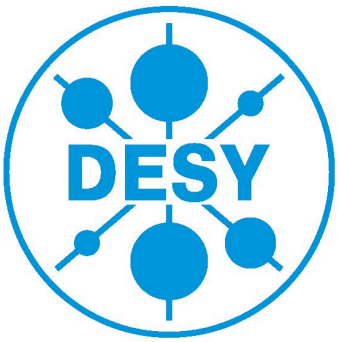


Search for supersymmetric Higgs bosons decaying into a pair of b-quarks

Edward Scott, University of Cambridge

Supervisor: Matthias Schröder

September 12, 2014

Abstract

The analysis procedure in the search for neutral Higgs bosons decaying into a pair of b quarks and produced in association with at least one additional b quark is presented. Events and the response of the CMS detector are simulated at a centre-of-mass energy of 13 TeV, which is expected to be achieved for the first time at the LHC in 2015. The kinematics of the Higgs, partons and jets are surveyed and the effect of varying the selection criteria applied on the samples is examined. The main conclusion is that the p_{T} cut on the third leading jet can be increased from the value of 20 GeV used in the previous analysis at a centre-of-mass energy of 8 TeV, potentially to a value as high as 120 GeV.

Contents

1 Introduction

The Minimal Supersymmetric Standard Model (MSSM) is the simplest viable supersymmetric extension to the Standard Model (SM) of particle physics. Supersymmetry (SUSY) theories introduce a symmetry between fundamental bosons and fermions that solves the so-called hierarchy problem. In the SM the Higgs boson mass has quadratically divergent loop corrections at high energy, which naturally cancel when the additional symmetry is included. Furthermore, SUSY has many other attractive properties, including improved unification of running gauge couplings and potential dark matter candidates.

The MSSM features two scalar Higgs doublets, in contrast to the single doublet that exists in the SM. Consequently the MSSM predicts five physical Higgs bosons. Three (h, H and A) are neutral and collectively denoted Φ. The other two, $H^{\pm}$, are charged. At tree level[1] only two parameters are required to define the MSSM Higgs sector and are normally chosen to be the mass m_A and the ratio of vacuum expectation values $\tan\beta = \nu_1/\nu_2$ [1].

The lightest neutral Higgs, h, can be identified with the Higgs particle discovered at the Large Hadron Collider (LHC) in 2012. The other two are almost mass degenerate and the search for their decay, shown below in Figure 1, is the basis of this project.

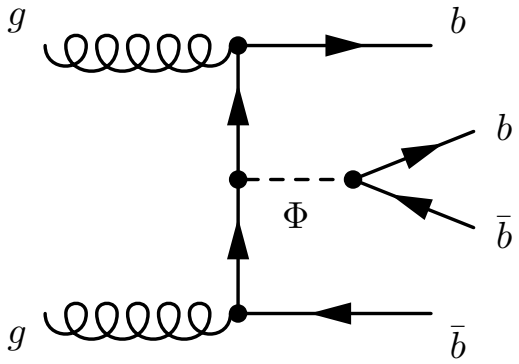


Figure 1: The decay of Φ into $b\bar{b}$. The process is associated with the production of two additional b quarks.

For $\tan\beta > 1$, the expected combined cross section for this Higgs boson production is enhanced by a factor of approximately $2\tan^2\beta$ relative to the SM prediction, and the branching fraction of the decay to a b-quark pair is approximately 90%. It may be possible to observe this process at the LHC and provide evidence for SUSY, or otherwise provide an upper limit on the cross-section times branching ratio for this process (as in [1]), and thereby restrict the possible parameter space of the MSSM.

[1]Tree level means to leading order, i.e. not including higher order loop corrections

In the upcoming LHC run in 2015, the centre-of-mass energy ($\sqrt{s}$) of the colliding protons will be increased to 13 TeV. In the project I worked on the analysis of data in the search for the Φ-to-$b\bar{b}$ process, using simulated events with $m_\phi = 500$ GeV and $\sqrt{s} = 13$ TeV. In particular the focus was on the offline selection criteria applied to preferentially select signal events over background events and how they can be optimised.

This report first describes the methods used in the analysis of Φ-to-$b\bar{b}$ production, and then the preliminary inspection of the signal and background processes. The main objective of the project follows, detailing the selection criteria applied to the simulated events, and how selection efficiency and the ratio of signal events to background events depend on the criteria used. Finally suggestions for further development of this work are presented and conclusions of the project are drawn.

2 The CMS Detector & Simulation of Events

The CMS (Compact Muon Solenoid) detector is one of the two multi-purpose detectors at the LHC, together with ATLAS [2], designed in particular to search for the now-observed Higgs boson [3] and signs of new physics. The main part of the detector is a superconducting solenoid with a diameter of 6 m, which produces a magnetic field of 3.8 T. Further information on the detector can be found in [4] and specific details on the upgrade for the $\sqrt{s} = 13$ TeV in [5].

Two key kinematic quantities measured by the detector that are used frequently in this report are the transverse momentum p_T and the pseudorapidity η. In a cylindrical co-ordinate system with polar angle θ measured with respect to the beam direction and azimuthal angle $\phi = 0$ corresponding to the radial direction, the pseudorapidity is defined as

$$\eta = -\ln\left[\tan\left(\theta/2\right)\right]$$

so that $\eta = 0$ is the transverse direction and $|\eta| \to \infty$ is directed along the beam. The transverse momentum is the component of the particle's four-momentum in the plane transverse to the beam direction. The motivation for using these quantities is described in [6].

For this study, the signal and background samples were produced using PYTHIA [7]. The CMS detector response was modelled using GEANT4 [8].

3 Analysis Strategy

The Φ-to-$b\bar{b}$ process shown in Figure 1 produces four b quarks - two that have come directly from the decay of the Higgs and two others. Free quarks cannot be observed because of confinement in quantum chromodynamics (QCD). Instead colour-neutral bound

states form once the energy stored in the colour field is sufficient to create two new quarks. These newly created quarks repeat the same process and consequently a highly collimated group of hadrons, called a jet, is formed. The jets that result from the hadronisation of b quarks have certain properties, such as a sufficiently long lifetime that the position of production and decay can be distinguished, which allow them to be identified with a certain probability. This identification process is known as b-tagging, and a full description of b-tagging with the CMS experiment is given in [9].

B-tagging is crucial in this search because it greatly reduces the otherwise enormous number of background events that produce jets originating from charm (c) quarks, light-flavour (u, d, s) quarks and gluons (g). The search is performed only on events in which at least three jets are b-tagged ("triple b-tag" sample). The two leading[2] quarks are expected to correspond to the decay products of the Higgs and therefore the invariant mass of these two leading quarks is reconstructed. Evidence for a signal would be a peak in this mass distribution (M_{12}), corresponding to the value of the Higgs mass. This is explored in detail in Section 4.

The primary source of background events in this analysis is heavy flavour multi-jet QCD [1]. Other sources of background events, such as $t\bar{t}$ and $Z + jets$, are small enough that they can be neglected. Calculating the relevant QCD production rates accurately is very difficult and results contain large uncertainties. Therefore when the analysis is done with observed data, as opposed to MC simulated events, an estimation of the background contribution is made directly from the data. This method, which uses a double b-tag sample, is described within the full $\sqrt{s} = 7$ TeV analysis [1]. For this project looking at $\sqrt{s} = 13$ TeV, the QCD background was simulated and its properties are examined in Section 5.

The final part of the analysis strategy involves further selection of events to increase the number of signal events relative to background events. Since the signal and background jets have different distributions in kinematic quantities such as p_{T} and η, cuts can be applied that remove more background jets than signal jets. In Section 6 these cuts and event selection criteria are explained and then optimised.

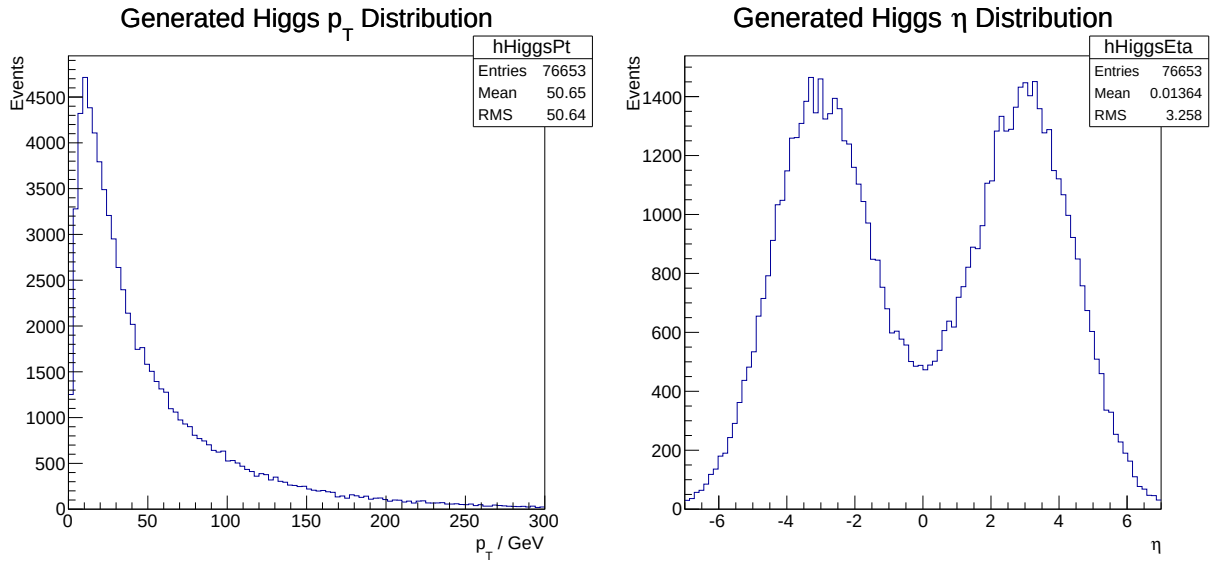
4 Signal Sample

4.1 Higgs, Parton & Jet Kinematics

In this section the p_{T} and η of the Higgs, partons (the daughter b quarks from the Higgs) and the reconstructed jets observed in the detector are plotted. The primary aim is to check that these distributions make sense and to give some context to the system of selection and cuts applied later on in the analysis. The p_{T} and η distributions of the

[2] Jets are ordered by descending p_{T}, which means the leading two jets are the two with the greatest p_{T} values.

generated Higgs boson are shown below in Figures 2a and 2b respectively.



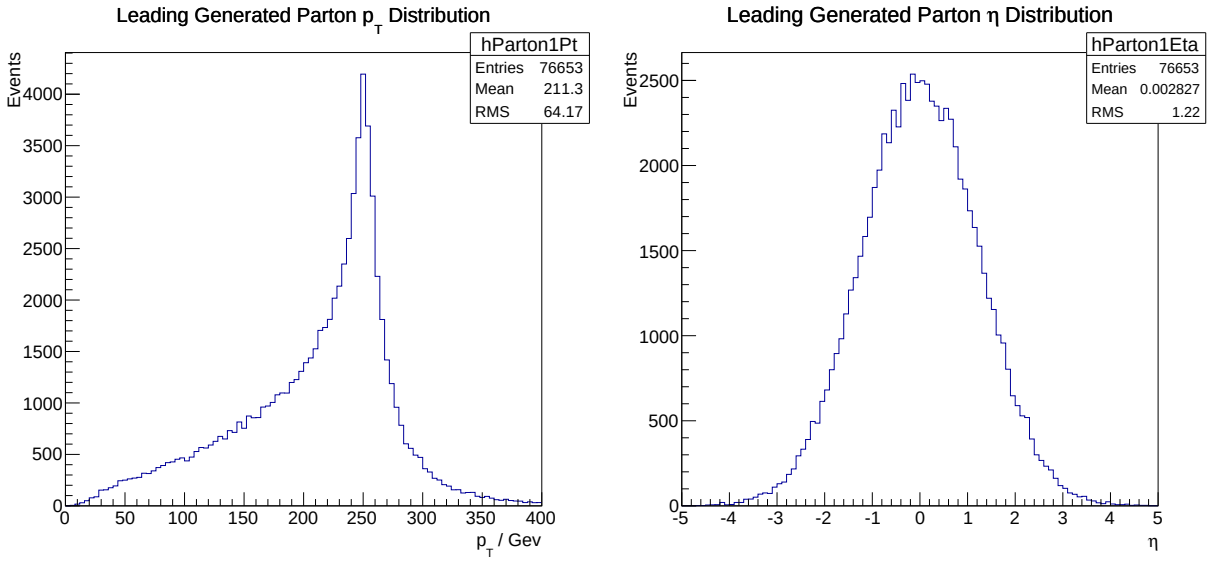
(a) p_{T} distribution

(b) η distribution

Figure 2: The kinematics of the generated Higgs boson.

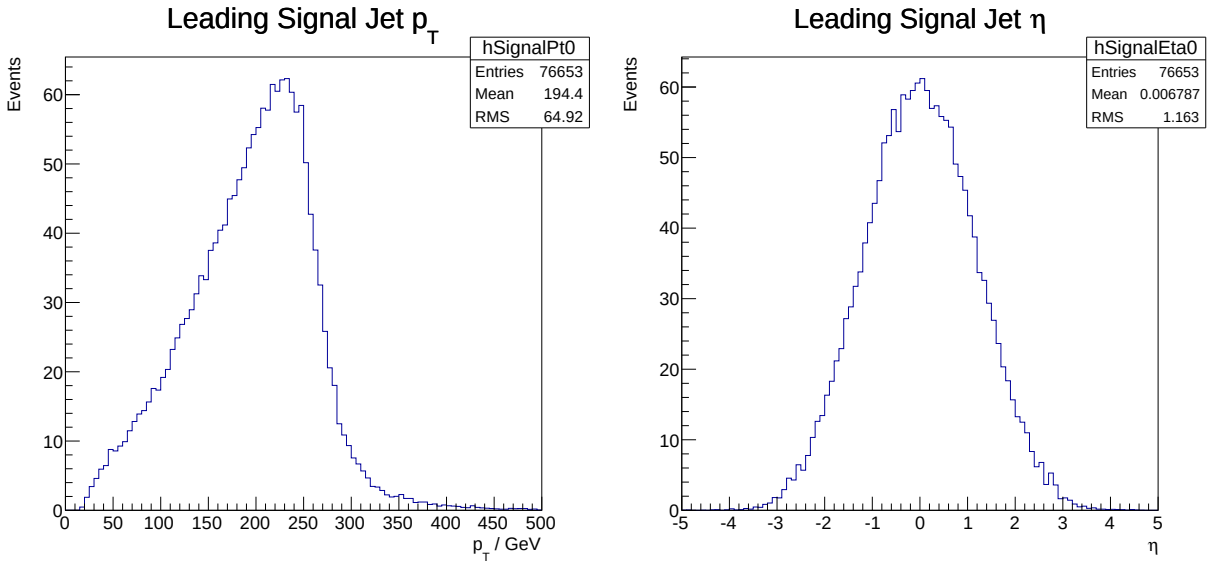
The Higgs is more likely to be produced in the forward direction and has typically has a relatively low p_{T} value. The equivalent distributions for the leading parton are displayed in Figures 3a and 3b.

The partons have a more central η distribution, which is expected since they should be emitted isotropically in the Higgs frame. The peak in the p_{T} distribution, at 250 GeV, also makes sense because it is equal to half the Higgs mass. This value is much greater than the bottom mass or the initial p_{T} of the Higgs so the partons should have a four momentum of magnitude $|p| \approx 250$ GeV. Since the parton η distribution is centred around zero, the most common direction is in the transverse plane, which means that $p_{\mathrm{T}} \approx |p|$, hence the peak at 250 GeV. This high average p_{T} value for the leading parton, which is only slightly lower for the other parton, is why the leading two jets are assumed to be due to the daughter b quarks from the Higgs. This assumption is explored in further detail next, after Figure 3 and Figure 4, showing the same quantities for the leading jet, below.



(a) p_{T} distribution (b) η distribution

Figure 3: The kinematics of the leading generated parton.



(a) p_{T} distribution (b) η distribution

Figure 4: The kinematics of the leading reconstructed jet.

One way to check if a jet matches a certain parton is to use the quantity ΔR, which is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

where $\Delta\eta$ and $\Delta\phi$ are the differences between the parton and jet η and ϕ values. If the jet does indeed come from the parton, the ΔR value between them should be less than approximately 0.3 [6]. Therefore we can look at how often both the leading two jets "match" the partons by imposing the condition $\Delta R < 0.3$, as well as how often the two leading jets are b type. This is useful since in the data we do not have the parton or flavour information, so have no way of estimating how accurate the assumption is. The results are summarised in Table 1 below:

	% Without B-tag	% With B-tag
Both B Type	78.8	95.6
And Both Match Partons	59.5	87.7

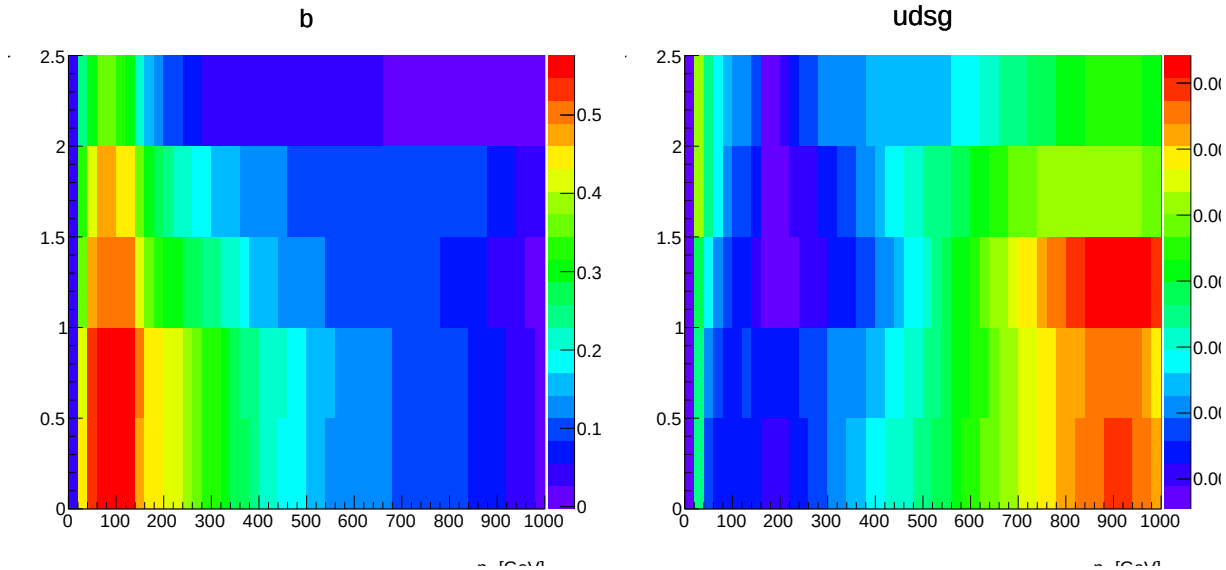
Table 1: Percentage of events in which both leading jets are b type and in which both are b type and both match the partons.

As you can see the b-tagging process is crucial in improving the quality of the signal sample, at least in terms of the validity of the assumption that the leading two jets correspond to the daughter b quarks of the Higgs. A discussion of b-tagging now follows in Section 4.2.

4.2 B-tagging

The number of events occurring at the CMS detector is too large for all of them to be stored permanently. For an event to be stored and its data to become available for analysis, it must pass a set of online trigger requirements. The first implementation of b-tagging in this search occurs at trigger level, where at least two jets must pass the b-tagging requirements online. The signal sample is then subjected to a further offline selection requirement such that all three leading jets must be b-tagged. A double b-tag sample is then used to estimate the shape of the background, as mentioned earlier. In this search the requirement used misidentifies light jets as b jets approximately 0.1% of the time. The details on the algorithms used to perform this process can be found in [9].

The b-tagging process was not fully functional in the simulated events used in this study, so an effective b-tagging process was implemented by weighting events with the probability of being b-tagged. These probabilities were read from b-tag efficiency maps obtained from $\sqrt{s} = 8$ TeV data, two of which are shown below in Figure 5.



(a) b type jet b-tag probability

(b) light flavour jet b-tag probability

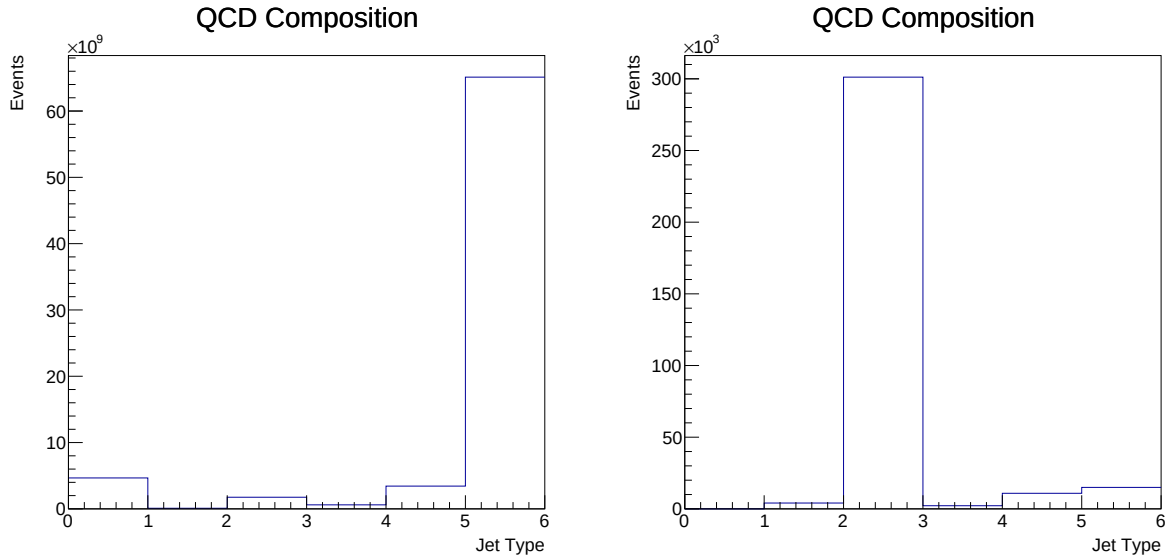
Figure 5: B-tag efficiency maps for b and light flavour jets, as a function of p_{T} and η.

These maps contain the probability that a jet of a given type will be b-tagged, as a function of p_{T} and η values. They only contain values for $|\eta| < 2.5$ and 20 GeV $< p_{\mathrm{T}} <$ 1000 GeV so these cuts are applied to all jets in this study out of necessity. Further cuts to preferentially select signal events are imposed later on and discussed in Section 6.

5 Background Sample

5.1 Composition

In the raw data, without any b-tagging or selection, QCD events occur over 10^7 times more frequently than signal events. The majority of these are light flavour jets, due to one of a gluon, up, down, or strange quark. Figure 6 shows the type of jet in the QCD background both with and without b-tagging. It is clear from these histograms that the b-tagging process is very effective, greatly reducing the number of background events, which are initially predominately light flavour jets. However despite the fact that the majority of QCD events remaining after b-tagging are b jets, they have very different kinematic properties from the leading signal jets, which will be illustrated in Section 5.2.



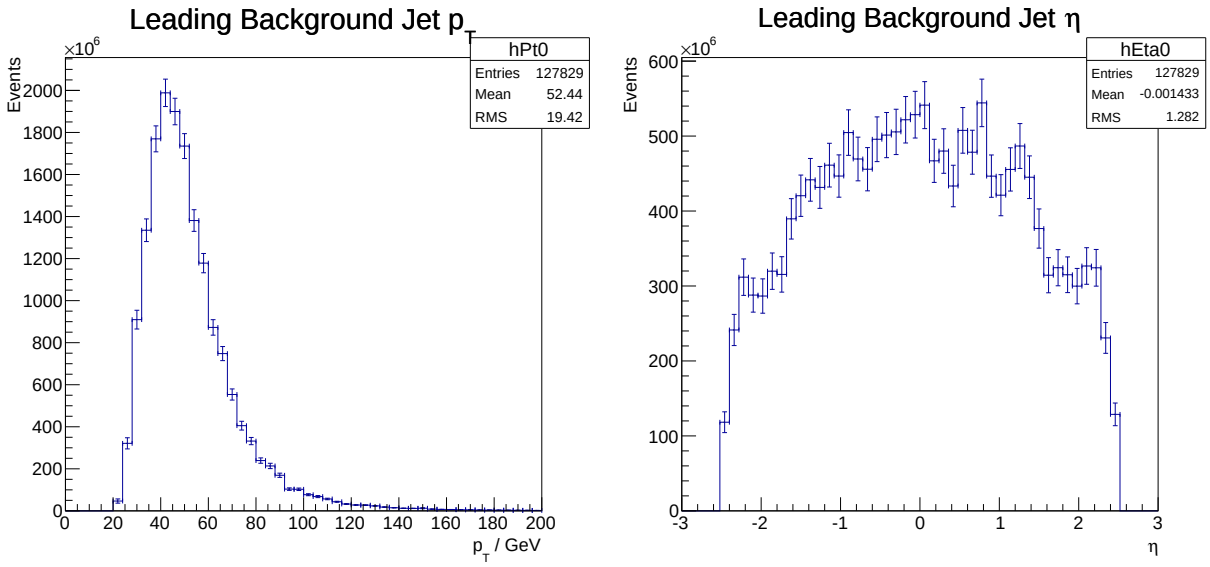
(a) Without b-tagging

(b) With b-tagging

Figure 6: The composition of the QCD background. Unidentified jets appear in the first bin, b jets are in the 2-3 bin and light flavour jets in the 5-6 bin.

5.2 Kinematics & Mass Distribution

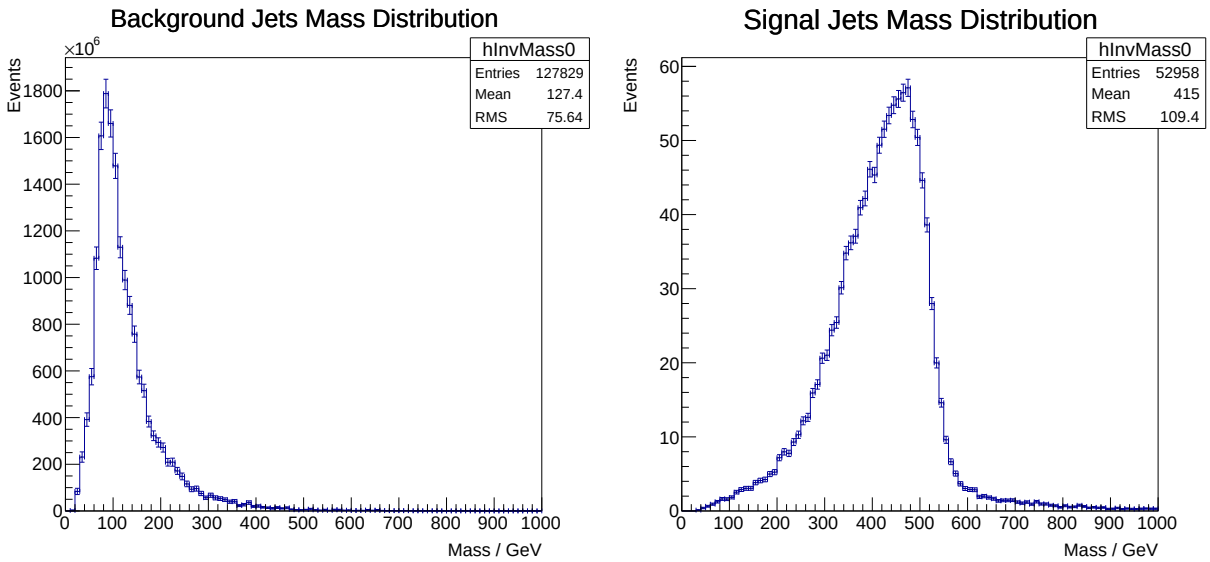
Figure 7 on the following page shows the p_{T} and η distributions for the leading background jet. The plots are not actually the true distributions of the b-tagged background sample. This is because the background statistics are limited so the plot with b-tagging weights applied is very noisy and does not clearly show the distribution. With the b-tagging weights applied, the distribution is very similar but has a slightly higher mean of 68 GeV. In any case it is clear that the average p_{T} is much lower for the background events.



(a) p_{T} distribution (b) η distribution

Figure 7: The kinematics of the leading background jet.

The invariant mass distribution of the leading two background jets is naturally also very different from that of the expected signal. Figure 8 shows the two distributions, once again without the b-tagging weights applied.



(a) Background (b) Signal

Figure 8: The mass distributions of the leading two jets.

With b-tagging, the mean of the background distribution is shifted up to 162 GeV. The aim of this search is to observe an excess in the mass distribution at the value of the Higgs mass. The main aim of my project was to investigate the selection criteria used to increase the number if signal events relative to background events in order to maximise the sensitivity of the search. These criteria are the subject of the next part of the report, Section 6.

6 Selection Optimisation

6.1 Overview

As well as the online trigger selection of events for this search, several further cuts are applied offline. The aim is to increase the ratio of signal events to background events to increase the sensitivity to the process. Two quantities are discussed in this section: the ratio of the number of selected signal events to the number of selected background events, denoted S/B, and the ratio $S/\sqrt{S+B}$. This second ratio is used because it also depends on the total number of events selected - the simple S/B ratio can take a very high value but only involve very few events and therefore not be very useful.

In this project I used the previous cuts used in the 8 TeV analysis as a reference point and then initially examined the effect of varying one at a time. Table 2 lists the cuts applied previously:

Variable	Selection Criterion
$p_{\mathrm{T}}^{1/2/3}$	$> 80/70/20$ GeV
$\lvert\eta^{1/2/3}\rvert$	$< 1.65/1.65/2.2$
$\Delta\eta^{1,2}$	< 1.4
$\Delta R^{1,2}$	> 1
$\Delta R^{1,3}$	> 1
$\Delta R^{2,3}$	> 1

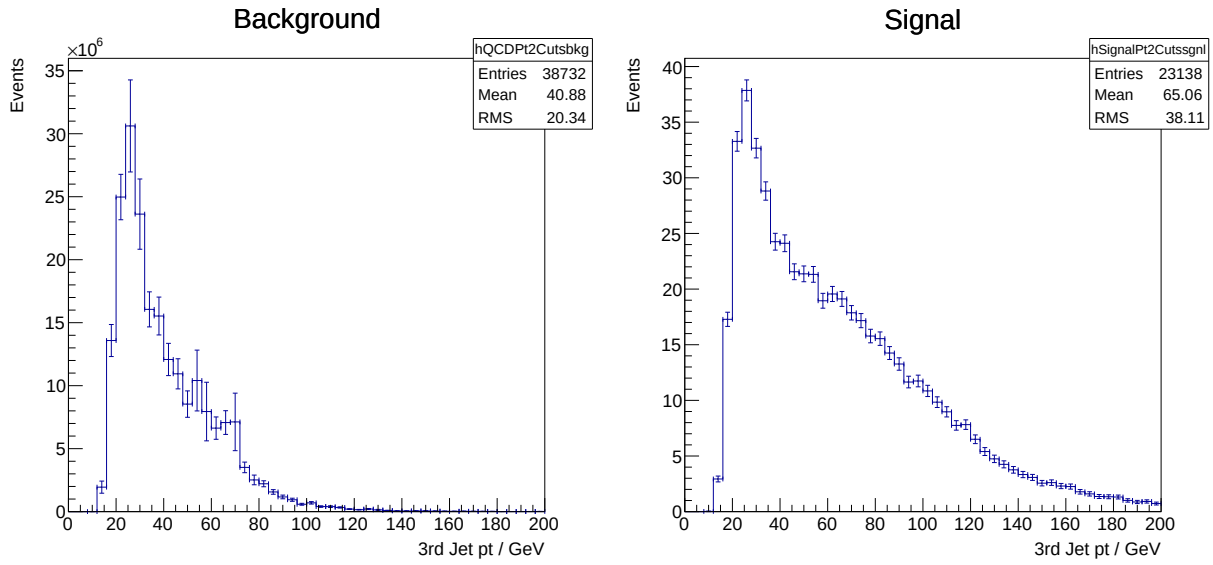
Table 2: The cuts used in the search with $\sqrt{s}$=8 TeV data.

The motivation for these cuts is as follows. The average p_{T} of the signal sample jets is greater than in the background, so selecting only jets above a certain p_{T} is a good idea. The η distribution of the background is flatter than the signal, hence the exclusion of events with η greater than some value. The $\Delta\eta$ cut is mainly used to reduce the rate of data meeting the trigger conditions, and the ΔR cuts ensuring the jets are well separated reduce the number of jets due to gluon splitting that are selected.

In this study I varied the third p_{T}, third η and the $\Delta\eta$ cuts, which are now discussed in more detail.

6.2 Third p_{T} Cut

The third jet p_{T} distributions shown in Figure 9 below illustrate the reason for imposing a cut:



(a) Background

(b) Signal

Figure 9: The third p_{T} distributions of signal and background events.

In the previous analysis a cut of 20 GeV was used, which seems rather low at first glance. Tables 3 and 4 show the selection efficiencies for a range of cut values for the background and signal events respectively. The values are normalised to an integrated luminosity of 1 fb^{-1}. The signal cross-section, calculated using FeynHiggs 2.10.2 with $\tan\beta$ (see feynhiggs.de) is 1.8 pb so there are 1800 expected signal events, from 76653 MC events. The background cross-sections are taken from the PYTHIA simulation and predict 7.6×10^{10} events from 226924 MC events.

Cut Value / GeV	Selected Events	Selection Efficiency Without B-Tagging	Total Selection Efficiency
20	17810	0.21 %	2.4×10^{-5} %
60	2829	0.04 %	3.7×10^{-6} %
100	146	4.2×10^{-3} %	1.9×10^{-7} %
140	2.5	6.3×10^{-4} %	3.2×10^{-9} %

Table 3: Background Selection Efficiencies

Cut Value / GeV	Selected Events	Selection Efficiency Without B-Tagging	Total Selection Efficiency
20	16.9	27.24 %	0.94 %
60	10.5	14.32 %	0.58 %
100	4.4	5.71 %	0.24 %
140	1.3	1.84 %	0.07 %

Table 4: Signal Selection Efficiencies

The expected result that the selection efficiency drops faster in the background sample than the signal sample is seen. This has important consequences for the S/B and $S/\sqrt{S+B}$ ratios, which are shown in Table 5.

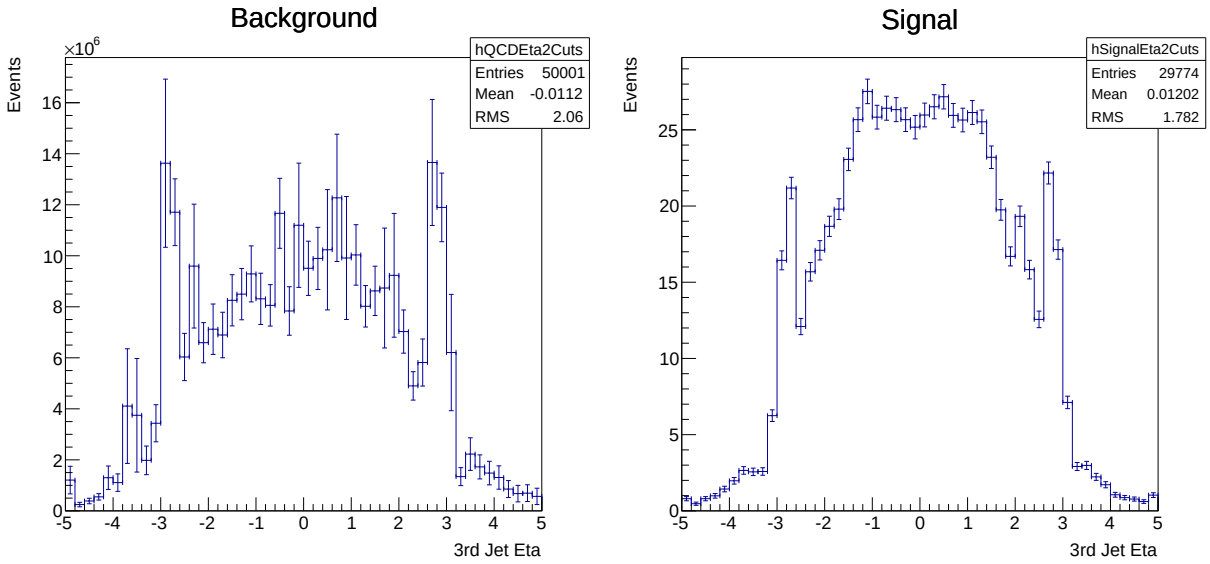
Cut Value / GeV	S/B	$\Delta S/B$	$S/\sqrt{S+B}$	$\Delta S/\sqrt{S+B}$
20	9.5×10^{-4}	4.1×10^{-4}	0.13	0.03
60	3.7×10^{-3}	1.8×10^{-3}	0.20	0.05
100	0.030	0.023	0.36	0.13
140	0.53	0.18	0.67	0.08

Table 5: The values and errors of S/B and $S/\sqrt{S+B}$.

There is a substantial increase in the value of both ratios when the cut value is increased. The errors calculated are only due to the uncertainty in the "data" provided by the simulation, not including for example the error in cross-section. Evidently the sensitivity of the search would be improved if the third p_{T} cut was increased.

6.3 Third η Cut

Figure 10 shows the third jet η distributions, in which the background sample is somewhat flatter than the signal.



(a) Background (b) Signal

Figure 10: The third η distributions of signal and background events.

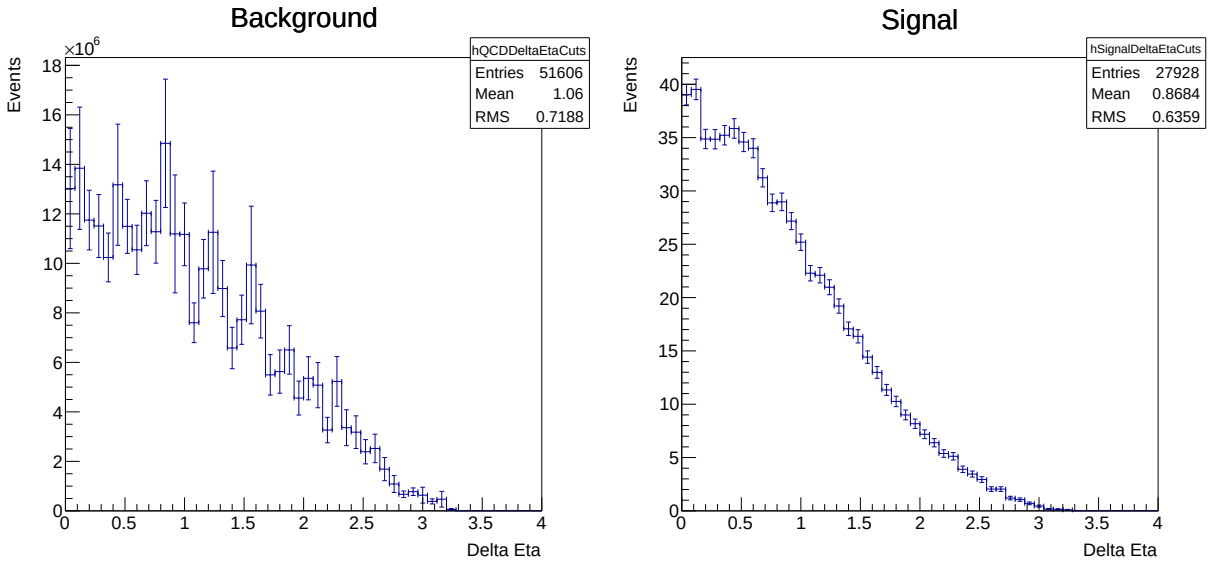
However in this case there is not any improvement in S/B or $S/\sqrt{S+B}$ as the cut value is altered, as shown in Table 6. Consequently there is no reason to move the value of the third η cut.

Cut Value	S/B	$\Delta S/B$	$S/\sqrt{S+B}$	$\Delta S/\sqrt{S+B}$
0.7	6.6×10^{-4}	4.4×10^{-4}	0.062	0.021
1.2	9.9×10^{-4}	5.7×10^{-4}	0.10	0.03
1.7	8.6×10^{-4}	4.0×10^{-4}	0.11	0.03
2.2	9.5×10^{-4}	4.1×10^{-4}	0.13	0.03

Table 6: The values and errors of S/B and $S/\sqrt{S+B}$.

6.4 $\Delta\eta$ Cut

The situation is very similar for the $\Delta\eta$ cut - the signal and background distributions are almost the same shape (Figure 11) and the values of the two ratios do not increase when the cut position is moved (Table 7).



(a) Background (b) Signal

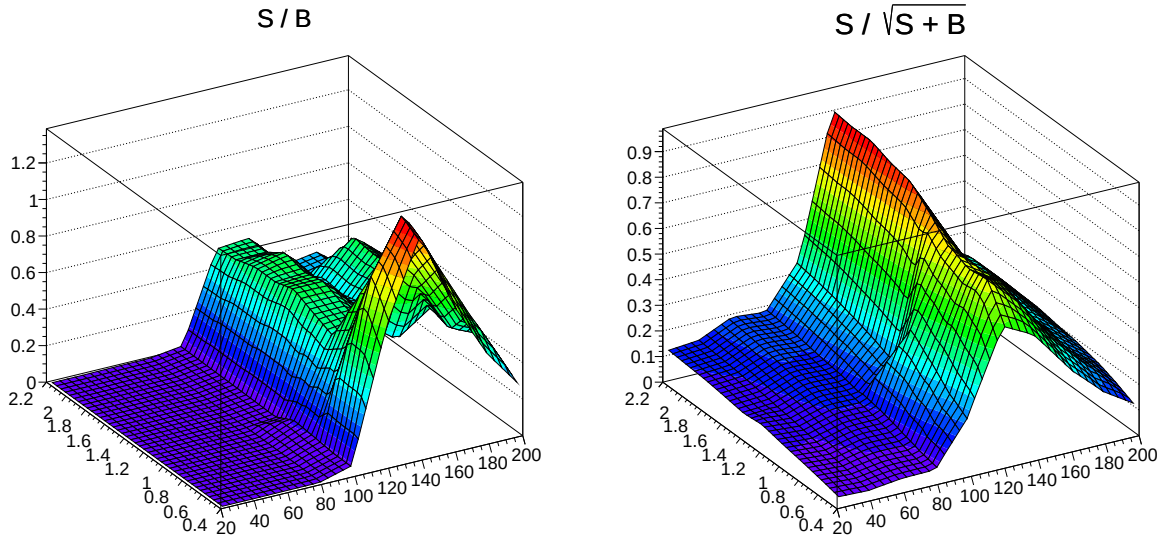
Figure 11: The $\Delta\eta$ distributions of signal and background events.

Cut Value	S/B	$\Delta S/B$	$S/\sqrt{S+B}$	$\Delta S/\sqrt{S+B}$
0.4	8.0×10^{-4}	5.6×10^{-4}	0.07	0.02
0.9	1.5×10^{-3}	0.9×10^{-3}	0.13	0.04
1.4	9.5×10^{-4}	4.1×10^{-4}	0.13	0.04
1.9	1.1×10^{-3}	0.4×10^{-3}	0.14	0.03

Table 7: The values and errors of S/B and $S/\sqrt{S+B}$.

6.5 Final 2D Scan

The final computation performed as part of this project was a two dimensional version of the previous scans, in which both the third p_{T} and third η cuts were allowed to vary. The greatest value of $S/\sqrt{S+B}$ was obtained when the values were 120 GeV and 2.2 respectively, further supporting the conclusion that the third p_{T} cut should be increased. The plots of both ratios are shown in Figure 12.



(a) S/B (b) $S/\sqrt{S+B}$

Figure 12: The ratios S/B and $S/\sqrt{S+B}$ as a function of the third p_{T} and η cuts.

7 Conclusion

The principal aim of this project was to optimise jet selection varying the p_{T} and η cuts applied to the third jet and the $\Delta\eta$ cut between the leading two jets and examining the effects on the two ratios S/B and $S/\sqrt{S+B}$. It was found that altering the third η and $\Delta\eta$ cuts did not increase either value, whilst increasing the third p_{T} cut from the value of 20 GeV previously used in the equivalent analysis at $\sqrt{s} = 8$ TeV improved both values. Specifically, the ratio $S/\sqrt{S+B}$ increased from to , clearly indicating that increasing the third p_{T} cut value will increase the sensitivity of this search.

The next step in this investigation would be to repeat the selection optimisation process with a few changes. A greater number of MC events, especially in the background sample, would reduce the uncertainty in the ratio values and allow a more precise calculation of the optimal point at which to cut the jet. Using samples in which the b-tagging was fully implemented in the simulation would also be preferable to using the efficiency maps taken from the 8 TeV data. Finally the effect of varying other cuts, particularly on the p_{T} of the leading two jets (since increasing the third cut beyond 70 GeV effectively increases them too), could be assessed. This would then help inform the choice of trigger and the analysis techniques used once the $\sqrt{s} = 13$ data recorded by the CMS detector becomes available.

Acknowledgements

I would like to thank my supervisor Matthias Schröder for all his help, enthusiasm and incredible knowledge of the subject that has made my summer here so enjoyable. I also greatly appreciated the guidance and encouragement from Rainer Mankel, Alexei Raspereza and Gregor Mittag and the company of Benoît Roland in the office.

I have had a fantastic time here at DESY and hope to be back one day!

References

[1] Search for a Higgs boson produced in association with b quarks and decaying into a b quark pair *CMS Collaboration*

[2] The ATLAS experiment at the CERN LHC *ATLAS Collaboration*

[3] Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC *CMS Collaboration*

[4] The CMS experiment at the CERN LHC *CMS Collaboration*

[5] Projected Performance of an Upgraded CMS Detector at the LHC and HL-LHC *CMS Collaboration*

[6] Quality of Jet Measurements and Impact on a Search for New Physics at CMS *Matthias Schröder*

[7] PYTHIA 6.4 Physics and Manual arXiv:hep-ph/0603175 *T. Sjostrand,S.Mrenna,and P.Z.Skands*

[8] GEANT4: A Simulation toolkit *GEANT4 Collaboration*

[9] Identification of b-quark jets with the CMS experiment *CMS Collaboration*

[10] study *name*