PROCEEDINGS OF SCIENCE

$t\bar{t}$+heavy flavour measurements with the full run 2 dataset using the ATLAS experiment

Ricardo Gonçalo on behalf of the ATLAS Collaboration[a,b,*]

[a] *Laboratório de Instrumentação e Física Experimental de Partículas – LIP Av. Prof. Gama Pinto 2, 1649-003 Lisbon, Portugal*

[b] *Departamento de Física, Universidade de Coimbra, Universidade de Coimbra Rua Larga, 3004-516 Coimbra, Portugal*

E-mail: jgoncalo@lip.pt, jgoncalo@uc.pt

The production of top quark pairs in association with heavy-flavour jets (b/c) is a difficult process to calculate and model, and is one of the leading sources of background affecting the measurement of processes such as the associated production of a Higgs boson and a top-quark pair, $t\bar{t}H$. To improve our understanding of this process, a new inclusive and differential measurement of the production of $t\bar{t}$ accompanied by a jet initiated by b quarks. The results presented were achieved using the full LHC Run-2 dataset collected by ATLAS.

ATL-PHYS-PROC-2024-089
16 October 2024

42nd International Conference on High Energy Physics (ICHEP2024)
18-24 July 2024
Prague, Czech Republic

[*] Speaker

1. Introduction

These proceedings summarise a new measurement [1] by the ATLAS experiment [2] of the $t\bar{t}+b$-jets fiducial cross sections at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of $\mathcal{L} = 140\,\text{fb}^{-1}$. The production of a top–antitop quark pair ($t\bar{t}$) in association with additional jets challenges our current ability to model complex quantum chromodynamics (QCD) processes. Measurements of such processes at the LHC provide essential tests of QCD and also crucial experimental constraints, that will allow to improve our modelling tools such as Monte Carlo (MC) generators. Among these processes, $t\bar{t}$ events produced in association with jets originating from b-quark hadronisation ($t\bar{t}+b$-jets) are particularly challenging due to the non-zero mass of the b-quark, as well as the very different scales of $t\bar{t}$ and $b\bar{b}$ production. Furthermore, this process constitutes an especially difficult (often irreducible) background to other important measurements at LHC experiments. One crucial example is the direct measurements of the top quark Yukawa coupling in associated production of the Higgs boson with one or two top quarks, tH or $t\bar{t}H$, with $H \to b\bar{b}$ decays [3–5]. In fact, modelling of the $t\bar{t}+b$-jets background has been the main source of systematic uncertainty affecting this analysis and severely limits its sensitivity. Similarly, the production of $t\bar{t}Z$ with $Z \to b\bar{b}$ faces significant challenges due to large backgrounds, making it crucial to enhance our understanding of QCD $t\bar{t}+b$-jets production.

Different MC simulations of the $t\bar{t}+b$-jets process implement different treatments of the b-quark mass, the calculation of b-quark production in the matrix element at next-to-leading order (NLO) or parton shower, and the merging of b-quark production into the inclusive $t\bar{t}$ prediction. These variations lead to different predictions for observables such as angular jet distributions, jet transverse momenta (p_{T}), etc. State-of-the-art QCD calculations give predictions for the $t\bar{t}$ production cross-section with up to two additional massless partons at NLO in perturbation theory matched to a parton shower algorithm. QCD predictions for $t\bar{t}b\bar{b}$ are also calculated at NLO matched to a parton shower algorithm.

Several inclusive and normalised differential fiducial cross sections were measured by unfolding observed data to particle level. These were compared with predictions from state-of-the-art MC generators (see Ref. [1] and references therein) in order to probe the accuracy of different aspects of the modelling. Only the $e\mu$ channel was considered in this analysis, in which both W bosons from $t\bar{t}$ decays, decay leptonically producing an electron and a muon. This channel allows a particularly performant event selection and reconstruction, minimising contamination from several sources of background as well as combinatorial reconstruction errors. Results shown here may be compared to previous measurements of the $t\bar{t}+b$-jets cross section by the ATLAS [6] and CMS collaborations [7]. Most of the measured cross section values were found to be higher than most MC predictions.

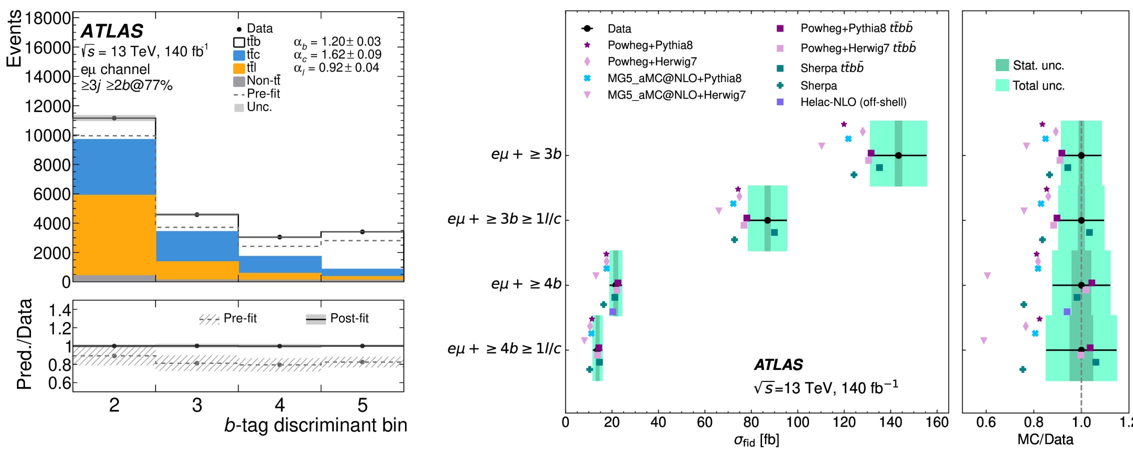
2. Analysis Description

Collision events are preselected if they pass a combination of single-electron and single-muon triggers, designed to have a high efficiency above p_{T} of 26 GeV. The offline selection proceeded with the identification of events containing exactly one electron and one muon candidate, both isolated and emerging from the primary vertex, with p_{T} above 28 GeV and pseudorapidity within $|\eta| < 2.5$.

Electron and muon must have opposite charge and their invariant mass, $m_{e\mu}$, was required to be above 15 GeV to suppress background from Drell-Yan events with two τ-leptons.

Selected events were required to contain at least two Particle-Flow jets [8] associated with the primary vertex. Hadronic jets were reconstructed with the anti-k_t algorithm [9], with a radius parameter of $R = 0.4$, and had p_{T}> 25 GeV and $|\eta| < 2.5$. Jets initiated by b-quarks (b-jets), were identified with the DL1r algorithm [10], using a baseline working point corresponding to a 77% b-jet tagging efficiency and rejection factors of around 170 against light-flavoured jets (initiated by u-, d-, or s-quarks, or gluons) and 5 against jets originating from c-quarks.

This event selection yields a sample largely dominated by $t\bar{t}$ production (including $t\bar{t}$+light jets and $t\bar{t}+c$-jets) with other sources of background contributing only at a lower level. Remaining background events either contain genuine prompt leptons from top, W, or Z decays, such as $t\bar{t}H$ or Wt, or arise from misidentified lepton candidate. The latter can be a non-prompt lepton from the decay of a b- or c-hadron, an electron from a photon conversion, jets misidentified as electrons, or muons from the in-flight decay of a pion or kaon. Electroweak processes such as $t\bar{t}H$ or Wt production were estimated from MC. The non-prompt lepton background is estimated from collision data by selecting a sample of events containing an electron and a muon with equal electrical charge in each analysis region. This assumes that jet misidentification as electrons or muons yields both charges with equal probability. The non-prompt lepton background in each bin is obtained by scaling this sample in each analysis region by a factor determined from MC. Correction factors for the yields of $t\bar{t}+b$-jets, $t\bar{t}+c$-jets and $t\bar{t}$+light jets are estimated from a template fit in the inclusive analysis region containing events with at least three jets, two or more of which are b-tagged ($\geq 3j, \geq 2b$). Templates are constructed from the distributions of the third highest b-tagging discriminant, taken as a proxy for an additional b-jet in $t\bar{t}$ production. Such templates are constructed for the $t\bar{t}+b$-jets, $t\bar{t}+c$-jets, and $t\bar{t}$+light jets samples (including backgrounds such as $t\bar{t}H$), and fitted to data after subtracting non-$t\bar{t}$ background, as shown in Figure 1(a).



(a) **(b)**

Figure 1: (a): template fit to the distribution of the third highest b-tagging discriminant in data, to determine background and signal normalisation. (b): measured inclusive fiducial cross section unfolded to particle level [1].

In parallel to the above selection, applied to reconstructed collision events, simulated events undergo a similar event selection at particle-level, which is used to define the fiducial phase spaces for unfolding of cross sections [1]. Fiducial phase spaces are defined in simulated events by requiring exactly one electron and one muon at particle level, each with $p_{\mathrm{T}} \geq$ 28 GeV and opposite charge, and with two, three or four b-jets with $p_{\mathrm{T}} >$ 25 GeV. Samples containing only two b-jets (noted $\geq 2j, 2b$) are used to determine the normalisation of different backgrounds from collision data. The signal $t\bar{t} + b$-jets signal cross sections are extracted in analysis regions containing either three b-jets ($\geq 3j, 3b$) or at least four b-jets ($\geq 4j, \geq 4b$). The $\geq 3j, 3b$ region is interpreted as containing events with a $t\bar{t}$ pair plus an extra b-jet, while the $\geq 4j, \geq 4b$ region is assumed to correspond to events containing a $t\bar{t}$ pair plus at least two additional b-jets.

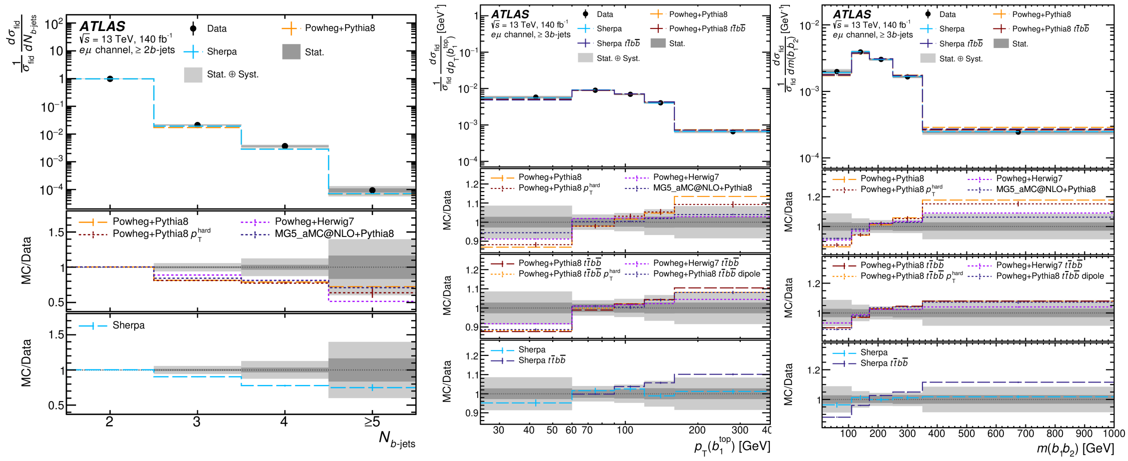
Normalised differential cross sections are calculated with respect to a large number of particle-level observables listed in Ref. [1]. These include observables representing overall event properties, sensitive to matrix elements and parton shower modelling, such as the number of b-jets in the event ($N_{b-\text{jets}}$, or the scalar sum $H_{\mathrm{T}}^{\text{had}}$ of the p_{T} of each jet. Other examples are observables such as the average angular distance $\Delta R_{\text{avg}}^{bb}$ between any two b-jets and the maximum separation $\Delta\eta_{\max}^{jj}$ in η between any two jets, which provide sensitivity to additional QCD emissions, or the p_{T} and η of up to four leading b-jets, (labelled b_1 to b_4 in order of decreasing p_{T}), with the leading b-jet kinematics dominantly reflecting the top quark kinematics. Observables sensitive to the dynamics of top quark decay products include the invariant mass $m(e\mu b_1 b_2)$ of the electron, muon, and two leading b-jets system. Additional b-jet activity is studied through observables such as the angular distance min$\Delta R(bb)$, invariant mass $m(bb^{\min\Delta R})$, and transverse momentum $p_{\mathrm{T}}(bb^{\min\Delta R})$ of the two closest b-jets.

Observables may be divided into two categories, in which b-jets are either ordered in p_{T}, or are identified as coming from the top-quark decays or additional radiation. This identification is based on an empirical weight function calculated from the angular separation between pairs of objects such as leptons and b-jets. Identified b-jets in the event are arranged into a set of permutations, each with a given weight, in which the first two b-jet positions are assigned to the top quark decays. The b-jet permutation with the largest weight is used to identify the origin of each jet.

2.1 Results

After obtaining the corrected event yields at detector and particle level, an iterative Bayesian method is employed to unfold to particle level the inclusive fiducial cross sections and several normalised differential fiducial cross sections [11]. The unfolded inclusive fiducial cross sections for events containing at least three b-jets, at least three b-jets plus one c-jet at least four b-jets, are shown in Figure 1(b). This unfolding employed the nominal $t\bar{t}$ simulated sample, obtained from the PowhegBox v2 heavy-quark event generator, calculated at NLO precision in QCD with the `NNPDF3.0NLO` PDF set in the five-flavour scheme (5FS). The statistical uncertainty is shown as a darker error band and the total uncertainty as a lighter band around the measured points. These are compared to predictions by several other MC generators as described in the figure. Systematic uncertainties arise primarily from b-tagging, the jet energy scale, or the modelling of $t\bar{t}$ production.

Normalised differential cross sections are obtained in a set of phase-space regions corresponding to events with different numbers of b-jets, light- or c-jets (l/c), with different observables being defined in each of these regions. The different observables are intended to provide sensitivity to



(a) **(b)** **(c)**

Figure 2: Three examples of normalised differential fiducial cross sections published in Ref. [1]: (a) the number of b-jets ($N_{b-\text{jets}}$), (b) the transverse momentum of the leading b-jet from top quark decays ($p_{\text{T}}(b_1^{\text{top}})$) in the $\geq 3b$-jets phase space, and (c) the invariant mass of the two leading b-jets ($m(b_1 b_2)$) in the same phase space region.

different (and complementary) aspects of the event modelling, including the matrix elements and parton showers, additional QCD radiation, recoil model, or the dynamics of top quark decays. All observables are defined at both the particle and detector levels. In the latter case, b-tagged jets are taken to correspond to b-jets. Figure 2 shows examples of normalised fiducial cross sections for three of these observables: (a) the number of b-jets, (b) the transverse momentum of the leading b-jet from top quark decays ($p_{\text{T}}(b_1^{\text{top}})$) in the $\geq 3b$-jets phase space, and (c) the invariant mass of the two leading b-jets ($m(b_1 b_2)$) in the same phase space region. As can be observed from these examples, calculations often fall outside of the experimental uncertainty.

2.2 Conclusions

The new ATLAS result [1] presented here publishes several inclusive and normalised differential fiducial cross sections for $t\bar{t} + b$-jets production, in the $e\mu$ channel at $\sqrt{s} = 13$ TeV. The significance of these results goes beyond a test of QCD since these processes constitute irreducible and dominant backgrounds to several important measurements at the LHC.

Particle-level cross sections are obtained in different phase spaces depending on the number of b- and l/c-jets, and achieve the best precision measurements in this channel, and also better than the NLO theoretical prediction for $t\bar{t}b\bar{b}$. Leading systematic uncertainties are from b-tagging, jet energy scale and $t\bar{t}$ modelling. Measured cross sections are compared to a number of state-of-the-art MC generators using a large number of observables sensitive to different aspects of the modelling. It is hoped that these results will allow to make progress in the theoretical understanding of these important processes for the LHC physics programme.

References

[1] ATLAS Collaboration, *Measurement of $t\bar{t}$ production in association with additional b-jets in*

the $e\mu$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, submitted to JHEP [2407.13473].

[2] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003 [2305.16623].

[3] ATLAS Collaboration, *Measurement of the associated production of a top-antitop-quark pair and a Higgs boson decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC*, *submitted to EPJC* [2407.10904].

[4] ATLAS Collaboration, *Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **06** (2022) 097 [2111.06712].

[5] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, Phys. Rev. Lett. **121** (2018) 121801 [1808.08242].

[6] ATLAS 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*, *JHEP* **04** (2019) 046 [1811.12113].

[7] CMS Collaboration, *Inclusive and differential cross section measurements of $t\bar{t}b\bar{b}$ production in the lepton+jets channel at $\sqrt{s} = 13$ TeV*, 2309.14442.

[8] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) 466 [1703.10485].

[9] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [0802.1189].

[10] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, *Eur. Phys. J. C* **83** (2023) 681 [2211.16345].

[11] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **761** (2016) 136 [1606.02699].