CDF/PHYS/EXOTIC/CDFR/8061
Version 4.0
July 23, 2007

Search for Higgs Bosons Produced in Association with b-Quarks

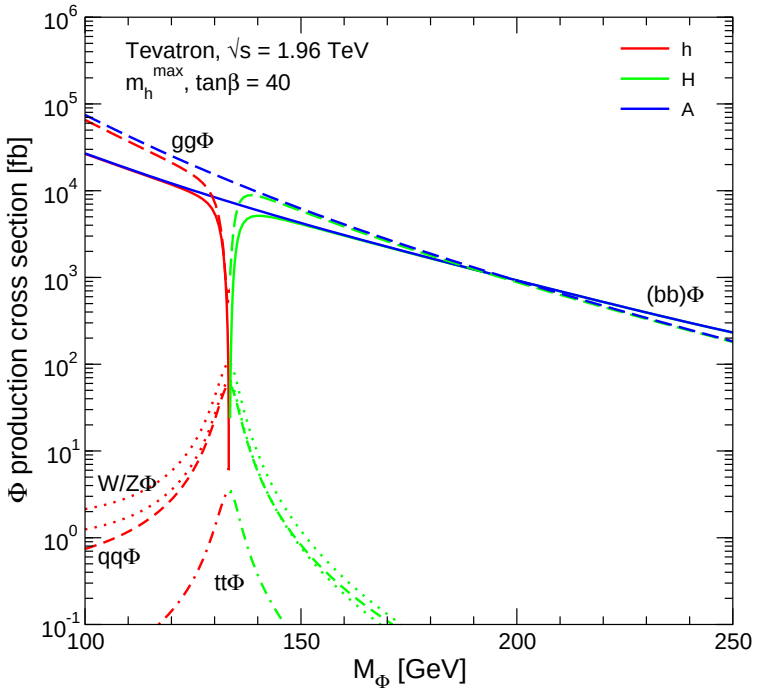
T. Wright, D. Amidei
University of Michigan

Abstract

We present a search for Higgs bosons produced in association with b-quarks in $p\bar{p}$ collisions. This process could be observable in supersymmetric models with high values of $\tan\beta$. We search for an enhancment in the mass of the two lead jets in triply b-tagged events, using 980 pb^{-1} of data from the HIGH_PT_BJET and HIGGS_HIGH_PT_BJET datasets. The dijet mass spectrum of the heavy flavor multi-jet background is derived from the double-tag data in a manner that accounts for tagging biases and kinematic differences introduced by the addition of the third tag. The procedure is verified using Monte Carlo and also negative tag control samples in the data. We set mass dependent limits on $\sigma \times BR$ and $\tan\beta$ in MSSM models.

Contents

Figure 1: MSSM Higgs cross sections for various production modes at $\tan\beta = 40$ in the m_h^{max} scenario, from the TeV4LHC Working Group [2].



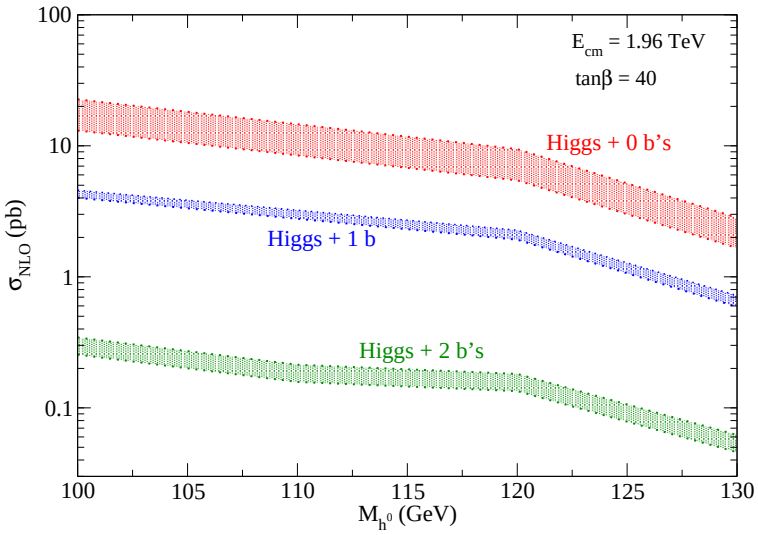
1 Introduction

The production rate of light Higgs bosons in association with b-quarks can be significantly enhanced in supersymmetric extensions of the standard model. This occurs for large values of $\tan\beta$, the ratio of the Higgs coupling to down-type versus up-type quarks. Figure 1 shows the cross section expected for $\tan\beta = 40$ in the m_h^{max} benchmark scenario [1], from the TeV4LHC Working Group [2]. The cross section for $(bb)\Phi$ is in the 10 pb range, which could potentially be observable with existing data samples. Also interesting is that at large $\tan\beta$ the pseudoscalar A becomes degenerate with either the light (h) or heavy (H) scalar, giving an effective factor of two enhancement to the cross section.

The cross sections shown in Figure 1 are for inclusive production [3], however only the case where at least of the b's accompanying the Higgs is at high p_T is relevant to these results, since we will require that it be b-tagged. Fortunately, as shown in Figure 2, cross section calculations are available for this case as well [4, 5], allowing for the interpretation of the results of the search described in this note.

Results for the Higgs+1b process in the case of Higgs decays to $b\bar{b}$ have been obtained by DØ [6, 7], and for inclusive Higgs production in the $\tau\tau$ decay mode by CDF [8, 9] and DØ [10, 11, 12].

Figure 2: MSSM Higgs cross sections at $\tan\beta = 40$ as a function of the number of high-p_T b quarks accompanying the Higgs (taken from Ref. [5]).



1.1 Analysis Plan

We study the mass spectrum of the two leading jets in three-jet events with triple tight SECVTX tags. We fit the dijet mass spectrum m_{12} to a background-only and background-plus-signal hypothesis and use a $1 - CL_S$ comparison to derive 95% C.L. upper limits on $\sigma(H^0 + b) \times BR(H^0 \to b\bar{b})$.

The sample is drawn from the HIGH_PT_BJET and HIGGS_HIGH_PT_BJET datasets, triggered on at least two SVT tagged jets with $E_T > 15$ GeV and, in the HIGGS path, a third untagged cluster with $E_T > 10$ GeV at Level2. All jets must have tight SECVTX tags, and the two leading jets must overlap with the L2 clusters and SVT tracks. Monte Carlo models are corrected for jet energy scales and other effects so as to provide accurate acceptance determinations for background and signal.

The three-tag sample is a mixture of multi-parton heavy flavor processes. The PYTHIA Monte Carlo predicts that the triple tags are overwhelmingly events with at least two real b-tags, with the additional tag being any of a mistagged light jet, a c-tag, or another b-tag. The double-tagged three jet events are also predominantly two real b-tags, which makes them a natural starting point for constructing background estimates. We use simulation to show that the dijet mass spectrum can be parameterized in terms of components that depend on the flavor mix, bbb, bbc, etc. Dijet mass templates for all five distributions can be estimated from double tags in the data (in some cases after applying small reshapings derived from Monte Carlo studies of flavor correlations). In addition to the dijet mass we define a new variable related to the secondary vertex masses in the three tags $m_{diff} = m_{SVX}(j1) + m_{SVX}(j2) - m_{SVX}(j3)$, which carries information on the flavor of the jets in the events. We show that a simultaneous fit to m_{12} and m_{diff} fixes the relative ratio of the five background templates and is also sensitive to the presence of signal events.

The performance of the method is studied in a data control sample derived from events with two positive and one negative tag. The dijet mass and vertex mass difference are simultaneously well fit by the superposition of the background templates with no residual Higgs signal. The largest sources of systematic uncertainty are the jet energy scale and the corrections applied to account for kinematic differences between the double-tagged and triple-tagged background events.

We use the MCLIMIT program [13] to set 95% C.L. upper limits on $\sigma(H_0 + b) \times BR(H_0 \to b\bar{b})$ as a function of M_H in the three-jet triple tagged sample. Efficiencies and scale factors are treated as nuisance parameters with gaussian constraints, and shape systematics are propagated into sampled distributions in large pseudoexperiments. The median expected limit asymptotically approaches 8 pb at $M_H = 200$ GeV/c^2. The cross-section limit can be reinterpreted as a limit on $\tan\beta$ vs M_H.

2 Data and Monte Carlo Samples

All of the data and Monte Carlo samples used were processed using CDF software release 6.1.4mc. Tracks were refitted with L00 hits added, and the refitted tracks were used to construct a new `ZVertexColl`. Jet reclustering was performed around the `ZVertex` with the highest summed p_T, using `JetCluModule` with cone size 0.7. Jet corrections were done using the `jetCorr06b` release. All jet E_T values are given with the L5 correction applied unless otherwise noted. For the Higgs Monte Carlo samples, the additional jet energy scale factor of 0.974 indicated by the $Z \to b\bar{b}$ results [14] was also applied to get the best possible match to the data.

The data used for these results came from the `ebjt0d`, `ebjt0h`, and `ebjt0i` datasets which are based on the HIGH_PT_BJET and HIGGS_HIGH_PT_BJET trigger paths. The DQM group version 13 ntuple was used to generate a good runs list. Starting from the "QCD with silicon" requirements, the online SVT field was required good, and the offline SVT field was required not to be marked bad (it is not always filled). The resulting list represents 980 pb^{-1} of integrated luminosity, including the correction factor of 1.019. It should be noted that this trigger path does not appear in the "HIGHLUM" physics tables. This was accounted for by removing those runs from the good runs list when computing the offline luminosity from the DHInput log files.

Several Monte Carlo samples were used and are listed in Table 1. The `t2t*` samples were generated specifically for this analysis and are stored in the DFC book `twright`. They consist of PYTHIA process MSEL=32, which is bg scattering into bH, where the incoming b quark comes out of an initial state shower. The Higgs boson in this process is the Standard Model one with very small width. A procedure for simulating the non-negligible width acquired by the Higgs at large values of $\tan\beta$ will be described in Section 9.2. The showering and underlying event setup was copied from the standard CDF Monte Carlo sample TCL files.

The dijet Monte Carlo samples were used only to study various effects and to validate the analysis procedure. They were not used to estimate backgrounds, which are derived from the data itself.

Table 1: Monte Carlo samples used for these results.

Dataset	Events	Notes
t2t100	230k	PYTHIA $bg \to bH$, $m_H = 90$ GeV/c^2
t2t140	1.15M	PYTHIA $bg \to bH$, $m_H = 90$ GeV/c^2
t2t110	230k	PYTHIA $bg \to bH$, $m_H = 120$ GeV/c^2
t2t150	460k	PYTHIA $bg \to bH$, $m_H = 120$ GeV/c^2
t2t120	230k	PYTHIA $bg \to bH$, $m_H = 150$ GeV/c^2
t2t160	230k	PYTHIA $bg \to bH$, $m_H = 150$ GeV/c^2
N/A	370k	PYTHIA $bg \to bH$, $m_H = 180$ GeV/c^2
N/A	230k	PYTHIA $bg \to bH$, $m_H = 210$ GeV/c^2
jqcd1f	5M	PYTHIA dijet $p_T > 18$ GeV/c
jqcdci	1M	PYTHIA dijet $p_T > 18$ GeV/c, b quark filter
jqcdmi	1M	PYTHIA dijet $p_T > 18$ GeV/c, c quark filter
jqcd2f	5M	PYTHIA dijet $p_T > 40$ GeV/c
jqcddi	1M	PYTHIA dijet $p_T > 40$ GeV/c, b quark filter
jqcdni	1M	PYTHIA dijet $p_T > 40$ GeV/c, c quark filter
jqcdgg	1M	PYTHIA dijet $p_T > 60$ GeV/c
jqcdei	500k	PYTHIA dijet $p_T > 60$ GeV/c, b quark filter
jqcdoi	500k	PYTHIA dijet $p_T > 60$ GeV/c, c quark filter

All of the Monte Carlo samples were simulated with release 5.3.3, using the "Summer 2004" runs list which extends to run 179056, except for the c- and b-filtered dijet samples which used a run list that extends up to 186598. The problem of the default SVT simulation using incorrect beamlines, discussed in Ref. [15], is present in these samples, so the fix procedure described in that note was applied and the handful of unfixable runs listed there were dropped.

3 Event Selection

The base selection is three cone-0.7 jets with L5 $E_T > 20$ GeV and $|\eta| < 2$ which are tagged with the loose SecVtx algorithm. In order to simplify the background calculations, only the three highest-E_T jets in the event matching a L2 cluster with $E_T > 10$ GeV and $|\eta| < 1.8$ are considered (all three must pass the jet cuts and be SECVTX-tagged). The two leading jets must also pass the trigger b-tag selection defined in Ref. [15] (matched L2 cluster with $E_T > 15$ GeV, $|\eta| < 1.5$, matched L3 jet with $E_T > 20$ GeV, and matched SVT and L3 silicon tracks with $|d_0| > 120$ μm), and be separated in ϕ by at least 1.5 radians. The selection is designed such that any event satisfying these requirements should have passed any version of HIGH_PT_BJET or HIGGS_HIGH_PT_BJET, so that no additional trigger efficiency or turn-on corrections are necessary.

3.1 Event Types

Throughout this note a shorthand notation for different types of events will be used. The notation takes the form `XXY`, where `XX` is the tag or flavor combination of the two leading jets and `Y` refers to the third jet. In all cases, the two lead jets must match all of the trigger objects (L2 cluster, L3 jet, SVT and L3 silicon tracks), and there is no distinction between the first and second leading jet. For inclusive observed tags, the notation is `+` for a positive SECVTX tag of any flavor. The letters b, c, and q are used to refer to bottom, charm, and light-flavor tags, while the letter j refers to a jet with no tag requirement.

In the study of Monte Carlo events, the jet flavors signifiers are used to refer to both the jet flavor and tag, so that for example *bcb* would mean an event with a b-jet and c-jet as the two lead jets, and a b-jet as the third jet, all tagged. When discussing background templates derived from the double-tagged samples, the flavor signifier refers only to the tag properties, so that *bbc* in this case would mean that the two lead jets are both b-tags (assumed for double tags), and the event is weighted by treating the third jet as if it is a charm tag (even though the majority of third jets in the double-tagged sample are actually light jets). Although there is the potential for confusion we will try to make it clear from the context just what is being described in each case where this notation appears.

4 Backgrounds

This search is performed against a large background of QCD heavy flavor multijet production, so that our sample contains all of the three-way combinations of true b-tags, true c-tags and mistagged light flavor jets, in ratios controlled by known tagging efficiencies and poorly known cross-sections. Other physics backgrounds such as $Zb, Z \rightarrow b\bar{b}$ also contribute but their cross sections are so small that they can be neglected.

The challenge for our method is to understand and model the m_{12} distribution for the triple-tagged multijets. As our Monte Carlo simulations are not sufficiently calibrated to provide an absolute prediction, we seek a data based model. The natural candidate is the double-tagged data events. These will automatically capture the subtleties of jet energy scales, trigger and tag biases, but we will need to establish that the double tags, which are predominantly two b-tags and a light jet, are a good model of triple-heavy-flavor processes like *bbb*. This can be studied in Monte Carlo simulation, comparing the m_{12} and m_{diff} distributions across the different multi-flavor components and the simulated double tag sample. We will show here how the m_{12} spectrum can be decomposed into five primary flavor combination templates: *bbb*, *bbc*, *bbq*, *bcb*, and *bqb*, that these templates can be constructed from double-tag events in the data using modest flavor-process based corrections derived from the simulation studies, and that a combined fit including m_{diff} can determine the relative normalizations of the five components necessary to recreate the m_{12} spectrum.

4.1 Monte Carlo Simulation Sample

We study the QCD multijet background using the 18, 40, and 60 GeV PYTHIA dijet Monte Carlo samples (including b- and c-filtered ones) listed in Table 1. We used the events from

each sample weighted according to the generated cross sections. The scattering frame momentum transfer $\hat{p}_T$ was reconstructed for each event and used to normalize the overlap between the samples.

Even though the MC samples are quite large, the selection requirements are such that very few events survive ($\sim$100). Therefore, instead of relying on the actual tagged events, we used a weighting procedure that captures the full statistics of the Monte Carlo.

The base requirement was three taggable jets with $E_T > 20$ GeV and with the two leading jets matching all of the trigger objects, but with no impact parameter cut on the SVT or L3 silicon tracks. We then use the sample to derive tagging probabilities for each kind of jet. The flavor of each jet was determined by searching for heavy flavor hadrons within the jet cone. Based on this, we derive our own "b/c/mistag-matrix" of the expected tag probabilities for each kind of jet, parameterized in terms of jet flavor, number of good SecVtx tracks in the jet, jet E_T, and assumed SecVtx tag mass.

Using these tag probabilities, we can calculate a net tagging probability for the different three-way tags of each event. For each jet, we check the flavor inside the cone, and assign the appropriate tag probability weight. The tag rate parametrizations were derived from all jets in all of the QCD dijet samples in Table 1. The combined weights for an event to be triple-tagged and to have a given value of m_{diff} were computed and used along with the cross section weights when histogramming event quantities such as m_{12}.

Table 2: Estimated composition of the QCD background in the triple-tagged event sample.

flavor	fraction
bbb	0.450
bbc	0.096
bbq	0.097
bcb	0.164
bqb	0.132
other	0.061

4.2 Multi-Flavor Components of the Triple Tagged Sample

As a first step we investigate what types of events pass the triple-tag event selection. The results are shown in Table 2 and indicate that almost all of the events contain two real b-tags, with the remaining tag also a b-tag 45% of the time. The study shows that we need to model only five components (really four because $bb + c$ and $bb + q$ are quite similar) in order to account for essentially all of the QCD background.

The shapes of the two discriminating variables m_{12} and m_{diff}, are shown in Figure 3. We find that *bqb* exhibits a harder m_{12} spectrum than the others and that *bbq* is the softest, with the others falling somewhere in between within large uncertainties. We expect that *bbc* should be similar to *bbq* because the effect of the third jet tag on m_{12} is not very large. For *bbb* and *bcb* we see a somewhat harder spectrum than for *bbq*, although the statistics are not

sufficient to determine whether the effect is real and how large is the difference. This will be followed up in more detail in Section 4.5.

In m_{diff} the events break into three categories, in fact it was constructed to produce this separation. Events with only one b-tag in the two lead jets (bcb and bqb) populate the lower end of m_{diff}, while events with two b-tags in the lead jets and no b-tag in the third (bbc and bbq) have the highest m_{diff} values. The bbb events fall somewhere in between, along with Higgs events if present in the sample. As such this variable does not separate signal from background so much as it separates the background components from one another, giving a firmer prediction for the shape of the QCD background m_{12} spectrum.

4.3 Models of bbx Processes

As a starting point for a model of the backgrounds other than bqb, we examine events where the two lead jets are tagged and there is a third jet with no tag requirement. The Monte Carlo indicates that in 88% of these events both tags are b-tags, with about 4% each of bc and cc. Comparisons of m_{12} and m_{diff} between the full double-tagged sample ($++j$) and only the bb component (bbj) are shown in Figure 4 (a constant value of 1.5 was used for the tag mass of jet 3 in this case). We find that the differences are small, so that we can treat this sample as if it were pure bbj.

The simplest way to convert the double-tagged events into background shapes for the triple-tagged sample is to weight the events based on the expected probability for a tag on the third jet for a given jet flavor. The probabilities come from the same tag efficiency parametrizations discussed above, and also give a spectrum of tag masses for the third jet which is needed to compute m_{diff}.

4.3.1 The bbq and bbc Processes

Figure 5 shows comparison between real bbq and bbc events, and the background shapes constructed from the double-tagged sample weighting the third jet as a positive mistag or charm tag. The agreement is good, which is expected since there is a large overlap in the events used to make the various curves (except for bbc which has a much smaller cross section than bbq).

4.3.2 The bbb Process

Figure 6 illustrates the result of the weighting procedure as applied to the bbb background, where we have weighted the third jet by the probability to tag it assuming it is a b-jet. The agreement in m_{diff} is good, but the bbb events show a harder m_{12} spectrum than the prediction from the double-tagged events. In Section 4.5, we will show that this effect is due to the presence of single gluon splitting to $b\bar{b}$ in bbq events, which does not contribute to bbb and which has a soft m_{12} distribution.

Figure 3: Distributions of m_{12} (top) and m_{diff} (bottom) for the QCD background components.

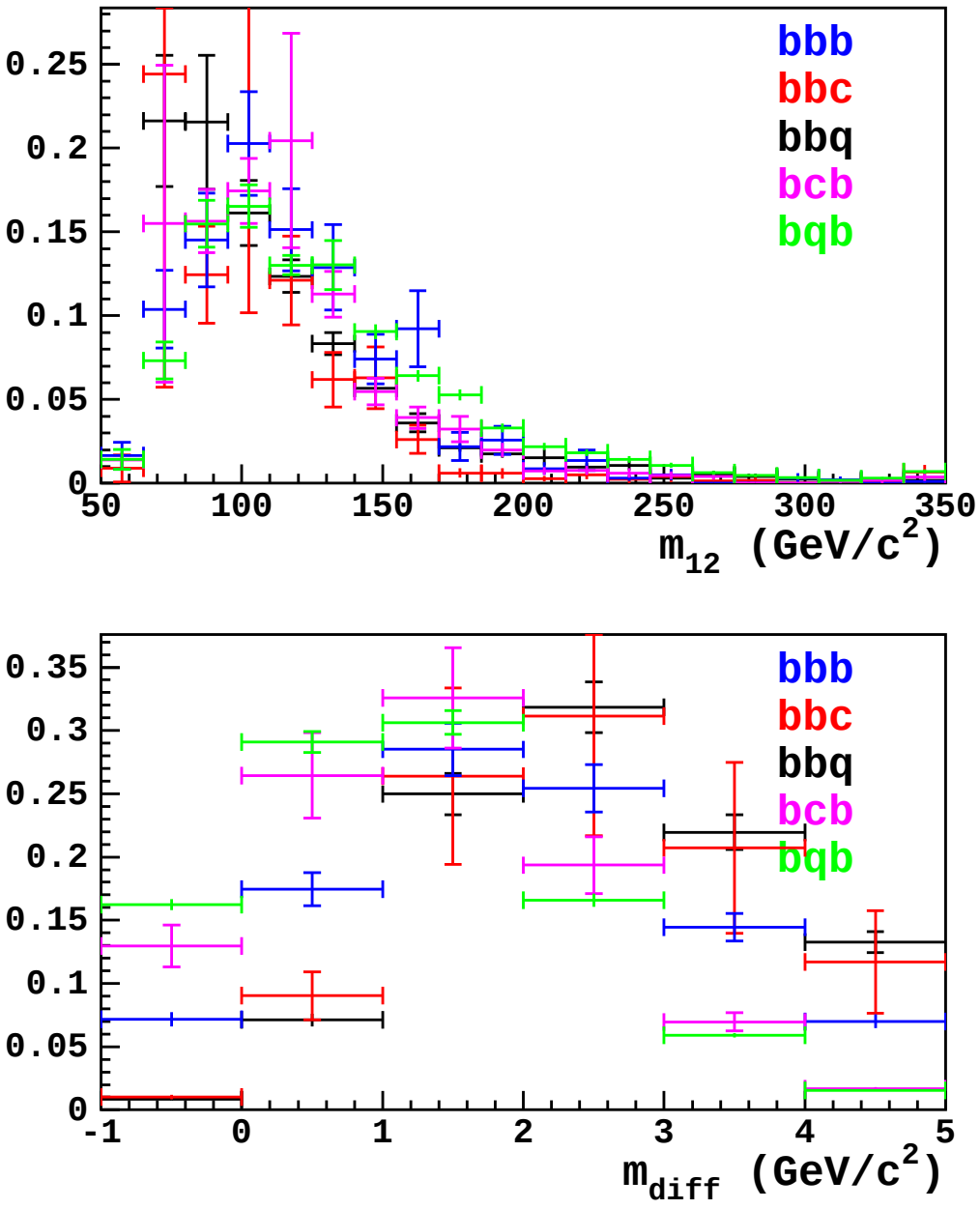
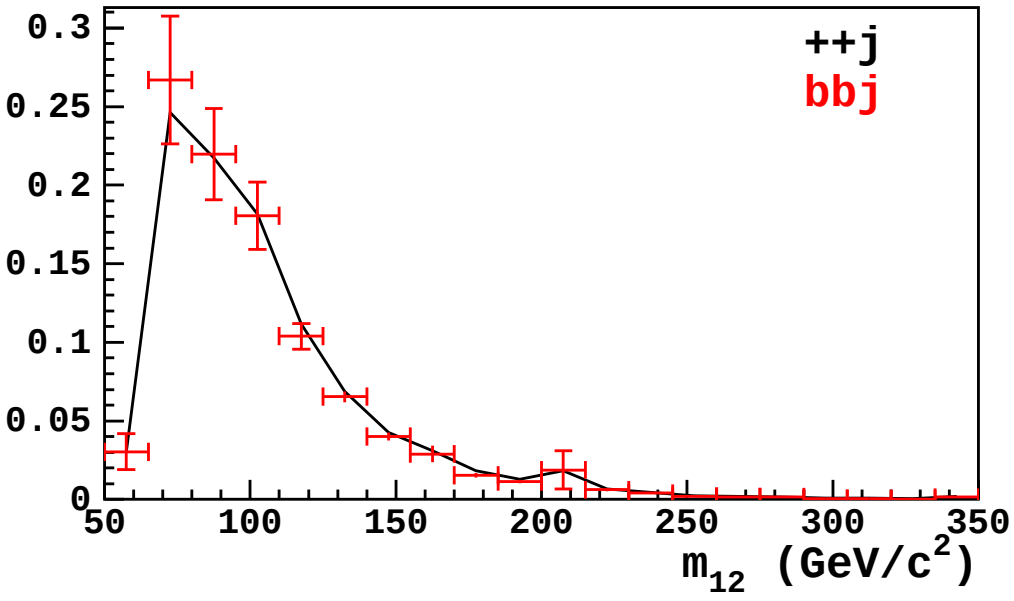


Figure 4: Distributions of m_{12} (top) and m_{diff} (bottom) for all double-tagged events ($++j$) and only the two-b-tag part (bbj). The distributions are normalized to the same area because only the shapes are of interest.



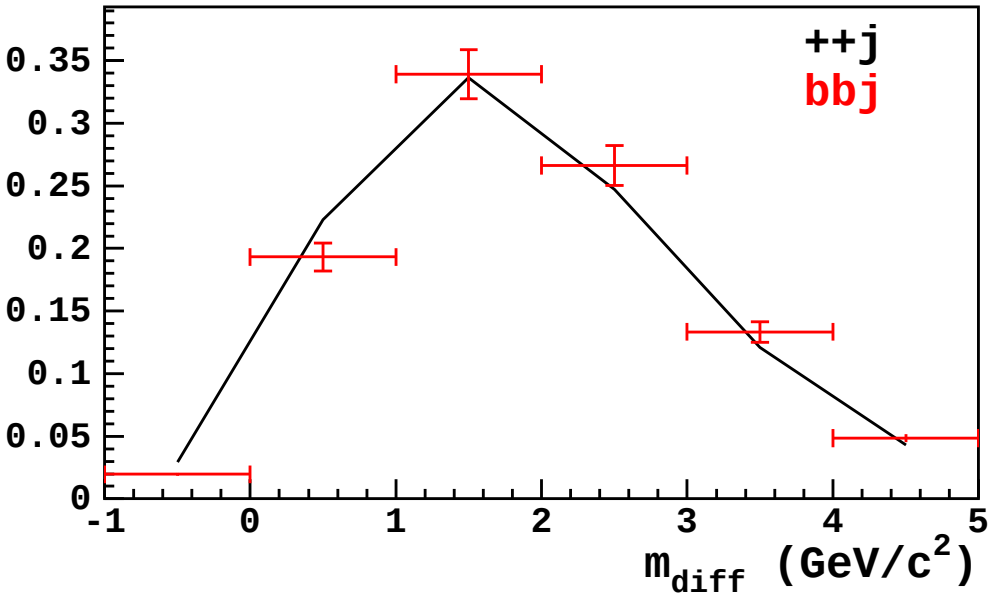


Figure 5: Distributions of m_{12} (top) and m_{diff} (bottom) for bbq and bbc events, compared to their respective backgrounds $++q$ and $++c$ derived from the double-tagged sample.

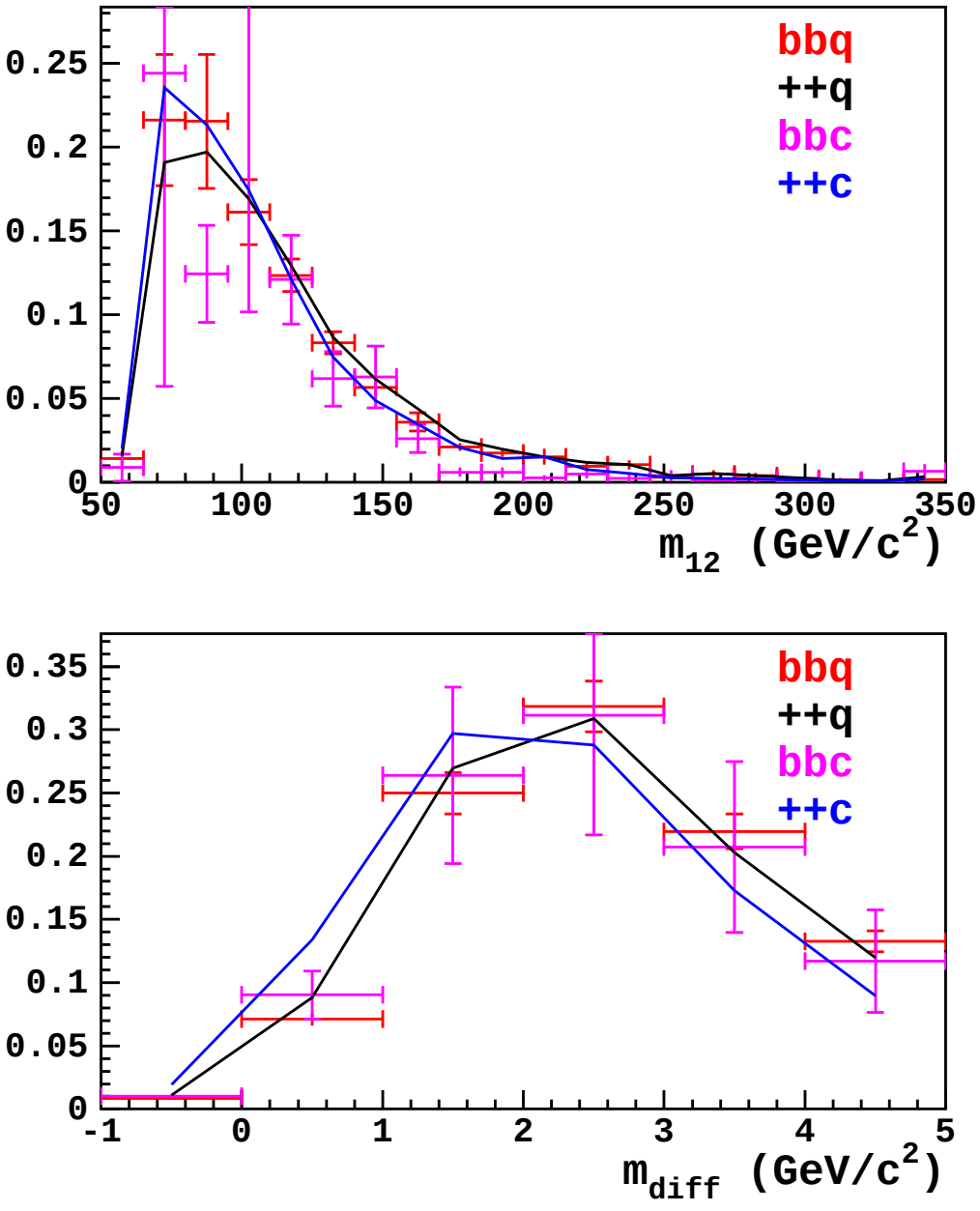


Figure 6: Distributions of m_{12} (top) and m_{diff} (bottom) for bbb events, compared to the background $++b$ derived from the double-tagged sample.

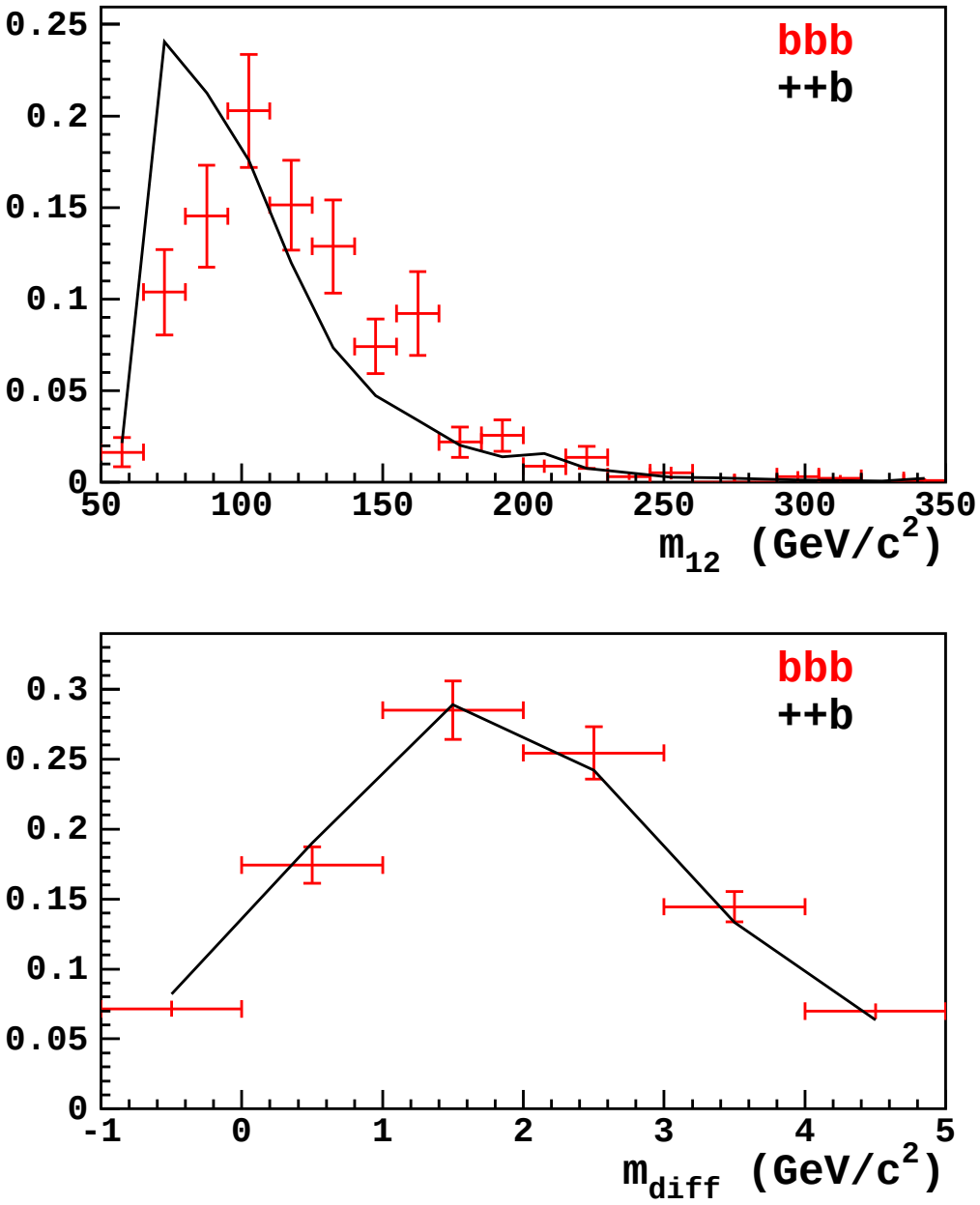
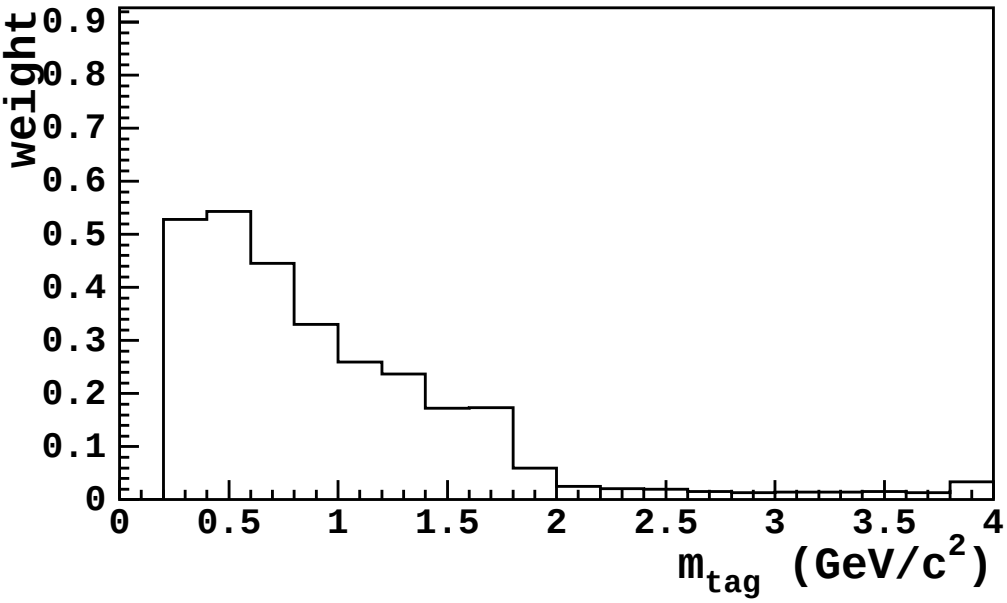


Figure 7: Ratio of charm to bottom tag masses, used to construct the *bcb* background shape. The y scale is arbitrary.



4.3.3 The *bcb* Process

With a b-tagged third jet and heavy flavor in the two lead jets, the process *bcb* is expected to have similar properties to *bbb*, but it is clear that m_{diff} needs a correction. We do this by weighting based on the tag masses in the two lead jets according to the ratio of the m_{tag} spectra in charm and bottom tags shown in Figure 7. For each event two entries are made, one weighted by the result of plugging the tag mass of jet 1 into Figure 7 and a second using jet 2. The tag efficiency versus jet E_T is similar for charm and bottom jets so no correction is needed there. The results of this procedure are shown in Figure 8. The m_{diff} agreement is very good, but as with *bbb*, the m_{12} spectrum appears a little too soft. We will return to this in Section 4.5.

4.4 Model of the *bqb* Sample

The background for *bqb* is derived in a similar way to the others, but uses a slightly different sample. The events are still double-tagged, but it is the third jet and either of the two leading jets which are required to be tagged in this case. The two lead jets are still required to match the trigger objects, but without the impact parameter cuts on the SVT and L3 silicon tracks in the untagged jet. The Monte Carlo study indicates that the two tags are b-tags in these events 86% of the time, similar to the other double-tagged sample. The events are weighted by a light-flavor jet efficiency parametrization to have a tag and displaced SVT and L3 silicon tracks in the jet. The results are shown in Figure 9.

Figure 8: Distributions of m_{12} (top) and m_{diff} (bottom) for bcb events, compared to the background $+cb$ derived from the double-tagged sample.

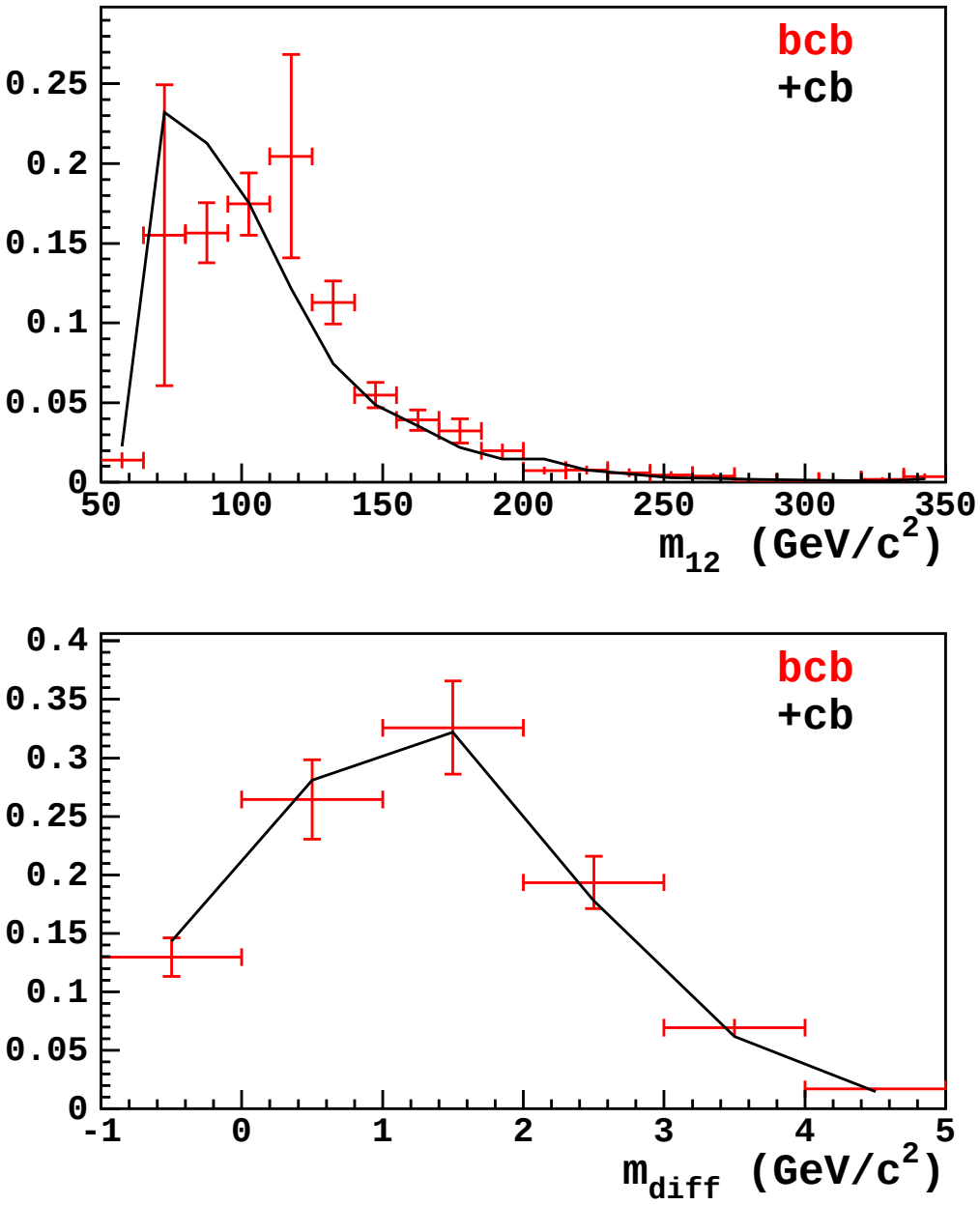


Figure 9: Distributions of m_{12} (top) and m_{diff} (bottom) for bqb events, compared to the background $+q+$ derived from the double-tagged sample.

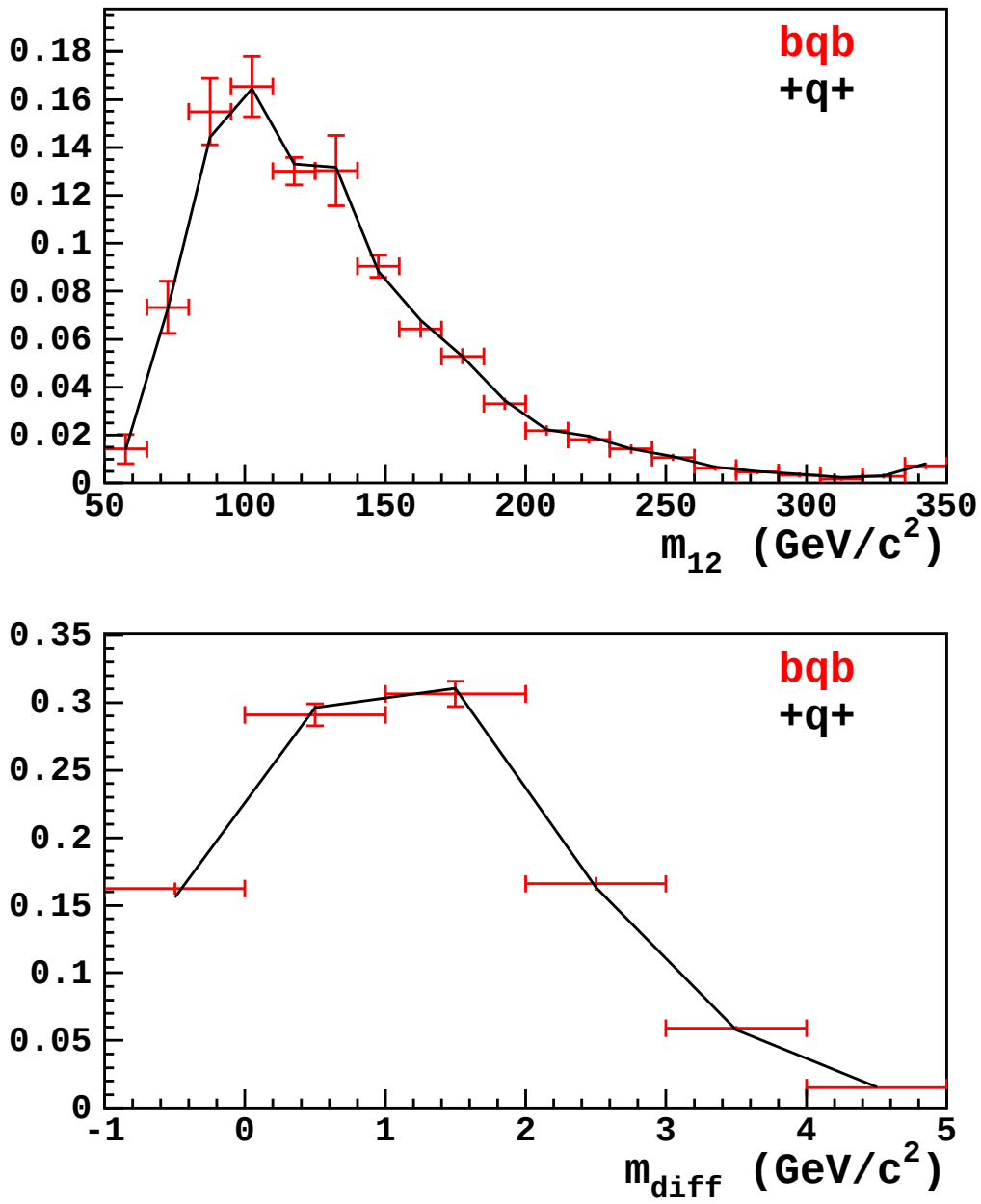
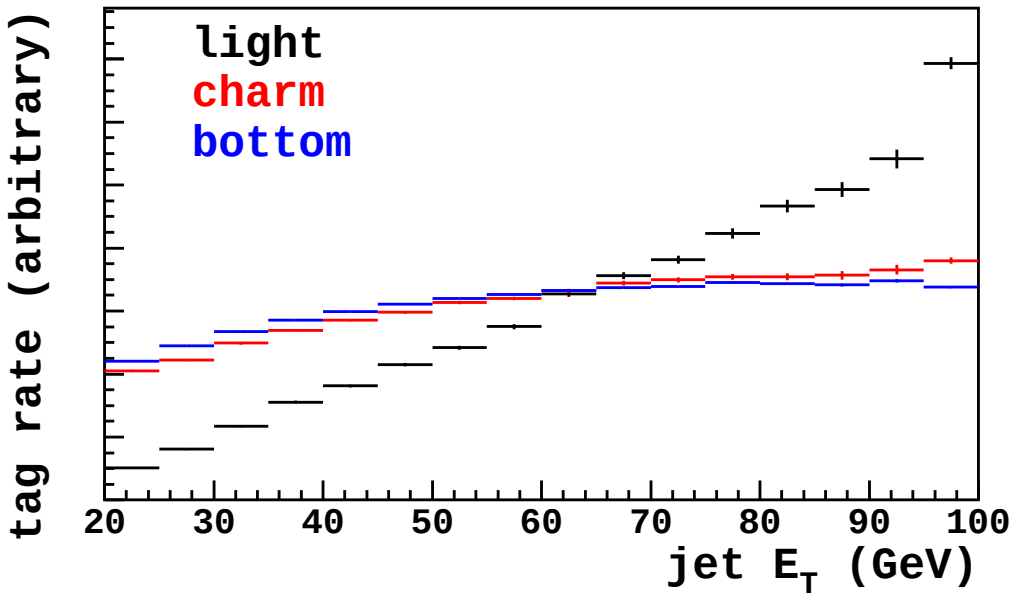


Figure 10: Tag rates versus jet E_T for light, charm, and bottom jets. The y scale is arbitrary and set so as to better illustrate the difference in E_T-dependence. The highest bin contains overflows.



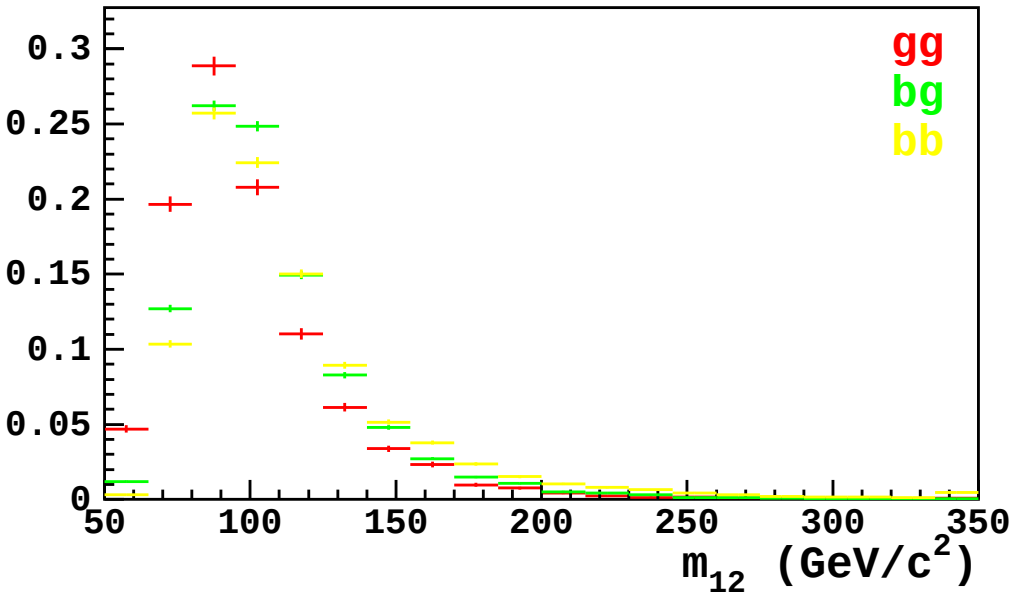
The m_{12} distribution is harder in this component of the background than in the others. The reason for this can be understood from Figure 10, which shows the E_T-dependence of the tag rates on an arbitrary scale. While the b-tag efficiency plateaus at high jet E_T, the mistag rate grows roughly linearly with E_T. The E_T of the tags, and thus the m_{12} spectrum, is therefore strongly affected by the presence of a mistag in the two leading jets.

4.5 Generator-Level Study of the *bbb* and *bcb* Processes

The discussion above showed that the *bbb* and *bcb* backgrounds have slightly higher means than our corresponding models built from the double-tagged events. Since the statistics in the fully-simulated QCD dijet samples are insufficient to answer the question, we use much larger generator-level samples to study the underlying physics.

The events are produced using Pythia 6.411 (MSEL=1) and the default tuning with PARP(67)=4, which corresponds to the "high-ISR" setting of Ref. [16]. The $\hat{p}_T$ cut on the hard scatter was set to 18 GeV/c; a smaller sample with $\hat{p}_T > 10$ was produced to verify this setting. The events are processed with the PYCELL jet finder (cone 0.7, 24 ϕ bins, η granularity 0.1) and the events were required to have two jets with $E_T > 35$ GeV and $|\eta| < 1$, and a third jet with $E_T > 20$ GeV and $|\eta| < 1.5$. These jets were matched to heavy-flavor hadrons to determine the jet flavors. Over six billion events were generated of which 383k passed the jet cuts and had at least one matched b-jet

Figure 11: Distributions of m_{12} in the bbj generator-level sample, for different production mechanisms.



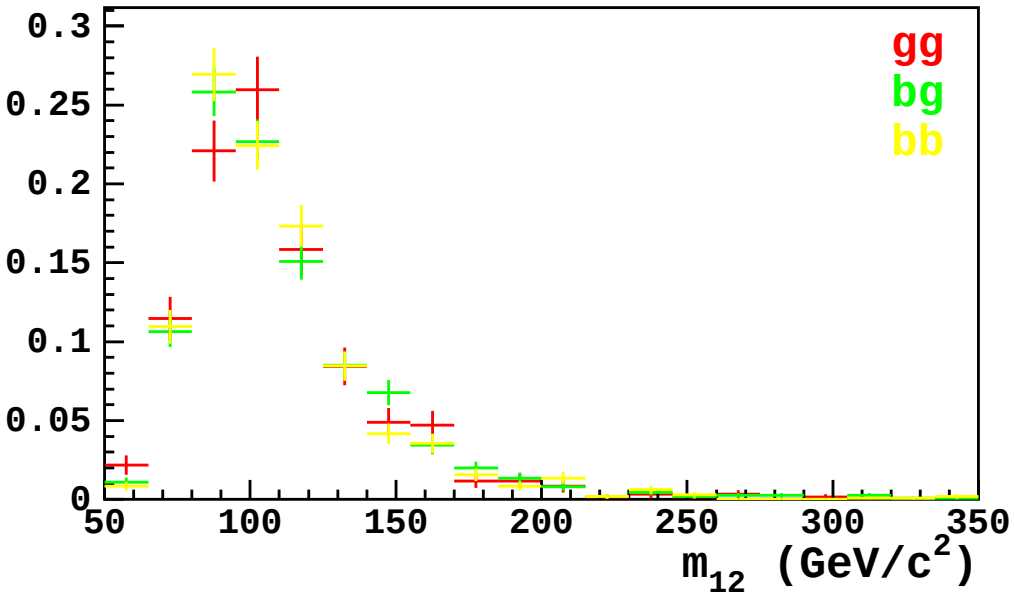
We divide the events by production mechanism, based on the numbers of heavy quarks entering and exiting the hard scatter.

- GS (gg): all heavy flavor is produced through gluon splitting, none in the hard scatter
- FE (bg, cg, bc): flavor excitation of heavy quarks in the initial state
- FC (bb, cc): flavor creation, no heavy quarks in the inital state

Distributions of m_{12} for events where both of the leading jets are b-jets and any flavor third jet are shown in Figure 11, separated by production process. Only processes which contribute at least 5% of the total are shown. We find 16% GS, 47% FE, and 36% FC in this sample. The GS events have a softer m_{12} spectrum than FE or FC.

Distributions of m_{12} for events where all three jets are b-jets are shown in Figure 12. In this sample we find 22% GS, 42% FE, and 36% FC, similar to the bbj sample. However, we see no difference in the GS m_{12} distribution compared to the others, they are all similar to the bbj FE and FC shapes. This indicates that the double-tagged sample really is softer in m_{12} due to the GS events. In double-tagged events, the two b-jets come from the same gluon splitting, while in triple-tagged events there have to be two gluon splittings and the two lead jets can come from separate gluons. Jets from a single gluon splitting tend to be closer in ϕ and therefore have a softer m_{12} spectrum than the other types of events, where the lead jets are more back-to-back.

Figure 12: Distributions of m_{12} in the bbb generator-level sample, for different production mechanisms.



Distributions of m_{12} for bcb events are shown in Figure 13. In this case there are more production mechanisms in play, such as charm flavor excitation and double-excitation involving both a charm and a bottom quark in the initial state. The totals are 28% GS, 63% FE, and 9% FC. Similar to bbb we do not find much dependence of m_{12} upon the production process, although the FC shape is perhaps a little softer than the others.

We now want to derive a correction to turn our observed bbj events into something that looks like bbb or bbc. We found that bbb and bcb are largely insenstive to relative populations of the various production mechanisms, however the m_{12} distribution in the bbj sample depends upon how much GS is present. In Table II of Ref. [16], the "high-ISR" tune of PYTHIA was found to undershoot GS by about a factor of two. This agrees with the recent high-p_T measurement of the $b\bar{b}$ cross section presented in Ref. [17]. Ref [16] also prefers very little FE, but since the FE and FC curves are similar in bbj it doesn't matter much where we put the remainder, so we simply scale the GS component by two and scale down the rest to keep the total area the same.

The ratio of m_{12} distributions between the bbb events and the rescaled bbj events is shown in Figure 14. We find that the difference can be simply represented as a rolling off of the lower end of the distribution. Also shown on the plot are fits of a simple hyperbolic tangent function to the displayed points (central), and to similar distributions after scaling the GS part of bbj by one (no correction) and by three. These alternate fits will be used to estimate systematic errors. A similar procedure was performed for the bcb events relative to bbj, the results are shown in Figure 15. The rolloff is slightly more pronounced in bcb compared to bbb.

Figure 13: Distributions of m_{12} in the bcb generator-level sample, for different production mechanisms.

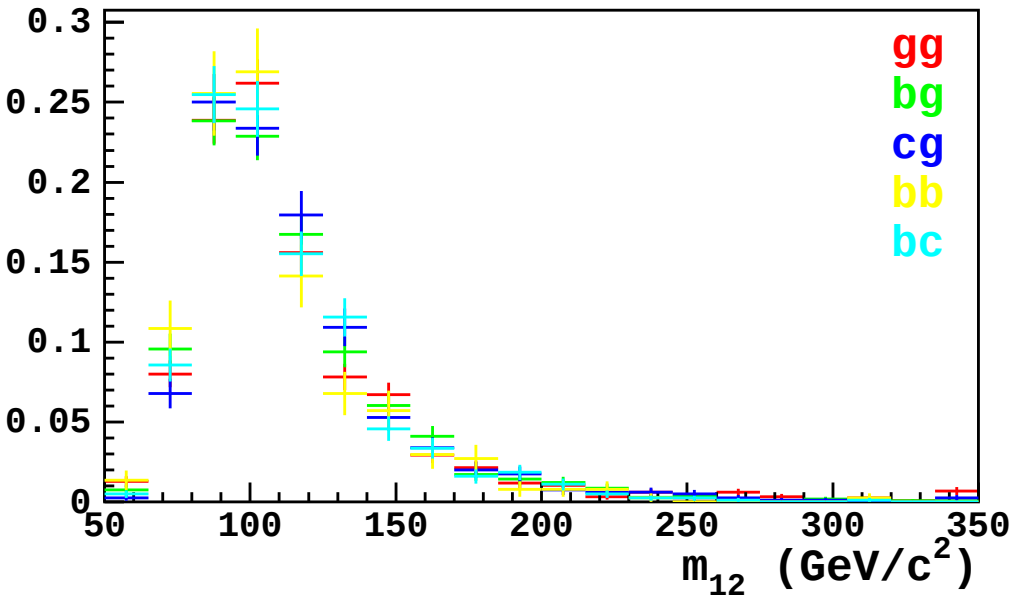


Figure 14: Ratio of the m_{12} distribution in bbb events relative to GS-scaled bbj events. The fits are described in the text. The y scale is arbitrary.

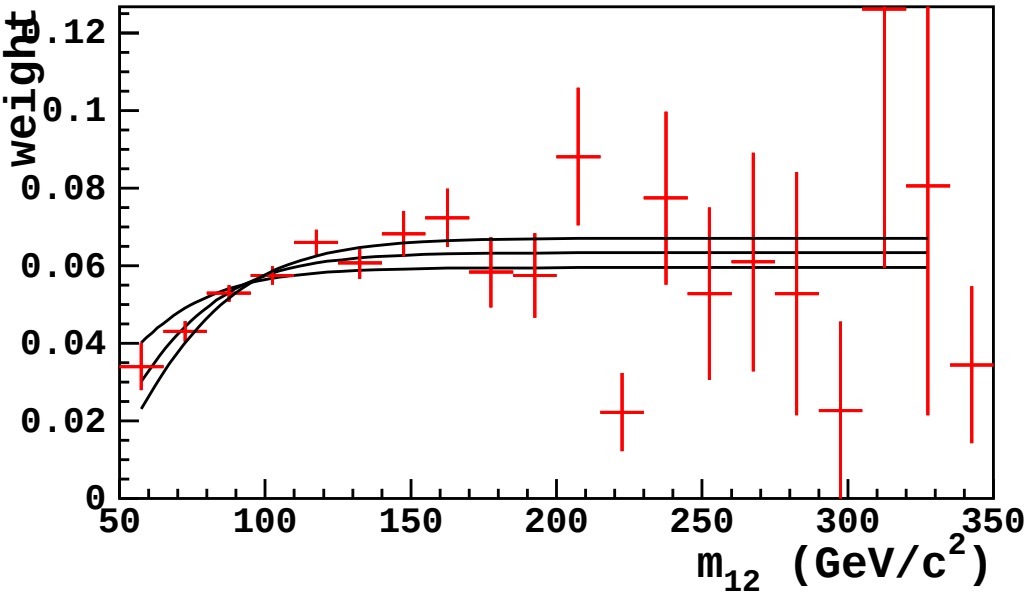
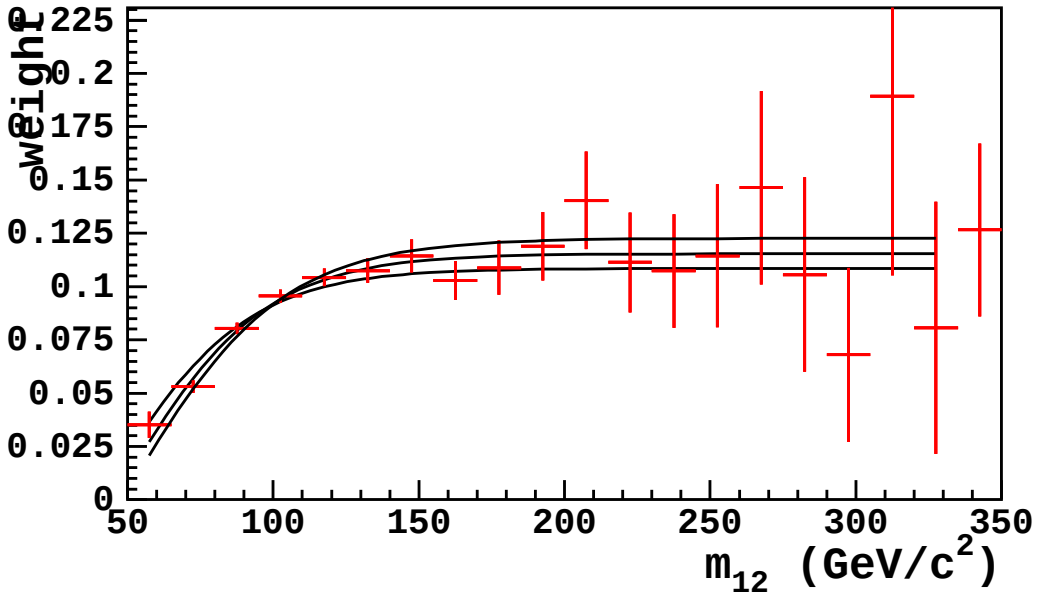


Figure 15: Ratio of the m_{12} distribution in bcb events relative to GS-scaled bbj events. The fits are described in the text. The y scale is arbitrary.



In order to check that the corrections derived above do not depend on some lucky coincidence in the particular PYTHIA tune chosen, we generated a similarly-sized sample with PARP(67)=1 to match the "low-ISR" results of Ref. [16]. The m_{12} distributions are shown in Figure 16 (bbj), Figure 17 (bbb), and Figure 18 (bcb). The bbj FE and FC components are somewhat harder than in the default case, otherwise the distributions are similar to what was seen earlier. The level of FE in bbj is predicted to be only 9% in this scenario compared to 47% in the other sample, with the remainder split about 2:1 between FC and GS.

Table II in Ref. [16] indicates that for this sample the FE component should be scaled up by a factor of three, while preserving the ratio between GS and FC. The ratio of bbb events to the rescaled bbj events is shown in Figure 19, along with the same curves from the GS-scaled samples. The same treatment applied to bcb events is shown in Figure 20. In both cases there is a falloff at high m_{12} that was not seen in the default sample. However, in the 70-150 GeV/c^2 region where the majority of the events are found the agreement is good with the points lying within the systematic error envelope, and only becomes poor above 200 GeV/c^2 where there are hardly any events.

We will use the central fits from Figures 14 and 15 to correct the double-tagged estimates $++b$ and $+cb$ in order to get final estimates for the bbb and bcb shapes.

4.6 Data-Driven Background Templates

With the corrections for bbb and bcb in hand, we can use the 35K double tag events in our data to derive all five background templates bbb, bbc, bbq, bcb, and bqb. The results are shown

Figure 16: Distributions of m_{12} in the bbj generator-level sample, for different production mechanisms.

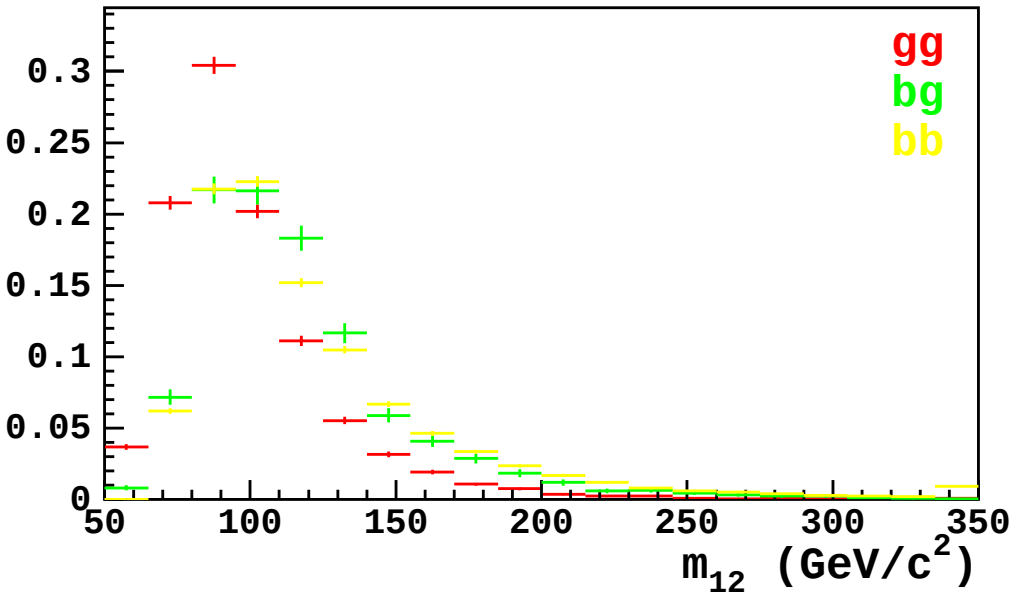


Figure 17: Distributions of m_{12} in the bbb low-ISR sample, for different production mechanisms.

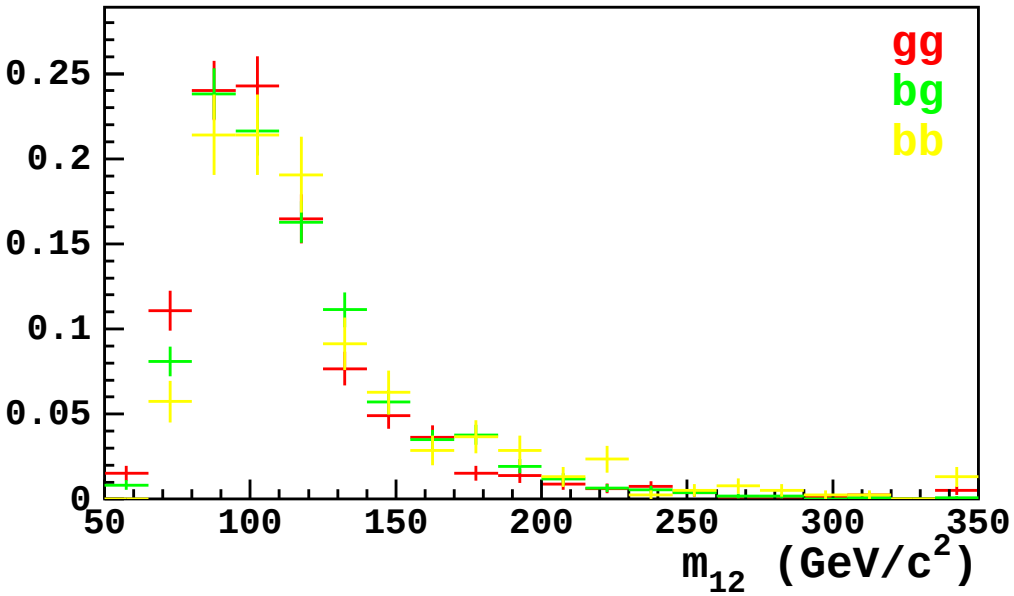


Figure 18: Distributions of m_{12} in the bcb low-ISR sample, for different production mechanisms.

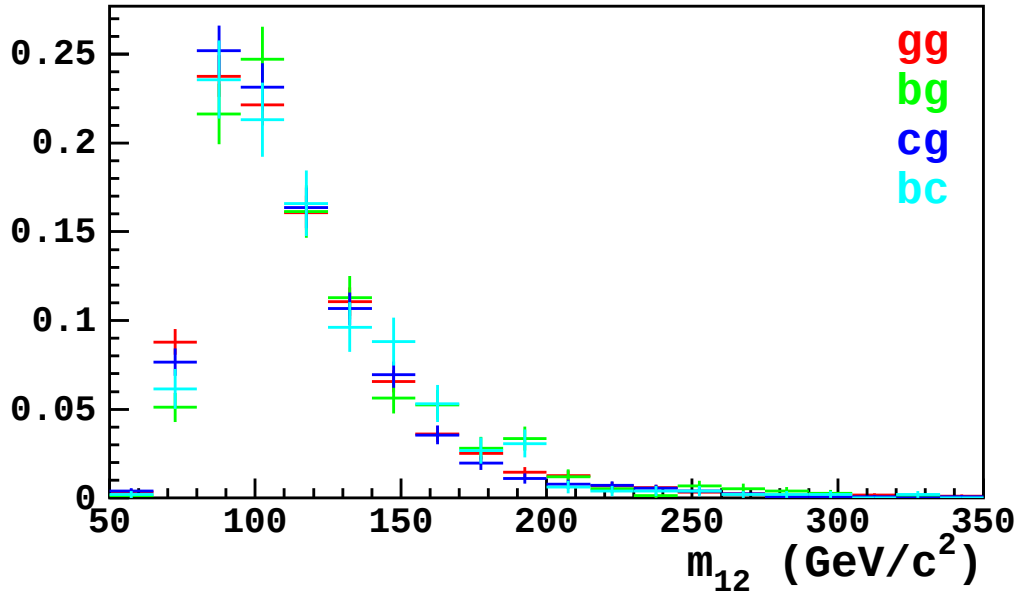


Figure 19: Ratio of the m_{12} distribution in bbb events relative to FE-scaled bbj events in the low-ISR sample. The fits are described in the text. The y scale is arbitrary.

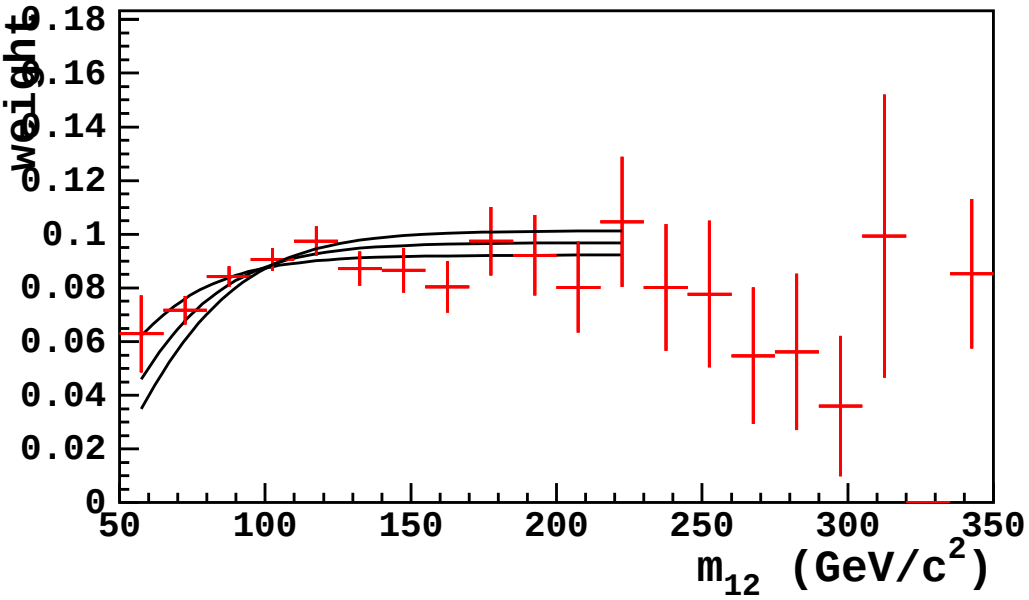
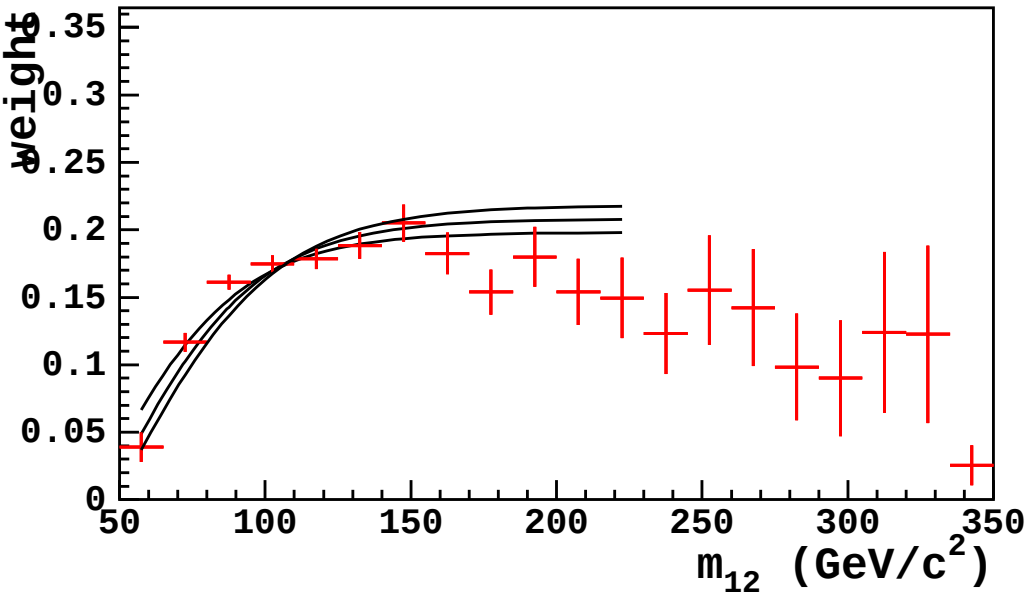


Figure 20: Ratio of the m_{12} distribution in bcb events relative to FE-scaled bbj events in the low-ISR sample. The fits are described in the text. The y scale is arbitrary.



in Figure 21. The effect of the m_{12} corrections on bbb and bcb can be seen by comparing to the bbc shape, which would otherwise be identical to bbb and bcb in m_{12}. Because bbq and bbc are so similar in m_{diff} they can't be separated by the fit. Instead, we use the average of the two shapes as a fit template (which we call bbx), and vary between them to estimate a systematic error.

The effect of the mistag bias to higher jet E_T, as discussed in Sec. 4.4 is evident in the m_{12} distribution for bqb. The m_{diff} distribution shows the desired property of separating the two-b vs. one-b in the two lead jets background components to opposite sides, leaving the bbb background and the Higgs signal in the middle.

The application of these templates to describe the observed +++ sample, with and without a Higgs signal, is discussed in Section 7.

5 Signal Model and Acceptance

In this section the fit templates and acceptance values for the hypothetical Higgs signal contribution are described. Several corrections were made to the Monte Carlo in order to match NLO theoretical cross section calculations and detector performance in the data. These take the form of weights, which can be either event-wide or per-jet. The total weight is the product of all event-wide weights and per-jet weights and is used when counting events and filling histograms.

Figure 21: Distributions of m_{12} (top) and m_{diff} (bottom) for the background templates derived from data.

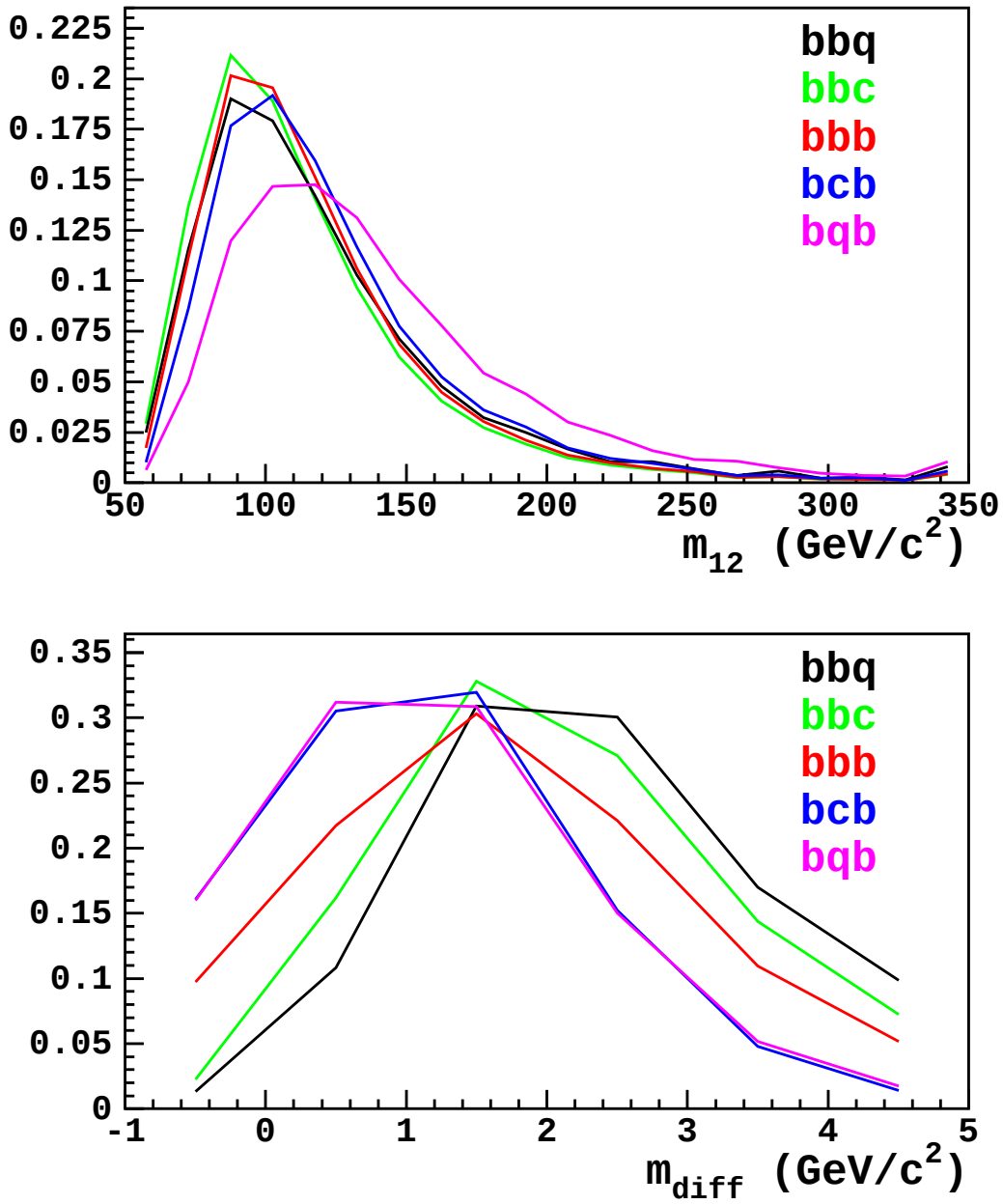
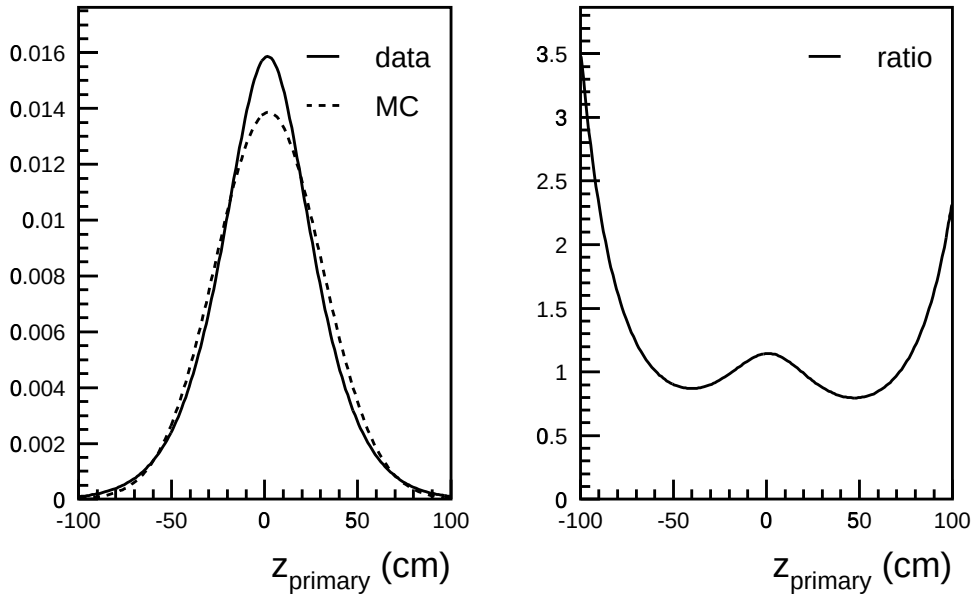


Figure 22: The z distribution of vertices for data and Monte Carlo (left) and the data/MC ratio (right).



The first correction is for the z width of the luminous region. The Monte Carlo used a Gaussian distribution with mean 2.07 cm and width 28.8 cm (determined from fitting the distribution of origins in one of the Higgs samples). This is a fairly good match to early data, however in more recent running periods the Tevatron optics have been changed to narrow the width in z. In Ref. [18] four run periods are defined and parametrizations of the z distribution of vertices in minimum-bias data are given for each. By weighting each of these by the luminosity recorded with HIGH_PT_BJET, an average beam profile for the full data sample can be constructed. This is shown in Figure 22, along with the Gaussian distribution used by the Monte Carlo. Also shown is the data/MC ratio, which is used to weight the Monte Carlo events so that they match the true distribution in z. This correction increases the acceptance by 2-3%.

The second set of corrections is related to the b-tagging and trigger requirements studied in Ref. [15]. For the two jets which satisfy the trigger b-tag cuts, the fit shown in Figure 17 of that note was used to derive a per-jet weight for each. We use the E_T-dependent scale factor parametrization $(1.064 - 0.0005E_T)$, as the default, and vary the uncertainties on the slope and intercept to derive systematic errors. In addition, the ratio of turn-on efficiencies shown in Figure 20 of [15] was applied to each of the trigger b-tagged jets. In the event that all three jets were trigger b-tagged, these weights were applied only to the two highest-E_T jets, and the third was considered to be simply a standard b-tag. For standard b-tags, the fit shown in Figure 13 of [15] $(0.872 + 0.00118E_T)$ was used as a per-jet weight, and the turn-on for the third L2 cluster requirement in HIGGS_HIGH_PT_BJET shown in Figure 20 of [15]

was applied. Because this requirement was highly efficient it was applied to all of the data to simplify the analysis, even the earlier periods when there was no such online requirement.

With all corrections specified the acceptance can be computed, however the denominator must first be chosen. This sets the physics process for which a cross section will be measured or for which limits will be set. Ideally, matching theoretical calculations should be available in order to facilitate conversion of a cross section into the corresponding $\tan\beta$ value. For these results the FeynHiggs program [19] v2.2.10 is used. It reports cross sections for inclusive $b\bar{b}A$ (based on the NNLO result [3]) and for the case when at least one of the quarks has $p_T > 15$ GeV/c and $|\eta| < 2.5$ (based on the NLO calculations [4]). The latter corresponds more closely to the final state studied here, and also provides kinematic distributions against which the Monte Carlo can be compared, which are not available from the NNLO results.

The PYTHIA Monte Carlo samples were generated without any requirement on the kinematics. In order to match them to the NLO calculations, the event record is searched for the $b/\bar{b}$ which came from the hard scatter $bg \to bH$, and also for the $\bar{b}/b$ which came from the same shower that produced the $b/\bar{b}$ which participated in the scattering. Only partons at the end of the showering process (status=2) were used. Events were retained only if at least one of the two passed the cuts given above ($p_T > 15$ GeV/c, $|\eta| < 2.5$). Around 2% of events which would otherwise have passed the event selection cuts fail this requirement, which is negligible compared to the eventual uncertainties.

An alternative method was also tried, which was to reconstruct $\hat{p}_T$, the p_T of the outgoing particles in the frame of the hard scatter, and to require $\hat{p}_T > 15$ GeV/c. The transverse momentum and pseudorapidity of the Higgs for the inclusive Monte Carlo and after the two sets of NLO matching requirements (all normalized to the same area) are shown in Figure 23. The two sets of NLO matching requirements are seen to be quite similar. A difference which is not apparent from the plot, however, is that the $\hat{p}_T$ method requires that it be the $b/\bar{b}$ which participated in the scatter that receives large p_T, whereas the $b/\bar{b}$ p_T cut can pass on either the scattered b or its partner from the initial shower. The result is that the $\hat{p}_T$ method only accepts roughly half as many events as does the $b/\bar{b}$ p_T method. Given that the results are so similar, the more efficient matching method was chosen to economize Monte Carlo statistics, however for future studies a $\hat{p}_T > 15$ GeV/c cut could safely be applied during Monte Carlo generation to save time.

Finally, with all corrections and generator-level cuts applied, the acceptance is calculated. The values for various Higgs masses are shown in Table 3. Because of the weighting of events, the statistical error cannot be computed simply as $\sqrt{N}$. However, the weights do not differ dramatically from one, so from the approximately 1200 events which survive the event selection for each mass a statistical error of around 3% can be expected.

Also shown in Table 3 are values for different PDF sets (including the default CTEQ5L), found using the reweighting method detailed in Ref. [20]. The acceptance using the CTEQ6 NLO sets (5M, 6M, 6.1M) is 6-8% lower than for the other sets, which are all consistent within 1%. It does not appear that this is an artifact of using NLO PDF's, as can be seen from the CTEQ5M1 results, nor is it due to inconsistent α_s or Λ_{QCD} values according to the CTEQ6L and CTEQ6L1 numbers.

Two of the PDF sets, CTEQ5L and CTEQ6.1M, were chosen for a comparison with the kinematic distributions shown in Figures 8 and 9 in Ref. [4]. The results are shown in

Figure 23: Higgs p_T and pseudorapidity (NOT rapidity) distributions, for the inclusive Monte Carlo sample and for the two matching criteria described in the text.

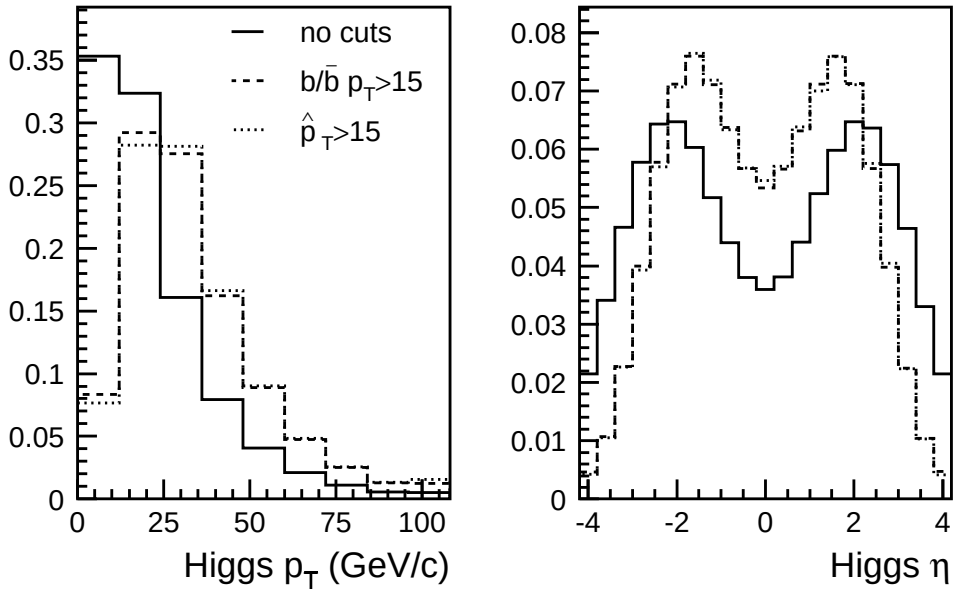


Table 3: Event selection acceptance for various Higgs masses and PDF sets.

PDF	$m_H = 90$	$m_H = 120$	$m_H = 150$	$m_H = 180$	$m_H = 210$
CTEQ5L	0.291%	0.509%	0.641%	0.723%	0.747%
CTEQ6L	0.299%	0.522%	0.653%	0.734%	0.755%
CTEQ6L1	0.297%	0.518%	0.651%	0.733%	0.755%
CTEQ5M1	0.297%	0.517%	0.651%	0.733%	0.757%
CTEQ6M	0.284%	0.490%	0.615%	0.687%	0.711%
CTEQ6.1M	0.279%	0.479%	0.602%	0.669%	0.694%
CTEQ6.5M	0.284%	0.487%	0.607%	0.672%	0.692%

Figure 24: Higgs p_T and η distributions, for two PDF sets and for the MCFM NLO calculation (for $m_H = 120$).

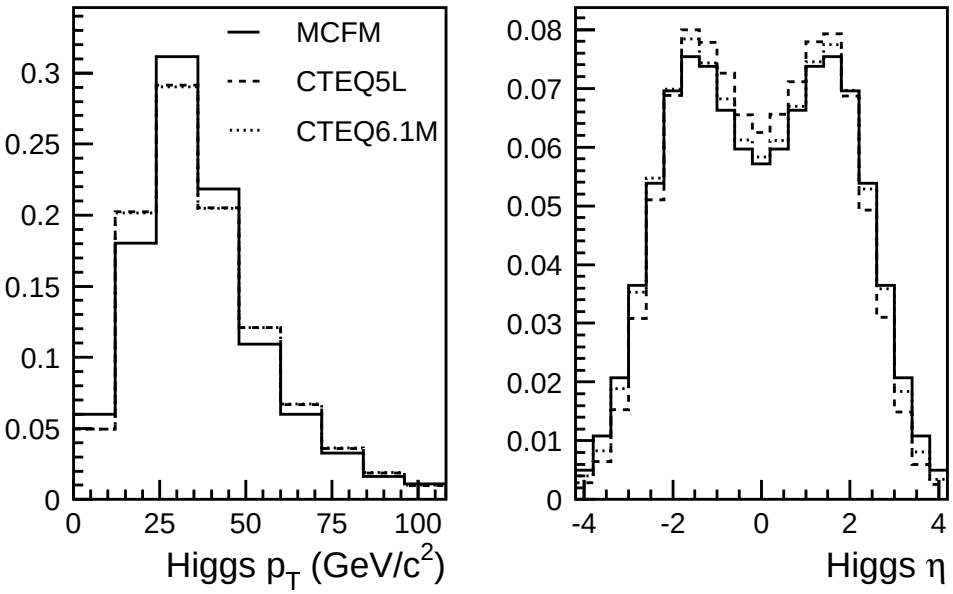
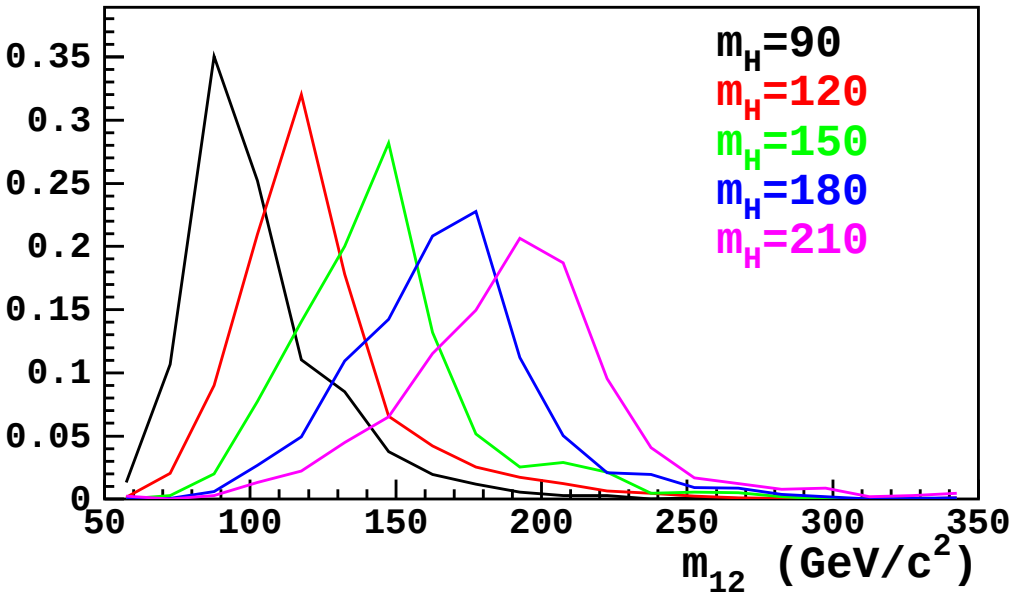


Figure 24. In this case the matching cuts on the $b/\bar{b}$ quarks were tightened to $p_T > 20$, $|\eta| < 2$ to match what was used in that paper. It should be noted that CTEQ6M was used in Ref. [4] and also for the cross sections given by FeynHiggs. The PDF choice has little effect on the Higgs p_T, however it does change the η distribution by a noticeable amount. Only events with $|\eta| < 2$ (for the most part) are accepted by the event selection, so the broader distribution of CTEQ6.1M is what gives the lower acceptance compared to CTEQ5L. Because it is what was used for the NLO calculation, the acceptance values derived from the CTEQ6M reweighting are chosen as the default values.

Compared to the Monte Carlo, the NLO p_T distribution shown in Figure 24 is neither harder nor softer but rather narrower around the peak. To check the effect of this, an event selection efficiency versus Higgs p_T function was convoluted with the MCFM and CTEQ6.1M distributions. The difference was less than 2%, so no correction is made.

One last issue is that due to a misunderstanding, the Higgs Monte Carlo samples were generated using both Tune A and the Run II W/Z p_T tune configurations. The main mistake is that PARP(64) ends up set to 0.2, far from the suggested default of 1.0 and even outside the "more ISR" value of 0.5. To investigate the impact of this, generator-level distributions of the Higgs p_T for the settings that were used and the suggested defaults were compared, for $\hat{p}_T > 15$. The difference between the two was very similar to the difference between the Monte Carlo and the MCFM calculation shown in Figure 24, which was already seen to have negligible effect on the acceptance.

Figure 25: Distributions of m_{12} for the Higgs signal samples.



The distributions of m_{12} for various Higgs masses after all corrections have been applied are shown in Figure 25. The $<m_{diff}>$ distribtions are shown in Figure 26. The latter do not show much dependence on Higgs mass because it is *bbb* in all cases, and the tag mass does not vary much with jet E_T. Based on studies within the high-p_T b-tag group the tag masses in the simulation are scaled down by 3% to obtain better agreement with the data.

6 Control Sample Studies of the Background Model

The background shapes were tested on a few control samples. Unfortunately, it is difficult to come up with a realistic sample that would not contain any Higgs signal. One useful sample (NEG) is events with two positive and one negative tag, where the negative tag can be on any of the three jets. A second sample (2pass2) with higher statistics can be selected by requiring positive tags on the two leading jets and at least two SECVTX pass2 tracks (but no positive tag) on the third jet. For the tag mass on the third jet in this sample the invariant mass of the pass2 tracks is used. Because these samples are not expected to have any *bbb*, *bcb*, or *bbc* in them, we fit them with the *bbq* and *bqb* background templates only.

Figure 27 shows the fit of the NEG sample using the two background templates, while Figure 28 shows the same for the 2pass2 sample. The background shapes appear to successfully describe the data. In order to increase statistics the same procedure was repeated using the loose tune of SECVTX. The results are shown in Figures 29 and 30, and again the fit quality looks good.

Figure 26: Distributions of $<m_{diff}>$ for the Higgs signal samples.

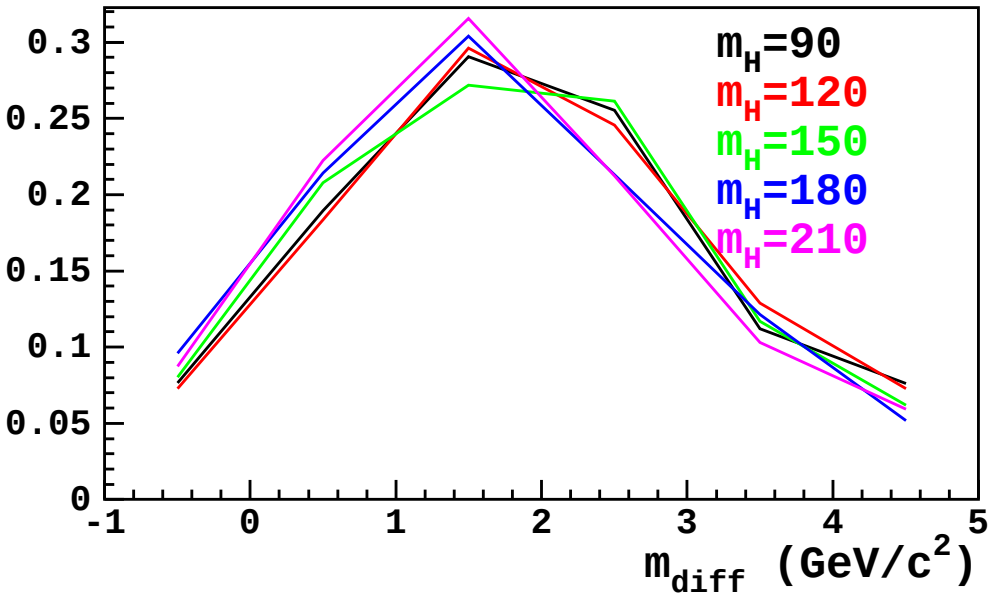


Figure 27: Background-only fit to the NEG data sample.

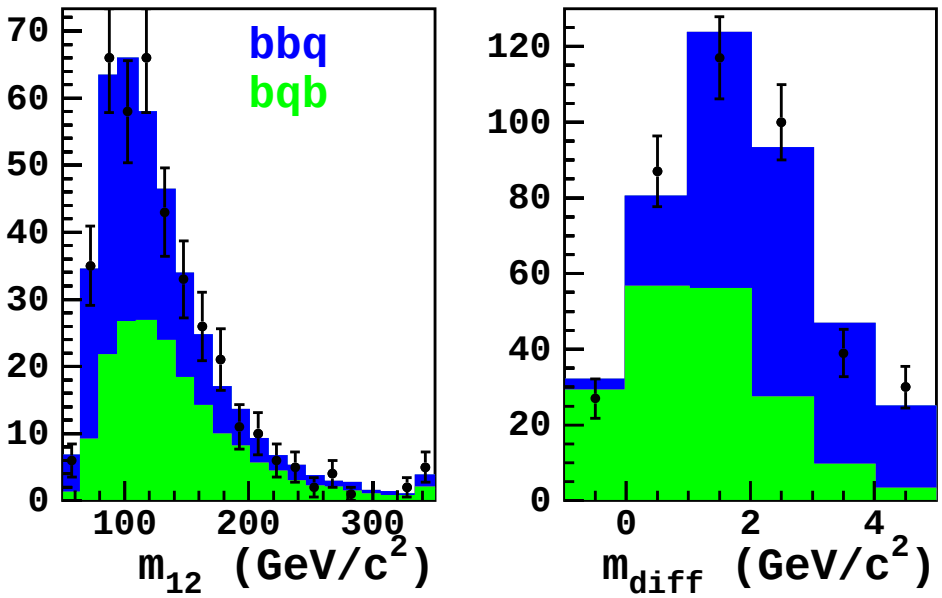


Figure 28: Background-only fit to the 2pass2 data sample.

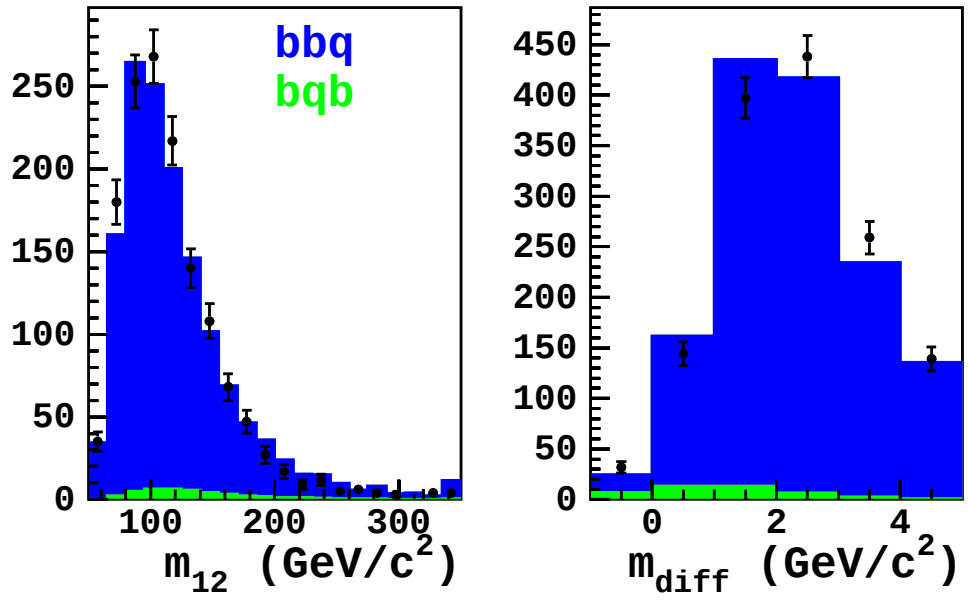


Figure 29: Background-only fit to the NEG data sample, using loose SecVtx.

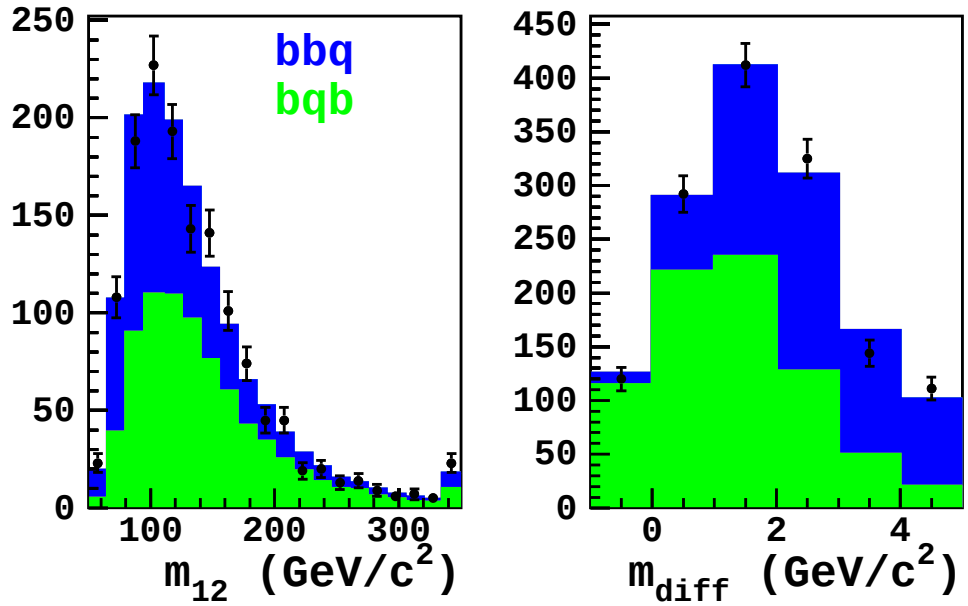
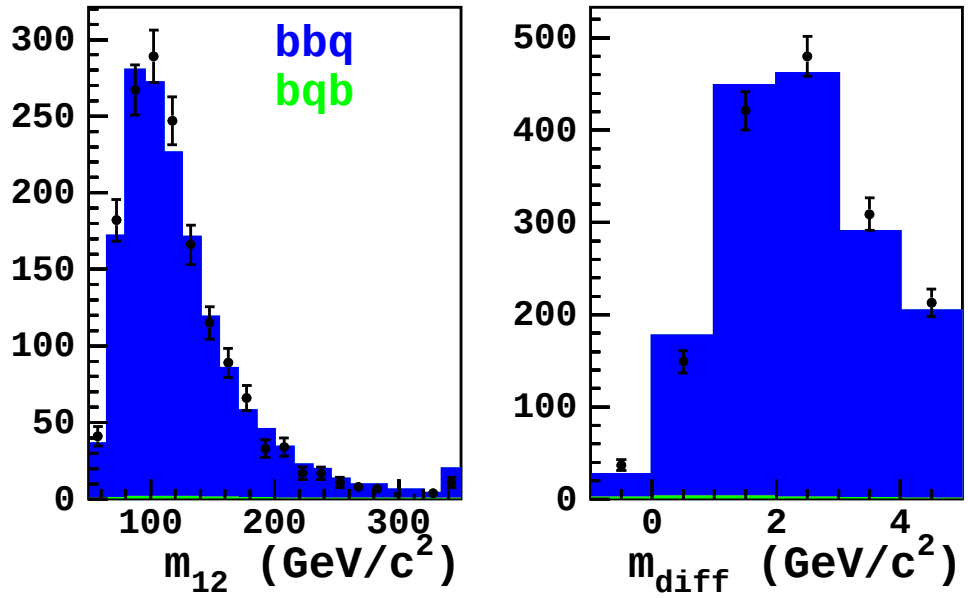


Figure 30: Background-only fit to the 2pass2 data sample, using loose SecVtx.



In order to make this a more quantitative study, the fits were repeated with Higgs signal templates included. Ideally the fits would return zero Higgs events for any mass in the case of no bias. Figure 31 shows the number of Higgs events found in each of the four control samples described above for Higgs masses from 90 to 210 GeV/c^2 in steps of 30 GeV/c^2. The fits are generally consistent with zero signal, and do not show any trend versus Higgs mass which would indicate that the backgrounds are too hard or too soft except for a bit of a downward trend with increasing mass in the tight tagged 2pass2 sample. The uncertainties decrease with Higgs mass because as the mass increases the signal moves into lower-background regions of m_{12}.

7 Simple Fits to m_{12} and m_{diff} in the Data

Although the limits will be calculated using the MCLIMIT package, it is interesting to first look at the +++ sample using a simple MINUIT fitter. The results of a fit to the 1582 observed +++ events using only the background templates is shown in Figure 32 and Table 4. The populations of the various components are similar to what was found in the Monte Carlo study. The fit χ^2/dof returned by MINUIT is 1.7, however many bins are non-Gaussian due to small numbers of entries so this number is not neccessarily distributed as a true χ^2.

Figure 33 and Table 33 show the results of a fit using the background templates and also one for a 150 GeV/c^2 Higgs, which shows the most significant excess of the mass points we have investigated (140 GeV/c^2 is about equally significant). In this mass range the fit prefers to put Higgs signal into the m_{12} region previously occupied by bqb, with the deficit at the

Figure 31: Number of fitted Higgs events for various Higgs masses in the control samples.

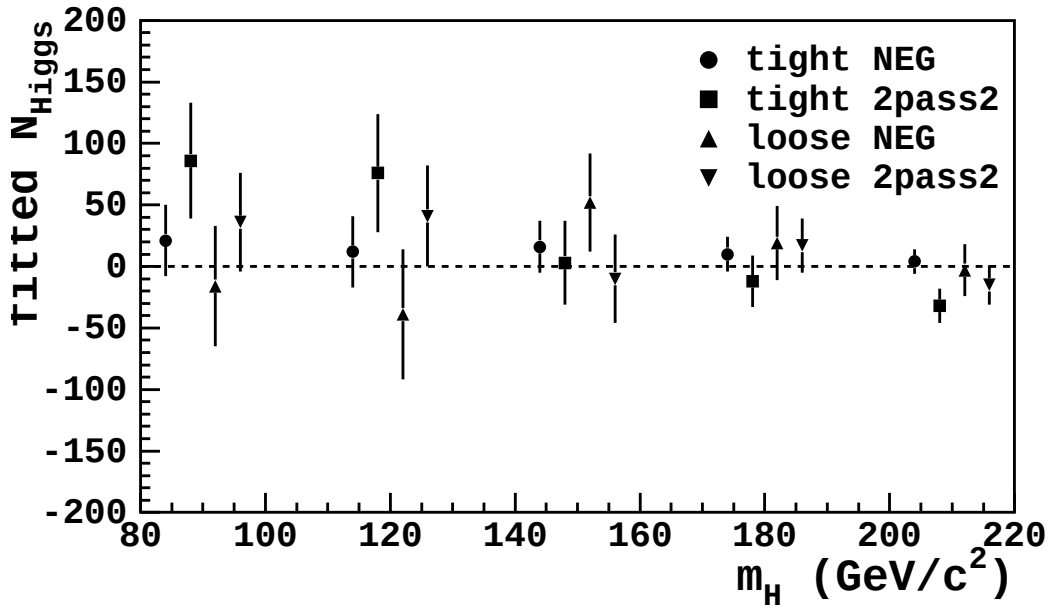


Figure 32: Fit of the +++ data sample using only the QCD background templates.

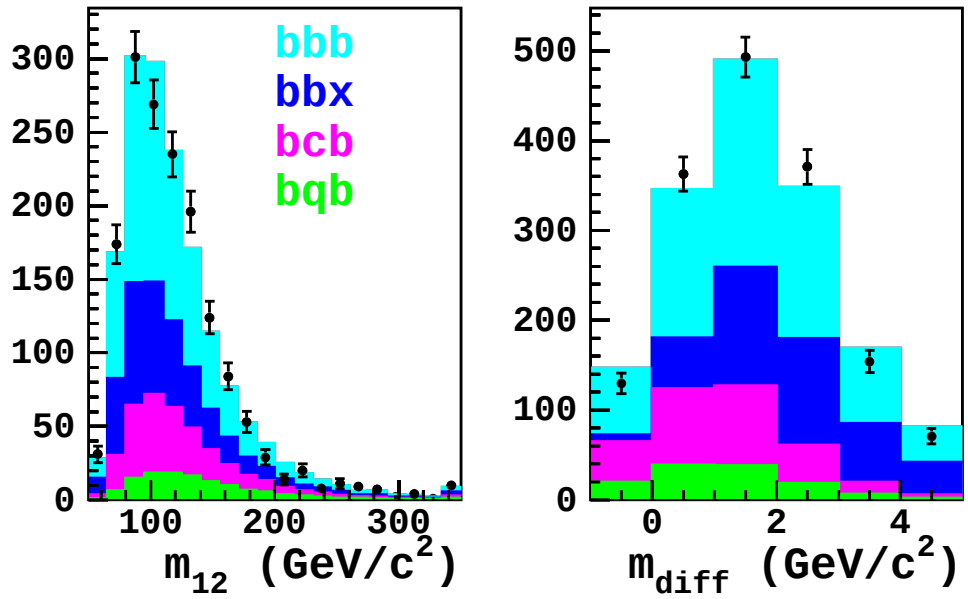


Table 4: Numbers of fitted events (top) and correlation matrix bottom for the background-only fit.

	N_{fit}
bbb	762 ± 131
bbx	415 ± 82
bcb	279 ± 114
bqb	125 ± 82

	bbb	*bbx*	*bcb*	*bqb*
bbb	1	-0.77	-0.56	+0.02
bbx	-0.77	1	+0.24	-0.04
bcb	-0.56	+0.24	1	-0.69
bqb	+0.02	-0.04	-0.69	1

low end of m_{diff} made up by increasing the amount of *bcb*. It is not the case that Higgs and *bqb* are completely interchangeable, however. With *bqb* fixed at the 125 events found in the background-only fit, the signal fit still returns 62 Higgs events.

It bears mentioning that because of the large anticorrelations between some of the background components, the likelihood has a fairly shallow minimum. This results in the fit being somewhat sensitive to the initial values of the parameters. The gradients are well-defined in the sense that the fit always ends up in the same neighborhood, within 10-20 events in the backgrounds and 1-2 events in the Higgs component, however the second derivatives are harder to compute accurately and so the fit uncertainties can vary by larger amounts. The limit-setting procedure uses pseudoexperiments to construct PDFs rather than relying on the curvature of the likelihood, so as long as the number of Higgs signal events returned by the fit is stable at the level observed we do not expect this to affect our results.

8 Sources of Systematic Error

Several sources of systematic uncertainty were considered. These can take two forms, either affecting only the number of Higgs events expected for a given cross section, or changing the shape of the signal or background fit templates.

Uncertainties which affect only the number of signal events expected for a given cross section include the integrated luminosity measurement (±6%) and the *b*-tag and trigger *b*-tag scale factors ($2 \times 3.3 + 4.0 = 10.6\%$, not including the E_T-dependence). Also in this category is the uncertainty due to PDF's. The 20 eigenvectors of the CTEQ6.1M set were used to check this. Only two were found to change the acceptance by more than 1%, 15 (±7.3%) and 16 (±2.9%). Because these results involve the *b* quark PDF, another set CTEQ6HQ was tried which includes heavy quark mass effects, which changed the acceptance by 1% relative to CTEQ6.1M. Summing the contributions from all eigenvectors in quadrature resulted in a total uncertainty due to PDF's of 8%.

Figure 33: Fit of the +++ data sample using the QCD background templates and one for $m_H = 150$ GeV/c^2.

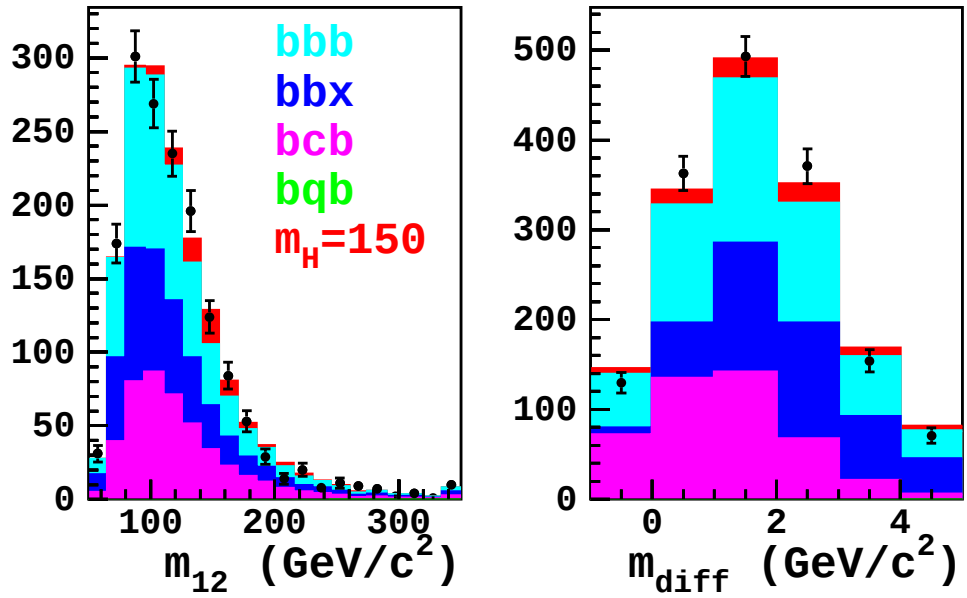


Table 5: Numbers of fitted events (top) and correlation matrix bottom for the background-only fit.

	N_{fit}
bbb	601 ± 130
bbx	453 ± 83
bcb	475 ± 113
bqb	-29 ± 87
Higgs	82 ± 37

	bbb	*bbx*	*bcb*	*bqb*	Higgs
bbb	1	-0.72	-0.54	-0.10	-0.01
bbx	-0.72	1	+0.18	+0.06	-0.16
bcb	-0.54	+0.18	1	-0.55	+0.05
bqb	-0.10	+0.06	-0.55	1	-0.35
Higgs	-0.01	-0.16	+0.05	-0.35	1

The E_T-dependence of the b-tag scale factor was varied according to the fit results given above. The changes in m_{12} are considerably smaller than for JES and can be neglected. However, the acceptance does change by $\pm 1.5\%$ for $m_H = 90$ up to $\pm 13\%$ for $m_H = 210$, and this variation is included. The total uncertainty on the Higgs event expectation including all contributions described above varies from 14.5% for $m_H = 90$ to 19.8% for $m_H = 210$.

The jet energy scale was varied by $\pm 1\sigma$ for the Monte Carlo samples only, and the resulting shifts in m_{12} are shown in Figure 34. Besides changing the shape, the JES also affects the acceptance, by 4% for $m_H = 90$ down to 2% for $m_H = 210$.

The mass of the SecVtx tags was varied by $\pm 3\%$ around the default point of -3%, as suggested by various studies within the high-p_T b-tag group. The effects on m_{diff} from varying m_{tag} in the signal samples are shown in Figure 35. Of course this has no effect on m_{12} or the acceptance.

For the *bbb* and *bcb* background templates, the correction applied to the double-tagged sample based on scaling the amount of gluon splitting was varied, using the fits shown in Figures 14 and 15. The effects on the m_{12} spectra are shown in Figure 36.

For the *bbx* background which is the average of *bbq* and *bbc*, the variation between the two extremes is taken as the systematic uncertainty on this component. The three templates are shown in Figure 37. Because the two bounds represent extreme cases of 100% *bbc* or 100% *bbq*, we treat this interval as containing 95% of the probability rather than 1σ.

The *bqb* template was studied under the effect of changing the tag mass and E_T-dependence of the tag rate applied to the untagged jet. While it is possible to change the number of *bqb* events returned by the fit, these variations had a small effect on the fitted number of Higgs events compared to the uncertainties on the other backgrounds. Therefore, we do not perform any variations of this template when setting limits.

9 Cross Section Limits and MSSM Interpretation

The fitting and limit calculations were performed using the MCLIMIT package [13]. It performs the fitting to either the observed distribution or to pseudoexperiments, and calculates confidence levels.

Pseudoexperiments were generated using the results of the background-only fit in Figure 32. The observed background fractions and errors were used to determine how many of each type of event to generate in each PE. The nuisance parameters were set up to reproduce the anticorrelations as closely as possible, so that the total expected number of events in each pseudoexperiment was the same within 20-30 events. For pseudoexperiments that include Higgs signal, the expected signal fraction was subtracted from the background fractions in order to keep the average number of events constant.

Shape variations in the signal and background templates are incorporated in the pseudoexperiments by providing central templates and $\pm 4\sigma$ variations. The 4σ variations are generated from the 1σ variations shown in the previous Section using the histogram interpolation/extrapolation functionality built into MCLIMIT. When throwing pseudoexperiments, the parameters in question (JES, GS scaling factor, etc.) are generated randomly according to unit gaussians, the templates are interpolated to those values, and then sampled. In

Figure 34: Jet energy scale variations for $m_H = 120$ (top) and 180 (bottom) GeV/c^2.

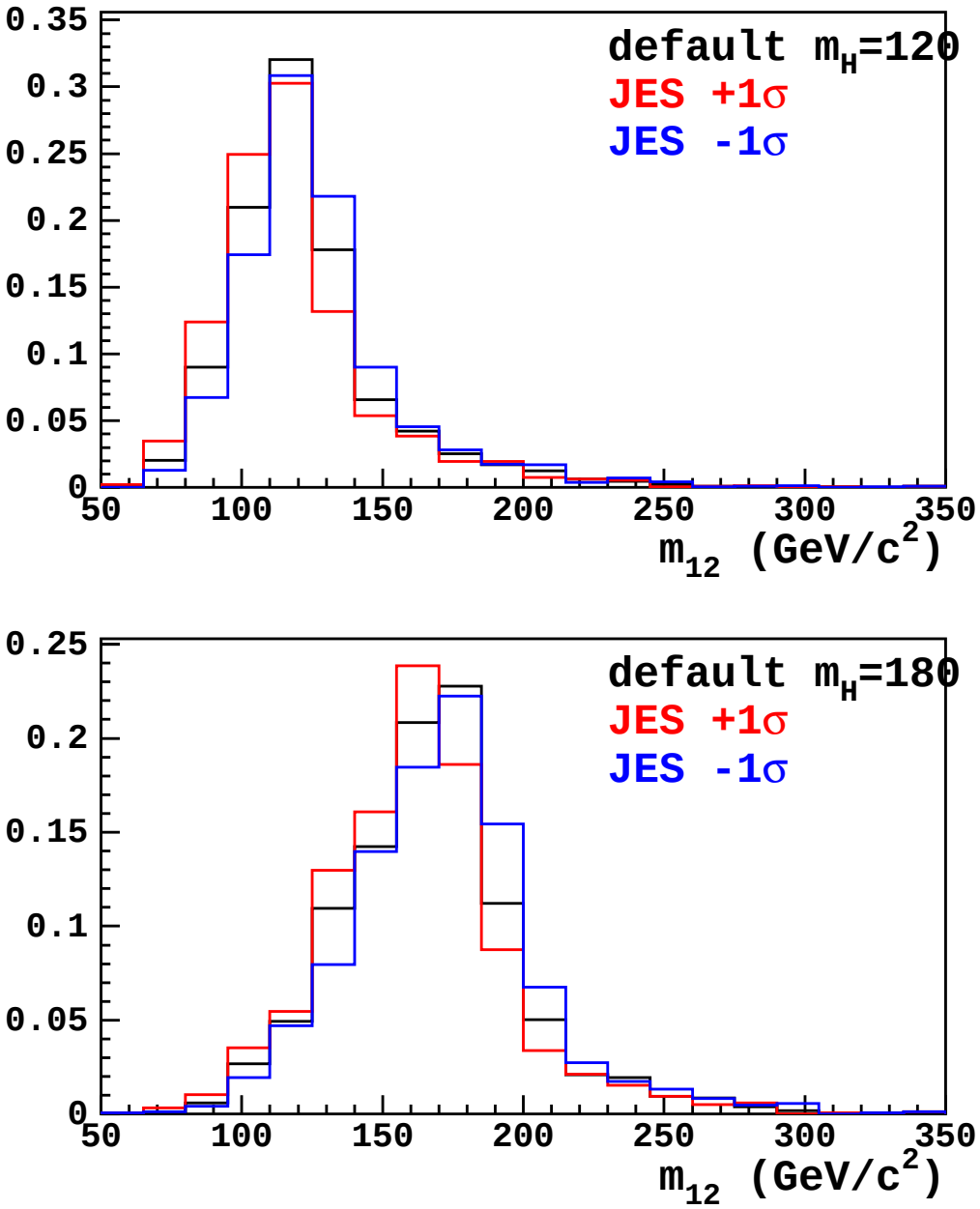


Figure 35: SecVtx tag mass variations for $m_H = 120$ (top) and 180 (bottom) GeV/c^2.

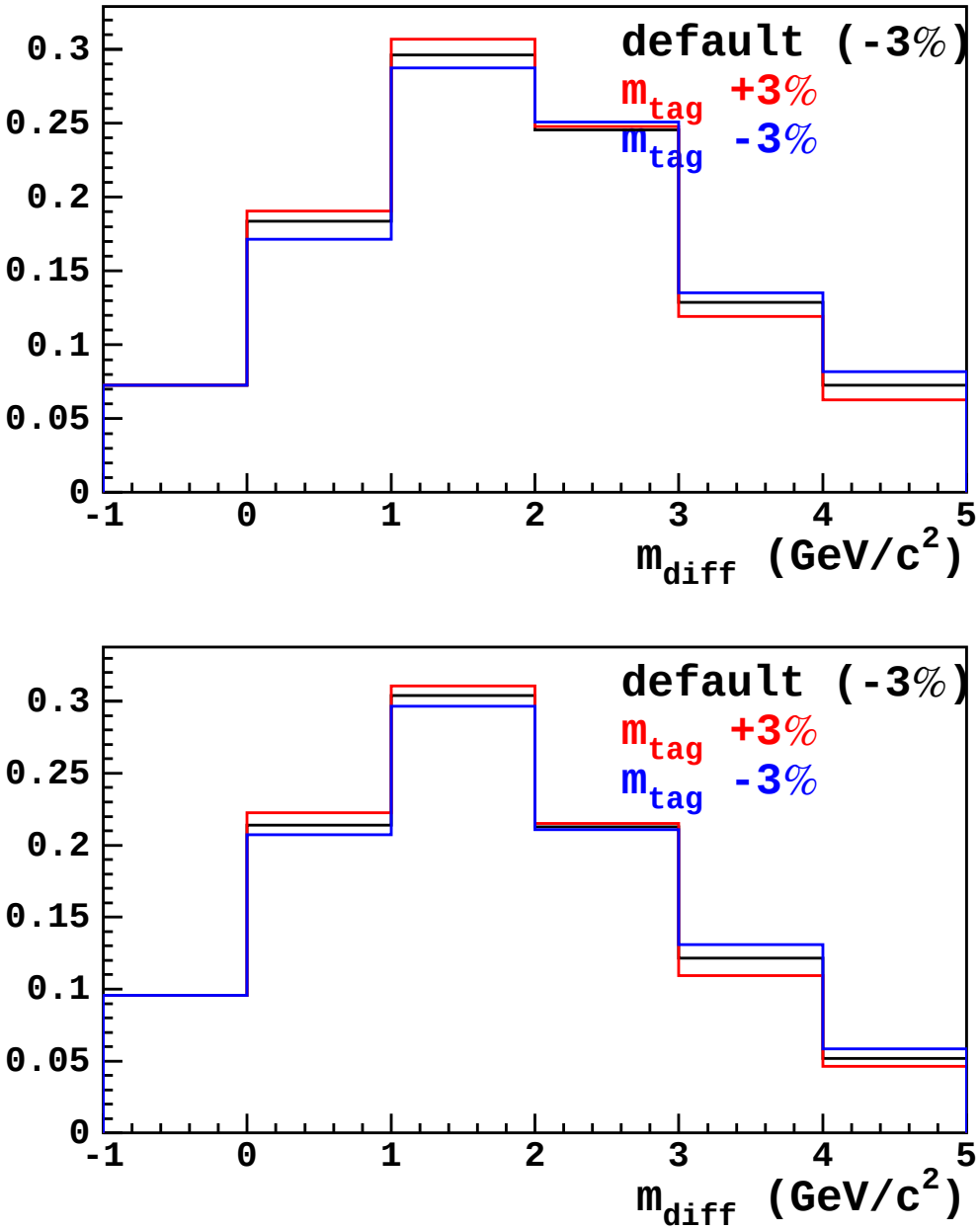


Figure 36: Effects of changing the correction function applied to the double-tagged sample, for *bbb* (top) and *bcb* (bottom).

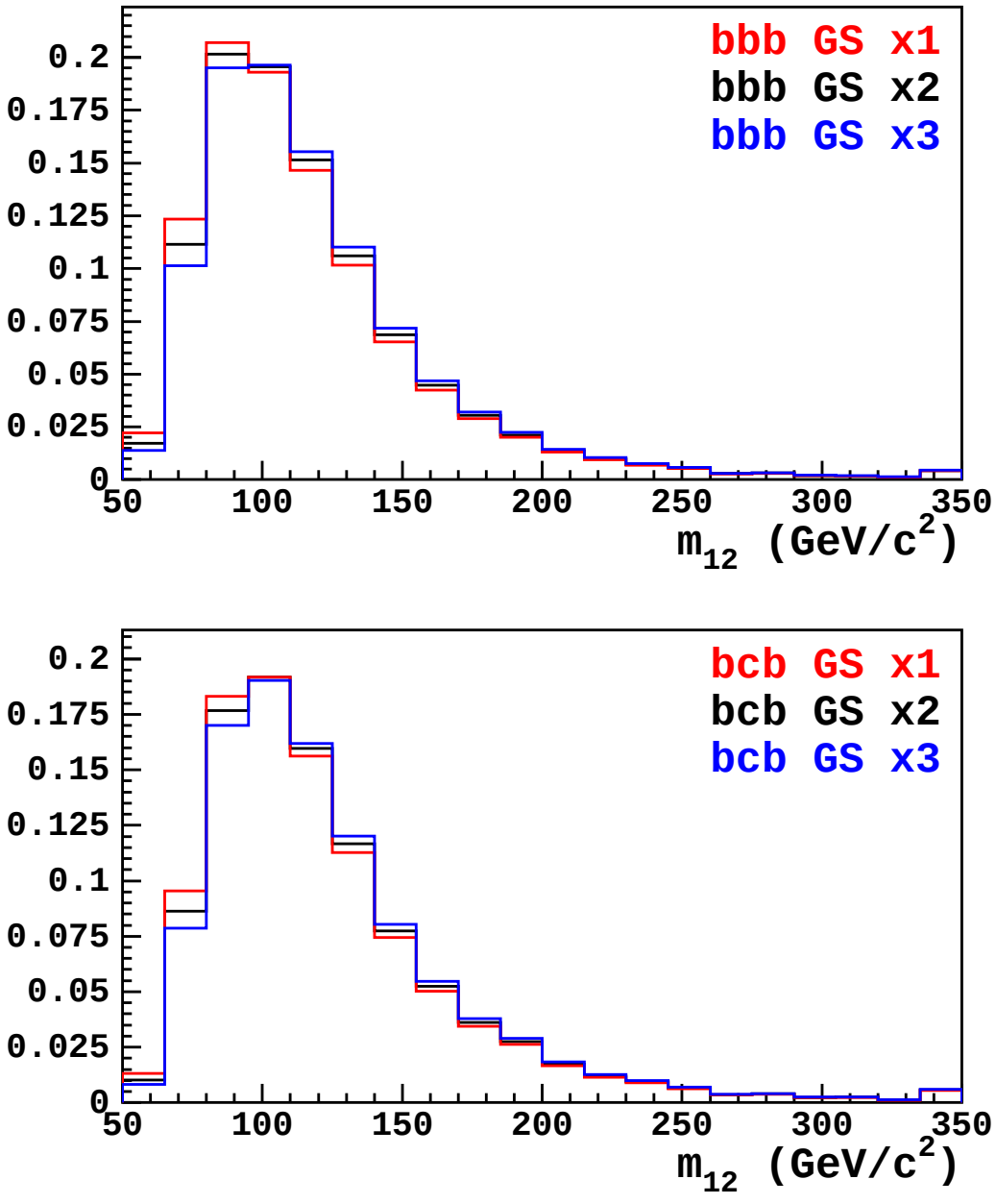
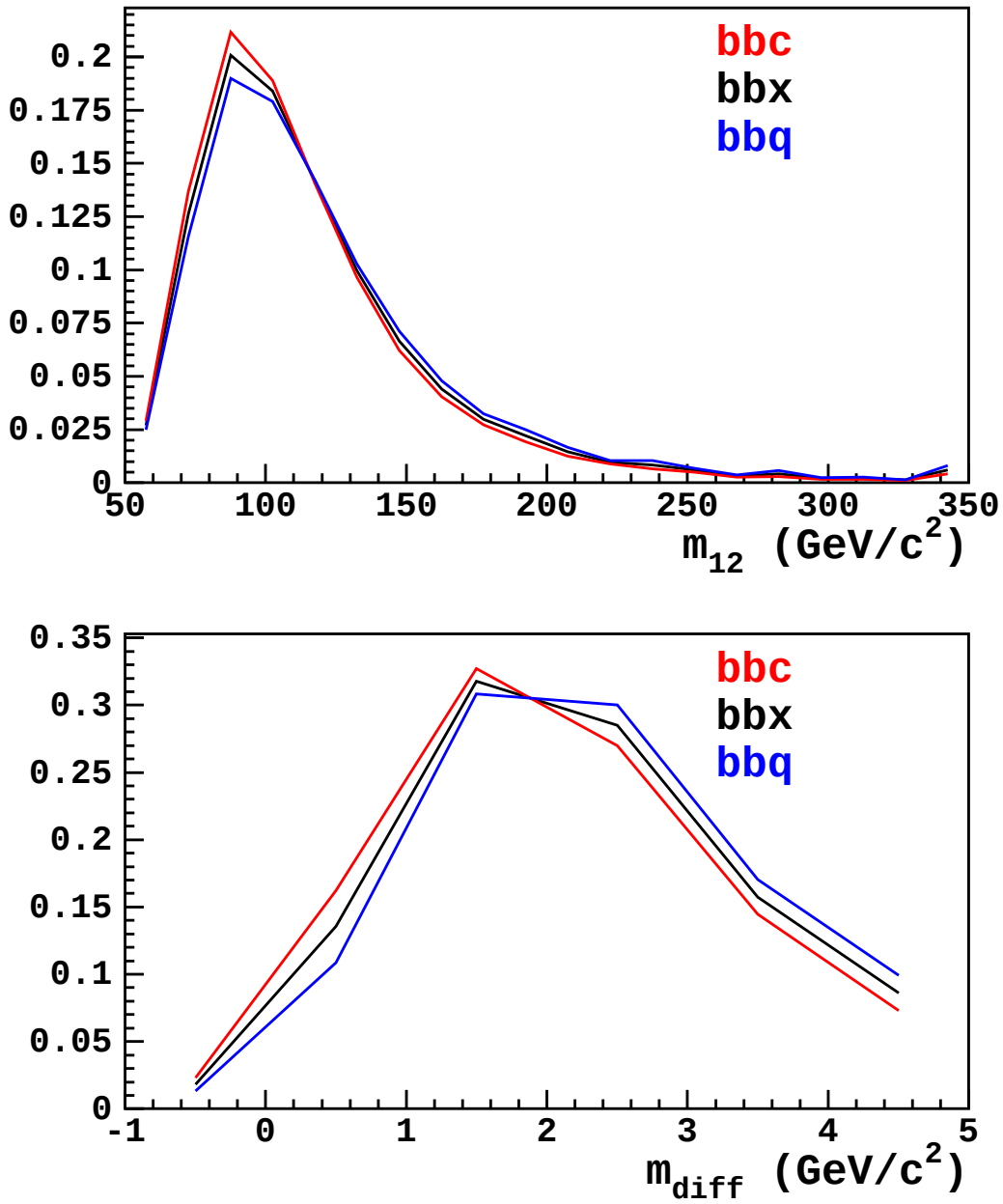


Figure 37: The bbx template and the two extreme cases bbq and bbc, in m_{12} (top) and m_{diff} (bottom).



this way the pseudoexperiments correctly represent the effects of possible variations in these parameters.

The limit-setting procedure uses a test statistic $-2\ln Q = \chi^2(s+b) - \chi^2(b)$, the difference in χ^2 between a fit involving both signal and background templates compared to a fit using the background templates only. For signal present, negative values of $-2\ln Q$ would be expected due to better fits with the signal template included, while for background only the values should be near zero. The confidence level CL_b is defined as one minus the probability of a background-only sample giving a value of Q less than that observed in the data (a "false positive"). The confidence level CL_{s+b} represents the probability that in a sample with a given amount of signal present, one would obtain a value of Q less than that observed in the data sample (a "false exclusion"). The procedure used by MCLIMIT is to construct the parameter $CL_s = CL_{s+b}/CL_b$ and to use it for setting limits. Regularizing by CL_b is thought to prevent exclusions beyond the sensitivity of the experiment, in cases where the data do not look like either the signal or background. Pseudoexperiments with varying levels of signal are generated and used to find the point at which $CL_s = 0.05$, which is considered the 95% CL excluded limit. The expected limit for no signal is calculated using the median value of the test statistic from background-only pseudoexperiments, while the observed limit uses the value of the test statistic obtained from fitting the data. The confidence levels were computed using 25000 pseudoexperiments for each iteration.

For the intermediate points where no Monte Carlo was generated, histogram interpolation was used between the two closest of the 90, 120, 150, 180, and 210 GeV/c samples. To test this, templates for 120 GeV/c^2 were derived from the 90 and 150 GeV/c^2 samples, and the resulting limits were within 5% of the ones obtained directly from the 120 GeV/c^2 sample. The acceptances for each mass point were derived from a fit to the values in Table 3 ($\epsilon = 0.0074445 * \tanh((m_H - 59.020)/77.125)$).

The median expected limits on $\sigma \times BR$ for statistics only with no systematic errors, with only the variations on the background levels and shapes, and with the full systematics including variations on the signal level and shape are shown in Table 6, along with the observed limits. Below 130 GeV/c^2 there is a considerable loss of senstivity due to systematics on the signal, primarily due to the JES variation allowing the signal to slide downwards into regions with much larger background. Above 130 GeV/c^2 this is less of an issue and the signal level variations are more important.

The expected and observed limits for the full systematics case are plotted as a function of the Higgs mass in Figure 38. Also shown are the bands resulting from calculating the expected limits using the $\pm 1\sigma$ and $\pm 2\sigma$ values of the test statistic from background-only pseudoexperiments. The observed limit is almost 2σ high around 140 GeV/c^2.

These limits can be trivially converted into limits on $\tan\beta$ by diving by the standard model cross section times branching ratio (90%) and taking the square root. The results of this are shown in Figure 39. The limits are not very realistic, however, because they do not include the effects of loop corrections which can enhance the cross section by more or less than $\tan^2\beta$ depending upon the SUSY scenario. They also do not include the effects of the Higgs width which can become significant when the down-type couplings are enhanced by such large factors.

Figure 38: Median, 1σ, and 2σ expected cross section limits, and the observed limits versus m_H.

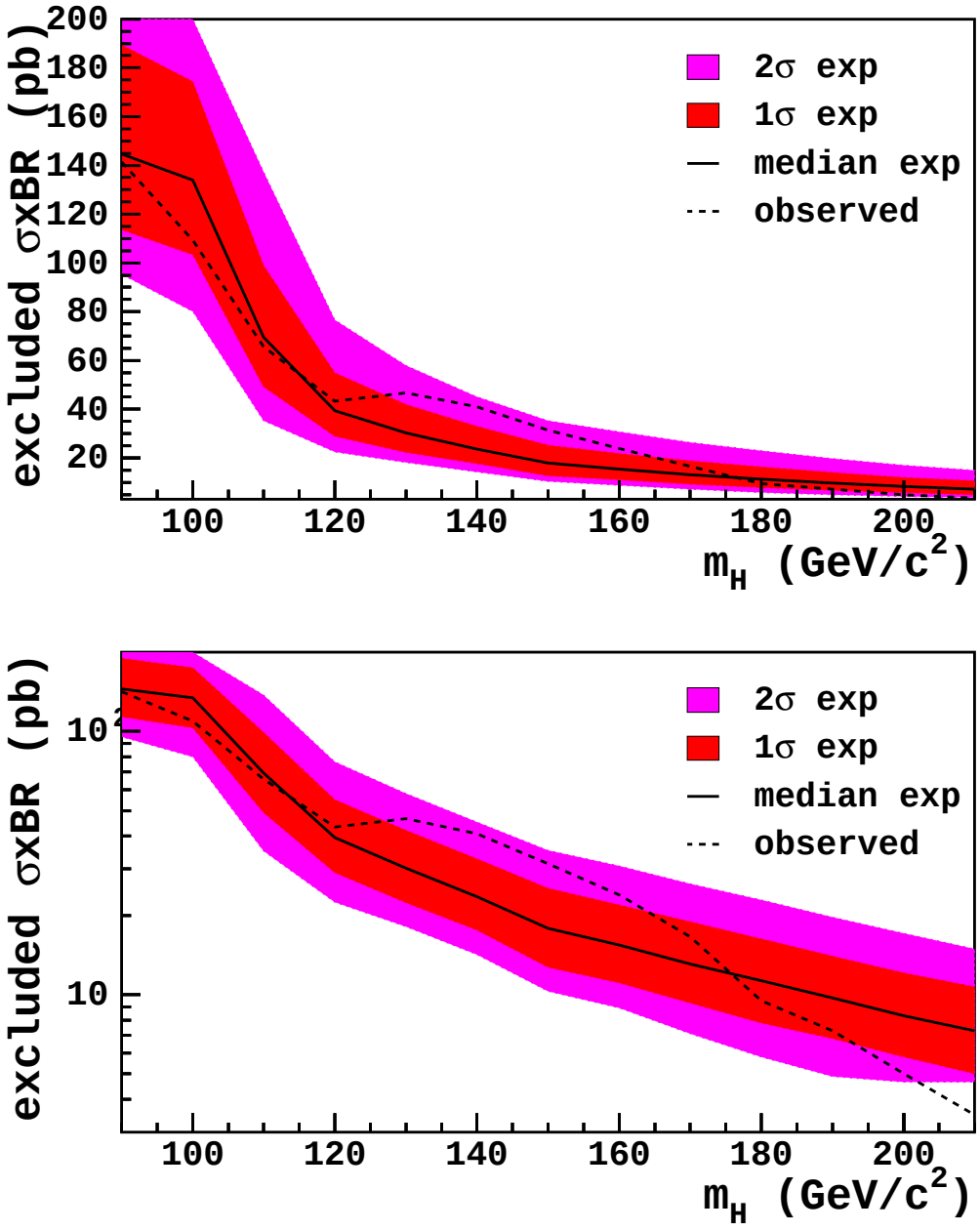
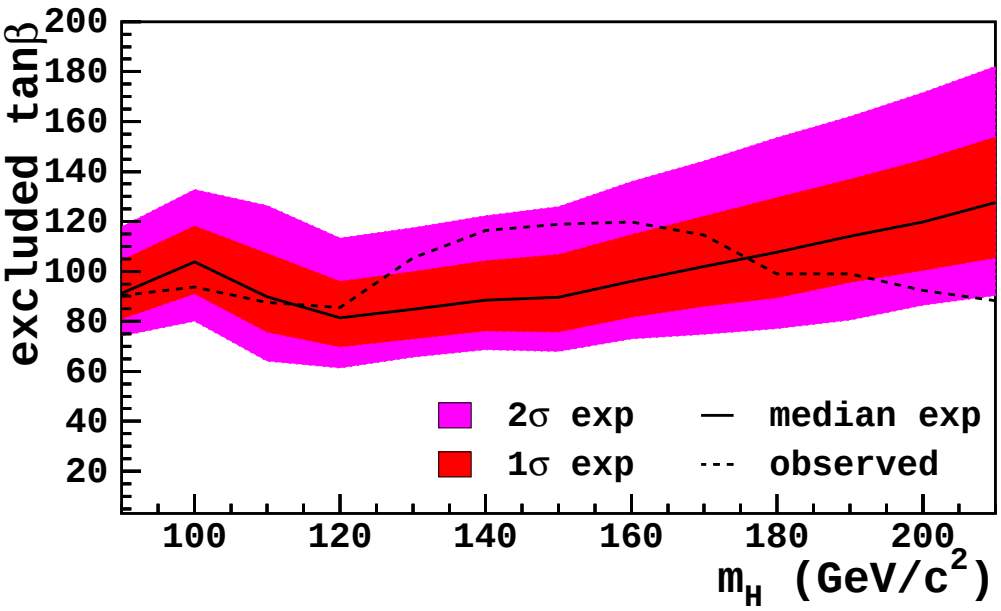


Table 6: Median expected and observed limits on $\sigma \times BR$ for bH production, in pb.

m_H	no systematics	bkgd systematics	full systematics	observed
90	72.1	70.9	144.7	141.6
100	63.3	72.5	133.9	109.3
110	37.9	48.6	69.5	65.7
120	23.4	29.6	39.4	43.4
130	20.3	25.7	30.2	46.6
140	16.5	20.3	23.6	40.9
150	12.7	14.5	17.9	31.4
160	12.1	13.4	15.4	23.9
170	10.7	11.4	13.1	16.6
180	8.7	9.1	11.3	9.5
190	8.3	8.5	9.7	7.3
200	7.3	7.3	8.3	5.0
210	6.0	6.0	7.3	3.5

Figure 39: Median, 1σ, and 2σ expected $\tan\beta$ limits (not including Higgs width effect or loop corrections), and the observed limits versus m_H.



9.1 MSSM Scenario Dependence

Scaling the SM cross section by $\tan^2\beta$ is correct at tree level, however loop effects can modify this relationship and introduce dependence on other parameters of the MSSM. In Ref. [21] an approximate expression for the cross section times branching ratio is given as:

$$\sigma(b\bar{b}\phi) \times BR(A \to b\bar{b}) \simeq 2\sigma(b\bar{b}\phi)_{SM} \frac{\tan^2\beta}{(1+\Delta_b)^2} \times \frac{9}{(1+\Delta_b)^2+9} \quad (1)$$

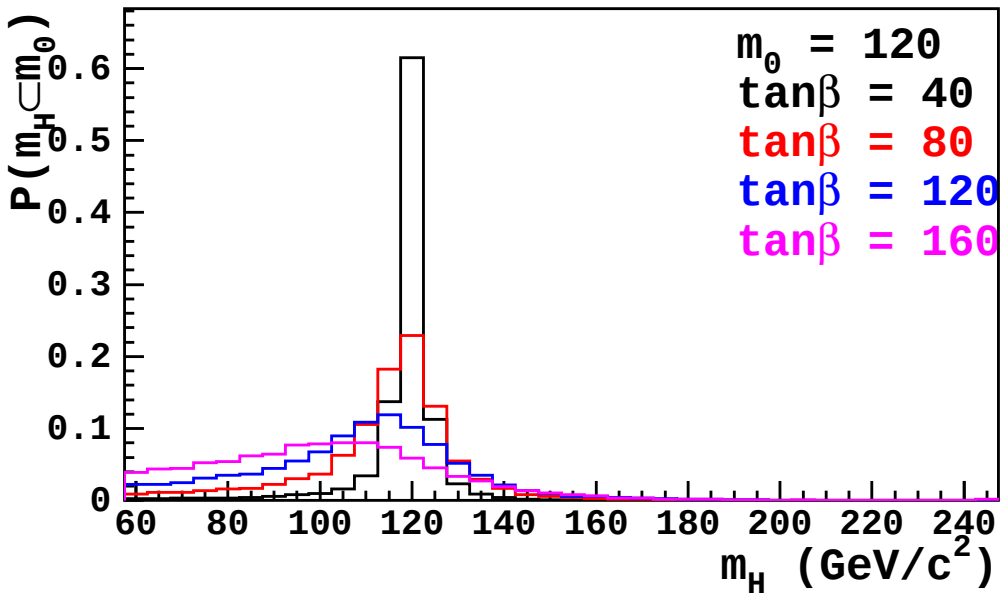
where ϕ is a Higgs boson (either the SM variety or one of $h/H/A$), $\sigma(b\bar{b}\phi)_{SM}$ is the SM cross section, the factor of 2 comes from the degeneracy of A with either h or H, and the loop effects are incorporated into the Δ_b parameter. For our purposes it is important only to note that Δ_b is proportional to the product of $\tan\beta$ and the Higgsino mass parameter μ. Sample values of Δ_b given in Ref. [21] are -0.21 for the m_H^{max} scenario and -0.1 for the no-mixing scenario (at $\mu = -200$ GeV and $\tan\beta = 50$). It is apparent that negative values of μ and hence of Δ_b will increase the MSSM Higgs yield at fixed $\tan\beta$ above the tree level values and result in stronger limits on $\tan\beta$, while scenarios with μ positive will produce the opposite effect. Using Eqn. 1 we can predict the Higgs yield for any value of $\tan\beta$ and Δ_b and therefore derive limits in any desired scenario.

9.2 Higgs Width

The limits shown in Figures 38 and 39 apply only to narrow Higgs like those in the standard model. If the cross section is increased by scaling the $b\bar{b}H$ coupling, as happens in the MSSM, then the width of the Higgs will increase as well. In order to account for this we used PYTHIA to produce m_H spectra for various values of the Higgs pole mass, $\tan\beta$, and Δ_b. Process 3 was used ($b\bar{b} \to H$) for this purpose, as process 32 which was used for the template samples does not properly take into account the $\hat{s}$-dependence of the width in the Breit-Wigner. The couplings to down-type quarks were scaled by $\tan\beta/(1+\Delta_b)$, to leptons (i.e. taus) by $\tan\beta$ (no loop effects here), and to up-type quarks and W/Z by zero. The initial state was forced to $b\bar{b}$ and no particular decay mode was selected in order for PYTHIA to report the full cross section. At least one of the b's accompanying the Higgs was required to have $p_T > 15$ GeV/c, $|\eta| < 2.5$, just like for the standard MC samples. Because the acceptance drops to zero, no events were generated below $m_H = 57.5$ Gev/c^2 in order not to rely on the PYTHIA cross section calculation in that region. Some distributions of m_H for various values of λ and pole masses are shown in Figure 40

Changing the width of the Higgs changes the total cross section as a function of the pole mass. The spectra derived from PYTHIA are divided by the PYTHIA cross section estimate for a Higgs with SM couplings and by $\tan^2\beta/(1+\Delta_b)^2$, to produce an enhancement factor. This factor ranges from 0.95-0.75 for pole mass of 90 GeV/c^2 to 1.05-1.40 for 180 GeV/c^2, for $\tan\beta$ from 40-150. The factor is below 1 for low pole mass because of the cutoff at 57.5 GeV/c^2. A graphical depiction is shown in Figure 41. This information is needed when computing the expected number of events for a given Higgs mass and $\tan\beta$ value in the limits calculator.

Figure 40: Distributions of m_H for varying $\tan\beta$ and $\Delta_b = 0$, for Higgs pole masses of 120 (top) and 180 GeV/c^2 (bottom). The area of each histogram indicates the cross section relative to that for a narrow Higgs.



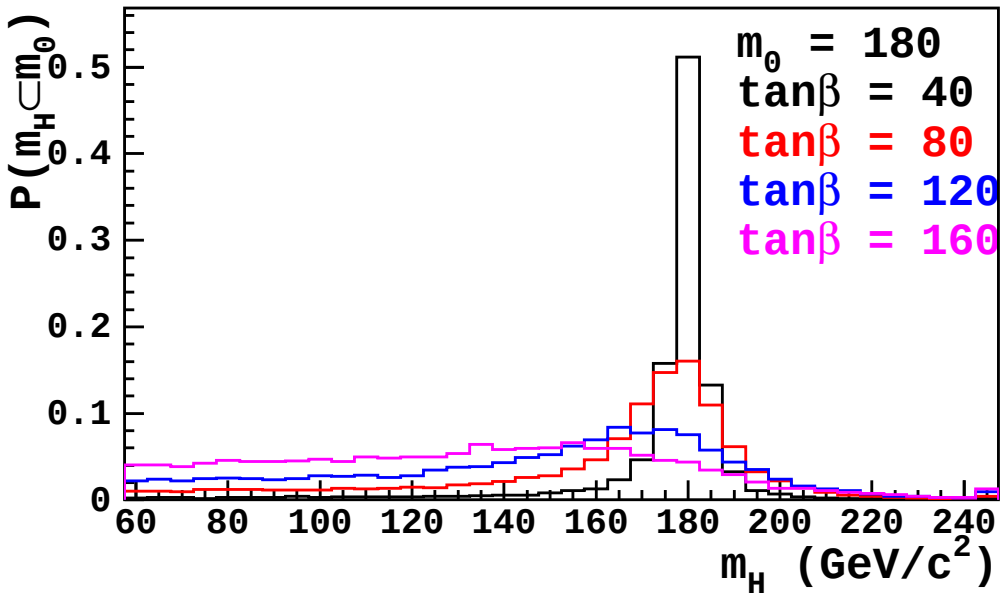
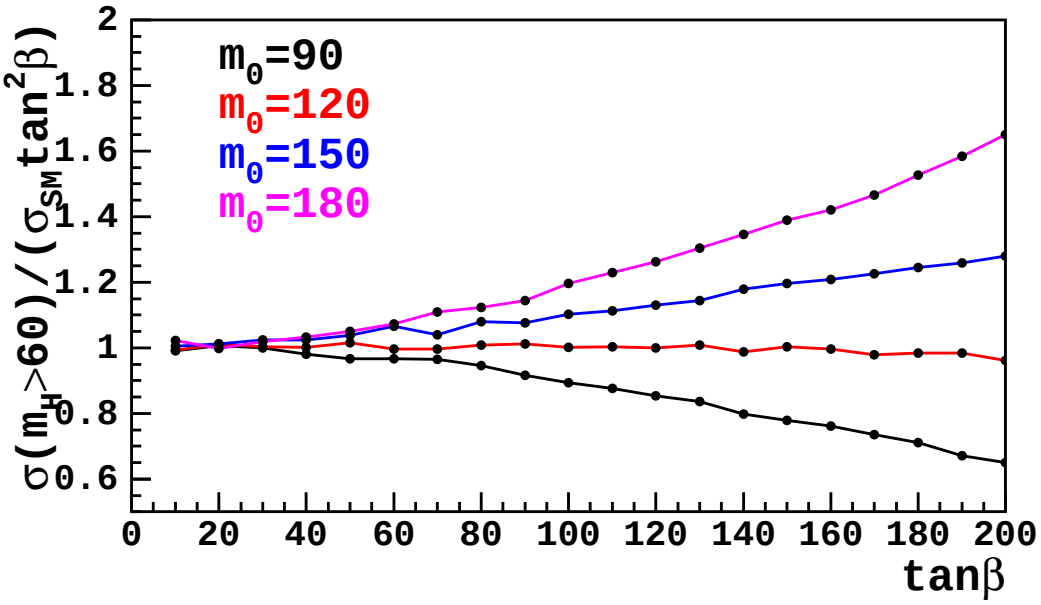


Figure 41: Ratio of the cross section above 57.5 GeV/c^2 relative to the no-width case for various values of the Higgs pole mass and $\tan\beta$ for $\Delta_b = 0$.



A $\tan\beta$-dependent effective acceptance is computed by convoluting the acceptance parameterization given earlier against the scaled spectra. The smaller acceptance for lower values of m_H tends to cancel out whatever enhancement of the cross section may be present, with the effective acceptance ranging from 0.95-0.61 (pole mass 90) to about 1.0 independent of $\tan\beta$ for pole mass 180. The event yield for a given pole mass and $\tan\beta$ can then be computed from the effective acceptance, the NLO SM cross section for bH production from FeynHiggs [19], the cross section ratio from Figure 41, the luminosity, and Eqn. 1.

Fit templates as a function of $\tan\beta$ were constructed by combining the narrow-width templates, weighted by the scaled m_H spectra obtained from PYTHIA and by the acceptance parametrization. Some examples are shown in Figure 42.

Setting limits on $\tan\beta$ requires assuming a SM cross section, so we must account for the PDF and scale dependence of the NLO calculations used to predict it. These uncertainties were taken from Figure 14 of Ref. [4]. The PDF uncertainty ranges from 10% at $m_H = 90$ up to 30% at $m_H = 210$, and for the scale dependence a constant value of 8% was used. The uncertainty on the cross section due to PDFs is anti-correlated with the acceptance PDF uncertainty, so that the uncertainty on the event yield due to PDFs is reduced compared to the uncertainty on the cross section itself.

Figure 42: Distributions of m_{12} for varying $\tan\beta$ and $\Delta_b = 0$, for Higgs pole masses of 120 (top), and 180 GeV/c^2 (bottom). The area of each histogram shows the effective event yield relative to the narrow Higgs case.

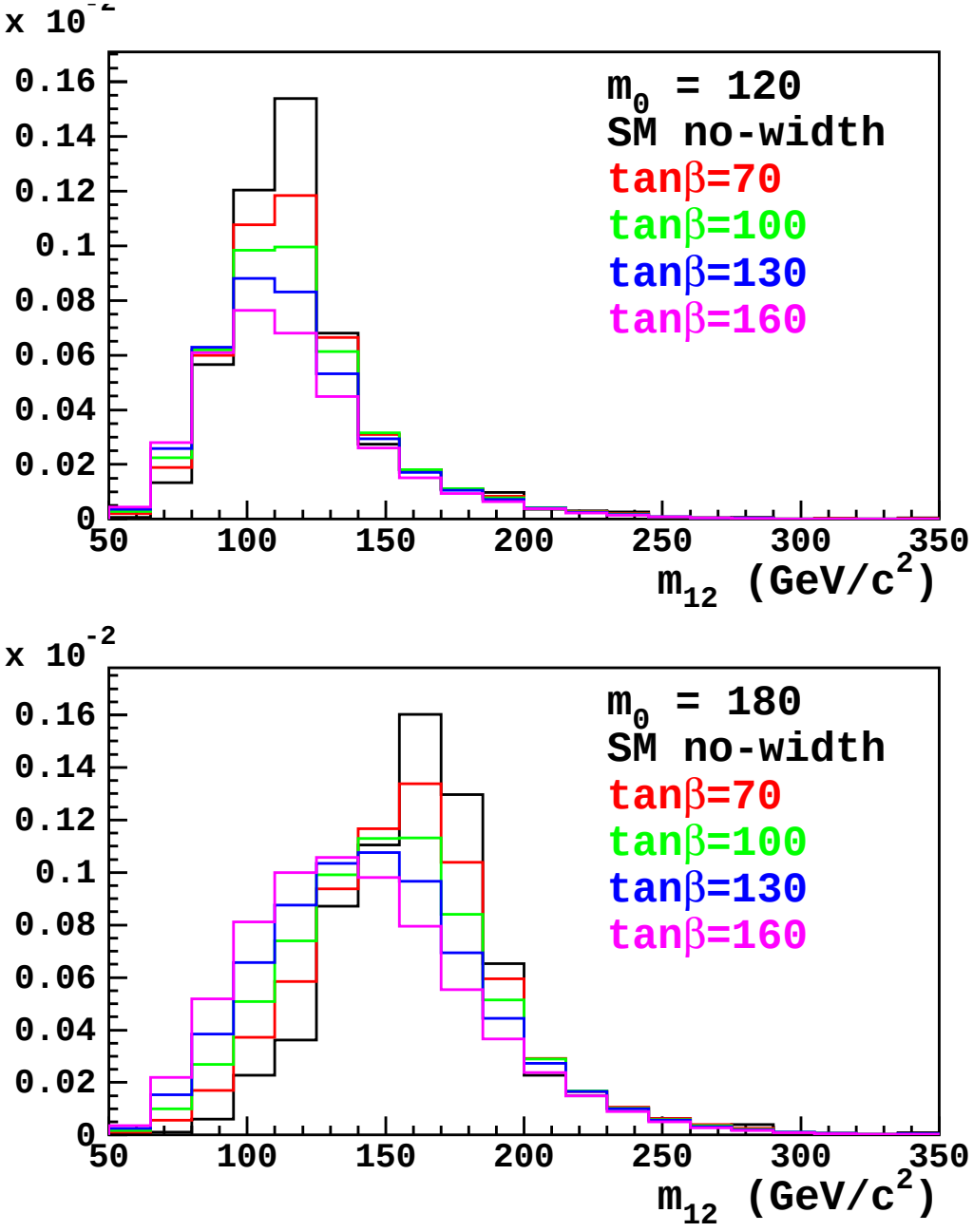
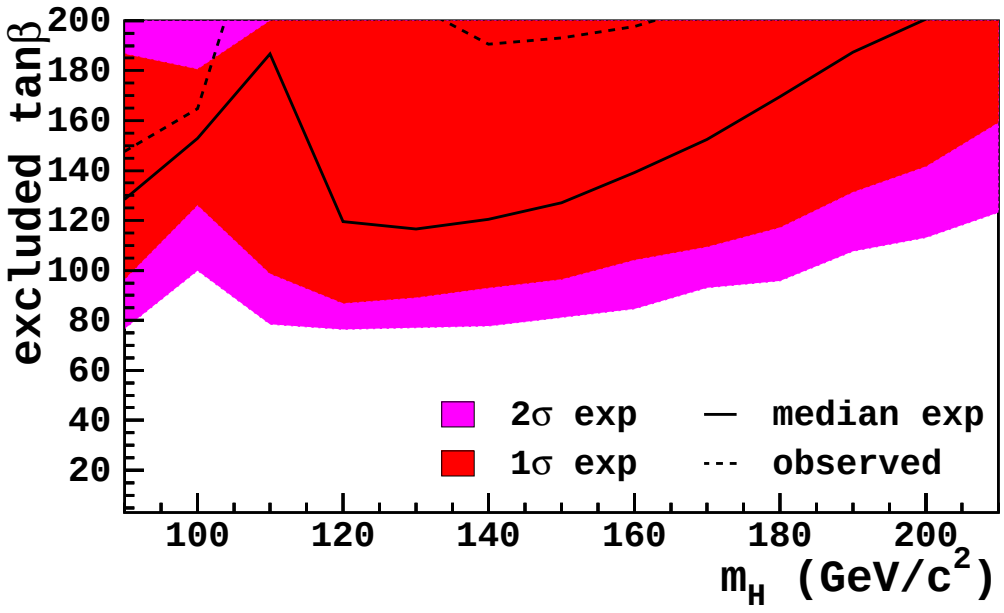


Figure 43: Median, 1σ, and 2σ expected limits, and the observed limits versus m_H, including the Higgs width and for $\Delta_b = 0$.



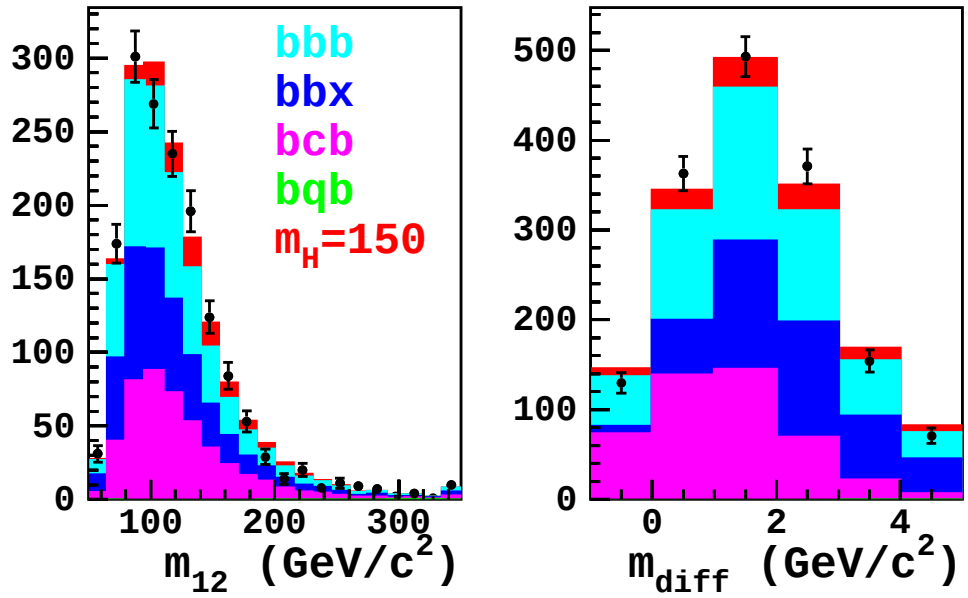
9.3 Full Limits with Width Included

We scan in $\tan\beta$ from 40 to 200 in steps of 5 and calculate CL_s at each point, and exclude regions with $CL_s > 0.05$. We chose 200 as the cutoff for $\tan\beta$ as the highest-possible limit that could be theoretically or experimentally interesting give the existing $\tau\tau$ results. The limits obtained are shown in Figure 43 for $\Delta_b = 0$. The limits get weaker in a strongly $\tan\beta$-dependent way, so that for example the -2σ contour hardly moves at all while the $+2\sigma$ one runs away to high values. This is because as $\tan\beta$ increases, the growing width spreads the events out over a larger region of m_{12}, reducing the fit power, and also tends to reduce the number of expected events due to the cross section lineshape extending downwards into regions with low or no acceptance. Looking towards the future, the good news is that you might expect the limits to increase by faster than $\sqrt{N}$ with more data because of this effect.

To illustrate why the limits worsen so quickly at high $\tan\beta$, Figure 44 shows the result of a simple fit in the data similar to what was shown in Figure 33, except the signal template includes the width effects. As shown in Figure 42, the net effect is for the signal to spread out into more bins, and also to shift towards lower values of m_{12} where the backgrounds are larger. Both of these effects reduce the statistical sensitivity of the search and require adding more signal to reach CL_s of 0.05, however that additional signal further broadens and shifts the m_{12} distribution, and so on.

Limits were also generated for the m_H^{max} scenario with $\mu = -200$ GeV and are shown in Figure 45. Because of the relatively large and negative values of Δ_b, the $\tan\beta$ limits are

Figure 44: Fit of the +++ data sample using the QCD background templates and one for $m_H = 150$ GeV/c^2 with $\tan\beta = 150$ and $\Delta_b = 0$.



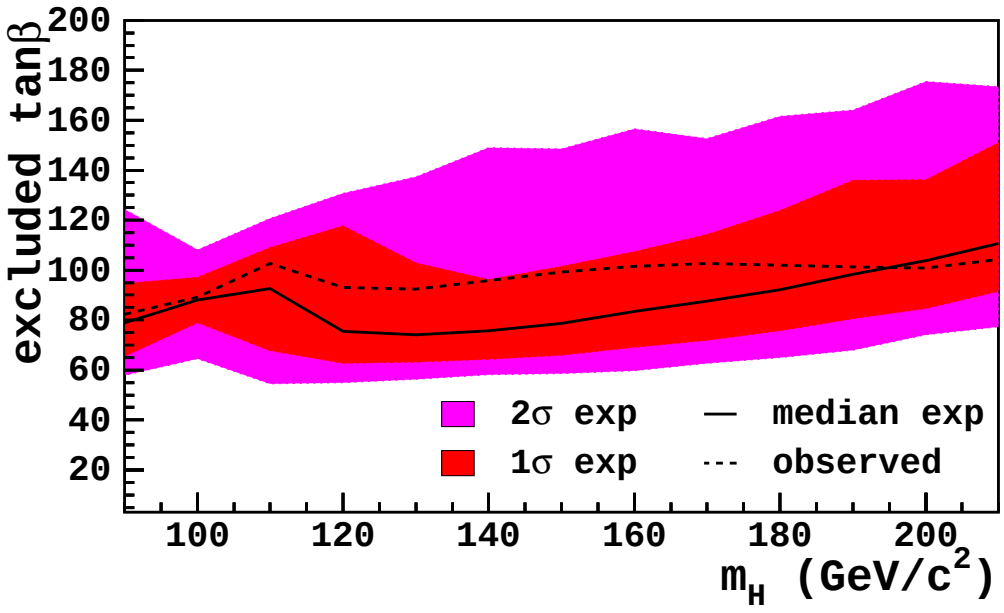
much tighter in this case. The improvement is especially good for higher values of $\tan\beta$, since Δ_b becomes even more negative as $\tan\beta$ increases.

10 Conclusion

A search for Higgs bosons produced in association with b-quarks was performed in 980 pb^{-1} of data. This process could be visible in supersymmetric models with high values of $\tan\beta$. The variable used was the mass of the two leading jets in triple-tagged events, with additional information from the SECVTX tag masses included to improve the background modeling.

The observed limits are within 2σ of expectations over the mass region from 90 to 210 GeV/c^2, with the largest excess occuring around 140 GeV/c^2. The results have been interpreted in two MSSM scenarios. In the case where loop effects are small, we find that the growth of the Higgs width as the couplings are enhanced permit only very weak limits on $\tan\beta$. In the m_H^{max} scenario with μ negative, the enhanced production through loop effects allows exclusion of $\tan\beta$ less than around 100.

Figure 45: Median, 1σ, and 2σ expected limits, and the observed limits versus m_H, including the Higgs width, for the m_H^{max} scenario with $\mu = -200$ GeV.



References

[1] M. Carena, S. Heinemeyer, C. E. M. Wagner, and G. Weiglein, "Suggestions for benchmark scenarios for MSSM Higgs boson searches at hadron colliders," *Eur. Phys. J.* **C26** (2003) 601–607, hep-ph/0202167.

[2] F. Maltoni. http://maltoni.home.cern.ch/maltoni/TeV4LHC/.

[3] R. V. Harlander and W. B. Kilgore, "Higgs boson production in bottom quark fusion at next-to- next-to-leading order," *Phys. Rev.* **D68** (2003) 013001, hep-ph/0304035.

[4] S. Dawson, C. B. Jackson, L. Reina, and D. Wackeroth, "Higgs production in association with bottom quarks at hadron colliders," hep-ph/0508293.

[5] S. Dawson, C. B. Jackson, L. Reina, and D. Wackeroth, "Hadronic Higgs Production with Heavy Quarks at the Tevatron and the LHC," hep-ph/0603112.

[6] **DØ** Collaboration, V. M. Abazov *et al.*, "Search for neutral supersymmetric Higgs bosons in multijet events at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. Lett.* **95** (2005) 151801, hep-ex/0504018.

[7] **DØ** Collaboration, "Search for Neutral Higgs Bosons at high $\tan\beta$ in Multijet Events," DONote 5503-CONF. http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/HIGGS/H23/H23.pdf.

[8] **CDF** Collaboration, A. Abulencia *et al.*, "Search for neutral MSSM Higgs bosons decaying to tau pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. Lett.* **96** (2006) 011802, hep-ex/0508051.

[9] A. Anastassov *et al.*, "Search for Neutral MSSM Higgs Boson(s) in the tau tau Decay Channel," CDF/8639.

[10] **DØ** Collaboration, V. M. Abazov *et al.*, "Search for neutral Higgs bosons decaying to τ pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev. Lett.* **97** (2006) 121802, hep-ex/0605009.

[11] **DØ** Collaboration, "Search for Neutral Higgs Boson Production in the Decay $h \to \tau_\mu \tau$ with the DØDetector at $\sqrt{s} = 1.96$ TeV," DØNote 5331-CONF. http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/HIGGS/H29/H29.pdf.

[12] **DØ** Collaboration, "Search for Neutral Higgs Bosons at high $\tan\beta$ in the $b(h/H/A) \to b\tau\tau$ Channel," DØNote XXXX-CONF. http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/HIGGS/H24/H24.pdf.

[13] T. Junk. http://www.hep.uiuc.edu/home/trj/cdfstats/mclimit_csm1/index.html.

[14] J. Donini *et al.*, "Measurement of the energy scale of b-jets using the $Z \to b\bar{b}$ signal," CDF/8746.

[15] D. Amidei and T. Wright, "Trigger and b-Tag Efficiency Measurements for the HIGH_PT_BJET Dataset," CDF/8060.

[16] **CDF** Collaboration, D. Acosta *et al.*, "Measurements of $b\bar{b}$ azimuthal production correlations in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV," *Phys. Rev.* **D71** (2005) 092001, hep-ex/0412006.

[17] S. Vallecorsa, "Measurement of the $b\bar{b}$ cross section using a SVT selected sample: comparison to Monte Carlo," CDF/8769.

[18] W. Sakumoto, "Event |Zvtx| <60 Cut Acceptance for Run II (to Sep05)," CDF/7935.

[19] S. Heinemeyer *et al.* http://www.feynhiggs.de/.

[20] O. Gonzalez and C. Rott, "Uncertainties due to the PDFs for the gluino-sbottom search," CDF/7051.

[21] M. Carena, S. Heinemeyer, C. E. M. Wagner, and G. Weiglein, "MSSM Higgs boson searches at the Tevatron and the LHC: Impact of different benchmark scenarios," *Eur. Phys. J.* **C45** (2006) 797–814, hep-ph/0511023.