PROCEEDINGS OF SCIENCE

The search for associated Higgs boson production with Higgs decaying to bb in CMS

Michele DE GRUTTOLA[*†]
University of Florida
E-mail: michele.degruttola@cern.ch

Results are presented on the study of a Higgs boson at 125 GeV decaying into final states consisting of b anti-b quark pair, based on the full statistics of about 25 fb^{-1}, collected in 2011 and 2012 at 7 and 8 TeV, respectively, with the CMS experiment. The $b\bar{b}$ decay channel is studied in associated production with W/Z bosons and in the vector boson fusion process. The combination of the latest CMS H $\rightarrow b\bar{b}$ search for a mass of 125 GeV is an excess of observed events of 2.1 standard deviations, and is consistent with the standard model prediction for Higgs boson production. The observed signal strength corresponding to this excess is 1.0 ± 0.5 relative to the standard model Higgs boson.

The European Physical Society Conference on High Energy Physics
18-24 July, 2013
Stockholm, Sweden

*Speaker.
†A footnote may follow.

PoS(EPS-HEP 2013)273

1. Introduction

The CMS and ATLAS collaborations have reported the discovery of a new boson [1] [2], with a mass, m_H, near 125 GeV and with properties compatible with those of the standard model Higgs boson [3]. To date, significant signals have been observed in channels where the Higgs boson decays into $\gamma\gamma$, ZZ and WW pairs. The interaction with the fermions and whether the Higgs field serves as the source of mass generation in the fermion sector through a Yukawa interaction between the Higgs fields and each fermion has not been firmly established. At a mass of 125 GeV the standard model Higgs boson decays predominantly into a $b\bar{b}$ pair. The observation and study of the H $\rightarrow b\bar{b}$ decay is therefore essential in determining the nature of the newly discovered boson. Recently, in a search for the standard model Higgs boson where the sensitivity below a mass of 130 GeV is dominated by the channels in which the Higgs boson is produced in association with a weak vector boson and decaying to $b\bar{b}$, the CDF and D0 collaborations have reported evidence for an excess of events at a local significance of 3.0 standard deviations for a mass of 125 GeV [4].

The content of this paper is divided in four main sections: after a brief introduction to the CMS reconstruction, we proceed in the description of the H $\rightarrow b\bar{b}$ analysis where the higgs is produced in association with a vector boson and in the vector boson fusion process. Finally in the summary section we'll state the combination of the results and implication for Higgs physics sector.

2. The CMS reconstruction

A detailed description of the CMS experiment can be found elsewhere [5]. The particle-flow algorithm (PF) [6] combines the information from all CMS sub-detectors to identify and reconstruct the individual particles emerging from all vertices: charged hadrons, neutral hadrons, photons, muons, and electrons. These particles are then used to extract the missing transverse energy ($\not{E}_T$), jets, leptons, hadronic τ decays, b jets and to quantify the isolation of leptons and photons. In addition, it allows the identification of the vertex corresponding to the hard-scattering process. Jets are reconstructed from all the PF particles using the anti-k$_T$ jet algorithm [7], with a distance parameter of R = 0.5. Jets originating from b quark hadronization are identified using the combined secondary-vertex b tagging algorithm [8]. This algorithm combines in an efficient 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, gluons and charm quarks. PF $\not{E}_T$ is reconstructed as the opposite of the vectorial sum of the transverse momenta of all particles. The SM Higgs boson events are generated with POWHEG [9] interfaced to PYTHIA [10]. Background processes are generated with MADGRAPH [11]. The generated events are processed through a detailed simulation of the CMS detector based on GEANT4 [12] and are reconstructed with the same algorithms as the ones used for data.

3. H $\rightarrow b\bar{b}$ in association with a vector boson search

Here a search for the standard model Higgs boson in the VH production mode is presented, where V is either a W or a Z boson and H $\rightarrow b\bar{b}$ [13]. The following six channels are included in the search: $W(\mu\nu)H$, $W(e\nu)H$, $W(\tau\nu)H$, $Z(\mu\mu)H$, $Z(ee)H$ and $Z(\nu\nu)H$, all with the Higgs boson decaying to $b\bar{b}$.

PoS(EPS-HEP 2013)273

3.1 Selection

The event selection is based on the kinematic reconstruction of the vector bosons in their leptonic decay modes and of the Higgs boson decay into two b tagged jets. Backgrounds are substantially reduced by requiring a significant boost of the p_T of the vector boson, $p_T(V)$, or the Higgs boson [14]. In that case the two particles recoil away from each other with a large azimuthal opening angle, $\Delta\phi(V,H)$, between them. For each channel, different regions of $p_T(V)$ boost are considered. The low, intermediate, and high boost regions for the $W(\mu\nu)H$ and $W(e\nu)H$ channels are $100 < p_T(V) < 130$ GeV, $130 < p_T(V) < 180$ GeV, and $p_T(V) > 180$ GeV. For the $W(\tau\nu)H$ a $p_T(V) > 120$ GeV region is considered. For the $Z(\nu\nu)H$ channel the low, intermediate and high boost regions are $100 < p_T(V) < 130$ GeV, $130 < p_T(V) < 170$ GeV and $p_T(V) > 170$ GeV, and for the $Z(\nu\nu)H$ channels, the low and high regions are $50 < p_T(V) <100$ GeV and $p_T(V) > 100$ GeV.

The reconstruction of the $\mathrm{H} \to b\bar{b}$ 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, requiring that the value of the CSV discriminator be above a certain threshold. After all event selection criteria the dijet invariant mass resolution of the two b jets from the Higgs decay is approximately 10%. The Higgs boson mass resolution is improved by applying regression techniques similar to those used at the CDF experiment [15]. In the final stage of the analysis, to better separate signal from background under different Higgs boson mass hypotheses a BDT algorithm is trained separately at each mass value using simulated samples for signal and all background processes that pass the event selection. The input variables used are kinematics (dijet transverse momentum and mass, vector boson transverse momentum, etc.), topological (azimuthal angle between the vector boson and dijet, distance in $\eta - \phi$ between Higgs daughters, number of additional jets etc.) and b tag of the two jets. The shape of the output distribution of this BDT discriminant is the final discriminant on which a fit is performed to search for events resulting from Higgs boson production.

3.2 Background control region and systematics

Backgrounds arise from production of W and Z bosons in association with jets (from all quark flavors), singly and pair-produced top quarks ($t\bar{t}$), dibosons and QCD multijet processes. Control regions in data are selected to adjust the event yields from simulation for the main background processes and to estimate their contribution in the signal region

We report also in brief the effects of the most important systematic uncertainties of the analysis. The jet energy scale is varied within one standard deviation as a function of jet p_T and η. The efficiency of the analysis selection is recomputed to assess the variation in yield. Depending on the process, a 2-3% yield variation is found. The effect of the uncertainty on the jet energy resolution is evaluated by smearing the jet energies according to the measured uncertainty. Depending on the process, a 3-6% variation in yields due to this effect is obtained. Data-to-simulation b-tagging scale factors, measured in $t\bar{t}$ events, are applied consistently to jets in signal and background events.These result into yield uncertainties in the 3-15% range, depending on the channel and the specific process. 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. Two calculations are available that estimate the NLO electroweak [16] and NNLO QCD [17] corrections to VH production in the boosted regime. Both the EWK and NNLO QCD corrections are applied to the signal samples.The estimated uncertainties from the NNLO electroweak corrections are 2% for ZH and 2% for WH. For the NNLO QCD correction, an uncertainty of 5% for both ZH and WH is estimated. The uncertainty in the background yields that results from the estimates from data is approximately 10%.

The combined effect of the systematic uncertainties results in an increase of about 15% on the expected upper limit on the Higgs boson production cross section and in a reduction of 15% on the expected significance of an observation when the Higgs boson is present in the data at the predicted standard model rate.

3.3 Results

Results are obtained from combined signal and background fits to the shape of the output distributions of the BDT discriminants trained separately for each channel and for each Higgs boson mass hypothesis in the 110−135 GeV range examined. Figure 1 (left) combines all these discriminants into a single distribution where all events, for all channels, are sorted in bins of similar expected signal-to-background ratio, as given by the value of the output of their corresponding BDT discriminant (trained with a Higgs boson mass of 125 GeV). The observed excess of events in the bins with the largest signal-to-background ratio is consistent with what is expected from the production of a standard model Higgs boson. Probabilities (p-values) that the observed excess is due to background fluctuations alone, as a function of the Higgs boson mass hypothesis, are shown in Fig. 1 (right). For a mass of 125 GeV the excess of observed events is 2.1 standard deviations, and is consistent with the standard model prediction for Higgs boson production. The observed signal strengths for the individual modes are consistent with each other and the value for the signal strength for all modes combined is 1.0 ± 0.5.

4. H $\rightarrow b\bar{b}$ in the vector boson fusion process

A search for the standard model Higgs boson in the vector boson fusion production channel with decay to bottom quarks is reported. A data sample comprising 19.0 fb^{-1} of proton-proton collisions at $\sqrt{s} = 8$ TeV collected during the 2012 running has been analyzed [18]. The prominent feature of the searched signal is the presence of four energetic hadronic jets generated from the qqH $\rightarrow$ qqbb final state. Two jets are expected to originate from a light-quark pair (u,d-type), that are typically two valence quarks from each of the colliding protons scattered away from the beam line and inside the detector acceptance by the VBF process. These VBF "tagging" jets are therefore roughly expected in the forward and backward direction with respect to the beam line. Two additional jets are expected from the Higgs boson decay to a b-quark pair, in more central regions of the detector with respect to the VBF tagging pair. Another important property of signal events is that, being produced in a VBF process, no QCD color is exchanged in the production. As a result very little additional QCD radiation and hadronic activity is expected in the space outside the color-connected regions, in particular in the whole rapidity interval (rapidity gap) between the two tagging jets, with the exception of the Higgs boson decay products.

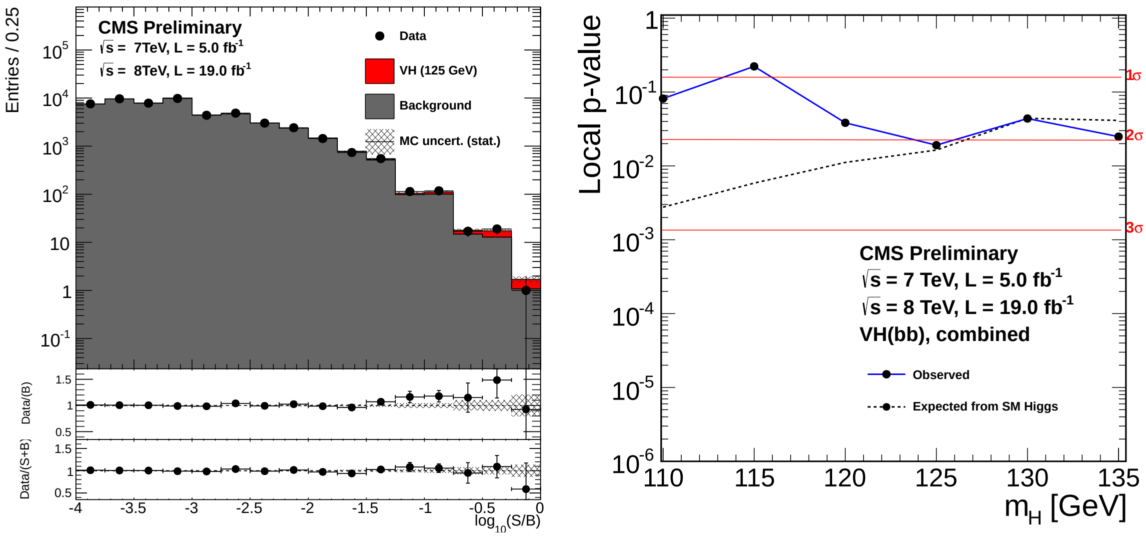

Figure 1: Left: Combination of all BDT discriminants of the VHbb analysis into a single distribution where all events, for all channels, are sorted in bins of similar expected signal-to-background ratio, as given by the value of the output of their corresponding BDT discriminant. Right: Observed and expected p-value 1-CL$_b$, and the corresponding significance in number of standard deviations for CMS VHbb analysis.

The analysis strategy is to select four-jet events with progressive p_T thresholds that are optimized on the jet p_T distributions expected for the signal. One or two jets should be significantly b-tagged, while the two less b-tagged jets should have a large invariant mass and be well separated in pseudorapidity.

The preselected events are further characterized by the response of an artificial neural network (ANN), trained to separate signal events from background ones, without making use of kinematic information of the two b-tagged jets. Besides the kinematics of the VBF tagging jets, the ANN uses the CSV b-tagging values of the two most b-tagged jets, the output of a quark-gluon jet discriminator applied to the two less b-tagged jets, and measurements of the additional hadronic activity in the event described. In order to improve the resolution of the b jets invariant mass, a jet-by-jet correction factor is derived by combining various jet properties in a multivariate regression analysis similar to the one used in the previous described search.

Events are selected requiring four PF jets with $p_T > 85$, 70, 60, 40 GeV. The four jets are ordered in pairs labeled "bb" and "qq" alternatively with b-tag ordering, where the "qq" pair is made with least CSV b-tagged jets, and with η ordering, where the "qq" pair is the most η-separated jet pair. For both "qq" orderings the event selection further requires $m_{qq} > 300$ GeV and $\Delta\eta_{qq} > 2.5$. Finally, to remove the large QCD contribution of back-to-back bb pairs, events are required to satisfy $\Delta\Phi_{bb} < 2$, for the b-tag ordered jet pair only.

4.1 Background control region and systematics

The search strategy relies on a background fit of the m_{bb} spectrum in four categories after the ANN cut. In each ANN event category the m_{bb} spectrum is fitted with a background template with three parts: a fifth degree Bernstein polynomial representing the non-peaking QCD background, a Z/W template taken from simulation, and a top template that combines the $t\bar{t}$ and single-top contributions, also taken from simulation. The parameters of the polynomial are floating, while the normalization of the Z/W and top components are fixed to the SM cross sections. The three parts form the model for the null hypothesis, as opposed to the signal hypothesis, which contains in addition the signal template with free normalization. The fits are performed in the 70 GeV < m_{bb} < 250 GeV mass range.

The by far most dominant background to the search is the QCD production of multijets. Other relevant backgrounds arise from: (i) hadronic decays of Z or W bosons in association with additional QCD jets, (ii) hadronic or semi-leptonic decays of top-pairs, and (iii) hadronic decays of single-top productions. The additional contribution of Higgs bosons produced in GF processes, with two or more associated QCD jets, also plays an important role in establishing the expected signal yields.

Due to the largely data driven way in which this analysis is conducted, the main sources of systematic uncertainties on the VBF Higgs boson signal are related to:

- the simulation of the trigger efficiency, resulting in an absolute trigger efficiency uncertainty of 5% in the least significant ANN category, increasing to 8% in the most significant ANN category.
- the signal acceptance being affected by jet energy scale uncertainty (10% effect), jet energy resolution uncertainty (2% effect, uncertainty on simulation of b-tagging, and quark/gluon jet discriminator.
- the shape of the simulated signal and Z+jet m_{bb} mass templates used in the final fit, affected by jet energy scale and resolution uncertainty

4.2 Results

The models representing the two hypotheses of background only and background+signal are fitted to the data simultaneously in the four categories. Figure /reffig:VBFbbmass (left) shows the most sensitive category. CLs 95% limits are computed, taking into account the systematic uncertainties. Figure /reffig:VBFbbmass (right) shows the expected and observed 95% C.L. limit on the signal strength, as a function of the Higgs boson mass, which ranges from ~ 2 at m_H = 115 GeV to ~ 4 at m_H = 135 GeV, together with the expected limits in the presence of a SM Higgs boson with mass 125 GeV. For a 125 GeV Higgs boson signal the fitted signal strength is $\mu = \frac{\sigma}{\sigma_{SM}} = 0.7 \pm 1.4$. Because of the small signal to background ratio, the limit is dominated by the statistical uncertainty on the background, while the inclusion of all systematic uncertainties worsens the limit by $\sim 15\%$.

The CMS VH and VBF searches for Higgs decaying to bottom quarks have been preliminary combined. The VBF contribution adds 5% to 10% sensitivity to the exclusion limits and significance to the VH results. For a mass of 125 GeV the excess of observed (expected) events is 2.0 (2.1) standard deviations for the combined results.

5. conclusions

A search for the standard model Higgs boson decaying to bb when produced in association with an electroweak vector boson and in the VBF topology has been reported. The search is

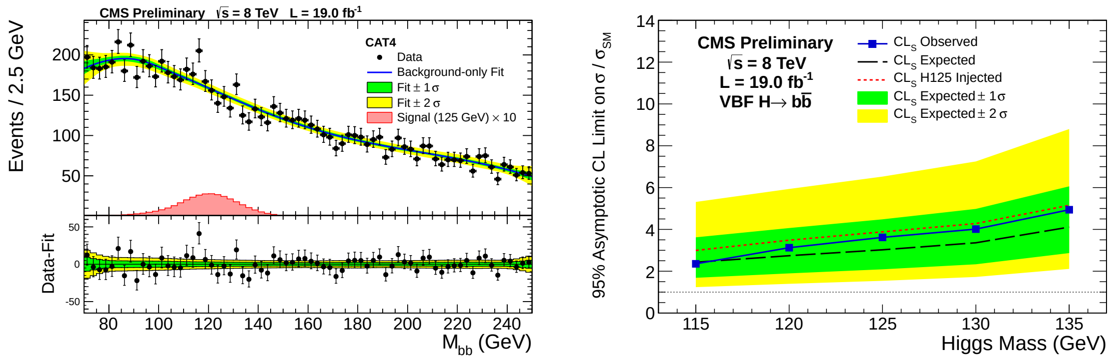

Figure 2: Left: fit of the background model to the data m_{bb} distributions of preselected events four. The top panels show the fitted curves and the amplified distribution of the searched SM signal, in red. The bottom panels show the data minus fit residuals, together with the fit one and two sigma uncertainty bands. Right: Expected and observed 95% confidence level limits on the signal cross section in units of the SM expected cross section, as a function of the Higgs boson mass, including all four higher ANN event categories. The limits expected in the presence of a SM Higgs boson with mass 125 GeV are indicated by the dotted curve.

performed in data samples corresponding to integrated luminosities of 5.0 fb^{-1} at $\sqrt{s} = 7$ TeV and up to 19.0 fb^{-1} at $\sqrt{s} = 8$TeV, recorded by the CMS experiment at the LHC. An excess of events is observed above the expected background with a significance of 2.0 standard deviations. The expected significance when taking into account the production of the standard model Higgs boson is 2.1 standard deviations. The signal strength corresponding to this excess, relative to that of the standard model Higgs boson, is $\mu = 1.0 \pm 0.5$.

References

[1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Phys.Lett.B (2012) doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Phys.Lett.B (2012)doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235

[3] P. W. Higgs, "Broken symmetries, massless particles and gauge Fields", Phys. Lett. 12 (1964) 132, doi:10.1016/0031-9163(64)91136-9.

[4] CDF Collaboration, D0 Collaboration "Higgs Boson Studies at the Tevatron", arXiv:1303.6346.

[5] CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 0803:S08004 2008.

[6] CMS Collaboration, "Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, (2009).

[7] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm", JHEP 04 (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[8] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", CMS Physics Analysis Summary CMS-PAS-BTV-12-001 (2012).

[9] P. Nason and C. Oleari, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG", JHEP 02 (2010) 037, doi:10.1007/JHEP02(2010)037.

[10] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", JHEP 05 (2006) 026, doi:10.1088/1126-6708/2006/05/026.

[11] J. Alwall et al., "MadGraph/MadEvent v4: the new web generation", JHEP 09 (2007) 028, doi:10.1088/1126-6708/2007/09/028, arXiv:0706.2334.

[12] GEANT4 Collaboration, "GEANT4â a simulation toolkit", Nucl. Instrum. Meth. A 506 (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[13] CMS Collaboration, "Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying to bottom quarks", CMS-PAS-HIG-13-012, 2013

[14] J. M. Butterworth et al. "Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider", Phys. Rev. Lett. 100 (2008) 242001, doi:10.1103/PhysRevLett.100.242001.

[15] T. Aaltonen et al. "Improved b-jet Energy Correction for H $\rightarrow b\bar{b}$ Searches at CDF", arXiv:1107.3026.

[16] M. Ciccolini et al., "Strong and electroweak corrections to the production of Higgs+2jets via weak interactions at the LHC", Phys. Rev. Lett. 99 (2007) 161803, doi:10.1103/PhysRevLett.99.161803, arXiv:0707.0381.

[17] G. Ferrera, M. Grazzini, and F. Tramontano, "Associated WH production at hadron colliders: a fully exclusive QCD calculation at NNLO", arXiv:1107.1164.

[18] CMS Collaboration, "Search for the standard model Higgs boson produced in the vector boson fusion, and decaying to bottom quarks", CMS-PAS-HIG-13-011, 2013

PoS(EPS-HEP 2013)273