

Search for new physics in events with jets, b-tagged jets, and large missing transverse momentum in the all-hadronic channel at $\sqrt{s} = 13$ TeV

Samuel Louis Bein for the CMS Collaboration^{1,3,a}

¹ *Universitaet Hamburg, Mittelweg 177, 20148 Hamburg, Germany*

Abstract. A search for new physics in events with jets, b-tagged jets, missing transverse momentum, and no leptons, corresponding to an integrated luminosity of 35.9 fb^{-1} collected by the CMS experiment at $\sqrt{s} = 13$ TeV, is presented. No significant excess of events on top of the standard model background expectation is observed. Results are interpreted in terms of a number of simplified supersymmetry models featuring squark and gluino production, and in terms of the pMSSM-19. For simplified models, lower limits on the gluino (squark) mass are established in the range 1.80-1.95 TeV (1.00-1.05 TeV), assuming a massless lightest supersymmetric particle.

1 Introduction

Many physics models beyond the SM predict an excess of final states with a large number of jets and missing transverse momentum to be seen in high-energy proton-proton (pp) collision events. Notably, R-parity conserving supersymmetry (SUSY) models featuring squarks and gluinos with masses of order 1 TeV would be expected to manifest with these signatures at the LHC. These models can famously resolve the fine tuning issue known as the large hierarchy problem [1] and potentially provide a particle-based account of dark matter. This note describes the methodology, results, and an interpretation of a search for evidence of new physics manifesting in such a manner in data collected by the CMS detector [2]. The search targets a broad range of signal processes, a few examples of which are shown in Fig. 1.

2 The CMS Detector

The Compact Muon Solenoid (CMS) is a nearly azimuthally symmetric particle detector centered on Interaction Point 5 of the LHC. Its geometry integrates a barrel plus endcap design with full (partial) pseudorapidity coverage up to $|\eta| < 2.4$ (5.0). Beginning with the innermost component and working outward, CMS features a silicon pixel tracker, silicon strips tracker, electromagnetic calorimeter, hadronic calorimeter, and muon detection system. Situated between the hadronic calorimeter and muon system is a superconducting solenoidal magnet with diameter 6 m which provides a uniform 3.8 T B-field inside the solenoidal volume. The B-field bends charged particles as they traverse the

^ae-mail: samuel.bein@desy.de

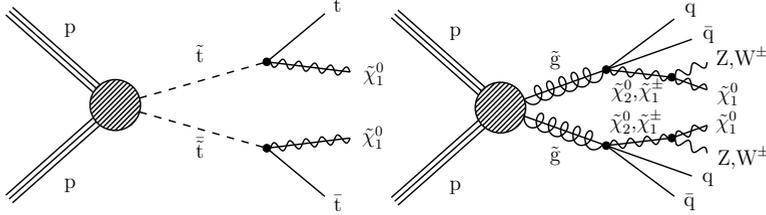


Figure 1. Two of several simplified model diagrams motivating the search strategy.

tracker and calorimeters, allowing an accurate reconstruction of the trajectories and momenta of electrons, photons, muons, and hadrons. A detailed description of the CMS detector is given in [3].

3 Object and event selection

Events considered for the search are required to pass a set of Level 1 and HLT triggers with triggering criteria based primarily on the missing transverse momentum p_T^{miss} (similarly, H_T^{miss}). These triggers require the online H_T^{miss} be larger than 120 GeV at most. A tighter requirement on the H_T^{miss} is placed offline to ensure negligible trigger inefficiencies in the signal region. Further offline event selection requires a minimum value of the total transverse energy H_T , a veto on identified leptons to suppress contributions from SM events with W bosons, and the requirement that the four highest- p_T jets have a significant azimuthal separation from the missing transverse momentum vector to suppress background contributions from multi-jet events with severely mis-measured jets. The object- and event-level selection criteria referred to as the baseline selection are specified below:

- $n_{\text{jet}} \geq 2$, where n_{jet} is the multiplicity of jets with $p_T > 30$ GeV and $|\eta| < 2.4$;
- $H_T > 300$ GeV, where $H_T = \sum_i (p_T)_i$, where i indexes jets with $p_T > 30$ GeV and $|\eta| < 2.4$;
- $H_T^{\text{miss}} > 300$ GeV, where $H_T^{\text{miss}} = |-\sum_j (\vec{p}_T)_j|$, where j indexes jets with $p_T > 30$ GeV and $|\eta| < 5.0$;
- $n_l = 0$, where n_l is the multiplicity of reconstructed and isolated electrons or muons with $p_T > 10$ GeV and $|\eta| < 2.5$ or 2.4, respectively;
- $\Delta\phi(\vec{H}_T^{\text{miss}}, (\vec{p}_T)_{j_1, j_2, j_3, j_4}) > 0.5, 0.5, 0.3, 0.3$.

Jets are clustered by applying the anti- k_r clustering algorithm [4] with distance parameter 0.4 to the reconstructed particles in each event. Events and jets are required to satisfy a set of quality criteria to ensure the suppression of anomalous events that arise from detector-related noise and inefficiencies. Events are rejected if they are found to contain an isolated track or if they satisfy $H_T > H_T^{\text{miss}}$.

Following the baseline selection, events are further categorized into 174 non-overlapping search regions, defined by rectangular cuts on the H_T , H_T^{miss} , n_{jet} , and the number of b-tagged jets n_b . The boundaries of the signal regions are indicated in figure 2. The 174 search regions comprise every combination of a numbered region in figure 2 (left), with each n_{jet} and n_b region shown in Fig. 2 (right) or $n_{\text{jet}} = 2$; regions 4 and 7 are dropped for bins with $n_{\text{jet}} \geq 7$. This binning increases the sensitivity of the search to a broad range of models that may manifest signal events in different observable phase space regions and provides a means of classifying signal models in the case of an observed excess.

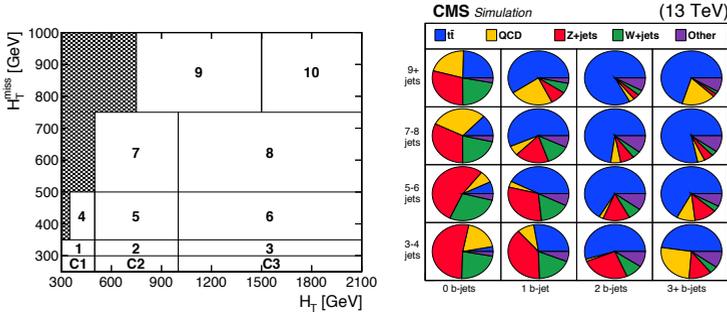


Figure 2. Left: boundaries of the signal regions (numbered) and control regions (C1, C2, C3) in the H_T – H_T^{miss} plane. Right: boundaries of the signal regions in the plane of n_{jet} and n_b , with the relative contribution to background events from the four primary sources, integrating over all H_T and H_T^{miss} . $n_{\text{jet}} = 2$ bins are not shown.

4 SM background estimation

Four SM processes constitute the primary source of background events in the search region: events containing leptons that fail some aspect of the selection, events with hadronically-decaying tau leptons produced in association with jets, events with invisibly-decaying Z-bosons produced in association with jets, and QCD multi-jet events in which the energy of one or more jets is significantly mis-measured. A description of the procedures used to estimate these backgrounds follows.

4.1 Lost lepton background

The lost lepton background arises from SM events in which a lepton is produced but is either not within the acceptance, not reconstructed, or not isolated. Expected event counts are estimated using a single-lepton control region constructed by applying the full baseline selection to events collected by the signal triggers but with an inverted lepton veto. Scaling factors based on lepton acceptance, reconstruction, an isolation efficiencies (ϵ_{acc} , ϵ_{rec} , and ϵ_{iso} , respectively), are applied to each lost lepton control region to give the expected count in the signal region. The estimated number of lost-lepton background events in the signal region is equal to

$$\hat{N}_{0\ell} = (\epsilon N_{1\ell}^{\text{obs}}) \frac{1-x}{x}, \quad x \equiv \epsilon_{\text{acc}} \epsilon_{\text{rec}} \epsilon_{\text{iso}}, \quad (1)$$

where $N_{1\ell}^{\text{obs}}$ is the observed number of events in the single-lepton control region. The acceptance efficiency ϵ_{acc} is taken directly from a simulated single-lepton event sample, and ϵ_{rec} and ϵ_{iso} are obtained from simulation that has been corrected by scale factors accounting for mis-modeling of the lepton identification efficiency. Additional terms are included to account for background arising from events in which two or more leptons fail to be identified. Figure 3 shows the comparison in each search bin between the prediction obtained by applying the above methodology to the simulated sample with the counts obtained directly from simulation. The dominant source of uncertainty in the most sensitive bins is the statistical uncertainty from the single-lepton control region event counts, which ranges as high as 70%.

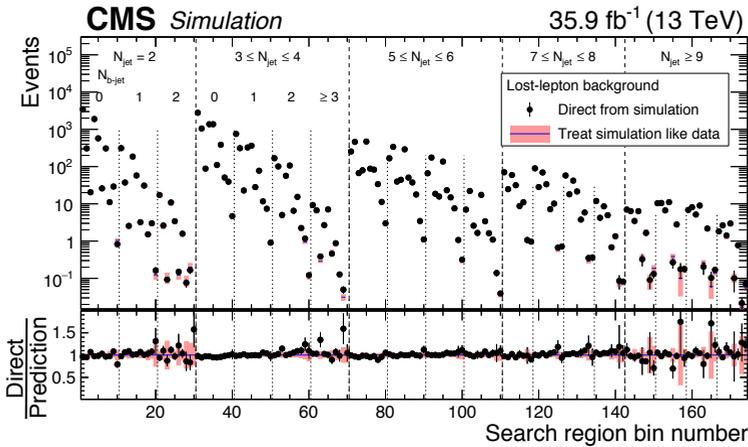


Figure 3. The result of the lost lepton background prediction method applied to simulated even samples compared with the counts obtained directly from simulation, as published in [2]. Uncertainties shown are statistical.

4.2 Hadronic τ background

The hadronic τ background consists of SM events in which a tau lepton is produced and decays into a single tau neutrino and hadrons. Such events evade the electron and muon vetoes, and can therefore populate the signal region. Lepton universality ensures that the kinematics of events with leptons will be lepton flavor independent, assuming the lepton energies are much greater than the τ mass. This fact motivates the use of real single-muon events to arrive at an estimate of the expected rate of the hadronically decaying tau events. Events in the single-muon control region are collected with a single-muon trigger with a lower threshold on the online muon p_T of approximately 30 GeV. The momentum of each muon is then artificially smeared according to a sampling of p_T - and η -dependent tau energy response distributions obtained from events simulated with a full CMS detector simulation. Events are weighted to account for the efficiency of the primary trigger as well as the known branching fraction of tau leptons into hadrons, yielding a prediction for the rate of this background in each signal region.

4.3 Z to invisible background

An irreducible source of background arises from events in which an invisibly-decaying Z boson is produced in association with two or more jets. This background is estimated using samples of events containing either two opposite-sign electrons, two opposite-sign muons, or a single photon. The close similarity between the kinematics of events with photons and events with Z bosons in the limit of large boson p_T ensures that the photon sample provides a good description of the background in the signal region. Single-photon events are collected with a trigger requiring a photon with online $p_T > 175$ GeV to be reconstructed in an event. Events are weighted to account for inefficiencies in the offline photon identification. The reconstructed photon is then artificially removed from each event, yielding a sample that closely resembles events with invisibly-decaying Z bosons, with only small differences arising from matrix element differences in the boson production modes. These differences are accounted for by applying the ratio of Z to γ event counts as a weight to control region events, where this ratio is derived in simulation and corrected using real data. The predicted number of

background events in a given signal region with no b-tags is given by

$$\hat{N}_{Z \rightarrow \nu\bar{\nu}} = \mathcal{R}\mathcal{R}_{\text{sim}}^{\text{obs}} \cdot \mathcal{R}^{\text{sim}}[Z(\nu\bar{\nu})/\gamma] \cdot (1/\epsilon) \cdot N_{\gamma}^{\text{obs}}, \quad (2)$$

where N_{γ}^{obs} is the number of observed events in the single-photon control region, ϵ is the efficiency of photon identification, $\mathcal{R}^{\text{sim}}[Z(\nu\bar{\nu})/\gamma]$ is the ratio of Z to γ events obtained from simulation, and $\mathcal{R}\mathcal{R}_{\text{sim}}^{\text{obs}}$ is the double ratio, or the ratio between real and simulated data of the ratio of Z to γ events. The prediction in bins with $n_b > 0$ is made by multiplying the predicted $n_{\text{jet}} = 0$, count by k-factors derived from real event samples with two opposite-sign, same-flavor leptons. The dominant source of uncertainty in the predicted counts is associated with the double ratio and ranges from 10 to 100%.

4.4 QCD multi-jet background

Additional background is attributable to QCD multi-jet events in which at least one of three scenarios occurs. First, if the energy of a jet is severely under-measured or over-measured, the jets in an event can exhibit a large momentum imbalance, and the event can have artificially large $H_{\text{T}}^{\text{miss}}$. The vast majority of such events are rejected by the requirement that the $\Delta\phi$ between the particle-level $H_{\text{T}}^{\text{miss}}$ and leading jets be substantial. However, QCD events in which the energy of two or more jets is mis-measured evade the $\Delta\phi$ -based rejection. Second, QCD events can feature heavy-flavor jets due to gluon splitting into $b\bar{b}$ pairs, and the ensuing hadronization and decay can result in neutrinos and therefore increased $H_{\text{T}}^{\text{miss}}$. Finally, the presence one or more jets that fails the jet acceptance criteria can lead to an inflated $H_{\text{T}}^{\text{miss}}$, further increasing contributions in the signal region. Two independent procedures are employed to estimate the QCD background.

The first procedure is the rebalance and smear method. This method builds a QCD event sample based on real data using a model of the jet energy response. The prediction accounts for cases in which one or more of the above scenarios occurs or occur simultaneously. Events are collected with a set of pure H_{T} triggers with thresholds ranging from 200 to 900 GeV, with only standard event filters and the lepton vetoes applied offline to the events. In the first step, each event is rebalanced by a rescaling of the various jet momenta by scale factors that maximize the posterior density for the ideally-measured QCD event; the posterior density is given by

$$P(\vec{J}_{\text{true}}|\vec{J}_{\text{meas}}) \sim P(\vec{J}_{\text{meas}}|\vec{J}_{\text{true}})\pi(\vec{H}_{\text{T}}^{\text{miss}}, \vec{p}_{\text{T},j_1}), \quad (3)$$

where \vec{J}_{true} ($\vec{J}_{\text{measured}}$) is the set of true (measured) four-vectors of jets in the event, and the right most factor is the prior probability density, given by

$$\pi(\vec{H}_{\text{T}}^{\text{miss}}, \vec{p}_{\text{T},j_1}) = P(H_{\text{T}}^{\text{miss}})P(\Delta\phi_{\vec{H}_{\text{T}}^{\text{miss}}, j_{1(b)}}), \quad (4)$$

where the prior is factorized as the product of the particle-level $H_{\text{T}}^{\text{miss}}$ and azimuthal separation between the $H_{\text{T}}^{\text{miss}}$ and the highest- p_{T} jet (in the case of 0 b-tags) or highest- p_{T} b-tagged jet (in the case of one or more b-tags), computed from samples of simulated QCD events. The prior is binned in of H_{T} and n_b to account for differences in jet activity and the presence of neutrinos.

The likelihood $P(\vec{J}_{\text{meas}}|\vec{J}_{\text{true}})$ is computed as the product of the jet response functions over all jets in the event, where a response function of a jet is the distribution of expected reconstructed p_{T} values given a true p_{T} and η of a jet:

$$P(\vec{J}_{\text{meas}}|\vec{J}_{\text{true}}) = \prod_{\text{jet}} P(p_{\text{T},\text{jet,meas}}|\vec{p}_{\text{T},\text{jet,true}}), \quad (5)$$

where the detector response to each jet is treated as independent from that to each other jet. The response distributions are computed in simulation that has been corrected by data-to-simulation scale

factors as the distribution of reconstructed p_T values for a given p_T and η of a matched, isolated generator-level jet. The effect of the rebalancing is to force each real event into a topology consistent with a QCD event without any detector smearing, and most importantly with characteristically low H_T^{miss} . In the smearing step, the rebalanced events are subjected to a smearing of the energy of every jet by a random sampling of each jet’s response function. The result of the smearing is sample of events that resemble QCD events at the level of the reconstruction, and is called the prediction sample. The prediction sample is more than 99% pure in real QCD events, as rebalancing removes the high- H_T^{miss} tail of the real event sample, which can be populated by events from electroweak processes. The prediction is derived by placing the baseline and search region cuts on the events in the prediction sample and taking the weighted count as the prediction in each signal region. Each rebalanced event is copied and smeared hundreds of times, each time with an independent random sampling of the jet response functions, allowing the prediction sample size to grow large. The weight on the prediction events, which incorporates the trigger prescale and the number of times the event is smeared, is typically 0.1. The largest systematic uncertainty is associated with the jet energy resolution scale factors,

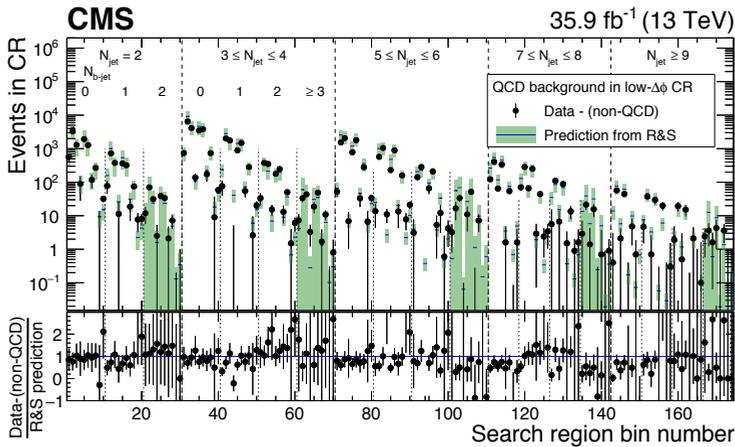


Figure 4. The result of the rebalance and smear prediction method applied to real events and evaluated in each inverted $\Delta\phi$ control region bin, compared with the observed count less the predicted non-QCD count, as published in [2]. Uncertainties shown are statistical and systematic.

and ranges from 20–100%. The results of a bin-by-bin test of the method is shown in figure 4, and this information projected into kinematic regions is shown in 5. The rebalance and smear prediction is used for the final background estimate in the interpretation.

The second procedure is referred to as the inverted $\Delta\phi$ method. This method uses a data control region obtained by applying the full baseline selection to real events but with the complimentary selection on the $\Delta\phi$ observables applied, as well as an independent control region defined by the application of the full baseline selection but in a low H_T^{miss} sideband from 250 to 300 GeV. Factors relating the counts in the low $\Delta\phi$ control region to those in the signal regions are derived in real events in the H_T^{miss} sideband region, and high-to-low H_T^{miss} ratios are computed using simulated events. Estimates on the QCD count in the signal regions are made by multiplying the observed count in the low $\Delta\phi$ control region by the high-low $\Delta\phi$ ratios, after contributions from other backgrounds are subtracted from all control regions. This prediction returns values consistent with the rebalance and

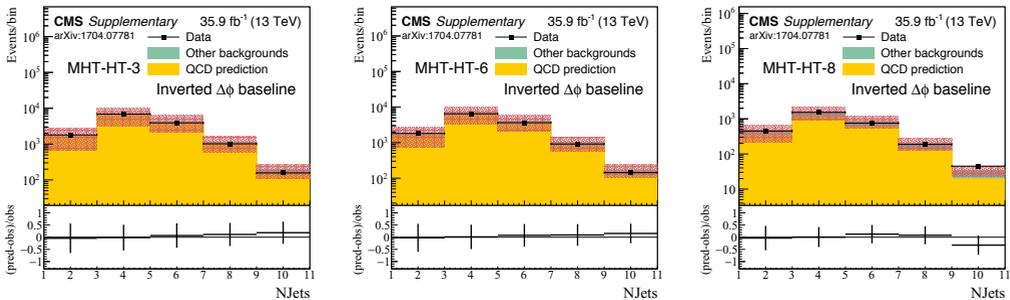


Figure 5. Comparison between the predicted and observed jet multiplicity distributions in the low $\Delta\phi$ control region in numbered kinematic regions defined in Fig. 2. The rebalance and smear QCD prediction is shown in yellow and the other backgrounds, determined by methods described herein, are in green. The red uncertainty band indicates the statistical and systematic uncertainties in the background prediction summed in quadrature.

smear prediction, and serves as a cross check for the QCD estimation procedure. Other backgrounds are small and are estimated using simulation.

5 Results and Interpretation

The observed counts in each search bin are shown in Fig. 6, along with the expected background counts. No significant excess of events over the expectation is observed. The results are interpreted in terms of various simplified model spectra (SMS), and in terms of the pMSSM.

For the SMS interpretation, 95% confidence level upper limits are computed on the signal cross section throughout a subset of the plane of the parent and LSP masses describing the model. The procedure for computing limits is described in detail in [2] and makes use of asymptotic formulae [5] as well as the CLs criterion [6, 7]. Correlations among the expected background and signal counts are incorporated in a global fit of a 174-bin likelihood. These results are shown for two simplified models in Fig. 7. Upper limits on the masses can be read off the plot by the lines where the nominal signal cross section is equal to the upper limit cross section value, and an upper limit of 1800 (950) GeV is observed on the mass of the gluino (top squark), assuming a massless LSP.

For the pMSSM interpretation, a 19-parameter scan of model space described in detail in [8] is used to compute the survival probability (SP) for various sparticle mass combinations, where the CLs limit setting procedure is used as the indicator of exclusion. The SP is shown in the planes of $m_{\tilde{g}}$ vs $m_{\tilde{\chi}_1^0}$ and m_{LCSP} vs $m_{\tilde{\chi}_1^0}$, where LCSP stands for the lightest colored supersymmetric particle. A comparison is given between the SP computed using the 7 and 8 TeV CMS search constraints as published in [8] and after the constraint from the multi-jet search described here. The 13 TeV search significantly impacts the SP, extending the region of minimal SP (yellow) to the maximum considered value of the gluino mass of 3 TeV—model points in this region often feature lighter SUSY particles with high cross sections. Additionally, the posterior density is computed using the simplified counts-based likelihood described in [8]. A comparison is given between the prior probability with only the pre-CMS constraints, the posterior density after the 7 and 8 TeV constraints as published in [8], and the posterior after the incorporation the multi-jet search described here. The posterior density is significantly impacted by the 13 TeV observations. The density is squeezed into the large gluino mass region, a consequence of the finite limits on the allowed range of particle masses in the scan.

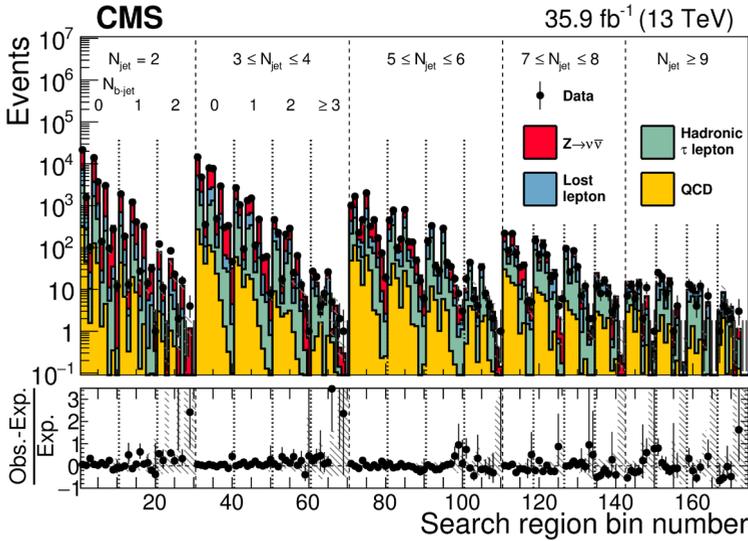


Figure 6. The observed count in each signal region, compared with the sum of the predictions for the various background contributions, as published in [2].

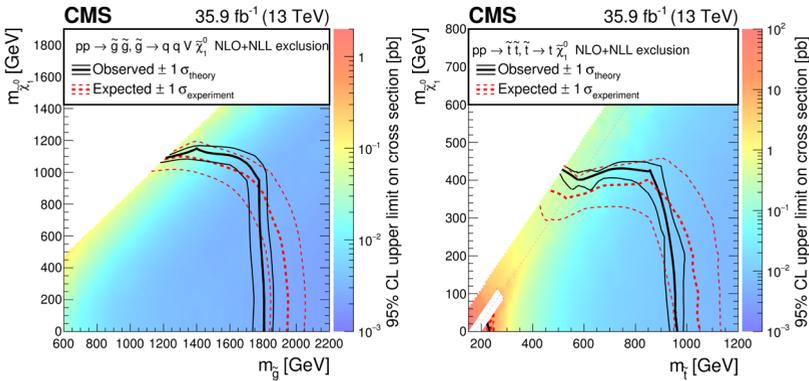


Figure 7. The color map gives the 95% CL upper limit on the cross section as calculated by the CLs procedure. Black (red) lines indicate the upper limit assuming the nominal cross section. Figures as published in [2].

6 Conclusion

A search for new physics in events with large missing transverse energy and various jet and b-tagged jet multiplicity ranges has been described in new detail. No significant evidence for BSM physics is observed. The sensitivity of the search is far reaching, and dramatically extended with respect to the 7 and 8 TeV SUSY searches. In terms of simplified models and the pMSSM, the search is able to exclude models with colored particles with masses below 1 TeV under the assumption of a massless LSP. The 174-bin search has been condensed into 12 aggregated search regions to ease recasting and re-interpreting of the analysis in a phenomenological context, and further details are provided in [2].

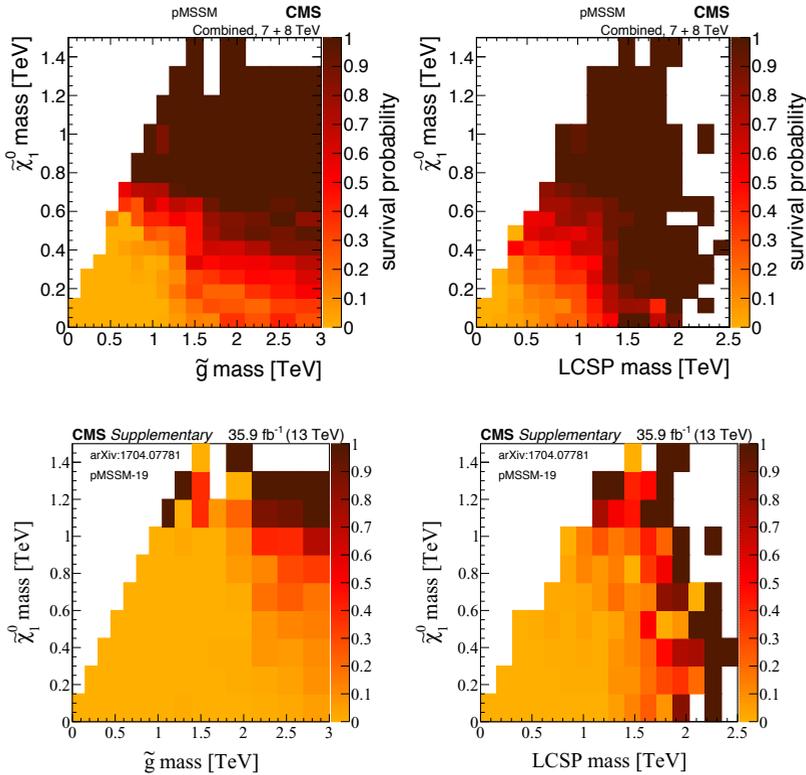


Figure 8. Survival probability (SP), calculated as the fraction of pMSSM model points that are not excluded by the multi-jet search as determined by the CLs criterion based on the full likelihood model. The top row shows the SP after the 7 and 8 TeV SUSY searches as published in [8] in the $m_{\tilde{g}}$ vs m_{LSP} (right) and m_{LCSP} vs m_{LSP} (right); the bottom row gives the SP after incorporating the observations of the 13 TeV multi-jet search.

References

- [1] Stephen P. Martin, *Advanced Series on Directions in High Energy Physics* **18**, 1-153 (2010)
- [2] CMS Collaboration, *Phys. Rev. D* **96**, 032003 (2017)
- [3] CMS Collaboration, *JINST* **3**, S08004 (2008)
- [4] M. Cacciari, G. P. Salam, and G. Soyez, *JHEP* **04**, 063 (2008)
- [5] Cowan, G., Cranmer, K., Gross, E. et al., *Eur. Phys. J. C* **71**, 1554 (2011)
- [6] T. Junk, *Nucl. Instrum. Meth. A* **434**, 435 (1999)
- [7] A. L. Read, *J. Phys. G* **28**, 2693 (2002)
- [8] CMS Collaboration, *JHEP* **10**, 129 (2016)

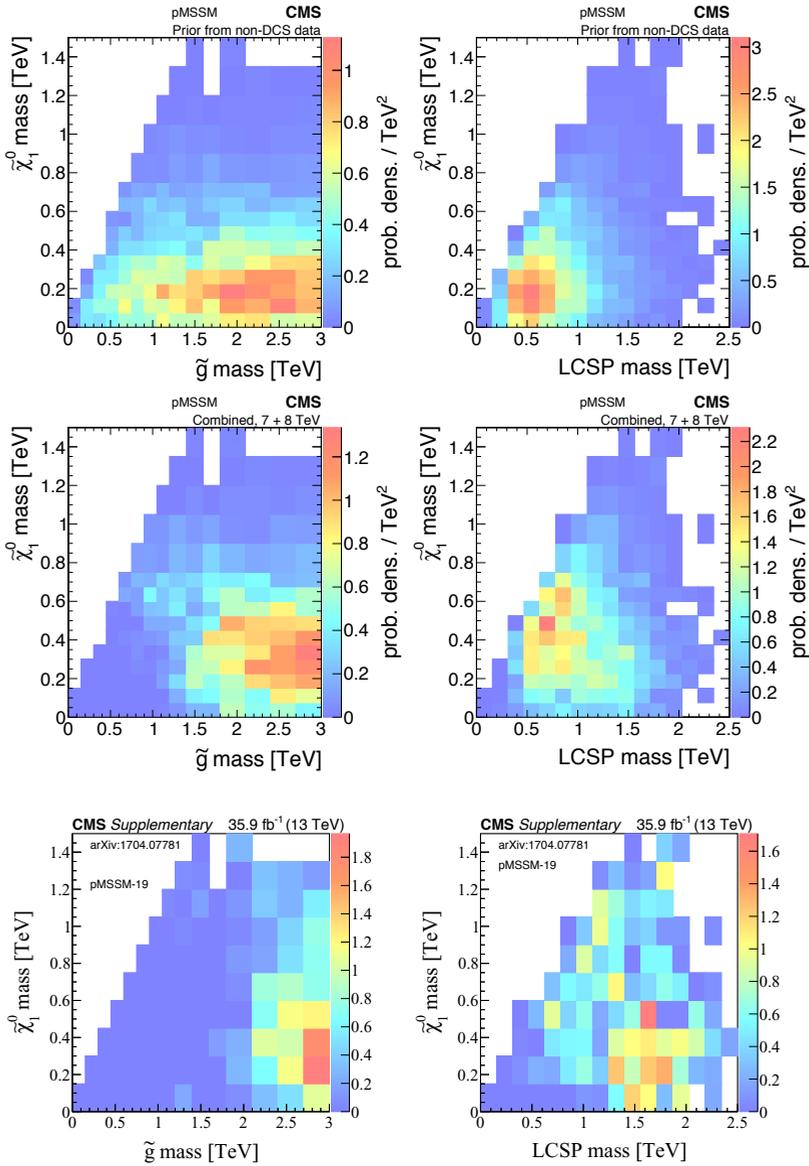


Figure 9. The pMSSM probability density corresponding to the prior (top) before any CMS constraints have been applied, the posterior density after the constraints of the 7 and 8 TeV SUSY searches have been incorporated as published in [8] (center), and the posterior density after all analyses including the 13 TeV multi-jet search have been incorporated.