PROCEEDINGS OF SCIENCE

Higgs boson production in association with a $t\bar{t}$ pair at the ATLAS experiment

Ana Elena Dumitriu *†
Author IFIN-HH Bucharest (RO), CPPM (Aix-Marseille UniversitÃl', CNRS/IN2P3, FR)
E-mail: ana.elena.dumitriu@cern.ch

The measurement of Higgs boson production in association with a $t\bar{t}$ pair is essential to understand the top-quark couplings to the Higgs boson. This talks presents the analyses using Higgs boson decays to $b\bar{b}$ pairs, to two Z bosons, to other multi-lepton final states, and to a pair of photons, using pp collision data collected at 13 TeV.

XXVII International Workshop on Deep-Inelastic Scattering and Related Subjects - DIS2019
8-12 April, 2019
Torino, Italy

*Speaker.
†On behalf of the ATLAS Collaboration.

PoS(DIS2019)086

1. Introduction

The coupling of the Higgs boson to top quark, the heaviest quark within the Standard Model (SM) framework, is the strongest of all Yukawa couplings. Indirect access on the the Yukawa coupling of the Higgs boson to the top quark is provided by the top quark contribution to gluon-gluon fusion production and diphoton decay loops. In this case, no Beyon Standard Model (BSM) contribution is assumed. However, in the direct measurement, any deviation in the top quark's Yukawa coupling due to couplings to new particles will lead to valid signs of new physics. A direct way of accessing this coupling is in the associated production of the Higgs boson with a top quark pair.

In the following sections a combined measurements of Higgs production in association with a $t\bar{t}$ pair cross sections [1] is presented, based on the proton-proton collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector, during Run II data-taking and accounts for 36-138 fb^{-1}. The results combine the analyses of the Higgs boson decay modes $H \to \gamma\gamma$ [1] (Section 2), $H \to ZZ* \to 4l$ [1] (Section 3), $H \to bb$ [2](Section 4) and $H \to multileptons$ [3](Section 5).

2. $ttH \to \gamma\gamma$

This channel presents the results of the ATLAS search for $t\bar{t}H$ production in the $H \to \gamma\gamma$ decay channel. To enhance the signal purity, events passing the diphoton selection and data quality requirements are further devised into two $t\bar{t}H$-enriched regions:

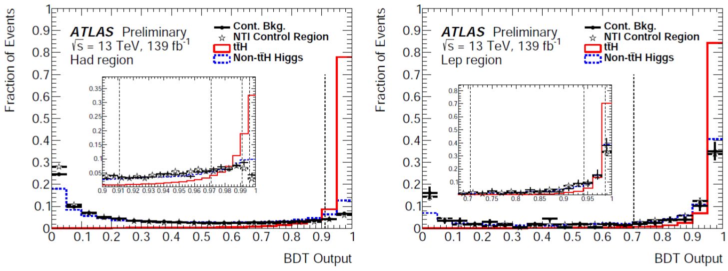
- The "Leptonic (Lep)" region, targeting $t\bar{t}$ decays in which at least one of the W bosons decays to a leptons (muon or electron) with transverse momentum $p_T > 25$ GeV. It requires also the presence of at least one jet with transverse momentum $p_T > 25$ GeV and tagged as containing a b-hadron.

- The "Hadronic (Had)" region targets hadronic top decays and requires no reconstructed leptons. Events which have at least one jet with p_T >25 GeV, tagged as containing a b-hadron, as well as contain at least two additional jets with $p_T > 25$ GeV are selected.

A dedicated boosted decision tree (BDT) to each "Lep" / "Had" is applied, trained with the XGBoost to create regions of high $t\bar{t}H$ signal purity. The BDT is targeted to disentangle the $t\bar{t}H$ signal from the major backgrounds which includes $\gamma\gamma$, tt+$\gamma\gamma$ (taken from data in control regions), other Higgs prod (taken from simulation) and uses as inputs discriminating variables like photon kinematics (p_T, mass of the diphoton system $m_{\gamma\gamma}$, η,Φ), missing transverse energy E_T^{miss}, the energy for up to 4(2) leading jets(lep) in p_T.

Finally, events in the $t\bar{t}H$-enriched pre-selection regions are sorted into one of the four "Had" or three "Lep" categories based on BDT score, labeled by their type ("Had" or "Lep") and by a number, the highest signal purity being labeled as category 1. In Figure 1 is shown the distributions of $t\bar{t}H$ signal MC events and background events as a function of BDT score, for each $t\bar{t}H$-enriched regions as well as the categorization based on the BDT selection.

The measurement of the cross-section is estimated with a simultaneous binned maximum-likelihood fit to the $m_{\gamma\gamma}$ distributions (in the range of 105-160 GeV) of the selected event categories,



(a) "Had" Region BDT Score Distributions (b) "Lep" Region BDT Score Distributions

Figure 1: The normalized fraction of events in bins of BDT score in the (a) "Had" and (b) "Lep" regions. The insets provide a zoomed-in picture of the BDT score distribution for events which are selected for the BDT categories.

using a double-sided crystal ball for the $t\bar{t}H$ signal while for the continuum background a smooth functions (power-law or exponential).

The observed signal strength is: $\mu_{t\bar{t}H} = 1.38^{+0.41}_{-0.36} = 1.38^{+0.33}_{-0.31}(stat)^{+0.13}_{-0.11}(exp)^{+0.22}_{-0.14}(theo)$ corresponding to the measured cross section times branching ratio of: $\sigma_{t\bar{t}H} \times B_{\gamma\gamma} = 1.59^{+0.43}_{-0.39}\, fb = 1.59^{+0.38}_{-0.36}(stat)^{+0.15}_{-0.12}(exp)^{+0.15}_{-0.11}(theo)\, fb$, in agreement with the Standard Model prediction of $t\bar{t}H \rightarrow \gamma\gamma = 1.15^{+0.09}_{-0.12}\, fb$. The combined observed significance for $t\bar{t}H \rightarrow \gamma\gamma$ channel is 4.9σ (with 4.2σ expected).

The analysis is statistically dominated followed by the theory uncertainties for which dominate the underlying event and parton shower, modeling of heavy flavor jets, higher-order QCD terms, while for the experimental uncertainty the dominating ones come from photon energy resolution, photon energy scale, jet and E_T^{miss} modeling.

3. $ttH \rightarrow ZZ* \rightarrow 4l$

In the $ttH \rightarrow ZZ* \rightarrow 4l$ analysis events with at least four isolated leptons (four electrons/muons, or two electrons and two muons) corresponding to two same-flavour opposite-charge (SFOS) pairs are selected, in the invariant mass window of 115 GeV$< m_{4l} <$130 GeV. In addition, at least one jet is required, with $p_T >$30GeV, b-tagged with an efficiency of 70%.

Following 2 procedure, to increase the expected $t\bar{t}H$ significance, events passing the preselection and data quality requirements are further devised into two $t\bar{t}H$-enriched signal regions:

- The "Leptonic (Lep)" region, enriched in semileptonic top-quark decays which requires at least one additional jet and at least one additional isolated lepton.

- The "Hadronic (Had)" region enriched in hadronic top-quark decays. It is formed by requiring at least three additional jets and zero additional isolated leptons. In this region a BDT is applied which includes as discriminating variables the invariant mass, the dijet p_T, and the difference in pseudorapidity of the two leading jets, E_T^{miss}, the angular separation between the four-lepton system and the leading jet, as well as between the dilepton pair with invariant mass closest to the Z boson mass and the leading jet, the scalar sum of the p_T of the jets in the event, the number of jets/b-tagged jets.

The main backgrounds in both regions are $t\bar{t}$W, $t\bar{t}$Z, and non-$t\bar{t}H$ Higgs boson production (ggF and tH for the "Had" and tH for the "Lep" region), estimated from simulation.

The observed events and expected background yields in a four-lepton invariant mass window of 115 GeV< m_{4l} <130 GeV devised into the two "Had" BDT bins and one "Lep" region, are used as input to a likelihood fit. No event is observed, with expected 1.1 events (0.6 in the $t\bar{t}H$ channel). The expected significance is 1.2σ. The dominant uncertainties arise from the QCD, the parton-shower modelling while the leading experimental uncertainty arises from the calibration of the jet energy scale.

4. $ttH \to bb$

The search targets the $H \to bb$ decay mode. The selected events are required to contain either one or two electrons/muons from the top-quark decays, and are then categorized according to the number of jets and how likely these are to contain b-hadrons. In the dilepton channel, three signal regions are defined, with different levels of purity for the $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ components, while in the single-lepton channel, five signal regions are formed from events passing the resolved selection, three with at least six jets requirement, and the other two with exactly five jets.

Multivariate techniques are used to discriminate between signal and background events, the latter being dominated by $t\bar{t}$ + jets production.

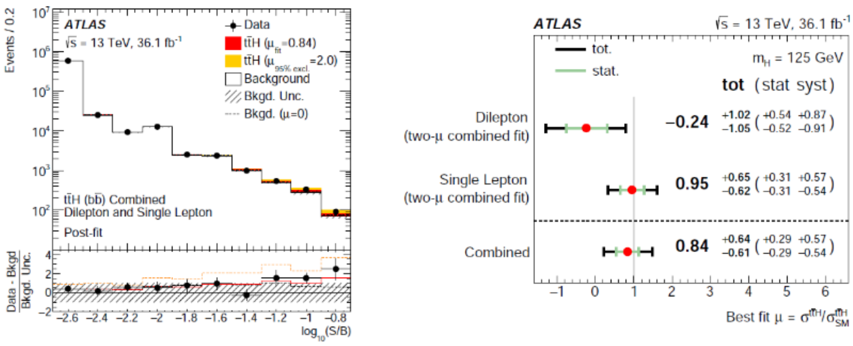


Figure 2: Post-fit yields of signal (S) and total background (B) as a function of log(S/B), compared to data (left); Summary of the signal-strength measurements in the individual channels and for the combination (right).

The best-fit signal-strength value is: $\mu = 0.84 \pm 0.29(stat)^{+0.57}_{-0.54}(sys) = 0.84^{+0.64}_{-0.61}$. The total systematic uncertainty is dominated by the uncertainties in modeling of the $t\bar{t}$ background, as well as from statistics in the simulated samples, followed by the uncertainties in the b-tagging efficiency, the jet energy scale and resolution and signal process modeling.

An excess of events over the expected SM background is found with an observed (expected) significance of 1.4σ(1.6σ). Figure 2 left shows the event yield in data compared to the post-fit prediction for all events passing the analysis selection, grouped and ordered by the signal-to-background ratio. Figure 2 right shows the comparison between the combined μ value and the value from the two independent individual channels, with their corresponding statistical and systematic uncertainties split.

5. $ttH \rightarrow multileptons$

This search focuses the multilepton final states and targets the Higgs boson decays to WW*, $\tau\tau$, and ZZ*. Seven final states (Figure 3), categorized by the number and flavor of charged-lepton candidates are examined .

For this decay channel, the irreducible backgrounds include events with selected light leptons produced in W or Z/γ* boson decays or leptonic τ decays and are estimated from MC and validated in data. The reducible backgrounds have at least one lepton arising from another source and are non-prompt e/μ and fake τ_{had}, estimated using data-driven approach. Specific multivariate techniques are applied in most channels to discriminate the signal from the background events.

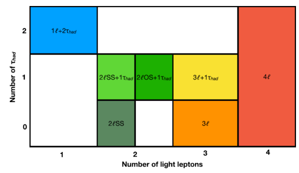


Figure 3: The channels used in the analysis organized according to the number of selected light leptons and τ_{had} candidates.

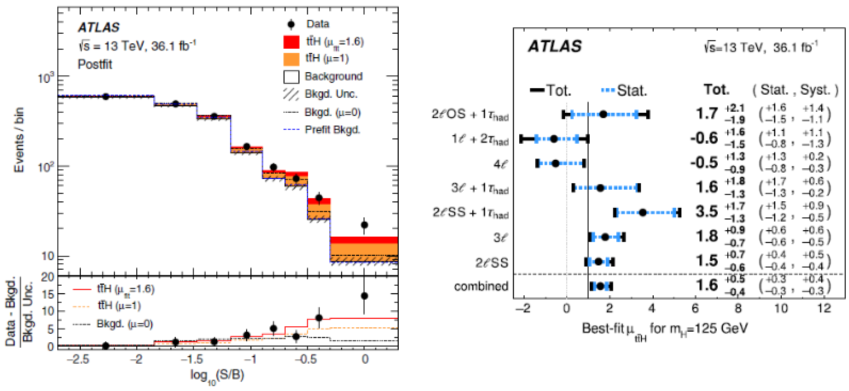


Figure 4: Event yields as a function of $log_{10}(S/B)$ for data, background and a Higgs boson signal with m_H = 125 GeV (left); The observed best-fit values of the $t\bar{t}H$ signal strength μ and their uncertainties by final-state category and combined (right).

An excess of events over the expected Standard Model background is found with an observed (expected) significance of $4.1\sigma(2.8\sigma)$. The observed (expected) best-fit value of μ (Figure 4 right) is: $\mu = 1.6^{+0.3}_{-0.3}(stat)^{+0.4}_{-0.3}(sys)$. The uncertainties with the largest impact are those associated with the signal modeling, the jet energy scale and the non-prompt light-lepton and fake τ_{had} estimates. Figure 4 left shows the data, background (B) and signal (S) yields, where the final-discriminant bins in all signal regions are combined into bins of log(S=B).

When assuming that the observed signal is due to the SM Higgs boson, the excess over the SM signal-plus-background hypothesis has a significance of 1.4σ. The measured cross section is $\sigma_{t\bar{t}H} = 790^{+150}_{-150}(stat)^{+170}_{-150}(sys) = 790^{+230}_{-210} fb$ comparable with to the prediction of $\sigma_{t\bar{t}H} = 507^{+35}_{-50} fb$.

6. Combined Results

The combination of the channels presented in previous sections lead to the first observation of Higgs boson production in association with a top quark with the ATLAS detector. Using 13 TeV data, the likelihood fit to extract the $t\bar{t}H$ signal yield provides an observed (expected) excess relative to the background-only hypothesis of $5.8\sigma(4.9\sigma)$, while a combined fit using the 7, 8, and 13 TeV analyses gives an observed (expected) significance of $6.3\sigma(5.1\sigma)$.

Assuming Standard Model branching fractions, the total $t\bar{t}H$ production cross section at 13 TeV is measured to be $670 \pm 90(stat.)^{+110}_{-100}(syst.)fb$, in agreement with the SM prediction $507^{+35}_{-50}fb$.

Figure 5 left shows the combined event yields in all analysis categories as a function of $log_{10}(S/B)$, where a clear excess of $t\bar{t}H$ signal-like events over the background is visible for high $log_{10}(S/B)$ values. The cross section extracted in the combined likelihood fit, as well as the results from the individual analyses ratios to the SM predictions are displayed in Figure 5 right.

The dominant uncertainties arise from the modelling of the $t\bar{t}$ + heavy-flavour processes in the $H \to b\bar{b}$ and the modelling of the $t\bar{t}H$ process, affected also by the statistical uncertainty. Relevant uncertainties also come from uncertainties in the estimate of leptons from heavy-flavour decays, conversions or misidentified hadronic jets, the jet energy scale and resolution, present in all analyses.

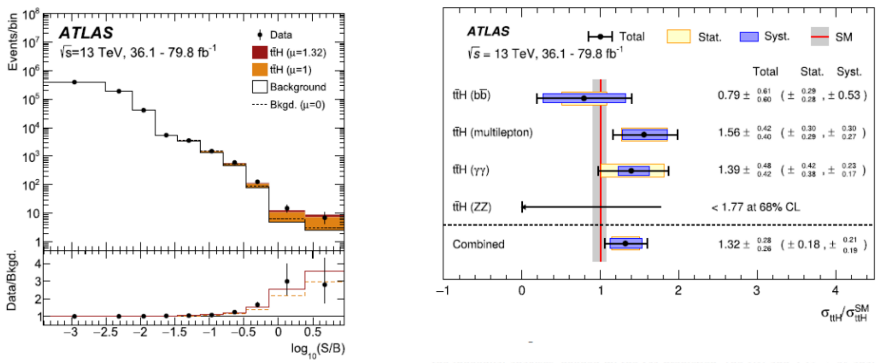


Figure 5: Observed event yields in all analysis categories. The background yields correspond to the observed fit results, and the signal yields are shown for both the observed results (μ =1.32) and the SM prediction (μ =1) (left); Combined $t\bar{t}H$ production cross section, as well as cross sections measured in the individual analyses, divided by the SM prediction (right).

References

[1] "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector", ATLAS collaboration, PHYSICS LETTERS B 784(2018)173-191

[2] "Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions $\sqrt{s}$ = 13 TeV with the ATLAS detector", ATLAS Collaboration, PHYSICAL REVIEW D 97, 072016 (2018)

[3] "Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector", ATLAS Collaboration, PHYSICAL REVIEW D 97, 072003 (2018)