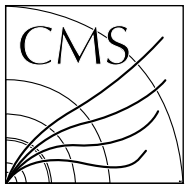
Available on CMS information server

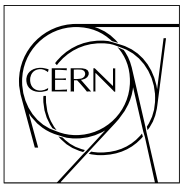
CMS AN -2009/050



The Compact Muon Solenoid Experiment

Analysis Note

The content of this note is intended for CMS internal use and distribution only



01 April 2009 (v6, 10 June 2009)

Expectations for observation of top quark pair production in the dilepton final state with the early CMS data

C. Campagnari, P. Kalavase, D. Kovalskyi, V. Krutelyov, J. Ribnik

University of California, Santa Barbara

W. Andrews, D. Evans, F. Golf, J. Mülmenstädt, S. Padhi, Y. Tu, F. Würthwein, A. Yagil

University of California, San Diego

L. Bauerdick, I. Bloch, K. Burkett, I. Fisk, O. Gutsche

Fermi National Accelerator Laboratory, Batavia, Illinois

Abstract

We present expectations for the observation of top anti-top quark production with two leptons (electrons or muons) in the final state in the early phase of CMS running. We define a possible event selection, and we survey the Standard Model sources of dileptons. Assuming $10\ \mathrm{pb}^{-1}$ of pp collision data collected at $\sqrt{s} = 10$ TeV, we find that a clear signal stands out when selecting two leptons with high transverse momentum in events with a high missing transverse energy and at least two jets.

CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

2009/06/09

Expectations for observation of top quark pair production in the dilepton final state with the early CMS data

Date: 2009/06/09 22:04:01 • RCSfile: AN.tex,v • Revision: 1.13 • Author: slava77

L. Bauerdick[1], I. Bloch[1], K. Burkett[1], I. Fisk[1], O. Gutsche[1], W. Andrews[2], D. Evans[2], F. Golf[2], J. Muelmenstaedt[2], S. Padhi[2], Y. Tu[2], F. Würthwein[2], A. Yagil[2], C. Campagnari[3], P. Kalavase[3], D. Kovalskyi[3], V. Krutelyov[3], and J. Ribnik[3]

[1] Fermilab
[2] University of California, San Diego
[3] University of California, Santa Barbara

Abstract

We present expectations for the observation of top anti-top quark production with two leptons (electrons or muons) in the final state in the early phase of CMS running. We define a possible event selection, and we survey the Standard Model sources of dileptons. Assuming 10 pb^{-1} of pp collision data collected at $\sqrt{s} = 10$ TeV, we find that a clear signal stands out when selecting two leptons with high transverse momentum in events with a high missing transverse energy and at least two jets.

1 What Is Missing From This Note

- In addition to the $N-1$ table add the cut flow table.
- Some text reasoning for the value of the MET cut

2 Introduction

Top quark pair production at the LHC is expected to be a major source of background to a number of physics searches beyond the standard model (SM). It is important to experimentally establish the $t\bar{t}$ production rate early on in order to test the QCD predictions. The dilepton $t\bar{t}$ final state, where both Ws from $t \to Wb$ and $\bar{t} \to W\bar{b}$ decay leptonically into e or μ, is the cleanest $t\bar{t}$ final state. In this note we outline a possible early analysis which aims to establish $t\bar{t} \to$ dileptons at CMS in pp collisions at $\sqrt{s} = 10$ TeV and to provide a first measurement of the $t\bar{t}$ cross-section.

The strategy for the analysis described in this note is that of a simple counting experiment: we define an event selection, we count the number of candidates, we compare this number to the number of expected candidates from all non-$t\bar{t}$ Standard Model (SM) sources, and we ascribe the excess of events to $t\bar{t}$. We discriminate the dileptonic $t\bar{t}$ final states based on their primary features: the charged leptons (e or μ) coming from the W decays are high-p_T, isolated and have opposite charges; the neutrinos from the W decays correspond to significant missing transverse energy ($\not\!\!E_T$); while the high-p_T b-quarks from the top quark decay can be identified as two hadronic jets in the event. In such a counting experiment, it is very important to be able to validate the estimation of the non-$t\bar{t}$ SM contributions. To do this we use control regions where the non-$t\bar{t}$ background is expected to be large and the $t\bar{t}$ signal is small. In practice this can be done most easily by dividing the event sample in jet multiplicity bins (N_{jets}). This is because $t\bar{t}$ events tend to have jets from the b-quarks in top decay, while most other backgrounds tend to have no jets in the final state. We will then test the background prediction in the $N_{\text{jets}} = 0$ and 1 bins, and extract the $t\bar{t}$ signal in the $N_{\text{jets}} \geq 2$ bins.

We also develop data-driven methods where we use control samples enriched in background events to estimate the $Z/\gamma^\star$ and fake lepton (leptons from QCD jets) contributions. In the $Z/\gamma^\star$ case, events in which the dilepton invariant mass is in the Z-mass region (which are excluded to supress the $Z/\gamma^\star$ contribution) are used to estimate the $Z/\gamma^\star$ contribution outside the Z-mass region. Almost all the events with fake leptons are coming from $W+$jets events. The $W+$jets events contribute to the dilepton signature by virtue of one of the jets misidentified as a lepton: this includes not only the pure $W+$jets, but also the case when the W is coming from one of the top quark decays. We account for these events by looking at the sample of events with leptons not passing the good isolated lepton selections and extrapolate from that region using a factor estimated from a sample with QCD jets.

Note that b-tagging is not used in this analysis. This is because the analysis is aimed at very early data, before the establishment of b-tags. Others in the top group are pursuing a similar analysis that exploits b-tags [1].

There are three dilepton final states that we consider: e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$. The e^+e^- and $\mu^+\mu^-$ final states are more difficult than the $e^\pm\mu^\mp$ final state because of the Drell-Yan (DY) background, $pp \to Z/\gamma^\star \to e^+e^-$ or $\mu^+\mu^-$. The ability to reject the large DY background depends crucially on $\not\!\!E_T$. The $\not\!\!E_T$performance of the detector at startup is very uncertain. Here we will carry out our analysis assuming that the current CMS Monte Carlo gives a good representation of the $\not\!\!E_T$. We are very aware that when real data arrives we may have to redesign the analysis

in the e^+e^- and $\mu^+\mu^-$ final states, and may be even forced to limit ourselves to $e^\pm\mu^\mp$. Once it becomes clear the behaviour of the CMS calorimeter in high-energy pp collisions is as expected from test-beam and the current simulation, we expect to have a substantial improvement from using the $\not\!\!E_T$ corrected for the charged particle response using the tracker as provided by the tcMET variable [2].

This analysis is an update to our previous analysis [3] of CMS simulated data in pp collisions at $\sqrt{s} = 14$ TeV. We found that with 10 pb^{-1} of collision data we should be able to observe $t\bar{t}$ production in the dilepton final state with the signal to noise ratio[1] of 4.4 to 1 (9.4 to 1) in all channels combined (in $e^\pm\mu^\mp$ alone) and we should be able to measure the cross section with 11% (13%) in all channels combined ($e^\pm\mu^\mp$ alone). Although the $t\bar{t}$ production cross-section is expected to be a factor of two lower at $\sqrt{s} = 10$ TeV, we still expect to measure the production cross-section in the dilepton channel with a statistical uncertainty of 15%

The scope of this note is the following:

- Survey all sources of SM dileptons.
- Identify tools, variables, and techniques that are useful in separating signal from background.
- Define a reasonable set of base requirements for this analysis. These will of course be re-evaluated with real data.
- Understand how to estimate the backgrounds in a data-driven way, whenever possible.
- Survey the various systematic uncertainties for the cross-section measurement, and provide a first estimate of their size.

This paragraph is an aside helpful to compare three analyses of the $t\bar{t} \to \ell\ell$ production Two more analyses described in Refs. [1, 4] of the top-quark pair production were performed in close collaboration with this analysis. A big fraction of the analysis workflow is shared with the other two: primarily in the samples and the software used, the lepton selections, some of the data-driven background predicion methods. Several minor discrepancies exist in event samples and the selection procedure. The discrepancies in the data driven methods

- We use the same MC samples, except for the WW and ZZ. We are using samples generated with all decay modes, while in the other two analyses these channels are modeled based only on generated leptonic decays. The discrepancies are tiny.
- We use the same software configuration (PAT layer 1 only) to extract data from the event records.
- We use the same trigger selections.
- We use the same lepton kinematic and identification selections.
- All three of the analyses identify a procedure to select only one dilepton in the event among the lepton pairs passing the same preselections. The preselections are different from the final selections only by the isolation requirement (it's looser). The dilepton assignment procedure is different in implementation and can give different results in particularly ambiguous cases. Note that at the level of the final lepton selections the procedure needs to be applied only to less than 0.5% of the events while the discrepancies between the procedures are much smaller than that.

[1] Note that the corresponding numbers reported in [3] do not include the single-top and low-mass DY contributions due to missing MC samples.

- The calorimeter jets used here and in Ref. [1] are counted in the same way with a small exception. The jets which are closer than $\Delta R = 0.3(0.4)$ to the electron are not counted in the jet multiplicity. A quick study shows that the discrepancy is below 0.5%.
- The MET selections in Ref. [1] are tighter (50 GeV vs. 30 GeV in the e^+e^- and $\mu^+\mu^-$final states, and 30 GeV vs. 20 GeV in the $e^\pm\mu^\mp$final state). In part this is intentional: the tighter selection is planned to be applied to the 100 data sample which will likely have the MET understood better. Selections in Ref. [4] do not include MET by design.
- The data driven methods (mathematical methods) to estimate both the DY and the fake lepton contribution are the same in this analysis as in Ref. [4]. Both of these methods were originally developed by the authors of this note (the ideas are primarily taken from the previous experiments and then implemented and studied for CMS).
- Systematic uncertainties for lepton identification and isolation are the same
- Systematics on the residual background (estimated from MC alone) is the same
- The jet energy scale uncertainty is derived in a similar way (the final selections where this uncertainty matters are substantially different to give different numerical results on this uncertainty).

The public summary of this and the other two analyses is given in Ref. [5].

The outline of this note is the following. First, we describe the MC samples we use for this analysis. We then describe the selections applied to the events followed by the tally of expected counts of events passing these selections. After that, we describe the data driven method we developed to estimate the $Z/\gamma^\star$ background and the method to estimate the contribution of events with fake leptons. We then identify the inputs needed to measure the cross-section of the $t\bar{t}$ production in the dilepton final state. The systematic uncertainties are discussed next. We then provide our expectation on the cross-section measurement given the inputs from the expectations for efficiencies, the systematic uncertainties and the data driven methods. Prior to the conclusion, we elaborate on the possible improvements to the signal selection that can be achieved by using the tcMET and the track-corrected jets [6].

3 Data Sets

We use fullsim MC data sets produced in late 2008 to early 2009 with CMSSW 2_1_X. These samples are refered to as Summer08 and Fall08 samples [7]. We estimate our expectations for $t\bar{t}$, single-top, $W+$jets, and $Z+$jets (DY) from data sets generated using the MadGraph [8] generator, while the diboson (WW, WZ, ZZ), the QCD and the low-mass DY (dilepton mass below 50 GeV/c^2) are predicted using the data sets generated with the Pythia generator. Note that from the Pythia M20 samples listed in Table 1 only events with generator level $Z/\gamma^\star$ mass below 50 GeV/c^2 were used for the main prediction to complement the high-mass (above 50 GeV/c^2) MadGraph samples. Also note that for the single top we always include t-channel, s-channel, and tW production processes. Since we find the combination to be totally dominated by the tW process after two leptons are selected, throughout the text we refer to the single-top contribution as tW. We use the ppMuX dataset and a combination of QCD EMenriched and BCtoE datasets (later referred to as EM) to estimate the contribution from pure QCD. While for the e^+e^- and $\mu^+\mu^-$ final states the ppMuX and the EM datasets are taken as is with their corresponding weights, for the $e^\pm\mu^\mp$ final state we assume the contributions overlap and in

this case both ppMuX and EM enter with an additional weight of 0.5. In addition, the VQQ sample, which consists of W or $Z/\gamma^\star$ events generated in association with two c- or b-jets, has some overlap with the ZJets and the WJets samples. As can be seen further in the text the contribution from this sample is much smaller than the expected uncertainty (either statistical or systematic) on the $W+$jets and $Z+$jets samples so that the double-count of events due to the overlap should not affect our predictions. All the datasets were created following the full detector response simulation, the high-level trigger (HLT) simulation, followed by the full reconstruction. The data sets are summarized in Table 1 together with the cross section values suggested for use by the top PAG [9]. The cross sections for each sample are scaled to their expected next-to-leading order (NLO) values except for the QCD multi-jet cross sections taken at the leading order. The $t\bar{t}$ cross section is taken from [10], the diboson cross sections are calculated with the MCFM program [11], the cross sections for $W+$jets and $Z/\gamma^\star+$jets are scaled from their leading order values by a factor of 1.14 [12], and the single-top cross sections are scaled from their NLO values at $\sqrt{s} = 14$ TeV reported in [13, 14] by a ratio of their leading order cross sections reported by MADGRAPH at 10 TeVand 14 TeV.

For our analysis we process these data sets using the physics analysis toolkit (PAT) [15] producing the PAT objects without additional selections and finally process this using a custom process making the edm-based nTuples which extracts the reduced set of information from the PAT objects which is done to simplify the access during the final analysis. This last stage is run in CMSSW_2_2_3 with additional tags summarized in Table 2.

The following two facts should be mentioned about the HLT step in the samples we use here. Unlike in the CSA07 samples used for the previous analysis [3] where at the HLT step the events were required to pass at least one HLT trigger, in the present samples the HLT is run in the tagging mode. The complete HLT configuration used for the samples we use will not be used for the LHC startup following the recently completed trigger review. We benefit from the fact that the lepton triggers identified for the startup are also included in the HLT configuration of the Summer-Fall08 samples.

It should also be noted that the Summer-Fall08 samples are simulated and reconstructed assuming ideal detector aligmnent and calibration. We assume the effects related to the detector miscalibration are similar to those simulated in the CSA07 samples and refer to the results observed in Ref. [3].

4 Event selection

Here we describe the selections applied to the simulated events to discriminate the $t\bar{t}$ production in the dilepton final state from the backgrounds. The main features used to identify the dilepton ttbar events are the two oppositely charged high-p_T isolated leptons in events with jets and a significant $\not\!\!E_T$. To the extent possible for the selections we are using variables available in the PAT objects. All the events in collision data will be selected by the CMS DAQ for storage based on the HLT trigger decision bits: we describe them first.

4.1 Triggers

The events collected by the CMS DAQ system for storage are selected using the two level triggering system. The first level (L1) is a hardware system used to identify muons based on the information only from the muon system (not the inner tracker which has a better resolution) and is used to identify electrons based on the calorimeter information alone. The HLT is the second layer of the trigger system. It uses improved reconstruction and identification algo-

Table 1: Datasets used for the analysis. Number of events, N_{ev}, corresponds to the number of successfully processed events. The cross section numbers, σ, are taken from a reference set of next to leading order cross section estimations maintained by the top physics analysis group [9]. The effective integrated luminosity of the each sample is denoted by $\mathcal{L}_e$. Note that all the dataset names end with GEN-SIM-RECO, which is not displayed in the table to save space.

Dataset name*	σ, pb	N_{ev}	$\mathcal{L}_e$, pb^{-1}
/TTJets-madgraph/Fall08_IDEAL_V9_v2/	412	1 023 322	2 483
/WJets-madgraph/Fall08_IDEAL_V9_v1/	45 600	8 799 192	193
/ZJets-madgraph/Fall08_IDEAL_V9_reco-v2/	4 218	1 158 479	275
/Zee_M20/Summer08_IDEAL_V9_reco-v3/	2 216	1 008 888	455
/Zmumu_M20/Summer08_IDEAL_V9_reco-v2/	2 216	1 275 840	576
/Ztautau_M20/Summer08_IDEAL_V9_v1/	2 216	994 800	449
/WW_Summer08_IDEAL_V9_v1/	74	203 591	2 754
/ZZ_Summer08_IDEAL_V9_v1/	10.4	200 564	19 217
/WZ_incl/Summer08_IDEAL_V9_v2/	32	214 100	6 611
/SingleTop_tWChannel/Summer08_IDEAL_V9_v1/	28.9	139 048	4 805
/SingleTop_sChannel/Summer08_IDEAL_V9_v1/	5	11 999	2 401
/SingleTop_tChannel/Summer08_IDEAL_V9_v1/	109	251 756	2 301
/VQQ-madgraph/Fall08_IDEAL_V9_v1/	329	999 772	3 035
/InclusiveMuPt15/Summer08_IDEAL_V9_v1/	121 651	6 238 383	51
/QCD_EMenriched_Pt20to30/Summer08_IDEAL_V9_v1/	3 200 000	5 165 892	1.614
/QCD_BCtoE_Pt20to30/Summer08_IDEAL_V9_v1/	192 000	2 217 417	11.5
/QCD_EMenriched_Pt30to80/Summer08_IDEAL_V9_v1/	4 700 000	14 633 866	3.11
/QCD_BCtoE_Pt30to80/Summer08_IDEAL_V9_v1/	240 000	2 030 444	8.46
/QCD_EMenriched_Pt80to170/Summer08_IDEAL_V9_v1/	285 000	5 661 833	19.9
/QCD_BCtoE_Pt80to170/Summer08_IDEAL_V9_v1/	22 800	798 039	35
/QCDpt30/Summer08_IDEAL_V9_v4/	109 057 228	2 805 000	0.0257

Table 2: List of additional packages used on top of CMSSW_2_2_3 release.

Package	Tag
`TopQuarkAnalysis/TopObjectProducers`	V04-06-01
`PhysicsTools/PatAlgos`	V04-14-19
`PhysicsTools/PatUtils`	V03-05-02
`DataFormats/PatCandidates`	V03-18-04
`CondFormats/JetMETObjects`	V01-06-06
`JetMETCorrections/Configuration`	V01-08-11
`JetMETCorrections/Modules`	V02-09-00
`JetMETCorrections/Algorithms`	V01-07-11
`JetMETCorrections/JetPlusTrack`	V03-02-06
`RecoMET/METProducers`	V02-08-02-14
`RecoMET/METAlgorithms`	V02-05-00-16
`DataFormats/METReco`	V00-06-02-09
`RecoMET/Configuration`	V00-04-02-15

rithms similar to the offline, and is capabale of reconstructing charged particle tracks in the inner tracker.

To select the dilepton signature one could use either the inclusive single lepton triggers or the dilepton triggers. The dilepton triggers become more important when the high rates of the single leptons require the single lepton thresholds to be high. The trigger thresholds are expected to be well below the requirement of $p_T > 20\,\mathrm{GeV}/c$ for the anticipated instantaneous luminosities (up to or around $\times 10^{31}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [16]) for the data used for this analysis [17].

In this Monte Carlo analysis, we use inclusive lepton triggers with no isolation, *i.e.*, the logical OR of `HLT_Ele15_SW_L1R` and `HLT_Mu9`. These triggers are present in the Summer08/Fall08 samples that we are using, as well as in the expected startup data taking trigger table [17].

To measure the efficiency of these two triggers in Monte Carlo, we select $WW \to e\mu$ events where both leptons are truth matched to $W \to e$ and $W \to \mu$ and where both leptons pass the requirements of Sections 4.2 and 4.3. We measure efficiencies of $96.2 \pm 0.6\%$ and $92.5 \pm 0.8\%$ for electrons and muons respectively. (Muons were required to have $|\eta| < 2.0$ to be within the single muon trigger acceptance).

In data we will measure the single lepton trigger efficiencies from tag-and-probe on $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ decays [18].

4.2 Muon selection

The muons selected for the $\mu^+\mu^-$ and $e^\pm\mu^\mp$ dilepton final states are required to have $p_T > 20\,\mathrm{GeV}/c$ and $|\eta| < 2.4$. They have to be reconstructed by the global muon algorithm (`isGlobalMuon`), have the number of tracker hits $N_{\mathrm{hit}}^{\mathrm{trk}} \geq 11$ and satisfy the following requirement on the global fit χ^2 normalized by the number of degrees of freedom ndof of the fit $\chi^2/\mathrm{ndof} < 10$.

We require the isolated muons to be isolated both in the tracker and in the calorimeter. The tracker isolation, I_{trk}, is defined as

$$I_{\mathrm{trk}} \equiv \frac{p_T}{p_T + \sum_{\Delta R<0.3} p_T^{\mathrm{trk}}},$$

where the sum is taken over tracks within a cone of $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ and the muon's track is excluded from the sum. The selections applied to the tracks are given in Ref. [19]. The calorimeter isolation, I_{cal}, is defined similarly as

$$I_{\mathrm{cal}} \equiv \frac{p_T}{p_T + \sum_{\Delta R<0.3}(E_T^{\mathrm{ecal}} + E_T^{\mathrm{hcal}})},$$

where the sum is taken over the electromagnetic and hadronic components of the calorimeter towers and the footprint of the muon energy deposit in the calorimeter is removed as described in Ref. [19]. The sums over the tracker and the calorimeter deposits are available in the `pat::Muon` objects we use. We define loosely isolated muons as muons passing $I_{\mathrm{trk}} > 0.5$ and $I_{\mathrm{cal}} > 0.5$. The isolated muons are required to have $I_{\mathrm{trk}} > 0.9$ and $I_{\mathrm{cal}} > 0.9$. We apply the cuts separately to the tracker and the calorimeter isolation variables in case the performance of the calorimeter isolation is worse than expected or changes with instantaneous luminosity or with detector conditions and might need to be adjusted separately. We have investigated performance of this isolation selection compared to the requiremnt on a single isolation variable using the sum over tracks and calorimeter together

$$I_{\mathrm{comb}} \equiv \frac{p_T}{p_T + \sum_{\Delta R<0.3}(p_T^{\mathrm{trk}} + E_T^{\mathrm{ecal}} + E_T^{\mathrm{hcal}})}.$$

198 As shown in Fig. 1, the differences between the two isolation variables are minor. This leaves
199 room for changes in case the calorimeter performance is worse in reality.

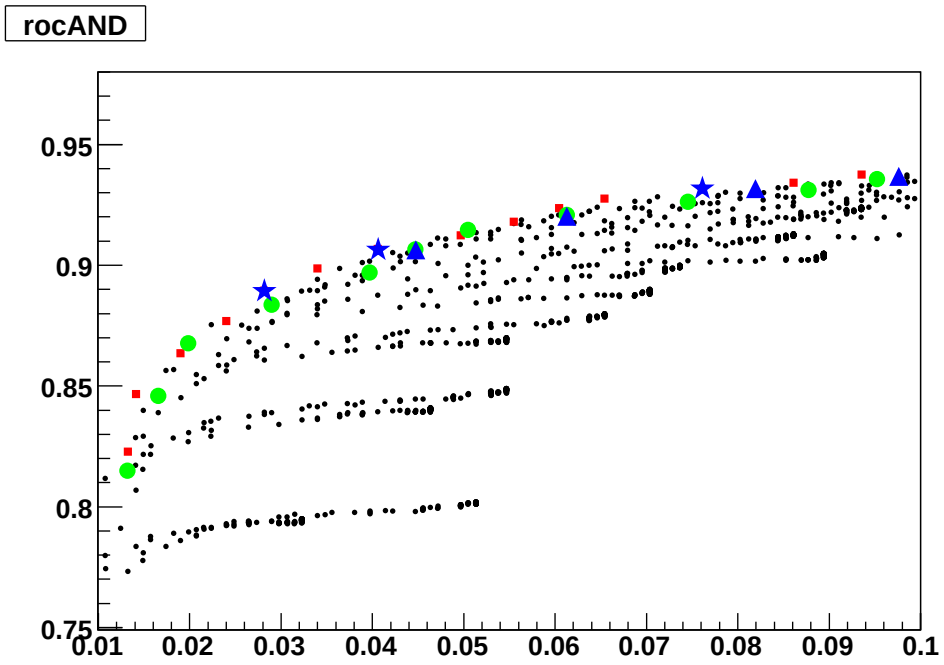


Figure 1: Comparison of isolation efficiency plotted along the vertical axis for muons from $t\bar{t}$ events where both Ws decay leptonically to the selection efficiency plotted along the horizontal axis for fake muons in the Pythia QCDpt30 sample. The lepton candidates selected for the denominator of the efficiency pass all identification requirements except isolation with the muons from $t\bar{t}$ events matched to the generator level muons from Ws. The points displayed are for a scan over $I_{\text{trk}} > a, I_{\text{cal}} > b$ (black smaller dots), $I_{\text{comb}} > a$ with a step of 0.01 (red boxes), $I_{\text{comb}} > 0.92, 0.9$, and 0.85 (blue stars, left to right respectively), $I_{\text{trk}} > a, I_{\text{cal}} > a$ with a step of 0.01 (green circles); and $(I_{\text{trk}} > 0.92, I_{\text{cal}} > 0.92)$, $(I_{\text{trk}} > 0.9, I_{\text{cal}} > 0.9)$, $(I_{\text{trk}} > 0.9, I_{\text{cal}} > 0.85)$, and $(I_{\text{trk}} > 0.9, I_{\text{cal}} > 0.8)$ (blue up triangles, left to right respectively).

200 4.3 Electron selection

201 Similar to muons, the electrons selected[2] for the e^+e^- and $e^\pm\mu^\mp$ dilepton final states are requred
202 to have $p_T > 20$ GeV/c and $|\eta| < 2.4$. To suppress contribution from other particles identified
203 as electrons (fakes electrons) we apply *categorised loose* electron ID requirement [20] available
204 in the `pat::Electron`.[3] To suppress contributions from photons converting asymmetrically[4]
205 and then reconstructed as electrons we require the electron to be consistent to be coming from
206 the interaction point: we require the electrons to have the impact parameter corrected for the
207 beam position d_0 to be $|d_0| < 400$ μm. In addition to suppress the contribution from a muon
208 faking an electron we require the electron to have $\Delta R > 0.1$ with respect to any muon. By

[2] We use the standard GSF electrons.

[3] We are aware that some variables used to define the electron ID flags have changed since CMSSW_1_6_X or CSA07 sample production while the code used to compute the ID flags did not. This doesn't have a significant effect on the electron ID efficiency though.

[4] The asymmetric conversions are more likely to pass the isolation cuts which we apply further than the symmetric ones.

muons faking electrons we mean cases where a relatively high-p_T photon was radiated along the muon and then together with the muon track gets reconstructed as an electron.

Similar to muons, we select isolated electrons by applying selections to the track and the calorimeter isolation variables separately. The track and the calorimeter isolation variables are based on the sums of the transverse momenta (transverse energies) in the tracker (calorimter) which are provided in the `pat::Electron`. The tracker isolation is defined as that for muons (also in the cone of $\Delta R < 0.3$), only with slightly different selections applied to the tracks entering the sum [21]. For the calorimeter isolation, the contributions from ecal are taken from the per-crystal readout (rec-hits, without additional thresholds as applied during the tower reconstruction and as used for muon isolation) while the hadronic calorimeter part is taken from the calorimeter towers. The footprint of the electron energy deposits in the ecal is removed using the Jurassic selection [21], which is done at the PAT object creation stage.

As in the case with muons, we define loosely isolated electrons by requiring $I_{\text{trk}} > 0.5$ and $I_{\text{cal}} > 0.5$. The isolated electrons are identified by passing $I_{\text{trk}} > 0.9$ and $I_{\text{cal}} > 0.8$. The looser numerical value of the cut on I_{cal} for electrons compared to that for muons reflects the fact that the thresholds used in the electron ecal isolation reconstruction are lower and more noise enters the sum while we require the efficiency to be roughly the same as that for muons. The choice of the cut on I_{cal} is illustrated in Fig. 2, in which the right-most three blue up triangle points represent a scan over $I_{\text{cal}} > 0.9, 0.85$, and 0.8 left to right respectively for $I_{\text{trk}} > 0.9$.

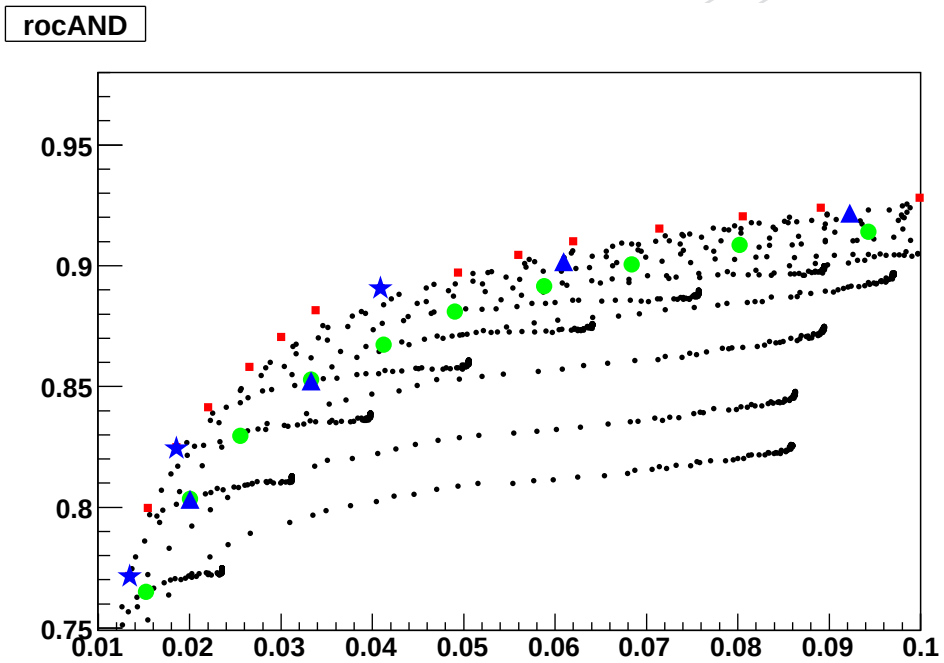


Figure 2: Comparison of isolation efficiency plotted along the vertical axis for electrons from $t\bar{t}$ events where both Ws decay leptonically to the selection efficiency plotted along the horizontal axis for fake electrons in the Pythia QCDpt30 sample. The lepton candidates selected for the denominator of the efficiency pass all identification requirements except isolation with the electrons from $t\bar{t}$ events matched to the generator level electrons from Ws. The choice of points is the same as in Fig. 1. Note that the right-most three blue up triangle points represent a scan over $I_{\text{cal}} > 0.9, 0.85$, and 0.8 left to right respectively for $I_{\text{trk}} > 0.9$.

4.4 Dilepton selection

For the three dilepton final state (e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$) we require the event to have a pair of oppositely charged leptons of the corresponding kind passing the requirement detailed in Sections 4.2 and 4.3. In the final event selection both the leptons in the pair are required to be isolated.

Each event is attributed uniquely to one of the three final states if there are multiple lepton pairs passing the lepton selections. It is preferable that the selection of a unique dilepton per event can be made early before the final event selection is made, so that the cut flow or the $N-1$ cut analysis can be made with a consistent set of events. It is important in this case that the dilepton assignment procedure does not mis-assign a significant fraction of signal events. It should also be clear that the rate of multiple dilepton pairs in the signal events should not be too large when the unique dilepton is selected. While the rate of multiple dileptons in the $t\bar{t}$ signal passing all selections is only around 0.1%, it grows to 5% if only loose isolation is required ($I_{\text{trk}} > 0.5$ and $I_{\text{cal}} > 0.5$). We have considered several assignment procedures and have found that the dileptons with higher momentum, better isolation, and better identification ca provide a sufficiently good dilepton assignment. Simpler choices were found not to give a sufficient separation in $t\bar{t}$ signal events passing the loose isolation selection. Out of the $t\bar{t}$ signal events with at least two dileptons passing the loose selections (1573 simulated events) the dilepton selection with the highest mass gives a correct decision in roughly 71% of the cases (1092 events) and will correspond to 1.5% loss in efficiency. Out of the same 1573 simulated events the dilepton selection of the lepton with the largest p_T(simple choice if one lepton of the two dileptons are shared) gives a correct decision in approximately 68% (1072 events). The event procedure chosen here gives a correct assignment in 95% of the cases (1493 corectly assigned).

The event assignement procedure is based on choosing the dilepton pair with the highest p_T of the leptons and the highest degree of their isolation which is typical of the leptons coming from the W decays compared to those coming from the QCD jets. The contribution from multiple leptons from WZ and ZZ or more multiboson production is small and no special care is taken to select the event-qualifying dilepton pair to address the multiboson contribution. To select the dilepton which qualifies the event we attribute a weight to all the dilepton candidates in the event and then select the one with the highest weight. The dilepton weight is defined as a product of single lepton weights which are defined as

$$w_\ell = (I_{\text{trk}} * I_{\text{cal}} - 0.25)[1 - (20/p_T)^2 + k_\ell^{\text{ID}}],$$

where the ID bonus k_ℓ^{ID} is equal to 0.4 for muons, to 0.2 for electrons passing the *category based tight* ID, and is equal to 0 in other cases. The ID bonus is introduced to account for relatively higher rate of good muons compared to electrons. The functional form of the weight was selected from a limited set of similar functions. It was optimised on the PYTHIA $t\bar{t}$ sample in 2_1_X and was later confirmed in the current sample. Most of the gain is achieved by combining the momentum and the isolation.

The dilepton selection procedure described above is applied to events with more than one pair of oppositely charged loosely isolated leptons, where leptons are required to pass the identification requirements described in Sections 4.2 and 4.3. This ensures that a single event does not contribute to more than one dilepton final state.

In order to suppress the $Z/\gamma^\star$ background we require the lepton pair selected for the e^+e^- and $\mu^+\mu^-$ modes to have the invariant mass $M_{\ell\ell}$ outside the *near-Z region* defined as $76\ \text{GeV}/c^2 < M_{\ell\ell} < 106\ \text{GeV}/c^2$. Note that the e^+e^- and $\mu^+\mu^-$ dileptons inside the *near-Z region* are used later in the data-driven method to estimate the $Z/\gamma^\star$ contribution.

4.5 Jet selection

In this analysis we use calorimeter jets recorded at the PAT stage. These jets are reconstructed with the SIS-cone algoritm with the cone size of 0.5. The jets are corrected for the energy response using L2 and L3 corrections [22]. We count jets having $p_T > 30$ GeV/c and $|\eta| < 2.4$. The choice of the jet pseudorapidity range corresponds to the tracker coverage and is in common with the selection used by other analyses in the group which rely on the tracker to either reconstruct the jets or apply b-tagging.

In addition we explore the possibility to use jets reconstructed with the JPT algorithm [6], which is shown to have a better resolution in the current simulation. The improvement in the jet energy resolution can help select the signal $t\bar{t}$ events better while potentially rejecting the softer jets radiated in the initial state in the $Z/\gamma^\star$ or single-top production. As shown in Section 11, the gain is not significant.

4.6 Missing transverse energy selection

Events with undetected neutrino energy in the $t\bar{t}$ dilepton final state are selected using the $\not\!\!E_T$ variable reconstructed at the PAT stage. This $\not\!\!E_T$ is based on the sum of calorimeter tower energies which is corrected for the calorimeter energy response to jets (*Type-1 jet correction*) and is also corrected for the momenta of muons in the event which are not measured by the calorimeter (*Type-1 muon correction*). As illustrated in Fig. 3, the contribution from the $Z/\gamma^\star$ background to the e^+e^-, and $\mu^+\mu^-$ final states is significant and a more stringent selection is required here compared to the $e^\pm\mu^\mp$ final state. We define the baseline $\not\!\!E_T$ selection to be $\not\!\!E_T > 20$ GeV for the $e^\pm\mu^\mp$ final state and $\not\!\!E_T > 30$ GeV for the e^+e^-, and $\mu^+\mu^-$ final states. This selection is in common with other ongoing analyses in the top group. While, e.g., in the analysis described in [1], it is possible to achieve a higher signal-to-background ratio by requiring b-tagged jets in the final event selection, a better rejection of $Z/\gamma^\star$ events with mismeasured $\not\!\!E_T$ is needed. Since in real data compared to this simulation the performance of $\not\!\!E_T$ can be worse, instead of placing a tighter cut and relying on good performance of $\not\!\!E_T$ in early data we leave a looser cut.[5] With these $\not\!\!E_T$ requirements the expected ratio of dilepton $t\bar{t}$ events to $Z/\gamma^\star$ events is roughly the same as we observed in our previous analysis [3].

It should be noted here that compared to our previous analysis we do not apply the selection dependent on the ratio $\not\!\!E_T/p_T^{\ell\ell}$ if the $\not\!\!E_T$ vector direction is anti-aligned with the dilepton vector direction. This requirement is no longer effective because we are using the $\not\!\!E_T$ corrected for the jet energy, which removes the correlation between the $\not\!\!E_T$ vector direction and the dilepton direction.

As mentioned above, this analysis makes use of the traditional calorimeter-based $\not\!\!E_T$ variable. In Section 11 we investigate the improvement in the analysis that could be achieved using tcMET [2]. We find that the use of tcMET would reduce the Drell Yan background by about a factor of three.

5 Expectations purely from Monte Carlo

The summary of the expected numbers of events passing the selections described in Section 4 is given in Table 3. These pure MC expectations are also illustrated in Fig. 4. The predictions of the data driven methods described in Sections 7 and 8. The values estimated with these data driven methods will supersede the predictions for DY+jets (in the e^+e^- and $\mu^+\mu^-$final states),

[5] This leaves a safety factor of up to about three in case the behavior of $\not\!\!E_T$ is worse in early data.

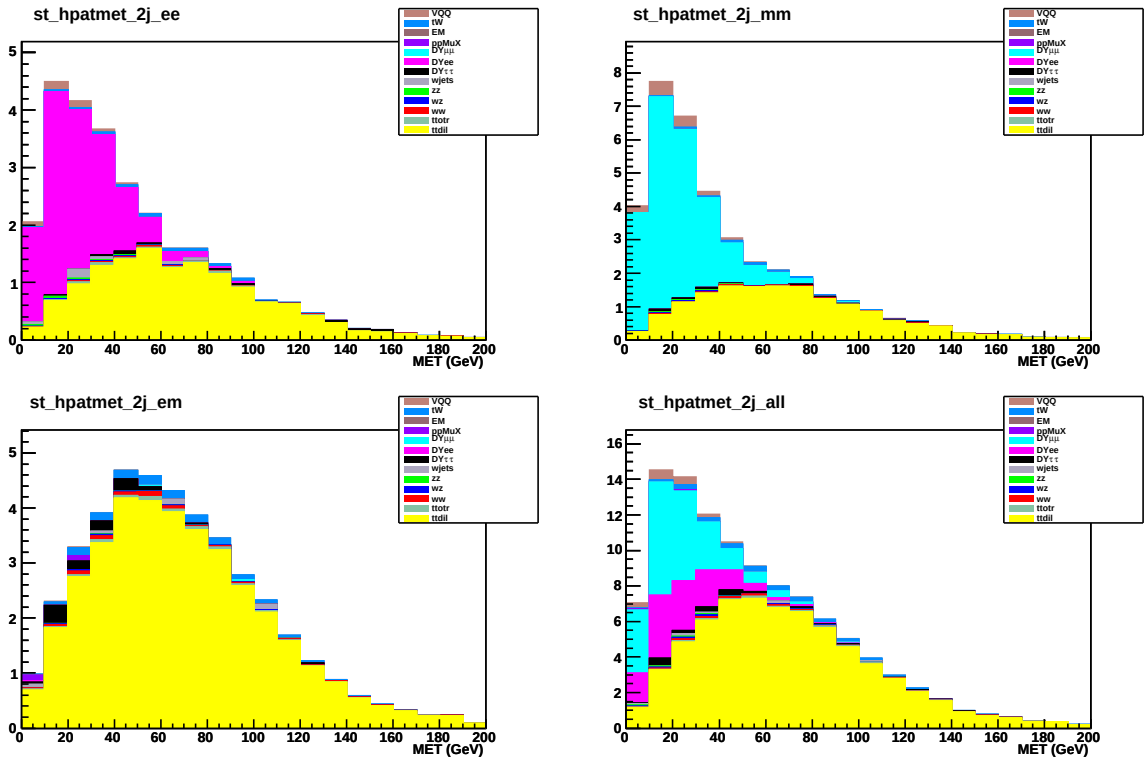


Figure 3: Distribution of $\not\!\!E_T$ in ee (top left), $\mu\mu$ (top right), $e\mu$ (bottom left), and all channels combined (bottom right) events. The events are passing all requirements except for $\not\!\!E_T > 30(20)$ GeV in ee, $\mu\mu$ ($e\mu$) cases. The first two contribution labeled as *ttdil* (yellow) and *ttotr* (light green) are shorthand notations for $t\bar{t}$ events with both Ws decaying leptonically and for all other $t\bar{t}$ events respectively.

W+jets, and pure QCD multi-jet. The contributions from the single-top, dibosons (excluding contributions where the dilepton comes from a Z), $Z/\gamma^\star \to \tau^+\tau^-$, and $W/Z+\gamma+$jets remain predicted by the simulation alone. Here, the $W/Z+\gamma+$jets corresponds to cases where the photon converts to an electron-positron pair that contributes an electron or a positron to the dilepton pair, while the second electron (or positron) of the conversion pair evades the event selection. Preliminary estimates of this contribution in the simulation suggest this contribution after the final event selection is 0.1 ± 0.1 (0.2 ± 0.2) events with at least two jets in the $e^\pm\mu^\mp$ (e^+e^-) final state and no contribution to the $\mu^+\mu^-$ final state as discussed in Appendix B.

The impact of the event selections on the number of expected events is summarized in Table 4. In addition to the full expectation of the number of events with $N_{\text{jets}} \geq 2$, the corresponding numbers are given for cases when one of the selections is disabled.

Similar to the observations in our previous analysis [3], here are some interesting points to note:

- As anticipated, the ee and $\mu\mu$ channels suffer from large Drell Yan backgrounds.
- In all channels, but particularly in the $e\mu$ channel, events with $N_{\text{jets}} \geq 2$ are dominantly $t\bar{t}$.
- The dominant background in the $e^\pm\mu^\mp$ final state is from the single-top production.
- The expected event yields are such that the statistical uncertainty on a cross-section measurement in 10 pb^{-1} will be around 15%.

ee final state				mm final state		
	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$
ttdil	0.70 ± 0.05	4.20 ± 0.13	11.6 ± 0.2	0.49 ± 0.04	3.92 ± 0.13	13.2 ± 0.2
ttotr	0.00 ± 0.01	0.03 ± 0.01	0.21 ± 0.03	0.00 ± 0.01	0.01 ± 0.01	0.04 ± 0.01
ww	1.15 ± 0.06	0.42 ± 0.04	0.13 ± 0.02	1.41 ± 0.07	0.53 ± 0.04	0.20 ± 0.03
wz	0.05 ± 0.01	0.16 ± 0.02	0.09 ± 0.01	0.18 ± 0.02	0.11 ± 0.01	0.08 ± 0.01
zz	0.03 ± 0.00	0.04 ± 0.01	0.04 ± 0.00	0.04 ± 0.00	0.04 ± 0.00	0.06 ± 0.01
W+jets	0.93 ± 0.22	0.47 ± 0.16	0.16 ± 0.09	0.00 ± 0.05	0.05 ± 0.05	0.00 ± 0.05
DY$\rightarrow \tau\tau$	0.40 ± 0.12	0.84 ± 0.17	0.33 ± 0.11	0.11 ± 0.06	0.62 ± 0.15	0.29 ± 0.10
DY$\rightarrow ee$	4.93 ± 0.41	5.66 ± 0.44	3.99 ± 0.37	0.00 ± 0.02	0.00 ± 0.02	0.00 ± 0.02
DY$\rightarrow \mu\mu$	0.00 ± 0.02	0.00 ± 0.02	0.00 ± 0.02	2.77 ± 0.29	7.26 ± 0.49	5.07 ± 0.42
muX	0.00 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.20	0.20 ± 0.20	0.00 ± 0.20
EM	0.00 ± 3.21	1.01 ± 3.21	0.00 ± 0.50	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
tW	0.21 ± 0.02	0.75 ± 0.04	0.46 ± 0.03	0.15 ± 0.02	0.63 ± 0.04	0.56 ± 0.03
VQQ	0.33 ± 0.03	0.37 ± 0.03	0.13 ± 0.02	0.22 ± 0.03	0.53 ± 0.04	0.27 ± 0.03
Total	8.7 ± 3.2	13.9 ± 3.2	17.2 ± 0.5	5.37 ± 0.23	13.8 ± 0.3	19.8 ± 0.3

em final state				all final state		
	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$	$N_{jets} = 0$	$N_{jets} = 1$	$N_{jets} \geq 2$
ttdil	1.66 ± 0.08	11.7 ± 0.2	35.6 ± 0.4	2.85 ± 0.11	19.8 ± 0.3	60.5 ± 0.5
ttotr	0.00 ± 0.01	0.04 ± 0.01	0.46 ± 0.04	0.00 ± 0.01	0.08 ± 0.02	0.71 ± 0.05
ww	4.06 ± 0.12	1.32 ± 0.07	0.52 ± 0.04	6.63 ± 0.16	2.28 ± 0.09	0.85 ± 0.06
wz	0.34 ± 0.02	0.37 ± 0.02	0.16 ± 0.02	0.57 ± 0.03	0.64 ± 0.03	0.33 ± 0.02
zz	0.02 ± 0.01	0.05 ± 0.01	0.03 ± 0.01	0.09 ± 0.01	0.12 ± 0.01	0.12 ± 0.01
W+jets	2.49 ± 0.36	0.88 ± 0.21	0.26 ± 0.12	3.42 ± 0.42	1.35 ± 0.26	0.41 ± 0.15
DY$\rightarrow \tau\tau$	1.38 ± 0.22	1.82 ± 0.26	0.69 ± 0.16	1.89 ± 0.26	3.28 ± 0.35	1.31 ± 0.22
DY$\rightarrow ee$	0.00 ± 0.02	0.00 ± 0.02	0.00 ± 0.02	4.93 ± 0.41	5.66 ± 0.44	3.99 ± 0.37
DY$\rightarrow \mu\mu$	0.29 ± 0.10	0.25 ± 0.09	0.07 ± 0.05	3.06 ± 0.31	7.52 ± 0.49	5.15 ± 0.42
μX	0.10 ± 0.10	0.29 ± 0.17	0.10 ± 0.10	0.10 ± 0.10	0.49 ± 0.26	0.10 ± 0.10
EM	0.00 ± 3.21	0.00 ± 3.21	0.00 ± 0.50	0.00 ± 3.21	1.01 ± 3.21	0.00 ± 0.50
tW	0.53 ± 0.04	2.05 ± 0.07	1.40 ± 0.06	0.89 ± 0.05	3.43 ± 0.09	2.42 ± 0.07
VQQ	0.11 ± 0.02	0.16 ± 0.02	0.04 ± 0.01	0.66 ± 0.05	1.06 ± 0.06	0.44 ± 0.04
Total	11.0 ± 3.2	18.9 ± 3.2	39.4 ± 0.5	25.1 ± 3.3	46.7 ± 3.5	76.3 ± 0.8

Table 3: Expected number of events passing the final event selections in 10 pb^{-1} of integrated luminosity. The first two contribution labeled as *ttdil* and *ttotr* are shorthand notations for $t\bar{t}$ events with both Ws decaying leptonically and for all other $t\bar{t}$ events respectively. The statistical error for samples for which the event count is zero is set to one weighted event. For those samples whose event weight is smaller than the precision displayed in the table ($t\bar{t}$ WW, WZ, ZZ, tW, VQQ), the error is set to 0.01. In the *ee* channel, the muX contribution is not counted as the relevant QCD sample is the EM sample. Similarily, in the $\mu\mu$ channel, the muX sample is the relevant QCD sample. In the $e\mu$ case, both samples appear with aweight of 0.5

ee final state					mm final state				
	base	nomet	noZveto	looseIso		base	nomet	noZveto	looseIso
ttdil	11.6	13.5	15.3	13.6	ttdil	13.2	15.4	17.4	16.6
ttotr	0.2	0.3	0.3	5.0	ttotr	0.0	0.0	0.1	4.9
WW	0.1	0.2	0.2	0.2	WW	0.2	0.2	0.3	0.2
WZ	0.1	0.1	0.5	0.1	WZ	0.1	0.2	0.6	0.1
ZZ	0.0	0.1	0.4	0.1	ZZ	0.1	0.1	0.4	0.1
W+jets	0.2	0.4	0.3	3.0	W+jets	< 0.1	< 0.1	< 0.1	1.1
DY$\rightarrow \tau\tau$	0.3	0.4	0.4	0.5	DY$\rightarrow \tau\tau$	0.3	0.4	0.3	0.4
DY$\rightarrow ee$	4.0	11.9	43.8	5.2	DY$\rightarrow ee$	0.0	0.0	0.0	0.0
DY$\rightarrow \mu\mu$	0.0	0.0	0.0	0.0	DY$\rightarrow \mu\mu$	5.1	20.0	53.2	7.0
μX	< 0.4	< 0.4	< 0.4	0.2	μX	< 0.4	< 0.4	< 0.4	41.0
EM	< 1.0	< 1.0	< 1.0	31.1	EM	< 1.0	< 1.0	< 1.0	9.0
tW	0.5	0.5	0.6	1.1	tW	0.6	0.7	0.7	1.9
VQQ	0.1	0.5	1.6	0.2	VQQ	0.3	1.2	2.4	0.5
Total	17.1	27.8	63.3	60.2	Total	19.8	38.3	75.3	82.7

em final state					all final state				
	base	nomet	noZveto	looseIso		base	nomet	noZveto	looseIso
ttdil	35.6	38.1	35.6	43.2	ttdil	60.5	67.0	68.3	73.4
ttotr	0.5	0.5	0.5	14.7	ttotr	0.7	0.8	0.8	24.6
WW	0.5	0.6	0.5	0.7	WW	0.8	1.0	0.9	1.2
WZ	0.2	0.2	0.1	0.2	WZ	0.3	0.5	1.2	0.5
ZZ	0.0	0.0	0.0	0.1	ZZ	0.1	0.3	0.8	0.2
W+jets	0.3	0.3	0.3	5.5	W+jets	0.4	0.7	0.5	9.6
DY$\rightarrow \tau\tau$	0.7	1.1	0.7	0.9	DY$\rightarrow \tau\tau$	1.3	1.8	1.4	1.7
DY$\rightarrow ee$	0.0	0.0	0.0	0.3	DY$\rightarrow ee$	4.0	12.0	43.8	5.6
DY$\rightarrow \mu\mu$	0.1	0.1	0.1	0.7	DY$\rightarrow \mu\mu$	5.1	20.1	53.3	7.7
μX	0.1	0.2	0.1	63.3	μX	0.1	0.2	0.1	104.4
EM	< 1.0	< 1.0	< 1.0	55.5	EM	< 1.0	< 1.0	< 1.0	95.6
tW	1.4	1.5	1.4	4.3	tW	2.4	2.7	2.7	7.2
VQQ	0.0	0.1	0.0	0.3	VQQ	0.4	1.7	4.0	1.0
Total	39.4	42.6	39.3	189.6	Total	76.3	108.7	177.9	332.5

Table 4: Expected number of events in 10 pb^{-1} of data for the full selection compared to cases when one of the requirement is relaxed. Uncertainties on the numbers are not reported. For reference, the numbers for pure QCD contributions with statistic uncertainty included in all dilepton final states combined are 104.4 ± 3.8 (95.6 ± 9.0) from μX (EM) samples. Note that μX and EM contributions are assumed to completely overlap in the $e^{\pm}\mu^{\mp}$ final state only where they are reported each scaled by a factor of two. The first two contribution labeled as *ttdil* and *ttotr* are shorthand notations for $t\bar{t}$ events with both Ws decaying leptonically and for all other $t\bar{t}$ events respectively.

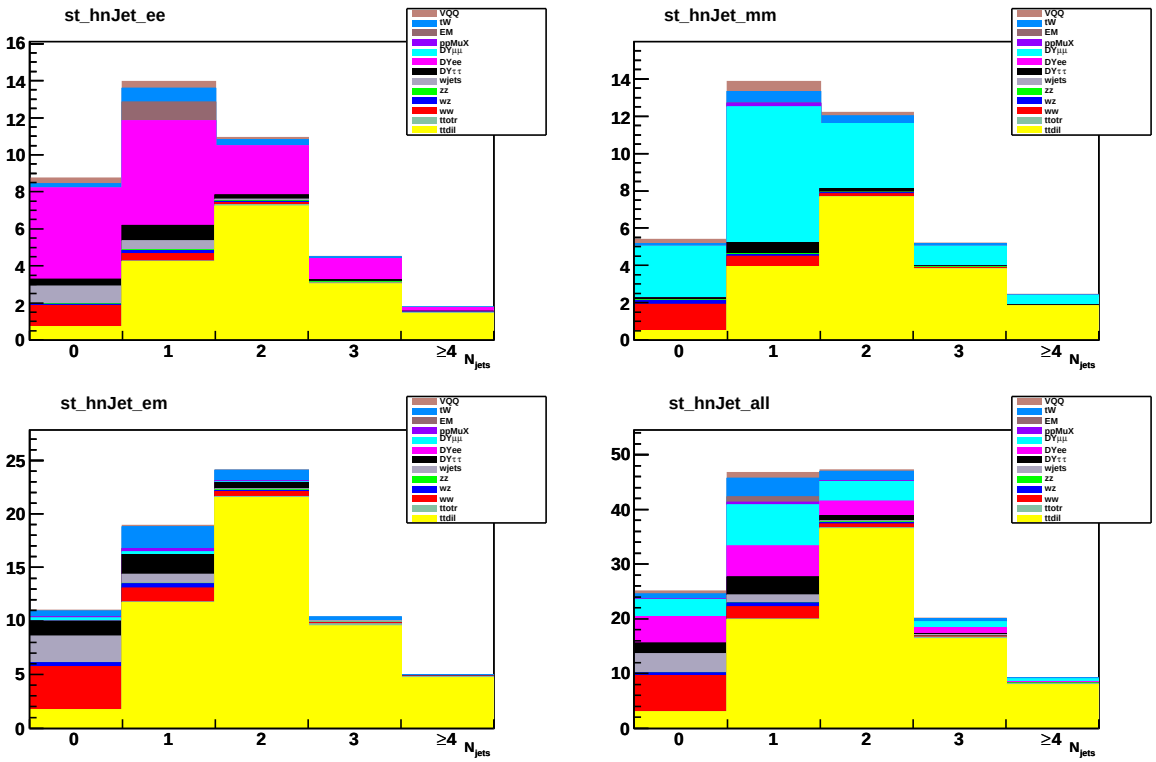


Figure 4: The expected number of dilepton events in 10 pb^{-1} as a function of jet multiplicity from various sources, with the selection of Section 4. Top left: ee; top right: $\mu\mu$; bottom left: $e\mu$; bottom right: all combined. The color coding is the following: yellow=$t\bar{t}$, red = WW, dark blue = WZ, green = ZZ, grey = W+jets, black = DY$\rightarrow \tau\tau$, magenta = DY$\rightarrow ee$, cyan = DY$\rightarrow \mu\mu$, dark blue = single-top. The first two contribution labeled as *ttdil* and *ttotr* are shorthand notations for $t\bar{t}$ events with both Ws decaying leptonically and for all other $t\bar{t}$ events respectively.

- The purely QCD backgrounds appear to be under control as long as both leptons are required to be isolated.
- The W+ jets background is small. It consists mostly of events with a fake electron.
- The diboson backgrounds are small.
- There is a small background in the $e\mu$ channel from Drell Yan $\rightarrow \mu\mu$. These are events where the muon comes from the $Z/\gamma^{\star}$, and the electron is fake.
- For the $t\bar{t}$ sample, 87% of the events with ≥ 2 jets are from the dilepton decay modes ($t\bar{t} \rightarrow ee, \mu\mu, e\mu$). The remainder are almost entirely lepton + τ (12%).

Note that the effect of multiple interactions in the same beam-beam collision (pile-up) have not been simulated or taken into account. It is forseeable that this effect will be negligible in the first 10 pb^{-1} of data considered for this analysis.

6 Strategy for Background Determination

The backgrounds to this analysis are from single-top production, diboson production (WW, WZ, ZZ), Drell-Yan with mismeasured $\not\!\!E_T$, and W+jets and QCD with fake leptons. As discussed in Section 2, the $N_{\text{jets}} = 0$ and 1 bins will be used to validate our background predictions.

The single-top, WW, WZ, and ZZ backgrounds will come almost entirely from Monte Carlo. The Monte Carlo acceptances for these processes will be corrected for differences in data and Monte Carlo lepton identification and trigger efficiencies, as determined from the tag-and-probe method. Further corrections may have to be applied if we find additional important discrepancies, *e.g.*, in $\not\!\!E_T$. The theoretical uncertainties in the cross-section calculations will be reflected in uncertainties in the background normalization.

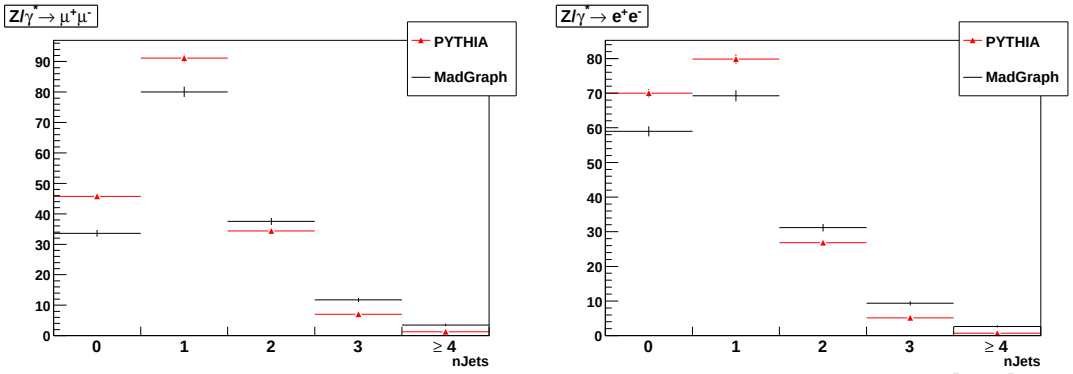


Figure 5: Jet multiplicity in $Z/\gamma^\star \to \mu^+\mu^-$ (left) and $Z/\gamma^\star \to e^+e^-$ (right) events from Pythia (red triangle) and MadGraph (black). All the selections except for the Z-mass rejection are applied. The distributions for events passing all selections have smaller MC statistics and are consistent with the distributions shown here. For a proper comparison to the MadGraph sample which was generated with the $Z/\gamma^\star$ mass above 50 GeV/c^2, the events from the Pythia samples used here are restricted to the $Z/\gamma^\star$ mass above 50 GeV/c^2.

Another set of uncertainties will arise from the modeling of the jet multiplicity in these events, as well as the uncertainty in the jet energy scale. For the moderate jet transverse momenta and the not too high jet multiplicities relevant to our analysis, we have seen that the Pythia and MadGraph model do not differ very much, at least for Drell-Yan, as shown in Fig. 5, which is in agreement with our previous observation [3].

Although the relative uncertainties on these backgrounds will be quite large, their impact on the uncertainty on the cross-section estimate will not be large. This is because these backgrounds are quite small, see Figure 3.

The Drell-Yan background is very important in the e^+e^- and $\mu^+\mu^-$ case. The level of background depends on the $\not\!\!E_T$resolution in CMS. We will use the $\not\!\!E_T$ distribution on Z events to verify and calibrate the $\not\!\!E_T$ resolution for Drell-Yan events. If we find good agreement between data and Monte Carlo for the $\not\!\!E_T$ in Z events, we will rely on Monte Carlo to model the $\not\!\!E_T$ for events away from the Z peak. The method to estimate $Z/\gamma^\star$ background contribution is described in Section 7

In the case of W+jets and QCD, the background predictions will be extracted in a data-driven method. This method is described in Section 8.

7 Data driven method to estimate $Z/\gamma^\star$ background

The method used to estimate the Drell-Yan contribution to the selected data sample is presented in detail in a dedicated note [23], since this method is being used by us for this analysisi as well as a related $WW \to$ dilepton analysis [24]. The goal of the method is to predict, in a data-driven way, the Drell-Yan yield after applying a Missing E_T cut and a Z-mass veto. This is

accomplished by using the near-Z mass region ($76 < m_{\ell\ell} < 106$ GeV) to normalize Monte Carlo to data. In this section we only give a brief summary of the method.

A scale factor between the number of Drell-Yan events predicted by simulation and measured in data may be computed in the near-Z control region. This is expressed in Equation 1, where the total number of Drell-Yan events selected in data or simulation, N_{DY}, is the sum of the number of events inside and outside the near-Z control region, $N^{in}_{DY} + N^{out}_{DY}$.

$$N^{out\ (est)}_{DY} = \frac{N^{in}_{DY\ DATA}}{N^{in}_{DY\ MC}} \cdot N^{out}_{DY\ MC} \tag{1}$$

For convenience and better factorisation of the relevant uncertainties, the two terms which come from simulation can be compressed into a single term, $R_{out/in} = N^{out}_{DY\ MC}/N^{in}_{DY\ MC}$. Thus $R_{out/in}$ is computed and the estimate of the Drell-Yan background outside of the near-Z control region comes from applying the ratio to the number of Drell-Yan events measured inside the control region in data.

Since the event selection requires a large MET, the number of Drell-Yan events in the control region may not be sufficiently larger than other sources of events to assume that the number counted in this region is Drell-Yan dominated. Non Drell-Yan processes which may contribute to the near-Z control region are split into two categories:

- **Peaking backgrounds**, such as WZ and ZZ give a peak in the reconstructed di-lepton invariant mass distribution at the Z mass if both selected leptons come from the Z in the case of WZ, or the same Z in the case of ZZ. We belive that the ZZ contribution can be estimated toghether with the DY contribution, as the ZZ contribution to the analysis is primarily due to $l^+l^-\nu\nu$ and the invariant mass of the two leptons will be in the Z window, giving a similar value of $R_{out/in}$ as DY. The same argument cannot be made for the WZ background as one of the leptons can be from the W. These backgrounds are very small in the $t\bar{t}$ analysis, and are neglected.
- **Non Peaking backgrounds**, such as WW, $t\bar{t}$, tW and $W + jets$ will give a continuum contribution in the reconstructed di-lepton invariant mass. The shape of this continuum is the weighted sum of the unknown shapes of each background, where the weighting of each contribution is unknown.

The non peaking backgrounds must be estimated from data and subtracted. This may be achieved by measuring the number of events in the control region in the $e - \mu$ final state, $N^{in}_{e\mu\ DATA}$. This number, scaled by a factor taking into account the combinatorics and efficiency to reconstruct the different flavor final state relative to each same flavor final state may be used as an estimate of the non peaking background in the same flavor final state. Thus, the estimate of the number of events outside of the near-Z control region due to the Drell-Yan can be expressed as in Equation 2.

$$N^{out\ (est)}_{DY} = (N^{in}_{ll\ DATA} - k \cdot N^{in}_{e\mu\ DATA}) \cdot R_{out/in} \tag{2}$$

The constant, k, is equal to 0.5 for combinatorics between the $e\mu$ and ll final states, multiplied by a correction due to the difference in efficiency to reconstruct and select a muon compared to an electron. This correction can be determined from the number of ee and $\mu\mu$ events in the control region with no MET requirement applied. This is expressed in Equation 3, where it is assumed that the true number of $q\bar{q} \to e^+e^-$ is equal to the true number of $q\bar{q} \to \mu^+\mu^-$, and that the purely geometric acceptances, A_{ee} and $A_{\mu\mu}$, are the same for electrons and muons.

$$\frac{n_{\mu\mu}^{obs}}{n_{ee}^{obs}} = \frac{N_{\mu\mu}^{true} \cdot A_{\mu\mu} \cdot \varepsilon_{\mu}^{2}}{N_{ee}^{true} \cdot A_{ee} \cdot \varepsilon_{e}^{2}} = \frac{\varepsilon_{\mu}^{2}}{\varepsilon_{e}^{2}} \tag{3}$$

$$\rightarrow k = \frac{1}{2}\sqrt{\frac{n_{\mu\mu}^{obs}}{n_{ee}^{obs}}} \tag{4}$$

Since the control region can be expected to be dominated by Drell-Yan events when no MET requirement is applied, the correction factor can be easily measured from data. The measurement of the correction from data can be compared with simulated Drell-Yan events.

We now check that the procedure works in Monte Carlo. Since the same Monte Carlo events are used to estimate $R_{out/in}$ and N_{DATA}, the test is effectively a test of the subtraction of the non peaking backgrounds from the $e\mu$ channel. We find $k_{\mu\mu} = 0.51, 0.55, 0.56$ $k_{ee} = 0.49, 0.46, 0.45$ for $N_{\text{jets}} = 0, 1, \geq 2$ respectively. Using these values of k, the comparison between the observed and actual number of Drell Yan events is shown in Table 7. The method works.

Final State	nJets	$R_{out/in}$	$N_{out\ (est)}$	$N_{DY}^{out\ (true)} + N_{ZZ}^{out\ (true)}$
ee	0	0.090 ± 0.008	5.40 ± 0.45	4.96 ± 0.41
ee	1	0.087 ± 0.007	6.19 ± 0.48	5.70 ± 0.44
ee	≥ 2	0.100 ± 0.010	4.23 ± 0.41	4.03 ± 0.37
$\mu\mu$	0	0.087 ± 0.01	3.16 ± 0.32	2.80 ± 0.3
$\mu\mu$	1	0.100 ± 0.007	7.85 ± 0.53	7.30 ± 0.5
$\mu\mu$	≥ 2	0.105 ± 0.009	5.33 ± 0.46	5.13 ± 0.4

Table 5: Results of the DY background calculation. Uncertainties are from MC statistics. The fourth column is the number of events estimated by the method to be outside the Z region using all data samples. The last column is actual number of events in the ZZ and DY samples.

7.1 Uncertainties

The expected uncertainties from this method are summarized below

- The statistical uncertainties on $N_{ll\ DATA}^{in}$ and $N_{e\mu\ DATA}^{in}$ in equation 2. For an integrated luminosity of 10 pb^{-1}, they translate into a relative statistical uncertainty of about 15%.
- The MC statistics uncertainty on $R_{out/in}$. This is now 10%, but can be made much smaller by using more Monte Carlo events if necessary.
- The theoretical uncertainty on $R_{out/in}$. This can be estimated by running different generators, etc.
- The experimental uncertainty on $R_{out/in}$ which comes from Z events spilling outside the Z mass window. In particular one worries about catastrophic mismeasurements of leptons from Z decays which would cause the invariant mass of the pair to be outside the Z mass window, as well as very high missing E_T. These can be estimated by selecting dileptons with no jets with large missing E_T pointing in the direction of one of the leptons.
- We assign a systematic uncertainty of 30% to this method based on the expected statistics in the N^{in} in the expected data sample, on the observed changes in $R_{out/in}$ with different calibration scenarios (CSA07 only) and MC generators (Pythia and

Madgraph in Summer08), as well as from the comparison of $R_{out/in}$ with the MET cut and with the cut inverted. All of the mentioned effects contribute similarly, close to 10% each, to the combined 30% systematics value.

8 Data Driven Method for Fake Lepton backgrounds

In the context of this analysis, only primary W, Z, and top decay leptons are considered to be true leptons (leptons from $W \to \tau \to \ell$ and $Z \to \tau \to \ell$ are also considered primary). All other reconstructed leptons are considered fake. Thus, in addition to instrumental lepton fakes, we also consider as fakes leptonic bottom and charm decays, electron from conversions, and muons from decays in flight of π and K.

As shown in Table 3, the background from fake leptons in this analysis is expected to be small, particularly in the signal region $N_{\text{jets}} \geq 2$. However, we still need a method to estimate the fake background in real data since it cannot be reliably predicted from Monte Carlo.

In Reference [25] we described a data-driven method to predict the fake background in a dilepton analysis. This method is being applied by us also in the WW analysis [24]. Here we briefly summarize the method, and we apply it to the $t\bar{t}$ analysis.

8.1 The fake rate definition

The method starts by defining a "fake rate" (FR) measured in QCD events. We use the Pythia QCD sample with $\hat{P}_T > 30$ GeV (aka "QCDPt30"). This fake rate is defined as the probability for a lepton passing loose cuts (aka, "Fakeable Object", FO) to pass the analysis cuts as a function of p_T and η. The basic idea is to then apply the FR to dilepton candidates passing loose cuts to obtain a prediction to the fake lepton contribution. The details of the applications of the FR are given in Section 8.3.

Fakeable Objects are defined as follows:

- Electron Fakeable Object, eFO:
  - PixelGSFElectron with $p_T > 20$ GeV;
  - $|\eta| < 2.4$;
  - $I_{\text{trk}} > 0.7$;
  - $I_{\text{cal}} > 0.6$;
  - No reconstructed muon within $\Delta R < 0.1$.
- Muon Fakeable Object, μFO:
  - Global muon with $p_T > 20$ GeV;
  - $|\eta| < 2.4$;
  - Global fit $\chi^2/\text{ndof} < 20$;
  - $I_{\text{trk}} > 0.7$;
  - $I_{\text{cal}} > 0.7$.

The FR for electrons and muons as determined from QCDPt30 are shown in Figure 6.

It is important to keep in mind that the absolute value of the FR is meaningless. This is **not** a fake rate per jet, rather it is simply the probability for a fake lepton passing loose identification and isolation cuts to also pass a tighter selection.

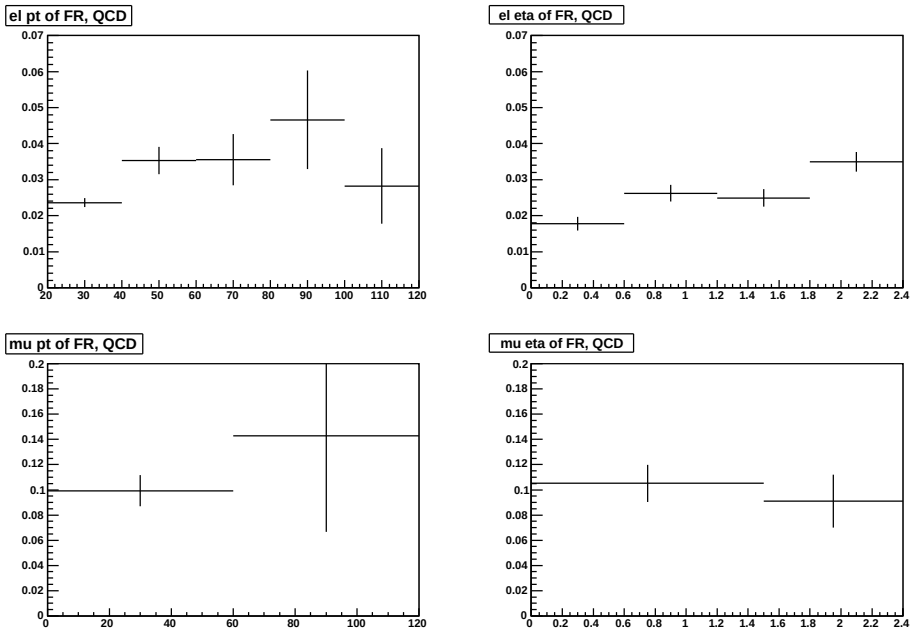


Figure 6: The projection on the p_T and η axis of $FR(p_T, \eta)$ for muons and electrons from the QCDPt30 sample. Error bars are statistical. The last p_T bin includes the overflow.

8.2 Monte Carlo test of the fake rate

We now perform a FR test in $W+$ jets Monte Carlo events. This test is meant to demonstrate that the FR as determined in QCD events can be applied to $W+$ jets. In order to perform this test we define the following four event selections:

1. $W \rightarrow \mu + e$:
   - Require a global muon, $p_T > 20$, truth matched to $W \rightarrow \mu$.
   - Require an opposite sign electron that passes all the standard identification and isolation requirements.
2. $W \rightarrow \mu + (eFO \times FR)$:
   - Require a global muon, $p_T > 20$, truth matched to $W \rightarrow \mu$.
   - Require an opposite sign eFO; weight each event by the FR for the corresponding eFO.
3. $W \rightarrow e + \mu$:
   - Require an electron, $p_T > 20$, truth matched to $W \rightarrow e$.
   - Require an opposite sign muon that passes all the standard identification and isolation requirements.
4. $W \rightarrow e + (\mu FO \times FR)$:
   - Require an electron, $p_T > 20$, truth matched to $W \rightarrow e$.
   - Require an opposite sign μFO; weight each event by the FR for the corresponding μFO.

The Monte Carlo test for the electron FR consists of comparing event yields and distributions for $W \rightarrow \mu + e$ and $W \rightarrow \mu + (eFO \times FR)$. Similarly, the Monte Carlo test for the muon FR consists of comparing event yields and distributions for $W \rightarrow e + \mu$ and $W \rightarrow e + (\mu FO \times FR)$.

There is a subtlety. Some μe events are due to $W \rightarrow \mu\nu\gamma$, where the opening angle between the γ and the μ is large, the γ converts, and it is then reconstructed as an electron. Clearly the FR from QCD is not meant to reproduce these events; thus, these events are identified at the GEN level and removed from this closure test. The fraction of $W \rightarrow \mu + e$ events in $W \rightarrow \mu +$ jets Monte Carlo that can be ascribed to this process is $10 \pm 3\%$.

Sample	Yield
$W \rightarrow \mu + e$	75
$W \rightarrow \mu + (eFO \times FR)$	66 ± 4

Table 6: Monte Carlo test of the electron FR. The uncertainty is from FR statistics. See text for details.

Sample	Yield
$W \rightarrow e + \mu$	5
$W \rightarrow e + (\mu FO \times FR)$	5.0 ± 0.6

Table 7: Monte Carlo test of the muon FR. The uncertainty is from FR statistics. See text for details.

Results of the Monte Carlo tests for event yields are given in Tables 6 and 7; results of the Monte Carlo tests for the N_{jet} distribution are displayed in Figures 7 and 8. From these studies we conclude that the QCD FR parametrization does a good job of reproducing the rate of fake electrons and muons in $W + jets$ events. Note that the electron fake contribution is about one order of magnitude larger than the muon fake contribution.

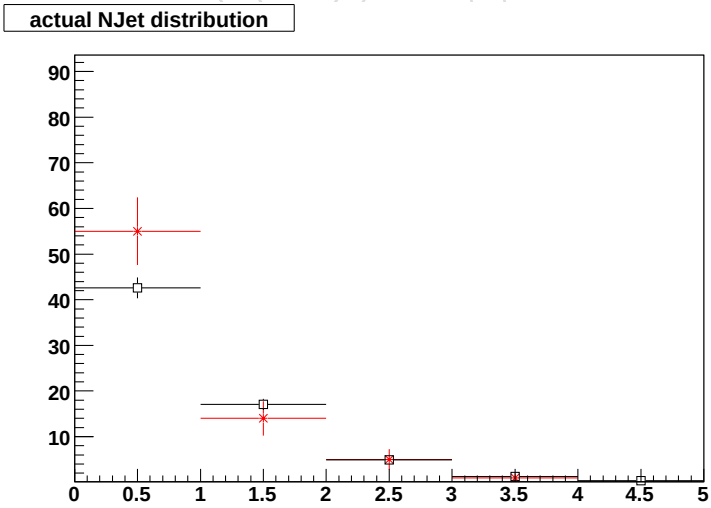


Figure 7: Number of $W \rightarrow \mu + (eFO \times FR)$ events (black) and $W \rightarrow \mu + e$ events (red) in $W +$ jets Monte Carlo as a function of the number of jets. For this test, jets are counted if they have $|\eta| < 3.0$ and $p_T > 15$ GeV without JES corrections.

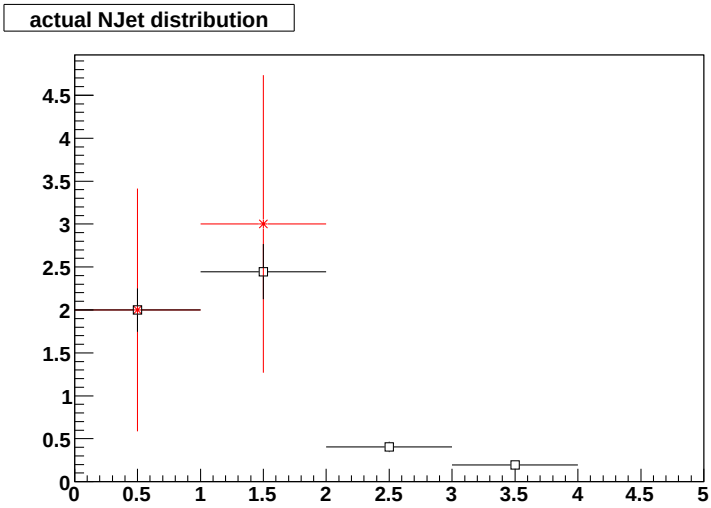


Figure 8: Number of $W \rightarrow e + (\mu FO \times FR)$ events (black) and $W \rightarrow e + \mu$ events (red) in W + jets Monte Carlo as a function of the number of jets. For this test, jets are counted if they have $|\eta| < 3.0$ and $p_T > 15$ GeV without JES corrections.

8.3 Application of the fake rate to our analysis

We are now ready to apply the FR to our analysis. The dilepton sample will consist of true dillepton events, *e.g.*, $t\bar{t} \rightarrow ee/\mu\mu/e\mu$, as well as events with fake leptons, *e.g.*, W+ jets. The existence of true dileptons in the sample complicates the calculation somewhat, since these true dileptons are a major source of fakeable objects.

Our procedure to estimate the fake contribution to our analysis is the following:

- Select lepton + FO events where
  - one of the leptons passes all the standard identification and isolation requirements
  - the other lepton is a FO but fails the standard identification and isolation requirements
- the event passes all the standard kinematical requirements on *e.g.* Missing E_T
- weigh each event by $FR/(1-FR)$ where FR is the fake rate for the FO under consideration
- add up all of the weights

The selection of events where one of the lepton is an FO that does not pass the standard requirements serves two purposes:

- it minimizes the impact of the existence of true dileptons in the sample. In fact, in the limit that the standard identification and isolation efficiencies for true leptons is 100%, true dileptons will not contribute to the lepton + FO selection at all
- it divided the data set into two statistically independent samples, *i.e.*, the sample of events passing dilepton requirements, and the sample of lepton + FO events.

However, since true lepton identification and isolation efficiencies are not 100%, true dilepton will contribute to the lepton + FO selection, and, when rescaled by $FR/(1-FR)$, will lead to an overestimate of the lepton + fake contribution. In addition the pure QCD contribution

where both leptons are fake will be double counted.

Before applying the procedure, it is instructive to estimate the size of this overestimation. The probability that a real lepton is a *FO* but fails the standard requirements is of order 10%; the factor $FR/(1-FR)$ is of order 5%, see Figure 6. Thus, the overestimation of the fake contribution due to true dileptons will be of order of $2 \times 10\% \times 5\% \approx 1\%$ of the total signal contribution (the factor of two comes from the fact that there are two leptons in a dilepton event). This introduces a negligible bias on the cross-section measurement at the present level of statistics. A data-driven method to estimate this bias is presented in Section 8.4.

ee final state			
	$N_{\text{jets}} = 0$	$N_{\text{jets}} = 1$	$N_{\text{jets}} \geq 2$
ttdil (observed)	0.7 ± 0.1	4.2 ± 0.1	11.6 ± 0.2
ttdil (predicted)	0.012 ± 0.001	0.051 ± 0.003	0.112 ± 0.004
W+jets (observed)	0.9 ± 0.2	0.5 ± 0.2	0.2 ± 0.1
W+jets (predicted)	1.08 ± 0.04	0.43 ± 0.03	0.22 ± 0.02
$\mu\mu$ final state			
	$N_{\text{jets}} = 0$	$N_{\text{jets}} = 1$	$N_{\text{jets}} \geq 2$
ttdil (observed)	0.5 ± 0.1	3.9 ± 0.1	13.2 ± 0.2
ttdil (predicted)	0.008 ± 0.002	0.057 ± 0.006	0.311 ± 0.014
W+jets (observed)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
W+jets (predicted)	0.05 ± 0.01	0.04 ± 0.01	0.03 ± 0.01
$e\mu$ final state			
	$N_{\text{jets}} = 0$	$N_{\text{jets}} = 1$	$N_{\text{jets}} \geq 2$
ttdil (observed)	1.7 ± 0.1	11.7 ± 0.2	35.6 ± 0.4
ttdil (predicted)	0.024 ± 0.002	0.173 ± 0.008	0.632 ± 0.018
W+jets (observed)	2.5 ± 0.4	0.9 ± 0.2	0.3 ± 0.1
W+jets (predicted)	1.89 ± 0.06	0.86 ± 0.05	0.36 ± 0.03
$ee + \mu\mu + e\mu$ final state			
	$N_{\text{jets}} = 0$	$N_{\text{jets}} = 1$	$N_{\text{jets}} \geq 2$
ttdil (observed)	2.9 ± 0.1	19.8 ± 0.3	60.5 ± 0.5
ttdil (predicted)	0.044 ± 0.004	0.281 ± 0.011	1.055 ± 0.024
W+jets (observed)	3.4 ± 0.4	1.3 ± 0.3	0.4 ± 0.1
W+jets (predicted)	3.02 ± 0.07	1.32 ± 0.06	0.61 ± 0.04

Table 8: Test of the *FR* procedure on the two main samples, ttdil and W+jets. The "observed" event rate is the event rate running the standard analysis. The "predicted" event rate is what is found running the *FR* procedure. Rates are normalized to 10 pb^{-1}. Uncertainties are purely from Monte Carlo statistics, *i.e.* the *FR* uncertainty has not been included.

The results of the application of the procedure outlined above is summarized in Table 8 for the most important samples with results for all samples shown in Table 9.

First, we find that the *FR* correctly predicts the W+jets background in all N_{jets} bins. This is a major sucess of the method.

The prediction for events with two fake leptons (contributions from μX and EM) are also consistent with the direct counts within statistical uncertainties after taking the double-counting into account. The double-counting occurs here because both of the lepton candidates have a chance to fail the numerator requirement with a rate of $(1 - FR) \approx 1$ relative to their total count.

ee final state				mm final state		
	$N_{\text{jets}} = 0$	$N_{\text{jets}} = 1$	$N_{\text{jets}} \geq 2$	$N_{\text{jets}} = 0$	$N_{\text{jets}} = 1$	$N_{\text{jets}} \geq 2$
ttdil	0.01 ± 0.00	0.05 ± 0.00	0.11 ± 0.00	0.01 ± 0.00	0.06 ± 0.01	0.31 ± 0.01
ttotr	0.01 ± 0.00	0.04 ± 0.00	0.26 ± 0.01	0.00 ± 0.00	0.01 ± 0.00	0.16 ± 0.01
ww	0.02 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00
wz	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
zz	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
W+jets	1.08 ± 0.04	0.43 ± 0.03	0.22 ± 0.02	0.05 ± 0.02	0.04 ± 0.01	0.03 ± 0.01
DY$\rightarrow \tau\tau$	0.07 ± 0.01	0.03 ± 0.01	0.02 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01
DY$\rightarrow ee$	0.11 ± 0.01	0.11 ± 0.01	0.07 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
DY$\rightarrow \mu\mu$	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.09 ± 0.02	0.21 ± 0.03	0.18 ± 0.03
μX	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.06 ± 0.04	0.02 ± 0.02
EM	2.38 ± 0.44	1.40 ± 0.30	0.41 ± 0.11	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
tW	0.02 ± 0.00	0.05 ± 0.00	0.04 ± 0.00	0.01 ± 0.00	0.03 ± 0.00	0.04 ± 0.00
VQQ	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.02 ± 0.00	0.01 ± 0.00
Total	3.71 ± 0.19	2.13 ± 0.09	1.14 ± 0.01	0.18 ± 0.00	0.44 ± 0.00	0.77 ± 0.00

em final state				all final state		
	$N_{\text{jets}} = 0$	$N_{\text{jets}} = 1$	$N_{\text{jets}} \geq 2$	$N_{\text{jets}} = 0$	$N_{\text{jets}} = 1$	$N_{\text{jets}} \geq 2$
ttdil	0.02 ± 0.00	0.17 ± 0.01	0.63 ± 0.02	0.04 ± 0.00	0.28 ± 0.01	1.06 ± 0.02
ttotr	0.01 ± 0.00	0.06 ± 0.00	0.64 ± 0.01	0.01 ± 0.00	0.11 ± 0.00	1.05 ± 0.02
ww	0.04 ± 0.00	0.02 ± 0.00	0.01 ± 0.00	0.06 ± 0.00	0.04 ± 0.00	0.02 ± 0.00
wz	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
zz	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
W+jets	1.90 ± 0.06	0.86 ± 0.05	0.36 ± 0.03	3.02 ± 0.07	1.32 ± 0.06	0.61 ± 0.04
DY$\rightarrow \tau\tau$	0.14 ± 0.01	0.07 ± 0.01	0.02 ± 0.01	0.20 ± 0.02	0.11 ± 0.01	0.05 ± 0.01
DY$\rightarrow ee$	0.00 ± 0.00	0.01 ± 0.01	0.03 ± 0.01	0.11 ± 0.01	0.12 ± 0.01	0.11 ± 0.02
DY$\rightarrow \mu\mu$	0.18 ± 0.01	0.09 ± 0.01	0.05 ± 0.01	0.27 ± 0.03	0.30 ± 0.03	0.24 ± 0.03
μX	0.22 ± 0.04	0.27 ± 0.05	0.19 ± 0.04	0.22 ± 0.04	0.34 ± 0.06	0.21 ± 0.05
EM	0.27 ± 0.13	0.29 ± 0.19	0.43 ± 0.21	2.65 ± 0.46	1.68 ± 0.35	0.84 ± 0.24
tW	0.07 ± 0.00	0.13 ± 0.01	0.14 ± 0.01	0.10 ± 0.01	0.21 ± 0.01	0.22 ± 0.01
VQQ	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.05 ± 0.00	0.05 ± 0.00	0.03 ± 0.00
Total	2.87 ± 0.02	2.01 ± 0.04	2.53 ± 0.05	6.76 ± 0.22	4.58 ± 0.13	4.44 ± 0.06

Table 9: Results of the fake rate calculation. Note that the QCD (μX and EM) contributions are double-counted by the method.

Next, we turn our attention to the overestimation of the fake rate due to true dileptons, which are mostly from $t\bar{t}$. From Table 8 this overestimation is 1.055 events out of 60.5 on the total event yield for $N_{\text{jets}} \geq 2$. This is a 1.7% effect, in rough agreement with the back-of-the-envelope estimate of $\approx 1\%$.

Interestingly, the FR prediction accounts reasonably well for the ttothr[6] contribution in the signal region: 0.7 ± 0.1 ttothr events observed vs. 1.0 events predicted with $N_{\text{jets}} \geq 2$ (Tables 3 and 9). This makes sense, as these events include fake leptons.

8.4 Data driven estimate of the small bias due to real leptons

According to the Monte Carlo, the oversubtraction of the fake contribution is a small effect and can be ignored when compared to the expected statistical uncertainty. However, one could worry about trusting the Monte Carlo for the size of this effect. Therefore, it is useful to have a data-driven way to check it. This then also opens up the possibility to correct the cross-section for this effect.

[6] ttothr is short hand for $t\bar{t}$ decays in a non-dilepton final state

The idea is simple: measure the effect on a clean true dilepton sample, *i.e.*, measure it on $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events. We can apply the lepton + FO selection to this sample, rescale by $FR/(1-FR)$, and see what fraction of these true dilepton events would be ascribed to lepton + fakes. Then, if we can convince ourselves that this fraction would be the same on $t\bar{t}$, we could use the information from Z events to quantify the size of the effect on $t\bar{t}$ events, and possibly correct for it.

To test this idea on Monte Carlo $Z+$ jets, we repeat the analysis inverting the Z veto and removing the Missing E_T requirement. We define a "bias per lepton":

$$\mathcal{B}(l) = \frac{1}{2}\frac{N_{sub}}{N(Z\rightarrow ll)},$$

where $N(Z \rightarrow ll)$ is the number of reconstructed $Z \rightarrow ll + \geq 2$ jets events, while N_{sub} is the number of these events which would have been ascribed to "fakes" when applying the fake rate procedure based on the lepton + FO selection described in the previous section.

The quantity $\mathcal{B}(l)$ is a measure of the bias per lepton introduced by the fake rate subtraction procedure. If $\mathcal{B}(l)$ is known, it can be used to correct for the bias by rescaling the $t\bar{t} \rightarrow l_1 l_2$ yield by $1/(1-\mathcal{B}(l_1)-\mathcal{B}(l_2))$.

The key question is whether $\mathcal{B}(l)$ measured in Z events is applicable to $t\bar{t}$ events.

In Monte Carlo Z events we find $\mathcal{B}(\mu) = 0.96 \pm 0.04\%$ and $\mathcal{B}(e) = 0.17 \pm 0.01\%$. From Table 8, we can extract the same quantity on $t\bar{t} \rightarrow ee$ and $t\bar{t} \rightarrow \mu\mu$ events. We find $\mathcal{B}(\mu) = 1.18 \pm 0.06\%$ and $\mathcal{B}(e) = 0.48 \pm 0.02\%$.

The values of $\mathcal{B}(\mu)$ in Z and $t\bar{t}$ events are in quite good agreement. There seems to be a discrepancy in the electron channel. However, it turns out that the $t\bar{t} \rightarrow ee$ Monte Carlo sample has a contribution of 0.1 out of 11.6 events where one of the electrons is fake[7]. This is actually in good agreement with the fake prediction in the $t\bar{t} \rightarrow ee$ sample of 0.112 ± 0.004 (second line of Table 8). Thus, in reality $\mathcal{B}(e)$ in the $t\bar{t}$ sample is very small, of order 0.1%, and in good agreement with the results of the Z sample.

8.5 Concluding remarks on the fake rate

Even though the contribution of fake leptons to this analysis is small, we have sucessfully developed a data-driven method to measure it.

- The method has been shown to work in Monte Carlo at the 10% level when the lepton is known to be fake at the generator level.
- The method introduces a bias in the $t\bar{t}$ cross-section measurement at the $\approx 1\%$ level, which is negligible compared to the statistical error.
- We have shown that the bias can be measured in a data-driven way from a study of Z events.
- The method double-counts the contribution from events with two fake leptons.
- Combining the uncertainty on applying the fake rate extracted from QCD samples to the data sample together with the potential biases of the method we assume 50% systematic uncertainty on the prediction of the method when it is applied to data.

[7] This was determined by truth matching

9 Systematic uncertainties

Here we summarize the sources of systematic uncertainties we find important to measure the production cross-section of the $t\bar{t}$ in the dilepton final states. Overall the sources of systematics can be categorized as detector effects, effects of theoretical modeling of the contributing processes, uncertainties of the data-driven background prediction methods, effects from multiple collisions, and finally the uncertainty on the determination of the integrated luminosity. While it is important to review the expected systematic uncertainties, all of them need to be considered relative to the expected statistical uncertainty of the measurement, which is expected to be approximately 15%.

The detector-related systematics we consider are from the jet energy scale uncertainty and from the determination of efficiencies of the lepton reconstruction, identification, and isolation. The effects of the theoretical modeling contribute to signal and background differently. While for the $t\bar{t}$ signal we are only interested in the changes of the fraction of events passing all selections relative to changes in the theoretical parameters, for the background prediction we are interested in the variation of the total number of predicted events.

The details of determining the uncertainties of the data driven methods are given in Sections 7 and 8. The uncertainty on the $t\bar{t}$ cross-section arising from the $Z/\gamma^{\star}$ prediction method is approximately 7% in e^+e^- and $\mu^+\mu^-$ final states combined using MC statistics only and approximately 14% from the expected Poisson statistics of the expected $Z/\gamma^{\star}$ contribution. There is no uncertainty from this method to the $e^{\pm}\mu^{\mp}$ final state. For the fake rate method, based on the systematic uncertainty of 50% and the statistical uncertainty expected from our MC samples as given in Table 9, the combined uncertainty from the fake lepton estimation method is 6% (7%) in $e^{\pm}\mu^{\mp}$($e^+e^-$and $\mu^+\mu^-$ combined).

Systematic uncertainties from the lepton selection, ID, and reconstruction efficiencies will be estimated based on the corresponding systematics of the tag-and-probe method used to determine these efficiencies in $Z \to \ell\ell$ data as well as on the systematics we assess from comparing the simulated values for the $Z \to \ell\ell$ sample and the signal (or a corresponding background to be determined from MC) MC samples. We assign the systematics from the dilepton selection to be 5% from the lepton identification, and 3% from the lepton isolation. We expect to be able to use the same tag and probe method to estimate these effects. The tag-and-probe data with Z will provide the efficiency for leptons from data. In the simulation we will get the values for Z and for ttbar (or other signal). The simulated efficiency for Z will be compared to that obtained with tag-and-probe. The level of agreement will give the measure of the systematics. In practice it will be the change of the ratio of efficiencies derived from the tag-and-probe for a range of kinematical variables including the number of jets.

The systematics due to the determination of the integrated luminosity is taken to be 10% as suggested in [26]. This systematics is the same for all backgrounds predicted from MC. It is reported separately from all other systematic uncertainties for the final cross-section measurement.

9.1 Systematics due to jet energy scale

We assess the uncertainty arising from the jet energy scale by varying the energies of all jets in the event by 10% up and down. Scaling by 10% corresponds to the expected uncertainty on the jet energy scale in the early period of data taking. The results of the procedure are summarized in Table 10. The observed changes for different dilepton channels are statistically consistent. We assign 5% (8%) to the JES systematics contribution to the $t\bar{t}$ signal selection acceptance

($N_{\text{jets}} \geq 2$) in the $e^{\pm}\mu^{\mp}$ (e^+e^- and $\mu^+\mu^-$) final states. The same-flavor final states are expected to have a slightly higher dependence on JES due to the more stringent requirement on $\not\!\!E_T$. We assign an 11% systematic uncertainty to the single-top background prediction due to JES. The JES systematics for the remaining backgrounds (WW, WZ, ZZ, and $Z/\gamma^\star \to \tau^+\tau^-$ combined) we predict from MC is 15%.

	$t\bar{t}$				single-top				$VV + (Z/\gamma^\star \to \tau^+\tau^-)$			
	$N_{\text{jets}} = 0,1$		$N_{\text{jets}} \geq 2$		$N_{\text{jets}} = 0,1$		$N_{\text{jets}} \geq 2$		$N_{\text{jets}} = 0,1$		$N_{\text{jets}} \geq 2$	
Channel	$-$	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$	$+$
e^+e^-	13	-9	-9	7	4	-2	-11	10	-12	7	-6	9
$\mu^+\mu^-$	14	-11	-8	7	6	-7	-10	11	-6	5	-13	27
$e^{\pm}\mu^{\mp}$	14	-10	-6	5	4	-5	-12	11	0	5	-16	16
All	14	-10	-7	6	4	-5	-12	11	-3	5	-13	17

Table 10: Relative changes in the number of selected events expressed in % for changes in the jet energy scale down and up by 10% denoted by $-$ and $+$ respectively.

9.2 Uncertainty from multiple pp collisions

We do not expect the multiple pp collisions to be an issue during the data taking for this early analysis. It is expected that the additional collisions will contribute to the estimates of the lepton isolation efficiency, to the MET and the jet counting. Only the calorimeter-based isolation is affected by pile-up whle the tracker-based isolation is expected to be largely independent from pile-up. This effect can be estimated from data by looking at the energy in the randomly directed cones corresponding to the calorimeter isolation cone size. Since the efficiency of the $\not\!\!E_T$ selection is rather high, the effects of pile-up on $t\bar{t}$ selection should be smaller than the effect on the jet multiplicity selection. Based on fast-sim samples generated with Pythia and MC@NLO [27] in CMSSW_1_6 $\sqrt{s} = 14$ TeV with 5 multiple interactions on average, we find the effect on the efficiency of $N_{\text{jets}} \geq 2$ is only 3%. Even in this case the effect can be estimated from data based on the energy integrated in a random cone with area covered by the jet.

9.3 Theoretical systematic uncertainties

Theoretical uncertainties contribute to the prediction of the $t\bar{t}$ selection efficiency in the dilepton final state as well as to the prediction of the total number of background events we estimate from MC. These uncertainties come from modeling the hard scattering process part with the subsequent fragmentation and hadronization as well as modeling of the underlying event. Compared to the relative values of uncertainties on the prediction of the total number of events, the uncertainties on the signal selection efficiency is expected to be smaller. Eventually it would be prudent to derive the uncertainties from MC samples produced using fast simulation. While we do not have all the relevant samples, we make several assumptions about the overall scale of the systematic uncertainties.

The uncertainty on the background prediction is assumed to be 50%. This corresponds to a conservative combination of uncertainties from the scale dependence of the leading order (LO) prediction and next-to-leading order (NLO) to LO difference of the tW production [13, 28], which is the dominant background in our case.

The theoretical uncertainty on the efficiency of the $t\bar{t}$ selection in the dilepton final state is expected to be dominated by the modeling of the hadronic activity in the event. As can be seen from Tables 3 and 4, the requirement of $N_{\text{jets}} \geq 2$ has the largest inefficiency compared to the MET and the isolation requirements. This suggests the largest systematics would come

from the $N_{\text{jets}} \geq 2$ cut. To assess this systematic, we use samples generated with the Alpgen, Pythia, and MC@NLO generators further processed through fast simulation and reconstruction for $\sqrt{s} = 14$ TeV using CMSSW_1_6.[8] In these samples the efficiency to select $N_{\text{jets}} \geq 2$ varies by 2.5%. The same comparison using full simulation Pythia and MadGraph samples produced for $\sqrt{s} = 10$ TeV gives statistically consistent results. The distributions of N_{jets} in these samples are shown in Fig. 9. In this case the difference in the total number of predicted events which amounts to about 3% (which is a rough average between the dilepton modes) can be an effect of MC statistics. We chose to treat this difference as an additional measure of systematics. The total systematic uncertainty on the $t\bar{t}$ selection efficiency from the event modeling is then 4%.

The dilepton assignment procedure applied at he loose lepton selection has an inefficiency of less than 0.3% due to the rate of incorrect dilepton lepton assignments. This performance was cross checked in the PYTHIA and in the MADGRAPH $t\bar{t}$ samples and was found to be consistent in both samples. No systematics is assigned to the dilepton assignment procedure.

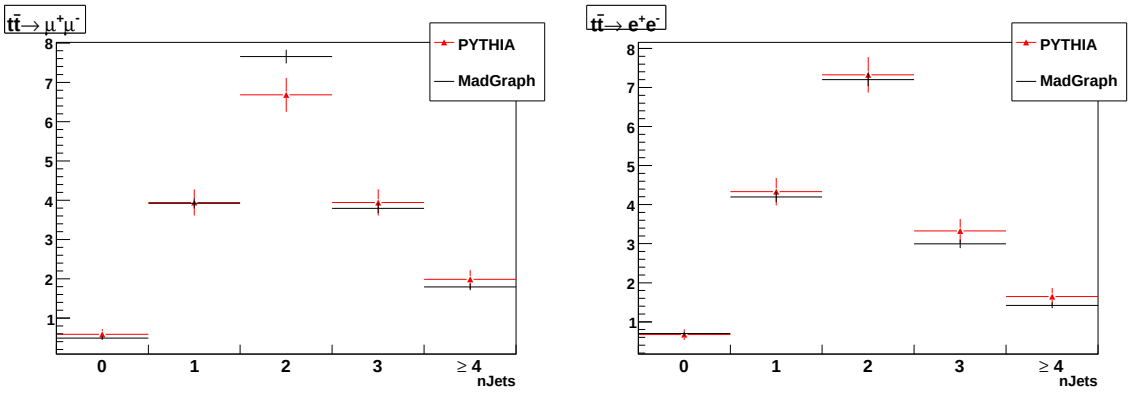


Figure 9: Jet multiplicity in $t\bar{t}$ event in $\mu^+\mu^-$ (left) and e^+e^- (right) final states from Pythia (red triangle) and MadGraph (black) in events passing the full selection. The entries in the histograms are normalized to the absolute $t\bar{t}$ cross section of the full sample and the difference corresponds to the difference in MC statistics and difference in selection efficiencies.

9.4 Summary of systematic uncertainties

We summarize the contributions to the total systematic uncertainty in Table 11. Excluding the uncertainty on the total integrated luminosity, we expect to have approximately 9% (11%) systematic uncertainty from the dilepton $t\bar{t}$ selection efficiency in $e^{\pm}\mu^{\mp}$ (e^+e^- and $\mu^+\mu^-$ combined) modes. The combined uncertainty from the background prediction and subtraction is expected to be 8% (17%) in $e^{\pm}\mu^{\mp}$ (e^+e^- and $\mu^+\mu^-$ combined) final states. The large uncertainty in the same flavor final states is predominantly due to the large uncertainty arising from the Poisson statistics of the expected $Z/\gamma^{\star}$ background events. Once it is clear the $\not\!\!E_T$ performance in data is as expected a more stringent requirement can be applied to the $\not\!\!E_T$ (or, better yet, we can use the tcMET), and the contribution from the $Z/\gamma^{\star}$ can be reduced. Combining the above, excluding the systematics on the integrated luminosity, we expect the systematic uncertainty on the $t\bar{t}$ cross-section in the $e^{\pm}\mu^{\mp}$ final state to be 10% and 16% in the e^+e^- and $\mu^+\mu^-$ combined.

[8] This study was done as an extension of our previous analysis.

Source	e^+e^- and $\mu^+\mu^-$	$e^\pm\mu^\mp$
Lepton ID	5%	5%
Lepton isolation	3%	3%
JES	8%	5%
Theory	4%	4%
All without backgrounds	11%	9%
$Z/\gamma^\star$	10%	N.A.
Fake	4%	4%
MC backgrounds	5%	4%
All w/o $\mathcal{L}$	16%	10%

Table 11: Summary of the systematic uncertainties relative to the expected signal yield in the given mode.

10 Cross-section determination

We define our signal to be $t\bar{t}$ events where both Ws decay leptonically, to an electron or a muon directly, or through a tau first with a subsequent leptonic tau decay. Based on the number of signal events $n_{\ell\ell'}$ ($\ell\ell'$ is e^+e^-, $\mu^+\mu^-$, or $e^\pm\mu^\mp$) identified in data and the total number of expected $t\bar{t}$ events in the $\ell\ell'$ final state $N_{\ell\ell'}$, the dilepton mode cross section can be defined as

$$\sigma_{\ell\ell'} = \frac{N_{\ell\ell'}}{\mathcal{L}} = \frac{n_{\ell\ell'}}{\mathcal{L} A^{\mathrm{MC}}_{\ell\ell'} SF_{\ell\ell'}},$$

where $\mathcal{L}$ is the integrated luminosity, $A^{\mathrm{MC}}_{\ell\ell'}$ is the fraction of $N_{\ell\ell'}$ passing our final event selection and estimated from our signal MC sample, and $SF_{\ell\ell'}$ is the scale factor derived from the comparison of MC selection efficiencies to their values estimated from data or other MC. The value of the product $A^{\mathrm{MC}}_{\ell\ell'} SF_{\ell\ell'}$ with its uncertainties gives us the estimate on the true value of the fraction of all produced $t\bar{t}$ events in the $\ell\ell'$ final state passing our final event selections. We find the value of $A^{\mathrm{MC}}_{\ell\ell'}$ to be[9] $17.3 \pm 0.3\%$, $19.8 \pm 0.3\%$, and $26.6 \pm 0.3\%$ in e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$ final states respectively. Without data we assume the central value of $SF_{\ell\ell'}$ to be 1 with uncertainty determined by the combination of systematic uncertainties summarized in Table 11 in the row excluding the contributions from the background estimates.

Measurements of the cross section in three dilepton final states can be combined into a measurement of the total $t\bar{t}$ cross section based on values of branching fractions for the Ws to decay leptonically (including leptonic tau decays).

11 Selections beyond the baseline

The analysis presented here makes use of standard calorimeter-based reconstruction techniques for jets and missing E_T. New techniques are being developed in CMS that include tracking information in the reconstruction of this basic quantities.

The simplest technique involves correcting the calorimeter using tracking on an average basis. Jets reconstructed with this technique are called *Jets Plus Track* jets [6]; the corresponding missing E_T is called *Track Corrected Missing ET* (tcMET)[2]. The more sophisticated technique,

[9] These values are actually estimates based on the expected number of dilepton events produced. Our nTuples were done with reconstruction-level filter applied and the exact number of dilepton events generated in the given mode is not available.

Particle Flow (pFlow), associates tracks with the measured energy deposition in the calorimeter on a track-by-track basis [29]

It is then very interesting to see how these new techniques could improve the analysis presented in this note. This analysis is presently based on CMSSW 21X Full Sim samples, and pFlow should only be used on CMSSW 22X (and beyond) samples. Thus, our first glimpse at the performance of these new tools is limited to JPT/tcMET; pFlow will be added soon.

The main effect of improving the Missing E_T in this analysis is to reduce the Drell Yan background. We find that if we substitute tcMet for Met in our analysis, leaving everything unchanged, the Drell Yan background in the ≥ 2 jets sample is reduced by about a factor of 2.8, leaving the signal essentially unchanged, see Figure 10.

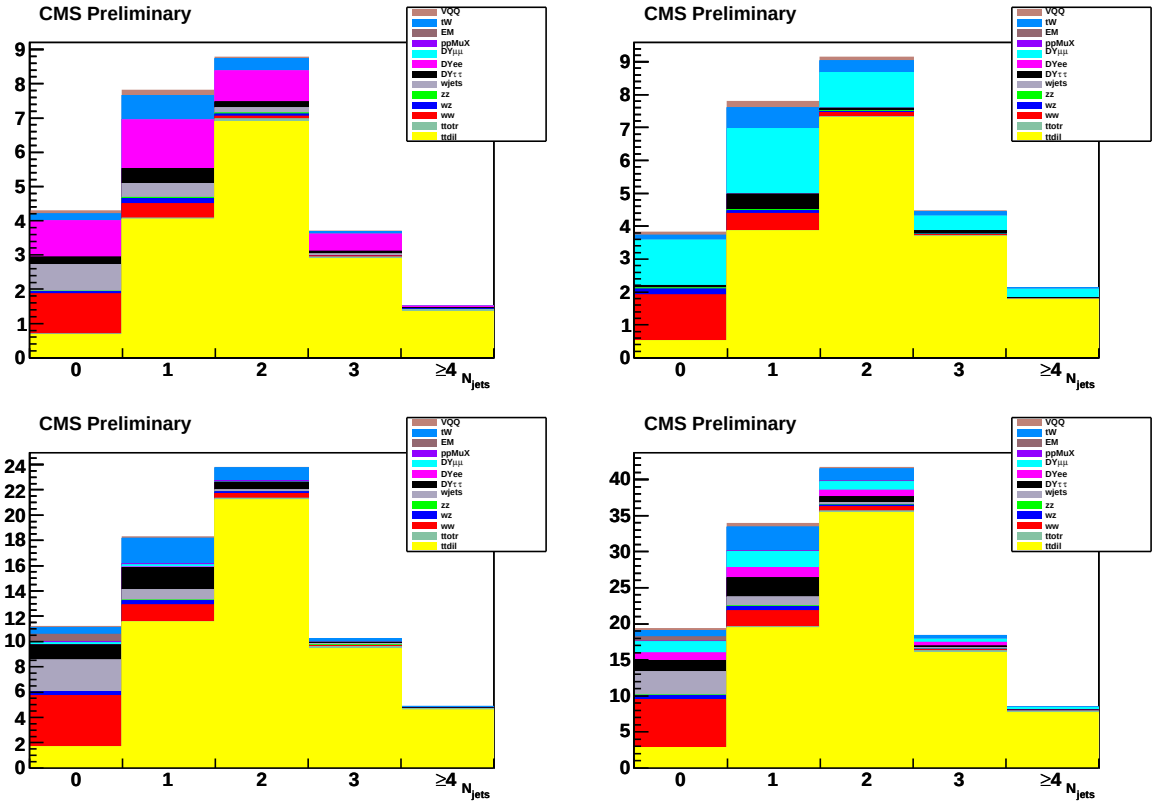


Figure 10: The expected N_{jets} distribution when the analysis is repeated using tcMET. This should be compared with the default results of Fig. 4. Top left plot: ee; Top left plot: $\mu\mu$; Bottom left plot: $e\mu$; Bottom right plot: all together.

Since this analysis only counts jets above some threshold, improvements to the jet energy resolution are not expected to have a major effect. However, background sources in this analysis have lower jet multiplicity than $t\bar{t}$. These background events can pass the $N_{\text{jets}} \geq 2$ selection only if jets from QCD radiation are counted; these QCD radiation jets tend to be soft. When changing the jet countind from caloJets to JPT jets or PF jets, the true jet threshold that is applied becomes sharper. This sharpening of the jet threshold could improve the signal-to-noise.

To test this hypothesis, we repeat the analysis by counting JPT jets instead of caloJets (we keep the default MET, *i.e.*, we do not use tcMET). To be consistent with the JPT treatment, which requires muon-jet cleaning, we add to the caloJet-based analysis a requirement that counted jets be separated by ΔR >0.4 from any of the muons in the dilepton solution. The results of

Sample	caloJET counting	JPT counting
ttdil	59.7 ± 0.5	59.6 ± 0.5
Drell Yan	9.0 ± 0.6	8.3 ± 0.6
WW	0.8 ± 0.1	0.7 ± 0.1
tW	2.3 ± 0.1	2.2 ± 0.1

Table 12: Expected event yields in 10 pb^{-1} with $N_{\text{jets}} \geq 2$, where either caloJet or JPT jets are counted. All dilepton types (ee, $e\mu$, $\mu\mu$) are included. The DrellYan line includes all leptonic modes (ee, $\mu\mu$, $\tau\tau$). The test is done on the same event sample, thus the statistical errors shown are strongly correlated.

743 this test are summarized in Table 12. As anticipated, there does appear to be a modest (5-10%)
744 improvement in signal-to-noise.

12 Conclusions

We have presented expectations of observing top quark pair production in a final state with two leptons with high transverse momentum and jets using the first 10 pb^{-1} of CMS data. Clear observation of the signal is expected in the sample with two or more jets with a signal-to-noise ratio of about 4 to 1 in all channels combined and about 9 to 1 in the $e\mu$ channel alone. With the first 10 pb^{-1} of data we expect to measure the top pair production cross section with the statistical uncertainty of 14% in a plain-sum combination of all dilepton channels or with an uncertainty of 18% using $e\mu$ channel alone. Based on the limited study of systematic effects we expect the systematic uncertainty of order 10% (excluding uncertainty on integrated luminosity of the sample).

In the sample of CMS data we anticipate to use for analysis we expect to have controls over effects not necessarily predicted well by the simulation. Events with 0 and 1 jets not dominated by contributions from top-quark pair production will be used to check our expectations of background events. The lepton selection efficiencies will be measured from $Z \to \ell\ell$ events in the data. We have also developed a method to estimate backgrounds arising from the W+jets or QCD multijet processes as well as a method to estimate the contribution from the $Z/\gamma^\star$ process.

References

[1] D. Gelé, J. Speck, J. Andrea, J. Cuevas, P. Lobelle, and J. Vizán, "Measurement of the $t\bar{t}$ cross section in the dilepton channels with $\mathcal{L} = 100$ pb^{-1} using the CMS detector at $\sqrt{s} = 10$ TeV,". CMS AN-2009/047.

[2] F. Golf, D. Evans, J. Mülmenstädt, S. Padhi, F. Würthwein, A. Yagil, J. Ribnik, P. Kalavase, D. Kovalskyi, V. Krutelyov, C. Campagnari, I. Bloch, O. Gutsche, I. Fisk, K. Burkett, and L. Bauerdick, "Correcting Missing Transverse Energy Using Tracks,". CMS AN 2008-089 and 2009-022.

[3] F. Golf, D. Evans, J. Mülmenstädt, S. Padhi, F. Würthwein, A. Yagil, J. Ribnik, P. Kalavase, D. Kovalskyi, V. Krutelyov, C. Campagnari, I. Bloch, O. Gutsche, I. Fisk, K. Burkett, and L. Bauerdick, "Expectations for $t\bar{t} \to$ dileptons in the early phase of CMS,". PAS TOP-08-001 and CMS AN-2008/015.

[4] A. Giammanco, J. Caudron, and V. Lemaitre, "Track-based measurement of top-quark pair events in the dileptonic channels with the first data of CMS,". CMS AN-2009/051.

[5] **CMS** Collaboration, "Expectations for observation of top quark pair production in the dilepton final state with the early CMS data at $\sqrt{s} = 10$ TeV,". CMS TOP-09-002.

[6] R. Bainbridge, N. Cripps, A. Magnan, S. Sharma, M. Tonjes, F. Chlebana, J. Weng, V. Sordini, A. Nayak, O. Kodolova, I. Vardanyan, N. Ilina, D. Kovalskyi, and A. Nikitenko, "Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS,". CMS PAS JME-09-002 and CMS AN-2009/031.

[7] `https://twiki.cern.ch/twiki/bin/view/CMS/ProductionSummer2008`.

[8] `http://madgraph.hep.uiuc.edu/`.

[9] `http://www.ge.infn.it/~tosi/cms/topMC.html`.

[10] 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.

[11] R. K. Ellis, "An update on the next-to-leading order Monte Carlo MCFM," *Nucl. Phys. Proc. Suppl.* **160** (2006) 170–174. doi:10.1016/j.nuclphysbps.2006.09.108.

[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] J. M. Campbell and F. Tramontano, "Next-to-leading order corrections to W t production and decay," *Nucl. Phys.* **B726** (2005) 109–130, arXiv:hep-ph/0506289. doi:10.1016/j.nuclphysb.2005.08.015.

[14] 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.

[15] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuidePAT.

[16] http://lhc-commissioning.web.cern.ch/lhc-commissioning/luminosity/09-10-lumi-estimate.htm.

[17] https://twiki.cern.ch/twiki/bin/view/CMS/TSG_18_II_09.

[18] G. Daskalakis, D. Evans, C. Hill, J. Jackson, P. Vanlaer, J. Berryhill, J. Haupt, D. Futyan, C. Seez, C. Timlin, and D. Wardrope, "Measuring Electron Efficiencies at CMS with Early Data,". PAS EGM-07-001 and AN-2007-019; https://twiki.cern.ch/twiki/bin/view/CMS/MuonTagAndProbe.

[19] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIsolation.

[20] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideCutBasedElectronID.

[21] https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideEgammaIsolationIn2_1And2_2AndPlansFor3_0.

[22] The Jet and MET Corrections Group. CMS PAS JME-07-02, CMS AN-2008/003, CMS AN-2008/031, and CMS AN-2007/055.

[23] F. Golf, D. Evans, J. Mülmenstädt, S. Padhi, F. Würthwein, A. Yagil, J. Ribnik, P. Kalavase, D. Kovalskyi, V. Krutelyov, C. Campagnari, I. Bloch, O. Gutsche, I. Fisk, K. Burkett, and L. Bauerdick, "A Method to Measure the Contribution of DY$\rightarrow l^+l^-$ to a di-lepton + MET Selection,". CMS AN-2009/023.

[24] F. Golf, D. Evans, J. Mülmenstädt, S. Padhi, F. Würthwein, A. Yagil, J. Ribnik, P. Kalavase, D. Kovalskyi, V. Krutelyov, C. Campagnari, I. Bloch, O. Gutsche, I. Fisk, K. Burkett, and L. Bauerdick, "Prospects for measuring the WW production cross section in pp collisions at $\sqrt{s}$ =10 TeV,". CMS AN-2009/042.

[25] F. Golf, D. Evans, J. Mülmenstädt, S. Padhi, F. Würthwein, A. Yagil, J. Ribnik, P. Kalavase, D. Kovalskyi, V. Krutelyov, C. Campagnari, I. Bloch, O. Gutsche, I. Fisk, K. Burkett, and L. Bauerdick, "Data-driven methods to estimate the electron and muon fake contributions to lepton analyses,". CMS AN-2009/041.

826 [26] Suggested by P. Sphicas. There is probably a document somewhere.

827 [27] S. Frixione and B. R. Webber, "The MC@NLO 3.3 event generator,"
828 arXiv:hep-ph/0612272.

829 [28] The uncertainty of 50% on the tW production is also used in the single-top analysis.
830 Private communication with A. Giammanco.

831 [29] "Particle Flow Reconstruction of Jets, MET, and Taus,". Prepared by The Particle Flow
832 Physics Object Group. CMS PAS PFT-09-01 and CMS AN-2009/039.

Appendix

A Dependance of Fake Rate on Heavy Flavor Content

It has been shown that semi-leptonic and hadronic $t\bar{t}$ processes, where one or more than one of the jets fakes a lepton, are major backgrounds to the dilepton $t\bar{t}$ final state (Table 3). As the semi-leptonic and hadronic final states of $t\bar{t}$ do not have the same flavor profile as QCD, it is not certain that one can accurately predict the rate of fake leptons in $t\bar{t}$ using the FR method described in Section 8.

In Figure 11 we plot the FR for electrons and muons respectively as a function of the PDG Id of the closest status==3 (documentation line) monte carlo particle in QCD and purely hadronic $t\bar{t}$ events. In the case of electrons, the FR from QCD matched to b quarks is considerably higher than the average FR ($2.6 \pm 0.1\%$). As $t\bar{t}$ events are naturally enriched in b-quarks, one can imagine that we underestimate the electron FR in $t\bar{t}$ events. However from the same figure we see that in purely hadronic $t\bar{t}$ events, the electron FR matched to b-quarks is much lower and close to the average FR in QCD. Thus we expect good agreement between the prediction from the FR method and the actual ttothr event counts for electrons.

For muons, the FR matched to heavy flavor in QCD is higher than in hadronic $t\bar{t}$, although below the average FR derived from QCD ($10.1 \pm 1.1\%$). The FR method will therefore overpredict the number of muon fakes from ttothr. In fact, we see that the FR method does indeed overpredict the number of fake muons from semi-leptonic and hadronic ttbar events (compare Table 9 and Table 8).

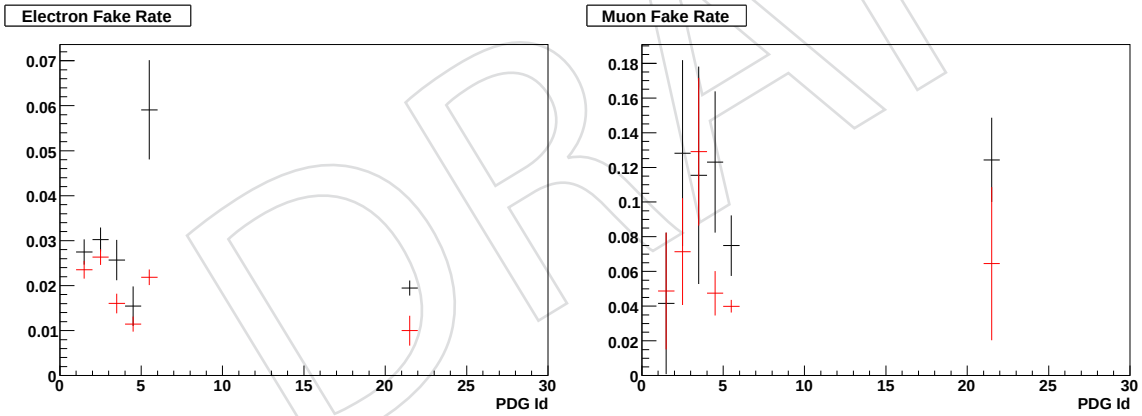


Figure 11: Fake Rate for electrons (left) and muons (right) measured in the QCD sample (black) and in pure hadronic TTbar events (red) as a function the PDG ID of the closest documentation line parton within a cone of $dR = 0.5$. The average FR in QCD is ($2.6 \pm 0.1\%$) and ($10.1 \pm 1.1\%$) for electrons and muons respectively. 1 is down quarks, 2 is up quarks, 3 is strange, 4 is charm, 5 is bottom and 21 is gluon.

For both the muons and electrons, the FR matched to heavy flavor jets in QCD is higher than for hadronic $t\bar{t}$. We postulate that the P_T spectrum of heavy flavor jets in QCD is steeply falling, and the high P_T (P_T¿20 GeV/c) lepton will carry a large fraction of the mother b-quark's momentum, resulting in a relatively isolated lepton. In the $t\bar{t}$ case, the P_T distribution of the b quark is higher, and therefore the fragmentation fraction will be lower, resulting in a less isolated lepton, and therefore a lower FR.

B Estimates of the contribution from $V + \gamma$

This Section/Appendix has been added after the pre-approval as a follow up to the questions of the ARC.

Events with a high-p_T isolated photon, where the photon converts to an electron-positron pair that contributes an electron or a positron to the dilepton pair, are not predicted by the fake-lepton method. We consider this contribution to be a part of the residual background and estimate it from the simulation. It is expected that most of these events are already simulated in the $W+$jets and DY+jets samples used in the analysis and the dominant contribution should come from the $W+$ jets events with an additional photon. As reported in Section 8, this contribution is only 10% of the total rate of $W+$ jets passing the final selections, which is negligible for the measurement of the signal cross section. The rate estimate in the $W+$ jets is based only on 13 simulated events with a generator level photon passing the final cuts. Also, it is known that the modeling of the photon radiation by PYTHIA is not precise. We perform further cross check on available higher statistics samples to confirm the conclusion that the contribution from the $V + \gamma$ events is not significant.

In the /Wgamma and /Zgamma PYTHIA samples combined we find 0.5 events expected in the $e^{\pm}\mu^{\mp}$ mode for 10 in all jet bins combined and before applying the MET cut. We obtain 0.03 events expected with at least two jets before applying the MET cut. These expected yields include $W + \gamma + (\rightarrow$ fake) together with $W + (\gamma \rightarrow$ fake). This is consistent with the 10% rate of events attributed to $W + \gamma$ in the $W+$ jets sample reported in Section 8. Based on observations in these samples we are at worst overestimating the contribution from $V + \gamma$.

In the AVJets MADGRAPH sample we find n the e^+e^- final state we find 0.7, 0.8, and 0.4 events with 0, 1, and more than 2 jets respectively. Upon closer inspection, approximately 60% of these events have an electron matching in dR to a high-$p_T$generator level photon whose mother is reported to be W, Z, a quark, or a gluon (the latter is likely a feature of the MADGRAPH generator). There are approximately 1.2 events expected in all jet bins combined (corresponding to 91 simulated events) that look like having a hard-scattering photon reconstructed as an electron. Out of these 1.2 events, 0.6, 0.4, and 0.18 events are with 0, 1, and at least 2 jets respectively. In the $e^{\pm}\mu^{\mp}$ final state we find 0.9, 0.7, and 0.3 events passing the main selections with 0, 1, and at least 2 jets respectively. After the electron is matched to a high-p_T genp-photon we find 0.6, 0.6, and 0.13 events with 0, 1, and at least two jets. Most of these eventes are $W(\rightarrow \mu\nu) + \gamma$, where the photon is substantially separated from the muon. In the same AVJets sample in the $\mu^+\mu^-$ mode we find 0.03, 0.3, and 0.6 events with 0, 1, and at least two jets. No events give a match of the muon to a genp-photon, which partially validates the parent matching logic.

Based on the AVJets sample, we conclude that the WJets underestimates the number of events where the hard-scattering photon is reconstructed as an electron by about a factor of 3 to 4 if the predictions from the AVJets sample are treated as is. Statistically, this conclusion is based on 13 simuilated events in the WJets sample and on about 200 simulated events in the AVJets sample. It is not yet clear that all the generator level settings are correct in the AVJets sample. According to the GEN configuration file, the ISR and FSR usually created by PYTHIA after the hard scattering is the same as the default configuration. This suggests that a fraction of the $V + \gamma$ events in the AVJets sample are double-counted inside the sample itself. Assuming this double-count rate is equal to the rate of events with a photon in the WJets sample, it is more likely that WJets itself does not underestimate the rate of $W + \gamma$ events by more than a factor of 2. With the (potential) double-counting suptracted we estimate that the contribution from $V + \gamma$ events, where the photon is reconstructed as an electron, is 0.1 ± 0.1 (0.2 ± 0.2) events with at least two jets in the $e^{\pm}\mu^{\mp}$ (e^+e^-) final states passing the final event selection. A 100%

uncertainty is assigned to the mean values. No contribution is expected in the $\mu^+\mu^-$ final state. In all of the final states in events with at least two jets the contribution from the $V + \gamma$ events is found to be negligibly small compared to the dominant backgrounds or the expected uncertainty of the signal cross section measurement.