SEARCH FOR A HIGGS BOSON IN ASSOCIATION WITH A PAIR OF TOP QUARKS WITH THE ATLAS DETECTOR*

Tim Michael Heinz Wolf

Nikhef, Science Park 105, 1098XG Amsterdam, The Netherlands

(Received February 5, 2018)

The production of a Higgs boson in association with a pair of top quarks ($t\bar{t}H$) is one of the main Higgs production channels at the LHC which is yet unobserved. In the recent publications, the ATLAS Collaboration claims evidence for the process using the data collected in 2015 and 2016 at the LHC. The process offers direct sensitivity to the top-Yukawa coupling which is the largest fermion coupling to the Higgs. Any deviation of the top-Yukawa coupling measured directly and compared to indirect constraints would provide a hint towards new physics. In the context of $t\bar{t}H$, the $H \to b\bar{b}$ decay is of particular interest due to the large branching ratio and it is, therefore, potentially offering high sensitivity to the top-Yukawa coupling. The goal of this paper is to summarize the recent progress in $t\bar{t}H(b\bar{b})$ analysis contrasting the strategy employed by ATLAS in 2016 with the new publication in 2017.

DOI:10.5506/APhysPolBSupp.11.257

1. Introduction

In this document, the status of the $t\bar{t}H(b\bar{b})$ search performed by ATLAS using 36.1 fb^{-1} of data collected in 2015 and 2016 [1] is summarized and contrasted with the previous search which only used 13.1 fb^{-1} [2]. The focus lies here in pointing out the differences in the analysis strategies comparing the currently existing analysis with the previous one.

The $t\bar{t}H(b\bar{b})$ search is interesting since the final state of $t\bar{t}H$ offers sensitivity to the top-Yukawa coupling which can be seen from the tree-level production diagram shown in Fig. 1 (a). The top-Yukawa is the largest coupling of the Standard Model (SM) Higgs to a fermion and a measurement of the $t\bar{t}H$-production cross section would offer direct sensitivity to it. It would

* Presented at the Final HiggsTools Meeting, Durham, UK, September 11–15, 2017.

provide a consistency check with indirect constraints on the top-Yukawa coupling. Furthermore, the $H \to b\bar{b}$ decay offers the highest branching ratio of all Higgs decay modes which makes this channel particularly attractive. The search is carried out in the single-lepton and dilepton channel referring to the $t\bar{t}$-decay modes where we focus on the single-lepton channel.

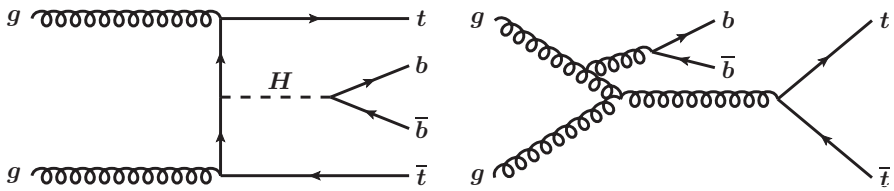


Fig. 1. Tree-level production diagrams for (a) $t\bar{t}H(b\bar{b})$ signal and (b) the main background $t\bar{t}b\bar{b}$ [1].

2. Analysis for $t\bar{t}H(b\bar{b})$ in single-lepton channel

The search for $t\bar{t}H(b\bar{b})$ is a challenging final state owing the high multiplicity of hadronic objects, especially jets originating from a b quark (b jets). As can be inferred from Fig. 1 (a), there are up to 4 b jets in the final state coming from the decay of the top quarks and the Higgs decay products. The difficulty in the analysis is driven by the overwhelming irreducible backgrounds ($t\bar{t}b\bar{b}$ *cf.* Fig. 1 (b)) which mimic the signal. Moreover, challenges in the modelling of the backgrounds in Monte Carlo (MC) simulation is a limiting factor in the analysis.

In this section, we will discuss the event selection, how to identify b jets, and the statistical treatment.

2.1. Event selection

The dataset was taken in pp collisions by the ATLAS detector in 2015 and 2016. After requiring standard quality measures from the data this amounts to 36.1 ± 0.8 fb^{-1}. Furthermore, requirements on the number of reconstructed jets and the number of reconstructed b jets as well as the number and type of leptons in the event define the event selection [1]. More details concerning the identification of b jets are given in Section 2.2. The semileptonic top decay which is studied here offers the possibility to trigger on a lepton, which is experimentally beneficial to select events.

The idea behind an event selection is to enrich the sample on which the analysis is carried out in the $t\bar{t}H$ process by selecting only potentially interesting events. The selection does not strictly require 4 b jets, since

there is the possibility that jets fall out of acceptance of b tagging, which can lead to a migration of signal events to lower b-tag multiplicities. In total, a fraction of 8.7% of the simulated $t\bar{t}H(b\bar{b})$ events pass the event selection.

2.2. Identification of jets originating from b quarks

Since already in the event selection b jets are required, this raises the question how a b jet is defined. b jets leave a distinct signature in the detector since b quarks have a considerable life-time and, therefore, a secondary displaced vertex is observable in the detector. This fact offers the possibility to study and extract different quantities for a jet: the location of the secondary vertex and the impact parameter. Furthermore, the decay chain inside a jet can be explored which gives another handle to identify b jets. These information are used and combined in a boosted decision tree (BDT) of which the response can be seen in Fig. 2 (a), where `MV2c10 BDT Output` corresponds to the BDT output which targets the discrimination between b jets, c jets and light-flavoured jets.

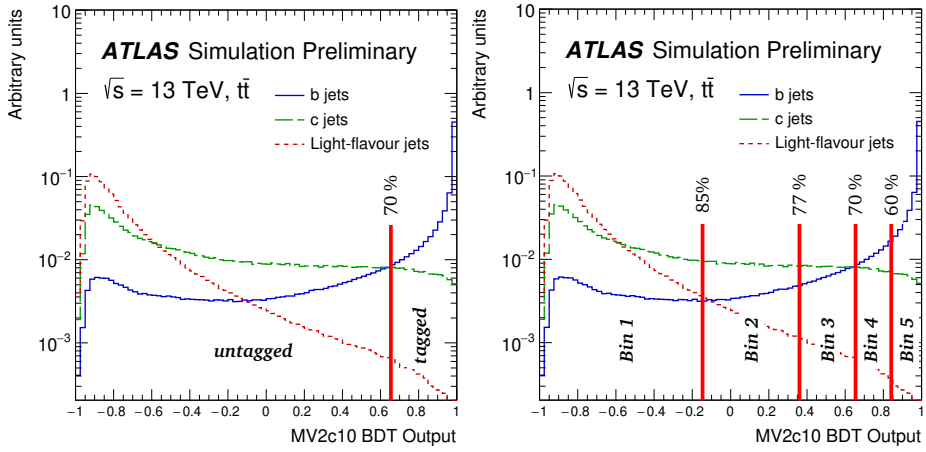


Fig. 2. Contrasting direct for b tagging (a) and pseudo-continuous approach (b) [3]. The quantity `MV2c10 BDT Output` is the output of a boosted decision tree (BDT), aiming to discriminate the flavour of a jet. Each vertical line corresponds to the cut values employed and the efficiency is indicated. *Tagged* in (a) means identified as a b jet and *untagged* means identified as a non-b jet.

It can be seen that for high values of the BDT output, the distribution of b-jet peaks. However, it falls off for c jets and light-flavoured jets, making it a good discriminant to identify b jets. If only 2 bins of the distribution (*tagged* and *untagged*) are used for the decision, such as in Fig. 2 (a), this is

called direct tagging. The decision where to place the cut which divides the distribution into two bins is driven by the efficiency to identify b jets (here 70%) and the rejection of c jets and light-flavoured jets.

Figure 2 (b) shows the same quantity but more bins are placed on the BDT output score which aims at better incorporating the shape of the distribution. This configuration is called pseudo-continuous (PC) b tagging. PC b tagging offers a finer granularity to decide if jets are considered as b jets or not. This has the consequence that less events are rejected, since events where jets are *untagged* in direct tagging might be recovered in *Bin 2* or *Bin 3* in the PC version.

2.3. $t\bar{t}H$ analysis contrasting strategy 2016 and 2017

The goal of the analysis is the extraction of the $t\bar{t}H$ signal strength defined as

$$\mu_{t\bar{t}H} = \frac{\sigma^{t\bar{t}H}_{\text{observed}}}{\sigma^{t\bar{t}H}_{\text{SM}}}\,, \tag{1}$$

where $\sigma^{t\bar{t}H}_{\text{observed}}$ refers to the observed $t\bar{t}H$ cross section and $\sigma^{t\bar{t}H}_{\text{SM}}$ denotes the SM expectation. A significant deviation of the expected SM cross section would result in a deviation of $\mu_{t\bar{t}H}$ from 1.

In order to achieve that goal, it is beneficial to define regions which are enriched in signal, $t\bar{t}H$. These regions are called signal regions (SR). However, since the backgrounds are large, there is the additional desire to define regions which extract information about these backgrounds as well as possible in order to control them. These regions are called control regions (CR).

In order to define SRs and CRs in the 2016 analysis and the 2017 analysis, different approaches were taken. In the 2016 version of the analysis, direct tagging was used as illustrated by Fig. 2 (a) and then SR and CR were defined according to their jet and b-jet multiplicity.

In the new analysis, PC b tagging is used which offers the possibility to target SRs with a high signal purity and CRs for dominant main backgrounds such as $t\bar{t}+\geq 1b$, $t\bar{t}+\geq 1c$, $t\bar{t}+$light. Figure 3 illustrates how the regions in the new analysis are formed where the focus lies on maximization of signal over background (S/B) ratio in the SRs and a homogeneous composition of the CRs, enriched in a dedicated background.

Figure 4 (a) and Fig. 4 (b) show the S/B ratio for both analysis. One sees that in the 2017 version of the analysis, there are more SRs (6, coloured in grey) compared to the 2016 version (3, coloured in black/red). Further, one can see that the purity in S/B is $\ll 10\%$ for all regions in both analysis owing to the low signal cross sections and the very similar topology of the backgrounds compared to the signal. However, by comparing in more detail

we see that the purity in the most sensitive SR, SR1, in the 2017 analysis increased to 5.4% from 5.2% in 2016. The $S/\sqrt{B}$ ratio is also slightly improved in SR1.

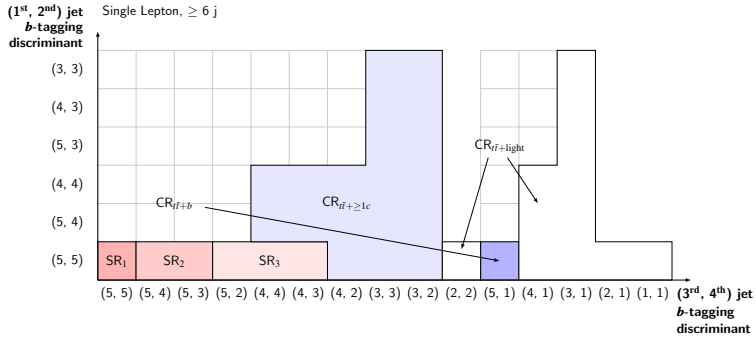
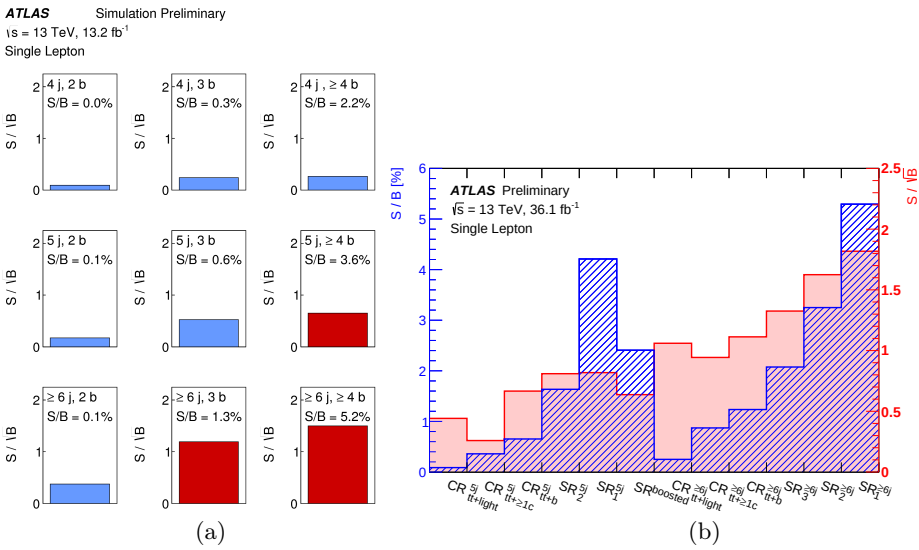


Fig. 3. Illustration of the region definition of the 2017 analysis for the categories involving ≥ 6 jets [1]. The b-tagging discriminant for each jet should be identified with the bins in Fig. 2 (b).

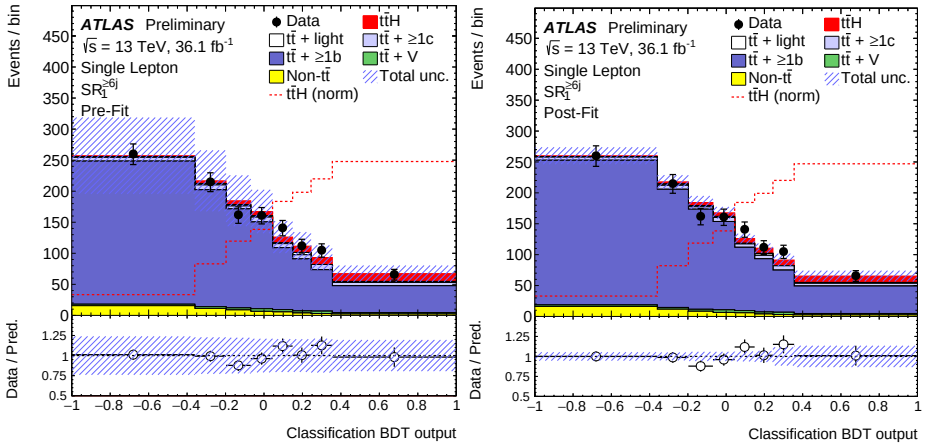


(a) (b)

Fig. 4. (Colour on-line) S/B of the analysis regions for the 2016 analysis [2] (a) and the 2017 analysis [1] (b). The signal regions in (a) are indicated in black/red.

In order to separate signal from background, a powerful discriminant is built which is referred to as the classification BDT. It combines both kinematic and topological information about the event and it is trained in different configurations for the different SRs to optimize its performance.

Figure 5 (a) shows the classification BDT in the most sensitive SR. It can be seen that the bins with high BDT values are more populated by the $t\bar{t}H$ signal.



(a) (b)

Fig. 5. Classification BDT in the most sensitive signal region (SR1) (a) pre-fit distribution and (b) post-fit distribution. The goal is the separation of signal and background. The dashed line shows the $t\bar{t}H$ signal distribution normalized to the total background prediction [1].

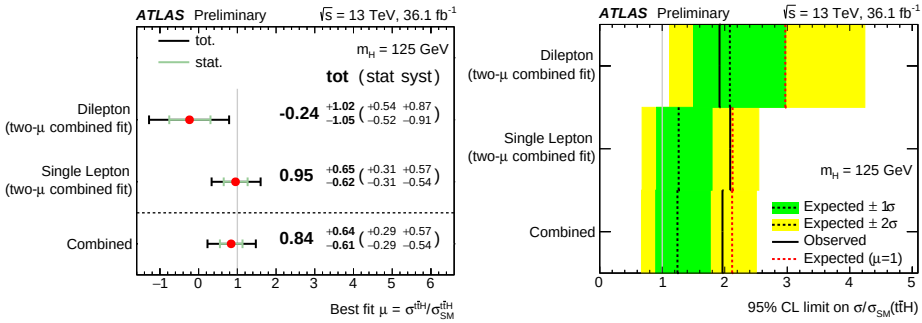
The stacked plot in Fig. 5 (a) incorporates predictions for the expected background yields which are taken from MC simulation. The choice of the nominal MC simulation in the analysis for the background predictions is Powheg + Pythia8. It is a $t\bar{t}$ MC simulation in the 5-flavour scheme meaning that b quarks are treated as massless particles in the matrix element. The nominal MC prediction was changed from the Powheg + Pythia6 prediction which was used in 2016 analysis. This change in the nominal MC simulation provided an improvement in the data-MC modelling.

In order to extract the signal strength as defined in Eq. (1), a profile likelihood fit is performed where $\mu_{t\bar{t}H}$ is the parameter of interest. Various sources for detector-related systematic uncertainties are considered. A set of systematic uncertainties which target theory, incorporating different predictions for MC matrix-element, parton-shower or differences in shape for the $t\bar{t}+ \geq 1b$-sample from 4-flavour or 5-flavour predictions, are taken into account. Systematic uncertainties are a nuisance to the measurement, hampering a better extraction of the parameter of interest. This fact is realized by adding the so-called nuisance parameters to the fit which correspond to the systematic uncertainties and model their impact on the measurement.

3. Results

In this section, the results of the 2017 analysis are presented. In order to extract the signal strength, the profile likelihood fit is performed in 12 regions simultaneously which also includes a boosted category which is new in the 2017 analysis[1]. In Fig. 5 (b), the post-fit distribution of the classification BDT in SR1 is shown. This means the nuisance parameters which were extracted from all regions of the fit are applied here. The pre-fit error bands are the quadratic sum of all systematic uncertainties without taking correlations into account. The correlations between nuisance parameters are taken into account in the post-fit distributions. It can be noted that the systematic uncertainties represented by the error bands are post-fit significantly smaller through the constraints of the CRs.

The extracted signal strengths for the different channels are summarized in Fig. 6 (a). The results for the single-lepton and dilepton channel stem from a fit where 2 uncorrelated signal strengths are fitted in each channel and afterwards the combination is performed in order to reach a combined result. One sees that the result of $\mu_{t\bar{t}H} = 0.95^{+0.65}_{-0.62}$ for single-lepton channel and $\mu_{t\bar{t}H} = 0.84^{+0.64}_{-0.61}$ for the combination with the dilepton channel are compatible with the SM expectation of $\mu_{t\bar{t}H} = 1$.



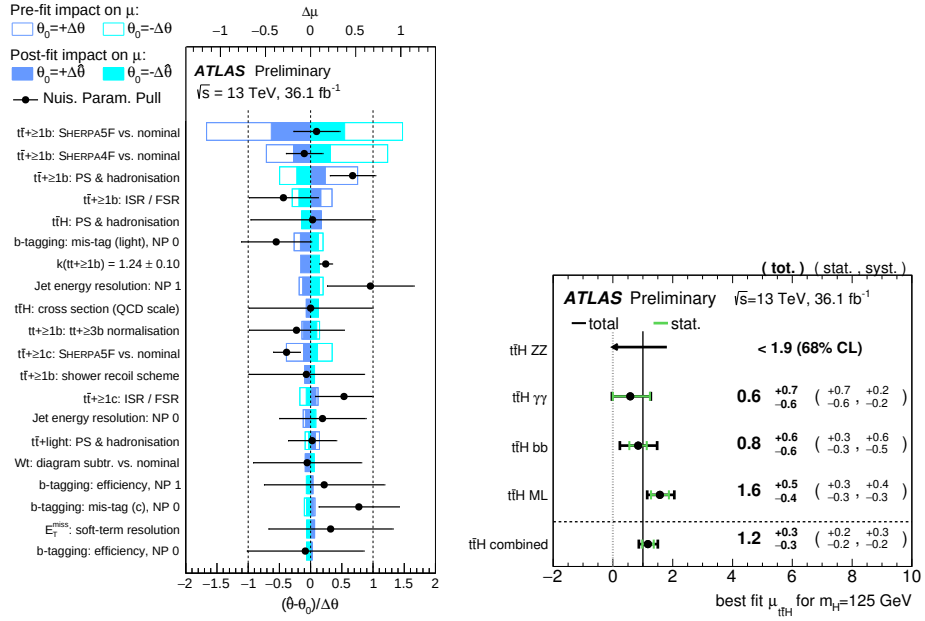
(a) (b)

Fig. 6. (a) Result from single-lepton and dilepton channel. (b) Limit on $\mu_{t\bar{t}H}$ from the fit from single lepton and dilepton [1].

The result can be translated into a limit on $\mu_{t\bar{t}H}$ which can be excluded to be higher than 2 at 95% confidence level as can be seen from Fig. 6 (b). From the result in Fig. 6 (a), one can see that the analysis is already dominated by systematic uncertainties. To give more details about the source of systematic uncertainties and their impact on the signal strength, Fig. 7 (a) shows what the dominant systematic uncertainties are. They are all connected to the

[1] The final result is extracted from a combined fit together with the dilepton categories. In the CRs mainly only one bin is used in the fit.

modelling of the $t\bar{t}+\geq 1b$ sample which also gives us a glimpse where the analysis can be improved. An accurate and precise modelling of the $t\bar{t}+\geq 1b$ backgrounds is indispensable to move forward in the analysis.



(a) (b)

Fig. 7. (a) Ranking plot of systematic uncertainties ordered by their impact on $\mu_{t\bar{t}H}$ for $t\bar{t}H(b\bar{b})$ in the single-lepton channel [1]. (b) Combination of $H \to b\bar{b}$, multilepton, ZZ, and diphoton final states [4].

4. Conclusion and outlook

To put the result in perspective of other channels, the combination with the other analysis of $t\bar{t}H$ production with different decay modes is performed, which is depicted in Fig. 7 (b). The combination of different analysis aims to unify the measurement of $\mu_{t\bar{t}H}$ and can be naively seen as a weighted sum of individual measurements[2]. The result is a measurement of $t\bar{t}H$-signal strength as $\mu_{t\bar{t}H} = 1.2 \pm 0.3$ with a significance of 4.2σ (obs.) and 3.8σ (exp.) based on which ATLAS claims evidence for the process. Compared to the result of the $t\bar{t}H(b\bar{b})$ channel, the error is almost reduced by 50% since the contribution of the $t\bar{t}H(b\bar{b})$ measurement to the combination is not dominant due to the large error. A better comparison is found if the

[2] A dedicated fit is performed for the combination incorporating again all nuisance parameters.

error from the combination is compared to the multilepton contribution ($t\bar{t}H$ ML) in Fig. 7 (b) which then only leads to moderate improvement in the combination. This is due to the fact that in the combination nuisance, parameters can be constraint better and their impact can be reduced.

In the $t\bar{t}H(b\bar{b})$ single-lepton analysis, the way forward is an increased amount of work on the modelling of the backgrounds in order to reduce systematic uncertainties of the measurement. Self-consistent theoretical uncertainties of the MC generators are highly desirable to avoid the comparison to other MC predictions, which currently are the dominant source of uncertainty.

Dedicated measurements of the main backgrounds such as $t\bar{t}b\bar{b}$ might provide a useful insight towards the modelling of the main background. The knowledge gained in these measurements might be ported in the derivation of dedicated 4-flavour $t\bar{t}b\bar{b}$ samples, potentially improving modelling. Care should be taken since it is *a priori* not clear what the effect of an extrapolation from a $t\bar{t}b\bar{b}$ analysis phase space into the $t\bar{t}H$ phase space would be.

REFERENCES

[1] M. Aaboud *et al.* [ATLAS Collaboration], *Phys. Rev. D* **97**, 072016 (2018) [arXiv:1712.08895 [hep-ex]].

[2] Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Technical Report ATLAS-CONF-2016-080, CERN, Geneva, August 2016.

[3] Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run, Technical Report ATL-PHYS-PUB-2016-012, CERN, Geneva, June 2016.

[4] M. Aaboud *et al.*, *Phys. Rev. D* **97**, 072003 (2018) [arXiv:1712.08891 [hep-ex]].