Search for associated production of a Higgs boson with a single top quark at $\sqrt{s} = 13$ TeV with the ATLAS detector

A. Tropina[a1], *I. Boyko*[a], *N. Huseynov*[a,b], *A. Dadashova*[a]

[a] Joint Institute for Nuclear Research, Dubna, Russia

[b] Institute of Physics of Azerbaijan National Academy of Sciences, Baku, Azerbaijan

An analysis is presented concerning the search for the associated production of a Higgs boson with a single top quark (tH), a process highly sensitive to the complex phase of the top-quark Yukawa coupling. The analysis uses the full 140 fb^{-1} of ATLAS data at $\sqrt{s} = 13$ TeV, targeting $H \to bb$, WW, ZZ, $\tau\tau$ decays. Advanced multivariate analysis techniques were employed to achieve optimal sensitivity and to separate the tiny tH signal from the large background. The measured signal strength is $\mu_{tH} = 8.1 \pm 2.6(\text{stat.}) \pm 2.0(\text{syst.})$. The observed (expected) significance of the tH signal over the background-only hypothesis is 2.8σ (0.4σ). The observed (expected) upper limit on the tH production cross-section is 13.9 (6.1) times the SM prediction at the 95% confidence level. Interpretations for both Standard Model and inverted top-quark Yukawa coupling scenarios also are presented.

PACS: 14.65.Ha; 14.80.Cp; 14.80.-j

Introduction

Following the Higgs boson's discovery in 2012 [1, 2], the research focus has shifted towards precision measurements of its properties. A central component of this program involves measuring the rates of its various production and decay channels and comparing them with Standard Model (SM) predictions.

The associated production of Higgs boson with a single top quark ($H \to tH$) is the fifth most probable SM production channel. However, with a predicted cross-section of approximately 74 fb [3], its rate is an order of magnitude smaller than that of associated production with a top-quark pair ($H \to t\bar{t}H$), which is forth most probable SM production mode. This suppression arises from a destructive interference between the two leading Feynman diagrams shown in Fig. 1.

As the most massive SM particle, the top quark possesses the largest Yukawa coupling, y_t. While its magnitude has been measured and found to be consistent with the SM via direct ttH measurements, the sign of y_t remains unconstrained. This property can only be probed in tH production. A simplest beyond-the-SM scenario is the Inverted Top Coupling (ITC) model, which postulates $y_t^{ITC} = -y_t^{SM}$. In this model, the destructive interference

[1]E-mail: atropina@jinr.ru

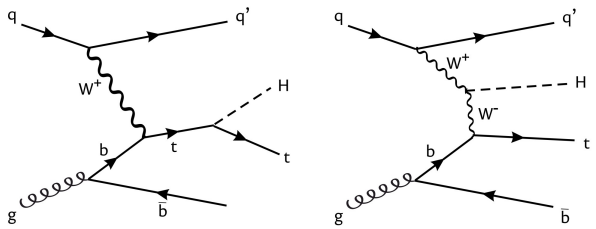


Fig. 1. The leading Feinman diagrams for the $pp \to tH$ production.

is transformed into a constructive one, enhancing the tH production cross-section to 848 fb [3] –more than an order of magnitude larger than the SM prediction. This makes the search for tH production a sensitive probe for physics beyond the Standard Model.

Analysis strategy

This search is conducted through three mutually orthogonal channels, categorized by the number of light-flavor leptons (electrons or muons) in the final state:

- The one-lepton ($1l$) channel.
- The two same-sign lepton ($2lSS$) channel.
- The three-lepton ($3l$) channel.

Events containing hadronically decaying tau leptons (τ) are explicitly vetoed in all channels, while those with τ decaying leptonically are included via their electron or muon decay products. The primary Higgs boson decay modes targeted are $H \to b\bar{b}$ in the $1l$ channel and $H \to WW^*$ in the $2lSS$ and $3l$ channels. The multilepton channels also receive sub-dominant contributions from $H \to \tau\tau$ and $H \to ZZ^*$ decays.

Given their distinct background compositions, the three channels are analyzed independently and combined afterwards. Each channel is split into multiple analysis regions: a Signal Region (SR) optimized for signal enrichment, and several Control Regions (CRs) designed to constrain the dominant backgrounds. The $H \to b\bar{b}$ channel defines one SR and one CR dedicated to $t\bar{t}$ production with additional b-jets. The $2lSS$ and $3l$ channels each define one SR, three CRs to constrain backgrounds from non-prompt and misidentified leptons, and one CR targeting $t\bar{t}W$ production.

The separation of signal from background and separation of different types of background from each other rely on the information from the pseudo-continuous flavour tagging (CBT). The CBT calibration constants are available for jets within certain p_T range. The maximum jet p_T is 400 GeV/c for the b-jets, 250 GeV/c for c-jets and 300 GeV/c for the light jets. CBT information can be safely used in the analysis if the fraction of jets above the

mentioned limits is less than 10%. Therefore, the distribution of the jet p_T was compared for signal and main sources of background and for jets of different flavours. The fractions of the high-p_T jets in $1l$ channel are summarized in Table 1. In all cases this fraction is well below the 10% threshold.

Flavour	Signal	Background
b-jets	0.6%	0.3%
c-jets	2.8%	0.7%
light jets	4.0%	1.5%

Flavour	Signal	Background
b-jets	0.3%	0.1%
c-jets	2.3%	1.3%
light jets	4.0%	2.0%

Table 1. Fraction of high-p_T jets for preselection (left) and signal (right) regions of $1l$ channel.

To maximize the separation of the tH signal from backgrounds, all channels employ multivariate analysis techniques based on Boosted Decision Trees (BDTs). These classifiers are used both to define the analysis regions and to enhance the sensitivity in the final statistical inference. A cross-check analysis using sequential cuts on the kinematic variables (C&C analysis) was performed to validate the results obtained with the BDT. The pre-fit signal significance obtained by C&C is 0.224σ for SM (with 0.156σ in preselection region) and 3.63σ for ITC (with 2.45σ in preselection region). Since the BDT shows better performance, it is chosen to obtain the final results.

The signal strength parameter, $\mu_{tH} = \sigma_{measured}/\sigma_{predicted}$, is extracted using a profile likelihood fit, that is performed simultaneously across all analysis regions and channels. It extracts the signal strength and allows the normalizations of the dominant background processes to be constrained by data. The results are derived first from channel-specific fits and subsequently from a combined fit of all channels.

Sources of systematic uncertainty

The measurement incorporates a comprehensive set of systematic uncertainties, which are categorized as follows: instrumental uncertainties arising from the reconstruction of physics objects; uncertainties related to the estimation of backgrounds from non-prompt and misidentified leptons; and theoretical modeling uncertainties in the simulated signal and background processes. The latter ones are evaluated by varying renormalization and factorization scales, comparing different parton shower and hadronization models, and assessing parton distribution functions. All mentioned uncertainties are incorporated into the analysis as nuisance parameters in the final statistical fit.

One of the theoretical modeling uncertainties arise from the choice between five-flavour-scheme (5FS) and four-flavour-scheme (4FS) of MC sample. The most dependant on the flavour-sheme is $t\bar{t}$ process. In the $1l$ channel, where $t\bar{t}$ production constitutes the dominant background, the associated modeling uncertainties are evaluated independently for its three

sub-categories: $t\bar{t}+ \geq 1b$, $t\bar{t}+ \geq 1c$, and $t\bar{t}+ \geq 0$ light jets. As the 4FS simulation includes a matrix element calculation of the $t\bar{t}+ \geq 1b$ process, it is more accurate than the nominal 5FS simulation, which has matrix element calculation of tt + jets. Table 2 shows the size of the difference (evaluated as the separation power) in the SR between the 4FS and 5FS samples for the five variables which discriminate the most between tH and the 5FS $t\bar{t}+ \geq 1b$ background. These variables are therefore expected to be most important for the signal BDT and hence to have the largest effect on the result. Table 2 proves that the signal BDT is learning most information from variables which differ between the 4FS and 5FS predictions.

Variable	S(tHq, 5FS) (%)	S(4FS, 5FS) (%)	Ranking w.r.t. S(4FS, 5FS)
Sphericity	6.54	0.279	2^{nd}
$M(b_{\text{top}}, j_{\text{tag}})$	4.97	0.276	3^{rd}
$p^T_{\text{light},0}$	4.64	0.110	9^{th}
$\Delta\eta_{\text{top-higgs}}$	4.64	0.0565	12^{th}
$m_{t_{\text{had}}}(\chi^2_{\text{min, ttl}})$	4.39	0.174	5^{th}

Table 2. The separation power (S) between signal and $t\bar{t}+ \geq 1b$ background in the SR for the most important for BDT training variables. The separation power between the 4FS and 5FS $t\bar{t}+ \geq 1b$ samples for these variables is also shown. The last column shows the order of importance of different variables in the separation between 4FS and 5FS.

Results

A central result of this search is the measurement of the tH signal strength μ_{tH}. Under the SM hypothesis, we found $\mu_{tH}^{SM} = 8.1 \pm 2.6(\text{stat.}) \pm 2.0(\text{syst.})$ (Figure 2), which corresponds to a measured cross-section of 720 ± 270 fb [4]. The significance of the observed signal above the background-only hypothesis is 2.8σ. At the 95% confidence level, we exclude tH production cross-sections larger than 13.9 times the SM prediction (Figure 2) [4]. The dominant source of uncertainty for the SM is statistical uncertainty, pointing to the potential for improved precision with future data.

Furthermore, was tested the ITC hypothesis, for which the measured signal strength is $\mu_{tH}^{\text{ITC}} = 1.2 \pm 0.4(stat.) \pm 0.5(syst.)$ (Figure 3), corresponding to a measured cross-section of 1020 ± 540 fb [4]. At the 95% confidence level, we exclude tH production cross-sections larger than 2.4 times the ITC prediction (Figure 3) [4]. Although the data show a preference for the ITC interpretation, both the SM and ITC signal hypotheses are consistent with the observed excess within uncertainties. It was calculated that from combined Run-2 and Run-3 statistics the ITC scenario will be discovered at 5σ level or disfavored at 4σ level.

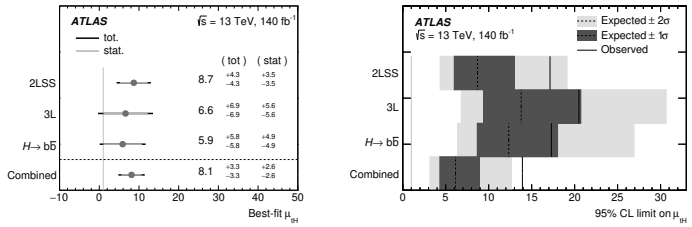


Fig. 2. The fitted values of the μ_{tH}^{SM} (left) and 95% CL upper limits in the individual channels and in the combined measurement in the SM scenario (right). The observed limits are shown (solid black lines), together with the expected limits under the SM hypothesis (dotted black lines) [4].

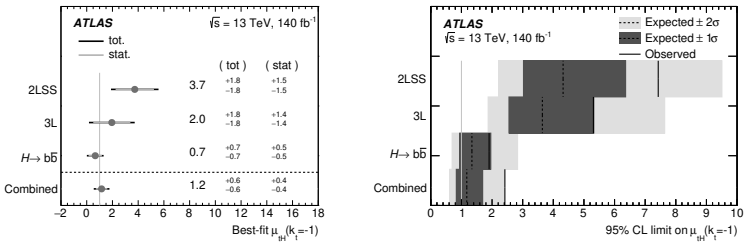


Fig. 3. The fitted values of the μ_{tH}^{ITC} (left) and 95% CL upper limits in the individual channels and in the combined measurement in the ITC scenario (right). The observed limits are shown (solid black lines), together with the expected limits under the SM hypothesis (dotted black lines) [4].

REFERENCES

1. *Aad G. et al.* [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. — 2012. — V. 716. — P. 1–29. — arXiv:1207.7214 [hep-ex].

2. *Chatrchyan S. et al.* [CMS Collaboration] Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC // Phys. Lett. B. — 2012. — V. 716. — P. 30–61. — arXiv:1207.7235 [hep-ex].

3. *de Florian D. et al.* [LHC Higgs Cross Section Working Group Collaboration] Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector // CERN Yellow Rep. Monogr. — 2017. — V. 2. — P. 1–869. — arXiv:1610.07922.

4. *Aad G. et al.* [ATLAS Collaboration] Search for the production of a Higgs boson in association with a single top quark in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector // JHEP. — 2025. — V. 10. — P. 093. — arXiv:2508.14695.