IL NUOVO CIMENTO **43 C** (2020) 35
DOI 10.1393/ncc/i2020-20035-2

Colloquia: IFAE 2019

Search for four-top-quark production in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

J. Magro([1])([2])

([1]) *Università degli Studi di Udine - Udine, Italy*
([2]) *INFN, Sezione di Trieste, Gruppo collegato di Udine - Udine, Italy*

received 8 June 2020

Summary. — Recent published results on a search for four-top-quark production, $t\bar{t}t\bar{t}$, are presented, based on a dataset collected at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector at the Large Hadron Collider, corresponding to an integrated luminosity of $36.1\,\text{fb}^{-1}$. The single-lepton and opposite-sign dilepton final states are analysed. The search exploits the high multiplicity of jets/b-tagged jets and the large total sum of jet transverse momenta, which characterise signal events and provide a good discrimination against the $t\bar{t}$+jets dominant background, which is estimated through a data-driven method. No significant excess above the Standard Model expectation is observed.

1. – Introduction

The top quark is the heaviest known elementary particle of the Standard Model (SM) with a mass of $m_t = 173.0 \pm 0.4\,\text{GeV}$ [1]. It has a large coupling with the SM Higgs boson [2] and it is predicted to have large couplings to new particles, playing a special role in many hypothetical models beyond the Standard Model (BSM). Because of this large mass, the production of $t\bar{t}t\bar{t}$ is one of the most rare mechanisms accessible at the Large Hadron Collider (LHC) [3]. In fact, the SM four-top-quark production cross-section ($\sigma_{SM}^{t\bar{t}t\bar{t}}$) at the next-to-leading order (NLO) in QCD at the center-of-mass energy $\sqrt{s} = 13\,\text{TeV}$ is predicted to be $\sigma_{SM}^{t\bar{t}t\bar{t}} = 9.2\,\text{fb}$ and it might be possibly enhanced in many BSM scenarios over the SM prediction. Previous searches for four-top-quark production were performed by both the ATLAS [4, 5] and CMS [6-8] Collaborations. Upper limits on the cross-section have been set, but the process has not been observed yet. The last result obtained by the ATLAS Collaboration [9] is described in the following sections. It exploits the data collected during 2015 and 2016, corresponding to and integrated luminosity of $36.1\,\text{fb}^{-1}$.

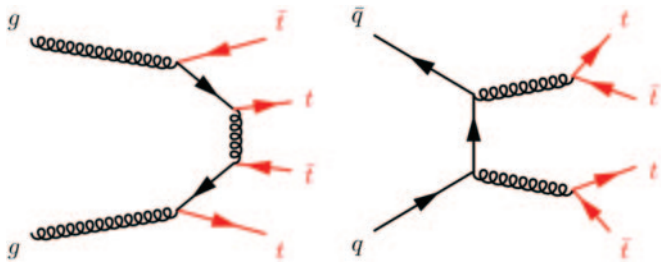

Fig. 1. – Examples of Feynman diagrams for the QCD production of $t\bar{t}t\bar{t}$ at LO.

2. – Event selection

At LHC the main $t\bar{t}t\bar{t}$ production process is through gluon-gluon fusion, with a small contribution from quark-antiquark annihilation. Two representative LO Feynman diagrams for both processes are shown in fig. 1. Since the top quark decays mostly into a W boson and a bottom quark, and the W boson can decay into a lepton and a neutrino or in two quarks, there are many possible final states: single-lepton (electron or muon) channel, dilepton with two opposite-sign charged leptons channel, dilepton with two same-sign charged leptons channel, multi-lepton channel and fully hadronic channel. In this paper only the single-lepton (1L) and dilepton opposite-sign (2LOS) final states are considered.

The four-top-quark signal is characterised by a high multiplicity of jets and b-tagged jets. Moreover, the event topology presents a high scalar sum of jet transverse momenta (H_T^{had}), which is used as the main discriminant against the dominant background: top quark production in association with jets ($t\bar{t}$+jets).

At the trigger level, events are selected using low-p_T thresholds associated to loose lepton-isolation requirements, or using tightened thresholds but looser identification and isolation criteria. Electron candidates are then reconstructed from an isolated deposit of energy in the electromagnetic calorimeter of the ATLAS detector [10] matched to a track inside the inner detector. The fiducial region is defined as $|\eta| < 1.37$ and $1.52 < |\eta| < 2.47$, with η the pseudorapidity of the energy deposit, in order to avoid electrons from the transition region between barrel and endcap electromagnetic calorimeters. Muon candidates are reconstructed by combining tracks in the inner detector and in the muon spectrometer. In this case the fiducial region is $|\eta| < 2.5$. Leptons are also required to be isolated, in order to reduce the background from non-prompt leptons, photon conversions and hadron decays. The isolation (I_R) is defined as the scalar sum of all tracks excluding the lepton one within a cone along the direction of the lepton. All leptons are required to satisfy $I_R/p_T^l < 0.06$. Lepton tracks have also to match the primary vertex of the event.

Jet candidates are reconstructed from topological calorimeter cluster, using an anti-k_t jet algorithm [11] with a radius parameter of 0.4. These are called "small-R jets". Calibrated jets are then required to have $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$. Quality criteria are also imposed to discard jets originating from non-collision sources or from pile-up. The identification of jets containing a b-hadron is performed via a multivariate algorithm [12]. The threshold value used to tag the jet as a b-quark originated jet has an average efficiency of 77%. An overlap removal procedure is applied to avoid single final-state objects being double counted: electrons close to a muon are discarded, as well as jets close to

leptons with less than three tracks originating from the primary vertex. The jets from $t\bar{t}t\bar{t}$ events are collimated, unlike $t\bar{t}$+jets ones, and therefore to exploit this particular feature, the small-R jets which pass the overlap removal procedure are reclustered using an anti-k_t algorithm with a radius parameter of 1.0. The so-called "RCLR jets" are then formed. RCLR jets which have $p_T > 200\,\text{GeV}$, $|\eta| < 2.0$ and a mass larger than 100 GeV (computed as the sum of the four-momenta of the associated small-R jets) are the "mass-tagged RCLR jets".

Events with exactly one lepton and at least five jets, of which at least two are b-tagged, enter in the 1L channel selection. Events with exactly two opposite-sign charged leptons and at least four jets, of which at least two are b-tagged, enter in the 2LOS channel.

3. – Event modelling and search strategy

As stated in the previous section, the dominant source of background is the $t\bar{t}$+jets process. Other important backgrounds are the single-top-quark production and the W/Z boson production associated with jets. Smaller contributions come from the associated production of boson (W, Z or H) with $t\bar{t}$, from diboson production and from fake and non-prompt leptons.

Monte Carlo (MC) simulated samples are employed to model the expected distributions of the signal and the majority of the backgrounds: the fake and non-prompt lepton background is estimated via a data-driven method, while for the main $t\bar{t}$+jets background a MC simulation is used assisted by a dedicated data-driven method described in the next section. $t\bar{t}t\bar{t}$ production is simulated at LO using MadGraph5_aMC@NLO 2.2.2 generator with the NNPDF2.3 LO PDF set, interfaced to Pythia 8.186. For the dominant $t\bar{t}$+jets background Powheg-Box v2 interfaced to Pythia 6.428 is used. Single-top-quark events are simulated with Powheg-Box v1 associated to Pythia 6.428, while the W/Z boson production is simulated using Sherpa 2.2. For the smaller backgrounds, Sherpa 2.1.1 is used for the diboson production and MadGraph5_aMC@NLO 2.3.2 interfaced to Pythia 8.210 generates the associated production of a boson with $t\bar{t}$.

In the single-lepton channel, the signal is characterised by the presence of one lepton, missing energy from the neutrino and a high multiplicity of high-p_T jets. At LO, the signal events are characterised by the presence of ten jets: six light-quark jets and four b-quark jets. In the opposite-sign dilepton channel, the signal is characterised by two leptons, missing energy and (at LO) eight jets: four b-quark jets and four light-quark jets. However, with such a large number of partons originating from the hard collision, the probability of some of the jets getting out of acceptance or overlapping with other jets becomes large. In addition, the b-tagging probability is significantly lower than 100%, and only a fraction of the hadronic top quark decays creates collimated enough jets to form a RCLR jet. Finally, events with a low multiplicity of jets and b-jets are useful to extract information about the dominant $t\bar{t}$+jets background. Therefore, events which pass the preselection cuts are classified taking into account the numbers of jets, b-tagged jets and mass-tagged RCLR jets. In this way 20 *signal* regions, 16 *validation* regions, 18 *source* region and 2 *efficiency extraction* regions are defined, as shown in fig. 2. Regions with at least nine (seven) jets in the 1L channel (in the 2LOS channel) and at least three b-jets have the largest signal-to-background ratio (S/B) and are used as signal regions. Validation regions have lower S/B and are designed to validate the data-driven estimate of the $t\bar{t}$+jets background. They require six or seven (six) jets and at least three b-jets. Source regions require at least seven (six) jets, but exactly two b-jets, and they are used to build pseudo-data events for the data-driven method. Finally, efficiency regions contain

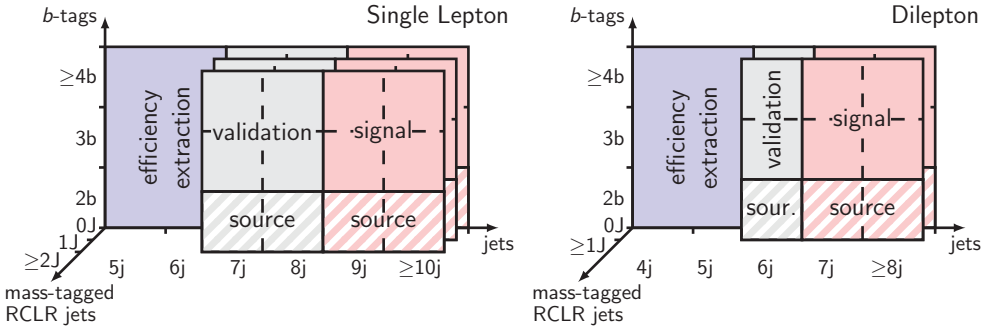

Fig. 2. – Schematic view of the analysis regions for the single-lepton channel on the left and the dilepton channel on the right [9].

events with low jet multiplicity: five or six (four or five) and at least two b-jets. They are used to extract the b-tagging probabilities.

4. – The $\mathrm{TRF}_{t\bar{t}}$ method

A purely MC-simulation–based approach at NLO accuracy in QCD for the prediction of the inclusive $t\bar{t}$+jets background is not expected to accurately model the high jet multiplicity regions used in this search. Consequently, a data-driven method is employed to estimate the dominant $t\bar{t}$+jets background: the "tag rate function for $t\bar{t}$+jets events" ($\mathrm{TRF}_{t\bar{t}}$) [13]. This method assumes that the probability of b-tagging an additional jet in $t\bar{t}$+jets events is independent of the number of additional jets. So, the tagging probability can be estimated in lower jet multiplicity regions and applied to data events in source regions, where the signal is negligible. The per-jet b-tagging probabilities are measured in the efficiency extraction regions after the subtraction of all non-$t\bar{t}$ events, evaluated with the MC simulations. These b-tagging probabilities are used to build pseudo-data samples in validation and signal regions, applying the derived information to the source regions, after subtracting the non-$t\bar{t}$+jets events. In this way, jets that were not b-tagged in the original data sample can be promoted to b-tagged in the pseudo-data sample.

The MC simulation is used to estimate the systematic uncertainties and evaluate the correction for each bin of H_T^{had} distribution. All the steps described above are thus applied to the MC-simulated $t\bar{t}$+jets sample, treating it as data, and comparing the resulting estimation with that coming from real data. This comparison gives a correction factor to apply to the initial data-driven estimation for each of the considered bins. The corrections result to be less than 20% on average. This procedure is then repeated for each of the systematic uncertainties affecting the $t\bar{t}$+jets MC simulation, giving rise to a set of systematic variations for the correction factors, and therefore a full set of MC-simulation–based systematic uncertainties on the data-driven estimation. Besides these systematic uncertainties, two additional sources of statistical uncertainty further affecting the estimate, coming from the limited number of data events in the source regions and from the limited number of simulated events.

5. – Systematic uncertainties

Two different sources of systematic uncertainties affect the measurement: the experimental uncertainties and the signal/background modelling ones.

The former includes uncertainties on the integrated luminosity, on the average number of simultaneous interactions (pile-up), on the lepton and jet energy scale and resolution, on the efficiency of the triggers, of the lepton identification and of the b-tagging algorithm. The latter includes those on the background process cross-sections and, in the case of the $t\bar{t}$+jets process, on the final object kinematical distributions, obtained through detailed comparisons using alternative MC simulation samples produced with different generators and parton-shower models or different values for various parameters. No uncertainty is assigned to the theoretical cross-section for $t\bar{t}t\bar{t}$ production.

6. – Results

The $H_T^{had}(SM)$ distributions are analysed jointly across all the signal regions, in order to test the prediction. The statistical interpretation uses a binned likelihood function, $\mathcal{L}(\mu, \vec{\theta})$, constructed as a product of Poisson probability terms over all considered bins. The likelihood function depends on several paramenters: the signal strength μ,

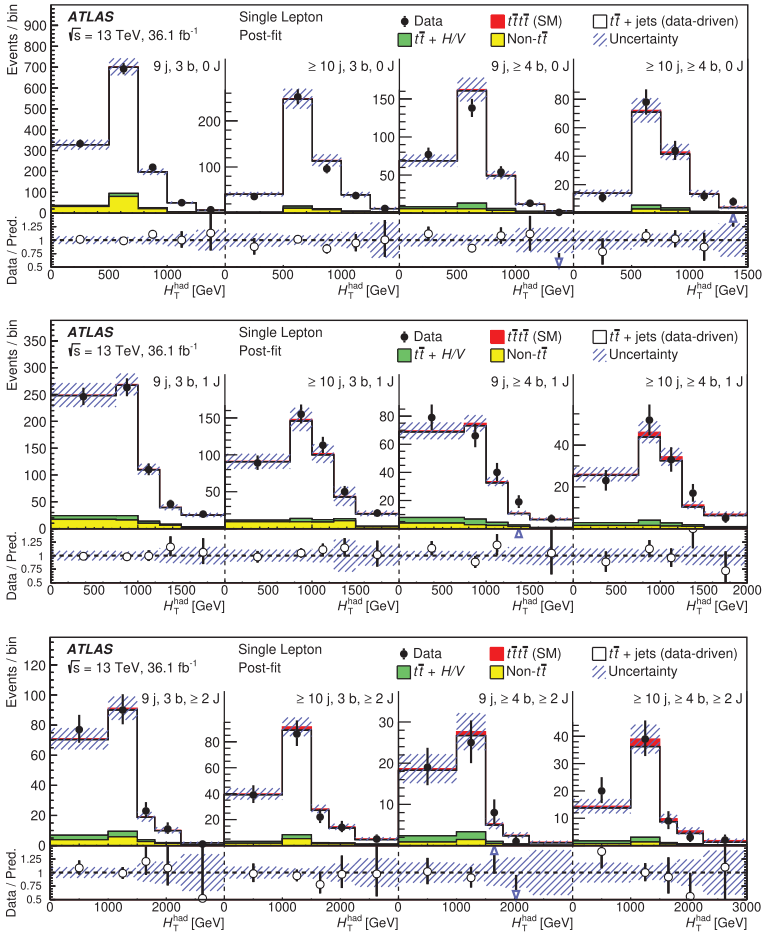

Fig. 3. – Comparison between data and prediction for the H_T^{had} distributions in the single-lepton channel after the combined fit to data [9].

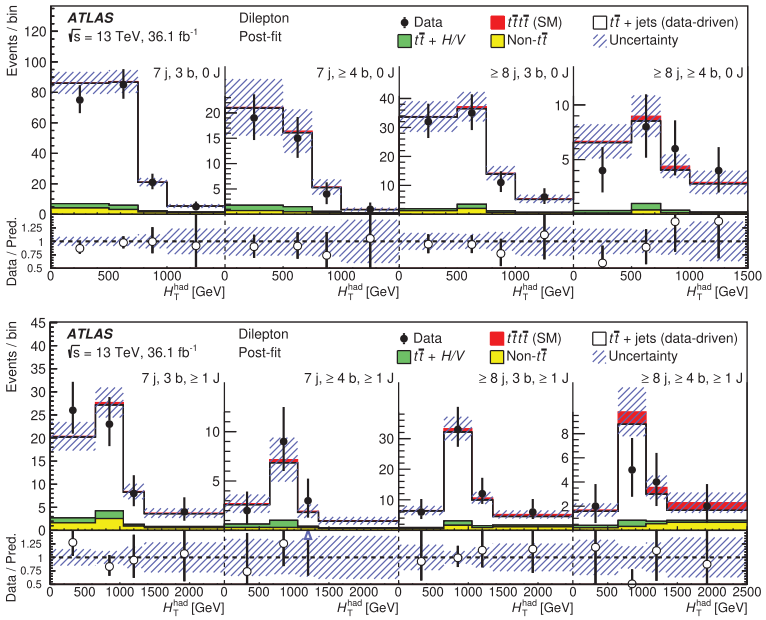

Fig. 4. – Comparison between data and prediction for the H_T^{had} distributions in the dilepton channel after the combined fit to data [9].

which scales the number of expected signal events, and the nuisance parameters $\vec{\theta}$, which take into account the systematic uncertainties on the expectations. A binned likelihood fit to the data is performed in the signal regions (12 for the 1L channel and 8 for the 2LOS channel), leading to a good agreement between data and predictions. Figures 3 and 4 show the comparison between data and the SM prediction for the H_T^{had} distributions in the signal regions. No significant excess of events with respect to the SM prediction is observed. An observed (expected) 95% confidence level upper limit on the production cross-section of four-top-quark is found to be 47 (33) fb, corresponding to an upper limit on $\sigma^{t\bar{t}t\bar{t}}/\sigma^{t\bar{t}t\bar{t}}_{SM}$ of 5.1 (3.6). The signal strength μ is measured to be $1.7^{+1.9}_{-1.7}$.

REFERENCES

[1] Particle Data Group (Tanabashi M. *et al.*), *Phys. Rev. D*, **98** (2018) 010001.
[2] ATLAS Collaboration, *Phys. Lett. B*, **784** (2018) 173, arXiv:1806.00425 [hep-ex].
[3] Breuning O., Collier P., Lebrun P., Myers S., Ostojic R., Poole J. and Proudlock P., *LHC Design Report*, Vol. **1**: *The LHC main ring* (CERN, Geneva) 2004.
[4] ATLAS Collaboration, *JHEP*, **09** (2017) 088, arXiv:1704.08493 [hep-ex].
[5] ATLAS Collaboration, *JHEP*, **07** (2018) 089, arXiv:1803.09678 [hep-ex].
[6] CMS Collaboration, *Eur. Phys. J. C*, **76** (2016) 439, arXiv:1605.03171 [hep-ex].
[7] CMS Collaboration, *Phys. Lett. B*, **772** (2017) 336, arXiv:1702.06164 [hep-ex].
[8] CMS Collaboration, *Eur. Phys. J. C*, **78** (2018) 140, arXiv:1710.10614 [hep-ex].
[9] ATLAS Collaboration, *Phys. Rev. D*, **99** (2019) 052009, arXiv:1811.02305 [hep-ex].
[10] ATLAS Collaboration, *JINST*, **3** (2008) S08003.
[11] Cacciari M., Salam G. and Soyez G., *JHEP*, **04** (2008) 063, arXiv:0802.1189 [hep-ph].
[12] ATLAS Collaboration, *JINST*, **11** (2016) P04008, arXiv:1512.01094 [hep-ex].
[13] ATLAS Collaboration, *JHEP*, **05** (2016) 160, arXiv:1604.03812 [hep-ex].