Available on the CERN CDS information server **CMS PAS TOP-08-005**

CMS Physics Analysis Summary

Contact: cms-pag-conveners-top@cern.ch 2009/08/07

Observability of Top Quark Pair Production in the Semileptonic Muon Channel with the first 10 pb^{-1} of CMS Data

The CMS Collaboration

Abstract

The feasibility for rediscovery of the top quark with the CMS detector, using the first data delivered by the Large Hadron Collider LHC is investigated. Employing the semi-leptonic decay channel of $t\bar{t}$ pairs into jets, missing transverse energy and a high transverse momentum muon, a simple and robust event selection is developed. Realistic detector conditions corresponding to the early data taking phase are included in the simulation. It is shown that already with 10 pb^{-1} of integrated luminosity, a $t\bar{t}$ signal can be established.

1 Introduction

At the Large Hadron Collider (LHC) top quark pairs ($t\bar{t}$) will be produced copiously. The next-to-leading order cross section at LHC energies has been estimated to be 833 pb [1], and the dominant production mechanism is gluon-gluon fusion. The large cross section implies that a top quark signal could be established already at fairly low integrated luminosities.

Top 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 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 add credibility to possible new physics signals.

Once the $t\bar{t}$ signal has been established, there will be a rich top quark physics program. Moreover, $t\bar{t}$ production constitutes an important source of background in searches for new physics beyond the Standard Model, such as Supersymmetry. Top quark events are also very useful to understand the performance of the detector, such as the b-tagging efficiency [2] or the absolute jet energy scale using the known mass of the $W^{\pm}$ boson [3]. The CMS potential 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 [4–6].

The scope of this analysis is to address the potential of the CMS detector to establish a top signal with the first 10 pb^{-1} of LHC data. The semi-leptonic decay channel into a muon plus jets and missing transverse energy is considered. The goal is to design a simple and robust analysis able to identify top quark pairs with the lowest possible integrated luminosity. The reconstruction of the objects in the final state used for the event selection should be based on simple observations in separate subdetectors, rather than employing complicated analysis techniques. Trigger as well as miscalibration and misalignment conditions in the early data taking phase must be taken into account.

2 Simulation

The simulation of $t\bar{t}$ signal events was performed using ALPGEN [7]. Top quark pairs are accompanied by up to four additional hard jets. The hard parton configurations generated by ALPGEN are matched with parton showers using PYTHIA [8] using the MLM matching prescription.

The production of W bosons in association with extra jets, where the W decays leptonically, has a similar signature and thus constitutes the main background to semileptonic $t\bar{t}$ events. Another source of background is Z boson production plus extra jets, where the Z decays into two leptons, one of which is not reconstructed. Both W/Z+jets are simulated using ALPGEN, which includes the production of W/Z bosons in association with up to five additional light jets. The ALPGEN cross sections are scaled to the NLO values using K-factors of 1.85 ($t\bar{t}$) and 1.12 (W/Z+jets) [9].

Another important source of background to be considered arises from QCD events with several jets and a lepton which passes the selection cuts, either a real lepton e.g. from semileptonic decays of hadrons containing a b or c quark, or fake muons. This background is very difficult to model using Monte Carlo because the huge cross section implies that a very large number of events need to be simulated. Furthermore, it is very sensitive to details in the tails of the simulated distributions. 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 QCD sample is used where the events are pre-filtered at the generator level with the requirement of a muon with $p_T > 15$ GeV. As the main isolated muon background originates from semileptonic decays of b/c-hadrons (around 70%), rather than fake muons and in-flight decays (around 30%), this sample should give an idea about the normalization as well as the shapes of the QCD background. It corresponds to an integrated luminosity of 8.7 pb^{-1}, which is adequate for this study.

All samples have been simulated and reconstructed within the scope of the 2007 *CMS Computing, Software and Analysis Challenge* CSA07. In order to take realistic detector conditions corresponding to the early data taking phase into account, the events were reconstructed applying miscalibration of the ECAL and HCAL calorimeters, as well as misalignment of the silicon tracker and muon detectors which is deemed to be realistic for the first 10 pb^{-1} of data taken by CMS.

During CSA07 the data- and workflow was set up to mimic the procedures during real data taking. Therefore, the generated events were mixed to a cocktail of events, which was reconstructed at the Tier-0. The High Level Trigger (HLT) response was simulated using a trigger table defined in [10], corresponding to a luminosity of $\mathcal{L} = 10^{32}\ \text{cm}^{-2}\text{s}^{-1}$. According to the HLT decision the events were split in *primary datasets* which were transferred to Tier-1 centres. Subsequently *skimmed datasets* were produced and transferred to Tier-2 centres, where the data were analyzed.

This analysis uses a skim which is defined by requesting that the event is accepted by the non-isolated single muon trigger. In addition, a loose preselection is performed at the offline level by requesting a reconstructed muon with $p_T > 20$ GeV and at least one jet with uncalibrated $E_T > 30$ GeV. The trigger efficiency is determined as 91% for muons with $|\eta| < 2.1$ and $p_T > 30$ GeV. In real data, the trigger efficiency must be determined from the data themselves using for example a sample triggered by an independent monitor trigger.

3 Event Reconstruction and Selection

Reconstructed muon candidates consist of muon track segments in the muon chambers, matched with a track reconstructed in the silicon tracker [11]. The momentum scale as well as the reconstruction efficiency will be determined in $Z \rightarrow \mu\mu$ events using the *tag and probe* method.

Isolation requirements are placed on the muon candidates, in order to distinguish leptons from W decays from leptons 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 lepton direction, excluding the lepton 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 lepton direction, excluding the lepton energy deposited in the calorimeter. The following isolation cuts are applied: $E_{iso}^{calo} < 5$ GeV and $p_{T,iso}^{tracker} < 3$ GeV. These cuts have been optimized in order to retain most of the leptons from W decays whilst rejecting a large fraction of leptons produced in jets.

Jets are reconstructed using the iterative cone jet algorithm with a cone radius of $R = 0.5$, using calorimeter towers as input [12]. By default, a response correction is applied to the reconstructed jet transverse energies, depending on jet E_T and position in the detector, which is derived by comparing with particle level jets. Similarly, the missing transverse energy ($\not\!\!E_{\text{T}}$) in the event is reconstructed using calorimeter towers [13]. $\not\!\!E_{\text{T}}$ is corrected using the corrections applied to the calorimetric energy contained in jets. The momenta of reconstructed muons are also taken into account in the $\not{E}_T$ reconstruction.

We consider the process $pp \to t\bar{t} + X \to bq\bar{q}b\mu\nu_\mu + X$, where each of the two top quarks decays into a W boson and a b-quark. Subsequently, one of the W bosons decays hadronically into a $q\bar{q}$ pair, whereas the other one decays into a muon and a neutrino. Thus, the experimental signature consists of one high transverse momentum muon, at least four jets and missing transverse energy.

Exactly one reconstructed isolated muon with $p_T > 30$ GeV and pseudorapidity $|\eta| < 2.1$ is requested, where the p_T cut is motivated in order to reduce the amount of non-W (i.e. QCD) background, and the η requirement is due to the acceptance of the muon trigger. Events with more than one high p_T 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 events.

At least four reconstructed jets in the range $|\eta| < 2.4$ are required. Due to higher order diagrams or parton showers, the four jets from the $t\bar{t}$ decay are often accompanied by additional jets. The η range corresponds to the acceptance of the silicon tracker, in order to allow the possible use of b-tagging. However, this analysis does not distinguish b-jets and non-b-jets. The transverse energy of the highest E_T jet must be $E_T > 65$ GeV, in order to be safe with respect to the skimming requirement of $E_T > 30$ GeV for uncalibrated jets. The transverse energies of the other three jets must satisfy $E_T > 40$ GeV.

The selection described above is called in the following *loose selection*. For 10 pb^{-1} of integrated luminosity, 163 $t\bar{t}$ signal events are selected, as well as 32 $t\bar{t}$ events from other decay channels, 57 W+jets and 8 Z+jets events. The number of QCD events is estimated as 110 from the muon enriched sample, where at the generator level the presence of a muon with $p_T > 15$ GeV is required. As mentioned in section 2 decays-in-flight or fake muons are not included in the QCD simulation.

As can be seen in Table 1 and Figures 1 and 2, the ratio of $t\bar{t}$ signal to QCD background events for this selection is about 1.5. Owing to the difficulty of estimating this background from data additional selection cuts are required in order to reduce the QCD background as much as possible. Several scenarios were studied, such as cuts on $\not{E}_T$ or H_T (the scalar sum of the transverse momenta of all jets and the muon), or tighter requirements on lepton and jet transverse momenta. While some improvement can be gained on the signal to QCD ratio, it turned out that a tight cut on the calorimeter isolation, together with the requirement that the closest jet is separated by at least 0.3 units in $r\phi$ from the muon improves the S/B considerably from to 11.5, while rejecting only 20% of the signal.

The final event selection is thus defined by requesting $E_{iso}^{calo} < 1$ GeV and $dR_{min} > 0.3$ (last line of Table 1). With these extra requirements, the number of $t\bar{t}$ signal events is still 128, while the QCD background is down to 11 events. The other backgrounds are 25 other $t\bar{t}$ decay channels, 45 W+jets and 7 Z+jets events. The overall selection efficiency (including acceptance) for semileptonic $t\bar{t}$ events is 10.3%.

It is presently not known how large the uncertainties on the shapes and normalizations of the various backgrounds are going to be for the early LHC data. Therefore no uncertainties on these backgrounds are quoted here. In real data, the shapes of the W/Z+jets backgrounds could be taken from simulation, whereas their normalizations would be fixed by comparing with a control data sample at low jet multiplicities.

On the other hand, the QCD background will most probably have to be determined from data

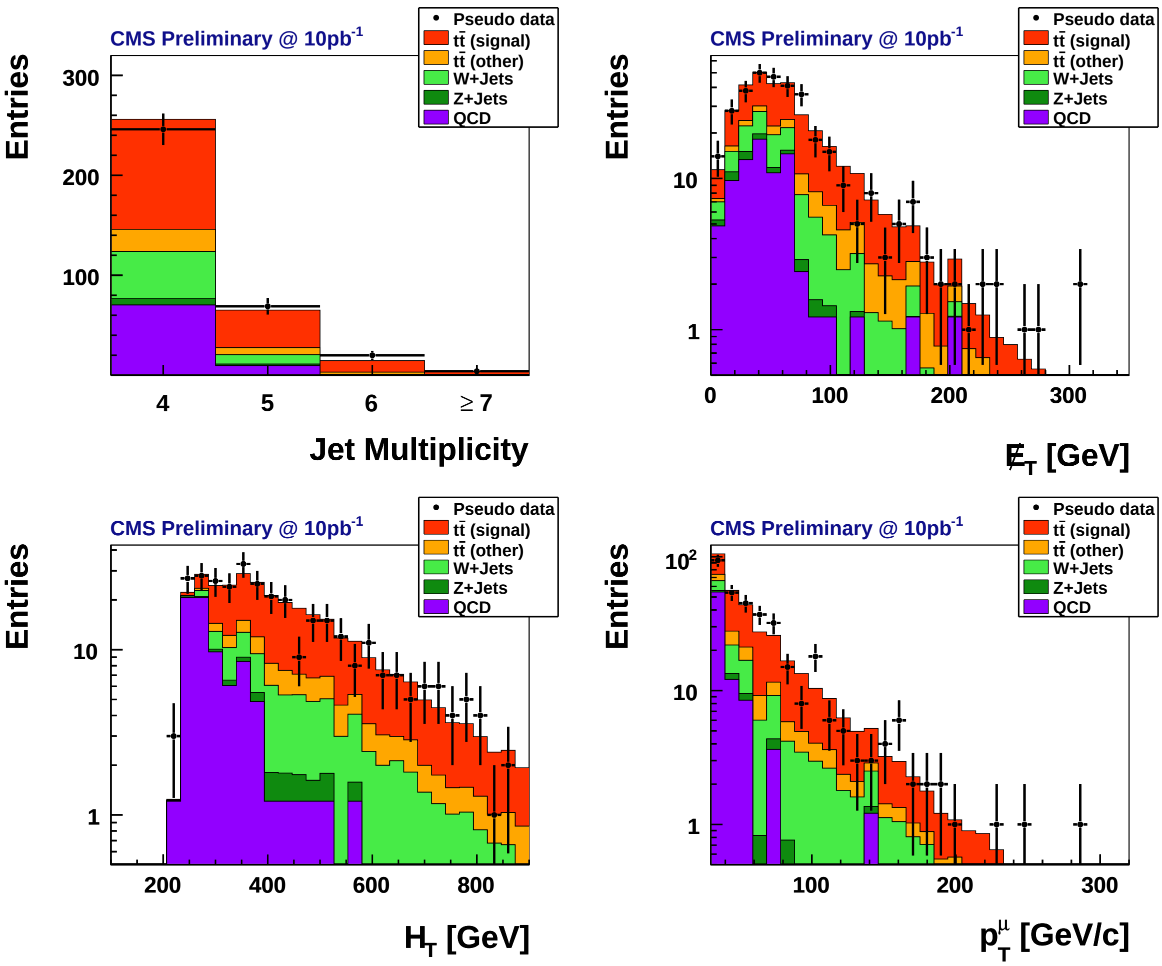

Figure 1: Jet multiplicity (top left), missing transverse Energy $\not\!\!E_{\mathrm{T}}$ (top right), transverse energy sum of all jets and the muon H_T (bottom left) and muon transverse momentum p_T^{μ} (bottom right), presented for the loose selection. Here and in the following figures, the event numbers are normalized to 10 pb^{-1} of integrated luminosity. The *pseudo data* correspond to a poissonian smeared distribution which is obtained from the summed MC distribution.

alone, as Monte Carlo modelling and simulation of details in the tails of detector response distributions are very difficult. Therefore, data-driven background determination will be essential. Several methods to estimate the QCD background from data could be applied:

- Fake rate method: The isolated lepton fake rate (both from mismeasurements as well as real non-isolated leptons) can be determined from a sample of QCD events by measuring the fraction (differentially in p_T, η etc.) of isolated leptons with respect to all (loose) lepton candidates. This fake rate can then be applied to a selected sample signal events with loose lepton cuts in order to determine the background fraction.
- Matrix method: This method is based on using two variables that characterize signal and QCD background and that are assumed to be uncorrelated for the QCD background, for example MET and lepton isolation. The number of QCD events in the signal region (large $\not\!\!E_{\mathrm{T}}$, isolated) can then in principle be determined by multiplying the efficiency for passing the $\not\!\!E_{\mathrm{T}}$ cut in the non-isolated sample with the number of isolated events below the $\not\!\!E_{\mathrm{T}}$ threshold.
- Template Method: The amount of QCD background can be in principle determined

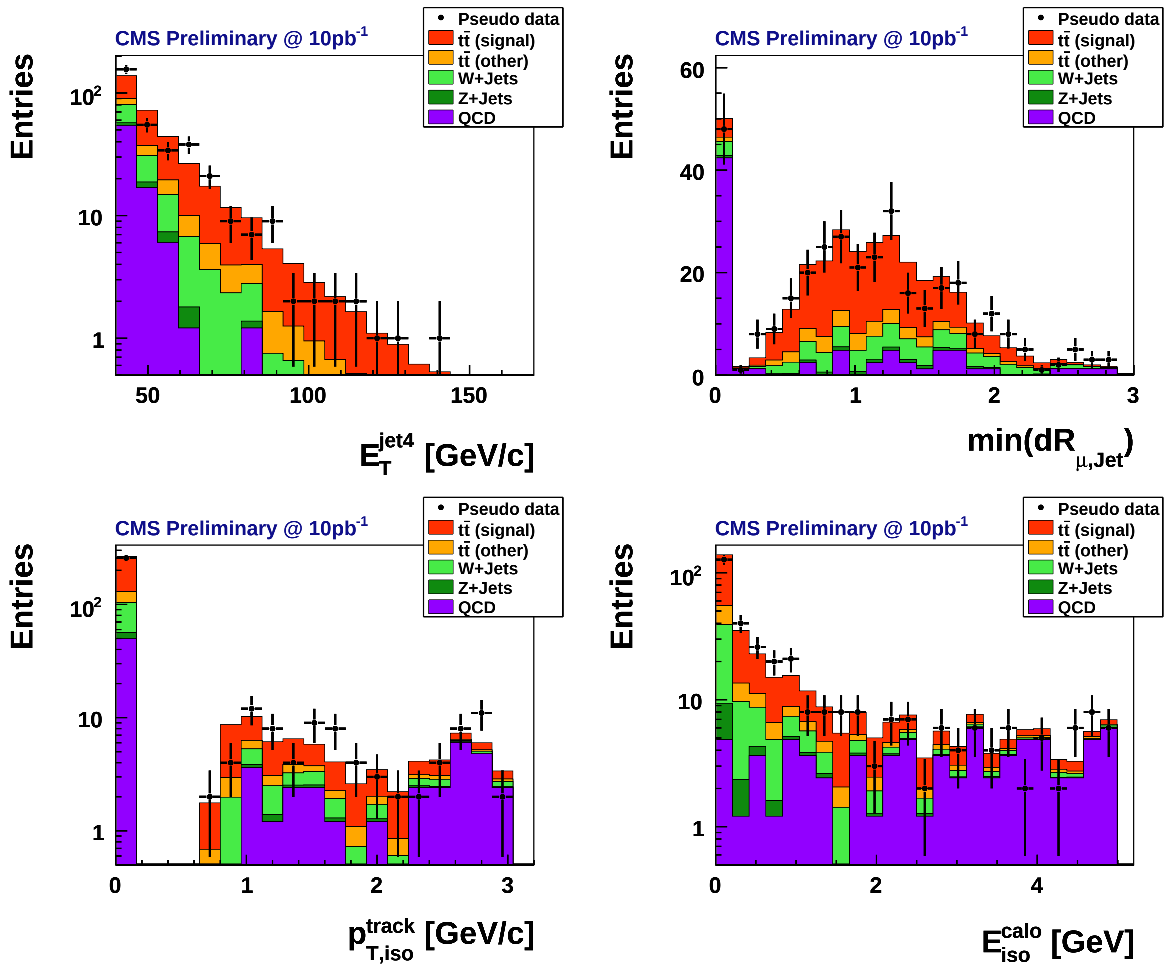

Figure 2: Transverse energy of the fourth highest E_T jet (top left), distance between muon and closest jet dR_{min} (top right), tracker isolation $p_{T,iso}^{tracker}$ (bottom left) and calorimeter isolation E_{iso}^{calo} (bottom right), presented for the loose selection.

by fitting a combination of templates for signal and background to a discriminating distribution such as $\not{E}_T$ or $M_{T,W}$ measured in data. For $t\bar{t}$ signal and W/Z+jets background templates can be obtained from Monte Carlo. For QCD, a sample of data events has to be selected which can be used as a template.

These methods are being studied in CMS at the moment.

4 Results

This section presents the results obtained with the final selection described in the previous section. All distributions and event numbers are rescaled to an integrated luminosity of 10 pb^{-1}.

The first distribution to look at is the jet multiplicity distribution for all events which pass the final selection, except for the cut on the number of reconstructed jets. It is presented in Figure 3 (left). As can be seen, $W + jets$ events dominate at low jet multiplicities, whereas the $t\bar{t}$ signal events typically contain a larger number of reconstructed jets. An important check of the normalization of the $W + jets$ background will come from a comparison of the selected number of events in data at low jet multiplicities with the MC prediction.

Table 1: Selected event numbers for 10 pb^{-1}. For the first three lines, the selections applied are cumulative. Line three corresponds to the *loose selection* as discussed in the text. For the remaining lines, each cut is applied on top of the loose selection. The last line corresponds to the final event selection. In the case of QCD, no numbers are quoted for the first two selection steps because the sample is filtered at the generator level. The two rightmost columns represent the ratio of signal to QCD background events S/B(QCD) as well as the overall signal to background.

	$t\bar{t}$ (signal)	$t\bar{t}$ (other)	W+jets	Z+jets	QCD	S/B(QCD)	S/B
Preselection	749	527	7474	1430	–	–	–
4 Jets $p_T > 65/40/40/40$ GeV	236	135	83	16	–	–	–
1 Muon $p_T > 30$ GeV	163	32	57	8	110	1.48	0.79
$\not{E}_T > 20$ GeV	151	31	53	7	91	1.66	0.83
$\not{E}_T > 30$ GeV	138	29	47	6	76	1.82	0.87
$\not{E}_T > 60$ GeV	87	23	28	2	29	3.04	1.07
$H_T > 300$ GeV	153	30	54	8	50	3.09	1.08
$H_T > 400$ GeV	104	22	39	6	14	7.27	1.27
$p_T^\mu > 40$ GeV	131	24	46	9	32	4.11	1.18
$p_{T,jet4} > 50$ GeV	94	19	27	4	20	4.76	1.35
$p_{T,iso}^{tracker} < 0.5$ GeV	134	26	47	7	61	2.22	0.95
$E_{iso}^{calo} < 3$ GeV	157	30	55	8	56	2.79	1.04
$E_{iso}^{calo} < 1$ GeV	131	25	47	7	17	7.91	1.37
$dR_{min} > 0.5$	152	30	52	8	44	3.44	1.14
$dR_{min} > 0.3$	159	31	54	8	48	3.28	1.12
$dR_{min} > 0.3$ & $E_{iso}^{calo} < 1$ GeV	128	25	45	7	11	11.62	1.47

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. 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 E_T are deemed to originate from the hadronic top decay, the invariant mass of these we denote $M3$. In around $30-40\%$ of the selected signal events, the three jets are correctly associated with the three quarks of the hadronic top decay.

Figure 3 (right) shows the distribution of the invariant mass $M3$ for the selected events. A clear peak can be observed around the nominal top mass value. The peak position is shifted from the expected value of 175 GeV to higher values for two reasons: Firstly, due to the miscalibration introduced in the hadron calorimeter, and secondly due to the fact that the jet energy scale corrections applied to the (quark induced) jets were determined from QCD dijet events which contain also gluon jets, for which the response is different. The width of the observed peak in the $M3$ distribution is due to the jet energy resolution, but also due to the fact that one or more of the selected jets may not originate from the hadronic top decay.

Figure 4 presents the transverse momentum p_T^μ and pseudorapidity η_μ of the muon. Figure 5 shows the transverse energy E_T^{jet} and pseudorapidity η^{jet} of the highest transverse momentum jet for the events passing the final selection. The transverse energies of the second, third and fourth jet are presented in Figure 6. Also shown there is the dijet invariant of those jets where the best match with the nominal value of M_W is obtained.

Figure 7 shows further distributions: the jet multiplicity, the transverse W mass, defined as $M_{T,W} = \sqrt{2p_T^\mu \not{E}_T - \vec{p}_T^\mu \cdot \vec{\not{E}}_T}$, the scalar sum of the transverse energies of all jets and the muon labelled H_T, and the missing transverse energy $\not\!\!E_T$.

5 Conclusions

A study of the potential to observe $t\bar{t}$ events with the CMS detector in the semileptonic muon channel with the first LHC data has been presented. A simple and robust event selection has been developed, requesting exactly one tightly isolated muon with $p_T > 30$ GeV and at least four jets with $E_T > 65$ GeV for the leading jet and $E_T > 40$ GeV for the other jets. For 10 pb^{-1} of integrated luminosity, 128 signal events are expected, corresponding to a selection efficency of 10.3%, together with 25 other $t\bar{t}$ final state events, as well as 45 (7) W+jets (Z+jets) events. The number of QCD background events is determined as 11, with a large uncertainty.

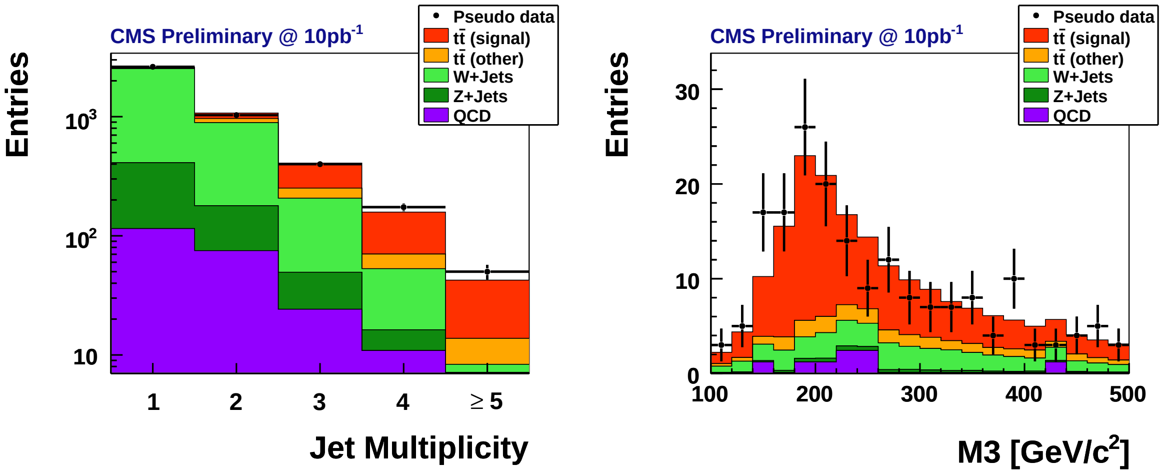

Figure 3: Jet multiplicity distribution for events passing the final selection except the requirement of $N_{jets} \geq 4$ (left); Invariant mass of the three jets with the highest vectorially summed E_T for the final selection (right).

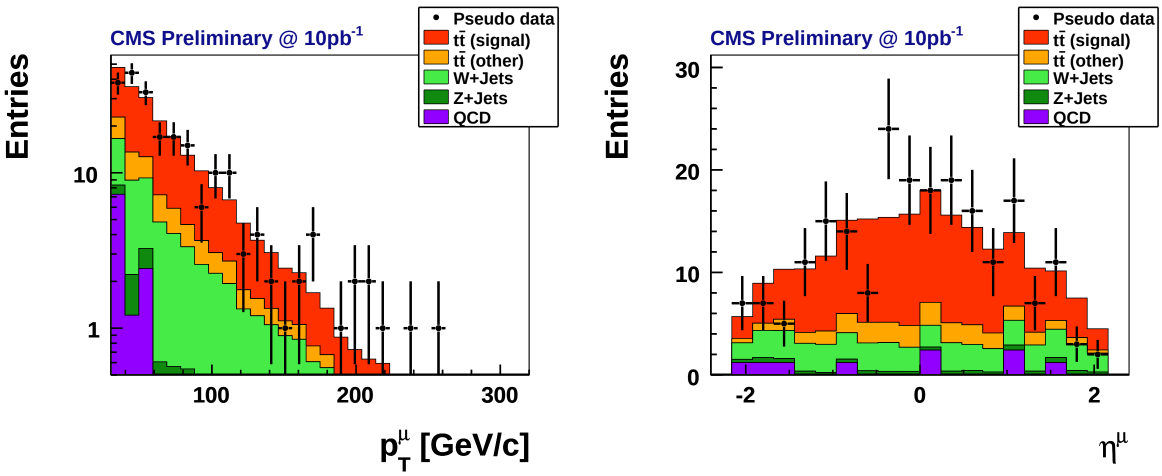

Figure 4: Transverse momentum p_T (left) and pseudorapidity η (right) of muons passing the final selection.

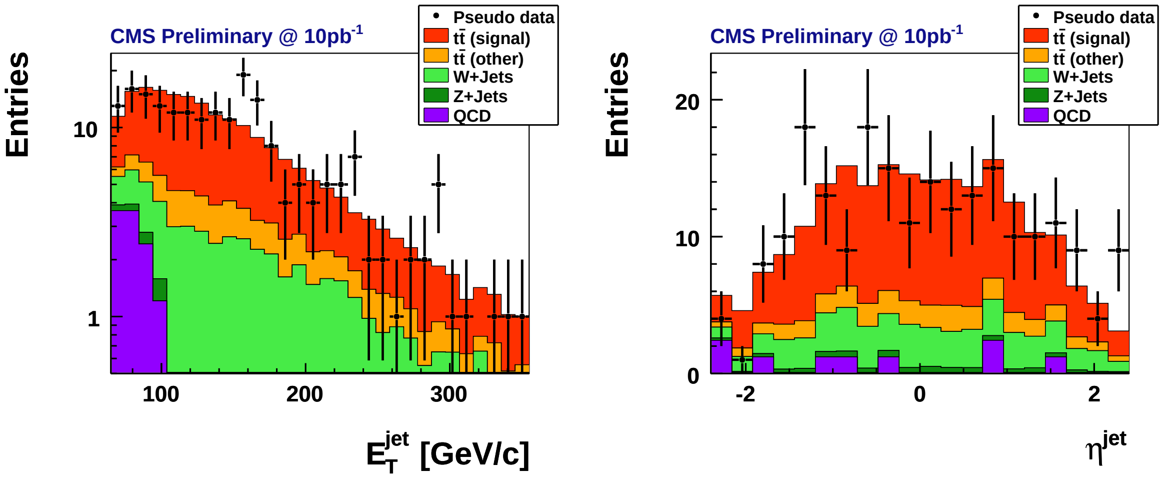

Figure 5: Transverse energy E_T (left) and pseudorapidity η (right) of the highest E_T jet for events passing the final selection.

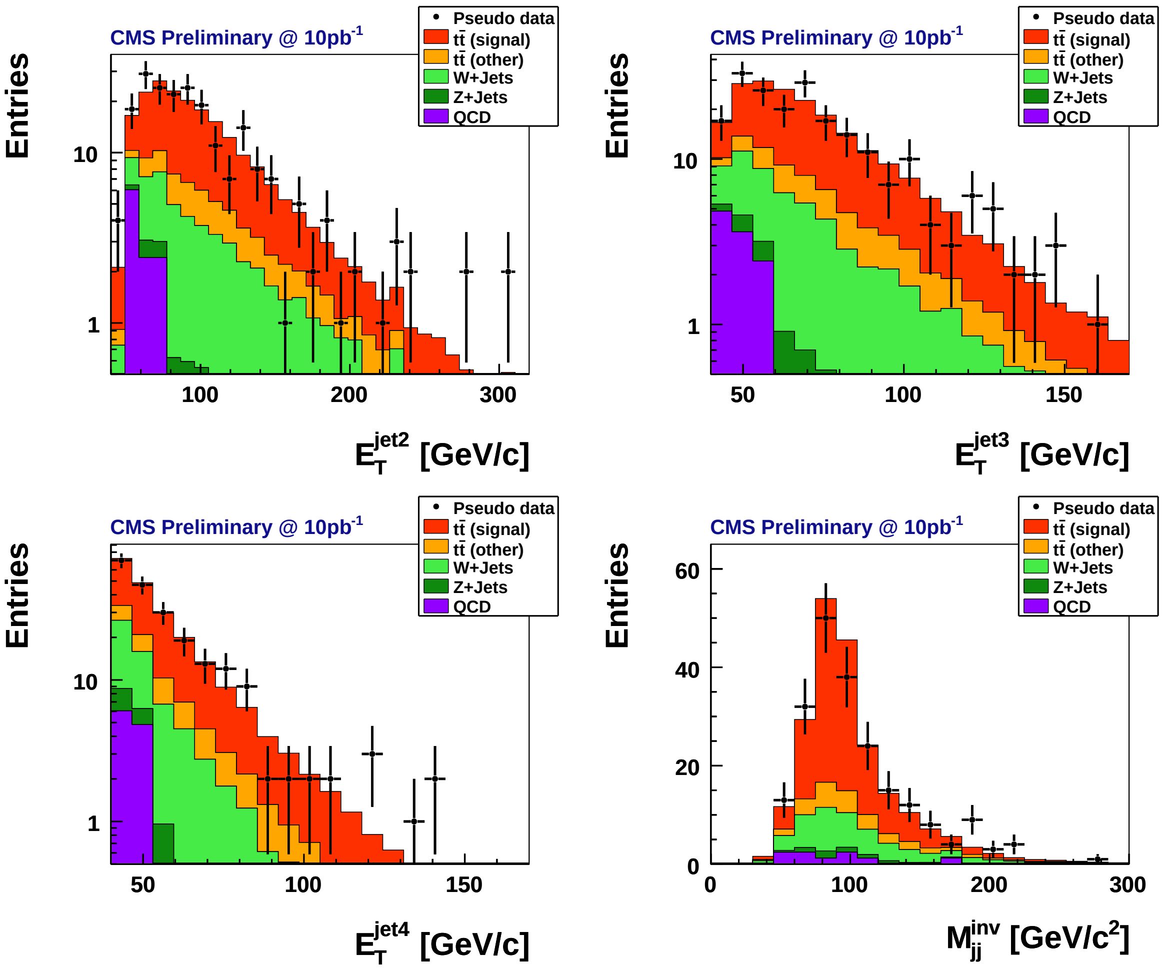

Figure 6: Transverse energies E_T of the 2nd (top left), 3rd (top right) and 4th (bottom left) jet, as well as the invariant mass of the two jets which best correspond to the nominal value of M_W (bottom right), shown for events passing the final selection.

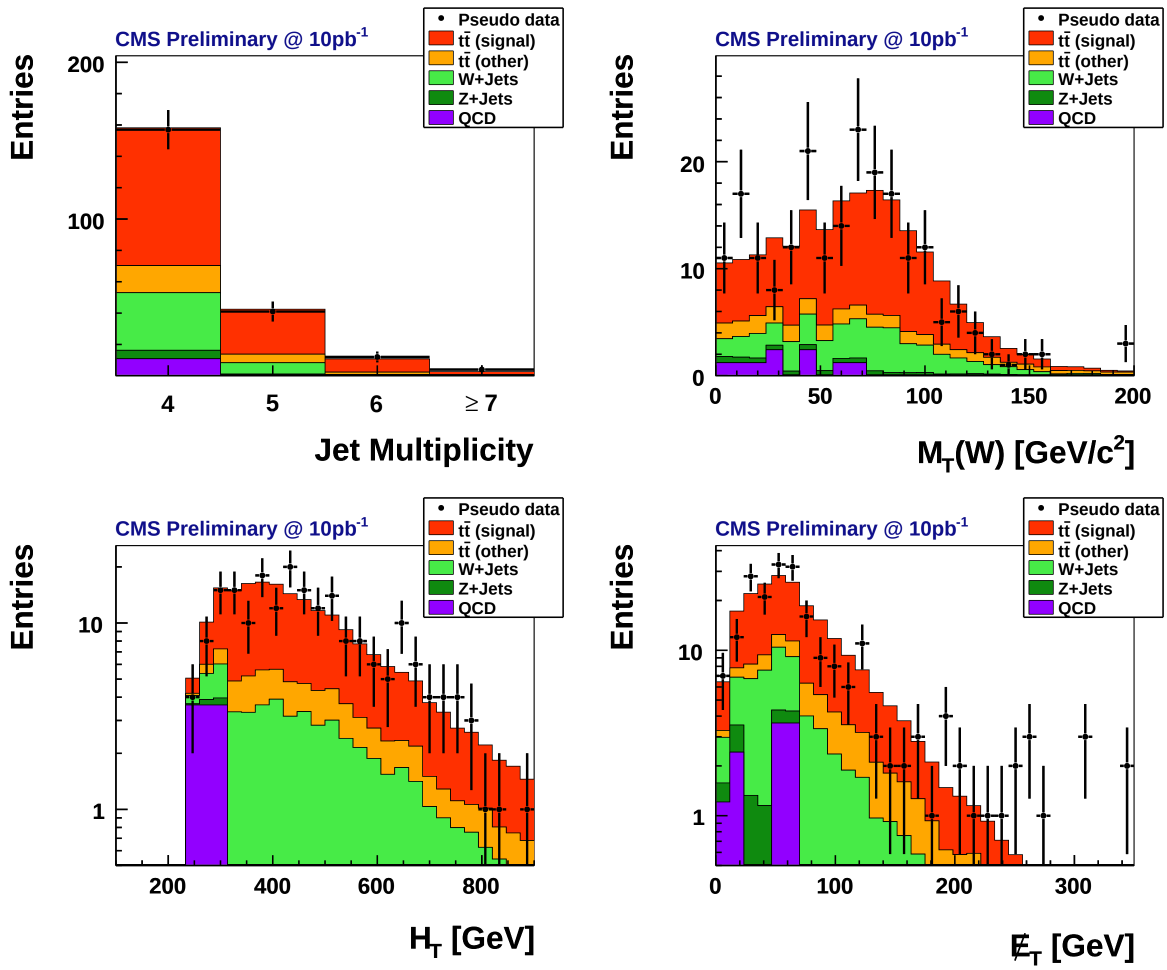

Figure 7: Jet multiplicity (top left), leptonic transverse mass $M_{T,W}$ (top right), transverse energy sum of all jets and the muon H_T (bottom left) and missing transverse energy $\not\!\!E_{\mathrm{T}}$ (bottom right) for events passing the final selection.

References

[1] R. Bonciani, S. Catani, M. L. Mangano, and P. Nason, "NLL resummation of the heavy-quark hadroproduction cross-section," *Nucl. Phys.* **B529** (1998) 424–450, arXiv:hep-ph/9801375.

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

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

[6] 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).

[7] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions," *JHEP* **07** (2003) 001, arXiv:hep-ph/0206293.

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

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

[10] **CMS** Collaboration, "CMS High Level Trigger," *CERN/LHCC* **2007-021** (2007).

[11] **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 PAS* **EWK-07-002** (2008).

[12] **CMS** Collaboration, "Performance of Jet Algorithms in CMS," *CMS PAS* **JME-07-003** (2008).

[13] **CMS** Collaboration, "$\not\!\!E_{\mathrm{T}}$ Performance in CMS," *CMS PAS* **JME-07-001** (2008).