SEARCHING FOR A HEAVY PARTNER TO THE TOP QUARK

JOSEPH VAN DER LIST

Abstract. We present a search for a heavy partner to the top quark with charge $\frac{5e}{3}$, where e is the electron charge. We analyze data from Run 2 of the Large Hadron Collider at a center of mass energy of 13 TeV. This data has been previously investigated without tagging boosted top quark (top tagging) jets, with a data set corresponding to 2.2 fb^{-1}. Here, we present the analysis at 2.3 fb^{-1} with top tagging. We observe no excesses above the standard model indicating detection of $X_{5/3}$, so we set lower limits on the mass of $X_{5/3}$.

1. Introduction

1.1. The Standard Model

One of the greatest successes of 20th century physics was the classification of subatomic particles and forces into a framework now called the Standard Model of Particle Physics (or SM). Before the development of the SM, many particles had been discovered, but had not yet been codified into a complete framework. The Standard Model provided a unified theoretical framework which explained observed phenomena very well. Furthermore, it made many experimental predictions, such as the existence of the Higgs boson, and the confirmation of many of these has made the SM one of the most well-supported theories developed in the last century.

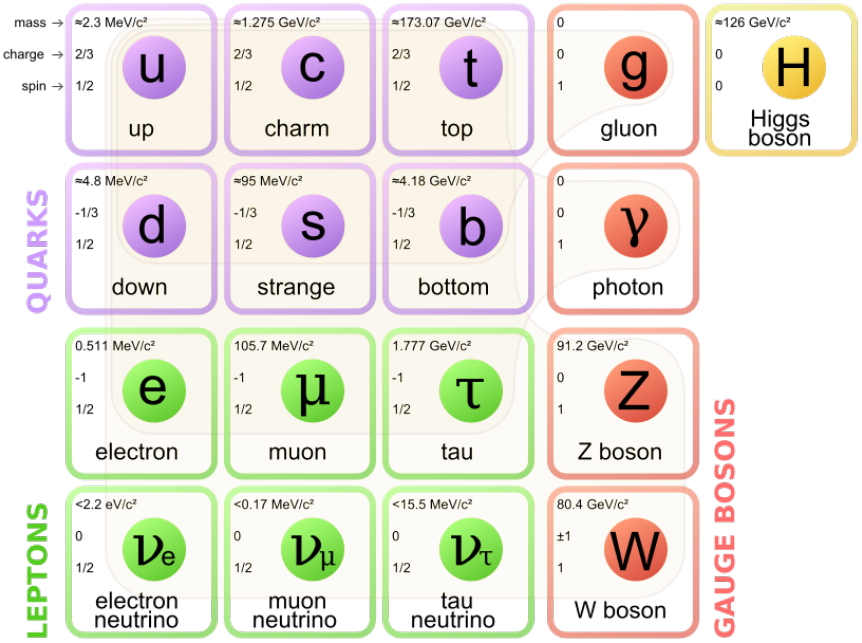

Figure 1. A table showing the particles in the standard model of particle physics. [7]

Broadly, the SM organizes subatomic particles into 3 major categories: quarks, leptons, and gauge bosons. Quarks are spin-$\frac{1}{2}$ particles which make up most of the mass of visible matter in the universe; nucleons (protons and neutrons) are composed of quarks. Quarks have color charge, and interact with each other via

the strong force. There are six quarks, three are up-type and three are down-type. The up-type quarks (up (u), charm (c), and top (t)) have a charge of $+\frac{2}{3}e$, and the down-type quarks (down (d), strange (s), and bottom (b)) have a charge of $-\frac{1}{3}e$. Composite particles which are made of quarks, like the proton, are called hadrons.

Leptons are also spin-$\frac{1}{2}$ particles. Notably, leptons do not interact via the strong force as quarks do. One of the leptons, the electron (e), is one of the three components of the atom. The muon (μ) and the tau (τ) are particles which are similar to the electron, and share its charge of -e, but are heavier. The other three leptons are neutrinos. The electron, muon, and tau each have an associated neutrino, denoted ν_e, ν_μ, and ν_τ respectively. Neutrinos are very light neutral particles which rarely interact with other matter. They are primarily seen as products of reactions involving the weak nuclear force.

The four gauge bosons are force-mediating particles with integer spins. The photon is massless and chargeless, and mediates the electromagnetic force. The gluon mediates the strong nuclear force while also carrying color charge, meaning that it also is affected by the strong force. The weak force is mediated by two different gauge bosons, the W and the Z. The W boson can have a charge of either +e or -e, while the Z boson is chargeless. The photon, W, and Z bosons can all also be considered the mediators of the eletroweak force, a unification of the electromagnetic and weak forces which occurs at high energies.

The Higgs boson is the most recently discovered particle in the SM. From experiments, it is known that the bosons have mass, but in order for this to happen, theory requires that there exist a field with which the bosons can interact that breaks gauge symmetry. This field is called the Higgs field, and the Higgs boson is the excitation of this field.

All of the particles in the SM have a corresponding particle of antimatter. These antiparticles have several important properties. Notably, the mass of any antiparticle is the same as the mass of the "regular" particle, while the electric and color charges are reversed. A particle and antiparticle can annihilate each other when they interact, and convert their mass to energy. Gluons, photons, and Z bosons are their own antiparticle, while W^+ and W^- are antiparticles of each other. Antiparticles (when they are not the same as the regular particle) are denoted by a bar over the particle symbol.

While the SM has great explanatory power, there are questions surrounding it which aren't yet answered. For example, there is no experimentally detected gauge boson which mediates the gravitational force. The search for physics which patch the holes in the SM is one of the most dynamic areas in high energy physics.

1.2. The $X_{5/3}$ Particle

In the summer of 2012, CMS and ATLAS announced the discovery of a new particle suspected to be, and now considered to be, the Higgs boson. The new particle, however, does not conform completely to the parameters expected from theory. Notably, it has a mass of about 125 GeV, while mass was expected to be much higher. The disparity between the expected and observed masses is an aspect of the hierarchy problem, one of the most fundamental problems yet to be solved surrounding the Higgs boson.

In current predictions of the Higgs mass, known particles contribute terms which are divergent, so the calculated mass becomes very large. Within known physics, the only way to bring the predictions for the Higgs mass in line with observations is to invoke fine-tuning, i.e adjusting parameters of theory to very precise levels. Because fine-tuning is generally not considered to be satisfactory, new physics is sought to solve the problem.

An appealing way to eliminate the problem of divergent terms from current particles is to invoke symmetry. The top quark, especially, contributes strongly to this effect and so symmetric partners to it are particularly interesting. One possible partner is a heavy quark with charge 5e/3 predicted by composite Higgs models, where the Higgs boson is not a fundamental particle but a composite one. This investigation deals with this heavy partner, which we call $X_{5/3}$.

Because we cannot directly detect the $X_{5/3}$ particle[1], we must look for it indirectly. Theoretical predictions tell us about the decay of the $X_{5/3}$, so we can analyze recorded events in the detector to search for signatures which could be produced by an $X_{5/3}$. The presence of such signatures would indicate that there may have been an $X_{5/3}$ created.

[1]The $X_{5/3}$ particle is very heavy and hence has a very short lifespan. Because of the short lifespan, an $X_{5/3}$ will not have the chance to interact with the detector.

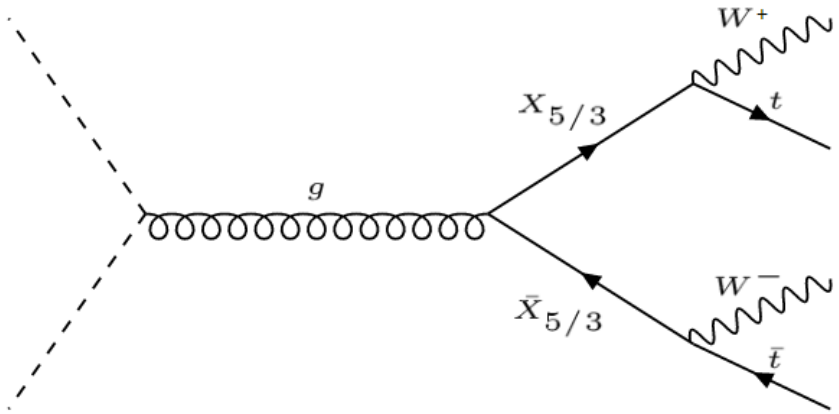

FIGURE 2. The pair production of an $X_{5/3}$ and an $\overline{X}_{5/3}$ from a gluon. The newly created particles then decay in $t + W^+$ and $\bar{t} + W^-$ respectively.

Figure 2 shows an example of the production and decay of $X_{5/3}$. Here, an $X_{5/3}$ and an $\overline{X}_{5/3}$ are pair produced by a gluon and then decay into $tW^+\bar{t}W^-$. Afterwards, the top decays to bW^+ and the anti-top decays to $\bar{b}W^-$.[2] There are multiple channels which the process can follow after the top decays, and there are advantages and disadvantages to analyzing each one. For example, the hadronic channel, where all of the W bosons decay hadronically (e.g. $W^+ \to c\bar{s}$) is the most likely, but intersects with a large amount of background which must be accounted for. On the other hand, the dileptonic channel, where two of the W's decay leptonically (into e, μ, or τ, plus a ν), is much less likely, but has a lower background. We investigate the "semi-leptonic" channel, where one of the W bosons decays leptonically while the rest decay hadronically. This channel is more likely than the dileptonic channel, but has less background than the hadronic channel, reaching a compromise between data volume and background presence. We further divide this channel into subchannels based on whether the lepton was an electron or muon.

1.3. **Compact Muon Solenoid**

The Compact Muon Solenoid (CMS) is one of the two largest experiments being conducted at the Large Hadron Collider (LHC). The LHC is a 27 kilometer particle accelerator built near Geneva, Switzerland by CERN which is designed to accelerate protons close to the speed of light and collide them. Currently, the LHC is performing its second run, with a combined energy of its two proton bunches of 13 TeV in the center of mass of collisions. CMS is designed to measure events caused by these collisions in order to probe the physics of the high energy realm. The CMS detector is built in many layers, each of which performs a different function. Figure 3 shows a slice of the detector with the various layers.

The innermost layer of the detector is composed of silicon trackers, essentially pieces of silicon between an anode and cathode. The trackers react to the passage of charged particles, giving information on their position. When a charged particle passes through a detector, it ionizes some of its atoms. The applied electric field then causes the electrons to drift to anode and cathode, giving a small current reading. The key feature of the trackers is that they are able to measure the position of particles passing through them very accurately. This means that the layer can be relatively light, and so interfere with the particles as little as possible. Ideally, the tracker layer measures the path of all charged particles through it without absorbing any energy from any of them.

The layer outside of the tracker is the electromagnetic calorimeter, or ECAL. This layer is designed to absorb electrons and photons in order to measure their energy. The ECAL is built out of Lead Tungstate crystals and photodetectors. When an electron or photon enters the crystals, they scintillate, and the resulting light is measured by the photodetectors. Because the light from the scintillation is more intense

[2]There are other possible decay channels for the tops, such as into a W and some other bottom-type quark, but they are far less likely than the bW mode, so they are not investigated.

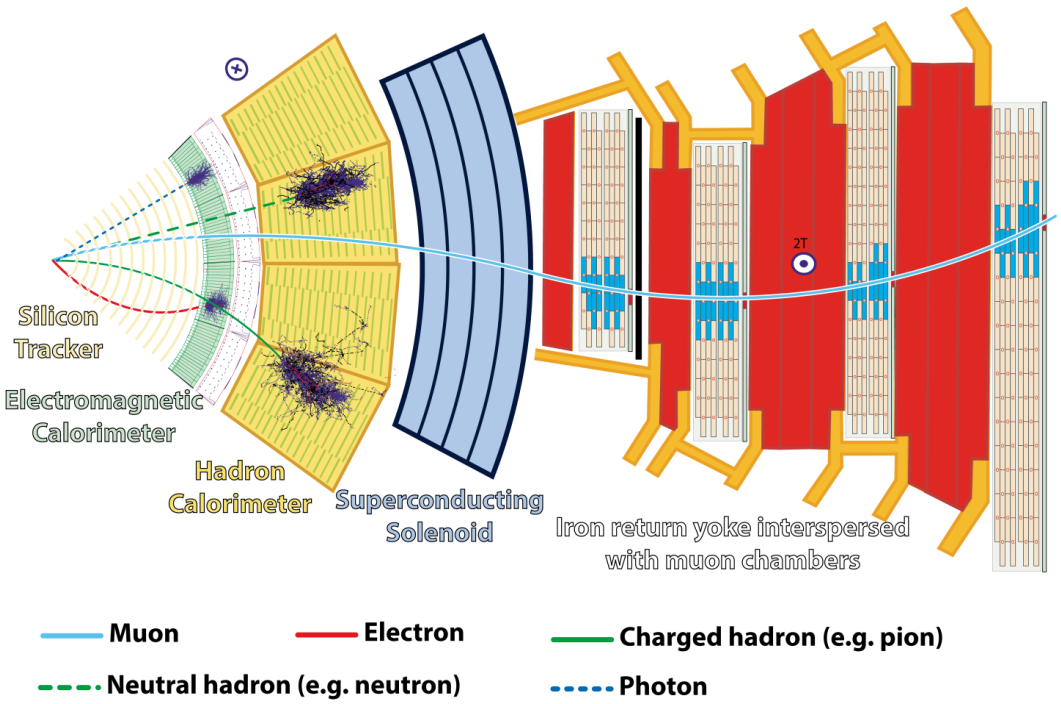

FIGURE 3. A slice of the CMS detector perpendicular to the beam axis, showing how various particles interact with the detector. [1]

the more energetic the absorbed particle is, the energy of incoming particles can be measured by the output of the photodetectors.

Outside of the electromagnetic calorimeter is the hadron calorimeter, or HCAL. The HCAL is designed to measure the energy of hadrons (e.g protons) which pass through it. To do so, the HCAL is constructed from alternating layers of absorbers and scintillators. The absorbers, brass and steel plates, stop the energetic hadrons and produce showers of other particles of lower energy. These lower energy particles can also interact with further out plates and produce even more particles. In this way, one high energy hadron creates a "tower" of many particles. The many particles in these towers pass through the scintillators, and their passage is measured similarly to the particles in the ECAL. The energy of all the particles in a tower can be summed to retrieve the energy of the original hadron.

The next layer, the solenoid after which CMS is named, does not actually detect particles, but does serve an important function. The solenoid is a large coil of superconducting wire which produces a 4T magnetic field parallel to the beam axis within its coils and a 2T magnetic field outside. The purpose of the magnetic field is to bend the paths of charged particles. Because the curve of a charged particle's path in a magnetic field is proportional to its momentum, the recorded particle paths in combination with their energies can be used to reconstruct their momentum and other information.

The outermost layer of CMS is made of two alternating parts. The first is the iron return yoke. The return yoke serves several purposes. First, it regulates the magnetic field outside of the solenoid, keeping it parallel to the beam axis. Second, the large mass of iron blocks almost all particles from passing through the muon chambers, letting only muons and neutrinos (which don't interact with the chambers) through[3]. Between the return yoke layers are the muon drift tubes. These devices are tubes filled with gas and a wire which serves as a cathode. When a charged particle passes through the gas, it ionizes the atoms, and the resulting electrons are drawn to the wire, resulting in a detectable current spike. By correlating signals from the different muon chambers, the path of a muon through the detector, and therefore its momentum, can be determined.

[3]Muons are relatively heavy, so they aren't slowed as easily as electrons, and can penetrate the large iron masses. Tauons are even heavier, but are much shorter-lived than muons and don't reach the outer layers.

1.4. **Motivations**

The CMS collaboration has performed a search for the $X_{5/3}$ particle before[2]. Using data from Run 2 of CMS at 13 TeV, investigation of the semileptonic channel of the $X_{5/3}$ decay has put an observed lower limit on the mass of right-handed $X_{5/3}$ particles of 700 GeV and on left-handed $X_{5/3}$ particles of 715 GeV.

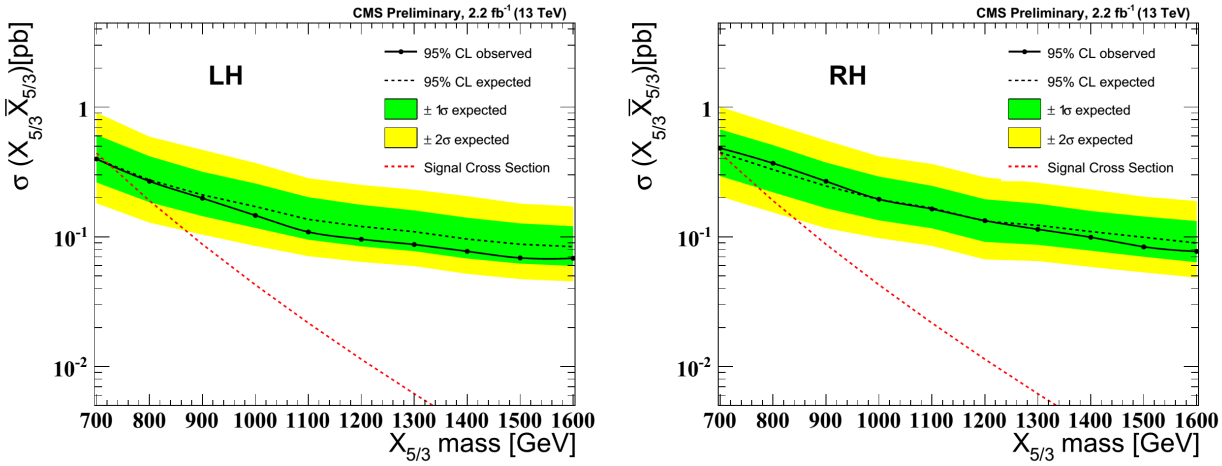

FIGURE 4. The expected and observed upper limits on the production cross section of $pp \rightarrow X_{5/3}\overline{X}_{5/3}$ in the semileptonic channel at a 95% confidence level as a function of the mass of $X_{5/3}$. Plots are shown for both left-handed and right-handed chiralities.[2] The intersection of the cross section limits with the cross section of signal gives the lower limit on mass.

This previous analysis, however, did not include every analytical method useful in the search for $X_{5/3}$. Notably, this previous analysis tagged jets which were produced from W's and b's, but did not tag jets sourced from top quarks. As noted above, top quarks are one of the major products of $X_{5/3}$ decay, so tagging their product jets has the potential to increase the sensitivity of the analysis.

In our investigation, we have included an algorithm to top tag jets. Furthermore, we analyze data at a slightly higher luminosity than the previous search. Both of these factors have given us a greater sensitivity to $X_{5/3}$ than previous analyses.

2. Analytical Methods

2.1. **Sample Sources**

The raw data used in our analysis can be split into two major parts: experimental data produced by CMS (referred to as "data"), and samples from simulated events created with various algorithms using Monte Carlo methods (referred to as "Monte Carlo" of MC). Monte Carlo simulations are used for two major purposes: first, signal samples for $X_{5/3}$ are produced in simulation so that they can be compared to observed data to assess whether it is consistent with an event containing an $X_{5/3}$. Second, background processes are modeled using simulation. This allows us to account for background in observed data to retrieve signals of interest.

The data is from Run 2 of the LHC, and was collected at a center of mass energy of 13 TeV and an integrated luminosity of 2.3 fb^{-1}. $X_{5/3}$ signals were generated with a combination of MadGraph5 and PYTHIA8. Signals were generated for various mass points and also for both left and right-handed coupling chiralities (see Table 2.1).

To simulate the response of the CMS detector to signals, GEANT4 was used. Background processes used also MadGraph and PYTHIA as well as POWHEG v2 and MG5_AMCNLO.

$X_{5/3}$ Mass (GeV)	Cross Section (pb)
700	.442
800	.190
900	.88
1000	.43
1100	.22
1200	.11
1300	.062
1400	.034
1500	.019
1600	.011

TABLE 1. The cross section of the $X_{5/3}$ signal samples at each of the mass points. Left and right-handed samples had the same cross section.

2.2. **Lepton and Jet Reconstruction**

In order to make use of the data from the detectors, we must be able to identify the signatures of particles and events. Some particles are easier to identify than others. In the first stage of identification, electrons, muons, and jets are identified.

Muons are very easy to identify in the data, as they are the only products of the proton collisions which interact significantly with the layers outside the calorimeters. The basic requirements for a track through the detector to be identified as a muon are: at least six hits in the silicon tracker, plus at least one hit in the muon chambers, and that the path fit to these hits is a good fit to predicted paths through the detector.

Electron identification is more complicated than muon identification because electrons do not interact with muon chambers, but are stopped in the ECAL. The basic requirements for electron selection are a recorded path through the silicon tracker and detection in the ECAL. The ECAL detection and the silicon track are required to be correlated (i.e. the ECAL event could reasonably be caused by a particle with the recorded track in the silicon).

Jets are groups of hadrons produced by the decay of high-energy particles. Because all of the particles in a jet are produced from the same source, which tends to have high momentum, they are grouped into a tight cone with its point at the source. So, the primary challenge in identifying jets is deciding when close hadrons are part of a jet and when they are not. For our analysis, we use the anti-k_T clustering algorithm[3] to cluster particles into jets. If a jet is identified and found to contain a lepton which is also identified, then the lepton energy is subtracted from the jet[4].

2.3. **Event Selection**

During a measurement session, the CMS detector produces an enormous volume of data. Some parts of this data are extremely valuable and may contain information about new physics, but some parts are not so useful. Signals from detectors may not represent events from the proton collisions: instruments and the surroundings may produce noise. Even if the data is useful, it may not be relevant to the current investigation. So, in order to be most efficient, cuts are applied to the raw data at various stages during analysis to clear away unneeded events.

The following pre-selection cuts were applied before the bulk of our analysis:

- #AK4 Jets ≥ 3
- Lepton $p_T > 50\ GeV$
- Missing $E_T > 100\ GeV$
- Leading Jet $p_T > 200\ GeV$
- Sub-Leading Jet $p_T > 90\ GeV$

The purpose of choosing these cuts is to ensure that the structure of events being analyzed matches the structure expected for the pair production and subsequent decay of $X_{5/3}$. The first requirement is that there

[4]Leptons and jets are considered separately during later stages of analysis. The lepton energy is subtracted from the jet so that it is not "double-counted."

are at least 3 jets constructed with the anti-k_T algorithm ("AK") with $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.4$ ("4")[5]. Because the final state of the semi-leptonic channel contains 3 hadronized W bosons as well as two b quarks, we expect several jets to be produced, so any event which might be from $X_{5/3}$ should have at least 3. $X_{5/3}$ particles are rather heavy, so their decay products should have fairly high energies. This is the reason for the 2nd, 4th, and 5th cuts, where we require the momentum perpendicular to the beam (transverse momentum) of leptons, the highest energy jet, and the second highest energy jet, to be above a threshold. The third cut requires that there be at least 100 GeV of transverse energy missing, i.e. at least 100 more GeV of energy entered the event than was accounted for by the detectors. Missing E_T signifies that energy was carried away in a form not registered by the detector. Semi-leptonic $X_{5/3}$ decay should not deposit all of its energy into the detector, e.g. the neutrinos produced by the leptonic W will not be detected. The final cut requires that at least one jet be tagged as originating from a b quark. There are two b quarks produced by the tops in the $X_{5/3}$ decay, so we expect to see b-tagged jets.

Before the final stages of analysis, more cuts were applied. These final selection cuts are tighter versions of the preselection cuts:

- #AK4 Jets ≥ 4
- Lepton $p_T > 80\ GeV$
- Missing $E_T > 100\ GeV$
- Leading Jet $p_T > 200\ GeV$
- Sub-Leading Jet $p_T > 90\ GeV$
- $\Delta R(\ell, \text{sub-leading jet}) > 1$

The last cut in this list serves to separate the $X_{5/3}$ signal from the background. When $\Delta R(\ell, \text{sub-leading jet}) > 1$, the ratio of signal to background is much higher than below. See section 2.6 for more.

2.4. **Tagging**

In our analysis, we tag some jets as having originated from one of three important sources: a W boson, a b quark, or a t quark. By identifying processes containing the three major constituents of the $X_{5/3}$ decay, we are able to better isolate processes which may be related to $X_{5/3}$.

2.4.1. *W Tagging.*

We W tagged jets with the following criteria:

- Jet $p_T > 200\ GeV$
- $|\eta| < 2.4$
- $65\ GeV <$ Pruned Mass $< 105\ GeV$
- $\frac{\tau_2}{\tau_1} < 0.55$

These conditions select jets which are consistent with the presence of a W boson. The W boson has a mass of about 80 GeV, so we expect the mass of the jet to be around this value[7]. Unlike general event selection where we use AK4 jets, for W and t tagging we use AK8 jets, which require $\Delta R < 0.8$. Jets produced from W bosons are also boosted, so we expect the jet momentum to be high. The jet should also not be directed too far from the beam axis, hence the limit on pseudorapidity. τ_2 and τ_1 are "2-subjettiness" and "1-subjettiness", respectively. N-subjettiness is a measure of how well a jet matches a model with N subjets. To calculate N-subjettiness, N axes are first drawn within the jet with an algorithm which tries to match the axes with the jet content. Then, τ_N is calculated using the following formula[4]:

$$\tau_N = \frac{\sum_{i \in particles} p_T^i min(\Delta R_{1,i}, \Delta R_{2,i}, \ldots, \Delta R_{N,i})}{\sum_{i \in particles} p_T^i R_{jet}} \tag{1}$$

[5] ϕ is the azimuthal angle of a particle in the plane perpendicular to the beam. $\eta = -ln(\frac{tan(\theta)}{2})$, where θ is the angle the particle makes with the beam, is the "pseudorapdity." ΔR measures how "spread out" a jet is.

[6] For more about these, see section 2.6.

[7] "Pruned" means that the jet has had constituents which aren't likely to be associated with the rest of the jet removed. Some elements of the jet may not originate from the same particle or process as the others; these are removed from the jet analysis.

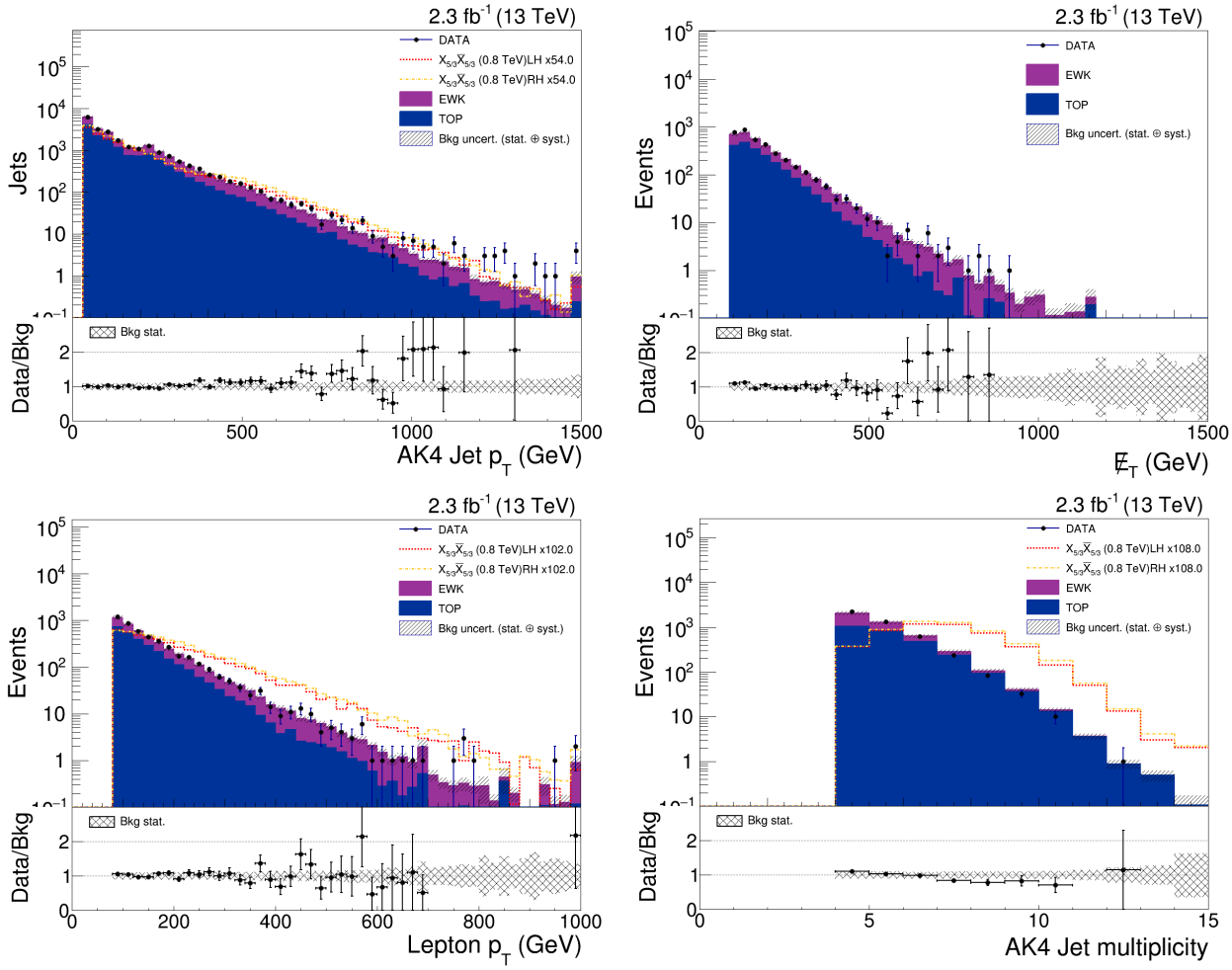

FIGURE 5. The distributions of the transverse momentum of jets (upper left), missing transverse energy (upper right), lepton transverse momentum (lower left), and number of AK4 jets (lower right) after cuts. Points in the histograms represent recorded data while the bars represent backgrounds as calculated by Monte Carlo simulation[6]. The red and yellow lines represent the simulated $X_{5/3}$ activity. The jet p_T plot does not show evidence of cuts because the cuts are only applied to the leading and sub-leading jets. The bottom bar of each plot shows the comparison between data and background for those points with recorded data events.

$\Delta R_{j,i}$ is measured between particle i and axis j. A low τ_N indicates that a jet is consistent with an N-subjet model. So, for W tagging, we require that a jet be more consistent with two subjets than one. Figure 6 shows the distribution of the number of events with each number of W tagged jets.

2.4.2. *b Tagging.* To b-tag jets, we used a medium working-point discriminator recommended for use with Run 2 data [5]. The discriminator describes the likelihood that a jet was produced by a b quark. Tighter working points mistag fewer jets but are less efficient. Looser working points are more efficient but mistag more jets. Medium working points balance these two factors. Figure 6 shows the distribution of the number of events with each number of b tagged jets.

2.4.3. *t Tagging.* The previous analysis performed on this data was done with W and b tagging, but not top tagging. Without previous investigation into different top tagging procedures, we needed to find the most effective one for the analysis.

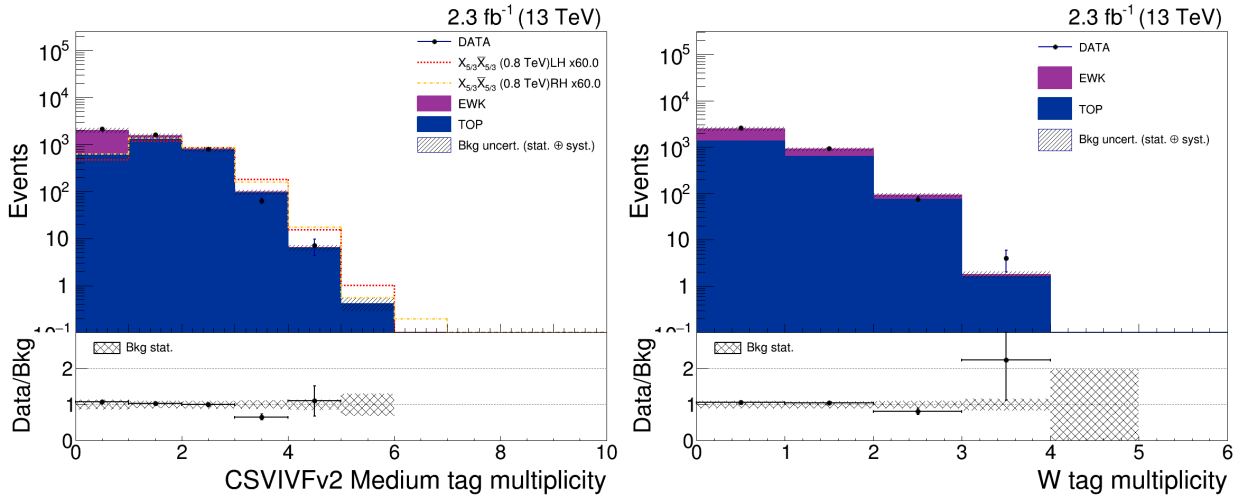

FIGURE 6. The distribution of the number of b tagged (left) and W tagged (right) jets.

Working Point	τ_3/τ_2	Max b-discriminant	Mistag Rate (%)	Expected Lower Limit of $X_{5/3}$ Mass (GeV)
1	<.44	>0	.1	794
2	<.54	>.79	.1	795
3	<.50	>0	.3	793
4	<.61	>.76	.3	794
5	<.59	>0	1	797
6	<.69	>.66	1	799
7	<.69	>0	3	803
8	<.75	>.39	3	797
9	<.86	>0	10	795
10	<.87	>.089	10	795

TABLE 2. Summary of the top tagging working points investigated in our analysis. The expected lower limit on the mass of $X_{5/3}$ serves as a measure of the sensitivity of the working point.

The top tagging method we employed put requirements on jet p_T, jet mass, N-subjettiness, and the presence of b quarks in subjets. We investigated ten recommended working points[6] in this method. All working points had identical restrictions on jet p_T and mass. Top quarks are heavy, so the jets they produced are quite boosted, and so the jet p_T was required to be above 400 GeV. The jet mass was required to be between 110 and 210 GeV. The top quark has a mass of about 175 GeV, so its jets should have roughly this mass.

Our working points differed in requirements on N-subjettiness and subjet b presence. Table 2.4.3 summarizes the points.

Each working point places a different restriction on τ_3/τ_2 (for discussion of N-subjettiness, see section 2.4.3). Five of the working points place restrictions on the maximum subjet b-discriminant. This discriminant serves a very similar purpose to the b-disciminant we used for b-tagging of jets. In this case, we measure the value for each subjet. This provides a measure of how likely it is that one of a jet's subjets was produced by a b quark. The most likely decay of the top quark produces a b, so we should expect to see b-produced subjets in a jet from a top. In order to measure the effectiveness of each working point, we performed a preliminary calculation of the expected value of the lower limit placed on the mass of $X_{5/3}$. While the values calculated here do not exactly match the values recovered from the final calculations, we were able to compare values and determine that working point 6 was the most sensitive set of restrictions; while working

point 7 gave a higher limit, the difference between the two was negligible and working point 6 is more similar to top tagging used in other analyses, so it is more comparable.

Figure 7 shows the distribution of the number of top tagged jets for our most sensitive working point, 6.

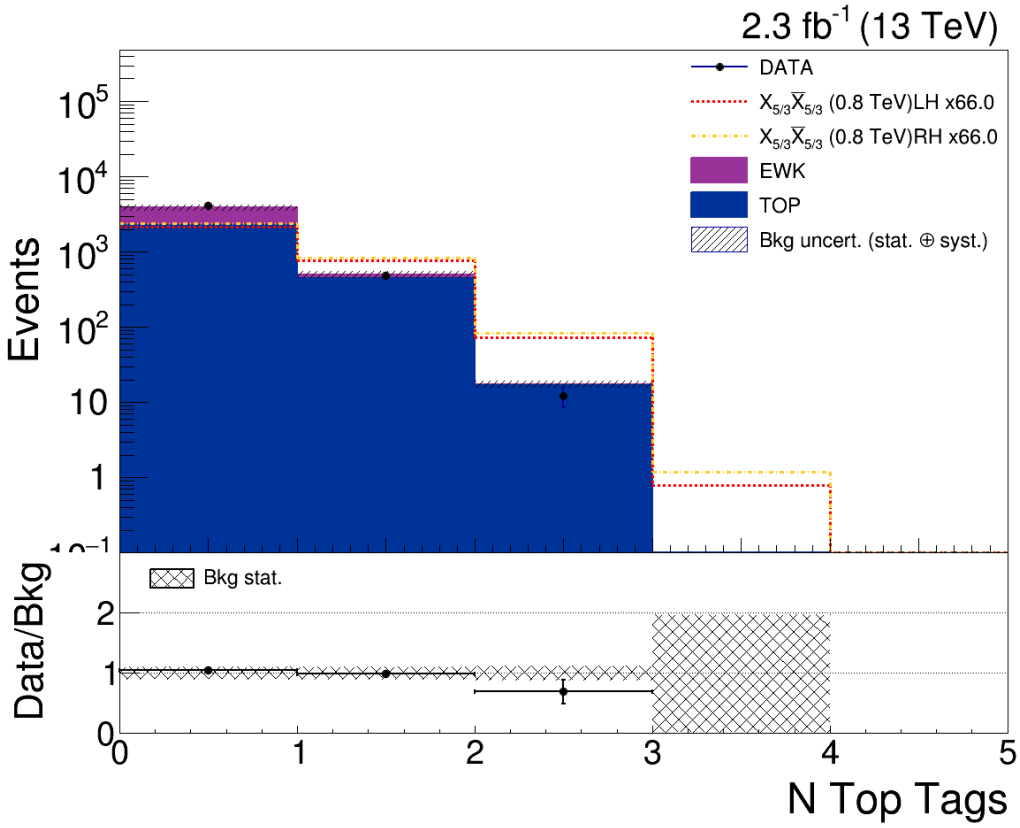

FIGURE 7. The distribution of the number of t tagged jets using our most sensitive working point.

2.5. **Categorization**

In order to achieve maximum sensitivity to the presence of $X_{5/3}$, we split events into categories defined by tags. Our categories are divided along each of the following parameters:

- Electron or Muon channel
- 0, 1, or 2+ b-tagged jets
- 0 or 1+ W-tagged jets
- 0 or 1+ t-tagged jets

It would be possible to divide events further, but we are limited by data. For example, we could add a category for 2+ top tagged jets, but while we do have some Monte Carlo available for 2 top tags (see Fig. 7), the number of events available to analyze is very low. Dividing into additional categories can add large amounts of computation time to the analysis, so we neglect categories which are unlikely to have many events or not contribute greatly to sensitivity.

In order to be able to examine the signals with minimal interference from background processes, we chose a discriminator which separates signals from the background well. For our analysis, we used Min[M(ℓ,b)], the minimum mass of lepton+b-tagged jet out of all pairs of lepton and b jet. This is the same discriminator used in the previous analysis, and was found to perform well.

Figure 8 demonstrates the effect of categorization of b-tagged jets.

For the full set of plots of the merged lepton channel, see Appendix A.

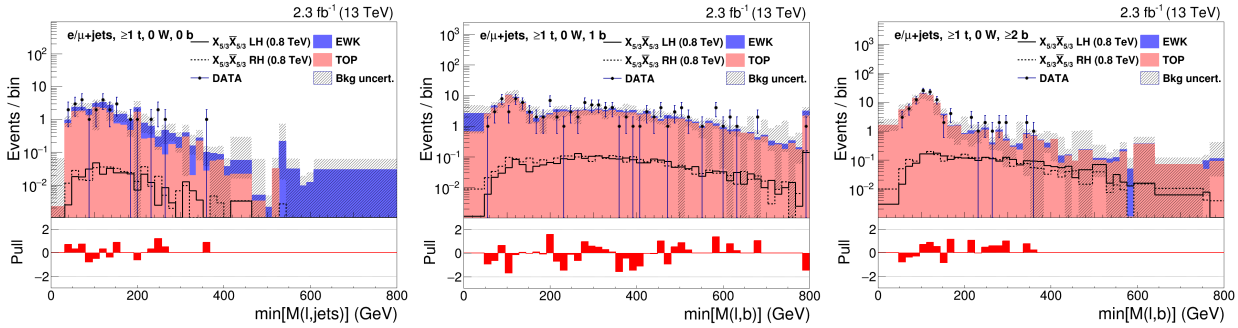

FIGURE 8. The distribution of Min[M(ℓ,b)] with merged lepton channels with 1 top-tagged jet, no W-tagged jets, and 0 (left) 1 (center), and 2 or more (right) b tagged jets. In the left plot, there are few events because not many events contain a t-tagged jet but not a b-tagged jet, because often top quarks are pair-produced and both decay into b quarks. The right plot has fewer events than the center because more events are singly b-tagged than are tagged more than once (see Fig. 6).

2.6. **Background Modeling**

We rely on our ability to model the background processes occurring in the data in order to perform analysis. Without knowledge of the characteristics of the background, we cannot be sure whether observed particles are from interesting events or already known SM interactions. In order to check that our background modeling is accurate, we perform new cuts on our data to define regions enriched in background processes, and then compare data in these regions to the Monte Carlo simulations.

There are two background processes which create states most like the one produced by the semi-leptonic $X_{5/3}$ decay channel. First, top quarks may be produced and then decay along a similar channel to those created by an $X_{5/3}$. For example, a high energy gluon similar to the one in Figure 2 may pair produce a t and a $\bar{\text{t}}$. The various backgrounds containing tops are labeled as TOP in our analysis. The second major group of background processes are the various electroweak interactions which produce W boson and jet activity. This background is labeled EWK in our analysis.

The same cuts used for final event selection are used, with two altered, to create the background regions. For both the TOP and EWK regions, we reverse the cut on the separation of the lepton and sub-leading jet, so we require $\Delta R(\ell, \text{sub-leading jet}) < 1$. Figure 9 shows the spearation between signal and background achieved by $\Delta R(\ell, \text{sub-leading jet})$.

To create the TOP region, we require at least 1 b-tagged jet, as background tops are still expected to decay mostly into b's. For the EWK region, we require instead that there be no b-tagged jets, as electroweak activity should not produce b quarks[8].

Figure 10 shows the Min[M(ℓ,b)] distributions in the TOP and EWK regions. In general, we have a good agreement between the data and Monte Carlo in both our regions. Disagreement between the two is taken into account as systematic uncertainty in the final analysis (see section 2.7).

2.7. **Uncertainties**

Our analysis is affected by several sources of uncertainty which need to be accounted for. Broadly, they can be divided into statistical and systematic uncertainties. Statistical uncertainties are caused by our inability to make perfect measurements and affect data broadly. Systematic uncertainties are caused by imperfections in data processing and analysis, and are somewhat easier to codify.

Systematic uncertainties can affect normalization (the magnitude of calculated distributions), shapes (the structure of calculated distributions), or both. For example, our simulated backgrounds do not exactly match the observed data, so we assign uncertainty related to the discrepancy. This can affect the normalization. As mentioned in section 2.4, our tagging algorithms are not perfect: they do not tag all possible jets, and mistag some jets which don't actually originate from the relevant particle. The uncertainty caused by these

[8]In this region, Min[M(ℓ,b)] becomes Min[M(ℓ,jets)], and we find the minimum mass of lepton+jet for all jets.

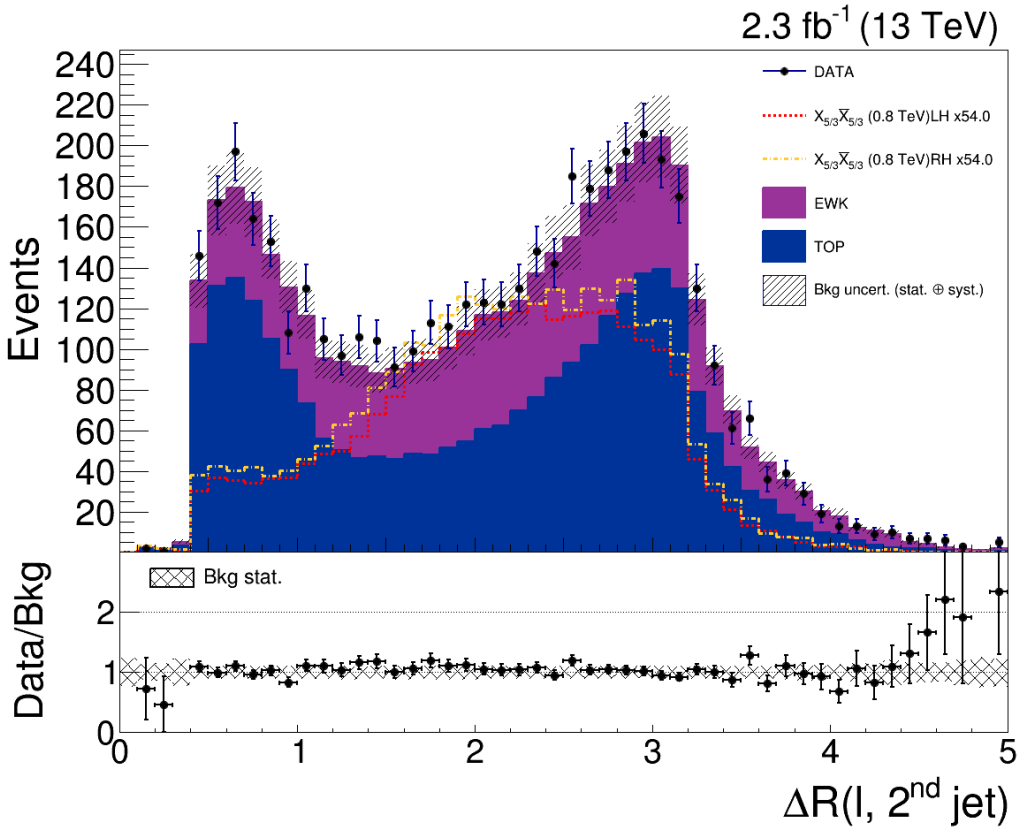

FIGURE 9. The distribution of $\Delta R(\ell, \text{sub-leading jet})$. Note that below 1, there is a large excess of background events over signal events.

tagging error can affect shapes and normalization. Table 2.7 summarizes that various sources of systematic uncertainty.

To observe how the errors in these parts of the analysis affect our results, we perform each of our calculations three times. First, we determine the results for the nominal values of all uncertain variables. Then, we recalculate the results for values +1 σ and -1 σ. The difference between the nominal and altered results gives the uncertainty.

For a more detailed discussion of the uncertainty parameters, see the CMS Physics Analysis Summary "Search for top quark partners with charge 5/3 at $\sqrt{s} = 13$ TeV" [2].

3. RESULTS

We did not observe any excesses in our data which would indicate the presence of $X_{5/3}$. Therefore, we calculate upper limits on the production cross section of $X_{5/3}$ in the semi-leptonic decay channel using Bayesian statistics. We determine both expected and observed values for these limits for both chiralities. Figure 11 summaries the results.

The cross section of the simulated $X_{5/3}$ signals at all mass points can be calculated with theory. By comparing the theoretical cross sections [8] with the upper limits set by our analysis, we set expected lower limits of 798 GeV and 786 GeV, and observed lower limits of 743 GeV and 722 GeV, on the left-handed and right-handed $X_{5/3}$, respectively.

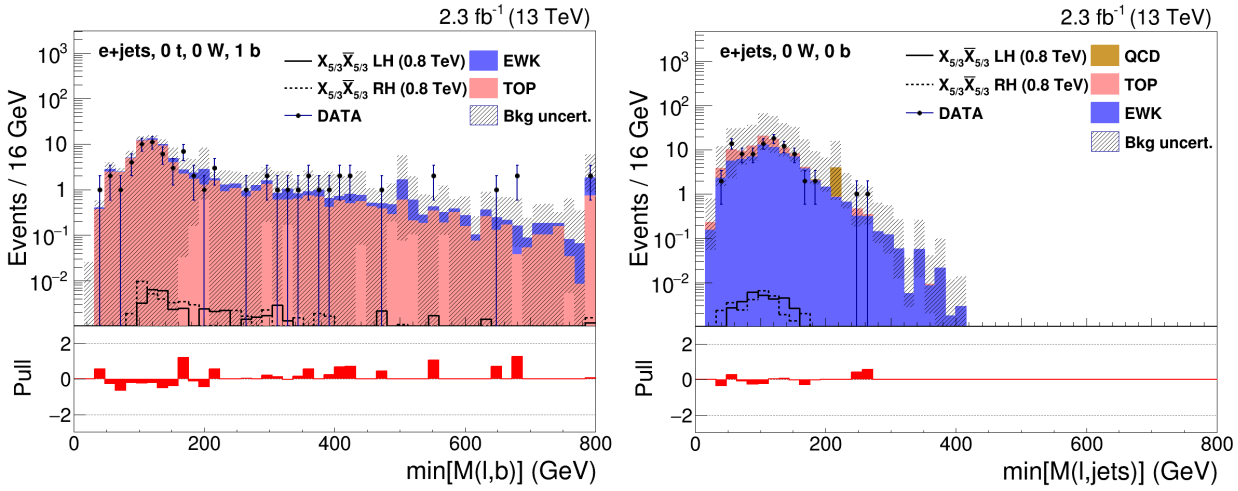

FIGURE 10. The Min[M(ℓ,b)] discriminator distributions in the TOP (left) and EWK (right) enriched regions. We have generally good agreement between data and Monte Carlo in these regions. Discrepancies are taken into account as systematic errors.

For the complete set of plots for each control region in the merged lepton channel, see Appendix B.

Parameter	Magnitude	Shape/Normalization	Signal	Background(s)
Lepton Trigger Efficiency	5%	Normalization	Yes	All
Lepton Isolation and ID Efficiency	1%	Normalization	Yes	All
Luminosity	2.7%	Normalization	Yes	All
TOP Modeling	6-17%	Normalization	No	TOP
EWK Modeling	11-17%	Normalization	No	EWK
b tagging	Gaussian	Both	Yes	All
Jet Energy Scale	Gaussian	Both	Yes	All
Jet Energy Resolution	Gaussian	Both	Yes	All
Jet Mass Resolution (W tagging)	Gaussian	Both	Yes	All
Jet Mass Scale (W tagging)	Gaussian	Both	Yes	All
Jet Reweighting	Gaussian	Both	Yes	All
PDF (Shape)	Gaussian	Shape	Yes	None
PDF (Both)	Gaussian	Both	No	All
Renormalization/Factorization (Shape)	Gaussian	Shape	Yes	None
Renormalization/Factorization (Both)	Gaussian	Both	No	All
Pileup	Gaussian	Both	Yes	All
Parton Shower Scale	Gaussian	Both	Yes	All
τ_2/τ_1 (W tagging)	Gaussian	Both	Yes	All
Top p_T	Gaussian	Both	No	TOP
top tagging	Gaussian	Both	Yes	TOP

TABLE 3. Summary of the systematic uncertainties in the samples used. "Gaussian" Magnitude indicates that the uncertainty values are different across signals and categories. Uncertainty from these is propagated by performing calculations for nominal and $\pm 1\sigma$ values.

4. CONCLUSION

We are able to get a significant increase in sensitivity due to top tagging. Table 4 shows our calculated results where jets are both top tagged and not top tagged. This isolates the effect of top tagging from the other sensitivity-increasing factors that are implemented in the analysis and are not discussed in this thesis,

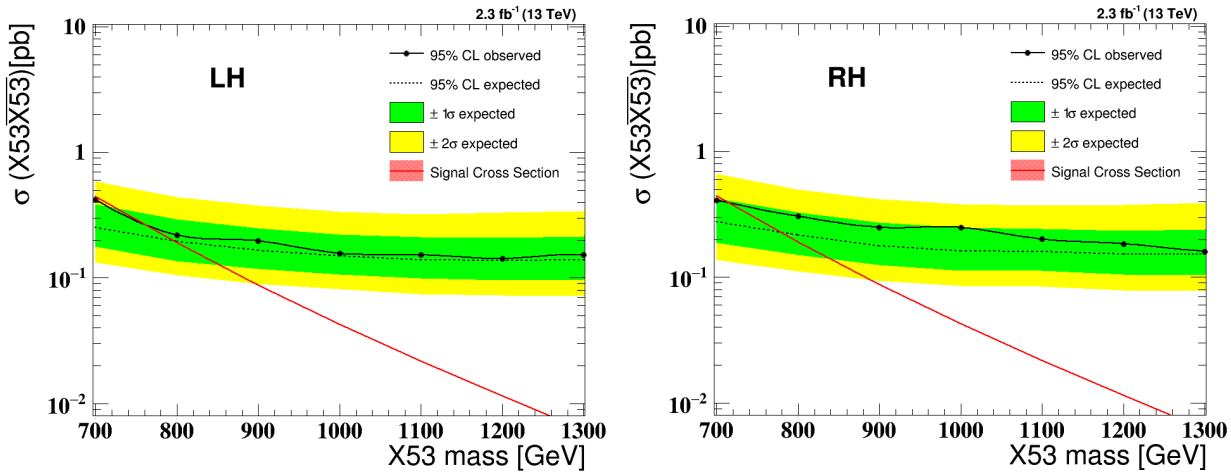

FIGURE 11. The calculated upper limits on the production cross section of $X_{5/3}$ as a function of mass. Also calculated are the cross sections of the various simulated $X_{5/3}$ signals. The intersection of these lines gives a lower limit on the mass of $X_{5/3}$.

Calculation	Expected (GeV)		Observed (GeV)	
	LH	RH	LH	RH
Top Tagging Used	798	786	743	722
No Top Tagging Used	761	729	<700	<700

TABLE 4. Summary of the $X_{5/3}$ mass limits obtained both with and without top tagging.

such as improvements in the W-tagging method and associated uncertainties. The difference between these calculations shows that top tagging alone increases the sensitivity of the analysis by roughly 50 GeV. This allows us to put stronger limits on the $X_{5/3}$ production cross-section, and hence its mass. We can conclude that top tagging is certainly an effective tool for searching for $X_{5/3}$. Future searches will no doubt benefit from implementing it.

REFERENCES

1. CERN (David Barney), *CMS Detector Slice*, https://cds.cern.ch/record/2120661?ln=en (2011).
2. The CMS Collaboration, *Search for top quark partners with charge 5/3 at* $\sqrt{s} = 13$ *TeV*, CMS Physics Analysis Summary B2G-15-006 (2015).
3. Cacciari, Salam, and Soyez, *The anti-k_T jet clustering algorithm*, arXiv:0802.1189 (2008).
4. Thaler and Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, arXiv:0802.1189 (2008).
5. The CMS Collaboration, *Usage of b Tag Objects for 13 TeV Data with 25ns bunch spacing and 74X reconstruction*, https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation74X (2016).
6. The CMS Collaboration, *Boosted Top Jet Tagging*, https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetTopTagging (2016).
7. Wikimedia User MissMJ, *The Standard Model of elementary particles*, `https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg` (2006).
8. Matsedonskyi, Panico, and Wulzer, *On the interpretation of Top Partners searches*, Journal of High Energy Physics Vol. 97, (2014).

Appendix A. Complete Category Plots

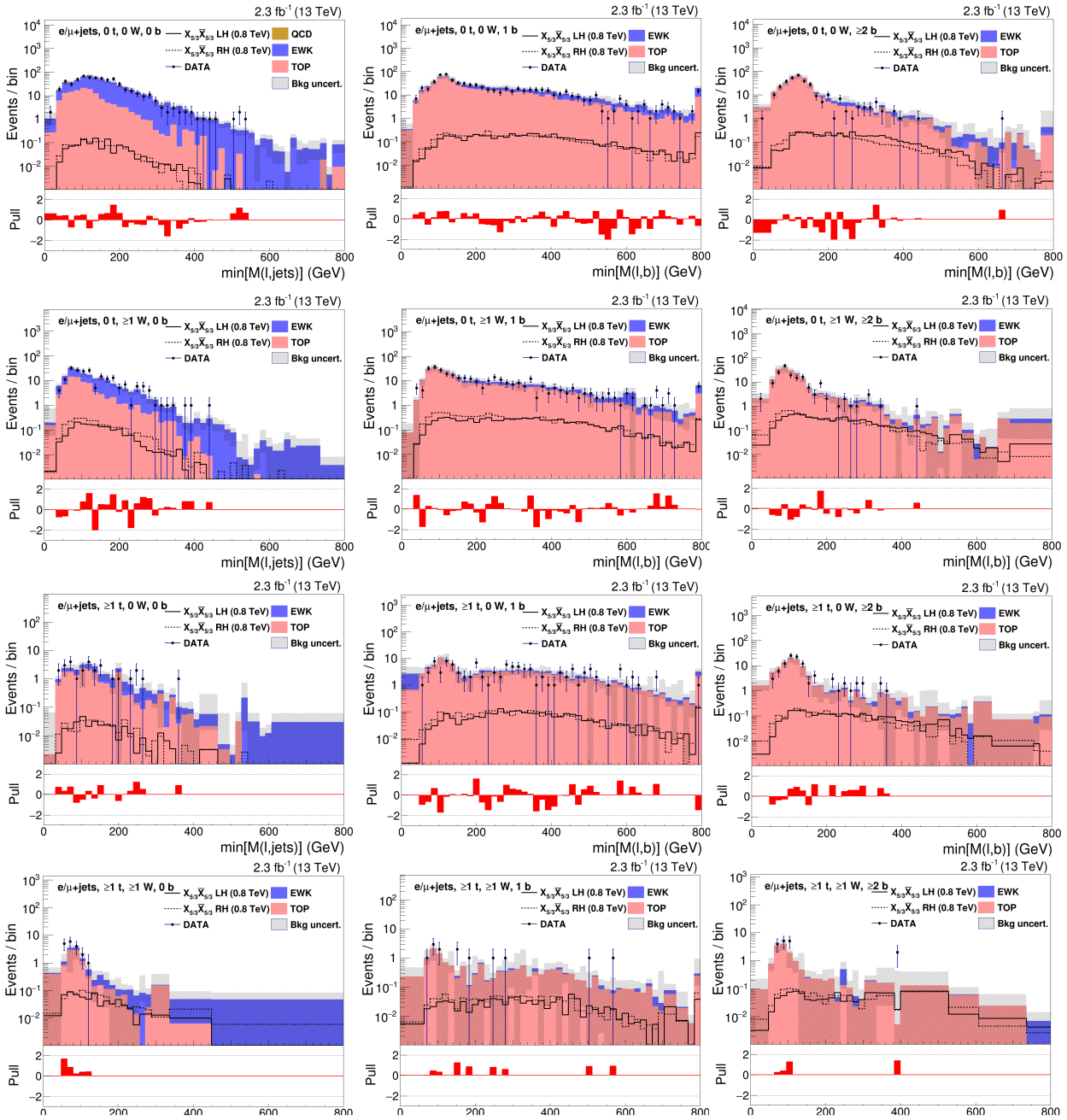

Figure 12. Plots of Min(M(l,b)) for all of the categories with the lepton channels merged.

Appendix B. Complete Control Region Plots

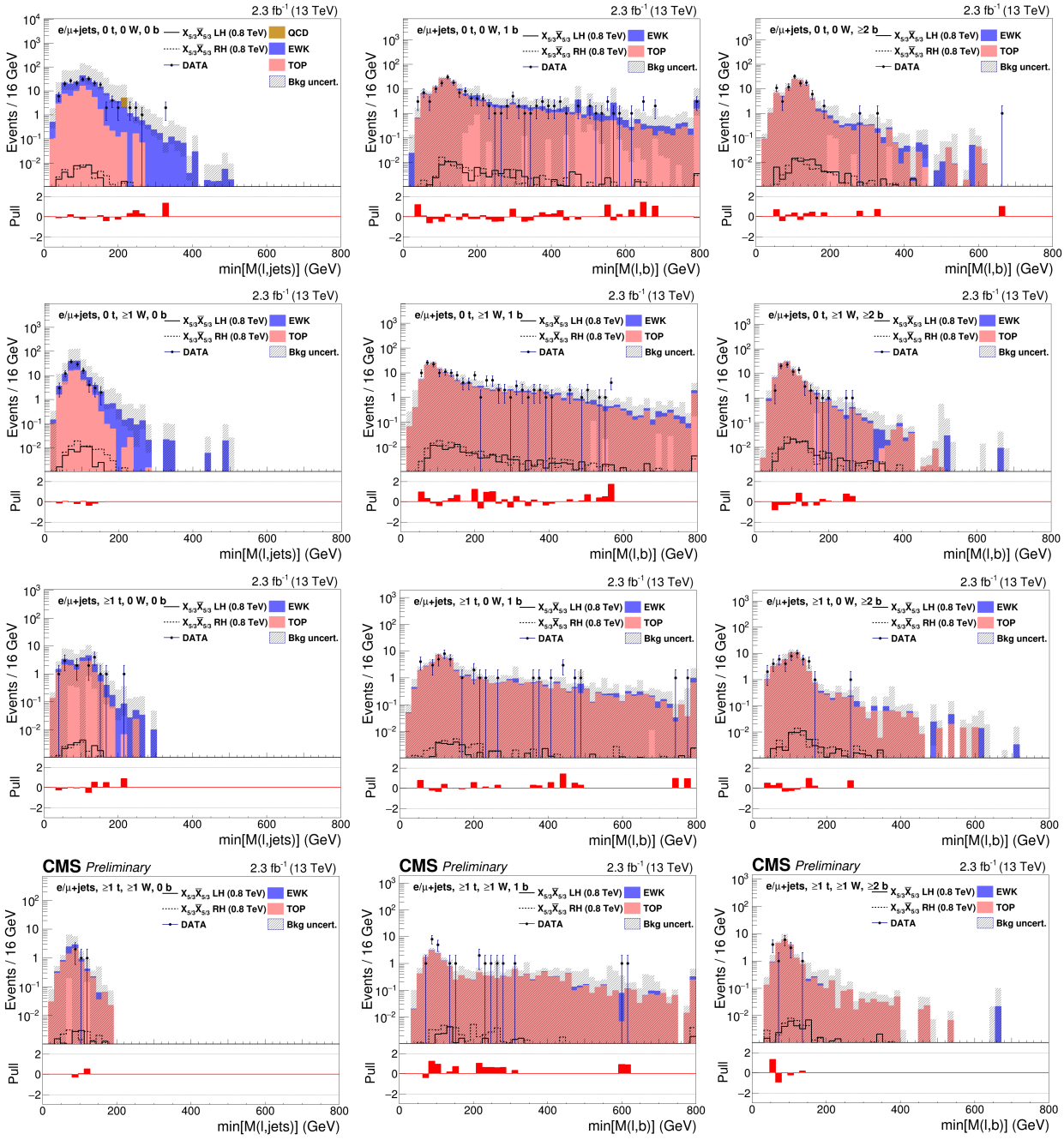

FIGURE 13. Plots of Min(M(l,b)) for all of the categories in the TOP background region with the lepton channels merged.

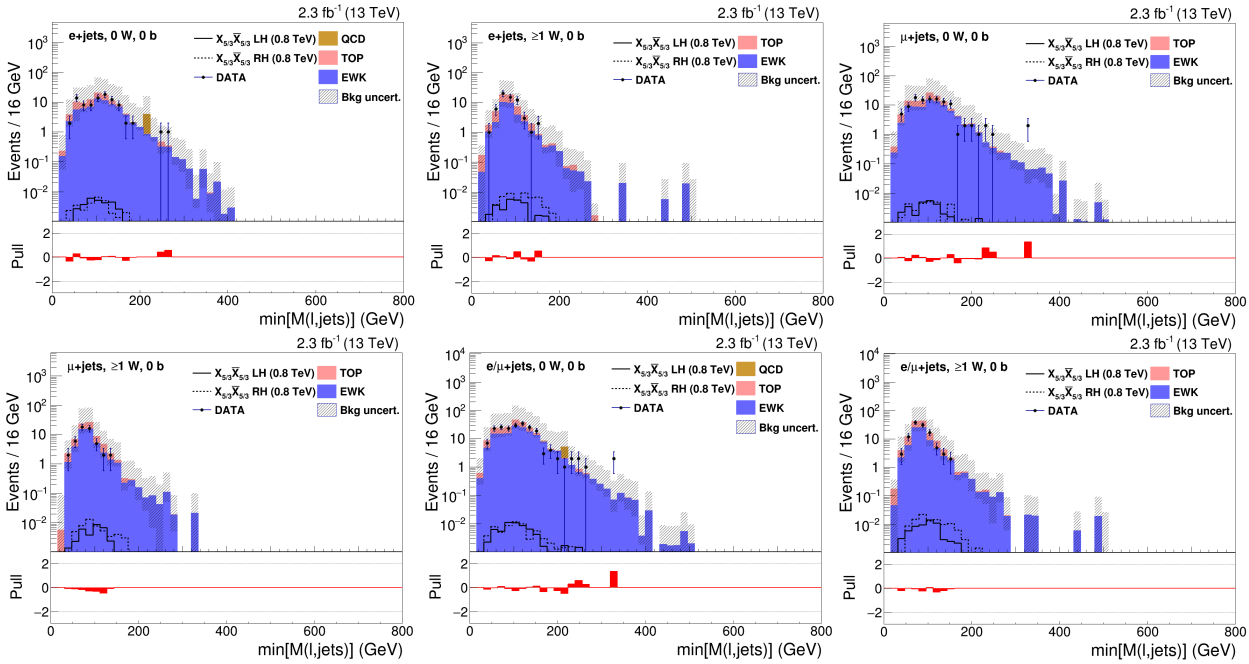

FIGURE 14. Plots of Min(M(l,b)) for all of the categories in the EWK background region with the lepton channels merged.