CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2013/09/30

Search for the standard model Higgs boson produced in association with top quarks in multilepton final states

The CMS Collaboration

Abstract

A search for the standard model Higgs boson produced in association with a top quark pair is presented, using 19.6 fb^{-1} of 8 TeV pp collision data collected by the CMS experiment at the LHC. Final states with a Higgs boson that decays to either ZZ^*,WW^*, or $\tau\tau$ are required to have a top quark pair that decays to either lepton plus jets ($t\bar{t} \to \ell\nu jjb\bar{b}$) or dileptons ($t\bar{t} \to \ell\nu\ell\nu b\bar{b}$), where ℓ represents an electron or a muon. The following signatures are selected: two isolated same-sign leptons (electrons or muons) plus b-tagged jets, three isolated leptons plus b-tagged jets, or four isolated leptons plus b-tagged jets. The expected 95% confidence level upper limit on the Higgs boson production cross section for a Higgs boson mass of 125.7 GeV/c^2 is 2.4 times the standard model expectation, to be compared to an observed limit of 6.6. The signal strength μ, relative to the expectation for the standard model Higgs boson, is measured to be $\mu = 3.7^{+1.6}_{-1.4}$.

1 Introduction

The newly-discovered Higgs boson [1, 2] provides a gateway for exploring physics beyond the standard model. The properties of the Higgs boson are uniquely predicted by the standard model for a given mass value. The CMS measurement of the new particle mass is $125.7 \pm 0.3\,(\text{stat.}) \pm 0.3\,(\text{syst.})$ GeV [3], which is accurate enough to create an opportunity to test the predictions of the standard model by measuring the other properties of the Higgs boson. So far, the analyses studying some of these properties have confirmed that this particle is compatible with the standard model Higgs boson [3]. These analyses were not yet exhaustive, which leaves ample room for new studies.

It is of particular interest to measure the coupling of the Higgs boson to the top quark ($t\bar{t}H$) because the top quark could play a special role in the context of electroweak symmetry breaking due to its large mass. The Higgs boson does not decay to top quarks. 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 as shown in Fig. 1. This process can be used to directly measure the top-Higgs coupling at tree level, with a cross section of 130 fb at next-to-leading order (NLO) [4] in 8 TeV pp collisions. At LHC energies the largest contribution to standard model Higgs boson production is a gluon-gluon induced loop dominated by virtual top exchange. The comparison of a direct measurement of the $t\bar{t}H$ coupling with the one inferred from the cross section measurement can put limits on the contribution of new physics to the gluon-gluon loop.

The first searches for $t\bar{t}H$ production were performed by the CDF and DØ experiments. The searches looked for the $b\bar{b}$ decays of the Higgs boson and set 95% confidence level (CL) upper limits on the $t\bar{t}H$ production cross section times $BR(H \to b\bar{b})$ at values in excess of 10 times the standard model prediction [5, 6]. More sensitive searches for $t\bar{t}H$ production in the same decay mode have been performed by ATLAS [7] and CMS [8]. The ATLAS analysis uses $5\,\text{fb}^{-1}$ of 7 TeV data. The CMS analysis uses 19.6 fb^{-1} of 8 TeV data and also includes the channel in which the Higgs boson decays to hadronic taus. The CMS analysis has an expected upper limit of 4.1 times the standard model $t\bar{t}H$ production cross section and observes a limit of 5.2. A search for $t\bar{t}H$ production in events where the Higgs boson decays to $\gamma\gamma$ was also performed by ATLAS [9] and CMS [10] with the 8 TeV data. The CMS results from the searches in $b\bar{b}$, $\tau\tau$, and $\gamma\gamma$ channels have been combined [8]. For a Higgs mass of 125 GeV, the combined expected limit is 2.7 times the SM expectation, while the observed limit is 3.4 times the SM expectation. The best-fit value for the combined signal strength is $0.74^{+1.34}_{-1.30}$ (68% CL).

In this note, 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 to decay leptonically is investigated for the first time. 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. 1. 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;
- four leptons plus b-tagged jets.

Multivariate analysis techniques are used to identify objects with high purity and to distinguish background from signal events. The amount of signal is fit to the multivariate discriminant output distribution in all the final states simultaneously.

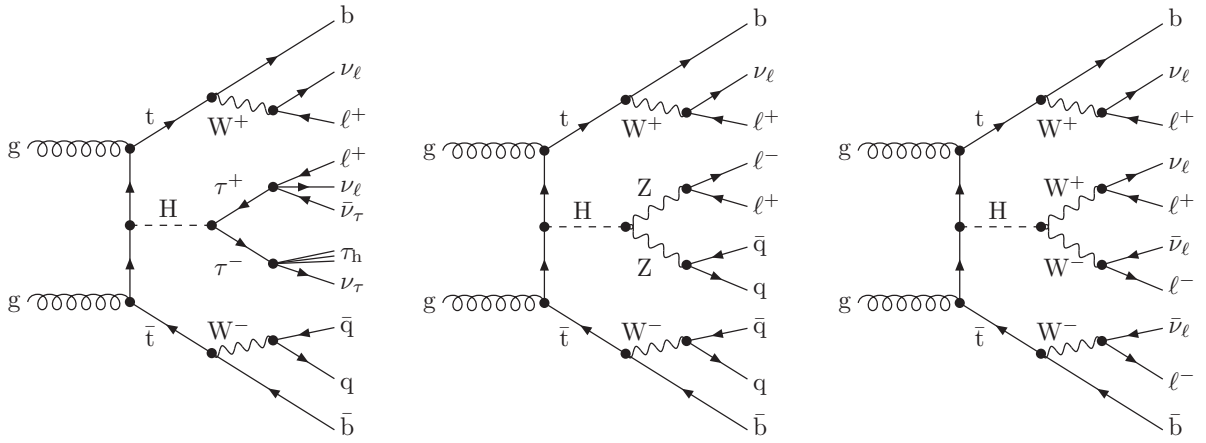

Figure 1: Examples of leading-order Feynman diagrams for $t\bar{t}H$ production at pp colliders, followed by Higgs boson decays to $\tau\tau$, ZZ^* and WW^* (from left to right). The first, second, and third diagrams are examples of the two same-sign lepton signature, the three lepton signature, and the four lepton signature, respectively.

2 The CMS detector

The CMS detector [11] consists of different components. A superconducting solenoid in the central region of the detector provides an axial magnetic field of 3.8 Tesla parallel to the beam direction. The silicon pixel and strip tracker, the crystal electromagnetic calorimeter (ECAL) and the brass/scintillator hadron calorimeter (HCAL) are located in concentric layers within the solenoid. These layers provide coverage out to $|\eta| < 2.5$, where pseudorapidity is defined as $\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$, and θ is the polar angle of the trajectory of the particle with respect to the beam direction. A quartz-fiber Cherenkov calorimeter (HF) extends the coverage to $|\eta| < 5.0$. Muons are detected by gas detectors embedded in the iron return yoke outside the solenoid. The first level of the CMS trigger system, composed of custom hardware processors, is designed to reduce the input rate by a factor of 1000 by selecting the most interesting events in less than 3 μs using information from the calorimeters and muon detectors. The High Level Trigger processor farm further decreases the event rate to a few hundred Hz for data storage. All of these components are used for the $t\bar{t}H$ search.

3 Data and simulation samples

We use the 2012 CMS dataset, corresponding to a total integrated luminosity of 19.6 fb^{-1}. The events are selected by the trigger requirement of the presence of either two leptons (electrons or muons), or a triplet of electrons. The minimal transverse momenta of the first and second lepton are 17 and 8 GeV for the double lepton triggers, and 15, 8, and 5 GeV for the triple electron trigger.

Simulated samples for the SM Higgs boson signal and for background processes are used to optimize the event selection and to evaluate the acceptance and systematic uncertainties. The $t\bar{t}H$ signal is modeled with the PYTHIA generator [12]. The background processes $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$+jets (which includes $t\bar{t}\gamma$+jets), Drell-Yan (DY) + jets (DY+γ+jets), W+jets (W+γ+jets), the diboson ZZ+jets, WW+jets, WZ+jets and the rare WWZ, WWW, and $t\bar{t}WW$ process are all simulated with the MADGRAPH [13] tree-level matrix element generator, combined with PYTHIA for the parton shower and hadronization. Single top production is modeled with the NLO generator POWHEG [14–19] combined with PYTHIA.

All events are processed through a detailed simulation of the CMS detector based on GEANT4 [20] and are reconstructed with the same algorithms that are used in data. The simulations include pileup interactions, the multiplicity of which matches the distribution observed with data. All events from data and simulated samples are required to pass the same trigger conditions.

4 Event reconstruction and objects identification

A global event description is obtained with the CMS particle-flow (PF) algorithm [21, 22], which optimally combines the information from all CMS sub-detectors to reconstruct and identify all individual particles in the collision event. The particles are classified into mutually exclusive categories: charged hadrons, neutral hadrons, photons, muons, and electrons.

The primary vertex is chosen as the vertex with the highest sum of $p_{\mathrm{T}}^2 = \sqrt{p_x^2 + p_y^2}$ of its constituent tracks. Jets are reconstructed by clustering PF particles with the anti-k_{T} algorithm with a distance parameter of 0.5, as implemented in the FASTJET package [23, 24]. The charged hadrons coming from the pileup interactions are subtracted from the PF particles considered in the clustering. Jet energy corrections are applied as a function of the jet p_{T} and η [25]. In addition, a multivariate discriminant is applied to distinguish between jets coming from the primary vertex and jets coming from pile-up vertices. The discrimination is based on the differences in the jet shapes, in the relative multiplicity of charged and neutral PF particles constituting the jet, and in the different fraction of transverse momentum which is carried by the leading particles. Jets are only considered if they have a transverse momentum above 25 GeV and $|\eta| < 2.4$. In addition, they have to be separated from any lepton candidates by requiring $\Delta\mathrm{R} = \sqrt{\Delta\phi^2 + \Delta\eta^2} > 0.5$.

A b-tagging algorithm [26] is used to identify jets that are likely to originate from the hadronization of b quarks. This algorithm combines both secondary vertex information and track impact parameter information in a likelihood discriminant called the Combined Secondary Vertex (CSV). The discriminating output value distinguishes between b jets and jets originating from light quarks, gluons and charm quarks. The efficiency to tag b jets and the rate of misidentification of non-b jets depend on the operating point chosen. Both the efficiency and the fake rate are parameterized as a function of the transverse momentum and pseudorapidity of the jets. These performance measurements are obtained directly from data in samples that are enriched in b jets, such as $\mathrm{t\bar{t}}$ and multijet events where a muon is found inside a jet. Two working points for the CSV output discriminant are used in the analysis. The *loose* working point has approximately 85% efficiency to tag jets from *b* quarks and 10% probability to tag jets from light quarks or gluons. The *medium* working point has approximately 70% efficiency for tagging jets from b quarks and 1.5% efficiency to tag jets from light quarks or gluons [26, 27]. Corrections that take into account the different performance of the CSV tagging in data and simulation are applied to the simulated samples used in the analysis.

The missing transverse energy vector is calculated as the negative of the vector sum of transverse momenta of all PF particles reconstructed in the event. The magnitude of this vector is referred to as $E_{\mathrm{T}}^{\mathrm{miss}}$. In order to recover from the degradation of performance of the missing transverse energy due to the pile-up interactions, the $H_{\mathrm{T}}^{\mathrm{miss}}$ is used. H_{T} is the scalar sum of p_{T} of all selected leptons and jets, and $H_{\mathrm{T}}^{\mathrm{miss}}$ is the magnitude of the negative vector sum of the transverse momenta of those objects. The $H_{\mathrm{T}}^{\mathrm{miss}}$ variable has worse resolution than the $E_{\mathrm{T}}^{\mathrm{miss}}$ but also a reduced dependency on the detector response to particles from pile-up and underlying event. In this analysis the event selection makes use of a linear discriminant, $E_{\mathrm{T}}^{\mathrm{miss}}LD$, that combines $E_{\mathrm{T}}^{\mathrm{miss}}$ and $H_{\mathrm{T}}^{\mathrm{miss}}$. It exploits the fact that $E_{\mathrm{T}}^{\mathrm{miss}}$ and $H_{\mathrm{T}}^{\mathrm{miss}}$ are less correlated in

events with instrumental missing transverse energy than in events with real missing transverse energy. A requirement of $E_{\mathrm{T}}^{\mathrm{miss}} LD > 0.2$ has a similar signal efficiency to a requirement of $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV, but rejects about a factor of two more Drell-Yan background.

The electrons are reconstructed within the geometrical acceptance of the tracker, $|\eta^{\mathrm{e}}| < 2.5$, and for transverse momentum $p_{\mathrm{T}}^{\mathrm{e}} > 7\,\mathrm{GeV}$. The reconstruction combines the information from clusters of energy deposits in the ECAL and the trajectory in the inner tracker [28–31]. The track-cluster matching is initiated either "outside-in" from energy cluster measurements, or "inside-out" from track reconstruction. Trajectories in the tracker volume are reconstructed using a dedicated modeling of the electron energy loss and fitted with a Gaussian sum filter [28]. The electron momentum is determined from the combination of ECAL and tracker measurements. Electron identification relies on a multivariate technique that combines observables sensitive to the amount of bremsstrahlung along the electron trajectory, the geometrical and momentum matching between the electron trajectory and associated clusters, and shower-shape observables.

Muons are reconstructed within $|\eta^{\mu}| < 2.4$ and for $p_{\mathrm{T}}^{\mu} > 5\,\mathrm{GeV}$ [32]. The reconstruction combines the information from both the silicon tracker and the muon spectrometer. The matching between the inner and outer tracks is initiated either "outside-in", starting from a track in the muon system, or "inside-out", starting from a track in the silicon tracker. The PF muons are selected among the reconstructed muon track candidates by applying minimal requirements on the track components in the muon and tracker system and taking into account matching with energy deposits in the calorimeters [33].

Corrections accounting for residual differences between data and simulation are applied to the muon momentum as well as on the ECAL energy before combining with the tracking momentum for electrons.

The isolation of individual electrons and muons is measured relative to their transverse momentum p_{T}^{ℓ}, by summing over charged and neutral particles in a cone $\Delta R < 0.4$ around the lepton direction at the interaction vertex:

$$R_{\mathrm{Iso}}^{\ell} \equiv \left(\sum_{\mathrm{charged}} p_{\mathrm{T}} + \mathrm{MAX}\left[0, \sum_{\mathrm{neutral}} p_{\mathrm{T}} + \sum_{\gamma} p_{\mathrm{T}} - 0.5 \sum_{\mathrm{charged,PU}} p_{\mathrm{T}} \right] \right) / p_{\mathrm{T}}^{\ell}, \tag{1}$$

where $\sum_{\mathrm{charged}} p_{\mathrm{T}}$, $\sum_{\mathrm{neutral}} p_{\mathrm{T}}$, and $\sum_{\gamma} p_{\mathrm{T}}$ are respectively the scalar sums of the transverse momenta of charged hadrons from the primary vertex, neutral hadrons, and photons located in the lepton cone. The contribution of pileup photons and neutral hadrons is estimated from the scalar sum of the transverse momenta of charged hadrons from pileup vertices in the cone, $\sum_{\mathrm{charged,PU}} p_{\mathrm{T}}$. This quantity is multiplied by a factor of 0.5, which corresponds approximately to the ratio of the energy of all neutral particles to the energy of charged hadrons in the hadronization process of pileup interactions, as estimated from simulation. The electrons or muons are considered isolated if $R_{\mathrm{Iso}}^{\ell} < 0.4$.

The efficiencies for the product of reconstruction, identification, and isolation of primary electrons and muons are measured in data, using a tag-and-probe technique [34] based on an inclusive sample of $\mathrm{Z} \to \ell^{+}\ell^{-}$ events. The measurements are performed in several bins of p_{T}^{ℓ} and $|\eta^{\ell}|$. The efficiencies to select electrons in the ECAL barrel (endcaps) varies from about 70% (60%) for $7 < p_{\mathrm{T}}^{\mathrm{e}} < 10\,\mathrm{GeV}$ to 85% (77%) at $p_{\mathrm{T}}^{\mathrm{e}} \simeq 10\,\mathrm{GeV}$, and reaches 95% (89%) for $p_{\mathrm{T}}^{\mathrm{e}} \geq 20\,\mathrm{GeV}$. It is about 85% in the transition region, $1.44 < |\eta| < 1.57$, from the ECAL barrel and endcaps, when averaged over the whole p_{T} range. The muons are selected with efficiencies that varies from about 85% for $5 < p_{\mathrm{T}}^{\mu} < 10\,\mathrm{GeV}$ to 90% at $p_{\mathrm{T}}^{\mu} \simeq 10\,\mathrm{GeV}$, and reaches 98%

for $p_T^{\mu} \geq 20\,\text{GeV}$. The same tag-and-probe technique is used to measure the efficiency of electrons and muons in simulated $Z \to \ell^+\ell^-$ events. Correction factors are used to account for the differences in the performance of the lepton identification in data and simulation.

Throughout this note, the lepton selection described above is called *lepton preselection*. Additional handles are used to tighten the lepton identification in order to suppress further the 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, while preserving high efficiency for signal leptons from W, Z, and τ decays.

In this analysis the most important source of misidentified leptons comes from the decay of b hadrons. A multivariate discriminant based on boosted decision tree (BDT) techniques was therefore tuned to discriminate signal leptons (from W, Z, or τ decays) from background leptons. This *lepton MVA discriminant* is trained with simulated $t\bar{t}H$ signal events and $t\bar{t}$+jets background events, separately for two p_T bins and three (two) η bins for electrons (muons). The input variables can be categorized into three groups: variables related to the impact parameter of the lepton computed with respect to the primary vertex, variables related to the isolation of the lepton considering separately the neutral and charged PF particle deposits in the lepton isolation cone, and variables related to the jet reconstructed in the event closest to the lepton. For this last set of variables the PF jets reconstructed around the leptons are used if ΔR is less than 0.5; charged hadrons from pile-up vertices are not removed prior to the jet clustering. The three discriminating variables relying on jets are the ΔR distance between the lepton and the closest jet, the ratio between the p_T of the lepton and the p_T of the jet, and the CSV b-tagging discriminant value of the jet. In the case of electrons, the multivariate discriminant used for the preselection and the number of missing hits in the innermost tracker layer are also used as inputs, to suppress backgrounds from charged hadrons misidentified as electrons or from photon conversions.

The agreement between data and simulation of the input variables and the final lepton multivariate discriminant is validated in dedicated control regions. For signal leptons, high-purity control samples are selected using Z boson events. High-purity muons and electrons are selected by requiring same-flavor, opposite-sign pairs of leptons with an invariant mass close to the Z boson mass in events with low amounts of missing transverse energy. In these events, tight isolation and p_T cuts are applied to the leading lepton, and the trailing lepton is used to check the agreement between simulation and data. High-purity τ leptons are selected by requiring opposite-flavor, opposite-sign pairs of electrons and muons with an invariant mass between 20 and 80 GeV/c^2. In these events, tight isolation, p_T, and impact parameter significance cuts are applied to one of the two leptons, and the other lepton is used to compare simulation and data. For background leptons, samples enriched in leptons from b-hadrons decay are selected with $Z + \ell$ and $t\bar{t} + \ell$ control regions. The agreement is good; small corrections to better match the data distributions of the input variables are applied to the simulation before training the MVA discriminant. Additional data-to-simulation scale factors are used account for the remaining differences in the performance of the discriminant. The scale factors are computed for signal leptons with the same tag-and-probe technique that was used to calculate the scale factors for the lepton preselection. Backgrounds with misidentified leptons are estimated directly from data, as described in Section 8.3.

Two working points for the lepton multivariate discriminant are defined: a tight working point, used for the search in the dilepton and trilepton final states, and a loose working point used for the four-lepton final state. The efficiencies for these working points are measured with respect to leptons passing the preselection. For the tight working point, the efficiency to select signal

muons in $|\eta^{\mu}| < 1.5$ ($1.5 < |\eta^{\mu}| < 2.4$) region is of the order of 60% (40%) for $p_{\mathrm{T}}^{\mu} \sim 10\,\mathrm{GeV}$ and reaches a plateau of 98% (95%) at $p_{\mathrm{T}}^{\mu} \sim 45\,\mathrm{GeV}$; for signal electrons in ECAL barrel (endcap), it is of the order of 40% (20%) for $p_{\mathrm{T}}^{e} \sim 10\,\mathrm{GeV}$ and reaches a plateau of 90% (70%) at $p_{\mathrm{T}}^{e} \sim 45\,\mathrm{GeV}$. The efficiency to select muons (electrons) from b-hadrons decays is less than 5% (between 5-10%) when averaged over the whole p_{T} and η ranges.

When using the tight working point, to further suppress background events with leptons arising from photon conversions, electrons either with missing hits in the innermost layer or associated with a successfully reconstructed conversion vertex [35], are rejected.

Finally, in the dilepton final state, additional requirements on the quality of the charge assignment are applied to suppress opposite-sign events in which the charge of one of the leptons is mismeasured. For the electrons, a consistency between the independent measurements of the charge from the ECAL energy deposit and the tracker is required. The ECAL energy deposit charge is estimated by comparing the position of the energy deposit to the extrapolated track trajectory. For the muons, the track transverse momentum is required to be well measured ($\Delta p_T / p_T < 0.2$).

5 Event selection

Candidate events that match the signal signatures are selected by requiring combinations of reconstructed objects. Three features are common to all three decay signatures:

- Each event is required to have one lepton with transverse momentum larger than 20 GeV/c and another lepton with transverse momentum larger than 10 GeV/c in order to satisfy the double-lepton trigger requirements.
- Since the simulation does not model accurately lepton pairs with low invariant mass, any event that has a pair of leptons with an invariant mass less than 12 GeV/c^2 is rejected. This applies to any combination of leptons regardless of flavor or charge.
- Since signal events have two top quarks, each event is required to have at least two jets, among which either two jets satisfy the loose CSV working point or one jet satisfies the medium CSV working point.

In addition, pairs of leptons with the same flavor whose invariant mass is consistent with the mass of the Z boson within 10 GeV/c^2 are used to reject background events with a Z boson decay. Same sign dilepton events are rejected if they contain any such pair. Events in the 3ℓ and 4ℓ categories are rejected only if the two leptons in the pair have opposite sign.

Same-sign dilepton events are required to have exactly two leptons with identical charges and at least four hadronic jets. Each lepton must pass the lepton preselection, the tight working point of the lepton MVA discriminant, and the charge-quality requirements. To reject events from backgrounds with a Z boson, $E_{\mathrm{T}}^{\mathrm{miss}} LD > 0.2$, is requested. To further suppress reducible backgrounds, especially non-$\mathrm{t\bar{t}}$ backgrounds, the threshold on the p_{T} of the second lepton is raised to 20 GeV, and the scalar sum of the transverse momenta of the two leptons and of the $E_{\mathrm{T}}^{\mathrm{miss}}$ is required to be above 100 GeV.

The three-lepton candidate selection requires exactly three leptons that pass the lepton preselection and the tight working point for the lepton MVA discriminant. To further reject events from backgrounds with a Z boson, a $E_{\mathrm{T}}^{\mathrm{miss}} LD$ requirement is applied, with a tighter threshold if the event has a pair of leptons with same flavour and opposite sign. For events with large jet multiplicity (≥ 4 jets), where the contamination from the Z background is smaller, the

requirement on the $E_{\mathrm{T}}^{\mathrm{miss}} LD$ is not applied.

The four-lepton candidate selection requires exactly four leptons that each pass the lepton preselection and the loose working point of the lepton MVA discriminant.

In both the three-lepton and four-lepton selections, the veto of same-flavor opposite-sign lepton pairs near the Z mass introduces an inefficiency for the $\mathrm{t\bar{t}H}$, $\mathrm{H} \to \mathrm{ZZ}^*$ with $Z \to \ell\ell$ events, but these events represent a small fraction of the expected signal.

The observed event yields in data for each final state and the expectations from the different physical processes are summarized in Table 1. The details of the calculations of the signal and background yields are discussed in the next sections.

	$\mu\mu$	ee	eμ	3ℓ	4ℓ
$\mathrm{t\bar{t}H}$, $\mathrm{H} \to \mathrm{WW}$	2.0 ± 0.3	0.9 ± 0.1	2.7 ± 0.4	3.2 ± 0.6	0.28 ± 0.05
$\mathrm{t\bar{t}H}$, $\mathrm{H} \to \mathrm{ZZ}$	0.1 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.2 ± 0.0	0.09 ± 0.02
$\mathrm{t\bar{t}H}$, $\mathrm{H} \to \tau\tau$	0.6 ± 0.1	0.3 ± 0.0	0.9 ± 0.1	1.0 ± 0.2	0.15 ± 0.02
$\mathrm{t\bar{t}\,W}$	8.2 ± 1.5	3.4 ± 0.6	13.0 ± 2.2	9.2 ± 1.9	-
$\mathrm{t\bar{t}\,Z}/\gamma^*$	2.5 ± 0.5	1.6 ± 0.3	4.2 ± 0.9	7.9 ± 1.7	1.25 ± 0.88
$\mathrm{t\bar{t}\,WW}$	0.2 ± 0.0	0.1 ± 0.0	0.3 ± 0.1	0.4 ± 0.1	0.04 ± 0.02
$\mathrm{t\bar{t}}\,\gamma$	-	1.3 ± 0.3	1.9 ± 0.5	2.9 ± 0.8	-
WZ	0.8 ± 0.9	0.5 ± 0.5	1.2 ± 1.3	4.2 ± 0.9	-
ZZ	0.1 ± 0.1	0.0 ± 0.0	0.1 ± 0.1	0.4 ± 0.1	0.45 ± 0.09
rare SM bkg.	1.1 ± 0.0	0.4 ± 0.0	1.5 ± 0.0	0.8 ± 0.0	0.01 ± 0.00
non-prompt	10.8 ± 4.8	8.9 ± 4.5	21.2 ± 8.1	33.2 ± 12.3	0.53 ± 0.32
charge flip	-	1.9 ± 0.6	2.4 ± 0.8	-	-
all signals	2.7 ± 0.4	1.2 ± 0.2	3.7 ± 0.6	4.4 ± 0.8	0.52 ± 0.09
all backgrounds	23.7 ± 5.2	18.0 ± 4.7	45.9 ± 8.6	58.9 ± 12.7	2.28 ± 0.94
data	41	19	51	68	1

Table 1: Expected and observed yields after the selection in all five final states. The rare SM backgrounds include triboson production, tbZ, $\mathrm{W^{\pm}W^{\pm}qq}$, and WW produced in double-parton interactions. A '-' indicates a negligible yield. Non-prompt and charge-flip backgrounds are described in Sec. 8.

6 Signal extraction

After the event selection described in the previous section, the overall yields are still dominated by background. It is not optimal to infer the presence of a $\mathrm{t\bar{t}H}$ signal on the basis of the yields alone. The strategy adopted in this search is to fit for the amount of signal from the distribution of a suitable discriminating variable.

In the dilepton analysis, a boosted decision tree (BDT) is used as discriminating variable. The BDT is trained with simulated $\mathrm{t\bar{t}H}$ signal and $\mathrm{t\bar{t}}$ background events, with six discriminating variables: the p_{T} and $|\eta|$ of the trailing lepton, the minimal angular separation between the trailing lepton and the closest jet, the transverse mass of the leading lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$, H_{T}, $H_{\mathrm{T}}^{\mathrm{miss}}$. The same training is used for the ee, eμ and $\mu\mu$ final states, as the gain in performance from dedicated trainings in each final state is found to be negligible.

In the trilepton analysis, a BDT is also used for the final discrimination. The BDT is trained with simulated $\mathrm{t\bar{t}H}$ signal and a mix of $\mathrm{t\bar{t}}$, $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}Z}$ background events, with seven discriminating variables: the multiplicity of hadronic jets, the p_{T} of the jet with the highest b-tagging

discriminant value, H_T, the fraction of H_T from jets and leptons within $|\eta| < 1.2$, the maximum of the $|\eta|$ values of the three leptons, the minimal ΔR separation between any pair of opposite-sign leptons, and the mass of the best candidate hadronically-decaying top quark reconstructed from the jets in the event.

As a cross-check in both the dilepton and the trilepton final states, the multiplicity of hadronic jets was also used as a discriminating variable. The gain in the precision of the signal strength measurement from the multivariate analysis compared to this simpler cross-check is about 10%.

In the four lepton analysis, only the multiplicity 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 addition, in the dilepton and trilepton final state, the events are further separated according to the sum of the electrical charges of the leptons, to exploit the charge asymmetry present in several standard model (SM) background cross section in pp collisions ($t\bar{t}W$, WZ, single top t channel, W+jets). The gain in the precision of the signal strength measurement from categorizing the events is approximately 5%.

The expected and observed distributions for the number of selected jets and for the BDT output, for the different final states of the dilepton analysis, are shown in Fig. 2. The same distributions are shown for the trilepton analysis in Fig. 3. The distribution for the number of selected jets is also shown for the four-lepton channel in Fig. 3. In the dilepton case the data are in good agreement with the predictions in the ee and eμ channels, while an excess of signal like events is visible in the $\mu\mu$ final state. The details of this excess are discussed in Section 9. In the trilepton channel the overall data yield matches the expectations. The average jet multiplicity in data is a bit higher but, as the kinematic properties of the events are similar to the predicted ones, the distribution of the BDT discriminator is well reproduced. In the four-lepton channel only one event is observed with respect to an overall prediction of about three events.

7 Signal modeling

The signal is modeled with simulated events. The simulation has two different sources of systematic uncertainty. A significant uncertainty originates from the correction factors applied to the simulation in order to better reproduce the detector conditions and performance in data. Another uncertainty arise from assumptions made in the theoretical models used to produce the simulation. These two sources are described in details in Sections 7.1 and 7.2.

7.1 Correction factors and experimental uncertainties

As discussed in Section 4, scale factors are used to correct for differences in lepton performance between data and simulation. The scale factors account for the differences in the trigger, lepton preselection, and lepton MVA discriminant. Each of these scale factors has an uncertainty associated with it. The average per-lepton correction factors for the different final states range between 0.93 and 0.85, with associated systematic uncertainties of about 5% per lepton.

The corrections applied to jet energies in simulation have associated uncertainties [25]. These jet energy scale uncertainties are parameterized as a function of p_T, η, and flavour of the jets. The impact of these uncertainties is assessed by shifting the jet energy correction factors for each jet up and down by $\pm 1\sigma$ before the calculation of all kinematic quantities. The uncertainty from the jet energy resolution was found to play a negligible role in this analysis.

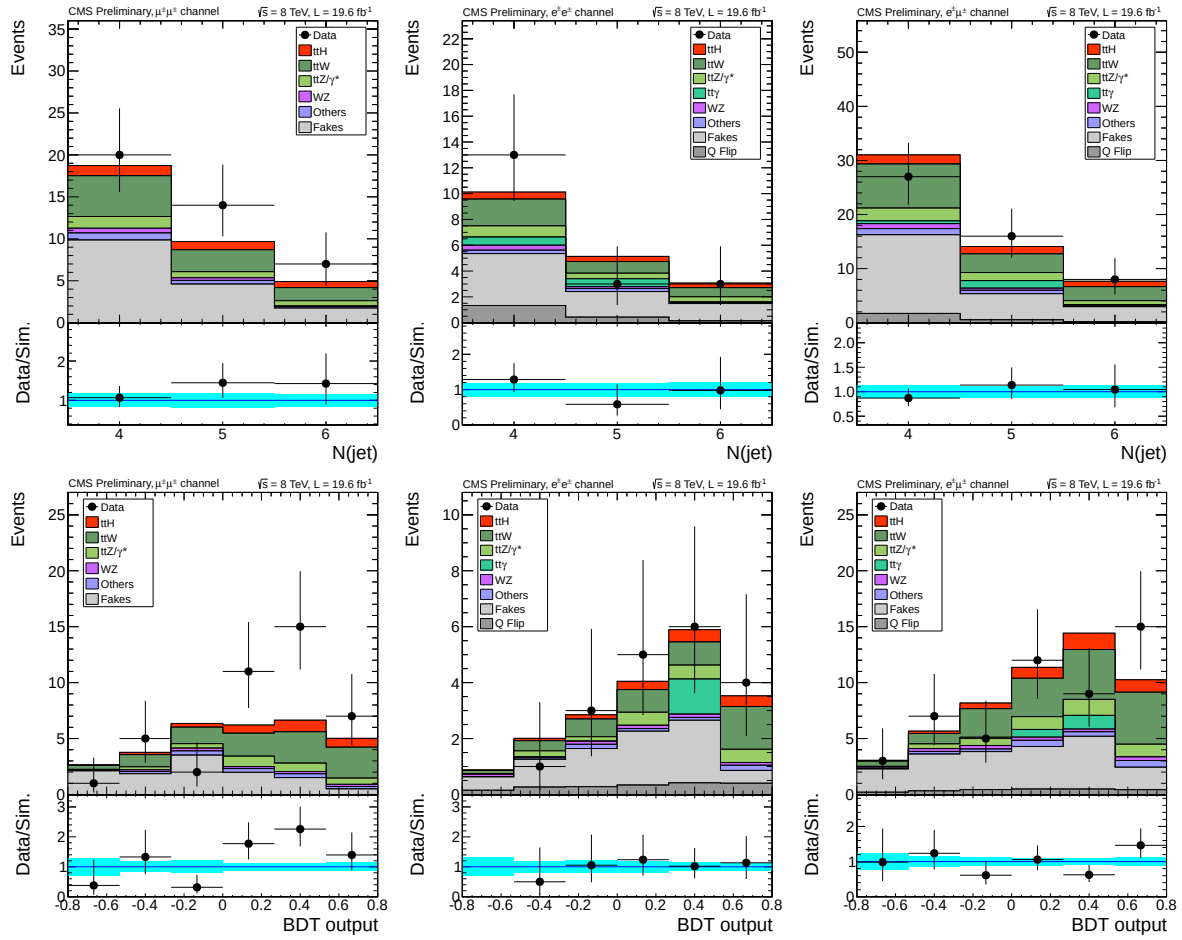

Figure 2: Distribution of the jet multiplicity (top row) and the BDT discriminant (bottom row) for the same-sign dilepton search, for the final states $\mu\mu$ (left), ee (center), and eμ (right). In these plots events with positive and negative charge are merged. The signal yield is the amount predicted 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. There is good agreement between the predicted and observed yields in the eμ and ee channels. There is an excess in the $\mu\mu$ channel.

The corrections for the b-tagging efficiencies for light, charm, and bottom flavour jets have associated uncertainties [26, 27]. These uncertainties are parameterized as a function of p_{T}, η, and flavour of the jets. Their effect on the analysis is evaluated by shifting the correction factor of each jet up and down by $\pm 1\sigma$ of the appropriate uncertainty.

In order to validate the agreement between data and simulation after these corrections, the same physics objects and methods are used to select well known processes with larger cross sections in dedicated control regions: $\mathrm{t\bar{t}} \to \mathrm{e}^{\pm}\mu^{\mp}\,\mathrm{b\bar{b}}\,\nu\bar{\nu}$, $\mathrm{WZ} \to 3\ell$, $\mathrm{ZZ} \to 4\ell$ and $\mathrm{Z} \to 4\ell$. The overall event yields in the control regions and the kinematic distributions are found to be in agreement with the predictions within the uncertainties (about 10%).

7.2 Theoretical uncertainties

The theoretical uncertainties on the NLO prediction for the inclusive $\mathrm{t\bar{t}H}$ production cross section amount to 6% from unknown higher orders in the perturbative series and 8% from the

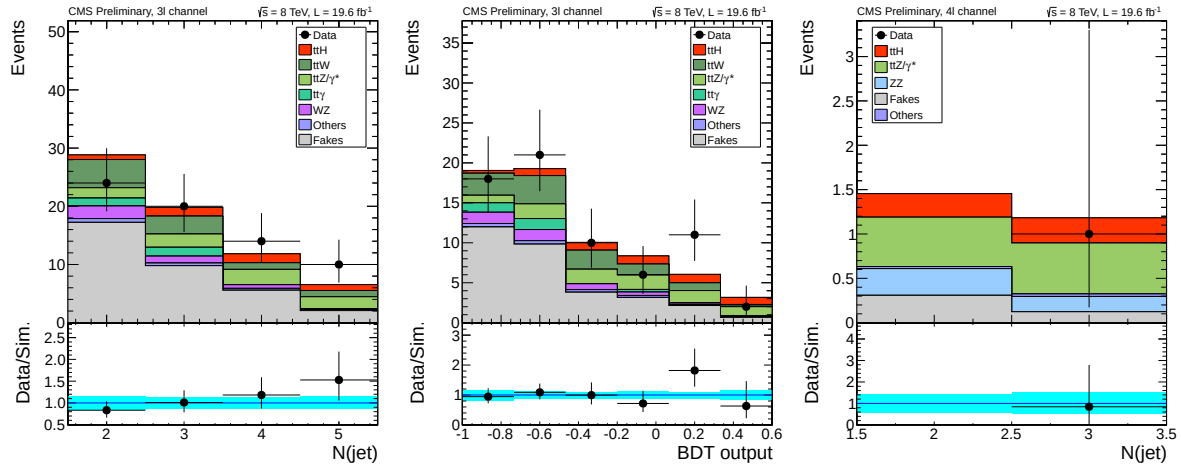

Figure 3: Distribution of the jet multiplicity (left) and BDT discriminant (center) for the trilepton search. Events with positive and negative charge are merged in these plots, but they are used separately in the signal extraction. The signal yield is the amount predicted by the standard model, ($\mu = 1$). The background yields are from the combined fit to the final discriminant at fixed $\mu = 1$. The plot on the right shows the jet multeplicity for the four lepton search. 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.

knowledge of the parton distribution functions (PDFs) [4].

Systematic uncertainties arising from the modeling of the process in PYTHIA are estimated by comparing the predictions obtained with different parameter configurations of the generator. The nominal prediction is taken from the Z2* tune, and an uncertainty band is obtained from the envelope of the predictions for the Z2*, Z2 [36], PROFESSOR Q_0^2 [37] and Perugia '11 [38] parameter sets. The relative differences observed in variables related to hadronic jet activity are of about 10%, while the uncertainties are smaller for variables related to leptons. Including also the older D6T parameter set, now disfavoured by LHC data, does not increase substantially the uncertainty.

Uncertainties from the PDFs beyond those affecting the overall normalization are estimated with the PDF4LHC prescriptions [39] using the thee PDF sets CT10 [40], NNPDF21 [41], MSTW2008 [42]. The resulting uncertainties on the shapes of the discriminating variables used for signal extraction amount to about 5%.

8 Background predictions

Three categories of backgrounds are identified in this search: $t\bar{t}V$ backgrounds from 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. These three classes of backgrounds are estimated separately with different methods, as described below. Additional minor backgrounds like triboson production are estimated directly from simulated events. The systematic uncertainties associated with each background estimate are discussed in the following sections and summarized in Table 2.

8.1 $t\bar{t}V$ backgrounds

Backgrounds from $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}WW$, are estimated from simulated events. Just like for the signal, corrections are applied for the different performance of individual physics objects in data and simulation measured in control regions in data. In all the analysed final states, the contribution of the $t\bar{t}WW$ process is found to be at least one order of magnitude smaller than $t\bar{t}W$ and $t\bar{t}Z$.

The inclusive production cross sections for the $t\bar{t}W$ and $t\bar{t}Z$ processes are taken from the NLO computation in [43], with theoretical uncertainties from unknown higher orders of 10% and 11% respectively, and uncertainties from the knowledge of the parton density functions of 7% and 9% respectively. The combined cross section of $t\bar{t}W$ and $t\bar{t}Z$ has been measured by CMS in 7 TeV data [44]. The results are consistent with theory but have larger uncertainties. In addition to the overall normalization, systematic uncertainties of theoretical origin on the distribution of the events in the final discriminating variables are considered. These uncerteinties are estimated conventionally by varying the renormalization and factorization scales up and down by a factor two and similarly varying the matching threshold between matrix element and parton shower. Uncertainties from parton distribution functions are also considered, as per the PDF4LHC prescriptions using the PDF sets CT10, NNPDF21, MSTW2008.

Experimental uncertainties originating from the knowledge of the lepton selection efficiencies, b-tagging efficiencies, and jet energy scales are also considereded. These uncertainties are estimated in the same way as for signal events following the descriptions in Section 7.1. They are naturally assumed to be completely correlated across the different physics processes.

The prediction for the $t\bar{t}Z$ process is tested directly in a trilepton control region requiring two of the leptons to have the same flavour, opposite electrical charge and the invariant mass pair of the pair to be within 10 GeV of the nominal Z boson mass. Agreement is observed in this control region, though the precision of the test is dominated by the statistical uncertainty of about 35%.

8.2 Diboson backgrounds

The WZ and ZZ production processes with the gauge bosons decaying to electrons, muons or taus can yield the same leptonic final states as the signal.

Without the requirement of additional hadronic jets in the final states, these processes are predicted theoretically at NLO accuracy, and the inclusive cross sections have been successfully measured at the LHC. This good agreement, however, does not translate automatically to the signal regions used in this search, which always require the presence of at least one additional b-tagged jet.

Since dibosons are preferentially produced in association with jets from light quarks or gluons, it is possible to isolate a clean control region of WZ or ZZ plus at least two jets by vetoing any event with a loose b tag, as well as inverting the $Z \to \ell\ell$ veto. The approach chosen for estimating this background is therefore to use simulated events but normalizing the overall event yield in control regions of WZ plus at least two non-b-tagged jets and ZZ plus at least one non-b-tagged jet. This procedure reduces the systematic uncertainty on the prediction. The theoretical uncertainty on the production cross section of diboson plus extra partons can be large. For instance, the uncertainty on the total rate of WZ from varying the renormalization/factorization scale by a factor two is $\sim$ 34%. The same uncertainty for ZZ is $\sim$ 32%. The extrapolation of event yields from the control region to the signal region is expected to have a smaller uncertainty: the majority of events from this background in the signal region contain

jets from gluons or light quarks mistagged as b-jets, for which the extrapolation is affected only by uncertainties of experimental origin, and for the remainder the uncertainties from unknown higher orders in QCD partially cancel out in the ratio.

The expected flavour composition in simulation for WZ events after the full selection in the trilepton final state is approximately 50% from WZ production in association with mistagged jets from light quarks or gluons, 35% from events with one jet originating from a charm quark, and the remaining from events with b quarks. For ZZ in the four-lepton final state, the expectation is about 40% of events with jets from gluons or light quarks, 35% from events with b quarks and 25% from events with c quarks.

The overall uncertainty on the normalization of the WZ background is 22%, of which 10% from the statistical uncertainty in the control region, 10% from the residual backgrounds in the control region, 15% from the uncertainties on the b-tagging efficiencies, 4% from the parton distribution functions and below 5% from the theoretical uncertainties on the extrapolation (dominated by the uncertainty on the flavor composition of the final state due to higher-order QCD terms).

The overall uncertainty on the normalization of the ZZ background is 19%, of which 12% from the statistical uncertainty in the control region, 4% from the residual backgrounds in the control region, 7.5% from the uncertainties on the b-tagging efficiencies, 3% from the parton distribution functions and 12% from the theoretical uncertainties on the extrapolation (dominated from the uncertainty on the ratio between the theoretical predictions for $ZZ + 2$ jets and $ZZ + 1$ jet).

8.3 Reducible backgrounds from misidentified or misreconstructed leptons

The reducible backgrounds in which at least one of the selected leptons does not originate from the decays of a W, Z or H boson are estimated from data. A control region dominated by reducible backgrounds is defined by selecting events with the same kinematics as the signal region, but for which at least one of the leptons fails the requirement on the multivariate lepton discriminant.

The extrapolation to the signal region is then performed by weighting the events in the control region as function of the probabilities for background-like leptons to pass the multivariate requirement, also measured from data in separate QCD dominated control regions as a function of the lepton p_{T} and η, separately for muons and electrons.

Events in which a single lepton fails the multivariate requirement enter the prediction with weight $\epsilon/(1-\epsilon)$, where ϵ denotes the aforementioned probability computed for the p_{T}, η and flavour of the lepton failing the selection. Events with two leptons failing the requirement are also used, but with a negative weight $-\epsilon_1\epsilon_2/[(1-\epsilon_1)(1-\epsilon_2)]$; this small correction is necessary to account for events with two background-like leptons contaminating the sample of events with a single lepton failing the requirement.

The measurement of the misidentificaiton probabilities and the application to the events in the control region are performed separately for events with at most one jet satisfying the medium operating point of the CSV b tagger and for events with at least two, to account for the different flavour composition and kinematic of the two samples.

The uncertainty for this background estimation is dominated by the systematic uncertainties on the modeling of the probabilities for background-like leptons to satisfy the multivariate selection, about 50%, estimated from a comparison of the measurements of the probabilities in different control regions in data, and from a closure test in which the procedure is applied to

simulated background events.

A similar procedure is applied in same-sign dilepton events to estimate the background originating from opposite-charge dileptons in which the charge of one lepton is misreconstructed. This step is relevant only for events with at least one electron, for which this background is 10-20% of that from misidentified leptons.

Charge misreconstruction probabilities are determined as function of the electron p_{T} and η from the observed yields of same-sign and opposite-sign dielectron pairs with mass within 10 GeV of the Z boson mass, and vary between 0.03% in the barrel to up to 0.3% for high-p_{T} electrons in the endcaps.

The prediction for background dilepton events with misreconstructed electron charge in the signal region is computed with opposite-sign dilepton events passing the full selection, except for the charge requirement: events with a single electron enter the prediction with a weight equal to the charge misreconstruction probability for that electron, while dielectron events enter the prediction with a weight equal to the sum of the charge misreconstruction probabilities for the two electrons. The systematic uncertainty on this estimate is 30%, determined from a closure test of the procedure on simulated events.

9 Results

The results of this search are interpreted by comparing the observed event yields and distributions to the expectations from background and a SM Higgs boson of mass 125.7 GeV. A common signal strength parameter $\mu = \sigma/\sigma_{\mathrm{SM}}$ is introduced, scaling the expected yields from $\mathrm{t\bar{t}H}$, without altering the branching fractions or the kinematics of the events.

Results are reported both in terms of upper limits on μ at 95% confidence level (CL), and in terms of the best fit value for μ and its associated uncertainty. The statistical procedures adopted in this analysis are the ones that have used for the observation of the Higgs boson candidate in CMS, and are described in detail in Ref. [45].

The observed upper limits on μ are shown in Fig. 4, compared with the expectations under the background-only hypothesis, ie assuming no $\mathrm{t\bar{t}H}$ production. In the absence of a $\mathrm{t\bar{t}H}$ signal, the median expected upper limit from the combination of all the decay modes is 2.4, at 95% CL; the corresponding median expectation under the hypothesis of SM $\mathrm{t\bar{t}H}$ production is 3.5. The observed upper limit is 6.6, larger than both predictions, driven by the excess of events seen especially in the same-sign dimuon category.

The best fit signal strengths from the individual channels are shown in the right hand panel of Fig. 4. The internal consistency of the five results with a common signal strength has been evaluated to be 16%, estimated from the asymptotic behaviour of the profile likeilhood function [3]. The fit to the combination yields $\mu = 3.7^{+1.6}_{-1.4}$: the deficit in the four-lepton channel partially compensate for the excess in the same-sign $\mu\mu$ one, bringing the combined result closer to the results of the other three channels (trilepton, same-sign ee, and same-sign eμ). The combined μ is compatible with the SM Higgs boson prediction $\mu = 1$ at the 3% level.

Cross-checks of the excess of data events in the same-sign dimuon category showed that they were unlikely to come from an underestimated background. The agreement between expected an observed yields in the ee and eμ channels suggests that the background estimates are reasonable. In particular, background from electrons with a mismeasured charge is already small in the ee and eμ channels. The rate of charge mismeasurement for muons would need to be

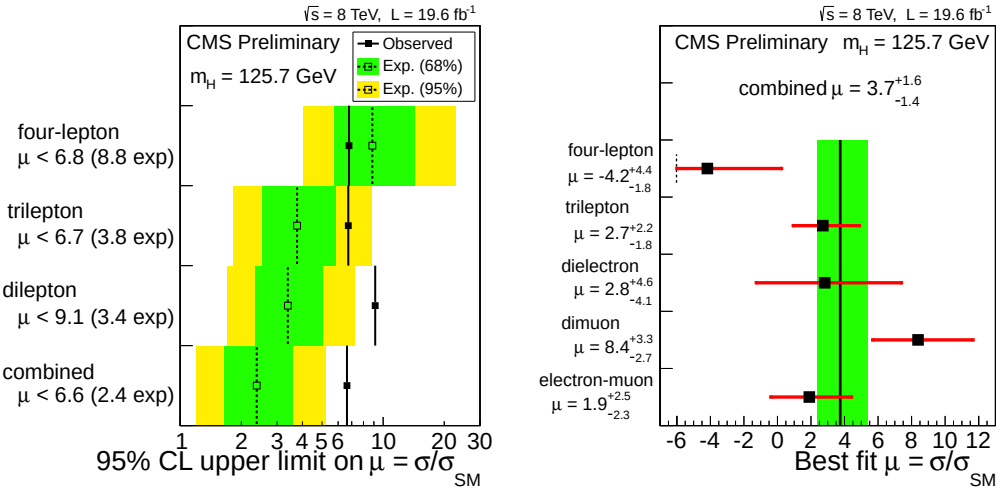

Figure 4: Results of the searches in the three final states and their combination, in terms of the signal strength parameter $\mu = \sigma/\sigma_{\mathrm{SM}}$. Left panel: 95% CL upper limit on μ, observed (solid markers), median expected under the background-only hypothesis (hollow markers), and intervals containing 68% and 95% of the expected outcomes under that hypothesis (green and yellow bands). Right panel: best fit values of μ and $\pm 1\sigma$ uncertainties, for the five individual final states (solid markers with red error bars) and the full combination (vertical line and green band). 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 any of the two bins of jet multiplicity.

10 times greater than the rate of charge mismeasurement for electrons in order to explain the excess. Detailed studies of various single- and dimuon distributions did not reveal any potential additional source of background. Moreover, the analysis of the dimuon final state has been repeated with different lepton selections, using looser working points for the multivariate discriminator and also with traditional cut based selections. These approaches have sensitivities 10-50% worse than the nominal analysis and give compatible results.

The results obtained with the cross-check analysis relying on the multiplicity of hadronic jets instead of the multivariate discriminator for the dilepton and trilepton final states are in good agreement with the ones of the nominal analysis: the expected and observed upper limits are 3.0 and 6.9, respectively, and the best fit signal strength is $\mu = 3.9^{+1.7}_{-1.5}$.

10 Conclusions

A search for the standard model Higgs boson produced in association with a top-quark pair has been performed at the CMS experiment using the full 2012 data sample, corresponding to an integrated luminosity of 19.6 fb^{-1} at $\sqrt{s} = 8$ TeV. Events are considered where the top-quark pair decays to either one lepton+jets ($\mathrm{t\bar{t}} \to \ell\nu\mathrm{q\bar{q}'b\bar{b}}$) or dileptons ($\mathrm{t\bar{t}} \to \ell^+\nu\ell^-\bar{\nu}\mathrm{b\bar{b}}$), ℓ being an electron or a muon. The search has been optimized for the $\mathrm{H} \to \mathrm{WW}^*$, $\mathrm{H} \to \mathrm{ZZ}^*$, and $\mathrm{H} \to \tau^+\tau^-$ decay modes.

Combining the results from the same-sign dilepton, three lepton, and four lepton channels, the observed and expected upper limits at 95% CL on the cross section for Higgs boson production in association with top-quark pairs for a Higgs boson mass of 125.7 GeV are 6.6 and 2.4 times the standard model expectation, respectively. The best-fit value for the signal strength μ is $3.7^{+1.6}_{-1.4}$ (68% CL).

Syst Name	Rate or Shape	Description
$t\bar{t}H$ higher orders	rate	Theoretical uncertainty on $t\bar{t}H$ cross section.
$t\bar{t}W$ higher orders	rate	Theoretical uncertainty on $t\bar{t}W$ cross section.
$t\bar{t}Z$ higher orders	rate	Theoretical uncertainty on $t\bar{t}Z$ cross section.
PDF	rate	Theoretical uncertainty on cross sections for $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$. Correlated in all channels for all processes sharing a dominant production mechanism.
$t\bar{t}H$ PDF Shape	shape only	Theoretical uncertainty from PDF on shape.
$t\bar{t}W$ PDF Shape	shape only	Theoretical uncertainty from PDF on shape.
$t\bar{t}Z$ PDF Shape	shape only	Theoretical uncertainty from PDF on shape.
$t\bar{t}H$ PYTHIA tune	shape only	Theoretical uncertainty on MC modeling.
$t\bar{t}W$ MADGRAPH tune	shape only	Theoretical uncertainty on MC modeling.
$t\bar{t}Z$ MADGRAPH tune	shape only	Theoretical uncertainty on MC modeling.
Non-prompt Fake Rate	envelope	Applied to reducible non-prompt backgrounds.
Charge-flip	envelope	Applied to charge flip background for 2ℓ channel.
WZ	rate	Uncertainty from fit in control region.
ZZ	rate	Uncertainty from fit in control region.
Jet Energy Scale	template	Applied to WZ,ZZ,$t\bar{t}W$,$t\bar{t}Z$,$t\bar{t}H$.
b-tagging efficiency	rate	Applied to WZ,ZZ,$t\bar{t}W$,$t\bar{t}Z$,$t\bar{t}H$.
b-tagging fake rate	rate	Applied to WZ,ZZ,$t\bar{t}W$,$t\bar{t}Z$,$t\bar{t}H$.
Lepton Trigger Scale factor	rate	Applied to WZ,ZZ,$t\bar{t}W$,$t\bar{t}Z$,$t\bar{t}H$.
Lepton preselection Scale factor	rate	Applied to WZ,ZZ,$t\bar{t}W$,$t\bar{t}Z$,$t\bar{t}H$.
Lepton MVA discriminator scale factor	rate	Applied to W,ZZ,$t\bar{t}W$,$t\bar{t}Z$,$t\bar{t}H$.
Luminosity	rate	Applied to WZ,ZZ,$t\bar{t}W$,$t\bar{t}Z$,$t\bar{t}H$.

Table 2: Summary of systematic uncertainty treatment in the fit to extract the signal. Systematics are correlated across all channels unless otherwise stated.

The results of this search show an excess of signal-like events compared to the expectations from the SM Higgs boson and the backgrounds, but are still compatible with those predictions. More data, in this and other $t\bar{t}H$ searches, is needed to assess more accurately whether the production cross section for $t\bar{t}H$ is that predicted by the SM, or if there are any deviations brought about by new physics beyond the SM.

References

[1] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021.

[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.

[3] CMS Collaboration, "Measurements of the properties of the new boson with a mass near 125 GeV", CMS Physics Analysis Summary CMS-PAS-HIG-13-005, (2013).

[4] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", *CERN-2011-002* (CERN, Geneva, 2011) arXiv:1101.0593.

[5] D∅ Collaboration, "Search for the Standard Model Higgs boson in the $t\bar{t}H \to t\bar{t}b\bar{b}$ channel", D∅ Conference Note CONF-5739, (2008).

[6] CDF Collaboration, "Search for the Standard Model Higgs Boson Produced in Association with Top Quarks Using the Full CDF Data Set", *Phys. Rev. Lett.* **109** (2012) 181802, doi:10.1103/PhysRevLett.109.181802.

[7] ATLAS Collaboration, "Search for the Standard Model Higgs boson produced in association with top quarks in proton-proton collisions at $\sqrt{s} = 7$TeV using the ATLAS detector", ATLAS Conference Report ATLAS-CONF-2012-135, (2012).

[8] 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 CMS-PAS-HIG-13-019, (2013).

[9] ATLAS Collaboration, "Search for $t\bar{t}H$ production in the $H \to \gamma\gamma$ channel at $\sqrt{s} = 8$ TeV with the ATLAS detector.", ATLAS Conference Note ATLAS-CONF-2013-080, (2013).

[10] CMS Collaboration, "Search for $t\bar{t}H$ production in events with $H \to \gamma\gamma$ at $\sqrt{s} = 8$ TeV collisions", CMS Physics Analysis Summary CMS-PAS-HIG-13-015, (2013).

[11] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[12] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **0605** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[13] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[14] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **0411** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[15] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **0711** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[16] 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* **1006** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[17] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J.* **C71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[18] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **0909** (2009) 111, doi:10.1007/JHEP02(2010)011,10.1088/1126-6708/2009/09/111, arXiv:0907.4076.

[19] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, "W+W-, WZ and ZZ production in the POWHEG BOX", *JHEP* **1111** (2011) 078, doi:10.1007/JHEP11(2011)078, arXiv:1107.5051.

[20] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.

[21] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, (2009).

[22] CMS Collaboration, "Commissioning of the Particle–Flow reconstruction in Minimum–Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-002, (2010).

[23] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[24] M. Cacciari, G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder", *Phys. Lett. B* **641** (2006) 57, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.

[25] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) 11002, doi:10.1088/1748-0221/6/11/P11002.

[26] CMS Collaboration Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[27] CMS Collaboration, "Results on b-tagging identification in 8 TeV pp collisions", CMS DP 2013/005, (2013).

[28] S. Baffioni et al., "Electron reconstruction in CMS", *Eur. Phys. J. C* **49** (2007) 1099, doi:10.1140/epjc/s10052-006-0175-5.

[29] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV", CMS Physics Analysis Summary CMS-PAS-EGM-10-004, (2010).

[30] CMS Collaboration, "Electron commissioning results at $\sqrt{s} = 7$ TeV", CMS Detector Performance Summary CMS-DP-2011-003, (2011).

[31] CMS Collaboration, "Electron performance with 19.6 fb^{-1} of data at $\sqrt{s} = 8$ TeV with the CMS detector", CMS Detector Performance Summary CMS-DP-2013-003, (2013).

[32] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[33] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with leptons from J/ψ and W decays at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-003, (2010).

[34] CMS Collaboration, "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **10** (2011) 132, doi:10.1007/JHEP10(2011)132.

[35] CMS Collaboration, "Studies of Tracker Material", CMS Physics Analysis Summary CMS-PAS-TRK-10-003, (2010).

[36] R. Field, "Min-Bias and the Underlying Event at the LHC", *Acta Phys. Polon.* **B42** (2011) 2631, doi:10.5506/APhysPolB.42.2631, arXiv:1110.5530.

[37] A. Buckley et al., "Systematic event generator tuning for the LHC", *Eur. Phys. J.* **C65** (2010) 331, doi:10.1140/epjc/s10052-009-1196-7, arXiv:0907.2973.

[38] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes", *Phys. Rev.* **D82** (2010) 074018, doi:10.1103/PhysRevD.82.074018, arXiv:1005.3457.

[39] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", arXiv:1101.0538.

[40] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys. Rev. D* **78** (2008) 013004, doi:10.1103/PhysRevD.78.013004.

[41] R. D. Ball et al., "A first unbiased global NLO determination of parton distributions and their uncertainties", *Nucl. Phys.* **B838** (2010) 136, doi:10.1016/j.nuclphysb.2010.05.008, arXiv:1002.4407.

[42] A. Martin, W. Stirling, R. Thorne, and G. Watt, "Parton distributions for the LHC", *Eur. Phys. J.* **C63** (2009) 189, doi:10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002.

[43] M. Garzelli, A. Kardos, C. Papadopoulos, and Z. Trocsanyi, "$t\bar{t}\,W^{+-}$ and $t\bar{t}\,$Z Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects", *JHEP* **1211** (2012) 056, doi:10.1007/JHEP11(2012)056, arXiv:1208.2665.

[44] CMS Collaboration, "Measurement of associated production of vector bosons and $t\bar{t}$ at $\sqrt{s}$ = 7 TeV", *Phys. Rev. Lett.* **110** (2013) 172002, doi:10.1103/PhysRevLett.110.172002, arXiv:1303.3239.

[45] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV", *JHEP* **06** (2013) 081, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.