PROCEEDINGS OF SCIENCE

Associated production of $t\bar{t}$+HF at CMS

Fabio Iemmi[a,*] **on behalf of the CMS Collaboration**

[a] *Institute of High Energy Physics (IHEP), Chinese Academy of Sciences, 19B Yuquan Road, Shijingshan District, Beijing, China*

E-mail: fabio.iemmi@cern.ch

A comprehensive set of measurements of top quark pair production in association with light, charm and bottom jets is presented and results are compared to theory predictions. The status of the search for four top quark production, to which the LHC experiments are starting to be sensitive, and that has important beyond the standard model interpretations, is also reported.

The European Physical Society Conference on High Energy Physics (EPS-HEP2021), 26-30 July 2021
Online conference, jointly organized by Universität Hamburg and the research center DESY

*Speaker

https://pos.sissa.it/

PoS(EPS-HEP2021)483

1. Introduction

At the CERN LHC, the production of top quark-antiquark pairs ($\mathrm{t\bar{t}}$) is often accompanied by a very abundant number of additional jets, including heavy flavor jets resulting from the hadronization of bottom (b) and charm (c) quarks. A Feynman diagram depicting one of such $\mathrm{t\bar{t}}$ plus heavy flavor ($\mathrm{t\bar{t}}$+HF) processes is shown in Figure 1.

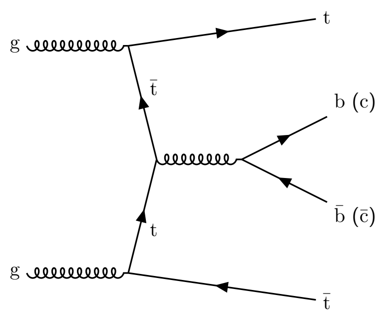

Figure 1: Leading order Feynman diagram for the associated production of a top quark-antiquark pair and a pair of heavy flavor jets.

Top quark pair production in association with a pair of b jets ($\mathrm{t\bar{t}b\bar{b}}$) is a challenging process to model. Even though calculations of the $\mathrm{t\bar{t}b\bar{b}}$ production cross section are available at next-to-leading order in quantum chromodynamics (QCD), they suffer from large uncertainties due to the choice of the renormalization and factorization scales. Such uncertainties come from the very different energy scales involved in the $\mathrm{t\bar{t}b\bar{b}}$ production, which range from the large scales associated with the top quark mass to the smaller scales associated with the additional b jets, which usually arise from gluon splitting. Improving the precision of perturbative calculations in QCD for this process is crucial, as it is a dominant background in many interesting measurements and searches at the LHC.

Up to now, the production of a $\mathrm{t\bar{t}}$ pair in association with a pair of c jets ($\mathrm{t\bar{t}c\bar{c}}$) has received less attention, both from the theoretical and experimental point of view. While the experimental signature of b jets is pretty different from the one of light jets, this is not the case for c jets. With the latest developments in charm jet identification algorithms, the $\mathrm{t\bar{t}c\bar{c}}$ signature can now be more efficiently disentangled from the $\mathrm{t\bar{t}b\bar{b}}$ and $\mathrm{t\bar{t}}$ plus light flavor ($\mathrm{t\bar{t}LL}$) topologies and measurements of the $\mathrm{t\bar{t}c\bar{c}}$ cross section are becoming possible.

Additionally, thanks to the high integrated luminosity provided by the LHC, searches of very rare standard model (SM) processes, such as the associated production of four top quarks ($\mathrm{t\bar{t}t\bar{t}}$) are now becoming accessible. Interestingly, $\mathrm{t\bar{t}t\bar{t}}$ production can be significantly enhanced by beyond-the-SM (BSM) models, so that measurements of the production cross section can lead to constraints on such models.

The $\mathrm{t\bar{t}}$+HF final states are characterized by high jet multiplicities and are thus challenging to analyze. Besides the previously mentioned charm jet identification, the jet-parton assignment is found to be a non trivial task too. In fact, the high number of heavy flavor jets makes it hard to correctly assign a jet either to a top quark decay or to a gluon splitting. The methods chosen by the different analyses to tackle these challenges will be highlighted in the following.

2. $\mathrm{t\bar{t}jj}$ and $\mathrm{t\bar{t}b\bar{b}}$ cross section measurements

The CMS Collaboration [1] performed a measurement [2] of the production cross section (σ) of $\mathrm{t\bar{t}b\bar{b}}$ events, of the production cross section of $\mathrm{t\bar{t}}$ pairs in association with two jets of any flavor ($\mathrm{t\bar{t}jj}$), and of their ratio $R_{\mathrm{t\bar{t}b\bar{b}/t\bar{t}jj}}$ using 35.9 $\mathrm{fb^{-1}}$ of data. The $\sigma_{\mathrm{t\bar{t}jj}}$ cross section and the ratio $R_{\mathrm{t\bar{t}b\bar{b}/t\bar{t}jj}}$

are measured separately in the dilepton (DL) and single-lepton (SL) channels. The measured $\sigma_{\mathrm{t\bar{t}b\bar{b}}}$ cross section is then obtained by multiplying $\sigma_{\mathrm{t\bar{t}jj}}$ by $\mathrm{R}_{\mathrm{t\bar{t}b\bar{b}/t\bar{t}jj}}$.

Events with a $\mathrm{t\bar{t}}$ pair and at least two additional jets in the simulation are split based on the flavor of the particle-level jets found from the Monte Carlo generator information, resulting in four categories: $\mathrm{t\bar{t}b\bar{b}}$, $\mathrm{t\bar{t}bj}$, $\mathrm{t\bar{t}c\bar{c}}$ and $\mathrm{t\bar{t}LF}$, where LF indicates light flavor jets. The category $\mathrm{t\bar{t}jj}$ comprises all the categories described above. Particle-level jets in the DL and SL channels are selected with different p_{T} cuts, in such a way that different visible phase spaces (VPS) are defined for the two independent measurements, providing additional information for testing the theoretical predictions. Events entering the analysis are required to pass several selection criteria designed to improve the purity in $\mathrm{t\bar{t}}$ plus jets events. Among these criteria, the request of at least two b tagged jets is made for both the DL and SL channels.

In order to perform jet-parton assignment, two different approaches depending on the final state are used. In the DL channel, where the jet multiplicity is lower, the two b tagged jets with the highest b tagging score are found to come from top quark decays in 85 (23)% of the selected $\mathrm{t\bar{t}jj}$ ($\mathrm{t\bar{t}b\bar{b}}$) events, based on the simulation. Thus, the jets with third- and fourth-largest b tagging score are considered to be additional jets.

In the SL channel, where the jet multiplicity is higher, the jet-parton assignment becomes more challenging; thus, a kinematic fit is employed. Starting from the four-momenta of the selected leptons and jets and from the missing transverse momentum, the kinematic fit is constrained to the $\mathrm{t\bar{t}} \rightarrow \mathrm{W^{+}bW^{-}\bar{b}}$. The algorithm computes a χ^2 value for each possible jet assignment; the solution showing the lowest χ^2 value is selected and, amongst all the jets that are not included in such solution, only the two with the highest b tagging discriminant value are considered to be additional jets and kept for further analysis.

The signal extraction is performed by means of a maximum likelihood (ML) fit to the two-dimensional distributions of the b tagging discriminants of the additional jets. While the DL and SL channels are fitted separately due to the different VPS definitions, within each channel the different final states are fitted simultaneously.

Several sources of systematic uncertainty are considered, the dominant ones being found to be the uncertainty in the amount of final-state radiation, the matrix element to parton shower matching uncertainty and the uncertainty in the b tagging efficiency.

Results obtained in the VPS are extrapolated to the full phase space using acceptance corrections and compared to the theoretical predictions, as shown in Figure 2. Predictions from POWHEG + HERWIG++ for $\mathrm{R}_{\mathrm{t\bar{t}b\bar{b}/t\bar{t}jj}}$, and thus for $\sigma_{\mathrm{t\bar{t}b\bar{b}}}$, are found to be slightly lower than the measured values. A measured $\mathrm{t\bar{t}b\bar{b}}$ cross section higher than the theoretical predictions has also been reported in a previous CMS measurement in the all-jet final state [3].

3. $\mathrm{t\bar{t}c\bar{c}}$ cross section measurement

The CMS Collaboration also performed the first measurement [4] of the inclusive cross section for the $\mathrm{t\bar{t}c\bar{c}}$ production. The analysis is performed in the dilepton final state of the $\mathrm{t\bar{t}}$ pair and uses $41.5\,\mathrm{fb}^{-1}$ of data. Besides $\sigma_{\mathrm{t\bar{t}c\bar{c}}}$, the analysis also aims at measuring $\sigma_{\mathrm{t\bar{t}b\bar{b}}}$, $\sigma_{\mathrm{t\bar{t}LL}}$ and their ratios to $\sigma_{\mathrm{t\bar{t}jj}}$.

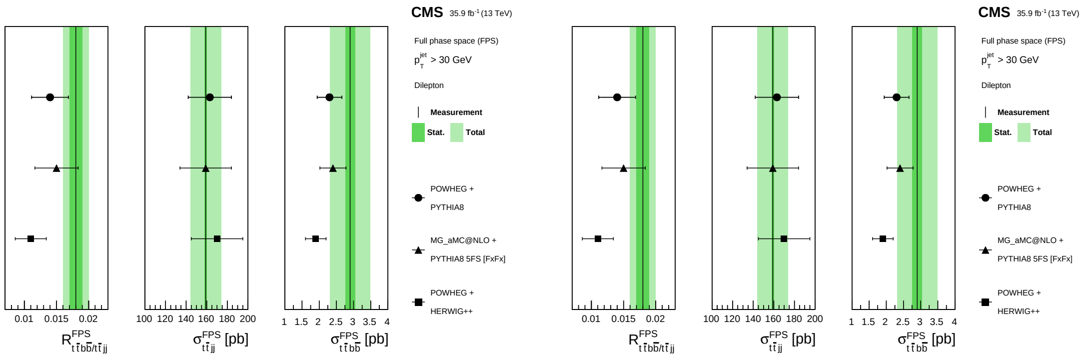

Figure 2: Measured values (vertical lines) for the $t\bar{t}b\bar{b}$ and $t\bar{t}jj$ cross sections and their ratio, along with their statistical and total uncertainties (dark and light bands) in the dilepton (left) and single-lepton (right) channels in the full phase space. The measured values are also compared with the predictions of several Monte Carlo generators.

In a similar fashion to what was done in the previous analysis, events with a $t\bar{t}$ pair and at least two additional jets in the simulation are split based on the flavor of the particle-level jets, resulting in five categories: $t\bar{t}b\bar{b}$, $t\bar{t}$bL, $t\bar{t}c\bar{c}$, $t\bar{t}$cL and $t\bar{t}$LL.

The charm jet identification is achieved through the use of DeepCSV [5], a novel HF tagging algorithm. The algorithm is based on a deep neural network (DNN) with multiclass output structure, which is able to predict the probabilities for each jet to contain a single b hadron ($P(\mathrm{b})$), two b hadrons ($P(\mathrm{bb})$), one or more c hadrons ($P(\mathrm{c})$) and no b or c hadrons ($P(\mathrm{udsg})$). Such probabilities are combined together to obtain two charm discriminators CvsL and CvsB that separate c jets from LF jets and c jets from b jets respectively:

$$\mathrm{CvsL} = \frac{P(\mathrm{c})}{P(\mathrm{c}) + P(\mathrm{udsg})}, \qquad \mathrm{CvsB} = \frac{P(\mathrm{c})}{P(\mathrm{c}) + P(\mathrm{b}) + P(\mathrm{bb})}.$$

The jet-parton assignment is performed by training a DNN, which considers all the possible permutations of jets in the event and identifies the correct assignment. To perform the choice, the algorithm uses several input variables, ranging from jet kinematic variables to b and c tagging discriminators and angular separations between jets and leptons. The network is set up to output the probability for a permutation of jets to belong to the following three categories: first, a permutation corresponding to the correct jet-parton assignment, P^{+}; second, a permutation corresponding to the correct assignment of additional jets, but for which the b jets from top quark decays are assigned in the reverse order, $P^{\times}$ (since the order of b jets assignment does not matter, this permutation is still considered appropriate); third, a permutation corresponding to a wrong assignment, P^{-}. The best jet-parton assignment is then identified by selecting the permutation of jets showing the highest value of

$$\max\left(\frac{P^{+}}{P^{+} + P^{-}}, \frac{P^{\times}}{P^{\times} + P^{-}}\right).$$

In order to perform the signal extraction, a second DNN is trained. Starting from the charm tagging discriminators CvsL and CvsB of the two additional jets, the angular separation between the additional jets and the DNN score for the best permutation of jets, the network predicts the

	Result	POWHEG	MADGRAPH5_AMC@NLO
$\sigma_{\mathrm{t\bar{t}c\bar{c}}}$ [pb]	8.0 ± 1.1 ± 1.3	9.1 ± 1.8	8.9 ± 1.5
R_{c} [%]	2.69 ± 0.36 ± 0.32	2.81 ± 0.20	2.72 ± 0.19

Table 1: Measured values for the cross section $\sigma_{\mathrm{t\bar{t}c\bar{c}}}$ and the cross section ratio $R_{\mathrm{c}} = \sigma_{\mathrm{t\bar{t}c\bar{c}}}/\sigma_{\mathrm{t\bar{t}jj}}$ with their statistical and systematic uncertainties listed in that order. The results are also compared to the predictions of the POWHEG and MADGRAPH5_AMC@NLO generators.

probabilities for an event to belong to each of the five signal categories: $P(\mathrm{t\bar{t}b\bar{b}})$, $P(\mathrm{t\bar{t}bL})$, $P(\mathrm{t\bar{t}c\bar{c}})$, $P(\mathrm{t\bar{t}cL})$ and $P(\mathrm{t\bar{t}LL})$. The signal is then extracted by means of a ML fit to a two-dimensional combination of such probabilities:

$$\Delta_{\mathrm{b}}^{\mathrm{c}} = \frac{P(\mathrm{t\bar{t}c\bar{c}})}{P(\mathrm{t\bar{t}c\bar{c}}) + P(\mathrm{t\bar{t}b\bar{b}})}, \qquad \Delta_{\mathrm{L}}^{\mathrm{c}} = \frac{P(\mathrm{t\bar{t}c\bar{c}})}{P(\mathrm{t\bar{t}c\bar{c}}) + P(\mathrm{t\bar{t}LL})}.$$

These discriminators can be interpreted to be topology-specific c tagger discriminators that enhance the information of the flavor of additional jets with kinematic information.

Several sources of systematic uncertainty are considered, the dominant ones being found to be the uncertainty in the charm tagging scale factors, the uncertainty in the energy scale of jets and the matrix element to parton shower matching uncertainty.

Results obtained in the VPS are extrapolated to the full phase space using acceptance corrections and compared to the theoretical predictions, as shown in Table 1. The inclusive $\mathrm{t\bar{t}c\bar{c}}$ cross secction and the ratio $R_{\mathrm{c}} = \sigma_{\mathrm{t\bar{t}c\bar{c}}}/\sigma_{\mathrm{t\bar{t}jj}}$ are measured here for the first time. They are found to be in agreement with the theoretical predictions, within the uncertainties.

In addition, this analysis is also able to extract the cross sections for the $\mathrm{t\bar{t}b\bar{b}}$ and $\mathrm{t\bar{t}LL}$ processes, together with the corresponding cross section ratios. Two-dimensional likelihood scans are performed in the VPS over different combinations of cross sections and ratios. Agreement is observed at the level of one to two standard deviations between the measured and predicted values. The most significant tension is observed in the $\mathrm{t\bar{t}b\bar{b}}$ cross section and in the corresponding cross section ratio, as it is shown in Figure 3. These results are consistent with the findings described in Section 2.

4. Search for $\mathrm{t\bar{t}t\bar{t}}$ events

The CMS Collaboration also performed a search [6] for the associated production of four top quarks. This is a very rare SM process (with a theoretical cross section $\sigma_{\mathrm{t\bar{t}t\bar{t}}}$ of about 12 fb at 13 TeV), which is still unobserved at the LHC; nevertheless, new particles coupling to the top quark in BSM scenarios, such as heavy scalars and pseudoscalars bosons predicted in Type-II-two-Higgs-doublet models [7] can contribute to $\sigma_{\mathrm{t\bar{t}t\bar{t}}}$ when their masses are larger that twice the top quark mass.

The analysis uses the full amount of data collected by the CMS experiment in the LHC RunII data taking period, corresponding to 137 fb^{-1}, and targets the two-same-sign-leptons (2SSL) and multileptons (ML) final states of the $\mathrm{t\bar{t}t\bar{t}}$ system, the ML label meaning at least three leptons.

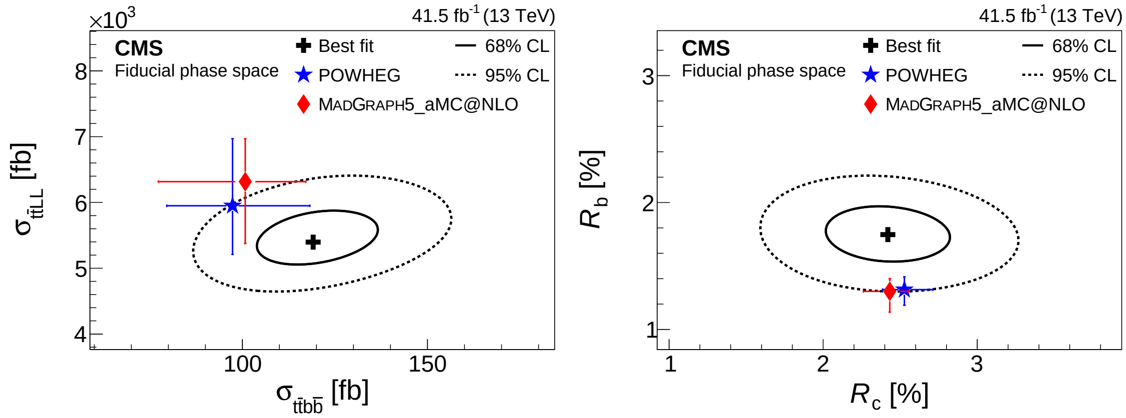

Figure 3: Results of the two-dimensional likelihood scans in the VPS for combinations of cross sections and cross section ratios. The best fit value (black cross) with corresponding 68% (full) and 95% (dashed) confidence level contours are shown, compared with the theoretical predictions of POWHEG and MADGRAPH5_aMC@NLO generators.

Events passing a baseline selection are then split into 17 signal regions (SRs) by discretizing the output of a boosted decision tree (BDT) trained to separate $t\bar{t}t\bar{t}$ events from SM backgrounds. The BDT classifier utilizes a gradient boosting algorithm and 19 input variables, the three most performing ones being the jet multiplicity, the b jet multiplicity and the number of leptons. The SRs are also complemented with a $t\bar{t}$Z-enriched category (CRZ), which is used to constrain the uncertainties on this background.

Several sources of systematic uncertainty are considered, the dominant ones being found to be the uncertainties related to the jet energy scale and resolution and the uncertainty in the b tagging scale factors.

The measured cross section and the corresponding significance of the observation are obtained by means of a profile maximum-likelihood fit to the data yields in the SRs and in the CRZ. The post-fit distribution of such yields is shown in Figure 4.

The observed $t\bar{t}t\bar{t}$ cross section is found to be $\sigma^{\mathrm{obs}}_{t\bar{t}t\bar{t}} = 12.6^{+5.8}_{-5.2}$ fb, a value that is consistent with the expectations from the standard model and corresponding to an observed significance of 2.6 standard deviations. To give an example of the many possible BSM interpretation of $t\bar{t}t\bar{t}$ searches, Figure 4 also shows the expected and observed upper limits on the cross section times branching ratio to $t\bar{t}$ for the production of a new heavy scalar H, as a function of the scalar mass. The measurement is able to exclude the mass range $350 < m_{\mathrm{H}} < 470$ GeV, which corresponds to a more than 100 GeV improvement with respect to previous CMS measurements.

5. Conclusions

Measurements of $t\bar{t}$+HF processes are important from both the theoretical and experimental point of view. First, the comparison with experiments can lead to improvement of the theoretical predictions; second, such processes are a background for many searches at the LHC, and third, they can provide interesting BSM interpretations. The CMS Collaboration has tackled $t\bar{t}$+HF

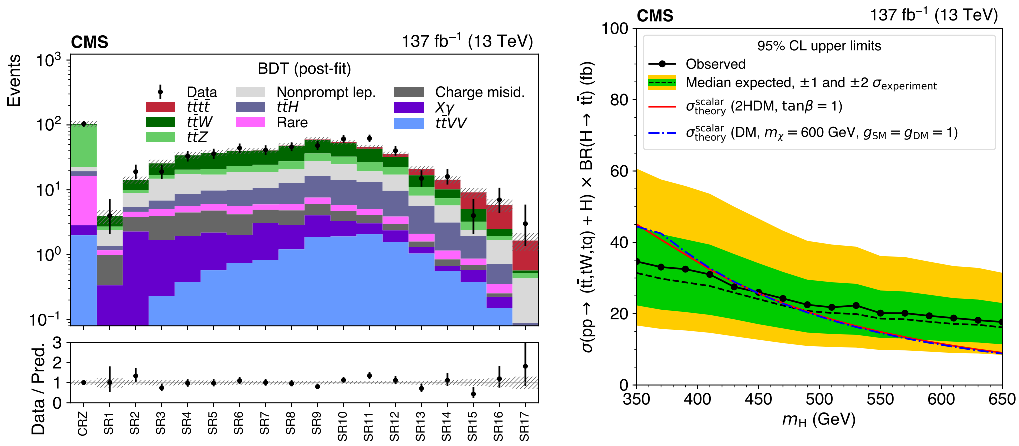

Figure 4: Observed yields in the SRs and CRZ, compared to the post-fit predictions for signal and background processes (left); observed (points) and expected (dashed line) 95% confidence level upper limits on the cross section times branching fraction to $t\bar{t}$ for the production of a new heavy scalar H (right).

measurements in many ways: it has provided the first measurement of $\sigma_{t\bar{t}c\bar{c}}$, precise measurements of $\sigma_{t\bar{t}b\bar{b}}$ and it has performed $t\bar{t}t\bar{t}$ searches which can be interpreted in terms of many BSM theories.

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 the cross section for $t\bar{t}$ production with additional jets and b jets in pp collisions at $\sqrt{s}$ = 13 TeV*, J. High Energ. Phys. **125** (2020), doi:10.1007/JHEP07(2020)125, arXiv:2003.06467

[3] 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*, Phys. Lett. B **803** (2020) 135285, doi:10.1016/j.physletb.2020.135285, arXiv:1909.05306

[4] 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 $\sqrt{s}$ = 13 TeV*, Phys. Lett. B **820** (2021) 136565, doi:10.1016/j.physletb.2021.136565, arXiv:2012.09225

[5] CMS Collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, JINST **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158

[6] 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*, Phys. Lett. B **333** (1994) 126, doi:10.1140/epjc/s10052-019-7593-7, arXiv:1908.06463

[7] D. Dicus, A. Stange, and S. Willenbroc, *Higgs decay to top quarks at hadron colliders*, Phys. Lett. B **333** (1994) 126, doi:10.1016/0370-2693(94)91017-0, arXiv:hep-ph/9404359