

Search for Single Top + Higgs Production with CMS

Stieger B.* on behalf of the CMS Collaboration

University of Nebraska–Lincoln

E-mail: benjamin.stieger@cern.ch

We present a search for the production of a Higgs boson in association with a single top quark, based on data collected in 2016 by the CMS experiment at the LHC at a center-of-mass energy of 13 TeV, which corresponds to an integrated luminosity of 35.9 fb^{-1} . The production cross section for this process is highly sensitive to the absolute values of the top quark Yukawa coupling, y_t , the Higgs boson coupling to vector bosons, g_{HVV} , and, uniquely, to their relative sign. Analyses using multilepton signatures, targeting $H \rightarrow WW$, $H \rightarrow \tau\tau$, and $H \rightarrow ZZ$ decay modes, and signatures with a single lepton and a $b\bar{b}$ pair, targeting the $H \rightarrow b\bar{b}$ decay, are combined with a reinterpretation of a measurement in the $H \rightarrow \gamma\gamma$ channel to constrain y_t . For a standard model-like value of g_{HVV} , the data favor positive values of y_t and exclude values of y_t below about $-0.9 y_t^{\text{SM}}$.

*ICHEP 2018, International Conference on High Energy Physics
4–11 July 2018
Seoul, South Korea*

*Speaker.

1. Introduction

The production of a single top quark and a Higgs boson (*tH*) is one of only a few processes that is directly sensitive to the sign of the Yukawa coupling of the top quark, y_t , without relying on assumptions on particles running inside loops [1]. In particular, this sensitivity comes from the interference of the leading order Feynman diagrams where the Higgs boson couples either to the top quark (proportional to y_t), or to a W boson (proportional to g_{HVV}). In the standard model (SM) the interference is maximally destructive and the resulting production cross section is small—only about 70fb in pp collisions at 13TeV. A non-SM like sign or phase between the two couplings would affect this interference and could enhance the cross section by an order of magnitude or more. In such scenarios the event yield from *tH* production would be comparable to processes such as the associated production of top quark-anti quark pairs and a Higgs boson ($t\bar{t}H$). Thus a measurement of *tH*-enhanced final states allows to directly assess the likelihood of a non-SM like sign of y_t .

There are three production modes for *tH* in pp collisions: the dominant *t*-channel (*tHq*), *t*-W associated production (*tHW*), and the much smaller *s*-channel which is neglected in the following.

This note presents a combined analysis of three channels targeting different Higgs boson decay modes: one focusing on final states with multiple leptons; one selecting single leptons and a pair of bottom quarks; and a reinterpretation of an analysis of diphoton events. All are based on the 2016 pp collision data set at 13TeV, corresponding to 35.9fb^{-1} , taken by the CMS experiment[2].

2. Multilepton channels

Signal *tH* events with Higgs boson decays to pairs of vector bosons or τ leptons and leptonic top quark decays can produce final states with multiple charged, light leptons (e^\pm or μ^\pm) [3]. Three channels are taken into account: exactly two same-charge muons, one electron and one muon of the same charge, and exactly three leptons of either flavor and charge. Events are selected to furthermore contain at least one hadronic jet that is tagged to result from the hadronization of a b quark, and at least one other jet that is anti-tagged. In such signatures, two classes of backgrounds can be identified: reducible backgrounds from events with nonprompt leptons (either genuine leptons from heavy-flavor hadron decays, or asymmetric photon conversions, or jets misidentified as leptons), from processes with relatively large cross sections such as $t\bar{t}$ production; and irreducible backgrounds of prompt leptons from processes with smaller cross sections, such as $t\bar{t} + \text{vector boson}$ production. Reducible backgrounds are estimated from the data by employing alternate selections with looser lepton identification requirements and extrapolating to the signal selection, whereas irreducible backgrounds are estimated using simulated events, corrected for efficiency effects and for mismodeling of the data.

To effectively separate the *tHq* signal from those two background classes, two separate boosted decision tree (BDT) classifiers are trained with either simulated $t\bar{t}$ events as the background class, or with a proportional admixture of $t\bar{t} + W$ and $t\bar{t} + Z$. The training is done separately for dilepton and trilepton event selections with input features related to the forward jet activity, the number of jets and b-tagged jets, and the kinematic properties of the leptons. Events in each of the three channels are then categorized into ten bins according to the output of the two BDT classifiers, and

this one-dimensional distribution is used in a simultaneous fit to the data to constrain the tH signal contributions.

3. Single lepton + $b\bar{b}$ channels

Signal events where the Higgs bosons decays to a pair of bottom quarks ($b\bar{b}$) produce events with at least three central b jets and a hard lepton from the top quark decay. The analysis described here defines two signal-enriched channels, both with exactly one lepton (e^\pm or μ^\pm), a minimum amount of missing transverse momentum (p_T^{miss}), and either exactly three or exactly four b tagged jets and one additional, anti-tagged jet [3]. This selection is dominated by events from $t\bar{t}$ production with additional gluons from initial or final state radiation which subsequently produce additional b and c quarks. To control such backgrounds directly in the data, an additional selection is defined that requires exactly two opposite-charge leptons, a minimum amount of p_T^{miss} , at least two b -tagged jets, and at least one additional jet tagged with a looser requirement. These background classes are constrained in the dilepton control selection in a simultaneous fit using a dedicated BDT classifier trained to separate $t\bar{t}$ + heavy flavor from $t\bar{t}$ + light flavor.

The tHq signal is separated from the backgrounds in two steps. First, a BDT classifier trained to pick the best jet-parton assignment in each event is used to calculate event characteristics. These are then used as input to a second BDT, trained to separate the tHq signal from the backgrounds.

4. $H \rightarrow \gamma\gamma$ reinterpretation

Signal tH events with $H \rightarrow \gamma\gamma$ are used in a dedicated measurement of Higgs boson properties in the diphoton channel [4]. They contribute predominantly to two categories targeting $t\bar{t}H$ production. The tHq and tHW inputs to Ref. [4] are corrected for the modified acceptance and efficiency of the various coupling scenarios under study before including them in the combined analysis described in this note.

5. Results

To derive constraints on the signal yields for a range of assumptions on the couplings and their implications on signal kinematics and normalization, a maximum likelihood (ML) fit is performed on the distributions of the various classifier outputs in the multilepton and single lepton- $b\bar{b}$ channels, and on the diphoton mass distribution in the $H \rightarrow \gamma\gamma$ channel. The different coupling scenarios are encoded in a modifier of the top quark Yukawa coupling y_t , defined as $\kappa_t \equiv y_t/y_t^{\text{SM}}$. For each value of κ_t , the affected tHq and tHW kinematics are taken into account when calculating acceptance, efficiency, and the classifier output shapes used in the fit. Furthermore, the Higgs branching fractions to $\gamma\gamma$, $Z\gamma$, and gluon-gluon are scaled accordingly with κ_t . The g_{HVV} coupling is assumed to be unmodified.

A scan of the profile likelihood ratio $\mathcal{L}(\kappa_t)/\mathcal{L}(\hat{\kappa}_t)$ is then performed, where $\hat{\kappa}_t$ is the best-fit value of κ_t , with the systematic uncertainties included in the form of nuisance parameters. The result is shown in Fig. 1, left, for scans on the observed data and on Asimov data sets with SM

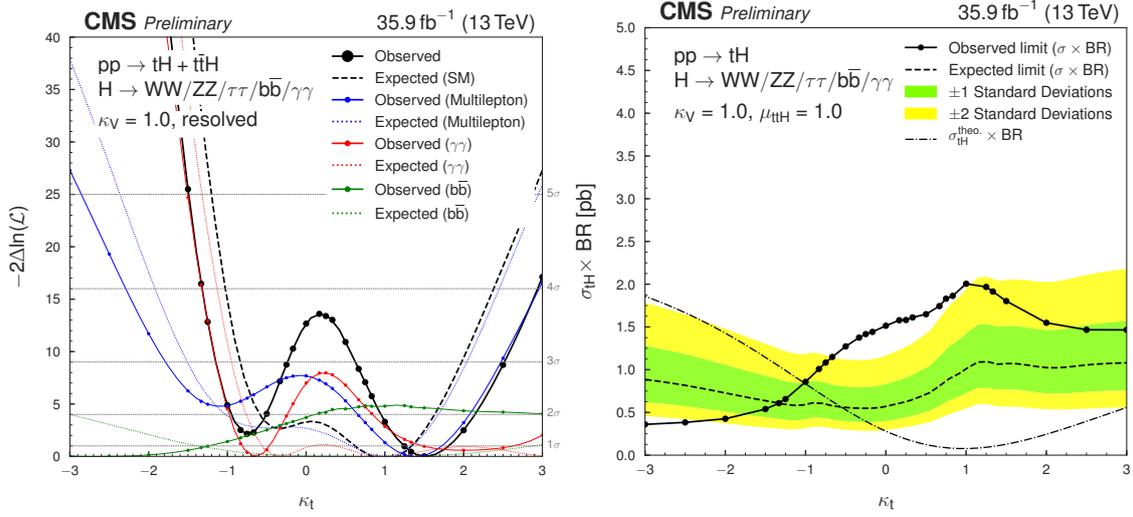


Figure 1: Left: scan of $-2\Delta\ln(\mathcal{L})$ versus κ_t for the data (black line) and the individual channels (blue, red, and green), compared to Asimov data sets corresponding to the SM expectations (dashed lines). Right: observed and expected 95% C.L. upper limit on the $t\bar{H}$ cross section times combined $H \rightarrow WW^* + \tau\tau + ZZ^* + b\bar{b} + \gamma\gamma$ branching fraction for different values of the coupling ratio κ_t . The expected limit is calculated on a background-only data set, i.e. without $t\bar{H}$ contribution, but including a κ_t -dependent contribution from $t\bar{t}\bar{H}$. The $t\bar{t}\bar{H}$ normalization is kept fixed in the fit, while the $t\bar{H}$ signal strength is allowed to float. From Ref. [3].

expectations for the $t\bar{H}$ and $t\bar{t}\bar{H}$ components, and separately for the different channels included in the combined analysis.

The data favor a positive value of κ_t over a negative value by about 1.5 standard deviations, where the expected performance in case of SM-like $t\bar{H}$ and $t\bar{t}\bar{H}$ signals would be to prefer a positive value by about four standard deviations. Values of κ_t outside of about $[-0.9, -0.5]$ and $[1.0, 2.1]$ are excluded at 95% C.L.

To set upper limits on $t\bar{H}$ production based on the observed data, a signal strength parameter is defined, modifying the $t\bar{H}q$ and $t\bar{H}W$ cross sections in the fit, while keeping $t\bar{t}\bar{H}$ fixed at its SM value. An ML fit is performed for this signal strength parameter at fixed points of κ_t , and upper limits on the signal strength are derived using the CL_s method. The resulting upper limits on the fiducial cross section of $t\bar{H}$ production are shown in Fig. 1, right. The observed and expected limits on SM-like $t\bar{H}$ production are at about 25 and 12 times the SM cross section, respectively. A simultaneous fit of a $t\bar{H}$ and a $t\bar{t}\bar{H}$ signal strength yields a result that is compatible with the SM within about two standard deviations.

References

- [1] F. Maltoni, K. Paul, T. Stelzer, S. Willebrock, PRD **64** 094023 (2001).
- [2] CMS Collaboration, JINST **3** (2008) S08004.
- [3] CMS Collaboration, CMS-HIG-18-009 (2018), [arXiv:1811.09696 [hep-ex]], subm. to PRD
- [4] CMS Collaboration, CMS-HIG-17-016 (2018), [arXiv:1804.02716 [hep-ex]], subm. to JHEP