Search for Supersymmetry, Leptoquarks and Study of Advanced Pileup Mitigation Techniques

Bibhuprasad Mahakud

Department of High Energy Physics
Tata Institute of Fundamental Research,
Homi Bhabha Road, Mumbai 400005, India

Abstract

We search for supersymmetry in all hadronic final states with proton-proton collision data recorded by the CMS detector at a center-of-mass energy of 13 TeV. We specifically target for pair produced gluinos in the 2015 data that correspond to an integrated luminosity of 2.3 fb^{-1}, and using more than five times data collected during 2016 where we also add stop and other squark pairs to our search. The data are examined in search regions of jet multiplicity, tagged bottom quark jet multiplicity, missing transverse momentum, and the scalar sum of jet transverse momenta. The observed numbers of events in all search regions are found to be consistent with the expectations from standard model processes. Exclusion limits are calculated for simplified supersymmetric models of gluino pair production. Depending on the assumed gluino decay mechanism, and for a massless, weakly interacting, lightest neutralino, lower limits on the gluino mass from 1440 to 1700 GeV are obtained, significantly extending previous limits.

Another search for the pair production of first generation scalar leptoquarks decaying to leptons and jets is presented. The data corresponding to 2.6 fb^{-1} were collected from proton-proton collisions at a center-of-mass energy of 13 TeV. No significant deviation is found with respect to the standard model predictions. Upper limits are set on cross section times branching fraction.

The increase of instantaneous luminosity for Run II LHC results in a large number of additional proton-proton collisions in a given event (pileup) leading to contamination of jets. We study advanced pileup removal techniques like trimming, prunning and soft-drop etc. The focus was on preparation for Run II for which we expected up to 40 additional pileup events in comparison to Run I LHC data that had typically half the pileup events on average.

Advisor: Prof. Gagan B. Mohanty

List of Publications

Journals

- **Search for supersymmetry in the multijet and missing transverse momentum final state in pp collisions at 13 TeV, PLB 758 (2016) 152.**

CMS Physics Analysis Summary (Public Results)

- **Search for supersymmetry in events with jets and missing transverse momentum in proton-proton collisions at 13 TeV, CMS-PAS-SUS-16-014.**
- **Search for supersymmetry in the multijet and missing transverse momentum channel in pp collisions at 13 TeV, CMS-PAS-SUS-15-002.**
- **Search for pair-production of first generation scalar leptoquarks in pp collisions at $\sqrt{s}$=13 TeV with 2.6 fb^{-1}, CMS-PAS-EXO-16-043.**
- **Study of Pileup Removal Algorithms for Jets, CMS-PAS-JME-14-001.**

Conference Proceedings

- **Search for SUSY in jets+MET final state, PoS(DIS2016)094.**

CMS Internal Notes

- **Search for supersymmetry in multijet final states in proton-proton collisions at 13 TeV, CMS-AN-16-188.**
- **Search for supersymmetry in the multijet and missing transverse momentum channel in pp collisions at 13 TeV, CMS-AN-15-003.**
- **Search for Pair-production of Scalar First Generation Leptoquarks in pp Collisions at $\sqrt{s}$ = 13 TeV, CMS-AN-15-294.**
- **Study of Pileup Removal Algorithms for Jets, CMS-AN-14-175.**

Contents

1 Introduction

The standard model (SM) of particle physics successfully describes a wide range of phenomena. However, it is unable to explain the stability of the Higgs boson mass in the face of higher-order corrections, suggesting that the model is incomplete. Many extensions to the SM have been proposed to provide a more fundamental theory. Supersymmetry (SUSY) [1], one such extension, postulates that each SM particle has a SUSY partner from which it differs in spin by half a unit. As examples, squarks and gluinos are the SUSY partners of quarks and gluons, respectively, while neutralinos arise from a mixture of the SUSY partners of neutral (charged) Higgs and electroweak gauge bosons. Radiative corrections involving SUSY particles can compensate the contributions from SM particles and thereby stabilize the Higgs boson mass. For this cancellation to be natural, the top and bottom squark, and gluino must have masses on the order of a few TeVs, possibly allowing them to be produced at the CERN LHC.

Amongst SUSY processes, the gluino pair production, typically yielding four or more hadronic jets in the final state, has the largest possible cross section, making it an apt channel at the recently started LHC Run 2. In R-parity [2] conserving SUSY models, as are considered here, the lightest SUSY particle (LSP) is stable and assumed to be weakly interacting, leading to potentially large undetected, or missing, transverse momentum. Supersymmetry events at the LHC might thus be characterized by significant missing transverse momentum, numerous jets, and in the context of natural SUSY [3], jets initiated by top and bottom quarks.

In this search, we consider SUSY scenarios in the context of simplified models [4] of new particle production. Diagrams for the three models are shown in Fig. 1. Simplified models contain the minimal particle content to represent a topological configuration. As for SUSY production scenarios, the four simplified models can be interpreted as follows. In the first one, shown in Fig. 1 (left), the gluino pair production is followed by the decay of each gluino to a bottom quark and an off-shell bottom squark. The off-shell bottom squark decays to a bottom quark and the LSP, where the LSP is assumed to be the lightest neutralino that escapes detection, leading to significant missing transverse energy. The second scenario, shown in Fig. 1 (middle), is the same as the first one except with top quarks and off-shell top squarks appearing in place of the bottom quarks and squarks. The third scenario, shown in Fig. 1 (right), is the corresponding situation with gluino decay to a light-flavored quark and off-shell squark up, down, strange, and charm with equal probability, for each gluino separately.

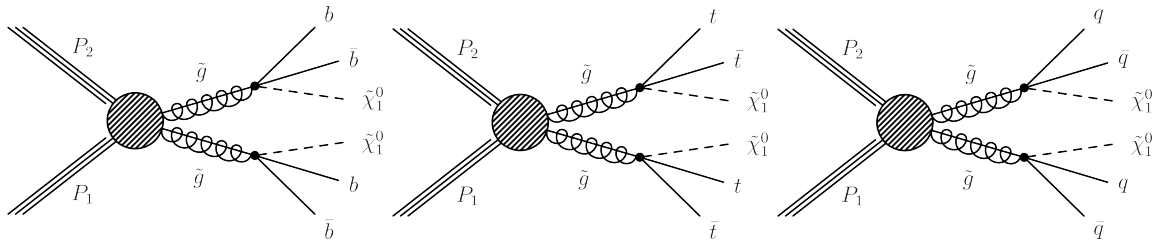

Figure 1: Event diagrams for the main new-physics scenarios considered in this study.

The SM hints that its lepton and quark sectors could be related by a fundamental symmetry. Many beyond the SM theories , include such a symmetry [5, 6, 7, 8], which gives rise to new class of bosons called leptoquarks (LQ). Leptoquarks couple to both leptons and quarks, carrying lepton and baryon numbers as well as fractional electric charge. They can be either scalar or vector particles (carrying zero or one unit of spin) and are color triplets. Current experimental searches for rare processes such as lepton number violation and flavor-changing neutral currents suggest that leptoquarks come in three generations which do not mix. Their pair-production at the LHC proceeds mainly via gluon-gluon (gg) fusion and quark antiquark ($q\bar{q}$) annihilation. Due to the gg dominance, along with the fact that only one q-Feynman diagram contains the LQ-quark-lepton vertex, the scalar LQ pair-production has a negligible dependence on the Yukawa LQ-quark-lepton coupling, usually denoted as λ. LQ searches are essentially independent of λ. In our search we consider pair produced leptoquarks with each of them decaying to an electron and a quark. Therfore we look for signatue in two electron and two jet final states in the data.

In the two physics topics discussed above as well as many other important studies at the LHC, the performance of jets is extremely important. One of the key challenges of the LHC run is the increase of instantaneous luminosity, which results in a large number of additional proton-proton collisions (pileup) in each event. In such high pileup environment, an accurate reconstruction of jet properties and shapes is become more and more demanding. The most common observables for jets at the LHC are primarily the jet p_T, η, and ϕ. In recent years, it is becoming increasingly popular to consider the internal structure of the jet. The applications include the discrimination of quark- and gluon-initiated jets as well as the identification of highly boosted hadronically decaying heavy resonances such as W/Z or Higgs boson that are contained in a single jet. In all such cases, contamination from pileup degrades our ability to effectively reconstruct the jet observables. Motivated by this, we study advanced techniques for pileup mitigation for jets in view of high pileup scenarios in Run II of the LHC.

2 CMS Detector

The CMS detector is built around a superconducting solenoid of 6-m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and microstrip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass-scintillator sandwich hadron calorimeter (HCAL). The tracking detectors cover a range of $|\eta| < 2.5$. The ECAL and HCAL, each composed of a barrel and two endcap sections, extend over $|\eta| < 3.0$. Forward calorimeters on either side of the interaction point encompass $3.0 < |\eta| < 5.0$. Muons are measured within $|\eta| < 2.4$ by gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. The detector is nearly hermetic, permitting accurate measurements of missing transverse energy. A more detailed description of the detector, together with a definition of the coordinate system and relevant kinematic variables, is given in Ref. [9]. The studies presented in the thesis use information from all parts of the CMS detector.

3 Event Reconstruction

Physics objects used in our studies are defined using the so-called particle flow (PF) algorithm [10], which reconstructs and identifies individual particles through an optimized combination of information from different detector components. The PF candidates are classified as photons, charged and neutral hadrons, electrons, or muons. Additional quality criteria are imposed on electron muon and photon candidates. For example, more restrictive conditions are placed on the shower shape and on the ratio of energies deposited in the HCAL and ECAL for electron and photon candidates, and similarly on the matching of track segments between the silicon tracker and muon detector for muon candidates. Photons being neutral particles do not produce tracks in the tracker. The event primary vertex is taken to be the one reconstructed with the largest sum of charged-track p_T^2 values and is required to lie within 24 cm (2 cm) of the center of the detector in the direction along (perpendicular to) the beam axis. Tracks from extraneous pp interactions within the same or a nearby bunch crossing (pileup) are removed . The PF objects serve as inputs for jet reconstruction, based on the anti-kt algorithm [11] with a distance parameter of 0.4. Jet quality criteria are applied to eliminate, for example, spurious events caused by calorimetric noise. Contributions to an individual jet's p_T from pileup interactions are subtracted, and corrections are applied as a function of jet p_T and η to account for residual effects of nonuniform detector response.

4 Search for Supersymmetry with 13 TeV pp Collision Data

Because of the large mass scale and their all-hadronic nature, the targeted SUSY events are expected to exhibit large values of H_T, where H_T is the scalar sum of the jet p_T. As a measure of missing transverse momentum, we use the variable H_T^{miss}, which is the magnitude of the vector sum of the jet p_T. We present a general search for gluino pair production leading to final states with large H_T, large H_T^{miss} as well as large jet multiplicity. The data are examined in bins of $\mathbf{N_{jet}}$, $\mathbf{N_{b-jet}}$, $\mathbf{H_T}$, and $\mathbf{H_T^{miss}}$, where N_{jet} is the number of jets and N_{b-jet} the number of tagged bottom quark jets (b jets). The search is performed in exclusive bins of these four observables.

The principal sources of background arise from the SM production of top quarks, a W or Z boson in association with jets (W+jets or Z+jets), and multiple jets through the strong interaction. We refer to the latter class of background as quantum chromodynamics (QCD) multijet events. Although events with top quarks mostly come from top quark-antiquark ($t\bar{t}$) production, a modest contribution is also from single top quark processes. The W and Z bosons in W+jets and Z+jets events can be either on- or off-shell. For top quark and W+jets events, significant H_T^{miss} can arise if the W boson decays leptonically, producing a neutrino and an undetected charged lepton, while Z+jets events can exhibit significant H_T^{miss} if the Z boson decays to two neutrinos. For QCD multijet events, significant H_T^{miss} can arise if the event contains a charm or bottom quark that undergoes a semileptonic decay; however the principal source is the mismeasurement of jet p_T. The signal vs. background composition plots in search variables are shown in Fig. 2 and 3.

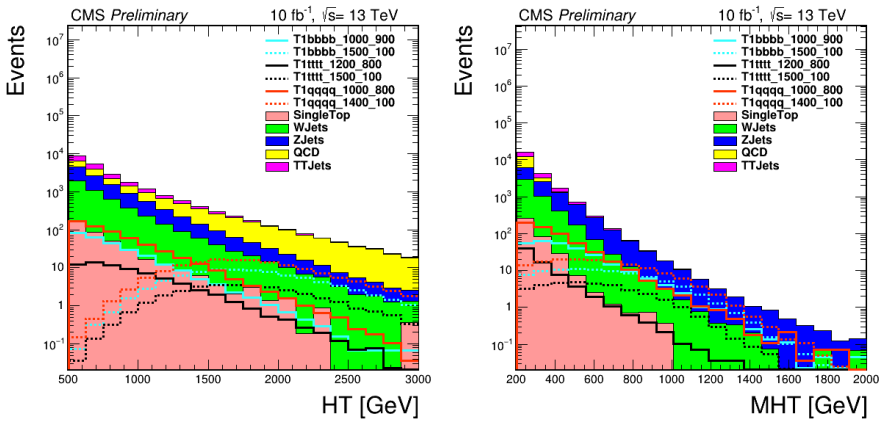

Figure 2: Signal vs. stacked backgrounds in H_T(HT, left) and H_T^{miss}(MHT, right)

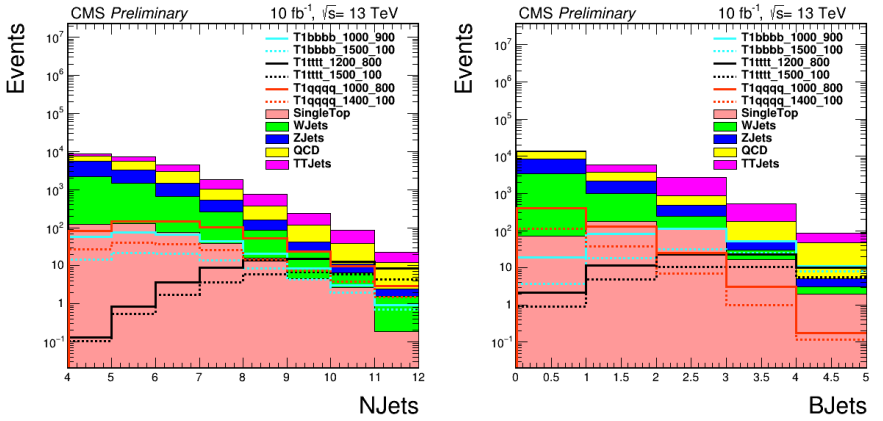

Figure 3: Signal vs. stacked backgrounds in N_{jet}(NJets, left) and N_{b-jet}(MHT, right)

4.1 Study on 2.3 fb^{-1} of 2015 Data

In this section, we discuss the analysis strategy and results of the search conducted using 2015 data. With as little as 2.3 fb^{-1} data we search for gluino pair production in terms of the above four variables. To get maximum signal and minimum background we select events with following criteria defined below

4.1.1 Event Selection and Search Region

The following requirements define the signal event candidates:

- $N_{jet} \geq 4$, where the jets must lie within $|\eta| < 2.4$; we require at least four jets because of our focus on gluino pair production;
- $H_T > 500$ GeV, where H_T is the scalar p_T sum of jets with $|\eta| < 2.4$;
- $H_T^{miss} > 200$ GeV, where H_T^{miss} is the magnitude of $\overrightarrow{H}_T^{miss}$, the negative of the vector

p_T sum of jets with $|\eta| < 5$; the η range is extended in this case so that $\overrightarrow{H}_{\mathrm{T}}^{\mathrm{miss}}$ better represents the total missing transverse momentum in a given event;

- No identified, isolated electron (muon) candidate with $p_T > 10$ GeV and $|\eta| < 2.5$ (<2.4);
- No isolated charged-particle track with $|\eta| < 2.4$, $\mathrm{m_T} < 100$ GeV, and $\mathrm{p_T} > 10$ GeV ($p_T > 5$ GeV if the track is identified as an electron or muon candidate, where $\mathrm{m_T}$ is the transverse mass formed from the $\overrightarrow{p_T}^{\mathrm{miss}}$ and isolated-track $\mathrm{p_T}$ vector, with $\overrightarrow{H}_{\mathrm{T}}^{\mathrm{miss}}$ the negative of the vector $\mathrm{p_T}$ sum of all PF objects;
- $\Delta\phi_{\overrightarrow{H}_{\mathrm{T}}^{\mathrm{miss}},\mathrm{j_i}} > 0.5$ (>0.3) for the two highest $\mathrm{p_T}$ jets $\mathrm{j_1}$ and $\mathrm{j_2}$ (the next two highest $\mathrm{p_T}$ jets $\mathrm{j_3}$ and $\mathrm{j_4}$), with the azimuthal angle between $\mathrm{H_T^{miss}}$ and the $\mathrm{p_T}$ vector of jet j_i.

The search is performed in the following 72 ($= 3 \times 4 \times 3 \times 3$) exclusive intervals of the four search variables:

- **3** $\mathrm{N_{jet}}$ bins: 4–6, 7–8 and ≥9;
- **4** $\mathrm{N_{b-jet}}$ bins: 0, 1, 2 and ≥3;
- **3** $\mathrm{H_T}$ bins: 500–800, 800–1200 and >1200 GeV;
- **3** $\mathrm{H_T^{miss}}$ bins : 200–500, 500–750 and >750 GeV.

4.1.2 Background Estimation

In this section, we describe the evaluation of the background from SM processes. The evaluation relies on data control regions (CRs) selected using similar criteria to the search regions. The backgrounds are divided into four different types, namely Z to neutrinos, lost lepton, QCD and hadronic tau.

Z to neutrinos: This is the most important backgrounds being an ireducible one. We are talking about events with a Z boson produced in association with jets when the Z decays to two neutrinos. The most straightforward way to measure this background is to exploit the decays $\mathrm{Z}(\ell^+\ell^-)$+jets in which the Z boson can be reconstructed from the observed pair of muons or electrons. The efficiency-corrected yields from these decays can be translated directly into the $\mathrm{Z}(\nu\bar{\nu})$+jets background yield by the known branching ratios. The limitation of this approach arises from the rather small branching ratio between the charged and neutral leptons, so that the transfer factor from the control sample measurement to the predicted background is larger than one (ignoring efficiencies, the branching ratio itself is approximately 3 when both muon and electron pairs are used).

The alternative approach is to exploit the similarity to Z boson radiation of the more copious radiation of photons. Here the challenge is to obtain validation in data of the MC predictions connecting the two processes.

Our baseline strategy is to use the γ+jets sample to determine the yields in the 18 bins corresponding to $\mathrm{N_{b-jet}}$=0. These are compared with the $\mathrm{Z}(\ell^+\ell^-)$+jets yields in the low-$\mathrm{N_{jet}}$bin to establish the systematic uncertainty of the physics modeling of γ+jets , and the

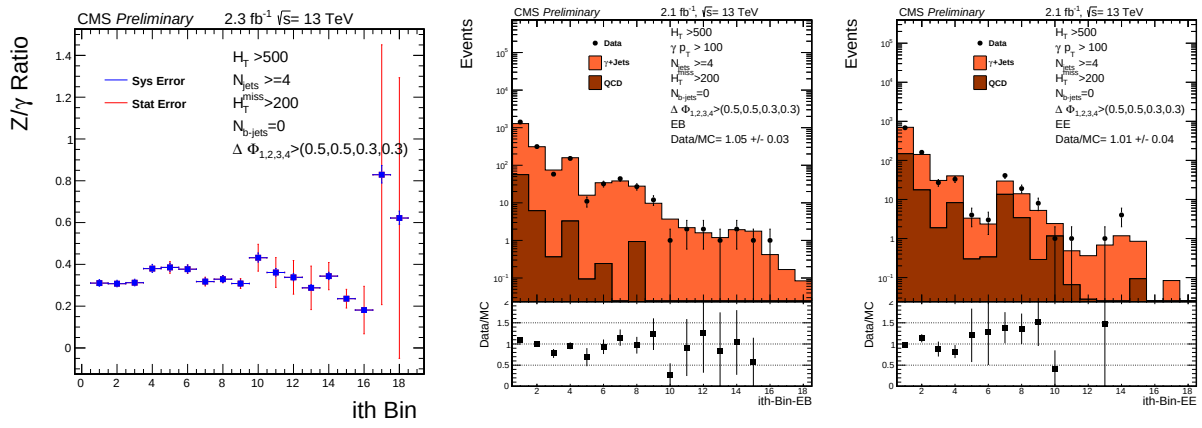

Figure 4: Physics ratio $R_{Z(\nu\bar{\nu})/\gamma}$ (left), numbers of observed events in the photon control samples in barrel(EB, middle) and endcap(EE, right) compared to simulation.

normalization corrected if necessary. The extrapolation to bins with $N_{b-jet}>0$ is performed to the extent possible with the $Z(\ell^+\ell^-)$+jets data sample, supplemented with MC information where necessary. We use simulation to correct for the resulting distribution to the higher N_{jet} bins.

As shown in Eq.(1), we predict the number of $Z(\rightarrow \nu\bar{\nu})$ + jets events contributing to each of the 18 0-btag analysis bins ($N^{pred}_{Z(\rightarrow\nu\bar{\nu})}$) from the number of events in the corresponding bin of the γ+jets control sample (N^{obs}_{γ}), the purity of the control sample (β), and the ratio of the numbers of $Z \rightarrow \nu\bar{\nu}$ + jets events and γ+jets events obtained from leading order MadGraph+Pythia ($\mathcal{R}_{Z(\nu\bar{\nu})/\gamma} \cdot \mathcal{F}_{\rm dir}$). Here $\mathcal{F}_{\rm dir}$ is the fraction of prompt photons that are direct.

$$N^{\rm pred}_{Z\rightarrow\nu\bar{\nu}} = DR \cdot \mathcal{R}_{Z(\nu\bar{\nu})/\gamma} \cdot \mathcal{F}_{\rm dir} \cdot \beta \cdot N^{\rm obs}_{\gamma} \quad (1)$$

DR, called the double ratio, is a correction factor to the physics ratio $R_{Z(\nu\bar{\nu})/\gamma}$. This is calculated by measuring the $R_{Z(\ell^+\ell^-)/\gamma}$ both in data and simulation and then taking the ratio.

Prompt photons can be distinguished from non-prompt photons by the shapes of their showers in the ECAL, as described by the well-known quantity $\sigma_{i\eta i\eta}$. The purity is determined with a two-component fit to the $\sigma_{i\eta i\eta}$ distribution in the photon control sample. The PDF for the prompt component is fitted directly in data using a gaussian distribution.

Fig 4 shows the data vs MC simulation for 18 kinematic bins (0-btag) where a fully photon method is employed.

Lost lepton: SM events (mostly $t\bar{t}$ and W+jets) with muons or electrons can satisfy the event selection and enter the signal sample as lost-lepton background if the requirements for any of the following analysis steps are not satisfied

- Kinematic acceptance,
- Reconstruction, or

- Isolation.

The basic idea behind our data-driven method to evaluate the lost-lepton background is to select single-lepton control samples (CS) in the data, through inversion of the lepton vetoes, and to weight each CS event by a factor that represents the probability for a lost-lepton event to appear with the corresponding search-variable values: H_T, H_T^{miss}, N_{jet}, and N_{b-jet}. The weights are determined through evaluation of the efficiencies for each analysis step. The weighted distributions of the search variables, summed over the events in the CS, define the predicted lost-lepton background in the respective search regions.

QCD: The H_T^{miss}in QCD multijet events is almost always due to a mismeasured jet in the event, thus the H_T^{miss}direction is usually close to the jet. The $\Delta\phi$variable is the minimum ϕ difference between $H_T^{miss}$and one of the four highest p_T jets.

The low $\Delta\phi$region is significantly enriched in QCD events. The sample of events with the $\Delta\phi$requirement inverted (i.e., $\Delta\phi_1 < 0.5$ or $\Delta\phi_2 < 0.5$ or $\Delta\phi_3 < 0.3$ or $\Delta\phi_4 < 0.3$) serves as the QCD control sample. The backgrond at high $\Delta\phi$, is estimated from the QCD yield at low $\Delta\phi$and a high/low ratio R^{QCD} for the QCD component. The $\Delta\phi$distribution shows that the high/low ratio has some dependence on the search variables H_T, H_T^{miss}, and N_{jet}. We model this dependence by assuming that it factorizes. That is, we assume the $H_T$dependence does not depend on H_T^{miss}or $N_{jet}$and similarly for $H_T^{miss}$and N_{jet}.

Hadronic tau: To evaluate the $t\bar{t}$, single-top and W+jets backgrounds that arises when a W boson decays to a neutrino and a hadronically decaying τ lepton (τ^h), we employ a tau-template method. In this approach, the τ^h background is estimated from a control sample (CS) of μ+jets events, which we select by requiring exactly one muon with $p_T > 20$ GeV and $|\eta| < 2.1$. This single-muon CS is mainly composed of $t\bar{t}(\rightarrow \mu\nu)$ and $W(\rightarrow \mu\nu)$+jets events. Since both μ+jets and τ^h+jets events arise from the same underlying process, the hadronic components of the two event classes are expected to be the same, aside from the response of the detector to a muon or a τ^h jet. The basic idea behind the method is to smear the muon p_T in the CS events, using MC-derived response functions (the "templates"), in order to emulate the τ^h jet response. Global hadronic variables such as N_{jet}, H_T, and $H_T^{miss}$are then recomputed, and the full analysis procedure is subsequently applied.

4.1.3 Uncertainties

Various kinds of systematic and statistical uncertainties that are considered in the analysis. The uncertainties that resultsfrom the background estimations include data control region statistics, Purity of control region , CR trigger efficiency etc. are discussed in the respective sections Ref. [13]. The signal systematics are discussed briefly below.

- **Luminosity:** A flat 4.6 % uncertainty on luminosity is propagated to the signal yield.
- **b-tag efficiency:** The b-tagging and mistagging scale factors are functions of the jet p_T and η. The scale factors are varied by their uncertainties and these variations are propagated as migrations between the different signal bins.

- **MC statistics:** The MC statistical uncertainties are propagated to the signal yield
- **Trigger efficiency:**The trigger efficiencies are measured in the data. The effect of the statistical and systematic uncertainties is at most 1.1 % at low H_T^{miss}.
- **Pileup reweighting:** The uncertainties in the pileup reweighting correction are derived from the uncertainties in the minimum bias cross section and the difference between the actual number of interactions and the observed number of interactions in the data. The minimum bias cross section in the 13 TeV is currently estimated to be 69 mb with an uncertainty of ±5%. The correction is varied according to these uncertainties, with a maximum effect of 0.5 %.
- **Scale:** The uncertainty is calculated from using the envelope of weights from varying the renormalization and factorization scales. The effect on the yield of non-compressed samples is less than 0.1 % and on compressed samples ranges from 1 % to 3 %.
- **ISR:** The effect on the yield of non-compressed samples is less than 0.1 % and on compressed samples ranges from 3 % to 11 %.
- **Jet Energy Corrections:** The jet energy corrections (JECs) are varied using the p_T and η dependent jet energy scale uncertainties from the official database, with a separate set of corrections for the fast simulation samples. The overall effect ranges from 0.5 % to 4%.
- **PDF:** The LHC4PDF prescription for the uncertainty on the total cross section is included as ±1 sigma bands in the results plots.

4.1.4 Results

The data in the signal regions are found to be in generally good agreement with the predicted backgrounds. Therefor we do not see any evidence for new physics. For the 72 search bins, the observed data and the pre-fit predictions for each background component are shown in Fig. 5. The 95% confidence-level(CL) upper limit is calculated on the production cross section taking all 72 bins. The upper limits on the signal cross section and the exclusion curves are shown in Fig. 6. For calculating the upper limits we use a test statistic $q_\mu = -2\ln(L_\mu/L_{max})$, where L_{max} is the maximum likelihood determined by allowing all parameters including the SUSY signal strength μ to vary and L_μ is the maximum likelihood for a fixed signal strength . The details of the statistical procedure can be found in Ref. [12]. For an explanation of the treatment of uncertainties, we refer to Ref. [13]. As can be seen from the plots in Fig. 6 the observed exclusion limits for low LSP masses lie around 1600 GeV both for four top and four b-quark final state, for four light quark final state, it is around 1450 GeV of gluino mass. The small disagreement of observed exclusion curves with the expected ones can be ascribed to the small insignificant excesses of events we see in various bins.

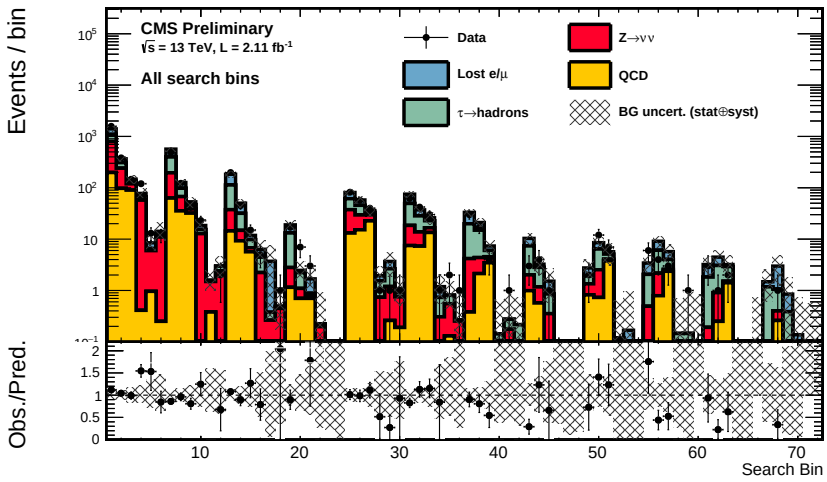

Figure 5: Data vs. the SM background before fit

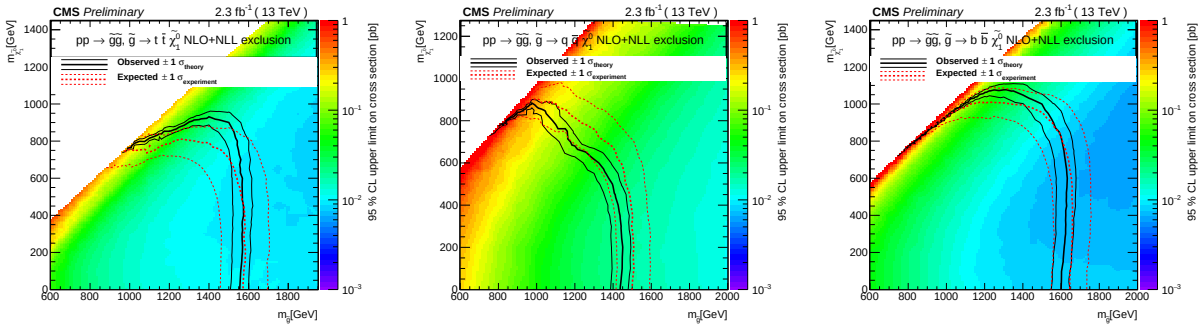

Figure 6: The 95% CL upper limits on the production cross sections for four top quark(left),four light quark(middle) and four b-quark(right) in the final state.

4.2 Study on 12.9 fb^{-1} of 2016 Data

In 2016 we lowered the $H_T$and N_{jet}thresholds to enhance the search sensitivity to some stop and squark production models. Also the number of bins is changed from 72 to 160 for an increased sensitivity. The important event selection and search region definitions that are different from the 2015 analysis are given below.

4.2.1 Search Region and Event Selection

The following requirements define the selection criteria for signal event candidates:

- $N_{jet} \geq 3$, where the jets must satisfy $|\eta| < 2.4$; we change the jet multiplicity threshold to 3 because of our change in focus to direct squark-pair production in addition to gluino-pair production;
- $H_T > 300$ GeV;

- $H_T^{miss} > 300$ GeV ; All other criteria remain almost the same as for the 2015 analysis (see Sec. 4.1.1).

The search is performed in the following 160 ($= 4 \times 4 \times 10$) exclusive intervals of the four search variables:

- **3** N_{jet} bins: 3-4, 5-6,7–8, ≥9;
- **4** N_{b-jet} bins: 0, 1, 2, ≥3;
- **10** bins in $H_T$and H_T^{miss}: defined below

Bin	H_T range [GeV]	H_T^{miss} range [GeV]
1:	300-500	300-350
2:	500-1000	300, 350
3:	>1000	300-350
4:	350-500	350-500
5:	500-1000	350-500
6:	>1000	350-500
7:	500-1000	500-750
8:	>1000	500-750
9:	750-1500	>750
10:	>1500	>750

4.2.2 Background Estimation

Similar methods to estimate various SM backgrounds are employed in this analysis even though their relative composition has changed from 2015 owing to the change of baselines. As the phase space has changed, these methods are optimized accordingly. Below we describe the changes for the $Z(\rightarrow \nu\bar{\nu})$+jets background. Details of other background estimations can be found in Ref. [14]

$Z(\rightarrow \nu\bar{\nu})$+jets:

- A new high statistics MC sample is used to calculate the transfer factor $R_{Z/\gamma}$ helping to reduce the related systematic uncertainty.
- For the calculation of photon purity, a charged isolation template is used instead of $\sigma_{i\eta i\eta}$ as we find it to have a better performance.
- Trigger for the photon control region is also changed.

  Note that due to the change in H_T^{miss} and N_{jet} threshold the contribution of Z background is increased in many bins.

4.2.3 Uncertainties

The uncertainties are treated in the sameway as it is done with 2015 data. The % uncertainties are also similar when compared with the previous analysis.

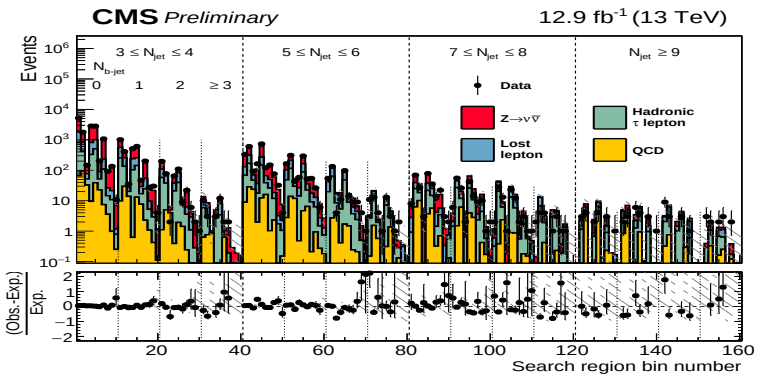

Figure 7: Data vs. the SM background before fit

4.2.4 Results

Fig. 7 shows the predicted background vs. the observed data. The background is found to be statistically compatible with data for all 160 regions. Thus, we do not see any evidence for new physics. The 95% CL uppper limits on the cross section are shown in Fig. 8. These plots show a lot of improvements in terms of expected limits when compared to the previous results with 2.3 fb^{-1} of data (Fig. 6). But the mismatch between the observed and expected curves owing to the deficit of events in some of the bins.

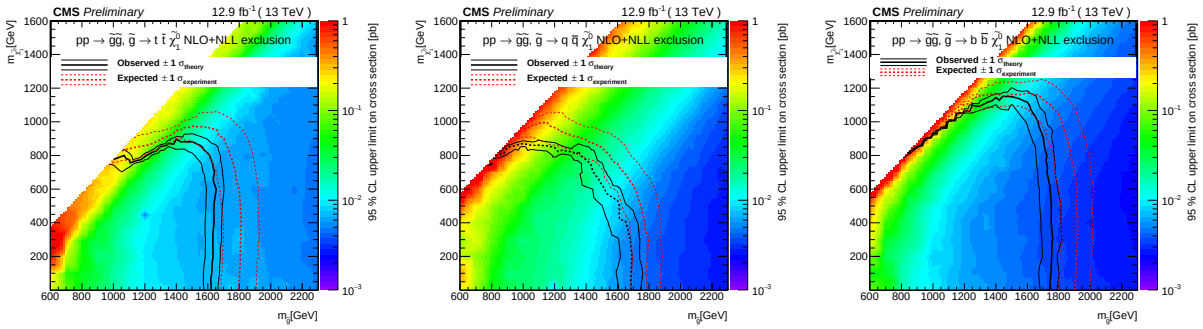

Figure 8: The 95% CL upper limits on the production cross sections for four top quark(left),four light quark(middle) and four b-quark(right) in the final state

5 Search for Pair Production of First Generation Leptoquarks

The eejj final state is the end product of pair-produced leptoquarks with each of them decaying to an electron and a jet. Events containing two electrons and atleast two jets are selected, where the two leading p_T electrons and jets are used in the analysis.

A set of cuts on three search variables are optimized for an improved signal sensitivity. The variables are:

- S_T is the scalar sum of p_T of two electrons and the leading two jets;
- M_{ee} is the invariant mass of two leading electrons; and
- $M^{min}_{\ell,j}$ is defined as the smaller lepton-jet invariant mass for the assignment of jets and leptons to leptoquarks that minimizes the LQ-$\overline{\text{LQ}}$ invariant mass difference.

  The cuts on the above variables are calculated for different leptoquark mass points. The details of selections and optimization procedures can be found in Ref. [15].

The data vs. backgrounds plots for two variables are shown in the following figure.

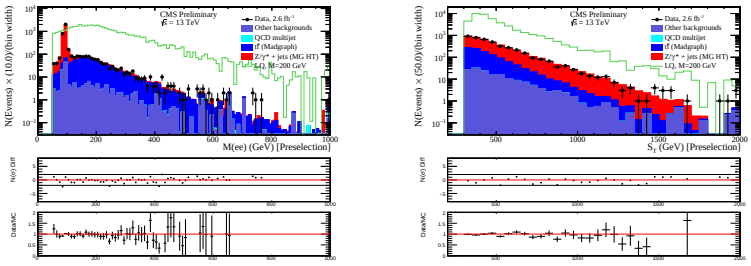

Figure 9: Data vs. background plots forM_{ee}(left) and S_T(right)

5.1 Background Estimation

The major backgrounds from SM processes are Z+jets and $t\bar{t}$, where single top, W+jets, diboson, and γ+jets contribute at a lower level. There is also an instrumental background from QCD events with jets faking electrons. Below we describe how the backgrounds are determined in this analysis.

- The Z+jets and $t\bar{t}$ background shapes are taken from MC simulation, and normalized to data using the eejj preselection. More details are given in Ref. [15].
- Single top, W+jets, diboson, and γ+jets backgrounds are derived completely from MC. The samples are scaled to the cross sections.
- QCD background is determined using a data-driven fake rate method, as described in Ref. [15].

5.2 Results

Similar to the two SUSY analyses, we observe no significant excess of events as compared to the SM backgrounds. The broad excess of the events that was seen in the 8 TeV analysis [16] has dissapeared with the 2015 data. We set the upper limits on the cross section times branching fraction using the same tool as before. Fig. 10 shows both the observed and expected limits for different leptoquark masses. We exclude leptoquark masses up to 1130 GeV from this study.

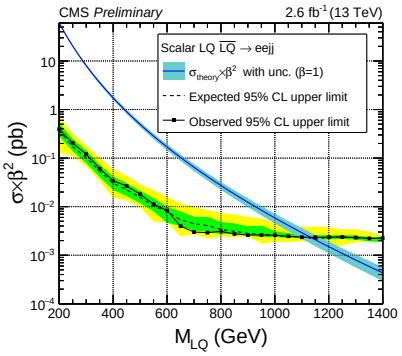

Figure 10: The 95% CL upper limits on the production cross sections as a function of leptoquark mass.

6 Advanced Pileup Mitigation Techniques

6.1 Jet Grooming

Grooming, introduced in Ref. [17], is intended to remove soft and wide-angle radiation from a jet. It is typically used to reduce the overall jet mass of QCD (quark- and gluon-initiated) jets while retaining the larger jet mass for jets originating from heavy particles such as the top quark and W/Z/H boson. Additionally, it can also help minimize the pileup dependence on jet mass. In general, grooming alters the soft structure of the jet while other jet structure observables may rely on this soft structure. Here we explore three grooming methods to mitigate the effects of pileup on large-R jets (R=0.8).

1. Pruning [18] reclusters the constituents of the jet through the Cambridge-Aachen (CA) algorithm [19] using the same distance parameter. At each step in the clustering algorithm, the softer of the two particles i and j to be merged is removed when the following conditions are met:

$z_{ij} = \frac{\min(\mathrm{p_{T_i}},\mathrm{p_{T_j}})}{\mathrm{p_{T_i}}+\mathrm{p_{T_j}}} < z_{\mathrm{cut}}$ and $\Delta R_{ij} > \frac{2\times r_{\mathrm{cut}}\times m_J}{\mathrm{p_T}}$,

where m_J and $\mathrm{p_T}$ are the mass and transverse momentum of the originally-clustered jet, and z_{cut} and r_{cut} are parameters of the algorithm.

2. Trimming [20] ignores particles within a jet that fall below a dynamic threshold in $\mathrm{p_T}$. It reclusters the constituents of the jet using the kt algorithm [11] with a radius parameter r_{sub}, accepting only the subjets that have $\mathrm{p_{T_{sub}}} > \mathrm{p_{Tfrac}}\lambda_{\mathrm{hard}}$, where $\mathrm{p_{Tfrac}}$ is a dimensionless

cutoff parameter, and λ_{hard} is some hard QCD scale chosen to be equal the p_T of the original jet.

3. Soft-drop [21] declusters the jet recursively to remove soft and wide-angle radiation from the jet. The jet is reclustered using the CA algorithm. Then the jet is declustered and at each step, subjets j_1 and j_2 are used to define the following condition:

$$\frac{\min(p_{T_{j1}}, p_{T_{j2}})}{p_{T_{j1}} + p_{T_{j2}}} > z_{\text{cut}} \times \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

where the algorithm parameters are z_{cut} and β. If the condition is met, the declustering continues, otherwise only the leading p_T subjet is kept. In the case when $\beta = 0$, soft drop can be considered a generalization of the modified mass drop tagger (MMDT) [19].

The (groomed) masses are corrected for pileup using a four-vector safe subtraction. In the cases of soft drop and trimming, the four-vector subtraction corrects the jet p_T and mass at each step in the algorithm. For pruning however the correction is applied to the final product using the pruned jet area. The parameters for the grooming algorithms explored in this study can be found in Ref. [22]

6.2 Results

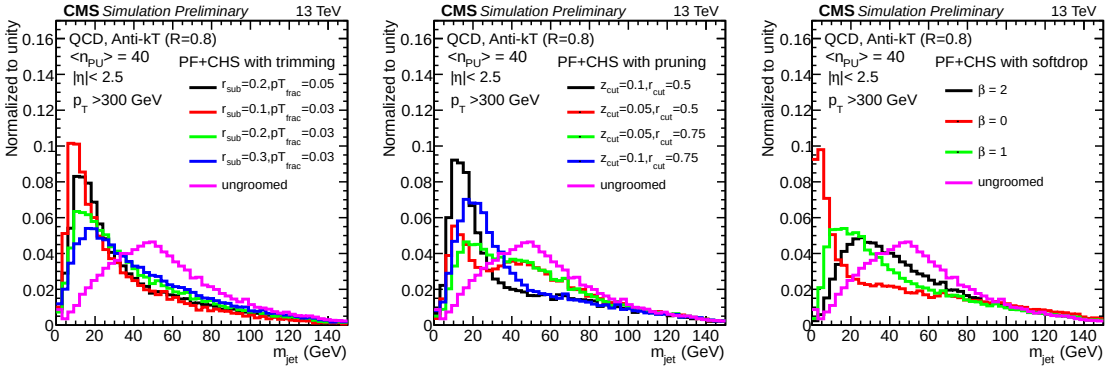

Figure 11: The mass distributions of PFCHS jets after application of trimming (left), pruning (pruning) and soft-drop (right) for different set of parameters. The ungroomed mass distribution is shown to compare the aggressiveness of grooming methods.

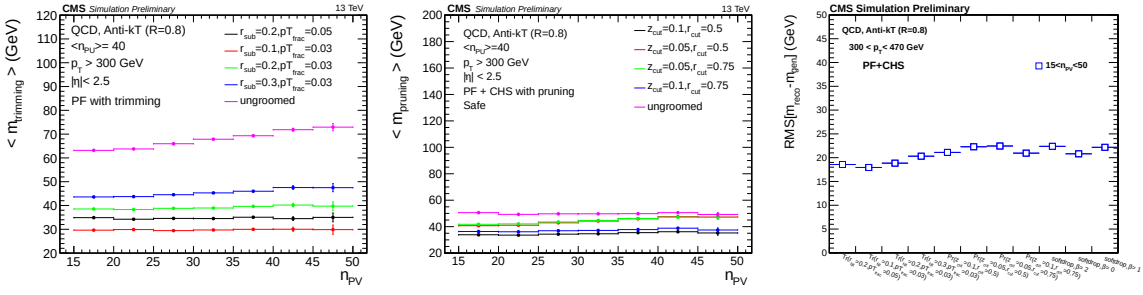

Figure 12: The plots shows variation of average jet mass vs. number of primary vertices for trimming (left), pruning (middle) and RMS vs. various grooming techniques(right).

The groomed mass distributions of charged hadron subtracted PF (PFCHS) [10, 22] jets are shown in Fig. 11. One sees that after the application of grooming owing to the reduction of pileup the peaks in the distributions shift to left whereas in the ungroomed case they have an unphysical looking like peak in the higher side. Even though these one dimensional mass distributions do not tell everything but it says because of the application of grooming the mass distributions change aggressively.

Average jet mass and jet resolutions are taken as quality parameters for determining the robustness of pileup mitigation algorithms. The plots of average jet mass vs. the number of primary vertices are shown in Fig. 12. The left plot shows the average jet mass is stable for trimmed jets but not for ungroomed jets. However once the charged hadron subtraction is applied jet mass becomes stable considerably that can be seen from the middle plot. The right plot summarizes the results from three different grooming techniques considered here.

7 Summary

We have presented the searches for SUSY and leptoquarks with 13 TeV pp collision data recorded with CMS. The SUSY search targets for direct gluino, stop and squark pair production in all hadronic final state. Two separate searches with 2.3 fb^{-1} of 2015 and 12.9 fb^{-1} of 2016 data yield null results of new physics. Similarly no signatures of any signal are found in the first generation scalar lepto quark search performed with 2.6 fb^{-1} of the 2015 data. Upper limits on the signal cross section are obtained in all cases significantly extending the previous 8-TeV limits.

Lastly a study on advanced pileup mitigation techniques is presented. Performance of various jet grooming methods like trimming, pruning and softdrop are compared. Optimized quality parameters of these techniques are recommended for use.

References

[1] H. Nilles, "Supersymmetry, supergravity and particle physics," *Phys. Rep.* **110** (1984) 1.

[2] G. Farrar and P. Fayet, "Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry," *Phys. Lett. B* **76** (1978) 575.

[3] G. G. S. Dimopoulos, "Naturalness constraints in supersymmetric theories with nonuniversal soft terms," *Phys. Lett. B* **357** (1995) 573.

[4] P. S. Alwall and N. Toro, "Simplified models for a first characterization of new physics at the LHC," *Phys. Rev. D* **79** (2009) 075020.

[5] B. Schrempp and F. Schremp, "Light Leptoquark," *Phys. Lett. B* **153** (1985) 101.

[6] S. Dimopoulos and L. Susskind, "Mass Without Scalar," *Nucl. Phys. B* **155** (1979) 237.

[7] J. C. Pati and A. Salam, "Lepton Number as the Fourth Color," *Phys. Rev. D* **10** (1974) 275.

[8] W. Buchmuller and D. Wyler, "Constraints on SU(5) Type Leptoquarks," *Phys. Lett.* **B177** (1986) 377–382.

[9] **CMS** Collaboration, S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC," *JINST* **3** (2008) S08004.

[10] **CMS** Collaboration, F. Beaudette, "The CMS Particle Flow Algorithm," in *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*, pp. 295–304. 2013. `arXiv:1401.8155 [hep-ex]`. https://inspirehep.net/record/1279774/files/arXiv:1401.8155.pdf.

[11] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm," *JHEP* **04** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[12] **The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group** Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011," Tech. Rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug, 2011. https://cds.cern.ch/record/1379837.

[13] **CMS** Collaboration, "Search for supersymmetry in the multijet and missing transverse momentum channel in pp collisions at 13 TeV," Tech. Rep. CMS-PAS-SUS-15-002, CERN, Geneva, 2015. https://cds.cern.ch/record/2114817.

[14] **CMS** Collaboration, "Search for supersymmetry in events with jets and missing transverse momentum in proton-proton collisions at 13 TeV," Tech. Rep. CMS-PAS-SUS-16-014, CERN, Geneva, 2016. https://cds.cern.ch/record/2205158.

[15] **CMS** Collaboration, "Search for pair-production of first generation scalar leptoquarks in pp collisions at $\sqrt{s} = 13$ TeV with 2.6 fb^{-1}," Tech. Rep. CMS-PAS-EXO-16-043, CERN, Geneva, 2016. https://cds.cern.ch/record/2205285.

[16] **CMS** Collaboration, "Search for Pair-production of First Generation Scalar Leptoquarks in pp Collisions at sqrt s = 8 TeV," Tech. Rep. CMS-PAS-EXO-12-041, CERN, Geneva, 2014. https://cds.cern.ch/record/1742179.

[17] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new Higgs search channel at the LHC," *Phys. Rev. Lett.* **100** (2008) 242001.

[18] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches," *Phys. Rev.* **D81** (2010) 094023.

[19] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, "Towards an understanding of jet substructure," *JHEP* **09** (2013) 029.

[20] D. Krohn, J. Thaler, and L.-T. Wang, "Jet Trimming," *JHEP* **02** (2010) 084.

[21] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop," *JHEP* **05** (2014) 146.

[22] **CMS** Collaboration, "Study of Pileup Removal Algorithms for Jets ," Tech. Rep. CMS-PAS-JME-14-001, CERN, Geneva, 2016. https://cds.cern.ch/record/1751454.