

Constraining QCD multijet background in the t -channel single-top quark production at $\sqrt{s} = 13 \text{ TeV}$

Georgios Konstantinos Krintiras* on behalf of the CMS Collaboration

Université catholique de Louvain, Louvain-la-Neuve, Belgium

E-mail: gkrintir@cern.ch

Precision measurement of the cross section for single top production is an important test of the Standard Model (SM). The purity of the collected data in single top events is limited by the understanding of the shape and yield of background contributions. Besides electroweak and $t\bar{t}$ processes, QCD multijet events constitute a non-negligible background for the considered signal $bq' \rightarrow tq$ (t -channel) process. The data-driven technique for constraining QCD contribution, employed in the measurement of the t -channel single top-quark cross section using the very first LHC proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ with the CMS detector, is described. The dataset corresponds to an integrated luminosity of $\mathcal{L} = 42 \text{ pb}^{-1}$.

8th International Workshop on Top Quark Physics, TOP2015

14-18 September, 2015

Ischia, Italy

*Speaker.

1. t -channel signature

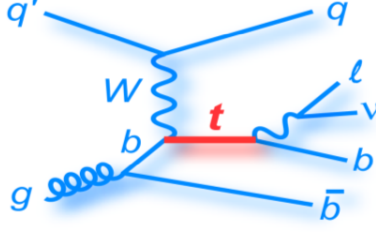


Figure 1: A representative diagram for the electroweak production of top quarks via t -channel.

The electroweak t -channel production of top quarks (Fig. 1) is the most abundant single top production mode in the Standard Model (SM) with a predicted cross section at approximate next-to-next-to-leading order (NNLO) that equals to $\sigma_{t\text{-ch.}} = 214.5_{-1.2}^{+1.0}$ (scale) pb at $\sqrt{s} = 13 TeV$ [1]. More specifically, the signature of single top quark events produced via t -channel and with a leptonic top-quark decay is dictated by the presence of exactly one **isolated lepton** (l), a **b -tagged jet** (b), and **escaping neutrinos** (ν) from the top quark decay chain, as well as a **light flavour jet** (q) recoiled mainly in the forward region. It should be noted that a second b quark coming from gluon splitting results in an **additional b -tagged jet** (\bar{b}) with a softer p_T spectrum and a broader η distribution.

Events are selected according to the above topology and divided in categories according to the number of jets and b -tagged jets using the naming convention of N -jet- M -tag, referring to events with N jets, M of which are b -tagged [2]. The category enriched with t -channel signal events is referred to as 2-jet-1-tag (2j1t). Exclusively for this category, an additional requirement is applied on the mass of the reconstructed top quark to lie within the $130 < m_{l\nu b} < 225 GeV$ range. The observed and expected event yield, inside (SR) and outside (SB) the $m_{l\nu b}$ window, is summarized below at the recorded $\mathcal{L} = 42 pb^{-1}$, after imposing the $m_T^W > 50 GeV$ requirement on the transverse W -boson mass (Table 1). The systematic uncertainty accounts for roughly 10% of the final uncertainty on the estimated QCD yield. Possible sources of *non-prompt* or misidentified leptons render QCD multijet events signal-like (cf. Sec. 2).

Process	SR	SB
$t\bar{t}$ & tW	157 ± 1	71.7 ± 0.4
W/Z+jets	40 ± 4	47 ± 4
QCD	10 ± 5	2 ± 1
t -channel	33 ± 1	7.2 ± 0.3
Total expected	240 ± 6	128 ± 4
Data	252	127

Table 1: Event yields for the signal and the main background processes in the 2-jet-1-tag sample inside (SR) and outside (SB) the $m_{l\nu b}$ window [3]. Only top quark decays to muons are considered during the current study. Except for the number of QCD multijet events, all yields are taken from simulation and their uncertainties correspond to the size of the simulated samples. The yield of QCD multijet events is determined from the data.

2. Multijet model from data

Given the associated theoretical uncertainties of QCD modelling, for instance stemming from a not precise enough knowledge of the cross section or the approximate description of the underlying hadronic activity, it is necessary to predict the size and properties of this process by data. A reliable model for determining QCD contamination is then derived by fitting templates of all physics processes to the x ($\equiv m_T^W$) discriminant (Fig. 2), i.e.

$$F(x) = N_{\text{QCD}} \times Q(x) + N_{\text{non-QCD}} \times W(x). \quad (2.1)$$

The $W(x)$ templates, representing the signal and other non-QCD background processes, are taken from the Monte Carlo simulation. After having subtracted non-QCD contamination, the $Q(x)$ template is constructed from a dedicated control sample that will be described in the following.

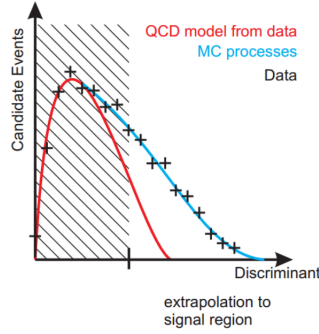


Figure 2: Fictitious example that establishes the basic principle behind QCD estimation. The QCD shape distribution is extracted by data over a phase space in the proximity of the signal region. The total amount of QCD events is then determined by a template fit having utilized a variable that discriminates QCD against non-QCD processes.

QCD multijet events become signal-like when an additional lepton is presented in the final state, where a lepton could simply be the decay product of a hadron formed within jets. Typically, such *non-prompt* leptons, as opposed to leptons having been produced from the hard scattering, are surrounded by many activities within the jet. On the other hand, leptons arisen out of the W-boson leptonic decay are expected to be isolated, meaning that they should be found in a restricted spatial region where they contribute with a major fraction in the deposited energy. Therefore a QCD enriched data sample, yet orthogonal with the signal region, is achieved by referring to leptons failing the isolation requirement.

Relative isolation, I_{rel} , is defined in this analysis as

$$I_{\text{rel}} = \frac{I^{\text{ch.h}} + \max((I^\gamma + I^{\text{n.h}} - I^{\text{PU}}), 0)}{p_T}, \quad (2.2)$$

where $I^{\text{ch.h}}$, I^γ and $I^{\text{n.h}}$ are the sum of the transverse energies of stable particles in a cone $\Delta R = 0.4$ around the lepton direction with momentum p_T ; $I^{\text{PU}} \equiv \Delta\beta \times \sum p_T^{\text{PU}} = 0.5 \times \sum p_T^{\text{PU}}$ is the sum of transverse momenta of tracks associated to pileup vertices. The $\Delta\beta$ prefactor corresponds to the naive expectation about the average of neutral to charge particle ratio and corrects the total contribution of neutral $I^\gamma + I^{\text{n.h}}$ candidates for pileup contamination. Such correction is validated by monitoring the stable behavior of $I^\gamma + I^{\text{n.h}}$ against the number of overlapping collisions.

3. Results

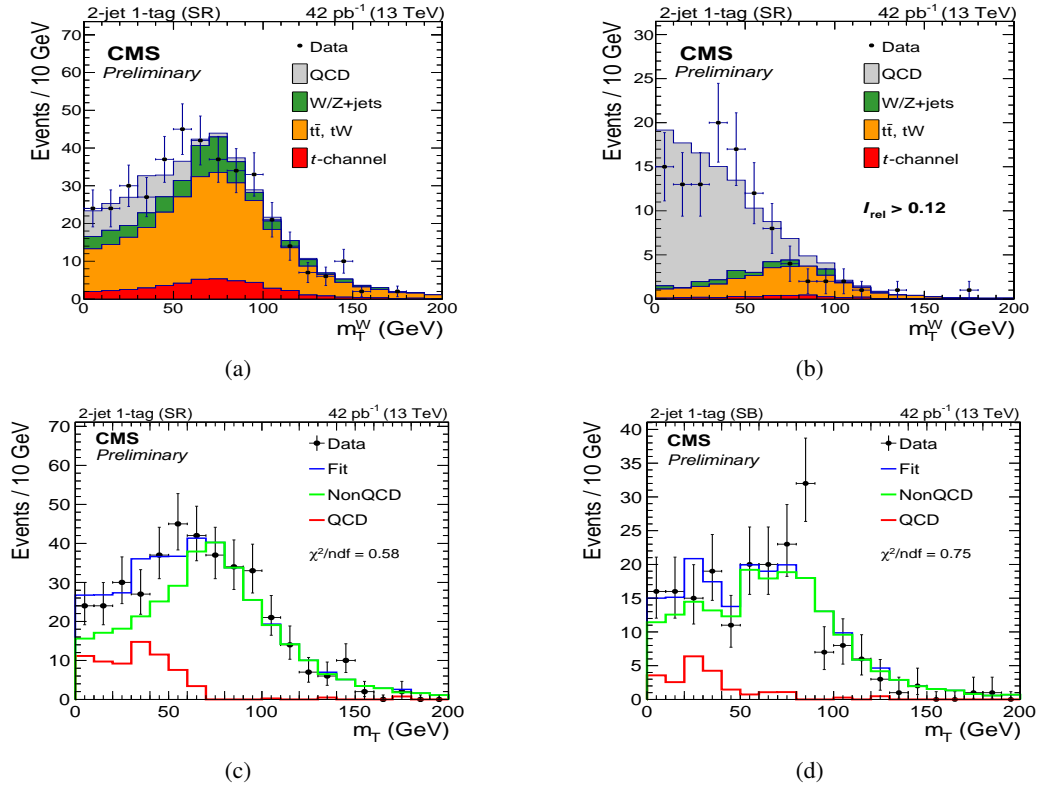


Figure 3: QCD estimation in 2-jet 1-tag region.

As important as rejecting QCD away is, constraining the amount of its residual contribution to the SR window is equally crucial. Indeed, the reconstructed m_T^W (Fig. 3, a) shows reasonable agreement between data and simulation, after both QCD and non-QCD components have been scaled to the estimated contribution, the latter determined by the fit process. The $I_{rel} > 0.12$ separation effectively results in a high statistics model for the QCD background, while non-QCD contamination is found less dominant (Fig. 3, b). To exploit the discrimination power of m_T^W as much as possible, the fit (Eq. 2.1) is performed without imposing the 50 GeV requirement. Thus the contribution from QCD multijet events is determined, separately in SR (Fig. 3, c) and SB (Fig. 3, d) windows, and the QCD yield is finally evaluated for the $m_T^W > 50$ GeV signal region [3].

References

- [1] M. Brucherseifer, F. Caola, and K. Melnikov, *On the NNLO QCD corrections to single-top production at the LHC*, Phys. Lett. **B 736**, 58 (2014) [arxiv:1404.7116]
- [2] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, JINST **8**, P04013 (2013) [arxiv:1211.4462]
- [3] CMS Collaboration, *Measurement of the t-channel single top-quark cross section at $\sqrt{s} = 13$ TeV*, CMS PAS TOP-15-004 (2015)