

Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ -pair in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector at the LHC

Leonid SERKIN* on behalf of the ATLAS Collaboration

*INFN Gruppo Collegato di Udine, Sezione di Trieste, Udine and ICTP, Trieste
Strada Costiera 11, Trieste 34151, Italy*

E-mail: Leonid.Serkin@cern.ch

A search for the Standard Model Higgs boson produced in association with a pair of top quarks ($t\bar{t}H$) and decaying into a pair of bottom quarks ($H \rightarrow b\bar{b}$) is presented. The search is focused on the semileptonic decay of the $t\bar{t}$ system and exploits different topologies given by the jet and b-tagged jet multiplicities of the event. A neural network is used to discriminate between signal and background events, the latter being dominated by $t\bar{t}$ + jets production. Using 20.3/fb of data at $\sqrt{s} = 8$ TeV collected with the ATLAS detector during Run 1 of the Large Hadron Collider, we obtain an observed (expected) 95% confidence-level upper limit of 3.6 (2.6) times the Standard Model cross section for a Higgs boson with a mass of 125 GeV.

*The European Physical Society Conference on High Energy Physics
22–29 July 2015
Vienna, Austria*

*Speaker.



1. Introduction

Three years after the discovery of a new particle in the search for the Standard Model (SM) Higgs boson at the LHC reported by the ATLAS and CMS collaborations [1], there is a clear signal of the observed particle in the $H \rightarrow \gamma\gamma$, $H \rightarrow ZZ^{(*)}$, $H \rightarrow WW^{(*)}$ and $H \rightarrow \tau^+\tau^-$ decay channels at a mass of around 125 GeV, while no significant excess has been found yet in searches targeting the dominant decay mode $H \rightarrow b\bar{b}$ [2].

The overwhelming multijet background precludes the possibility of a search for Higgs bosons produced via gluon fusion followed by the $H \rightarrow b\bar{b}$ decay. However, the search for the SM Higgs boson in association with a top quark pair ($t\bar{t}H$) with subsequent Higgs decay into bottom quarks ($H \rightarrow b\bar{b}$) reduces this background significantly. Besides, the observation of the $t\bar{t}H$ production mode would allow the direct measurement of the top-Higgs Yukawa coupling, to which other analyses are mostly sensitive through loop effects.

In the following, we review a search [3] for $t\bar{t}H(H \rightarrow b\bar{b})$ production in proton-proton collisions at $\sqrt{s} = 8$ TeV using data corresponding to an integrated luminosity of 20.3 fb^{-1} collected by the ATLAS experiment [4]. This search is simultaneously sensitive to the Yukawa coupling between the top quark and the Higgs boson and the $H \rightarrow b\bar{b}$ branching ratio, and the only assumption made is that the Higgs boson is a narrow scalar particle.

2. Analysis overview

The present review is focused on the single lepton decay mode of the $t\bar{t}$ system, where the W boson from one top quark decays to a charged lepton and its associated neutrino, and the W boson from the other top quark decays to a quark-antiquark pair. The experimental signature of a $t\bar{t}H(H \rightarrow b\bar{b})$ event is determined by one high transverse momentum (p_T) electron or muon, a characteristic that is crucial for triggering, missing transverse momentum (E_T^{miss}) from the undetected neutrino, and six jets: 2 light jets from the W boson decay, 2 b -tagged jets from the top pair decay and 2 b -tagged jets from the Higgs boson decay.

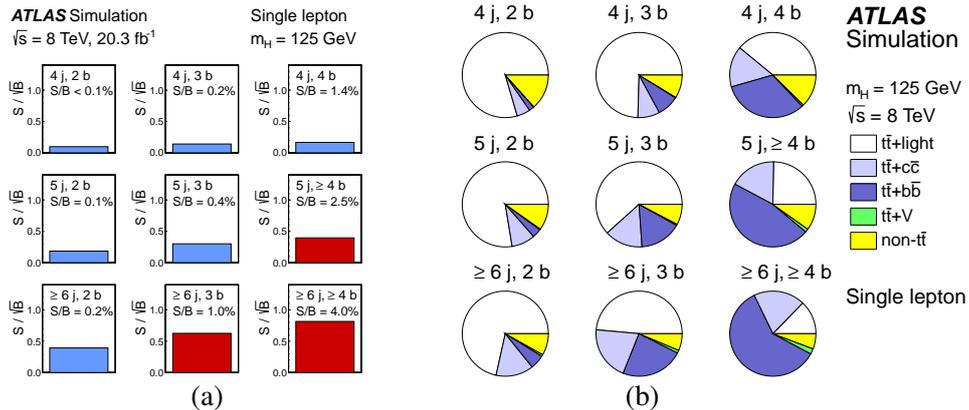


Figure 1: (a) S/\sqrt{B} ratio for all the various regions, *signal-enriched regions* are shaded in dark red, while all the others are shown in light blue. The S/B ratio for each region is also noted, where S and B denote the expected signal for a SM Higgs boson with $m_H = 125$ GeV and background, respectively. (b) Fractional contributions of the various backgrounds to the total background prediction in each considered region [3].

This signature determines the event selection of the analysis: events are required to have exactly one central ($|\eta| < 2.5$) identified electron or muon with $p_T > 25$ GeV and at least four central jets with $p_T > 25$ GeV with at least two of them b -tagged. Selected events are classified into exclusive categories according to the number of reconstructed jets and b -tagged jets. They are analysed separately and combined statistically to maximise the overall sensitivity.

The main background to the search comes from $t\bar{t}$ events which have at least two extra jets produced in association with the top quark pairs. In the following, a given region with m jets of which n are b -tagged jets is referred to as (mj, nb) . The three regions with the largest signal-to-background (S/B) ratio are $(5j, \geq 4b)$, $(\geq 6j, 3b)$ and $(\geq 6j, \geq 4b)$. These regions are the so-called *signal-enriched regions*, as they dominate the sensitivity to the signal: they were blinded at the first stage of the analysis. The remaining six regions considered are referred to as *signal-depleted regions* and consist of $(4j, 2b)$, $(4j, 3b)$, $(4j, 4b)$, $(5j, 2b)$, $(5j, 3b)$ and $(\geq 6j, 2b)$. They are dominated by different backgrounds and are used to constrain systematic uncertainties, thus improving the background prediction in the *signal-enriched regions*. Figure 1(a) shows the S/\sqrt{B} and S/B ratios for the different regions under consideration and the expected proportions of different backgrounds in each region are shown in Fig. 1(b).

3. Signal and background modelling

After the event preselection described above, the main background is due to $t\bar{t}$ + jets production. Other background contributions come from the production of a W -boson in association with jets (W + jets), single top quark, Z + jets and diboson (WW , WZ , ZZ) production, as well as from the associated production of a vector boson and a $t\bar{t}$ pair, with multijet events contributing to a lesser extent. Table 1 provides a summary of some of the relevant parameters of the Monte Carlo (MC) samples used in the analysis.

Sample	Generator	PDF	Shower	Normalisation
$t\bar{t}H$	PowHel	CT10	Pythia 8.1	NLO
$t\bar{t}$ + jets	Powheg	CT10	Pythia 6.425	NNLO+NNLL
W + jets	Alpgen 2.14	CTEQ6L1	Pythia 6.425	NNLO
Z + jets	Alpgen 2.14	CTEQ6L1	Pythia 6.425	NNLO
Single top (s -channel, Wt)	Powheg	CT10	Pythia 6.425	aNNLO
Single top (t -channel)	Powheg	CT10	Pythia 6.425	aNNLO
$t\bar{t}$ + W/Z	Madgraph 5	CTEQ6L1	Pythia 6.425	NLO
Diboson	Alpgen 2.14	CTEQ6L1	Herwig 6.520	NLO

Table 1: A summary of some of the relevant generator parameters used to simulate various processes.

To improve the agreement between data and simulation in the top pair production dominated regions, the $t\bar{t}$ events simulated by Powheg + Pythia are corrected to reproduce the top quark p_T and $t\bar{t}$ system p_T distributions as measured in data [5]. To gain more confidence in describing jet multiplicity and heavy flavour content of $t\bar{t}$ + jets production, additional Madgraph samples were produced to provide the best possible description of the data. Moreover, $t\bar{t}$ + $b\bar{b}$ modelling has

been compared and corrected to a sample generated with SherpaOL NLO [6], which is expected to model the $t\bar{t} + b\bar{b}$ contribution more accurately.

4. Multivariate approach

Multivariate discriminants are employed in the regions with a significant expected contribution from $t\bar{t}H$ signal, to separate it from the backgrounds. Meanwhile, a simple kinematic variable (H_T^{had}) is used in the *signal-depleted regions*: its primary function is to constrain the uncertainties on the background prediction.

In the *signal-enriched regions* a 3-layer feed-forward Neural Network (NN) built using the NeuroBayes package [7] is constructed, trained to separate the $t\bar{t}H$ signal from the main $t\bar{t} + \text{jets}$ background. The choice of the variables that enter the NN discriminant is made through the ranking procedure implemented in the NeuroBayes package which considers the statistical separation power and the correlation of variables. Several classes of variables are used: object kinematics, global event variables, event shape variables and object pair properties. In addition to the kinematic variables, two variables calculated using the Matrix Element method are included in the NN training in the ($\geq 6j, 3b$) and ($\geq 6j, \geq 4b$) regions

Figure 2 shows the distribution of the NN discriminant for the $t\bar{t}H$ signal and background in the *signal-rich regions*. In particular, Fig. 2(a) shows the separation between the $t\bar{t} + \text{heavy flavour}$ and $t\bar{t} + \text{light-jet}$ production achieved by a dedicated NN in the ($5j, 3b$) region.

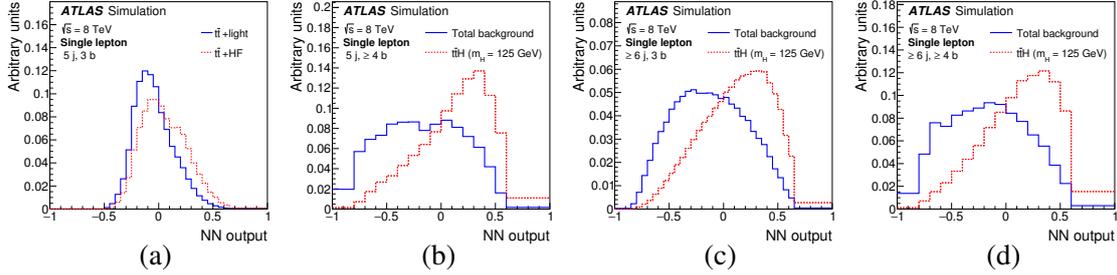


Figure 2: Neural network outputs for the ($5j, 3b$) region (a), ($5j, \geq 4b$) region (b), ($\geq 6j, 3b$) region (c) and ($\geq 6j, \geq 4b$) region (d). The distributions are normalised to unit area. Taken from [3].

5. Systematic uncertainties

Several sources of systematic uncertainties have been considered that can affect the normalisation of signal and background and/or the shape of their corresponding discriminant distributions. The dominant systematics arises from the uncertainty in the normalisation of the irreducible $t\bar{t} + b\bar{b}$ background, followed by the $t\bar{t} + \text{heavy- and light-flavour}$ modelling, b -, c - and light-tagging efficiencies, $t\bar{t} + W/Z$ normalisation and jet energy scale.

Table 2 summarises the pre-fit and post-fit contributions of the different categories of uncertainties (expressed in %) for the $t\bar{t}H$ signal and main background processes in the ($\geq 6j, \geq 4b$) region. The ‘‘Lepton efficiency’’ category includes systematic uncertainties on electrons and muons. The ‘‘Jet efficiency’’ category includes uncertainties on the jet vertex fraction and jet reconstruction. The ‘‘ $t\bar{t}$ heavy-flavour modelling’’ category includes uncertainties on the $t\bar{t} + b\bar{b}$ NLO shape

and on the $t\bar{t}+c\bar{c}$ p_T reweighting and generator. The ‘‘Theoretical cross sections’’ category includes uncertainties on the single top, diboson, V +jets and $t\bar{t} + W/Z$ theoretical cross sections. The ‘‘ $t\bar{t}H$ modelling’’ category includes contributions from $t\bar{t}H$ scale, generator, hadronisation model and PDF choice.

	Pre-fit				Post-fit			
	$t\bar{t}H$ (125)	$t\bar{t} + \text{light}$	$t\bar{t} + c\bar{c}$	$t\bar{t} + b\bar{b}$	$t\bar{t}H$ (125)	$t\bar{t} + \text{light}$	$t\bar{t} + c\bar{c}$	$t\bar{t} + b\bar{b}$
Luminosity	± 2.8	± 2.8	± 2.8	± 2.8	± 2.6	± 2.6	± 2.6	± 2.6
Lepton efficiencies	± 1.4	± 1.4	± 1.4	± 1.5	± 1.3	± 1.3	± 1.3	± 1.3
Jet energy scale	± 6.4	± 13	± 11	± 9.2	± 2.3	± 5.3	± 4.7	± 3.6
Jet efficiencies	± 1.7	± 5.2	± 2.7	± 2.5	± 0.7	± 2.3	± 1.2	± 1.1
Jet energy resolution	± 0.1	± 4.4	± 2.5	± 1.6	± 0.1	± 2.3	± 1.3	± 0.8
b -tagging efficiency	± 9.2	± 5.6	± 5.1	± 9.3	± 5.0	± 3.1	± 2.9	± 5.0
c -tagging efficiency	± 1.7	± 6.0	± 12	± 2.4	± 1.4	± 5.1	± 10	± 2.1
l -tagging efficiency	± 1.0	± 19	± 5.2	± 2.1	± 0.6	± 11	± 3.0	± 1.1
High p_T tagging efficiency	± 0.6	–	± 0.7	± 0.6	± 0.3	–	± 0.4	± 0.3
$t\bar{t}$: p_T reweighting	–	± 12	± 13	–	–	± 5.1	± 5.8	–
$t\bar{t}$: parton shower	–	± 13	± 16	± 11	–	± 3.6	± 10	± 6.0
$t\bar{t}$ +HF: normalisation	–	–	± 50	± 50	–	–	± 28	± 14
$t\bar{t}$ +HF: modelling	–	–	± 11	± 8.3	–	–	± 8.1	± 7.1
Theoretical cross sections	–	± 6.3	± 6.3	± 6.3	–	± 4.1	± 4.1	± 4.1
$t\bar{t}H$ modelling	± 2.7	–	–	–	± 2.6	–	–	–
Total	± 12	± 32	± 59	± 54	± 6.9	± 9.2	± 23	± 12

Table 2: Normalisation uncertainties (expressed in %) on signal and main background processes for the systematic uncertainties considered, before and after the fit to data in the ($\geq 6j, \geq 4b$) region.

6. Results and conclusions

The various discriminant distributions from each of the channels and regions considered are combined to test for the presence of a signal, assuming a Higgs boson mass of $m_H = 125$ GeV. To obtain the final result, a combined fit to *signal-enriched* and *signal-depleted regions* is performed to search for the signal while simultaneously obtaining a background prediction with substantially reduced uncertainties, resulting in an improved search sensitivity compared to the one obtained by fitting the signal region alone.

Figure 3 shows a comparison of data and prediction for the neural network outputs after the fit which was performed on data under the signal-plus-background hypothesis. The uncertainties decrease significantly in all regions due to constraints provided by data and correlations between different sources of uncertainty introduced by the fit to the data.

No significant excess of events above the background expectation is observed and 95% confidence level (CL) upper limits on $\sigma(t\bar{t}H) \times BR(H \rightarrow b\bar{b})$ are derived. In the single lepton channel we obtain an observed (expected) 95% CL upper limit of 3.6 (2.6) times the Standard Model cross section for a Higgs boson with a mass of 125 GeV.

The fitted values of the signal strength and its uncertainties for the single lepton channel, as well as its combination with the dilepton channel are shown in Fig. 4(a). The observed limits, those

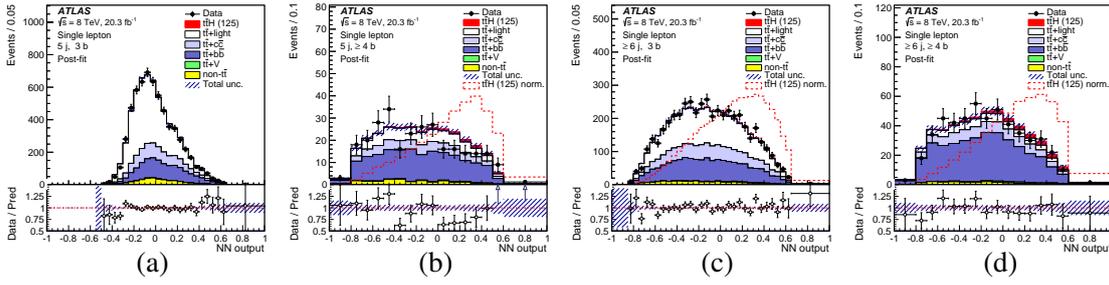


Figure 3: Comparison of data and prediction for the discriminant variable used in the $(5j, 3b)$ (a), $(5j, \geq 4b)$ (b), $(\geq 6j, 3b)$ (c) and $(\geq 6j, \geq 4b)$ (d) regions. The fit is performed on data under the signal-plus-background hypothesis. The bottom panel displays the ratio of data to the total prediction. Taken from [3].

expected with and without assuming a SM Higgs boson with $m_H = 125$ GeV, for each channel and their combination are shown in Fig. 4(b). For the combined result, a signal 3.4 times larger than predicted by the SM is excluded at 95% CL using the CL_s method. A signal 2.2 times larger than for the SM Higgs boson is expected to be excluded in the case of no SM Higgs boson, and 3.1 times larger in the case of a SM Higgs boson.

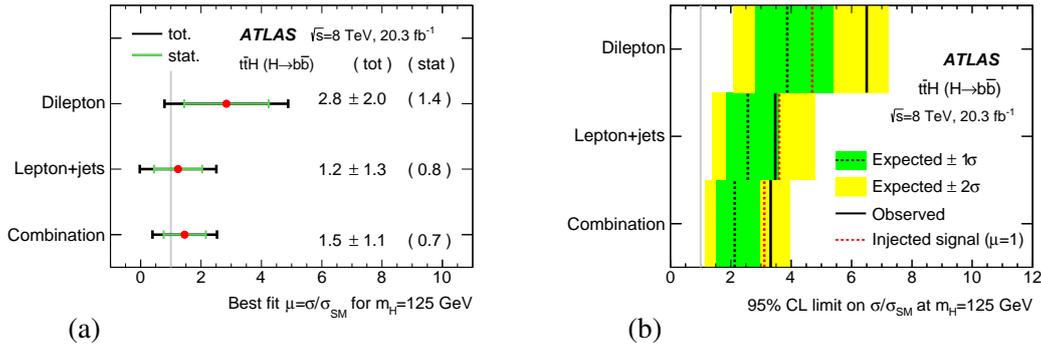


Figure 4: (a) The fitted values of the signal strength and their uncertainties for the individual channels and their combination. (b) 95% CL upper limits on $\sigma(\bar{t}tH)$ relative to the SM prediction, σ/σ_{SM} , for the individual channels as well as their combination. The observed limits (solid lines) are compared to the expected (median) limits under the background-only hypothesis and under the signal-plus-background hypothesis assuming the SM prediction for $\sigma(\bar{t}tH)$ and pre-fit prediction for the background [3].

References

- [1] ATLAS Collaboration, Phys. Lett. **B716** 1, (2012) [hep-ex/1207.7214].
CMS Collaboration, Phys. Lett. **B716** 30, (2012) [hep-ex/1207.7235].
- [2] ATLAS and CMS Collaborations, Phys. Rev. Lett. **114** 191803, (2015) [hep-ex/1503.07589].
- [3] ATLAS Collaboration, Eur. Phys. J. **C75** 349, (2015) [hep-ex/1503.05066].
- [4] ATLAS Collaboration, JINST **3** S08003, (2008).
- [5] ATLAS Collaboration, Phys. Rev. **D90** 072004, (2014) [hep-ex/1407.0371].
- [6] F. Cascioli *et al.*, Phys. Lett. **B734** 210, (2014) [hep-ex/1309.5912].
- [7] M. Feindt and U. Kerzel, Nucl. Instrum. Meth. **A559** 190, (2006).