
CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch

2016/08/04

Search for $H \rightarrow b\bar{b}$ in association with a single top quark as a test of Higgs boson couplings at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

A direct search for the production of a Higgs boson in association with a single top quark is performed. The analysis considers single top quark production either via the t channel or via the associated production with a W boson and uses Higgs boson decays to a bottom quark-antiquark pair and semileptonic top quark decays. Data recorded by the CMS detector in 2015 in pp collisions at a center-of-mass energy of 13 TeV, amounting to an integrated luminosity of about 2.3 fb^{-1} , are analyzed. Upper limits in the two-dimensional plane spanned by the scaling factors for the coupling strength of the Higgs boson to top quarks and to weak gauge bosons with respect to the standard model predictions, κ_t and κ_V , are determined. The observed (expected) limit on the cross section for the production of a Higgs boson in association with a single top quark in the scenario with a negative Yukawa coupling of the top quark ($\kappa_t = -1.0$ and $\kappa_V = +1.0$) is 6.0 (6.4) times the predicted value.

1 Introduction

In 2012 the ATLAS and CMS collaborations announced the discovery of a new boson [1, 2]. Up to now, all measurements of its properties are consistent with those of the Higgs boson of the standard model (SM). Nevertheless, small deviations from SM predictions could point to physics beyond the standard model (BSM) and precise property measurements are therefore crucial for answering the question whether the found particle is really the Higgs boson predicted by the SM.

One prominent property is the strength of the Yukawa coupling, y_t , of the Higgs boson to the top quarks. Most measurements that probe this property are only sensitive to the magnitude of y_t , rather than its sign, as quantities such as the rate of Higgs boson production in association with top quark pairs depend only on $|y_t|^2$. The sign of y_t provides information about the relative phase between the fermion and boson couplings of the Higgs boson. A few measurements, such as the decay rate of the Higgs boson to photon pairs [3] and the cross section of the Higgs and Z boson associated production via gluon fusion [4], are however sensitive to the sign of y_t , and recent results [5–7] disfavor $y_t < 0$. Yet negative values for y_t are still allowed in theory models where no assumption is made about the absence of new particle contributions to the loop amplitudes [7, 8].

The ratio of the actual coupling strengths to the SM predictions are defined as κ_t for the Higgs-top coupling, and κ_V for the coupling of the Higgs boson to massive vector bosons. The production of a single top quark in association with a Higgs boson (tH) is sensitive to the magnitude and the sign of κ_t [9–11]. At leading order, this process can be separated into three production modes, namely t channel (hereafter referred to as "tHq"), associated tW production (hereafter referred to as "tHW"), and s channel. The latter process has a negligible cross section at the LHC [12] and therefore is not considered in this analysis. In both processes, t -channel and tW-channel production, the Higgs boson can be emitted either from the top quark or the intermediate W boson; the corresponding Feynman diagrams can be found in Fig. 1. The amplitudes of the two possibilities for the emission of the Higgs boson interfere — and the resulting amplitude depends strongly on the values of κ_t and κ_V . In the SM scenario, with $\kappa_t = \kappa_V = +1.0$, the interference is destructive and the predicted cross sections, calculated using MG5_aMC@NLO [13] and the method described in Ref. [12], are 71 fb for tHq and 16 fb for tHW [14] at 13 TeV center-of-mass energy.

Changes in the relative phases can occur in BSM scenarios, either through different couplings of fermions and bosons to the Higgs boson, or by introducing new particles entering loops that cause \mathcal{CP} -violation. Negative Yukawa couplings of the top quark for example would result in an increased cross section and different kinematic properties. The scenario with $\kappa_t = -1.0$ and $\kappa_V = +1.0$ differs only in the sign of κ_t from the SM case, while preserving the magnitudes of the couplings. This scenario will be named ITC (inverted top coupling) in the following. In this case the interference is constructive and the predicted cross sections are 739 fb for tHq and 147 fb for tHW at 13 TeV center-of-mass energy (again calculated using MG5_aMC@NLO and the method described in Ref. [12]).

In the analysis described in this document limits on the cross section of tH production (for both tHq and tHW processes) are derived as a function of the coupling strength factors κ_t and κ_V . See Fig. 2 for an illustration of the impact on the cross section due to variations of the coupling strengths.

For three values of κ_V ($\kappa_V = +0.5$, $\kappa_V = +1.0$, and $\kappa_V = +1.5$), κ_t is varied between -3.0 and $+3.0$ (51 points in total). The κ_V values are chosen because they are not excluded by Run 1

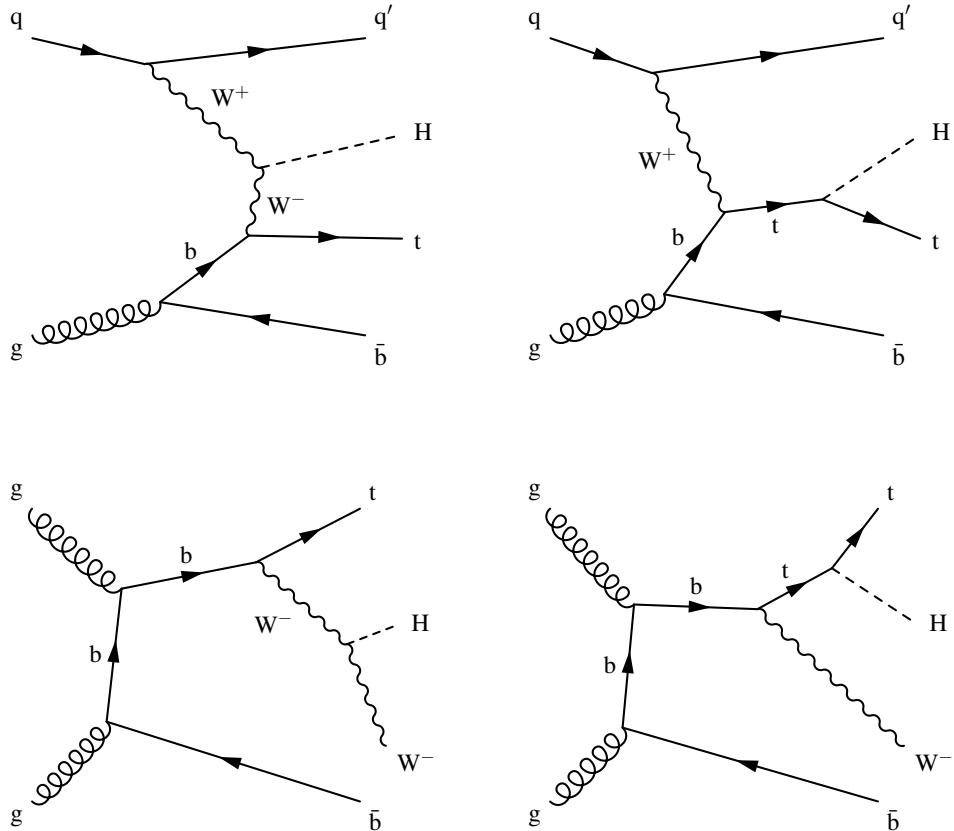


Figure 1: Representative Feynman diagrams for the associated production of a single top quark and a Higgs boson in the t channel (upper row) and in the tW channel (bottom row).

results [7]. These points extend the scope of the predecessor analysis at 8 TeV center-of-mass energy [15], which considered only the ITC scenario and yielded an observed (expected) limit of 7.6 (5.2) times the cross section σ_{ITC} .

The analysis is based on the 2015 dataset collected at 13 TeV center-of-mass energy. The semileptonic decay of the produced top quark and the decay of the Higgs boson into a bottom quark-antiquark pair are considered. Consequently, events whose final state contains exactly one isolated lepton and at least three b-tagged jets (corresponding to the b quark from the top quark decay and the two b quarks from the Higgs boson decay) are selected. As the fourth b jet present in the event (see Fig. 1) features on average a softer transverse momentum spectrum and might therefore not be selected, two signal regions, one with three b-tagged jets and one with four b-tagged jets, are defined. Both signal regions are dominated by backgrounds from top quark pair production ($t\bar{t}$). A characteristic feature of the signal process is the light-flavor jet from the quark recoiling against the top quark and the Higgs boson (hereafter referred to as "recoil jet"). Therefore at least one additional non b-tagged jet is required in each event.

To achieve a better discrimination between the signal processes and the dominating background from $t\bar{t}$ production, kinematic variables of reconstructed objects like top quarks and Higgs bosons are exploited. Each event is reconstructed under both the tHq signal and $t\bar{t}$ background hypotheses. The assignment of jets to final state particles of these two hypotheses is performed using two separate sets of reconstruction boosted decision trees (BDT). The properties of the reconstructed particles are then used, together with additional variables that are independent from the reconstruction, to train classification BDTs, separating the signal and the

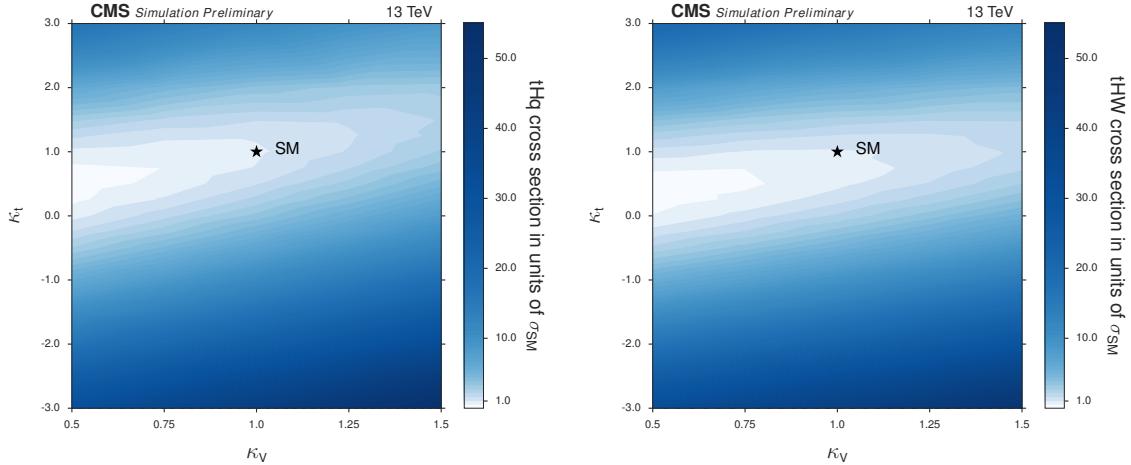


Figure 2: Cross sections in the $\kappa_t - \kappa_V$ plane at 13 TeV for tHq (left) and tHW (right) production. Right figure adapted from [14].

background processes. The classification BDT response distributions are then used to derive upper limits on the signal cross sections for each signal scenario. The limits are derived on the sum of the cross sections for tHq and tHW production, since both have a similar dependency on the studied coupling parameters.

2 Data and simulation

The analysis described in this document uses the dataset recorded by CMS in 2015 at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 2.3 fb^{-1} with an estimated overall uncertainty of 2.7% [16]. A description of the CMS detector can be found in Ref. [17].

The simulation of the tHq and tHW signal processes is performed using MG5_aMC@NLO [13] at LO precision and using the NNPDF3.0 PDF set [18]. The tHq process is generated within the four-flavor scheme and with dynamical factorization and renormalization scales, while for the generation of tHW events fixed scales (40 GeV) and the five-flavor scheme are used. For the tHW process the five-flavor scheme has been chosen because this allows to eliminate interference with the $t\bar{t}H$ production at leading order, which would otherwise complicate the generation. The samples are generated such that they can be reweighted to produce different combinations of κ_t and κ_V .

The backgrounds from single top quark and $t\bar{t}$ production as well as from $t\bar{t}H$ production are generated using the POWHEG event generator [19–21]. The production of $t\bar{t}$ in association with a W boson ($t\bar{t}W$) or a Z boson ($t\bar{t}Z$) is simulated using MG5_aMC@NLO. The tZq , $Z \rightarrow b\bar{b}$ background, which has a cross section similar to the SM tHq process, is not included; its impact on the result is expected to be small, due to the discriminating power of the kinematic observables used in the analysis. The production of a W boson together with jets (W+jets) is simulated using MADGRAPH [22], while Z+jets events are simulated using MG5_aMC@NLO. For all samples the modeling of the parton shower is performed using PYTHIA 8 [23]. A top quark mass of 172.5 GeV and a Higgs boson mass of 125 GeV are used for the simulations.

The effect of additional proton-proton interactions (pileup) is simulated by superimposing minimum bias events generated with PYTHIA, taking into account in-time and out-of-time pileup contributions. All generated events undergo a full GEANT 4 [24, 25] simulation of the detector

response.

3 Event selection

In order to distinguish signal events from the overwhelming background, the analysis focuses on semileptonically decaying top quarks with an electron ("electron channel") or muon ("muon channel") in the final state exclusively. The analysis is performed in the combined "muon+electron channel". No attempt is made to reconstruct τ -leptons, and the corresponding decays are included through decays of τ leptons to muons or electrons.

The analysis uses collision events collected by two trigger paths. The muon trigger path requires a muon candidate with transverse momentum (p_T) $> 20\text{ GeV}$ and the electron trigger path requires one electron candidate with $p_T > 27\text{ GeV}$ and a pseudorapidity (η) < 2.1 .

Additional selection criteria require an event in the muon (electron) channel to contain exactly one muon (electron) candidate with p_T larger than 25 GeV (30 GeV) and within $|\eta| < 2.4$ ($|\eta| < 2.1$). Strict selection criteria are applied to reject leptons produced inside jets. In order to suppress the contribution from Drell-Yan and other processes in which multiple prompt leptons are produced, events are rejected if an additional muon or electron is found passing lower-quality criteria. These electron or muon candidates are required to have $p_T > 15\text{ GeV}$.

Particle candidates are reconstructed with the Particle Flow (PF) algorithm [26, 27] and clustered into jets using the anti- k_T algorithm with a distance parameter of 0.4 [28]. The measured jet energy is corrected in data and simulated events [29] to account for known detector effects and pileup interactions. The jet energy resolution is modified in simulated events to match that found in data. All central ($|\eta| < 2.4$) jets with $p_T > 20\text{ GeV}$ and all forward ($2.4 < |\eta| < 4.7$) jets with $p_T > 40\text{ GeV}$ are considered in the analysis.

Table 1: Event yields for tHq and tHW signal (for the SM and ITC scenarios) as well as the various background processes in the two signal regions. The uncertainties include both systematic and statistical uncertainties. Additionally, the numbers of observed events in data are shown.

	3 tag	4 tag
t \bar{t} +LF	2119 ± 651	21.3 ± 21.6
t \bar{t} +c \bar{c}	852 ± 624	39 ± 46
t \bar{t} +b	324 ± 203	18.7 ± 15.3
t \bar{t} +b \bar{b}	333 ± 298	71 ± 67
t \bar{t} +2b	177 ± 102	13.0 ± 9.6
Single top	156 ± 44	6.0 ± 2.4
t \bar{t} H	20.3 ± 9.7	5.3 ± 2.9
t \bar{t} Z	9.4 ± 2.3	1.8 ± 1.7
t \bar{t} W	8.0 ± 2.5	0.4 ± 0.4
W+jets	42 ± 35	0.0 ± 0.0
Z+jets	10.2 ± 5.0	0.0 ± 0.0
Sum of Backgrounds	4051 ± 978	177 ± 86
tHq (SM)	0.77 ± 0.21	0.12 ± 0.04
tHW (SM)	0.61 ± 0.09	0.09 ± 0.02
tHq (ITC)	11.2 ± 3.1	1.7 ± 0.6
tHW (ITC)	6.7 ± 1.0	1.1 ± 0.3
Observed	3603	171

As the final state of the signal process contains in total four b quarks, the identification of jets

stemming from the hadronization of b quarks is crucial for this analysis. The medium working point of the CSVv2 b-tagging algorithm [30], which has a tagging efficiency of approximately 69% for b jets and a misidentification probability of about 1% for light-flavor jets, is used for the analysis. The CSV distribution in simulation is corrected in order to match the one observed in data using a tag-and-probe method in $Z+jets$ (for light-flavor jets) and $t\bar{t}$ (for heavy-flavor jets) events, which is further described in Ref. [30].

The fourth b quark in the event is on average produced with a softer p_T spectrum and the corresponding jet might therefore not pass the jet p_T threshold. Two signal regions are defined based on the total number of jets and the number of b-tagged jets in the event. Central jets contribute to the number of jets only if their p_T is larger than 30 GeV. The first signal region contains exactly three b-tagged jets and the second signal region contains exactly four b-tagged jets. Requiring at least three b-tagged jets significantly reduces the amount of $W+jets$ and QCD multijet events, with a moderate impact on the signal. Both regions contain at least one additional jet.

To further reject QCD multijet contributions a requirement on the amount of missing transverse energy (E_T^{miss}) [31] in the event is applied, with thresholds optimized individually for each channel: $E_T^{\text{miss}} > 35$ GeV in the muon channel and $E_T^{\text{miss}} > 45$ GeV in the electron channel. This requirement reduces the QCD multijet contributions to a negligible level and this background source is therefore not considered further in the analysis.

Table 1 summarizes the expected yields of background and signal processes after applying the full set of event selection requirements. The simulated $t\bar{t}$ sample is split according to the flavor of the additional partons in the final state. Five different categories of $t\bar{t}$ production with additional heavy-flavor ($t\bar{t}+HF$) or light-flavor ($t\bar{t}+LF$) jets are distinguished and treated separately: $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, and $t\bar{t}+LF$. The $t\bar{t}+2b$ category comprises those events where two b quarks end up being clustered into a single jet, as it often happens with b quarks produced in a gluon splitting. For the signal processes, numbers are given for the SM and ITC scenarios.

4 Signal extraction

The high jet multiplicity and small signal-to-background ratio after the event selection make it essential to exploit multivariate analysis techniques, both to reconstruct events and to discriminate the signal process from the background. For this purpose, the analysis utilizes boosted decision trees (BDTs) implemented in TMVA [32].

First, every event is reconstructed under two different hypotheses, where the jets are assigned to the quarks in the final states of both the tHq signal process and the $t\bar{t}$ background process. The tHW process is not reconstructed separately as it kinematically lies between the tHq and $t\bar{t}$ process and hence can also be separated from the background with these two reconstructions. The choice of the assignment is based on the responses of the aforementioned BDTs. A number of observables are then derived for each event, from both hypotheses. Together with reconstruction-independent variables, they are exploited to discriminate signal events from background in the two signal regions.

4.1 Jet assignment

For each of the two reconstruction hypotheses several possibilities exist to assign the jets to the respective final-state quarks. The number of possible assignments is reduced by utilizing the b-tagging information of the jets. In the tHq reconstruction, the jets assigned to the final-

state b quarks need to be within the tracker acceptance ($|\eta| < 2.4$), and the jet assigned to the light-flavor quark must not be b-tagged. In the $t\bar{t}$ reconstruction, only those combinations are considered in which the jets assigned to the b quarks from the top quark decays are b-tagged.

Two dedicated BDTs are then trained to find for each event and each hypothesis the best possible jet-quark assignment. Each BDT is trained with correct and wrong jet-quark assignments from simulated events of the corresponding process. Each simulated event contributes one correct and one wrong assignment, where the latter is randomly chosen from all wrong assignments in the event. If an event does not have a correct interpretation, it is not used for training. The correct assignment is defined as the configuration in which every jet is within a cone of the size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.3$ around the final-state quark it is assigned to. If at least one of the final-state quarks fails this requirement, the assignment is considered to be wrong. The best possible jet-quark assignment is subsequently chosen by selecting the assignment that produces the highest BDT response in an event.

Since the kinematics of the Higgs boson and the top quark produced in the signal process depend on the couplings, the tHq assignment is done separately for each of the 51 different $\kappa_t - \kappa_V$ combinations. Each of the trainings is evaluated and checked for possible signs of overtraining by splitting the sample into dedicated training, testing, and evaluation sets. Only the latter are then used in the analysis. Tables 2 and 3 list the variables used in the two BDTs for the jet-quark assignment under the tHq and the $t\bar{t}$ hypotheses, respectively, sorted by their importance in the training. Kinematic variables of reconstructed objects like top quarks or Higgs bosons refer to the kinematic variables of the respective object candidates present in each possible assignment.

Figure 3 shows the distribution of the response of the tHq reconstruction BDT for two coupling scenarios (SM and ITC), obtained from simulated tHq events. The response distribution of the reconstruction BDT under the hypothesis of a $t\bar{t}$ event, obtained from simulated $t\bar{t}$ events, is shown in Fig. 4. Figure 5 shows the reconstruction efficiencies under the two different event hypotheses, obtained from the corresponding simulated events. The efficiencies are calculated with respect to the number of events for which a correct assignment exists. The matching efficiency for a complete tHq event is 47% (32%) in the 3 tag (4 tag) region, the matching efficiency for the complete $t\bar{t}$ event is 68% (40%).

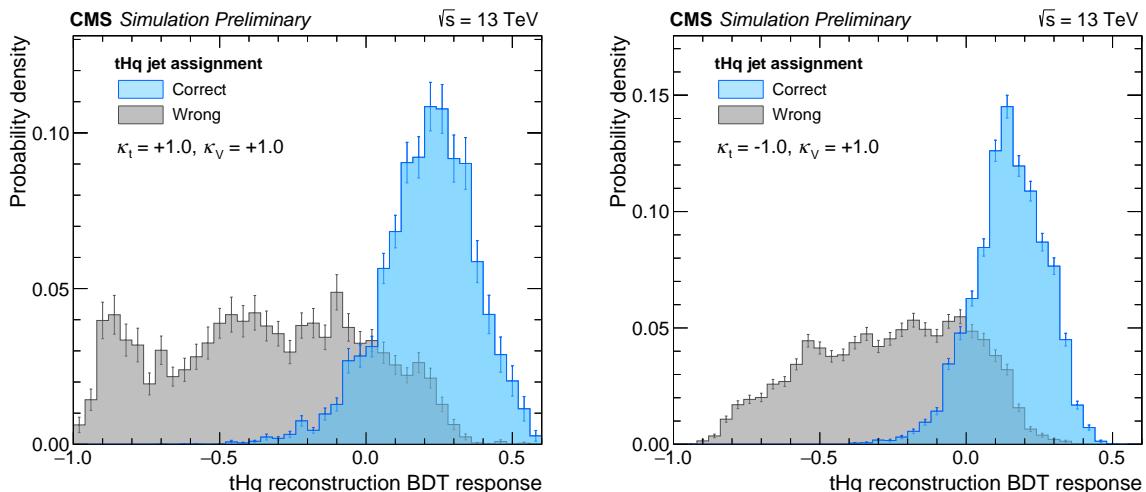


Figure 3: Output values of the tHq reconstruction BDTs for correct and wrong jet assignments for the coupling case of the SM prediction (left) and the ITC scenario (right).

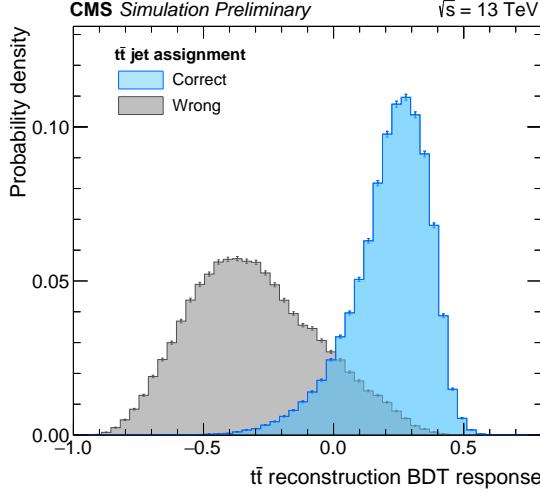


Figure 4: Output values of the $t\bar{t}$ reconstruction BDTs for correct and wrong jet assignments. Only $t\bar{t}$ events are used.

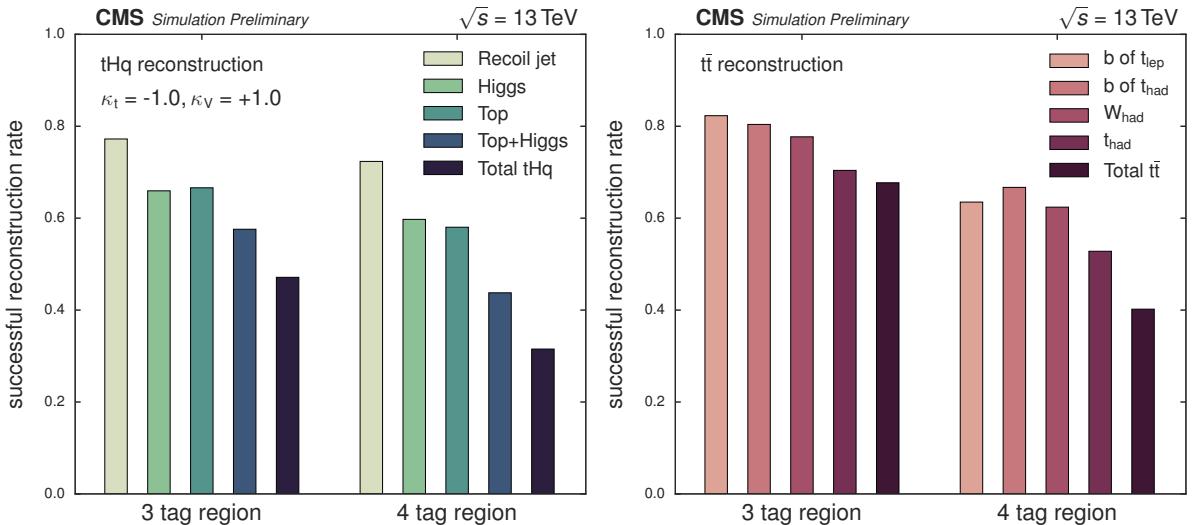


Figure 5: Reconstruction efficiencies for different objects of the tHq process obtained from the tHq reconstruction in the 3 and 4 tag regions (left). Reconstruction efficiencies for different objects of the $t\bar{t}$ process obtained from the $t\bar{t}$ reconstruction in the 3 and 4 tag regions (right).

Table 2: Input variables for the jet-assignment BDT under the tHq hypothesis sorted by their importance in the training.

Variable	Description
$\log m(H)$	Invariant mass of the reconstructed Higgs boson
$\log m(t)$	Invariant mass of the reconstructed top quark
$\Delta R(\text{Higgs jets})$	ΔR between the two jets from the Higgs boson decay
$\Delta R(b_t, W)$	ΔR between the jet assigned to the b quark from the top quark decay and the W boson
relative H_T	Ratio of $p_T(H) + p_T(t) + p_T(\text{recoil jet})$ to the scalar sum of p_T of all jets, charged lepton, and E_T^{miss}
$\cos \theta(t, \ell)$	Cosine of the angle between the top quark momentum and the sum of momenta of top quark and charged lepton, in their common rest frame
CSV(Higgs jet 2)	Output of the CSVv2 b-tagging algorithm for the second hardest jet assigned to the Higgs boson
CSV(b_t)	Output of the CSVv2 b-tagging algorithm for the jet assigned to the b quark from the top quark decay
$ \eta(\text{recoil jet}) - \eta(b_t) $	Absolute difference of pseudorapidities of the recoil jet and of the b jet from the top quark decay
CSV(Higgs jet 1)	Output of the CSVv2 b-tagging algorithm for the hardest jet assigned to the Higgs boson
$ \eta(b_t) $	Absolute pseudorapidity of the jet assigned to the b quark of the top quark decay
$ \eta(t) - \eta(H) $	Absolute difference of pseudorapidities of reconstructed top quark and the reconstructed Higgs boson
$\log \min(p_T(H \text{ jets}))$	Lower transverse momentum of the two jets assigned to the Higgs boson decay products
$ \eta(\text{recoil jet}) $	Absolute pseudorapidity of the recoil jet
$\Delta E(\text{recoil jet}, b_t)$	Energy difference between the recoil jet and the jet assigned to the b quark from the top quark decay

Table 3: Input variables for the jet-assignment BDT under the $t\bar{t}$ hypothesis sorted by their importance in the training. The hadronically and leptonically decaying W bosons and top quarks are labeled with W_{had} , W_{lep} , t_{had} , and t_{lep} .

Variable	Description
$\log m(W_{\text{had}})$	Invariant mass of the two jets assigned to the W boson of t_{had}
$\log (m(t_{\text{had}}) - m(W_{\text{had}}))$	Difference between the invariant masses of reconstructed t_{had} and W_{had}
$\log m(t_{\text{lep}})$	Invariant mass of the reconstructed t_{lep}
$\text{CSV}(W_{\text{had}} \text{ jet 1})$	CSVv2 output of the hardest jet assigned to W_{had}
$\Delta R(b_{t_{\text{lep}}}, W_{\text{lep}})$	ΔR between the b quark of the reconstructed t_{lep} and W_{lep}
$\text{CSV}(W_{\text{had}} \text{ jet 2})$	CSVv2 output of the second hardest jet assigned to W_{had}
$\Delta R(W_{\text{had}} \text{ jets})$	ΔR between the two jets assigned to the W boson of t_{had}
relative H_T	Ratio of $p_T(t_{\text{had}}) + p_T(t_{\text{lep}})$ to the scalar sum of p_T of all jets, charged lepton, and E_T^{miss}
$\Delta R(b_{t_{\text{had}}}, W_{\text{had}})$	ΔR between the b quark of the reconstructed t_{had} and W_{had}
$\log p_T(t_{\text{had}})$	Transverse momentum of the reconstructed t_{had}
$\log p_T(t_{\text{lep}})$	Transverse momentum of the reconstructed t_{lep}

4.2 Discrimination between signal and background

The discrimination of signal and background is performed by employing three sets of variables: one set describing the reconstructed objects of the tHq jet assignment, one set describing the reconstructed objects of the $t\bar{t}$ assignment and one set of variables independent of any assignment. The classification is performed separately for each point in the $\kappa_t - \kappa_V$ plane. While the same set of variables is used in each separate classification, the definitions of observables relying on the tHq reconstruction differ for every probed coupling point. A total of 15 different variables is used; a description of the variables can be found in Table 4.

The training is performed in the 3 tag region, where enough simulation events are available, and only the tHq process is used as signal input. The classification BDT is trained with background events from a mixture of semileptonic $t\bar{t}$, dileptonic $t\bar{t}$, and $t\bar{t}H$ events. The background events are scaled to their predicted cross sections and the signal is scaled such that the integral coincides with the number of the background events. Figure 6 shows the ROC curves of the event classification for the SM and the ITC scenarios.

The classification BDT responses show a clear separation between signal events and background events. For the analysis, the classification BDT is applied to simulation samples and data samples. The distributions in Fig. 7 show the most important variables in the training of the classification BDT for the ITC scenario and the BDT response for the SM and ITC scenario in the $t\bar{t}$ control region, which requires exactly 2 b -tagged jets (2 tag region).

Table 4: Input variables used in the classification, ranked by their importance within each category.

Variable	Description
Variables independent of any reconstruction	
aplanarity	Aplanarity of the event
log m3	Invariant mass of three hardest jets in the event
Fox-Wolfram #1	First Fox-Wolfram moment of the event
q(ℓ)	Electric charge of the lepton
Variables based on objects reconstructed under the $t\bar{t}$ hypothesis	
log m(t_{had})	Invariant mass of t_{had}
CSV(W_{had} jet 1)	CSVv2 output of the hardest jet assigned to W_{had}
ΔR(W_{had} jets)	ΔR between the two jets from the decay of W_{had}
CSV(W_{had} jet 2)	CSVv2 output of the second hardest jet assigned to W_{had}
Variables based on objects reconstructed under the tHq hypothesis	
η(recoil jet) 	Absolute pseudorapidity of the recoil jet
CSV(Higgs jet 2)	CSVv2 output of the second hardest jet assigned to the Higgs boson
CSV(Higgs jet 1)	CSVv2 output of the hardest jet assigned to the Higgs boson
log p_T(recoil jet)	Transverse momentum of the recoil jet
log p_T(Higgs)	Transverse momentum of the Higgs boson
η(Higgs) 	Absolute pseudorapidity of the Higgs boson
$\cos \theta(t, \ell)$	Cosine of the angle between the top quark momentum and the sum of top quark and charged lepton, in their common rest frame

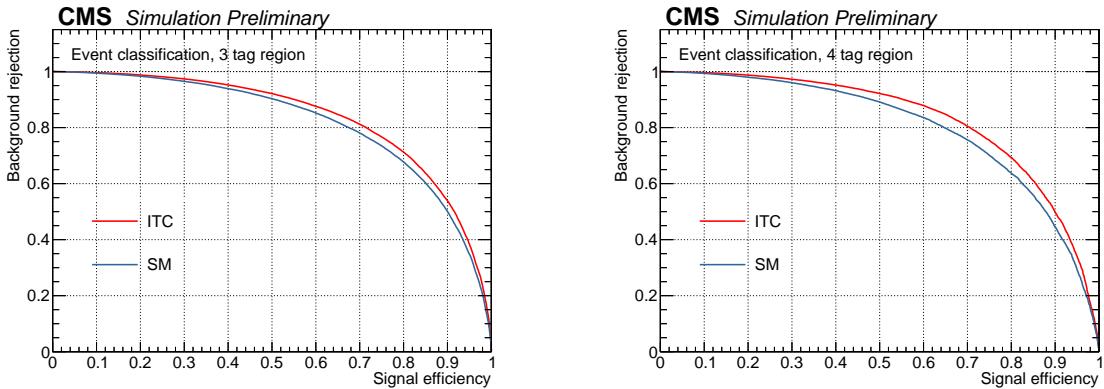


Figure 6: ROC curve of the classification BDT for the SM and the ITC scenario in the 3 tag (left) and 4 tag (right) regions as examples for the 51 different classification BDTs.

5 Systematic uncertainties

The upper limit on the cross section of the $t\bar{t}$ production is affected by several sources of systematic uncertainties. These uncertainties can be grouped into experimental and theoretical uncertainties and they are parametrized with nuisance parameters in the statistical inference performed in the final analysis step.

The uncertainty on the measurement of the luminosity (2.7% [16]) affects the normalization of all processes. Uncertainties in the estimation of the corrected lepton efficiencies are taken into account by varying the efficiencies by 2%. The uncertainty in the distribution of the number of pileup interactions is applied by rescaling the target profile of the instantaneous luminosity used to correct simulation by $\pm 5\%$. The applied jet energy corrections are varied within their uncertainties. A smearing is applied to account for the known difference in jet energy resolution with respect to data [33], increasing or decreasing the resolutions by their uncertainties. For both jet energy uncertainty categories the resulting changes are propagated to the calculation of E_T^{miss} . In addition, the contributions of unclustered particles to E_T^{miss} are varied within their respective energy resolutions [31]. The evaluation of the systematic uncertainties on the CSVv2 b-tagging scale factors follows what has been done in Ref. [34].

The $t\bar{t}+HF$ categories are important sources of irreducible background since neither higher-order theoretical calculations nor control region studies are able to constrain the normalization of these processes to better than 50%. Therefore an additional 50% rate uncertainty is assigned for the $t\bar{t}+HF$ categories. Variations of the factorization and renormalization scales (μ_R, μ_F) at the matrix element are taken into account by reweighting the simulated events to produce samples with doubled or halved scales in a correlated manner. These weights are available for all processes except for the single top samples. The latter are assigned a 4.0% rate uncertainty to cover the effect of the variation of these scales. For the largest background contribution, stemming from $t\bar{t}$ production, additional μ_R and μ_F variations in the parton shower are taken into account. Uncertainties from the choice of the parton distribution functions (PDF) in the simulation are considered only as rate uncertainties since the shape variations have been found to be negligible.

Figure 8 lists the relative changes in the expected limit when a specific uncertainty is omitted from the computation of the expected limit (while all the rest is considered), as well as the change in the limit when only the given uncertainty is considered. The largest systematic uncertainty arises from variations in the jet energy scale and from variations in the μ_R and μ_F

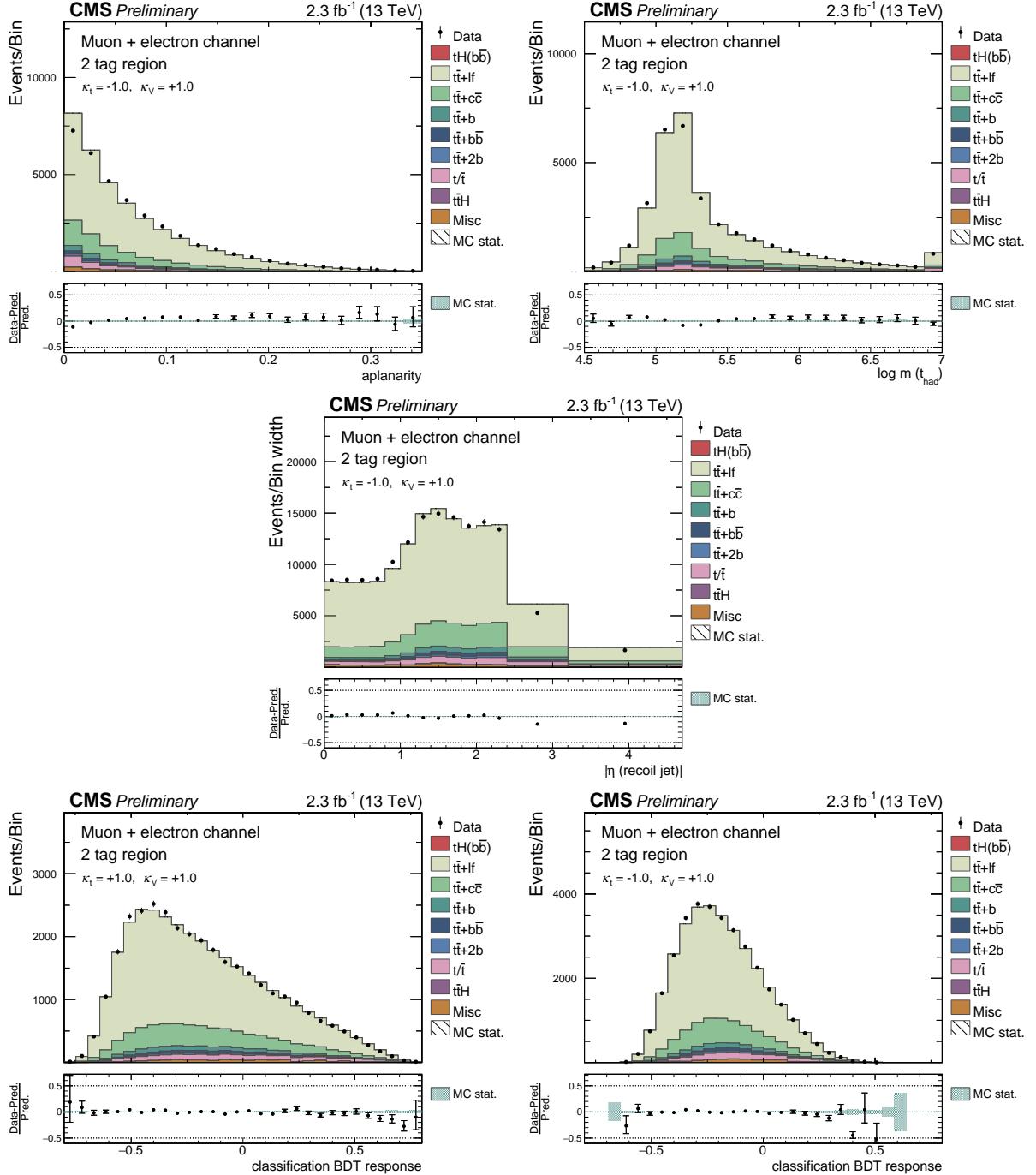


Figure 7: Simulation to data comparisons for the most discriminating variable of each category as shown in Table 4 (1st and 2nd row) and the response distribution of the classification BDTs (3rd row) in the 2 tag region at 13 TeV for the SM (left) and the ITC (right) case. The simulation is scaled to match the event yield observed in data. Contributions from $t\bar{t}Z$, $t\bar{t}W$, $W+jets$, and $Z+jets$ are summarized under the category "Misc".

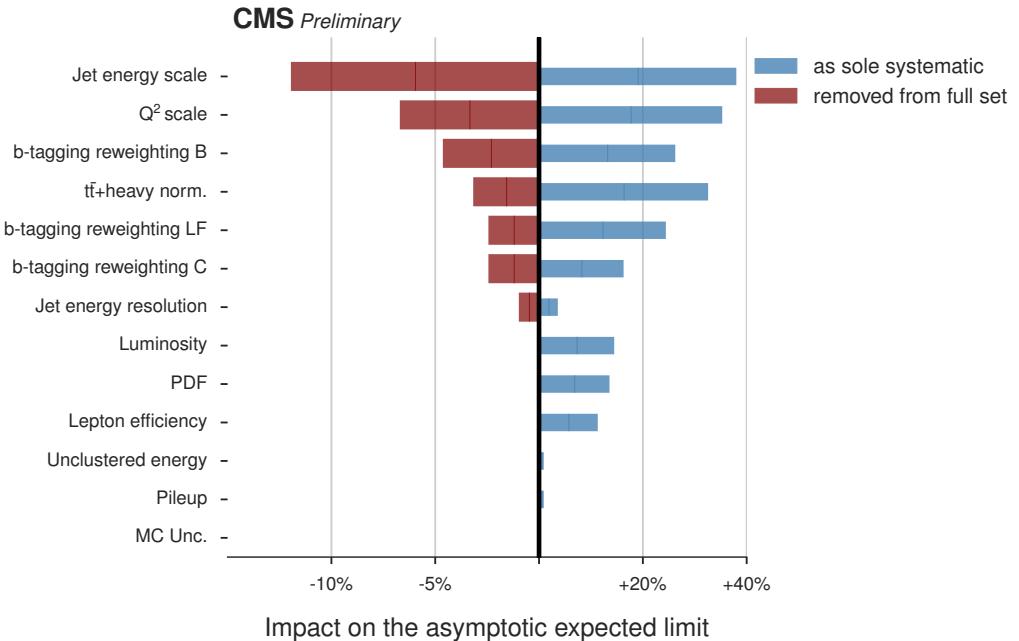


Figure 8: Impact of groups of systematic uncertainties on the expected asymptotic limit. The groups of systematic uncertainties are either removed from the fit by fixing them to their postfit values, or used as single systematic uncertainty by fixing all other uncertainties to their post-fit values. The changes displayed in this diagram are calculated relative to the limit with all systematic uncertainties included (red bars) and to the limit where all uncertainties are fixed to their best fit values (blue bars).

scales used in the generation of the $t\bar{t}$ and tHq samples.

6 Results

The upper asymptotic CL_S limits [35] at 95% CL are calculated for each of the 51 coupling configurations by fitting simultaneously the corresponding BDT response distributions in the 3 tag and 4 tag regions. In Fig. 9 these distributions after the fit are shown for the SM and ITC coupling scenarios.

The expected and observed upper limits on the combined production rates of the tHq and tHW processes (tH) for all 51 studied couplings are determined and can be found in Fig. 10. Table 5 lists the limits for the SM and ITC coupling scenarios. The upper limit for the SM tH production is $113.7 \times \sigma_{SM}$ with an expected sensitivity of $98.6 \times \sigma_{SM}$. The observed limit is well within one standard deviation of the expected limit. For the ITC scenario the limit is $6.0 \times \sigma_{ITC}$, which is within one standard deviation of the expected limit of $6.4 \times \sigma_{ITC}$.

The sensitivity of this analysis is already comparable to that of the Run 1 analysis which yielded an expected limit of $5.4 \times \sigma_{ITC}$ for the ITC scenario. This is explained by the inclusion of the tHW process, the increase of the tH production cross section, and a higher selection efficiency due to the use of a looser b-tagging working point.

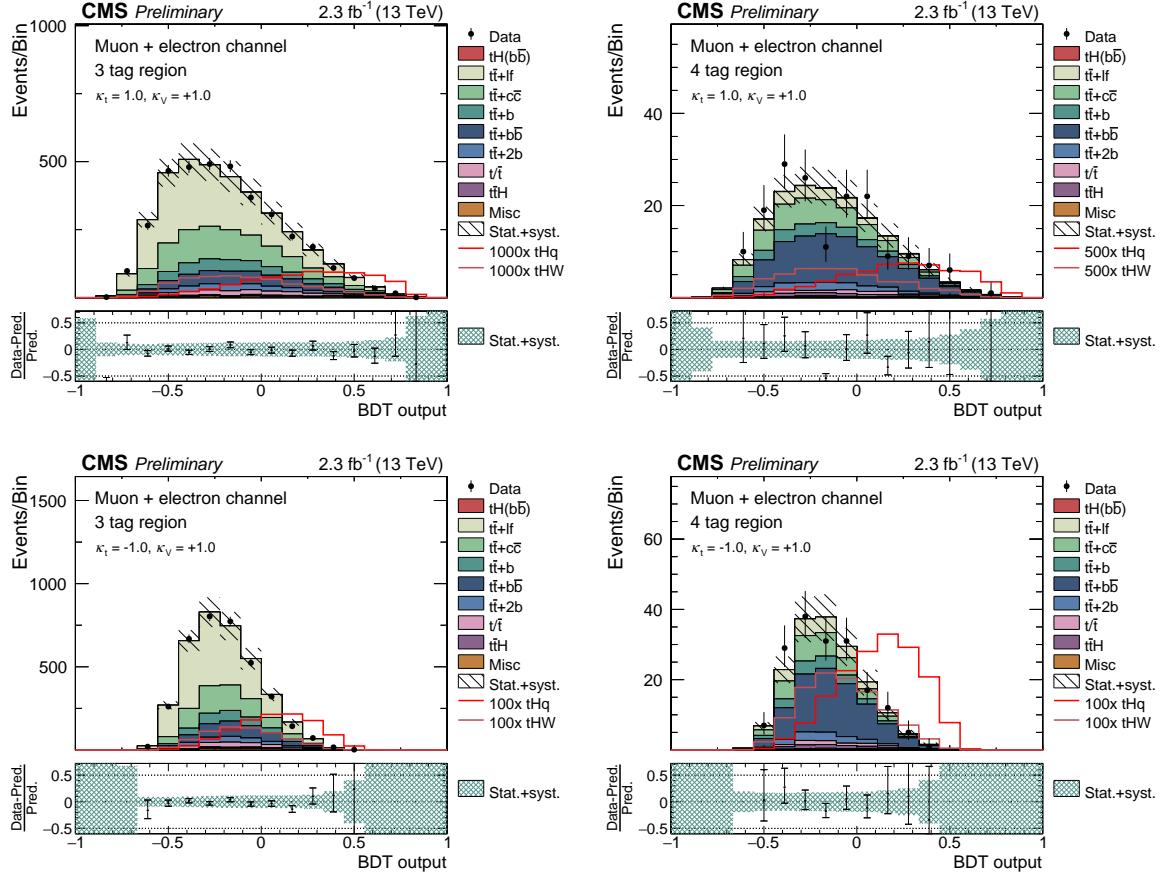


Figure 9: Postfit distributions of the classification BDT response in the 3 tag (left) and 4 tag (right) regions for the SM (top) and ITC (bottom) coupling scenarios. The uncertainty bands contain both systematic and statistical contributions. The signal distributions correspond to the expected contributions scaled by the factors given in the legends.

Table 5: Expected and observed asymptotic CL_s limits at 95% CL in the 3 tag and 4 tag regions and their combination for the SM and ITC coupling scenarios. Also the 68% and 95% uncertainty band values are shown.

	Region	Observed Limit	Expected Limit		
			Median	$\pm 1\sigma$	$\pm 2\sigma$
SM scenario	3 tag	124.0	114.3	[73.6, 184.4]	[52.0, 295.2]
	4 tag	195.8	174.6	[112.9, 287.4]	[78.8, 464.4]
	Combination	113.7	98.6	[64.0, 159.2]	[45.3, 254.8]
ITC scenario	3 tag	7.4	7.4	[4.9, 11.6]	[3.5, 17.8]
	4 tag	9.2	10.0	[6.5, 16.3]	[4.5, 26.3]
	Combination	6.0	6.4	[4.2, 10.1]	[3.0, 15.7]

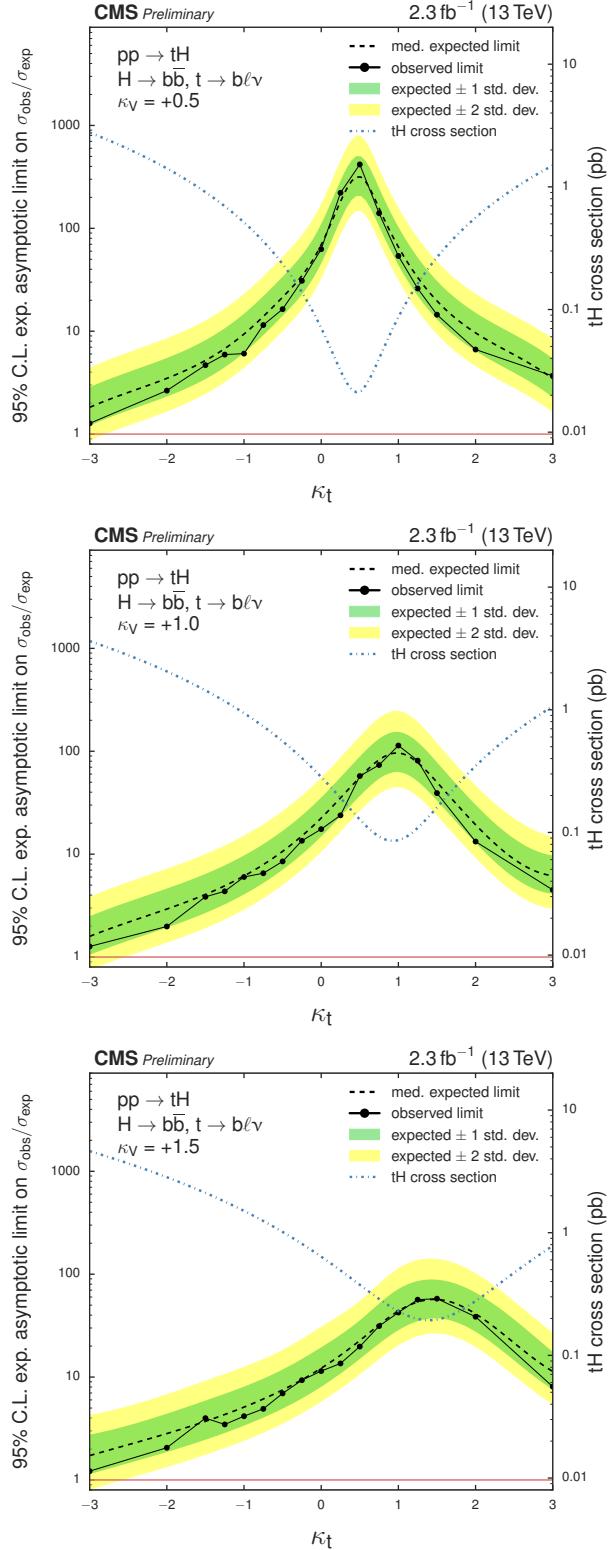


Figure 10: Upper limits on $t\bar{H}$ scenarios with different $\kappa_t - \kappa_V$ configurations, for $\kappa_V = +0.5$ (top), $\kappa_V = +1$ (middle), and $\kappa_V = +1.5$ (bottom). In addition also the $t\bar{H}$ cross sections are given (right y axis).

7 Summary

A search for t - and tW -channel single top quark production in association with the 125 GeV Higgs boson has been conducted using the 2015 CMS dataset at a center-of-mass energy of 13 TeV. For the first time the tW channel has been included. The analysis focuses on possible variations in the coupling between the Higgs boson and the top quark or the W boson to search for deviations from the standard model. Exclusion limits are calculated as a function of the coupling strength factors κ_t and κ_V .

Two different signal regions are selected: one region with at least 4 jets, among them exactly 3 b-tagged jets; the other region with at least 5 jets, exactly 4 of which are b-tagged. An identification of jets in the final state under the tHq or $t\bar{t}$ hypothesis is performed using two separate sets of reconstruction BDTs. A consecutive set of classification BDTs uses variables derived in these assignments to enhance the signal-to-background ratio and discriminate against main backgrounds, including the predominant $t\bar{t}$ production. Limits are then calculated using the distributions of the responses of the classification BDTs.

The observed (expected) 95% confidence level upper limit for the standard model is $113.7 \times \sigma_{\text{SM}}$ ($98.6 \times \sigma_{\text{SM}}$) and the 95% confidence-level upper limit for the inverted top coupling scenario is 6.0 times the predicted cross section, with an expected upper limit of 6.4.

References

- [1] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *JHEP* **06** (2013) 081, doi:10.1007/JHEP06(2013)081.
- [2] 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** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [3] S. Biswas, E. Gabrielli, and B. Mele, “Single top and Higgs associated production as a probe of the $Ht\bar{t}$ coupling sign at the LHC”, *JHEP* **01** (2013) 088, doi:10.1007/JHEP01(2013)088, arXiv:1211.0499.
- [4] B. Hespel, F. Maltoni, and E. Vryonidou, “Higgs and Z boson associated production via gluon fusion in the SM and the 2HDM”, *JHEP* **06** (2015) 065, doi:10.1007/JHEP06(2015)065, arXiv:1503.01656.
- [5] “Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC”, *Phys. Lett. B* **726** (2013), no. 1.3, 88, doi:htp://dx.doi.org/10.1016/j.physletb.2013.08.010.
- [6] CMS Collaboration, “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV”, *Eur. Phys. J. C* **75** (2015), no. 5, 212, doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.
- [7] ATLAS, CMS Collaboration, “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 $\sqrt{s} = 7$ and 8 TeV”, arXiv:1606.02266. Submitted to *JHEP*.
- [8] J. Ellis and T. You, “Updated Global Analysis of Higgs Couplings”, *JHEP* **06** (2013) 103, doi:10.1007/JHEP06(2013)103, arXiv:1303.3879.

- [9] G. Bordes and B. van Eijk, “On the associate production of a neutral intermediate mass Higgs boson with a single top quark at the LHC and SSC”, *Phys. Lett. B* **299** (1993) 315, doi:10.1016/0370-2693(93)90266-K.
- [10] T. M. P. Tait and C. P. Yuan, “Single top quark production as a window to physics beyond the standard model”, *Phys. Rev. D* **63** (2000) 014018, doi:10.1103/PhysRevD.63.014018, arXiv:hep-ph/0007298.
- [11] M. Farina et al., “Lifting degeneracies in Higgs couplings using single top production in association with a Higgs boson”, *JHEP* **1305** (2013) 022, doi:10.1007/JHEP05(2013)022, arXiv:1211.3736.
- [12] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, “Higgs production in association with a single top quark at the LHC”, *Eur. Phys. J. C* **75** (2015), no. 6, 267, doi:10.1140/epjc/s10052-015-3475-9.
- [13] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [14] F. Demartin et al., “tWH associated production at the LHC”, arXiv:1607.05862.
- [15] CMS Collaboration, “Search for $H \rightarrow b\bar{b}$ in association with single top quarks as a test of Higgs couplings”, CMS Physics Analysis Summary CMS-PAS-HIG-14-015, CERN, Geneva, 2014.
- [16] CMS Collaboration, “CMS Luminosity Measurement for the 2015 Data Taking Period”, CMS Physics Analysis Summary CMS-PAS-LUM-15-001, CERN, Geneva, 2016.
- [17] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [18] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [19] P. Nason, “A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms”, *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.
- [20] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.
- [21] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.
- [22] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [23] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1”, *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

- [24] GEANT4 Collaboration, “GEANT4: A Simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [25] J. Allison et al., “Geant4 developments and applications”, *IEEE Transactions on Nuclear Science* **53** (Feb, 2006) 270, doi:10.1109/TNS.2006.869826.
- [26] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T ”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, Geneva, 2009.
- [27] CMS Collaboration, “Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV”, CMS Physics Analysis Summary CMS-PAS-PFT-10-003, CERN, Geneva, 2010.
- [28] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [29] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, arXiv:1607.03663. Submitted to *JINST*.
- [30] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2”, CMS Physics Analysis Summary CMS-PAS-BTV-15-001, CERN, Geneva, 2016.
- [31] CMS Collaboration, “Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. 02, P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.
- [32] A. Hoecker et al., “TMVA - Toolkit for Multivariate Data Analysis”, (2007) arXiv:physics/0703039.
- [33] CMS Collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *JINST* **6** (2011), no. 11, P11002.
- [34] 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 Physics Analysis Summary CMS-PAS-HIG-16-004, CERN, Geneva, 2016.
- [35] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)].