EVIDENCE FOR $t\bar{t}H$ PRODUCTION AT $\sqrt{s} = 13$ TeV?

G. PETRUCCIANI
on behalf of the ATLAS and CMS Collaborations
CERN, European Organization for Nuclear Research, Geneva, Switzerland

Higgs boson production in association with a top quark pair ($t\bar{t}H$) is important for constraining the Yukawa coupling between the Higgs boson and top quark fields. Recent results by the ATLAS and CMS collaborations on the search for $t\bar{t}H$ are briefly presented, discussing the main challenges, strengths and weaknesses of the analyses in the different Higgs boson decay modes.

1 Introduction

In the standard model (SM), the Yukawa coupling between the Higgs boson and the top quark plays an important role in many observables, such as the cross section for Higgs boson production in gluon fusion, the branching ratio for Higgs boson decays to photon pairs, and cross section for associated production of a Higgs boson with a $t\bar{t}$ pair or with a single top quark. The first two observables are more easily accessible, but they are all loop-induced and thus could receive contributions from additional heavy particles not predicted by the SM. A direct test of the Yukawa coupling through a tree-level process is therefore a necessary complement to those two measurements, and $t\bar{t}H$ is the most promising channel since in the SM the cross section for associated production with a single top quark is an order of magnitude smaller than the $t\bar{t}H$ one.

Searches for $t\bar{t}H$ production from LHC run 1 by the ATLAS\textsuperscript{1} and CMS\textsuperscript{2} Collaborations show some excess compared to the SM predictions in multiple final states, leading to a combined\textsuperscript{3} signal strength of $2.3^{+1.2}_{-1.0}$. This result, and the almost-fourfold increase in the production cross section from $\sqrt{s} = 8$ TeV to 13 TeV, make the first results from LHC run 2 especially interesting.

At the LHC, searches for $t\bar{t}H$ can be broadly classified into three groups, depending on the Higgs boson decay mode. At one end lay searches in the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ decay modes, where the Higgs boson can be precisely reconstructed, characterized by a good purity but very low expected event yields. At the opposite end, searches in the $H \rightarrow b\bar{b}$ final state can benefit from the largest decay branching fraction, but have also the largest backgrounds and lowest purity. The third group, with intermediate event yields and purities, targets $H \rightarrow WW^*$ and $H \rightarrow \tau\tau$ decays in final states with multiple leptons.

2 Bosonic decay modes

In the long run, the resonant diphoton and four lepton final states could provide a clean measurement of $t\bar{t}H$ since the events can be selected with high purity, and the decay products from the Higgs can be separated from those of the $t\bar{t}$ system. The main challenge is the small product of cross section and branching fraction for these channels, only about 1 fb and 0.14 fb for
The two measurements have similar sensitivities, and the observed signal strengths are slightly above one event in both categories, and a signal to background ratio of about 1/3. The two measurements have similar sensitivities, and the observed signal strengths are $-0.3^{+1.2}_{-1.0}$ for ATLAS and $1.9^{+1.5}_{-1.2}$ for CMS. Both results are compatible with the SM prediction at the 1$\sigma$ level, and the dominant uncertainty is statistical. Data from the more pure leptonic category is shown in Fig. 1.

Figure 1 – Searches for $t\bar{t}H$ in the $H \rightarrow \gamma\gamma$ final state: leptonic category of the ATLAS (left) and CMS (right) analyses.

So far, $t\bar{t}H$ searches in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay mode at $\sqrt{s} = 13$ TeV have been performed only by CMS$^7$. The search features a single $t\bar{t}H$ category, including both hadronic and leptonic events, and has an even looser requirement of $\geq 4$ jets instead of $\geq 5$ for the hadronic part. In the analyzed data, 35.9 fb$^{-1}$, the expected event yields are approximately 0.5 events for the signal and 0.3 for the sum of all the backgrounds. Having observed zero events, the reported result for the signal strength is $0.0^{+1.2}_{-0.0}$. This uncertainty figure is to be taken with some measure of caution, as for very low event yields the confidence intervals derived in the asymptotic approximation from the negative log-likelihood function are known to undercover.

3 Fermionic decay modes

Searches for $t\bar{t}H$ in the $H \rightarrow b\bar{b}$ final state can benefit from a $\sigma_{t\bar{t}H} \cdot B_{b\bar{b}} \sim 0.3$ pb, more than two orders of magnitude larger than in the diphoton final state. The main challenge however is the background from $t\bar{t}$ production in association with jets, and especially $b$-jets. The $t\bar{t}$ production cross section at $\sqrt{s} = 13$ TeV is about 830 pb, and the irreducible background from associated $t\bar{t}b\bar{b}$ production has cross section of $O(10)$ pb, depending on the kinematic cuts used its definition. The combinatoric from the presence of many $b$ partons in the final state and the jet energy resolution do not allow a good discrimination between signal and $t\bar{t}b\bar{b}$, and so the sensitivity of the search is limited by the large theoretical uncertainties associated with the modelling of this background.
The overall strategy of the ATLAS\textsuperscript{8} and CMS\textsuperscript{9} searches in this final state is similar. Starting from semileptonic and dileptonic t\overline{t} decays, events are categorized according to the jet and b-jet multiplicity. In each categories discriminators are defined to separate signal from background, and combined fit to all categories is used to constrain the backgrounds and allow the extraction of a signal. However, the two analyses differ in several respects in the way this strategy is realized. Both analyses are quite complex, and an exhaustive comparison of the two would go well beyond the scope of this writeup.

The dominant t\overline{t} background is modelled by simulations, and the two experiments use similar definitions for splitting it into different components depending on the presence of extra generated b and c hadrons not from the top quark decays, and whether those heavy flavour hadrons are clustered into separate jets or not. In the case of ATLAS, the simulation of t\overline{t} plus heavy flavour is corrected relying on parton-level computations, while the CMS analysis relies on the parton shower, but assigns larger uncertainties on the different components of the t\overline{t}+ 1b background.

While both analyses use a similar categorization of the events in jet and b-jet multiplicities, the ATLAS analysis is extended to categories with lower multiplicity: in the more sensitive semi-leptonic final state, the CMS analysis uses only events with \( \geq 4 \) b-tags, or \( \geq 6 \) jets and 3 b-tags, while the ATLAS one includes events down to 4 jets and 2 b-tags as control regions to constrain more the backgrounds. In the categories with the highest jet and b-tag multiplicity, signal to background ratios of a few percent are achieved, and the background is mostly from t\overline{t}b\overline{b}.

Different approaches are used to build discriminators to separate the signal from the background in each category. For the higher multiplicity categories, the ATLAS analysis relies on two boosted decision trees (BDTs), the first is used to reconstruct the event, i.e. find the best assignment of reconstructed jets to partons, and its output is used in a second BDT to classify events as more or less signal-like. The CMS analysis instead uses first a BDT to split each category in two with different purities, and then uses the matrix element method for the t\overline{t}H vs t\overline{t}b\overline{b} hypotheses to build the discriminating variable used in the final fit.

Qualitatively, a final fit of all categories exhibits the same behaviour: the initial large uncertainties on the background are reduced to \( O(10\%) \) or less after the fit, and good agreement is observed between data and predictions from simulations (Fig. 2). The two analyses, both based on the summer 2016 data sets, have similar sensitivity. The measured signal strengths are \( (2.1^{+0.5+0.9}_{-0.5-0.7}) \) for ATLAS and \( -0.2^{+0.5+0.7}_{-0.4-0.7} \) for CMS, where the first uncertainty term is statistical and the second is systematic.

4 Multilepton final states

Multilepton searches for t\overline{t}H select final states with two leptons, electrons or muons, with the same electric charge, or at least three leptons, plus jets and b-tags. Signal events are expected
mostly from Higgs boson decays to WW* and $\tau\tau$, with at least one W or $\tau$ decaying leptonically, and semileptonic or dileptonic $t\bar{t}$ decays. The main background with prompt leptons is from $t\bar{t}V$ associated production, and is estimated from theoretical predictions with $O(10\%)$ uncertainties. The other main background in this final state is the reducible one, mostly from $t\bar{t}$ events with non-prompt leptons or charge mis-assignment, estimated from data with larger systematic uncertainties, $O(30\%)$.

In this final state, there is some difference in the approaches used by the two collaborations: the CMS analysis\textsuperscript{11} makes extensive use of multivariate methods both for the lepton selection and for the signal extraction, while the ATLAS analysis\textsuperscript{10} relies on more traditional methods. Also, the ATLAS analysis is performed on the 13.2 fb$^{-1}$ summer 2016 data set, while in the CMS result presented at Moriond 2017 is on 35.9 fb$^{-1}$ from the full 2016 data set. Another difference is in the treatment of final states with hadronically decaying $\tau$ leptons ($\tau_h$): they are included in the ATLAS analysis, while in the CMS case they are vetoed in the multilepton analysis, to be covered in the future in a dedicated analysis.

Events are categorized according to lepton multiplicity, flavour, and the presence of $\tau_h$'s, with various requirements on the jet and b-tag multiplicity. The ATLAS analysis separates events into four main categories: (i) $\ell^+\ell^-0\tau_h$ events with $\geq$ 5 jets, further split in $ee$, $e\mu$, $\mu\mu$; (ii) $\ell^+\ell^-1\tau_h$ events with and $\geq$ 4 jets; (iii) $3\ell$ events with $\geq$ 3 jets; and (iv) $4\ell$ events with $\geq$ 2 jets. In all final states, at least one b-tag is required. The CMS multilepton analysis has the same categorization except for the $\ell^+\ell^-1\tau_h$ state, but is more inclusive in jet multiplicity, allowing events with $\geq$ 4 in the $\ell^+\ell^-0\tau_h$ category and events with $\geq$ 3 in the $3\ell$ category. Also, in the CMS case these two categories are then split further by b-tag multiplicity and by sum of the charges of the leptons, in addition to separating $\ell^+\ell^-$ events by lepton flavour.

In the ATLAS case, each category is modelled as a counting experiment. The same approach is followed for the $4\ell$ final state of the CMS analysis, where the event yields are small. In the more populated categories, the CMS uses two BDTs, trained separately to discriminate the $t\bar{t}H$ signal from the $t\bar{t}V$ background and from the reducible backgrounds; the bidimensional plane defined by the two discriminators is then binned by grouping regions with similar signal to background ratio.

Data from both analyses is in slight excess compared to the SM $t\bar{t}H$ predictions, at the level of about one standard deviation (Fig. 3). The CMS analysis measures a signal strength for a SM $t\bar{t}H$ signal is 2\textsuperscript{\textpm}1.3\textsuperscript{\textpm}0.3 fb$^{-1}$, while the ATLAS one measures $2.5^{+0.7}_{-0.7}\; fb^{-1}$; as before, uncertainty is split in statistical and systematic components, respectively. The observed significance of the signal with respect to the background only hypothesis is 3.3\textsigma for CMS and 2.2\textsigma for ATLAS, while the expected significances for a SM $t\bar{t}H$ signal are 2.5\textsigma for CMS and 1.0\textsigma ATLAS.

![Figure 3](image.png)

Figure 3 – $t\bar{t}H$ search in the multilepton final state: comparison between data and predictions for ATLAS analysis\textsuperscript{10} (left plot), and for the $\ell^+\ell^-$ and $3\ell$ categories of the CMS analysis\textsuperscript{11} (centre, right).

It has to be noted that most ATLAS and CMS $t\bar{t}H$ searches in the multilepton final state have measured signal strengths larger than SM predictions, the only exception being the first
CMS result on 2015 data ($\mu = 0.6^{+1.4}_{-1.1}$). The same is true for all measurements of $t\bar{t}W$ production at the LHC so far, but not for measurements of $t\bar{t}Z$, or in searches for beyond the SM physics or SM measurements in similar final states. Future measurements, and possibly progress on the theoretical predictions (e.g. offshell effects, NNLO QCD), should tell whether this is simply a statistical coincidence or not.

5 Summary and outlook

Results for all $t\bar{t}H$ searches using 13 TeV LHC data are summarized in Fig. 4, including the ATLAS combination of $t\bar{t}H$ results on the summer 2016 data set$^5$. Even in the absence of a proper statistical combination of all results, it is clear that the absence of $t\bar{t}H$ production is rejected at more than 3$\sigma$, i.e. there is statistical evidence for $t\bar{t}H$ production. It is nonetheless important to consider the future prospects for these searches, as there is not yet an analysis with a strong and unambiguous signal.

![Figure 4 – Summary of the measured $t\bar{t}H$ signal strengths in the ATLAS and CMS searches, plotted (left) and tabulated (right). The summary plot includes also the result from the LHC run 1 combination, 2.3$^{+1.2}_{-0.6}$.](image)

The sensitivity of all the $t\bar{t}H$ searches in bosonic final state is currently dominated by statistical uncertainties, and so improvements are expected with larger data sets. This is especially true for the $H \rightarrow \gamma\gamma$ searches, as they have not yet been updated to the full 2016 data set, and the $4\ell$ channel will probably not have much statistical power until the end of Run 2.

The current analyses in the $H \rightarrow b\bar{b}$ final state are limited by systematic uncertainties, so progress is expected to come mostly from incremental improvements in the understanding and modelling of the $t\bar{t}$ background, in collaboration with the theory community, and possibly by improving further the signal extraction techniques, which are however already quite advanced. Improved sensitivity may also come from the investigation into boosted topologies where the event kinematic is easier to reconstruct, possibly allowing for higher purity signal regions less affected by uncertainties on the background model. Experimentally, so far this has been explored only in the first CMS $H \rightarrow b\bar{b}$ at 13 TeV with 2015 data$^{13}$: in this dataset the improvement in sensitivity is modest, but the approach should be more promising for larger data set.

Multilepton searches are characterized by comparable statistical and systematic uncertainties, and thus progress is expected both from larger data samples and from improvements in the analysis techniques, e.g. for the estimation of reducible backgrounds and signal extraction, and in the $t\bar{t}W$ background predictions. Final states with $\tau$s are also quite promising as they allow a better discrimination between signal and $t\bar{t}V$ background, and the current analyses are limited by statistical uncertainties.
References

3. ATLAS and CMS Collaborations, Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ TeV and 8 TeV, JHEP 08, 045 (2016), DOI: 10.1007/JHEP08(2016)045.
4. ATLAS Collaboration, Measurement of fiducial, differential and production cross sections in the $H \rightarrow \gamma\gamma$ decay channel with 13.3 fb$^{-1}$ of 13 TeV proton-proton collision data with the ATLAS detector, ATLAS CONF-2016-067 (Aug 2016), http://cdsweb.cern.ch/record/2206210.
5. ATLAS Collaboration, Combination of the searches for Higgs boson production in association with top quarks in the $\gamma\gamma$, multilepton, and $b\bar{b}$ decay channels at $\sqrt{s} = 13$ TeV with the ATLAS Detector, ATLAS CONF-2016-068 (Aug 2016), http://cdsweb.cern.ch/record/2206211.
9. CMS Collaboration, Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with 2016 pp collision data at $\sqrt{s} = 13$ TeV, CMS PAS-HIG-16-038 (Nov 2016), http://cdsweb.cern.ch/record/2231510.
13. CMS Collaboration, Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with $\sqrt{s} = 13$ TeV pp collisions at the CMS experiment, CMS PAS-HIG-16-004 (Mar 2016), http://cdsweb.cern.ch/record/2139578.