CMS Physics Analysis Summary

Contact: cms-pag-conveners-b2g@cern.ch 2013/11/22

Search for pair production of resonances decaying to a top quark plus a jet in final states with two leptons

The CMS Collaboration

Abstract

Results are reported from a search for pair production of new physics resonances decaying into a top quark and another parton. Two specific models, spin-3/2 excited top quark and *R*-parity violating bottom squark pair production, are considered. Data recorded by the CMS experiment at the LHC from proton-proton collisions at $\sqrt{s} = 8$ TeV corresponding to a total integrated luminosity of 19.5 fb^{-1} are used. The experimental signature in this search is two leptons (e or μ), two jets coming from decays of bottom quarks, and two jets coming from light flavor quarks or gluons. Resonant decay candidates reconstructed from these objects are studied. The observations are consistent with the standard model. Upper limits on the pair production cross section are set for excited top quarks with masses between 300 GeV and 1000 GeV and for bottom squarks with masses between 250 GeV and 600 GeV. Excited top quark (bottom squark) pair production is excluded at the 95% confidence level between 300 GeV and 703 GeV (250 GeV and 326 GeV).

1 Introduction

The origin of elementary particle masses in the standard model (SM) of particle physics is ascribed to spontaneous breaking of the electroweak symmetry [1–3]. It has long been suspected that the top quark plays a special role in this mechanism and that its coupling to hypothetical new particles anticipated in theories beyond the standard model (BSM) could potentially be large. We search for pair produced BSM resonances that couple to a top quark and another parton. Two specific models are studied in detail:

- a spin-3/2 top excitation (denoted as t^*) that decays to a top quark and a gluon [4–9].
- the R-parity violating minimal supersymmetric model (RPV MSSM) constrained to have minimal flavor violation [10], which allows the bottom squark ($\tilde{b}$) to be the lightest supersymmetric particle and restricts its decays to a top quark and a down-type quark. In this model, the dominant decay mode is to a top quark and a strange quark. This search is sensitive to the λ''_{tbs} and λ''_{tbd} parameters of the RPV Lagrangian described in Ref. [10].

These two models predict similar kinematics for the final state, but differ in production cross sections, e.g., the excited top model predicts cross sections of order 100 pb for a t^* with a mass of 300 GeV and the RPV MSSM model predicts a cross section of order 1 pb for a $\tilde{b}$ of the same mass. When considering the RPV MSSM model, we assume that the $\tilde{b}$ decays promptly. For some parameters in this particular model, e.g., a 100% left-handed $\tilde{b}$ with decay width highly suppressed by mass insertion as discussed by Csaki, Grossman, and Heidenreich [10], the $\tilde{b}$ may have a lifetime that is too large to be covered by this search. The t^* model has previously been studied at CMS in the single lepton channel [11] where it has been excluded at 95% confidence level (C.L.) between 450 GeV and 790 GeV.

We restrict this search to dilepton final states, where each top quark decays to $W^{\pm}$ bosons which in turn decay leptonically. Feynman diagrams of these processes are shown in Fig. 1. To discriminate signal from background events, we analyze the reconstructed resonance mass and the transverse momenta of jets identified as coming from light quarks or gluons.

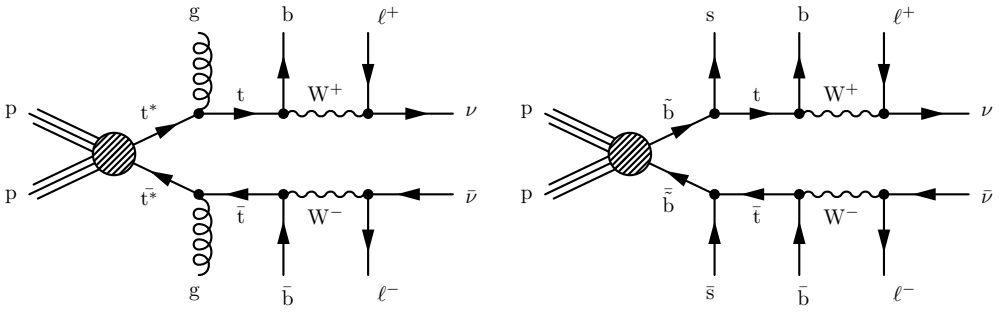

Figure 1: Leading order Feynman diagrams for pair production of (left) excited top quarks and (right) R-parity violating bottom squarks.

2 Event reconstruction, selection, and simulation

Signal events are characterized by the presence of two isolated leptons (e or μ), missing transverse energy (E_T^{miss}), at least two jets originating from b quarks, and at least two jets originating from light flavor quarks or gluons (denoted in the text as light jets). The E_T^{miss} is assumed to be due to the undetected neutrinos from the $W^{\pm}$ decays.

Events are measured by the Compact Muon Solenoid (CMS) detector. The central feature of the CMS apparatus is a superconducting solenoid with an internal diameter of 6 m, which generates a 3.8 T uniform magnetic field along the axis of the LHC beams. A silicon pixel and strip tracker, a crystal electromagnetic calorimeter (ECAL) and a brass/scintillator hadron calorimeter (HCAL) are located within the magnet. Muons are identified and measured in gas-based detectors embedded in the outer steel magnetic flux return yoke of the solenoid. In addition to the barrel and endcap detectors, CMS has extensive forward calorimetry. The inner tracker measures charged particles within the pseudorapidity range $|\eta| < 2.5$, where $\eta = -\ln[\tan(\theta/2)]$ and θ is the polar angle with respect to the counterclockwise proton beam. It provides an impact parameter resolution of 15 μm and a transverse momentum (p_{T}) resolution between 1.5% and 2.5% for 100 GeV particles. A detailed description of CMS can be found elsewhere [12].

Collision events are selected by the trigger system that requires the presence of a pair of charged leptons: two electrons, two muons, or one electron and one muon. The requirements on the minimum p_{T} of the two leptons are 17 and 8 GeV. Using simulations, the trigger is estimated to be more than 95% efficient in signal events passing the final selection.

All physics objects used in this analysis are reconstructed by the particle-flow algorithm [13], which performs a global event reconstruction and provides the full list of particles identified as electrons, muons, photons, and charged or neutral hadrons. We require at least one primary vertex reconstructed from at least four tracks with longitudinal (radial) distance from the center of the detector less than 24 (2) cm. The primary vertex with the highest sum of transverse momenta squared of the associated tracks is taken as the interaction vertex of the hard collision. Charged objects associated with primary vertices that are not the interaction vertex are not used in the event reconstruction.

At least two leptons passing identification and isolation criteria are required in the event. Muon candidates are reconstructed using a global fit to hits in both the silicon tracker and muon systems; we require them to have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$. Electron candidates are reconstructed from clusters of energy deposits in the ECAL, which are then matched to hits in the silicon tracker, taking into account the energy losses due to bremsstrahlung. They are identified using a multivariate-based selection [14] and are required to have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$.

Muons and electrons are required to be consistent with originating from the primary vertex in the event and to have a transverse impact parameter with respect to the interaction point of less than 0.02 cm. The relative isolation is defined as the sum of the p_{T} of charged and neutral objects not including the lepton in a cone centered on the lepton direction, divided by the lepton p_{T}. The cone size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is 0.3 for electrons and 0.4 for muons. We require electrons (muons) to have relative isolation less than 0.15 (0.12). In addition to the single object requirements, in events with two same flavor leptons, we require this dilepton system to have a mass either between 15 GeV and 76 GeV or greater than 106 GeV in order to suppress events with Z-bosons or low-mass resonances decaying to two leptons.

Jets are defined by clustering reconstructed objects according to the anti-k_{T} algorithm [15] with a distance parameter of 0.5. We subtract energy from overlapping proton-proton interactions (pileup) and the underlying event using the FASTJET technique [16–18]. We restrict the selection to jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$. Each jet is required to be constructed from more than one identified particle (at least one of which is charged) and the distance ΔR between the jet and the nearest isolated lepton is required to be larger than 0.3.

The Combined Secondary Vertex algorithm [19] is used to identify jets coming from the de-

cay of a b quark. Two working points corresponding to different misidentification rates are used [20]. At the loose working point, the b-tagging efficiency for jets with p_T between 80 and 120 GeV is approximately 85% and the misidentification rate is 9.9%. At the medium working point, the b-tagging efficiency for jets in the same p_T range is 70% and the misidentification rate is 1.4%. To account for the difference in performance of this algorithm between data and simulation, scale factors depending on p_T and η are computed for jets with p_T above 20 GeV and applied to the simulation. These scale factors range from 0.90 to 0.97 with uncertainties below 0.04.

We require at least two selected jets to pass the loose b-tagging selection and in addition at least one of these to pass the medium b-tagging selection. Additionally, we require at least two jets to fail the loose b-tagging selection. This allows for unambiguous categorization of light jets from the BSM resonance decay and b-jets from the top decay.

We simulate several mass points of t^* and $\tilde{b}$ events. Simulation is performed using the MADGRAPH 5 event generator [21]. PYTHIA 6.4 [22] using tune Z2star [23] simulates hadronization and fragmentation of t^* events; PYTHIA 8.1 [24] using tune 4C [25] simulates the same for $\tilde{b}$ events. We use PYTHIA 8.1 for the $\tilde{b}$ events because it is designed to correctly handle baryon number violating decays. The CTEQ6L parton distribution function (PDF) sets are used for the t^* and $\tilde{b}$ samples [26]. All events are simulated using the full simulation of the CMS detector and trigger implemented in GEANT 4 [27]. We include additional minimum bias events in the simulation that match the distribution of the number of such interactions observed in data to account for pileup. Cross sections for pair production of the t^* model are calculated for a spin-3/2 particle; however the simulation assumes a spin-1/2 particle, as no generator of spin-3/2 particles exists at present. This may affect the signal efficiency as discussed in Section 4. The differences between the spin-0 and spin-1/2 pair reconstruction efficiencies can be seen in Figs. 5 and 6. We use the more conservative spin-1/2 efficiency for the t^* model.

3 Signal and background discrimination

The dominant background to the t^* and $\tilde{b}$ signals originates from SM top quarks pair-produced in association with jets from initial or final-state radiation (ISR or FSR respectively). Other SM processes account for a small ($\approx$ 5%) contribution; they are single-top production, diboson production, Drell–Yan production and top pair production in association with vector bosons. Signal events contain a mass resonance that produces top quarks in association with light jets. The light jets from this decay have a relatively high p_T. We use both of these properties to discriminate between signal and background.

We associate the two highest p_T light flavor tagged jets with light partons from the BSM resonance decays, the two highest p_T b-tagged jets with bottom quarks from top quark decays, and the two highest p_T leptons with leptons from $W^{\pm}$ boson decays. In total, 6478 events pass selection requirements: 1365 in the $\mu\mu$ channel, 1723 in the ee channel, and 3390 in the eμ channel. There are only eleven events with more than two leptons.

3.1 Light jet p_T spectrum

To model the light jet spectrum of SM processes, we assume that the light jets are produced predominantly by ISR or FSR and therefore have a steeply falling spectrum. Signal events, on the other hand, are more likely to contain light jets with relatively high p_T.

Letting $p_T^{(1)}$ and $p_T^{(2)}$ denote the transverse momenta of respectively the highest-p_T and sec-

ond highest-p_T light jets in the event, we use simulated SM events to help choose the form of a two-dimensional probability density function $\rho_{2D}^{SM}(p_T^{(1)}, p_T^{(2)})$ with sufficient flexibility to fit the data well. This distribution is constructed as a sum of three 2D densities with index i, $\rho_i^{2D}(p_T^{(1)}, p_T^{(2)})$, (multiplied by a Heaviside step function to enforce the ordering and a factor of two to normalize the function):

$$\rho_{2D}^{SM}(p_T^{(1)}, p_T^{(2)}) = 2\sum_{i=1}^{3} f_i \rho_i^{2D}(p_T^{(1)}, p_T^{(2)})\Theta(p_T^{(1)} - p_T^{(2)}), \quad (1)$$

where $\sum_{i=1}^{3} f_i = 1$. Each component 2D density is factorized into 1D densities,

$$\rho_i^{2D}(p_T^{(1)}, p_T^{(2)}) = \rho_i^{jet}(p_T^{(1)})\rho_i^{jet}(p_T^{(2)}). \quad (2)$$

The 1D densities all have the same form,

$$\rho_i^{jet}(p_T) = \lambda_i \alpha \exp\left(-\lambda_i p_T^{\alpha}\right) p_T^{\alpha-1}, \quad (3)$$

where α is a parameter between 0 and 1 common to all components, while the λ_i are parameters differing in each component. This function has the steeply falling behavior that we expect from ISR or FSR, and potentially a longer tail than a pure exponential distribution, becoming identical to an exponential distribution if $\alpha = 1$ and approaching a uniform distribution as $\alpha \to 0$.

The parameters of these densities are determined by fitting to data, maximizing the likelihood function defined in Section 5. This function is sufficiently flexible to accommodate observed light jet p_T behavior while not over-fitting. Figure 2 shows background-only hypothesis fits to the data, illustrating good agreement. In the background-only hypothesis, the best fit value for alpha is 0.93 and the best fit value for the dominant ρ_i^{2D} component's λ_i is 1.7×10^{-2}.

To parameterize the signal distribution, we model jets as being sampled from the sum of two two-dimensional log-normal distributions and ordered by p_T. The parameters of these two distributions and their relative fractions are determined by fitting to the signal simulation. We call this distribution $\rho_{2D}^{signal}(p_T^{(1)}, p_T^{(2)})$.

Figure 3 illustrates the difference between light jet shapes. It shows the two-dimensional distributions of the background-only hypothesis fit to data and the signal distributions for a high mass point (750 GeV t^*) and a low mass point (350 GeV $\tilde{b}$).

3.2 Resonance reconstruction

In order to extract the mass of the potential resonance, we reconstruct the BSM particle decay chain. First we reconstruct top quark pair candidates using the lepton momenta, b-tagged jet momenta, and E_T^{miss}. Then we combine these candidates with the light jets to reconstruct possible BSM resonance pairs.

Given the mass of the top quark and $W^{\pm}$ boson, the four-momenta of the neutrinos in $t\bar{t}$ decays resulting in two charged leptons can be determined by solving a quartic equation [28, 29]. We use the measured four-momenta of the leptons and b-tagged jets, the measured E_T^{miss} in each event, and the PDG world average [30] top quark and $W^{\pm}$ boson mass (173 GeV and 80.4 GeV, respectively). We assume that all E_T^{miss} in the event comes from neutrinos associated with leptonic $W^{\pm}$ decays and that all resonances occur on-shell. We solve this quartic equation for both pairings of leptons and b-tagged jets yielding eight (possibly unphysical) solutions for the $t\bar{t}$ candidates in the event.

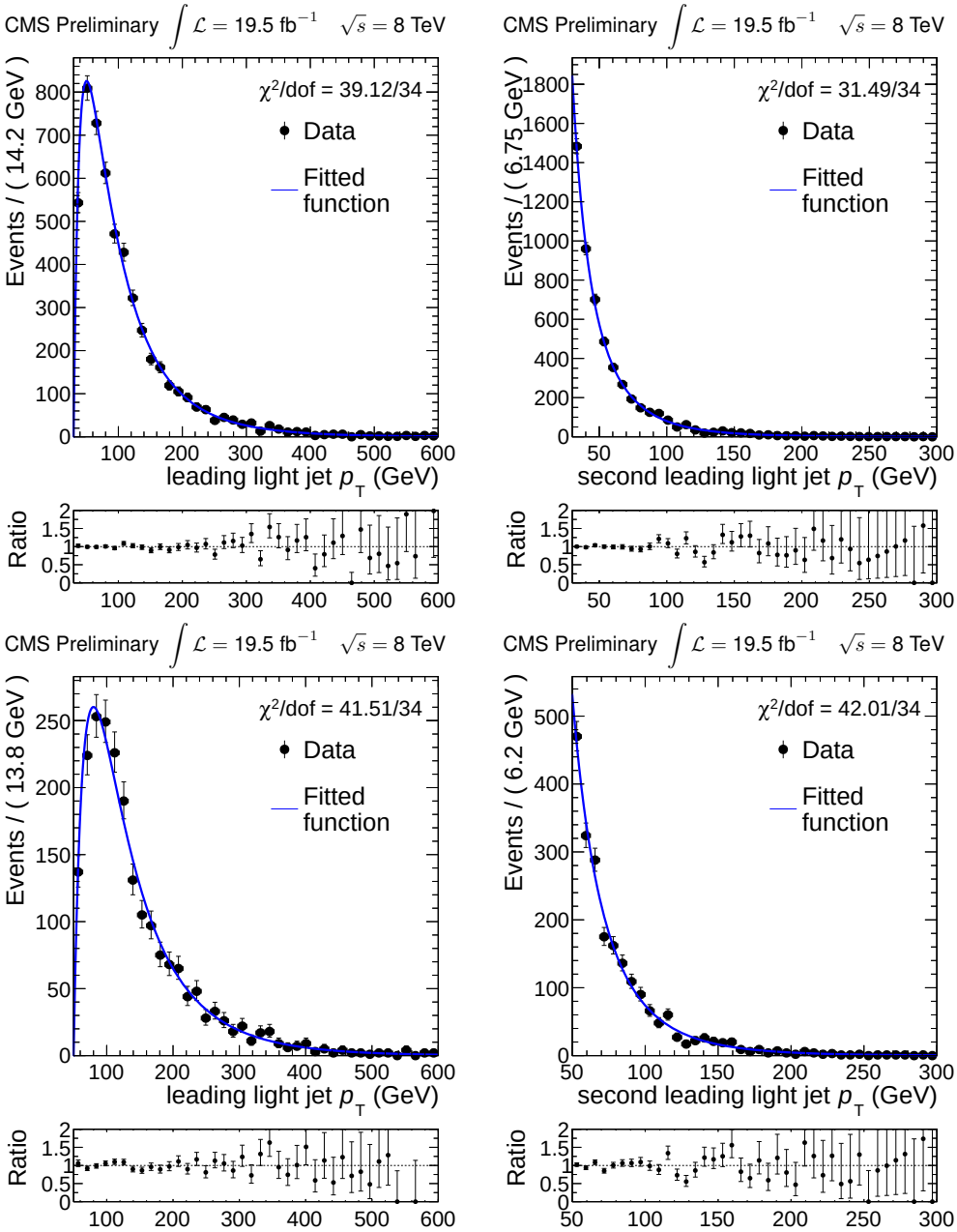

Figure 2: Light jet p_{T} distributions and the likelihood fit, with signal cross section set to zero. The line represents the fitted function and the points represent the data. The ratio of the data to the fitted function is also shown. The top row shows the (left) leading light jet distribution and (right) second leading light jet distribution. The bottom row is restricted to events with second leading light jet $p_{\mathrm{T}} > 50\,\mathrm{GeV}$.

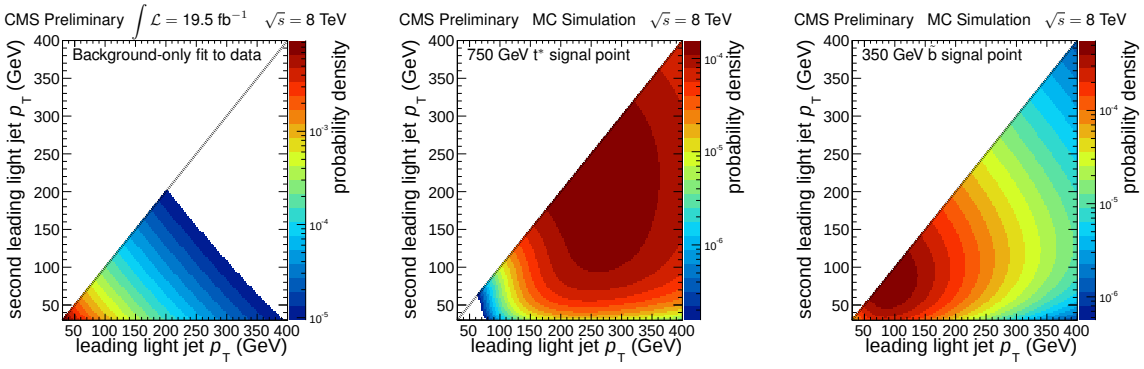

Figure 3: Two-dimensional light jet p_T distributions for (left) the background-only hypothesis fit to data, (center) the t* signal with a mass of 750 GeV, and (right) the $\tilde{b}$ signal with a mass of 350 GeV. The scales are logarithmic and a line has been drawn along the diagonal to illustrate the ordering of the jets by p_T.

If none of these solutions are physical, we vary the measured jet momenta within their resolutions using the procedure described in Ref. [31]. To perform this variation, we use a Monte Carlo (MC) sampling of the parameter space defining the jet resolution. Each jet has a resolution around its measured p_T, η, and ϕ coordinates from which we sample a random number and once again solve the quartic equation. The E_T^{miss} in the event is recomputed for each sampling. We repeat the sampling 1000 times per event and choose the set of parameters with at least one physical solution for $t\bar{t}$ candidates in the event that has the least χ^2 with respect to the measured jets. This procedure improves the event selection efficiency by a factor of approximately 1.4 for both background and simulated signal events. The efficiency of finding real solutions in signal events that pass all other selection criteria is between 55% and 85%, depending on the particular signal model and mass point. The background efficiency is approximately 81%. Additional data events retrieved using this procedure improve the signal sensitivity by 15%. Repeating this sampling procedure more than 1000 times results in negligible further gains in sensitivity.

The physical solutions for $t\bar{t}$ candidates (up to eight in each event) are combined with the light jets to form candidate t* or $\tilde{b}$ resonances. For each solution, there are two such pairings and hence up to sixteen solutions for the candidate BSM resonance pair in the event. We select the pair with the smallest absolute value of the logarithm of the ratio of the two candidate resonance masses. For the remainder of the analysis, we substitute the average of these two masses, denoted by $M_{t,jet}$, for the mass of each candidate within the chosen pair. It is bounded from below by the top quark mass.

The resulting mass distributions for background and signal are parameterized by different functions. The background distribution displays a turn-on behavior due to p_T requirements on the leptons and jets followed by a falling tail. It is modeled by a gamma distribution added to a log-normal distribution. Both functions are required to peak at the same value.

For simulated signal, the success rate for correctly matching the reconstructed and true objects increases from 30% to 50% with increasing signal mass. Other reconstructions display a mismatch either in the b-jet and lepton pairing or in the light jet and top candidate pairing. Events that are correctly paired have a narrower mass width than those that are incorrectly paired, which also have a large tail extending to high mass. The signal is parameterized as the sum of two gamma distributions with parameters determined by a fit to the signal simulation. One of

the gamma functions models the correctly paired events and the other models the incorrectly paired events. Examples of these signal distributions can be seen in Figs. 9 and 10.

Figure 3 illustrates the differences between SM and signal light jet p_T distributions. In order to increase the sensitivity of the analysis, events are categorized according to the second leading light jet p_T, as described in Table 1. We label these as three signal enhanced regions (denoted as SR1–3) and one signal depleted control region (denoted in the text as CR). Each region's mass spectrum is fit separately. The mass distribution is conditional on which second leading light jet p_T region the event falls into. We write the background distribution as $\rho^{\text{SM}}_{\text{mass}}(m|p_T^{(2)})$ and the signal distribution as $\rho^{\text{signal}}_{\text{mass}}(m|p_T^{(2)})$.

Table 1: Definition of signal and control regions.

Second leading light jet p_T	Region
30 to 50 GeV	control region (CR)
50 to 80 GeV	signal region 1 (SR1)
80 to 110 GeV	signal region 2 (SR2)
$>$ 110 GeV	signal region 3 (SR3)

Figure 4 shows the background-only hypothesis fits of the invariant mass spectra to the data in the four light jet regions (SR1–3 and CR).

4 Systematic uncertainties

Fitted parameters of the background model are determined exclusively from data. Potential systematic uncertainties in the background model could arise from the choice of functional forms in Section 3.1. We perform studies to ensure that the model is sufficiently general to approximate the true shape of the distributions, with any differences negligible compared to statistical uncertainties.

The signal, on the other hand, is modeled by a MC simulation and we consider several systematic uncertainties that can affect the spectrum of light jet p_T and the mass reconstruction procedure. For each source of systematic uncertainty, we determine the change in the signal distribution parameters and encode this as a covariance matrix of all parameters. A multivariate normal distribution is constructed from the sum of all such covariance matrices and used to constrain the parameters of the signal distribution. Allowing the invariant mass parameters to vary within this constraint has a negligible effect on signal sensitivity. The final constraint function sets these parameters to their maximum likelihood values in signal simulation and does not allow them to vary. The light jet parameters and selection efficiencies are allowed to vary within their uncertainties.

The central value of the total selection efficiency after all selection requirements and systematic uncertainty is shown in Figs. 5 and 6. These figures include the 9.4% branching fraction [30] of each of the top quarks to a leptonic final state, resulting in an initial efficiency of about 1% in the same flavor (ee, $\mu\mu$) lepton channels and 2% in the opposite flavor (eμ) channel. These initial efficiencies include production of tau leptons that decay to electrons or muons. We estimate that in BSM resonance pair production with a dileptonic final state, about 60% of events produce leptons, b quarks, and light partons within the fiducial volume of the detector. About 85% of these events have the physics objects reconstructed and about 65% of those events have the jets from b quarks correctly tagged. And about 90% of the remaining events pass all

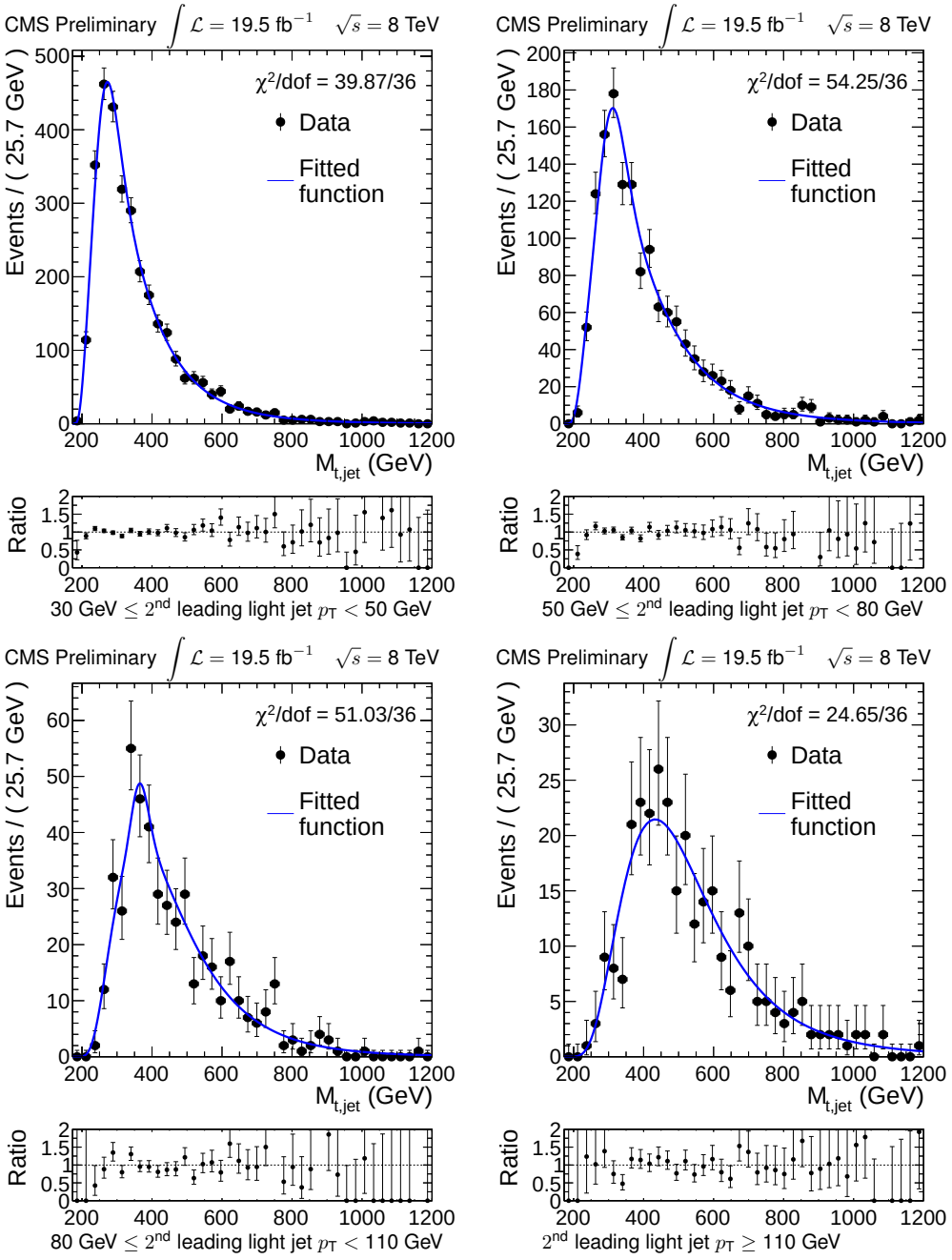

Figure 4: Reconstructed invariant mass distributions in the signal (SR1–3) and control (CR) regions. The line corresponds to the result of the likelihood function maximization with signal cross section set to zero. The points represent the data. The ratio of the data to the fitted function is also shown. Top left: second leading light jet p_T between 30 and 50 GeV (CR). Top right: second leading light jet p_T between 50 and 80 GeV (SR1). Bottom left: second leading light jet p_T between 80 and 110 GeV (SR2). Bottom right: second leading light jet p_T greater than 110 GeV (SR3).

selection requirements. Efficiency differences between t^* and $\tilde{b}$ models of the same mass arise because the pair produced particles have different spins. The p_T spectra of the light quarks and gluons, leptons, and b quarks lead to different efficiencies for the different models. The figures present neither correlations between the efficiencies nor changes to the relative signal efficiency in each light jet p_T bin. These effects are taken into account in the full likelihood construction.

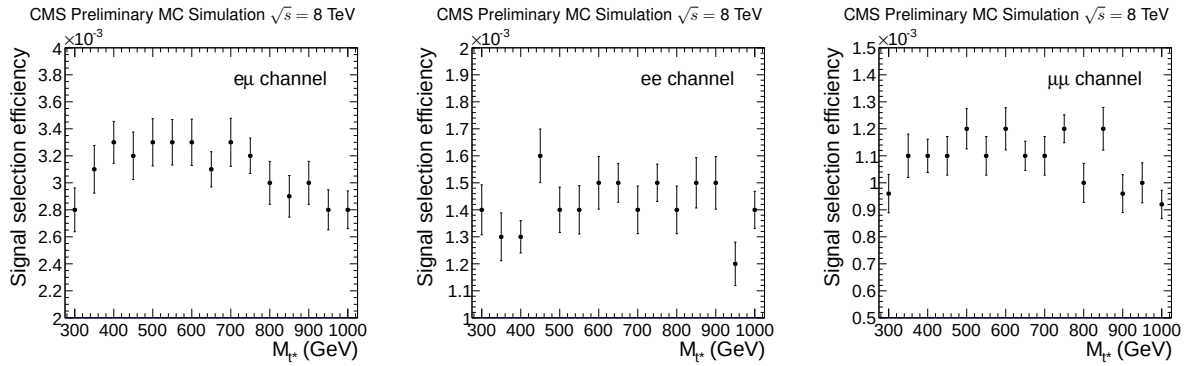

Figure 5: Signal selection efficiency and relative systematic uncertainty on the efficiency in each channel for the t^* model. Left: eμ channel, middle: ee channel, right: $\mu\mu$ channel.

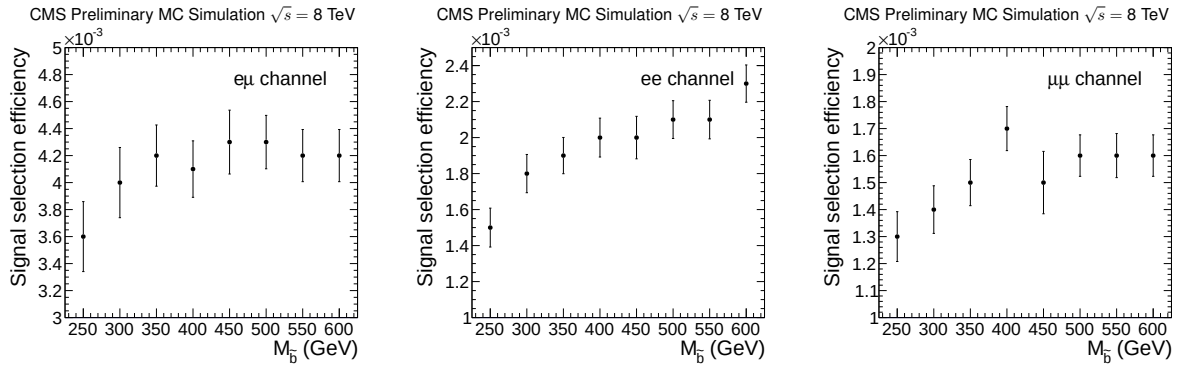

Figure 6: Signal selection efficiency and relative systematic uncertainty on the efficiency in each channel for the $\tilde{b}$ model. Left: eμ channel, middle: ee channel, right: $\mu\mu$ channel.

Sources of systematic uncertainty and the methods used to estimate the effect on the signal distribution parameters and selection efficiency are briefly described below.

Jet four-momenta are varied using standard CMS p_T and η-dependent correction factors [32] to account for the uncertainty on the jet energy scale measurement. The uncertainty on the jet energy resolution is accounted for using p_T and η-dependent data-to-simulation resolution ratios based on the difference between reconstructed and matched generator jet p_T. We account for any additional discrepancy between simulation and data E_T^{miss} [33] from unclustered objects using standard CMS scale factors and find this to be a negligible effect.

Potential discrepancies between lepton objects from data and simulation are taken into account using energy scale factors [14, 34]. For electrons this scale factor is 1.5% in the range $1.5 < |\eta| < 2.5$ of the detector and 0.6% in the range $|\eta| < 1.5$; for muon this scale factor is 0.2%. Uncertainties from lepton reconstruction and isolation effects are taken into account as 0.5% uncertainties on the signal selection efficiency of each selection flavor which are 100% correlated between each pair of flavors.

The b, c, and light flavor tagging efficiencies are varied separately within their statistical uncertainties, and standard CMS data-to-simulation scale factors are applied and varied within the measured uncertainties [19, 20]. The b and c quark scale factors are treated as correlated, and the light flavor scale factors are treated as uncorrelated with the heavy flavor scale factors.

The PDF uncertainty is varied according to the prescription of the PDF4LHC Working Group [35]. The uncertainty on the luminosity is 2.6% [36], and the number of pileup interactions in the simulation is weighted to match the measured distribution in data. The distribution is varied using the standard CMS procedure which changes the shape and varies the mean by 5%. The trigger efficiency is modeled in the simulation and the uncertainty is accounted for by a 1% systematic on the signal selection efficiency of each lepton pair flavor (ee, eμ, and $\mu\mu$). Finally, we account for signal selection efficiency uncertainties due to the finite size of the signal MC sample using Poisson statistics.

Table 2 reports the uncertainty on the signal selection efficiency due to the different sources of systematic uncertainty for two signal models and mass points. In combination these systematic uncertainties change the calculated upper limit of the cross section value between 1% and 10%, depending on the mass point, compared to the upper limit calculated only with statistical errors. The dominant systematic effect comes from variations of the jet energy scale.

Table 2: Relative systematic uncertainty on the signal selection efficiency broken down by source of signal systematic uncertainty for two signal models and mass points.

Simulation uncertainty	750 GeV t*	350 GeV $\tilde{b}$
Heavy flavor scale factor for b-tagging	5.3%	4.9%
Light flavor scale factor for b-tagging	3.2%	4.7%
Jet Energy Scale	1.1%	4.6%
Signal MC Statistics	0.22%	2.1%
Jet Energy Resolution	0.44%	1.8%
Pile up	1.1%	1.5%
Parton density function	1.9%	1.0%
MC b-tagging efficiency for b-jets	1.3%	0.43%
MC b-tagging efficiency for c-jets	0.31%	0.25%
MC b-tagging efficiency for light jets	0.51%	0.52%
Electron Energy Scale	0.62%	0.19%
Muon Energy Scale	0.12%	0.04%

5 Constructing confidence intervals

For both signal and background, we define a three-dimensional probability density function which is constructed using the two-dimensional light jet distributions defined in Section 3.1 and the invariant mass distributions defined in Section 3.2. These three-dimensional distributions can be written as:

$$\rho^{\mathrm{SM}}_{\mathrm{3D}}(m, p_{\mathrm{T}}^{(1)}, p_{\mathrm{T}}^{(2)}) = \rho^{\mathrm{SM}}_{\mathrm{mass}}(m|p_{\mathrm{T}}^{(2)})\rho^{\mathrm{SM}}_{\mathrm{2D}}(p_{\mathrm{T}}^{(1)}, p_{\mathrm{T}}^{(2)}) \tag{4a}$$

$$\rho^{\mathrm{signal}}_{\mathrm{3D}}(m, p_{\mathrm{T}}^{(1)}, p_{\mathrm{T}}^{(2)}) = \rho^{\mathrm{signal}}_{\mathrm{mass}}(m|p_{\mathrm{T}}^{(2)})\rho^{\mathrm{signal}}_{\mathrm{2D}}(p_{\mathrm{T}}^{(1)}, p_{\mathrm{T}}^{(2)}), \tag{4b}$$

and the complete distribution is:

$$\rho_{3D}^{\text{total}} = (\mu^{\text{SM}}\rho_{3D}^{\text{SM}} + \mathcal{E}\sigma_{\text{signal}}\rho_{3D}^{\text{signal}})/(\mu^{\text{SM}} + \mathcal{E}\sigma_{\text{signal}}). \tag{5}$$

Here μ^{SM} is the SM yield, $\mathcal{E}$ is the product of signal efficiency and total luminosity, and σ_{signal} is the signal cross section.

Constraints on the signal shape parameters are derived as described in Section 4. We write the constraint distribution as ρ_{syst}. There are no constraints on the parameters describing the SM distribution. We construct an extended unbinned likelihood function from our data and these distributions as:

$$\mathcal{L}(\sigma,\theta) = \rho_{\text{syst}}\frac{(\mu^{\text{SM}} + \mathcal{E}\sigma_{\text{signal}})^N \exp\left(-\mu^{\text{SM}} - \mathcal{E}\sigma_{\text{signal}}\right)}{N!}\prod_{i=0}^{N}\rho_{3D}^{\text{total}}(m_i, {p_{\text{T}}^{(1)}}_i, {p_{\text{T}}^{(2)}}_i), \tag{6}$$

where N is the number of events in our sample; m_i, ${p_{\text{T}}^{(1)}}_i$, and ${p_{\text{T}}^{(2)}}_i$ are, respectively, the mass, leading light jet p_{T}, and second leading light jet p_{T} of the i^{th} event; σ is the cross section for production of the resonance pair; and θ is the set of all nuisance parameters.

To perform hypothesis tests on a particular model, we treat the parameter of interest (cross section) using the unified approach advocated by Feldman and Cousins [37]. With our observed data, the two-sided intervals from this approach all have lower endpoints at zero, and we use the upper endpoint as a proxy for an "upper limit". For comparison, we also compute (conservative) upper limits using the CL_s criterion [38, 39]. In both approaches, we treat the nuisance parameters using a profile likelihood technique, as described by the LHC Higgs Combination Group [40] and references therein. The profile likelihood ratio test statistic is

$$\lambda_p(\sigma) = \frac{\mathcal{L}(\sigma,\hat{\hat{\theta}}(\sigma))}{\mathcal{L}(\hat{\sigma},\hat{\theta})}, \tag{7}$$

where σ is the cross section and θ is the set of nuisance parameters. The denominator is the usual global maximum of the likelihood function $\mathcal{L}$, with $\{\hat{\sigma},\hat{\theta}\}$ denoting the parameter values for which $\mathcal{L}$ is maximized, with $\hat{\sigma}$ required to be non-negative. The numerator is the profile likelihood function of σ, obtained for each σ by finding the values of the nuisance parameters that maximize the likelihood function for that value of σ, denoted by $\hat{\hat{\theta}}(\sigma)$. The profile likelihood ratio is evaluated for the observed data set and for synthetic data sets obtained from pseudo-experiments.

In both the unified approach and the CL_s approach, pseudo-experiments are generated by MC simulation to obtain the distribution of $\lambda_p(\sigma)$ for a fine grid in σ. For each σ in this grid, these pseudo-experiments are generated using the frequentist approach of Ref. [40], in which the nuisance parameters are set to the $\hat{\hat{\theta}}(\sigma)$ values obtained from conditionally maximizing the likelihood function of the observed data, using the σ for which the distribution of $\lambda_p(\sigma)$ is being constructed.

In the unified approach [37], we associate with each value of σ to be tested a (possibly different) critical value λ_p^{Cr} such that 95% of the pseudo-experiments have $\lambda_p(\sigma) > \lambda_p^{\text{Cr}}$. If the profile likelihood ratio evaluated in data is less than λ_p^{Cr}, the test then rejects that σ at 95% C.L. The union of all cross sections that are not rejected is the unified interval. Treatment of nuisance parameters in the manner used has been found to have good coverage by Cranmer [41].

In the CL_s approach [38, 39], the construction is intrinsically one-sided, which is enforced by modifying $\lambda_p(\sigma)$ so that it is set to 1 if $\hat{\sigma}$ is greater than the cross section under test; $\lambda_p^{\mathrm{mod}}(\sigma)$ denotes this modified version. For each σ under test, we find the one-tailed p-value of the observed data in the signal-plus-background hypothesis, denoted p_σ. This is the fraction of pseudo-experiments with $\lambda_p^{\mathrm{mod}}(\sigma)$ less than the value measured in data. We also generate pseudo-experiments with the signal cross section set to zero to construct the distribution of $\lambda_p^{\mathrm{mod}}(\sigma)$, where σ is the tested cross section value. From these pseudo-experiments we obtain the distribution of the test statistic in the background-only hypothesis, and from this distribution we obtain the p-value of data in the background-only hypothesis, denoted p_0. Then CL_s is defined as $p_\sigma/(1-p_0)$. If $CL_s < 0.05$, we reject that σ at 95% C.L. The largest cross section not rejected corresponds to the the CL_s upper limit.

6 Results

We observe consistency with the SM expectation and set limits on the cross section of each of the signal models. We construct unified intervals [37] on the signal cross section and observe only intervals with lower edges of zero. The upper edge of these intervals is considered as the upper limit. For each signal mass, pseudo-experiments generated using the background-only model are used to determine the distribution of upper limits in the absence of signal. The median of each distribution, along with intervals ("bands") containing the central 68% and 95% of each distribution, is found.

Table 3: Expected and observed upper limits for the production cross section of t^* pairs decaying to a top quark and gluon using unified intervals.

t^* Mass (GeV)	95% CL obs. limit (pb)	95% CL exp. limit (pb)		
		Median	68% band	95% band
300	0.95	1.6	[0.93,2.3]	[0.26,2.3]
350	0.84	1.3	[0.68,1.8]	[0.19,2.5]
400	0.89	1.1	[0.55,1.6]	[0.16,2.3]
450	0.80	0.65	[0.37,1.0]	[0.083,1.4]
500	0.73	0.52	[0.31,0.80]	[0.13,1.0]
550	0.46	0.37	[0.21,0.57]	[0.092,0.74]
600	0.42	0.30	[0.17,0.48]	[0.088,0.61]
650	0.38	0.23	[0.15,0.40]	[0.051,0.50]
700	0.29	0.18	[0.10,0.30]	[0.054,0.40]
750	0.24	0.14	[0.088,0.23]	[0.025,0.33]
800	0.22	0.13	[0.070,0.22]	[0.025,0.28]
850	0.16	0.11	[0.060,0.17]	[0.013,0.23]
900	0.13	0.087	[0.052,0.14]	[0.012,0.20]
950	0.14	0.082	[0.043,0.13]	[0.011,0.19]
1000	0.16	0.075	[0.043,0.12]	[0.0076,0.18]

Figure 7 shows the observed and expected 95% CL upper limits on the production cross section of t^* pairs as a function of the quark mass using unified intervals. Using the intersection of this cross section limit and the cross section predicted by the model, we find the median expected limit is 763 GeV with the central 68% of the limit distribution falling in the range [699,833] GeV. The observed limit is 703 GeV. Figure 8 shows the observed and expected 95% CL upper limits on the production cross section of $\tilde{b}$ pairs using unified intervals. Using the intersection

of this cross section limit and the cross section predicted by the model minus its theoretical uncertainty, we find the median expected limit is 298 GeV with the central 68% of the limit distribution falling in the range [283,347] GeV. The observed limit is 326 GeV.

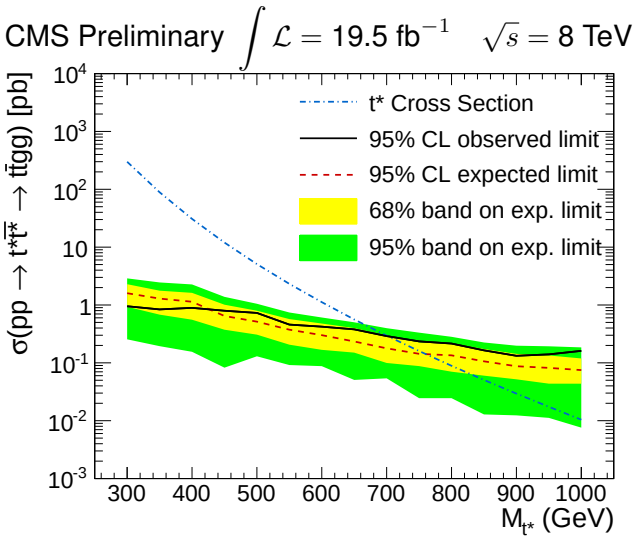

Figure 7: Observed and expected 95% CL upper limits on the production cross section of t^* pairs as a function of t^* mass using unified intervals. The difference between observed and expected limits is correlated between neighboring signal points.

We also construct CL_s intervals [38, 39] on the signal cross section and observe similar results. We find that the median expected CL_s-upper limit on the production cross section of t^* pairs is 754 GeV with the central 68% of the limit distribution falling in the range [716,779] GeV. The measured CL_s-upper limit is 717 GeV. We find that the median expected CL_s-upper limit on the production cross section of $\tilde{b}$ pairs is 295 GeV with the central 68% of the limit distribution falling in the range [282,304] GeV. The measured CL_s-upper limit is 307 GeV.

We show example plots of signal distributions created by fixing the signal cross section to the upper limit shown in Tables 3 and 4 and maximizing the likelihood function over the remaining parameters. The particular regions displayed show that the light jet shapes and invariant mass shapes have different discrimination power for different signal models. Figure 9 shows the

Table 4: Expected and observed upper limits for the production cross section of $\tilde{b}$ pairs decaying to a top quark and light quark using unified intervals.

$\tilde{b}$ Mass (GeV)	95% CL obs. limit (pb)	95% CL exp. limit (pb)		
		Median	68% band	95% band
250	2.1	2.2	[1.0,3.1]	[0.50,4.1]
300	1.0	1.8	[1.1,2.5]	[0.35,3.3]
350	1.3	1.5	[0.73,2.0]	[0.25,2.6]
400	0.92	0.76	[0.47,1.2]	[0.18,1.6]
450	0.68	0.48	[0.27,0.72]	[0.067,0.94]
500	0.53	0.37	[0.22,0.56]	[0.051,0.77]
550	0.37	0.27	[0.15,0.41]	[0.043,0.55]
600	0.30	0.21	[0.11,0.32]	[0.031,0.43]

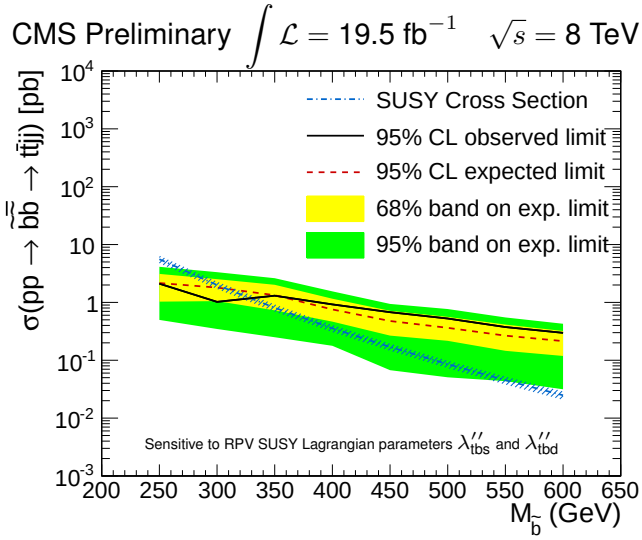

Figure 8: Observed and expected 95% CL upper limits on the production cross section of $\tilde{b}$ squark pairs as a function of $\tilde{b}$ mass using unified intervals. The difference between observed and expected limits is correlated between neighboring signal points. The parameters λ''_{tbs} and λ''_{tbd} are from the RPV SUSY Lagrangian [10] and control the branching fraction from $\tilde{b}$ squarks to top quarks and down-type quarks.

750 GeV t^* point where the search sensitivity is coming almost entirely from the invariant mass shape in SR3. Figure 10 shows the 350 GeV $\tilde{b}$ point where the invariant mass distributions peak in similar locations to the background and the search sensitivity is mainly due to the two-dimensional light jet distribution.

The limits at large masses are consistently above the median expected limit because those signal distributions are all concentrated in the SR3 region and the invariant mass shapes have a large overlap. An upwards fluctuation in the data accommodates more signal for all of those models, creating a correlated difference between observed and expected limits.

7 Summary

This note describes a search for pair production of BSM resonances decaying into a top quark and another parton. Two models, spin-3/2 excited top quark (t^*) and R-parity violating bottom squark ($\tilde{b}$), pair production, are considered in detail. We use the full dataset collected by the CMS experiment from proton-proton collisions at $\sqrt{s} = 8$ TeV. The experimental signature consists of two leptons (e or μ), two jets identified as coming from the decay of a b quark, and two jets identified as coming from light flavor quarks or gluons. The dominant background is due to top quark pair production with additional jets from initial or final-state radiation. We reconstruct pairs of top quarks and light jets to extract an upper limit on the mass and cross section of t^* or $\tilde{b}$ pair production, using an unbinned maximum likelihood fit to the data. The observation is in agreement with the standard model-only expectation. The most discrepant point is the 1000 GeV t^* which is outside the 1σ expected limit band but inside the 2σ band. Upper limits on the production cross section of t^* pairs are set for masses between 300 and 1000 GeV. For $\tilde{b}$ squarks, upper limits on the pair production cross section are set for masses between 250 and 600 GeV. Lower limits on the masses are set by comparing the results to predicted cross sections. At the 95% confidence level, t^* masses between 300 GeV and 703 GeV

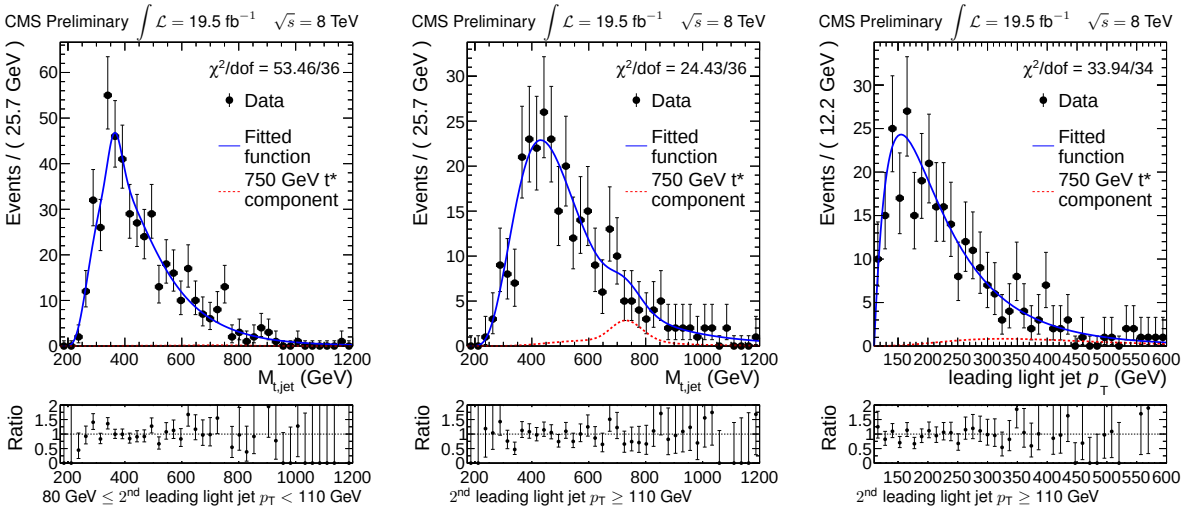

Figure 9: Results of likelihood maximization with signal cross section set to the calculated upper limit of the 750 GeV t* point. The solid line shows the fitted function, the dashed line shows the signal component, and the points show the data. The ratio of the data to the fitted function is also shown. From left to right, these are the invariant mass distribution in SR2, the invariant mass distribution in SR3, and the leading light jet p_T distribution for events in SR3.

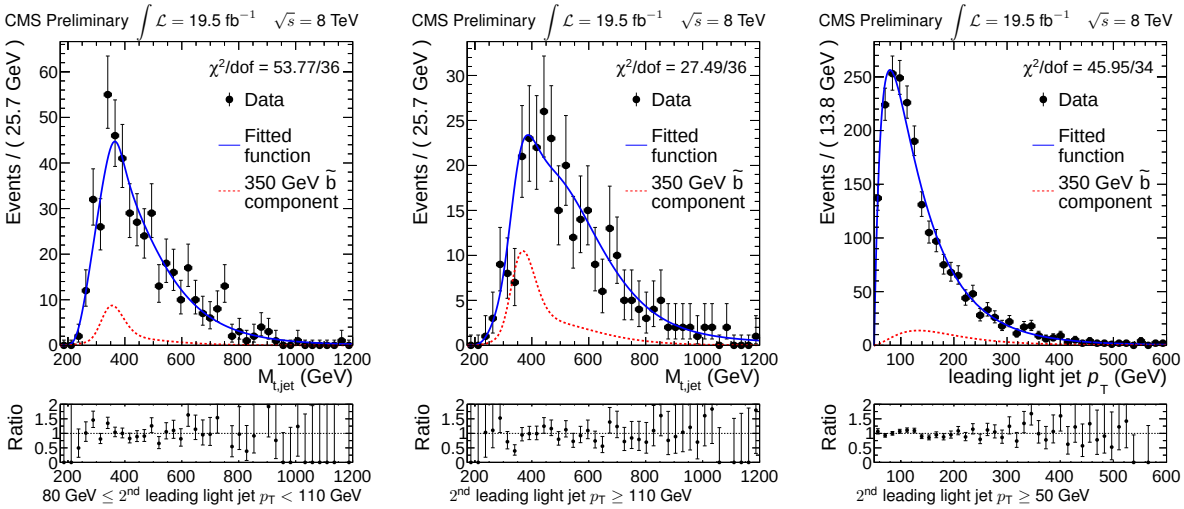

Figure 10: Results of likelihood maximization with signal cross section set to the calculated upper limit of the 350 GeV $\tilde{b}$ point. The solid line shows the fitted function, the dashed line shows the signal component, and the points show the data. The ratio of the data to the fitted function is also shown. From left to right, these are the invariant mass distribution in SR2, the invariant mass distribution in SR3, and the leading light jet p_T distribution for events with second leading light jet greater that 50 GeV.

and $\tilde{b}$ masses between 250 GeV and 326 GeV are excluded.

References

[1] S. L. Glashow, "Partial-symmetries of weak interactions", *Nucl. Phys.* **22** (1961) 579, doi:10.1016/0029-5582(61)90469-2.

[2] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264, doi:10.1103/PhysRevLett.19.1264.

[3] A. Salam, "Weak and electromagnetic interactions", in *Elementary particle physics: relativistic groups and analyticity*, N. Svartholm, ed., p. 367. Almqvist & Wiskell, 1968. Proceedings of the eighth Nobel symposium.

[4] C. Burges and H. J. Schnitzer, "Virtual Effects of Excited Quarks as Probes of a Possible New Hadronic Mass Scale", *Nucl. Phys. B* **228** (1983) 464, doi:10.1016/0550-3213(83)90555-2.

[5] B. Moussallam and V. Soni, "Production of heavy spin 3/2 fermions in colliders", *Phys. Rev. D* **39** (1989) 1883, doi:10.1103/PhysRevD.39.1883.

[6] R. M. Harris, "Discovery mass reach for excited quarks at hadron colliders", in *Proceedings of DPF/DPB Summer Study on New Directions for High Energy Physics*, pp. 1010–1014. June 25-July 12, 1996. arXiv:hep-ph/9609319.

[7] B. Hassanain, J. March-Russell, and J. Rosa, "On the possibility of light string resonances at the LHC and Tevatron from Randall-Sundrum throats", *JHEP* **07** (2009) 077, doi:10.1088/1126-6708/2009/07/077, arXiv:0904.4108.

[8] W. Stirling and E. Vryonidou, "Effect of spin-3/2 top quark excitation on $t\bar{t}$ production at the LHC", *JHEP* **01** (2012) 055, doi:10.1007/JHEP01(2012)055, arXiv:1110.1565.

[9] D. A. Dicus, D. Karabacak, S. Nandi, and S. K. Rai, "Search for spin-3/2 quarks at the Large Hadron Collider", *Phys. Rev. D* **87** (2013) 015023, doi:10.1103/PhysRevD.87.015023, arXiv:1208.5811.

[10] C. Csaki, Y. Grossman, and B. Heidenreich, "MFV SUSY: A Natural Theory for R-Parity Violation", *Phys. Rev. D* **85** (2012) 095009, doi:10.1103/PhysRevD.85.095009, arXiv:1111.1239.

[11] CMS Collaboration, "Search for pair production of new physics resonances decaying to a top quark and jet in the lepton+jets decay channel", CMS Physics Analysis Summary CMS-PAS-B2G-12-014, (2013).

[12] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **03** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[13] CMS Collaboration, "Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, (2009).

[14] CMS Collaboration, "Properties of the Higgs-like boson in the decay H to ZZ to 4ℓ in pp collisions at $\sqrt{s} = 7$ and 8 TeV", CMS Physics Analysis Summary CMS-PAS-HIG-13-002, (2013).

[15] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[16] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[17] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[18] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005, arXiv:0802.1188.

[19] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **08** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[20] 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, (2013).

[21] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[22] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **0605** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[23] R. Field, "Early LHC Underlying Event Data - Findings and Surprises", in *22nd Hadron Collider Physics Symposium (HCP 2010)*. 2010. arXiv:1010.3558.

[24] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[25] R. Corke and T. Sjöstrand, "Interleaved parton showers and tuning prospects", *JHEP* **03** (2011) 032, doi:10.1007/JHEP03(2011)032, arXiv:1011.1759.

[26] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **07** (2002) 012, doi:10.1088/1126-6708/2002/07/012.

[27] J. Allison et al., "Geant4 developments and applications", *IEEE Transactions on Nuclear Science* **53** (February, 2006) 270–278, doi:10.1109/TNS.2006.869826.

[28] L. Sonnenschein, "Analytical solution of ttbar dilepton equations", *Phys. Rev. D* **73** (2006) 054015, doi:10.1103/PhysRevD.78.079902,10.1103/PhysRevD.73.054015, arXiv:hep-ph/0603011.

[29] H.-C. Cheng et al., "Mass determination in SUSY-like events with missing energy", *JHEP* **12** (2007) 076, doi:10.1088/1126-6708/2007/12/076, arXiv:0707.0030.

[30] Particle Data Group, J. Beringer et al., "Review of Particle Physics (RPP)", *Phys. Rev. D* **86** (2012) 010001, doi:10.1103/PhysRevD.86.010001.

[31] CMS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ events with dilepton final states in pp collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C* **72** (2012) 2202, doi:10.1140/epjc/s10052-012-2202-z, arXiv:1209.2393.

[32] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **06** (November, 2011) 11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[33] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-002, (2010).

[34] CMS Collaboration, "Performance of CMS muon reconstruction in pp collisions at $\sqrt{s} = 7$ TeV", *JINST* **07** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002.

[35] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", arXiv:1101.0538.

[36] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS Physics Analysis Summary CMS-PAS-LUM-13-001, (2013).

[37] G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals", *Phys. Rev. D* **57** (April, 1998) 3873–3889, doi:10.1103/PhysRevD.57.3873, arXiv:physics/9711021.

[38] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[39] A. L. Read, "Presentation of search results: the CL_s technique", *J Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

[40] LHC Higgs Combination Group, the ATLAS Collaboration, and the CMS Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report ATL-PHYS-PUB-2011-11, CMS-NOTE-2011-005., (2011).

[41] K. Cranmer, "Statistical Challenges for Searches for New Physics at the LHC", in *Statistical Problems in Particle Physics, Astrophysics And Cosmology: Proceedings of Phystat05*, L. Lyons, ed., pp. 112–124. Oxford, UK, September 12-15, 2005. doi:10.1142/9781860948985_0026.