PROCEEDINGS OF SCIENCE

# Measurements of $t\bar{t}$+jets, $t\bar{t}b\bar{b}$ and $t\bar{t}t\bar{t}$ production with CMS

**Sébastien Wertz, on behalf of the CMS collaboration***
*University of Zurich*
*E-mail:* sebastien.wertz@cern.ch

Recent inclusive and differential cross section measurements of top quark pair production in association with jets and b jets are presented. The measurements are confronted with several theoretical predictions. In addition, the latest searches for four top quark production are presented. The results of these searches are used to place bounds on scenarios of physics beyond the standard model. All results have been obtained by the CMS experiment at the LHC, based on pp collision data collected at $\sqrt{s} = 13$ TeV.

*European Physical Society Conference on High Energy Physics - EPS-HEP2019 -*
*10-17 July, 2019*
*Ghent, Belgium*

*Speaker.



https://pos.sissa.it/

PoS(EPS-HEP2019)665

## 1. Introduction

Top quark pair production in association with jets or jets produced by the hadronization of b quarks (b jets) constitutes an important background for numerous searches and measurements at the LHC. Measuring these processes and comparing the results with theoretical predictions constitutes an important test of perturbative QCD calculations and helps validate and improve the modelling of these processes, which in turns helps to reduce the uncertainties in analyses affected by these backgrounds. In particular, the modelling of $\mathrm{t\bar{t}b\bar{b}}$ production is especially challenging and constitues a leading systematic uncertainty in searches for associated production of a top quark pair and a Higgs boson ($\mathrm{t\bar{t}H}$), where the Higgs boson decays to a b quark pair, as well as in searches for four top quark production ($\mathrm{t\bar{t}t\bar{t}}$).

Both $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}t\bar{t}}$ production provide a direct access to the top quark Yukawa coupling, a crucial parameter of the standard model (SM). Additionally, numerous scenarios of physics beyond the SM predict an increased $\mathrm{t\bar{t}t\bar{t}}$ cross section, compared to its value in the SM. This increase can be due to the production of hypothetical particles in association with a top quark pair, that then decay to a further top quark pair, or due to indirect effects from new physics beyond the direct reach of the LHC, which can be parameterized in an Effective Field Theory (EFT) approach.

In this document, I review the latest measurements of the properties of additional jets in $\mathrm{t\bar{t}}$ production and of the $\mathrm{t\bar{t}b\bar{b}}$ production cross section, and I report on the latest searches for $\mathrm{t\bar{t}t\bar{t}}$ production. All of these results, obtained by the CMS experiment [1], are based on pp collisions delivered by the LHC at $\sqrt{s} = 13\,\mathrm{TeV}$.

## 2. Differential cross sections for additional jets in $\mathrm{t\bar{t}}$ production

The modelling of additional jets in $\mathrm{t\bar{t}}$ production is a challenging task as it involves multi-parton matrix elements, and is sensitive to the choice and tuning of parton-shower generator and of matching and merging schemes.

Measurements of the jet multiplicity in $\mathrm{t\bar{t}}$ events, unfolded to stable particle level, have been carried out by CMS in the single-lepton and dilepton decay channels of the top quark pairs using a dataset corresponding to $35.9\,\mathrm{fb}^{-1}$ [2, 3], and are shown in Fig. 1. In the single-lepton channel, the properties of additional jets in $\mathrm{t\bar{t}}$ production have also been measured as both absolute and normalized differential cross sections. These include the jet multiplicity for varying threshold of particle-level jet $p_\mathrm{T}$, ranging from from 25 to 100 GeV, the gap fraction $f_n(p_\mathrm{T})$, defined as the fraction of events with fewer than $n$ additional jets as a function of the jet $p_\mathrm{T}$ threshold (with $n = 1$ and 2), and for the four $p_\mathrm{T}$-leading additional jets, their $p_\mathrm{T}$, their absolute pseudorapidity, their minimal angular separation $\Delta R$ with a jet coming from a top quark decay, and their minimal $\Delta R$ with either the t or the $\mathrm{\bar{t}}$ quark.

The measured distributions are compared with predictions from several Monte-Carlo event generators. None of the tested generators manages to describe all of these distributions simultaneously if only experimental and no modelling uncertainties are considered. Predictions from POWHEG interfaced with PYTHIA8 generally agree with the measurements once modelling uncertainties are taken into account.



PoS(EPS-HEP2019)665

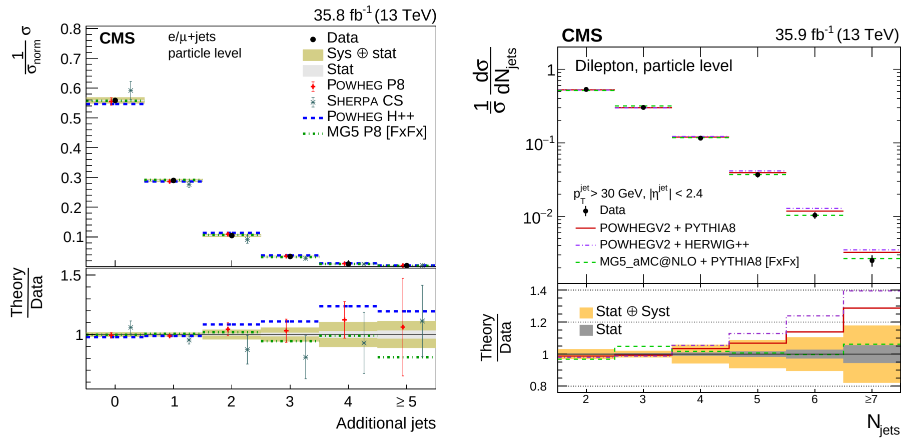


**Figure 1:** Normalized differential distributions of the multiplicity of additional jets in the single-lepton channel (left) [2] and of the jet multiplicity in the dilepton channel (right) [3] in $t\bar{t}$ events, unfolded to particle level.

## 3. Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state

The first measurement of the $t\bar{t}b\bar{b}$ cross section in the all-jet decay channel of the top quark pairs has been performed using 35.9 fb$^{-1}$ of data [4]. Events with at least eight jets are selected, of which at least two are b-tagged. The combinatorial self-background is reduced using a Boosted Decision Tree (BDT) trained to identify the jets originating from top quark decays in simulated $t\bar{t}$+jets events with exactly seven reconstructed jets. The main background in this channel is multijet production, which is reduced using a combination of multivariate techniques. Exploiting the fact that $t\bar{t}$+jets production contains on average more quark-produced jets than multijet production, a quark-gluon likelihood ratio (QGLR) is built from the quark-gluon likelihood (QGL) discriminators of non-b-tagged jets. The QGL discriminates quark-induced from gluon-induced jets using jet substructure information. In addition, a second BDT is trained to discriminate $t\bar{t}$+jets from multijet production. This BDT is trained using collision data with events containing exactly seven jets, using a classification without labels (CWoLa) [5].

Based on the QGLR and CWoLa BDT discriminators, four orthogonal event regions are defined: one signal-enriched region (SR) and three background-enriched control regions (CRs). The $t\bar{t}b\bar{b}$ signal is extracted with the two-dimensional distribution (2DCSV) of the highest and second-highest b-tagging discriminators among all jets which have not been identified by the permutation BDT as originating from top quark decays. The multijet background is estimated using the ABCD method, applied independently to each bin of the 2DCSV distribution, owing to the fact that the QGLR and CWoLa BDT are uncorrelated for multijet production.

The 2DCSV distributions of the four event regions are used to build a binned likelihood from which the $t\bar{t}b\bar{b}$ cross section is measured using a maximum-likelihood fit. Leading systematic uncertainties arise from the calibration of the QGL and of the b-tagging discriminators, and from the modelling of the $t\bar{t}b\bar{b}$ signal and $t\bar{t}c\bar{c}$ and $t\bar{t}jj$ backgrounds (renormalisation, factorisation and parton shower scales, underlying event, colour reconnection). The $t\bar{t}b\bar{b}$ cross section is measured for two different definitions of the fiducial $t\bar{t}b\bar{b}$ phase space, which either do or do not use parton-level information to track the origin of the b jets. The fiducial cross section is also corrected by the experimental acceptance to yield the total $t\bar{t}b\bar{b}$ cross section, which is found to be $5.5 \pm 0.3$ (stat) $^{+1.6}_{-1.3}$ (syst) pb. The measurements are compared with predictions from several event generators, as shown on Fig. 2, and are found to be larger than theoretical predictions by a factor of 1.5-2.4, corresponding to 1-2 standard deviations.

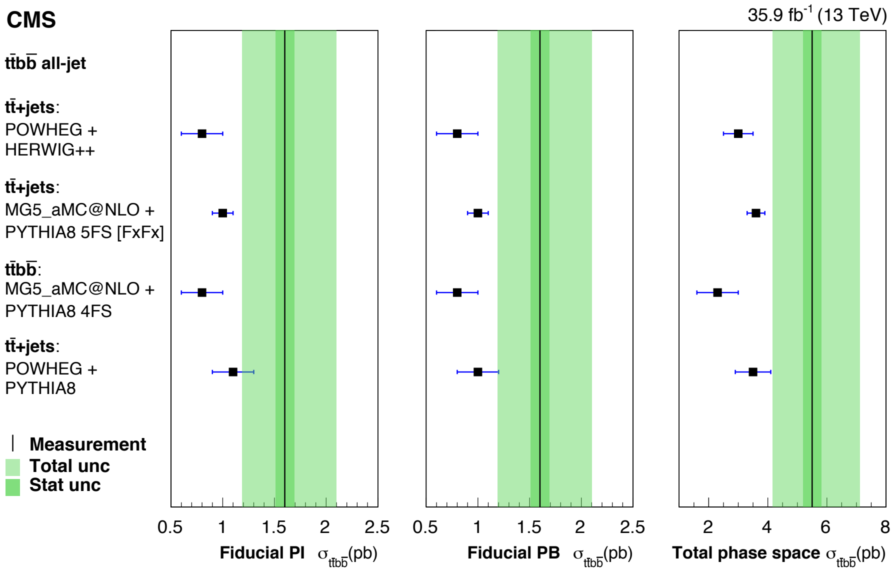


**Figure 2:** Measured (vertical lines) and predicted (squares) $t\bar{t}b\bar{b}$ production cross sections in the all-jet final state [4], for three definitions of the $t\bar{t}b\bar{b}$ phase space: fiducial parton-independent (left), fiducial parton-based (middle), total (right). The dark (light) shaded bands show the statistical (total) uncertainties in the measured value. Uncertainty bands in the theoretical cross sections show the statistical uncertainty as well as the uncertainties in the PDFs and the renormalization and factorization scales.

More recently, the $t\bar{t}b\bar{b}$ and $t\bar{t}jj$ cross sections and their ratio have also been measured in the single-lepton and dilepton final states [6]. These results provide important input to searches for $t\bar{t}H$ production with $H \to b\bar{b}$ and will help reduce the systematic uncertainties from the modelling of the irreducible $t\bar{t}b\bar{b}$ background.

## 4. Searches for $t\bar{t}t\bar{t}$ production

With a predicted cross section of 12 fb, $t\bar{t}t\bar{t}$ production is a very rare process which hasn't been observed yet. It features a variety of experimental final states, depending on the number and the charges of leptons coming from decaying top quarks. CMS has searched for $t\bar{t}t\bar{t}$ in final states with either one or two opposite-sign (OS) leptons (e or $\mu$), for which $t\bar{t}$+(b) jets constitutes the main background, and in final states with two same-sign (SS) or more than three leptons (e or $\mu$), which suffer from backgrounds such as $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$ production.

### 4.1 Single-lepton and dilepton (opposite sign) channels

The search for $t\bar{t}t\bar{t}$ in the one or two OS leptons final state has been performed using 35.9 fb$^{-1}$ of data [7]. In the single-lepton channel events are selected if the contain one isolated muon and at least eight jets, or one isolated electron and at least seven jets. In both cases, at least two of those jets must be b-tagged. In the dilepton channel at least four jets are required, of which at least two are b-tagged. Events with two muons or two electrons whose invariant mass lies close to the Z mass are vetoed. Since in these channels either two or three hadronically decaying top quarks are expected in the signal, a dedicated BDT is trained using simulated $\mathrm{t\bar{t}}$ events to identify combinations of three jets which originate from top quark decays. The largest (in the dilepton channel) and second-largest (single-lepton channel) value of this BDT among all possible jet combinations is used as input, alongside kinematic and event-level information, to further BDTs trained to discriminate the $\mathrm{t\bar{t}t\bar{t}}$ signal from the +jets background. Several BDTs are trained depending on the number of leptons, lepton flavour and jet multiplicity. Selected events are categorized as a function of jet and b jet multiplicities, and output shapes of the BDTs in these categories are used to build a binned likelihood from which the signal is extracted. Figure 3 shows the distribution of one of these BDTs in three of the event categories.

The leading systematic uncertainties affecting this channel are related to the modelling of the $\mathrm{t\bar{t}}$+jets background in high jet and b jet multiplicities. An uncertainty of 35% is considered for the rate of the $\mathrm{t\bar{t}b\bar{b}}$ background, which corresponds to the uncertainty in the $\mathrm{t\bar{t}b\bar{b}}$ cross section obtained by CMS in the dilepton channel using a reduced dataset [8].

The dilepton search channel is the most sensitive. By combining both channels, under the SM hypothesis an expected significance of 0.4 standard deviations versus the background-only hypothesis is reached. The observed significance is 0. The combined observed (expected) upper limit at 95% confidence level (CL) on the $\mathrm{t\bar{t}t\bar{t}}$ signal strength, defined as the $\mathrm{t\bar{t}t\bar{t}}$ cross section divided by the SM prediction, is 5.2 (5.7). This result is reinterpreted in an EFT framework to place bounds on four effective operators that could potentially impact $\mathrm{t\bar{t}t\bar{t}}$ production.

### 4.2 Dilepton (same sign) and multilepton channels

In the channel with two SS leptons or more than three leptons, the full dataset of $137\,\mathrm{fb}^{-1}$ collected by CMS during LHC Run II has been used [9]. Advanced lepton identification and isolation criteria are used to maximise the lepton selection efficiency and reduce the probability for jets to be misidentified as leptons, and to reject leptons produced within jets (nonprompt). The isolation criteria are based on lepton $p_\mathrm{T}$-dependent isolation cones, the ratio of the lepton $p_\mathrm{T}$ to that of the closest jet, and the lepton $p_\mathrm{T}$ relative to the direction of the nearby hadronic activity. In addition, great care is taken to reduce the probability for the leptons' charge to be misidentified.

Events are selected if they contain two SS leptons and at least two b-tagged jets. Events containing a third lepton and a pair of OS leptons of same flavour with an invariant mass close to that of the Z boson are used to populate a control region enriched in $\mathrm{t\bar{t}Z}$ events. A BDT is trained to discriminate the $\mathrm{t\bar{t}t\bar{t}}$ signal from the backgrounds, and the signal is extracted using the shape of the BDT output score, combined with the $\mathrm{t\bar{t}Z}$ control region. As a cross-check, the signal is also extracted by constructing 14 signal regions based on the number of jets, b jets, and leptons, combined with the $\mathrm{t\bar{t}Z}$ control region and a further control region enriched in $\mathrm{t\bar{t}W}$ events.

Backgrounds involving prompt leptons, such as $\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}Z}$ ($\mathrm{t\bar{t}X}$), are estimated using the simulation. The jet multiplicity in simulated $\mathrm{t\bar{t}X}$ events is corrected using an independent sample of events containing a single lepton, and the rate of b jet production in $\mathrm{t\bar{t}X}$ is corrected based on an earlier measurement of the ratio between the $\mathrm{t\bar{t}b\bar{b}}$ and $\mathrm{t\bar{t}jj}$ cross sections [8]. The background

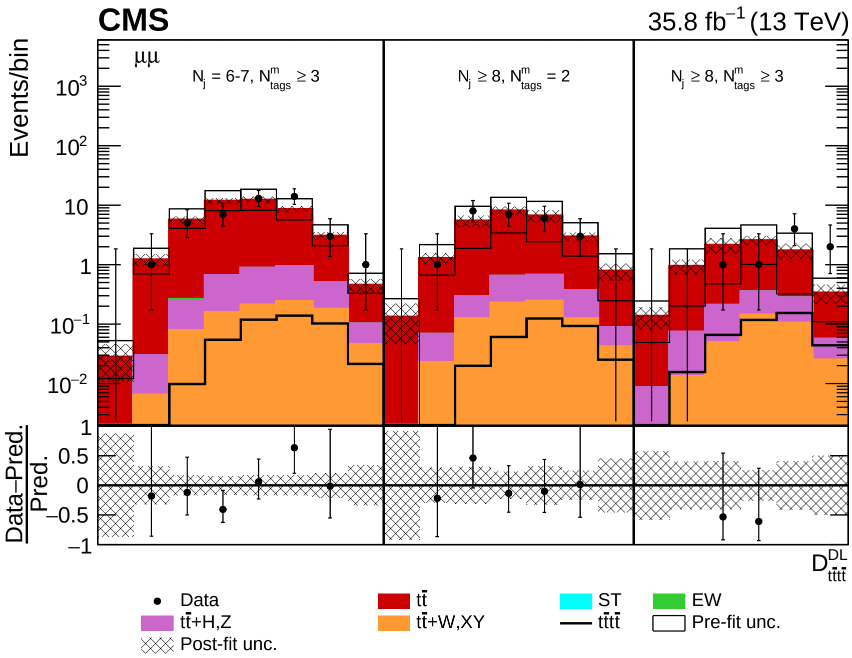


**Figure 3:** Post-fit distribution of the BDT trained to discriminate $t\bar{t}t\bar{t}$ from $t\bar{t}$+jets events [7], for events with two muons and either 6-7 jets and at least 3 b-tagged jets (left), at least 8 jets and 2 b-tagged jets (middle), and at least 8 jets and at least 3 b-tagged jets (right). The expected signal contribution, scaled to the SM cross section, is shown as a black line. Hatched bands indicate post-fit uncertainties.

from nonprompt leptons and jets misidentified as leptons is estimated from the data by inverting the lepton identification and isolation criteria, and by measuring the misidentification rate in an independent sample of events. Finally, the electron charge misidentification rate is measured using a sample of $Z \to e^+e^-$ events and corrected in the simulation.

Dominant systematic uncertainties affecting this analysis are related to the modelling of additional b jets in $t\bar{t}X$ events, the jet energy scale and resolution, the calibration of the b tagging, and the experimental uncertainty in the $t\bar{t}H$ cross section. Using the BDT discriminant, an observed (expected) significance of 2.6 (2.7) standard deviations is achieved in this channel. The measured $t\bar{t}t\bar{t}$ cross section is $12.6^{+5.8}_{-5.2}$ fb, which is in agreement with SM predictions and with the cross-check analysis that does not rely on a BDT. These results can be used to constrain the top quark Yukawa coupling $y_t$ relative to its SM value $y_t^{SM}$, yielding a limit of $|y_t/y_t^{SM}| < 1.7$ at 95% CL. The production of new scalar or pseudoscalar particles with mass $m > 2m_t$, produced in association with top quarks and decaying to to $t\bar{t}$, is also probed, and the resulting limits are interpreted in the context of simplified dark matter models and of Type-II two-Higgs doublet models.

## References

[1] CMS Collaboration, "The CMS Experiment at the CERN LHC," JINST **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[2] CMS Collaboration, "Measurement of differential cross sections for the production of top quark pairs

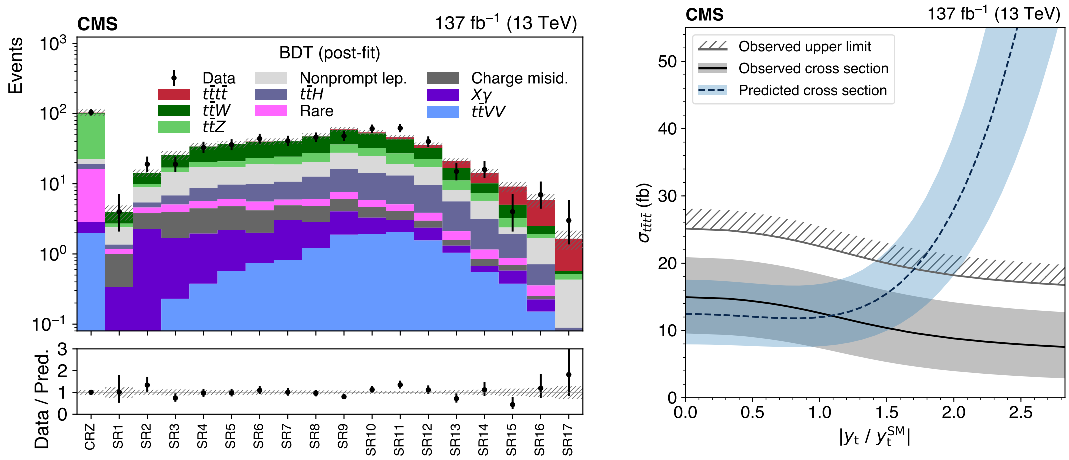


**Figure 4:** Left: post-fit distribution of the BDT used to extract the $t\bar{t}t\bar{t}$ signal in events with two same-sign leptons or at least three leptons [9]. The leftmost bin shows the post-fit and observed yields in the $t\bar{t}Z$ control region. Right: measured (grey band) and predicted (blue band) $t\bar{t}t\bar{t}$ cross section as a function of $|y_t/y_t^{SM}|$. The measured cross section is not constant as this ratio also impacts the $t\bar{t}H$ background.

and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV," Phys. Rev. D **97** (2018) 112003 doi:10.1103/PhysRevD.97.112003 [arXiv:1803.08856 [hep-ex]].

[3] CMS Collaboration, "Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons," JHEP **1902** (2019) 149 doi:10.1007/JHEP02(2019)149 [arXiv:1811.06625 [hep-ex]].

[4] CMS Collaboration, "Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV," Submitted to PLB. [arXiv:1909.05306 [hep-ex]].

[5] E. M. Metodiev, B. Nachman and J. Thaler, "Classification without labels: Learning from mixed samples in high energy physics," JHEP **1710** (2017) 174 doi:10.1007/JHEP10(2017)174 [arXiv:1708.02949 [hep-ph]].

[6] CMS Collaboration, "Measurement of the cross section for $t\bar{t}$ production with additional jets and b jets in proton-proton collisions at $\sqrt{s} = 13$ TeV," CMS-PAS-TOP-18-002.

[7] CMS Collaboration, "Search for the Production of Four Top Quarks in the Single-Lepton and Opposite-Sign Dilepton Final States in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV," Submitted to JHEP. [arXiv:1906.02805 [hep-ex]].

[8] CMS Collaboration, "Measurements of $t\bar{t}$ cross sections in association with $b$ jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV," Phys. Lett. B **776** (2018) 355 doi:10.1016/j.physletb.2017.11.043 [arXiv:1705.10141 [hep-ex]].

[9] CMS Collaboration, "Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV," Submitted to EPJC. [arXiv:1908.06463 [hep-ex]].