PROCEEDINGS OF SCIENCE

Observation of the Higgs boson production in association with a $t\bar{t}$ pair with the ATLAS detector

Antonio Policicchio*, **on behalf of the ATLAS Collaboration**
Sapienza Università di Roma and INFN Roma1, Italy
E-mail: antonio.policicchio@roma1.infn.it

The measurement of Higgs boson production in association with a $t\bar{t}$ pair ($t\bar{t}H$) is essential to understand the top-quark couplings to the Higgs boson: it allows for a direct measurement of the top Yukawa coupling and completes the indirect measurement based on the production of the Higgs boson in gluon-gluon fusion and in the decay of the Higgs boson to pair of photons. These proceedings present the results of the recent $t\bar{t}H$ analyses using up to 139 fb^{-1} of pp collision data collected at a center-of-mass energy of $\sqrt{s}$ = 13 TeV by the ATLAS detector at the Large Hadron Collider. The results of the searches in different Higgs boson decay channels are presented. The combination of the results is also presented. It uses up to 79.8 fb^{-1} of data recorded at $\sqrt{s}$ = 13 TeV in combination with searches performed at $\sqrt{s}$ = 7 TeV and 8 TeV, corresponding to an additional 4.9 fb^{-1} and 20.8 fb^{-1} integrated luminosity of pp collision data, respectively. The production of the Higgs boson in association with a top quark pair is observed with a significance of 6.3 standard deviations relative to the background-only hypothesis. The expected significance is 5.1 standard deviations.

European Physical Society Conference on High Energy Physics - EPS-HEP2019 -
10-17 July, 2019
Ghent, Belgium

*Speaker.

PoS(EPS-HEP2019)325

1. Introduction

One of the great successes of runs at $\sqrt{s}$ = 7 TeV and 8 TeV of the Large Hadron Collider (LHC) was the discovery of the Higgs boson by the ATLAS and CMS experiments [1, 2]. Since then many properties of the boson have been measured, and no significant deviations from the predictions of the Standard Model (SM) for the couplings with vector bosons have been found. One area of great interest is the origins of the fermion masses, and how the elementary fermions couple to the Higgs boson. In the SM, these are predicted to be generated by Yukawa mass terms, with coupling constants proportional to fermion masses. Of particular interest is the coupling of the top quark to the Higgs boson. The top quark is predicted to have the largest Yukawa coupling in the SM due to the large top quark mass. Current measurements of the top Yukawa coupling, y_t, are driven by constraints from loop processes, in particular the production of the Higgs boson in gluon-gluon fusion and in the decay of the Higgs boson to pair of photons, and rely on assumptions of particles entering loops. A direct measurement of y_t is possible in the production of the Higgs boson in association with a pair of top quarks ($t\bar{t}H$), where there is a direct coupling between the top quark and the Higgs. Any significant deviations from indirect measurements would be a strong suggestion of the presence of physics beyond the Standard Model (BSM).

This article summarizes the results of the recent analyses that use up to 139 fb^{-1} of pp collision data collected at a center-of-mass energy of $\sqrt{s}$ = 13 TeV by the ATLAS detector [3] at the LHC. The small production cross section of $t\bar{t}H$, 0.507 pb, and the large irreducible backgrounds to the signal make $t\bar{t}H$ a challenging process to measure. The measurement is performed in four separate analyses, each targeting different Higgs boson decay modes. Finally, the combination of the results is also presented.

2. Analysis strategy and results

The decay modes of the Higgs boson are used to categorize events into four different analyses. The decays of the top quark pair is then used to further categorize events within each analysis. The four analyses are: the search for $t\bar{t}H$ with resonant $H \to \gamma\gamma$ which uses the full 139 fb^{-1} dataset collected between 2015 and 2018; the searches for $H \to b\bar{b}$ and $H \to$ multilepton, which targets leptonic final states of Higgs boson decays to WW^*, ZZ^* and $\tau^+\tau^-$, each using 36.1 fb^{-1} of pp data; and the search for the resonant $H \to ZZ^* \to 4l$ which uses 79.8 fb^{-1} of data. The $H \to b\bar{b}$ analysis benefits from the highest branching ratio, corresponding to 58% of all Higgs boson decays, and targets events where the top quark pairs decay leptonically. However, in this channel the signal is dominated by a large irreducible background from $t\bar{t}$ + jets production, especially where the jets originating from heavy flavour quarks (c and b). The signal to background ratios in $H \to b\bar{b}$ range from 1.8% to 5.5% compared to values greater than 100% in the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ searches. In the case of the multilepton channel, several regions exhibit high expected statistical yields but are similar to $H \to b\bar{b}$ in purity, whereas other regions have higher purity, up to 50%.

2.1 $H \to \gamma\gamma$

The $H \to \gamma\gamma$ analysis [4] benefits from a clean signature with a falling continuous background, though it has a low rate. The Higgs boson can be reconstructed as a narrow peak in the di-photon invariant mass $m_{\gamma\gamma}$, with side bands on the falling background used to estimate the background contribution. The main backgrounds come from non resonant di-photon production and other Higgs boson production modes, with the Higgs boson decaying into two photons. Events with two isolated photon candidates are selected. The di-photon $m_{\gamma\gamma}$ invariant mass is chosen to be in the range $105-160$ GeV and at least one b-tagged jet is required. Two signal regions (SRs) are defined: a hadronic SR with at least two jets and zero isolated leptons, and a leptonic SR with at least one isolated lepton. A Boosted Decision Tree (BDT) is trained in each region using the four-momenta of photons and jets in the event. In order to remove dependence on the invariant mass $m_{\gamma\gamma}$, the transverse momentum of the photons is divided by $m_{\gamma\gamma}$. A BDT is trained using a fraction of the available $t\bar{t}H$ Monte Carlo events as a signal sample. Data events in which one or both of the photon candidates fail the isolation and/or identification requirement are used as a background training sample. The BDT output is shown in Figure 1. The events are then categorized depending on the value of the BDT response in four (three) categories for the hadronic (leptonic) channel. This is done to optimize the sensitivity to the $t\bar{t}H$ signal. Figure 2 (left) shows the weighted global fit of the di-photon mass. The main systematic uncertainties are signal modelling, photon isolation and energy scale and resolution. The event yields are presented in Figure 2 (right), where a signal strength (μ) of 1.4 is assumed. With an observed (expected) significance of 4.9 (4.2) standard deviations (σ) this analysis provides evidence for $t\bar{t}H$ production.

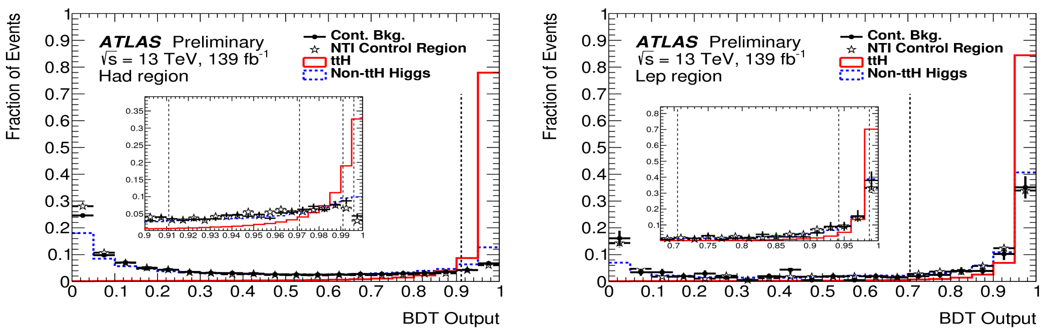

Figure 1: BDT output for the hadronic and leptonic signal regions for the $H\to\gamma\gamma$ decay channel. Events to the left of the vertical dashed line are rejected. The distributions are normalized to unity [4].

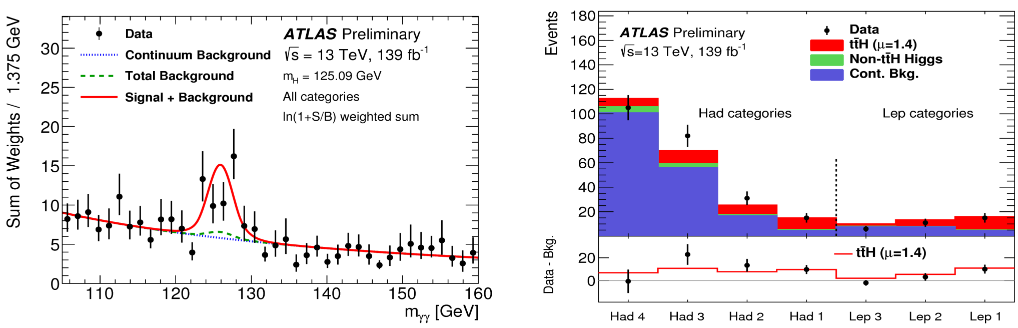

Figure 2: Weighted di-photon invariant mass spectrum (left) and number of events in the different regions for the $H\to\gamma\gamma$ decay channel (right) [4].

2.2 $H \to$ multilepton

The multilepton channel [5] is designed to be sensitive to many leptonic Higgs boson decay modes. The top-quark pair can decay into a single or di-lepton final state. In order to reduce background contributions from $t\bar{t}$ production, events with at least three leptons or with same-sign leptons are chosen. Seven orthogonal channels depending on the number of light leptons and hadronically decaying τ's are considered and are split into eight SRs and four Control Regions (CRs) as shown in Figure 3. Background contributions with prompt leptons originate mainly from top production in association with a vector boson, $t\bar{t}W$, $t\bar{t}Z/\gamma^*$, and di-boson production, VV. Data-driven methods are used to estimate non-prompt light leptons and hadronic τ fakes. To separate signal from background and to suppress backgrounds, different BDTs are used. One dedicated BDT is used to reduce the misidentification of the electron charge, another BDT reduces the non-prompt electrons or muons. Background contributions with prompt leptons are estimated from MC simulations. A maximum likelihood fit of the 12 categories is performed simultaneously to extract the $t\bar{t}H$ signal cross-section normalized to the prediction from the SM, as shown in Figure 4 (left). For five SRs a BDT shape is used as the final discriminant, whereas the total yield is used in regions with low statistics. The main systematic uncertainties are from signal modelling, jet energy scale and resolution and the uncertainty related to the non-prompt light lepton estimation. As shown in Figure 4 (right), a combined signal strength of 1.6 is found. This corresponds to an observed (expected) significance of 4.1 σ (2.8 σ) and gives evidence for $t\bar{t}H$ production.

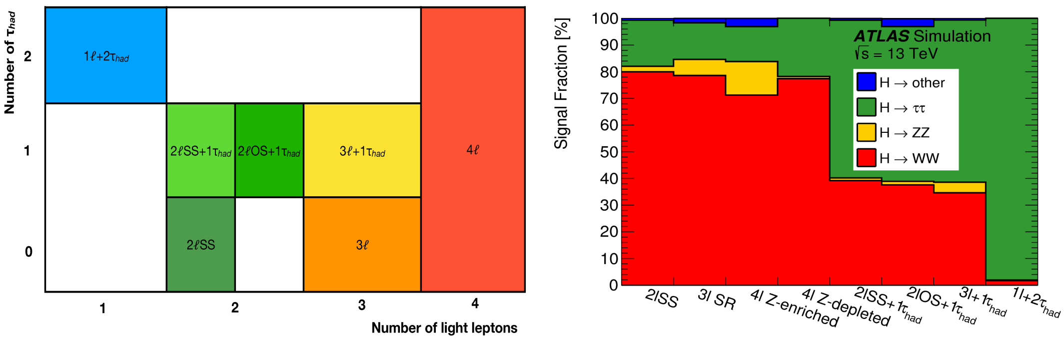

Figure 3: The channels of the $H \to$ multilepton analysis with the corresponding Higgs boson decay modes [5].

2.3 $H \to b\bar{b}$

The analysis [6] targets events in which the Higgs boson decays to pairs of b quarks and one or both top quarks decay semi-leptonically, producing an electron or a muon. Events are categorized first by number of charged light leptons. Events with exactly one lepton fall in the single lepton channel and those with two opposite sign leptons enter the di-lepton channel. The events are then further categorized in nine SRs and ten CRs either enriched in $t\bar{t}H$ or $t\bar{t}\ +\ \geq 1b$, $t\bar{t}\ +\ \geq 1c$, $t\bar{t}$ + light quarks. These regions are defined such that they are dominated by different flavour components of the $t\bar{t}$ + jets background and are shown in Figure 5. A two-steps multivariate strategy is employed in the signal regions to enhance the sensitivity of the analysis. First, reconstruction techniques are employed to solve object combinatorics and reconstruct the final state of the event.

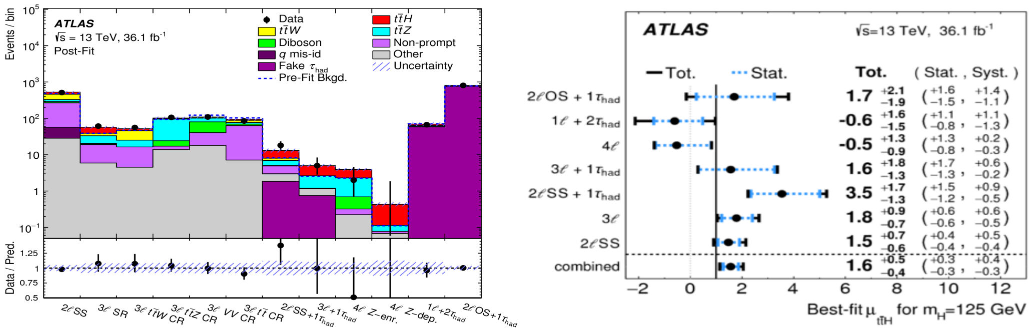

Figure 4: Observed number of events in data compared to the background and signal yields after the fit in the twelve fit regions (left) and observed best-fit values of the $t\bar{t}H$ signal strength μ and their uncertainties by final-state category and combined for the $H \to$ multilepton analysis (right) [5].

Three different techniques are used, a likelihood discriminant, the matrix element method, and a BDT constructed to assign jets to the final state partons in $t\bar{t}H$ events. In the second step, these reconstruction variables are combined with additional kinematic observables in BDTs trained to classify events as signal or background. The BDTs are optimised separately for each signal region, and are used as the final discriminant in the statistical analysis. A binned profile likelihood fit is performed simultaneously over all regions, with floating normalization factors for $t\bar{t}$ + heavy flavor jets. The uncertainty with the largest contribution is the $t\bar{t}\ +\ \geq 1b$ modelling ($\sim$50%). Overall, the analysis is dominated by systematic uncertainties but the statistical uncertainties of the Monte Carlo background modelling are also large. A combined signal strength of $0.84^{+0.64}_{-0.61}$ is found, corresponding to an observed (expected) significance of 1.4 σ (1.6 σ).

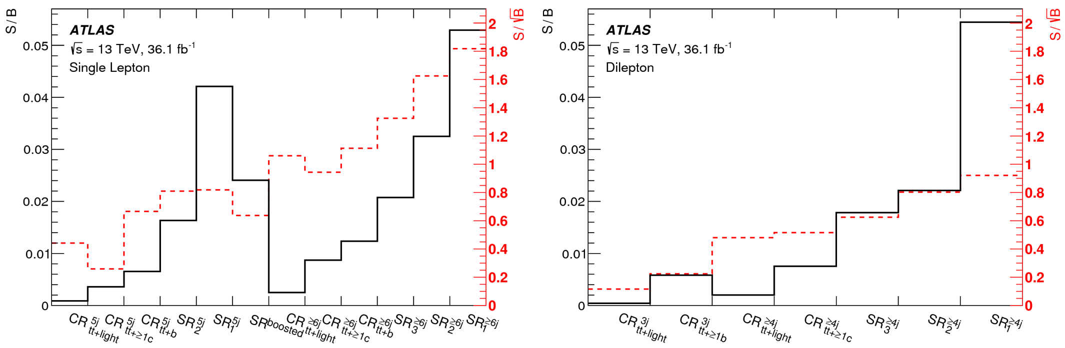

Figure 5: The ratio S/B (black solid line, referring to the vertical axis on the left) and $S/\sqrt{B}$ (red dashed line, referring to the vertical axis on the right) in the single-lepton channel (left) and in the di-lepton channel (right) of the $H \to b\bar{b}$ analysis [6]. S (B) is the number of selected signal (background) events.

2.4 $H \to ZZ^* \to 4l$

In the resonant $H \to ZZ^* \to 4l$ analysis [7], the Higgs boson is searched for by looking for same flavor opposite sign pairs of four electrons, four muons, or two electrons and two muons. A Higgs candidate is considered within the four lepton invariant mass range 115 – 130 GeV, a region that is excluded from the $H \to$ multilepton analysis. Two SRs enriched in $t\bar{t}H$ are selected by requiring at least one b-tagged jet: a hadronic SR requires in addition at least four jets and a leptonic SR requires in addition at least two jets and at least one lepton. In the hadronic region a BDT is employed to separate signal from background. No event is observed and as an upper limit on signal strength $\mu < 1.8$ can be set at 68% C.L.. The expected significance is 1.2 σ.

3. Combination of the results

All four analyses are combined using the profile likelihood method [7]. There is a negligible overlap of events across all four analyses, which are designed to be orthogonal to each other. In addition, the correlation scheme of common sources of systematic uncertainties have been studied in detail. The $t\bar{t}H$ with resonant $H \to \gamma\gamma$ used in the combination is based on the 79.8 fb^{-1} dataset collected between 2015 and 2017. The combination includes also searches performed at $\sqrt{s} = 7$ TeV and 8 TeV, corresponding to an additional 4.9 fb^{-1} and 20.8 fb^{-1} integrated luminosity respectively. The extracted signal strengths in each analysis, as well as the value obtained in the combined fit are shown in Figure 7 (left). The observed signal strength for $t\bar{t}H$ has a significance of 5.8 σ, compared to 4.9 σ expected, and corresponds to a cross section of 670 $\pm$ 90 (stat) $^{+110}_{-100}$ (syst) fb, to be compared with the SM predicted value of 507 $^{+35}_{-50}$ fb. The sensitivity is driven by the multilepton and $H \to \gamma\gamma$ analyses, with $H \to b\bar{b}$ systematically dominated due to the modelling of $t\bar{t}$ + jets and with $H \to ZZ^* \to 4l$ statistically limited, with zero events observed. The two cross sections measured at $\sqrt{s} = 8$ and 13 TeV in the combined analysis are shown in comparison to the SM prediction in Figure 7 (right).

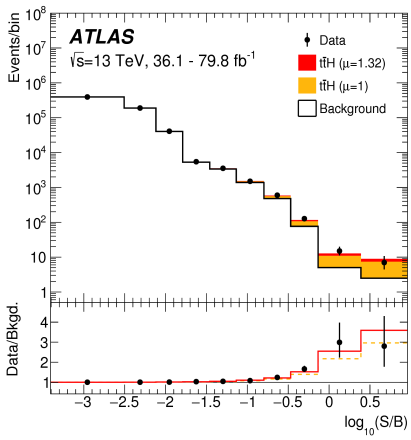

Figure 6: Observed event yields in all analysis categories in up to 79.8 fb^{-1} of 13 TeV data [7]. The background yields correspond to the observed fit results, and the signal yields are shown for both the observed results (μ = 1.32) and the SM prediction (μ = 1). The discriminant bins in all categories are ranked by $\log_{10}(S/B)$, where S is the signal yield and B the background yield extracted from the fit with freely floating signal, and combined such that $\log_{10}(S+B)$ decreases approximately linearly. For the $H \to \gamma\gamma$ analysis, only events in the smallest $m_{\gamma\gamma}$ window containing 90% of the expected signal are considered. The lower panel shows the ratio of the data to the background estimated from the fit with freely floating signal, compared to the expected distribution including the signal assuming μ = 1.32 (full red) and μ = 1 (dashed yellow). The error bars on the data are statistical.

PoS(EPS-HEP2019)325

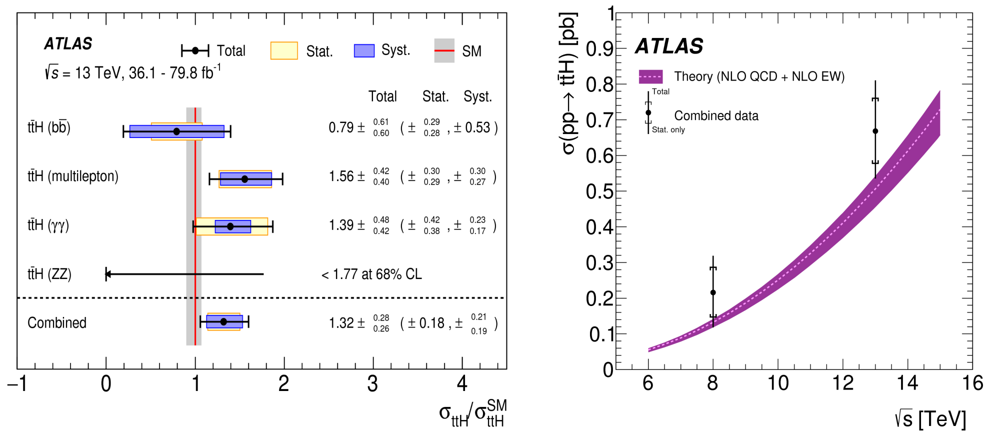

Figure 7: Observed signal strengths of the individual $t\bar{t}H$ channels and in the combination (left) and the observed $t\bar{t}H$ production cross section at $\sqrt{s}$ = 8 and 13 TeV compared to the SM prediction (right) [7].

4. Conclusions

The measurement of $t\bar{t}H$ production is a combination of several challenging analyses. Performing the analyses on $\sqrt{s}$ = 13 TeV pp data recorded by the ATLAS experiment at the LHC, the $t\bar{t}H$ production mode is observed with a significance of 5.8 σ. Additionally, when combined with data collected at $\sqrt{s}$ = 7 and 8 TeV, the observed significance increases to 6.3 σ (5.1 σ expected). This constitutes a direct observation of the Yukawa coupling between the Higgs boson and the top quark.

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 716, 1 (2012)

[2] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B 716, 30 (2012)

[3] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 S08003 (2008)

[4] ATLAS collaboration, Measurement of Higgs boson production in association with a $t\bar{t}$ pair in the diphoton decay channel using 139 fb^{-1} of LHC data collected at $\sqrt{s}$ = 13 TeV by the ATLAS experiment, ATLAS-CONF-2019-004, CDS record: https://cds.cern.ch/record/2668103

[5] ATLAS collaboration, Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector, Phys. Rev. D 97 (2018) 072003

[6] ATLAS collaboration, 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 $\sqrt{s}$ = 13 TeV with the ATLAS detector, Phys. Rev. D 97 (2018) 072016

[7] ATLAS collaboration, Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector, Phys. Lett. B 784 (2018) 173