# Search For Trijet Resonance

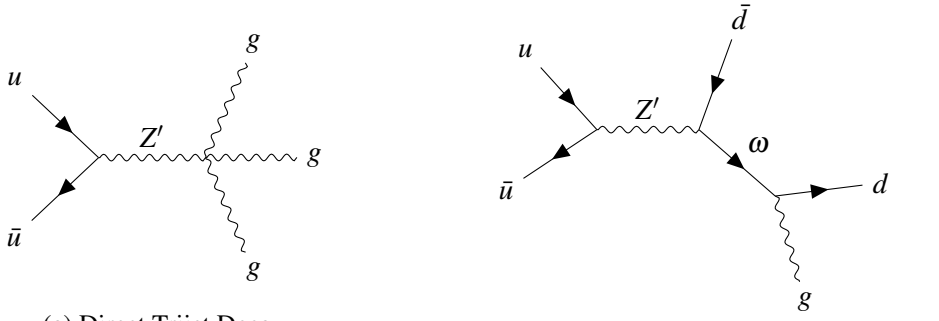
Ciaran Godfrey

**Abstract**

A search for resolved trijet production is performed in proton-proton collisions at 13 TeV. The search is performed over the entire run 2 dataset collected by the CMS detector which corresponds to an integrated luminosity of 150 $fb^{-1}$. Decay modes include direct decay and decay involving an intermediate dijet state.

## 1 Introduction

Up to this point in the search for beyond the standard model physics there has been a lot of effort put into the search for dijet decays but none into a trijet model. In this paper we look at the trijet invariant mass spectrum in order to find a resonance that implies a deviation from current theory. The specific signal models considered include states with and without intermediate resonances. These models are depicted in Fig. 1 in case of a Z' production, where Z' is a spin-1 particle, similar to the standard model Z boson, but much heavier.




(a) Direct Trijet Decay

(b) Trijet Decay with Intermediate Resonance

Figure 1: Trijet Signal Models

## 2 CMS Detector

The central component of the CMS detector is a superconducting solenoid with an internal diameter of 6 meters. The solenoid generates a fields of 3.8 Tesla. There are 4 layers to the detector. The first 3 which are contained inside the solenoind are, from the innermost to the outermost: the silicon tracker, the electromagnetic calorimeter, and

the hadron calorimeter. The 4th layer which is located outside the solenoid is the muon system. Each layer is sensitive to different particles and are used together to build a picture of of the collision products.

## 2.1 Silicon Tracker

The tracker detector is used to measure the momentum of charged particles from the collision. The strong magnetic field generated by the solenoid causes charged particles to curve. Lower momentum particles curve more so the tracker measures the position of the particles at several points as they leave the collision site. In order to disturb the particles as little as possible it takes extremely precise measurements so as to require as few points as possible. The detector is made of silicon pixels at the center near the high intensity collisions surrounded by silicon microstrips. These measure when a charged particle passes through. Particles detected include muons, electrons, and hadrons.

## 2.2 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is used to measure the energy of electrons and photons. This is done using lead tungstate scintilating crystals. When electrons and photons pass through the crystals they produce a short photon burst proportional to the energy of the particle. This photon burst is the measured by photodetectors. The ECAL is made up of a barrel and two endcaps. The barrel is made of 61,200 crystals split into 36 supermodules. The endcaps seal the ends and contain another 15,000 crystals.

## 2.3 Hadron Calorimeter

The hadron calorimeter (HCAL) measures the energy of hadrons as well as providing indirect evidence of noninteracting uncharged particles. HCAL is designed to - as completely as possible - detect every particle emerging from the collisions. This way if we see an asymmetry in the recorded momentum we can infer the existence of "invisible" particles. The HCAL uses alternating layers of absorber and scintillator materials to measure position, energy, and arrival time. When the energy is summed over many layers in a region this gives the total energy. This means what we see is not a single reading for each particle but "jets" of energy spread over an area. HCAL is split into the barrel (HB and HO), endcap (HE), and forward (HF) sections. The barrel is made of 36 wedges which form the last layer before the solenoid. HO is located outside the solenoid to catch any particles that escape HB. The two HF sections are located at either end of CMS to catch any particles that leave the collision at very shallow angles.

## 2.4 Muon Detector

Muons interact very weakly so they penetrate the first three layers of the the detector without being stopped. The Muon detectors are the outermost layer and are located outside the solenoid. Muons are the only particle likely to make it that far that will be detected. The subdector is made of 4 muon stations interweaved with "return yoke" plates. These operate on the same principle as the tracker in that they measure the path of the muons in a strong magnetic field in order to determine their momentum. The

muon chambers are aligned to the tracker to less than a millimeter so they can work together to reconstruct the path of the particle. In total there are 14000 muon chambers.

# 3 Data and Monte Carlo used

In addition to the the signal models we used simulated QCD multijet production binned in transverse momentum ($p_T$) to allow covering the entire $p_T$ spectrum of interest. This was used to train the BDT and as a stand in for data to blind ourselves while developing the analysis. For background estimation we just require that we get a smoothly falling function so an exact match between QCD and data is not required to the extent it is in other analyses.

For data we used the full Run 2 JetHT dataset and the SingleMuon dataset to run our trigger efficiency studies.

# 4 Event Selection

We applied 3 requirements to the data before processing it. The first was the requirement that the 3 leading jets have transverse momentum greater than 50 GeV. This selection was used to filter out uninteresting events. The second cut required all jets to be at |eta| less than 2.5 where eta is the pseudorapidity of the jet which encodes the angle of a jet's momentum with respect to the beam axis. Larger eta corresponds to smaller angles. This cut is required because the triggers PFHT 800, 900, and 1050 don't exist past this point. The third was to limit the maximum distance in eta between any two jets to 2.5. Without this cut the trigger turn on curve is too gradual and irregular making an accurate fit difficult. This cut gives minimal significance loss while simultaneously vastly increasing the accuracy of our trigger parameterization.

# 5 BDT

Due to the relative weakness of the signal compared to the background in data we next make cuts in an effort to make the signal more prominent. Rather than make standard cuts we used a boosted decision tree (BDT). A BDT is a multivariate classifier that is trained to distinguish signal from background. It takes known signal and QCD samples as input and creates a set of trees based on preset variables. The first tree is created to correctly identify as many events as possible. Every tree after that gives greater importance to events that were mis-classified by previous trees. The final decision for each event is a weighted vote of all the trees. The tree is applied to each event and based on the vote assigns a number (MVA value) between -1 and 1 to the event. A value close to -1 means the event is background-like while a number close to 1 means the event is signal-like. The variables used were:

- Leading jet $p_T$ over trijet mass
- Second leading jet $p_T$ over trijet mass
- Third leading jet $p_T$ over trijet mass
- Dijet mass (jets 1 and 2) over trijet mass

- Dijet mass (jets 1 and 3) over trijet mass
- Maximum eta difference between two of the three leading jets
- Minimum eta difference between two of the three leading jets
- Distance (Delta R) between second and third leading jets
- Leading jet eta
- Second leading jet eta

Some variables are normalized by the trijet mass so in order to express the ratios of energy contained in each jet rather than prioritizing based on the mass of the signal sample. This makes it more useful when training on signals of various masses at the same time as well as reducing correlation between the variables used. The performance of the BDT can be summed up by the ROC curve. A larger area under the curve corresponds to a more efficient BDT. Our ROC curve has an area of 0.819 and can be seen in figure 2. The curve is parameterized by the cut on the MVA value.

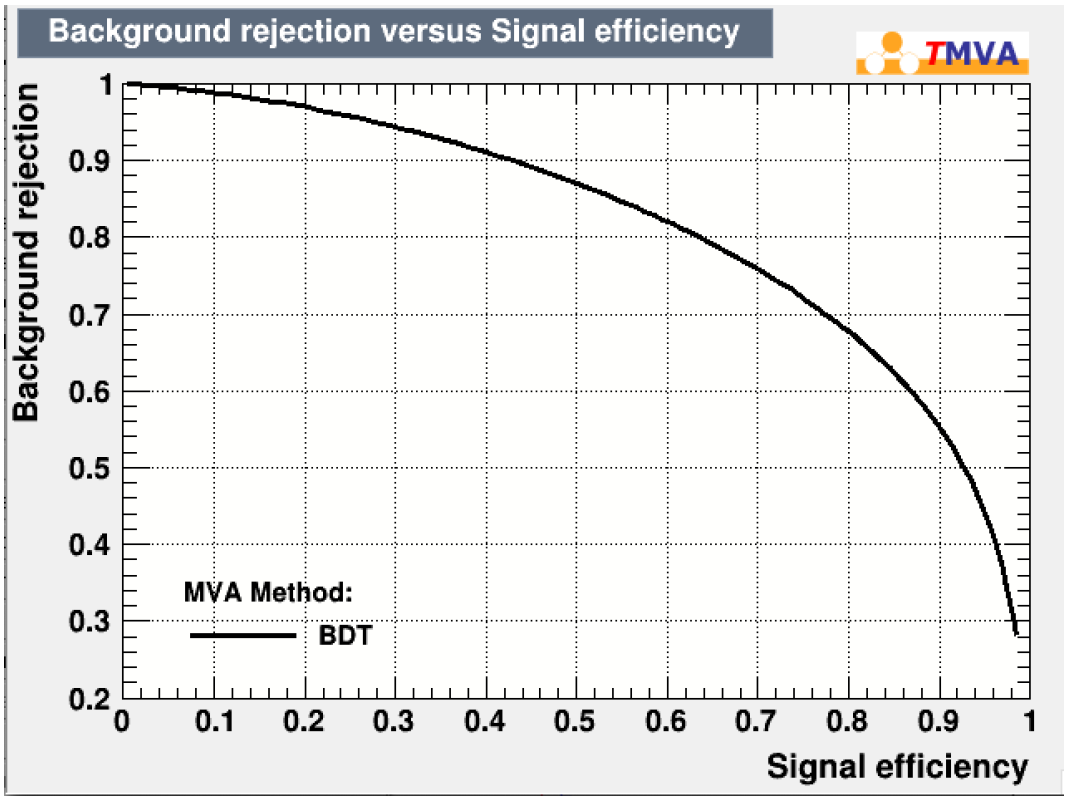



Figure 2: BDT signal efficiency vs. background rejection

While having a large area under the curve is good sign for the BDT as a whole, what determines how helpful it is is the significance gained from the optimal cut where we define signifcance as:

$$S = \frac{S}{\sqrt{S+B}} \tag{1}$$

Significance is plotted along with signal efficiency and background efficiency in figure 3. From this we can see that at the optimal cut value we have a signal efficiency of 0.9 and background efficiency of 0.5.

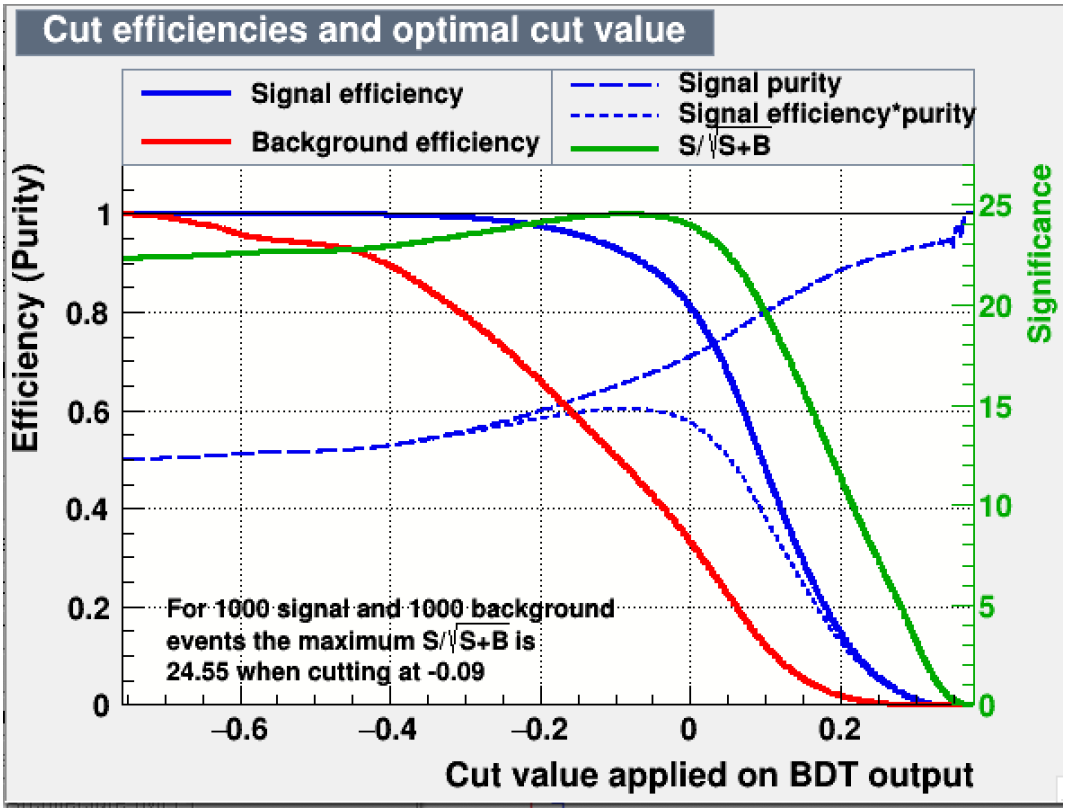



Figure 3: BDT signal efficiency vs. background rejection

The other important consideration is that the BDT will perform as well on the data sample as it did in training. To measure this when the BDT is trained the sample is split in half, one half is used for training the trees and the other for testing them. The risk is that if the BDT is trained too much it will start to pick up the statistical fluctuations of the training sample casuing it to perform much more poorly on the testing sample. This is checked for by plotting the MVA value of the testing events on top of the training events as can be seen in figure 4. We can see that the BDT performed equaly well on both samples so it was not overtrained.

## 6 Trigger Efficiency Parameterization

We require all events pass a set of jet triggers to ensure they are interesting. The triggers changed between 2016 and 2017. The full menu of high level triggers used for 2016 was: PFHT_800, PFHT_900, PFJet_450, PFJet_500, AK8PFJet_450, and CaloJet500_NoJetID. For 2017 and 2018 the menu was PFHT_1050, PFJet_500, PFJet_550, AK8PFJet_500, AK8PFJet_550, CaloJet500_NoJetID, and CaloJet550_NoJetID.

PFHT triggers place a minimum requirement on the scalar sum of transverse momenta. The PFJet triggers place a minimum requirement on Jet $P_t$. CaloJet triggers

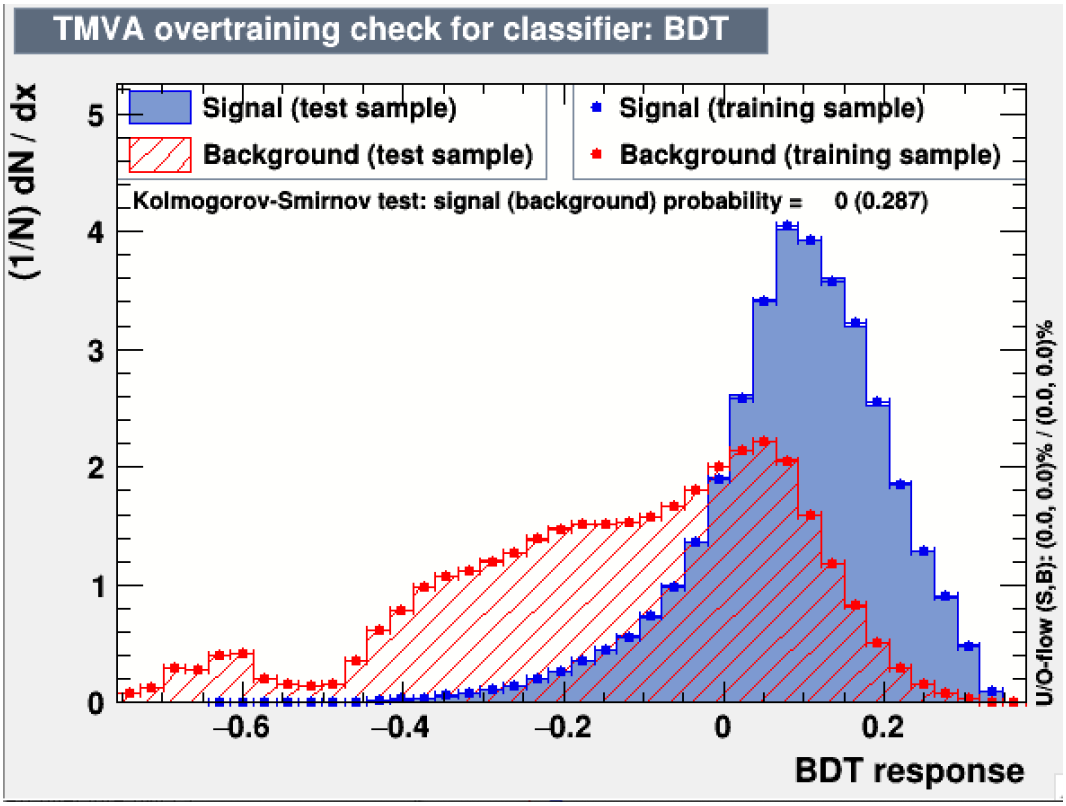



Figure 4: MVA distribution for training and testing samples (ovetraining check)

place a similar $P_t$ requirement but it is applied before the jet reconstruction algorithm (particle flow) is applied. AK8 triggers are the same as their counterparts but the jet reconstruction algorithm used allows for wider jets.

The triggers we are using are not fully efficient across the full range of masses we want to study. We correct for this by parameterizing the trigger efficiency function and including that in our background model. Since the triggers varied by year we parameterized each year separately. The form of our final model is then:

$$trigger\ efficiency = \frac{1}{total\ events} * \sum_{years} (number\ of\ events) * (efficiency\ function) \tag{2}$$

To perform this study we used the SingleMuon dataset. We then required events to pass the orthogonal triggers IsoMu_24 and IsoMu_27 before using them in the study. The results for 2016 are shown in figure 5.

The fit function used was:

$$\frac{a}{2} * (1 + (\tanh(x-b)/c) \tag{3}$$

where a, b, and c are free parameters.

## 7 Background Estimation

After applying cuts we model the background as a smoothly falling function with several free parameters times the trigger efficiency function. For each signal model we perform

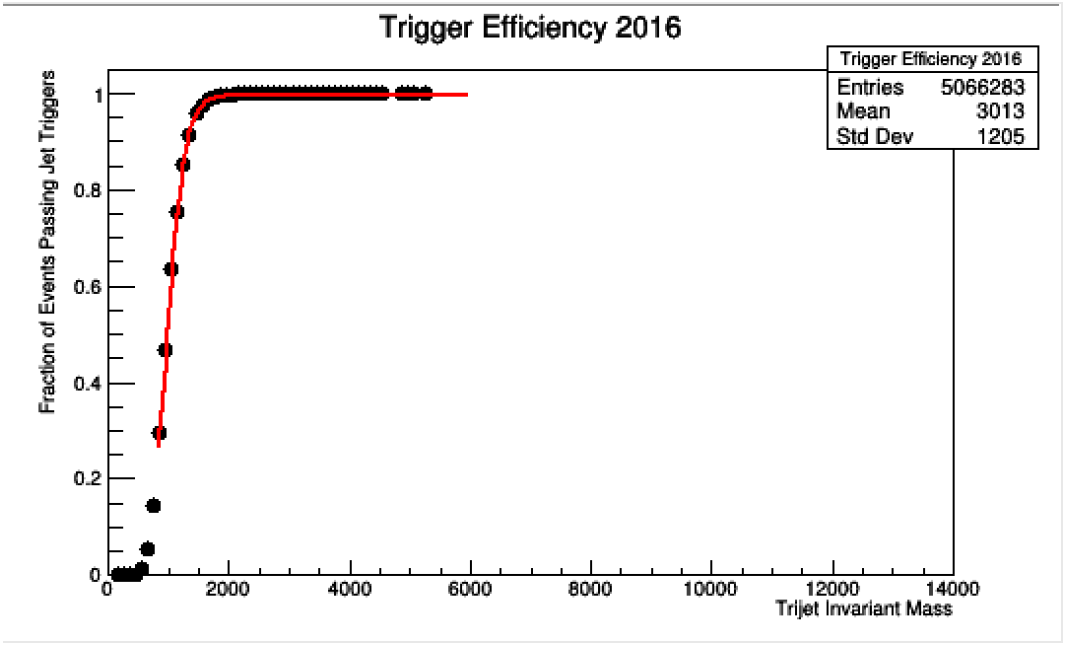



Figure 5: Trigger Efficiency fit for 2016 SingleMuon dataset.

a fit with the background model and a fit with the signal plus background model. We then compare these results to either declare a discovery or put limits on the cross section of the model in question. Each fit is performed on both data and Asimov datasets at varying signal strengths in order to find observed and expected limits. The Asimov dataset is a toy dataset with statistical fluctuation suppressed. Instead of randomly generating a distribution of different datasets from our model and seeing where in the distribution we fall, computational power can be saved by generating a single "typical (Asimov) dataset" and comparing our results to that. The most accurate function for the trijet mass spectrum ended up being the 4 parameter function given below.

$$e^{-(p_2*x^9+p_3*x^{10}+p_4*x^{11})}*x^{-p_1} \tag{4}$$

The results of this fit can be seen in figure 6.

# 8 Limits

With this fit we placed limits on the signal cross section for the decay with intermediate resonance. The results can be seen in figure 7.

We can see from the figure that there is no observed significant excess over the expected background. We have excluded the existence of a particle with a trijet decay signature between 2500 and 4000 GeV to a with cross section for each mass as show in figure 7.

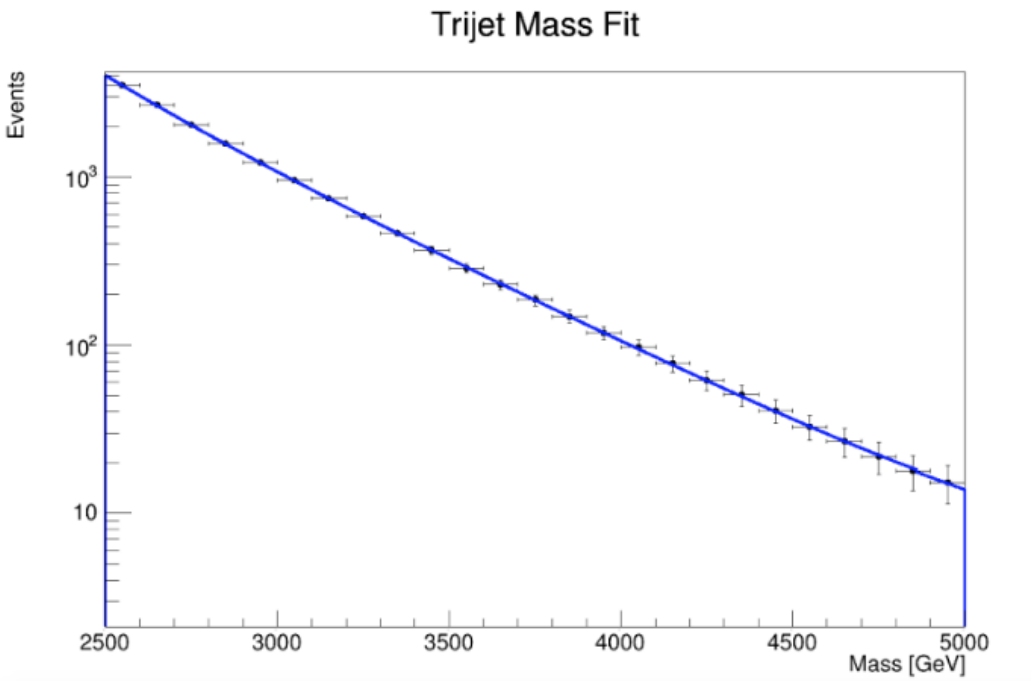



Figure 6: Signal plus background fit to Trijet mass spectrum

# 9 Conclusion

This paper has set limits on resolved trijet decay but there is a lot of phase space still to cover. This study can be extended with a two-dimensional fit in trijet and dijet resonances to search for the intermediate resonance state. This is the first step in a collaboration with other groups in an attempt to map out the possibility of a trijet resonance.

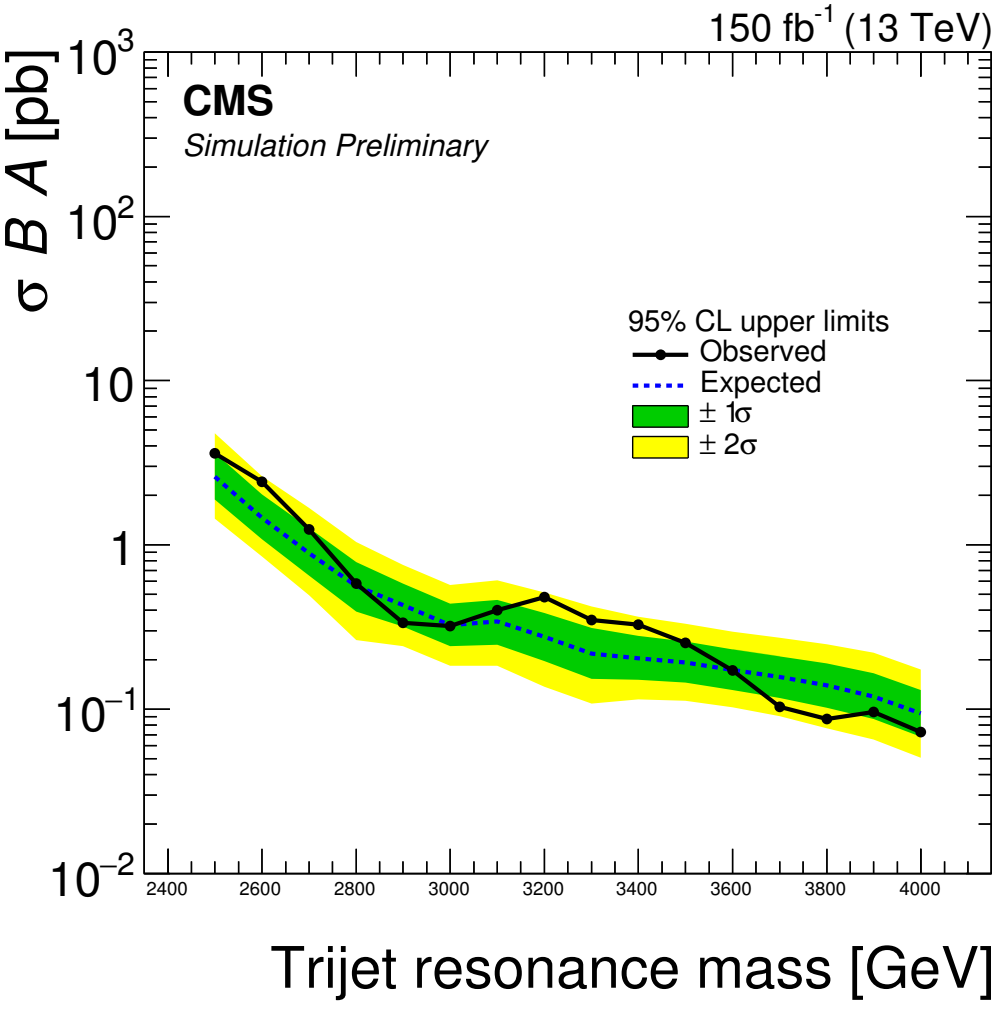



Figure 7: Limits on cross section of particle decaying into 3 jets