PROCEEDINGS OF SCIENCE

Top quark pair production with heavy-flavor jets

Ana Luísa Carvalho for the ATLAS and CMS Collaborations[a,*]

[a] *DESY,*
Notkestrasse 85, Hamburg, Germany

E-mail: ana.luisa.carvalho@desy.de

The top-quark pair production in association with heavy-flavour jets (b- or c-jets) is a difficult process to calculate and model and is one of the leading sources of background to Standard Model measurements of $t\bar{t}H$ and $t\bar{t}t\bar{t}$ in final states with one or two leptons. To improve our understanding of these processes, $t\bar{t} + b\bar{b}$ inclusive and differential cross-section measurements were performed by the ATLAS and CMS Collaborations, and an inclusive $t\bar{t} + c\bar{c}$ cross-section measurement was performed by the CMS Collaboration.

12th Large Hadron Collider Physics Conference (LHCP2024)
3-7 June 2024
Boston, USA

*Speaker

PoS(LHCP2024)134

1. Introduction and motivation

The production of top quark pairs in association with heavy-flavor jets is a very abundant process at the LHC. In this context, a heavy-flavor jet is defined as a jet initiated by a b- or c-quark. Hadrons containing b- or c-quarks have a relatively long lifetime of a few picoseconds and therefore travel for a few milimeters inside the tracking system before decaying, producing a secondary vertex. Using the properties of the secondary vertex, as well as the kinematic properties of the jet, multivariate algorithms can distinguish (tag) b- and c-jets from light-flavored jets. To define b- and c-tagging working points, selections are applied on the output of these algorithms with thresholds chosen to have a fixed efficiency to select true b- and c-jets. A tighter working point corresponds to a smaller b- or c-jet efficiency and larger rejection factor for the remaining jet flavors.

Precisely measuring the cross-sections of $t\bar{t}+b$-jets and $t\bar{t}+c$-jets, both inclusively and differentially, constitutes an important test of quantum chromodynamics. Additionally, these processes are often leading background contributions in searches for beyond the Standard Model physics and in precision SM measurements, such as $t\bar{t}H(H \rightarrow b\bar{b})$ and four top quark production. Depending on the chosen final state, they can even be irreducible backgrounds, as it is the case of $t\bar{t}+b\bar{b}$ in $t\bar{t}H(H \rightarrow b\bar{b})$ analyses, and therefore extremely hard to isolate.

There are multiple challenges affecting the measurements of $t\bar{t}+b$-jets and $t\bar{t}+c$-jets. Theoretically, the very different energy scales associated with the $t\bar{t}$ and $b\bar{b}$ or $c\bar{c}$ pairs, and the non-negligible mass of the b-quark complicate the calculations. Experimentally, flavor-tagging algorithms have associated inefficiencies and often large systematic uncertainties and the measurements usually have to rely on Monte Carlo simulation which is known to not reproduce data well, especially in remote regions of the phase space.

As an example, Figure 1 shows the leading order Feynamn diagram of $t\bar{t}$ production with two additional c-jets. The top quark decays to a b-quark and a W-boson with a branching fraction of 99.8%. The W-boson can then decay to two quarks, or to a charged lepton and a neutrino. If both W-bosons decay to leptons, as in the diagram of Figure 1, there are two isolated, high-energy leptons in the event and this is referred to as dileptonic channel. If one W-boson decays to leptons and the other to quarks, there is only one lepton in the final state and this topology is referred to as single-lepton channel.

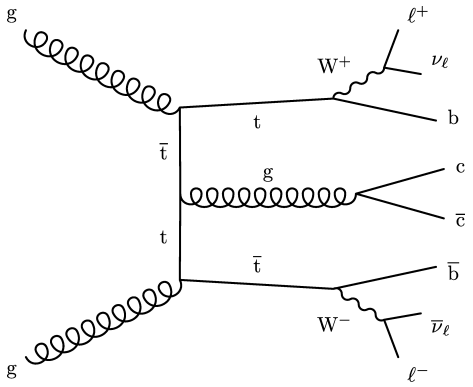

Figure 1: Example of a Feynman diagram at the lowest order in QCD, describing the dileptonic decay channel of a top quark pair with an additional c-quark pair produced via gluon splitting.

Table 1: Summary of the analyses covered in these proceedings, including the integrated luminosity, $\mathcal{L}$, and the channels used.

Measurement	ATLAS	CMS
$t\bar{t}$ + b-jets (inclusive and differential)	[1] $\mathcal{L}$ = 36.1 fb^{-1} Dilepton ($e\mu$) and single-lepton	[2] $\mathcal{L}$ = 138 fb^{-1} Single-lepton
$t\bar{t}$ + c-jets (inclusive)	—	[3] $\mathcal{L}$ = 45.1 fb^{-1} Dilepton

The analyses covered in these proceedings are summarized in Table 1. They include $t\bar{t}+b$-jets inclusive and differential cross-section measurements performed by the ATLAS [1] and CMS [2] Collaborations, detailed in Section 2, and an inclusive $t\bar{t}$ + c-jets measurement performed by the CMS Collaboration [3], discussed in Section 3.

2. Measurements of $t\bar{t}$ + $b-$jets

Measurements of inclusive and differential $t\bar{t}$ + b-jets cross-sections are performed by the ATLAS [1] and CMS [2] Collaborations in multiple fiducial phase space regions defined based on the number of truth-level leptons and jets. These regions target different aspects of $t\bar{t}$ + b-jets production and their definition follows as closely as possible the requirements imposed on reconstructed objects.

Truth-level electrons are required to have $p_{\mathrm{T}} > 25(29)$ GeV while for muons the requirements are $p_{\mathrm{T}} > 25(26)$ GeV for ATLAS (CMS). Truth-level jets are defined by clustering stable particles using the anti-k_{t} algorithm with radius parameter $R = 0.4$, and required to have $p_{\mathrm{T}} > 25$ GeV.

In the ATLAS dilepton channel, fiducial regions with exactly one electron and one muon with opposite charges, and at least three (four) b-jets target $t\bar{t}$ + b ($t\bar{t}$ + $b\bar{b}$) production. In the ATLAS and CMS single-lepton channel, fiducial regions with at least five jets and at least three b-jets, or at least six jets and at least four b-jets, target $t\bar{t}$ + b and $t\bar{t}$ + $b\bar{b}$ production, respectively. In the CMS measurement, two additional fiducial phase space regions are defined, with at least six jets, at least three b-jets and at least three light-jets, or with at least seven jets, at least four b-jets and at least three light-jets. These regions target $t\bar{t}+bj$ and $t\bar{t}+b\bar{b}j$, respectively, where j stands for a light-jet.

The signal and most background contributions are estimated from Monte Carlo simulation. In the ATLAS measurement, the $t\bar{t}$ + b-jets component is estimated through a five-flavor scheme inclusive $t\bar{t}$ sample produced using Powheg+Pythia8 [4–7], with factorization (μ_{F}) and renormalization (μ_{R}) scales set to the top quark transverse mass. The same sample is used to estimate the contributions from $t\bar{t}$ + c-jets and $t\bar{t}$+light jets by both ATLAS and CMS. In the CMS measurement, the $t\bar{t}$ + b-jets contribution is estimated using a four-flavor scheme $t\bar{t}b\bar{b}$ sample produced using Powheg+OpenLoops+Pythia8 [8, 9], with $\mu_{\mathrm{F}} = \frac{1}{4}H_T$ and $\mu_{\mathrm{R}} = \frac{1}{2}\sum_i m_{T,i}^{1/4}$, where $H_T = \sum_{t,\bar{t},b,\bar{b},g} m_{T,i}$ and $m_{T,i} = \sqrt{m_i^2 + p_{\mathrm{T},i}^2}$ is the transverse mass.

Particle-level differential cross-sections are measured for a large variety of observables, including global quantities such as H_T and kinematic properties such as the transverse momentum of the leading b-jet. Variables are also chosen for their sensitivity to different aspects of the modeling of

$t\bar{t}$ production with additional b-jets. For example, the angular opening between the two additional b-jets, $\Delta R(b\bar{b})$, is sensitive to the modeling of gluon splitting to $b\bar{b}$.

For each observable, the cross-section is measured by a dedicated fit. In the CMS measurement, a binned maximum likelihood fit, which simultaneously yields the inclusive cross-section and the unfolded normalized differential distribution, is performed. The number of b-jets passing the 60% working point is used as an ancillary variable to define signal- and background-enriched regions which are fitted simultaneously. The inclusive cross-section in each fiducial region is reported for the observable for which the measured value of the cross-section is closest to the mean of all measured cross-section. In the ATLAS measurement, a two-step signal extraction process is used. The normalizations of $t\bar{t}$ + b-, c- and light-jets are extracted from binned maximum likelihood fits to the b-tagging scores in regions with at least three b-jets tagged with the 77% working point, in dilepton region, or with at least five jets and at least two b-jets tagged at 60% working point, in the single-lepton channel. In each fiducial region, the non-$t\bar{t}$ background is subtracted from data in each bin, at detector level. The flavor composition is corrected for using the normalization factors derived from the likelihood fit. The fiducial acceptance, fiducial- and detector-level matching and reconstruction efficiency are also corrected for. Particle-level unfolding is then performed in each fiducial region.

The inclusive cross-sections measured by the ATLAS and CMS Collaborations in each fiducial region are shown in Figures 2 and 3, respectively. The predictions from multiple generators are also shown. In the single-lepton regions with at least five jets and at least three b-jets, and with at least six jets and at least four b-jets, the measured values of the inclusive cross-sections are directly comparable between the two experiments and they are in agreement within the uncertainties. The measurements are well described by the generator predictions, the notable exception being the Powheg+OpenLoops+Pythia8 sample in the CMS measurement, which predicts larger cross-sections in all fiducial regions except in the one with at least five jets and at least three b-jets.

In both the ATLAS and CMS measurements, the b-tagging efficiency is a leading systematic uncertainty.

3. Measurement of $t\bar{t} + c\bar{c}$

This analysis [3], performed by the CMS Collaboration, presents the cross-section measurement of $t\bar{t}$ pairs with two additional c-jets ($t\bar{t} + c\bar{c}$), in the fiducial and generator phase spaces, defined using truth-level stable particles and following as closely as possible the detector-level selections. The fiducial phase space is defined by the presence of two oppositely charged electrons, muons or tau-leptons with $p_{\mathrm{T}} > 25$ GeV. Each lepton is required to originate from a W-boson that in turn originates from a top quark decay. The particle-level jets originating from the top quark decay must be b-jets and have $p_{\mathrm{T}} > 20$ GeV. At least two additional particle-level jets must be present in the event. At detector-level the assignment of jets to partons is performed using a neural network.

The $t\bar{t} + c\bar{c}$ signal and the $t\bar{t}$ + b- and light-jets background contributions are estimated from an inclusive $t\bar{t}$ five-flavor scheme Monte Carlo sample produced with Powheg+Pythia8. The factorization and renormalization scales are set to $\sqrt{m_t^2 + p_{\mathrm{T},t}^2}$, where $m_t = 172.5$ GeV is the top quark mass and $p_{\mathrm{T},t}$ its transverse momentum in the $t\bar{t}$ rest frame. The $t\bar{t}$ cross-section is

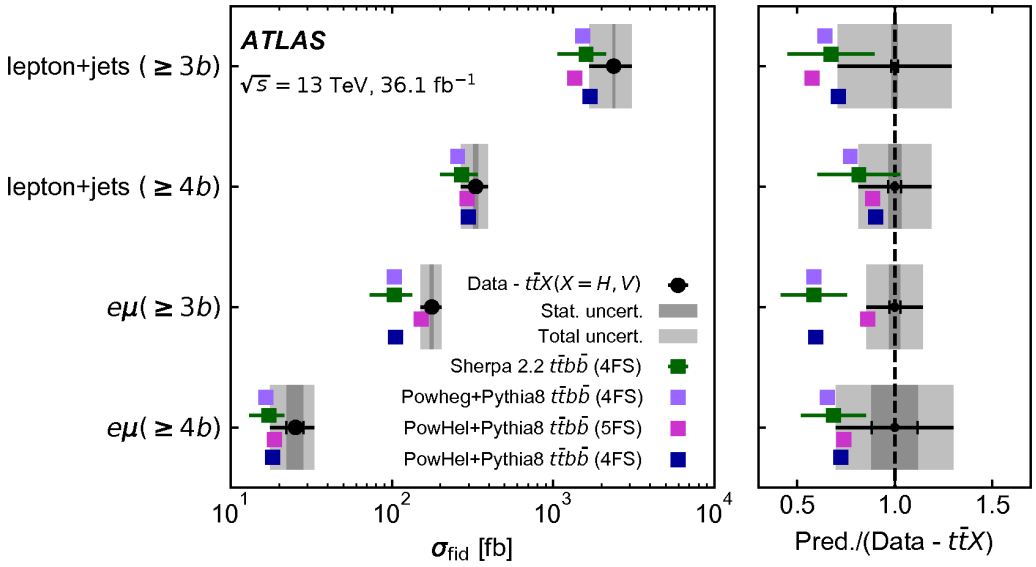

Figure 2: The measured fiducial cross-sections, with $t\bar{t}H$ and $t\bar{t}V$ contributions subtracted from data, compared with $t\bar{t}b\bar{b}$ predictions obtained using SHERPA 2.2 $t\bar{t}b\bar{b}$ with uncertainties obtained by varying the renormalisation and factorisation scales by factors of 0.5 and 2.0 and including PDF uncertainties. Comparisons with the central values of the predictions of POWHEG+PYTHIA8 and POWHEL+PYTHIA88 $t\bar{t}b\bar{b}$ are also made. No uncertainties are included in the subtraction of the $t\bar{t}H$ or $t\bar{t}V$ predictions [1], where V stands for a Z or $W^{\pm}$ boson.

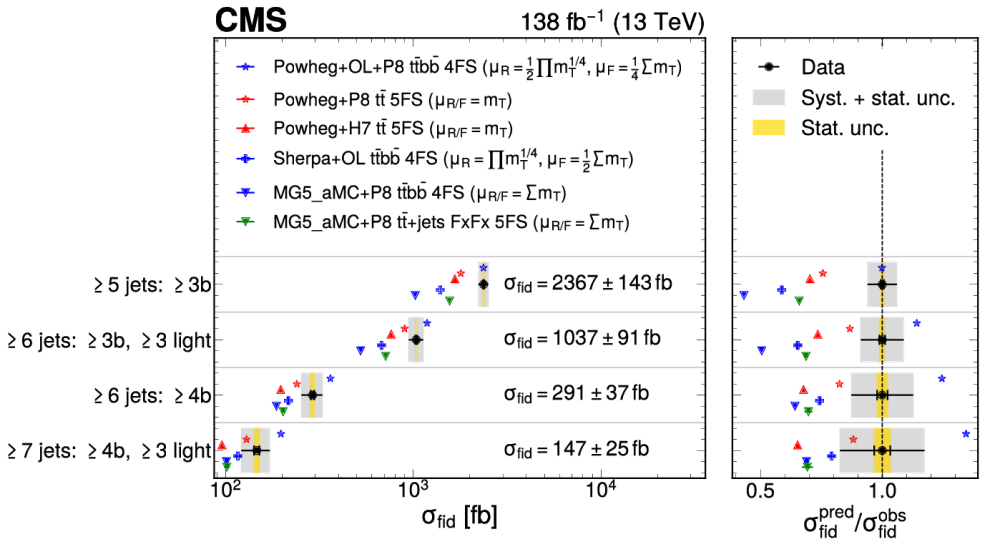

Figure 3: Measured inclusive cross sections for each considered phase space, compared to predictions from different $t\bar{t}b\bar{b}$ simulation approaches shown as coloured symbols. The predictions include uncertainties (horizontal bars) due to the limited number of simulated events. The blue colour is reserved for models using massive b quarks and NLO QCD $t\bar{t}b\bar{b}$ matrix elements, while red is used for the inclusive $t\bar{t}$ generators at NLO in QCD with massless b quarks. The right panel shows the ratios between the predicted and measured cross sections, with the black bars showing the relative uncertainties in the measurements [2].

	Result	POWHEG	MADGRAPH5_aMC@NLO
	Fiducial phase space		
$\sigma_{t\bar{t}c\bar{c}}$ [pb]	$0.207 \pm 0.025 \pm 0.027$	0.187 ± 0.038	0.189 ± 0.032
$\sigma_{t\bar{t}b\bar{b}}$ [pb]	$0.132 \pm 0.010 \pm 0.015$	0.097 ± 0.021	0.101 ± 0.023
$\sigma_{t\bar{t}LL}$ [pb]	$5.15 \pm 0.12 \pm 0.41$	5.95 ± 1.02	6.32 ± 0.94
R_c [%]	$3.01 \pm 0.34 \pm 0.31$	2.53 ± 0.18	2.43 ± 0.17
R_b [%]	$1.93 \pm 0.15 \pm 0.18$	1.31 ± 0.12	1.30 ± 0.16
	Full phase space		
$\sigma_{t\bar{t}c\bar{c}}$ [pb]	$10.1 \pm 1.2 \pm 1.4$	9.1 ± 1.8	8.9 ± 1.5
$\sigma_{t\bar{t}b\bar{b}}$ [pb]	$4.54 \pm 0.35 \pm 0.56$	3.34 ± 0.72	3.39 ± 0.66
$\sigma_{t\bar{t}LL}$ [pb]	$220 \pm 5 \pm 19$	255 ± 43	261 ± 37
R_c [%]	$3.36 \pm 0.38 \pm 0.34$	2.81 ± 0.20	2.72 ± 0.19
R_b [%]	$1.51 \pm 0.11 \pm 0.16$	1.03 ± 0.08	1.03 ± 0.09

Table 2: Measured parameter values in the fiducial (upper rows) and full (lower rows) phase spaces with their statistical and systematic uncertainties listed in that order. The last two columns display the expectations from the simulated $t\bar{t}$ samples using the POWHEG or MADGRAPH5_aMC@NLO matrix element generators. The uncertainties quoted for these predictions include the contributions from the theoretical uncertainties, as well as the uncertainty in the $t\bar{t}$ cross section [3].

scaled to a theoretical prediction at next-to-next-to-leading order in QCD including resummation of next-to-next-to-leading logarithmic soft-gluon terms, yielding $\sigma_{t\bar{t}} = 832^{+40}_{-46}$ pb.

A multiclass neural network is trained to predict the probability of five event classes: $P(t\bar{t}+c\bar{c})$, $P(t\bar{t}+cj)$, $P(t\bar{t}+b\bar{b})$, $P(t\bar{t}+bj)$ and $P(t\bar{t}+jj)$, using flavor-tagging information, the ΔR separation between the two additional jets and the score of the parton-jet assignment neural network for the best permutation. The signal extraction is performed via a two-dimensional binned maximum likelihood fit to two discriminators, targeting the separation between c- and b-jets (Δ^c_b) and between c- and light-jets (Δ^c_j):

$$\Delta^c_b = \frac{P(t\bar{t}+c\bar{c})}{P(t\bar{t}+c\bar{c})+P(t\bar{t}+b\bar{b})}, \quad \Delta^c_j = \frac{P(t\bar{t}+c\bar{c})}{P(t\bar{t}+c\bar{c})+P(t\bar{t}+j\bar{j})}. \tag{1}$$

In addition to the $t\bar{t}+c\bar{c}$ signal cross-section, the $t\bar{t}+b\bar{b}$ and $t\bar{t}+jj$ cross-sections are also free-floating parameters of the fit. An alternative fit is performed to extract the ratios of the $t\bar{t}+c\bar{c}$ and $t\bar{t}+b\bar{b}$ cross-sections to the inclusive $t\bar{t}$ cross-section with two additional jets.

The measured values of the cross-sections in the fiducial and generator phase spaces are summarized in Table 2. For $t\bar{t}+c\bar{c}$, the measured value of the cross-section in the fiducial phase space is $\sigma_{t\bar{t}c\bar{c}} = 0.207 \pm 0.025(\text{stat.}) \pm 0.027(\text{syst.})$, which is in agreement with the predictions from POWHEG and MADGRAPH5_aMC@NLO, within the uncertainties. The uncertainties associated with the calibration of the c-tagging algorithm are the leading source of systematic uncertainty.

References

[1] 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].

[2] 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*, *JHEP* **05** (2024) 042 [2309.14442].

[3] CMS collaboration, *First measurement of the cross section for top quark pair production with additional charm jets using dileptonic final states in pp collisions at s=13TeV*, *Phys. Lett. B* **820** (2021) 136565 [2012.09225].

[4] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [hep-ph/0409146].

[5] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [0709.2092].

[6] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126 [0707.3088].

[7] 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* **06** (2010) 043 [1002.2581].

[8] T. Ježo and P. Nason, *On the Treatment of Resonances in Next-to-Leading Order Calculations Matched to a Parton Shower*, *JHEP* **12** (2015) 065 [1509.09071].

[9] F. Buccioni, J.-N. Lang, J.M. Lindert, P. Maierhöfer, S. Pozzorini, H. Zhang̃ et al., *OpenLoops 2*, *Eur. Phys. J. C* **79** (2019) 866 [1907.13071].

PoS(LHCP2024)134