Available on the CERN CDS information server **CMS PAS HIG-12-026**

CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2012/07/06

Search for a Higgs boson produced in association with b quarks and decaying into a b-quark pair

The CMS Collaboration

Abstract

A search for a neutral Higgs boson produced in association with at least one additional b quark and decaying into a pair of b quarks, is performed in a data sample corresponding to an integrated luminosity of 4.0 fb^{-1}, recorded by the CMS detector in proton-proton collisions at the LHC with a center-of-mass energy of 7 TeV. This search is particularly sensitive to Higgs bosons in MSSM scenarios with large values of $\tan\beta$. Upper limits on the $pp \to bH + X$, $H \to b\bar{b}$ cross section are derived, and interpreted in terms of bounds in the MSSM (M_A, $\tan\beta$) parameter space.

1 Introduction

A search is performed for a Higgs boson decaying into b quarks in multi b jet final states. In the Standard Model (SM), a light Higgs boson decays dominantly to b quarks. Due to the large background from QCD processes with b-quark final states the expected sensitivity for a SM Higgs signal in the current data set is still low.

In the Minimal Supersymmetric Model (MSSM), two scalar Higgs doublets are proposed and the symmetry is spontaneously broken twice giving three neutral, $\phi = h, H$ and A, and two charged, H^+ and H^-, Higgs bosons. The masses of M_A and either M_h or M_H are nearly degenerate within the experimental resolution. At tree level, two parameters, conventionally chosen to be the mass of the pseudoscalar Higgs M_A and the ratio of the two vacuum expectation values $\tan\beta = v_1/v_2$, are needed to define the Higgs sector in the MSSM.

For relatively large values of $\tan\beta$, the Higgs couplings to u-type particles are suppressed while the couplings to d-type particles are enhanced by a factor $\tan\beta$, relative to the SM. Therefore, in the MSSM, the combined cross section of Higgs boson production in association with b quarks is enhanced by a factor $\approx 2\tan^2\beta$. Moreover, the decay into b quarks has a very high branching fraction ($\approx 90\%$), even at large values of the Higgs mass.

Similar searches for heavy resonances decaying into, and produced in association with, b quarks have already been performed by the CDF [1] and D0 [2] experiments at the Tevatron collider. Excesses of $\sim 2\sigma$ in the observed limits with respect to the expectations from SM background have been reported by both experiments for a resonance in the $100 - 150$ GeV$/c^2$ mass range.

In this note we describe a search for a resonance decaying into b quarks produced in association with at least one more b quark. The predominant background is the QCD production of heavy-flavour multi-jet events, with either three b jets, or two b jets plus a third jet, charm or light flavour mis-identified as a b jet. Specialised triggers that exploit online algorithms for the identification of b jets are required to tackle the large hadronic interaction rate at the LHC.

1.1 Analysis strategy

The search is performed in a sample with events with at least three jets tagged as b jets ("triple-btag" sample). A significant excess in the di-jet mass distribution (M_{12}) of the two leading jets could be an evidence of a signal.

The main background to the signal topology pursued in this analysis arises from the heavy flavour multi-jet QCD, whose production rate prediction involves complex calculations, receiving individual or combined contributions of high-order flavour creation, flavour excitation and gluon splitting, and contains large uncertainties. In order to avoid the theoretical complications and technical limitations of predicting this background with good accuracy, a method to estimate its contribution directly from data, similar to the procedure employed by the CDF Collaboration [1], has been pursued. The other backgrounds, $t\bar{t}$ and $Z + jets$, give significantly smaller contributions.

This data-driven approach relies on the fact that the vast majority of the events in the triple-btag sample contain at least two real b jets. From Monte Carlo (MC) predictions this is the case in about 98% of the events. This fact is important because it implies a reduction of possible flavour combinations among the jets in background events.

It also allows for the backgrounds to be derived directly from a sample with two b tags ("double-btag" sample) that is selected with the same triggers as the signal, and in this way kinematical

biases due to the trigger are automatically taken into account. A third untagged jet is assumed to be either light, c or b jet. Templates reflecting the different flavour compositions of the background are constructed for two variables, the di-jet mass M_{12} and the EventBTag variable (described later in this note), that allow distinction of kinematics and heavy flavour content of the events between the background contributions and also with the Higgs signal.

The size of the background contributions and a possible Higgs signal yield are obtained from a two-dimensional fit of the data, in the space of the M_{12} and the EventBTag variables.

2 The CMS Detector and Simulated Samples

The central feature of the Compact Muon Solenoid (CMS) apparatus is a superconducting solenoid of 6 m internal diameter, providing a field of 3.8 T. Within the field volume are a silicon pixel and strip tracker, a crystal electromagnetic calorimeter (ECAL) and a brass/scintillator hadron calorimeter (HCAL). The ECAL has an energy resolution of better than 0.5% for unconverted photons with transverse energies above 100 GeV. The HCAL, when combined with the ECAL, measures jets with a resolution $\Delta E/E \approx 100\%/\sqrt{E\,[\mathrm{GeV}]} \oplus 5\%$. Within each tower, the energy deposits in ECAL and HCAL cells are summed to define the calorimeter tower energies, subsequently used to provide the energies and directions of hadronic jets. Muons are measured in gas-ionization detectors embedded in the steel return yoke. Extensive forward calorimetry complements the coverage provided by the barrel and endcap detectors. A more detailed description can be found in Ref. [3].

Simulated samples of signal and backgrounds were produced using various event generators, with the CMS detector response modelled in extensive detail with `GEANT4` [4], and including pileup (PU) events. The Higgs signal samples were produced with `Pythia6` [5], $t\bar{t}$ background with `MadGraph` [6], and QCD background with `Pythia6` and `Alpgen` [7]. The simulated samples were reweighted to represent the PU distribution measured in the data.

3 Trigger and Event Reconstruction

The large hadronic interaction rate at the LHC poses a major challenge for triggering in the context of this analysis. Events are accepted if either two or three jets are centrally produced and have transverse momenta (p_T) above certain thresholds. Depending on the data-taking period, the p_T threshold for the leading jet was 46 or 60 GeV/c, and the one for the second leading jet 38 or 53 GeV/c, respectively. In part of the data sample, the presence of a third jet with $p_T > 20$ GeV/c was required to keep the trigger rates low as the instantaneous luminosity increased.

In order to further reduce the trigger rate, b-tag requirements are applied on the jets at the trigger level. Among the first four leading jets with $p_T > 20$ GeV/c, at least two must satisfy the following b-tag criteria. The first b tag step is based on tracks reconstructed only from pixel detector hits and used to reconstruct the pp primary vertex. If more than one primary vertex is found, the one with highest p_T^2-sum of its associated tracks is considered as the vertex originated from the hard interaction. The tracks are associated to a jet if they lie within a cone of radius $R = 0.5$ around the jet axis. For each jet the associated tracks are ordered in descending 3D impact parameter significance $d_{3D}/\sigma(d_{3D})$, where the impact parameter d_{3D} is the minimum distance between the primary vertex and the trajectory of the track measured in three dimensions and its estimated uncertainty is $\sigma(d_{3D})$. The second track on this list is considered, and if its 3D impact parameter significance is above a certain threshold, either 3 or

4 depending on the data-taking period, the jet is considered as b tagged. The second online b tag step proceeds in a similar way, however the tracks are reconstructed using both pixel and strip tracker hit information within reach of the corresponding jet. Again, a jet is b tagged if the 3D impact parameter significance of its second selected track in order of decreasing significance satisfies the criterion $d_{3D}/\sigma(d_{3D}) > 6$.

In the offline reconstruction, primary vertices are identified using the Deterministic Annealing (DA) clustering of tracks [8]. The vertices are required to have a z position within 24 cm of the nominal detector centre and a radial position within 2 cm from the beamspot. As in the case for the trigger, the primary vertex with the largest value of the p_T^2-sum of tracks associated with it is chosen as the one originating from the hard interaction, and is used as reference for all other objects reconstructed in the event.

Jets are reconstructed using the anti-kt algorithm [9] from particle-flow objects [10, 11] with the cone radius parameter $R = 0.5$. The extra PU interactions affect the jet momentum reconstruction. To mitigate this effect, a track-based algorithm that filters all charged hadrons that do not originate from the primary interaction is used. In addition, a calorimeter-based algorithm evaluates the energy density in the calorimeter from interactions not related to the primary vertex and subtracts it from the reconstructed jets in the event. Standard jet-energy corrections are applied [12].

Secondary vertex (SV) reconstruction is performed using the Adaptive Vertex Finder [13], which performs a fully inclusive vertex search in a list of given tracks. The SV candidates must not share more than 65% of the tracks with the primary vertex and their radial distances to the primary vertex must exceed 3σ. Vertices from long-lived particles are suppressed rejecting SV candidates with radial distance to the primary vertex larger than 2.5 cm or whose mass is compatible with that of the K_S^0 or greater than 6.5 GeV/c^2. The flight direction of the SV candidates must lie within a cone of $\Delta R < 0.5$ around the jet direction, where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, $\Delta\eta = \eta_{jet} - \eta_{SV}$ and $\Delta\phi = \phi_{jet} - \phi_{SV}$ are, respectively, the pseudorapidity (η) and the azimuthal angle (ϕ) separations between the jet and the SV flight direction.

The combined secondary vertex (CSV) algorithm [14] is used in the offline identification of b jets. The CSV algorithm uses track impact parameter and secondary vertex information in a jet combined in a likelihood discriminant that provides a good separation between b jets and jets of other flavours.

4 Event Selection and the EventBTag variable

Events are required to have at least three reconstructed jets with $|\eta| < 2.2$. In the low-mass scenario ($M_\phi < 180$ GeV/c^2), events accepted by the low jet p_T threshold triggers are selected, making a sample corresponding to 2.7 fb^{-1} of integrated luminosity. The offline selection demands that the first three leading jets have to pass p_T cuts of 46, 38 and 20 GeV/c, respectively. In the medium-mass scenario ($M_\phi \geq 180$ GeV/c^2), a combination of di-jet triggers with low and high jet p_T thresholds forms an event sample whose integrated luminosity corresponds to 4.0 fb^{-1}. As in the low-mass scenario, at least three jets are required to pass an offline p_T selection but for the first two leading jets the cuts are raised to 60 and 53 GeV/c, respectively.

For the "triple-btag" selection, in which the signal is ultimately searched for, all three leading jets are required to pass a tight CSV b tag selection requirement, in which the misidentification probability for light-flavour jets is about 0.1%, at an average jet p_T of about 80 GeV/c.

The total number of events that pass the selections are 106626 and 89637 for the low- and

medium-mass scenarios, respectively.

For the "double-btag" selection, which is instrumental in estimation of the main background, only two of the three leading jets have to pass the above-mentioned criteria, while the third jet remains untagged, i.e. no b tag requirement is made on that jet.

The vertex invariant mass, calculated from all tracks that form the secondary vertex provides an additional separation between b, c and udsg jets beyond the CSV requirement. Figure 1 shows the additional flavour separation of the SV mass of b tagged jets obtained from simulated $t\bar{t}$ events. In the case where the SV is not reconstructed in a jet the SV mass assumes a fixed negative value.

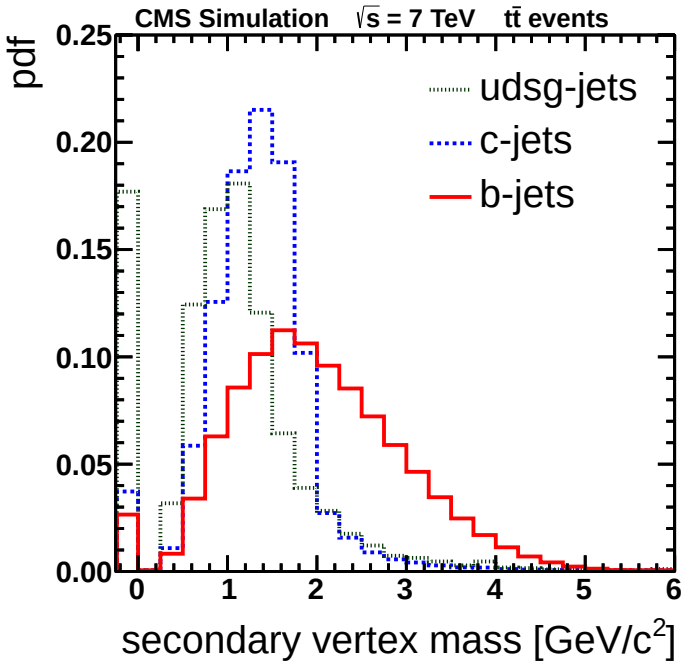


Figure 1: Normalised distributions of the mass of secondary vertices in jets for different flavours, udsg, charm and bottom, in simulated $t\bar{t}$ events.

In order to construct an independent, compact overall event b tag variable, while preserving large statistics in all bins of the di-jet mass distribution, the SV mass range is divided into three bins, $B_j = 0, 1, 2$ for each jet j. The quantity B_j is referred to as the secondary vertex mass index of the respective jet.

- $B_j = 0$, if $M_{SV,j} \leq 1$ GeV/c^2, including the cases when an SV is not reconstructed and a negative value for $M_{SV,j}$ is assumed.
- $B_j = 1$, if $1 < M_{SV,j} \leq 2$ GeV/c^2,
- $B_j = 2$, if $M_{SV,j} > 2$ GeV/c^2,

The values of B_1, B_2 and B_3 are combined in the EventBTag variable defined as follows:

$$\text{EventBTag} = T_{12} + T_3 \,,$$

where

- $T_{12} = 0$, if $B_1 + B_2 < 2$
- $T_{12} = 1$, if $2 \leq B_1 + B_2 < 3$

- $T_{12} = 2$, if $B_1 + B_2 \geq 3$

and

- $T_3 = 0$, if $B_3 < 2$
- $T_3 = 3$, if $B_3 = 2$

By construction, the EventBTag can take six possible values ranging from 0 to 5. Events with a strong triple-btag signature typically give the largest contribution for EventBtag values of 2 and 5.

5 Background Model

The events in the double-btag sample, described in Section 1.1, are organised in three categories, bbx, bxb and xbb, depending on the rank of the untagged jet which is represented by the lower-case letter x. The flavour of the untagged jet can be either light (udsg)[1], charm or bottom. The ranking in descending p_T of the three jets is incorporated in the nomenclature[2] adopted here, e.g. bbx means a sample of events where the two leading jets are b tagged and the third jet is the untagged jet.

From these three double-btag categories, bbx, bxb and xbb, nine background templates are constructed by weighting each untagged jet with the b tag probability assuming that its true flavour corresponds to either a light flavour (Q = udsg), a charm (C) or a beauty (B) parton. The b tag probability for each flavour is determined as a function of jet p_T and η with `Pythia` simulated QCD events. Data/MC scale factors for the b tag efficiencies of b, c and light jets are applied where appropriate [14].

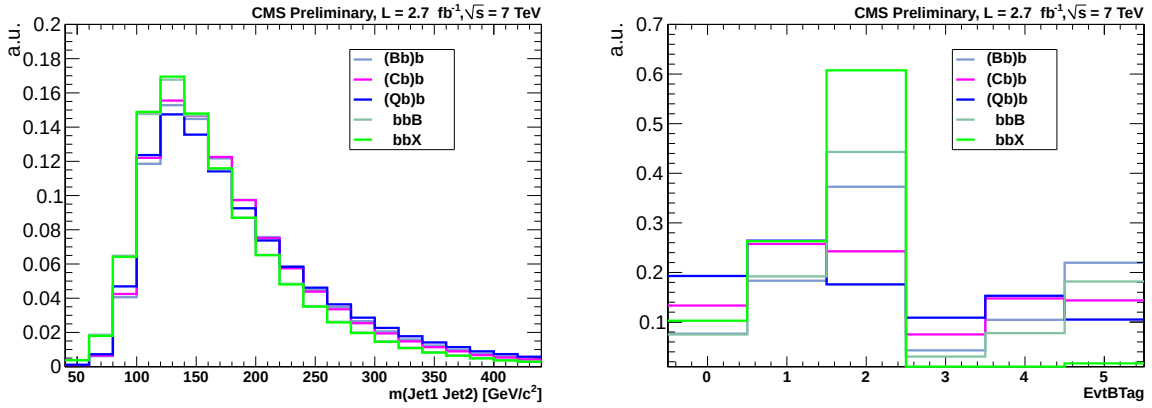
With the convention of the capital letter indicating the assumed flavour of the untagged jet, the nine templates Qbb, Cbb, Bbb, bQb, bCb, bBb, bbQ, bbC and bbB are created. As pointed out in Ref. [1], the template bbB models mainly bbb events in which the two leading b quarks originate from the same bb pair, while Bbb and bBb are important to cover cases where the two leading b quarks originate from different bb pairs.

The EventBTag variable is modelled in a similar way. Each of the three possible values of the secondary vertex mass index of the untagged jet is taken into account with a weight according to the probability that a jet will end up in a given bin of the SV mass distribution. These probabilities parametrised as a function of the jet p_T and η have been determined for each flavour with jets from simulated $t\bar{t}$ events. Thus each background template is a distribution in the di-jet mass and EventBTag two-dimensional space.

Some of the nine templates are rather similar to each other in shape, in such cases they are combined. In the cases where one of the two leading jets is untagged, e.g. Qbb and bQb, the templates are combined resulting in a merged template (Qb)b = Qbb + bQb. In analogy, also (Cb)b and (Bb)b are obtained. Similarly, when the third leading jet is the untagged one and the assumptions of its flavour are either Q or C, the bbQ and the bbC templates are combined to form the template bbX. Eventually, the total number of templates to be fitted to the data is reduced from nine to five, (Bb)b, (Cb)b, (Qb)b, bbB and bbX, whose M_{12} and EventBTag projections are shown in Figure 2 for the low-mass scenario.

Other templates whose di-jet mass spectra also resemble each other can be clearly distinguished

[1] Up, down, strange and gluon jets are denoted by udsg or simply called light flavour jet.
[2] Unless stated otherwise, the jet p_T ranking is always respected.

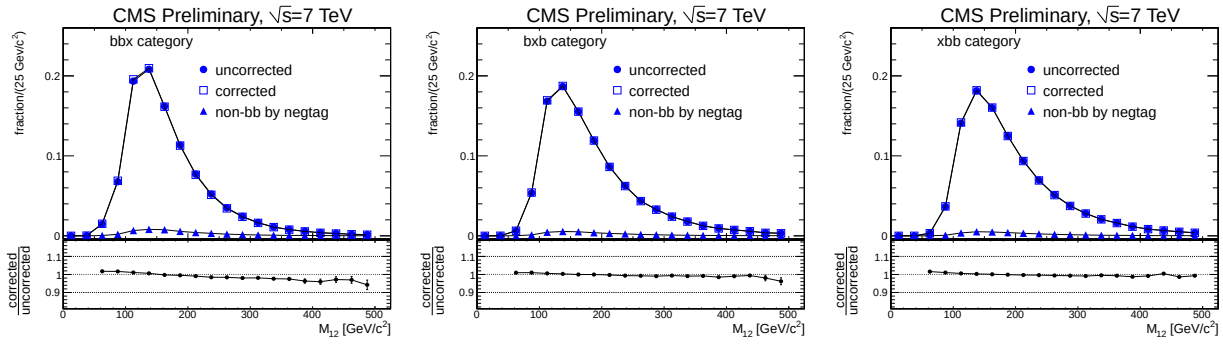


(a) (b)

Figure 2: The M_{12} (left) and EventBTag (right) projections of the five background templates, (Bb)b, (Cb)b, (Qb)b, bbB and bbX, for the low-mass scenario.

with the introduction of the EventBTag. This is the case for example between (Bb)b and (Qb)b. In general, the EventBTag improves significantly the discrimination between all the modelled flavour components. The template projections shown in Figure 2 include two additional corrections, which will be detailed in the following.

The basic assumption of the background model of the double-btag sample to consist entirely of events with at least two real b jets is only approximately correct. Although the remaining contamination from non-bb events is indeed very small, the application of improper b tag efficiency weights could lead to distortions of the background model and must be corrected for. This contamination is estimated directly from the data using a "negative" b tag discriminator [14]. The background from non-bb jets is estimated to be proportional to the negative b tag rate, i.e. the rate of events with the negative b tag discriminator above a certain threshold at least for one of the tagged jets in a template category. The threshold is calibrated for each tagged jet of the three template categories as a function of jet p_T with simulated `Pythia` QCD events, such that the ratio between the mistag rate and the negative tag rate is equal to unity.



(a) (b) (c)

Figure 3: Normalised M_{12} projection spectrum of the categories bbx (left), bxb (centre) and xbb (right) before (circles) and after (squares) the corrections, and the non-bb contribution (triangles) for the low-mass scenario. The ratio of the normalised corrected and uncorrected spectra are shown in the plots at the bottom.

The resulting correction effect for the three template categories is shown in Figure 3 from a subset of the available data sample. The solid circles represent the normalised invariant mass

distribution before correction, the triangles indicate the estimated non-b contribution, which is at the level of ∼3–4%. The open squares show the shape after subtraction of the estimated non-b component and subsequent normalization. At most a marginal change in shape is visible.

Another correction due to the online b tag is necessary because the double- and the triple-btag samples differ in their online b tag patterns. In the double-btag sample, the two offline b tagged jets coincide with the online b tagged jets in 90–95% of the events. Therefore the online and the offline b tag patterns of the templates are essentially the same, i.e. in the bbx category, for example, the two leading jets were also b tagged online. In the triple-btag sample, a bbc background is not entirely described by the online b tag pattern in which the two leading jets are online b tagged but also the cases with the first and the third, or the second and the third jets online b tagged are possible. To account for the other trigger patterns, relative online b tag weights are applied.

Background contributions from $t\bar{t}$ and Z + jets production are estimated from Monte Carlo and found to be negligible, $\sim$ 100 and 1 events per fb^{-1}, respectively. Still, those contributions are present in the double-btag sample and therefore absorbed in the background templates.

6 Signal Extraction

The signal templates are derived for each mass point from the simulated Higgs datasets. Examples for the projections of the signal templates for three different Higgs masses with the low-mass selection are shown in Figure 4.

A linear combination of the signal and the background templates are fitted to the data using a binned least-squares fit in M_{12} and EventBTag space. The χ^2 function is defined as

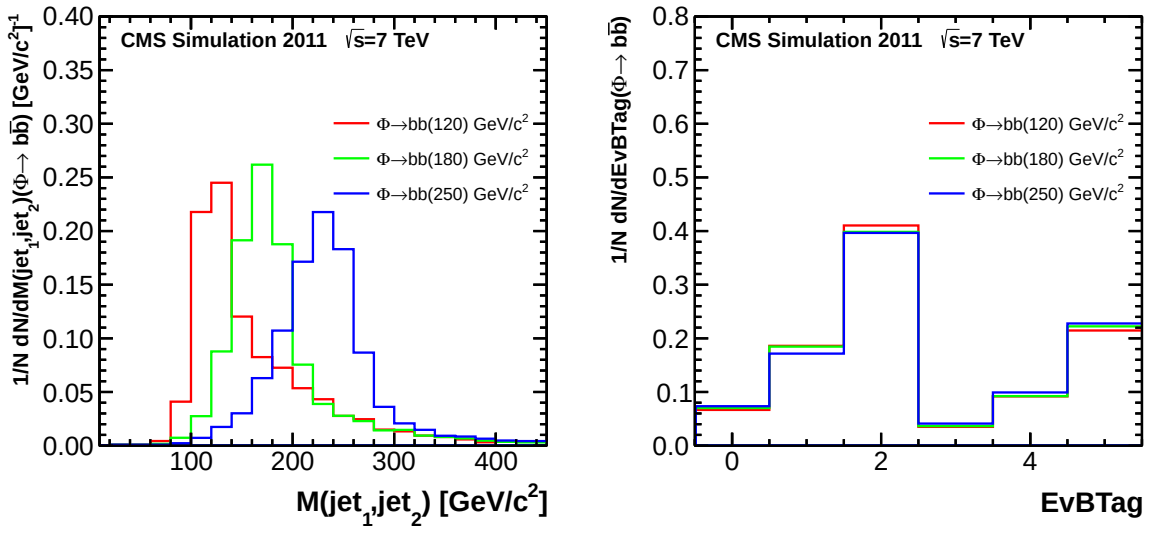
$$\chi^2 = \sum_{ij} \frac{\left[N_{ij}^{obs} - \left(\sum_b f_b N_{ij}^b + f_s N_{ij}^s\right)\right]^2}{(\sigma_{ij}^{obs})^2 + (\sigma_{ij}^{templ})^2}$$

where N_{ij}^{obs} is the number of observed events in the bin i, j, N_{ij}^b and N_{ij}^s are the bin contents of the five b background templates and of the signal template symbolised by s, scaled to the total number of observed events. The denominator sums the errors of the number of observed events and the corresponding estimation from the templates in quadrature. The parameters f_b and f_s are the free parameters of the fit and represent the fractions of each component, with f_b constrained to be non-negative.

The fitting procedure has been validated by artifical injection of MC signal into the data and subsequent comparison of the fit result.

6.1 Background-only Fit

In a first step, an unconstrained fit is performed without inclusion of a signal template, involving a linear combination of the background templates only. Results are shown in Figure 5 and Table 1 for the low- and medium-mass scenarios. The templates fit well to the data within the uncertainty of the template fits (shaded area). Events with three b jets have the largest contribution to the background $(74.3 \pm 2.5)\%$ in the low-mass scenario) as qualitatively expected and consistent with the $(72 \pm 4)\%$ obtained with `Pythia`.



(a) (b)

Figure 4: The M_{12} (left) and EventBTag (right) projections of signal templates for three different masses of the Higgs boson, $M_A = 120\ \mathrm{GeV}/c^2$, $M_A = 180\ \mathrm{GeV}/c^2$, $M_A = 250\ \mathrm{GeV}/c^2$ with the low-mass selection.

template	f_b low-mass	f_b medium-mass
(Bb)b	0.48 ± 0.02	0.46 ± 0.02
(Cb)b	0.13 ± 0.01	0.14 ± 0.01
(Qb)b	0.00 ± 0.01	0.00 ± 0.01
bbB	0.27 ± 0.02	0.29 ± 0.03
bbX	0.13 ± 0.01	0.11 ± 0.01

Table 1: The fractions of the various background templates for the low- and medium-mass scenarios from the background-only fit.

6.2 Combined fit of signal and background templates

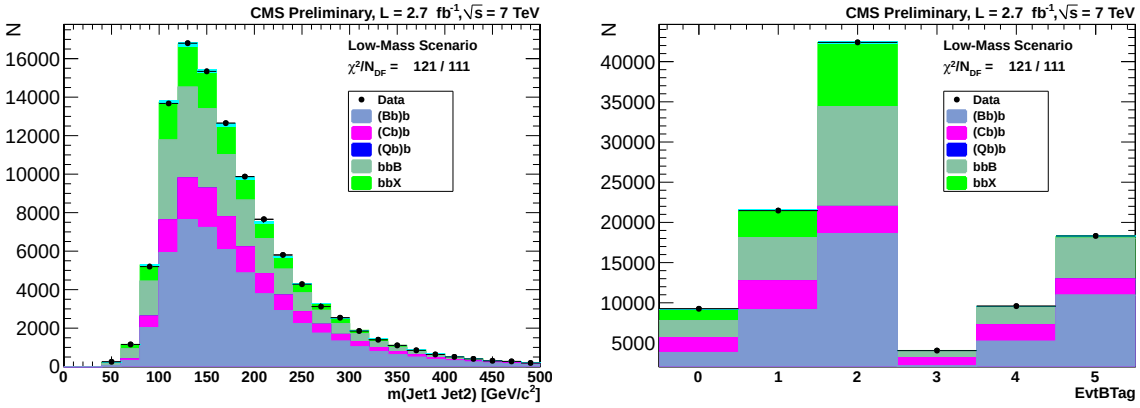
In the second step, a signal template is included together with the background templates in the fit with the fractions allowed to vary freely. The fit is performed for various Higgs mass points from 90 to 350 GeV/c^2. Results of the fit for a Higgs mass of 200 GeV/c^2 in the medium-mass scenario is shown in Figure 6. At this mass point the largest upward fluctuation of the fitted Higgs cross section is observed, corresponding to $\sim 1.4\ \sigma$ on the cross section.

7 Systematic Uncertainties

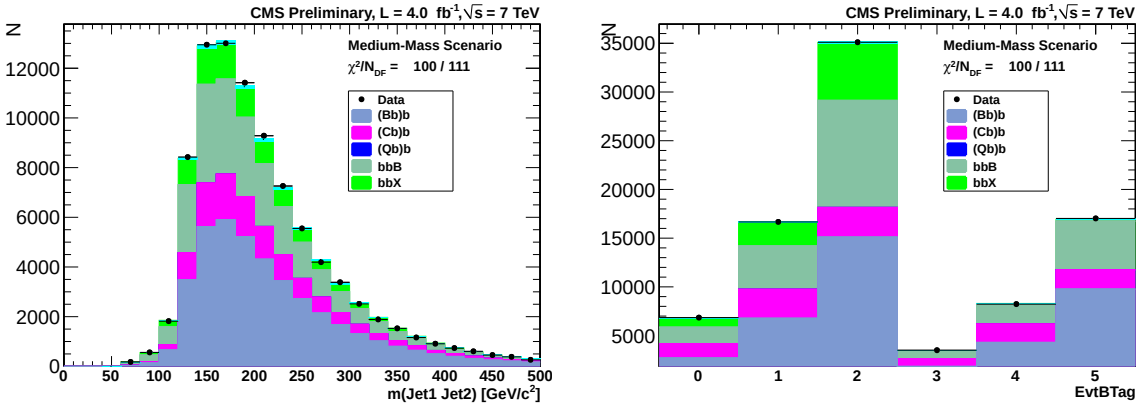
Various systematic uncertainties on the expected signal and background estimates affect the cross section measurement and/or its interpretation within MSSM. They are summarized in Table 2, and discussed in more detail in the following.

7.1 Uncertainties of the signal rate

Various uncertainties affect the signal efficiency and thus the measured cross section numbers. The percentages given in the "value" column of Table 2 represent the size of these uncertainties. It includes the statistical uncertainties of the signal Monte Carlo samples, whose size is typically between 0.5 and 1 million events per mass point. The uncertainty of the kinematical trigger

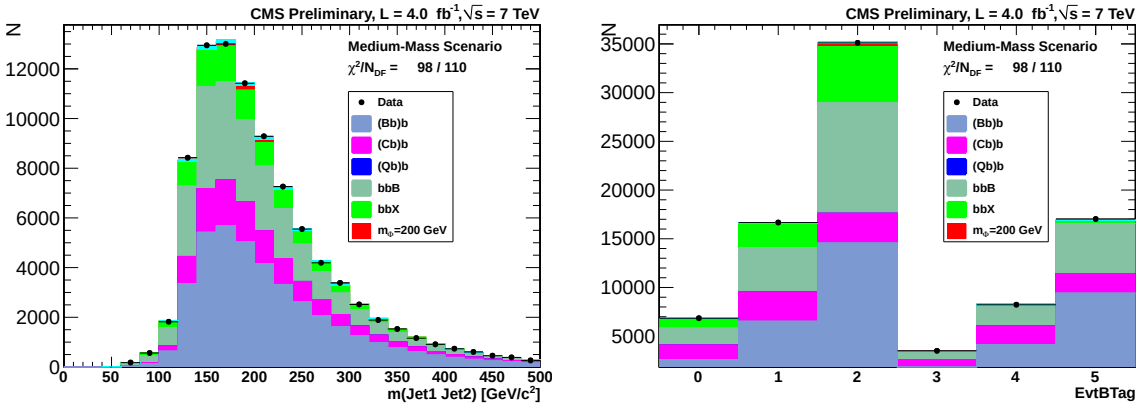


(a) (b)



(c) (d)

Figure 5: Results of the background-only fit in the triple-btag sample, on top (bottom) for the low-(high-)mass scenario. The cyan area corresponds to the templates uncertainties. The left plot shows the projection in di-jet mass, the right plot the projection on the EventBTag axis.



(a) (b)

Figure 6: Results of the fit including a signal template for a Higgs boson with a mass of 200 GeV/c^2 in the triple-btag sample, for the medium-mass scenario. The left plot shows the projection in di-jet mass, the right plot the projection on the EventBTag axis.

Source	Value	Applies to	Type
luminosity	2.2%	signal	rate
signal Monte Carlo statistics	$1.1 - 2.6\%$	signal	rate
trigger efficiencies	10%	signal	rate
b-tag efficiency	$3 - 10\%$	signal/background	rate/shape
b-tag efficiency dependence on topology	6%	signal	rate
mistag rate	$10 - 20\%$	background	shape
relative online b tag (per jet)	15%	signal	rate
bb purity corrections	(see text)	background	shape
online b tag template correction	(see text)	background	shape
jet energy scale	$1 - 4\%$	signal	rate/shape
jet energy resolution	$6 - 9\%$	signal	rate/shape
μ_r, μ_f $(bb\phi)$	$6 - 28\%$	signal	rate
parton density functions and α_s	$3 - 6\%$	signal	rate
underlying event/parton shower	4%	signal	rate

Table 2: Summary of systematic uncertainties.

efficienciency is 10%. For the offline btag efficiency, the uncertainty per beauty and charm jet is between $3 - 10\%$ depending on the transverse momentum of the jet. The mistag rate uncertainty for light flavor jets is higher and ranges between $10 - 20\%$. The uncertainty of the additional efficiency factor arising from the online btag algorithm relative to the offline btag is accounted for as 15% for each of the two online btags required per event. The jet energy scale uncertainty ranges from $1 - 4\%$ and the jet energy resolution uncertainty is between $6 - 9\%$; the impact of both is evaluated along with the corresponding shape effects which are addressed below.

A dependence of the b-tag efficiency on the event topology, in events with 2 or more b-tagged jets, is observed. This dependence amounts to $5 - 8\%$ depending on the jet rank and p_T. The actual value taken as uncertainty is 6%. The effect on the template shapes is negligible and only accounted for in the signal yields.

7.2 Uncertainties in the MSSM interpretation

The cross sections used for the MSSM interpretation are subject to the QCD scale (μ_r, μ_f) uncertainties, the uncertainties due to the PDFs and α_s, and the uncertainties from the underlying event and parton shower modelling. These uncertainties affect only the translation of the signal cross section into $\tan\beta$ in the MSSM interpretation. They have no effect on the signal cross section measurements.

7.3 Shape uncertainties

Various uncertainties affect the shapes of the signal and background templates used in the fit, and thus affect also the background estimate. These include the uncertainties on the jet energy scale, jet energy resolution, b-tag efficiencies and mistag rates. The latter two also affect the shape of the background. These four sources of systematic uncertainties were accounted for in the fits via nuisance parameters. This procedure allows for a continuous variation of the 2D templates for shape variation uncertainties in the fit. The original χ^2_0 is modified to

$$\chi^2 = \chi^2_0 + \sum_{k=1}^{n_p} p_k^2$$

where n_p is the number of nuisance parameters p_k which are essentially pulls. A variation of the shape of the distributions of up to $\pm 2\sigma$ is considered, therefore

$$S_{\pm} = S_0 + \sum_{k=1}^{n_p} \frac{S^k_{\pm 2\sigma} - S_0}{2} p_k$$

where S_0 is the unaltered shape, $S^k_{\pm 2\sigma}$ is the shape after $\pm 2\sigma$ variation of the nuisance parameter p_k. The uncertainty arising from jet energy scale and b tag on the template shape are found to increase the error of the fitted fraction by typically 0.001–0.004 in quadrature, the corresponding effect from jet energy resolution uncertainty is typically 0.001–0.003.

Additional systematic uncertainties arise from the impurity of the double-btag sample and the online b tag correction to the templates shapes. They are estimated as 30% of the respective change of the fitted signal yield. The uncertainty of the bb purity correction ranges between 0.001-0.003 in terms of the measured Higgs signal fraction in the mass range 90-130 GeV^2, and is below 0.001 elsewhere. The uncertainty of the online b tag shape correction in terms of the measured Higgs signal fraction ranges from 0.001-0.004 in the mass range 90-160 GeV/c^2, and is below 0.001 elsewhere.

The statistical uncertainty of the used offline b tag efficiency values is propagated directly into the templates and accounted for in the fitting procedure. The impact on the error of the fitted fractions is typically in the range 0.001–0.006. The effects of systematic errors relevant for the shape of the background are summarized in Table 3.

Uncertainty	Δf_s
JES and b tag	0.001–0.004
JER	0.001–0.004
bb purity correction	0.000–0.003
Online b tag correction	0.001–0.004
Offline b tag statistical	0.001–0.006

Table 3: Summary of background-driven systematic shape uncertainties, expressed in their effect on the fitted signal fraction.

8 Results

8.1 Signal efficiencies

The signal efficiencies ϵ_s are summarised in Table 4 and shown in Figure 7 as a function of the mass of the MSSM pseudoscalar neutral Higgs for the low- and medium-mass scenarios. The smaller efficiency in the medium-mass scenario is compensated by the larger luminosity compared to the low-mass scenario. Depending on the Higgs mass the efficiency varies from $0.13 - 1.4\%$, therefore from 350 to 5600 events are expected for a cross section of 100 pb.

8.2 Cross sections

The cross sections for the pseudoscalar neutral Higgs boson with mass M_A are obtained with the formula

$$\sigma(M_A) = \frac{f_s(M_A) N^{data}}{\epsilon_s(M_A) \mathcal{L}}$$

where $f_s(M_A)$ and $\epsilon_s(M_A)$ are the signal fractions returned by the fit and the signal efficiencies (Sect. 8.1), respectively, N^{data} and $\mathcal{L}$ are the total number of data events in the triple-btag sample and the integrated luminosity, respectively.

M_A (GeV/c^2)	ϵ_s (low-mass)	ϵ_s (medium-mass)
90	0.129%	0.070%
100	0.179%	0.096%
120	0.342%	0.183%
130	0.430%	0.247%
140	0.522%	0.334%
160	0.707%	0.536%
180	0.876%	0.717%
200	1.001%	0.901%
250	1.256%	1.227%
350	1.416%	1.401%

Table 4: Signal efficiencies ϵ_s for various masses of the MSSM pseudoscalar neutral Higgs in the low- and medium-mass scenarios.

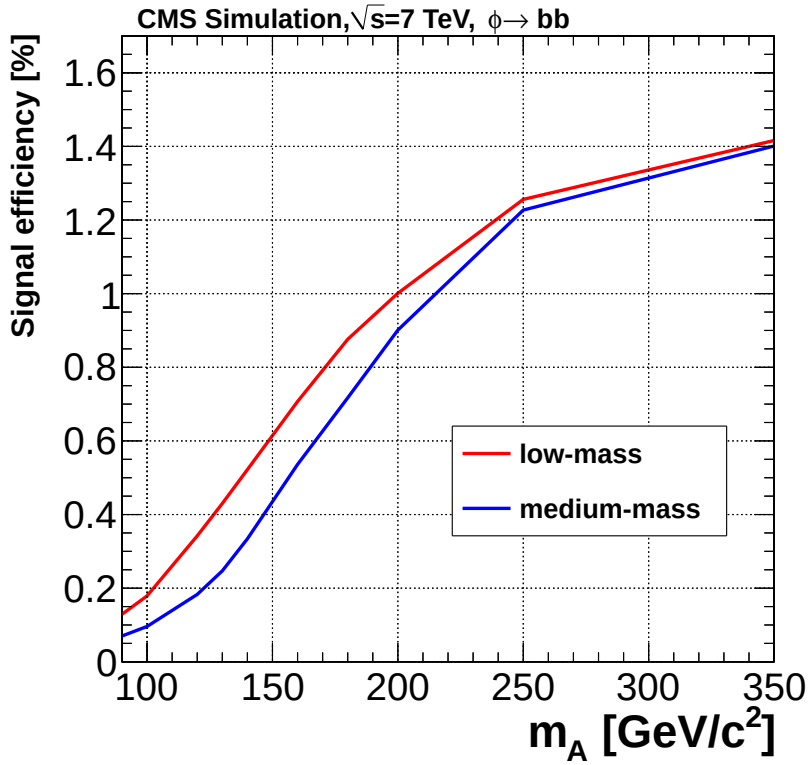


Figure 7: Signal efficiencies as a function of the mass M_A of MSSM pseudoscalar neutral Higgs in the low-mass scenario.

Systematic uncertainties affecting normalisation are taken into account when calculating the cross section. The shape altering systematic uncertainties are already accounted for in the fit procedure.

The measured cross sections are shown in Figure 8 and in Table 5. The measurements are consistent with no observation of a signal, with the largest observed upward deviation of $\sim$ 1.4 σ at $M_A = 200\,\mathrm{GeV}/c^2$

M_A(GeV/c^2)	f_s	observed events	σ (pb)
90	-0.0205 $\pm$ 0.0138	-2194 $\pm$ 1677	-630 $\pm$ 481
100	-0.0383 $\pm$ 0.0164	-4094 $\pm$ 2308	-845 $\pm$ 477
120	-0.0082 $\pm$ 0.0104	-873 $\pm$ 1158	-94 $\pm$ 125
130	-0.0034 $\pm$ 0.0096	-365 $\pm$ 1031	-31 $\pm$ 89
140	-0.0044 $\pm$ 0.0097	-474 $\pm$ 1050	-34 $\pm$ 75
160	-0.0022 $\pm$ 0.0074	-235 $\pm$ 795	-12 $\pm$ 42
180	0.0021 $\pm$ 0.0077	189 $\pm$ 684	7 $\pm$ 24
200	0.0109 $\pm$ 0.0069	968 $\pm$ 706	27 $\pm$ 20
250	0.0041 $\pm$ 0.0043	360 $\pm$ 405	7 $\pm$ 8
350	-0.0057 $\pm$ 0.0026	-508 $\pm$ 297	-9 $\pm$ 5

Table 5: The signal fractions returned by the fit, the number of observed events and the measured cross sections (σ) for different masses of the neutral Higgs bosons. The uncertainties of the fractions include statistical and shape altering systematics, whereas the cross section, number of observed, events, uncertainties also include normalisation uncertainties.

8.3 Limits on cross sections and MSSM tan β

The results are translated into upper limits on the cross section times the branching ratio, $\sigma(pp \to b\phi) \times BR(H \to b\bar{b})$, in the mass range $90 - 350$ GeV/c^2. For calculations of exclusion limits, we adopt the modified frequentist criterion CL$_s$ [15, 16] using the RooStats package [17]. The chosen test statistic, used to determine how signal- and background-like the data are, is based on the profile likelihood ratio. Systematic uncertainties are incorporated in the analysis via nuisance parameters and treated as pseudo-observables, following the frequentist paradigm. In the procedure of statistical inference the normalisation of the background templates are kept unconstrained.

The observed limits and the median expected 95% C.L. limits as a function of the scalar mass are shown in Table 6 and Figure 9. In the latter the 1σ and 2σ bands of the test statistic including systematic uncertainties are also shown. All points of the observed limit are well within the expected 2σ band. There is no indication of a statistically significant excess.

M_A(GeV/c^2)	median expected	observed
90	897	537
100	708	451
120	251	174
130	175	149
140	138	116
160	84.1	74.7
180	49.8	45.7
200	36.2	55.5
250	19.4	28.1
350	10.8	7.55

Table 6: Median expected and observed limits at 95% C.L. on $\sigma(pp \to b\phi) \times BR(H \to b\bar{b})$, in pb.

The interpretation within the MSSM m_h^{max}, $\mu = +200$ GeV, scenario [18] is performed using the NNLO cross sections calculated by bbh@nnlo [19] for b-associated h/H/A production in the 5 flavour scheme and the branching ratios computed with the FeynHiggs program [20–23].

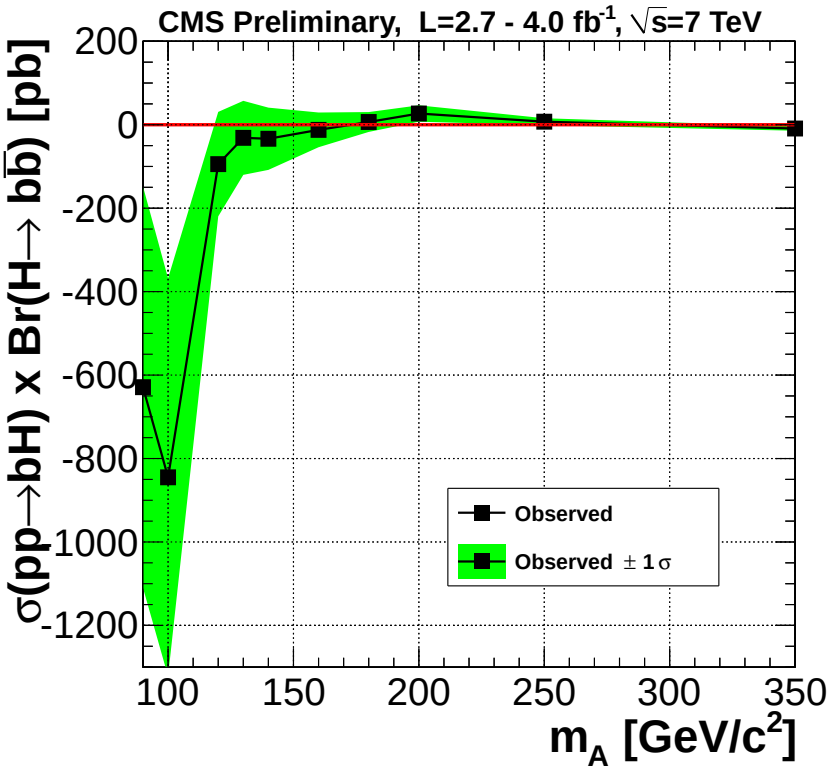
The observed and expected medians 95% C.L. upper limits on tan β versus M_A are shown in

Figure 10. Bands corresponding to 1σ and 2σ are also shown. The observed $\tan\beta$ upper limit ranges from about 25 to 50 in the 90–350 GeV/c^2 Higgs boson mass range.

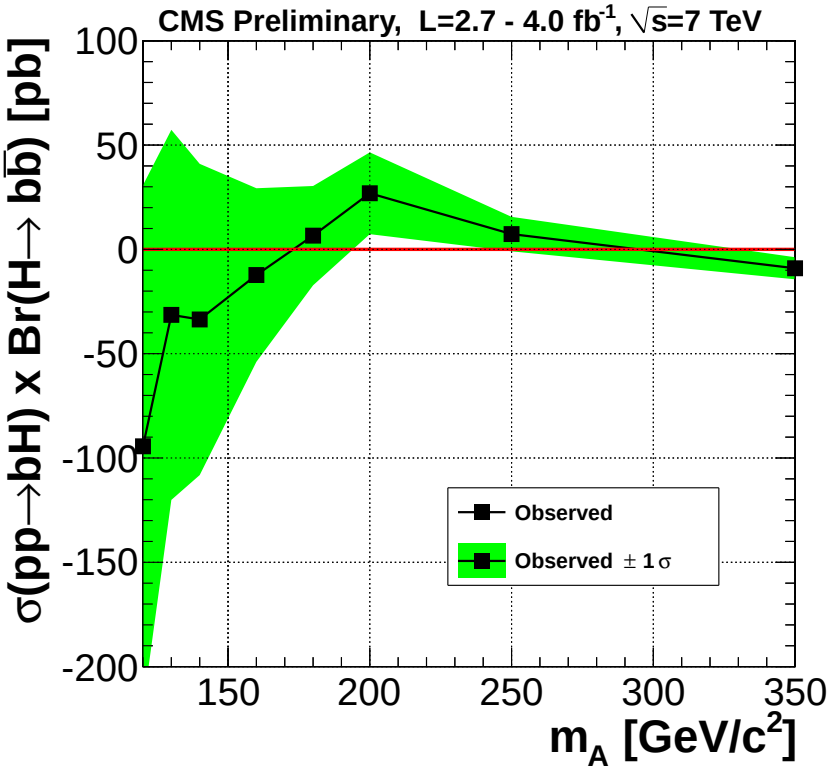
9 Conclusions

A search for a Higgs boson decaying into a pair of b quarks and produced in association with at least one additional b quark is performed with up to 4.0 fb^{-1} in proton-proton collisions at a centre-of-mass energy of 7 TeV at the LHC. The data were taken with dedicated jet triggers in combination with an online b-tag algorithm. The search is performed on a triple b-tagged sample using a simultaneous fit in two discriminant variables: the invariant mass of the two leading jets and the EventBTag variable that reflects the heavy flavour content of the event.

No signal is observed above SM background expectations and upper limits on the $pp \to bH + X$, $H \to b\bar{b}$ cross section times branching ratio are derived in the 90–350 GeV/c^2 mass range. These results are interpreted within the MSSM model in the m_h^{max} scenario in terms of bounds in the MSSM (M_A, $\tan\beta$) parameter space. The observed $\tan\beta$ upper limit ranges from about 25 to 50 in the same Higgs boson mass range.

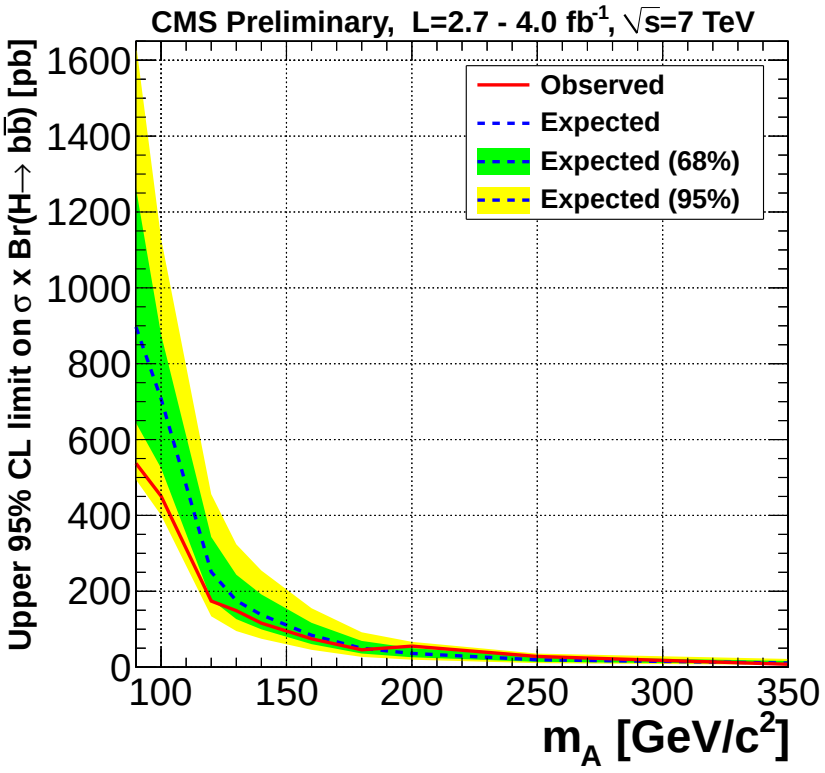


(a)

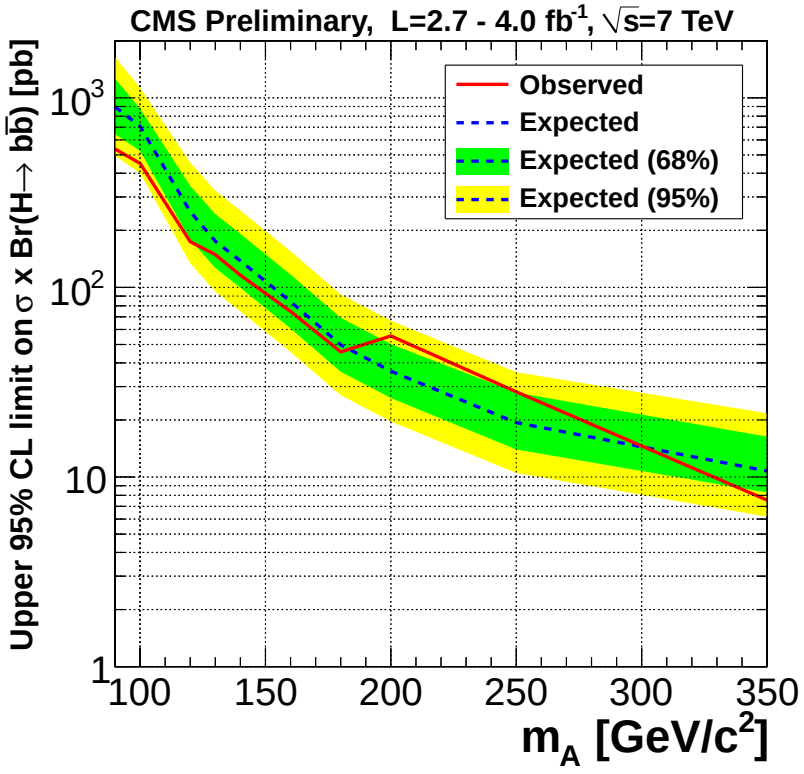


(b)

Figure 8: Measured cross sections as a function of the mass the Higgs mass M_A. On the top for the full mass range and on the bottom for $M_A \geq 130\,\mathrm{GeV/c^2}$.



(a)



(b)

Figure 9: The observed 95% C.L. upper limits on the cross section times the branching ratio. Linear-(log-)scale on top (bottom).

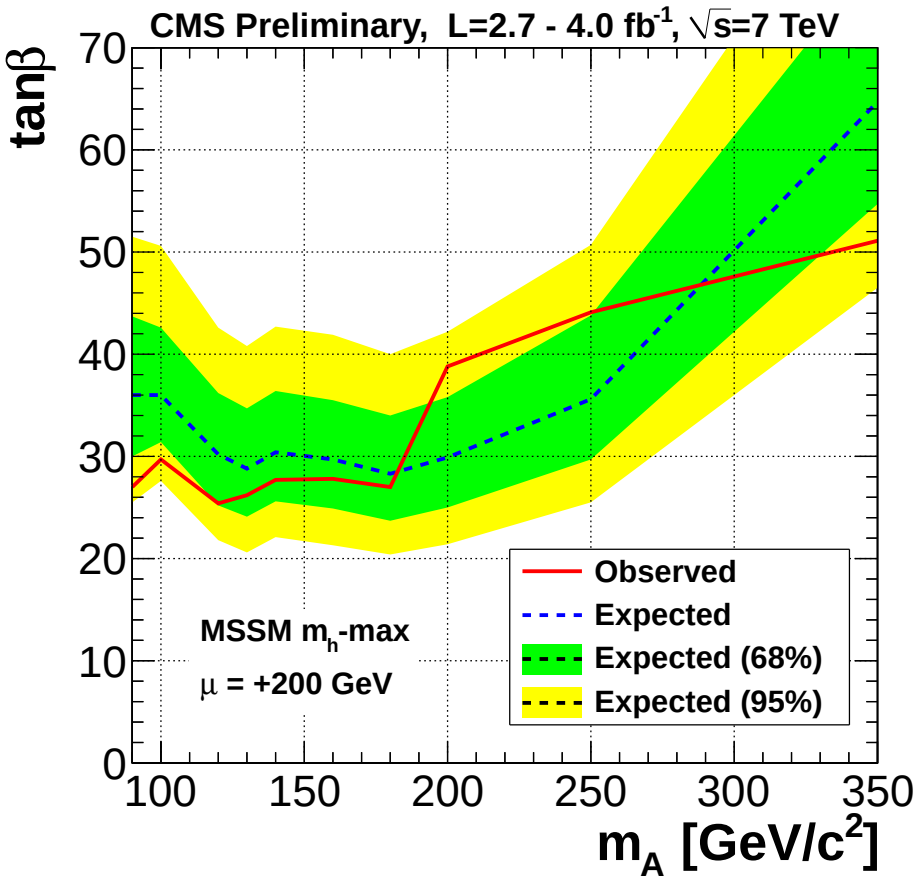


Figure 10: Observed and median expected 95% C.L. upper limits on $\tan\beta$ versus M_A in the m_h^{max} scenario for $\mu = +200$ GeV. The expected 1σ and the 2σ bands are also shown.

References

[1] CDF Collaboration Collaboration, "Search for Higgs Bosons Produced in Association with b-quarks", *Phys.Rev.* **D85** (2012) 032005, arXiv:1106.4782.

[2] D0 Collaboration Collaboration, "Search for neutral Higgs bosons in the multi-b-jet topology in 5.2fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys.Lett.* **B698** (2011) 97–104, arXiv:1011.1931.

[3] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[4] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl.Instrum.Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[5] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **0605** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[6] J. Alwall, P. Demin, S. de Visscher et al., "MadGraph/MadEvent v4: The New Web Generation", *JHEP* **0709** (2007) 028, doi:10.1088/1126-6708/2007/09/028, arXiv:0706.2334.

[7] M. L. Mangano, M. Moretti, F. Piccinini et al., "ALPGEN, a generator for hard multiparton processes in hadronic collisions", *JHEP* **0307** (2003) 001, arXiv:hep-ph/0206293.

[8] W. Erdmann, "Offline Primary Vertex Reconstruction with Deterministic Annealing Clustering", *CMS INTERNAL NOTE* **2011/014** (2011).

[9] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **0804** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[10] CMS Collaboration Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET",.

[11] CMS Collaboration Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV",.

[12] CMS Collaboration Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, arXiv:1107.4277.

[13] W. Waltenberger, "Adaptive Vertex Reconstruction", *CMS Note* **2008/33** (2008).

[14] CMS Collaboration, "b-Jet Identification in the CMS Experiment", CMS Physics Analysis Summary CMS-PAS-BTV-11-004, (2011).

[15] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl.Instrum.Meth.* **A434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[16] A. L. Read, "Presentation of search results: The CL(s) technique", *J.Phys.G* **G28** (2002) 2693–2704, doi:10.1088/0954-3899/28/10/313.

[17] L. Moneta, K. Belasco, K. Cranmer et al., "The RooStats Project", in *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. SISSA, 2010. arXiv:1009.1003. PoS(ACAT2010)057.

[18] M. S. Carena, S. Heinemeyer, C. Wagner et al., "MSSM Higgs boson searches at the Tevatron and the LHC: Impact of different benchmark scenarios", *Eur.Phys.J.* **C45** (2006) 797–814, doi:10.1140/epjc/s2005-02470-y, arXiv:hep-ph/0511023.

[19] 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, doi:10.1103/PhysRevD.68.013001, arXiv:hep-ph/0304035.

[20] G. Degrassi, S. Heinemeyer, W. Hollik et al., "Towards high precision predictions for the MSSM Higgs sector", *Eur.Phys.J.* **C28** (2003) 133–143, doi:10.1140/epjc/s2003-01152-2, arXiv:hep-ph/0212020.

[21] M. Frank, T. Hahn, S. Heinemeyer et al., "The Higgs Boson Masses and Mixings of the Complex MSSM in the Feynman-Diagrammatic Approach", *JHEP* **0702** (2007) 047, doi:10.1088/1126-6708/2007/02/047, arXiv:hep-ph/0611326.

[22] S. Heinemeyer, W. Hollik, and G. Weiglein, "FeynHiggs: A Program for the calculation of the masses of the neutral CP even Higgs bosons in the MSSM", *Comput.Phys.Commun.* **124** (2000) 76–89, doi:10.1016/S0010-4655(99)00364-1, arXiv:hep-ph/9812320.

[23] S. Heinemeyer, W. Hollik, and G. Weiglein, "The Masses of the neutral CP - even Higgs bosons in the MSSM: Accurate analysis at the two loop level", *Eur.Phys.J.* **C9** (1999) 343–366, doi:10.1007/s100529900006, arXiv:hep-ph/9812472.