m^2 of silicon detectors, with 6.2 million readout channels. The spatial resolution is 16 μm in $R - \phi$ and 580 μm in z. Tracks can be distinguished if separated by more than 200 μm.

The Transition Radiation Tracker

The TRT [50] consists of proportional drift tubes (straws). These straw detectors can operate at very high rates, needed by virtue of their small diameter and the isolation of the sense wires within individual gas envelopes. Each straw is 4 mm in diameter, giving a fast response and good mechanical properties for a maximum straw length of 150 cm.
The barrel contains about 50000 straws, each divided in two at the centre in order to reduce the occupancy and read out at each end. The end-caps contain 320000 radial straws, with the readout at the outer radius. The total number of electronic channels is 420000. Each channel provides a drift-time measurement, giving a spatial resolution of 170 μm per straw, and two independent thresholds. The two independent thresholds allow the detector to discriminate between tracking hits, which pass the lower threshold, and transition-radiation hits, which pass the higher.

3.2.2 The Calorimetric system

All the particles exiting the ID, except muons and neutrinos, are stopped in the ATLAS calorimeters which thus measure their energy. The calorimeters implied by ATLAS are sampling calorimeters, detectors in which particles cross alternatively active and inert materials.

The active material can be a plastic scintillator or liquid argon: particles interact with the active part of calorimeters through various mechanisms. In the case of a plastic scintillator, particles interacting electromagnetically can excite valence electrons, whose de-excitation produces photons: the number of photons produced is proportional to the deposited energy by the initial particle. When passing through the liquid argon, charged particles can ionize the material: the electrons and positive ions drift towards the electrodes that measure the deposited charge. The inert material is usually composed of heavy absorber material that interact with charged and neutral particles but does not measure the energy deposits.

Electrons and photons lose their energy via bremsstrahlung and e^+e^- pair production, respectively. Their radiation length X_0 is equivalent to $7/9$ of the mean free path of a photon or the mean distance over which the electron loses all but $1/e$ of its energy.
Hadrons lose their energy through inelastic hadronic collisions in matter, causing showering of particles. The mean free path of a hadron and the characteristic length of the hadronic showers is given by the nuclear interaction length λ. As a result, the calorimeters must be adequately large in order to fully capture interactions, to precisely measure energies, and to avoid losing energy into the muon spectrometer. Muons deposit a small amount of energy in the calorimeters while neutrinos do not

interact, escaping the detection.

The calorimeter resolution is expressed as follows

$$\frac{\sigma_E}{E} = \frac{N}{E} \oplus \frac{S}{\sqrt{E}} \oplus C, \tag{3.4}$$

where N stands for the measurement of the noise due to background and electronics which is dominant at low energies, S parameterizes the stochastic uncertainty caused by the random sampling nature, and C is a constant term that reflects the non-uniformities in the detector and is dominant at higher energies. These terms are added in quadrature to obtain the fractional resolution.

The ATLAS calorimeter system, shown in Figure 3.6, covers the pseudorapidity range up to $|\eta| < 4.9$. The calorimeters are segmented into towers in both η and ϕ, pointing towards the center of the detector. The active material used is liquid ar-

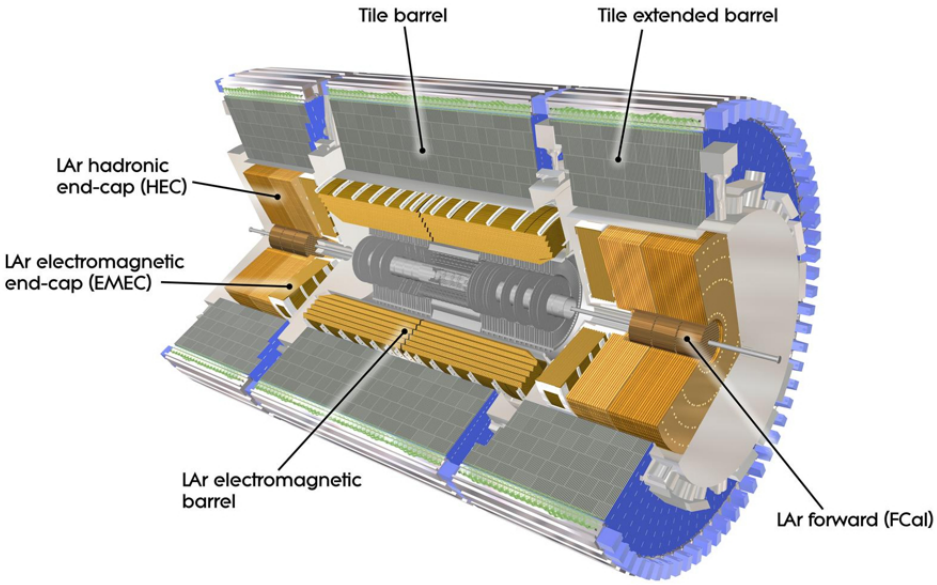


FIGURE 3.6: ATLAS calorimeter system: the electromagnetic calorimeter is shown in yellow while the hadronic calorimeter is shown in grey. Both the barrel and forward modules are visible [5].

gon (LAr) for most of the calorimeters while the absorber material depends on the region in which the detector is located. The choice of LAr is connected to its performance and in particular to the linearity of the signal, the fast response and the radiation hardness.
The calorimeter system is divide into an electromagnetic part, which measures the energy of electrons and photons, and the hadronic part, which measures the energy of strongly interacting particle.

The EM calorimeters [51, 52] are divided in barrel (EMB) and endcap (EMEC). The EM calorimeter is segmented into cells of $\Delta\eta \times \Delta\phi = 0.003 \times 0.025$ with three layers in depth. The first layer has the best granularity in η while the second provides the best position in ϕ. This fine segmentation is important to distinguish single photons

from $\pi^0 \rightarrow \gamma\gamma$ decays. The EMB uses LAr as active material and lead as absorber and all the system covers a pseudorapidity of $|\eta| < 1.475$. The number of radiation lengths in the EMB is 24 X_0. The EMEC calorimeters are made with the same material of the EMB but the number of radiation lengths in the endcap is larger, 26 X_0. In the endcaps the EM is composed of two concentric wheels covering the range $1.375 < |\eta| < 3.2$.
The response resolution of the stochastic and constant terms were measured to be

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E[GeV]}} \oplus 1\%. \tag{3.5}$$

The hadronic calorimeter system [53] is located after the EM calorimeters. As the EM it is divided in a barrel (TileCal) and two endcaps (HEC). TileCal is composed by steel, as absorber, and plastic scintillators as sampling material. The number of interaction lengths in the TileCal is 8. The HEC is composed of LAr and copper and it covers up to $|\eta| < 3.2$. It is divided in two wheels for each endcap. In the endcap regions the total number of interactions lengths (including the EMEC) is 12 λ. To cover pseudorapidity larger than $|\eta| > 3.2$ another calorimetric is used, the forward calorimeter (FCAL). FCAL is composed by one EM and two hadronic calorimeters and covers between $3.1 < |\eta| < 4.9$. The number of interaction lengths is about 10 λ. In test beams [54] the response resolution to isolated charged pions of the combined HEC and TileCal of the stochastic and constant terms is

$$\frac{\sigma_E}{E} = \frac{53\%}{\sqrt{E[GeV]}} \oplus 3\%. \tag{3.6}$$

3.2.3 The Muon spectrometer

The MS [55, 56], shown in Figure 3.7, is the outermost detector of the ATLAS apparatus.

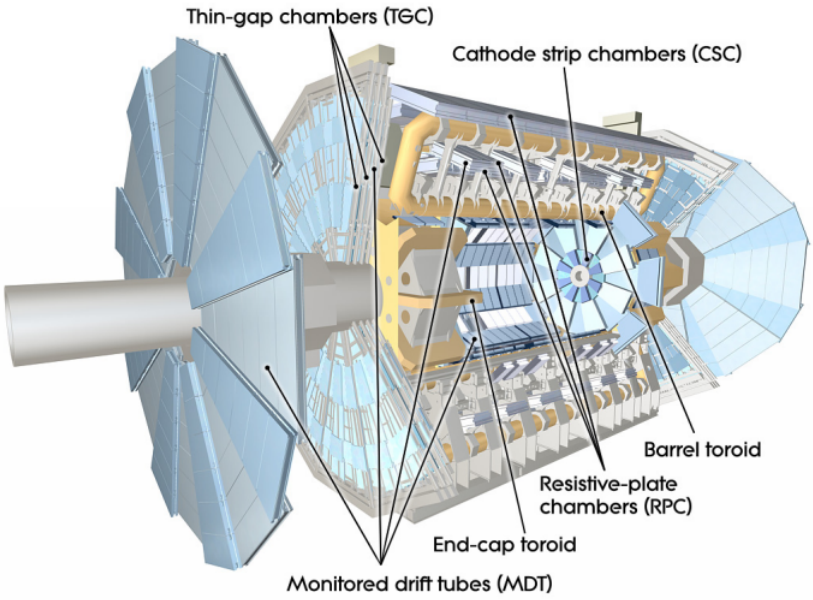


FIGURE 3.7: ATLAS Muon Spectrometer longitudinal view showing the MDTs, CSCs, RPCs and TGCs detectors in light blue and grey. The yellow parts are the toroidal magnets [55].

It is composed of four different technologies, two connected to the trigger and two connected to the precise tracking. The MS is built to reconstruct the muon trajectories and measure the muon momentum independently from the ID and to provide muon trigger signals. The spectrometer operates under a magnetic field orthogonal to the muon trajectory. The four sub-detector technologies composing the MS are described below.

Muon Drift Chamber

The Muon Drift Chamber (MDT) system is composed by 1088 chambers: each chamber is made by two multi-layers of three or four (only in the innermost chambers) layers of tubes with 3 cm diameter and 400 μm thick aluminum walls. The gas mixture is Ar+CO_2+H_2O operating at 3 bar pressure and at 3040 V. The MDT are located in both the barrel and endcap regions. In the barrel, $|\eta| < 1.3$ the chamber are divided in 16 sectors along ϕ. In each sector there are large and small chambers. This allows a full coverage and an overlap between chambers that ensure a robust muon momentum measurement. In the endcap the MDTs cover the region $1.3 < |\eta| < 2.4$. The average single tube resolution is around $100\mu m$, corresponding to a spatial resolution of the chambers of $\approx 40\ \mu m$ as required for 10% momentum resolution of 1 TeV muons [57].

Resistive Plate Chamber

The Resistive Plate Chambers (RPC) [58] are used in the barrel for $|\eta| < 1.05$. 544 chambers are located in three concentric layers connected to the MDT. Every chamber has 2-layers of gas gap filled with a gas mixture of $C_2H_2F_4$ + $isoC_4H_{10}$ + SF_6, where the last one is added to limit the charge avalanches in the chamber. The chambers are made with bakelite plates of 2 mm and readout strip with pitches of about 3 cm. The RPCs work at 9.8 kV and have a time resolution of 1.5 ns.

Cathode Strip Chamber

The Cathode Strip Chambers (CSC) [59] are multiwire proportional chambers located between $2.0 < |\eta| < 2.7$. They are designed to provide high granularity in an apparatus region near to the beam pipe. The CSCs are divided in 16 sectors for each of the two wheels, 8 small and 8 large. The chambers, composed by four layers, are in overlap to ensure no loss of information. The chatode strips are mounted in the $\eta - \phi$ plane such that the muon track position will be measured by the interpolation of the induced charges in different strips of the layers. The gas mixture is Ar + CO_2 and the typical spatial resolution is 40 μm in the magnet field direction and 5 mm in the azimuthal direction while the time resolution is about 7 ns.

Thin Gap Chamber

The Thin Gap Chambers (TGC) [60] are multiwire proportional chambers dedicated to the trigger system on the endcap part of the ATLAS detector. They cover the forward region in the pseudorapidity range $1.05 < |\eta| < 2.7$. The TGCs, like the RPCs, provide also a measurement of the muon track coordinate orthogonal to the one provided by the precision tracking chambers. The nominal spatial resolution for the TGC it is 3.7 mm in the $R - \phi$ plane. The TGC system is divided in 4 layers, one innermost (TGI) and three in the endcap (TGC1, TGC2 and TGC3). The TGCI covers

$1.05 < |\eta| < 1.92$ while the others TGC layers cover up to $|\eta| = 2.7$. TGC1 is composed by three chambers while TGC2 and TGC3 are composed by two chambers. The gas mixture used for these chambers is $CO_2 + nC_5H_{12}$ and they work at 2.9 kV. The time resolution is about 4 ns.

3.2.4 The Trigger system

In ATLAS each event digitised needs a storage size of about 1.7 MB, this translates into an enormous amount of information to be recorded during the collision period. Only a fraction of these informations can be recorded, due to the limited data storage capacity and rates. The ATLAS trigger system [61] performs a run-time event selection, resulting in a reduction of the amount of recorded informations.
The trigger system consists of a hardware Level-1 (L1) and a single software-based high-level trigger (HLT). This new two-stage system will reduce the event rate from the bunch-crossing rate of 40 MHz to 100 kHz at L1 and to an average recording rate of 1 kHz at the HLT (Figure 3.8).

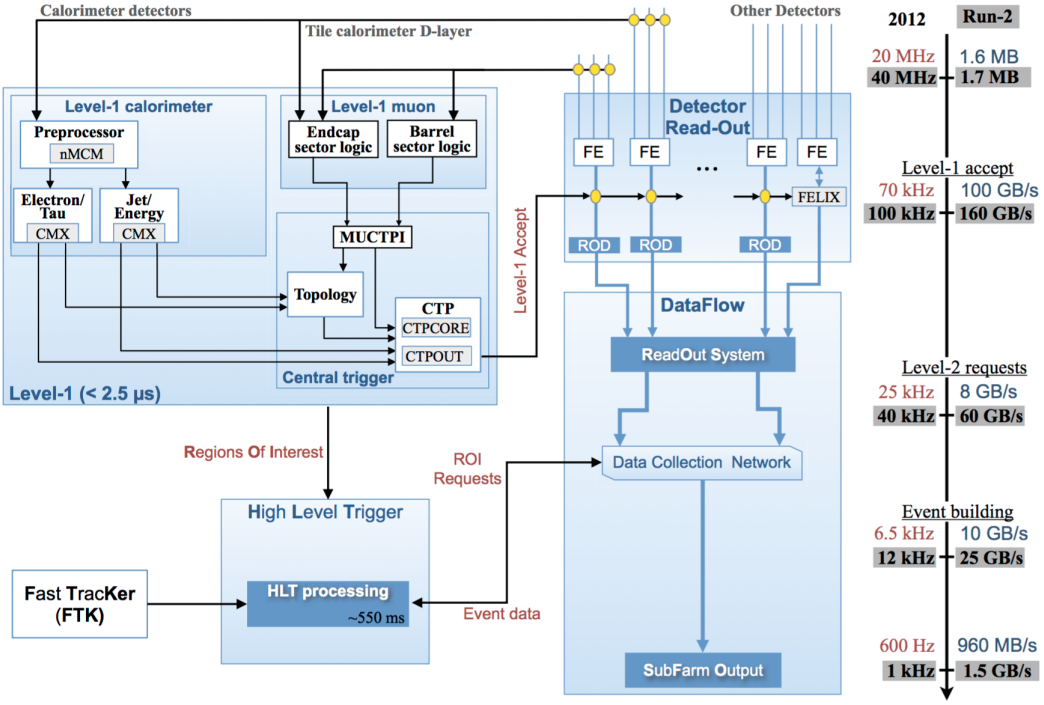


FIGURE 3.8: Scheme showing the ATLAS Trigger system together with the operating rate of each stage [61].

The L1 trigger system performs the initial event selection and accepts events at a 100 kHz rate. It is optimised to provide a fast decision. It searches for high energy leptons, photons, and jets using a combination of information from the calorimeter and MS. Electrons and photons are triggered on energy deposits in the EM calorimeter. In the hadronic calorimeter, jet candidates are constructed at L1 from coarse calorimeter towers made of trigger-elements using a sliding window algorithm. A trigger-element is determined by the sum of cells in a $0.2 \times 0.2(\eta - \phi)$ region, and the sliding window examines the total E_T against a trigger threshold value in a 4×4 region of trigger-elements. Muon triggers are based on a coincidence of hits among

several layers of the trigger chambers.
During Run-1, this was a three-stage system with two stages in the HLT. At L1, fast custom-made electronics find regions of interest (RoI) using the calorimeter and muon data with coarse information within a latency of 2.5 μs. The L1 system in Run-2 consists of the L1 calorimeter trigger system (L1Calo), the L1 muon trigger system (L1Muon), new L1 topological trigger modules (L1Topo) and the Central Trigger Processors (CTP). At the HLT, fast algorithms accessing data from an RoI, or offline-like algorithms using the full-event information run on a unique PC farm within a processing time of 0.2 s on average.

3.3 Physics objects: reconstruction and calibration

The measurement of R_b makes use of events containing electrons, muons and jets. The reconstruction and identification of these objects is discussed in this section. Using the signals in the detectors described before, tracks emerging from the collisions are reconstructed together with the vertices of interaction, as described in Section 3.3.1. The reconstruction and identification of electrons and muons is given in Section 3.3.2 and 3.3.3, respectively. The algorithm used to build hadronic jets is discussed in Section 3.3.4 together with the identification performance. Given its crucial role in the R_b measurement, a detailed description of the of the algorithm and performance for the identification of b-jets is reported in Chapter 4.

3.3.1 Tracks and vertices

Charged particles traveling in the ATLAS detector leave several hits in the different sub-detectors of the ID. The hits of IBL, Pixel, SCT and TRT are reconstructed into tracks. Tracks satisfying particular characteristics identify vertices: the Primary Vertex (PV) is the one that more likely can be associated to the hard scatter interaction. The PV is used for calculation of the main physics objects in this analysis.

Hits recorded in the individual ID layers are used to reconstruct the trajectories of charged particles inside the tracker and to estimate their kinematic parameters. Particles trajectories are parametrised with a five parameter vector

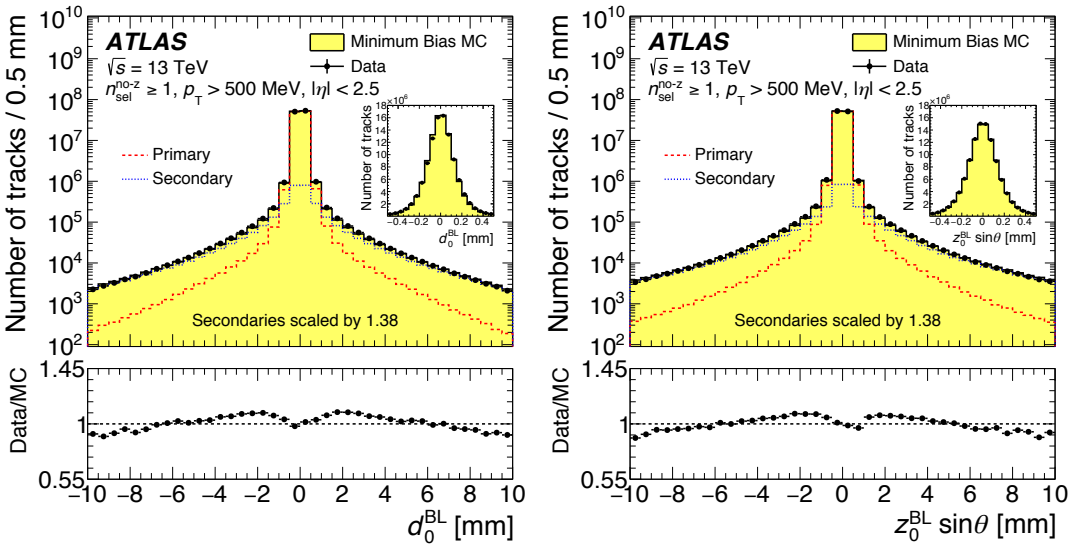
$$\tau = \tau(d_0, z_0, \phi_0, \theta, q/p), \tag{3.7}$$

where d_0 is the transverse impact parameter defined as the distance of closest approach in the transverse plane of the track to the primary vertex; z_0 is the longitudinal impact parameter, ϕ_0 is the azimuthal angle of the track and θ is its polar angle. q/p is the inverse of the particle momentum multiplied by its electric charge.
Track reconstruction in ATLAS [62] is developed in two-stages procedure:

- **Track finder**: Hits left in the detector by charged particles traversing active detector elements are grouped together to form a track candidate.
- **Track fitter**: The hits are used to reconstruct the trajectories by performing a fit to the track kinematic parameters. The track fitting is based on the minimisation of the track-hit residuals.

The accuracy of the particle kinematics estimation is limited by a combination of: the resolution of the detector elements, the not detailed map of the magnetic field, the misalignment of the detector elements and the multiple scattering and energy losses

due to the material in the detector. A deterioration of reconstruction performance can also come from the increasing detector occupancy with pileup, which results in nearby hits from other particles confusing the pattern recognition algorithm.
Track reconstruction performance have been evaluated using data from pp collisions [63]. Figure 3.9a shows the comparison between simulation and data of the transverse impact parameter distribution of tracks calculated with respect to the average beam position d_0^{BL}. In Figure 3.9b the data and simulated distributions of the difference between the longitudinal position of the track along the beam line, at the point where d_0^{BL} is measured, and the longitudinal position of the primary vertex projected to the plane transverse to the track direction, $z_0^{BL}sin\theta$, are compared. A small discrepancy between data and simulated events is found for tracks in the range of $|d_0^{BL}| < 4\ mm$ and $z_0^{BL}sin\theta < 3mm$. The track reconstruction efficiency is measured

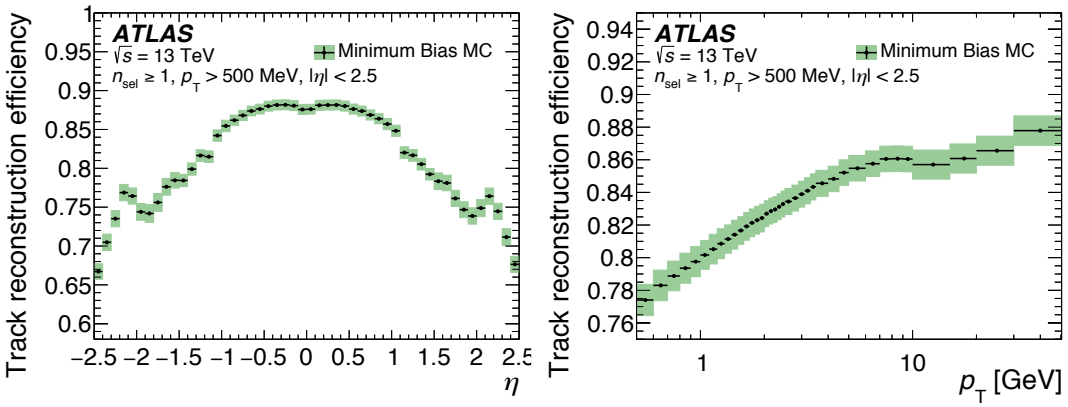


(A) (B)

FIGURE 3.9: Distribution of the transverse impact parameter d_0^{BL} (A). Distribution of $z_0^{BL}sin\theta$ (B). Simulated PYTHIA 8 A2 (yellow) events are compared to data (black dots) [62] .

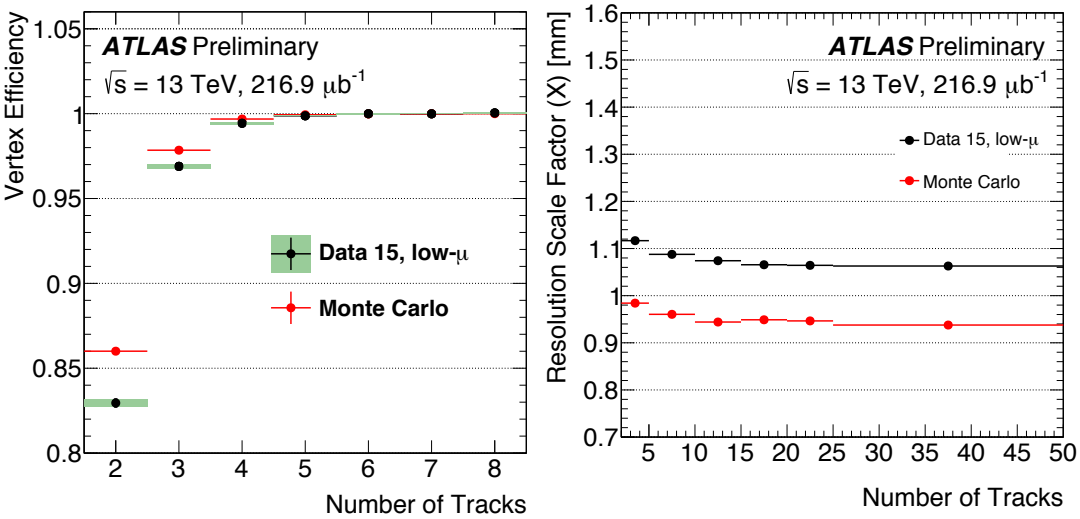
as function of the track $|\eta|$ and p_T and shown in Figure 3.10a and 3.10b, respectively. Here the simulated efficiency is shown after the application of a correction factor derived in data. The efficiency shows a dependency on both the pseudorapidity and the transverse momentum.

A vertex requires at least two tracks passing selection criteria to be formed as a candidate vertex [64]. Between all these candidates the vertex with the higher Σp_T^2 is choose to be the PV.
The vertex reconstruction efficiency is determinate from data by taking the ratio between events with a reconstructed vertex and events with at least two reconstructed tracks [65]. The measured vertex efficiency is shown in Figure 3.11a as function of the number of tracks in low-μ. The plot suggests that the vertex reconstruction efficiency reaches the 100% if the number of tracks is larger than 5. Scale factors for the reconstructed vertex resolution in the x-axis σ_x, are shown in Figure 3.11b.
The resolution is around 1.1mm in data and decreases with the number of tracks.



(A) (B)

FIGURE 3.10: Track reconstruction efficiency after the data driven correction as function of the track pseudorapidity (A) and p_T (B). The simulation is PYTHIA 8 A2 [62].



(A) (B)

FIGURE 3.11: Vertex reconstruction efficiency as function of the number of track in data (black dots) and simulated events (red dots) (A). Scale factors for the reconstructed vertex resolution σ_x as function of the number of tracks in data events (black dots) and simulated events (red dots) (B) [65].

3.3.2 Electrons

The characteristic signature of electrons consists in a track in the ID together with a narrow shower in the EM calorimeter. Electrons are identified within the region $|\eta|<$ 2.47, where the ID coverage ends. The region 1.37 $<|\eta|<$ 1.52 corresponds to the *crack region*, situated between barrel and end-cap calorimeters. Here, a precise simulation of the material is difficult, due to the presence of infrastructures for cooling, support and services. Electrons in this region are therefore excluded from the analyses. The ID provides not only tracking capabilities but also discriminates between electrons and charged hadrons via the detection of X-rays produced by transition radiation.

Reconstruction

Electron reconstruction in the central region of the ATLAS detector ($|\eta|<$ 2.47) proceeds in several steps [66]:

- **Seed-cluster reconstruction:** A sliding window with a size of 3 $\times$ 5 in units of 0.025 $\times$ 0.025, corresponding to the granularity of the EM calorimeter middle layer, in $\eta \times \phi$ space is used to search for electron cluster "seeds" as longitudinal towers [1] with total cluster transverse energy above 2.5 GeV. The clusters are then formed around the seeds using a clustering algorithm that allows for duplicates to be removed.

- **Track reconstruction:** Track reconstruction proceeds in two steps: pattern recognition and track fit. If a track seed (consisting of three hits in different layers of the silicon detectors) with a transverse momentum larger than 1 GeV can not be successfully extended to a full track a second attempt is performed with the new pattern recognition using an electron hypothesis that allows for larger energy loss. Track candidates are then fit either with the pion hypothesis or the electron hypothesis (according to the hypothesis used in the pattern recognition). If a track candidate fails the pion hypothesis track fit (for example, due to large energy losses), it is refit with the electron hypothesis. In this way, a specific electron-oriented algorithm has been integrated into the standard track reconstruction. It improves the performance for electrons and has minimal interference with the main track reconstruction.

- **Electron specific track fit:** The tracks thus obtained are loosely matched to EM clusters using the distance in η and ϕ between the position of the track, after extrapolation, in the calorimeter middle layer and the cluster ceter of gravity. The matching conditions account for energy-loss due to bremsstrahlung and the number of precision hits in the silicon detector.

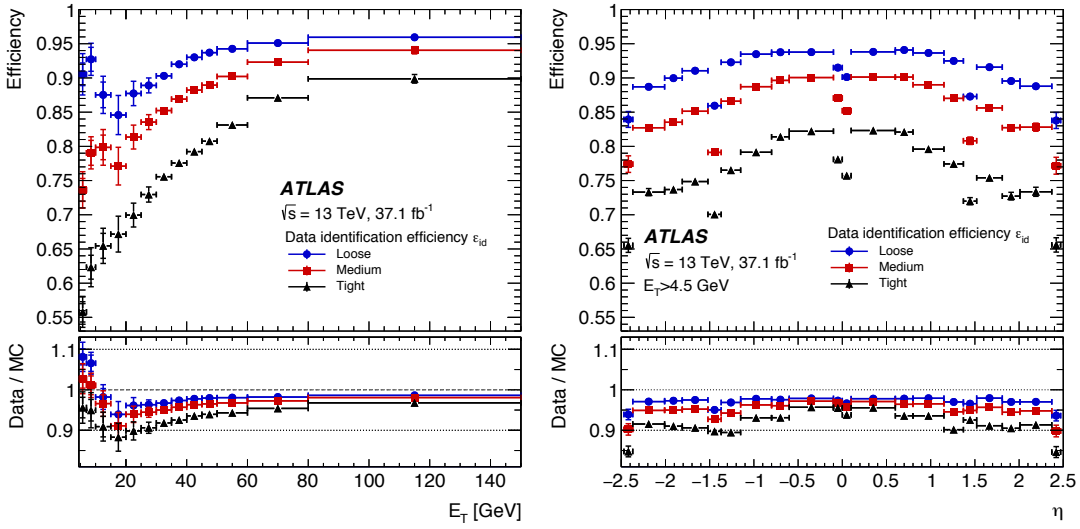
- **Electron candidate reconstruction:** The matching of the track candidate to the cluster seed completes the electron reconstruction procedure. A similar matching as the one described above is repeated for the refit track with stricter conditions. If several tracks fullfil the matching condition, one track is chosen as "primary" track.

[1] The $\eta \times \phi$ space of the EM calorimeter is divided into a grid of $N\eta \times N\phi$ = 200 $\times$ 256 elements of size $\Delta\eta$ *tower* $\times$ $\Delta\phi$ *tower* = 0.025 $\times$ 0.025, called towers. Inside each of these elements, the energy of cells in all longitudinal layers is summed into the tower energy, with the energy of cells spanning several towers distributed uniformly among the participating towers.

The four-momentum of the electrons is computed using information from both the final calibrated energy cluster and the best track matched to the original seed cluster. The energy is given by the final calibrated cluster, while the ϕ and η directions are taken from the corresponding track parameters with respect to the beam-line.

Identification

Three levels of identification operating points are typically provided for electron ID. These are referred to, in order of increasing background rejection, as *Loose*, *Medium*, and *Tight*. In general, each cut adds to the previous some additional requirements. Thus, electrons selected by Medium are all selected by Loose, and Tight electrons are all selected by Medium. Electron efficiencies are measured in pp collision data taken at a center-of-mass energy of $\sqrt{s}$ = 13 TeV in 2015 with spacing of subsequent LHC bunches of 50 ns and compared to efficiencies measured in a simulated $Z \rightarrow ee$ sample. These efficiencies are derived in data events using the Z *tag-and-probe method* and shown as function of the electron E_T and η in Figure 3.12a and 3.12b, respecitvely. The *tag-and-probe method* consists in reconstructing the Z boson mass using one tagged electron and one probe of which one wants to estimate the identification efficiency.



(A) (B)

FIGURE 3.12: Measured electron identification efficiencies in $Z \rightarrow ee$ events for the Loose (blue circle), Medium (red square), and Tight (black triangle) operating points as a function of E_T (A) and η (B) [66].

Isolation

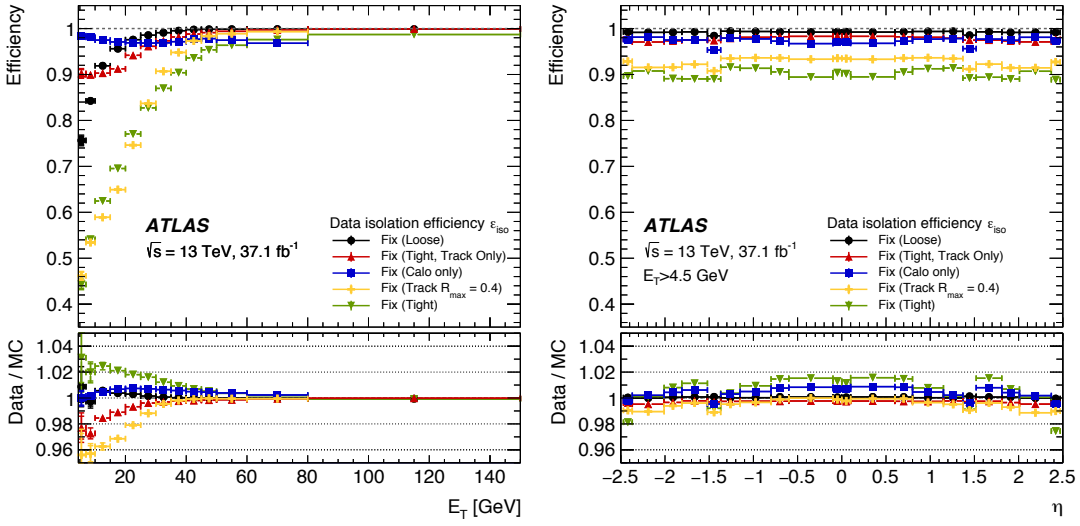
The ability to identify electron candidates isolated from any other local activity is based on two varieties of isolation variables, calorimeter- and tracking-based. Since the implementation of isolation criteria is specific to the physics analysis needs, several operating points are developed; their typical isolation efficiencies are measured in data and determined from simulation, ranging from approximately 90% for the tightest operating points to nearly 99% for the loosest, for electrons with E_T = 40 GeV. Several components enter into building the isolation variables: identifying the

candidate object itself, its direction, and its contribution to the activity within the cone, and summing, in a pileup and underlying- event robust way, the other activity found within the cone. Table 3.13 summarizes the definition of the electron-isolation operating points and isolation efficiency ε_{iso}. For the *Gradient* operating points, the units of p_T are GeV. All operating points use a cone size of ΔR=0.2 for calorimeter isolation and R_{max}=0.2 for track isolation except for the final entry *Fix (Track)* which uses R_{max}=0.4

Operating point	$E_{\mathrm{T,cone}}^{\mathrm{isol}}$ $(\Delta R = 0.2)$	$p_{\mathrm{T,var}}^{\mathrm{isol}}$ $(R_{\mathrm{max}} = 0.2)$	Total ϵ_{iso}
Loose (Track Only)	-	$\epsilon_{\mathrm{iso}} = 99\%$	99%
Loose	$\epsilon_{\mathrm{iso}} = 99\%$	$\epsilon_{\mathrm{iso}} = 99\%$	98%
Gradient	$\epsilon_{\mathrm{iso}} = 0.1143 \times p_{\mathrm{T}} + 92.14\%$	$\epsilon_{\mathrm{iso}} = 0.1143 \times p_{\mathrm{T}} + 92.14\%$	90(99)% at 25(60) GeV
Gradient (Loose)	$\epsilon_{\mathrm{iso}} = 0.057 \times p_{\mathrm{T}} + 95.57\%$	$\epsilon_{\mathrm{iso}} = 0.057 \times p_{\mathrm{T}} + 95.57\%$	95(99)% at 25(60) GeV
Fix (Loose)	$E_{\mathrm{T,cone}}^{\mathrm{isol}}/p_{\mathrm{T}} < 0.20$	$p_{\mathrm{T,var}}^{\mathrm{isol}}/p_{\mathrm{T}} < 0.15$	-
Fix (Tight)	$E_{\mathrm{T,cone}}^{\mathrm{isol}}/p_{\mathrm{T}} < 0.06$	$p_{\mathrm{T,var}}^{\mathrm{isol}}/p_{\mathrm{T}} < 0.06$	-
Fix (Tight, Track Only)	-	$p_{\mathrm{T,var}}^{\mathrm{isol}}/p_{\mathrm{T}} < 0.06$	-
Fix (Calo Only)	$E_{\mathrm{T,cone}}^{\mathrm{isol}} < 3.5$ GeV	-	-
Fix (Track $R_{\mathrm{max}} = 0.4$)	$E_{\mathrm{T,cone}}^{\mathrm{isol}}/p_{\mathrm{T}} < 0.11$	$p_{\mathrm{T,var}}^{\mathrm{isol}}/p_{\mathrm{T}} < 0.06$	-

FIGURE 3.13: Definition of the electron-isolation operating points and isolation efficiency for the different working points [66].

The isolation efficiency is measured in data events using electron emerging from a Z boson decay as function of the E_T and η, as shown in Figure 3.14a and 3.14b



(A) (B)

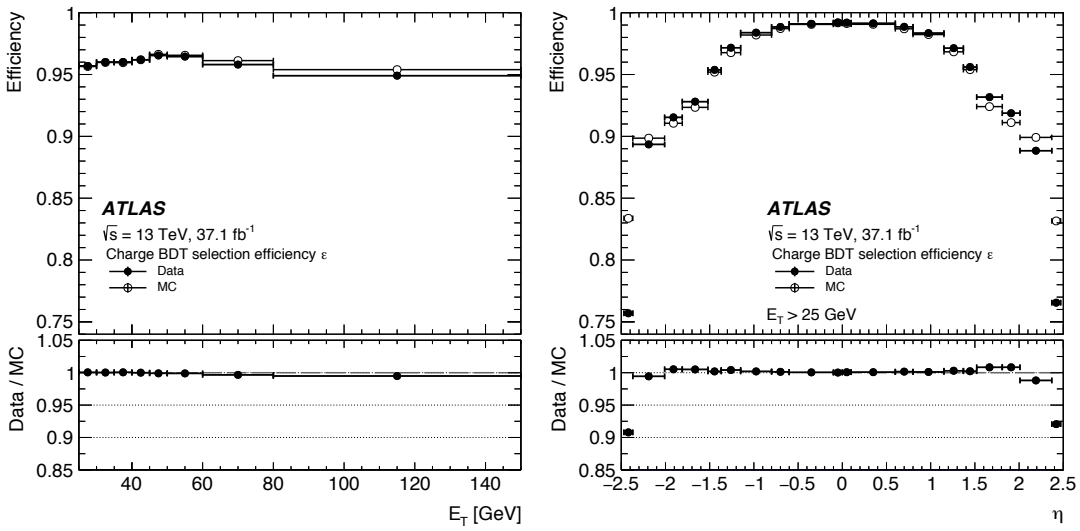
FIGURE 3.14: Isolation efficiencies for data for the different isolation operating points, as function of electron E_T (A) and η (B) [66].

The efficiency shows a marked dependency from the electron transverse energy for all the operating points, while an overall difference between the efficiency is present as function of the electron pseudorapidity.

Charge Identification

The electric charge of an electron is determined from the curvature of the associated track reconstructed in the inner detector. The mis-identification of electron charge

can result from the matching of an incorrect track to the electron candidate or from a mis-measurement of the curvature of the primary electron track. Pair production resulting from bremsstrahlung photon is the main reason why a wrong track can be matched to the electron candidate; three tracks in close proximity are present, two of which have the correct charge assignment, causing an ambiguity in the selection of the primary electron track.
The charge mis-identification rate for reconstructed electron candidates is reduced with an additional selection criterion based on the output discriminant of a boosted decision tree (BDT).
The efficiency of applying BDT selection is measured in data using $Z \to ee$ events as function of the E_T and η, as shown in Figure 3.15a and 3.15b



(A) (B)

FIGURE 3.15: Efficiency of the BDT selection criterion to choose Medium-identified electrons reconstructed with the correct charge, as evaluated in $Z \to ee$ events, as function of E_T (A) and η (B).

3.3.3 Muons

Muon reconstruction is first performed independently in the ID and MS [67]. Muons deposit a small amount of energy in the calorimeter system. The information from individual sub-detectors is then combined to form the muon tracks that are used in physics analyses.

Reconstruction

In the ID, muons are reconstructed like any other charged particles as previously described. Muon reconstruction in the MS starts with a search for hit patterns inside each muon chamber to form segments. In each MDT chamber and nearby trigger chamber, a Hough transform is used to search for hits aligned on a trajectory in the bending plane of the detector. The MDT segments are reconstructed by performing a straight-line fit to the hits found in each layer. The RPC or TGC hits measure the coordinate orthogonal to the bending plane. Segments in the CSC detectors are built using a separate combinatorial search in the η and ϕ detector planes. The search algorithm includes a loose requirement on the compatibility of the track with the

luminous region.
Muon track candidates are then built by fitting together hits from segments in different layers. The hits associated with each track candidate are fitted using a global χ^2 fit. A track candidate is accepted if the χ^2 of the fit satisfies the selection criteria. Hits providing large contributions to the χ^2 are removed and the track fit is repeated. A hit recovery procedure is also performed looking for additional hits consistent with the candidate trajectory. The track candidate is refit if additional hits are found.

The combined ID-MS muon reconstruction is performed according to various algorithms based on the information provided by the ID, MS and calorimeters. Four muon types are defined depending on how the information from the various subdetectors is combined:

- **Combined (CB) muons**: track reconstruction is performed independently in the ID and MS, and a combined track is formed with a global fit that uses the hits from both the ID and MS sub-detectors. Usually, muons are firstly reconstructed in the MS, where the track density is much smaller, and then extrapolated inward and matched to an ID track. They are the most commonly used muons in physics analysis since they have the highest purity and the best resolution on the kinematic properties.

- **Segment-tagged (ST) muons**: a muon is classified like ST muon if a track in the ID, once extrapolated to the MS, it is associated with at least one local track segment in the MDT or CSC chambers. ST muons are used when muons cross only one layer of MS chambers, either because of their low p_T or because they fall in regions with reduced MS acceptance.

- **Calorimeter-tagged (CT) muons**: a track in the ID is identified as a muon if it can be matched to an energy deposit in the calorimeter compatible with a minimum-ionizing particle. This kind of muons have the lowest purity with respect to all the other muon types but it recovers acceptance in the region where the ATLAS Muon Spectrometer is partially instrumented (to host cabling and services), close to $|\eta| = 0$.

- **Extrapolated (ME) muons**: the muon trajectory is reconstructed based only on the MS track and a loose requirement on compatibility with originating from the IP. ME muons are mainly used to extend the acceptance for muon reconstruction into the region $2.5 < |\eta| < 2.7$, which is not covered by the ID.

Identification

Four muon identification selections (*Medium*, *Loose*, *Tight*, and *High-pT*) are provided to address the specific needs of different physics analyses. *Loose*, *Medium*, and *Tight* are inclusive categories in that muons identified with tighter requirements are also included in the looser categories. These categories are listed here:

- **Tight muons**: Tight muons are selected to maximise the purity of muons at the expenses of the efficiency for real muons. Only CB muons with hits in at least two stations of the MS and satisfying the Medium selection criteria are considered.

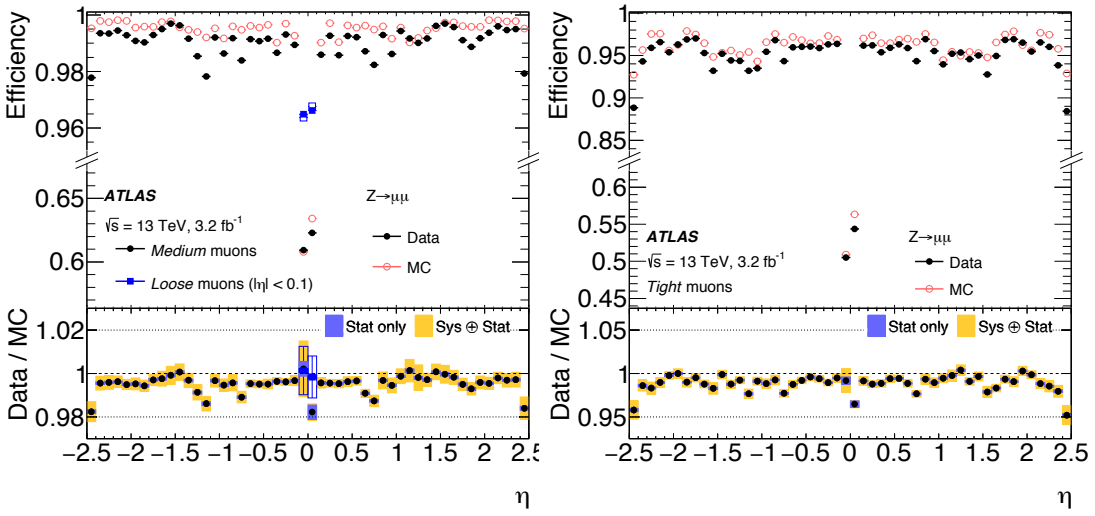
- **Medium muons**: The Medium identification criteria provide the default selection for muons in ATLAS. Only CB and ME tracks are used. The former are

required to have ≥ 3 hits in at least two MDT layers, except for tracks in the $|\eta| < 0.1$ region, where tracks with at least one MDT layer but no more than one MDT hole layer are allowed. The latter are required to have at least three MDT/CSC layers, and are employed only in the $2.5 < |\eta| < 2.7$ region to extend the acceptance outside the ID geometrical coverage. A loose selection on the compatibility between ID and MS momentum measurements is applied to suppress the contamination due to hadrons misidentified as muons.

- **Loose muons** The Loose identification criteria are designed to maximise the reconstruction efficiency, providing, however, good-quality muon tracks. All muon types are used in the loose definition. All CB and ME muons satisfying the Medium requirements are included in the Loose selection. CT and ST muons are restricted to the $|\eta| < 0.1$ region.

- **High-pT muons**: The High-pT selection aims to maximise the momentum resolution for tracks with transverse momentum above 100 GeV. The selection is optimised for searches for high-mass Z' and W'resonances.

As described for electrons, muon efficiency is evaluated using a tag-and-probe method on $J/\Psi \to \mu\mu$ and $Z \to \mu\mu$ samples. In Figure 3.16a and 3.16b the *Medium* and *Tight* working points efficiencies are shown as function of the muon η.



(A) (B)

FIGURE 3.16: Muon reconstruction efficiency as a function of η measured in $Z \to \mu\mu$ events for muons with $p_T > 10$ GeV. The Medium (A) and Tight (B) working points are shown [67].

Isolation

Muons originating from the decay of heavy particles, such as W, Z, or Higgs bosons, are often produced isolated from other particles. Unlike muons from semileptonic decays, which are embedded in jets, these muons are well separated from other particles in the event. The measurement of the detector activity around a muon candidate, referred to as muon isolation, is therefore a powerful tool for background rejection in many physics analyses.
Two variables are defined to assess muon isolation: a track-based isolation variable

and a calorimeter- based isolation variable. Seven isolation selection criteria (isolation working points) are defined, each optimised for different physics analyses and calibrated implying $Z \to \mu\mu$ events. For the sake of simplicity, only the results from the calibration of the so-called Gradient working point is shown in Figure 3.17 as it is the one used in the R_b measurement. This working point is defined cutting on the

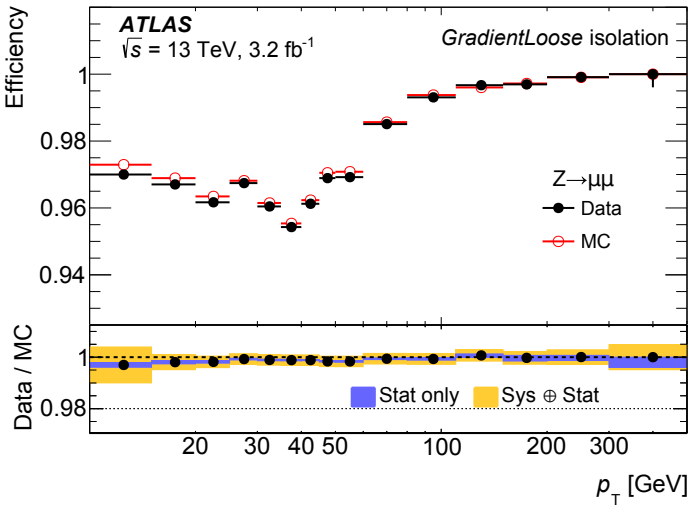


FIGURE 3.17: Isolation efficiency for the GradientLoose muon isolation working points. The efficiency is shown as a function of the muon transverse momentum p_T and is measured in $Z \to \mu\mu$ events. The full (empty) markers indicate the efficiency measured in data (MC) samples [67].

isolation variables $p_T^{varcone30}/p_T^{\mu}$ and $E_T^{topocone20}/p_T^{\mu}$ ensuring an isolation efficiency of 95% for muons of at least 25 GeV, and 99% for muons of at least 60 GeV.

The track-based isolation variable, $p_T^{varcone30}$, is defined as the scalar sum of the transverse momenta of the tracks with p_T >1 GeV in a cone of size ΔR = min(10 GeV/p_T^{μ},0.3) around the muon of transverse momentum p_T, excluding the muon track itself. The cone size is chosen to be p_T-dependent to improve the performance for muons produced in the decay of particles with a large transverse momentum.

The calorimeter-based isolation variable, $E_T^{topocone20}$, is defined as the sum of the transverse energy of topological clusters in a cone of size $\Delta R = 0.2$ around the muon, after subtracting the contribution from the energy deposit of the muon itself and correcting for pileup effects.

3.3.4 Hadronic Jets

A jet is a collimated set of hadrons and other particles produced by the hadronisation of a quark or gluon. At hadron colliders, jets are produced as result of the elastic scattering or the annihilation of partons. Quarks and gluons carry color charge, but they cannot exist in colored state because of the QCD confinement. Therefore, when they fragment, each fragmentation product carries away some of the color charge in order to obey confinement. The ensemble of these objects is what it is usually referred to as *hadronic jet*.

Reconstruction

Hadronic jets are reconstructed from the ATLAS detectors signals using the tracks reconstructed in the ID or the energy clusters in the calorimeter. Three different algorithms are currently supported by the ATLAS collaboration: a track-based algorithm, a cluster-based algorithm [68, 69] and the particle flow algorithm [70], in which the informations from clusters and tracks are combined.
In this work, jets reconstructed using the calorimeter energy deposit clusters, hereafter called *EMTopo* jets, are used and described in following.
EMTopo jets are reconstructed using the *anti-k_t* algorithm [71] with radius parameter $R = 0.4$ or $R = 0.6$, using as input positive-energy topological clusters (topo-clusters) [69]. The topo-clusters are built from topologically connected calorimeter cells that contain a significant signal above noise. The topo-clusters are reconstructed at the EM scale, which correctly measures the energy deposited by particles produced in electromagnetic showers in the calorimeter.

The *anti-k_t* jet clustering algorithm is collinear and infrared safe:

- The collinear safety is connected to how the transverse momentum is distributed among the collinear decay products. For example a reconstructed jet should not depend on the decay of a particle into two collinear particles.
- The infrared safety is connected to the sensitivity of the algorithm to the presence of an additional soft particle not related to the fragmentation of the hadronised particles. For example, if there is a low energy cluster between two high energy clusters separated, the algorithm should have the possibility to discriminate between the two high energy clusters and define it as two separated objects.

As first step the two distances are introduced:

- d_{ij} between entities (particles, pseudojets) i and j
- d_i between entity i and the beam

The clustering proceeds by identifying the smallest of the distances: if it is a d_{ij} recombining entities i and j, while if it is d_i calling i a jet and removing it from the list of entities. The distances are recalculated and the procedure repeated until no entities are left. These distances are defined as follows:

$$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \tag{3.8}$$

$$d_i = k_{ti}^{-2}, \tag{3.9}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti}, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth angle of particle i.

Calibration

The sample of jets obtained from the clustering procedure are then calibrated with a procedure described in the following. Jets are calibrated to the energy scale of truth jets created using the same clustering algorithm from stable interacting particles in Monte Carlo [72]. The calibration has to account for several different effects:

- Calorimeter non-compensation: correction for the different scales of the energy measured from hadronic and electromagnetic showers.
- Dead material: energy lost in inactive areas of the detector.
- Leakage: showers reaching the outer edge of the calorimeters.
- Out of calorimeter jet: energy of particles which are included in the truth jet but which are not included in the reconstructed jet.
- Energy deposits below noise thresholds: clusters are only formed by energy deposits which are well above the background noise. Therefore the correction is required to correct for particles that do not form clusters. Additionally some part of a shower may fall outside of the topological clusters such that this also needs to be corrected for.
- Pileup: energy deposition in jets is affected by the presence of multiple pp collisions in the same bunch crossing as well as residual signals from other bunch crossings.

The calibration is derived using a combination of methods based on Monte Carlo simulation and data-driven techniques. The data are used through data-driven methods to derive a small residual calibration correction and constrain the uncertainty in the calibration. As a first step of the calibration procedure, a jet is corrected to point back to the identified hard-scatter vertex. Next, the effect of pileup is removed using an area-based subtraction procedure and residual corrections. The jet energy is then calibrated by applying a p_T and η-dependent correction derived from the nominal simulation. In addition to these steps, further corrections are applied to the jets that reduce the dependence of the jet energy measurement on the longitudinal and transverse structure of the jets and also correct for jets that are not fully contained in the calorimeter. Finally, for jets in the data, an additional correction is applied that changes their calibration to its correct value based on in situ studies [73].

The calibration provides a jet energy scale (JES) and resolution (JER). The relative uncertainties for these two quantities are shown in Figure 3.18 and 3.19 for central jets, as function of their p_T.

Pileup suppression techniques

As shown in the uncertainty associated to the JES, pileup contributes to the total uncertainty with a leading contribution. Therefore, a pileup suppression technique is developed and presented here.
The performance of the jet reconstruction are affected by pileup in two ways: hard QCD jets originating from a pileup vertex and detector signals which occurred a bunch crossing before the event of interest. The pileup QCD jets are genuine jets and are tagged and rejected using the vertex-pointing information of charged-particle tracks. Since tracks can be precisely associated with specific vertices, track-based observables can provide information about the pileup structure and vertex composition of jets within the tracking detector acceptance ($|\eta| < 2.5$). This information is therefore used to build a discriminating quantity for hard scatter jets. The composition of pileup jets depends on both $< \mu >$ and p_T.
The Jet Vertex Tagger algorithm [74] matches jet tracks to the primary reconstructed

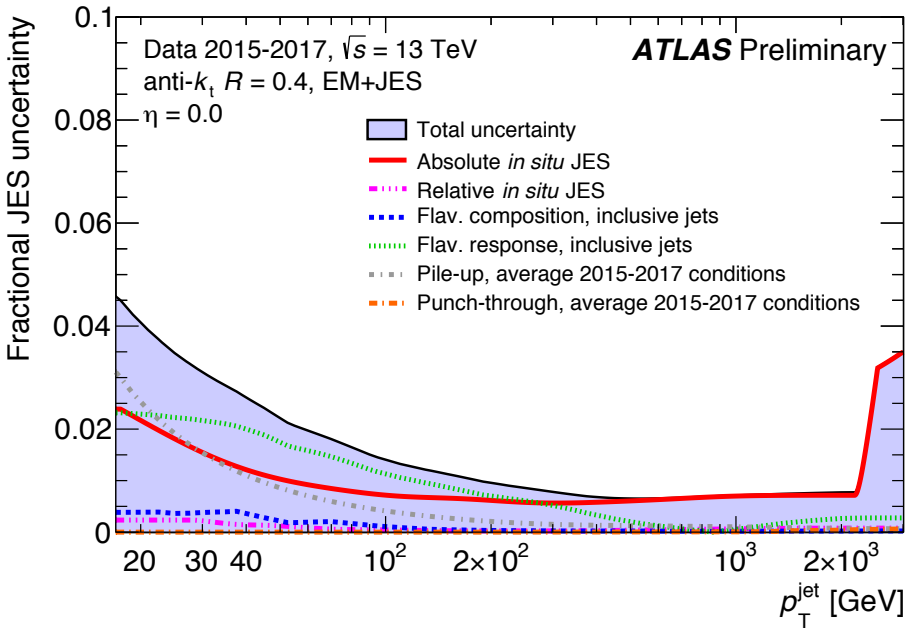


FIGURE 3.18: Fractional jet energy scale systematic uncertainty components as a function of jet p_T at $\eta = 0.0$. The total uncertainty (all components summed in quadrature) is shown as a filled region topped by a solid black line. Topology-dependent components are shown under the assumption of a dijet flavour composition.

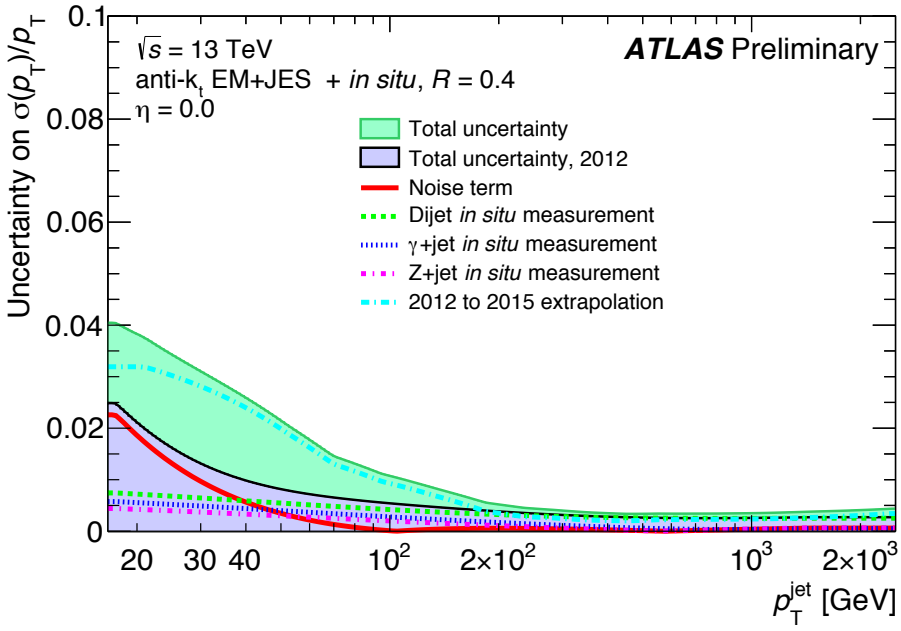


FIGURE 3.19: Fractional jet energy resolution systematic uncertainty components as a function of jet p_T at $\eta = 0.0$. The total uncertainty (all components summed in quadrature) related to 2015 data is shown as a filled region topped by a solid green line and compared to the total uncertainty in 2012 (black solid line) [73].

vertex. A selection on the output distribution, the JVT discriminant, define the working point selecting more than 90% of hard scatter jets. The efficiency distribution in data and simulated events is shown in Figure 3.20.

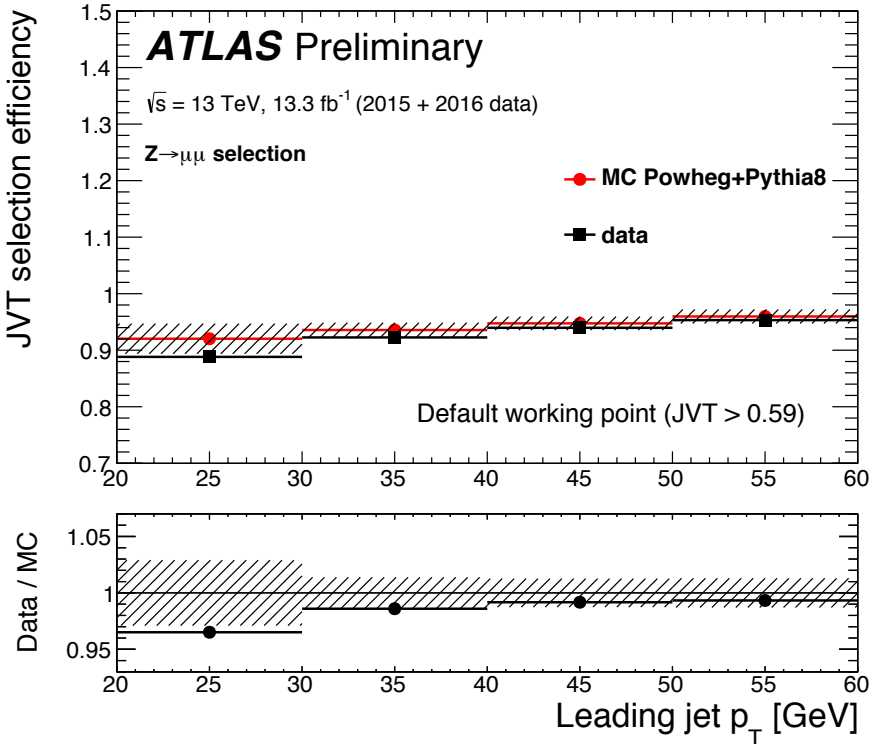


FIGURE 3.20: The hard-scatter jet selection efficiency, in Powheg+Pythia8 MC (red) and in 2015+2016 data (black) as function of the jet p_T.

3.3.5 Missing Transverse Energy

In collider experiments, the conservation of momentum in the plane transverse to the beam axis implies that the transverse momentum of the collision products should sum to zero. Any imbalance is known as *Missing Transverse Momentum*, or E_T^{miss}, and may be indicative of weakly-interacting, stable particles in the final state. Within the Standard Model, this arises from neutrinos. Fake E_T^{miss} can also result from interacting Standard Model particles which escape the acceptance of the detector, are badly reconstructed, or fail to be reconstructed altogether. Thus E_T^{miss} can also serve as an important measure of the overall event reconstruction performance.
The E_T^{miss} reconstruction uses selected calibrated hard objects to measure the missing transverse momentum in an event [75]:

$$-E_T^{miss} = \Sigma p_T^e + \Sigma p_T^\gamma + \Sigma p_T^\tau + \Sigma p_T^\mu + \Sigma p_T^{jet} + \Sigma p_T^{soft}. \tag{3.10}$$

The hard objects used in the E_T^{miss} computation are electrons, photons, hadronically decaying tau-leptons, muons and jets. The soft term is reconstructed from detector signal objects not associated with any hard object passing selection cuts. These can be ID tracks (track-based soft term TST) or calorimeter signals (calorimeter-based soft term CST)

- Calorimeter-based soft term (CST) is constructed from the energy deposits in the calorimeter not associated with hard objects: selected and reconstructed electrons, photons, hadronically decaying tau-leptons, muons or jets. Contributions to the soft term arise from underlying event activity and soft radiation from the hard event. The calorimeter-based approach is intrinsically vulnerable to additional proton-proton interactions overlapping the hard-scatter process (pileup interactions).

- Track-based methods offers greater resilience under the conditions of increased pileup expected during Run 2. Tracks, differently from calorimetric clusters, can be associated to vertices. Thus this soft term (TST) definition based on the momenta of ID tracks gives a measure which is largely independent of the pileup. A purely track-based quantity is, however, insensitive to neutral particles (which do not leave tracks in the ID) and has an acceptance limited by the tracking volume of the ATLAS tracker. So this soft term is combined this with calorimeter-based measurements for the hard objects in order to reduce the bias on the overall missing momentum from the lost low-energy neutral component. The primary vertex associated tracks from jets which fail these quality requirements are included in the TST soft term.

The track based soft term is employed in the E_T^{miss} calculation used in this work.

Performance

The performance of E_T^{miss} are evaluated in $Z \to \mu\mu$ samples. These events provide an ideal final state for the evaluation of E_T^{miss} performance, due to the limited backgrounds and precise measurement of the kinematics of the Z boson. Neutrinos are produced only through heavy-flavour meson decays, so this channel has very little genuine E_T^{miss}, making the resolution measured by the width of the E_T^{miss} distribution indicative of the E_T^{miss} reconstruction quality. The distribution of E_x^{miss} is shown in Figure 3.21 while its resolution in Figure 3.22.

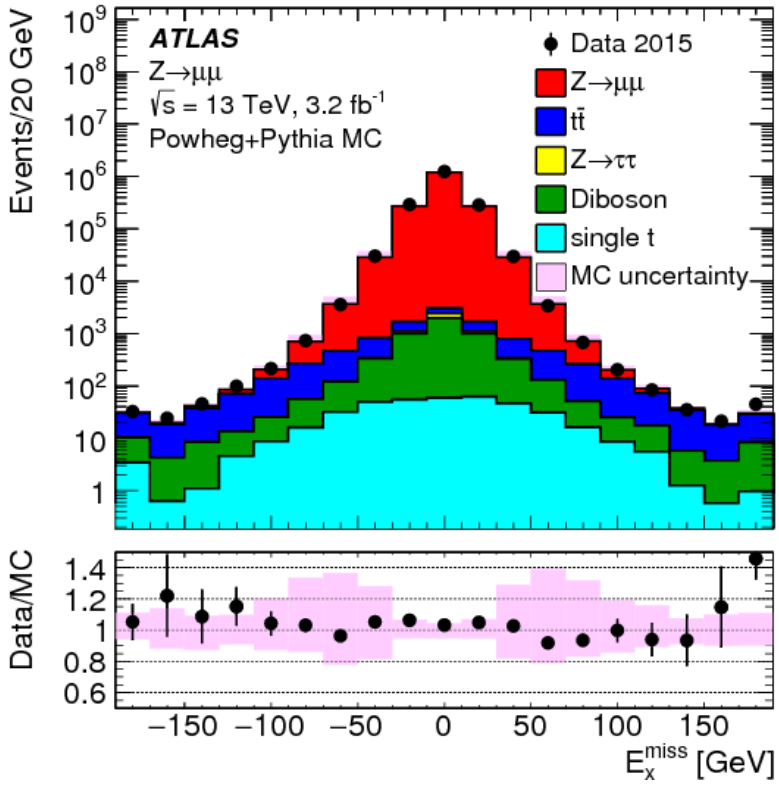


FIGURE 3.21: Distributions of the $E_{x,y}^{miss}$ in $Z \rightarrow \mu\mu$ events. The expectation from MC simulation is superimposed and normalised to data, after each MC sample is weighted with its corresponding cross-section [75].

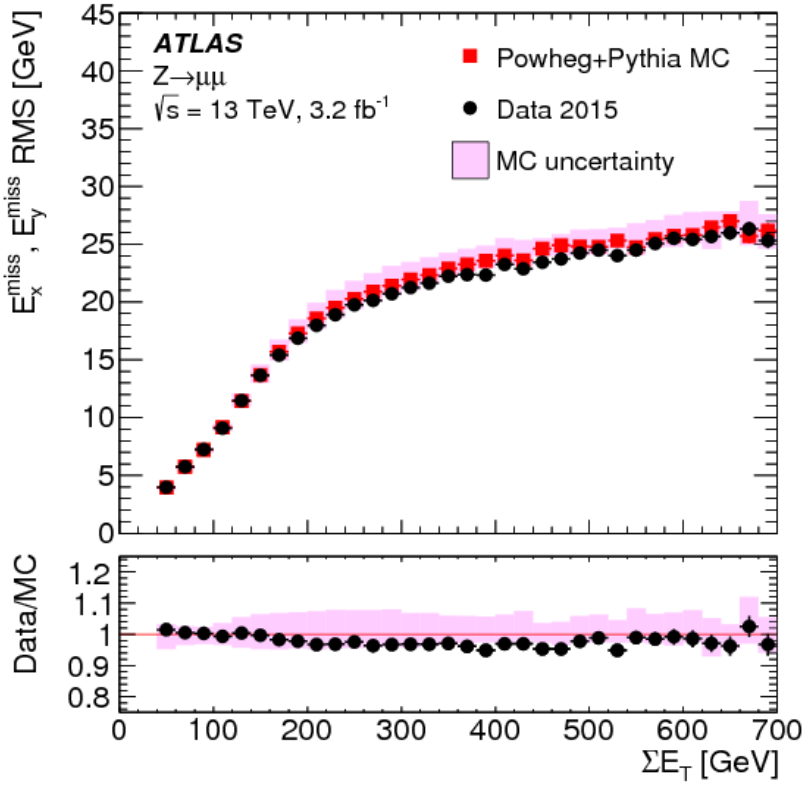


FIGURE 3.22: The RMS width of the $E_x(y)^{miss}$ distributions in bins of ΣE_T in an inclusive sample of $Z \rightarrow \mu\mu$ events. Predictions from MC simulations are overlaid on the data points, and the ratios are shown below the respective plot. The shaded bands indicate the combined statistical and systematic uncertainties of the resolution measurements [75].

Chapter 4

Identification of *b*-jets

The identification of jets containing *b*-hadrons, known as b-*tagging*, aims to discriminate these particular jets from the larger background of jets containing other quark flavours.
The interest of identifying *b*-jets arises from several physics signatures among which: the Higgs boson decay into a bottom quarks pair, its associated production to top quarks, precision measurements of the SM such as the differential cross sections of the top quarks pairs, the determination of the top quark mass and width, and searches for new phenomena.
The ATLAS Collaboration developed various *b*-tagging algorithms [40] that exploit the properties of *b*-hadrons and *b*-quark fragmentation.
Given the long lifetime of the order of 1.5 ps, *b*-hadrons have a significant mean flight length in the detector before the decay. In principle, this allows to resolve the decay vertices from the hard-scatter point and reconstruct the full path of production and decay of the *b*-hadrons. A schematic view of the *b*-hadron production and decay is shown in Figure 4.1.

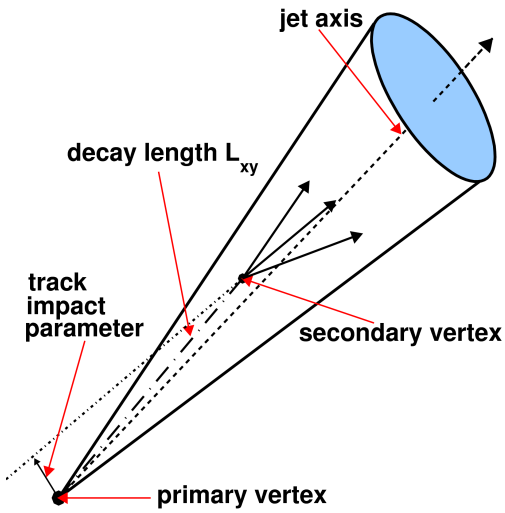
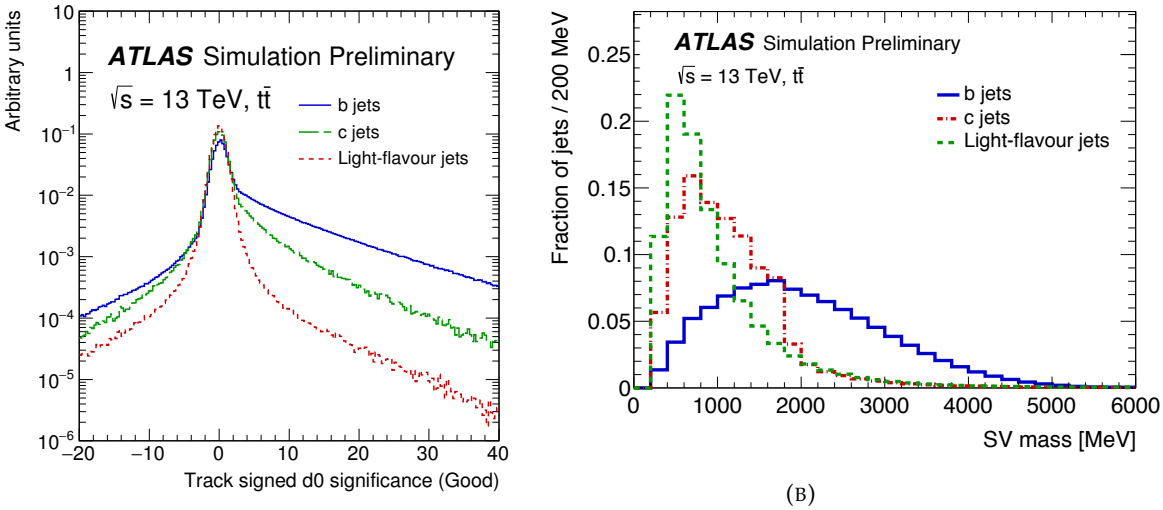


FIGURE 4.1: Schematic view of a *b*-hadron decay inside a jet resulting in a secondary vertex with three charged particle tracks. The track impact parameter, which is the distance of closest approach between the extrapolation of the track and the primary vertex, is shown in addition for one of the secondary tracks.

Moreover, displaced charged-particle tracks originating from b-hadron decays usually have large transverse d_0 (Figure 4.2a) and longitudinal z_0 impact parameters. In addition to topological information, kinematics is also implied. For example, on account of the high b-quark mass $m_b = 4.18^{+0.04}_{-0.03}$ GeV [7], the decay products of b-hadrons are on average more energetic than other hadrons: as a result kinematical variables, such as the secondary vertex mass (Figure 4.2b), are used to b-tag jets.

Finally, b-hadrons have an higher decay tracks multiplicity which is also used by



(A) (B)

FIGURE 4.2: The transverse (A) signed impact parameter significance of tracks in $t\bar{t}$ events for b-jets (blue), c-jets (green) and *light*-jets (red) [**?**]; Secondary vertex total invariant mass (B) distribution in $t\bar{t}$ events with b-jets (blue), c-jets (red) and light jet (green) [76]

the algorithms to discriminate against charm and light-flavour jets.

The algorithms, described in Section 4.1, are developed using Monte Carlo simulations: any imperfection in the description of the detector response or physics modeling needs to be corrected trough a measurement of the algorithm performance using data, the so-called *calibration*.

The calibration procedure is performed separately for each jet flavour (bottom, charm and light) where the b-tagging efficiency is derived as function of the jet kinematics. The calibration of the b-tagging efficiency for b-jets using the $p_{\mathrm{T}}^{\mathrm{rel}}$ method is reported in Section 4.2: as motivated in Section 2.2, a model independent calibration, not based on top events, is one of the needed ingredients for the R_{b} measurement.

4.1 Algorithms for the identification of b-jets

The development of the b-tagging algorithms is based on a two-stage approach that makes use of tracks, vertices and hadronic jets as they have been described in Section 3.3.

The jet flavour label classifies the jets as bottom, charm or light using the truth MC simulation information and a spatial match between hadrons and jets. As the algorithms make use of tracks, also these objects are associated to jets through a spatial match.

Once all the needed ingredients are in place, basic algorithms (Section 4.1.1) reconstruct the characteristic features of the *b*-jets via two complementary approaches, one that uses the individual properties of charged-particle tracks and a second which combines the tracks to explicitly reconstruct displaced vertices.
In order to maximise the *b*-tagging performance, the results of the basic *b*-tagging algorithms are combined in algorithms consisting of multivariate classifiers. The final algorithms are two: MV2 [77], a tagger based on a boosted decision tree (BDT) discriminant, and DL1 [77], which is based on a deep feed-forward neural network. These two final taggers show similar performance in the region of interest of the R_{b} measurement: for the sake of simplicity only MV2 is described in Section 4.1.2 as it is the one used in this work.

4.1.1 Basic Taggers

There are several basic taggers which fall into two classes: the ones exploiting the large impact parameters of the tracks originating from the *b*-hadron decay (IP2D and IP3D) and the ones aimed to explicitly reconstruct the displaced vertices (SV1 and JetFitter).

The IP2D and IP3D algorithms

Two complementary impact parameter based algorithms are developed: IP2D and IP3D [78]. The IP2D tagger makes use of the signed transverse impact parameter significance d_0/σ_{d_0} as discriminant variable whereas IP3D uses both the signed transverse and the longitudinal impact parameter significance $z_0 sin\theta/\sigma_{z_0 sin\theta}$ in a two-dimensional template to account for their correlation.
Probability density functions (PDF) obtained from the reference histograms of the signed transverse and longitudinal impact parameter significances of tracks associated with *b*-jets, *c*-jets and *light*-jets are derived from the MC simulation. The PDFs are used to calculate ratios of the *b*-jets, *c*-jets and light jets probabilities on a per-track basis. Log-likelihood ratio (LLR) discriminants are then defined and used as inputs to the final taggers.

Secondary Vertex Finder algorithm

The Secondary Vertex Finder algorithm SV1 [76] reconstructs a single displaced secondary vertex (SV) in a jet. Firstly, the algorithm finds all two-track vertices then it identifies and reject as many two-track vertices as possible that are unrelated to the vertices of interested and finally then merges them into a multi-track vertex. The rejected tracks are compatible with the decay of long-lived particles, photon conversions and hadronic interactions with the detector material.
The SV1 algorithm uses a likelihood ratio test to discriminate between b-, c- and *light*-jets. The corresponding PDFs are constructed using the vertex mass, the energy fraction and the number of two-track vertices distributions. The SV1 output obtained is given as input to the final taggers.

Topological Multi-Vertex Algorithm (JetFitter)

The topological multi-vertex algorithm, JetFitter [79], reconstructs the full *b*-hadron decay chain. This is possible thanks to the topological structure of the weak decays

of b- and *c*-hadrons inside the jet. The algorithm makes use of a modified Kalman filter [80] to find a common line on which the primary, bottom and charm vertices lie under the approximation that the vertices positions are placed along the *b*-hadron flight path. Eight discriminating variables, including the track multiplicity at the JetFitter displaced vertices, the invariant mass of tracks associated with these vertices, their energy fraction and their average three-dimensional decay length significance, are used as inputs to the final taggers.

4.1.2 The MV2 Tagger

The MV2 algorithm consists of a boosted decision tree (BDT) algorithm that combines the outputs of the basic taggers together with the kinematic properties of the jets, namely p_T and η. The kinematic properties are added in the training in order to take advantage of the correlations with the other input variables. The output discriminant of the MV2 algorithm for *b*-jets, *c*-jets and *light*-jets evaluated using $t\bar{t}$ simulated events is shown in Figure 4.3.

The distribution of the MV2 output discriminant is dominated by *b*-jets for values

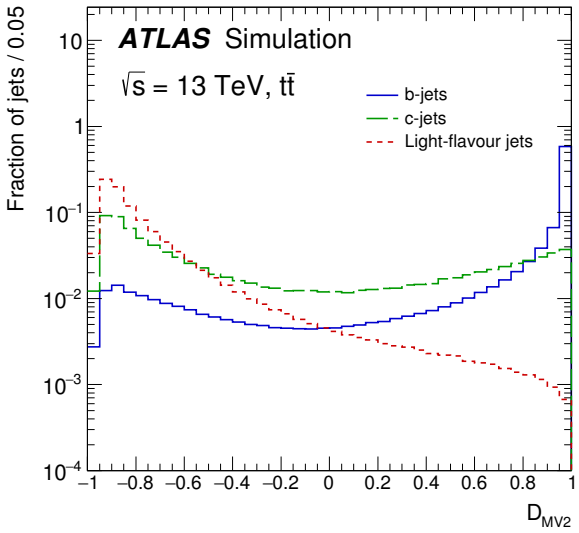
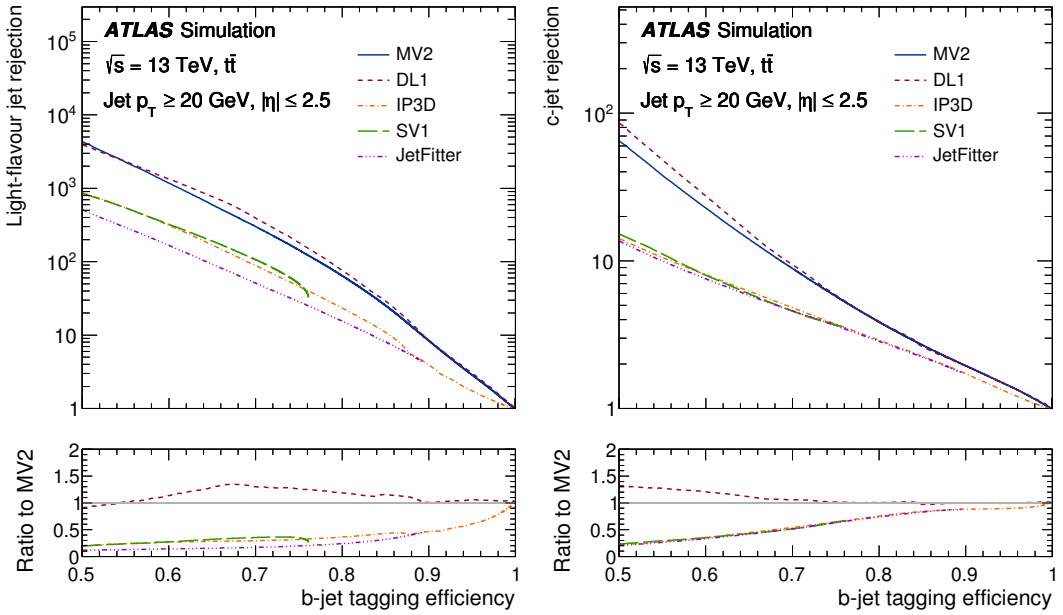


FIGURE 4.3: Distribution of the output discriminant of the MV2 *b*-tagging algorithm for *b*-jets (blue), *c*-jets (green) and light-flavour (red) jets in the baseline $t\bar{t}$ simulated events. [40]

around 1 while charm and light-flavour jets are more concentrated around -1. The *light*-jet 4.4a and *c*-jets 4.4b rejections are shown as function of the *b*-tagging efficiency for the various taggers: from the distribution of the output discriminant the rejection of charm and light jets is bigger for lower values of the *b*-jets tagging efficiency.

The performance of the *b*-tagging algorithms is based on a fixed selection on the *b*-tagging algorithm discriminant distribution. This ensures a specific *b*-jets tagging efficiency ε_b for *b*-jets in $t\bar{t}$ simulated events. Four working points (WP) are defined at 85%, 77%, 70% and 60% for the MV2 tagger.

The calibration of the *b*-tagging efficiency at the 77% WP for *b*-jets is shown in the



(A) (B)

FIGURE 4.4: The light-flavour jet (A) and c-jets (B) rejections versus the b-jets tagging efficiency for the IP3D, SV1, JetFitter, MV2 and DL1 b-tagging algorithms evaluated on the baseline $t\bar{t}$ events [40].

next section as it is the WP used in R_b measurement. This WP is chosen as it represents a good compromise between the number of b-jets selected and the uncertainty associated to their b-tagging efficiency.

4.2 Calibration of b-tagging efficiency

In order to have a thorough understanding of the b-tagging performance, the b-tagging efficiency of the MV2 tagger is measured in collision data. The calibration is performed in a sample of events characterised by a strong predominance of b-jets, whose fractional abundance can be measured from data before and after the b-tagging has been applied.
A sample enriched in b-jets can be obtained by selecting jets containing a muon inside, from which the b-tagging efficiency can be measured taking advantage of the kinematical properties of the semileptonic decay of b-hadrons. The p_T^{rel} method [41] uses templates of the muon momentum transverse with respect to the axis defined by the muon-plus-jet system. The fraction of b-jets before and after the b-tagging is estimated from the p_T^{rel}distribution and used to extract the b-tagging efficiency.
An introduction to the method is given in Section 4.2.1, followed by a description of the samples used given in Sections 4.2.2 and the events and objects selection of Section 4.2.3. Section 4.2.4 describes the derivation of the templates used in the fit. The description and impact of the systematic effects on the final uncertainty is exposed in Section 4.2.5. Section 4.2.6 contains the results of the calibration.

4.2.1 Introduction to the p_T^{rel} method

The number of b-jets before and after tagging can be obtained for a subset of all b-jets, namely those containing a reconstructed muon, using the variable $p_{\mathrm{T}}^{\mathrm{rel}}$ (Figure 4.5). This method relies on the muon channel as it ensures better experimental conditions

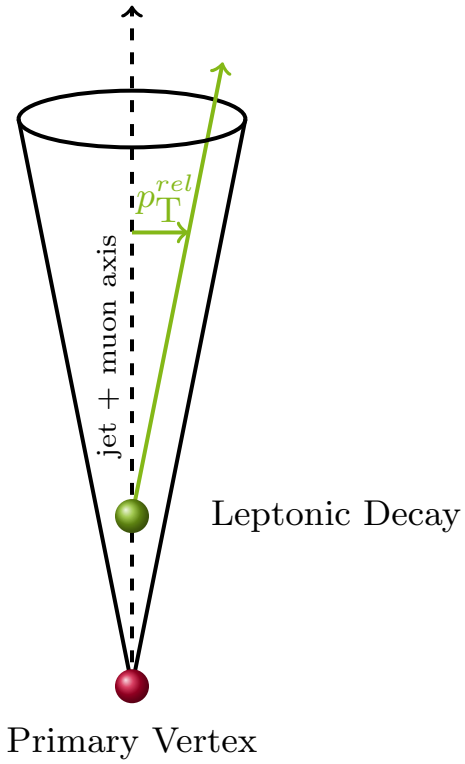


FIGURE 4.5: Schematic drawing showing a jet cone in black with a b-hadron in the jet, decaying semileptonically at a secondary vertex shown in green. The final state muon is shown in green as well as the projection of its momentum transverse to the jet+muon axis, the $p_{\mathrm{T}}^{\mathrm{rel}}$.

than the other charged leptons generated in b-hadrons semileptonic decays: electrons and τ leptons are reconstructed relying on the information from the calorimeters, while muons are reconstructed as tracks in the Inner Detector and in the Muon Spectrometer resulting in an higher isolation, better momentum resolution and reconstruction efficiency.
Given the larger mass of b-hadrons, relative to hadrons containing only charm and light quarks, their decay products are more energetic. Therefore, the final state particles have larger momenta in the rest frame of the decaying hadron, hereafter referred to as p^*, compared to the decays of less massive hadrons. As it has been defined, $p_{\mathrm{T}}^{\mathrm{rel}}$ can be used to measure the p^* in the laboratory frame: it follows that its range and peak is also determined by the mass of the parent quark. However, the discrimination power of the muon $p_{\mathrm{T}}^{\mathrm{rel}}$ decreases with the jet p_{T} as muons emerging from the decay of energetic charm and light hadrons can more frequently reach high $p_{\mathrm{T}}^{\mathrm{rel}}$ values.
In the $p_{\mathrm{T}}^{\mathrm{rel}}$ calibration templates for bottom, charm and light jets are derived and the normalisation of these templates is then fitted to data. The fit determines the fraction of b-jets passing, *tagged*, and failing, *untagged*, the b-tagging criterion and hence the

b-tagging efficiency ε_b as follows:

$$\varepsilon_b^{data} = \frac{N_b^{tagged}}{N_b^{tagged} + N_b^{untagged}}. \tag{4.1}$$

Here $N_b^{tagged} = f_b^{tagged} \cdot N_{jets}^{tagged}$ where f_b^{tagged} is the fraction of tagged b-jets fitted in the selected sample of tagged jets. The same holds for untagged jets, that is for jets which explicitly do not pass the b-tagging requirement
The b-tagging efficiency in data is then compared to the efficiency predicted by the simulation, defined by counting directly the number of tagged matched to a b-hadron (usually referred to as *truth-level* information)

$$\varepsilon_b^{MC} = \frac{N_{truth-b}^{tagged}}{N_{truth-b}}. \tag{4.2}$$

It is customary to define a *scale factor* k_b as:

$$k_b = \frac{\varepsilon_b^{data}}{\varepsilon_b^{MC}}. \tag{4.3}$$

In this way the MC simulation, used to extract the physics result of interest, is re-scaled by the efficiency extracted from data.
Because in general one can expect a dependency of the efficiency on the kinematics of the b-jets, such scale factors are derived in bins of jet p_T. The bins used in this calibration are 20-30 GeV, 30-40 GeV, 40-50 GeV, 50-70 GeV, 70-90 GeV, 90-110 GeV, 110-140 GeV, 140-170 GeV and 170-200 GeV.

4.2.2 Data and simulated samples

The calibration here presented is performed in a sample of multi-jets events enriched in $g \to b\bar{b}$. The dataset and simulated samples used are described in the following.

Data events selection

This calibration uses a set of data collected in 2016 and 2017, corresponding to the accelerator conditions described in Section 3.1.1. Only a fraction of these data is used for the physics analysis, consistent to periods in which all the sub-detectors were fully functional. These events are included in a Good Run List (GRL): about the 10% of the events in 2016 and 2017 do not satisfy the GRL.
Data events are selected requiring that the event satisfy a muon-in-jet triggers and then applying kinematic cuts on jets and muons. Muon-in-jet triggers require a muon spatially matched to a jet, with a varying p_T threshold. The spatial matching is made within a cone of $\Delta R < 0.5$ around the jet axis. A $|\Delta z| < 2mm$ is also required, where $|\Delta z|$ is the distance of the closest approach along the beam axis of the muon track to the primary vertex. These muon-in-jet triggers are part of the second-level software trigger system and they are supported by first-level hardware muon and jet triggers.

Due to the steeply falling jet p_T spectrum and the large cross-section of multi-jet events these triggers are pre-scaled: the pre-scale factors used depends on the jet and muon p_T and decreases as the jet p_T threshold gets higher. This means that only a fraction of the events fulfilling the muon-in-jet trigger requirement is recorded.

Jet p_T Bin [GeV]	Jet p_T Threshold [GeV]	Muon p_T Threshold [GeV]	Effective Luminosity [pb^{-1}]	
			2016	2017
[20,30], [30,40]	15	4	4.66	3.61
[40,50]	25	4	5.35	4.32
[50,70]	35	4	6.83	6.64
[70,90], [90,110]	55	4	40.18	16.60
[110,140]	85	6	272.97	173.92
[140,170]	110	6	500.27	382.62
[170,200]	150	6	1492.73	1249.31

TABLE 4.1: Muon-in-jet triggers used in the different jet p_T bins of the calibration. The effective luminosity collected in 2016 and 2017 data has been calculated taking into account the pre-scale factors.

To maximize the event yield, multiple muon-in-jet triggers are used with each one defining an exclusive jet p_T bin. The choice of the jet p_T binning in this work is thus motived by the available triggers. Table 4.1 illustrates which triggers are used in the different bins together with the effective luminosity collected for 2016 and 2017 data.

Simulated samples

This calibration makes use of simulated samples to derive p_T^{rel} template of bottom and charm jets together with the prediction of the jet flavours and b-tagging efficiency. To reproduce the pileup conditions of the real data taking, two simulations are produced for 2016 and 2017 where the distribution of $< \mu >$ is re-weighted to match the data distributions, presented in Figure 3.3a and 3.3b of Section 3.1.1 for the two years respectively.
Two multi-jets simulations are used, both generated by PYTHIA8 [81, 82] using the A14 tune with the NNPDF23LO set of PDFs [83]. The EVTGEN simulation package [84] is used to simulate the b-hadron production fractions and decay. The simulation of the detector response is handled by the GEANT4 toolkit [85].
Different samples are generated depending on the jet p_T range: this provides sufficient statistic across the entire p_T spectrum.
The so-called *muon filtered* simulation requires the presence of a muon having $p_T >$ 4 GeV at the generator level. The muon filtering requirement changes the relative amount of bottom, charm and light jets in the simulation. In particular, this simulation is enriched in bottom and charm jets as muons inside a light jet are generated at a later stage in the simulation chain and do not pass the generator level filter. Given the predominance of bottom and charm jets, this simulation is used to the derive p_T^{rel} templates for these flavour. The range in p_T of each muon filtered sample is reported in Table 4.2 together with the available number of events in the two simulations, which are referred to the 2016 and 2017 data taking years.

No muon-filter is applied in the so-called *unfiltered* simulation, which is thus used to determine the relative fraction of jet flavours. Table 4.3 summarises the available samples, their range in jet p_T and the number of generated events in the two simulation referred to 2016 and 2017, respectively.

Sample name	Jet p_T range [GeV]	Events (2016)	Events (2017)
JZ1WA	$[20-40]$	3577323	4041700
JZRW1B	$[40-60]$	4272861	4489927
JZRW2	$[60-160]$	1830464	1881318
JZRW3	$[160-400]$	1907006	1929887

TABLE 4.2: Muon filtered samples divided by the range in p_T of the simulated jets. The number of events is reported for the two simulations, divided by the 2016 and 2017 data taking years.

Sample name	Jet p_T Range [GeV]	Events (2016)	Events (2017)
JZ1W	$[20-60]$	5339106	8802080
JZ2W	$[60-160]$	8350535	10645935
JZ3W	$[160-400]$	10757379	12273564

TABLE 4.3: Muon unfiltered samples divided by the range in p_T of the simulated jets. The number of events is reported for the two simulations, divided by the 2016 and 2017 data taking years.

4.2.3 Objects selection and events categorisation

Objects selection

The jets used in this calibration are reconstructed as has been described in Section 3.3.2 using the anti-k_t algorithm with R = 0.4. The jets must pass the p_T cut of $p_T > 20$ GeV and the $|\eta| < 2.5$ cut.
Muons have to satisfy the $|\eta| < 2.5$ and $p_T > 5$ GeV cuts and the Tight quality working point (defined in Section 3.3.3). A cut on the transverse impact parameter is applied $|d_0| < 2\ mm$ together with a requirement on the longitudinal impact parameter cut of $|z_0 sin\theta| < 4\ mm$. These cuts ensure that the muon is consistent with the production at the primary vertex and the posterior displaced decay of a b-hadron.
The association of muons to jets is based on a $\Delta R(muon, jet) < 0.4$ requirement and on the transverse momentum of the two objects. All the selections applied to jets and muons are listed in Table 4.4.

Jet Selections	$p_T > 20$ GeV $\|\eta\| \leq 2.5$ JVT < 0.4 (if $p_T^{Jet} < 60$ GeV)
Muon Selections	$\|\eta\| \leq 2.5$ $\|d_0\| < 2$ mm $\|z_0 \cdot \sin\theta\| < 4$ mm Tight Quality
Muon to jet matching	$\Delta R(muon, jet) < 0.4$ $p_T^{Muon} > 5$ GeV (if $p_T^{Jet} < 40$ GeV) $p_T^{Muon} > 6$ GeV (if $40 < p_T^{Jet} < 90$ GeV) $p_T^{Muon} > 8$ GeV (if $p_T^{Jet} > 90$ GeV)

TABLE 4.4: Kinematic and quality selections applied to jets and muon together with the matching requirements between the two.

The muon p_T threshold is varied from 5 to 8 GeV as the jet p_T increases; this requirement has been optimised to suppress the contribution of muons produced by light hadrons decay. Muons generated by the decay of pions and kaons contained in a *light*-jet can fly inside the detector, mimicking the behaviour of a muon merging from the semileptonic decay of a *b*-hadrons: the two cases become more and more similar as the jet p_T increases. However, the p_T spectrum of muons produced in the decay of a *b*-hadron is harder in the entire jet p_T spectrum: tightening the muon p_T threshold keeps the bottom fraction around the same value in the different jet p_T bins.

Regions definition

Two event regions are used in this calibration:

- the *bottom-enhanced* region: enriched in bottom jets, this sample of events is the one used to extract ε_b ;
- the *light-enhanced* region: given the big portion of *light*-jets this region is used to derive the *light*-jets template from data.

In *pp* collisions *b*-quarks are mostly produced in pairs from a gluon splitting, with a cross section of ≈ 0.5 mb. It follows that, if one jet in the event is identified as a *b*-jets, this enhances the probability that there is another *b*-jets in the event. For this reason, the *b-enhanced* region is defined requiring at least one jet tagged by MV2 at the 85% working point, hereafter called *tag* jet. At this point:

- if multiple candidate *b*-jets are identified, the *tag* jet is the one without a matched muon;
- if multiple jets without a matched muon are tagged, or if all tagged jets contain muons, the *tag* jet is randomly chosen;
- the *tag* jet is then removed from the sample of jets used to extract the ε_b, as it would contribute only in the tagged jets sample distorting the ε_b computation.

The *light-enhanced* region is defined requiring that all jets in the event fail the MV2 85% working point cut. The resulting set of events is thus enriched in *light*-jets and statistically independent of the events used to measure the *b*-tagging efficiency. This region has a low contamination of bottom and charm jets that needs to be taken into account during the fitting procedure (more details in Section 4.2.4).

4.2.4 The p_T^{rel} fitting method

The extraction of the *b*-tagging efficiency in data events is performed in the *bottom-enhanced* region: following the Equation 4.1 two quantities need to be determined: N_b^{tagged} and $N_b^{untagged}$.
A template fit method is applied to the p_T^{rel} data distribution to extract the fraction of bottom jets respectively in the sample of jets passing and failing the MV2 77% working point cut.
As the calibration is performed in bins of jet p_T, different templates are derived independently for the various bins.

Construction of templates

Templates for the $p_{\mathrm{T}}^{\mathrm{rel}}$ spectrum of heavy flavour (bottom and charm jets) are generated from the muon-filtered simulation, while a template for light jets is derived using data events.

Bottom template: The b-jets template is derived from jets truth-labeled as bottom and illustrated in Figure 4.6 for some intervals of low, medium and high p_{T} respectively. Given the different shapes in the various jet p_{T} bins, templates for b-jets are

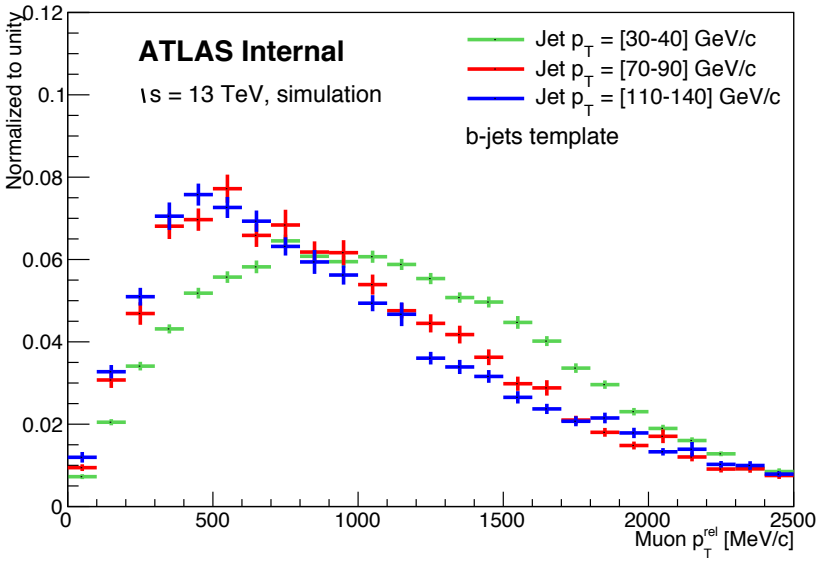
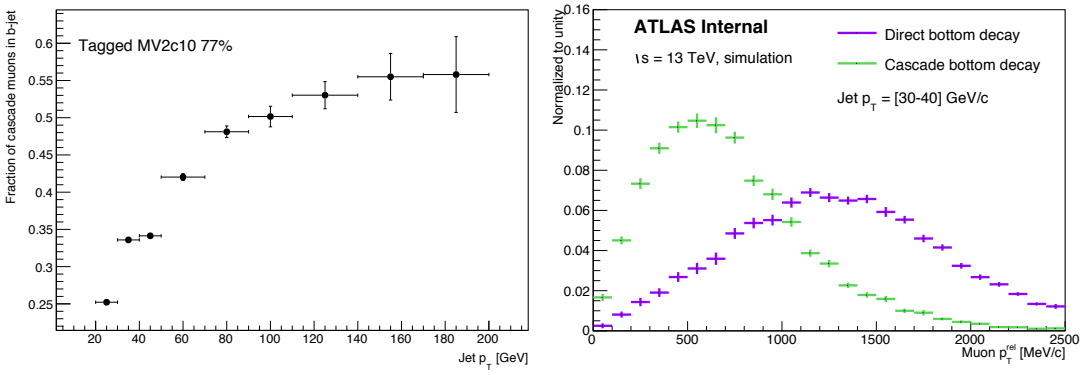


FIGURE 4.6: Templates of b-jets derived in the muon-filtered simulation for jet having p_{T} = [30-40] GeV (green), p_{T} = [70-90] GeV (red) and p_{T} = [110-140] GeV (blue).

derived for each jet p_{T} bin exclusively. The cause for the different shapes seen here is the amount of cascade and direct decays. Bottom hadrons can produce muons from a semileptonic decay in two ways:

- from a *direct* decay, in which the muon is directly produced from the semileptonic decay of the b-hadron, implying $b \to \mu\nu + X$. In this case the muon $p_{\mathrm{T}}^{\mathrm{rel}}$ reaches its maximum discriminating power as it is proportional to the bottom quark mass;
- from a *sequential* or *cascade* decay, where the muon is produced from the semileptonic decay of a charm-hadron generated by the b-hadron decay, implying $b \to c \to \mu\nu + X$. Here the muon has been produced by the decay of a charm hadron, thus the muon $p_{\mathrm{T}}^{\mathrm{rel}}$ discriminating power is diluted.

The fraction of cascade decays in Figure 4.7a shows an increasing trend as function of the jet p_{T}. The distribution of the muon $p_{\mathrm{T}}^{\mathrm{rel}}$ for the cascade and direct decay (Figure 4.7b) reflects the different decay properties explained above: the $p_{\mathrm{T}}^{\mathrm{rel}}$ spectrum in the direct decay case is harder than the cascade case. It follows that the increase of cascade muons in the high jet p_{T} range causes the bottom template to shrink to lower values.



(A) (B)

FIGURE 4.7: Templates of b-jets derived in the muon-filtered simulation for muons produced in a direct decay (violet) and muons produced in a cascade decay (green). The jet p_T considered is p_T = [30-40] GeV (A). Fraction of muons from a cascade bottom decay matched to a tagged b-jets as function of its p_T (B).

There are two factors affecting the amount of direct and cascade decays in the different jet p_T slices: both of them are related to the energy carried away by muons in the decay. On average, the transverse momentum of muon generated in a direct bottom decay is bigger than in the cascade case, as shown in Figure 4.8. It follows

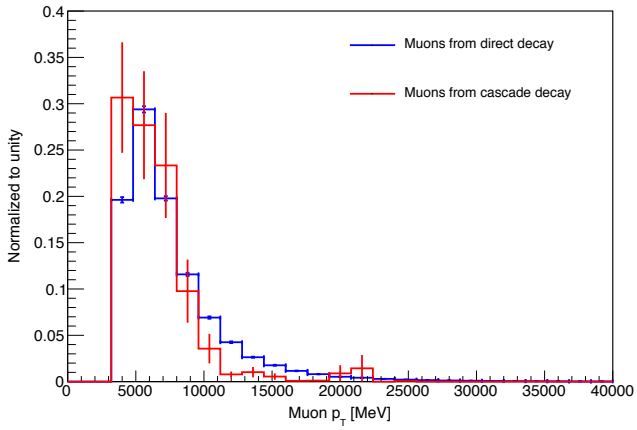


FIGURE 4.8: Transverse momentum of muons produced in direct (blue) and cascade (red) decays of b-hadrons. The distributions are produced from events of the filtered MC samples.

that muons from a direct decay are more likely to pass the p_T requirement, while muons from a cascade decay are more suppressed in the low jet p_T region. With the increase of the b-hadron momentum, the muons emerging from a cascade decay will more easily pass the p_T requirement. This explains the predominance of muons from cascade in the low-jet p_T range.
In addition to this effect, it has to be considered that jets containing a b-hadron decaying directly will appear with a smaller p_T than jets containing a b-hadron of the same energy but decaying through a cascade. This is due to the fact that the muon

momentum is not summed to the jet. Therefore, for a b-hadron of the same energy, the jet initiated by a directly decaying b-hadron can fall in lower jet p_{T} than a jet from a sequential decay, resulting in a bigger contribution of the direct decays in the first jet p_{T} slices.
Given the impact of the direct and cascade decay fraction in the b-jets template, a systematic uncertainty is evaluated in order to account for any mis-modeling in the simulation.

Charm template: The charm-jet template is derived using jet truth matched to a charm hadron in the muon-filtered simulation. The c-jets template is illustrated for a low, medium and high jet p_{T} category in Figure 4.9.
The differences in the three templates can be ascribed to statistical fluctuations in

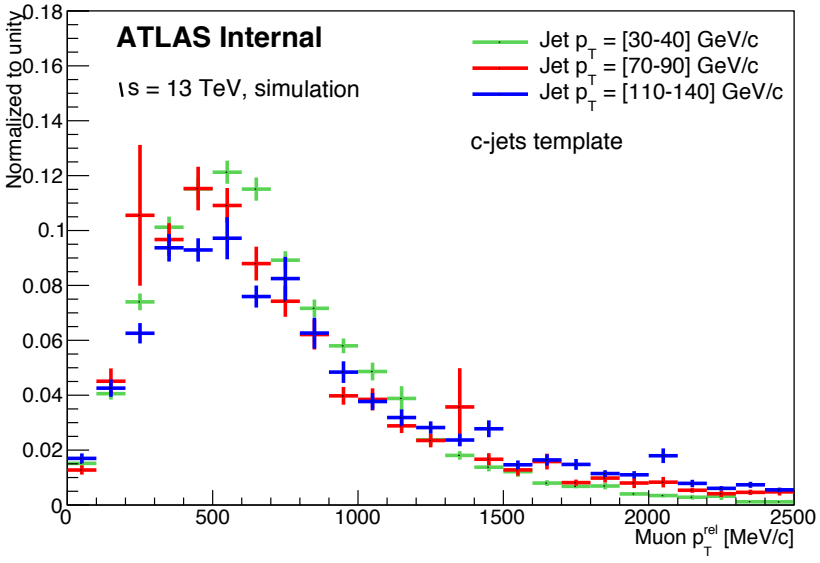


FIGURE 4.9: Templates of charm-jets derived in the muon-filtered simulation in jet p_{T} slices p_{T} = [30-40] GeV (green), p_{T} = [70-90] GeV (red) and p_{T} = [110-140] GeV (blue).

the muon $p_{\mathrm{T}}^{\mathrm{rel}}$bins. Given the negligible dependency on the jet p_{T}, the c-jets template is derived inclusively in jet p_{T}. A template derived using all the c-jets passing the b-tagging requirement is used to fit on the sample of b-tagged jets in the different p_{T} slices. The same thing is done for jets failing the b-tagging requirement.

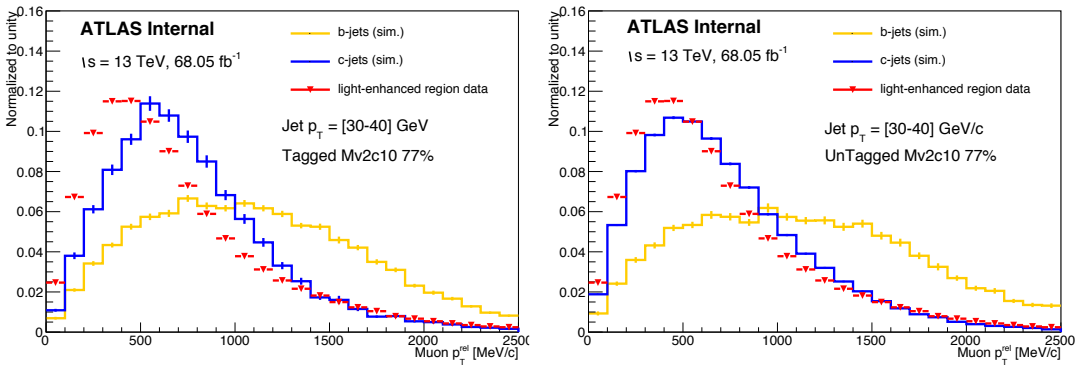
Light template: The *light*-jet template is derived using data. The unfiltered simulation contains too few light jets spatially matched muons while, in the muon-filtered simulation, muons from light-hadrons decay are filtered out.
The *light*-jet template is thus derived from data events in the *light-enhanced* region, where a remaining contamination of bottom and charm jet exists. The fraction of bottom and charm jets are calculated in the unfiltered simulation and shown in Table 4.5. The subtraction of the heavy flavour $p_{\mathrm{T}}^{\mathrm{rel}}$ templates from data in the light-enhanced region results in the *light*-jet template. This correction is performed by the fit and described in the following section.

Jet p_T bin [GeV]	b-jets %	c-jets %
[20, 30]	10.79 ± 0.15	39.05 ± 0.24
[30, 40]	8.33 ± 0.14	37.77 ± 0.24
[40, 50]	8.31 ± 0.17	37.55 ± 0.29
[50, 70]	5.02 ± 0.10	33.38 ± 0.21
[70, 90]	4.21 ± 0.13	33.23 ± 0.31
[90, 110]	5.54 ± 0.30	38.65 ± 0.65
[110, 140]	4.62 ± 0.36	37.26 ± 0.82
[140, 170]	4.53 ± 0.57	35.33 ± 1.31
[170, 200]	4.64 ± 0.87	34.02 ± 1.95

TABLE 4.5: Bottom and charm-jets fractions in the jet p_T bins calculated in the light-enhanced region using the unfiltered simulation.

The fitting strategy

The case of jets having p_T = [30-40] GeV is taken as reference to illustrate the fitting strategy of the p_T^{rel} method. The normalised muon p_T^{rel} distributions of bottom



(A) (B)

FIGURE 4.10: Muon p_T^{rel} spectra for bottom, charm and data in the light-enhanced region jets in yellow, blue and red respectively having p_T = [30-40] GeV. The heavy flavour templates are generated from the simulation using jets passing (A) or failing (B) the b-tag requirement.

and charm jets, together with the data distribution in the light enhanced region, are shown in Figure 4.10a and 4.10b for jets passing and failing the MV2 77% working point respectively. The b-template shows an harder spectrum compared to the charm and *light*-jet templates, while the charm and light jets templates have a similar shape in the p_T^{rel}=[1.5-2.5] GeV range.
Using these p_T^{rel} distributions as prediction, a binned-template likelihood fit [86] is performed to the b-tagged and untagged data distributions in the b-enhanced region. In each p_T^{rel} bin i, the likelihood $\mathcal{L}_i$ for the expectation value λ_i given k_i observed events follows a Poissonian distribution:

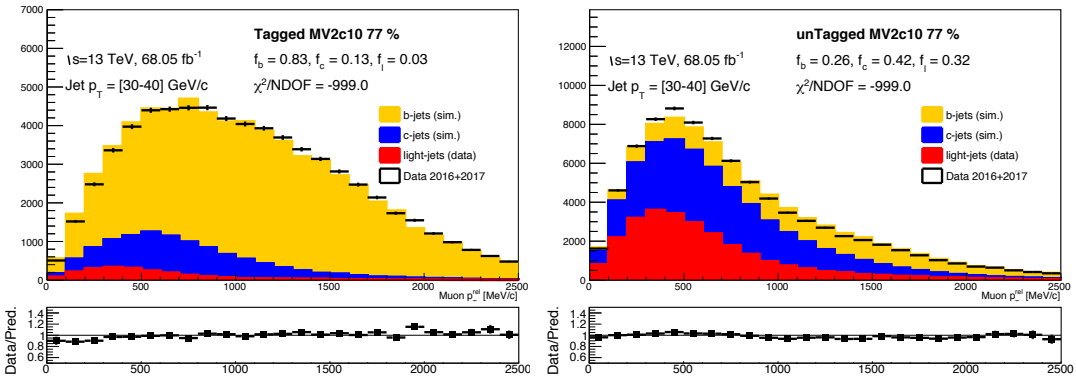
$$\mathcal{L}_i(\lambda_i|k_i) = \frac{\lambda_i^{k_i}}{k_i!} e^{-\lambda_i}. \tag{4.4}$$

Using the natural logarithm the likelihood of multiple Poisson observations becomes a sum over $p_{\mathrm{T}}^{\mathrm{rel}}$ bins i with the log-likelihood

$$\begin{aligned} \log \mathcal{L}_i(\lambda_i(SF_b, SF_c)|N_i^{\mathrm{data}}) &= N_i^{\mathrm{data}} \log \lambda_i(SF_b, SF_c) - \lambda_i(SF_b, SF_c), \\ \text{with} \quad \lambda_i(SF_b, SF_c) &= f_b \cdot SF_b \cdot N_i^b + f_c \cdot SF_c \cdot N_i^c + f_l \cdot N_i^{\mathrm{light}}, \\ \text{where} \quad N_i^{\mathrm{light}} &= N_i^{\text{data-l-enhanced}} - \hat{f}_b \cdot SF_b \cdot N_i^b - \hat{f}_c \cdot SF_c \cdot N_i^c, \\ f_l &= 1 - f_b \cdot SF_b - f_c \cdot SF_c, \end{aligned}$$

where N_i^{data} is the number of observed events, N_i^b is the number of events in the b template, N_i^c is the number of events in the c-template and $N_i^{\text{data-l-enhanced}}$ is the number of data events in the light-enhanced region. The fractions f_b and f_c represent the bottom and charm jets fractions in the b-enhanced region, while $\hat{f}_b$ and $\hat{f}_c$ are the fractions in the light-enhanced region. The free parameters of the fit are SF_b and SF_c, which are correction factors of the flavour fractions estimated by the unfiltered simulation. The fit assumes that the correction of the heavy flavour fractions is the same in the light- and bottom-enhanced region. The overflow bin is not considered during fitting. The long tails in the $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution in data cause the overflow bin to become the most statistically significant bin. This tail, however, is a reconstruction and detector effect; $p_{\mathrm{T}}^{\mathrm{rel}}$ values larger than half of the b-hadron mass are nonphysical, and these effects are not reproduced accurately in MC. Because of this, the $p_{\mathrm{T}}^{\mathrm{rel}}$ range of the templates is confined to 0-2.5 GeV.

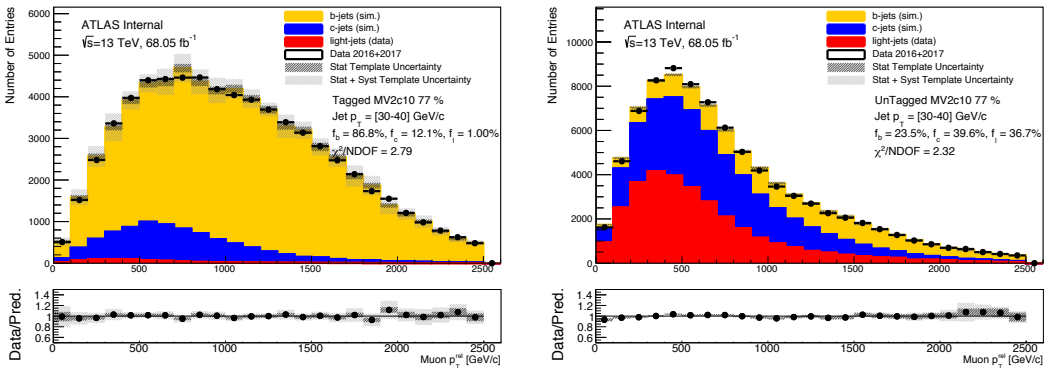
Pre-fit plots for tagged (Figure 4.11a) and untagged (Figure 4.11b) jets are shown where the fractions of the different jet flavours are fixed to the predictions given by the unfiltered simulation and reported in Table 4.6. The fit to the b-tagged $p_{\mathrm{T}}^{\mathrm{rel}}$



(A) (B)

FIGURE 4.11: Pre-fit distribution of b-tagged (A) and untagged (B) jets. The data is shown in black while the templates are shown as a stack. The uncertainty is statistical only.

distribution is shown in Figure 4.12a while the untagged fit result is shown in Figure 4.12b. Here the predicted jet flavour fractions are corrected for the correction factors derived by the fit: the resulting values are reported in Table 4.6.
The fit on the sample of tagged jets gives a bigger f_b than the expected value, while the amount of b-jets failing the b-tagging requirement is found to be smaller if compared to the prediction.
The two fits also estimate the heavy flavour contamination in the light-enhanced region: they give two different values of $\hat{f}_b$ and $\hat{f}_c$ that are reported in Table 4.7.



(A) (B)

FIGURE 4.12: Post-fit distribution of b-tagged (A) and untagged (B) jets. The data is shown in black while the templates are shown as a stack. The uncertainty is the sum of the statistical and systematic contributions.

Tagged	f_b	f_c	f_l
Pre-Fit	84.0±0.4%	13.0±0.3%	3.0±0.5%
Post-Fit	86.9±0.6%	12.0±1.3%	0.97±1.4%
UnTagged	f_b	f_c	f_l
Pre-Fit	26.0±0.5%	42.0±0.6%	32.0±0.8%
Post-Fit	23.5±0.6%	39.7± 3.4%	36.6±3.5%

TABLE 4.6: Pre-Fit and Post-Fit fractions of bottom, charm and light jets are shown together with the uncertainty evaluated by the fit. The flavour fractions are calculated for jets in the b-enhanced regions passing and failing the b-tagging requirement.

The fractions of charm jets in the light-enhanced region estimated by the two fits are

	f_b	f_c
Pre-Fit	8.3±0.2%	37.8±0.3%
Post-Fit (Tagged)	8.7±0.1%	34.3±3.7%
Post-Fit (UnTagged)	7.6±0.2%	35.4±3.0%

TABLE 4.7: Fraction of b-jets and charm jets in the light-enhanced region estimate by the fit on tagged jets (left) and untagged jets (right).

in agreement within the uncertainty, while a small difference in the bottom fraction is found for the two fits. However, this difference can be assigned to the lowest sensitivity of the fit on the tagged sample, in which the fraction of light is estimated to be $\approx 1\%$.
The post-fit plots shown also contain the uncertainty band in all the $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution bins evaluated summing up all the contributions: a description of all the sources of uncertainties is given in Section 4.2.5.
The same procedure is repeated in all the jet p_{T} bins considered in this work, the pre-fit and post-fit plots are reported in Appendix B and C, respectively. The fitted values of b-jets are then used to derive the b-tagging efficiency in data and thus the scale factor curves.

4.2.5 Uncertainties

Uncertainties affecting the b-tagging efficiency determination, and thus the scale factors, originate from statistical and systematic sources. Multiple sources of systematic uncertainties affect the shape of the templates: they are related to the detector calibration and physics modeling.
In general, two kinds of systematics are used. Two-sided systematics are evaluated by considering the up- and down- variations of the parameter under study by its uncertainty. The effect is then propagated to the scale factors. The systematic is then assumed to be symmetric and half the difference between the two variations is taken as a systematic. Single-sided systematics are evaluated by doing a variation and using the full difference to the nominal value as the uncertainty. The total uncertainty is calculated from the sum in quadrature of all uncertainties.
Table 4.8 shows the percentage uncertainty on the scale factor of the MV2 77% working point in the different jet p_{T} slices. The total uncertainty on the scale factors in-

Uncertainty Source	Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}[GeV]$ Bins								
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
MC Stat.	0.8	1.1	1.0	0.9	1.5	1.8	2.0	2.9	2.5
Data Stat.	1.0	0.7	0.7	0.8	0.5	1.0	0.6	0.7	0.3
Detector	0.4	0.4	1.0	0.5	0.8	2.3	3.0	0.6	1.4
Modeling	0.9	1.0	1.0	1.3	2.1	2.3	2.7	3.9	4.2
Total Systematic	1.0	1.1	1.4	1.4	2.2	3.2	4.0	4.0	4.4
Total Uncertainty	1.3	1.5	1.7	1.6	2.7	3.7	4.5	5.0	5.1

TABLE 4.8: Percentage uncertainty on the scale factor in the different jet p_{T} slices. The uncertainties are grouped in those pertaining to the statistical fluctuations, the detector performance and the physics modeling. The total uncertainty is also reported.

creases with the jet p_{T} from the 1.3% of the first bin to the 5.1% of the last bin. In the range of jet p_{T} = [20-70] GeV the uncertainty is dominated by the contribution coming from the systematics while in the p_{T} = [70-200] GeV range the statistical and systematic uncertainties have similar impact. The statistical uncertainty on the data sample has never dominant contribution.

Statistical uncertainties

The three sources of statistical uncertainties are shown in Table 4.9.
The *statistical uncertainty* takes into account the statistical fluctuation of the sample of simulated events used to calculate ε_b^{MC}. This quantity is the denominator of the scale factor, thus its uncertainty is propagated to the scale factor curve.

Stat. Uncert.	Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}[GeV]$ Bins								
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Simulation Stat.	0.2	0.1	0.1	0.1	0.1	0.2	0.1	0.2	0.2
Template Stat.	0.8	1.1	1.0	0.9	1.5	1.8	2.0	2.9	2.5
Data Stat.	1.0	0.7	0.7	0.8	0.5	1.0	0.6	0.7	0.3

TABLE 4.9: Percentage statistical uncertainty on the scale factor in the different jet p_{T} slices

The *template statistical uncertainty* takes into account any change in the $p_{\mathrm{T}}^{\mathrm{rel}}$ templates shape due to the statistical fluctuations. To evaluate this uncertainty, ten thousand pseudo-experiments are performed. In each pseudo-experiment a new template is

generated by extracting the value of the each $p_{\mathrm{T}}^{\mathrm{rel}}$ bin from a Gaussian distribution, whose mean value is the nominal bin content and whose width is the statistical uncertainty of the bin. For each iteration the template fit is repeated, building a distribution of f_b filled ten thousand times: its mean value is used to compute the b-tagging efficiency in data and compared to the nominal value. The difference between these two values is assigned as the template statistical uncertainty. The template statistical uncertainty represents the main contribution to the scale factor uncertainty of the jets having a p_{T} greater than 70 GeV. This is due to the limited statistics available in the JZRW2 simulation: the events generated in this filtered sample are half the events generated in the lower jet p_{T} slices, as reported in Table 4.2. This set of events is implied in the four bins in the jet p_{T} range [70-170] GeV that are thus more affected by the template uncertainty.

The *data statistical uncertainty* is the uncertainty associated to the fraction of b-jets estimated by the fit due to the statistical fluctuations of the data sample.

Detector uncertainties

This calibration takes into account several uncertainties related to the detector response, listed in details in Table 4.10.

Detector Uncert.	Uncertainty in $p_{\mathrm{T}}^{\mathrm{jet}}[GeV]$ Bins								
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
Muon Sagitta RES	<0.1	<0.1	<0.1	<0.1	0.3	0.5	1.4	<0.1	0.1
Muon Sagitta Bias	<0.1	0.1	<0.1	<0.1	<0.1	0.1	<0.1	<0.1	0.2
Muon ID	<0.1	<0.1	<0.1	<0.1	0.1	<0.1	0.2	<0.1	<0.1
Muon MS	<0.1	0.1	0.1	<0.1	0.6	0.2	<0.1	<0.1	0.4
Muon Scale	<0.1	<0.1	<0.1	<0.1	<0.1	0.1	0.1	<0.1	0.9
Jet pileup	0.2	0.2	0.7	0.1	0.3	1.4	1.5	0.4	0.6
Jet JES	0.2	0.2	0.5	0.1	0.2	0.9	1.5	<0.1	0.2
Jet JER	<0.1	0.1	<0.1	<0.1	0.1	1.2	0.8	0.2	0.2
BJES Response	0.2	0.1	0.3	0.1	0.1	0.7	0.3	<0.1	0.1
Flavour Composition	0.2	<0.1	0.3	0.3	0.2	0.3	0.4	0.4	0.1
Flavour Response	<0.1	<0.1	0.2	0.2	<0.1	0.4	1.1	0.3	0.1
JVT Calib	<0.1	<0.1	<0.1	<0.1	<0.1	0.1	<0.1	<0.1	0.6
Total Muon	0.1	0.2	0.2	0.1	0.7	0.5	1.4	0.1	1.0
Total Jet	0.4	0.3	0.9	0.5	0.4	2.2	2.6	0.6	0.8

TABLE 4.10: Percentage detector uncertainty on the scale factor in the different jet p_{T} slices. The different components to the muon and jet uncertainty are shown together with their total contribution.

Uncertainties associated to muons arise from the reconstruction in the Inner Detector and Muon Spectrometer, the identification as well as the momentum scale and resolution of the sagitta. The uncertainties related to jets arise from the estimation of their energy (JES, JER) which depends also on the jet flavour (BJES, Flavour Response and Composition). Also the uncertainty on the JVT cut has been evaluated taking the up and down variation provided by the dedicated calibration of the tagger efficiency.
The detector uncertainty is dominated by the contribution of the calibration of the jet energy scale, resolution and pileup correction.

Modeling uncertainties

The physics modeling uncertainties accounted for in this analysis have been grouped into those pertaining to muons and those to jets. The details of the different components are reported in Table 4.11.

Modeling Uncert.	Uncertainty in $p_T^{jet}[GeV]$ Bins								
	[20, 30]	[30, 40]	[40, 50]	[50, 70]	[70, 90]	[90, 110]	[110, 140]	[140, 170]	[170, 200]
B-Decay Fractions	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	0.1	0.1	0.1
Fake Muons	0.1	0.1	0.1	0.2	0.1	0.4	0.7	<0.1	0.3
Muon p^*	0.1	0.3	0.3	0.8	1.2	0.8	0.9	0.7	0.6
Jet Axis Deter.	0.1	0.1	0.2	0.1	1.3	1.7	0.3	3.7	3.6
Gluon SplitB	<0.1	0.1	<0.1	<0.1	0.3	0.5	1.3	0.4	1.5
Gluon SplitC	0.1	<0.1	0.2	0.2	0.3	0.2	1.7	<0.1	0.2
B-Frag. Fractions	<0.1	<0.1	<0.1	<0.1	0.1	<0.1	0.1	<0.1	<0.1
B-Frag. Function	0.1	0.1	0.1	0.1	0.2	0.3	0.5	0.7	1.0
Extrapolation	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9

TABLE 4.11: Percentage modeling uncertainties on the scale factor in the different jet p_T slices. The different contributions to the modeling of muons and jets are reported separately.

***b*-hadrons decay fractions:** The muons studied in this calibration originate from direct and cascade decays of *b*-hadrons. The branching fraction of *b*-hadrons decaying directly is measured to be $\mathscr{B}(b \to \mu + X) = 10.86 \pm 0.16\%$ [7]: the relative uncertainty associated to the latter measurement is used to perform an up and down variation of the direct decay fraction of jets entering the bottom template.

"Fake" muons: In order to reduce contributions from muons not produced in the decay of a *b*-hadron the *tight* muon reconstruction quality criteria are required. Despite this, misidentified particles or decay-in-flight muons, so-called "fake" muons, can pass the selections affecting the templates. The muon-filtered simulation underestimates the amount of these "fake" muons due to the action of the muon-filter. To estimate their impact heavy flavour templates have been generated using a conservative approach: the fraction of "fake" muons is increased by a factor two and the fit is repeated using the resulting templates. The result is compared to the nominal case and the difference assigned as systematic uncertainty.

Muon p^*: The momentum of the lepton in the *b*-hadron rest-frame p^* has been measured by the BABAR and DELPHI collaborations [87, 88]. The agreement of the MC simulation with current measurements has been tested as shown in Figure 4.13. B-mesons are the most commonly produced *b*-hadrons for the considered selection. A set of B-mesons has been simulated and the p^* spectrum for the final state muons is shown in Figure 4.13 in comparison with the available measurements. It can be observed that the simulation matches the DELPHI result better. However, in order to account for the limited knowledge of the p^* spectrum, a systematic uncertainty is assigned re-weighting the simulation to the BABAR results. In the low p^* region no values are available for the BABAR measurement. Therefore, the weight is calculated for the lowest p^* value available and used for all lower p^* values.

Jet axis determination: The muon p_T^{rel} depends on the knowledge of the relative directions of the muon and the *b*-hadron. While the muon direction is measured with high precision, two aspects influence the knowledge of the *b*-hadron direction.

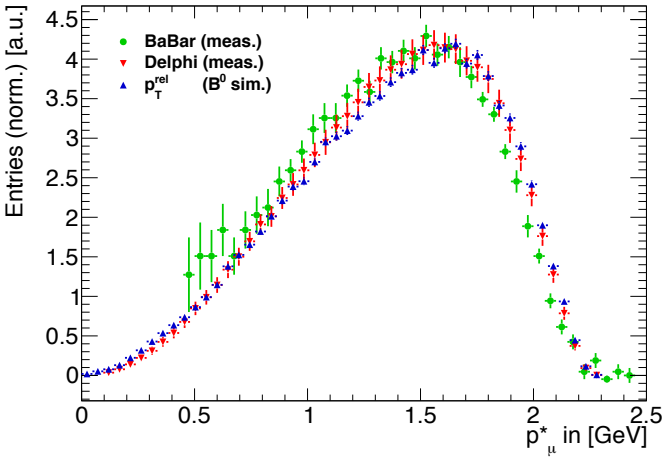
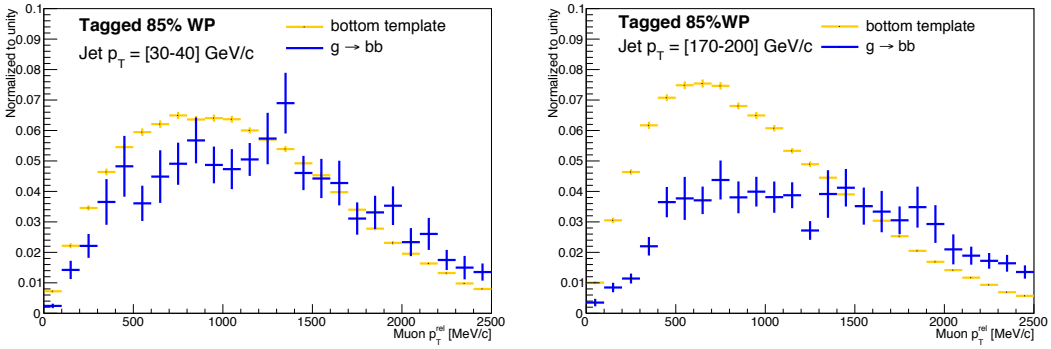


FIGURE 4.13: Final state lepton p^* from B-meson decays generated in simulation as used in the p_T^{rel} measurement shown in blue. In comparison measurements from BABAR and DELPHI are shown in green and red, respectively [87, 88].

First, the difference between the b-hadron flight direction and the reconstructed jet direction is studied using the simulation. Second, the precision of the jet direction reconstruction is estimated by comparing calorimeter and track jets. These systematic effect have been found to be 0.004 and 0.008 in ϕ and η, respectively. In order to estimate a jet angular resolution uncertainty, the jet direction has been varied within a Gaussian distribution with the width set to the respective values. These jets are then used to redo the calibration and the difference with the nominal case is taken as systematic uncertainty.

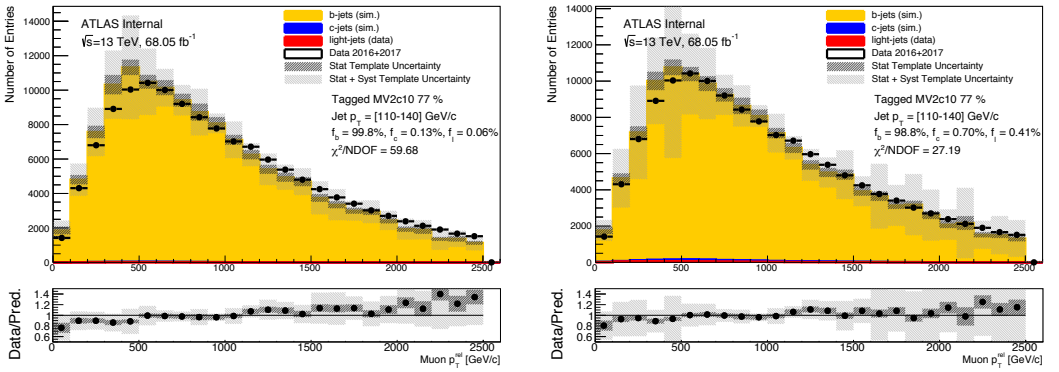
Gluon splitting to $b\bar{b}$ and $c\bar{c}$: Energetic gluon splitting events may result in jets with two or more heavy flavour hadrons inside the jet cone. These jets have a different tagging efficiency, different probability to contain a muon inside and a different muon p_T^{rel} distribution.
In the unfiltered simulation, the fraction of b-jets containing two b-hadrons inside a $\Delta R = 0.4$ cone increases with the jet p_T and goes from 2% of the [20-30] GeV bin to the 20% of the [170-200] GeV bin. The same trend is observed in the charm-jets case, where the fraction of jets containing two c-hadrons goes from 5 to 40 % in the first and last bins respectively.
A dependency of the gluon splitting fraction from the jet momentum is expected: high energy gluons recoiling against other partons produce two hadrons that will be spatially closer, and thus they will be more easily reconstructed as a single jet. On the other hand, two hadrons produced from a low energy gluon splitting event will tend to have opposite directions: as a result they are more easily resolved as two separate jets.
Figures 4.14a and 4.14b show the comparison between the bottom template computed using all the b-jets and the bottom template of b-jets matched to two b-hadrons for the [30-40] GeV and [170-200] GeV jet p_T bins respectively. The muon p_T^{rel} distribution is different in the case of b-jets containing two b-hadrons if compared to the inclusive case: this effect is bigger for high p_T jets. It follows that a larger or



(A) (B)

FIGURE 4.14: Muon $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution of b-jets (yellow) and b-jets matched to two b-hadrons (blue) having p_{T} = [30-40] GeV (A) and [170-200] GeV (B) and tagged at the 85% working point of MV2.

smaller fraction of gluon splitting into two bottom quarks modifies the shape of the overall bottom template: the effect on the fit results is expected to be bigger in the *tagged* sample which is dominated by b-jets. It has been found that increas-



(A) (B)

FIGURE 4.15: Stacked post-fit plots using the nominal b-jets template (A) and the one with an increased $g \to b\bar{b}$ fraction (B). The jets are b-tagged and have a p_{T} = [110-140] GeV. The ratio between data and prediction is shown in the bottom part, which includes the contribution of the systematic uncertainty.

ing the gluon splitting fraction in the bottom template has a negligible effect in the $p_T^{Jet} = [20-70]GeV$ region, while it provides a better agreement of the prediction to data in the region of $p_T^{Jet} = [70-200]GeV$. The post-fit plot for *tagged* jets in the [110-140] GeV p_{T} range in the nominal case (Figure 4.15a) and in the case of a gluon splitting fraction increased by a factor two (Figure 4.15b) are compared. When the fraction of gluon splitting into $b\bar{b}$ is increased, the agreement between data and prediction improves, particularly for high $p_{\mathrm{T}}^{\mathrm{rel}}$ values. This can be interpreted as an indication that the predicted fractions of gluon splitting is underestimated by the simulation.

An uncertainty taking into account for this effect is evaluated by with a two-sided variation where the gluon splitting fraction for bottom and charm jets is increased and lowered by a factor of two.

***b*-quark fragmentation function:** During the fragmentation process the energy of the initial b-quarks is transferred to the generated b-hadron. The energy fraction X_b that is transferred to the b-hadron is altered by 5% in an up-down variation to estimate the systematic uncertainty due to the limited knowledge of the hadronisation process.

***b*-quark fragmentation fraction:** The results of the $p_{\mathrm{T}}^{\mathrm{rel}}$ calibration also depends on the modeling of the b-quark fragmentation as it modifies the energy of the final state particles. Therefore, any mis-modeling in the fractions of hadrons ($B^0, B^{\pm}, B_S^0, b-baryons$) produced during the fragmentation as well as their functions needs to be quantified. To evaluate a systematic uncertainty on the b-fragmentation fractions

Hadron Fraction	Z^0-decays (MC)	Tevatron	HFLAV-combination
B^+ fraction	0.407 ± 0.007	0.344 ± 0.021	0.405 ± 0.006
B^0 fraction	0.407 ± 0.007	0.344 ± 0.021	0.405 ± 0.006
B_s^+ fraction	0.101 ± 0.008	0.115 ± 0.013	0.103 ± 0.005
b-baryon fraction	0.085 ± 0.011	0.198 ± 0.046	0.088 ± 0.012

TABLE 4.12: Fractions of b-hadrons as produced in simulated Z boson decays, inclusive measurement performed at CDF and compared to the average provided by HFLAV group [89].

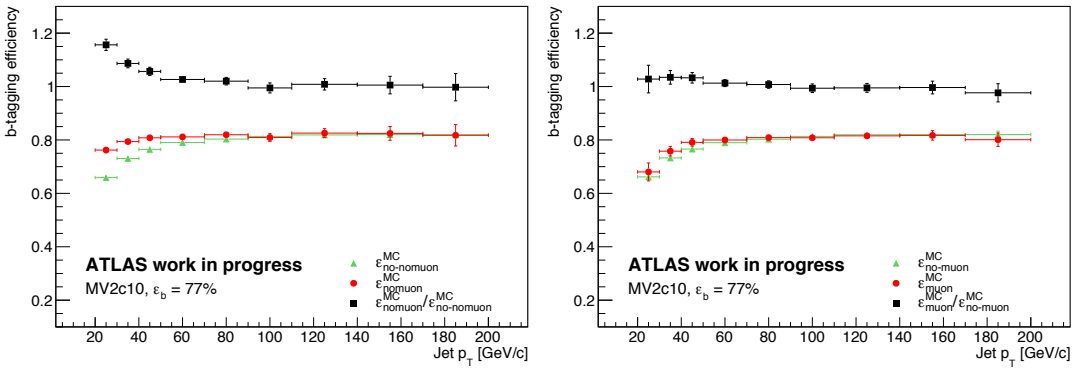
the events have been re-weighted to match the flavour fractions obtained combining Tevatron and LEP measurements [89] reported in Table 4.12. The fractions of b-hadrons calculated using the unfiltered simulation are reported in Table 4.13 show a smaller amount of b-baryons in favor of b-mesons which are more abundant if compared to the measured value previously reported.

b-hadron	Fraction
B^+ fraction	0.423 ± 0.001
B^0 fraction	0.432 ± 0.001
B_s^+ fraction	0.116 ± 0.001
b-baryon fraction	0.028 ± 0.001

TABLE 4.13: Fractions of b-hadrons calculated using the unfiltered simulation. The uncertainty is statistical coming from the available simulated sample.

Semileptonic to inclusive extrapolation: The p_T^{rel} method provides the b-tagging efficiency calibration for jets initiated by semileptonic decays of b-hadrons. It is not immediate that the scale factor derived with the p_T^{rel} method can be applied to an inclusive sample. The differences between the semileptonic and hadronic cases stand in the intrinsic properties of the decay and in the way jets are experimentally selected. Concerning the intrinsic properties, jet initiated by semileptonically decaying b-hadrons differ from the inclusive b-jets in the number of charged particles generated in the secondary vertex. Then semileptonic jets have to satisfy the request of containing a high momentum and well-reconstructed muon.
The extrapolation to inclusive b-jets has been studied in order to test the compatibility and assign a systematic uncertainty to account for any discrepancy observed.

This study has been performed also in the Run I analysis, resulting in a flat 4% value for this systematic uncertainty [41]. This was the leading source of uncertainty of the Run I result.
It is observed (Figure 4.16a) that low momentum hadronic b-jets have a 15% loss of b-tagging efficiency if compared to the semileptonic case; the two efficiencies smoothly become the same after jet $p_T > 60$ GeV. The reason lies in the mis-reconstruction of



(A) (B)

FIGURE 4.16: Simulated b-tagging efficiency as function of the jet transverse momentum of jets containing muons (red curve), jet not containing muons (green curve) and their ratio (black curve). The semileptonic correction to the jet momentum is applied in (B).

low momentum particles from the secondary vertex that causes the b-tagging algorithm to fail the identification. It is shown in Figure 4.16b how the ratio between the two efficiencies becomes closer to unity when a correction on the momentum of semileptonic jets is applied. The correction consists in adding the muon momentum to the jet momentum while, at the same time, removing the muon calorimeter energy deposits. The corrected jet represents a better proxy for the b-hadron. However, a residual mismatch is present because of the momentum carried away by the neutrinos.
Despite these considerations, the scale factor can still be employed also in the inclusive case, if the simulation reproduces the efficiency adequately for the two cases.
In order to check this and estimate the semileptonic-to-inclusive extrapolation uncertainty, b-tagging efficiency scale factors have been evaluated separately for jets containing a muon in a $\Delta R < 0.4$ and for jets failing this requirement. The estimation of these scale factors has been performed using the tag-and-probe technique in a high purity sample of b-jets of dileptonic $t\bar{t}$ events. A total integrated luminosity of $36.1 fb^{-1}$ of data collected during 2015 and 2016 has been used for this purpose.
The following selections are applied to define sample for the study of the extrapolation:

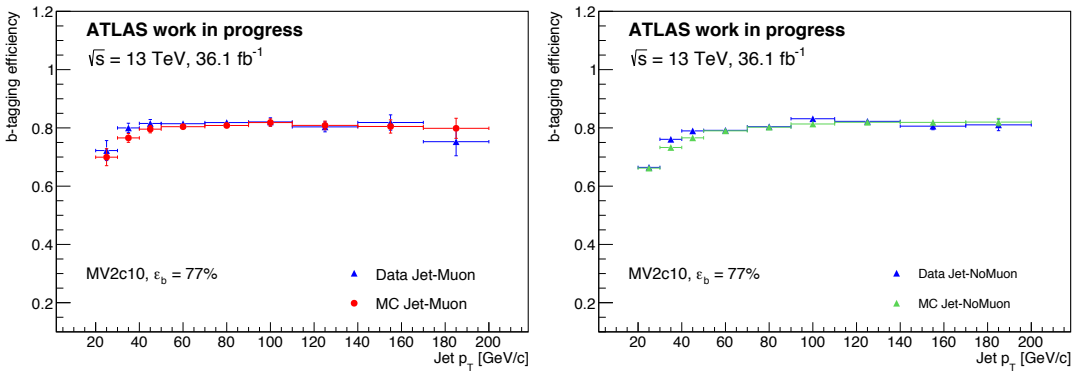
- one prompt electron of $p_T > 25$ GeV and $|\eta| < 2.47$, excluding the $1.37 < |\eta| < 1.57$ region;
- one prompt muon of $p_T > 25$ GeV and $|\eta| < 2.4$;
- opposite electric charge between the prompt electron and the prompt muon;
- two jets of $p_T > 20$ GeV and $|\eta| < 2.5$;
- at least one b-tagged jet at the 85% working point of the MV2c10 tagger.

It is estimated from the Monte Carlo simulation that these selections retain a 91% of $t\bar{t} \to e\mu + 2j$ in the signal region. The main background processes are the single-top and diboson productions, contributing in the signal region with the 8% of events. The production of W and Z bosons in association to jets and to top-quark pairs represent a 1% contribution. A good agreement between data and simulation is found in all the kinematical distributions.
In the tag-and-probe method the 85% b-tagged jet is fixed as tag while the other jet is used as probe to measure the b-tagging efficiency. In the case in which the two jets in the event are b-tagged at the 85%, then both of them are used as probe.
The b-tagging efficiency is calculated in data events using Equation (4.6). As a first step subtracting the contribution of the background Equation (4.5) and then correcting by the number of probe jets coming from b-quarks and the mis-tag rate estimated in the $t\bar{t}$ simulation.

$$\varepsilon_b^{uncorr} = \frac{N_{data}^{pass} - N_{bkg}^{pass}}{N_{data} - N_{bkg}}, \tag{4.5}$$

$$\varepsilon_b^{data} = \frac{\varepsilon_b^{uncorr} - (1 - f_b) \cdot \varepsilon_q}{f_b}. \tag{4.6}$$

The resulting scale factors in both the semileptonic and hadronic case are found to be consistent with unity in all the jet p_T bins and so is their ratio. Figure 4.17a shows the comparison of b-tagging efficiency for data and simulated events.
Figure 4.18 shows the ratio between the scale factors for jets with or without muons



(A) (B)

FIGURE 4.17: b-tagging efficiency in data and simulation as function of the jet transverse momentum, for jet containing muons (A) and jet without muons (B)

for the 77% working point: its flatness can ben interpreted as the goodness of the simulation to reproduce the different topologies of semileptonic and hadron decays of all b-hadrons. The distribution of the scale factors ratio as function of the jet p_T has been fitted using a constant function. The result of the fit is compatible with unity: the uncertainty on the constant value from the fit is 0.94% and it is taken as the semileptonic-to-inclusive extrapolation uncertainty.
The extrapolation uncertainty is the leading modeling uncertainty in the [20-70] GeV jet p_{T} range: the reduction from the 4% to the 1% of this uncertainty provides a correspondingly linear gain for the R_{b} measurement. About the 93% of the jets produced in $t\bar{t}$ events have a p_{T} = [20-140] GeV.

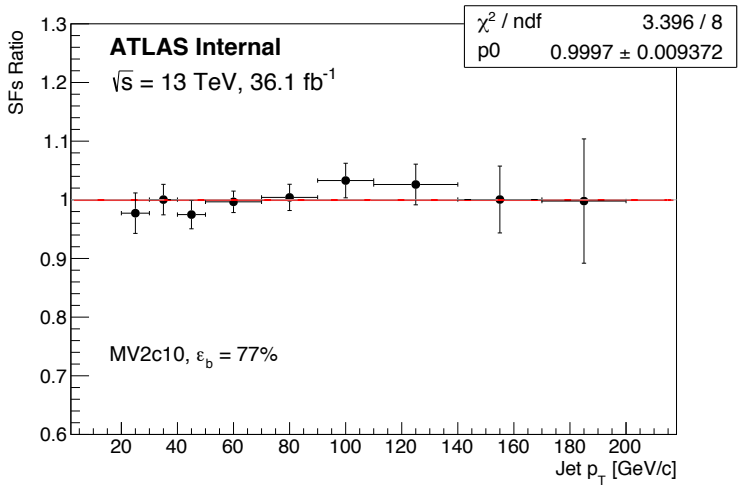


FIGURE 4.18: Ratio between the b-tagging efficiency scale factor of jet with and without muons for the 77% working point as function of jet p_T bins. The uncertainty is statistical only.

As the extrapolation uncertainty is derived using events in which b-jets are obtained from the decay of top quarks, one could wonder whether a recursive dependence of the measured R_b value is introduced. The ratio between inclusive and semi-muonic scale factors depends on the fraction of b-jets, as per Equation 4.7. This, in turn, depends on R_b. This is set to 1 in the MC simulation used, while it will be whatever nature decides in data. That notwithstanding, the entity of the parameters in Equation 4.7 is such that the dependence on r from R_b is very small.

$$r = \frac{k^{\mu}}{k^{\not\mu}} = \left(\frac{\varepsilon_b^{uncorr-\mu} - (1 - f_b^{\mu}) \cdot \varepsilon_q^{\mu} f_b^{\not\mu}}{\varepsilon_b^{uncorr-\not\mu} - (1 - f_b^{\not\mu}) \cdot \varepsilon_q^{\not\mu} f_b^{\mu}} \right) \times \frac{\varepsilon_b^{MC-\not\mu}}{\varepsilon_b^{MC-\mu}}. \tag{4.7}$$

Here the expression of the scale factors k_{μ} for jets matched to muons and $k_{\not\mu}$ for jets no associated to muons are written explicitly. As a result, two terms are found to be dependent from f_b:

- the one having the structure $(1 - f_b) \times \varepsilon_q$;
- the ratio $f_b^{\not\mu} / f_b^{\mu}$.

The first term contributes to the correction with a factor smaller than 1% in both the case of jets matched or not to muons as shown in Figure 4.19. This is due to the multiplication by the mis-tag rate, which is of the order of 0.001. Concerning the second term, the fraction of b-jets generated in the a top quark decay is calculated and shown in Figure 4.20 for jets matched or not a muon. The biggest difference between the two curve appears in the first bin, in which the ratio of the two fractions is different from one. For jets having a p_T greater than 30 GeV the two curves get closer, thus the ratio of two fractions cancel out.
For this reason, in the R_b measurement only jet satisfying $p_T > 30$ GeV are used.

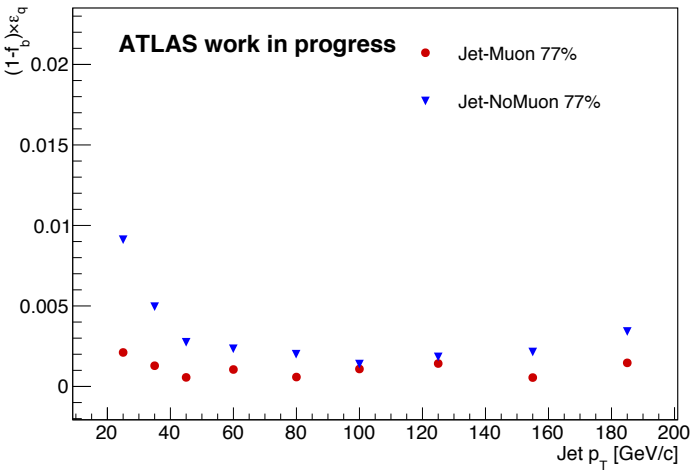


FIGURE 4.19: Value of $(1 - f_b) \times \varepsilon_q$ in the case of jets associated to muons (red) and jets not associated to muons (blue) as function of the jet p_{T} as calculated form the $t\bar{t}$ simulation.

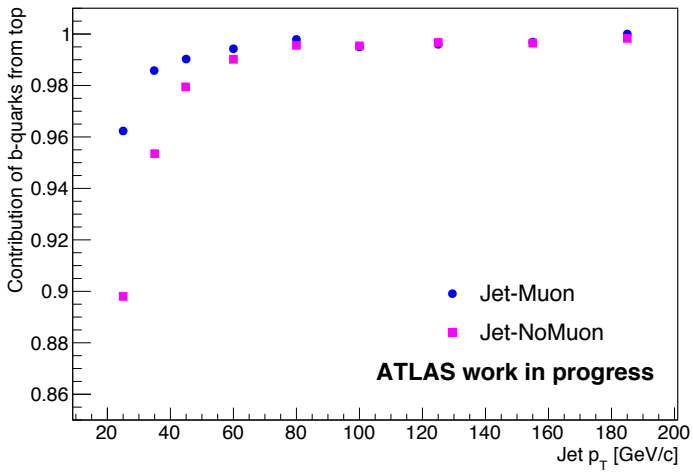


FIGURE 4.20: Fraction of b-jets coming from a top quark decay as function of jet p_{T}. Blue curve is relate to jets matched to muons while the magenta curve is

4.2.6 Results

Figure 4.21 and 4.22 show respectively the b-tagging efficiency in data and simulated events and the resulting scale factor as function of the jet transverse momentum. Both statistical and systematic uncertainties are shown.

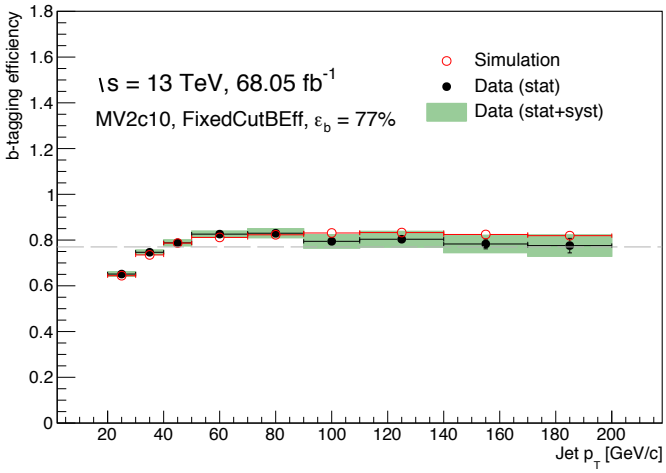


FIGURE 4.21: Efficiency of b-tagging in data (black dots) and simulated (red dots) events as function of the jet p_{T}. The green band shows the total uncertainty calculated as the sum of the statistical and systematics uncertainties.

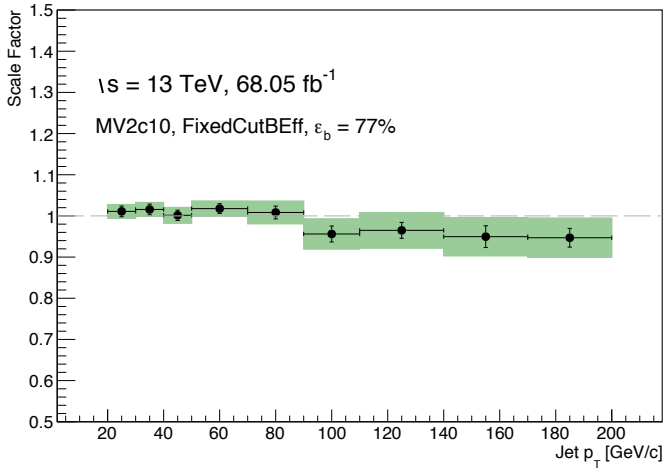


FIGURE 4.22: Scale factor of the b-tagging efficiency as function of the jet p_{T}. The green band shows the total uncertainty calculated as the sum of the statistical and systematics uncertainties.

The b-tagging efficiency in data and simulation are in agreement within the uncertainty band: as a result the scale factor is compatible with one in all the jet p_{T} spectrum. A change in the trend of the scale factor for jets having $p_{\mathrm{T}} > 70$ GeV appears: this is the result of the poor modeling of the heavy flavour templates due to a limited simulated statistics. In this region also the uncertainty is bigger, if compared to the

lower jet p_{T} slices, as an effect of the statistical fluctuations of templates derived using the MC, which affects also the detector systematic uncertainty which are derived using the same samples.

4.2.7 Application of scale factors to R_{b} analysis

The scale factor of the *b*-tagging efficiency derived in this calibration are used to correct the simulated sample implied in the R_{b} measurement. In particular, for each b-tagged jet in the simulation a weight is applied to match the tagging rate as measured in data. The weight is jet-flavour dependent, it means a different weight is applied to bottom, charm and light-flavour jets using the same procedure.
The calibration described in this work is applied to *b*-jets. If the jet is tagged at 77% of MV2c10 the weight w_{jet} is simply the scale factor itself:

$$w_{jet} = k_b(p_T), \tag{4.8}$$

where $SF(p_T)$ is the *b*-tagging efficiency scale factor evaluated at a give p_{T}. If the jet is not tagged the weight becomes:

$$w_{jet} = \frac{1 - \varepsilon_b^{data}(p_T)}{1 - \varepsilon_b^{MC}(p_T)} = \frac{1 - \varepsilon_b^{MC}(p_T) \cdot k_b(p_T)}{1 - \varepsilon_b^{MC}(p_T)}. \tag{4.9}$$

Every simulated event used in the R_{b} measurement is weighted by an event weight computed as the product of all jet weights.

Chapter 5

Measurement of top quark R_b

The top quark branching fractions ratio, R_b, is defined as

$$R_b := \frac{\mathscr{B}(t \to Wb)}{\mathscr{B}(t \to Wq)} = \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2}. \tag{5.1}$$

This chapter describes the measurement of R_b using proton-proton collisions data delivered by the LHC at a center of mass energy of 13 TeV and collected by the ATLAS detector during 2016 and 2017.
Details about the data and simulated samples are given in Section 5.1.
The analysis makes use of the b-tagging efficiency scale factors derived using the p_T^{rel} calibration method and reported in Section 4.2.6. R_b is extracted performing a likelihood fit on the distribution of b-tagged jets multiplicity in a sample of $t\bar{t}$ dileptonic events. The selection of the objects and events is explained in Section 5.2, followed by the description of the event categorization. The latter is performed by a kinematically reconstruction of the top quark pairs, as reported in Section 5.3. The event categorization allows to extract the information on the association of the jet to the top quark decay: this is one of the ingredients needed in the R_b extraction (Section 5.4). A description of the uncertainties is given in Section 5.5 while the final results is shown in Section 5.8.

5.1 Data and simulated samples

5.1.1 Data sample

This analysis uses a set of data collected in 2016 and 2017, corresponding to the accelerator conditions described in Section 3.1.1. Only a fraction of these data is used for the physics analysis, corresponding to periods in which all the sub-detectors were fully functional. These events are included in a Good Run List (GRL): about 10% of the events in 2016 and 2017 do not satisfy the GRL.
The dataset corresponds to an integrated luminosity of $32.9 \pm 0.7\ fb^{-1}$ recorded in 2016 and $44.3 \pm 1.1\ fb^{-1}$ recorded in 2017. The uncertainty in the 2016 and 2017 integrated luminosity is 2.2% and 2.4% respectively [90], obtained using the LUCID-2 detector [91] for the primary luminosity measurements.
The R_b measurement makes use of events in the GRL where the detector read-out is triggered by the presence of one electron or one muon, referred to as single-lepton triggers. In these analysis the $t\bar{t}$ dilepton final state is probed, therefore the aim of the triggers is to select events in which at least one lepton produced in a W boson decay is present, namely *prompt* leptons. Different settings of the triggers are used to rejects events having *fake* and *non-prompt* leptons. A "fake" lepton is a signal, particle or a jet mis-identified as a lepton. A "non-prompt" lepton is a real lepton generated

by the decay of heavy hadrons, a photon conversion or the decay in flight of kaons and pions.
The lepton p_T threshold required by the triggers is 26 GeV. Leptons triggers with

Trigger	p_T [GeV]	Isolation	Efficiency 2016 [%]	Efficiency 2017 [%]
Electron	26	Loose	91	93
Electron	60	No	43	44
Electron	140	No	4	4
Muon	26	Medium	81	81
Muon	50	No	51	52

TABLE 5.1: Single lepton triggers used in the analysis for the 2016 and 2017: the lepton flavour together with the p_T threshold, the isolation requirement and the percentage efficiency for $t\bar{t}$ di-leptonic $e\mu + 2jets$ events are reported in the different columns.

a low-p_T threshold of 26 GeV also include a requirement on the lepton isolation in order to reject the contribution of fake and non-prompt leptons. For the high-p_T thresholds this background is less significant, therefore the isolation requirement is dropped to increase the trigger efficiency. A summary of the triggers used is given in Table 5.1: the events used in the measurement are required to pass the logical OR of these triggers. The table contains the efficiency of the trigger, calculated as the fraction of simulated $t\bar{t}$ events in the $e\mu + 2jets$ final state passing the trigger requirement.

5.1.2 Simulation

Events simulated with MC event generators are used in this measurement. To reproduce the pileup conditions of the real data taking, two simulations are produced for 2016 and 2017 where the distribution of $< \mu >$ is re-weighted to match the data distributions, presented in Figure 3.3a and 3.3b of Section 3.1.1 for the two years respectively. Thanks to this re-weighting, variables sensitive to the pileup conditions, particularly the calorimeter based ones, like the jet-related variables, are better modeled in the simulation.
A sample of signal events is used to predict the fraction of events with jets emerging from a top quark decay, to estimate the b-tagging efficiency of b-jets and the b-tagging mis-tag rates of charm and light jets. The simulations of background processes are used to estimate their contributions in the different event regions of this analysis.

Signal samples

Nominal sample: The nominal $t\bar{t}$ MC simulations is generated using POWHEG-BOX V2 [92–94] a parton-level event generator based on NLO QCD calculations. The matrix-element calculation uses the NNPDF3.0NLO PDF set with a top-quark mass of 172.5 GeV. For the simulation of the parton shower, fragmentation and the underlying event, the matrix-element generatos is interfaced to PYTHIA 8 [81]. The A14 [82] tune together with the *NNPDF2.3LO PDF* set [83] is applied for the PYTHIA 8 showering. The h_{damp} parameter is set to $1.5m_{top}$. The events are normalised to the NNLO cross-section including the resummation of soft-gluon emissions at the next-to-next-to-leading-logarithmic (NNLL) accuracy using the TOP++2.0 program [95].

In this sample $|V_{tb}|$ is fixed to one, it means the simulation only allows the decay of the top quark to a bottom quark in association to a W boson. However, the kinematics of the top quark decay is not dependent on the flavour of the quark produced. It follows that this simulation can be used to derive the prediction for the b-tagging efficiency of b-jets produced a top quark decays and for the mis-tag rates of charm and light jets: these values already encode the corrections derived by the *b*-tagging algorithm calibrations and they are used in the likelihood fit.

Parton and hadronisation varied sample: In order to derive the uncertainty on the parton shower and hadronisation model, an alternative $t\bar{t}$ sample is used in this analysis. It is generated with POWHEG-BOX, having the same h_{damp} setting as the nominal sample, and it is interfaced with HERWIG 7 [96, 97] using the H7UE set of tuned parameters, and the MMHT2014LO PDF set [98].

Background samples

Single top samples: Three different simulations are produced for the s-channel, t-channel and tW-channel single-top quark processes. All of them are generated using POWHEG-BOX-v2 generator interfaced to PYTHIA 8 using the A14 tune. The PDF set used in the parton shower NNPDF2.3LO. Due to the overlap with the $t\bar{t}$ process (see Section 1.5.1), two tW-channel samples are produced: one using the diagram-removal technique, which is choose to be the nominal sample, the other one using the diagram subtraction technique [28]. All single-top quark samples are normalised to the cross-section predicted by the calculation at NLO in QCD with NNLL soft gluon correction [99, 100].

Diboson samples: The production of two vector bosons (WW, ZZ, WZ), *diboson*, is generated using SHERPA 2.2.2 with the PDF set NNPDF3.0NNLO. The samples are normalised to the NLO in QCD theoretical cross-sections [101].

V+jets samples: Events with a vector boson produced with additional jets ($W + jets, Z + jets$) are simulated with SHERPA 2.2.1 generator. The matrix element generates up to two jets at NLO [102] while higher jet multiplicities are included at LO. The NNPDF3.0NLO PDF set is used. The samples are normalized to the NNLO cross-section [103].

$t\bar{t}$+X samples: The associated production of $t\bar{t}$ with a vector boson ($t\bar{t}V$) or Higgs boson ($t\bar{t}H$) is simulated using the MADGRAPH5_*aMC@NLO* generator at NLO [104] interfaced to PYTHIA 8 with the A14 tune. The PDF set is NNPDF3.0NLO for the ME. The $t\bar{t}V$ and $t\bar{t}H$ samples are normalised to the NLO in QCD+EW cross-section calculation [105, 106].

5.2 Object and event selection

The object and event selection described in this section is applied to both the data events and the simulated samples.

5.2.1 Object selection

Electrons: Electrons are identified with a likelihood method (see Section 3.3.2) and are required to pass the TIGHT working point selection. Only electrons with a calorimeter cluster having $|\eta| < 2.47$ are considered and the transition region between barrel and end-cap $1.37 < |\eta| < 1.52$ is vetoed. In addition, impact parameter criteria $|d_0/\sigma(d_0)| < 5$ and $|z_0 sin\theta| < 0.5$ mm are applied. Isolation requirements with 90% efficiency on additional energy in the surrounding calorimeter cells within $\Delta R < 0.2$ are applied to the electron candidate. The sum of the transverse momenta of additional tracks within a cone of $\Delta R < 0.3$ is required.

Muons: Muons used in this analysis are Combined Muons, which are reconstructed combining the information from the Inner Detector and the Muon Spectrometer (See Section 3.3.3). Muons are required to have $|\eta| < 2.5$ and to pass track-quality requirements. The impact parameters must satisfy $|d_0/\sigma(d_0)| < 3$ and $|z_0 sin\theta| < 0.5\ mm$. A tighter requirement on the transverse impact parameter significance is needed to reject muons from the semileptonic decays of b-hadrons. Only isolated muons having calorimeter cluster energy in $\Delta R < 0.2$, $E^{0.2}_{cluster}$ divided by the muon p_{T}, $E^{0.2}_{cluster}/p^{\mu}_{T} < 0.15$ are used. An additional isolation cut is applied on the ratio between the summed transverse momenta of addition tracks within a cone of $\Delta R < 0.3$ over the muon p_{T} which has to be smaller than 0.04.

Hadronic jets: Hadronic jets are reconstructed from calorimetric clusters using the topological anti-k_T algorithm (see Section 3.3.4) with a radius of $R = 0.4$. Jets candidates are required to have p_{T} greater than 30 GeV and $|\eta| < 2.5$. To suppress jets originating from pile-up collisions, cuts on the Jet Vertex Tagger discriminant are applied.

Overlap Removal: In order to avoid double counting of the same energy clusters or tracks as different objects types, an overlap removal procedure is applied. First electron candidates sharing a track with any muon candidate are removed. Secondly, if the distance between a jet and an electron candidate is $\Delta R < 0.2$, the jet is removed. If multiple jets are found with this requirement, only the closest jet is removed. If the distance between a jet and an electron candidate is $0.2 < \Delta R < 0.4$, the electron candidate is removed. If the distance between a jet and any muon candidates is $\Delta R < 0.4$, the muon candidate is removed if the jet has more than two associated tracks, otherwise the jet is removed.

5.2.2 Event selection

Events in the dilepton $t\bar{t}$ channel are required to have exactly two reconstructed leptons of opposite electric charge. In this work only events having one electron and one muon are selected. A single lepton trigger is required in both data and simulation, as described in Section 5.1.1. For each applicable trigger, scale factors are applied to the simulation in order to correct for known differences in trigger efficiencies between the simulation and collision data [107–109]. Events with exactly two jets are selected. The variable H_T is defined as the scalar sum of the objects p_{T}. A cut on this variable is used, requiring $H_T > 120\ GeV$ to reject $Z \to \tau\tau$ events.

Moreover, a cut on the invariant mass of the two leptons is applied $M_{\ell\ell} < 15\ GeV$: this allows to remove events with two leptons originated by a low mass hadronic resonance, are found to be inaccurately modeled in the simulation. Expected and observed event yields obtained after applying the above selections are shown in Table 5.2. Here the contribution of $t\bar{t}$ events is expected to be the 88.7%, the leading

Sample	Yield ± Stat. Uncert.	% Contribution
Top pairs	175293 ± 90	88.7
Single-top	11318 ± 42	5.7
Diboson	5744 ± 13	2.9
$V + jets$	2533 ± 105	1.3
$t\bar{t}X$	2151 ± 31	1.1
Fakes leptons	545 ±6	0.28
Total Prediction	197583 ±148	
Data	199130	

TABLE 5.2: Events yields obtained after the event selection with their statistical uncertainties. The production of vector boson in association to jets is combined in the single V+jets category. $t\bar{t}V$ and $t\bar{t}H$ are combined into $t\bar{t}X$. In the last column the percentage contribution of the different processes is shown.

background processes is the single-top production in the tW-channel with the 5.7%. The total yield of the diboson production represents the 2.9%, while the production of a single vector boson in association to jets is the 1.3%. The 97% of the latter contribution is composed of $Z \to \tau\tau + jets \to e\mu + jets$ events. The associated production of the Z boson to a top quark pairs is the bigger part of the $t\bar{t} + X$ category, which is 1.1%. The smallest background process is the *fake leptons* background.
The ratio between the data and the total prediction is 1.00783± 0.00076, where the uncertainty is coming from simulation statistics. Considering also systematic uncertainties coming from the signal modeling and detector calibration, which typically are around few percents, the predicted yields are in good agreement with data.

5.3 Top pairs kinematical reconstruction

In the sample of dileptonic $t\bar{t}$ events selected in this analysis, there is a non-negligible probability that at least one of the two jets produced in the top decays is either missed because it was not reconstructed or it did not pass the selection criteria. The probability to have at least one jet not from top in the event is 37%, as estimated from the $t\bar{t}$ simulation. In this scenario, a jet could be selected which comes from the emission of gluons from initial or final state particles (ISR or FSR).
In order to measure R_b from the jets generated in top quarks decay we proceed in two steps:

- define a region enriched in events having both the selected jets coming from a top quark decay;
- estimate on data the remaining fraction of events in which the jets have been generated from other sources.

Both these points are possible by performing a categorisation of the events based on a kinematical reconstruction of the top quark pairs.

In dileptonic $t\bar{t}$ events two jets, two leptons and two neutrinos are produced in the top quarks decay. As neutrinos are not detected, it is not possible to fully reconstruct the top quark decay imposing the conservation of the momentum. However, the kinematic of the lepton and the jet produced in a top decay is bound to the finite mass of the top quark and the W boson. In particular, when the lepton and the jet are generated in the decay of the same top quark, the distribution of their invariant mass shows an endpoint at $\sqrt{m_{top}^2 - m_W^2} \approx 153\ GeV$. If such kinematical constraint

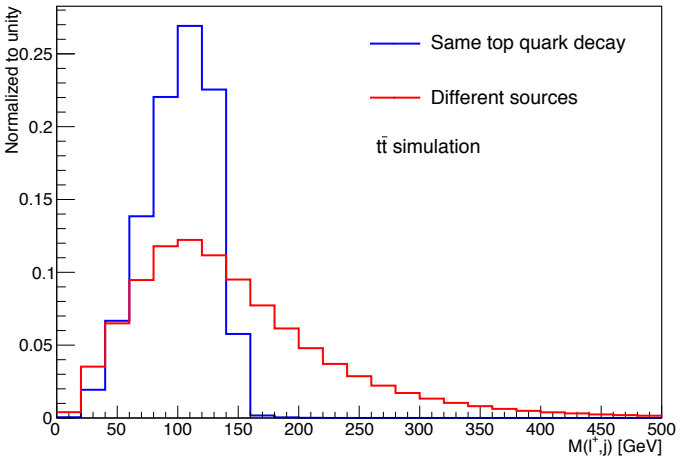


FIGURE 5.1: Invariant mass of the lepton having positive electric charge and one of the two jets in the $t\bar{t}$ simulation when: both are coming from the same top quark decay (blue), the lepton and the jet are not coming from the decay of the same top quark (red).

is not applicable, i.e. if the lepton and the jet are not emerging from the same top quark decay, their invariant mass can reach higher values, instead.
Figure 5.1 shows the invariant mass of the lepton having positive electric charge and one of the two jets in the event: the blue curve is filled if the lepton and the jet are coming from the same top quark decay, while the red curve is filled pairing the lepton with the jet not generated by the same top quark. The jet is associated to the top quark decay by spatially matching it to the bottom quark emerging from the top quark decay vertex stored in the truth record of the $t\bar{t}$ simulation within a $\Delta R < 0.4$ cone.
In the blue curve the invariant mass distribution reaches higher values up to 170 GeV due to resolution effects on the jet momentum estimation.
It follows that the region $M(\ell, jet) > 175\ GeV$ is mainly dominated by events in which the lepton and the jet are not correctly paired.

Stemming from such considerations, an event categorisation is used to characterise and estimate the various contributions in the top decay sample considered exploiting the behaviour of the $M(\ell, j)$ distribution:

- events having both jets coming from a top quark decay are labeled *top-top*;
- events with one selected jet from the top quark decay are labeled *top-notop* if the jet is also the leading one in p_{T}, and *notop-top* if the jet is the sub-leading one in p_{T} ;

- If both jets are not coming from top quarks, the event is labeled *notop-notop*.

The ratio between these $t\bar{t}$ categories and the totality of $t\bar{t}$ events is defined *jet-to-top assignment* fraction, α_i, where the index-i indicates the number of jets from the top quark decay. These fractions are calculate in $t\bar{t}$ simulated events as follows:

$$\begin{aligned} \alpha_2 &= \frac{N_{t\bar{t}}(\text{top-top})}{N_{t\bar{t}}}, \\ \alpha_1(leading) &= \frac{N_{t\bar{t}}(\text{notop-top})}{N_{t\bar{t}}}, \\ \alpha_1(subleading) &= \frac{N_{t\bar{t}}(\text{top-notop})}{N_{t\bar{t}}}, \\ \alpha_0 &= \frac{N_{t\bar{t}}(\text{notop-notop})}{N_{t\bar{t}}}, \end{aligned} \tag{5.2}$$

where by construction $\sum_{i=0}^{2} \alpha_i = 1$. Here the α_1 category is split in two: the case in which the jet not coming from the top decay is the leading in jet p_{T} and the case in which it is the sub-leading.
The different $t\bar{t}$ categories in the simulation are defined exploiting the fact that no light or charm jet is produced in the top quark decay ($R_b = 1$ in the MC). It follows that if a jet is not a bottom at truth level, then it has been generated by ISR or FSR activity and classified as *notop*. This approach implicitly assumes that the amount of b-jets generate by ISR/FSR gluons can be neglected. This assumption has been cross-checked looking at the fraction of b-jets in a $t\bar{t}$ simulation having $R_b = 0.5$: it has been found that the fraction of b-jets in this simulation is exactly half the amount of b-jets generated in the nominal simulation.
This allows to conclude that, in the $t\bar{t}$ enriched events region defined by this analysis, the fraction of b-jets from ISR/FSR can be neglected.

A criticism has been raised that, being this cross-check based on the MC simulation, its validity is limited to the correct modelling of the additional heavy flavour production rate in the simulations. The assumption is as good as the agreement between the observed and the predicted amount of b-jets produced in association with $t\bar{t}$ exists. A dedicated measurement of the inclusive and differential $t\bar{t}$ +bb cross-section is available, that has been performed by ATLAS using the partial Run 2 dataset [110]. Here two final states are probed: the dileptonic and semileptonic $t\bar{t}$. The measurement is performed in different bins of the number of b-jets in the event, comparing data with different $t\bar{t}$ +bb simulations. As shown in Figure 8 of this article, where the distribution of number of b-jets is presented for the dileptonic channel, a good agreement between the data and the prediction is found for the 2 b-jets bin. This is an indication that the simulation is well modeling the events in the phase space of this work, allowing us to conclude that the bias introduced by assuming a neglicible amount of ISR/FSR b-jets is small.

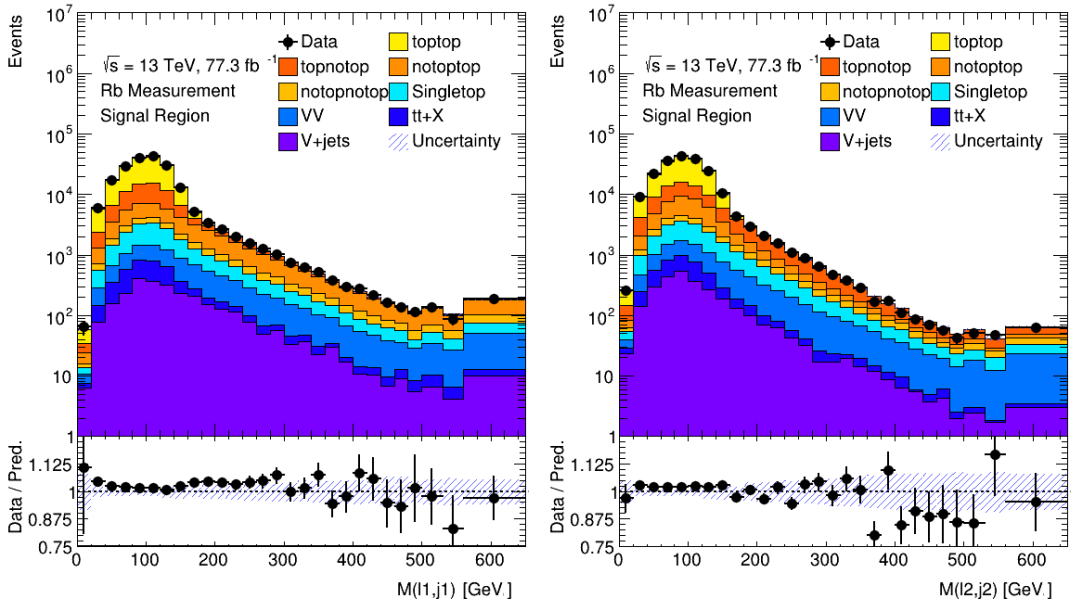
In order to classify the sample in *top-top*, *top-notop*, *notop-top* and *notop-notop* enriched regions, each of the two leptons is paired with a jet in an exclusive way to determine whether they come from the same top quark decay.
The pairing is performed minimising the following quantity:

$$M^2(\ell_1, j_1) + M^2(\ell_2, j_2), \tag{5.3}$$

where j_1 and j_2 are the leading and sub-leading jet respectively, ordered by their p_T, while ℓ_1 and ℓ_2 are the two leptons. The $M(\ell_1, j_1)$ is the invariant mass of the system including the leading jet and its associated leptons. In the same way one can define $M(\ell_2, j_2)$.
Choosing the pair that minimises this quantity relies on the fact that, if pairs of objects from the same original particles are correctly applied, then they should have similar masses. Using the minimum of the squared masses penalises asymmetric pairings with one high mass lepton-jet pair, which should be forbidden for those arising from top decay.

5.3.1 Event classification

The data distributions of $M(\ell_1, j_1)$ and $M(\ell_2, j_2)$ are shown in Figure 5.2a and 5.2b where the background events are displayed together with the $t\bar{t}$ events. The latter events are split in the four different categories *top-top*, *top-notop*, *notop-top* and *notop-notop*.



(A) (B)

FIGURE 5.2: Distribution of $M(\ell_1, j1)$ (A) and $M(\ell_2, j2)$ (B). In different shades of orange are reported the $t\bar{t}$ categories while background events are respectively the V+jets (violet), tt+X (dark blue), VV (azure) and single top (cyan). The prediction histograms are stacked and compared to the data distribution. The ratio plot shows the fraction of the data over the total prediction with the uncertainty band containing statistical and systematic uncertainties.

In the case of $t\bar{t}$ events with two jets from the top quark, both $M(\ell_1, j_1)$ and $M(\ell_2, j_2)$ have an upper limit at ≈ 170 *GeV*. This is not the case for *top-notop*, *notop-top* and *notop-notop* $t\bar{t}$ categories which result also in high $M(\ell_1, j_1)$ and $M(\ell_2, j_2)$ values. The leading $t\bar{t}$ contribution in the $M(\ell_1, j_1) > 175$ *GeV* is the *notop-top* category, while the majority of $t\bar{t}$ events in the $M(\ell_2, j_2) > 175$ *GeV* region fall in the *top-notop* class. It can be concluded that the two invariant masses discriminate between the different $t\bar{t}$ categories.

Thanks to the discrimination power of the $M(\ell_1, j_1)$ and $M(\ell_2, j_2)$ distributions, the selected events are classified into four different regions by cutting on these two variable. One Signal Region (SR) and three Control Regions (CR) are defined as follows:

- **SR(top,top)**: this region is defined cutting on $M(\ell_1, j_1) < 175\ GeV$ and $M(\ell_2, j_2) < 175\ GeV$ and it is enhanced in the $t\bar{t}$ *top-top* events;
- **CR(top,notop)**: this region is defined cutting on $M(\ell_1, j_1) < 175\ GeV$ and $M(\ell_2, j_2) > 175\ GeV$ and it is enhanced in the $t\bar{t}$ *top-notop* events;
- **CR(notop,top)**: this region is defined cutting on $M(\ell_1, j_1) > 175\ GeV$ and $M(\ell_2, j_2) < 175\ GeV$ and it is enhanced in the $t\bar{t}$ *notop-top* events;
- **CR(notop,notop)**: this region is defined cutting on $M(\ell_1, j_1) > 175\ GeV$ and $M(\ell_2, j_2) > 175\ GeV$ and it has a mixed contribution of *top-notop*, *notop-top* and *notop-notop*.

The definition of the regions is summarised in Table 5.3 together with the relevant dominant $t\bar{t}$ category and the percentage contribution of $t\bar{t}$ events.

Region	$M(\ell_1, j_1)$ cut	$M(\ell_2, j_2)$ cut	Dominant Category	N($t\bar{t}$)/N(all)
SR(top,top)	<175 GeV	<175 GeV	top-top	92%
CR(top,notop)	<175 GeV	>175 GeV	top-notop	73%
CR(notop,top)	>175 GeV	<175 GeV	notop-top	71%
CR(notop,notop)	>175 GeV	>175 GeV	notop-top top-notop notop-notop	54%

TABLE 5.3: Definition of the analysis regions based on the cut on $M(\ell_1, j_1)$ and $M(\ell_2, j_2)$ together with the contribution of the dominant $t\bar{t}$ categories.

Given the large contribution of events having both the jets associated to the top quark decay, the events in the SR(top,top) are used to extract R_b. The fraction of $t\bar{t}$ events pertaining to the other categories, approximately the 29%, are corrected using the information extracted from the CRs. The remaining 8% of events in the SR(top,top) is due to the backgrounds processes: their contributions are extracted from the simulation.
Among the three CRs, the CR(top,notop) has the biggest contribution of its relevant $t\bar{t}$ category: this suggests that between the two jets in the event the leading one is more likely to come from the top decay. Moreover, none of the $t\bar{t}$ categories shows a percentage larger than 23% in the CR(notop,notop). However, thanks to an iterative approach in which the correction factors are derived and applied to the different CRs, it is still possible to derive a correction of the different $t\bar{t}$ categories.

5.4 Statistical model for R_b extraction

Following the model presented in Section 2.2, R_b is extracted from the distribution of the b-tagged jets multiplicity: this model provides a description of each b-jet multiplicity bin as function of R_b, encoding also a dependence on the b-tagging efficiency of b-jets ε_b, the b-tagging mis-tag rate of charm and light jets η_q and the jet-to-top

assignment fractions α_i.

The statistical analysis developed for this measurement implies the usage of two profile likelihood fits [86] run simultaneously. From the first correction factors to the α_i fractions are derived in the three CRs: the procedure is shown in Section 5.4.1. The second, using the input correction factors from the CR fit, returns the parameter of interest R_b from the b-tagged jets multiplicity distribution in the SR(top,top) region, as reported in Section 5.4.2.
The results presented in this work are obtained using a minimisation procedure as implemented in the MINUIT2 package of the ROOFIT framework [111].

The uncertainty associated to the α_i correction factors is propagated to the uncertainty on R_b in the context of the profiling procedure. The effect of the systematic uncertainties on R_b is included in the statistical model as Gaussian-constrained *nuisance parameters* (NP), as described in Section 5.4.3. In the following, $\boldsymbol{\theta}$ is used for the set of nuisance parameters $(\theta_i, ..., \theta_n)$.
Due to the strong correlation between the b-tagging efficiency and R_b it was not possible to introduce the systematic variations of ε_b as nuisance parameters of the fit. The introduction of the ε_b related NPs prevents the fit from converging. Therefore, in order to evaluate the impact of the knowledge of ε_b on R_b, the fitting procedure has been repeated varying the value of ε_b within its uncertainty. The same stands for the mis-tag rate η_q.

5.4.1 Correction of jet-to-top assignment fractions

The jet-to-top assignment fractions α_i are input parameters of the model providing the expectation value of the b-tagged jets multiplicity bins: they are therefore used in the likelihood fit extracting R_b as shown in Section 5.4.2. The prediction for the α_i fractions is derived using the $t\bar{t}$ simulation, following Equation 5.2. The starting values of the assignment probabilities in the most relevant region (SR) are reported in Table 5.4.

Events Region	α_2	$\alpha_1(leading)$	$\alpha_1(subleading)$	α_0
SR(top,top)	68.6±0.1	9.8±0.1	19.4±0.2	2.2±0.1

TABLE 5.4: Predicted values of the α_i jet-to-top assignment fractions using the $t\bar{t}$ simulation in the SR(top,top) regions. Uncertainty is MC statistical only.

In order to obtain the α_i actually present in collision data, correction factors to such MC values are derived from the fit. Each CR provides a value of the correction factor relative to the dominant $t\bar{t}$ category, coherent to what is reported in Table 5.3. To this end, a binned-template profile likelihood fit is perfomed on the $M(\ell_1, j_1)$ and $M(\ell_2, j_2)$ distributions. The fits happen simultaneously in the three control regions, allowing to constrain the $t\bar{t}$ categories, even for the cases when their contribution is the CR does not exceed the 50%.
Correction factors μ_{α_i} for the $t\bar{t}$ categories are defined as the ratio between data events and the predicted yields. The likelihood in the different CR dedicated to

the α_i is the following:

$$\mathscr{L}_{\alpha_i}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{nbins} \mathscr{P}[N_j^{exp}(\mu_{\alpha_i}, \boldsymbol{\theta}), N_j^{data}] \cdot \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\theta_t^2/2}, \tag{5.4}$$

where the index j goes from the first to the last bin of the $M(\ell_1, j_1)$ and $M(\ell_2, j_2)$ distributions. The bin content in data N_j^{data} is compared to the expected value $N_j^{exp}(\mu_{\alpha_i}, \boldsymbol{\theta})$ following a Poissonian distribution $\mathscr{P}$. The contribution of the nuisance parameters $\boldsymbol{\theta}$ is expressed using a gaussian distribution. The expected number of events in each bin j is defined as:

$$N_j^{exp}(\mu, \boldsymbol{\theta}) = \mu_{\alpha_i} \cdot N_{j,\alpha_i}^{exp}(\boldsymbol{\theta}) + \sum_g N_{j,g}^{exp}(\boldsymbol{\theta}),$$

where the expected number of events of the $t\bar{t}$ category under exam $N_{j,\alpha_i}^{exp}(\boldsymbol{\theta})$ is multiplied by the correction factor μ_{α_i} while the other processes contributing are summed and indicated with the *g-index*.

The contribution of the "mixed" cases (*notop-top* and *top-notop*) is big even in the fourth region, CR(notop-notop). Therefore, these two contributions are scaled before considering them in the fit of the CR(notop-notop) to extract μ_{α_0}.

5.4.2 Extraction of R_b

The distribution of b-tagged jets multiplicity in SR(top,top) is employed in a profile likelihood fit to estimate R_b. For each bin k of the number of b-tagged jets distribution the number of data events N_k^{data} are compared to the expected number $N_k^{exp}(R_b, \varepsilon_b, \eta_q, \alpha_i, \theta)$. In the following $\boldsymbol{\beta}$ is used to indicated the parameters of the fit $(\varepsilon_b, \eta_q, \alpha_i)$.

$$N_k^{exp}(R_b, \boldsymbol{\beta}, \boldsymbol{\theta}) = N_{k,sig}^{exp}(R_b, \boldsymbol{\beta}, \boldsymbol{\theta}) + \sum_{g \in bkg} N_{k,g}^{exp}(\boldsymbol{\theta}).$$

The total number of expected events is expressed as the sum of the signal events ($t\bar{t}$), depending on R_b and the contribution of the different background processes, indicated with the index g. The extended expression of $N_{2\ b-tags,sig}^{exp}(R_b, \boldsymbol{\beta}, \boldsymbol{\theta})$ is:

$$\begin{aligned}
N_{2\ b-tags,sig}^{exp}(R_b, \boldsymbol{\beta}, \boldsymbol{\theta}) &= N_{2\ jets,sig}^{data}(\boldsymbol{\theta}) \times P_2(R_b, \boldsymbol{\beta}) \\
&= (N_{2\ jets}^{data} - \sum_{g \in bkg} N_{2\ jets,g}^{exp}(\boldsymbol{\theta})) \times \\
&\{\ \alpha_2 \cdot [R_b^2 \varepsilon_b^2 + 2R_b(1-R_b)\varepsilon_b \eta_q + (1-R_b)^2 \eta_q^2] \\
&+ \alpha_1 \cdot [R_b \varepsilon_b \eta_q + (1-R_b)\eta_q^2] \\
&+ \alpha_0 \cdot \eta_q^2 \},
\end{aligned}$$

where the number of signal events having two b-tagged jets is proportional to the number of signal events with two jets, obtained subtracting the background to the observed data events, and the probability of b-tagging both of them. Using the same logic, the expressions of the 0 and 1 b-tagged jet bins are derived and documented in Appendix A.

These equations are used in a profile likelihood fit in which ε_b, η_q are fixed, the α_i fractions have been corrected using dedicated control regions and R_b is a free

floating parameter. The resulting likelihood expression is:

$$\mathscr{L}(R_b, \boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{k=0}^{2\ b-tags} \mathscr{P}[N_k^{exp}(R_b, \boldsymbol{\beta}, \boldsymbol{\theta}), N_k^{data}] \cdot \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\theta_i^2/2}, \tag{5.5}$$

which is the product of the Poisson distribution $\mathscr{P}$ having as expected number of events $N_{k\ b-tags}^{exp}$ and the Gaussian distribution of each systematic effect. The best estimate for R_b is obtained by minimising the negative log-likelihood *-log*$\mathscr{L}$.

5.4.3 Nuisance parameters in profile likelihood fit

Each systematic variation is associated to a NP θ_i that modifies the shape and/or the normalisation of the $M(\ell, j)$ template, in the CRs, and the b-jets multiplicity template of the backgrounds in the SR(top,top).
The value $\theta = 0$, by construction, corresponds to the best knowledge of a specific parameter (nominal value). Uncertainty variations up to $\pm 1\sigma$ correspond to the 1σ uncertainty. The nuisance parameters are defined by the extrapolation ($|\theta| > 1$) and interpolation ($|\theta| < 1$) functions with constrains that $\theta = 0$ corresponds to no corrections and $\theta = \pm 1$ shifts the distribution by $\pm 1\sigma$ systematic uncertainty. The best fit value of the NP is $\hat{\theta}$. Nuisance parameters are implemented using Gaussian constrains reflecting the prior knowledge of the systematic uncertainty.

5.5 Systematic uncertainties

The measurement presented in this work relies on simulated samples in three ways:

- the signal simulation is used to derive the predicted values of the parameters ε_b, η_q and α_i in the SR;
- the signal and background simulations are used to predict the $M(\ell, j)$ distributions in the CRs;
- the background simulation provides the prediction of the b-tagged jets multiplicity in the SR.

Therefore, several sources of systematic uncertainties may affect the distributions in the SR and CRs together with the estimation of ε_b, η_q, α_i. They are classified into those pertaining to experimental effects and those related to the physics modeling of the signal and background processes. The experimental uncertainties include the contributions from the measurement of the luminosity, the pileup correction, and the identification and reconstruction of the physics objects. The signal and background modeling uncertainties are related to the theoretical framework used to produce the MC simulations.
In the following, a brief description of the sources of systematic uncertainty is provided. Some of the systematic uncertainties are evaluate by varying the originating effect within 1σ. The difference with respect to the nominal result is taken as systematic uncertainty. For some sources, only one variation is available and the difference is symmetrised.

The effect of the systematic variations on ε_b and η_q is evaluated by varying the signal MC: the impact on R_b is then estimated by repeating the fit and comparing the

resulting value with the nominal value.

The impact of the systematic variations on the α_i is obtained varying the signal and background simulations in the control regions, in the context of the profile likelihood fit.

The final uncertainty on R_b is evaluated combining the uncertainty coming from α_i, ε_b, η_q and the uncertainty on the background distribution of b-tagged jets in the SR.

5.5.1 Experimental Uncertainties

Luminosity and Pileup

The uncertainty on the integrated luminosity for the combined 2016 and 2017 data set is 2.0% [90]. This uncertainty is applied on the normalisation of all the processes determined by the MC simulation.
An uncertainty due to the pile-up is evaluated by varying the average number of interactions by a 9% in the simulation, common practice in ATLAS.

Objects uncertainties

The final state selected in this measurement contains muons, electrons and jets. The identification efficiencies of these objects are derived in simulation and then corrected with scale factors extracted from data. Therefore, the uncertainties on these corrections have to be considered. The corrections related to the scale factors of the trigger efficiency, the reconstruction and identification are applied. The corrections related to the energy scales and resolutions are applied by smearing or re-scaling the energies of the objects.

This measurement accounts for the systematic effect of the electron and muon reconstruction, identification, trigger, isolation efficiencies, and the resolution of the determination of the energy and momentum.

Uncertainties on the jets are related to the smearing of the jet energy resolution, jet energy scale, and jet vertex tagger.

The uncertainty on the jets identified by the MV2 b-tagging algorithm as b-jets is evaluated by varying the scale factors of b-, c- and *light*-jets within their uncertainty. The systematic uncertainty on the calibration of the b-jets is the leading source of uncertainty in the determination of ε_b and R_b.

5.5.2 Signal Modeling

The uncertainty related the fragmentation and hadronisation model is evaluated using the alternative $t\bar{t}$ simulation generated by POWHEG and interfaced to HERWIG 7. The extraction of R_b is performed using the α_i, ε_b and η_q predicted by this alternative $t\bar{t}$ model. The resulting value is compared to the nominal case and their difference is taken as fragmentation and hadronisation uncertainty. This uncertainty is the third leading contribution to the final R_b uncertainty.

Three variations of the $t\bar{t}$ nominal simulation are performed and explained in the following. In all these three cases, the impact on the α_i is evaluated in the CRs by the profile fit.

In order to account for the activity of extra jets produced in the event, an uncertainty on the ISR and FSR is evaluated. The ISR variation is performed by doubling and dividing by a factor two the renormalisation and factorisation scales and then varying the showering. The systematic effect due to FSR activity is quoted by varying the internal parameters of PYTHIA 8.

The uncertainties originating from the parton distribution functions are quantified following the PDF4LHC recommendations [112].

5.5.3 Background

Single top modeling

The main background process in all the regions defined by this measurement is the single top production via the tW-channel. An overall uncertainty of 5.3% on the normalisation of the single top is assigned: this uncertainty derives from the theoretical knowledge of the single top cross-section [113, 114].

The cross-section of the single top production depends on $|V_{\mathrm{tb}}|$ and on the parton distribution function for the incoming quark. In particular, for the tW-channel production cross section, it is:

$$\sigma(pp \to tW) \propto |V_{tb}|^2\sigma_b^{tW} + |V_{ts}|^2\sigma_s^{tW} + |V_{td}|^2\sigma_d^{tW}, \tag{5.6}$$

where $\sigma_{b/s/d}^{tW}$ is the tW-channel cross section for an incoming b/s/d-quark.
It follows that under the assumption of $|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2 = 1$, if $R_b < 1$, and so is $|V_{\mathrm{tb}}|$, the single top cross section increases, modifying its contribution in the various regions of this analysis correspondingly. More in detail, given the large values of the down and strange quarks PDFs inside the proton, the smaller the value for $|V_{\mathrm{tb}}|$ the higher the single top cross-section becomes. To account the impact of this $|V_{\mathrm{tb}}|$ dependency, a one side uncertainty is evaluated comparing the nominal single top normalisation with a varied value. The nominal single top normalisation is computed using the standard simulation, in which $|V_{\mathrm{tb}}| = 1$. The varied value is obtained by computing the $gq \to tW$ cross section in the case of $|V_{\mathrm{tb}}| = 0.95$. This value has been chosen as it represents a scenario in which $|V_{\mathrm{tb}}|$ is not compatible with unity within the uncertainty reported in the literature, which is 2.5% (see Equation 2.2 in Section 2.1.1). The cross-section for the varied case is obtained as function of $|V_{\mathrm{tb}}|$ and $|V_{td}/|V_{ts}|$ as follows:

$$\sigma(gq \to tW) = (|V_{tb}|^2\sigma_b^{tW} + \frac{1-|V_{tb}|^2}{1+|V_{td}/V_{ts}|^2}\sigma_s^{tW} + \frac{|V_{td}/V_{ts}|^2(1-|V_{tb}|^2)}{1+|V_{td}/V_{ts}|^2}\sigma_d^{tW}), \tag{5.7}$$

where the theoretical prediction of $|V_{td}/V_{ts}|$ [7] is

$$|V_{td}/V_{ts}| = 0.216 \pm 0.001 \pm 0.011, \tag{5.8}$$

and the σ_q^{tW} are calculated under the assumption $V_{tq} = 1$. The resulting cross-sections calculated using POWHEG-BOX V2 are reported in Table 5.5. The single

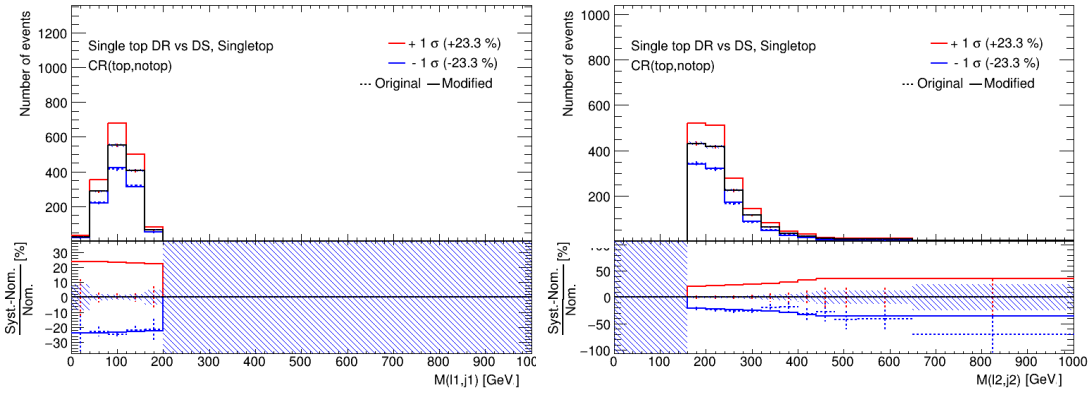
Process	Cross-Section[pb]
$gb \to tW$	73±0.03
$gs \to tW$	175±0.07
$gd \to tW$	486±0.4

TABLE 5.5: tW-channel cross sections for bottom, strange or down incoming quarks under the assumption of $V_{tq} = 1$. The uncertainty is statistical only.

top varied cross section in the case of $|V_{tb}| = 0.95$ is 1.13 times the nominal value. The single top contribution is varied by this factor in all the regions of analysis. At the same time, a migration of single top events in the different b-tagged jets bins is performed to account for the same $|V_{tb}|$ variation in the decay vertex.

As in the $t\bar{t}$ case, uncertainties on the parton and hadronisation model has been evaluated comparing the POWHEG+PYHTIA 8 and POWHEG+HERWIG 7 simulations. The ISR and FSR variations are performed in the same way as for the $t\bar{t}$ case.

The difference between the diagram subtraction and diagram removal tW-channel simulations (defined in Section 1.5.1) is also assigned as systematic uncertainty. This uncertainty represents one of the dominant contributions to the α_i uncertainty and thus on R_b. The reason for this stands in the difference between the DR and DS simulation in the CRs of the analysis. Figures 5.3a and 5.3a show the distribution



(A) (B)

FIGURE 5.3: Distribution of $M(\ell_1, j_1)$ (A) and $M(\ell_2, j_2)$ (B) in the CR(top,notop). The nominal tW-channel (DR in black) is compared to the alternative simulation (DS), where the one side variation is symmetrised with an up (red) and down (blue) variation.

in the CR(top,notop) of the $M(\ell_1, j_1)$ and $M(\ell_2, j_2)$ variable, respectively. Here the nominal sample (DR) is compared to the alternative sample (DS): the alternative sample prediction is 23% larger than the nominal. This is due to the definition of the CRs: the cut on $M(\ell_1, j_1)$ and $M(\ell_2, j_2)$ defining the regions ensure that one or both of the top in event is not resonant. As a result the CRs are more sensitive to the interference term with respect to the SR, where the contribution of double resonant

top quarks is dominant. The measurement of the interference term is reported in this paper [115], where the minimum mass between of the lepton and b-jets is used as discriminant variable for various interference models.

Other backgrounds

Uncertainties due to the theoretical cross-sections of the remaining backgrounds are evaluated by varying them by $\pm 6\%$ for the diboson production [116], by a $\pm 5\%$ for the W+jets and Z+jets production [117], and by a $\pm 13\%$ for the associate production of $t\bar{t}$ to a boson [118].

5.6 Closure test on the fitting procedure

In order to validate the robustness of the statistical model, the fitting procedure is applied to the nominal $t\bar{t}$ simulation, in which R_b is fixed to one. The results are summarised in Table 5.6 and show that the fitted values of R_b and μ_{α_i} agree with the prediction within their uncertainty. The uncertainty reported on R_b is the out-

Parameter	Fitted Value
R_b	1.00 ± 0.01
$\mu_{\alpha_1}(leading)$	1.00 ± 0.08
$\mu_{\alpha_1}(subleading)$	1.00 ± 0.06
μ_{α_0}	1.00 ± 0.17

TABLE 5.6: Resulting R_b and μ_{α_i} from the closure test: here the parameters are extracted from a fit performed on the simulation. The uncertainty is obtained summing the systematic and statistical contributions from the profile fit.

come of the profiled fit on the four regions and it does not include the effect on the uncertainty on the variation of ε_b.

5.7 Expected performance

The expected performance of this analysis are discussed in this section. Thanks to these studies it has been possible to fine tune the analysis strategy, understand the constrains on the NP and their impact on the R_b sensitivity and define the final binning of the distributions in the CRs.
The expected uncertainty on R_b is 3.0% and it is evaluated as the sum in quadrature of the uncertainty coming from the profile fit, the variation of ε_b and η_q, and the variation of the $t\bar{t}$ model.

5.7.1 Statistical Model

The expected performances of the statistical model are estimated performing the profile likelihood fit on the Asimov data-set [119]. The Asimov data-set is built from the predicted distribution, assuming a Poisson statistical error in each bin. The Asimov fit is constructed in such a way that the nuisance parameters corresponding to the systematic uncertainties are all centered at zero and the normalization factor is

centered at 1. The Asimov fit gives

$$R_b = 1.00 \pm 0.003\ (stat) \pm 0.013\ (syst), \tag{5.9}$$

where the uncertainty is the sum of the statistical fluctuations of both data and simulation, and the systematic uncertainty is the result of the variations of the simulations in the CRs and SR.
Figure 5.4 shows the ten most important systematic uncertainties, ranked by their impact on R_b. The impact of each nuisance parameter, ΔR_b, is computed by compar-

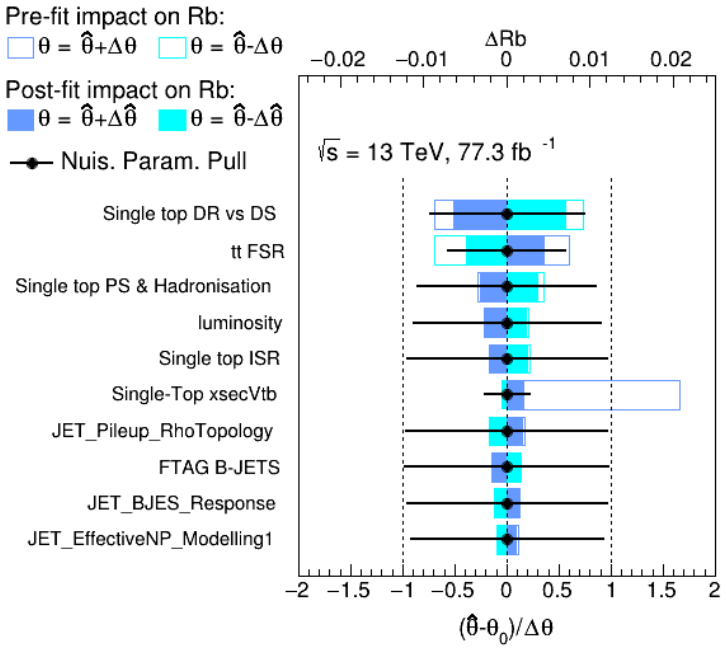


FIGURE 5.4: Ranking plot of the ten most important nuisance parameters included in the Asimov fit, ordered by their impact on R_b. The empty blue rectangles correspond to the pre-fit impact and the filled blue ones to the post-fit impact. The black points show the pull of the NP with respect to the nominal value θ_0.

ing the nominal best-fit with the result of the fit when fixing the considered nuisance parameter to its best fit value $\hat{\theta}$ shifted by its pre-fit and post-fit uncertainties.
The leading contribution to the R_b uncertainty is the comparison between the DR and DS models of the tW-channel single top simulation. The reason of this large uncertainty stands in the differences in the modeling of the $M(\ell, j)$ distributions in the CRs, as already discussed in Section 5.5. The second leading contribution is the variation of the FSR activity in the $t\bar{t}$ simulation in the CRs: the presence of jets coming from FSR, and thus not produced in the top quark decay, affects the correction of the jet-to-top assignment fractions. The other following uncertainties are: the comparison between the parton shower models in the single top simulation, the knowledge of the integrated luminosity, the variation of the single top ISR activity, the variation of the single top cross-section according to its dependence from $|V_{tb}|$, the variation of the jet quantities due to pileup effects, the calibration of the b-tagging efficiency of b-jets, the JES response for b-jets and one the variations of the jet calibration.

5.7.2 Variation of ε_b and η_q

The systematic variations of the $t\bar{t}$ simulation also affect the determination of ε_b and η_q. However, it is not possible to perform these variations making use of the NPs in the profile fit: the strong correlation with the parameter of interest causes the fit not to converge. Therefore, these two parameters have been fixed to their nominal values in the profile fit. Their uncertainty has been evaluated varying the $t\bar{t}$ simulation and summing in quadrature all the contributions, reported in Table 5.7. Here

Uncertainty Source	$\sigma(\varepsilon_b)/\varepsilon_b$[%]	$\sigma(\eta_q)/\eta_q$[%]
Calibration of b-tagging	2.4	17.8
Top Pair FSR	0.7	8.8
Top Pair ISR	0.1	0.3
Top Pair PDF	0.1	0.1
Jet Reco. and Calib.	0.1	0.6
Leptons Reco. and Calib.	<0.1	<0.1
Pileup	<0.1	0.2
Total Uncertainty	2.5	20.1

TABLE 5.7: Breakdown of the relative percentage systematic uncertainties on ε_b and η_q.

the systematic variation have been grouped: the calibration of the b-tagging contains the variations of the b-jets, c-jet and light-jets b-tagging efficiencies, the $t\bar{t}$ nominal simulation contributions from the FSR, ISR and PDF are reported separately, the variation of the reconstruction and calibration of jets and leptons are also grouped and the pileup variation is reported as a single term. The uncertainty on ε_b is dominated by the calibration of the b-tagging efficiency provided by the p_T^{rel} method, while the uncertainty on η_q is dominated by the calibration of the mis-tag rate of light jets.
The fit model is applied to the $t\bar{t}$ simulation for the up and down variation of ε_b and η_q. The resulting uncertainty on R_b is 2.5% and 0.07% for the ε_b and η_q variations, respectively.

5.7.3 Variation of the $t\bar{t}$ model

The R_b measurement relies on the $t\bar{t}$ simulation generated using POWHEG and interfaced to PYTHIA 8. The variation of the parton shower model and hadronisation effects the ε_b, η_q and the α_i fractions. This variation is studied comparing the nominal $t\bar{t}$ simulation with the one interfaced to HERWIG 7.
The comparison of the parameters values provided by the two models are reported in Table 5.8. The difference in the predicted values of the b-tagging efficiency and mis-tag rate can be assigned to the different fragmentation models implemented in the two parton showers. The differences in the α_i fractions can be explained by the different kinematics of the top quark decays products provided by the two simulations: this can modify the fraction of jets reconstructed and selected and then assigned to the top quark decay.

Parameter	PYTHIA 8	HERWIG 7
ε_b	0.79	0.77
η_q	0.019	0.017
α_2	0.69	0.67
$\alpha_1(leading)$	0.10	0.10
$\alpha_1(subleading)$	0.19	0.20
α_0	0.02	0.02

TABLE 5.8: Predicted values of the fit parameters in the SR(top,top) by the POWHEG+PYTHIA8 and POWHEG+HERWIG7 simulations.

To quantify the uncertainty on the choice of the $t\bar{t}$ POWHEG+PYTHIA8 simulation as the nominal, the profile fit is performed on the nominal simulation using the values predicted by POWHEG+HERWIG7. The result gives an uncertainty of 1.0% on R_b.

5.8 Results

The value of R_b measured from this analysis is

$$R_b = 1.022 \pm 0.031, \tag{5.10}$$

where the uncertainty is the sum in quadrature of the uncertainty coming from the profile fit, the variation of ε_b and η_q and the variation of the $t\bar{t}$ model. This value is compatible with unity and represents the most precise single measurement of R_b.

The post fit distribution of the b-tagged jets multiplicity is reported in Figure 5.5, where the prediction is in agreement with data. The distribution of the prediction is shown in a stacked plot, where the b-tagged jets multiplicity template for the background processes is derived from the simulation while the template for $t\bar{t}$ events is derived following the model exposed in Section 5.4, making use of the distribution of the number of jets observed in data, the fitted values of R_b and α_i and the central values of ε_b and η_q.

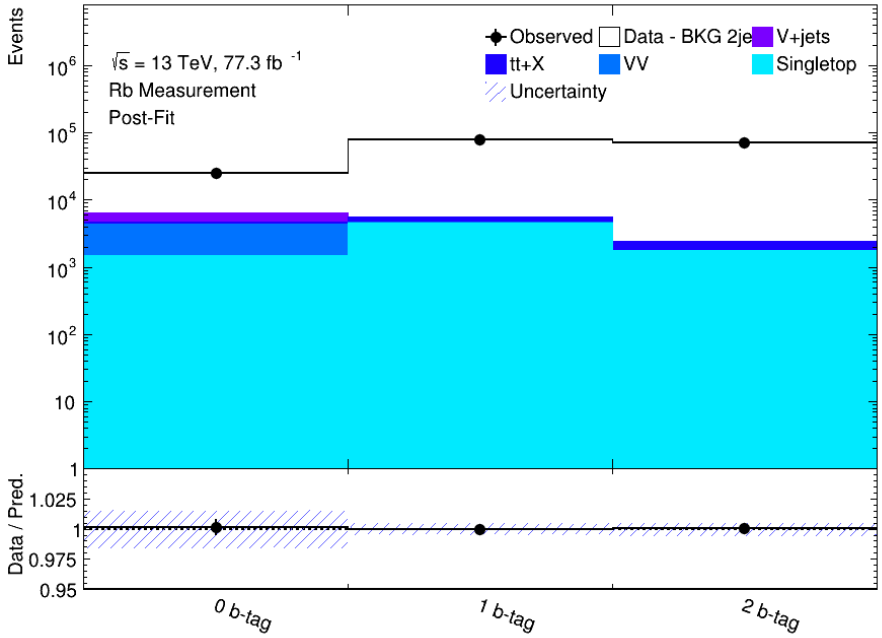


FIGURE 5.5: Distribution of the b-tagged jets multiplicity comparing the data with the prediction. The distribution for the single top is shown in cyan, for the diboson in azure, for the tt+X in blue and for the Z+jets in violet. The white distribution is built from the data distribution of number of jets, subtracting the background contribution and multiplying the probabilities to have 0, 1 or 2 b-tagged jets for the fitted value of R_b = 1.022.

A summary of the different contributions to the final uncertainty is reported in Table 5.13. The result is coherent with the expected performances, with a leading contribution contribution coming from the variation of the b-tagging efficiency. This variation is performed comparing the nominal value with the value obtained from a fit on data with a varied ε_b value. The uncertainty coming from the profile likelihood fit is 0.012. The uncertainty related to the $t\bar{t}$ model is evaluated fitting the data using

Uncertainty Source	$\sigma(R_b)$	$\sigma(R_b)/R_b$[%]
ε_b	0.026	2.5
Profile fit	0.012	1.2
$t\bar{t}$ model	0.011	1.0
η_q	<0.001	0.01
Statistical	0.003	0.3
Total Uncertainty	0.031	3.0

TABLE 5.9: Breakdown of the absolute and relative uncertainties on the R_b value fitted in data events, together with the total value.

the parameters and the $M(\ell, j)$ distributions predicted by the POWHEG+HERWIG7 simulation. The values of the α_i fractions obtained are reported in Table 5.10, where they are compared to the values extracted from the fit using the nominal $t\bar{t}$ simulation. The values obtained are in agreement within their uncertainties.

Post-Fit	α_2	$\alpha_1(leading)$	$\alpha_1(subleading)$	α_0
POWHEG+PYTHIA8	66.6±6.9	11.1±0.8	19.4±0.5	2.8±0.2
POWHEG+HERWIG7	69.1±7.1	9.8±0.8	18.8±0.5	2.3±0.2

TABLE 5.10: Values of the α_i jet-to-top assignment fractions obtained from the fit on data using as a $t\bar{t}$ model the POWHEG+PYTHIA 8 simulation and the POWHEG+HERWIG 7 simulation. The uncertainty contains both the systematic and statistical effects.

The ranking plot for the fit on data has been derived in the same way as in the Asimov fit. The ten most important contribution to the R_b uncertainty from the likelihood fit are reported in Figure 5.6, where the pull of the NPs is now different from zero as it corresponds to the observed value on data.
The leading contributions are again the modeling of the single top process and the additional jet activity in $t\bar{t}$ events due to ISR and FSR. It is interesting to notice that the variation of the single-top cross-section, as effect of the $|V_{tb}|$ dependency, shows a big pre-fit value which is strongly reduced after the constrain on data. An important contribution to the final uncertainty is also the calibration of jets and the luminosity.

The results of the fits in the CRs are summarised in Table 5.11. The uncertainty on

Correction Factor	Fit Result
$\mu_{\alpha_1}(leading)$	1.13 ± 0.07
$\mu_{\alpha_1}(subleading)$	1.01 ± 0.05
μ_{α_0}	1.26 ± 0.17

TABLE 5.11: Resulting correction factors derived from the fit in the CRs. Uncertainty is statistical and systematic.

the μ_{α_0} is bigger than the other due to the limited statistics in the CR(notop,notop). These values have been used to correct in α_i fractions in the SR(top,top): pre-fit and post-fit values are shown in Table 5.11. As no correction factors is derived for α_2, this quantity is calculated imposing $\sum_{i=0}^{2} \alpha_i = 1$. The pre-fit and post-fit values are

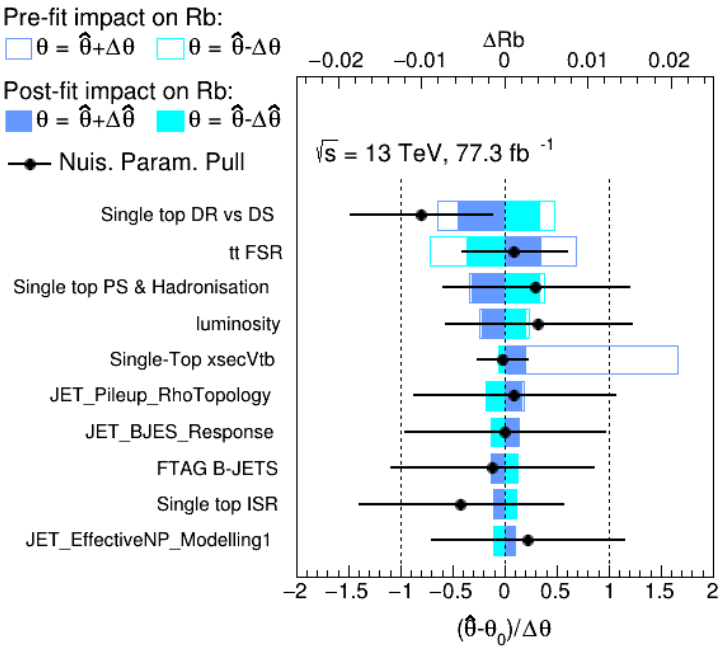


FIGURE 5.6: Ranking plot of the ten most important nuisance parameters included in the fit on data, ordered by their impact on R_b. The empty blue rectangles correspond to the pre-fit impact and the filled blue ones to the post-fit impact. The black points show the pull of the NP with respect to the nominal value θ_0.

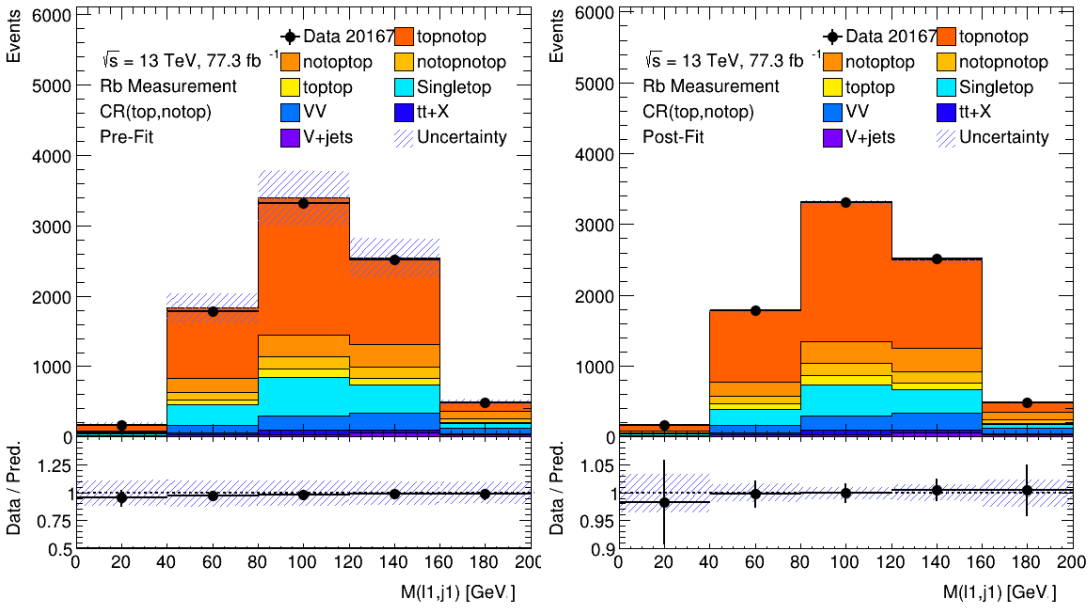
shown in Table 5.12. From the result of the fit in the CRs it can be concluded that a

Events Region	α_2	$\alpha_1(leading)$	$\alpha_1(subleading)$	α_0
Pre-Fit	68.6±0.1	9.8±0.1	19.4±0.2	2.2±0.1
Post-Fit	66.6±6.9	11.1±0.8	19.4±0.5	2.8±0.2

TABLE 5.12: Values of the α_i jet-to-top assignment fractions using the $t\bar{t}$ simulation in the SR(top,top) region before and after the fit on the CRs. The Post-Fit fractions include the systematic uncertainty.

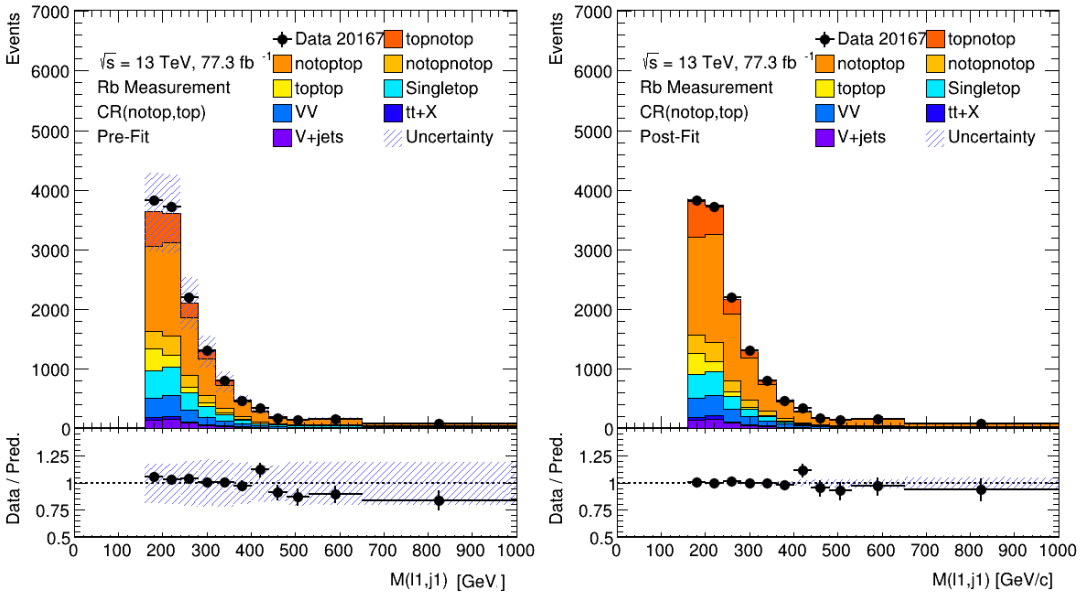
bigger fraction of events having at least one jet not from the top decay is observed in data. As a result, the fraction of events having both jets from the top decay, α_2, in the SR(top,top) has a post-fit value of 66.6 ± 6.9, with the uncertainty including both the statistical and systematic effects. The post-fit and pre-fit values agree within their uncertainty.

The effect on the $M(\ell_1, j_1)$ distribution in the CRs before performing the fit, *"pre-fit"*, and after the effect of the fit, *"post-fit"*, is discussed in the following. Figure 5.7a and 5.7b show the comparison between data and prediction in the CR(top,notop). The effect of the fit is in the change of the $t\bar{t}$ normalisation and the reduction of the systematic uncertainty, due to the pull of the NPs in data. The pre-fit and post-fit plots of the $M(\ell_1, j_1)$ distribution in the CR(notop,top) region for data compared to the simulation is presented in Figure 5.8a and 5.8b, respectively. The pre-fit and post-fit plots of the $M(\ell_1, j_1)$ distribution in the CR(notop,notop) region for data compared to the simulation is presented in Figure 5.9a and 5.9b, respectively. The



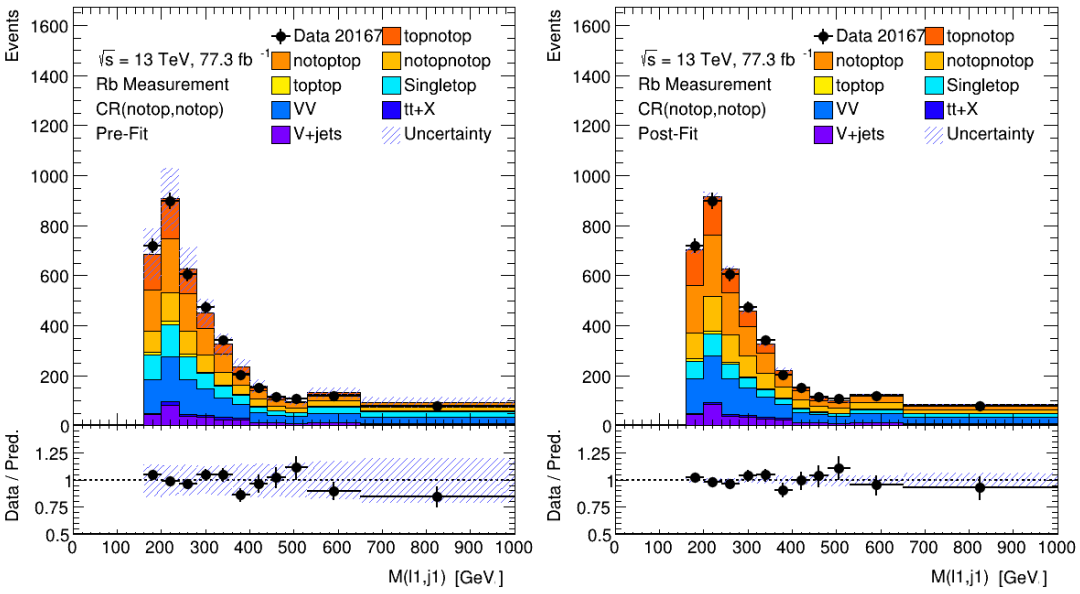
(A) (B)

FIGURE 5.7: Pref-Fit (A) and Post-Fit (B) distribution of the $M(\ell_1, j_1)$ variable in the CR(top,notop). The $t\bar{t}$ categories top-notop, notop-top, notop-notop and top-top are shown in dark orange, light orange, dark yellow and light yellow, respectively. The single top contribution is shown in cyan, the diboson in azure, the tt+X consitribution in blue and the V+jets in violet.



(A) (B)

FIGURE 5.8: Pref-Fit (A) and Post-Fit (B) distribution of the $M(\ell_1, j_1)$ variable in the CR(notop,top). The $t\bar{t}$ categories top-notop, notop-top, notop-notop and top-top are shown in dark orange, light orange, dark yellow and light yellow, respectively. The single top contribution is shown in cyan, the diboson in azure, the tt+X consitribution in blue and the V+jets in violet.



(A) (B)

FIGURE 5.9: Pref-Fit (A) and Post-Fit (B) distribution of the $M(\ell_1, j_1)$ variable in the CR(notop,notop). The $t\bar{t}$ categories top-notop, notop-top, notop-notop and top-top are shown in dark orange, light orange, dark yellow and light yellow, respectively. The single top contribution is shown in cyan, the diboson in azure, the tt+X consitribution in blue and the V+jets in violet.

$M(\ell_1, j_1)$ distribution is well modeled in all the CRs. The same conclusion can be drawn for the $M(\ell_2, j_2)$, of which plots are reported in Appendix D.

5.9 Discussion on the result and its limitations

The measurement of the ratio of the top quark branching fractions presented in this work is limited by the systematic uncertainty. The main source of uncertainty is the calibration of the b-tagging efficiency of b-jets. From the results of the $p_{\mathrm{T}}^{\mathrm{rel}}$ calibration shown in Section 4.2.6, it appears that the uncertainty on the calibration in the jet p_{T} range [70-200] GeV is limited by the available MC statistics. Generating more simulated events could result in smaller fluctuations of the templates and would reduce the template statistical fluctuations. This would translates into gain on the sensitivity on R_{b}.
For illustrative purpose, the different components of the profile fit uncertainty are evaluated and reported in Table 5.13. For each group of systematic uncertainties, the effect on the parameter of interest is evaluated performing a Asimov fit that includes only the NPs under study. The total uncertainty is the one coming from the Asimov fit that includes all the NPs: this fit exploits the correlations between them and gives a final value different from the sum in quadrature of the different components.
From this useful study one can have an idea of the impact of the different components. It is possible to conclude that a gain in sensitivity can be obtained if with a better modeling of the top processes simulations.

Uncertainty Source	$\sigma(R_b)$
Single Top Modeling	0.012
Top Pair PDF	0.007
Luminosity	0.006
Top Pair ISR/FSR	0.005
Jet Reco. and Calib.	0.005
Leptons Reco. and Calib.	0.005
Other Backgrounds	0.004
Pileup	0.003
Statistical	0.003
Total Uncertainty	0.013

TABLE 5.13: Contribution of the systematic and statistical uncertainties from the profile fit grouped in those pertaining to the detector response on the reconstruction of jet, muons and electron, the determination of the luminosity and the pileup activity correction, the modeling of the signal and the background events.

Chapter 6

Conclusions and outlook

A large amount of top quarks is produced in the proton-proton collisions delivered by the Large Hadron Collider. This has opened the possibility to perform precision measurements of the properties of this fundamental particle. These measurements are motivated by a variety of reasons: first of all, given its peculiar properties, this particle is considered a good candidate to probe for physics BSM; moreover, the top quark is a background in many searches for physics BSM.
So far, all the results obtained at the Large Hadron Collider are compatible with the SM predictions.

The Standard Model gives no prediction on the coupling of the top quark to the bottom quark, V_{tb}, therefore is necessary to measure it. Under different assumptions, the module of V_{tb} has been extracted in two ways from hadron colliders data: through the measurement of the single-top cross-section and from the estimation of the decay rate of $t\bar{t}$ pairs into b quarks, relative to all down-type quarks: R_{b}.
The current status of the art shows that both these estimations of $|V_{\mathrm{tb}}|$ are currently limited by the knowledge of the systematic effects.

This thesis presented the measurement of R_{b} performed using the data collected by the ATLAS experiment at the LHC in 2016 and 2017, corresponding to an integrated luminosity of $78 fb^{-1}$ of proton-proton collisions at center-of-mass energy of 13 TeV. Events with two isolated charged leptons, of which one electron and one muon of opposite electric charge, and two jets have been considered.
A kinematical reconstruction has been performed, with the purpose of associating the final state objects to the decay of two top quarks in the event. Based on this, events have been classified into four statistical independent regions: one enriched in events having two jets most likely coming from the top decay, three regions enriched in events having at least one jet coming from ISR and/or FSR activity.
The control regions have been used to correct the predicted jet-to-top assignment fractions using collision data. The parameter of interest has been extracted performing a profile likelihood fit on the b-jet multiplicity distribution in the signal region.
The result is:

$$R_{\mathrm{b}} = 1.022 \pm 0.031, \tag{6.1}$$

which is compatible with the $R_{\mathrm{b}} = 1$ hypothesis and represents the most precise single measurement of this quantity.
The leading source of uncertainty is the one associated to the b-tagging algorithm performance, contributing with a 2.5% to the total uncertainty.
A dedicated calibration of the b-tagging efficiency of b-jets has been performed in this work. This calibration relies on the $p_{\mathrm{T}}^{\mathrm{rel}}$ method, applied on a sample of multi-jets events with at least on jet spatially matched to a muon.

The final uncertainty on the b-tagging efficiency has been strongly reduced by a new estimation of the semi-leptonic to inclusive extrapolation uncertainty of the scale factors. This uncertainty was, in the Run 1 calibration, 4% in the whole jet kinematical range: it has been lowered to the 1% thanks to a new estimation performed in bins of jet p_{T}, using the larger data set of the Run 2.

The present measurement is systematically limited. The main systematic uncertainties come from the b-tagging calibration and the modeling of $t\bar{t}$ production and decay. An immediate measure that would mitigate the effect of the first is the production of a larger sample of simulated di-jet events. The second requires inputs from the theory community: most importantly, a more accurate treatment of the overlap with the $t\bar{t}$ and single-top simulation would result in a higher sensitivity of the measurement presented in this thesis. Moreover, different methods for the kinematical reconstruction could be investigated to improve the measurement here presented. One possibility is the Matrix Element Method [120,121]: this analytical method could be implied to build a discriminant variable for the different $t\bar{t}$ categories and thus estimate correction factors of the jet-to-top assignment fractions. It is computationally demanding and needs specific resources and a detailed validation to be brought to fruition.

Appendix A

Expression used in R_b likelihood

Expressions of the number of $t\bar{t}$ events having k b-tagged jets (k = 0,1,2) as function of the ratio of the top quark branching fractions, the b-taggin efficiency ε_b, the b-tagging mis-tag rate and the jet-to-top mis-assignment fractions. This expressions are used as prediction in the profile likelihood fit on data to extract R_b.
The number of $t\bar{t}$ events with 2 b-tagged jets is:

$$\begin{aligned}
&N_{events}(2\ b-tags)\\
&\quad = N_{events}(2jets) \times P_2(R_b, \varepsilon_b, \eta_q, \alpha_i)\\
&\quad = N_{events}(2jets) \times \{\\
&\quad\ \alpha_2 \cdot [R_b^2\varepsilon_b^2 + 2R_b(1-R_b)\varepsilon_b\eta_q + (1-R_b)^2\eta_q^2] +\\
&\quad\ \alpha_1 \cdot [R_b\varepsilon_b\eta_q + (1-R_b)\eta_q^2] +\\
&\quad\ \alpha_0 \cdot \eta_q^2\\
&\quad\ \}
\end{aligned}$$

The number of $t\bar{t}$ events with 1 b-tagged jet is:

$$\begin{aligned}
&N_{events}(1\ b-tags)\\
&\quad = N_{events}(2jets) \times P_1(R_b, \varepsilon_b, \eta_q, \alpha_i)\\
&\quad = N_{events}(2jets) \times \{\\
&\quad\ \alpha_2 \cdot [2R_b^2\varepsilon_b(1-\varepsilon_b) + 2R_b(1-R_b)[\varepsilon_b(1-\eta_q)\\
&\quad\ + \eta_q(1-\varepsilon_b)] + 2(1-R_b)^2\eta_q(1-\eta_q)] +\\
&\quad\ \alpha_1 \cdot [R_b\varepsilon_b(1-\eta_q) + R_b(1-\varepsilon_b)\eta_q + 2(1-R_b)\eta_q(1-\eta_q)] +\\
&\quad\ \alpha_0 \cdot [2\eta_q(1-\eta_q)]\\
&\quad\ \}
\end{aligned}$$

The number of $t\bar{t}$ events with 0 b-tagged jets is:

$$\begin{aligned}
&N_{events}(0\ b-tags)\\
&\quad = N_{events}(2jets) \times P_0(R_b, \varepsilon_b, \eta_q, \alpha_i)\\
&\quad = N_{events}(2jets) \times \{\\
&\quad\ \alpha_2 \cdot [R_b^2(1-\varepsilon_b)^2 + 2R_b(1-R_b)(1-\varepsilon_b)(1-\eta_q)\\
&\quad\ + (1-R_b)^2(1-\eta_q)^2] +\\
&\quad\ \alpha_1 \cdot [R_b(1-\varepsilon_b)(1-\eta_q) + (1-R_b)(1-\eta_q)^2] +\\
&\quad\ \alpha_0 \cdot (1-\eta_q)^2\\
&\quad\ \}
\end{aligned}$$

Appendix B

Distributions of $p_{\mathrm{T}}^{\mathrm{rel}}$ using predicted fractions

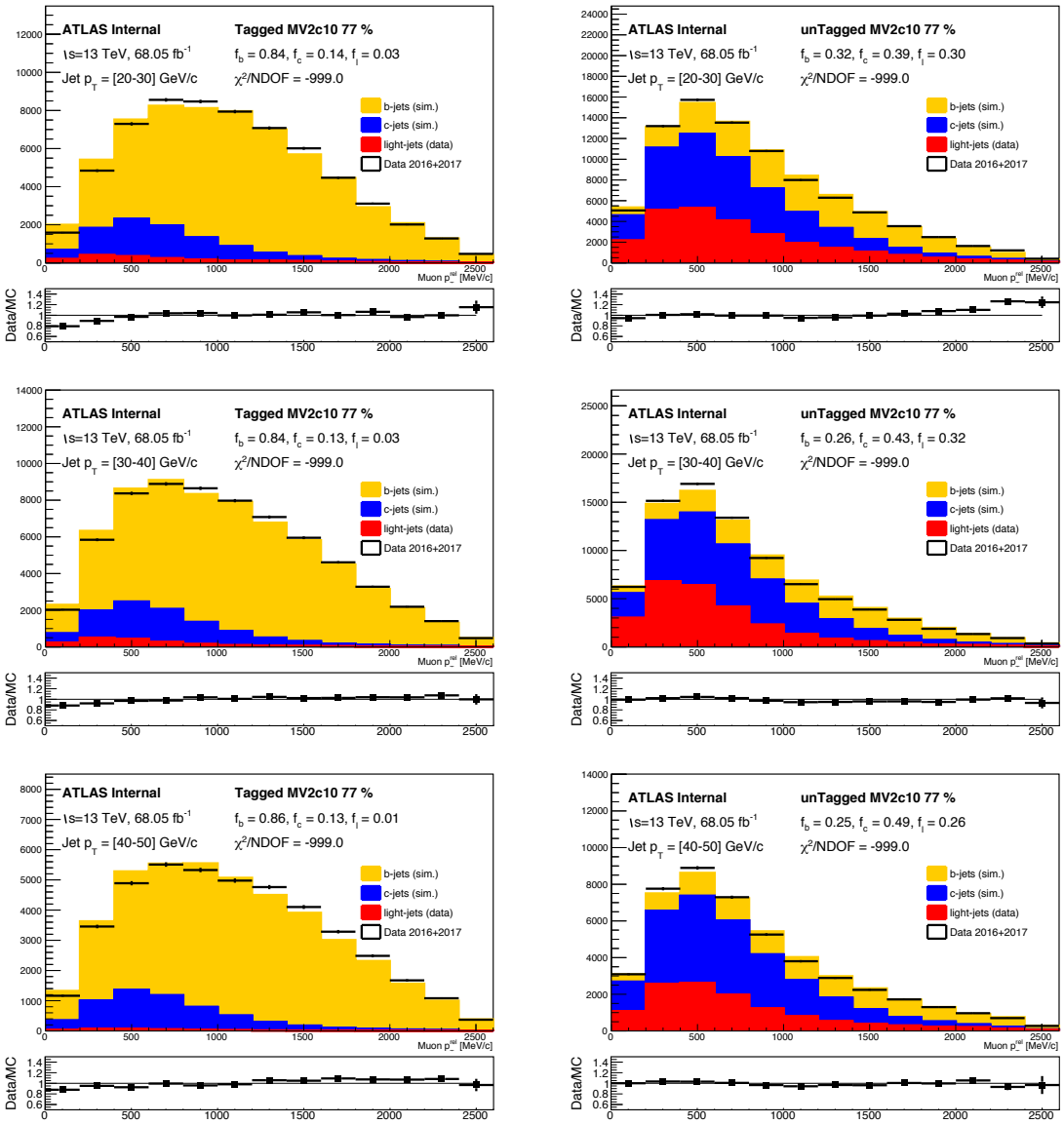


FIGURE B.1: Tagged (left) and untagged (right) $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions using flavour fractions obtained by the unfiltered simulation.

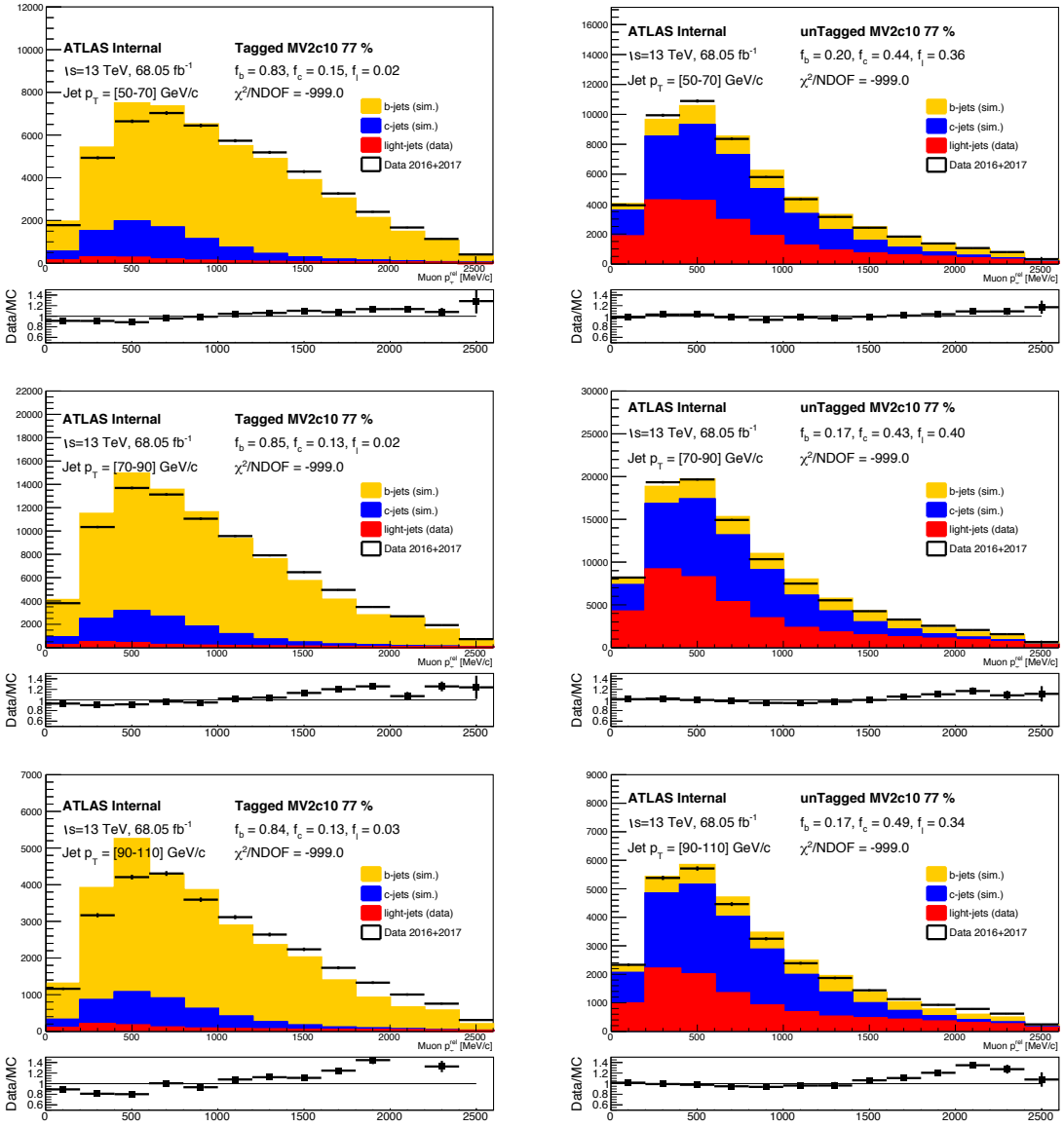


FIGURE B.2: Tagged (left) and untagged (right) p_T^{rel} distributions using flavour fractions obtained by the unfiltered simulation.

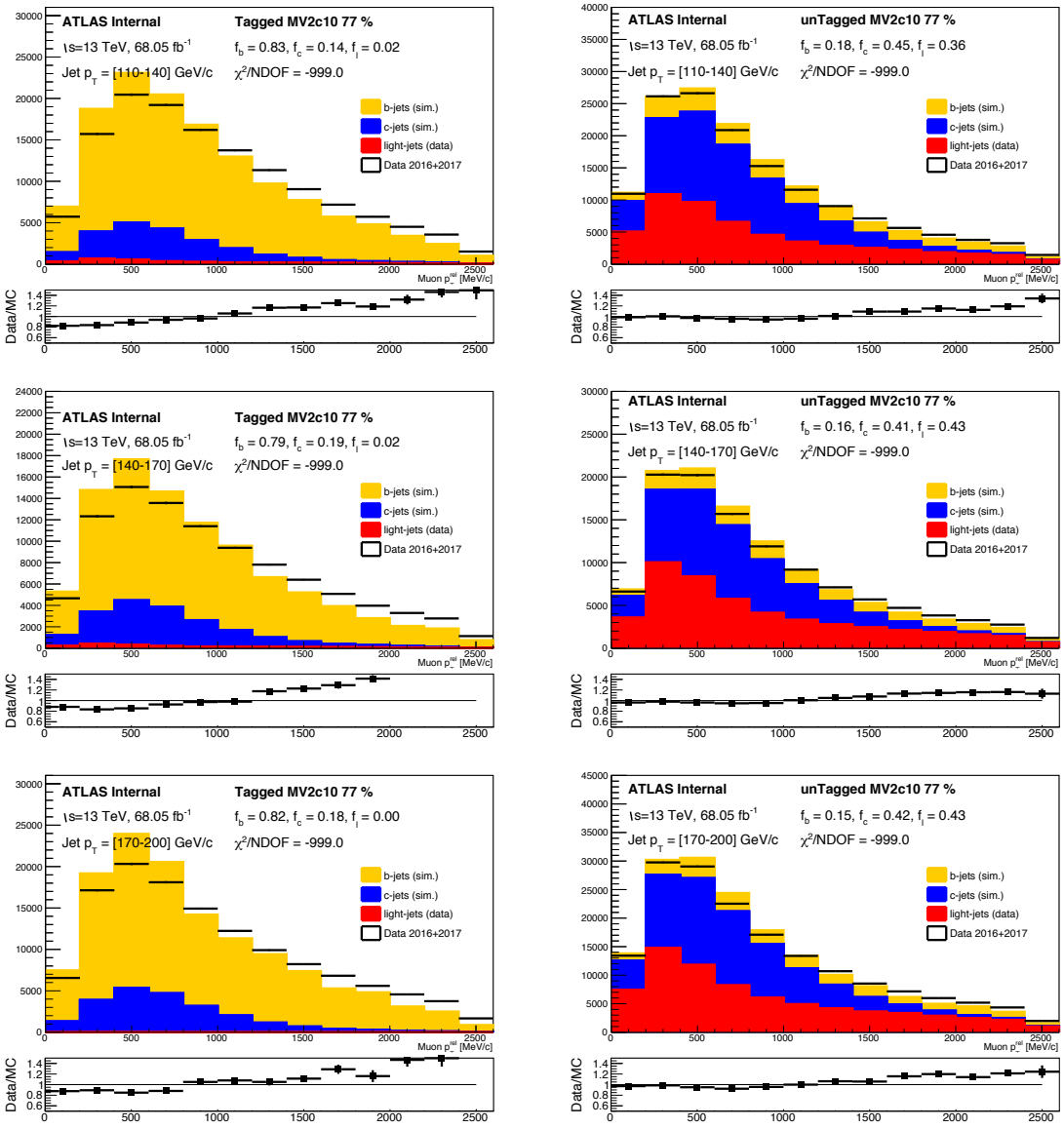


FIGURE B.3: Tagged (left) and untagged (right) p_T^{rel} distributions using flavour fractions obtained by the unfiltered simulation.

Appendix C

Distributions of $p_{\mathrm{T}}^{\mathrm{rel}}$ using fit results

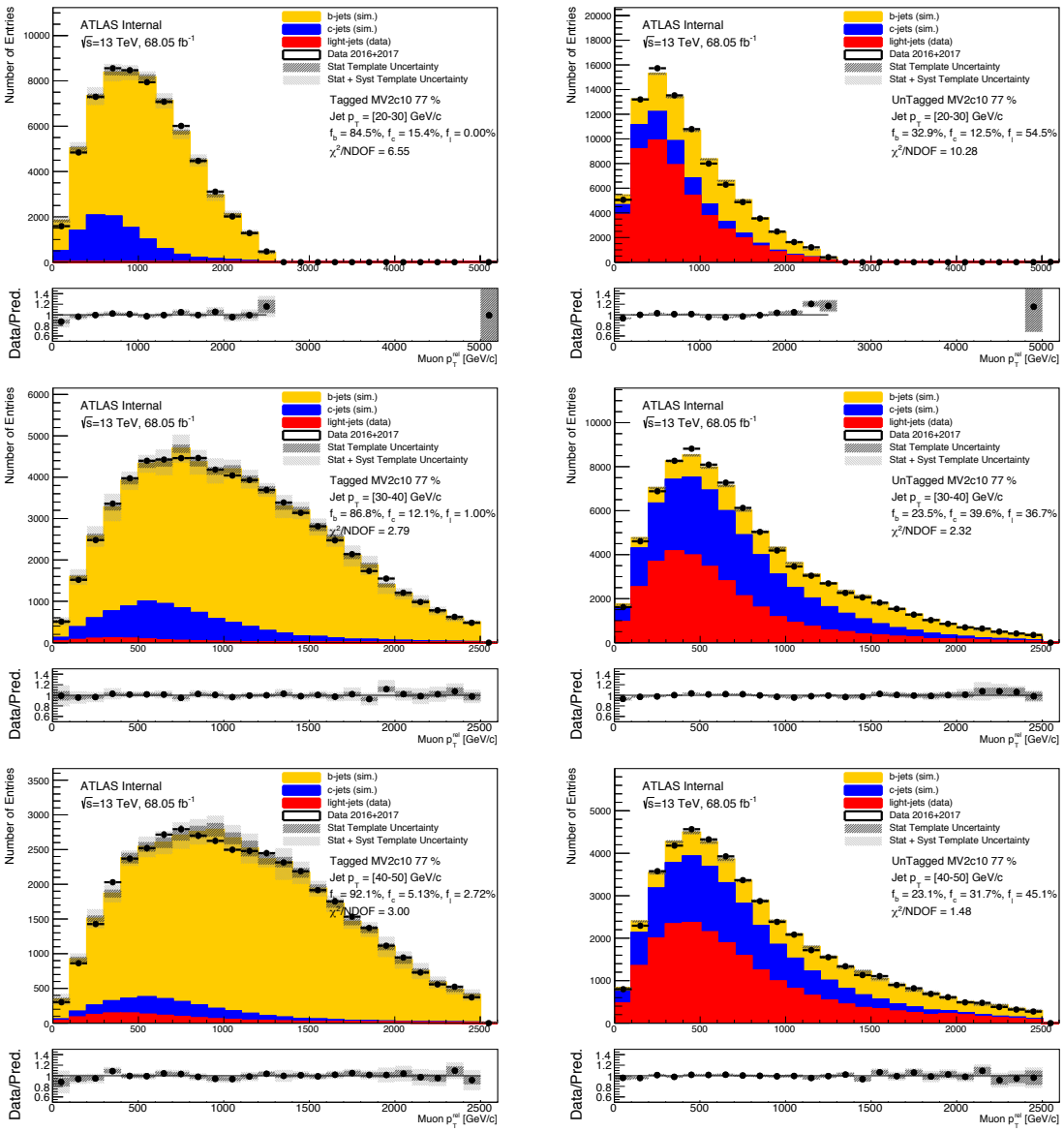


FIGURE C.1: Tagged (left) and untagged (right) $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions using flavour fractions obtained by the log-likelihood fit.

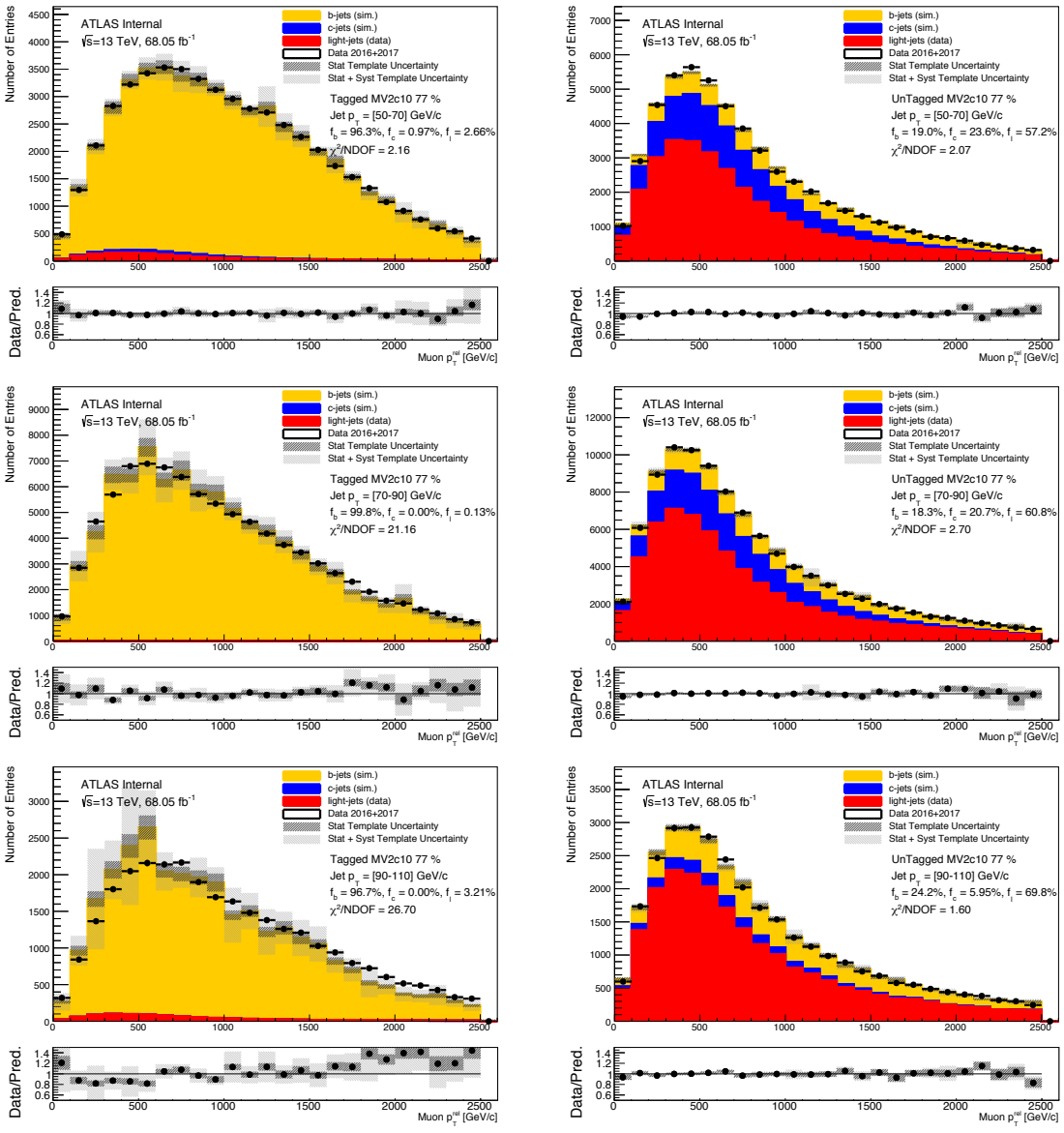


FIGURE C.2: Tagged (left) and untagged (right) p_T^{rel} distributions using flavour fractions obtained by the log-likelihood fit.

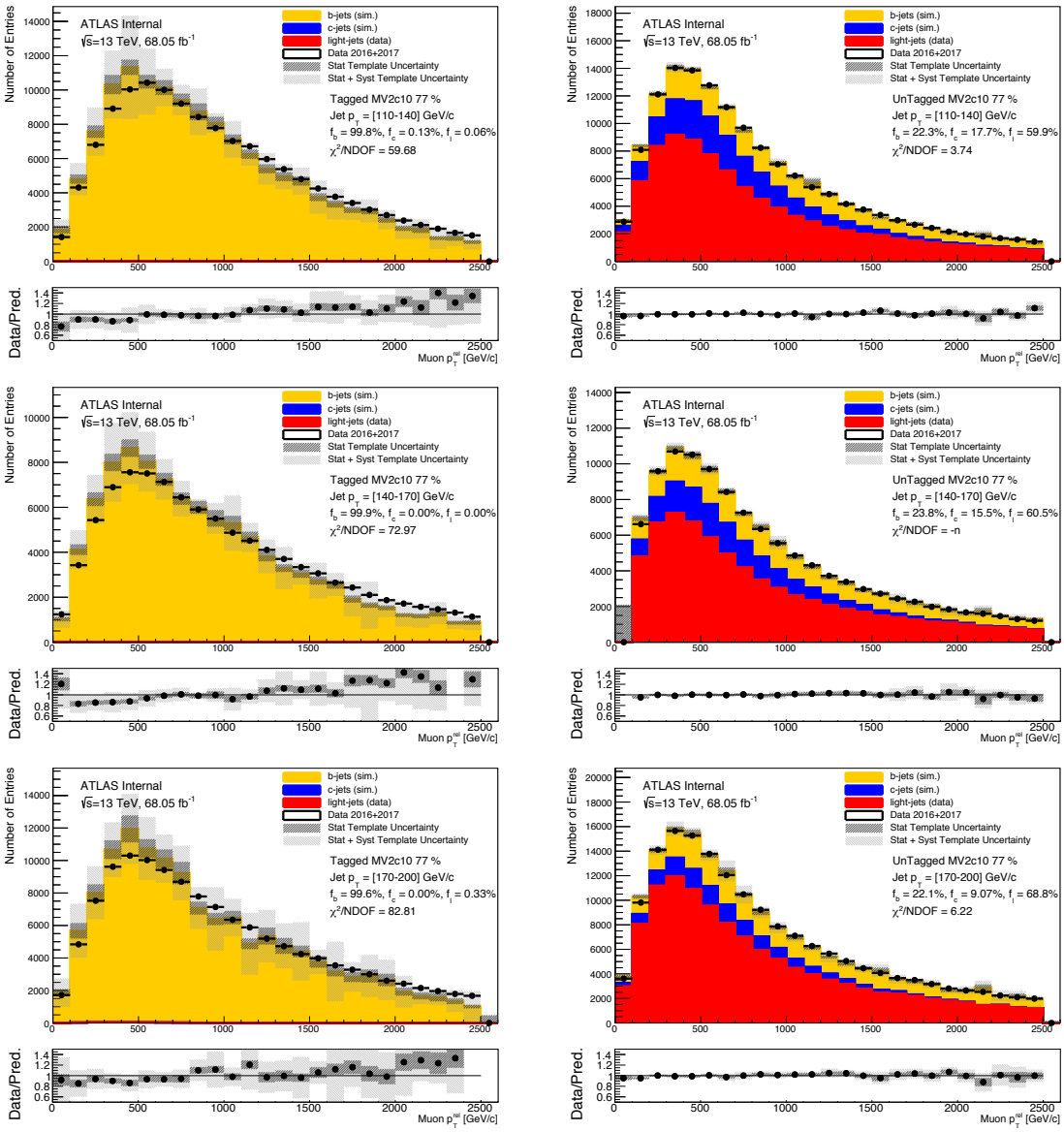


FIGURE C.3: Tagged (left) and untagged (right) $p_{\mathrm{T}}^{\mathrm{rel}}$ distributions using flavour fractions obtained by the log-likelihood fit.

Appendix D

Jet-to-top assignment pre- and post-Fit plots

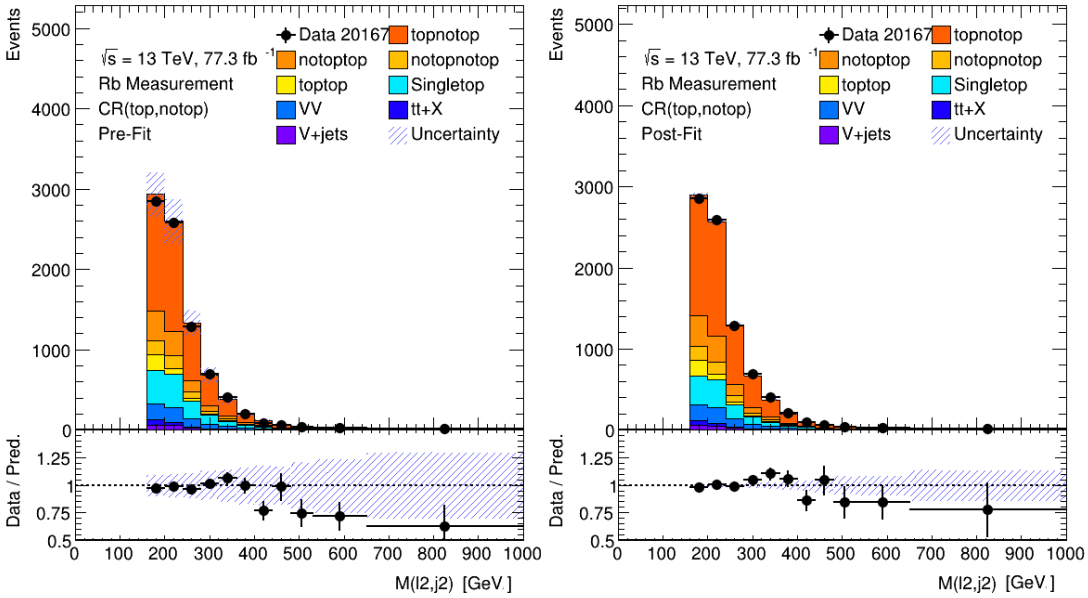


FIGURE D.1: Pref-Fit (A) and Post-Fit (B) distribution of the $M(\ell_2, j_2)$ variable in the CR(top,notop). The $t\bar{t}$ categories top-notop, notop-top, notop-notop and top-top are shown in dark orange, light orange, dark yellow and light yellow, respectively. The single top contribution is shown in cyan, the diboson in azure, the tt+X consitribution in blue and the V+jets in violet.

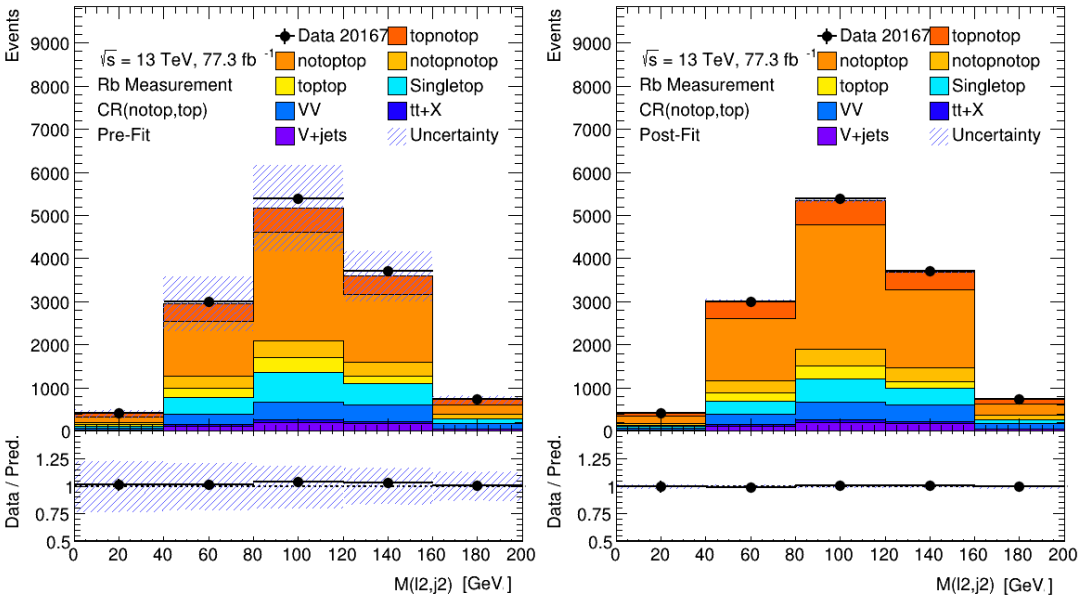


FIGURE D.2: Pref-Fit (A) and Post-Fit (B) distribution of the $M(\ell_2, j_2)$ variable in the CR(notop,top). The $t\bar{t}$ categories top-notop, notop-top, notop-notop and top-top are shown in dark orange, light orange, dark yellow and light yellow, respectively. The single top contribution is shown in cyan, the diboson in azure, the tt+X consitribution in blue and the V+jets in violet.

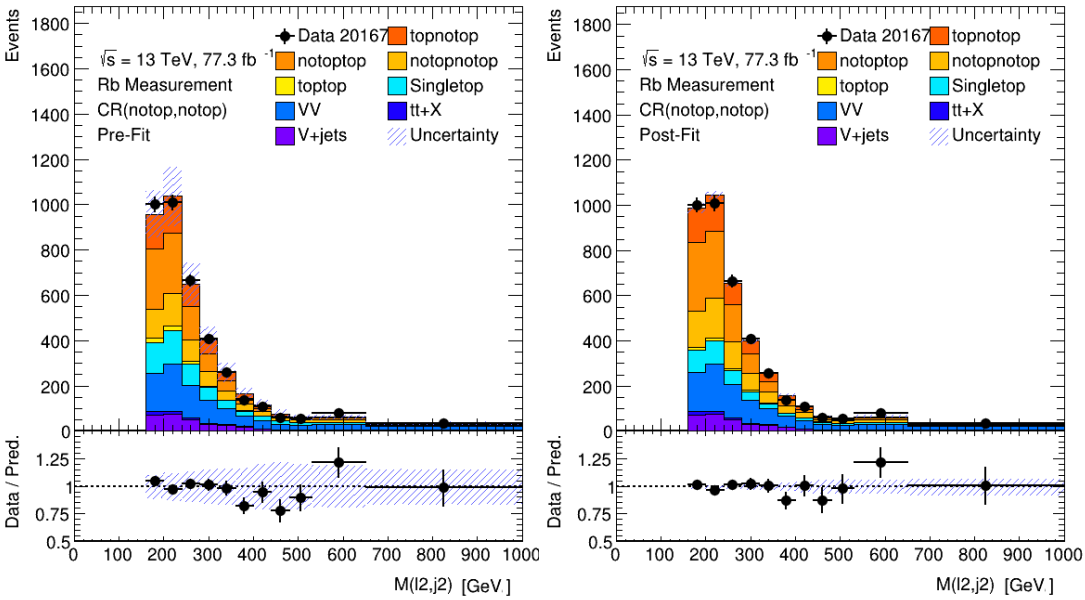


FIGURE D.3: Pref-Fit (A) and Post-Fit (B) distribution of the $M(\ell_2, j_2)$ variable in the CR(notop,notop). The $t\bar{t}$ categories top-notop, notop-top, notop-notop and top-top are shown in dark orange, light orange, dark yellow and light yellow, respectively. The single top contribution is shown in cyan, the diboson in azure, the tt+X consitribution in blue and the V+jets in violet.

Bibliography

[1] Cabibbo, Nicola, "Unitary symmetry and leptonic decays," *Phys. Rev. Lett.*, vol. 10, pp. 531–533, Jun 1963.

[2] Kobayashi, Makoto and Maskawa, Toshihide, "CP-Violation in the Renormalizable Theory of Weak Interaction," *Progress of Theoretical Physics*, vol. 49, pp. 652–657, 02 1973.

[3] Evans, Lyndon and Bryant, Philip, "LHC Machine," *JINST*, vol. 3, p. S08001, 2008.

[4] R S Moore and A Jansson and V Shiltsev, "Beam instrumentation for the tevatron collider," *Journal of Instrumentation*, vol. 4, pp. P12018–P12018, dec 2009.

[5] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, p. S08003, 2008.

[6] CMS Collaboration, "The CMS experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[7] Particle Data Group, "Review of particle physics," *Phys. Rev. D*, vol. 98, p. 030001, Aug 2018.

[8] CDF Collaboration, "Measurement of $b(t \to wb)/b(t \to wq)$ in top-quark-pair decays using dilepton events and the full cdf run ii data set," *Phys. Rev. Lett.*, vol. 112, p. 221801, Jun 2014.

[9] CDF Collaboration, "Measurement of $r{=}(t \to wb)/\mathscr{B}(t \to wq)$ in top-quark-pair decays using lepton+jets events and the full cdf run ii dataset," *Phys. Rev. D*, vol. 87, p. 111101, Jun 2013.

[10] D0 Collaboration, "Precision measurement of the ratio $\mathscr{B}(t \to wb)/\mathscr{B}(t \to wq)$ and extraction of V_{tb}," *Phys. Rev. Lett.*, vol. 107, p. 121802, Sep 2011.

[11] CMS Collaboration, "Measurement of the ratio $\mathscr{B}(t \to wb)/\mathscr{B}(t \to wq)$ in pp collisions at s=8 tev," *Physics Letters B*, vol. 736, pp. 33 – 57, 2014.

[12] CDF Collaboration, "Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab," *Phys. Rev. Lett.*, vol. 74, pp. 2626–2631, Apr 1995.

[13] D0 Collaboration, "Observation of the Top Quark," *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, Apr 1995.

[14] DONUT Collaboration, "Observation of tau neutrino interactions," *Physics Letters B*, vol. 504, no. 3, pp. 218 – 224, 2001.

[15] Higgs, Peter W., "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964. [,160(1964)].

[16] Englert, F. and Brout, R., "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, 1964. [,157(1964)].

[17] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[18] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 1 – 29, 2012.

[19] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 30 – 61, 2012.

[20] The Super-Kamiokande Collaboration, "Measurement of the Flux and Zenith-Angle Distribution of Upward Throughgoing Muons by Super-Kamiokande," *Phys. Rev. Lett.*, vol. 82, pp. 2644–2648, Mar 1999.

[21] SNO Collaboration, "Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory," *Phys. Rev. Lett.*, vol. 89, p. 011301, Jun 2002.

[22] KamLAND Collaboration, "First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance," *Phys. Rev. Lett.*, vol. 90, p. 021802, Jan 2003.

[23] Sheldon L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics*, vol. 22, no. 4, pp. 579 – 588, 1961.

[24] Weinberg, Steven, "A Model of Leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967.

[25] Salam, Abdus, "Weak and Electromagnetic Interactions," *Conf. Proc.*, vol. C680519, pp. 367–377, 1968.

[26] ATLAS, CDF, CMS and D0 collaborations, "First combination of Tevatron and LHC measurements of the top-quark mass," *arXiv:1403.4427*, 2014.

[27] Ulrich Husemann, "Top-Quark Physics: Status and Prospects," *arXiv:1704.01356*, 2017.

[28] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method," *Eur. Phys. J. C*, vol. 71, p. 1547, 2011.

[29] Chris D White and Stefano Frixione and Eric Laenen and Fabio Maltoni, "Isolating Wt production at the LHC," *Journal of High Energy Physics*, vol. 2009, pp. 074–074, nov 2009.

[30] Y. Amhis, S. Banerjee, R. Bernhard, S. Blyth, A. Bozek, C. Bozzi, A. Carbone, A. Campos, R. Chistov, G. Cibinetto, J. Coleman, J. Dingfelder, W. Dungel, M. Gersabeck, T. Gershon, L. Gibbons, B. Golob, R. Harr, K. Hayasaka, and R. Van Kooten, "Average of b- hadron and τ-lepton properties as of early 2012," 07 2012.

[31] L. Collaboration vol. 15, p. 053021, may 2013.

[32] J. Alwall, R. Frederix, J.-M. Gérard, A. Giammanco, M. Herquet, S. Kalinin, E. Kou, V. Lemaitre, and F. Maltoni, "Is $v_{tb} \approx 1$?," *The European Physical Journal C*, vol. 49, p. 791–801, Dec 2006.

[33] J. Swain and L. Taylor, "First determination of the quark mixing matrix element V_{tb} from electroweak corrections to Z decays," *Phys. Rev. D*, vol. 58, p. 093006, Sep 1998.

[34] CDF Collaboration and D0 Collaboration, "Tevatron Combination of Single-Top-Quark Cross Sections and Determination of the Magnitude of the Cabibbo-Kobayashi-Maskawa Matrix Element V_{tb}," *Phys. Rev. Lett.*, vol. 115, p. 152003, Oct 2015.

[35] ATLAS and CMS collaborations, "Combinations of single-top-quark production cross-section measurements and |fLVVtb| determinations at $\sqrt{s}$ = 7 and 8 TeV with the ATLAS and CMS experiments," *Journal of High Energy Physics*, vol. 2019, May 2019.

[36] LHC Top Working Group, "LHC Top Working Group summary plots, single top quark production," May 2018.

[37] B. Clerbaux, W. Fang, A. Giammanco, and R. Goldouzian, "Model-independent constraints on the ckm matrix elements | vtb | , | vts | and | vtd | ," *Journal of High Energy Physics*, vol. 2019, Mar 2019.

[38] H. Lacker, A. Menzel, F. Spettel, D. Hirschbühl, J. Lück, F. Maltoni, W. Wagner, and M. Zaro, "Model-independent extraction of $|v_{tq}|$ matrix elements from top-quark measurements at hadron colliders," *The European Physical Journal C*, vol. 72, June 2012.

[39] Pedro Silva, Michele Gallinaro, "Probing the flavor of the top quark decay," *Il Nuovo Cimento B*, vol. 125, pp. 983–998, Sept. 2010.

[40] ATLAS Collaboration, "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$=13 TeV," 2019.

[41] ATLAS Collaboration, "Performance of b-jet identification in the ATLAS experiment," *Journal of Instrumentation*, vol. 11, pp. P04008–P04008, apr 2016.

[42] E. Mobs, "The CERN accelerator complex. Complexe des acelerateurs du CERN," Jul 2016. General Photo.

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

[44] J. Wenninger, "LHC Status and Performance," in *PoS(CHARGED 2018)001*.

[45] ATLAS Luminosity Group, "https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2,"

[46] ATLAS Collaboration, "The ATLAS Inner Detector commissioning and calibration," *Eur. Phys. J. C*, vol. 70, p. 787, 2010.

[47] Gc Aad et al., "ATLAS pixel detector electronics and sensors," *Journal of Instrumentation*, vol. 3, pp. P07007–P07007, jul 2008.

[48] M. Capeans, G. Darbo, K. Einsweiler, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet, "ATLAS Insertable B-Layer Technical Design Report," 2010.

[49] ATLAS Collaboration, "Operation and performance of the ATLAS semiconductor tracker," *JINST*, vol. 9, p. P08009, 2014.

[50] M. Aaboud *et al.*, "Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties," *JINST*, vol. 12, no. 05, p. P05002, 2017.

[51] H. Wilkens and the ATLAS LArg Collaboration, "The ATLAS liquid argon calorimeter: An overview," *Journal of Physics: Conference Series*, vol. 160, p. 012043, apr 2009.

[52] ATLAS Collaboration, "Monitoring and data quality assessment of the ATLAS liquid argon calorimeter," *JINST*, vol. 9, p. P07024, 2014.

[53] ATLAS Collaboration, "Readiness of the ATLAS Tile Calorimeter for LHC collisions," *Eur. Phys. J. C*, vol. 70, p. 1193, 2010.

[54] A. Henriques, "The ATLAS tile calorimeter," in *Proceedings, 4th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA 2015): Lisbon, Portugal, April 20-24, 2015*, p. 7465554, 2015.

[55] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS, Geneva: CERN, 1997.

[56] ATLAS Collaboration, "Commissioning of the ATLAS Muon Spectrometer with cosmic rays," *Eur. Phys. J. C*, vol. 70, p. 875, 2010.

[57] J. Dubbert, S. Horvat, F. Legger, O. Kortner, H. Kroha, R. Richter, C. Valderanis, F. Rauscher, and A. Staude, "Performance of the ATLAS Muon Drift-Tube Chambers at High Background Rates and in Magnetic Fields," 2016.

[58] G. C. and, "The resistive plate chambers of the ATLAS experiment: performance studies," *Journal of Physics: Conference Series*, vol. 280, p. 012001, feb 2011.

[59] K. Nikolopoulos *et al.*, "Cathode strip chambers in ATLAS : Installation, commissioning and in situ performance," in *Proceedings, 2008 IEEE Nuclear Science Symposium, Medical Imaging Conference and 16th International Workshop on Room-Temperature Semiconductor X-Ray and Gamma-Ray Detectors (NSS/MIC 2008 / RTSD 2008): Dresden, Germany, October 19-25, 2008*, pp. 2819–2824, 2008.

[60] K. Nagai, "Thin gap chambers in ATLAS," *Nucl. Instrum. Meth.*, vol. A384, pp. 219–221, 1996.

[61] M. zur Nedden, "The LHC run 2 ATLAS trigger system: design, performance and plans," *Journal of Instrumentation*, vol. 12, pp. C03024–C03024, mar 2017.

[62] Aaboud, M., Aad, G., Abbott, B. et al., "Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2," *Eur. Phys. J. C*, vol. 77, no. 673, 2017.

[63] ATLAS Collaboration, "Charged-particle distributions in $\sqrt{s} = 13$TeV pp interactions measured with the ATLAS detector at the LHC," *Physics Letters B*, vol. 758, pp. 67 – 88, 2016.

[64] Aaboud, M., Aad, G., Abbott, B. et al., "Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton?proton collisions at the LHC," *Eur. Phys. J. C*, vol. 77, no. 332, 2017.

[65] "Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s}$ = 13TeV"," Tech. Rep. ATL-PHYS-PUB-2015-026, CERN, Geneva, Jul 2015.

[66] M. Aaboud *et al.*, "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.*, vol. C79, no. 8, p. 639, 2019.

[67] G. Aad *et al.*, "Muon reconstruction performance of the ATLAS detector in protonproton collision data at $\sqrt{s}$ =13 TeV," *Eur. Phys. J.*, vol. C76, no. 5, p. 292, 2016.

[68] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, "Calorimeter clustering algorithms: Description and performance," 2008.

[69] G. Aad *et al.*, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1," *Eur. Phys. J.*, vol. C77, p. 490, 2017.

[70] M. Aaboud *et al.*, "Jet reconstruction and performance using particle flow with the ATLAS Detector," *Eur. Phys. J.*, vol. C77, no. 7, p. 466, 2017.

[71] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[72] "Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS," Tech. Rep. ATLAS-CONF-2015-037, CERN, Geneva, Aug 2015.

[73] "Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV," Tech. Rep. ATL-PHYS-PUB-2015-015, CERN, Geneva, Jul 2015.

[74] "Tagging and suppression of pileup jets with the ATLAS detector," Tech. Rep. ATLAS-CONF-2014-018, CERN, Geneva, May 2014.

[75] M. Aaboud *et al.*, "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s}$ = 13 TeV," *Eur. Phys. J.*, vol. C78, no. 11, p. 903, 2018.

[76] ATLAS Collaboration, "Secondary vertex finding for jet flavour identification with the ATLAS detector," 2017.

[77] ATLAS Collaboration, "Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run," 2017.

[78] ATLAS Collaboration, "Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run," 2016.

[79] ATLAS Collaboration, "Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC," 2018.

[80] R. Frühwirth, "Application of kalman filtering to track and vertex fitting," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2, pp. 444 – 450, 1987.

[81] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.

[82] A. Buckley, "ATLAS Pythia 8 tunes to 7 TeV data," Tech. Rep. ATL-PHYS-PROC-2014-273, CERN, Geneva, Dec 2014.

[83] R. D. Ball *et al.*, "Parton distributions with LHC data," *Nucl. Phys.*, vol. B867, pp. 244–289, 2013.

[84] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth.*, vol. A462, pp. 152–155, 2001.

[85] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.

[86] G. Cowan, "Statistical Data Analysis," *Oxford science publications*, vol. Clarendon Press, 1998.

[87] BABAR Collaboration, "Measurement of the electron energy spectrum and its moments in inclusive $B \to Xev$ decays," *Phys. Rev. D*, vol. 69, p. 111104, Jun 2004.

[88] DELPHI Collaboration, "Determination of heavy quark non-perturbative parameters from spectral moments in semileptonic B decays," *Eur. Phys. J.*, vol. C45, p. 35, 2006.

[89] HFLAV Group, "Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016," 2017.

[90] ATLAS Collaboration, "Luminosity Determination in pp Collisions at sqrt(s) = 13 TeV using the ATLAS Detector at the LHC,"

[91] G. Avoni *et al.*, "The new lucid-2 detector for luminosity measurement and monitoring in atlas," *JINST*, vol. 13, no. 07, p. P07017, 2018.

[92] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP*, vol. 06, p. 043, 2010.

[93] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP*, vol. 11, p. 040, 2004.

[94] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method," *JHEP*, vol. 11, p. 070, 2007.

[95] M. Czakon and A. Mitov, "Top++: A program for the calculation of the top-pair cross-section at hadron colliders," *Computer Physics Communications*, vol. 185, no. 11, pp. 2930 – 2938, 2014.

[96] J. Bellm *et al.*, "Herwig 7.0/Herwig++ 3.0 release note," *Eur. Phys. J. C*, vol. 76, no. 4, p. 196, 2016.

[97] M. Bahr *et al.*, "Herwig++ Physics and Manual," *Eur. Phys. J. C*, vol. 58, p. 639, 2008.

[98] Harland-Lang, L.A., Martin, A.D., Motylinski, P. et al., "Parton distributions in the LHC era: MMHT 2014 PDFs," *Eur. Phys. J. C*, vol. 75, p. 204, 2015.

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

[100] Nikolaos Kidonakis, "Next-to-next-to-leading logarithm resummation for *s*-channel single top quark production," *Phys. Rev. D*, vol. 81, p. 054028, Mar 2010.

[101] J. M. Campbell and R. K. Ellis, "Update on vector boson pair production at hadron colliders," *Phys. Rev. D*, vol. 60, p. 113006, Nov 1999.

[102] Hoche, S., Krauss, F., Schonherr, M. et al., "QCD matrix elements + parton showers. The NLO case," *J. High Energ. Phys.*, vol. 04, p. 027, 2013.

[103] C. Anastasiou, L. Dixon, K. Melnikov, and F. Petriello, "High-precision qcd at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order," *Phys. Rev. D*, vol. 69, p. 094008, May 2004.

[104] Alwall, J., Frederix, R., Frixione, S. et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *J. High Energ. Phys.*, vol. 07, p. 079, 2014.

[105] Frixione, S., Hirschi, V., Pagani, D. et al., "Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons," *J. High Energ. Phys.*, vol. 06, p. 184, 2015.

[106] D. de Florian et al., "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector," *arXiv:*, vol. 1610.07922 [hep-ph], 2016.

[107] ATLAS Collaboration, "Performance of the ATLAS Trigger System in 2015," *Eur. Phys. J.*, vol. C77, no. 5, p. 317, 2017.

[108] ATLAS Collaboration, "Performance of electron and photon triggers in ATLAS during LHC Run 2," 2019.

[109] ATLAS Collaboration, "Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s} = 8$ TeV," *Eur. Phys. J.*, vol. C75, p. 120, 2015.

[110] T. A. Collaboration, "Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton-proton collisions at $\sqrt{s}$ = 13 tev with the atlas detector," *J. High Energ. Phys.*, vol. 46, no. 46.

[111] Wouter Verkerke, David Kirkby, "The RooFit toolkit for data modeling," *arXiv/physics/0306116v1*.

[112] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II," *J. Phys.*, vol. G43, p. 023001, 2016.

[113] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-," *Phys. Rev. D*, vol. 82, p. 054018, Sep 2010.

[114] P. Kant, O. Kind, T. Kintscher, T. Lohse, T. Martini, S. Malbitz, P. Rieck, and P. Uwer, "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions," *Computer Physics Communications*, vol. 191, pp. 74 – 89, 2015.

[115] ATLAS Collaboration, "Probing the quantum interference between singly and doubly resonant top-quark production in pp collisions at $\sqrt{s} = 13$ TeV with the atlas detector," *Phys. Rev. Lett.*, vol. 121, p. 152002, Oct 2018.

[116] ATLAS Collaboration, "Multi-boson simulation for 13 TeV ATLAS analyses," *ATL-PHYS-PUB-2016-002*, 2016.

[117] ATLAS Collaboration, "Monte Carlo Generators for the Production of a W or Z/γ^* Boson in Association with Jets at ATLAS in Run 2," *ATL-PHYS-PUB-2016-003*, 2016.

[118] A. Collaboration, "Measurement of the $t\bar{t}$z and $t\bar{t}$w cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the atlas detector," *Phys. Rev. D*, vol. 99, p. 072009, Apr 2019.

[119] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J.*, vol. C71, p. 1554, 2011. [Erratum: Eur. Phys. J.C73,2501(2013)].

[120] J. S. Gainer, J. Lykken, K. T. Matchev, S. Mrenna, and M. Park, "The matrix element method: Past, present, and future," 2013.

[121] M. Meinhard, "Reconstruction of $t\bar{t}$h, h $\rightarrow$ b$\bar{b}$ events using the matrix element method and substructure techniques," 2018.

Acknowledgements

This work would not have been possible without the love of my familiy and friends and the support of all those who surrounded me professionally.

I am deeply grateful to my supervisor, Giuseppe Salamanna, for his endless support. I will remember the afternoons spent discussing new ideas at the whiteboard as the most intellectually stimulating of my PhD years. Thanks for the confidence in my abilities, the motivation and guidance.

A special thank you to all the members of the Roma Tre ATLAS group. This work is also the result of your support, both professional and personal.

I would like to thank all my colleagues in the FTAG and TOP working groups for their help. A special thank you to Bingxuan Liu for the support.

Thank you to all the amazing human beings I had the chance to meet during this journey. Thomas, Dylan, Francesco, Veronica, Blake and Camilla: my months at CERN will be forever linked to all the moments, songs, gigs and experiences I shared with you.

A special thank you to Francesco, Ludovica and Andrea the best friends I could have wished for.

And most crucially I would like to express my deepest gratitude to my family for their continuous patience, support and love.