Available online at www.sciencedirect.com

ScienceDirect

Nuclear Physics B Proceedings Supplement 00 (2014) 1–6

Nuclear Physics B Proceedings Supplement

Search for the associated Higgs boson production with top quarks in CMS

Lorenzo Bianchini
on behalf of the CMS Collaboration

Institute for Particle Physics, ETH Zürich
Otto-Stern-Weg 5, 8093 Zürich

Abstract

A precise determination of the Yukawa coupling to top quarks is crucial for the characterization of the recently discovered Higgs boson. A direct measurement of this coupling can be provided only by those processes where the Higgs boson is produced in association with top quarks. In the Standard Model (SM), the production of the Higgs boson together with a top-antitop pair ($t\bar{t}H$) has the largest cross section, followed by the production with a single top quark (tH), with roughly one order of magnitude smaller rate. As of summer 2014, the CMS experiment at LHC has performed searches for either processes using the full Run-I data set in a variety of decay channels of the Higgs boson, namely $H \to WW^*, ZZ^*, \tau^+\tau^-, b\bar{b}, \gamma\gamma$ for the $t\bar{t}H$ production, and $H \to \gamma\gamma$ for tH. Due to the small cross section, the sensitivity to the SM tH production is still far from being achieved, yet stringent constraints can be already set for non-SM couplings. Conversely, the combination of the various $t\bar{t}H$ searches has already attained sensitivity to measure a SM-like signal, providing a best-fit value of the signal strength modifier $\hat{\mu} = 2.76^{+1.05}_{-0.92}$, compatible with the SM expectation at the 2σ level. An improved $t\bar{t}H$ search in the $H \to \gamma\gamma$ channel using the full 7 and 8 TeV data set, as well as a new $t\bar{t}H, H \to b\bar{b}$ analysis using a matrix element technique, are here presented for the first time.

Keywords:
Higgs, top quark, CMS, matrix element

1. Introduction

The precise measurement of the coupling of the recently discovered Higgs boson [1, 2] to fermions is a milestone in particle physics. Due to its heavy mass, the top quark is the only fermion with a full-strength (order 1) Yukawa coupling (y_t). In the standard model (SM) indirect insights on the Higgs-top coupling can be inferred from measuring top-quark dominated, loop-induced processes like $gg \to H$ or $H \to \gamma\gamma$; however, beyond-standard model particles may also participate in these loops. A direct probe of y_t can be provided only by those processes where the Higgs boson is radiated from a top quark. Among such reactions, the production in association with a top-antitop pair, denoted as $t\bar{t}H$ in the following, has the largest cross section (proportional to $|y_t|^2$), with a theoretical prediction $\sigma_{t\bar{t}H} = 129.3^{+3.8}_{-9.3}(\text{scale}) \pm 8.1(\text{PDF} + \alpha_s)$ fb at NLO in QCD for a Higgs boson mass of 125 GeV [3]. The cross section for Higgs boson production in association with a single top quark (mostly through t-channel exchange of a W boson) is one order of magnitude smaller ($\sigma_{tH} \approx 18$ fb), as a result of the interference between diagrams where the Higgs boson is attached either to the top-quark leg or to the W boson propagator. The interference is indeed destructive as a consequence of the opposite sign of the Higgs couplings to top quarks and W bosons [4]. A sign flip of the former would enhance considerably both the cross section and the branching ratio to photons, which is subject to the same interference effect.

In this talk, the most recent results on the search for the associated Higgs boson production with top quarks using the CMS detector are presented.

2. Top plus Higgs searches in CMS

Searches for Higgs boson production with top quarks performed by the CMS experiment [5] can be grouped into three main analysis streams by exclusive Higgs boson decay: $H \to$ hadrons [7, 8], $H \to$ leptons [9], and $H \to$ photons [10, 11].

Hadronic jets are defined by clustering particle-flow particles with an anti-k_T algorithm using a radius parameter $R = 0.5$. Jets with a corrected momentum transverse to the beam line (p_T) as low as 20 GeV are considered. The combined secondary vertex (CSV) algorithm [13] is employed to tag jets that originate from the fragmentation of b quarks (b tagging). The algorithm assigning a continuous discriminant to each jet, with large values indicating more compatibility with a b quark than with a light quark or a gluon. The loose and medium working points of the CSV algorithm (CSVL and CSVM, respectively) are usually adopted to define tagged jets. Due to the tracker coverage, the b tagging is possible only for jets with pseudorapidity $|\eta| < 2.5$. Standard CMS algorithms are used to reconstruct electrons (e), muons (μ), hadronic taus (τ_h), and photons (γ).

2.1. $H \to$ hadrons

The analyses documented in Ref. [7, 8] search for events where the Higgs boson is produced in association with a top-antitop pair, and subsequently decays into a pair of b quarks. At least one of the top quark is required to decay leptonically into either electrons, muons, or leptonically-decaying taus. These signatures result in busy final states and are contaminated by the overwhelming $t\bar{t}$ + jets background, with typical signal-to-background ratios of a few percent only.

2.1.1. BDT analysis

The analysis documented in Ref. [7] complements the search with a dedicated channel where the Higgs boson decays to tau leptons, and both taus subsequently decay to hadrons ($H \to \tau_h\tau_h$). In the $H \to b\bar{b}$ channel, events are collected using single or double-lepton triggers, and are required to contain offline one isolated lepton $\ell = e, \mu$ with $p_T > 30$ GeV (LJ channel), or two isolated and opposite-sign leptons ($ee, e\mu, \mu\mu$) with p_T in excess of 20 and 10 GeV respectively (DIL channel). For the LJ (DIL) channel, at least four (three) jets with $p_T > 30$ GeV and $|\eta| < 2.4$ not overlapping with the lepton(s) are further requested, with at least two of them passing the CSVM working point. The $H \to \tau_h\tau_h$ channel (labelled $\tau_h\tau_h$ in the following) selects single-lepton events plus at least four jets, of which exactly two are identified as hadronic taus and two or more are tagged. In each channel, events are further separated into exclusive categories based on the number of jets and tags, with the highest multiplicity bin defined as inclusive. A total of 7, 3, and 6 bins are considered for the LJ, DIL, and $\tau_h\tau_h$ channels, respectively. Figure 1 reports the expected and observed event yields per category. In order to separate signal events from the larger $t\bar{t}$+jets background, Boosted Decision Trees (BDT) are trained separately in each category. The input variables are chosen among a pull of well simulated discriminating observables encoding event shape information, single jet or jet-pair kinematics, and b tagging. For the LJ and DIL channels, a separate BDT is first trained to separate signal events from a simulation of $t\bar{t}$ plus heavy-flavor jets, and then its output is used as an additional input variable to the final BDT. The shape and rate of the backgrounds is estimated from the Monte Carlo simulation. A data-driven correction to the `MadGraph` $t\bar{t}$+jets simulation is applied to correct for discrepancies observed in the top p_T spectrum. Events that at generator level contain one (two) jets matched to b quarks not coming from a top decay are labelled as $t\bar{t} + b$ ($t\bar{t} + b\bar{b}$). Likewise, events with at least one jet matched to a c quark are labelled $t\bar{t} + c\bar{c}$. The three sub-processes are treated as independent backgrounds, and their normalisation is assigned an *a priori* uncertainty of 50% to account for the large theoretical uncertainties on their modeling. The output of the BDTs in the various categories is used for the statistical interpretation. The best-fit value of the signal strength modifier from the combined LJ+DIL fit is $\hat{\mu} = 0.65^{+1.85}_{-1.81}$. The $\tau_h\tau_h$ channel is less sensitive and provides a fit $\hat{\mu} = -1.3^{+6.1}_{-3.6}$.

2.1.2. ME analysis

The analysis recently presented in Ref. [8] searches for $t\bar{t}H$ events in the $H \to b\bar{b}$ final state using an analytical matrix element (ME) method. Events are collected using single-electron, single-muon, or double-electron triggers, and are required to contain offline one isolated lepton $\ell = e, \mu$ with $p_T > 30$ GeV, or two isolated and opposite-sign leptons ($ee, e\mu, \mu\mu$) with p_T in excess of 20 GeV. For single- (double-) lepton events, at least five (four) jets with $p_T > 30$ GeV and $|\eta| < 2.5$ are further requested. Based on the per-jet CSV output, the likelihood $\mathcal{L}_{bbbb}$ ($\mathcal{L}_{bbjj}$) that exactly four (two) of the selected jets originate from b quarks and the remaining jets from lighter quarks is calculated for each jet-flavor assignment, and then summed over all possible assignments. The likelihood ratio $(1 + \frac{\sum \mathcal{L}_{bbjj}}{\sum \mathcal{L}_{bbbb}})^{-1}$ is then used as a cut variable to select those events most compatible with a $t\bar{t}$+ heavy-flavor hypothesis. Two work-

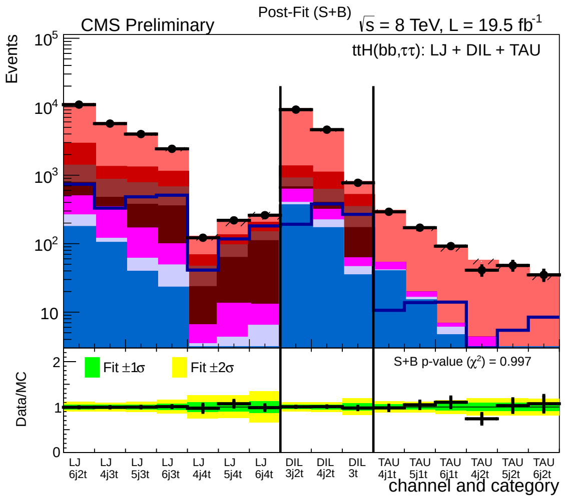

Figure 1: Expected and observed event yields in the BDT analysis, broken-up by category. The nuisance parameters affecting the signal and background normalisation are fixed to their best-fit value as obtained from a signal plus background fit with the constraint $\mu = 1$.

ing points are defined, one tight and one loose, which split events into two exclusive categories, denoted as low and high purity, respectively. For the sake of performing the event interpretation, the jet-flavor assignment that maximises $\mathcal{L}_{bbbb}$ is used to define which jets are matched to b quarks (tagged jets) and which to other kind of partons (untagged jets). Each event is then assigned a unique event interpretation, under either the $pp \to t\bar{t}H, H \to b\bar{b}$ (signal) or the $pp \to t\bar{t}b\bar{b}$ (background) hypothesis. Depending on the number of leptons, on the jet multiplicity, and on the invariant mass of the untagged jet pair(s) m_{jj}, an event can fulfil either of the following distinct Born-level interpretations:

1. single lepton plus $\geq$ 6 jets, and $|m_W - m_{jj}| \leq 20$ GeV $\Leftrightarrow$ one top quark decays leptonically and the other hadronically; all 6 quarks in the event are reconstructed as separate jets. If more than six jets are reconstructed in the event, the additional jets are interpreted as extra gluon radiation.
2. single lepton plus $\geq$ 6 jets, and $|m_W - m_{jj}| > 20$ GeV $\Leftrightarrow$ one top quark decays leptonically and the other hadronically; one quark from the W decay has fallen outside of the detector acceptance, while the other 5 quarks are reconstructed as separate jets. The additional jet(s) is (are) interpreted as extra gluon radiation.
3. single lepton plus 5 jets $\Leftrightarrow$ one top quark decays leptonically and the other hadronically; one quark from the W decay has fallen outside of the detector acceptance, while the other 5 quarks are reconstructed as separate jets.
4. double lepton plus $\geq$ 4 jets $\Leftrightarrow$ both top quark decay leptonically; all 4 quarks in the event are reconstructed as separate jets. If more than six jets are reconstructed in the event, the additional jets are interpreted as extra gluon radiation.

For a given event interpretation, the probability density under either the signal ($\mathcal{P}_S$) or background hypothesis ($\mathcal{P}_B$) is calculated. The narrow-width approximation is used to factorise the matrix elements for the hard-scattering (calculated at LO accuracy using the `OpenLoops` [14] program) from the top and Higgs decay amplitude (calculated analytically without spin correlations). A numerical integration over the final state particles allows to marginalise over unreconstructed or poorly measured observables. Single or double Gaussian transfer functions are used to relate the measured jet energy to the true parton energy, whereas the parton direction is constrained to the associated jet direction through delta distributions. The vector sum of the neutrino momenta is constrained to the measured missing transverse vector through a binormal transfer function. All possible quark-jet assignments are considered, and the corresponding probabilities summed up. By combining the matrix-element weight with the b tagging likelihood, three event weights (w_i) are defined: $w_S = \mathcal{L}_{bbbb} \cdot \mathcal{P}_S$, $w_{B_1} = \mathcal{L}_{bbbb} \cdot \mathcal{P}_B$, and $w_{B_2} = \mathcal{L}_{bbjj} \cdot \mathcal{P}_B$, The final signal/background discriminant is defined as:

$$P_{s/b} = \left[1 + \frac{\lambda \cdot w_{B_1} + (1-\lambda) \cdot w_{B_2}}{w_S} \right]^{-1}, \qquad (1)$$

where λ is the ratio between $t\bar{t} + b\bar{b}$ and $t\bar{t} + jj, j \neq b$, as predicted by the MC simulation. The background estimation and the treatment of systematics is the same as in the BDT analysis. The distribution of the discriminant defined by Eq. (1) is used for the statistical interpretation. A total of eight independent categories (one low- and one high-purity for each of the four event interpretations) are considered. For illustration, the $P_{s/b}$ distribution for the most sensitive category is shown in Fig. 2. The expected (observed) 95% CL limit on μ is 2.9 (3.3) times the SM expectation, corresponding to a best-fit value $\hat{\mu} = (0.7 \pm 1.4)$.

2.2. *$H \to$ leptons*

The analysis documented in Ref. [9] searches for Higgs bosons decaying to final states with at least one lepton $\ell = e, \mu$, namely $H \to ZZ^*, WW^*, \tau\tau$, and produced in coincidence with a top-antiquark pair. Three

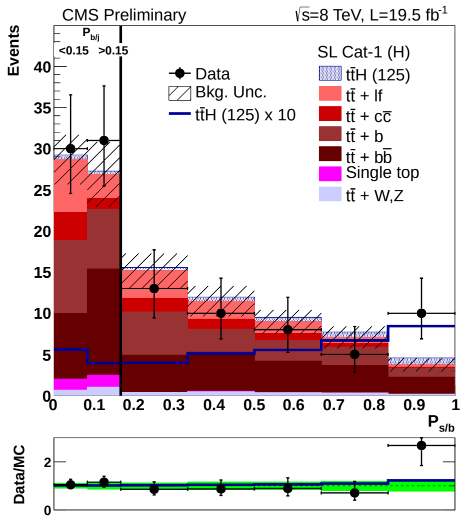

Figure 2: Distribution of the $P_{s/b}$ discriminant of Eq. (1) for the most sensitive event category. The nuisance parameters affecting the signal and background normalisation are fixed to their best-fit value as obtained from a signal plus background fit with the constraint $\mu = 1$.

distinct multi-lepton signatures are exploited: same-sign dilepton, trileptons, and four leptons. Events featuring any of these distinctive signatures are further required to have at least two jets and at least one CSVM tag, or two CSVL tags. For the dilepton channel, two more jets are required. To suppress the Drell-Yan production, a cut on a linear discriminant built using E_T^{miss} and H_T^{miss} (the magnitude of the p_T vector sum of all particles, and of the selected jets and leptons, respectively) is applied. In addition, events with a same-flavor lepton pair with mass less than 10 GeV away from the Z boson mass are vetoed. Instrumental background coming from misidentified leptons (mostly non-prompt leptons from b hadron decay) is reduced by a dedicated multivariate lepton identification, which is calibrated from data using a tag-and-probe technique. In the dilepton and trilepton channels, events with total positive or total negative charge are treated separately, as to exploit the charge asymmetry typical of some of the SM backgrounds. The $t\bar{t} + V$ background is predicted using a Monte Carlo simulation validated in control regions. The diboson background is also predicted using the simulation, and normalised to the observed event yields in control regions without the b tagging. Finally, for each category a BDT is trained using jet and lepton related variables, the jet multiplicity being the single most discriminating variable. The best-fit value of the signal strength modifier from the combined fit is $\hat{\mu} = 3.9^{+1.7}_{-1.4}$. The large value of the fitted signal yield is driven by a broad excess of signal-like events in the $\mu^{\pm}\mu^{\pm}$ channel, clearly visible in Fig. 3.

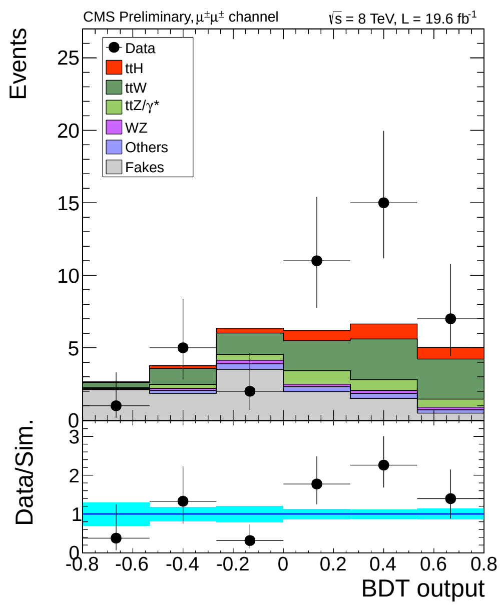

Figure 3: Distribution of the BDT output in the $\mu^{\pm}\mu^{\pm}$ channel. The nuisance parameters affecting the signal and background normalisation are fixed to their best-fit value as obtained from a signal plus background fit with the constraint $\mu = 1$.

2.3. $H \to$ photons

The analyses documented in Ref. [10, 11] search for $t\bar{t}H$ or tH events where the Higgs boson decays into a photon pair.

2.3.1. $t\bar{t}H$ tag

In the 8 TeV analysis, events are categorised into two separate classes by requiring the presence of either zero or at least one charged lepton. The leading and trailing photons are required to have $E_T > m_{\gamma\gamma}/2$ and 25 GeV, respectively, as to comply with the trigger requirement. Jets are required to have $p_T > 25$ GeV, and both classes require the presence of at least one CSVM tag. The lepton tag channel is then defined by requiring at least one more jet in the event and at least one electron or muon with $p_T > 20$ GeV, while the multijet channel is defined by the requirement of at least four more jets in the event and no leptons. Requirements are also made on the minimum diphoton BDT classifier score [10] for entry into the two classes: 0.17 for the lepton class, and 0.48 for the multijet class. For the 7 TeV dataset the events in the two classes are combined after selection to form a single event class. The contamination from other Higgs production mechanisms (gluon fusion or VH) is estimated from MC to be of about 10%. The signal is extracted by an unbinned fit to the $m_{\gamma\gamma}$ spectrum. Figure 4 shows the $m_{\gamma\gamma}$ spectrum with its signal plus background fit, as observed in the 8 TeV multijet channel. The best-fit value of the signal strength modifier from the combined fit is $\hat{\mu} = 2.7^{+2.4}_{-1.7}$.

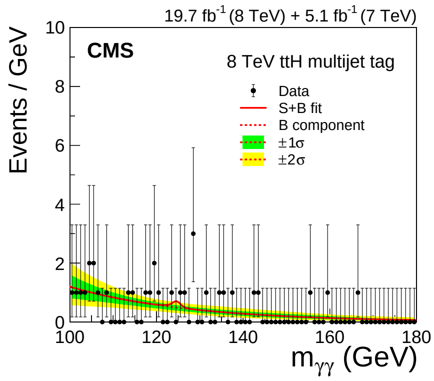

Figure 4: Events tagged in the multijet channel of the 8 TeV dataset, binned as a function of $m_{\gamma\gamma}$.

2.3.2. *tH tag*

The event selection starts from the same diphoton definition used for the $t\bar{t}H$ tag, and is then optimised to tag the $qb \to tHq'$ signature, which at leading order proceeds via a t-channel scattering of a valence quark via W boson exchange (see Fig. 5). Exactly one lepton with $p_T > 10$ GeV and one tagged jet in excess of 20 GeV are required, consistently with a single-top decay. Furthermore, the hardest jet in the event must have $p_T > 20$ GeV and $|\eta| > 1.0$. To suppress the contamination from $t\bar{t}H$ production, a likelihood discriminant involving the jet and lepton properties is used. The signal region is defined by requiring the diphoton mass $m_{\gamma\gamma}$ to be less than 3 GeV away from the nominal Higgs boson mass. Resonant backgrounds stemming from other Higgs production modes can contaminate the signal region, and are estimated from MC, while the non-resonant background is estimated by counting events in the [100, 180] GeV sideband. The extrapolation factor between the mass sideband and the signal region is measured from data using looser data selection or control regions with inverted photon identification. No events are counted neither in the sideband nor in the signal region, and 95% CL upper limits are set on the signal strength modifier. Assuming $y_t = -y_t^{SM}$, the expected (and observed) limit is 4.1 times the corresponding cross section × branching ratio.

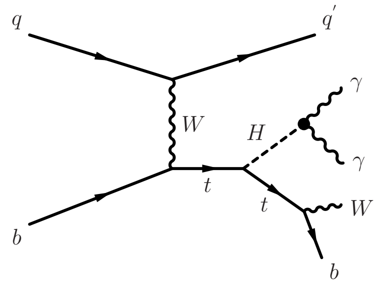

Figure 5: Example of a LO diagram contributing to the $qb \to tHq'$, $H \to \gamma\gamma$ process in a five-flavor scheme.

2.4. *Combination of $t\bar{t}H$ searches*

The results from the various $t\bar{t}H$ searches are summarised in Table 1. The analyses documented in Ref. [7, 9, 10] select mutually exclusive sets of events: they can be therefore easily combined, thus profiting from inter-correlations for some of the nuisance parameters (e.g. luminosity, b tagging). The best-fit value of the signal strength modifier from the combined fit (assuming SM branching ratios) is $\hat{\mu} = 2.76^{+1.05}_{-0.92}$ [12]. The larger value compared to the SM expectation is driven by the excess measured in the $H \to$ leptons channel and, to a lesser extent, in the $H \to$ photons channel. Assuming asymptotic properties for the profiled likelihood test statistic, the compatibility of this measurement with the SM expectation ($\mu = 1$) is estimated to be 2.0σ.

$t\bar{t}$ tag	H tag	$\hat{\mu} \pm \Delta\hat{\mu}$	Ref.
LH or LL	$b\bar{b}$	$0.65^{+1.85}_{-1.81}$	[7]
LH or LL	$b\bar{b}$	$0.67^{+1.35}_{-1.33}$	[8]
LH	$\tau_h\tau_h$	$-1.32^{+6.08}_{-3.6}$	[7]
LH,LL,HH	VV^*, $\tau_\ell\tau_{\ell/h}$	$3.94^{+1.70}_{-1.43}$	[9]
LH,LL or HH	$\gamma\gamma$	$2.67^{+2.41}_{-1.73}$	[10]

Table 1: Summary of the different $t\bar{t}H$ searches in CMS. The first column indicates which top final state is tagged by each analysis, with L (H) indicating a leptonic (hadronic) top-quark decay. Top final states separated by a comma are not distinguished at the analysis level. The second column reports the dominant Higgs decay tag. The third column summarises the best-fit value of the signal strength modifier with its 68% confidence interval.

3. Conclusion

Using the full Run-I data set, the CMS experiment has reached sensitivity to the SM $t\bar{t}H$ production. This major achievement has been made possible by the combination of different analyses optimised for the most relevant (in terms of rate or purity) decay channels of the Higgs boson: $H \to b\bar{b}$, $H \to VV^*$ ($V = Z, W$) with leptonic vector boson decays, $H \to \tau^+\tau^-$ with fully-hadronic or leptonic τ lepton decays, and $H \to \gamma\gamma$. The best-fit value of the signal strength modifier from the combined fit is $\hat{\mu} = 2.76^{+1.05}_{-0.92}$. The observation of a larger rate of signal-like events in the $H \to$ leptons and $H \to$ photons channel drives the measured excess compared to the SM expectation. A new $t\bar{t}H$ analysis in the $H \to b\bar{b}$ decay channel that uses a matrix element technique has been deployed on the 8 TeV data. This analysis has proved to be performing, with up to 30% improvement compared to previously published results, thus providing a promising technology for Run-II. Finally, a search for single-top plus Higgs production in the $H \to \gamma\gamma$ channel has been carried out, yielding results compatible with the SM expectation.

References

[1] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LH", *Phys.Lett.B* (2012), doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235

[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* (2012), doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214

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

[4] G. Bordes and B. van Eijk : "On the associate production of a neutral intermediate-mass Higgs boson with a single top quark at the LHC and SSC', *Phys. Lett. B* **299**, 315 (1993).

[5] CMS Collaboration, "The CMS experiment at the CERN LHC", *JHEP* **03** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[6] 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.

[7] 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).

[8] CMS Collaboration, "Search for ttH events in the H to bb final state using the Matrix Element Method", CMS Physics Analysis Summary CMS-PAS-HIG-14-010 (2014).

[9] CMS Collaboration, "Search for the SM Higgs produced in association with top quarks and decaying to leptons", CMS Physics Analysis Summary CMS-PAS-HIG-13-020 (2013).

[10] CMS Collaboration, "Observation of the diphoton decay of the Higgs boson and measurement of its properties", submitted to Eur. Phys. J. **C**.

[11] CMS Collaboration, "Search for associated production of a single top quark and a Higgs boson in events where the Higgs boson decays to two photons at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-HIG-14-001 (2014).

[12] CMS Collaboration, "Precise determination of the mass of the Higgs boson and studies of compatibility of its couplings with the Standard Model", CMS Physics Analysis Summary CMS-PAS-HIG-14-009 (2014).

[13] CMS 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.

[14] F. Cascioli and others, "Scattering Amplitudes with Open-Loops", *Phys. Rev. Lett.* **108** 111601 (2009)