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_T^{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 $M_T < M_W$. 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.
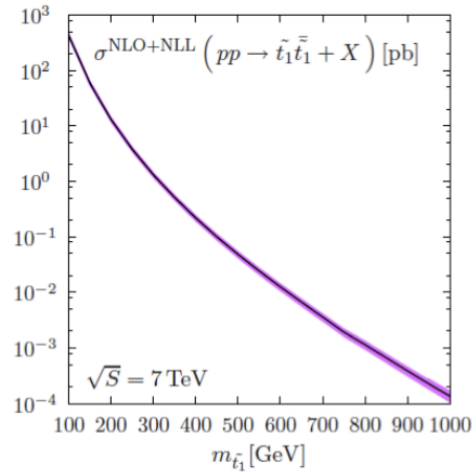

Figure 1: The stop quark pair production cross section in pb, as a function of the stop quark mass.

2 Overview and Analysis Strategy

We are searching for a $t\bar{t}\chi^0\chi^0$ or $W\ell b W\ell\bar{b}\chi^0\chi^0$ 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 $\not{E}_T$".

We work in the $\ell+$ jets final state, where the main background is $t\bar{t}$. We look for $\not{E}_T$ 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 $\not{E}_T$ 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 $\not{E}_T$ 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 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

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

- out of acceptance ($|\eta| > 2.50$)

4

77   • $P_T < 10$ GeV

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

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

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

83 There are two types of effects that can influence the MC dilepton prediction: physics effects and instrumental
84 effects. We discuss these next, starting from physics.

85 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 BR$(W \to$
86 $\ell\nu) = 0.1080 \pm 0.0009$. This difference matter, so the $t\bar{t}$ MC must be corrected to account for this.

87 Second, our selection is $\ell+$ 4 or more jets. A dilepton event passes the selection only if there are two additional
88 jet from ISR, or one jet from ISR and one jet which is reconstructed from the unidentified lepton, *e.g.*, a three-
89 prong tau. Therefore, all MC dilepton $t\bar{t}$ samples used in the analysis must have their jet multiplicity corrected (if
90 necessary) to agree with what is seen in $t\bar{t}$ data. We use a data control sample of well identified dilepton events
91 with $\displaystyle{\not}E_T$ and at least two jets as a template to "adjust" the N_{jet} distribution of the $t\bar{t} \to$ dileptons MC samples.

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

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

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

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

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

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

110   • Add an explicit three prong tau veto

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

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

3 Background Estimation Technique

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. We consider three samples:

- Dilepton sample (exactly 2 selected leptons): used to correct the N_{jets} distribution in $t\bar{t} \to \ell\ell$ MC, which is not necessarily well-modelled since ISR jets are needed to satisfy the $N_{jets} \geq 4$ requirement defining the signal region;

- Inclusive sample (at least 1 selected lepton): used to define a scale factor which corrects for effects of integrated luminosity, $t\bar{t}$ cross section, jet energy scale and jet selection efficiencies, lepton selection and trigger efficiencies; **THING 1 we are uncertain about: should this sample be fully inclusive and require AT LEAST 1 lepton, or should we veto events with a 2nd lepton and remove only the isolated track veto**.

- Signal sample (exactly 1 selected lepton): this is the sample used to define the signal region. In addition, this sample is used to determine a scale factor accounting for possible data vs. MC discrepancies in the isolated track fake rate for backgrounds which have only 1 genuine lepton.

3.1 Step 1: Use dilepton control sample to correct the N_{jets} distributon in $t\bar{t} \to \ell\ell$ MC

The dilepton control sample is defined by the following requirements:

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

- $E_T^{miss} > 50$ GeV

- ≥ 1 b-tagged jet

This sample is dominated by $t\bar{t} \to \ell\ell$. The distribution of N_{jets} for data and MC passing this selection is displayed in Fig. **??**. We use this distribution to derive scale factors which reweight the $t\bar{t} \to \ell\ell$ MC N_{jets} distribution to match the data. We define the following quantities

- N_2 = data yield minus non-dilepton $t\bar{t}$ MC yield for $N_{jets} \leq 2$

- N_3 = data yield minus non-dilepton $t\bar{t}$ MC yield for $N_{jets} = 3$

- N_4 = data yield minus non-dilepton $t\bar{t}$ MC yield for $N_{jets} \geq 4$

- M_2 = dilepton $t\bar{t}$ MC yield for $N_{jets} \leq 2$

- M_3 = dilepton $t\bar{t}$ MC yield for $N_{jets} = 3$

- M_4 = dilepton $t\bar{t}$ MC yield for $N_{jets} \geq 4$

We use these yields to define 3 scale factors, which quantify the data/MC ratio in the 3 N_{jets} bins:

- $SF_2 = N_2/M_2$

- $SF_3 = N_3/M_3$

- $SF_4 = N_4/M_4$

And finally, we define the scale factors K_3 and K_4:

- $K_3 = SF_3/SF_2$

- $K_4 = SF_4/SF_2$

The scale factor K_3 is extracted from dilepton $t\bar{t}$ events with $N_{\text{jets}}= 3$, which have exactly 1 ISR jet. The scale factor K_4 is extracted from dilepton $t\bar{t}$ events with $N_{\text{jets}}\geq 4$, which have at least 2 ISR jets. Both of these scale factors are needed since dilepton $t\bar{t}$ events which fall in our signal region (including the $N_{\text{jets}}\geq 4$ requirement) may require exactly 1 ISR jet, in the case that the second lepton is reconstructed as a jet, or at least 2 ISR jets, in the case that the second lepton is not reconstructed as a jet. These scale factors are applied to the dilepton $t\bar{t}$ MC only. For a given MC event, we determine whether to use K_3 or K_4 by counting the number of reconstructed jets in the event (N_{jets}^R) , and subtracting off any reconstructed jet which is matched to the second lepton at generator level (N_{jets}^ℓ); $N_{\text{jets}}^{\text{cor}} = N_{\text{jets}}^R - N_{\text{jets}}^\ell$. For events with $N_{\text{jets}}^{\text{cor}} = 3$ the factor K_3 is applied, while for events with $N_{\text{jets}}^{\text{cor}} \geq 4$ the factor K_4 is applied. For all subsequent steps, the scale factors K_3 and K_4 have been applied to the $t\bar{t} \to \ell\ell$ MC.

3.2   Step 2: Use the pre-veto sample to define a data-to-MC scale factor

The pre-veto sample is defined by the following requirements

- At least 1 selected electron (muon) with $p_T > 30$ GeV and $|\eta| < 2.5$ ($|\eta| < 2.1$)

- At least 4 selected jets, with at least 1 b-tagged jet

- $E_T^{\text{miss}} > 50$ GeV

Thus all selection criteria are applied with the exception of the veto on events containing an isolated track. This sample is dominated by $t\bar{t} \to \ell + \text{jets}$ with secondary contributions from $W + \text{jets}$ and $t\bar{t} \to \ell\ell$. This sample is used to define an overall data over MC scale factor (SF) in the peak control region, that accounts for differences between the data and the MC due to the luminosity estimate, the lepton efficiencies and so on and is thus applied to the full MC cocktail. To do so we define for the pre-veto sample (labeled 'all'):

- $N_{\text{peak}}^{\text{all}}$ = data yield in the peak region $60 < M_T < 100$ GeV

- $M_{\text{peak}}^{\text{all}}$ = MC yield in the peak region $60 < M_T < 100$ GeV

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

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

3.3   Step 3: Isolated Track Veto Efficiency Correction

The signal sample is defined by the same requirements as the pre-veto sample, except that we veto events containing an isolated track. The background is the inclusive sample can be split into 2 contributions:

- Dilepton background: mostly $t\bar{t} \to \ell\ell$.

- Single lepton background: mostly $t\bar{t} \to \ell + \text{jets}$ and $W + \text{jets}$;

The isolated track veto impacts these 2 contributions in different ways. For the dilepton background, the veto rejects events which have a genuine 2nd lepton, so applying the isolated track veto scales the dilepton background by the efficiency $\epsilon(trk)$ to identify the isolated track. For the single lepton background, the veto rejects events which do not have a genuine 2nd lepton but which have a "fake" isolated track, so the isolated track veto scales the single lepton background by (1-FR), where FR is the "fake rate" to identify an isolated track in events which contain no genuine 2nd lepton.

The isolated track efficiency ϵ can be measured in data and MC using $Z \to \ell\ell$ tag-and-probe studies. We therefore use the tag-and-probe studies to apply a correction to the $t\bar{t} \to \ell\ell$ MC. A measurement of the FR in data is non-trivial, but we can account for differences between the data and the MC by scaling only the single lepton background in the M_T peak region after applying the isolated track veto.

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→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 $t\bar{t} \to \ell\ell$ MC event, we determine the p_T of the 2nd lepton and translate this to find the corresponding requirement on the **absolute** track isolation, which is simply $0.1 \times p_T$.

- 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 $t\bar{t} \to \ell\ell$ MC event.

- **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 $\gamma \to e^+e^-$.**

At this point we are done applying scale factors to the dilepton background. We next apply a scale factor to the single lepton background. We use the following quantities:

- $N_{\text{peak}}^{\text{veto}}$ = data yield in the peak region $60 < M_T < 100$ GeV

- $M_{\text{peak}}^{\ell,\text{veto}}$ = single lepton background MC in the peak region $60 < M_T < 100$ GeV

- $M_{\text{peak}}^{\ell\ell,\text{veto}}$ = dilepton background MC in the peak region $60 < M_T < 100$ GeV

- $SF_{\ell}^{\text{veto}} = (N_{\text{peak}}^{\text{veto}} - M_{\text{peak}}^{\ell\ell,\text{veto}} SF_{\epsilon(trk)})/M_{\text{peak}}^{\ell,\text{veto}}$

The scale factor SR_{ℓ} is applied to the single lepton background to account for potential data vs. MC discrepancies in the isolated track veto.

To summarize, the dilepton and single lepton background prediction in the signal region ($M_T > 150$ GeV) are given by

- $P_{\text{sig}}^{\ell\ell} = SF^{\text{all}} \times SF_{\epsilon(\text{trk})} \times M_{\text{sig}}^{\ell\ell}$

- $P_{\text{sig}}^{\ell} = SF^{\text{all}} \times SF_{\ell}^{\text{veto}} \times M_{\text{sig}}^{\ell}$

where M_{sig} correspond to the Monte Carlo predictions in the tail of the distributions and the SF terms are data over MC scale factors to account for differences between the data and the MC for the following effects

- $SF^{\text{all}} \to$ effects impacting the normalization with the exception of the isolated track veto i.e. luminosity, lepton identification and trigger efficiencies . . .

- $SF_{\epsilon(trk)} \to$ the track veto efficiency for real second leptons

- $SF_{\ell}^{\text{veto}} \to$ the inefficiency introduced by the track veto on events without a second lepton (1-FR defined previously)

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, *Performance of b-jet identification in CMS*, PAS BTV-11-001

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

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

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

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

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

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

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

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

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

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

[17] https://twiki.cern.ch/twiki/bin/viewauth/CMS/CrossSections_3XSeries, https://twiki.cern.ch/twiki/bin/view/CMS/ProductionSpring

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

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

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

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

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

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

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

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

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

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