CMS Physics Analysis Summary

Contact: cms-pag-conveners-susy@cern.ch 2017/08/11

Search for R-parity-violating supersymmetry in proton-proton collisions at $\sqrt{s} = 13$ TeV in events with a single lepton and large jet and bottom quark jet multiplicity

The CMS Collaboration

Abstract

Results are reported from a search for physics beyond the standard model in proton-proton collisions at a center-of-mass energy $\sqrt{s} = 13$ TeV, focusing on the signature of a single lepton, large jet multiplicity, and large bottom quark jet multiplicity, without a requirement on the missing transverse momentum in an event. The data sample corresponds to an integrated luminosity of 35.9 fb^{-1} recorded by the CMS experiment at the LHC. No excess beyond the prediction from standard model processes is observed. The results are interpreted in terms of limits on the parameter space for R-parity-violating supersymmetric extensions of the standard model using a benchmark model of gluino pair production in which each gluino decays promptly via $\tilde{\text{g}} \to \text{tbs}$. Gluinos with a mass below 1610 GeV are excluded at 95% confidence level.

1 Introduction

Searches for physics beyond the standard model (SM) are motivated by several considerations including astrophysical evidence for dark matter and theoretical problems associated with explaining the observed particle mass hierarchies in the presence of quantum corrections, i.e., the hierarchy problem [1, 2]. While the SM has been successful in describing a vast range of phenomena, its inability to describe these experimental and theoretical issues makes it an incomplete description of fundamental particles and their interactions.

Supersymmetry (SUSY) can provide a solution to these problems [3–10]. The hierarchy problem can be addressed by supersymmetry models with a sufficiently low mass top squark and gluino, and the lightest supersymmetric particle (LSP) is a potential dark matter candidate if it is stable. That stability is assured in R-parity-conserving (RPC) SUSY models, where the R-parity of a particle is defined as $(-1)^{2s+3(B-L)}$ and s, B, and L are the spin, baryon number, and lepton number of the particle, respectively.

Recent searches at the LHC have set stringent limits on RPC SUSY production. For example, within the models studied mass limits for the top squark are reaching $\sim$ 1 TeV [11, 12] and for the gluino $\sim$ 2 TeV [13–17], which creates tension with the ability of these models to explain the hierarchy problem with little fine tuning. These searches typically require signatures with a large amount of missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) resulting from the escaping LSPs. In R-parity-violating (RPV) SUSY, however, the LSP is not stable and decays to SM particles, which does not give rise to large $p_{\mathrm{T}}^{\mathrm{miss}}$ in an event. Thus, RPV SUSY models are not covered by these searches.

Given that there is no purely theoretical reason that R-parity be conserved, RPV SUSY is an important class of models that can ease the tension between natural solutions to the hierarchy problem and current experimental limits, at the cost of a SUSY dark matter candidate. In addition, the absence of a $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement can allow RPV SUSY searches to be sensitive to parameter space of RPC SUSY where only a small amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ is expected, such as in models where the mass splitting between the next-to-lightest supersymmetric particle and the LSP is small. As such, RPV SUSY searches help to complete the coverage of SUSY model space.

If R-parity violation is permitted, additional terms in the superpotential are allowed and the R-parity-violating superpotential is

$$W = \frac{1}{2}\lambda^{ijk} L_i L_j \bar{\mathrm{e}}_k + \lambda'^{ijk} L_i Q_j \overline{\mathrm{d}}_k + \mu'^{i} L_i H_u + \frac{1}{2}\lambda''^{ijk} \overline{\mathrm{u}}_i \overline{\mathrm{d}}_j \overline{\mathrm{d}}_k. \tag{1}$$

Here L_i, Q_j, and H_u are $SU(2)$ doublets corresponding to leptons, quarks, and the Higgs boson, respectively. The fields $\bar{\mathrm{e}}_k$, $\overline{\mathrm{u}}_i$, and $\overline{\mathrm{d}}_j$ are the charged lepton, up-type quark, and down-type quark $SU(2)$ singlets. Color indices are suppressed and Latin letters denote family indices. A review of RPV SUSY can be found in Ref. [18].

In this search, we focus on a particular model of R-parity violation, minimal flavor-violating (MFV) SUSY [19], in which the R-parity-violating couplings arise from the SM Yukawa couplings. This makes the third generation couplings large and those of the first two generations small, which is consistent with the strong constraints on baryon and lepton number violation involving the lightest two generations. In MFV scenarios, the gluino decays primarily via $\widetilde{\mathrm{g}} \to \mathrm{t}\widetilde{\mathrm{t}} \to \mathrm{tbs}$. The coupling λ''^{ijk} must be antisymmetric in the last two indices because of color conservation, excluding decays to tbb. The coupling to a top, bottom, and strange quark therefore provides the largest allowed coupling. Pair production of gluinos that decay in this

way is used as a benchmark signal for this analysis.

The simplified model that is used in the interpretation of this analysis makes several assumptions about the SUSY mass spectrum. It is assumed that squarks other than the top squark are much heavier than the gluino so that their effect is negligible. The top squark is assumed to be heavier than the gluino and therefore is off-shell in this decay. The off-shell decay of the top squark in this channel results in a three body decay, so searches for dijet resonances (i.e., $\tilde{\bar{t}} \to$ bs) are not applicable in this scenario. We further assume that λ''^{tbs} is large enough that the gluino decay is prompt. Although this benchmark is used for interpreting the results, we strive to structure the search to be generically sensitive to high mass signatures with large jet and b-jet multiplicity and either little or no p_T^{miss}, which are potential features of other models of physics beyond the standard model.

Previous limits on such models were obtained by the ATLAS and CMS collaborations using the 8 TeV dataset [20–23], and the ATLAS collaboration recently reported [24] a search in the same running period analysed in this paper.

We search for an excess of events with a large number of b-tagged jets, N_b, in signal regions determined as a function of the jet multiplicity, N_{jet}, and the sum of the masses of large radius jets, M_J. The quantity M_J was proposed in phenomenological studies [25–27] and was used for R-parity conserving SUSY searches by the ATLAS collaboration in the all-hadronic signature [28, 29] and by the CMS collaboration in the single-lepton signature [30].

2 Samples and Event Selection

The search uses a sample of proton-proton collision data at a center-of-mass energy of $\sqrt{s} =$ 13 TeV corresponding to an integrated luminosity of 35.9 fb^{-1}, which was collected by the CMS experiment during 2016. The central feature of the CMS detector is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are the charged particle tracking systems, composed of silicon-pixel and silicon-strip detectors, and the calorimeter systems, consisting of a lead tungstate crystal electromagnetic calorimeter (ECAL) and a brass and scintillator hadron calorimeter (HCAL). Muons are identified and measured by gas-ionization detectors embedded in the magnetic flux-return yoke outside the solenoid. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, is given in Ref. [31].

Events are selected with triggers that require either at least one jet with transverse momentum, $p_T > 450$ GeV or the scalar sum of all reconstructed jet transverse energies above 900 GeV. Trigger efficiencies are over 99% for signal events passing the analysis selection defined below.

The background predictions use Monte Carlo simulation samples with data-driven corrections. Madgraph5_aMC@NLO is used in leading-order mode [32, 33] to generate the signal samples and the dominant $t\bar{t}$, W+jets, and QCD multijet background processes. Comparison to a POWHEG [34] sample generated at next-to-leading order (NLO) shows that the NLO corrections do not have a significant impact after including data-driven corrections. The Drell-Yan, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}t\bar{t}$, and t-channel single top quark production backgrounds are generated with Madgraph5_aMC@NLO in NLO mode [32, 35], while the tW, $\bar{t}W$, and s-channel single top quark processes are generated with POWHEG. The generated samples are interfaced with PYTHIA 8.2 [36] for fragmentation and parton showering, and the detector response is simulated with GEANT4 [37]. Simulated samples are processed through the same reconstruction algorithms as the data.

The signal sample is generated by Madgraph5_aMC@NLO in leading-order mode [32, 33]. It follows the same procedure for fragmentation, parton showering and event reconstruction as the background samples.

The reconstruction of objects in an event proceeds from the candidate particles identified by the particle-flow (PF) algorithm [38], which uses information from the tracker, calorimeters, and muon systems to identify the candidates as charged or neutral hadrons, photons, electrons, or muons. Charged particle tracks are required to originate from the event primary vertex (PV), which is the reconstructed vertex with the largest value of summed physics-object p_{T}^2. The physics objects are the objects returned by a jet finding algorithm [39, 40] applied to all charged tracks associated with the vertex, plus the corresponding associated missing transverse momentum.

The charged PF candidates associated with the PV and the neutral PF candidates are clustered into jets using the anti-k_{T} algorithm [39] with distance parameter $R = 0.4$, as implemented in the FASTJET package [40]. The estimated contribution to the jet p_{T} from neutral PF candidates produced by additional proton-proton collisions in the same crossing (pileup) is removed with a correction based on the area of the jet and the average energy density of the event [41]. The jet energy is calibrated using p_{T}- and η-dependent corrections; the resulting calibrated jet is required to satisfy $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| \leq 2.4$. Each jet must also meet loose identification requirements [42] to suppress, for example, calorimeter noise. Finally, jets that have PF constituents matched to the selected lepton, as defined below, are removed from the jet collection.

A subset of the jets are "tagged" as originating from b quarks using the combined secondary vertex (CSV) algorithm [43, 44]. The tagging efficiency for b jets in the range $p_{\mathrm{T}} = 30$ to $50\,\mathrm{GeV}$ is 60–67% (51–57%) in the barrel (endcap), increasing with p_{T}. Above $p_{\mathrm{T}} \approx 150\,\mathrm{GeV}$ the efficiency decreases. The probability to misidentify jets arising from c quarks is 13–15% (11–13%) in the barrel (endcap), while the misidentification probability for light-flavor quarks or gluons is 1–2%.

Electrons are reconstructed by associating a charged particle track with an ECAL supercluster [45]. The resulting candidate electrons are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$, and to satisfy identification criteria designed to remove light-parton jets, photon conversions, and electrons from heavy flavor hadron decays. Muons are reconstructed by associating tracks in the muon system with those found in the silicon tracker [46]. Muon candidates are required to satisfy $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$.

To preferentially select leptons that originate in the decay of W bosons, leptons are required to be isolated from other PF candidates. Isolation is quantified using an optimized version of the "mini-isolation" variable originally suggested in Ref. [47], in which the transverse energy of the particles within a cone in η-ϕ space surrounding the lepton momentum vector ($\vec{p}^{\,\ell}$) is computed using a cone size that scales as $1/p_{\mathrm{T}}^{\ell}$. The mini-isolation, $I_{\mathrm{mini}}^{\mathrm{rel}} = I_{\mathrm{mini}}/p_{\mathrm{T}}^{\ell}$, is defined as the transverse energy I_{mini} of particles in a cone of radius $R^{\text{mini-iso}}$ around the lepton, divided by p_{T}^{ℓ}. The transverse energy I_{mini} is computed as the scalar sum of the p_{T} values of the charged hadrons from the PV, neutral hadrons, and photons, with a correction that estimates the average amount of energy contributed by pileup. The cone radius $R^{\text{mini-iso}}$ varies with the p_{T}^{ℓ} according to

$$R^{\text{mini-iso}} = \begin{cases} 0.2, & p_{\mathrm{T}}^{\ell} \leq 50\,\mathrm{GeV} \\ \frac{10\,\mathrm{GeV}}{p_{\mathrm{T}}^{\ell}}, & p_{\mathrm{T}}^{\ell} \in (50\,\mathrm{GeV}, 200\,\mathrm{GeV}) \\ 0.05, & p_{\mathrm{T}}^{\ell} \geq 200\,\mathrm{GeV}. \end{cases} \tag{2}$$

The p_T-dependent cone size reduces the rate of accidental overlaps between the lepton and jets in high-multiplicity or highly Lorentz-boosted events, particularly overlaps between b jets and leptons originating from a boosted top quark. Muons (electrons) must satisfy $I^{rel}_{mini} < 0.2$ (0.1).

The combined efficiency for the electron reconstruction and isolation requirements is about 50% at p^{ℓ}_T of 20 GeV, increasing to 65% at 50 GeV and reaching a plateau of 80% above 200 GeV. The combined reconstruction and isolation efficiencies for muons are about 70% at p^{ℓ}_T of 20 GeV, increasing to 80% at 50 GeV and reaching a plateau of 95% at 200 GeV.

We cluster $R = 0.4$ ("small-R") jets and the isolated lepton into $R = 1.2$ ("large-R") jets using the anti-k_T algorithm. Clustering small-R jets instead of PF candidates incorporates the jet pileup corrections, thereby reducing the dependence of the mass on pileup. The variable M_J is defined as the sum of all large-R jet masses:

$$M_J = \sum_{J_i \in \text{large-}R\text{ jets}} m(J_i). \tag{3}$$

Leptons are included in the large-R jets to include the full kinematics of the event. The technique of clustering small-R jets into large-R jets has been used previously in, for example, Refs. [30, 48].

Events are selected with a baseline requirement of one electron or muon, $M_J > 500$ GeV, and $H_T > 1.2$ TeV where H_T is the scalar sum of the transverse momenta of all the small-R jets passing the selection. The number of small-R jets, N_{jet}, is required to be at least 4, and the number of those jets that are b-tagged, N_b, is required to be at least 1.

3 Background prediction

After the baseline selection, the dominant background contribution is from the $t\bar{t}$ + jets process, with small contributions from W+jets and QCD multijet (QCD) production. Rare background contributions, classified below as "Other", come from single-top quark, $t\bar{t}$ + (W, Z, H, or $t\bar{t}$), and Drell-Yan production.

To identify signal events from high mass new particles decaying with large jet and b-quark multiplicity, the events are separated into regions based on N_{jet} and M_J. The dominant background is normalized to data in each region because they are difficult to model reliably in the kinematic tails of these variables. The N_{jet} bins are 4–5, 6–7, and ≥ 8. The M_J bins are $500 < M_J < 800$ GeV, $800 < M_J < 1000$ GeV and $M_J > 1000$ GeV, with the two highest-M_J bins merged for the $4 \leq N_{jet} \leq 5$ case because of the limited data sample size in the $M_J > 1000$ GeV region. We then search for contributions from b-quark rich signal events by examining the N_b distribution in these regions using $N_b = 1$, $N_b = 2$, $N_b = 3$, and $N_b \geq 4$ bins. The signal is expected to populate the two highest N_b bins, while the lower N_b bins provide background constraints.

The signal component is extracted by a global, binned maximum likelihood fit using the N_b distribution in each bin of N_{jet} and M_J. The N_b shape is modeled with simulation, but varied to assess the impact of mismodelling of parameters that affect it, including tagging efficiencies for heavy and light flavor jets [43, 44] and the rate of gluon splitting to $b\bar{b}$. The appropriate ranges for these parameters are determined based on the measurements in dedicated control samples and then constrained by a simultaneous fit across all bins of N_{jet} and M_J. The signal poor $N_b \leq 2$ bins allow the fit to determine the background normalization in each (N_{jet}, M_J) bin as described below.

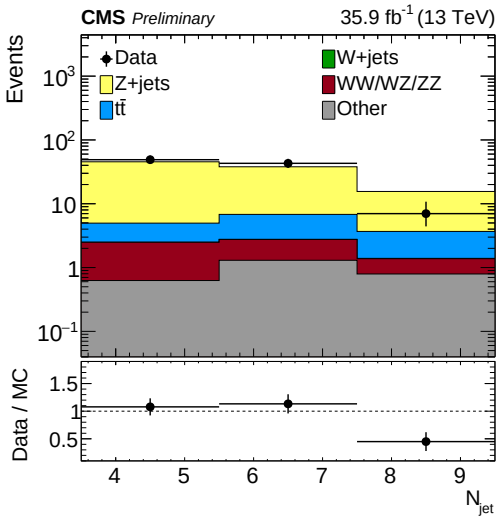

Figure 1: Comparison of data and MC simulation yields in a Z+jets control sample selected by requiring $N_{\text{lep}} = 2$, $H_{\text{T}} > 1200$ GeV, $N_{\text{b}} = 1$, and $80 < m_{\ell\ell} < 100$ GeV. The total yield of MC is normalized to the number of events in data. The uncertainty on the ratio of data to total MC (Data/MC) is from the statistics of the data sample.

The $\text{t}\bar{\text{t}}$ and QCD normalizations are allowed to float in each (N_{jet},M_{J}) bin. Contributions from QCD multijet events with a misidentified lepton are constrained using control regions with no identified leptons ($N_{\text{lep}} = 0$). These control regions are integrated in N_{b} for $N_{\text{b}} \geq 1$, and higher N_{jet} bins in the $N_{\text{lep}} = 0$ regions are used to constrain the next lower N_{jet} bins to account for jets being misidentified as a lepton. The small contribution of $\text{t}\bar{\text{t}}$ to the $N_{\text{lep}} = 0$ control regions is included using the normalization from the corresponding $N_{\text{lep}} = 1$ bins.

The N_{jet} shape of W+jets events is taken from simulation and allowed to vary based on the data-to-simulation agreement in a kinematically similar Z+jets sample selected with $N_{\text{lep}} = 2$, $H_{\text{T}} > 1200$ GeV, $N_{\text{b}} = 1$, and $80 < m_{\ell\ell} < 100$ GeV where $m_{\ell\ell}$ is the invariant mass of the two leptons. Figure 1 shows this sample's N_{jet} distribution and the data/MC scale factors. The W+jets background is then determined in the fit with one global normalization parameter and two parameters to adjust the bin to bin normalization based on the difference between scale factors in adjacent N_{jet} bins (17% between $4 \leq N_{\text{jet}} \leq 5$ and $6 \leq N_{\text{jet}} \leq 7$ and 62% between $6 \leq N_{\text{jet}} \leq 7$ and $N_{\text{jet}} \geq 8$.) After correcting the N_{jet} spectrum, residual M_{J} mismodelling is expected to be small, and so the M_{J} shape is taken from simulation.

The "Other" component is estimated from simulation. Its contribution is less than 15% of the total backgrounds in all kinematic regions considered.

4 Systematic Uncertainties

4.1 Background systematic uncertainties

The nominal simulated shape of the N_{b} distribution is allowed to vary by the inclusion of systematic uncertainties. Each uncertainty is incorporated in the fit with template N_{b} histograms to account for the effects of the systematic variation and a nuisance parameter to control the variation amplitude. The nuisance parameters are subject to Gaussian constraints, normalized so that $\mu = 0$ corresponds to the nominal N_{b} shape and $\mu = \pm 1$ corresponds to ± 1 standard deviation variation of the systematic uncertainty. These uncertainties affect only the N_{b} shape

for $\mathrm{t\bar{t}}$, QCD, and W+jets, as their normalizations are determined from data, while for the other sub-leading backgrounds the uncertainties affect both the N_b shape and normalization.

The primary source of systematic uncertainty is due to the uncertainty on the modeling of gluon splitting, as gluon splitting to $\mathrm{b\bar{b}}$ provides additional b quarks in $\mathrm{t\bar{t}}$ and QCD events and may not be properly simulated. An uncertainty on the gluon splitting rate is determined using a fit to the $\Delta R_{\mathrm{b\bar{b}}}$ distribution, defined as the ΔR between two b-tagged jets in the event, in a control sample selected with $N_\mathrm{lep} = 0$, $N_\mathrm{b} = 2$, $N_\mathrm{jet} \geq 4$ and $M_J > 500\,\mathrm{GeV}$. Events where both jets originating from a gluon splitting are b-tagged populate the low $\Delta R_{\mathrm{b\bar{b}}}$ region, while events without gluon splitting or where the splitting yields one or fewer b-tagged jets populate both the low and high $\Delta R_{\mathrm{b\bar{b}}}$ regions roughly equally. Gluon splittings can contribute fewer than 2 b-tagged jets either because the quarks are collimated into a single jet, one of the b-jets is not tagged, or because one of the quarks is too soft to form a reconstructed jet.

The fit to the $\Delta R_{\mathrm{b\bar{b}}}$ distribution is used to extract the relative contributions of events with and without gluon splitting and is performed in four bins in the range $0 \leq \Delta R_{\mathrm{b\bar{b}}} < 4.4$. In simulation the instances of gluon splitting are selected by requiring a gluon with $p_\mathrm{T} > 30\,\mathrm{GeV}$ that decays to b quarks, which is used to define three categories: events with gluon splitting resulting in two b-tagged jets (denoted GSbb), with gluon splitting resulting in one or fewer b-tagged jets (GSb), and without any gluon splitting (no GS). In the fit, the GSbb and GSb contributions are varied together with a single normalization parameter.

The $\Delta R_{\mathrm{b\bar{b}}}$ fit extracts a weight of 0.77 ± 0.09 for gluon splitting events and a weight of 1.21 ± 0.08 for non-gluon splitting events. Figure 2 shows the post-fit $\Delta R_{\mathrm{b\bar{b}}}$ distribution compared to data where both are plotted with finer binning and the GSbb and GSb categories are plotted separately to demonstrate the difference in shapes. These weights are propagated to the N_b fit as a systematic uncertainty, not a correction to the central value, with $\pm 1\sigma$ variations formed by applying weights of 1 ± 0.25 to gluon splitting events and 1 ∓ 0.22 to non-gluon splitting events in an anti-correlated manner. In addition, to test the dependence of the gluon splitting weights on the kinematic region, the fit is repeated both with a higher M_J threshold and with different N_jet bins. The resulting weights are consistent with those of the nominal fit.

Another significant systematic uncertainty is due to the uncertainty on the data-to-simulation b-tagging scale factors for efficiency and mistag rates. These scale factors are derived from data in various QCD and $\mathrm{t\bar{t}}$ control samples and are binned in jet p_T and jet flavor (light + g, c, and b). Uncertainties due to these scale factors are assessed by varying them by the uncertainties on their measurements.

Other experimental uncertainties are small and include lepton efficiency, jet energy resolution, jet energy scale, and luminosity. The uncertainty associated with lepton efficiency is determined by varying the efficiency to reconstruct a lepton within its uncertainty determined from data. Jet energy scale uncertainties are calculated by varying the p_T of $R = 0.4$ jets as a function of p_T and η. The uncertainty arising from jet energy resolution is determined by applying an $|\eta|$-dependent factor to the jet p_T to match the jet energy resolution observed in data. Lastly, the luminosity is varied according to its uncertainty of 2.5% [49], affecting only sub-leading backgrounds. No uncertainty is applied for the amount of pileup as its effect is negligible. The uncertainties due to the limited size of simulation samples are incorporated as nuisance parameters in the fit.

Systematic uncertainties due to theoretical uncertainties are applied and include variations of the renormalization scale, factorization scale, and correlated variations of the renormalization and factorization scales. Additionally, uncertainties on the parton distribution function (PDF)

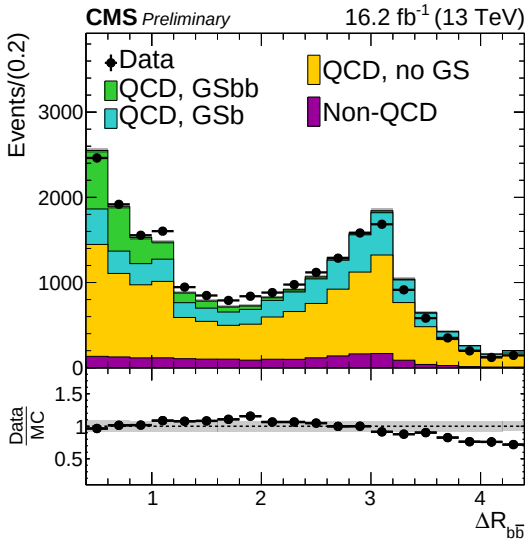

Figure 2: Post-fit $\Delta R_{b\bar{b}}$ distributions in a selection with $N_{lep} = 0$, $N_b = 2$, $N_{jet} \geq 4$ and $M_J > 500\,GeV$ and $H_T > 1500\,GeV$. The ratio of data to simulation is shown at the bottom, and the fit uncertainty is represented by the gray band.

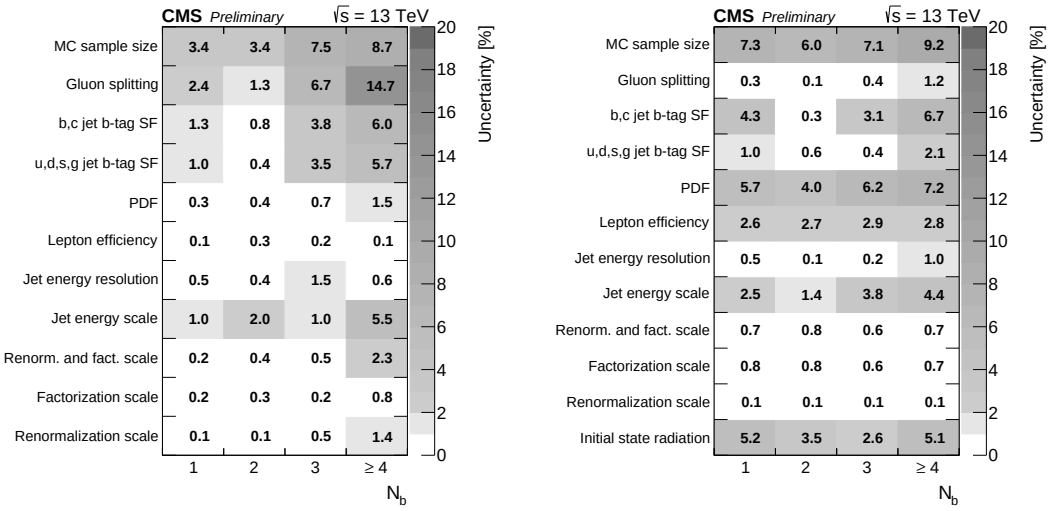

Figure 3: Background (left) and $m_{\tilde{g}} = 1600\,GeV$ signal (right) systematic uncertainties on the $N_{jet} \geq 8$ and $M_J \geq 1000\,GeV$ bin. SF in the label means scale factor.

are incorporated by considering variations in the NNPDF 3.0 scheme [50]. The size of these uncertainties is typically small as the effect of these variations is largely to vary the cross sections of processes, which is constrained by data for the main background processes.

The background systematics for the most sensitive search bin are shown in Fig. 3 (left).

4.2 Signal systematic uncertainties

Several of the systematic uncertainties affecting the signal yield are evaluated in the same way as the background yield. These are the uncertainties due to gluon splitting, lepton efficiency, jet energy scale, jet energy resolution, b-tag scale factors, simulation sample size, and theoretical

uncertainties. All systematic uncertainties affect both the N_b shape and normalization, except for the gluon splitting uncertainty which only affects the N_b shape.

The number of jets from initial state radiation (ISR) produced in signal simulation is reweighted based on comparisons between data and Madgraph-generated $t\bar{t}$ samples. The reweighting factors vary between 0.92 and 0.51 for the number of ISR jets between 1 and ≥ 6. We take one half of the deviation from unity as the systematic uncertainty on these reweighting factors.

The systematic uncertainties are shown in Fig. 3 for the most sensitive bin in a model with $m_{\tilde{g}} = 1600\,\text{GeV}$. The dominant signal systematic uncertainties arise from the limited simulation sample size, the b-tagging efficiency scale factor, and the ISR modeling. There is no systematic uncertainty taken for pileup reweighting, as the signal efficiency is found to be insensitive to the number of pileup interactions.

5 Results

The results of a background-only fit are shown in Figs. 4 and 5. These figures separately show the control and signal regions; however the fit includes all bins simultaneously. The N_b distributions in data are well described by the fit, and examination of the nuisance parameters shows that none of the nuisance parameters is significantly changed by the fit. The post-fit yields are presented in Table 1.

A signal + background fit is performed for gluino masses ranging from 1000 to 2000 GeV. In all cases, the post-fit N_b distribution describes the data well, and the fit extracts at most a small and insignificant signal contribution. For example with a 1600 GeV gluino, the signal strength is $r = 0.18^{+0.41}_{-0.18}$. The nuisance parameter pulls are small and consistent with those of the background-only fit.

Limits on the signal production cross section are calculated at 95% confidence level (CL) using the CL_s criterion [51–53] and shown in Fig. 6. Comparing the observed limit to the gluino pair production cross section [54], we exclude $m_{\tilde{g}} < 1610$ GeV in this benchmark $\tilde{g} \to tbs$ model. This observed mass limit is slightly lower than the expected limit, $m_{\tilde{g}} < 1640$ GeV, because of the small observed excess in the $M_J > 1000$ GeV bins.

6 Summary

We have searched for evidence of new phenomena with a single lepton and high jet and b-quark multiplicity without a missing transverse momentum requirement. The background is predicted using a combined fit in bins of the number of jets, number of b-tagged jets, and the sum of masses of large radius jets, using Monte Carlo simulated predictions with data driven corrections for the normalizations of the dominant backgrounds and nuisance parameters for theoretical and experimental uncertainties. Statistical uncertainties dominate in the signal regions, while the most important systematic uncertainties arise from modeling of gluon splitting and the b-quark tagging efficiency and mistagging rate. The data are consistent with a background-only fit. Cross section limits of approximately 10 fb are derived using a benchmark R-parity-violating supersymmetry model of gluino pair production with a prompt three-body decay to tbs as predicted in minimal-flavor-violating models. Gluino masses below 1610 GeV are excluded in this model.

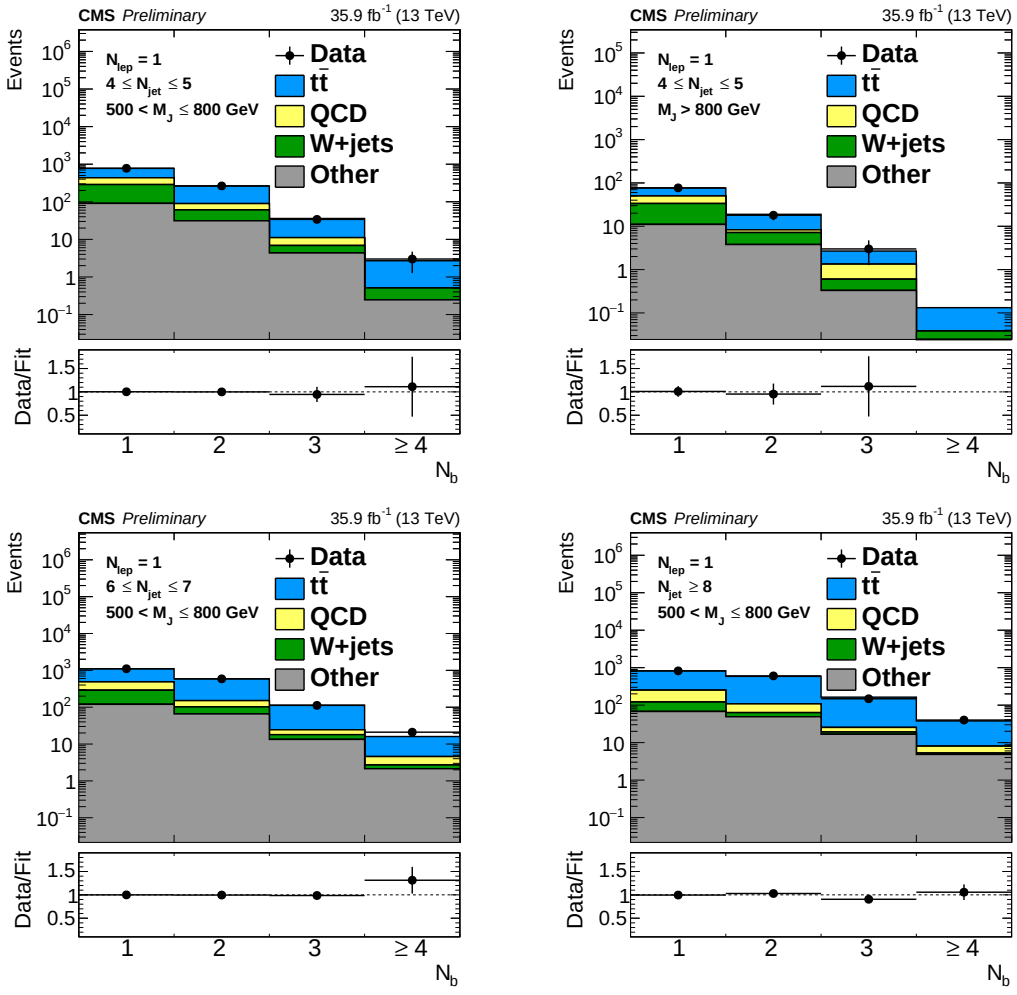

Figure 4: Data and the background-only post-fit N_{b} distribution for bins with low expected signal contribution: $500 < M_{\mathrm{J}} \leq 800$ GeV, $4 \leq N_{\mathrm{jet}} \leq 5$ (top-left), $M_{\mathrm{J}} > 800$ GeV, $4 \leq N_{\mathrm{jet}} \leq 5$ (top-right), $500 < M_{\mathrm{J}} \leq 800$ GeV, $6 \leq N_{\mathrm{jet}} \leq 7$ (bottom-left), and $500 < M_{\mathrm{J}} \leq 800$ GeV, $N_{\mathrm{jet}} \geq 8$ (bottom-right). The uncertainty on the ratio of data to total background (Data/Fit) is from the statistics of the data sample.

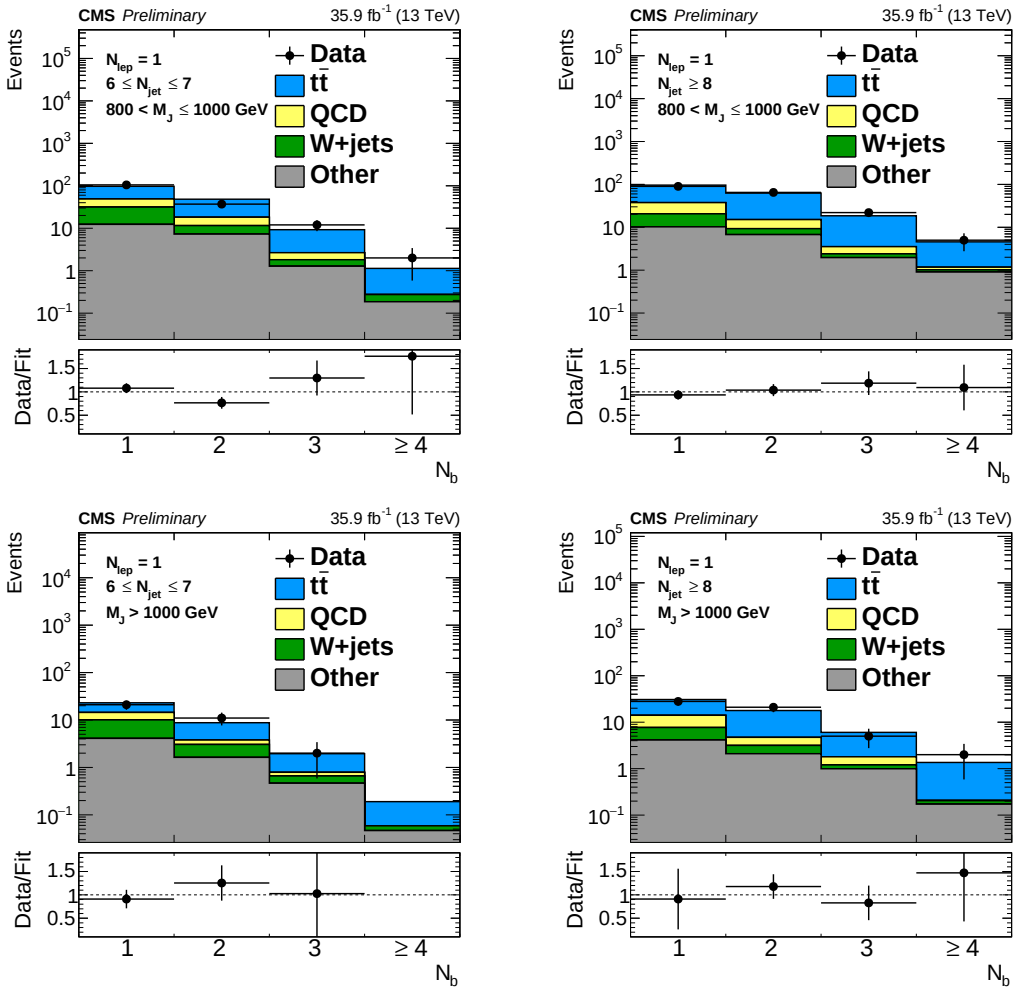

Figure 5: Data and the background-only post-fit N_b distribution for bins with large expected signal contribution: $800 < M_J \leq 1000$ GeV, $6 \leq N_{jet} \leq 7$ (top-left), $800 < M_J \leq 1000$ GeV, $N_{jet} \geq 8$ (top-right), $M_J > 1000$ GeV, $6 \leq N_{jet} \leq 7$ (bottom-left), and $M_J > 1000$ GeV, $N_{jet} \geq 8$ (bottom-right). The uncertainty on the ratio of data to total background (Data/Fit) is from the statistics of the data sample.

Table 1: Table of the post-fit yields for the background-only fit, observed data, and expected yields for $m_{\tilde{g}} = 1600$ GeV in each search bin.

N_b	QCD	$t\bar{t}$	W+jets	Other	All bkg.	Data	Expected $m_{\tilde{g}} = 1600$ GeV
$4 \leq N_{\text{jet}} \leq 5$, $500 < M_J < 800$ GeV							
1	148	340	196	91	775 ± 43	777	0.50 ± 0.13
2	29	175	30	31	264 ± 17	264	0.39 ± 0.11
3	4.3	24.8	2.5	4.4	36 ± 4	34	0.18 ± 0.08
≥ 4	0.0	2.2	0.3	0.2	2.7 ± 0.4	3	0.04 ± 0.04
$4 \leq N_{\text{jet}} \leq 5$, $M_J > 800$ GeV							
1	16.5	26.3	22.5	11.0	76 ± 6	77	0.32 ± 0.11
2	1.1	10.6	3.4	3.8	19 ± 2	18	0.40 ± 0.12
3	0.7	1.3	0.3	0.3	2.7 ± 0.5	3	0.13 ± 0.06
≥ 4	0.00	0.09	0.03	0.01	0.13 ± 0.03	0	0.03 ± 0.03
$6 \leq N_{\text{jet}} \leq 7$, $500 < M_J < 800$ GeV							
1	197	620	169	120	1106 ± 48	1105	2.5 ± 0.3
2	49	440	36	66	591 ± 21	588	3.1 ± 0.3
3	6.4	89.2	4.6	13.4	114 ± 8	112	1.4 ± 0.2
≥ 4	1.9	11.4	0.6	2.1	16 ± 2	21	0.25 ± 0.09
$N_{\text{jet}} \geq 8$, $500 < M_J < 800$ GeV							
1	130	574	53	68	825 ± 38	821	3.5 ± 0.3
2	45	478	14	49	586 ± 20	603	5.4 ± 0.4
3	6.3	138.1	2.5	16.7	164 ± 9	148	3.0 ± 0.3
≥ 4	2.8	29.8	0.4	4.8	38 ± 4	40	1.4 ± 0.2
$6 \leq N_{\text{jet}} \leq 7$, $800 < M_J < 1000$ GeV							
1	17.3	48.4	19.2	12.3	97 ± 8	105	1.2 ± 0.2
2	6.6	30.1	4.3	7.3	48 ± 4	37	2.0 ± 0.3
3	0.8	6.6	0.5	1.3	9.3 ± 1.0	12	1.0 ± 0.2
≥ 4	0.0	0.9	0.1	0.2	1.1 ± 0.2	2	0.31 ± 0.09
$N_{\text{jet}} \geq 8$, $800 < M_J < 1000$ GeV							
1	17.0	58.7	10.3	10.2	96 ± 8	90	4.2 ± 0.4
2	5.8	47.5	2.5	6.8	63 ± 5	65	5.3 ± 0.4
3	1.1	15.0	0.4	2.0	19 ± 2	22	2.6 ± 0.3
≥ 4	0.2	3.4	0.1	0.9	4.6 ± 0.6	5	1.3 ± 0.2
$6 \leq N_{\text{jet}} \leq 7$, $M_J > 1000$ GeV							
1	4.4	8.7	6.0	4.1	23 ± 2	21	2.0 ± 0.3
2	0.7	5.0	1.4	1.6	8.8 ± 1.2	11	2.3 ± 0.3
3	0.1	1.2	0.2	0.5	1.9 ± 0.3	2	1.0 ± 0.2
≥ 4	0.00	0.13	0.01	0.05	0.19 ± 0.04	0	0.23 ± 0.08
$N_{\text{jet}} \geq 8$, $M_J > 1000$ GeV							
1	6.4	16.7	3.5	4.1	31 ± 3	28	5.4 ± 0.4
2	1.6	13.1	1.1	2.1	18 ± 2	21	8.2 ± 0.5
3	0.6	4.2	0.2	1.0	6.0 ± 0.8	5	5.7 ± 0.4
≥ 4	0.0	1.2	0.0	0.2	1.4 ± 0.3	2	3.2 ± 0.3

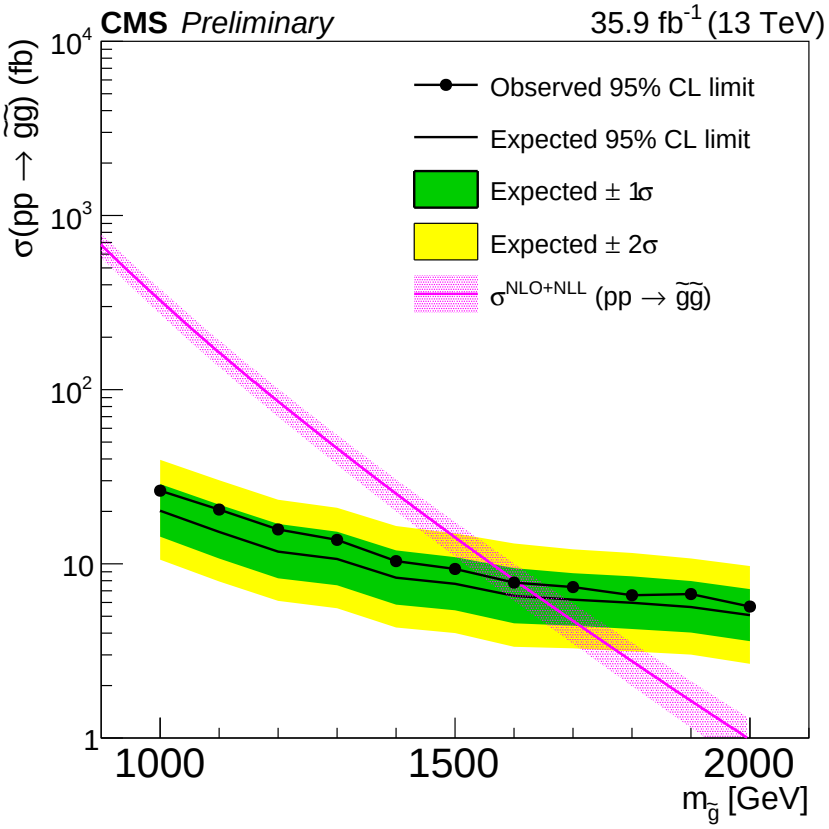

Figure 6: Cross section upper limits at 95% CL compared to the gluino pair production cross section (magenta). The theoretical uncertainties on the cross section are shown as a band around the line [54]. The expected limits (black solid line) and their $\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) variations are shown. The observed limit is the black solid line with dots.

References

[1] E. Witten, "Dynamical breaking of supersymmetry", *Nucl. Phys.* **B 188** (1981) 513.

[2] S. Dimopoulos and H. Georgi, "Softly broken supersymmetry and su(5)", *Nucl. Phys.* **B 193** (1981) 150.

[3] P. Ramond, "Dual theory for free fermions", *Phys. Rev. D* **3** (1971) 2415, doi:10.1103/PhysRevD.3.2415.

[4] Y. A. Golfand and E. P. Likhtman, "Extension of the algebra of Poincaré group generators and violation of P invariance", *JETP Lett.* **13** (1971) 323.

[5] A. Neveu and J. H. Schwarz, "Factorizable dual model of pions", *Nucl. Phys. B* **31** (1971) 86, doi:10.1016/0550-3213(71)90448-2.

[6] D. V. Volkov and V. P. Akulov, "Possible universal neutrino interaction", *JETP Lett.* **16** (1972) 438.

[7] J. Wess and B. Zumino, "A Lagrangian model invariant under supergauge transformations", *Phys. Lett. B* **49** (1974) 52, doi:10.1016/0370-2693(74)90578-4.

[8] J. Wess and B. Zumino, "Supergauge transformations in four dimensions", *Nucl. Phys. B* **70** (1974) 39, doi:10.1016/0550-3213(74)90355-1.

[9] P. Fayet, "Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino", *Nucl. Phys. B* **90** (1975) 104, doi:10.1016/0550-3213(75)90636-7.

[10] H. P. Nilles, "Supersymmetry, supergravity and particle physics", *Phys. Rep.* **110** (1984) 1, doi:10.1016/0370-1573(84)90008-5.

[11] ATLAS Collaboration, "Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector", *Phys. Rev. D* **94** (2016), no. 5, 052009, doi:10.1103/PhysRevD.94.052009, arXiv:1606.03903.

[12] CMS Collaboration, "Search for top squark pair production in pp collisions at $\sqrt{s} = 13$ TeV using single lepton events", arXiv:1706.04402.

[13] ATLAS Collaboration, "Search for gluinos in events with an isolated lepton, jets and missing transverse momentum at $\sqrt{s}$ = 13 Te V with the ATLAS detector", *Eur. Phys. J. C* **76** (2016) 565, doi:10.1140/epjc/s10052-016-4397-x, arXiv:1605.04285.

[14] M. Aaboud et al., "Search for squarks and gluinos in final states with jets and missing transverse momentum at $\sqrt{s} = 13$ TeV with the ATLAS detector", *The European Physical Journal C* **76** (2016) 392, doi:10.1140/epjc/s10052-016-4184-8.

[15] ATLAS Collaboration, "Search for new phenomena in final states with large jet multiplicities and missing transverse momentum with ATLAS using $\sqrt{s} = 13$ TeV proton-proton collisions", *Phys. Lett. B* **757** (2016) 334–355, doi:10.1016/j.physletb.2016.04.005, arXiv:1602.06194.

[16] CMS Collaboration, "Search for new phenomena with the MT2 variable in the all-hadronic final state produced in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:1705.04650.

[17] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s} = 13$ TeV in the single-lepton final state using the sum of masses of large-radius jets", arXiv:1705.04673.

[18] R. Barbier et al., "R-parity violating supersymmetry", *Phys. Rept.* **420** (2005) 1, doi:10.1016/j.physrep.2005.08.006, arXiv:hep-ph/0406039.

[19] C. Csaki, Y. Grossman, and H. B., "MFV SUSY: A Natural Theory for R-Parity violation", *Phys. Rev. D* **85** (2012) 095009, doi:11.1103/PhysRevD.85.095009, arXiv:1111.1239.

[20] ATLAS Collaboration, "A search for top squarks with R-parity-violating decays to all-hadronic final states with the ATLAS detector in $\sqrt{s}$ = 8 TeV proton-proton collisions", *JHEP* **06** (2016) 067, doi:10.1007/JHEP06(2016)067, arXiv:1601.07453.

[21] ATLAS Collaboration Collaboration, "A search for pair produced resonances in four jets final states in proton-proton collisions at $\sqrt{s}$=13 TeV with the ATLAS experiment", Technical Report ATLAS-CONF-2016-084, CERN, Geneva, Aug, 2016.

[22] CMS Collaboration, "Searches for *R*-parity-violating supersymmetry in ppcollisions at $\sqrt{s} = 8$ TeV in final states with 0-4 leptons", *Phys. Rev. D* **94** (2016), no. 11, 112009, doi:10.1103/PhysRevD.94.112009, arXiv:1606.08076.

[23] CMS Collaboration, "Search for R-parity violating supersymmetry with displaced vertices in proton-proton collisions at $\sqrt{s}$ = 8 TeV", *Phys. Rev. D* **95** (2017), no. 1, 012009, doi:10.1103/PhysRevD.95.012009, arXiv:1610.05133.

[24] ATLAS Collaboration, "Search for new phenomena in a lepton plus high jet multiplicity final state with the ATLAS experiment using $\sqrt{s} = 13$ TeV proton-proton collision data", arXiv:1704.08493.

[25] A. Hook, E. Izaguirre, M. Lisanti, and J. G. Wacker, "High Multiplicity Searches at the LHC Using Jet Masses", *Phys. Rev. D* **85** (2012) 055029, doi:10.1103/PhysRevD.85.055029, arXiv:1202.0558.

[26] T. Cohen, E. Izaguirre, M. Lisanti, and H. K. Lou, "Jet Substructure by Accident", *JHEP* **03** (2013) 161, doi:10.1007/JHEP03(2013)161, arXiv:1212.1456.

[27] S. El Hedri, A. Hook, M. Jankowiak, and J. G. Wacker, "Learning How to Count: A High Multiplicity Search for the LHC", *JHEP* **08** (2013) 136, doi:10.1007/JHEP08(2013)136, arXiv:1302.1870.

[28] ATLAS Collaboration, "Search for massive supersymmetric particles decaying to many jets using the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev. D* **91** (2015) 112016, doi:10.1103/PhysRevD.91.112016, arXiv:1502.05686.

[29] ATLAS Collaboration, "Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS experiment", *JHEP* **10** (2013) 130, doi:10.1007/JHEP10(2013)130, 10.1007/JHEP01(2014)109, arXiv:1308.1841. [Erratum: *JHEP* **01** (2014) 109].

[30] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s} = 13$ TeV in the single-lepton final state using the sum of masses of large-radius jets", arXiv:1605.04608.

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

[32] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[33] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473–500, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[34] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[35] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[36] 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.

[37] GEANT4 Collaboration, "GEANT4 — a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[38] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", (2017). arXiv:1706.04965. Submitted to *JINST*.

[39] 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.

[40] 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.

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

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

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

[44] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", CMS Physics Analysis Summary CMS-PAS-BTV-15-001, CERN, 2016.

[45] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[46] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[47] K. Rehermann and B. Tweedie, "Efficient Identification of Boosted Semileptonic Top Quarks at the LHC", *JHEP* **03** (2011) 059, doi:10.1007/JHEP03(2011)059, arXiv:1007.2221.

[48] ATLAS Collaboration, "Search for direct pair production of the top squark in all-hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *JHEP* **09** (2014) 015, doi:10.1007/JHEP09(2014)015, arXiv:1406.1122.

[49] CMS Collaboration, "Cms luminosity measurements for the 2016 data taking period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, 2017.

[50] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[51] 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.

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

[53] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: doi:10.1140/epjc/s10052-013-2501-z].

[54] D. Alves et al., "Supersymmetry production cross sections in pp collisions at $\sqrt{s} = 7$ TeV", arXiv:1206.2892.