The Compact Muon Solenoid Experiment
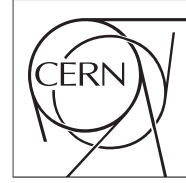
Analysis Note

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

Search for Direct Stop Quark Pair Production in the Single Lepton Channel with the Full 2011 Dataset

L. Bauerdick, K. Burkett, I. Fisk, Y. Gao, O. Gutsche, B. Hooberman, S. Jindariani, J. Linacre, V. Martinez Outschoorn

Fermilab National Accelerator Laboratory, Batavia, USA

D. Barge, C. Campagnari, D. Kovalskyi, V. Krutelyov

University of California, Santa Barbara, Santa Barbara, USA

W. Andrews, G. Cerati, D. Evans, F. Golf, I. MacNeill, S. Padhi, Y. Tu, F. Würthwein, A. Yagil, J. Yoo

University of California, San Diego, San Diego, USA

Abstract

This note describes a search for direct stop quark pair production in the single lepton channel using 4.98 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV taken with the CMS detector in 2011. A search for an excess of events over the Standard Model prediction is performed in a sample with a single isolated electron or muon, several jets, missing transverse energy and large transverse mass.

Contents

1   Introduction

This note presents a search for the production of supersymmetric (SUSY) stop quark pairs in events with a single isolated lepton, several jets, missing transverse energy, and large transverse mass. We use the full 2011 data sample, corresponding to an integrated luminosity of 4.98 fb^{-1}. This search is of theoretical interest because of the critical role played by the stop quark in solving the hierarchy problem in SUSY models. This solution requires that the stop quark be light, less than a few hundred GeV and hence within reach for direct pair production. We focus on two decay modes $\tilde{t} \to t\chi_1^0$ and $\tilde{t} \to b\chi_1^+$ which are expected to have large branching fractions if they are kinematically accessible, leading to:

- $pp \to \tilde{t}\bar{\tilde{t}} \to t\bar{t}\chi_1^0\chi_1^0$, and

- $pp \to \tilde{t}\bar{\tilde{t}} \to b\bar{b}\chi_1^+\chi_1^- \to b\bar{b}W^+W^-\chi_1^0\chi_1^0$.

Both of these signatures contain high transverse momentum (p_T) jets including two b-jets, and missing transverse energy ($E_\mathrm{T}^{\mathrm{miss}}$) due to the invisible χ_1^0 lightest SUSY particles (LSP's). In addition, the presence of two W bosons leads to a large branching fraction to the single lepton final state. Hence we require the presence of exactly one isolated, high p_T electron or muon, which provides significant suppression of several backgrounds that are present in the all-hadronic channel. The largest backgrounds for this signature are semi-leptonic $t\bar{t}$ and W+jets. These backgrounds contain a single leptonically-decaying W boson, and the transverse mass (m_T) of the lepton-neutrino system has a kinematic endpoint requiring $m_\mathrm{T} < M_W$. For signal stop quark events, the presence additional LSP's in the final states allows the m_T to exceed M_W. Hence we search for an excess of events with large m_T. The dominant background in this kinematic region is dilepton $t\bar{t}$ where one of the leptons is not identified, since the presence of two neutrinos from leptonically-decaying W bosons allows the m_T to exceed M_W. Backgrounds are estimated from Monte Carlo (MC) simulation, with careful validation and determination of scale factors and corresponding uncertainties based on data control samples.

The expected stop quark pair production cross section (see Fig. 1) varies between O(10) pb for $m_{\tilde{t}} = 200$ GeV and O(0.01) pb for $m_{\tilde{t}} = 500$ GeV. The critical challenge of this analysis is due to the fact that for light stop quarks with a mass close to the top quark, the production cross section is large but the kinematic distributions, in particular m_T, are very similar to SM $t\bar{t}$ production, such that it becomes very difficult to distinguish the signal and background. For large stop quark mass the kinematic distributions differ from those in SM $t\bar{t}$ production, but the cross section decreases rapidly, reducing the signal-to-background ratio.
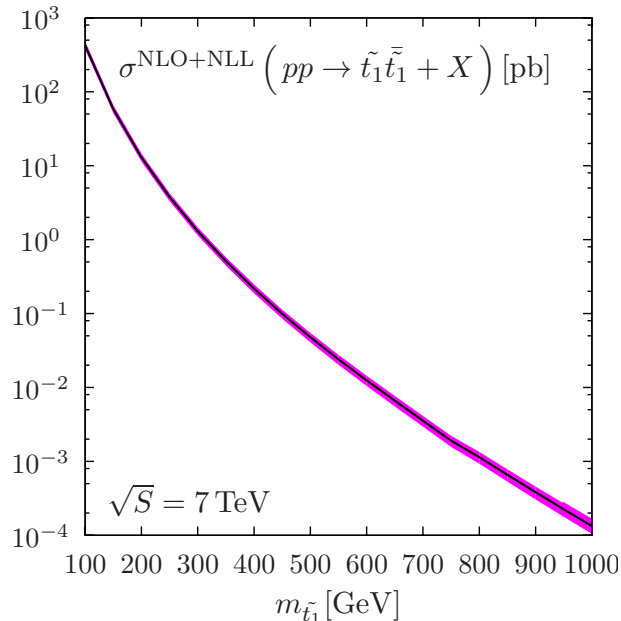

Figure 1:   The stop quark pair production cross section in pb, as a function of the stop quark mass.

2 Overview and Analysis Strategy

We are searching for a $t\bar{t}\chi^0\chi^0$ or $W\ell b W\bar{\ell}\bar{b}\chi^0\chi^0$ final state (after top decay in the first mode, the final states are actually the same). So to first order this is "$t\bar{t}$+ extra $E_{\mathrm{T}}^{\mathrm{miss}}$".

We work in the $\ell+$ jets final state, where the main background is $t\bar{t}$. We look for $E_{\mathrm{T}}^{\mathrm{miss}}$ inconsistent with $W \to \ell\nu$. We do this by concentrating on the $\ell\nu$ transverse mass (M_T), since except for resolution effects, $M_T < M_W$ for $W \to \ell\nu$. Thus, the initial analysis is simply a counting experiment in the tail of the M_T distribution.

The event selection is one-and-only-one high P_T isolated lepton, four or more jets, and some moderate $E_{\mathrm{T}}^{\mathrm{miss}}$ cut. At least one of the jets has to be btagged to reduce $W+$ jets. The event sample is then dominated by $t\bar{t}$, but there are also contributions from $W+$ jets, single top, dibosons, etc.

In order to be sensitive to $\widetilde{t\bar{t}}$ production, the background in the M_T tail has to be controlled at the level of 10% or better. So this is (almost) a precision measurement.

The $t\bar{t}$ events in the M_T tail can be broken up into two categories: (i) $t\bar{t} \to \ell+$ jets and (ii) $t\bar{t} \to \ell^+\ell^-$ where one of the two leptons is not found by the second-lepton-veto (here the second lepton can be a hadronically decaying τ). For a reasonable M_T cut, say $M_T > 150$ GeV, the dilepton background is of order 80% of the total. This is because in dileptons there are two neutrinos from W decay, thus M_T is not bounded by M_W. This is a very important point: while it is true that we are looking in the tail of M_T, the bulk of the background events end up there not because of some exotic $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction failure, but because of well understood physics processes. This means that the background estimate can be taken from Monte Carlo (MC), after carefully accounting for possible data/MC differences. Sophisticated fully "data driven" techniques are not really needed.

Another important point is that in order to minimize systematic uncertainties, the MC background predictions are always normalized to the bulk of the $t\bar{t}$ data, ie, events passing all of the requirements but with $M_T \approx 80$ GeV. This removes uncertainties due to $\sigma(t\bar{t})$, lepton ID, trigger efficiency, luminosity, etc.

The $\ell+$ jets background, which is dominated by $t\bar{t} \to \ell+$ jets, but also includes some $W+$ jets as well as single top, is estimated as follows:

1. We select a control sample of events passing all cuts, but anti-btagged. This is sample is now dominated by $W+$ jets. The sample is used to understand the M_T tail in $\ell+$ jets processes.

2. In MC we measure the ratio of the number of $\ell+$ jets events in the M_T tail to the number of events with $M_T \approx 80$ GeV. This ratio turns out to be pretty much the same for all sources of $\ell+$ jets.

3. In data we measure the same ratio but after correcting for the $t\bar{t} \to$ dilepton contribution, as well as dibosons etc. The dilepton contribution is taken from MC after the correction described below.

4. We compare the two ratios, as well as the shapes of the data and MC M_T distributions. If they do not agree, we try to figure out why and fix it. If they agree well enough, we define a data MC scale factor (SF) which is the ratio of the ratios defined in step 2 and 3, keeping track of the uncertainty.

5. We next perform the full selection in $t\bar{t} \to \ell+$ jets MC, and measure this ratio again (which should be the same as that in step 2).

6. We perform the full selection in data. We count the events with $M_T \approx 80$ GeV, we subtract off the dilepton contribution, we multiply the subtracted event count by the ratio from step 5 (or from step 2), and also by the data/MC SF from step 4. The result is the prediction for the $\ell+$ jets BG in the M_T tail.

The dilepton background can be broken up into many components depending on the characteristics of the 2nd (undetected) lepton

- 3-prong hadronic tau decay

- 1-prong hadronic tau decay

- e or μ possibly from τ decay

78 We have currently no veto against 3-prong taus. For the other two categories, we explicitly veto events
79 with additional electrons and muons above 10 GeV , and we veto events with an isolated track of $P_T > 10$
80 GeV. This also rejects 1-prong taus (it turns out that the explicit e or μ veto is redundant with the isolated
81 track veto). Therefore the latter two categories can be broken into

82 • out of acceptance ($|\eta| > 2.50$)

83 • $P_T < 10$ GeV

84 • $P_T > 10$ GeV, but survives the additional lepton/track isolation veto

85 Monte Carlo studies indicate that there is no dominant contribution: it is "a little bit of this, and a little
86 bit of that".

87 The high M_T dilepton backgrounds come from MC, but their rate is normalized to the $M_T \approx 80$ GeV
88 peak. In other to perform this normalization in data, the $W+$ jets events in the M_T peak have to be
89 subtracted off. This introduces a systematic uncertainty.

90 There are two types of effects that can influence the MC dilepton prediction: physics effects and instru-
91 mental effects. We discuss these next, starting from physics.

92 First of all, many of our $t\bar{t}$ MC samples (eg: MadGraph) have BR($W \to \ell\nu$) $= \frac{1}{9} = 0.1111$. PDG says
93 BR($W \to \ell\nu$) $= 0.1080 \pm 0.0009$. This difference matter, so the $t\bar{t}$ MC must be corrected to account for
94 this.

95 Second, our selection is $\ell+$ 4 or more jets. A dilepton event passes the selection only if there are two
96 additional jet from ISR, or one jet from ISR and one jet which is reconstructed from the unidentified
97 lepton, $e.g.$, a three-prong tau. Therefore, all MC dilepton $t\bar{t}$ samples used in the analysis must have
98 their jet multiplicity corrected (if necessary) to agree with what is seen in $t\bar{t}$ data. We use a data control
99 sample of well identified dilepton events with E_T^{miss} and at least two jets as a template to "adjust" the
100 N_{jet} distribution of the $t\bar{t} \to$ dileptons MC samples.

101 The final physics effect has to do with the modeling of $t\bar{t}$ production and decay. Different MC models
102 could in principle result in different BG predictions. Therefore we use several different $t\bar{t}$ MC samples
103 using different generators and dfferent parameters, to test the stability of the dilepton BG prediction.
104 All these predictions **after** corrections for branching ratio and N_{jet} dependence, are compared to each
105 other. The spread is a measure of the systematic uncertainty associated with the $t\bar{t}$ generator modeling.

106 The main instrumental effect is associated with the underefficiency of the 2nd lepton veto. We use tag-
107 and-probe to compare the isolated track veto performance in $Z + 4$ jet data and MC, and we extract
108 corrections if necessary. Note that the performance of the isolated track veto is not exactly the same
109 on e/μ and on one prong hadronic tau decays. This is because the pions from one-prong taus are often
110 accompanied by π^0's that can then result in extra tracks due to phton conversions. We let the simulation
111 take care of that. Similarly, at the moment we also let the simulation take care of the possibility of a
112 hadronic tau "disappearing" in the detector due to nuclear interaction of the pion.

113 The sample of events failing the last isolated track veto is an important control sample to check that we
114 are doing the right thing.

115 Note that JES uncertainties are effectively "calibrated away" by the N_{jet} rescaling described above.

116 Finally, there are possible improvements to this basic analysis strategy that can be added in the future:

117 • Move from counting experiment to shape analysis. But first, we need to get the counting experiment
118 under control.

119 • Add an explicit three prong tau veto

120 • Do something to require that three of the jets in the event be consistent with $t \to Wb, W \to q\bar{q}$.
121 This could help rejecting some of the dilepton BG; however, it would also loose efficiency for the
122 $\widetilde{t} \to b\chi^+$ mode

123 • Consider the $M(\ell b)$ variable, which is not bounded by M_{top} in $\widetilde{t} \to b\chi^+$

4

3 Data Samples

4 Event Selection

5 Background Estimation

In order to search for a possible signal from stop decays giving rise to a signature of $t\bar{t}$ with additional E_T^{miss} from the LSPs, it is necessary to determine the composition of the SM backgrouns in the signal region. This section details the methods pursued to estimate the background in the signal sample and describes the procedure to estimate the systematic uncertainties. The general strategy is to use the MC prediction for the backgrounds after applying corrections derived from data.

The most important background to a stop signal arises from SM $t\bar{t}$. The $t\bar{t}$ background may be separated into contributions containing a single lepton $t\bar{t} \rightarrow \ell + \text{jets}$ and two leptons $t\bar{t} \rightarrow \ell\ell$. As described in this section, the $t\bar{t} \rightarrow \ell\ell$ background is the dominant process satisfying the event selection, contributing $\sim 80\%$ of the signal sample defined with the benchmark selection of E_T^{miss} 100 GeV and m_T 150 GeV. This background has large true E_T^{miss} and consequently larger m_T due to the presence of two neutrinos. Additional contributions to the single lepton sample arise from W+jets and single top. The combination of all single lepton backgrounds, $t\bar{t} \rightarrow \ell + \text{jets}$, W+jets and single top, comprises $\sim 15\%$ of the signal sample. Finally, other background sources such as dibosons, $Z/\gamma^* + \text{jets}$, in addition to rarer processes such as $t\bar{t}$ produced in association with a vector boson and tribosons, provide a combined contribution to the signal sample at the level of $\sim 5\%$. Finally, the QCD background contribution is small, particularly in the signal sample, with a large E_T^{miss} requirement.

5.1 Single Lepton Backgrounds

This category of backgrounds includes processes with a single leptonic W decay, giving rise to one lepton and E_T^{miss} from a single neutrino. As a result, the m_T variable, constructed from the lepton and the E_T^{miss}, exhibits a kinematic edge at $m_T \sim m_W$. The main contributors to this background are $t\bar{t} \rightarrow \ell + \text{jets}$, W+jets and single top, though in the latter case there is a contribution from tW that can give rise to two leptons. As shown in Figure. 2 (left), these backgrounds exhibit a similar m_T shape and are thus combined into a single background estimate.

It should be underlined that single lepton events entering the signal sample are in the far m_T tail for these processes and contribute mainly due to the limited E_T^{miss} resolution that smears the m_T peak, particularly in the presence of multiple jets. Since these types of effects are challenging to model in simulation, the background estimate for this category of processes is cross checked in a control sample in data. The control sample is obtained by applying the full selection criteria with the exception of the b-tagging requirement, that is reversed, vetoing events with b-tagged jets. The b-tag veto greatly reduces the contamination from $t\bar{t}$, which is particularly important in the case of $t\bar{t} \rightarrow \ell\ell$ which otherwise populates the m_T tail. The resulting sample is dominated by W+jets events. The derivation of the background estimate in this control sample serves to validate the method. In addition, the level of agreement between the prediction and the data in the m_T tail provides an estimate of the systematic uncertainty for this background prediction.

The single lepton background contribution to the tail of the m_T is derived from MC and normalized to the data in the m_T peak region, defined by the range $60 < m_T \le 100$ GeV. In particular, the total yield for the combination of the single lepton + jets samples in MC satisfying $60 < m_T \le 100$ GeV is scaled to match the entries in data satisfying the same m_T requirement. The data is corrected for the expected contamination from non-single lepton + jets processes, which is obtained from MC. The derivation of this MC to data scale factor is shown in Figure. 2 (right), where the contribution from non-single lepton + jets processes in the m_T peak is also shown to be small. Scaling to the data yield in the m_T peak largely reduces the dependence on the $t\bar{t}$ cross section and cancels systematic uncertainties associated with effects such as the luminosity, selection efficiencies, etc...

Figure. 3 shows a comparison of the m_T distribution in data and MC in the b-veto control sample for two values of the E_T^{miss} cut after scaling the single lepton contribution to the m_T peak region. The simulation models the m_T distribution in data reasonably well. Table. 1 shows a comparison of the single lepton + jets prediction and the data in the signal region defined as $m_T > 150$ GeV. The prediction and the data are in agreement within the statistical uncertainty, which is 30% for the $E_T^{\mathrm{miss}} > 100$ GeV case.
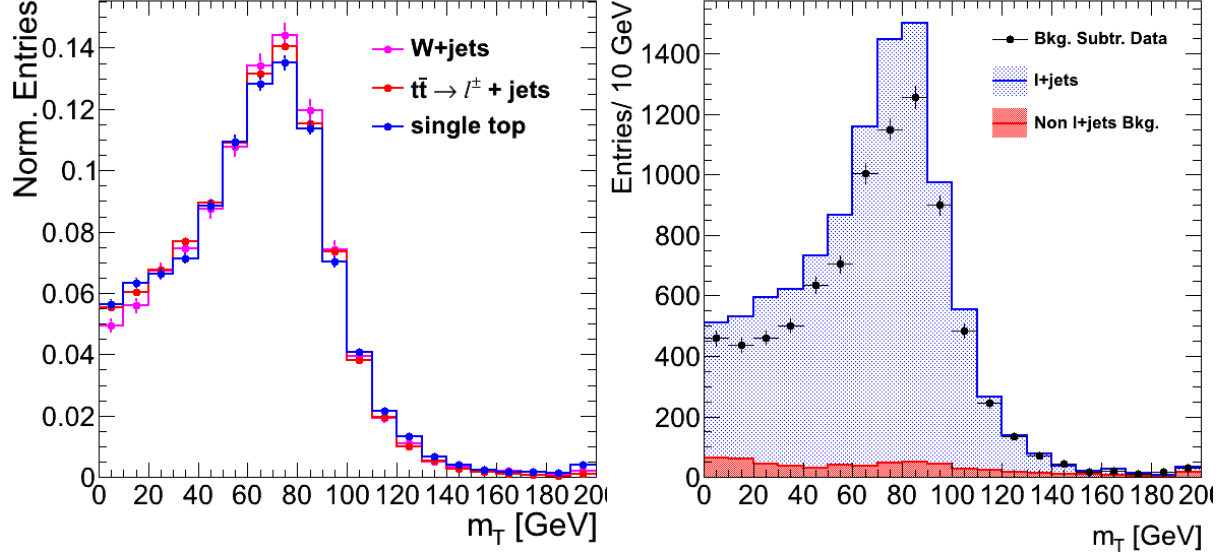
Figure 2: Monte Carlo comparison of the shapes (left) of the m_{T} distribution for the main background processes containing a single leptonic W decay (left). The m_{T} shape for $t\bar{t} \to \ell + \mathrm{jets}$, W+jets and single top are similar and thus combined into a single background estimate. Comparison of the MC prediction (right) for single lepton processes (blue) and the data after subtracting the non-single lepton background MC prediction (also shown in red). The m_{T} distribution in MC is scaled to the data in the m_{T} peak region ($60 - 100$ GeV). Since the control sample is dominated by $W+ \geq 4\mathrm{jets}$ events, the MC is not expected to provide an estimate of the overall normalization to better than the $\sim 15\%$ difference observed.

Sample	$E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV
Lepton + Jets Prediction	88 ± 12	38 ± 8
Data in Signal Region	106 ± 13	40 ± 8
Data/MC Closure	1.20 ± 0.22	1.03 ± 0.30

Table 1: Summary of closure test in b-veto control sample for two values of the $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement. The closure serves to estimate systematic uncertainty for the prediction of the single lepton + jets background.

No correction is necessary given the agreement observed and the uncertainty serves as an estimate of the systematic uncertainty on the lepton + jets background. It may be noted that the 30% uncertainty applies to the lepton + jets background, which constitues 15% of the sample. The resulting uncertainty on the total background is 5%.

Given the similarities in the shapes of the m_{T} distributions for the various samples, the resulting dependence on the sample composition is small. Figure. 4 shows the m_{T} distribution for the single lepton + jets components in MC. The distribution on the right shows the impact of varying the relative sample composition on the total single lepton + jets prediction. In this study, each component is varied independently by up to 20% and the corresponding change on the prediction is at the level of a few percent and therefore negligible compared to the statistical uncertainty of the closure test. It should be noted that despite differences in the m_{T} shape for the various single lepton + jets processes, expected for example due to differences in the p_{T} of the W, this test provides a check for the impact of reconstruction effects, such as the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, that are the dominant contributors to the tail of the m_{T} and similar for the $t\bar{t} \to \ell + \mathrm{jets}$ and W+jets processes.
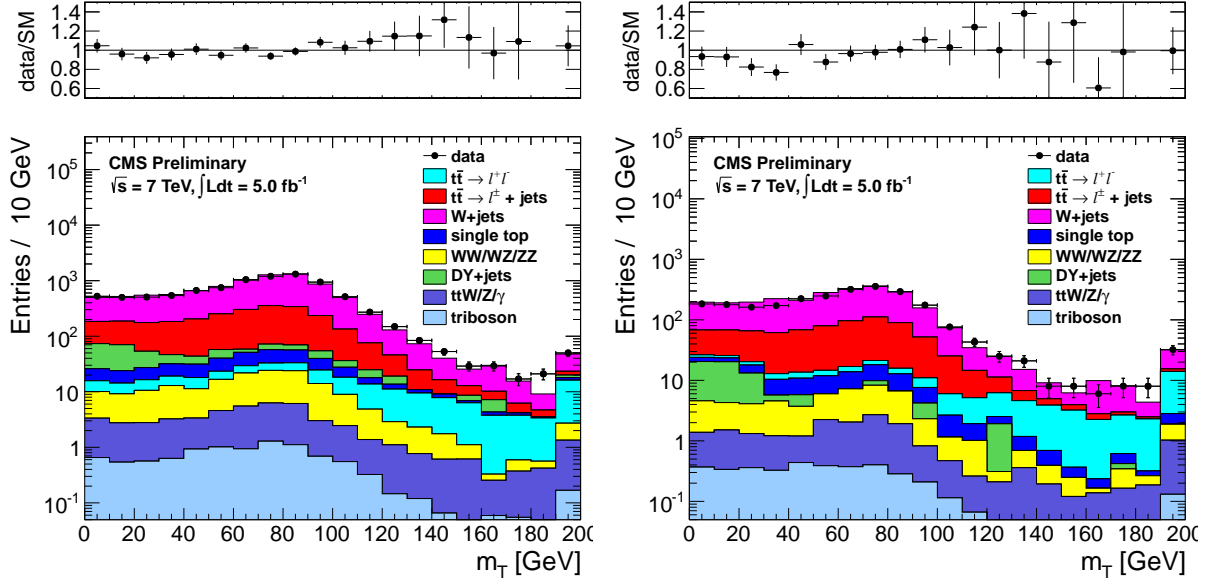
6

Figure 3: Comparison of data and MC in the b-veto sample after scaling the single lepton samples in the m_T peak region $(60 - 100$ GeV) for two $E_\mathrm{T}^\mathrm{miss}$ requirements $E_\mathrm{T}^\mathrm{miss} > 50$ GeV (left) and $E_\mathrm{T}^\mathrm{miss} > 100$ GeV (right). The simulation shows reasonable agreement with the data both in the peak and the m_T tail, as can also be seen the ratios.
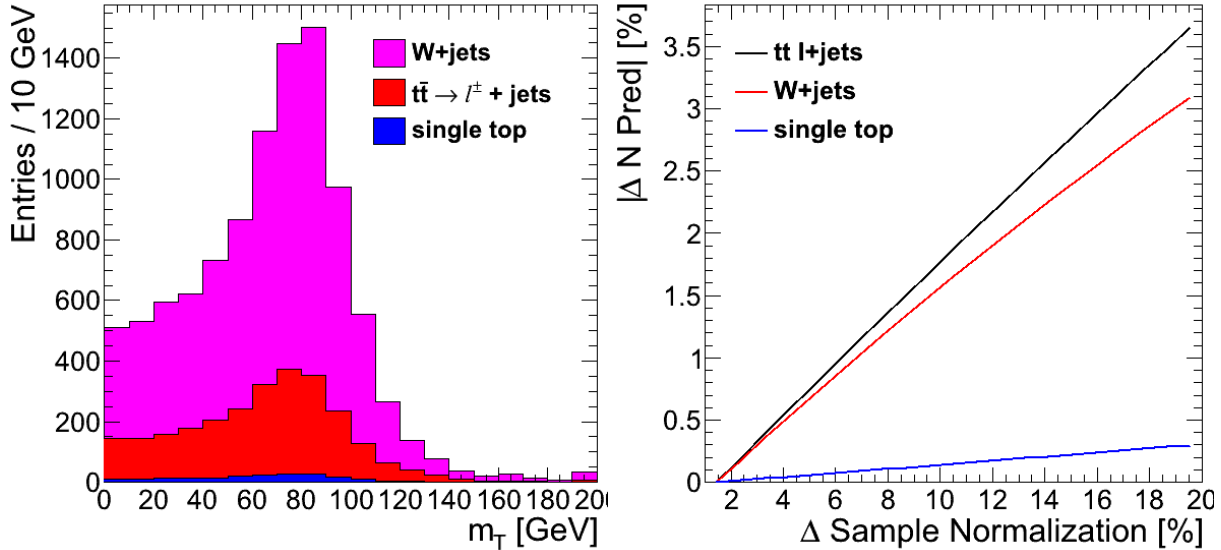

Figure 4: Stacked plot (left) showing the relative sample composition in MC for the main single lepton + jets components, $t\bar{t} \to \ell + $ jets, W+jets and single top sample for the b-veto control selection. Note no scalings are applied and the $E_\mathrm{T}^\mathrm{miss} > 50$ GeV is shown. Percent change in the total single lepton + jets background prediction for variations in the normalization of each single lepton + jets background component (right). Each sample is varied independently and the full background prediction is performed.

7

5.2 Top Dilepton Background

The dominant background to the signal sample comes from $t\bar{t} \to \ell\ell$ events. Due to the presence of a second neutrino, $t\bar{t} \to \ell\ell$ events do not have a kinematic edge at $m_T \sim m_W$. These events satisfy the selection criteria due to real E_T^{miss} and do not depend on detector resolution or E_T^{miss} mis-measurement effects. As a result, the $t\bar{t} \to \ell\ell$ background is expected to be well modeled in the MC. The prediction for this background is thus derived from MC and normalized to the data in the m_T peak region. However, there are two aspects that require dedicated corrections and detailed in this section

- the modeling of additional jets from radiation, required to satisfy the 4-jet selection criteria.

- the modeling of the isolated track veto efficiency, which is applied to explicitly reject leptons from W and $W \to \tau$ decays and single prong τ decays.

The systematic uncertainty associated with the MC prediction is derived by comparing various generators and from the uncertainties on the various correction factors used. These are described at the end of this section.

5.3 Modeling of Additional Hard Jets in Top Dilepton Events

Dilepton $t\bar{t}$ events have 2 jets from the top decays, so additional jets from radiation or higher order contributions are required to enter the signal sample. The modeling of addtional jets in $t\bar{t}$ events is checked in a a $t\bar{t} \to \ell\ell$ control sample, selected by requiring

- exactly 2 selected electrons or muons with $p_T > 20$ GeV

- $E_T^{miss} > 50$ GeV

- ≥ 1 b-tagged jet

Figure. 5 shows a comparison of the jet multiplicity distribution in data and MC for this two-lepton control sample. After requiring at least 1 b-tagged jet, most of the events have 2 jets, as expected from the dominant process $t\bar{t} \to \ell\ell$. There is also a significant fraction of events with additional jets. The 3-jet sample is mainly comprised of $t\bar{t}$ events with 1 additional emission and similarly the ≥ 4-jet sample contains primarily $t\bar{t} + \geq 2$ jet events. Even though the primary $t\bar{t}$ Madgraph sample used includes up to 3 additional partons at the Matrix Element level, which are intended to describe additional hard jets, Figure. 5 shows a slight mis-modeling of the additional jets.

It should be noted that in the case of $t\bar{t} \to \ell\ell$ events with a single reconstructed lepton, the other lepton may be mis-reconstructed as a jet. For example, a hadronic tau may be mis-identified as a jet (since no τ identification is used). In this case only 1 additional jet from radiation may suffice for a $t\bar{t} \to \ell\ell$ event to enter the signal sample. As a result, both the samples with $t\bar{t} + 1$ jet and $t\bar{t} + \geq 2$ jets are relevant for the signal sample.

Jet Multiplicity Sample	Data/MC Scale Factor
N jets = 3 (sensitive to $t\bar{t} + 1$ extra jet from radiation)	0.92 ± 0.03
N jets ≥ 4 (sensitive to $t\bar{t} + \geq 2$ extra jets from radiation)	0.83 ± 0.04

Table 2: Data/MC scale factors used to account for differences in the fraction of events with addtional hard jets from radiation in $t\bar{t} \to \ell\ell$ events.

Table. 2 shows scale factors to correct the fraction of events with additional jets in MC to the observed fraction in data. These are applied to the $t\bar{t} \to \ell\ell$ MC in the signal sample. In order to do so, it is first necessary to count the number of additional jets from radiation and exclude leptons mis-identified as jets. A jet is considered a mis-identified lepton if is matched to a generator-level second lepton with sufficient energy to satisfy the jet p_T requirement ($p_T > 30$ GeV).

In the signal sample, leptons mis-identified as jets are not rare. Figure. 6 shows the MC jet multiplicity distribution for $t\bar{t} \to \ell\ell$ events satisfying the full selection criteria before and after subtracting
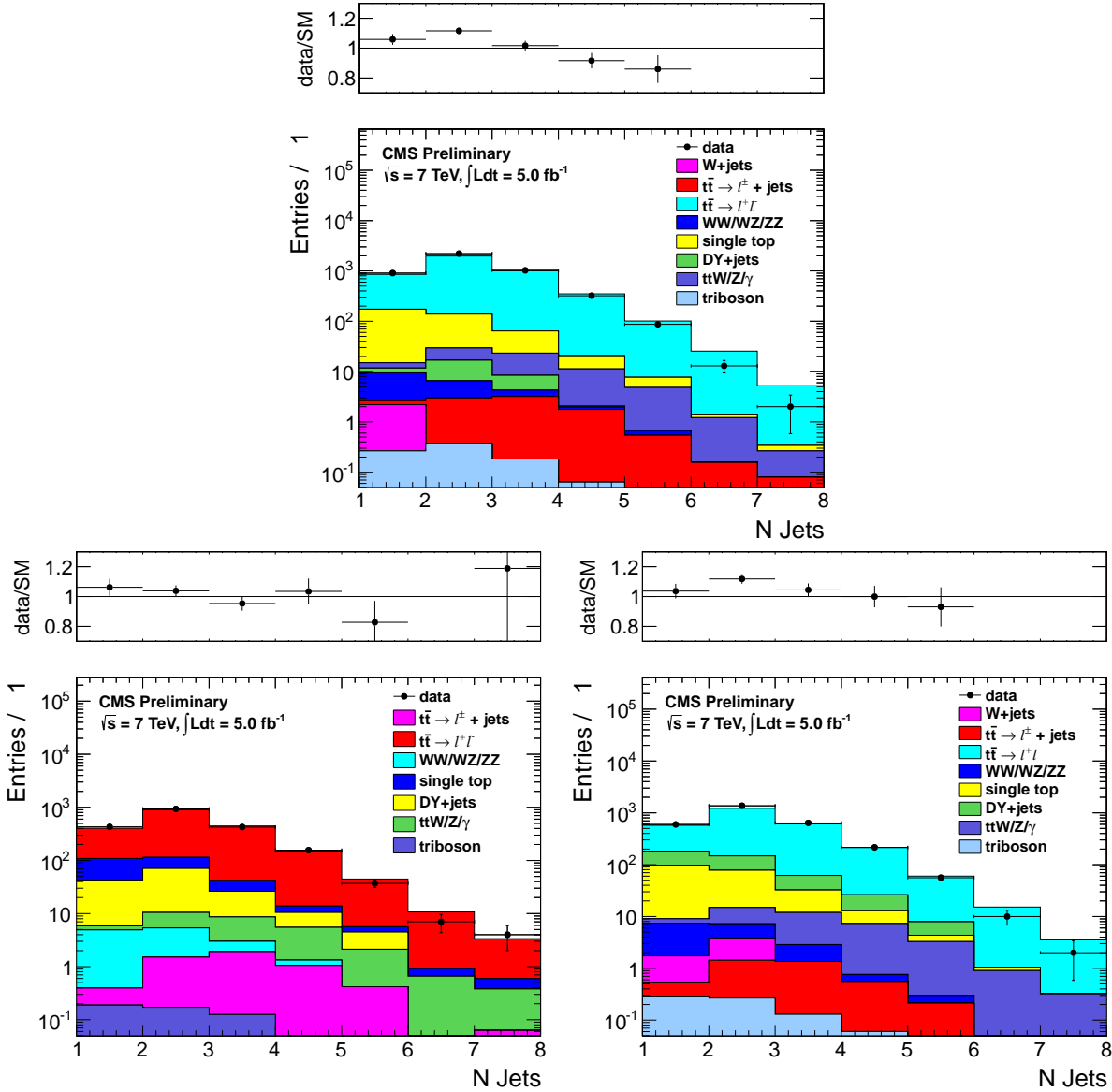
Figure 5: Comparison of the jet multiplicity distribution in data and MC for dilepton events in the e-μ (top), e-e (bottom left) and μ-μ (bottom right) channels.

leptons mis-identified as jets. Approximately a quarter of the sample is comprised of 4-jet events that actually correspond to a 2-lepton + 3 jet event where the second lepton is counted as a jet. Lepton mis-identification depends strongly on the type of second lepton, occuring more frequently in the case of hadronic τs than leptonic objects. According to the $t\bar{t} \to \ell\ell$ MC, for hadronic τs, $\sim 85\%$ of multi-prong τs and about half single-prong τ are mis-identified as jets. In the case of leptonic objects, the fractions are smaller, comprising about a third of e/μ from a W decay and $< 20\%$ for leptonic τs, mainly because of the softness of the decay products. The scale factors listed in Table. 2 are applied to the "cleaned" jet counts in the signal sample (shown in blue in Figure. 6). The impact of applying the scale factors on the $t\bar{t} \to \ell\ell$ is about a 10% reduction in the prediction for the signal sample.

Ultimately, the interesting quantity for reweighting is the number of additional hard jets from radiation and this information is accessed using the number of reconstructed jets. Figure. 7 compares the number of additional jets from truth matching probed by N reconstructed jets, in this case 3 and ≥ 4 jets. In order to do so, jets that are truth-matched to the top decay products (the b-quarks and additional leptons) are removed. The 3-jet distribution shows excellent agreement and the differences in the $\geq$ 4-jet distribution are of at most 5%. The impact of possible differences in the underlying distribution of extra jets between
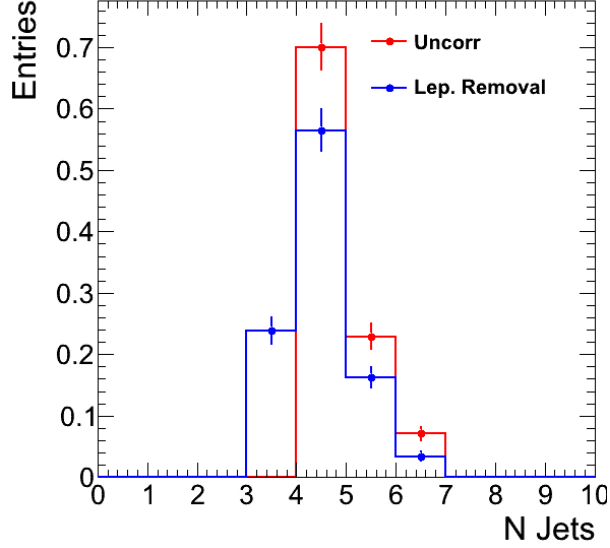
9

Figure 6: Comparison of the jet multiplicity distribution for $t\bar{t} \to \ell\ell$ events in MC in the signal sample before (red) and after (blue) applying the lepton-jet overlap removal. Note only the first 6 jets are shown.

the signal and control $t\bar{t} \to \ell\ell$ samples is estimated by varying the scale factor contributions by 10% and calculating the change in the dilepton prediction. This effect is found to have a negligible impact on the prediction, well below 1%.

Other effects that have been examined include the impact of additional jets from pileup that may bias the jet multiplicity distribution, which is found to be a negligible effect in this dataset. The impact of the non-$t\bar{t} \to \ell\ell$ background on the jet fraction scale factors has also been studied. In particular, given the large uncertainty on the $Z/\gamma^* + HF$ MC prediction, this component has been varied by a factor of 2 and the resulting change on the dilepton prediction is $< 1\%$. As a result, the dominant source of uncertainty is the statiscal uncertainty, primarily from the two-lepton control sample size, that corresponds to a 3% uncertainty on the dilepton prediction.

The scale factors for the fraction of additional jets in the dilepton sample are applied throughout the analysis. It may be noted that this scaling is also performed consistently for the alternative $t\bar{t}$ samples, always reweighting the jet multiplicity distribution to the data in the $t\bar{t} \to \ell\ell$ control sample. In this way, effects truly arising from using different MC samples and settings can be examined, separately from issues related to the modeling of additional jets.

5.4 Normalization of the Top Prediction

The overall normalization of the $t\bar{t}$ sample is determined by scaling to the m_{T} peak control region, following a procedure similar to that described in Section. 5.1. This control region is dominated by $t\bar{t}$ albeit in its single lepton decay mode. The basic idea is that after adjusting the modeling of additional jets from radiation in the $t\bar{t} \to \ell\ell$ sample and correcting the leptonic branching fractions in the $t\bar{t}$ sample, the MC prediction for the $t\bar{t} \to \ell + \mathrm{jets}$ and $t\bar{t} \to \ell\ell$ samples is subject to the same sources of uncertainty: the $t\bar{t}$ cross section, the luminosity, the selection efficiencies, etc. . . The exception is the veto on events containing an isolated track, since this last requirement has a different impact on the $t\bar{t} \to \ell + \mathrm{jets}$ and $t\bar{t} \to \ell\ell$ samples. The impact of this requirement is addressed separately in Section. 5.5.

The m_{T}-peak scale factor is thus determined after applying the full analysis selection with the exception of the isolated track veto. Specifically, the pre-veto sample is defined by the following requirements

- At least 1 selected electron (muon) with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.5$ ($|\eta| < 2.1$)

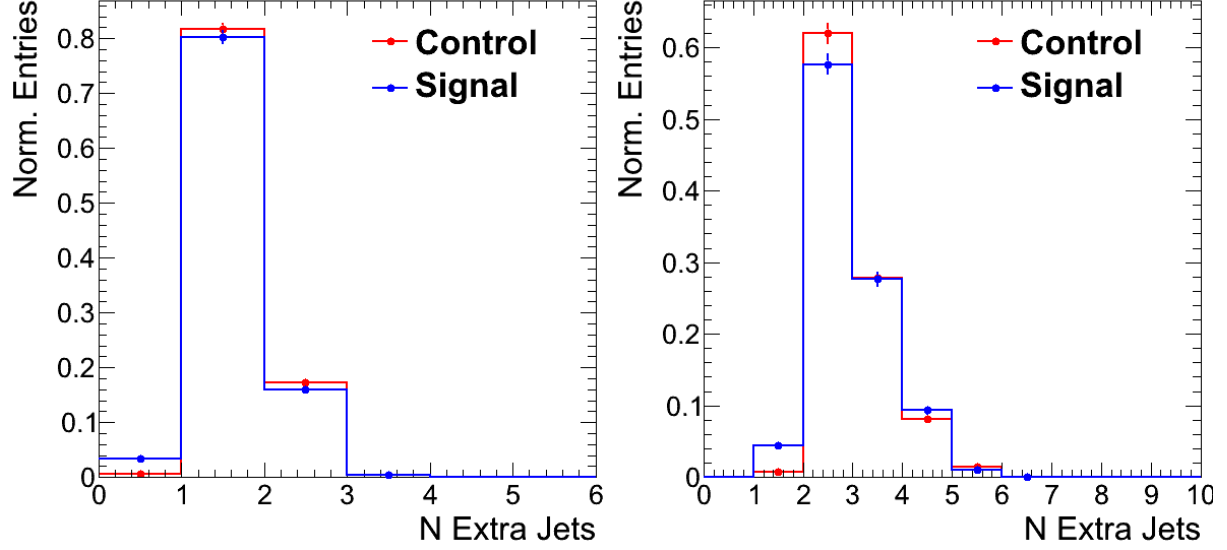- At least 4 selected jets, of which at least 1 is b-tagged

10

Figure 7: Comparison of the number of additional jets from radiation in the 3-jet (left) and $\geq$ 4-jet (right) bins for the control $t\bar{t} \to \ell\ell$ sample (with two reconstructed leptons) and the signal sample (with one reconstructed lepton). The distributions show good agreement, indicating that the usage of the reconstructed jet multiplicity in one sample to reweight the signal sample is indeed justified.

271   • $E_T^{\mathrm{miss}} > 50$ GeV

272 As in the case of the single lepton + jets sample, scaling the overall normalization to the m_T peak
273 largely reduces the dependence on the $t\bar{t}$ cross section and cancels systematic uncertainties associated
274 with effects such as the luminosity, selection efficiencies, etc... However, the m_T peak control sample is
275 contaminated by non-$t\bar{t}$ processes, particularly W+jets that contributes at the $5-10\%$ level, even after
276 requiring a b-tagged jet. The W+jets+HF process is a particular concern given the large theoretical
277 uncertainties associated with their production. Therefore a systematic uncertainty is derived to account
278 for the uncertainty in this background component. The normalization of the W+jets sample is scaled
279 up and down by 50% and the full background estimate recomputed. The impact of such a variation on
280 the dilepton sample is of 5%, as shown in Table. 3. However, since the single lepton + jets prediction
281 decreases because of the large variation in the W+jets component, the impact on the total background
282 prediction is significantly smaller, at the level of $\sim 1\%$.

283 It should be noted that the background prediction obtained using the same normalization scale factor

Sample	Nominal	W+Jets x1.5	W+Jets x0.5
$t\bar{t} \to l^+l^-$	85.0 ± 3.8 (4 %)	81.4 ± 3.7	89.0 ± 3.8
$t\bar{t} \to l^\pm +$ jets	12.3 ± 0.6 (5 %)	11.7 ± 0.6	12.8 ± 0.6
W+jets	7.6 ± 3.9 (51 %)	10.9 ± 5.5	4.0 ± 2.0
single top	5.7 ± 0.6 (11 %)	5.4 ± 0.6	5.9 ± 0.7
WW/WZ/ZZ	0.5 ± 0.2 (33 %)	0.5 ± 0.2	0.6 ± 0.2
ttW/Z/γ	5.6 ± 0.4 (7 %)	5.4 ± 0.4	5.9 ± 0.4
triboson	0.1 ± 0.0 (40 %)	0.1 ± 0.0	0.1 ± 0.0
Total	116.8 ± 5.5 (5 %)	115.5 ± 6.7 (6 %)	118.2 ± 4.4 (4 %)
Non-tt dilepton	31.8 ± 4.0 (12 %)	34.0 ± 5.6 (16 %)	29.3 ± 2.2 (8 %)
Data/MC ctr SF	0.90 ± 0.03	0.86 ± 0.03	0.94 ± 0.03
MC sig/ctr SF	0.079 ± 0.003	0.078 ± 0.004	0.080 ± 0.002

Table 3: Comparison of the nominal background predictions and those under the assumption that the W+jets normalization is 50% higher and 50% lower than the nominal. NUMBERS TO BE UPDATED

based on the m_T peak for the single lepton and dilepton samples is subject to smaller uncertainties than the prediction obtained by normalizing the $t\bar{t} \to \ell\ell$ component alone to the overall yield in the two-lepton control sample. The reason is that the normalization of the two-lepton control yield depends on effects that do not impact the $t\bar{t} \to \ell\ell$ sample with one reconstructed lepton (i.e. the signal sample of interest) in the same way. Example of these effects include the dilepton trigger and second lepton reconstruction efficiencies.

In conclusion, the pre-veto sample is used to define an overall data over MC scale factor (SF^{all}) in the m_T peak control region, that is applied to all background predictions and is simply defined as

- $N_{\mathrm{peak}}^{\mathrm{all}}$ = data yield in the peak region $60 < m_\mathrm{T} < 100$ GeV

- $M_{\mathrm{peak}}^{\mathrm{all}}$ = MC yield in the peak region $60 < m_\mathrm{T} < 100$ GeV

- $SF^{\mathrm{all}} = N_{\mathrm{peak}}^{\mathrm{all}}/M_{\mathrm{peak}}^{\mathrm{all}}$

For all subsequent steps, the scale factor SF^{all} is applied to all MC contributions.

5.5 The Isolated Track Veto

The $t\bar{t} \to \ell\ell$ background is further suppressed after the 4-jet requirement by removing events with any non-isolated track with $p_\mathrm{T} > 10$ GeV. The isolated track veto rejects events with an e or a μ, as well as single-prong τ-decays. This veto is very effective at reducing the dilepton background. In particular, according to the $t\bar{t} \to \ell\ell$ MC, the veto removes about three-quarters of events with an e or μ from the W decay and almost half the leptonic and single prong τ decays. The veto has no impact on multi-prong τs, though this is a smaller component overall. Since the $t\bar{t} \to \ell\ell$ background includes different types of processes, it is useful to first characterize the composition of this background.

5.5.1 Top Dilepton Sample Composition

The $t\bar{t} \to \ell\ell$ background may be categorized based on the type of second lepton, as shown in Table. 4. The main component is from electrons and muons from a W decay or through an intermediate τ decay. The second largest component arises from single-prong hadronic τ decays, followed by multi-prong τs. Finally an additional contribution arises from leptons falling in the forward region, outside the Tracker acceptance $|\eta| > 2.5$ (refered to as 'lost').

Sample	Yield	Fraction [%]
$t\bar{t} \to l^+l^-$ (lost)	7 ± 1	6
$t\bar{t} \to l^+l^-$ (e/μ)	30 ± 3	26
$t\bar{t} \to l^+l^-$ (τ_{lep})	21 ± 2	18
$t\bar{t} \to l^+l^-$ ($\tau_{\mathrm{had}} \to 1-\mathrm{prong}$)	39 ± 3	34
$t\bar{t} \to l^+l^-$ ($\tau_{\mathrm{had}} \to 3-\mathrm{prong}$)	19 ± 2	16
total $t\bar{t} \to l^+l^-$	117 ± 5	100

Table 4: Dilepton events satisfying the full selection criteria separated by decay modes.

The isolated track veto does not apply to the components where the second lepton falls outside the acceptance or where it decays to a hadronic tau that is not explicitly rejected. For the cases where the second lepton includes an electron or muon or a charged π/K, it is possible to further distinguish cases when the relevant particle targeted by the veto is below the p_T threshold. Matching the truth-level particle to reconstructed tracks shows that in $t\bar{t} \to \ell\ell$ MC

- for $t\bar{t} \to l^+l^-$ (e/μ), about a third of the sample falls below the p_T threshold of the track veto, and the remaining two thirds fail the isolation

- for $t\bar{t} \to l^+l^-$ (τ_{lep}), about 80% are soft $p_\mathrm{T} < 10$ GeV and about 20% are non-isolated

- for $t\bar{t} \to l^+l^-$ ($\tau_{\mathrm{had}} \to 1-\mathrm{prong}$), about 70% are soft, as expected from a τ decay product and the rest fail the isolation criteria.

In summary, the combination of these fractions with the relative sample composition listed in Table. 4 shows that only about a third of the $t\bar{t} \rightarrow \ell\ell$ background sample is affected by the track veto[1]. The performance of the isolation used in the track veto requirement is the subject of the next section.

It should also be noted that according to the MC, track reconstruction inefficiencies affect a few percent ($\sim 1-2\%$) of the leptonic and single prong τ events. The tracking efficiency in this analysis is taken from MC, which is expected to provide good modeling of isolated tracks with $p_{\mathrm{T}} > 10$ GeV. The impact of possible differences between data and MC is found to be negligible. In particular, the case of single-prong taus is the most challenging to model due to the effect of nuclear interactions in the tracker material. Past studies of the tracking efficiency for pions [6] provide a data/MC uncertainty in the tracking efficiency of 3.9%[2]. Propagating this uncertainty to the total background estimate yields a total uncertainty of $< 0.5\%$. The reason is that the tracking efficiency uncertainty only applies to single prong τ decays with $p_{\mathrm{T}} > 10$ GeV, which are under 30% of the dilepton component, which in turn is $\sim 80\%$ of the total sample.

To conclude, the $t\bar{t} \rightarrow \ell\ell$ background arises from events where the second lepton falls outside the acceptance (both in η and p_{T}), because the event contains a hadronic tau that is not explicitly rejected or because the second lepton fails the isolation requirement. Even though the $t\bar{t} \rightarrow \ell\ell$ sample is quite heterogenous and comprises multiple types of second lepton events, there are two main sources of uncertainty in this estimate:

- Acceptance effects, which are estimated by using alternative MC samples

- Detector effects, mainly arising from understanding the performance of the isolated track veto, which impacts only about a third of the total $t\bar{t} \rightarrow \ell\ell$ sample.

5.5.2 Performance of the Isolation Requirement

The last requirement used in the analysis is an isolated track veto. This selection criteria rejects events containing a track of $p_{\mathrm{T}} > 10$ GeV with relative track isolation $\sum p_{\mathrm{T}}/p_{\mathrm{T}}(trk)$ in a cone of size $R = 0.3 < 0.1$. It may be noted that only tracks consistent with the vertex with highest $\sum p_{\mathrm{T}}^2$ are considered in order to reduce the impact of spurious tracks, for example from pileup interactions. This requirement has very good performance. Figure. 8 shows the efficiency for rejecting dilepton events compared to the efficiency for selecting single lepton events for various cone sizes and cut values. The chosen working point provides a signal efficiency of $\epsilon(sig) = 92\%$ for a background rejection of $\epsilon(bkg) = 53\%$ in MC.

It should be emphasized that the isolated track veto has a different impact on the samples with a single lepton (mainly $t\bar{t} \rightarrow \ell + \mathrm{jets}$ and $W+\mathrm{jets}$) and that with two leptons (mainly $t\bar{t} \rightarrow \ell\ell$). For the dilepton background, the veto rejects events which have a genuine second lepton. Thus the performance may be understood as an efficiency $\epsilon(trk)$ to identify the isolated track. In the case of the single lepton background, the veto rejects events which do not have a genuine second lepton, but rather which contain a "fake" isolated track. The isolated track veto thus effectively scales the single lepton sample by (1-FR), where FR is the "fake rate" to identify an isolated track in events which contain no genuine second lepton. It is thus necessary to study the isolated track efficiency $\epsilon(trk)$ and the "fake rate" FR in order to fully characterize the veto performance.

The veto efficiency for dilepton events is calculated using the tag and probe method in Z events. A good lepton satisfying the full ID and isolation criteria and matched to a trigger object serves as the tag. The probe is defined as a track with $p_{\mathrm{T}} > 10$ GeV that has opposite charge to the tag and has an invariant mass with the probe consistent with the Z mass.

The variable used to study the performance of the veto is the absolute track isolation, since it removes the dependence of the isolation variable on the p_{T} of the object under consideration. This is particularly useful because the underlying p_{T} distribution is different for second leptons in $t\bar{t} \rightarrow \ell\ell$ events compared

---

[1] Explicitly, the fraction of events that give rise to a sufficiently energetic lepton or single prong τ is: 70% of e-μ events which are 26% of the sample, 20% of leptonic tau events which are 18% of the sample and 30% of single prong τ events which are 10% of the sample.

[2] This tracking efficiency uncertainty estimate is conservative for this analysis since it includes tracks of p_{T} down to 250 MeV, where material effects are larger and so are the corresponding uncertainties.
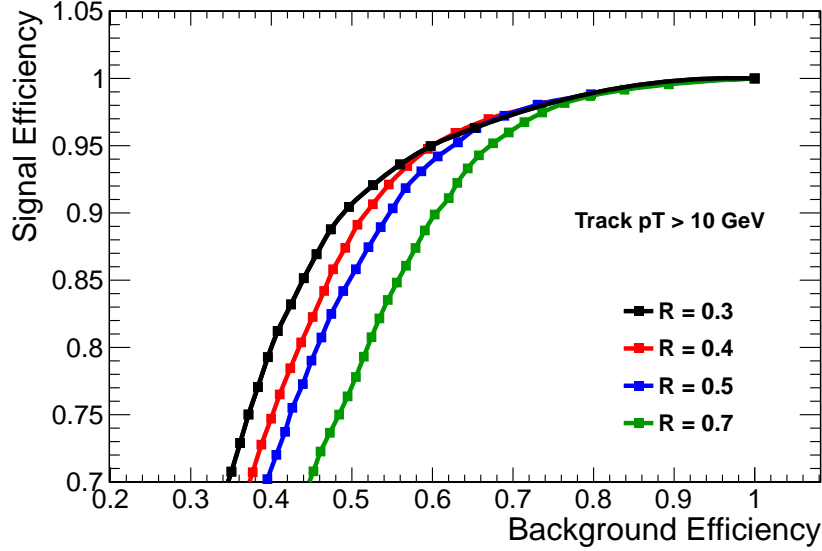
Figure 8: Comparison of the performance in terms of signal (single lepton events) efficiency and background (dilepton events) rejection for various cone sizes and cut values. The current isolation requirement uses a cone of size $\Delta R = 0.3$ and a cut value of 0.1, corresponding to $\epsilon(sig) = 92\%$ for $\epsilon(bkg) = 53\%$. ADD ARROW OR LINE TO INDICATE WORKING POINT.

to Z events, particularly due to the presence of τs that have softer decay products. As shown in Figure. 9, the absolute isolation is consistent between $Z+4$ jet events and $t\bar{t} \rightarrow \ell\ell$ events, including leptons from W and τ decays. This supports the notion that the isolation, defined as the energy surrounding the object under consideration, depends only on the environment of the object and not on the object itself. The isolation is thus sensitive to the ambient pileup and jet activity in the event, which is uncorrelated with the lepton p_{T}. It is thus justified to use tag and probe in $Z+4$ jet events, where the event topologies are similar to $t\bar{t} \rightarrow \ell\ell$ events, in order to estimate the performance of the isolation requirement for the various leptonic categories of $t\bar{t} \rightarrow \ell\ell$ events.

It may be noted that tracks from single prong τ decays are slightly less isolated compared to electrons and muons. The reason is that single prong τs can have π^0 associated with the single charged track. These decay into γs that in turn convert $\gamma \rightarrow e^+e^-$ and spoil the isolation. As also shown in Figure. 9, the isolation distribution for charged tracks from τ decays that are not produced in association with π^0s are consistent with that from es and μs. Since events from single prong τ decays produced in association with π^0s comprise a small fraction of the total sample, the isolation measured for leptons is used for all single prong τ events. A systematic uncertainty is assigned to account for the difference in the underlying isolation distribution for this sample.

BEGIN SECTION TO WRITE OUT In detail, the procedure to correct the dilepton background is:

- Using tag-and-probe studies, we plot the distribution of **absolute** track isolation for identified probe electrons and muons **TODO: need to compare the e vs. μ track iso distributions, they might differ due to e$\rightarrow$eγ.**

- We verify that the distribution of absolute track isolation does not depend on the p_{T} of the probe lepton. This is due to the fact that this isolation is from ambient PU and jet activity in the event, which is uncorrelated with the lepton p_{T}**TODO: verify this in data and MC..**

- Our requirement is **relative** track isolation < 0.1. For a given $t\bar{t} \rightarrow \ell\ell$ MC event, we determine the $p_{\mathrm{T}}$of the 2nd lepton and translate this to find the corresponding requirement on the **absolute** track isolation, which is simply $0.1 \times p_{\mathrm{T}}$.

- We measure the efficiency to satisfy this requirement in data and MC, and define a scale-factor $SF_{\epsilon(trk)}$ which is the ratio of the data-to-MC efficiencies. This scale-factor is applied to the $t\bar{t} \rightarrow \ell\ell$ MC event.
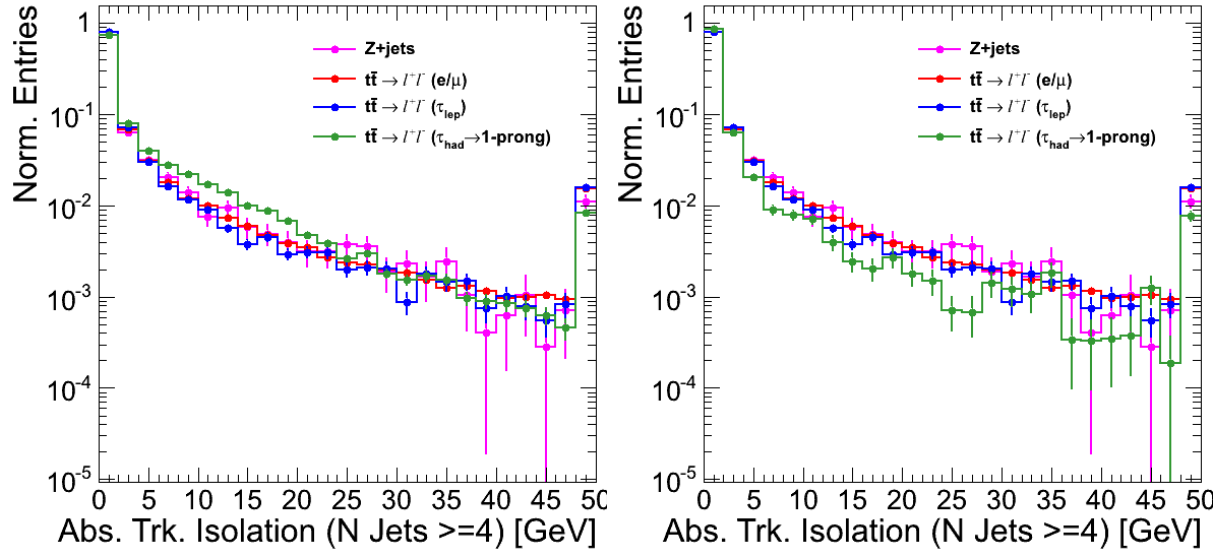
14

Figure 9: Comparison of absolute track isolation for track probes in $Z + 4$ jet and $t\bar{t} \to \ell\ell$ events for different lepton types. The isolation variables agree across samples, except for single prong τs, that tend to be slightly less isolated (left). The agreement across isolation distributions is recovered after removing single prong τ events produced in association with π^0s from the sample (right).

- **THING 2 we are unsure about: we can measure this SF for electrons and for muons, but we can't measure it for hadronic tracks from τ decays. Verena has showed that the absolute track isolation distribution in hadronic τ tracks is harder due to $\pi^0 \to \gamma\gamma$ with $\gamma \to e^+e^-$.**

END SECTION TO WRITE OUT

A measurement of the FR in data is non-trivial. However, it is possible to correct for differences in the FR between data and MC by applying an additional scale factor for the single lepton background alone, using the sample in the m_T peak region. This scale factor is determined after applying the isolated track veto and after subtracting the $t\bar{t} \to \ell\ell$ component, corrected for the isolation efficiency derived previously. As shown in Figure. 10, the efficiency for selecting an isolated track in single lepton events is independent of m_T so the use of an overall scale factor is justified to estimate the contribution in the m_T tail.

6   Summary of the Background Estimation Procedure

The SM background in the signal region, defined by requirements of large E_T^{miss} and m_T, is estimated using MC. The MC is validated using data control samples, which are used to derive data-to-MC scale factors and corresponding uncertainties.

The procedure to estimate the background prediction may be summarized as

- Apply the state-of-the-art corrections to the MC, reflecting the best knowlege of the detector performance, in order to improve the agreement with the data. This includes effects such as the modeling of the pileup, the jet energy scale, E_T^{miss} corrections, etc...

- Correct the leptonic branching fractions in the $t\bar{t}$ MC

- Use the dilepton sample with two selected leptons to reweight the N_{jets} distribution in $t\bar{t} \to \ell\ell$ MC, which is not necessarily well-modeled due to the presence of additional jets from radiation.

- Use the pre-veto sample (i.e. applying the full analysis selection with the exception of the isolated track veto) to define a scale factor in the m_T peak region. This scale factor corrects for effects of integrated luminosity, $t\bar{t}$ cross section, lepton selection and trigger efficiencies.
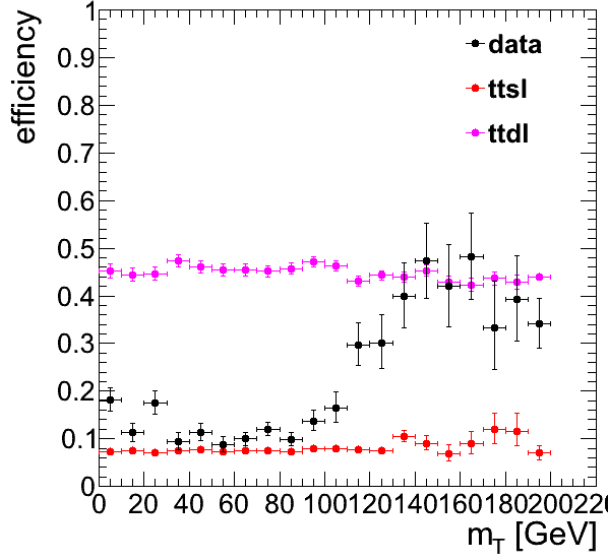
15

Figure 10: Efficiency for selecting an isolated track comparing single lepton $t\bar{t} \to \ell + \text{jets}$ and dilepton $t\bar{t} \to \ell\ell$ events in MC and data as a function of m_T. The efficiencies in $t\bar{t} \to \ell + \text{jets}$ and $t\bar{t} \to \ell\ell$ exhibit no dependence on m_T while the data ranges between the two. This behavior is expected since the low m_T region is predominantly $t\bar{t} \to \ell + \text{jets}$, while the high m_T region contains mostly $t\bar{t} \to \ell\ell$ events.

- Correct the $t\bar{t} \to \ell\ell$ sample for differences between data and MC in the isolation for events with a second lepton. This correction is derived using $Z + 4$ jet events and applied to the $t\bar{t} \to \ell\ell$ sample.

- In the signal sample, after applying the full selection including the isolated track veto, derive a scale factor to account for possible data vs. MC discrepancies in the isolated track fake rate for backgrounds which have a single genuine lepton. This scale factor is applied to the single lepton backgrounds only.

6.1 Other Backgrounds

7 Systematic Uncertainties

7.1 Uncertainty on the $t\bar{t} \to \ell\ell$ Acceptance

The $t\bar{t}$ background prediction is obtained from MC, with corrections derived from control samples in data. The uncertainty associated with the theoretical modeling of the $t\bar{t}$ production and decay is estimated by comparing the background predictions obtained using alternative MC samples. It should be noted that the full analysis is performed with the alternative samples under consideration, including the derivation of the various data-to-MC scale factors. The variations considered are

- Top mass: The alternative values for the top mass differ from the central value by 5 GeV: $m_\text{top} = 178.5$ GeV and $m_\text{top} = 166.5$ GeV.

- Jet-parton matching scale: This corresponds to variations in the scale at which the Matrix Element partons from Madgraph are matched to Parton Shower partons from Pythia. The nominal value is $x_q > 20$ GeV. The alternative values used are $x_q > 10$ GeV and $x_q > 40$ GeV.

- Renormalization and factorization scale: The alternative samples correspond to variations in the scale $\times 2$ and $\times 0.5$. The nominal value for the scale used is $Q^2 = m_\text{top}^2 + \sum_\text{jets} p_\text{T}^2$.

- Alternative generators: Samples produced with different generators include MC@NLO and Powheg (NLO generators) and Pythia (LO). It may also be noted that MC@NLO uses Herwig6 for the hadronisation, while POWHEG uses Pythia6.

16

443 • Modeling of taus: The alternative sample does not include Tauola and is otherwise identical to the
444 Powheg sample.

445 • The PDF uncertainty is estimated following the PDF4LHC recommendations[CITE]. The events
446 are reweighted using alternative PDF sets for CT10 and MSTW2008 and the uncertainties for each
447 are derived using the alternative eigenvector variations and the "master equation". In addition,
448 the NNPDF2.1 set with 100 replicas. The central value is determined from the mean and the
449 uncertainty is derived from the 1σ range. The overall uncertainty is derived from the envelope of
450 the alternative predictions and their uncertainties.

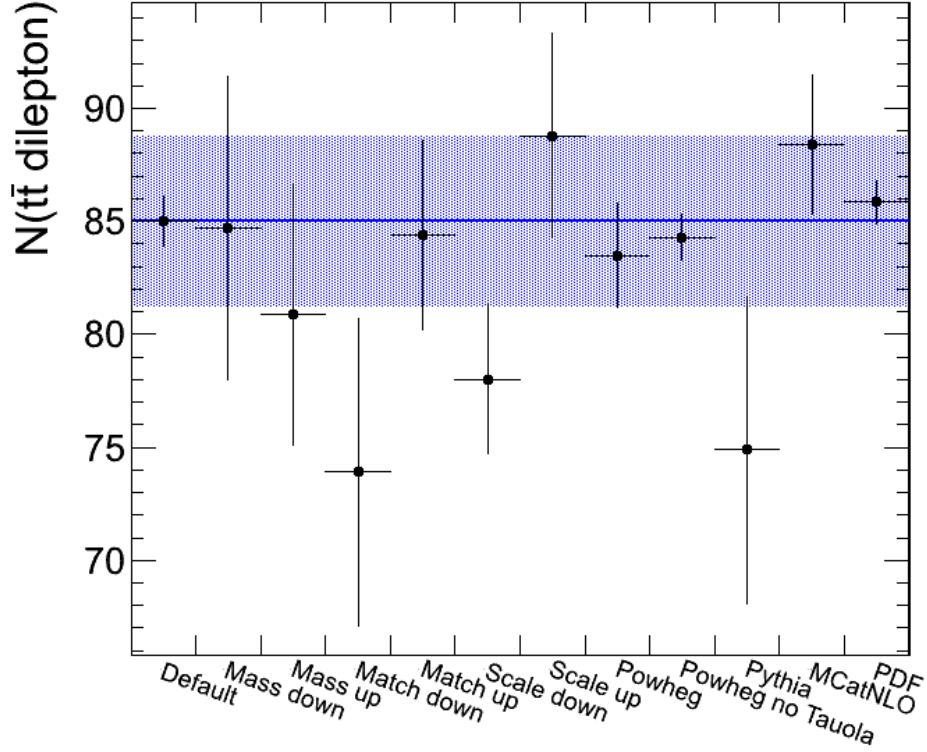

Figure 11: Comparison of the $t\bar{t} \to \ell\ell$ central prediction with those using alternative MC samples. The blue band corresponds to the total statistical error for all data and MC samples. The alternative sample predictions are indicated by the datapoints. The uncertainties on the alternative predictions correspond to the uncorrelated statistical uncertainty from the size of the alternative sample only.

8   Results

9   Conclusion

References

[1] https://hypernews.cern.ch/HyperNews/CMS/get/SUS-12-007/32.html

[2] arXiv:1204.3774v1 [hep-ex]

[3] D. Barge, CMS AN-2011/464

[4] CMS Collaboration, *Search for supersymmetry in events with a Z boson, jets and momentum imbalance* PAS SUS-11-021

[5] CMS Collaboration, *"Measurement of the b-tagging efficiency using* $t\bar{t}$ *events"*, PAS BTV-11-003, in preparation.

[6] CMS Collaboration, *Measurement of Tracking Efciency*, PAS TRK-10-00

[7] CMS Collaboration, *Performance of b-jet identification in CMS*, PAS BTV-11-001

[8] M. Narain for BTV POG, `https://indico.cern.ch/getFile.py/access?contribId=0&resId=1&materialId=slides&confId=163892`

[9] arXiv:1103.1348v1, D. Barge *at al.*, CMS AN-CMS2011/269.

[10] V. Pavlunin, Phys. Rev. **D81**, 035005 (2010).

[11] V. Pavlunin, CMS AN-2009/125

[12] A reference to the top paper, once it is submitted. Also D. Barge *et al.*, AN-CMS2010/258.

[13] Changes to the selection for the 38x CMSSW release are given in `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopDileptonRefAnalysis2010Pass5`.

[14] https://twiki.cern.ch/twiki/bin/viewauth/CMS/SimpleCutBasedEleID

[15] https://twiki.cern.ch/twiki/bin/viewauth/CMS/EgammaWorkingPointsv3

[16] D. Barge *at al.*, AN-CMS2009/159.

[17] B. Mangano *et al.*, AN-CMS2010/283.

[18] https://twiki.cern.ch/twiki/bin/viewauth/CMS/CrossSections_3XSeries, https://twiki.cern.ch/twiki/bin/view/CMS/ProductionSpring2011

[19] CMS Collaboration, "Measurement of CMS luminosity", *CMS-PAS* **EWK-10-004** (2010).

[20] D. Barge *at al.*, AN-CMS2009/130.

[21] W. Andrews *et al.*, AN-CMS2009/023.

[22] D. Barge *at al.*, AN-CMS2010/257.

[23] L. Bauerdick *at al.*, AN-CMS2011/155.

[24] CMS-PAS-JME-10-010.

[25] arXiv:1103.6083v1, J. T. Ruderman, D. Shih

[26] H. Haber, G. Kane, Phys. Reports 117, Nos. 2-4 (1985) 75-263.

[27] http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/UserCode/SusyAnalysis/SLHAFILES/TChiwz/

[28] http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/UserCode/SusyAnalysis/SLHAFILES/TChizz/