PROCEEDINGS OF SCIENCE

Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a b-quark pair with the ATLAS detector at the LHC

Giacinto Piacquadio*
On behalf of the ATLAS Collaboration
CERN
E-mail: giacinto.piacquadio@cern.ch

These proceedings report on recent progress in the search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a b-quark pair using 4.6-4.7 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC. No evidence for Higgs boson production is observed. Exclusion limits on Higgs boson production times decay branching ratio, at the 95% confidence level, of 2.5 to 5.5 times the Standard Model expectation are obtained in the mass range 110–130 GeV. The expected exclusion limits range between 2.6 and 5.1 for the same mass interval.

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

*Speaker.

PoS(ICHEP2012)061

1. Introduction

The observation of decays of Higgs bosons to a pair of b-quarks is an important step for establishing whether the new observed particle with an invariant mass of ≈ 125 GeV [1] is the Higgs boson predicted by electroweak symmetry breaking in the Standard Model or not. These proceedings report on the search for the Higgs boson in this decay mode based on a dataset of 4.6$-$4.7 fb^{-1} of $\sqrt{s} = 7$ TeV collision data provided by the LHC [2] and recorded by the ATLAS detector [3]. A more detailed description of the analysis can be found in [4]. Despite the high branching ratio of Higgs boson decays to a pair of b-quarks predicted in the Standard Model (*BR* 58% for $m_H = 125$ GeV), the observation of a signal in this decay channel is very challenging, predominantly because of the copious amount of backgrounds with b-jets in the final states. To reduce such backgrounds to a managable level and at the same time provide a way to trigger the events of interest, only the production modes where the Higgs boson is produced in association with a vector boson are considered, with three signatures, meant to select $Z \to \ell\ell$, $W \to \ell\nu$, $Z \to \nu\nu$: events are triggered based on the leptonic signature and the corresponding channels are labelled as 2-, 1- and 0-lepton channels in the present article. In addition to the leptonic part of the event selection, the analysis relies on identifying two b-quark jets and looking for an excess over the background expectation corresponding to a broad peak in the $b\bar{b}$ invariant mass distribution ($\sigma(b\bar{b}) \approx 15$ GeV). The signal-over-background ratio can be enhanced by requiring a W or a Z boson candidate with high transverse momentum: different categories are therefore considered, corresponding to different intervals of the vector boson transverse momenta, p_T^W or p_T^Z.

2. Data and selection strategy

Events with leptons (muons or electrons) are triggered by single electron and muon triggers with offline p_T thresholds of 25 GeV and the pseudo-rapidity limited to $\eta < 2.5$. In the two lepton channel, these triggers are supplemented with a di-electron trigger with a threshold of 12 GeV, which are then exploited offline by reducing the p_T threshold to 20 GeV. Efficiencies with respect to offline event selection are close to 100% for the electron based channels, while for the muon based channels they are $\approx 90\%$ for the 1-lepton channel and $\approx 95\%$ for the 2-lepton channel. The 0-lepton channel relies on a missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) trigger, which has an efficiency above 50% for the offline selection cut of $E_{\mathrm{T}}^{\mathrm{miss}} > 120$ GeV and is nearly fully efficient for $E_{\mathrm{T}}^{\mathrm{miss}} > 170$ GeV. The basic selection cuts for all three channels are summarised in Table 1. The definitions of signal and loose leptons vary across channels, and a more detailed description can be found in [4]. Jets are reconstructed from calorimeter energy deposits, using an anti-kt clustering algorithm with radius parameter $R = 0.4$. Jets from pile-up are suppressed by requiring at least 75% of the summed momenta of tracks matched to the jet to be associated with the primary event vertex.

The jet and (additional) lepton vetoes play an important role in reducing the copious top pair background, in both 0- and 1-lepton channels. A similar role is played in the 2-lepton channel by the $83 < m_{\ell\ell} < 99$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} < 50$ GeV requirements, which suppress top pair events with 2 leptons in the final state. The p_T^V intervals corresponding to the signal categories defined in the three channels and additional topological cuts which depend on p_T^V are shown in Table 2. In the 0-lepton channel, in addition to the calorimeter based definition of $E_{\mathrm{T}}^{\mathrm{miss}}$, a track-based definition (relying

Table 1: The basic event selection of the three channels.

Object	0-lepton	1-lepton	2-lepton
Leptons	0 loose leptons	1 signal lepton + 0 loose leptons	2 signal leptons
Jets	2 b-jets ($p_T^1 > 45$ GeV,$p_T^2 > 25$ GeV,$\|\eta\| < 2.5$) $+ \leq 0$ loose jets ($p_T > 25$ GeV,$\|\eta\| < 2.5$)	2 b-jets ($p_T^1 > 45$ GeV,$p_T^2 > 25$ GeV,$\|\eta\| < 2.5$) $+ \leq 0$ loose jets ($p_T > 20$ GeV,$\|\eta\| < 4.5$	2 b-jets - -
MET	$E_{\mathrm{T}}^{\mathrm{miss}} > 120$ GeV $p_T^{miss} > 30$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} < 50$ GeV
Vector Boson Mass	-	$m_T^W > 40$ GeV	$83 < m_{\ell\ell} < 99$ GeV

Table 2: Categories in p_T^W or p_T^Z intervals (in GeV) and additional topological cuts in each category.

0-lepton channel				
p_T^Z [GeV]	120-160	160-200	>200	
$\Delta R(j_1, j_2)$	$0.7 < \Delta R$	$0.7 < \Delta R$	-	
$\Delta\phi(V,H)$	> 2.7	> 2.9	> 2.9	
$\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, p_T^{miss})$	$< \pi/2$			
$\mathrm{Min}[\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, jets)]$	> 1.8			
1- and 2-lepton channels				
p_T^V [GeV]	0-50	50-100	100-200	>200
$\Delta R(j_1, j_2)$	$0.7 < \Delta R$	$0.7 < \Delta R$	$0.7 < \Delta R$	-

on charged particles only) is considered, which is labeled as p_T^{miss}. Cuts on $\Delta R(j_1, j_2) > 0.7$, i.e. on the pseudo-rapidity difference between the two b-jets, are placed to cut out the phase space region which is dominated by close-by gluon splittings in $b\bar{b}$ and therefore is not particularly easy to model by the Monte Carlo generators, except for the highest p_T^V region, where this cut would reduce a good part of the signal and the modelling issue is less relevant: one of the reasons is that the b-quark mass plays a less important role at such high transverse momenta. In the 0-lepton analysis additional cuts are placed to further reduce the backgrounds, in particular multi-jet backgrounds from non real $E_{\mathrm{T}}^{\mathrm{miss}}$.

3. Backgrounds

The analysis relies on testing the signal hypothesis based on the $m_{b\bar{b}}$ distribution for the signal and the backgrounds within the $(80-150)$ GeV mass region (this region is referred to as signal region in the following). These distributions are considered separately for the three channels and the different p_T^V categories, to exploit their different sensitivities. Fig. 1 shows the $m_{b\bar{b}}$ invariant mass distribution after applying the full analysis selection, for all three channels. For the sake of simplicity, in these plots no distinction is made in p_T^V categories. The Monte Carlo predictions for the various backgrounds are scaled to their data-driven estimates, which will be described in the following, while the shape is obtained from Monte Carlo predictions, except for the multi-jet

background in the 1-lepton channel, for which both normalisation and shapes are derived from data.

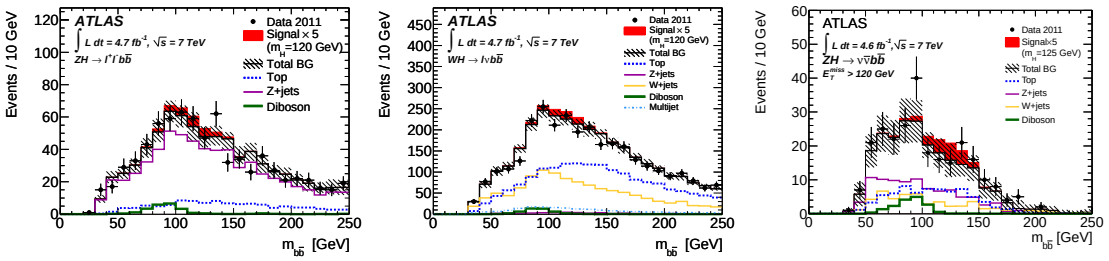

Figure 1: The invariant mass $m_{b\bar{b}}$ for the 2-lepton (left), 1-lepton (middle) and 0-lepton (right) channels [4]. The signal distributions are shown for $m_H = 120$ GeV and are enhanced by a factor of 5 for visibility. The shaded area indicates the total uncertainty on the background prediction. For better visibility, the signal histogram is stacked onto the total background, unlike the various background components which are simply overlaid in the figure.

As can be observed in the figure, the main background for the 2-lepton analysis is the Z+jet continuum, while for the 1-lepton analysis both the W+jet and top backgrounds are very important, where top includes both top pair and single top production. In the 0-lepton channel, the most important backgrounds are Z+jet, W+jet and top pair production, the first being irreducible, the second and third passing the selection due to leptons escaping detection. In addition to these, a sub-leading background for all channels is also the diboson production (WW, ZZ, WZ) and in the 1-lepton channel a significant amount of multi-jet background, mostly in the lowest p_T^W categories.

Given that the expected signal-over-background ratios are low (from 1% to 15% depending on channel and p_T^V category), the correct estimation of the backgrounds, both in terms of shapes ($m_{b\bar{b}}$ and p_T^V) and overall normalisation, and moreover of their related uncertainties, is crucial. The overall normalisation of the main backgrounds is entirely based on data. To make this possible, additional control regions are defined:

- The sidebands of the $m_{b\bar{b}}$ distribution of the 1- and 2-lepton channels (in the 20-80 GeV and 80-250 GeV regions).
- A top control region for the 1-lepton channel, where exactly three jets are required rather than exactly two.
- A top control region for the 2-lepton channel, where $m_{\ell\ell}$ is required not to be compatible with m_Z and $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV is required.

In the 1-lepton channel, the left sideband of the $m_{b\bar{b}}$ distribution is dominated by W+jet events, while the right sideband is dominated by top pair and single-top events. In the 2-lepton channel, both the left and right sideband of the same distribution is dominated instead by Z+jet events, with a minor presence of di-leptonic top pair events. A simultaneous fit to the data of the overall normalisations of the W+jet, Z+jet and top (top pair + single-top) backgrounds in all defined control regions is performed and the result is used to rescale the Monte Carlo predictions in the signal region accordingly. These rescaling factors are also used for the 0-lepton channel, after cross-checks performed on additional control regions.

4. W/Z+jet backgrounds: flavour fractions and modelling uncertainties

The identification of b-jets is crucial to suppress backgrounds with light and c-jets in the final state. The analysis relies on a Neural Network based discriminator which combines several b-jet identification algorithms, a 3d impact parameter, an inclusive vertex finder and a multi-vertex $b \to c$-hadron decay chain fit. Taking top pair events as a benchmark, a light-jet (u,d,s) and c-jet fake rate of respectively 0.6% and 20% is obtained for a b-jet efficiency of 70%.

The W/Z+jet backgrounds are known to be theoretically not perfectly modelled. This potentially affects both normalisation and differential distributions of their separate flavour components. Besides the overall normalisation, the fractions of $W(Z)$+b,$W(Z)$+c and $W(Z)$+light-jet events are therefore estimated from data: the b- and c-jet fractions are defined by the presence of at least one b- or, if the b-jet is missing, one c-jet among the two jets used to build the Higgs boson candidate. The flavour fractions for W+jet are determined within the 1-lepton channel (using events from the left sideband of the $m_{b\bar{b}}$ distribution), while the flavour fractions for Z+jet are determined within the 2-lepton channel (exploiting both the left and right $m_{b\bar{b}}$ sidebands). In both cases, the fractions are extracted with a fit to the data of the flavour weight (b-tagging discriminator) templates expected for the different flavour fraction contributions, either on events with one only b-tagged jet, based on the distribution of the flavour weight of the second jet, or in events where b-tagging has been released completely, based on the flavour weight of the first two jets.

After b-tagging is applied, the W+jet and Z+jet backgrounds are mostly given by the irreducible $W+b\bar{b}$ and $Z+b\bar{b}$ contributions. Apart from the overall normalisation, a possible mismodelling of the $m_{b\bar{b}}$ and p_T^V distributions has been considered as well: systematic variations for these distributions have therefore been derived, for $W+b\bar{b}$ based on the difference among various models (Alpgen+Herwig, Powheg+Pythia, Powheg+Herwig, aMC@NLO+Herwig) and for $Z+b\bar{b}$ based on the difference between data and Monte Carlo predictions in the $m_{b\bar{b}}$ sidebands of the signal region.

5. Results

The number of selected data events in the signal region is shown in Table 3 for each channel and each p_T^V category. The expected number of signal events for $m_H = 120$ GeV is also shown, along with the corresponding estimated number of background events. The signal sample has been normalised to the inclusive NNLO QCD+NLO EW computation. Differential EW corrections as a function of p_T^V, that reduce the cross section at high p_T^V, have been applied as well.

The statistical analysis of the data employs a likelihood function constructed as the product of Poisson probability terms for each channel and each p_T^V category and of the $m_{b\bar{b}}$ binned PDF (probability density function) in the $80-150$ GeV region for each category. In order to test the presence of a signal, the test statistics is constructed according to the profile likelihood ratio, which is then used to measure the compatibility of the data with the background only hypothesis and to derive exclusion intervals with the CL_s method.

The analysis sensitivity depends crucially on the background shape and normalisation uncertainties in the signal region. The main background normalisation uncertainties are induced by theory modelling uncertainties, dominated by the W+jet and Z+jet modelling uncertainty ($\Delta B \approx$

Table 3: Number of data, simulated signal, and estimated background events in each bin of p_T^V for the 1-, 2- and 0-lepton channels [4].

bin	2-lepton p_T^V [GeV]				1-lepton p_T^V [GeV]				0-lepton p_T^V [GeV]		
	0-50	50-100	100-200	>200	0-50	50-100	100-200	>200	120-160	160-200	>200
Number of events for $80 < m_{b\bar{b}} < 150$ GeV											
signal	1.3 ± 0.1	1.8 ± 0.2	1.6 ± 0.2	0.4 ± 0.1	5.0 ± 0.6	5.1 ± 0.6	3.7 ± 0.4	1.2 ± 0.2	2.0 ± 0.2	1.2 ± 0.1	1.5 ± 0.2
top	17.4	24.1	7.3	0.2	229.9	342.7	201.3	8.2	35.2	8.3	4.1
W+jets	–	–	–	–	285.9	193.6	85.8	17.5	13.2	7.8	4.8
Z+jets	123.2	119.9	55.9	6.1	11.1	10.5	2.8	0.0	31.5	11.9	7.1
diboson	7.2	5.6	3.6	0.7	12.6	11.9	7.8	1.4	4.6	4.3	3.6
multijet	–	–	–	–	55.5	38.2	3.6	0.2	–	–	–
total BG	148 ± 10	150 ± 6	67 ± 4	6.9 ± 1.2	596 ± 23	598 ± 16	302 ± 10	27 ± 5	85 ± 8	32 ± 3	20 ± 3
data	141	163	61	13	614	588	271	15	105	22	25

$1-15\%$), by the b-tagging uncertainty ($\Delta B \approx 1-7\%$), by the limited statistics in the $m_{b\bar{b}}$ sidebands and top control regions ($\Delta B \approx 2-5\%$) and jet related uncertainties ($\Delta B \approx 2-12\%$), resulting in a total background uncertainty of $\approx 3-20\%$, where the exact value depends on the channel and p_T^V category considered. The total signal uncertainty, dominated by b-tagging, jet and theory related uncertainties, ranges from 9% to 17%, depending again on channel and p_T^V category.

The resulting exclusion limits are listed in Table 4 for each channel and for the combination of the three channels. The limits are expressed as the multiple of the SM Higgs boson production cross section which is excluded at 95% CL for each value of the Higgs boson mass.

Table 4: The observed and expected 95% CL exclusion limits on the Higgs boson cross section for each channel, expressed in multiples of the SM cross section as a function of the hypothesized Higgs boson mass [4].

Mass	2-lepton		1-lepton		0-lepton		Combined	
[GeV]	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
110	7.7	6.0	3.3	4.2	3.7	4.0	2.5	2.5
115	7.7	6.2	4.0	4.9	3.6	4.2	2.6	2.7
120	10.4	8.0	4.9	5.9	4.8	5.0	3.4	3.3
125	11.6	9.1	5.5	7.5	7.3	6.0	4.6	4.0
130	14.4	11.6	5.9	9.2	10.3	7.6	5.5	4.9

References

[1] ATLAS Collaboration, **Phys. Lett. B** 716 (2012) 1-29.

[2] L. Evans, P. Bryant (editors), **JINST** (2008) 3 S08001.

[3] ATLAS Collaboration, **JINST** (2008) 3 S08003.

[4] ATLAS Collaboration, **Phys. Lett. B** 718 (2012) 369-390.