Search for Scalar Top-Quark Production in the all Hadronic Channel at 13 TeV

S Norberg On behalf of the CMS Collaboration
University of Puerto Rico Mayaguez, Peurto Rico, USA
E-mail: snorberg@cern.ch

Abstract. A search for supersymmetry in all-hadronic events with missing transverse momentum is presented. The data were collected in proton-proton collisions at a center-of-mass energy of 13 TeV with the CMS detector at the LHC and correspond to an integrated luminosity of 2.3 fb$^{-1}$. Search regions are defined using the properties of reconstructed jets, the presence of b quark and t quark candidates, and missing transverse momentum. No statistically significant excess of events above the expected contribution from standard model processes is observed. Exclusion limits are set on the masses of potential new particles in the context of simplified models of direct and gluino-mediated top squark production.

1. Introduction

A direct top squark production search is described in this document and refers to the high efficiency top tagger (HETT) portion of Ref. [1]. The T2tt and T2tb models explored here are illustrated in Fig. 1. In the T2tt model the stop decays to a top quark (t quark) and neutralino. In the T2tb model the stop decays to a bottom quark (b quark) and chargino.

![Figure 1](image.png)

Figure 1. The top squark pair production with the top squark decaying into a t quark and neutralino or into a b quark and chargino.

2. CMS Detector

The detector used is the Compact Muon Solenoid (CMS) and is a multi-purpose detector of cylindrical design. More details of the main features of the detector can be found in Ref. [2]. For a jet to be considered a b-quark jet (b-tagged) it has to pass the medium working
point requirements of the 'Combined Secondary Vertex' (CSV) method [3] and have transverse momentum ($p_T$) greater than 30 GeV and be within $|\eta| < 2.4$. The efficiency to identify the b-quarks is 67% overall. The probability of a jet originating from a light quark or gluon to be mis-identified as a b-quark jet is 1.4%, averaged over $p_T$ in t$t$ events [3].

3. Search Strategy
Top tagging in this analysis is done using all jets with a cone of radius 0.4 in $(\eta, \phi)$ space that satisfy $p_T > 30$ GeV and $|\eta| < 5$. There are three categories in this algorithm for these jets: trijet, dijet, and monojet candidates. In the trijet category we require the jets to lie within a cone of radius 1.5 in $(\eta, \phi)$ space. Within this cone they are subject to a set of conditions that are kinematically consistent with a t quark decay, more details are available in Refs. [1, 4]. Next the dijets category where one jet originates from the decay products of a W boson (W jet) that is merged, the mass is required to be between 70–110 GeV. Also the ratio of the Wjet and dijet masses are set to be between the ratio of the W boson and the t quark masses. Last but not least is the monojet category, for this we have a single jet which we require to have a jet mass between 110–220 GeV. After all the categories are looked at and all the top candidates are found the final step is to require that there be one b jet and that the mass of the candidates be within this range 100–250 GeV. If these candidates pass these requirements they are considered a top jet Ref. [1].

In this analysis the 37 search regions are defined in terms of the magnitude of the missing transverse momentum($E_T$), 'stransverse' mass $M_T^2$ [5, 6], and the number of b-tagged jets and top-tagged objects which is shown in Fig. 2. $M_T^2$ is defined using a pair of top candidates and $E_T$. When there are more than two top candidates the combination of all possible tops are found and the smallest $M_T^2$ is used. When only one top candidate is found the second top candidate is taken from the remanent of the event using the bottom-tagged jet. If this is the case than we find the jet closest to the bottom-tagged jet and if that jet has a value between 50 GeV and the top quark mass than we use that jet as the top candidate. If no combination is found that satisfies this requirement than the bottom tagged jet is considered the only remanent and used as the second top candidate. Fig. 3 demonstrates the background composition following the pre-selection cuts for $N_t$, and $E_T$. There are three main backgrounds in this analysis. There is the $\tau$ background, the QCD multijet background, and the $Z\rightarrow \nu\nu$ background, in the next section techniques to understand these backgrounds are discussed.

4. Backgrounds

4.1. Estimation of the $\tau$ background
The biggest contribution to the SM background comes from $t\bar{t}$ and W+jets events with leptonic W decays. In one case where the W boson decays to an electron or muon, then the veto on leptons are satisfied unless the lepton is 'lost'. By 'lost' we mean not isolated, not identified/reconstructed, or out of the acceptance region. In the other case a $\tau$ lepton decays hadronically from the W boson which is reconstructed as a jet. In this case the lepton would pass all of our veto requirements, this we call the hadronic $\tau$ background.

To model these lost leptons (LL) a weighted data sample that consists mainly of t$t$ events is used. The search trigger and the pre-selection are the same in this region but the muon veto on leptons is satisfied unless the lepton is 'lost'. By 'lost' we mean not isolated, not identified/reconstructed, or out of the acceptance region. In the other case a $\tau$ lepton decays hadronically from the W boson which is reconstructed as a jet. In this case the lepton would pass all of our veto requirements, this we call the hadronic $\tau$ background.

The weight factor is defined as taking into account acceptance, reconstruction and
Figure 2. Search bin definitions. The highest $E_T$ and $M_{T2}$ bins are open-ended. An example of this is in bin 10 which requires $E_T > 450$ GeV and $M_{T2} > 400$ GeV.

Figure 3. Background composition as a function of $N_t$, and $E_T$ between SM backgrounds (filled histograms) and two mass points of the T2tt signal model (dashed lines) and one mass point of the T2tb model.
identification, and isolation efficiencies multiplied by a factor that accounts for dilepton events where both leptons are lost is applied to the sum over the events in the single-muon control region, where this includes events with lost leptons from \(t\bar{t}\), \(W+\)jets, and single-top processes. A further correction is applied in order to take into account the efficiency of the \(M_T < 100\) GeV requirement. The final correction that is applied is to take into account the isolated track veto efficiency.

The estimate of the remaining hadronic \(\tau\)'s background is calculated by taking the signal region after applying the isolated track veto and is based on a control sample of \(\mu+\)jets events selected from data using a muon and \(H_T\)-based trigger, and requiring exactly one muon with \(p_T^\mu > 20\) GeV and \(|\eta| < 2.4\). A cut on the transverse mass of the W boson, \(M_T < 100\) GeV, is required to select events containing a \(W \rightarrow \mu\nu\) decay and to suppress potential signal events from being present in the \(\mu+\)jets sample.

The muon \(p_T\) is smeared by response template distributions derived for a hadronically-decaying \(\tau\) lepton to correct the leptonic part of the event. The response templates are derived using \(t\bar{t}\) and \(W+\)jets MC by comparing the true \(\tau\) lepton \(p_T\) with the reconstructed hadronic \(\tau\)'s jet \(p_T\). The kinematic variables of the event are recalculated with this hadronic \(\tau\)'s jet, and the search selections are applied. The probability to mistag a hadronic \(\tau\)'s jet as a b jet is significant and affects the \(N_b\) distribution of hadronic \(\tau\)'s background events. This effect is taken into account in the same way as in Ref. [1].

The hadronic \(\tau\)'s background prediction is calculated as a sum over all events in the \(\mu+\)jets control sample weighted by the \(\tau_h\) response. Additional corrections are applied. A few corrections are the muon acceptance, the \(M_T\) selection efficiency, the isolated track veto efficiency, and others.

### 4.2. Estimation of the \(Z\rightarrow \nu\nu\) background

For the \(Z \rightarrow \nu\nu\) background the \(Z \rightarrow \nu\nu\) simulation is used to have more statistics in the interpretation. This background is calculated in a way that any difference seen between data and simulation is taken into account through a correction factor that is applied. Two scale factors are applied to simulated events and are called \(R_{\text{norm}}\) and \(S_{\text{DY}}(N_j)\), and they correct the normalization of the simulation and the shape of the simulated \(N_j\) distribution. To calculate the factors a dimuon control region is used.

The \(R_{\text{norm}}\) scale factor, is derived in data using the same selection as the search region pre-selection, apart from the muon requirement and the requirement on the b-tagged jets. The expected yield in the DY simulation in this region is compared to the observed event yield in data after subtraction of the other SM processes for the scale factor calculation.

The factor, \(S_{\text{DY}}\), depends on the number of jets \((N_j)\) in the event and is derived in a loose dimuon control region in which the signal region requirement on \(H_T\) is relaxed to \(H_T > 200\) GeV and the requirements on \(E_T\), \(N_1\) and \(M_{T2}\) are removed. This scale factor is derived for each \(N_j\) bin as the ratio between the data, with non-DY backgrounds subtracted, and the DY simulation.

### 4.3. Estimation of the QCD multijet background

In order to predict the QCD multijet background a QCD multijet-rich data control region with minimum signal is achieved by inverting the pre-selection requirements on \(\Delta\phi(E_T,J_{1,2,3})\). A little of the other SM backgrounds remain, such as \(t\bar{t}\), \(W+\)jets, and \(Z+\)jets so they are subtracted out. The method used for lost leptons and hadronic \(\tau\)'s in this analysis are also used to subtract out those backgrounds in this region. For \(Z \rightarrow \nu\nu\) simulation is used because the contribution is so small. The number of QCD multijet events is measured in the data control region and a translation factor is applied to predict the amount of the QCD multijet background in each search bin. The translation factor is calculated in data and partly by simulation. It is computed as a simulated ratio between the signal region and the inverted-\(\Delta\phi\) control region, in
bins of $E_T$ and $M_{T2}$ where the bin boundaries follow those of the signal bins. The overall shape of this translation factor is kept from simulation. The value is normalized to data. Where the data measurement is in a sideband of the pre-selection region and is defined by the requirement $175 < E_T < 200$ GeV, where the amount of data is sufficiently large to make an accurate measurement.

5. Results

In Fig. 4 the number of events observed in data and the SM background predictions for the search region defined above are shown. In general the most significant background across the search regions comes from the SM tt production or W-boson production where the lepton is not detected or where the W boson decays hadronically. The next largest contribution is $Z \rightarrow \nu \nu$ production in association with jets, which also including heavy-flavor jets, where the neutrino pair gives large $E_T$ and the t quark conditions are satisfied by a combination of jets accidentally. The rest of the backgrounds are small contributions.

![Figure 4](image.png)

**Figure 4.** The black points are observed event yields in data and the filled solid areas are the predicted SM backgrounds for the 37 search bins of this analysis. The ratio of data over total background prediction in each search bin is shown in the latter half of the plot. In the ration only statistical uncertainties are propagated.

The binned likelihood fit to the observed data is the statistical interpretation of the results for the exclusion limits for both T2tt and T2tb models. This takes into account the predicted background and expected signal yields with their uncertainties in each bin. Exclusion limits are extracted based on a modified frequentist approach [1] using a profile likelihood ratio as test statistic. If the 95% upper limit on the production cross section falls below the theoretical cross section the signal models are considered to be excluded by the analysis.

Figures 5 show the 95% CL exclusion limits obtained, for both the pure T2tt scenario, and in the mixed T2tb scenario assuming a 50% branching fraction for each of the two decay modes ($\tilde{t} \rightarrow t\tilde{\chi}_i^0/\tilde{t} \rightarrow b\tilde{\chi}_1^\pm$). In the T2tb model the $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_1^0$ have a 5 GeV mass difference this is because there masses are assumed to be nearly degenerate. Using the 2.3 fb$^{-1}$ dataset, for the T2tt model stop masses are up to 780 GeV and LSP masses up to 260 GeV are probed. In the T2tb model, stop masses up to 750 GeV and LSP masses up to 200 GeV are probed. Observed weaker limits than expected for the LSP mass in the T2tb model because the most sensitive bins have a small excess.

6. Conclusions

An over view of the search results for direct top squark production in the all-hadronic final state have been presented here. In this analysis a top tagger is used and events with $E_T$, and $M_{T2}$ are
Figure 5. The solid black curves represent the observed exclusion contours and the corresponding ±1 standard deviations. The expected exclusion contour is the dashed red curves and includes a ±1 standard deviations with experimental uncertainties.

selected from a data sample corresponding to an integrated luminosity of 2.3 fb⁻¹ collected in proton-proton collisions at a center-of-mass energy of 13 TeV with the CMS detector. No statistically significant excess of events above the expected standard model background is observed. Exclusion limits are set at the 95% confidence level in the context of two simplified models of direct top squark pair production. For simplified models in which both top squarks decay to a top quark and a neutralino (T2tt), top squark masses up to 780 GeV and neutralino masses up to 250 GeV are probed. In the case of models that assume 50% branching fractions for top squark decays to a top quark and a neutralino, or to a bottom quark and a chargino that is nearly degenerate in mass with respect to the neutralino (T2tb), top squark masses up to 625 GeV and neutralino masses up to 200 GeV are probed.

References
[1] CMS Collaboration. Search for direct production of top squark pairs decaying to all-hadronic final states in pp collisions at \(\sqrt{s} = 13\) TeV. Technical Report CMS-PAS-SUS-16-007, CERN, Geneva, 2016.
[2] CMS Collaboration. CMS technical design report, volume II: Physics performance. J. Phys. G, 34:995, 2007.
[3] CMS Collaboration. Performance of b tagging at \(\sqrt{s} = 8\) TeV in multijet, \(t\bar{t}\) and boosted topology events. CMS Physics Analysis Summary CMS-PAS-BTV-13-001, CERN, 2013.
[4] T Plehn, M Spannowsky, M Takeuchi, and D Zerwas. Stop Reconstruction with Tagged Tops. JHEP, 1010:078, 2010.
[5] C G Lester, and D J Summers. Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders. Phys.Lett., B463, 99-103, 10.1016/S0370-2693(99)00945-4, 1999.
[6] A Barr, C Lester, and P Stephens. \(m(T2)\): The Truth behind the glamour. J.Phys., G29, 2343-2363, 10.1088/0954-3899/29/10/304, 2003.