Search for the standard model Higgs boson produced in association with top quarks at CMS

C. Botta on behalf of the CMS collaboration
CERN

The search for the standard model Higgs boson produced in association with a top-quark pair is presented, using up to 5 fb^{-1} and 19.6 fb^{-1} of 7 and 8 TeV pp collision data collected by the CMS experiment at the LHC. The search is performed in several channels characterised by different combinations of the decay products of the top-quark pair and the Higgs boson. The decay channels of the Higgs boson into $b\bar{b}$, $\tau\tau$, $\gamma\gamma$, ZZ^*, and WW^* are investigated. The signal strength μ, relative to the expectation for the standard model Higgs boson production is measured to be $\mu = 2.5^{+1.1}_{-1.0}$ for a Higgs boson mass of 125 GeV.

1 Introduction

Given the measured Higgs boson mass, the top quark Yukawa coupling cannot be assessed by measuring Higgs boson decays to top quarks. But if we assumed there is no new physics beyond the standard model, the Higgs boson's coupling to top quarks can be experimentally constrained through measurements involving the gluon fusion production mechanism or the Higgs boson decay to photons. These processes proceed via a fermion loop in which the top quark provides the dominant contribution. Resolving the loop-induced couplings to gluons and photons in terms of the tree-level couplings assuming only SM particles in the loop, current results from CMS provide an indirect measurement of the top Yukawa coupling consistent with the SM expectation and with a precision of the order of 25% [1].

The $t\bar{t}H$ interaction vertex however, is present in a rare production mechanism where the Higgs boson is produced in association with a top quark-antiquark pair. A measurement of the rate of this process provides a direct test of the coupling between the top quark and the Higgs boson. The comparison of the direct measurement of the $t\bar{t}H$ coupling with the one derived indirectly from the loop-induced couplings to gluons and photons is sensitive to new physics contributions in the loops.

The results of a search for $t\bar{t}H$ production at CMS [2] at the Large Hadron Collider (LHC) are described. The $t\bar{t}H$ production cross section is very small at the LHC, roughly 130 fb at $\sqrt{s} = 8$ TeV [3], but the signal has very specific characteristics that make the measurement possible. In addition to the decay products of the Higgs boson the events are very crowded due to the presence of the additional jets and leptons from the top-quark pair decay which is determined by the decay of the W bosons. If both W bosons decay hadronically the final state is characterised by the presence of six additional jets, two of which are b-jets. If one of the W bosons decays leptonically, the presence of a charged lepton, a neutrino, and four jets, two of which are b-jets, determines the experimental signature. Finally, when both W bosons decay leptonically, the final state has two charged leptons, two neutrinos and two b-jets. These three decay modes of the top-quark pair are referred to as all hadronic, lepton + jets, and dilepton channels respectively. These signatures are then further categorized according to the Higgs decay products. All the

possible decay channels that provide experimentally accessible signatures are considered in order to increase as much as possible the rate of selected signal events. Therefore the dominant decay mode for the Higgs boson, which is $\mathrm{H} \to \mathrm{b\bar{b}}$, contributing almost to 60% of the total Higgs decay width, and the next largest contribution, which comes from $\mathrm{H} \to \mathrm{WW}$ with a branching fraction around 20% are considered, as well as the the Higgs decay channels with significantly smaller branching fractions, such as $\mathrm{H} \to \gamma\gamma$, $\mathrm{H} \to \tau\tau$, and $\mathrm{H} \to \mathrm{ZZ}$.
The experimental searches for $\mathrm{t\bar{t}H}$ production presented here can be divided into three broad categories based on the Higgs boson decay modes: $\mathrm{H} \to$ hadrons, $\mathrm{H} \to$ photons, and $\mathrm{H} \to$ leptons. All of the searches make use of the full CMS 8 TeV dataset, with an integrated luminosity of 19.6 fb^{-1}. The $\mathrm{t\bar{t}H}$ search in the $\mathrm{H} \to \mathrm{b\bar{b}}$ final state using 5.0 fb^{-1} of 7 TeV data, described in Ref. [4], is also combined to determine the final $\mathrm{t\bar{t}H}$ result.

2 $\mathrm{H} \to$ Hadrons

2.1 Signatures

There are two main Higgs boson decay modes that contribute to the $\mathrm{H} \to$ hadrons searches: $\mathrm{H} \to \mathrm{b\bar{b}}$ and $\mathrm{H} \to \tau\tau$, where both τ leptons decay hadronically. Note that $\mathrm{H} \to \tau\tau$ includes both direct $\mathrm{H} \to \tau\tau$ decays and those where the τ leptons are produced by the decays of W or Z bosons from $\mathrm{H} \to \mathrm{WW}$ and $\mathrm{H} \to \mathrm{ZZ}$ decays. Events used in the $\mathrm{H} \to$ hadrons searches have one or more isolated charged leptons from the W boson decays from the top quarks, which means that these searches focus on the lepton + jets and dilepton $\mathrm{t\bar{t}}$ final states, using single lepton or dilepton triggers respectively. In particular three channels are investigated: the lepton + jets channel ($\mathrm{t\bar{t}} \to \ell\nu \mathrm{q\bar{q}}^{'}\mathrm{b\bar{b}}$, $\mathrm{H} \to \mathrm{b\bar{b}}$), the dilepton channel ($\mathrm{t\bar{t}} \to \ell^{+}\nu\ell^{-}\bar{\nu}\mathrm{b\bar{b}}$, $\mathrm{H} \to \mathrm{b\bar{b}}$), and the tau channel ($\mathrm{t\bar{t}} \to \ell\nu \mathrm{q\bar{q}}^{'}\mathrm{b\bar{b}}$, $\mathrm{H} \to \tau\tau$), where a lepton is an electron or muon. Examples of leading-order Feynman diagrams describing the three channels are shown in Fig. 1.
In the lepton + jets and tau channels, events are selected if they contain an isolated lepton (electron or muon) and at least four jets. In addition, in the lepton + jets channel two or more of these jets should be b-tagged, while in the tau channel, two of these jets must be consistent with hadronic decays of the τ lepton and one or two must be b-tagged. For the dilepton channel two opposite sign isolated leptons are required as well as three more jets, with at least two of them being b-tagged. The event selections are designed to be mutually exclusive.
In general the jet p_T cut is at 30 GeV, in the lepton + jets channel the leading 3 jets must have $p_T > 40$ GeV. Multivariate analysis (MVA) techniques are employed to tag the jets coming from b or τ decays. In the tau channel the τ must have $p_T > 20$ GeV. To match the trigger thresholds electrons and muons are required to have $p_T > 30$ GeV in the single lepton channels, and $p_T >$ 20, 10 GeV in the dilepton one.

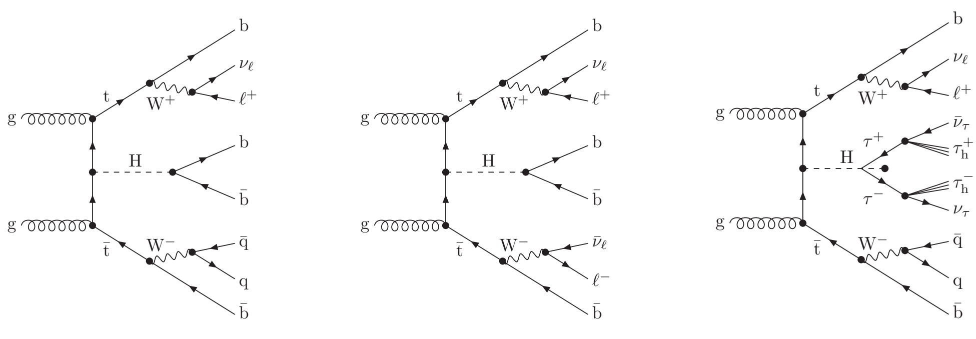

Figure 1 – Example of leading-order Feynman diagrams for $\mathrm{t\bar{t}H}$ production at pp colliders followed by top quarks and Higgs boson decays typical of the $\mathrm{H} \to$ Hadrons search.

2.2 Analysis strategy

First of all, to improve the sensitivity of the analysis, the events are separated into different categories based on the number of jets and b-tagged jets (b-tags). Signal $t\bar{t}H$ events are generally characterised by having more jets and more b-tags than the background processes. In the lepton + jets channel, the events are separated into the following categories: $\geq$6 jets + 2 b-tags, 4 jets + 3 b-tags, 5 jets + 3 b-tags, $\geq$6 jets + 3 b-tags, 4 jets + 4 b-tags, 5 jets + $\geq$4 b-tags, and $\geq$6 jets + $\geq$4 b-tags. For the dilepton channel, the events are separated into the categories 3 jets + 2 b-tags, $\geq$4 jets + 2 b-tags and $\geq$3 jets + $\geq$3 b-tags. For the tau channel, the following categories are used: 2, 3 or $\geq$4 jets with exactly 2 additional τ-tagged jets and 1 or 2 b-tags.
All the background rates and distributions in the H $\rightarrow$ hadrons analysis are extracted from MC simulations. The main background, $t\bar{t}$+jets, is generated inclusively, with tree-level diagrams for up to $t\bar{t}$+3 extra partons. These extra partons include both b and c quarks. To properly take into account the systematics related to this background, as there are significantly different uncertainties on the production of additional light-flavor jets compared to heavy-flavor, the $t\bar{t}$+jets sample is separated into subsamples based on the quark flavor associated with the reconstructed jets in the event, as it can be seen in the figures of this section.
After the event selection and the categorisation in numbers of jets and b-tags, MVA techniques are employed to separate $t\bar{t}H$ events from the large $t\bar{t}$+jets backgrounds. For the lepton + jets and dilepton channels, Boosted Decision Trees (BDTs) use information related to object kinematics, event shapes, and the output of the b-tagging discriminant. The choice of the input variables is optimized for each category. Ten variables are used as inputs to the final BDT in all lepton + jets categories. The dilepton channel uses 4 variables for the 3 jets + 2 b-tags category and 6 in the other categories. In the tau channel, the BDT makes use of 10 input variables related to τ isolation, τ kinematics, and event topology. For the lepton + jets and dilepton channels, a separate BDT is trained for each category, resulting in 7 and 3 BDTs respectively. For the tau channel, as the input variables do not change between the different categories, a single BDT is used.
The BDT outputs provide better discrimination between signal and background than any of the input variables individually. Their distributions are used to set limits on the Higgs boson production cross section. Two examples of BDT output distributions for the lepton + jets and the dilepton channels are provided in Fig. 2.
All the details of the analysis can be found here [5].

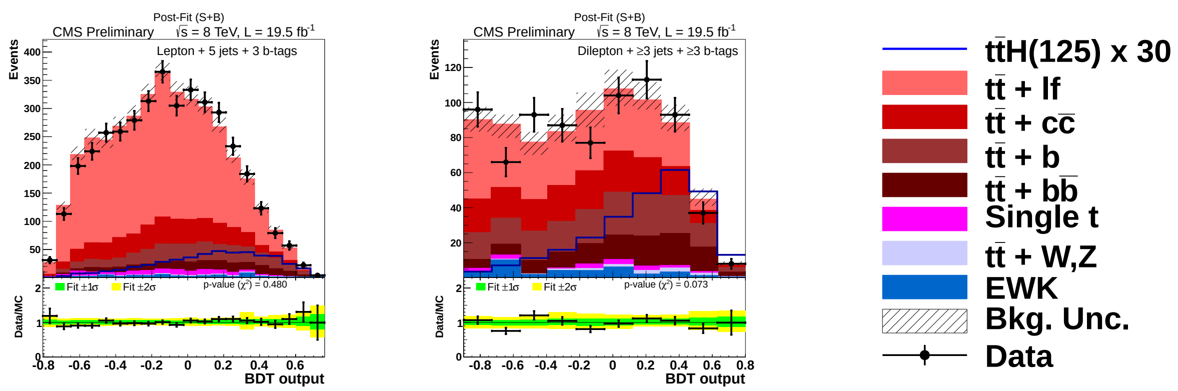

Figure 2 – Final BDT output for the lepton + jet category with 5 jets and 3 b-tag (left) and final BDT output for the dilepton category with $\geq$3 jets + $\geq$3 b-tags. Background-like events have a low BDT output value. Signal-like events have a high BDT output value. The background is normalized using the best fit value of all nuisance parameters and the uncertainty bands are constructed using the post-fit nuisance parameter uncertainties. The $t\bar{t}H$ signal (m_H = 125 GeV) is normalised to the best fit signal strength, μ, multiplied by a factor of 30.

3 H $\rightarrow$ **Photons**

3.1 Signatures

The diphoton analysis selects events by identifying a diphoton system consistent with a Higgs boson decay and by applying a loose selection on the remaining objects to tag the $\mathrm{t\bar{t}}$ decays and reject other Higgs boson production processes.
Two sets of event selection criteria are defined, which aim to search for hadronic ($\mathrm{t\bar{t}} \rightarrow \mathrm{q\bar{q}'q\bar{q}'b\bar{b}}$, $\mathrm{H} \rightarrow \gamma\gamma$) and leptonic ($\mathrm{t\bar{t}} \rightarrow \ell\nu\mathrm{q\bar{q}'b\bar{b}}/\mathrm{t\bar{t}} \rightarrow \ell^+\nu\ell^-\bar{\nu}\mathrm{b\bar{b}}$, $\mathrm{H} \rightarrow \gamma\gamma$) top quark pair decays in $\mathrm{t\bar{t}H}$ events. Example of leading-order Feynman diagrams describing the three channels are shown in Fig. 3.
The two channels share the same photon selection, which requires the leading photon to have transverse momentum greater than $m_{\gamma\gamma}/2$ GeV, and the subleading photon to have $p_T > 25$ GeV. The adoption of a variable threshold on the leading photon is aimed at increasing efficiency. The p_T threshold at 50 GeV, corresponding to the lower end of the mass spectrum, is still larger that the trigger threshold on the leading photon as the analysis makes use of the diphoton trigger with p_T thresholds at 36 and 22 GeV. In addition to this, both channels require the presence of at least one b-tag. The hadronic channel is defined by the requirement of at least four more jets in the event and no lepton, while the leptonic channel is defined by requiring at least one more jet in the event and at least one lepton with $p_T > 20$ GeV, where the lepton can be an electron or a muon.

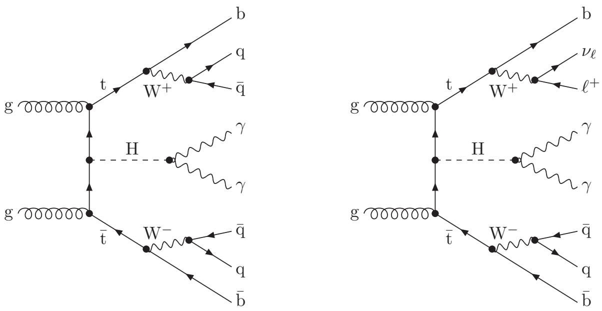

Figure 3 – Example of leading-order Feynman diagrams for $\mathrm{t\bar{t}H}$ production at pp colliders followed by top quarks and Higgs boson decays typical of the H $\rightarrow$ Photons search.

3.2 Analysis strategy

The analysis has a very low number of expected events but it has a distinctive signature with two energetic photons and a narrow Higgs boson peak over the falling background in the distribution of the diphoton invariant mass. The strategy adopted in this search is to fit for the amount of signal from the diphoton mass spectrum in the region surrounding the Higgs mass. The background is obtained by fitting the observed diphoton mass distributions in each channel (hadronic or leptonic) over the range 100 GeV $< m_{\gamma\gamma} <$ 180 GeV. The diphoton invariant mass spectra for data, the expected signal contribution, and the data-driven background are shown in Fig. 4. The data are fitted with a simple exponential for the leptonic channel, and a second order polynomial for the hadronic channel. The result of the fit is shown on the plots, together with the uncertainty bands corresponding to 68% and 95% probability. The expected contribution of a SM Higgs boson is also shown as a blue line.
All the details of the analysis can be found here [6].

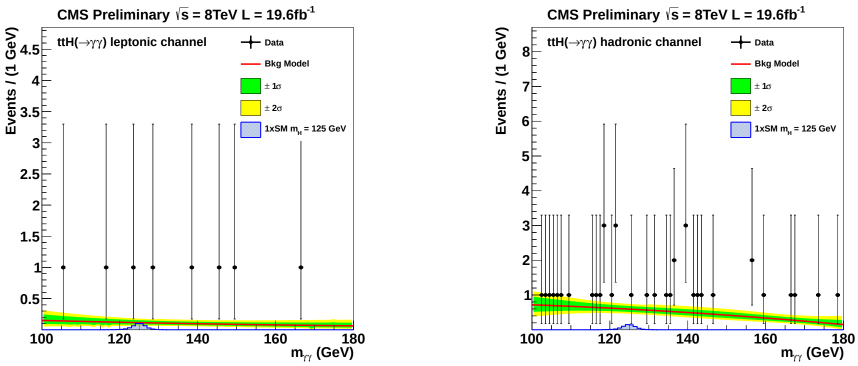

Figure 4 – Diphoton invariant mass distribution for candidate $t\bar{t}H$ events passing the leptonic selection (left plot) and the hadronic selection (right plot).

4 $H \to$ Leptons

4.1 Signatures

In this search, the production of $t\bar{t}H$ where the Higgs boson decays into ZZ^*, WW^*, or $\tau\tau$, with at least one Z, W or τ decaying leptonically is investigated. Despite the small branching ratio, the presence of one or two additional leptons from the top quark pair decays leads to the following clean experimental signatures: two same-sign leptons (electrons or muons) plus b-tagged jets, three leptons plus b-tagged jets, and four leptons plus b-tagged jets. Examples of Feynman diagrams for $t\bar{t}H$, followed by the decays of the top quark and the Higgs boson that lead to the signatures described above are shown in Fig. 5.
In the two-same-sign-lepton channel the events are selected requiring leptons with $p_T > 20$ GeV and at least four jets with at least one b-tag. In the three- and four-lepton final states the leptons p_T thresholds are 20, 10 GeV for the two leading leptons and 7 (5) GeV for the additional electrons (muons), and the events are required to have at least two jets and one b-tag. In both the three-lepton and four-lepton channels the events are rejected if there is a least one opposite-sign same-flavour lepton pair with invariant mass compatible with a Z candidate, in order to reduce $t\bar{t}Z$, ZZ, and events where the Higgs bosons is decaying into $ZZ \to 4\ell$. The jet p_T cut is 25 GeV for all the final states.

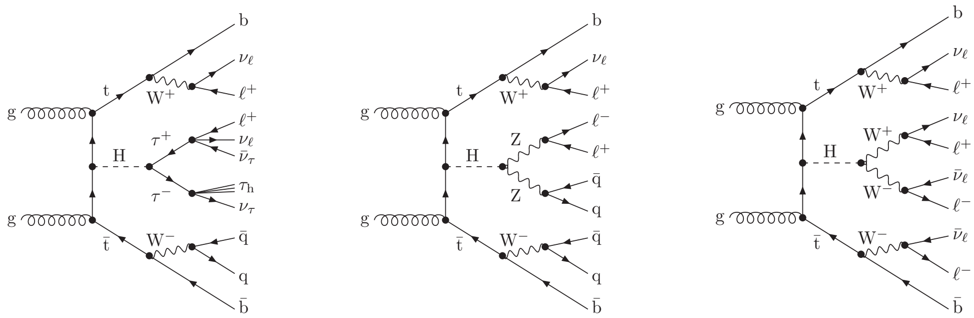

Figure 5 – Examples of leading-order Feynman diagrams for $t\bar{t}H$ production at pp colliders, followed by Higgs boson decays to $\tau\tau$, ZZ^*, or WW^* (from left to right). The first, second, and third diagrams are examples of the two same-sign lepton signature, the three lepton signature, and the four lepton signature, respectively.

4.2 Analysis strategy

The main backgrounds for this search can be subdivided into three categories: the $t\bar{t}V$ process from the associated production of a $t\bar{t}$ pair and one or more electroweak bosons; electroweak diboson or multiboson production associated with multiple hadronic jets; and reducible backgrounds from events with non-prompt leptons misidentified as prompt ones, or opposite-sign dilepton events in which the charge of one of the leptons is mismeasured.

Given the low number of expected events with the current statistic, in designing the analysis strategy the main effort was put on suppressing and controlling the reducible background, rather than to develop a dedicated selection to separate the irreducible $t\bar{t}V$ process from the signal. The reducible background contributes up to two thirds of the total background and is mainly composed by $t\bar{t}$ events in all the three channels.

A dedicated lepton identification algorithm was developed to enhance the separation power between signal leptons, meaning leptons from Z, W, and τ decays, from background leptons arising from b-hadrons decays, the dominant source of reducible background. The algorithm relies both on traditional isolation and impact parameter variables and on additional information from jet and b-tagging reconstruction applied to particles in the vicinity of the lepton, combined using multivariate methods. Correspondingly, a data-driven approach was developed to estimate the reducible background and its uncertainty, using multiple control regions and sophisticated methods to address the dependency of the lepton fake rate on the origin of the lepton, and the contamination from prompt leptons in the sample where the fake rate is measured.

The irreducible background is modelled using MC simulations.

After the event selection, to further improve the sensitivity of the analysis multivariate techniques are used to find a suitable discriminating variable to fit for the amount of signal and background from its distribution. In the dilepton and trilepton analysis, a BDT output is used as discriminating variable. In the first case the BDT is trained using simulated $t\bar{t}H$ signal and $t\bar{t}$ background events, with six discriminating variables. In the trilepton analysis the BDT is trained with simulated $t\bar{t}H$ signal and a mix of $t\bar{t}$, $t\bar{t}Z$, and $t\bar{t}W$ background events, with seven discriminating variables. These variables are related to the leptons kinematics, the hadronic activity of the event, and the missing transverse energy. The gain in the precision of the signal strength measurement from the multivariate analysis compared to the analysis which simply uses the distribution of the number of jets as discriminating variable is about 10%. In the four lepton analysis, only the number of hadronic jets is used: the sensitivity of this channel is anyway limited by the very small branching ratio, and the estimation of the kinematic distributions of the reducible backgrounds from data is also challenging due to the low event yields.

In the dilepton and trilepton final states, the events are further categorized with respect to the sum of the electrical charge of the leptons. This is done in order to exploit the charge asymmetry present in several SM background cross sections in pp collisions ($t\bar{t}W$, WZ, single top quark t channel, W+jets). The gain in the precision of the signal strength measurement from this categorization is approximately 5%.

The expected and observed distributions for the BDT output, for the different final states of the dilepton and trilepton analysis, are shown in Fig. 6. The distribution of the number of selected jets is also shown for the four-lepton channel.

All the details of the analysis can be found here [7].

5 Results

The statistical method used to extract the results is the same as the one used for other CMS Higgs analyses [8]. A binned likelihood spanning all analysis channels included in a given result is constructed. To quantify the amount of signal the strength parameter μ, which is defined as the ratio of the observed cross section for $t\bar{t}H$ production to the SM expectation, is used.

The results are reported in terms of the best fit value for μ and its associated uncertainty both independently for each of the distinct $t\bar{t}H$ signatures ($b\bar{b}$, $\tau\tau$, $\gamma\gamma$, same-sign $2l$, $3l$, and $4l$) and for the combination over all channels.

The best fit signal strengths are given in Fig. 7. The internal consistency of the six results

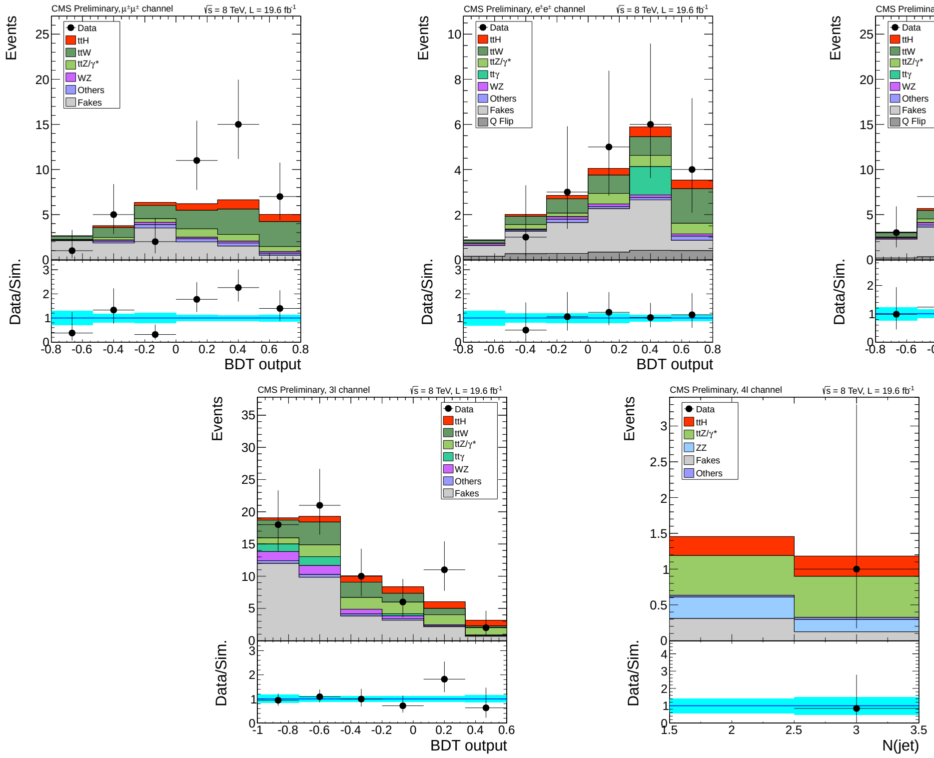

Figure 6 – Distribution of the BDT discriminant for the same-sign dilepton search, for the final states $\mu\mu$ (left), ee (center), and $e\mu$ (right). In these plots events with positive and negative charge are merged. The signal yield is the amount predicted by the standard model, ($\mu = 1$). The background yields are from the combined fit to the final discriminant at fixed $\mu = 1$. The bottom panel of each plot shows the ratio between the observed events and the expectation from simulation, with statistical and systematical uncertainties on the expectations after the fit.

with a common signal strength has been evaluated to be 22%, estimated from the asymptotic behavior of the profile likelihood function [8]. From the combination of the channels, the best fit value of the common signal strength is $\mu = 2.5^{+1.1}_{-1.0}$ (68% CL). Although the fit result shows an excess, within uncertainties, the result is consistent with SM expectations. The p-value under the SM hypothesis ($\mu = 1$) is 5.5%. The p-value for the background-only hypothesis ($\mu = 0$) is 0.3%, corresponding to a combined local significance of 2.7 standard deviations. Assuming SM Higgs boson production with $m_{\mathrm{H}} = 125$ GeV, the expected local significance is 1.2 standard deviations.

The observed signal strength is larger than the the expectation from the SM by about 2σ. The channel that drives the excess is the same-sign two-lepton channel as shown in Fig. 7. Within that channel, the same-sign two-muon subsample has the largest measured signal strength of $\mu = 8.4^{+3.3}_{-2.7}$, to be compared with $\mu = 2.8^{+4.6}_{-4.1}$ for the same-sign two-electron channel and $\mu = 1.9^{+2.5}_{-2.3}$ for the same-sign electron-muon channel. Many studies of the excess of data events in the same-sign dimuon channel have been carried out and the final outcome of the investigations has been that the excess is consistent with a statistical fluctuation of the background or with a genuine signal-like excess. More details about this can be found here [7].

References

1. CMS Collaboration, Measurements of the properties of the new boson with a mass near 125 GeV, CMS Physics Analysis Summary HIG-13-005 (2013).
2. CMS Collaboration, The CMS experiment at the CERN LHC, JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
3. LHC Higgs Cross Section Working Group et al., Handbook of LHC Higgs Cross Sections:

ttH Channel	$\mu = \sigma/\sigma_{SM}$ ($m_H = 125.7$ GeV)
$\gamma\gamma$	$-0.2^{+2.4}_{-1.9}$
$b\bar{b}$	$+1.0^{+1.9}_{-2.0}$
$\tau\tau$	$-1.4^{+6.3}_{-5.5}$
4l	$-4.8^{+5.0}_{-1.2}$
3l	$+2.7^{+2.2}_{-1.8}$
Same-sign 2l	$+5.3^{+2.2}_{-1.8}$
Combined	$+2.5^{+1.1}_{-1.0}$

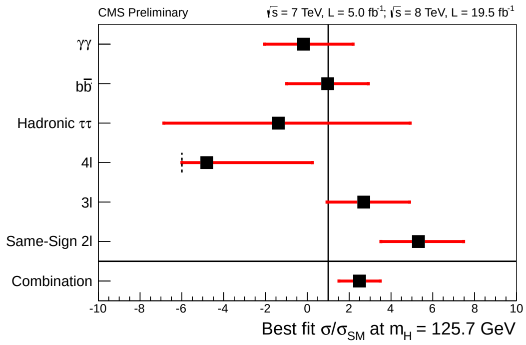

Figure 7 – The best-fit values of the signal strength parameter $\mu = \sigma/\sigma_{SM}$ for each $\mathrm{t\bar{t}H}$ channel at $m_{\mathrm{H}} = 125$ GeV. The signal strength in the four-lepton final state is not allowed to be below approximately -6 by the requirement that the expected signal-plus-background event yield must not be negative in either of the two jet multiplicity bins.

1. Inclusive Observables, CERN-2011-002 (CERN, Geneva, 2011) arXiv:1101.0593.

4. CMS Collaboration, Search for the standard model Higgs boson produced in association with a top-quark pair in pp collisions at the LHC, JHEP 1305 (2013) 145, doi:10.1007/JHEP05(2013)145, arXiv:1303.0763.
5. CMS Collaboration, Search for Higgs Boson Production in Association with a Top-Quark Pair and Decaying to Bottom Quarks or Tau Leptons, CMS Physics Analysis Summary HIG-13-019 (2013).
6. CMS Collaboration, Search for $\mathrm{t\bar{t}H}$ production in events with $\mathrm{H} \to \gamma\gamma$ at $\sqrt{s} = 8$ TeV , CMS Physics Analysis Summary HIG-13-015 (2013).
7. CMS Collaboration, Search for the standard model Higgs boson produced in association with top quarks in multilepton final states, CMS Physics Analysis Summary HIG-13-020 (2013).
8. CMS Collaboration, Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV, JHEP 1306 (2013) 081, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.