

## Search for SM Higgs decaying to bb at CMS

---

**David Lopes-Pegna**\*On behalf of the CMS Collaboration

*Princeton University*

*E-mail:* [dlopes@princeton.edu](mailto:dlopes@princeton.edu)

Results are presented on the search for the standard model Higgs decaying into two b-quark jets, using associated production of the Higgs with W and Z bosons and with a ttbar pair. The search is performed in a data sample corresponding to an integrated luminosity of 5.0 fb<sup>1</sup>, recorded by the CMS detector in proton-proton collisions at the LHC with a center-of-mass energy of 7 TeV during 2011, and in a data sample of up to 5.1 fb<sup>1</sup> of 8 TeV collisions during 2012.

*36th International Conference on High Energy Physics,  
July 4-11, 2012  
Melbourne, Australia*

---

\*Speaker.

## 1. Introduction

The search for the Higgs boson is currently one of the most important endeavors of experimental particle physics, as in the standard model (SM) the Higgs mechanism is considered at the explanation for the electroweak symmetry breaking mechanism. Before Summer 2012, direct searches at the Large Hadron Collider (LHC) exclude at 95% CL production of a SM Higgs boson with a mass between 112.9-115.5 GeV and 127-600 GeV [1, 2]. Direct searches at the Tevatron exclude at 95% CL the 100-106 GeV and the 147-179 GeV mass ranges, and observe an excess of events over expected background in the 115-135 mass range [3]. In the mass region not currently excluded by these searches, the Higgs boson decays predominantly into  $b\bar{b}$ . Therefore, the observation of the  $H \rightarrow b\bar{b}$  decay is essential in determining the nature of the Higgs boson. At the LHC the main SM Higgs boson production mechanism is gluon fusion. However, in this production mode, the detection of the  $H \rightarrow b\bar{b}$  decay is considered nearly impossible due to overwhelming dijet production expected from quantum-chromodynamic (QCD) interactions. The same holds true for the next most copious production mode, through vector-boson fusion. Searches for  $H \rightarrow b\bar{b}$  are therefore performed in processes in which a low-mass Higgs boson is produced in association with a vector boson (W or Z) or a  $t\bar{t}$  pair.

We report on a search for the standard model Higgs boson in the  $pp \rightarrow VH$  and  $t\bar{t}H$  production mode where V is either a W or a Z boson. The analysis is performed using a data sample corresponding to an integrated luminosity of  $5.0 \text{ fb}^{-1}$ , collected in 2011 by the CMS experiment at a center-of-mass energy of 7 TeV (for both the VH and  $t\bar{t}H$  analyses), and also using a data sample of up to  $5.1 \text{ fb}^{-1}$  at 8 TeV collected during 2012 (VH).

For the VH analysis, the following final states are included in the search:  $W(\mu\nu)H$ ,  $W(e\nu)H$ ,  $Z(\mu\mu)H$ ,  $Z(ee)H$  and  $Z(\nu\nu)H$ . For the  $t\bar{t}H$  analysis, the top pair decays semi-leptonically ( $t\bar{t} \rightarrow l\nu jjb\bar{b}$ ) or fully leptonically ( $t\bar{t} \rightarrow l\nu l\nu b\bar{b}$ ).

## 2. Triggers

For the WH channels the trigger paths consist of several single-lepton triggers with tight lepton identification. Leptons are also required to be isolated from other tracks and calorimeter energy depositions to maintain an acceptable trigger rate. For the  $W(\mu\nu)H$  channel and  $Z(\mu\mu)H$ , for the 2011(2012) data set, the trigger thresholds for the muon transverse momentum,  $p_T$ , are in the range of 17(24) to 40 GeV, where the highest threshold is used without additional isolation requirements. For the  $W(e\nu)H$  channel, for the 2011 data set, the electron  $p_T$  threshold ranges from 17 to 30 GeV. The lower-threshold paths require two jets and a minimum requirement on the norm of an online estimate the missing transverse energy vector. For the 2012 data set, the single isolated-electron trigger uses a 27 GeV threshold. For the  $Z(ee)H$  channel, dielectron triggers with lower  $p_T$  thresholds (17 and 8 GeV) and tight isolation requirements are used. For the  $Z(\nu\nu)H$  channel, combinations of several triggers are used. A trigger with missing transverse energy  $> 150$  GeV is used for the complete dataset in both 2011 and 2012. During 2011 this trigger was used in conjunction with triggers that require the presence of two central jets ( $|\eta| < 2.6$ ) with  $p_T > 20$  GeV and missing transverse energy thresholds of 80 and 100 GeV, depending on the luminosity. During 2012 this trigger was used in conjunction with a trigger that required two central jets above

$p_T > 30$  GeV (60 and 25 GeV for instantaneous luminosity above  $3 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ ) and a missing transverse energy threshold of 80 GeV, with additional requirements <sup>1</sup>.

For the  $t\bar{t}H$  analysis, events are triggered based on the presence of isolated, energetic, charged leptons and jets consistent with the presence of a top pair decay. Lepton + jets events are selected at the trigger level by requiring either the presence of a single, isolated muon with  $p_T > 24$  GeV or an isolated electron with  $p_T > 25$  GeV and three jets with  $p_T > 30$  GeV. In the dilepton case, events are triggered by having any combination of electrons and muons that gives one with  $p_T > 17$  GeV and the other with  $p_T > 8$  GeV.

### 3. $VH \rightarrow b\bar{b}$ Analysis

The analysis presented here has an improved sensitivity of about 50% with respect to the previous CMS Higgs boson search in these final states [4]. The main improvements result from the treatment of the signal and background BDT discriminant, from an increase in the kinematical phase space studied and from improvements in the measurement of the energy associated with jets originating from b quarks.

Details on the event reconstruction and selection are provided in [5]. The event selection is based first on the kinematic reconstruction of the vector bosons and the Higgs boson decay into two b-tagged jets. Backgrounds are then substantially reduced by requiring a significant boost of the  $p_T$  of the vector boson and the Higgs boson [6], which tend to recoil away from each other with a large azimuthal opening angle between them. Two regions of  $p_T(V)$  boost are used for each mode, with one region reaching lower in boost and the other remaining similar to the previous  $VH$  analysis [4]. The boost regions for the  $WH$  channels are  $120 < p_T(V) < 170$  GeV and  $p_T(V) > 170$  GeV, for the  $Z(\nu\nu)H$  channel the regions are  $120 < p_T(V) < 160$  GeV and  $p_T(V) > 160$  GeV, and for the  $Z(\ell\ell)H$  channels  $50 < p_T(V) < 100$  GeV and  $p_T(V) > 100$  GeV. Due to different signal and background composition, each boost region has different sensitivity and the analysis is performed separately on each region. The results of each region are then combined for each mode.

The reconstruction of the  $H \rightarrow b\bar{b}$  decay is made by requiring the presence of two central ( $|\eta| < 2.5$ ) jets above a minimum  $p_T$  threshold, and tagged by the CSV algorithm. This b-tagging algorithm combines in an optimal way the information about track impact parameters and secondary vertices within jets in a likelihood discriminant to provide separation of b jets from jets originating from light quarks and gluons, and charm quarks. Several working points for the CSV output discriminant (which can have values between zero and one) are used in the analysis, with different efficiencies and misidentification rates for b jets. For a CSV  $> 0.50$  requirement the efficiencies to tag b quarks, c quarks, and light quarks, are approximately 72%, 23%, and 3%. If more than two such jets are found in the event, the pair of jets with the highest total dijet transverse momentum,  $p_T(jj)$ , is selected.

The mass resolution for the two b jets from the Higgs decay is approximately 10% with a few percent bias on the mass. The mass resolution and bias are improved with respect to the analysis in [4] by applying a regression technique. A specialized BDT algorithm is trained on simulated

<sup>1</sup>At least one pair of jets with  $p_T$  greater than 100 GeV, and no jet with  $p_T$  greater than 40 GeV closer than 0.5 in azimuthal angle to the missing transverse energy direction.

signal  $H \rightarrow b\bar{b}$  events with inputs that include detailed information about the jet structure and that help differentiate jets from  $b$  quarks from light-flavor jets. These include variables containing information about several properties of a secondary vertex when present, information about tracks, charged constituents, and other variables related to the energy reconstruction of the jet. The resulting improvement on the mass resolution is approximately 15%, resulting in an increase in the analysis sensitivity of 10-20%, depending of the specific channel.

In the final stage of the analysis, and to better separate signal from background under different Higgs boson mass hypotheses, another BDT algorithm is trained separately at each mass value using simulated samples for signal and background that pass the event selection described above. The shape of the output distribution of this BDT algorithm is the final discriminant in which a fit is performed to search for events resulting from Higgs boson production.

Backgrounds arise from production of  $W$  and  $Z$  bosons in association with jets (from all quark flavors), singly and pair-produced top quarks, dibosons and QCD multijet processes. Appropriate control regions are identified in data and used to correct the Monte Carlo yield estimates for several of the most important background processes, including production of  $W$  and  $Z$  bosons in association with jets (light- and heavy-flavor), and  $t\bar{t}$  production. A set of simultaneous fits is then performed to the distributions of discriminating variables in the control regions, separately in each channel, to obtain consistent scale factors by which the Monte Carlo yields are adjusted.

Dominant systematic uncertainties due to detector effects come from the  $b$ -tagging simulation and the jet energy scale and resolution. Data-to-simulation  $b$ -tagging scale factors, measured in  $t\bar{t}$  events, are applied consistently to jets in signal and background events and the associated uncertainty is on the order of 3-15% depending on the channel and the specific process. The jet energy scale is varied within one standard deviation as a function of jet  $p_T$  and  $\eta$ , giving a 2-3% yield variation. The effect of the uncertainty on the jet energy resolution is evaluated by smearing the jet energies according to the measured uncertainty, giving a 3-6% variation in yields.

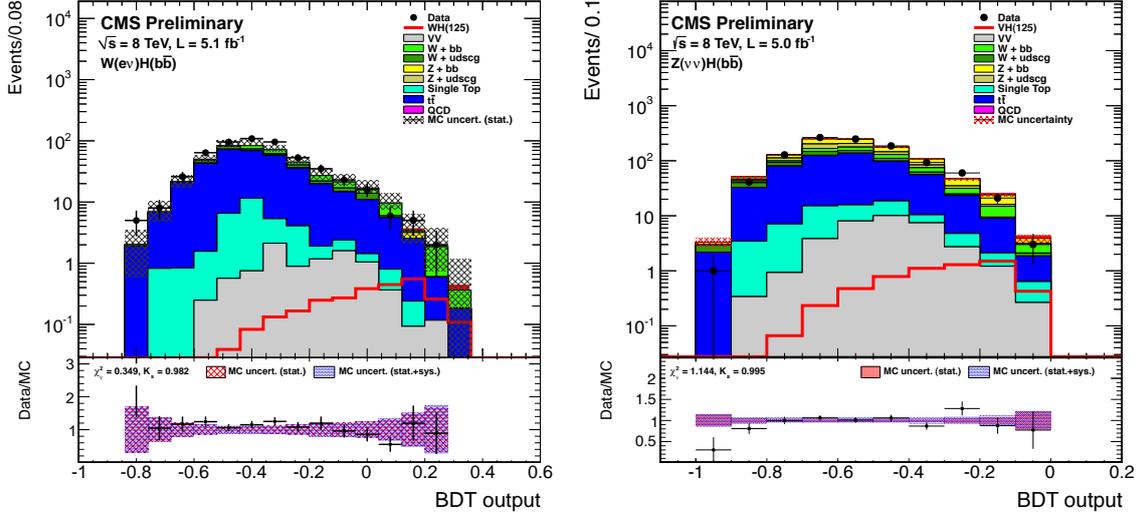
The uncertainty in the background yields that results from the estimates from data is approximately 10%. For the single-top and diboson backgrounds, estimated from simulation, an uncertainty of 30% is assigned.

Theoretical Uncertainties affect the total and differential  $VH$  cross-section. The total  $VH$  signal cross section has been calculated to next-to-next-to-leading (NNLO) order accuracy, with a total uncertainty of 4%. This analysis is performed in the boosted regime, and thus, potential differences in the  $p_T$  spectrum of the  $V$  and  $H$  between data and Monte Carlo generators could introduce systematic effects in the signal acceptance and efficiency estimates. Uncertainties of the order of 10-15% are included due next-to-leading-order (NLO) electroweak and NNLO QCD [5] corrections to  $VH$  production in the boosted regime.

Figure 1 show some examples of the BDT distributions for the 8 TeV data.

#### 4. $t\bar{t}H \rightarrow b\bar{b}$ Analysis

Details on the event reconstruction and selection are provided in [9]. For the lepton + jets case, we select events containing an energetic, isolated, charged electron or muon, and at least four energetic jets, two or more of which should be identified as containing a bottom quark hadron ( $b$ -tagged). For the dilepton case, we require a pair of oppositely charged energetic leptons (two



**Figure 1:** BDT output distributions for  $W(ev)H$  (left) and  $Z(vv)H$  (right) in the high  $p_T(V)$  bin, for data (points with errors), all backgrounds, and signal, after all selection criteria have been applied.

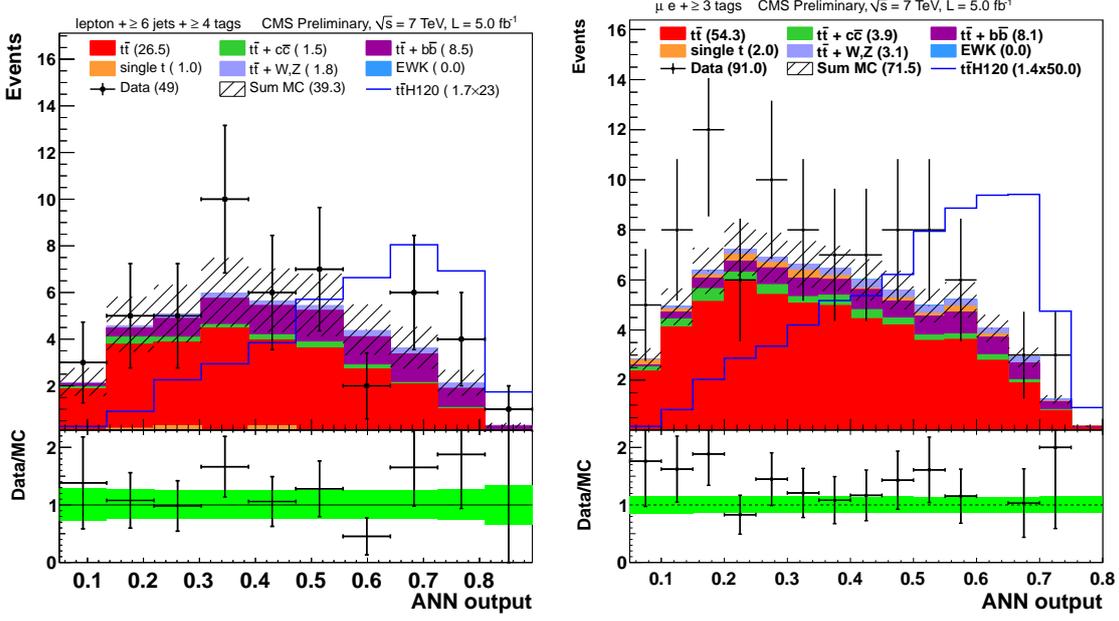
electrons, two muons or one electron and one muon) and two or more jets, with at least two of the jets being  $b$ -tagged. In both channels, a significant amount of missing transverse energy should be present as the result of one or more energetic neutrinos.

Artificial neural networks (ANNs) are used to further improve signal-to-background sensitivity in all channels of the analysis. Separate ANNs are trained for different channels, utilizing information related to object kinematics, event shape, and the discriminant output from the  $b$ -tagging algorithm. Events are divided into different categories based on the number of jets and the number of  $b$ -tagged jets that they contain. For the lepton + jets channel, we separate events by 4 jets (3,  $\geq 4$   $b$ -tags), 5 jets (3,  $\geq 4$   $b$ -tags), and  $\geq 6$  jets (2, 3,  $\geq 4$   $b$ -tags). For the dilepton channel, we separate events into the categories  $\geq 3$  jets +  $\geq 3$   $b$ -tags and 2 jets + 2  $b$ -tags. In both lepton + jets and dilepton channels, the variables that discriminate best between signal and background directly involve  $b$ -tagging information, such as the average CSV output value for  $b$ -tagged jets. The variables that distinguish best between signal and  $t\bar{t} + b\bar{b}$ , a background very similar to the signal, are variables that include  $b$ -tagging information, kinematic information, and angular correlation information, such as the minimum  $\Delta R$  between all pairs of  $b$ -tagged jets. The modeling of these variables were checked before using them as inputs to the ANN.

Figure 2 show some examples of the ANN distributions.

## 5. Results

For the  $VH$  analysis, upper limits at the 95% CL on the  $pp \rightarrow VH$  production cross section are obtained for Higgs boson masses in the 110-135 GeV range. These limits result from fitting the shape of the distribution of the output discriminant of a boosted-decision-tree (BDT) algorithm for the presence of a Higgs boson signal above what is expected from all background components. for the  $t\bar{t}H$  analysis, we set limits on the Higgs boson production cross section in the 110-140



**Figure 2:** ANN output distributions for the 6 jet and 4 tags in the lepton+jet analysis (left) and the 3 tags in the dilepton analysis (right), for data (points with errors), all backgrounds, and signal, after all selection criteria have been applied.

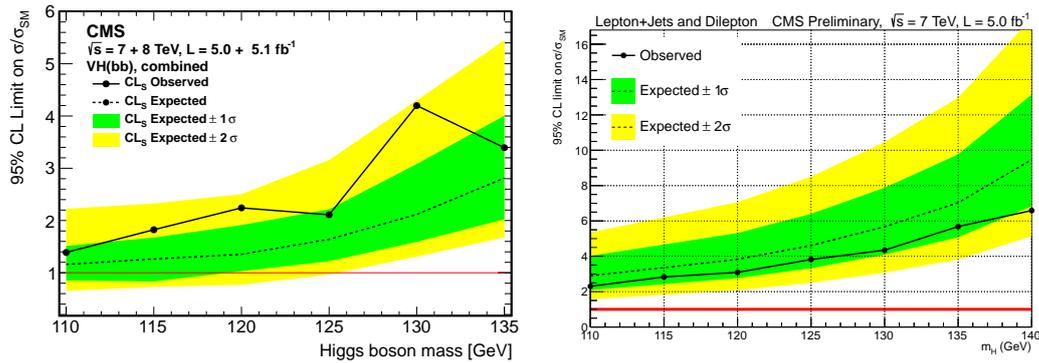
GeV range by a simultaneous maximum-likelihood fit to the nine lepton+jets and dilepton neural network output distributions.

The observed limits at each mass point, the median expected limits, and the  $1\sigma$  and  $2\sigma$  bands are calculated using the modified frequentist method CLs [10] and displayed in Fig. 3 and Table 1. In both analyses, no significant excess of events over the expected background contributions is observed in the fit of the BDT/ANN output distributions. Further studies of these channels with higher integrated luminosities will help to determine if the newly found boson at 125 GeV couples with the b quarks <sup>2</sup>.

**Table 1:** Expected and observed 95% CL upper limits on the product of the Higgs production cross section times the  $H \rightarrow b\bar{b}$  branching ratio for the VH and  $t\bar{t}H$  analyses, with respect to the expectations for a standard model Higgs boson.

m(H)(GeV)	110	115	120	125	130	135	140
Exp. (VH)	1.16	1.26	1.35	1.64	2.12	2.81	-
Obs. (VH)	1.39	1.82	2.24	2.11	4.20	3.39	-
Exp. ( $t\bar{t}H$ )	2.9	3.5	3.8	4.6	5.7	7.0	9.5
Obs. ( $t\bar{t}H$ )	2.3	2.8	3.1	3.8	4.4	5.7	6.6

<sup>2</sup>A recent update for the VH analysis on a data sample of  $12 \text{ fb}^{-1}$  at  $\sqrt{s} = 8 \text{ TeV}$  shows an excess of 2.2 standard deviations at  $m(H)=125 \text{ GeV}$  [11].



**Figure 3:** Expected and observed 95% CL upper limits on the product of the Higgs production cross section times the  $H \rightarrow b\bar{b}$  branching ratio for the VH analysis (left) and  $t\bar{t}H$  one (right), with respect to the expectations for a standard model Higgs boson.

## References

- [1] ATLAS Collaboration, “Combined search for the Standard Model Higgs boson using up to  $4.9 \text{ fb}^{-1}$  of pp collision data at  $\sqrt{s} = 7 \text{ TeV}$  with the ATLAS detector at the LHC”, Physics Letters B 710 (2012), no. 1, 49-66.
- [2] CMS Collaboration, “Combined results of searches for the standard model Higgs boson in pp collisions at  $\sqrt{s} = 7 \text{ TeV}$ ”, Physics Letters B 710 (2012), no. 1, 26-48.
- [3] The TEVNP Working Group for the CDF and D0 Collaborations Collaboration, “Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to  $10.0 \text{ fb}^{-1}$  of Data”, 2012.
- [4] CMS Collaboration, “Search for the standard model Higgs boson decaying to bottom quarks in pp collisions at  $\sqrt{s} = 7 \text{ TeV}$ ”, Physics Letters B 710 (2012), no. 2, 284-306.
- [5] CMS Collaboration, “Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying to bottom quarks for ICHEP 2012.”, CMS Physics Analysis Summary CMS-PAS-HIG-12-019 (2012).
- [6] J. M. Butterworth et al., “Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider”, Phys. Rev. Lett. 100 (2008) 242001.
- [7] CMS Collaboration, “Measurement of the b-tagging efficiency using tt events”, CMS Physics Analysis Summary CMS-PAS-BTV-11-003 (2011).
- [8] CMS Collaboration, “Algorithms for b Jet Identification in CMS”, CMS Physics Analysis Summary CMS-PAS-BTV-09-001 (2009).
- [9] CMS Collaboration, “Search for Higgs boson production in association with a top quark pair in pp collisions”, CMS Physics Analysis Summary CMS-PAS-HIG-12-025 (2012).
- [10] ATLAS and CMS Collaborations, LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, ATL-PHYS-PUB/CMS NOTE 2011-11, 2011/005 (2011).
- [11] CMS Collaboration, “Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying to bottom quarks”, CMS Physics Analysis Summary CMS-PAS-HIG-12-044 (2012).