# CMS Physics Analysis Summary

Contact: cms-pag-conveners-top@cern.ch 2009/07/07

# Prospects for the first Measurement of the $t\bar{t}$ Cross Section in the Muon-plus-Jets Channel at $\sqrt{s} = 10$ TeV with the CMS Detector

The CMS Collaboration


## Abstract

A feasibility study is presented of the first measurement of the $t\bar{t}$ pair production cross section in the muon-plus-jets channel at a center-of-mass energy of 10 TeV, using the CMS detector at the LHC. Events are selected which contain one highly energetic muon and at least four jets. Several techniques are used to identify the three jets originating from the hadronic top decay. The $t\bar{t}$ cross section is extracted by means of template fits to discriminating distributions. The amount of QCD background is determined in a data-driven way using two complementary approaches. Statistical and systematical uncertainties are evaluated using ensemble tests. With 20 $\text{pb}^{-1}$ of data, the $t\bar{t}$ pair production cross section in the muon plus jets channel is expected to be measurable with 12-18% statistical and 20-25% systematic error, where the latter is dominated by the jet energy scale uncertainty.

# 1 Introduction

At the Large Hadron Collider (LHC) top quark pairs will be produced copiously. Around 8M $t\bar{t}$ pairs will be produced per year of nominal data taking (10 fb$^{-1}$). At LHC the $t\bar{t}$ production is predominantly gluon induced, in contrast to the Tevatron where the top quark pairs are mostly produced from $q\bar{q}$ in the proton and anti-proton. The estimated increase of the $t\bar{t}$ pair production cross section at the LHC compared to the Tevatron is about two orders of magnitude. The $t\bar{t}$ cross section has been estimated as $\sigma(t\bar{t}) = 414 \pm 40(\text{scale}) \pm 20(\text{PDF})$ pb [1] at a center-of-mass energy of $\sqrt{s} = 10$ TeV, which is envisaged for the 2009-2010 run.

Top quark events contain almost all relevant experimental signatures which need to be understood in order to claim successful commissioning of CMS, such as jets, missing transverse energy and leptons. Therefore, the observation of top quark events in CMS can be considered as a milestone in the physics commissioning of the experiment. An early confirmation of the observation of top quark pair events will also be essential to validate possible new physics signals. Once the $t\bar{t}$ signal has been established, there will be a rich top quark physics program.

Top quark events are also very useful to measure the performance of the detector. In particular, $t\bar{t}$ events can be used to determine the efficiency for identifying jets originating from a $b$-quark (*b-tagging*) [2]. They also serve as input for calibrating the absolute jet energy scale, making use of the known mass of the $W^{\pm}$ boson [3, 4].

The potential of CMS to measure the $t\bar{t}$ cross section in the semileptonic channel for integrated luminosities of 1 fb$^{-1}$ or higher has already been studied in [5–7]. In contrast to previous studies, the scope of this analysis is to address the potential of the CMS detector to establish a top quark signal within the first 20 pb$^{-1}$ of LHC data at $\sqrt{s} = 10$ TeV. The semileptonic decay into one highly energetic muon plus jets and missing transverse energy is considered. A first investigation of the $t\bar{t}$ rediscovery in the muon-plus-jets channel at $\sqrt{s} = 14$ TeV was presented in [8].

The goal of this analysis is to design a simple and robust method able to identify top quark pairs with the lowest possible integrated luminosity, using the very early dataset of about 20 pb$^{-1}$. Events are selected which contain one high transverse momentum muon and at least four jets. Several techniques are employed which try to identify the three jets originating from the hadronic top decay. Two different methods are presented in which the amount of QCD background is extracted in a data-driven way. One is using two uncorrelated variables which distinguish between signal and QCD background (ABCD method), whereas the other one employs an extrapolation of the muon isolation variable. The $t\bar{t}$ cross section is extracted using a binned likelihood template fit to observables which are sensitive to the composition of the sample in terms of of $t\bar{t}$ signal and background processes. The statistical and main systematical uncertainties are addressed employing ensemble tests. Finally, a novel alternative method is proposed to subtract the background from $W$ boson + jets production from the selected event sample using the charge asymmetry in $W$ events.

# 2 Simulation

The simulation of $t\bar{t}$ events was performed using MadGraph [9]. Top quark pairs are accompanied by up to four additional hard jets, and the hard parton configurations generated by MadGraph are matched to parton showers utilizing PYTHIA [10] using the MLM matching prescription. For systematic studies, an alternative $t\bar{t}$ sample generated with PYTHIA was used.

Table 1: Data samples used in the analysis. All samples were generated with MadGraph, except the QCD multijet sample, which was made using PYTHIA. The $t\bar{t}$ signal cross section has been normalized to 414 pb$^{-1}$ [1]. The W+jets and Z+jets cross sections have been scaled by a k-factor of 1.14 [12]. For the single top samples, NLO cross sections have been used. In the vector boson(s) plus jets and single top s- and t-channel samples, only leptonic boson decays are simulated. The QCD sample is filtered at the generator level and corresponds to an integrated luminosity of around 50 pb$^{-1}$.

| Process | $\sigma$ [pb] | Events used |
|---|---|---|
| $t\bar{t}$ +jets | 414 | 1M |
| W+jets ($W \to l\nu_l$) | 45,600 | 10M |
| Z+jets ($Z \to l^+l^-$) | 4,218 | 1M |
| WW/ZZ/WZ+jets ($Z \to l^+l^-$, $W \to l\nu_l$) | 11.8 | 100K |
| single top s-Channel ($W \to l\nu_l$) | 1.67 | 12K |
| single top t-Channel ($W \to l\nu_l$) | 43.3 | 280K |
| single top tW-Channel | 28.9 | 170K |
| QCD multijet | $509 \cdot 10^6$ | 6.2M |

The electroweak production of single top quarks is taken into account as a background process. More information about the simulation of the single top samples can be found in [11].

The production of $W$ and $Z$ bosons in association with up to four extra jets, where the vector boson decays leptonically, has a similar signature and thus constitutes the main background to semileptonic $t\bar{t}$ events. These processes are also simulated using MadGraph. Di-boson production (WW, WZ, ZZ) in association with extra jets is also simulated using MadGraph.

QCD events with several jets and a muon which passes the selection cuts (either a real muon e.g. from semileptonic decays of hadrons containing a $b$ or $c$ quark, fake muons, decays-in-flight or punch-throughs) may constitute another significant source of background. This background has a very large cross section and is very difficult to model. Therefore, this background contribution has to be determined from data. Nevertheless, in the absence of data a first idea about the size of the QCD background can be obtained from simulation. A high-statistics sample generated with PYTHIA is used in the analysis, which is pre-filtered at the generator level for the presence of a muon, including decays-in-flight. Table 1 summarizes the simulated data samples used in the analysis.

# 3 Reconstruction and Event Selection

Semileptonic $t\bar{t}$ events contain one high $p_T$ lepton in the final state. Thus, single muon triggers are well suited in muon+jets analyses where the $W$ boson from one top quark decays into $W \to \mu + \nu_\mu$, as they provide clean signatures and have comparatively low trigger thresholds. In this analysis, the events are triggered by an inclusive single muon trigger with a threshold of $p_T > 9$ GeV, which is defined in the CMS start-up trigger table for luminosities of $8 * 10^{29}$ cm$^{-2}$s$^{-1}$. The trigger efficiency, defined as the fraction of events passing all cuts which satisfy the muon trigger requirement, is evaluated as 92%. In real data, the trigger efficiency must be determined using for example an independently triggered sample or by using the "Tag-and-Probe" [13] method.

Offline, muons are reconstructed with information from both the muon system and the silicon tracker [13]. The momentum scale as well as the reconstruction and identification efficiencies will be determined using e.g. a "Tag-and-Probe" method in $Z \to \mu\mu$ events. In order to fur-

ther enhance the muon purity, additional quality cuts are applied, and the calorimeter energy deposit around the muon track must be consistent with a minimum ionizing particle.

Isolation requirements are placed on the muon candidates, in order to distinguish muons from $W$ decays from muons in jets, which originate mostly from semileptonic b/c-hadron decays. A tracker isolation variable $p_{T,iso}^{tracker}$ is defined by forming the sum of the transverse momenta of all tracks found within a cone of size R=0.3 around the muon direction, excluding the muon track. Similarly, a calorimeter isolation variable $E_{iso}^{calo}$ is defined by summing the energies of all calorimeter towers within $R < 0.3$ around the muon direction, excluding the muon energy deposited in the calorimeter. A combined isolation variable can be defined as RelIso $= (E_{iso}^{calo} + p_{T,iso}^{tracker})/p_{T,\mu}$, where the isolation is calculated with respect to the transverse momentum of the muon $p_{T,\mu}$.

Exactly one isolated muon with $p_{T,\mu} > 20$ GeV/c, pseudorapidity $|\eta| < 2.1$, relative isolation RelIso $< 0.05$ and transverse impact parameter with respect to the beam-spot $d_0 < 200$ $\mu$m is required. The $\eta$ cut is motivated by the acceptance of the muon trigger. The $p_T$ cut, the isolation requirement and the requirement on the transverse impact parameter are motivated in order to reduce the amount of non-W (i.e. QCD) background.

A cut on the muon impact parameter significance was also studied and achieves a similar QCD background rejection. However, applying this cut would require a good knowledge of the tracking and beam spot position uncertainties, which may not be well known in the early data taking phase.

Events with more than one muon are rejected in order to reduce the contamination from dileptonic top decays, which are treated as background here, as well as from $Z$+jets and Di-boson events. In order to become statistically independent from the semileptonic electron channel [14], events with a good, isolated electron with transverse energy $E_T > 30$ GeV/c are rejected.

In addition to the rejection of events with a tight selection on a second muon or an electron, events with a loose ($p_T > 10$ GeV/c for muons, $E_T > 15$ GeV for electrons, RelIso $< 0.2$ for both) second muon or a loose electron are rejected. This is motivated by the rejection of $Z$+jets and Di-boson events. On the other hand, events from $t\bar{t}$ decay modes other than semileptonic muon decays, e.g. dilepton $e + \mu$ or $\mu + \mu$, can also be rejected.

For the clustering of the jets the Seedless Infrared-Safe Cone (SISCone) algorithm [15] with an cone size of $R = 0.5$ is used. A detailed study on the performance of this algorithm for the CMS detector can be found in [16]. CMS plans to determine the Jet Energy Scale (JES) with a multi step approach [17]. The jet energies used in the analysis are corrected to the ones of jets formed from final state hadrons.

The events need to meet the requirement of at least four reconstructed jets in the range $|\eta| < 2.4$ with a transverse momentum of $p_T > 30$ GeV/c. Due to higher order diagrams or parton showers, the four jets from the $t\bar{t}$ decay are often accompanied by additional jets. The $\eta$ range corresponds to the acceptance of the silicon tracker, in order to potentially facilitate the use of $b$-tagging. However, the analysis does not distinguish $b$-jets and non $b$-jets.

The event yields obtained with the selection described above are summarized in Table 2, scaled to an integrated luminosity of 20 pb$^{-1}$. The final selection yields 320 $t\bar{t}$+jets events, out of which 43 events originate from decay modes other than semileptonic muon events. The background event yield is 171 events leading to a signal-to-background ratio of $S/B = 1.9$, a pseudo-significance $S/\sqrt{B} = 24.5$, and $S/\sqrt{S+B} = 14.4$. Among the background events there are

Table 2: Cut-flow table. Event yields normalized to an integrated luminosity of 20 pb$^{-1}$, where cuts are sequentially applied. The second row, labelled 'Trigger', corresponds to the number of all generated events which satisfy the trigger requirement.

| | $t\bar{t}$+jets s.l. $\mu$ | $t\bar{t}$+jets other | Single top s-Ch. | Single top t-Ch. | Single top $tW$ | $W$+jets | $Z$+jets | $VV$+jets | QCD |
|---|---|---|---|---|---|---|---|---|---|
| AllEvents | 1,220 | 7,060 | 32 | 832 | 580 | 912,000 | 76,240 | 236 | 2,546,279 |
| Trigger | 978 | 1,418 | 10 | 260 | 147 | 168,633 | 20,952 | 100 | 2,032,021 |
| $\geq 1$ tight $\mu$ | 620 | 345 | 5 | 140 | 69 | 110,509 | 15,296 | 73 | 7,200 |
| $< 2$ tight $\mu$ | 620 | 309 | 5 | 140 | 66 | 110,509 | 9,300 | 62 | 7,200 |
| no tight $e$ | 620 | 264 | 5 | 140 | 62 | 110,508 | 9,292 | 53 | 7,200 |
| veto on loose $\mu$ | 618 | 228 | 5 | 140 | 60 | 110,503 | 5,492 | 44 | 7,192 |
| veto no loose $e$ | 616 | 183 | 5 | 140 | 56 | 110,469 | 5,415 | 34 | 7,188 |
| $\geq 1$ jet | 614 | 180 | 4 | 125 | 55 | 16,998 | 1,325 | 18 | 2,701 |
| $\geq 2$ jets | 593 | 158 | 3 | 63 | 47 | 3,076 | 256 | 5 | 387 |
| $\geq 3$ jets | 489 | 99 | 1 | 18 | 27 | 651 | 51 | 1 | 60 |
| $\geq 4$ jets | 277 | 43 | 0 | 5 | 9 | 140 | 10 | 0 | 7 |

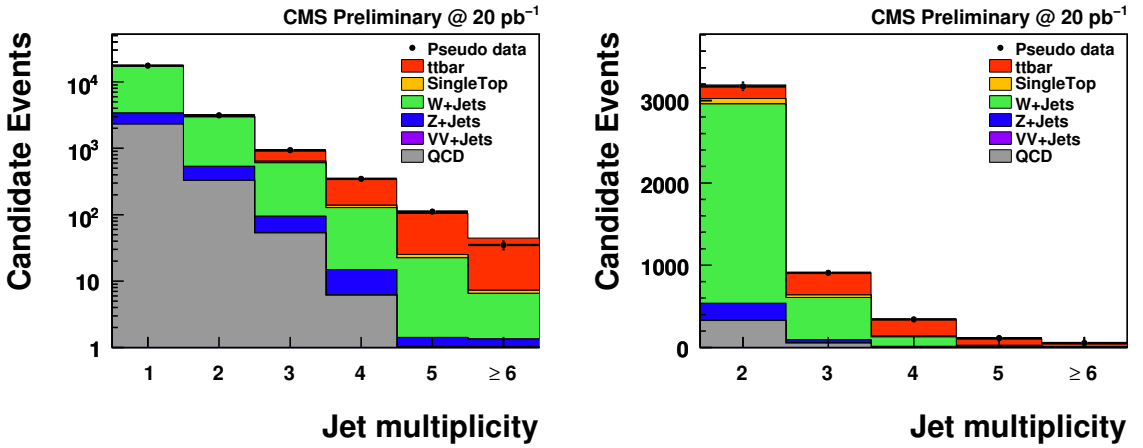


Figure 1: Expected event numbers for an integrated luminosity of 20 pb$^{-1}$ as a function of jet multiplicity on a logarithmic (left) and on a linear (right) scale. For jet multiplicities of four and higher, the sample is dominated by $t\bar{t}$ signal events, while the lower jet bins are dominated by background from $W$+jets and multi-jet QCD events. Here and in the following figures, the *pseudo data* distribution is obtained by applying a bin-by-bin smearing based on Poisson statistics.

14 single top, 140 $W$+jets, 10 $Z$+jets and 7 QCD events. An overall selection efficiency of 22.7% (including acceptance and trigger) is obtained for semileptonic $t\bar{t}$ events in the muon channel. Of the selected $t\bar{t}$ events, around 86% originate from true $pp \to t\bar{t} + X \to bq\bar{q}b\mu\nu_\mu + X$ decays.

Figure 1 shows the jet multiplicity with the final event selection applied, except for the cut on the number of reconstructed jets. Figure 2 shows other kinematic distributions for the final selected dataset normalized to an integrated luminosity of 20 pb$^{-1}$. Also shown in Figure 2 is the distribution of the relative isolation RelIso, before the cut RelIso $< 0.05$ is applied. It clearly is able to separate QCD multi-jet events from W-like events, which will be further exploited in Section 5.2.

Figure 3 shows the total number of $b$-tagged jets in the selected sample with at least four jets. Even though $b$-tagging is not applied during the event selection, the distribution illustrates that the event sample is enriched in b-jets. One of the possibilities to demonstrate with the first data

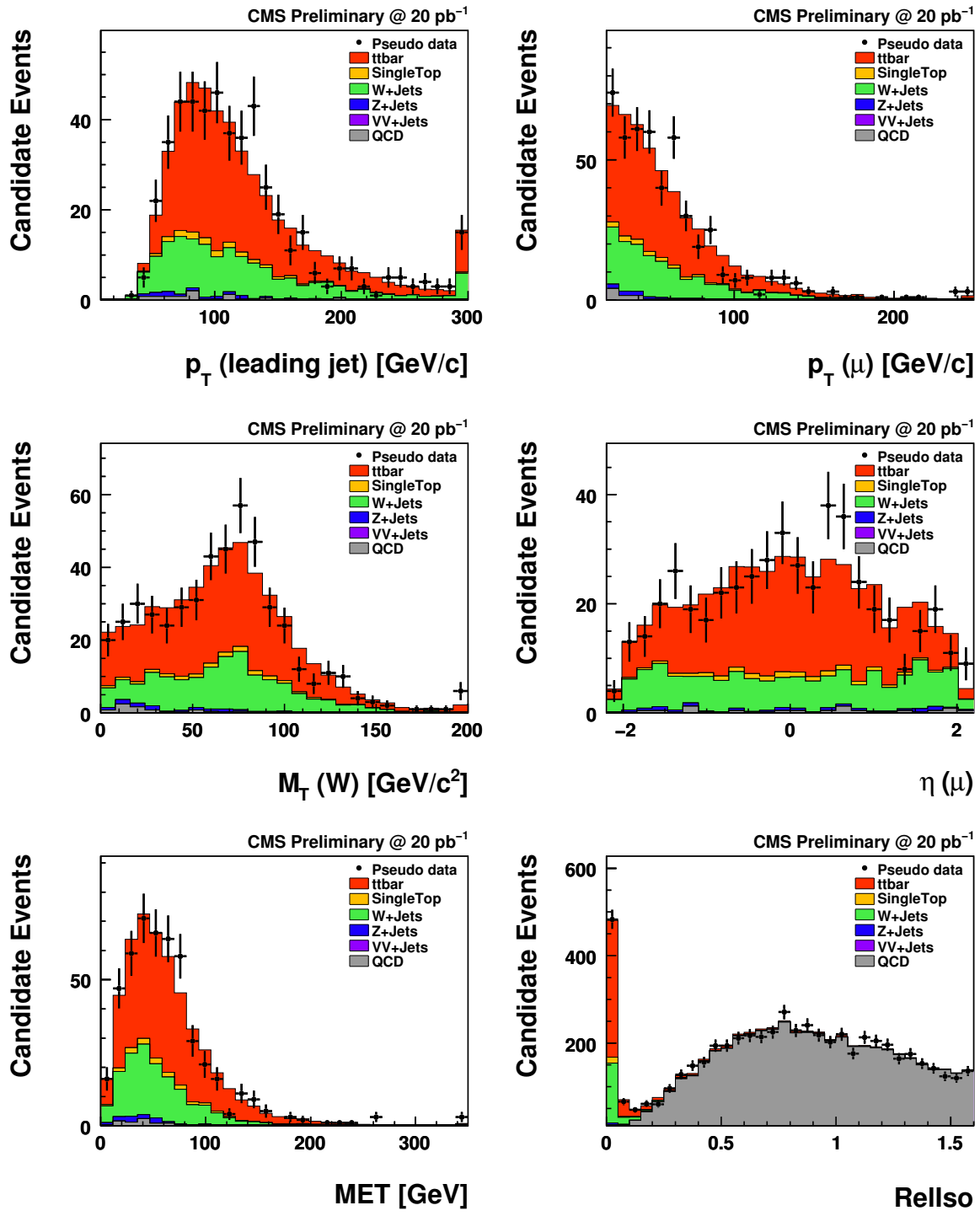


Figure 2: Kinematic distributions for the final event selection, scaled to an integrated luminosity of 20 pb$^{-1}$. Shown are the $p_T$ distributions of the leading jet and of the muon, the transverse $W$ mass, the pseudorapidity of the muon, and the missing transverse energy. Finally, the RelIso distribution is shown, where the cut RelIso $< 0.05$ is not applied.

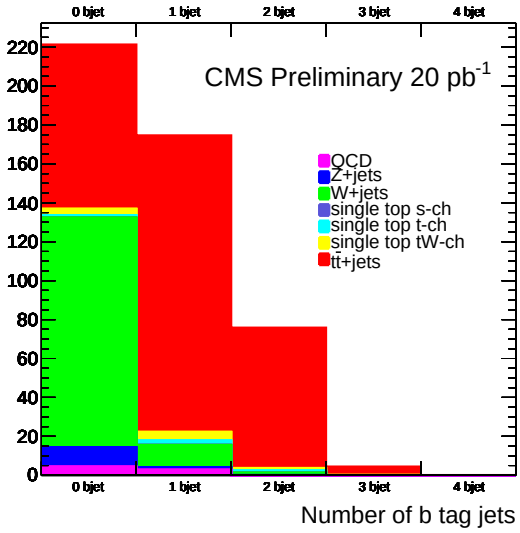


Figure 3: Number of $b$-tagged jets in the selected sample with at least four reconstructed jets. Even though this analysis does not use $b$-tagging in the event selection, the distribution illustrates that the event sample is enriched in b-jets.

that the selected events are indeed dominated by $t\bar{t}$ events will be to study if the observed b-jet multiplicity is in line with expectations. b-jets are identified using an algorithm which requires the presence of a reconstructed secondary vertex in the jet. It has been demonstrated in [18] that this algorithm is fairly insensitive to the effects of tracker misalignment expected in the early data. Under ideal conditions, its b-tagging efficiency is around 63% for a mis-tagging probability of around 1%.

# 4 Top Quark Reconstruction

Confidence that an excess observed at high jet multiplicities is indeed due to $t\bar{t}$ events can be obtained from looking at a distribution which is sensitive to the mass of the hadronically decaying top quark.

## 4.1 M3 Method

A simple way to identify the three out of four jets which originate from the hadronic top decay is to calculate the vectorially summed transverse momentum of any combination of three jets. The jets of the combination with the highest summed $p_T$ are deemed to originate from the hadronic top decay, and their invariant mass is denoted $M3$. Only the hadronic leg of the $t\bar{t}$ decay is reconstructed with this procedure.

Figure 4 shows the M3 distribution for the selected event sample. It can be seen that the distribution peaks at a value close to the nominal top mass. The peak value is shifted to higher values due to the fact that the jet energies are corrected using correction factors derived from QCD dijet events (mostly gluon jets) which are here applied to quark jets. The dominant W+jets background exhibits a broader distribution compared with the signal.

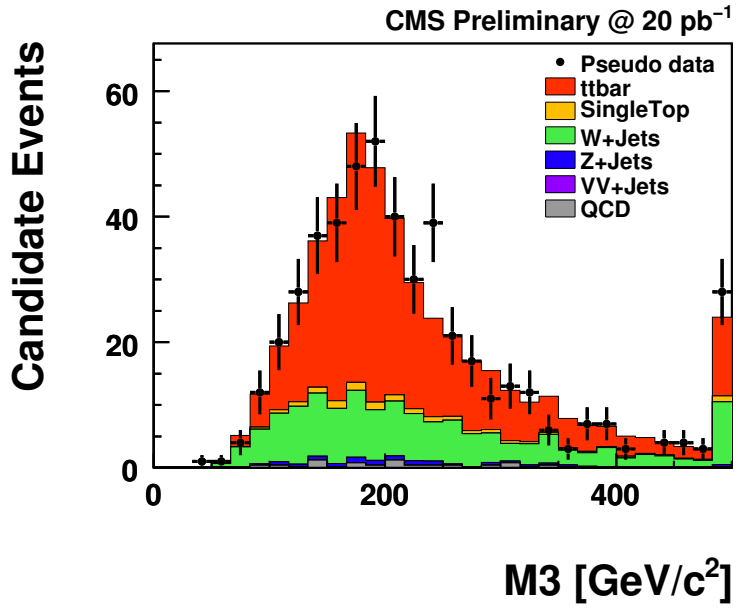


Figure 4: Invariant mass M3 of the three jets with the highest vectorially summed transverse momentum.

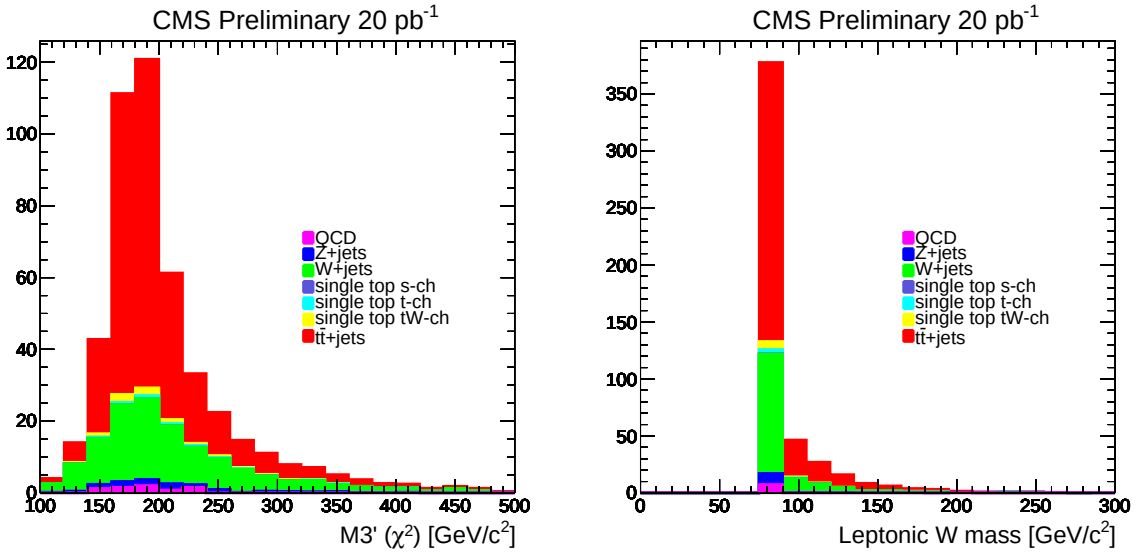


Figure 5: Left: Distribution of the M3′ variable, defined using the best $\chi^2$ jet combination, as explained in the text. Right: Leptonic W-mass distribution.

## 4.2 Reconstruction using a $\chi^2$-sorting method

A second method is based on a $\chi^2$ distribution which is defined as

$$\chi^2 = \frac{(m_{j_1 j_2} - m_W)^2}{\sigma_{jj}^2} + \frac{(m_{j_1 j_2 j_3} - m_t)^2}{\sigma_{jjj}^2} + \frac{(m_{\mu\nu j_4} - m_t)^2}{\sigma_{\mu\nu j}^2}, \quad (1)$$

where $m_{j_1 j_2}$ and $m_{j_1 j_2 j_3}$ are the dijet and tri-jet invariant masses, $m_W$ and $m_t$ are the nominal values of the W boson and top quark masses, and $m_{\mu\nu j_4}$ is the invariant mass of the $\mu + \nu$+jet combination. The $\sigma_i$ are the resolutions for each case of jet combination. The resolutions used in this analysis are obtained from MC and have fixed values for all events. The numerical values are $\sigma_{jj} = 10.5$, $\sigma_{jjj} = 19.2$, and $\sigma_{\mu\nu j} = 24.2$ GeV/$c^2$ . The $\chi^2$ is calculated for each permutation of jets, using at most seven of the reconstructed jets. The permutations are sorted in $\chi^2$ and the best combination of jets is selected as the one with the lowest $\chi^2$. This method distributes the jets that are associated to the hadronic and leptonic legs, and also assigns the jets that are assumed to be $b$-jets in each decay branch. Figure 5 (left) shows the invariant mass of the hadronic leg for the jet combination with the lowest $\chi^2$. This distribution is called M3'($\chi^2$), and the method is called $\chi^2$-*sorting*.

The full reconstruction of the neutrino is needed to estimate a $\chi^2$ for each permutation. The $p_Z$ of the neutrino is calculated using a quadratic equation which relates the W-mass to the muon momentum and the missing transverse energy (MET). MET is calculated from calorimeter towers, including corrections due to the jet energy scale as well as due to muons. The solution closest to the $p_Z$ of the muon is chosen. In the case of complex solutions, the real part is chosen as the solution. About 35% of the solutions are complex. This is mostly due to misreconstructed missing transverse energy, leading to a tail at high values of reconstructed leptonic W masses (Figure 5 right). Therefore, we apply an additional cut $m_{\mu\nu} < 150$ GeV/$c^2$ to reduce badly reconstructed events.

In order to evaluate the fraction of events where the jets used to calculate the M3 or M3' variables are correctly matched to the quarks originating from the hadronic top decay, a sub-set of generated $t\bar{t}$ events is studied where the $t\bar{t}$ decay products are in principle reconstructible. This is performed by applying detector-matched generator level cuts on the rapidities and transverse momenta of the $t\bar{t}$ decay partons ($p_T > 30$ GeV/c, $|\eta| < 2.4$). Using this subsample, the fraction of events with correct matching of hadronic top decay partons and the reconstructed jets used in calulating the mass estimator is estimated as around 30% for the M3 variable, whereas around 70% is obtained for the M3'($\chi^2$) method.

The benefit of the $\chi$2-sorting method is that it can be used to select a sample which is enriched in well-matched $t\bar{t}$ events. In principle, the number of well-matched events can be extracted in a data-driven way. Wrong permutations of jets are expected to reproduce the shape of the background in the M3' distribution. As an example, the jet combination with the 3rd lowest $\chi^2$ is considered. In Figure 6 (left), the M3' shape obtained from these jet combinations is compared with the shape of the M3'(lowest $\chi^2$) for the non-$t\bar{t}$ backgrounds. A good agreement is observed, suggesting that using the jet combination with the 3rd lowest $\chi^2$ can indeed be used as a background model. Subsequently the background is parameterized using the sum of a Gaussian and a Landau function.

The M3' distribution is then fitted by a sum of background and signal components, where the background is parameterized as described above, and the signal ($t\bar{t}$ events with correct jet combination) is parameterized with a Gaussian function. The result of a binned maximum likelihood fit to the pseudo-data distribution is shown in Figure 6 (right). The number of fitted

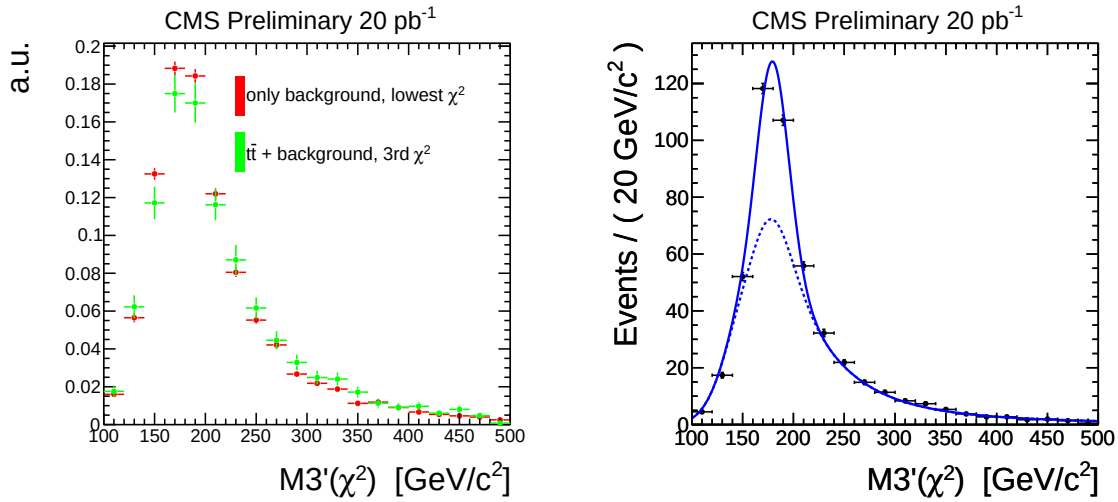


Figure 6: Left: Comparison between the shape of the M3′ background-only distribution and the distribution obtained by using wrong jet combinations for signal and background. Right: Fit to the reconstructed M3′ distribution using a background template obtained from wrong jet combinations, plus a Gaussian function for the signal, which corresponds to well-matched $t\bar{t}$ events.

$t\bar{t}$ signal events is in reasonable agreement with the true value (those reconstructed $t\bar{t}$ signal events where the jets have been correctly matched to the generated partons from the top decay).

# 5 Estimation of QCD Background

In this section, two data-driven techniques to estimate the number of QCD events are discussed: a method using two uncorrelated variables which separate signal and QCD background (so-called ABCD method), and an extrapolation fit using the muon isolation variable.

## 5.1 ABCD Method

The ABCD method provides a way to estimate the number of background events in the signal region by taking advantage of the discrimination power in the phase space region of two independent variables. The method assumes that the two variables are at most weakly correlated. Four regions are defined in the phase space given by the two variables. Region A is the region dominated by signal events while regions B, C, and D are mostly dominated by background events. Therefore, the ratio of events $N_A/N_B$ should be equal to the ratio of events $N_C/N_D$. The number of events in the signal region can then be estimated as $N_A = (N_B \cdot N_C)/N_D$ .

For this study, two weakly correlated variables are used:

- The modified relative isolation $\text{RelIso}' = 1/(1 + \text{RelIso})$, which is defined in the interval $[0, 1]$. W+jet like events with isolated muons are peaked at values close to one in this variable.
- The muon impact parameter significance $\text{sig}(d_0) = d_0/\sigma(d_0)$, where the impact parameter error $\sigma(d_0)$ takes the uncertainty in the beam spot position into account.

Figure 7 shows the two-dimensional distribution of the $\text{sig}(d_0)$ and RelIso′ variables, as well as the four regions A, B, C, and D. It has been verified that the two variables are only weakly

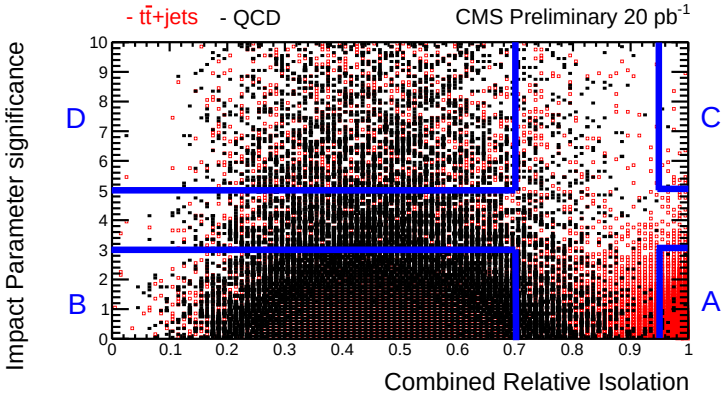


Figure 7: Two-dimensional distribution of the muon impact parameter significance sig$(d_0)$ and RelIso' variables, as defined in the text.

Table 3: Results of the ABCD method as a function of jet multiplicity. The results are in agreement with the expected number of events within errors for the signal region N-jets$\geq 4$, and in the control regions $\leq 3$.

| Jets | N(QCD) Predicted | $N_B$ | $N_C$ | $N_D$ | N(QCD) Estimated |
|---|---|---|---|---|---|
| 2 | 327 | 86625 | 61 | 16240 | $325 \pm 26$ |
| 3 | 53 | 24216 | 10 | 5058 | $48 \pm 9$ |
| $\geq 4$ | 7 | 5345 | 3 | 1148 | $12 \pm 5$ |

correlated and can thus be used for the ABCD method (see Appendix A).

The results of the ABCD method as a function of jet multiplicity are shown in Table 3. They are in good agreement with the expected values within statistical errors for all jet bins. The stability of the ABCD method has been cross checked by changing the boundaries of the background regions B, C, and D, while keeping the signal region A fixed. The results are also stable as a function of jet multiplicity. From these studies, the uncertainty of the method is conservatively estimated as 50%.

## 5.2 Extrapolation of the Isolation Variable

In order to estimate the number of QCD events in the final event selection the RelIso distribution is employed before the cut on this variable is applied, see Figure 2. W-like events which contain an isolated muon are strongly peaked towards very small values of RelIso. On the other hand QCD jet events are broadly distributed. These features can be exploited in order to estimate the number of QCD events in the signal region (RelIso $< 0.05$) in a data driven way.

In a first step an appropriate fit function has to be found which models the isolation distribution in the background region dominated by QCD events, and can be extrapolated to the signal region. Using the multi-jet distribution obtained from the QCD Monte Carlo sample, the best performance was found using a Landau function [19].

For a data-driven estimation of the number of QCD events in the final selection, the Landau function is fitted to the RelIso distribution in a side-band region. After the fit is performed in the side-band region (RelIso $> 0.3$), an extrapolation of the function is employed to estimate the number of QCD events in the signal-region (RelIso $< 0.05$) as the integral of the Landau

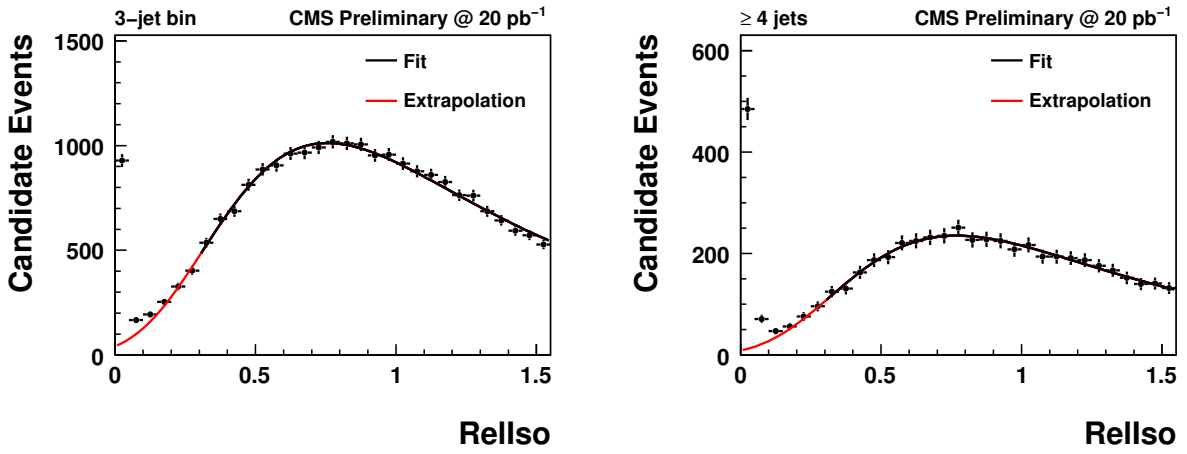


Figure 8: Fit of the Landau function in the side band region of the RelIso distribution and extrapolation of the fit function into the signal region for the 3-jet (left) and $\geq$ 4-jet bins (right).

Table 4: Results of the estimation of the number of QCD events using the RelIso extrapolation fits.

| Jets | N(QCD) Predicted | N(QCD) Estimated |
|---|---|---|
| 2 | 327 | $378 \pm 82$ |
| 3 | 53 | $47 \pm 24$ |
| $\geq 4$ | 7 | $13 \pm 7$ |

function in this range. An example for the fit and extrapolation in the 3-jet and $\geq$ 4-jet bins is given in Figure 8.

The stability of the method is studied by varying the upper and lower boundaries for the sideband region, as well as the bin width. It is observed that the peak position of the RelIso distribution depends on the jet multiplicity, and thus in general the optimal fit range will be different for each jet multiplicity bin. A method to automatically determine the optimal fit range by means of a $\chi^2$ minimization has been investigated as well. Similar fit studies have also been performed using the RelIso' distribution, where a Gaussian function is used to model the sideband region. The results of the method can be found in Table 4. The uncertainty of this data-driven method for the estimation of QCD events in the signal region can be estimated employing the results from the studies with varied fit boundaries as explained above. From these variations the uncertainty is conservatively estimated as 50%.

## 6 Determination of the $t\bar{t}$ Cross Section

The theoretical prediction for the $t\bar{t}$ production cross section at center-of-mass energy of 10 TeV at NLO is 414 pb. Experimentally the $t\bar{t}$ cross section can be obtained from the number of measured $t\bar{t}$ events ($N_{t\bar{t}}$) as

$$\sigma(t\bar{t}) = \frac{N_{t\bar{t}}}{A \cdot \epsilon \cdot L}\,, \tag{2}$$

where $A \cdot \epsilon$ is the efficiency (trigger and reconstruction) times acceptance, determined from Monte Carlo in this study, and $L$ is the integrated luminosity of the analyzed data set. The method to obtain the number of $t\bar{t}$ events will be discussed in the following.

In order to extract the number of $t\bar{t}$ events from data, a binned likelihood fit method to an appropriate variable is used. Appropriate in this sense is a variable which discriminates between

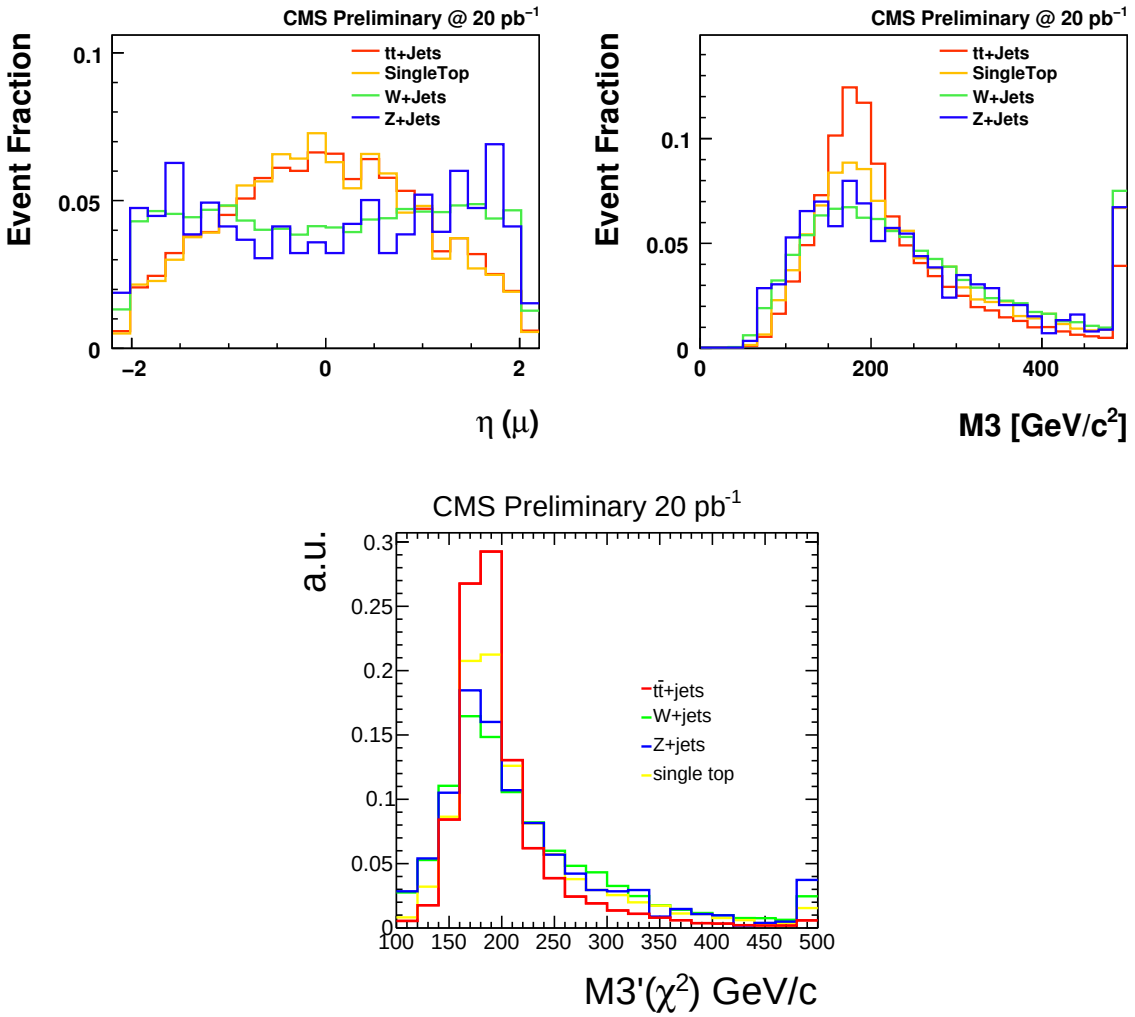


Figure 9: Comparison of template shapes for signal and backgrounds for the $\eta(\mu)$, M3 and M3' variables.

signal and background. Three variables are considered for this exercise, the pseudorapidity of the lepton $\eta(\mu)$, M3 and M3'. A comparison of the signal and background shapes for these variables can be found in Figure 9.

Since the shapes of $W$+jets and $Z$+jets are almost indistinguishable and the amount of expected QCD events is very small, only three fit templates are used, namely $t\bar{t}$+jets, single top and $W$+jets. Due to the fit of several background processes ($W$+jets, $Z$+jets and QCD) with one single template, only the sum of these contributions can be extracted, in addition to the number of $t\bar{t}$+jets and single top events. A standard binned likelihood fit procedure is employed. A Gaussian constraint is used for the amount of single top events, where a standard deviation of 40% is assumed, estimated from the uncertainty of the single top NLO cross section [20, 21] and the statistical error of the used MC sample.

When collision data are available, the observed distributions will be validated with a mixture of Monte Carlo simulations for $t\bar{t}$ signal and backgrounds (QCD will be estimated from data). In particular, control samples will be studied which are dominated by the dominant W+jets background, e.g. at lower jet multiplicity, in order to validate the simulation. The modelling of the shapes of the signal and background distributions is also included in the systematic error

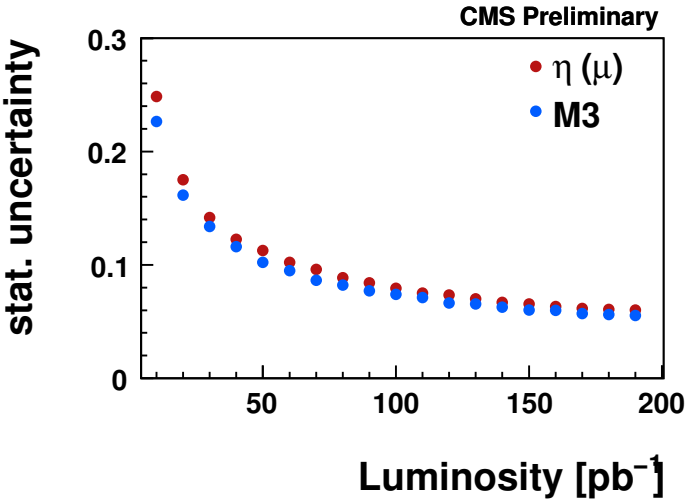


Figure 10: Statistical error of the template fits to eta and M3 as a function of luminosity

of the measurement.

For sanity checks of the method, the calculation of its sensitivity and also for the estimation of its systematical uncertainty, ensemble tests are employed. An ensemble test consists of a set of simulated pseudo experiments, where the number of events of a particular process are drawn from a Poisson distribution with a mean value which corresponds to the number of expected events for this particular process. The uncertainty due to the limited number of simulated events is incorporated by fluctuating the number of entries in each bin with a Gaussian distribution. Events are randomly drawn from the template distributions of the physical processes and filled into the pseudo data distribution which is subject to the fit. For each pseudo experiment the likelihood function is maximized and the number of $t\bar{t}$+jets events is extracted.

In order to estimate the statistical uncertainty of the method an ensemble test is applied with 5000 pseudo experiments. The pseudo data are drawn from templates for each process present in the final selection, i.e. from $t\bar{t}$, single top, $W$+jets, $Z$+jets and QCD. From these ensemble tests, the statistical uncertainty for an integrated luminosity of 20 pb$^{-1}$ is estimated as 16% for the M3 fit and 18% for the $\eta(\mu)$ fit. The sensitivity of the method is also estimated for different integrated luminosities. Figure 10 shows the dependence of the statistical error on the integrated luminosity. It is observed that the fit to the M3 distribution performs slightly better compared with the one to $\eta(\mu)$. It is estimated that the cross section can be estimated with a statistical uncertainty of 10% with a data sample corresponding to about 50-60 pb$^{-1}$.

The method described above is similarly applied to the M3' distribution. 1000 pseudo experiments are performed in order to evaluate the performance. For an integrated luminosity of 20 pb$^{-1}$, the statistical uncertainty on the number of $t\bar{t}$ events is estimated as 12%.

## 7 Systematic Uncertainties

The sources of systematic uncertainty which have been considered are listed below. The effects on the shapes of the templates used, as well as on the signal normalization have been evaluated.

- It is assumed that for the early data, the jet energy scale (JES) will only be known

Table 5: Summary of statistical and systematic uncertainties in the template fits to the $\eta(\mu)$, M3 and M3′ variables.

| Source | Uncertainty [%] | | |
|---|---|---|---|
| | Fit to $\eta(\mu)$ | Fit to M3 | Fit to M3′ |
| Statistical Uncertainty (20 pb$^{-1}$) | 17.7 | 16.3 | 11.5 |
| Jet Energy Scale | 16.7 | 15.1 | 19 |
| $t\bar{t}$ MC Generator | 1.9 | 14.9 | 14 |
| $t\bar{t}$ ISR/FSR | 3.3 | 7.7 | 2 |
| W+jets Factorization scale | 4.4 | 4.7 | 4 |
| W+jets Matching threshold | 5.5 | 2.8 | 4 |
| Single Top Shape | 0.1 | 0.8 | 1 |
| PDF Uncertainty | 5.0 | 5.0 | 5.0 |
| Total Systematic Error | 19.2 | 23.8 | 25.0 |
| Luminosity Error | 10.0 | 10.0 | 10.0 |

to about 10%. The resulting cross section uncertainty is estimated by scaling all jet momenta by $\pm 10\%$, and subsequently correcting the MET accordingly.

- The uncertainty in the modeling of the $t\bar{t}$ signal is estimated by using PYTHIA instead of MadGraph, and by using PYTHIA samples with modified amounts of initial and final state radiation.
- The uncertainty in the modeling of the dominant W+jets background is estimated by using MadGraph samples with varied factorization scale or varied thresholds for the matching of matrix element and parton shower contributions.
- The theoretical uncertainty in the normalization of the single top contribution is estimated as 30% according to the NLO calculations [20, 21]. Together with the statistical uncertainty due to the size of the Monte Carlo sample, the single top normalization is varied by about 40% in the fit using a Gaussian constraint. The resulting uncertainty is already included in the fit error. In addition, the shape of the single top template used has been varied by changing the relative weights of the individual t- and tW-channel contributions used in building the template.
- The uncertainty on the cross section measurement arising from the imperfect knowledge of the parton density functions (PDF) is estimated as 5%, using the CTEQ6.6 [22] PDF set and the LHAPDF [23] package using a re-weighting procedure also described in [5].
- The luminosity is assumed to be known to 10%.

The resulting systematic uncertainties are summarized in Table 5. The systematic error is dominated by the jet energy scale error and by the uncertainty on the shape of the $t\bar{t}$ template and on the signal efficiency determined from using PYTHIA instead of MadGraph. It is observed that the uncertainty on the shape of the $t\bar{t}$ signal template is much reduced when fitting $\eta(\mu)$, compared with using the M3 variable.

# 8 Background Determination using Charge Asymmetry

In this section, a novel alternative method to estimate the background contribution from $W$+jets events, as well as all other events with charge asymmetries, is presented. The method is based on the charge asymmetry of these processes in proton-proton collisions, while the signal (semileptonic $t\bar{t}$ decays) are charge-symmetric by nature. The cross sections and geometrical accep-

tances are different for $W^+$ and $W^-$ events in proton-proton collisions. The number of selected events that contain a selected lepton candidate (from $W^-$) production is then different from the number of selected events that contain an anti-lepton candidate (from $W^+$). This property can be used to estimate the $W$+jets background as well as all other kinds of events that have charge asymmetries after a $t\bar{t}$ lepton+jets selection as described above. For convenience, these processes are called "Events leading to Charge Asymmetry" (ECA). In addition to $W$+jets, $Vbb$ (mainly $Wb\bar{b}$), single top $s$ and $t$ channels and $WZ$ events are considered in this study.

The total number of ECA events, which is assumed to be dominated by $W$+jets events, can be estimated by measuring the difference between the numbers of events selected with an anti-lepton $N_+$ and with a lepton $N_-$ in data according to

$$(N_+ + N_-)_{data} \;=\; R_{\pm}(W) \times (N_+ - N_-)_{data}\,, \tag{3}$$

where $(N_+ - N_-)_{data}$ is measured in data. The factor $R_{\pm}$(W), which corresponds to the inverse of the $W$ charge asymmetry $A$, is defined as:

$$R_{\pm}(W) \;=\; \frac{N_{W^+} + N_{W^-}}{N_{W^+} - N_{W^-}} \;=\; \frac{A_+\sigma_{W^+} + A_-\sigma_{W^-}}{A_+\sigma_{W+} - A_-\sigma_{W-}}\,, \tag{4}$$

assuming that $N_{-(+)}$ can be expanded as $A_{-(+)}\varepsilon_{-(+)}L\sigma_{-(+)}$ and that $\varepsilon_+ \;=\; \varepsilon_-$ (where $\varepsilon_{-(+)}$ is the global selection efficiency, $\sigma_{-(+)}$ is the $W^{-(+)}$ cross section, $A_{-(+)}$ the geometrical acceptance and $L$ the integrated luminosity).

$R_{\pm}(W)$ is assumed to be independent of the reconstruction and selection efficiency (except from the acceptance component). In the approach described here, $R_{\pm}$ is estimated using simulated events. This method should be robust since possible errors in the modelling of the detector will tend to cancel in the determination of the charge asymmetry. However, methods to measure $R_{\pm}$ from an independent set of event data are also being considered. The asymmetry method suffers from a large statistical uncertainty related to the term $(N_+ - N_-)_{data}$ in Equation 3. For this reason it has been demonstrated using an event sample corresponding to an integrated luminosity of 100 pb$^{-1}$, which is larger than the sample used elsewhere in this note.

## 8.1 Systematic Uncertainties

The $R_{\pm}$ including all ECA events is assumed to be the same as for $W$+jets events only. A systematic uncertainty is associated with the relative difference of $R_{\pm}$ between the $W$+jets events and other ECA sources as predicted by the simulation. In addition, a 50% uncertainty on the relative contribution of each individual source of ECA, excluding the $W$+jets events, has been taken into account.

The jet energy scale as well as the resolution are each varied by $\pm$10%. The PDF uncertainty is estimated as discussed in Section 7. The uncertainty coming from the mis-identification of the muon charge is assumed to be small and is neglected in this study.

For high jet multiplicities ($\geq$4), the total uncertainty is statistically dominated (approximately 25% for $\mathcal{L}$ =100 pb$^{-1}$). The main sources of systematical uncertainty are the PDF uncertainty (approximately 8%) and the ECA uncertainty (about 6%), the total systematic uncertainty being around 11%. It should be noted, that, even though the systematic uncertainty related to the modeling of initial and final-state radiation in $W$+jets events is expected to be small, a complete analysis of this method should also consider these effects.

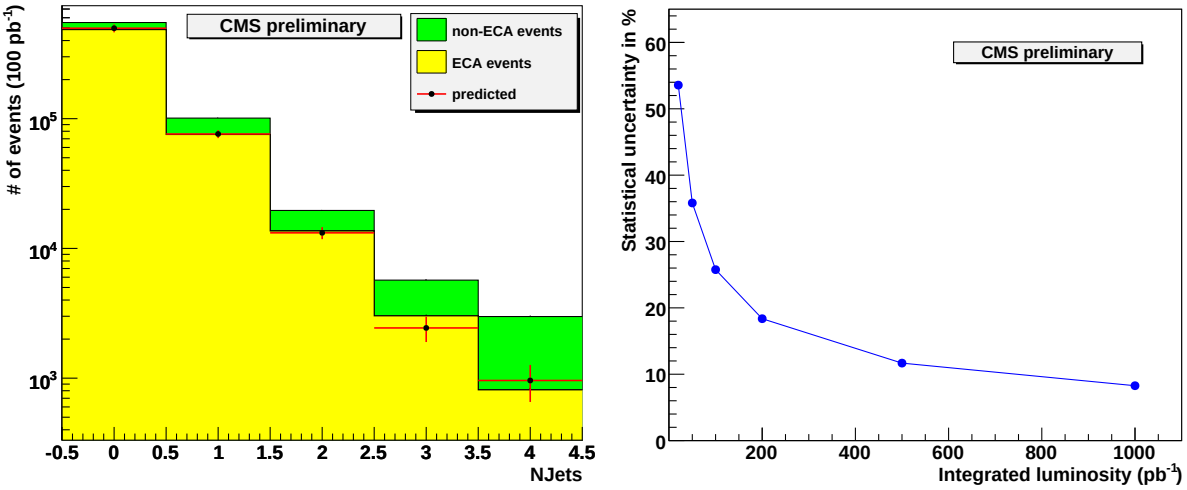


Figure 11: Left: Number of predicted (black dots), true ECA and non-ECA events as a function of the jet multiplicity for an integrated luminosity of 100 pb$^{-1}$. The errors bars correspond to statistical and systematical uncertainties summed in quadrature. The last bin also contains the overflow, i.e. $N(\text{jet}) \geq 4$. Right: Statistic uncertainty on the estimate of the ECA contribution as a function of the integrated luminosity.

### 8.2 Results and expected precision

The method is tested on a set of simulated events after the selection described above, but corresponding to an integrated luminosity of 100 pb$^{-1}$. All the sub-samples are used as pseudo-data in the closure test. The quantity $(N_+ - N_-)_{data}$ is then estimated in the pseudo-data by computing the difference between the number of events containing a selected anti-muon ($N_+$) and the number containing a selected muon ($N_-$).

A statistically independent sample of $W$+jets events is used to calculate $R_\pm$. We find $R_\pm = 4.34 \pm 0.14$ (MC stat.) $\pm$ 0.47 (syst.) for $N_{jet} \geq 4$. The number of ECA events is then determined using Equation 3 and presented in Figure 11 (left), together with the Monte Carlo truth expectation. For $N_{jet} \geq 4$ and 100 pb$^{-1}$, the total uncertainty is expected to be $\approx 30\,\%$.

The dependence on the number of data events is shown in Figure 11 (right), where the statistical uncertainty is presented as a function of the integrated luminosity. The systematical uncertainty may be reduced by measuring $R_\pm$ directly from data using an independent data sample (e.g. at low jet multiplicities).

## 9 Summary and Conclusions

A feasibility study for a first $t\bar{t}$ pair production cross section measurement in the muon-plus-jets channel using 20 pb$^{-1}$ of data at a center of mass energy of 10 TeV has been presented. Events are selected which contain exactly one isolated, high transverse momentum muon and at least four jets. After all selection cuts, about 320 $t\bar{t}$ events are selected, compared with 171 background events. Several techniques are employed which try to identify the three jets originating from the hadronic top decay.

An estimation of the QCD multi-jet contribution in the final selection was performed in a data-driven way using two complementary approaches: an extrapolation method using the muon isolation distribution, and the ABCD method. Both methods yield an uncertainty of about 50%.

For the measurement of the cross section a binned likelihood fit to various distributions is employed. The statistical and systematic uncertainties are evaluated using ensemble tests. With a data set corresponding to an integrated luminosity of 20 $\text{pb}^{-1}$, it is estimated that the $t\bar{t}$ cross section can be measured with $12-18\%$ statistical and around $20-25\%$ systematic error, dominated by the jet energy scale uncertainty.

A novel method for estimating the background from $W$+jets events (as well as single-top $t$ and $s$ channels and WZ) using charge asymmetry has been demonstrated. It has been shown to yield an uncertainty on the number of events with charge asymmetry of around 30% for 100 $\text{pb}^{-1}$ and has the potential to provide an independent cross-check of the background estimation.

In conclusion, it has been demonstrated that the $t\bar{t}$ cross section can be measured already with 20 $\text{pb}^{-1}$ of good LHC data at 10 TeV. The event selection does not rely on tools such as missing transverse energy measurement or b-tagging, which might not yet be reliable in the early running phase. The systematic uncertainty will be dominated by the knowledge of the jet energy scale.

## References


[1] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi, "Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC," *JHEP* **09** (2008) 127, arXiv:0804.2800. doi:10.1088/1126-6708/2008/09/127.

[2] S. Lowette, J. D'Hondt, J. Heyninck, and P. Vanlaer, "Offline Calibration of b-Jet Identification Efficiencies," *CMS Note* **2006/013** (2006).

[3] J. D'Hondt, S. Lowette, J. Heyninck, and S. Kasselmann, "Light quark jet energy scale calibration using the W mass constraint in single-leptonic $t\bar{t}$ events," *CMS Note* **2006/025** (2006).

[4] **CMS** Collaboration, "Measurement of jet energy scale corrections using top quark event," *CMS Physics Analysis Summary* **TOP-07-004** (2007).

[5] **CMS** Collaboration, "Physics Technical Design Report, Vol. 2: Physics Performance," *J. Phys G: Nucl. Part. Phys.* **34** (2007) 995–1579.

[6] J. D'Hondt et al., "Electron and muon reconstruction in single leptonic $\text{t}\bar{\text{t}}$ events," *CMS Note* **2006/024** (2006).

[7] J. D'Hondt, J. Heyninck, and S. Lowette, "Measurement of the cross section of single leptonic $t\bar{t}$ events," *CMS Note* **2006/064** (2006).

[8] **CMS** Collaboration, "Observability of Top Quark Pair Production in the Semileptonic Muon Channel with the first 10 $\text{pb}^{-1}$ of CMS Data," *CMS Physics Analysis Summary* **TOP-08-005** (2008).

[9] J. Alwall et al., "MadGraph/MadEvent v4: The New Web Generation," *JHEP* **09** (2007) 028, arXiv:0706.2334.

[10] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

[11] **CMS** Collaboration, "Prospects for the measurement of the single-top t-channel cross sectionin the muon channel with 200 pb$^{-1}$ of CMS data at 10 TeV," *CMS Physics Analysis Summary* **TOP-09-005** (2009).

[12] S. Frixione and M. L. Mangano, "How accurately can we measure the W cross section?," *JHEP* **05** (2004) 056, arXiv:hep-ph/0405130.

[13] **CMS** Collaboration, "Towards a measurement of the inclusive $W \to \mu\nu$ and $Z \to \mu^+\mu^-$ cross sections in pp collisions at $\sqrt{s} = 14$ TeV," *CMS Physics Analysis Summary* **EWK-07-002** (2008).

[14] **CMS** Collaboration, "Early $t\bar{t}$ Cross Section in the Electron plus Jets Channel at 10 TeV," *CMS Physics Analysis Summary* **TOP-09-004** (2009).

[15] G. P. Salam and G. Soyez, "A Practical Seedless Infrared-Safe Cone jet algorithm," *JHEP* **05** (2007) 086.

[16] **CMS** Collaboration, "Performance of Jet Algorithms in CMS," *CMS Physis Analysis Summary* **JME-07-003** (2008).

[17] **CMS** Collaboration, "Plans for Jet Energy Corrections at CMS," *CMS Physics Analysis Summary* **JME-07-002** (2008).

[18] **CMS** Collaboration, "Impact of Tracker Misalignment on the CMS b-Tagging Performance," *CMS Physics Analysis Summary* **BTV-07-003** (2008).

[19] K. S. Kölbig and B. Schorr, "A program package for the Landau distribution," *Computer Physics Communications* **31** (1984), no. 1, 97 – 111.

[20] B. W. Harris, E. Laenen, L. Phaf, Z. Sullivan, and S. Weinzierl, "The Fully differential single top quark cross-section in next to leading order QCD," *Phys. Rev.* **D66** (2002) 054024, arXiv:hep-ph/0207055. doi:10.1103/PhysRevD.66.054024.

[21] T. M. P. Tait, "The $tW^-$ mode of single top production," *Phys. Rev.* **D61** (2000) 034001, arXiv:hep-ph/9909352. doi:10.1103/PhysRevD.61.034001.

[22] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables," *Phys. Rev.* **D78** (2008) 013004, arXiv:0802.0007. doi:10.1103/PhysRevD.78.013004.

[23] M. R. Whalley, D. Bourilkov, and R. C. Group, "The Les Houches Accord PDFs (LHAPDF) and Lhaglue," arXiv:hep-ph/0508110.

# A Additional Distributions for QCD Background Estimate

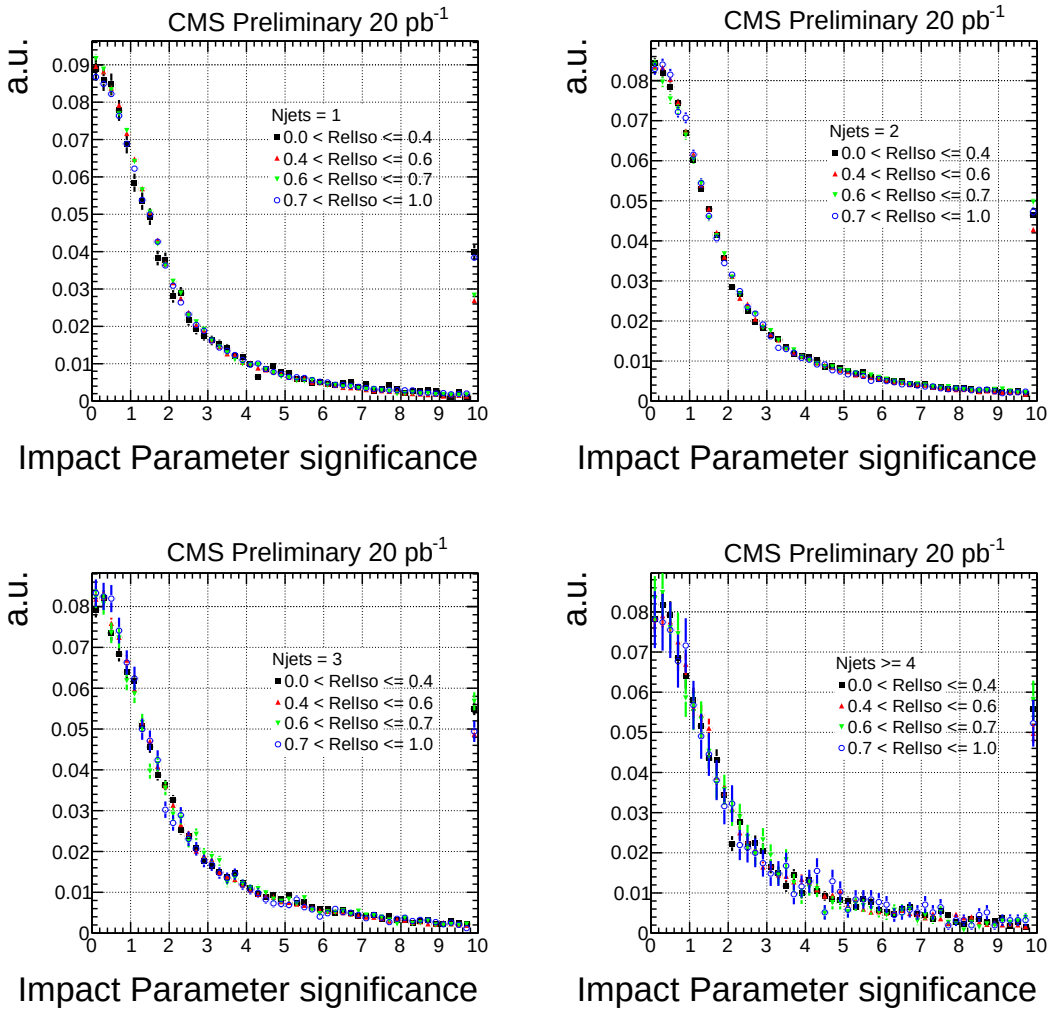


Figure 12: Distributions of the impact parameter significance for different jet multiplicities. For each plot, the impact parameter significance is shown for several intervals of combined relative isolation $RelIso'$. The distributions are normalized to unit area. The last interval contains the overflow events. The shape of the distributions is invariant for the different $RelIso'$ intervals and jet multiplicities, which indicates that the sig$(d_0)$ and combined relative isolation $RelIso'$ are only weakly correlated.