IL NUOVO CIMENTO Vol. 36 C, N. 6 Novembre-Dicembre 2013
DOI 10.1393/ncc/i2014-11626-5

Colloquia: LaThuile13

Search for SM Higgs boson in the WH production at CMS with $H \to \tau\tau$ final state

R. Venditti

Dipartimento Interateneo di Fisica "M. Merlin" and INFN, Sezione di Bari - Bari, Italy

ricevuto il 20 Giugno 2013; approvato l'1 Luglio 2013

Summary. — The Standard Model (SM) Higgs boson is mainly produced from gluon-gluon and vector boson fusion at LHC. The associated production with vector bosons, although with a lower cross section, can be also considered a sensitive channel because a signicant background rejection can be achieved using the presence of highly energetic charged leptons coming from the decays of W/Z. In the light mass region, the SM Higgs boson decay into τ-lepton pairs has the second highest branching ratio, after the decay into $b\bar{b}$. For these reasons, a search for WH process is performed, in which the W boson decays into muon or electron, and the Higgs boson into τ pair, both decaying hadronically. The analysis is based on data from proton-proton collisions collected with CMS detector in 2011 and 2012 corresponding to an integrated luminosity of $4.9\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 7\,\mathrm{TeV}$ and $19.5\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 8\,\mathrm{TeV}$ respectively. A data-driven technique, the fake rate method, has been used for background estimation. The results are consistent with the expected SM background, so upper limits are set at 95% CL for the SM Higgs boson production cross section.

PACS 07.05.Hd – Data acquisition: hardware and software.
PACS 07.05.Kf – Data analysis: algorithms and implementation; data management.
PACS 29.85.fJ – Data analysis.

1. – Introduction

At the LHC, the SM Higgs boson is mainly produced from gluon-gluon and vector boson fusion processes. The production in association with vector bosons ($q\bar{q} \to W^{\pm}H^0$), although characterized by one order of magnitude lower cross section, provide promising channels for Higgs boson searches at the LHC because a significant background rejection can be achieved using the presence of the additional highly energetic leptons in the event coming from the $W^{\pm}$ decays.

In the light mass region ($M_H < 130\,\mathrm{GeV}/c^2$), the dominant Higgs boson decays are into pairs of b-quarks and τ-leptons. Detecting inclusive production using b or τ decays is difficult due to the overwhelming multi-jet background processes. The associated

production process, where the Higgs boson is produced together with a $W^{\pm}$ boson, can be considered in this case because the presence of the additional highly energetic lepton in the event coming from the vector boson decay provides additional handles to suppress the SM background. Therefore, this analysis focuses on exploiting this channel to search for a low mass SM Higgs boson produced in association with $W^{\pm}$, where W decays into a light lepton (electron or muon) and Higgs decays into a pair of τ-leptons decaying hadronically (in the following indicated with τ_h) [1].

The analysis is performed on the 2011 and 2012 CMS [2] dataset that corresponds to a total integrated luminosity of $24\,\mathrm{fb}^{-1}$.

2. – Selections

Candidate WH events are selected with trigger paths which require either one isolated muon or one electron and one tau depending on the channel. All final leptons are required to be associated to the reconstructed primary vertex with highest transvers momentum (p_T) sum. At offline level requirements on minimum transverse momentum and pseudorapidity, identification and isolation are applied.

The τ_{had} identification is performed using the Hadrons Plus Strips algorithm in which one or three charged hadrons are combined with photons to reconstruct τ decay modes individually [3].

Then only events with at least one pair of opposite charge hadronic tau candidates are preserved for further analysis. In order to reduce the contribution of $t\bar{t}$+jets background, a veto is put for events where a jet tagged as coming from a b-quark is found.

In both channels the final state is characterized by a certain amount of missing energy in the transverse plane coming from neutrinos contribution. Thus only events with $E_T^{miss} \geq 20\,\mathrm{GeV}$ are preserved for further analysis

The remaining topological selections are specific to each of the two channels considered in this analysis and are described separately in what follows.

$\mu\tau\tau$ Final State. –

– $Z \to \mu\mu$ veto: A $Z/\gamma^* \to \mu^+\mu^-$ process accompanied by additional jet production can mimic the signal events. This type of contributions can be removed with a veto against the additional muon. If a second muon is found with $p_T > 15\,\mathrm{GeV}/c$ and $|\eta| < 2.1$ coming from the same primary vertex as the leading τ ($|z_{\tau_{lead}} - z_\mu| \leq 0.14$ cm) the event is discarded.

– Electron veto: For similar reasons, if an electron with $p_T > 10\,\mathrm{GeV}/c$ which falls in the acceptance of the electromagnetic calorimeter and passing the tight electron identification criteria with $|z_{\tau_{lead}} - z_e| \leq 0.14$ cm is found, the event is rejected.

– $Z \to \tau\tau$ veto: An event is discarded if it contains a muon and τ_h candidates of opposite charge signs that have an invariant mass $m^{vis}_{\tau_h^{OS}\mu} < 80\,\mathrm{GeV}/c^2$ and if the p_T of the di-tau system is less than $50\,\mathrm{GeV}/c$.

– Overlap removal: The transverse mass M_T of the muon and the missing transverse energy vector is required to be greater than 20 GeV. This cut makes sure that there is no event overlap with the $\mu\tau_h$ channel of the $H \to \tau\tau$ analysis in CMS [4] and therefore allows the two to be easily combined.

$e\tau\tau$ Final state. –

– $Z \to ee$ veto: $Z \to ee$ events are rejected if there are two opposite sign electrons in the event whose invariant mass falls within the window of $|M(e_1, e_2) - M_Z| < 25\,\text{GeV}$ around the nominal Z mass. Furthermore, the event is rejected if the invariant mass of the selected electron candidate and the hadronic tau with opposite sign falls within the window of $|M(e, \tau) - M_Z| \leq 6\,\text{GeV}$. This requirement recovers events where an electron has been misidentified as a hadronic τ decay. Finally, events where the selected electron candidate and the opposite sign hadronic tau are separated in ΔR by less than 0.01 are rejected.

– $Z \to \tau\tau$ veto: If the value of the transverse mass of the identified electron and E_T^{miss}, $M_\text{T}\left(e, E_T^{miss}\right)$, is less than $50\,\text{GeV}/c^2$ the event is discarded.

3. – Background estimation

The backgrounds events surviving all selections of this analysis can be classified into two categories. The irreducible background comes from WZ and ZZ events which contain three isolated leptons in the final state. The reducible backgrounds contain at least one quark or gluon jet which is incorrectly identified as an isolated e, μ or τ_h.

The irreducible ZZ and WZ backgrounds are estimated using PYTHIA Monte Carlo simulations, and normalized using the NLO theoretical prediction. The main sources of reducible background events are due to W + jets and Z + jets processes, where at least one of the jets or their constituents are misidentified as isolated leptons.

The misidentification probabilities are driven by the performance of the jet fragmentation models in describing rare fluctuations in the regime that is far from the design limits of applicability of such models. Even though the simulation predictions tend to be fairly accurate, relying on simulation in predicting the misidentification probabilities is no prudent. Instead, we use a data-driven approach based on the fake rate calculation.

The Z + jets events satisfying all analysis selections fall into two categories. One contribution stems from the $Z \to \tau_h l$ + jets events, where a recoil jet fakes one of the two hadronic tau candidates. Another contribution is due to the $Z \to ee$ + jets events, where one of the electrons is misidentified as one tau and a recoil jet is misidentified as the second tau. Note that due to the requirement that the two tau candidates are oppositely charged, in both of these two cases the tau candidate stemming from jet misidentification has the same electrical charge as the lepton. Thus, in these events the fake object is the one which has the same charge as the lepton candidate. W + jets events predominantly pass analysis selections when W decays to a light lepton and two recoil jets are misidentified as tau candidates. Note that in this case the tau candidate with the same charge as the lepton is likely a misidentified jet (although, the other tau is also likely a fake). We will utilize this property in defining the method for estimating the background contributions of these processes. Contributions from the processes with a misidentified light lepton are small.

The method is applied as follows:

– A background enriched region is selected. In this region the probability $f(p_T)$ for a jet to pass the final object criteria, parametrized as a function of the p_T of the jet, is measured.

– The fake background estimation is performed by defining a "side-band" region by selecting events with all standard selections except that the tau candidate which

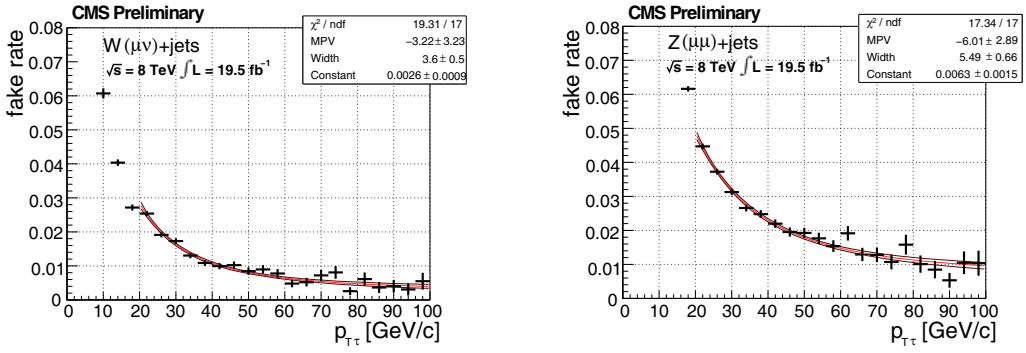

Fig. 1. – Jet-to-leading-tau misidentification rates in the muon channel in 8 TeV, in $W \to \mu\nu$+jets events (left) and $Z \to \mu\mu$ + jets (right), *versus* τ_h p_T for τ_h candidates with $p_T \geq 20\,\mathrm{GeV}/c$.

has the same charge as the light lepton is not required to be isolated (it is the fakeable object).

– Each event in the "side-band" region is weighted by the corrected probability $p = f(p_T)/(1 - f(p_T))$. The resulting weighted spectrum provides the estimate for the expected background contribution in the signal region due to jets misidentified as tau candidates.

The Fake Rate is measured as a function of jet p_T using two separate samples, one enriched with $W \to \mu\nu$ + jets and the other with $Z \to \mu\mu$ + jets events, which are used to evaluate the systematic uncertainty for the measured fake rate. With the exception of the isolation, the τ_h candidates must fulfill the same requirements as for the direct analysis.

The Fake Rates are measured individually in the 2011 and the 2012 dataset and for the leading and sub-leading τ_h candidates in the two channels. All available tau candidates in an event are used in both samples and the fake rate is measured per fakeable object. The measured fake rate distributions as a function of p_t are fit with the Landau function with peak position and width as free parameters and an additive constant.

As the Fake Rates in the two regions are different due to a different fraction of quark-induced and gluon-induced jets, we will use a weighted average of the two Fake Rate measurements determined in W + jets and Z + jets samples. Figure 1 show the Fake Rate functions determined in 2012 data in the two measurement regions.

4. – Systematics

The systematics that have been considered in the analysis are summarized in table I. The major systematic effect is due to the fake rate normalization that is calculated taking into account different kinds of contributions. The first one is coming from the statistical uncertainty of the fit of the measured fake rate as a function of p_T. To account for the uncertainty of the fit and its systematic effect on the analysis an uncertainty of 10% has been determined by propagating the error on the fit parameter to the predicted number of background events. Another source of systematic uncertainty is the difference between the calculations of the fake rate function using the W + jets and Z + jets events. The

TABLE I. – *Summary of the systematic uncertainties that are taken into account as nuisance parameters into the final fit.*

Systematic uncertainty	Affected sample	Value
Luminosity	Simulation	2.2% (2011) or 4.4% (2012)
Parton Distribution Functions	Simulation	4.3%
Trigger Efficiency	Simulation	3.5%
Electron ID Efficiency	Simulation	2.9%
Muon ID Efficiency	Simulation	1.4%
Tau ID Efficiency	Simulation	12%
Tau Energy Scale	Simulation	3%
MET Energy Scale	Simulation	3.7%
Additional Electron Veto	Simulation	3.8%
Additional Muon Veto	Simulation	0.7%
Fake rate normalization	Fake background estimate	20%

two functions are different, because of the different fraction of quark-induced and gluon-induced jets. Thus a 10% is attributed for the difference between the two regions when requiring two jets in both of them. Finally a 10% is attributed for the determination of the weights when averaging the Fake Rate function measured in a $W+$jets or a $Z+$jets enriched region. By combining this three contributions, assumed independent one from the other, one obtains the conservative value of 20% quoted in table I. The second main source of uncertainty comes from the tau identification systematic uncertainty measured in [3] that is 6%. In a conservative approach, considering that the two selected taus are correlated, the total systematic uncertainty due to tau identification is 12%. Moreover bin by bin shape uncertainties on the shape of the visible mass distribution used as input for the limit computation have been considered. The statistical uncertainty on the fake rate and irreducible background can lead to differences in the shape of the background distribution. To estimate the systematic influence on the exclusion limits different shapes are used where each bin in the fake rate, WZ and ZZ visible mass distribution is scaled up and down by the statistical error individually. The exclusion limit is calculated for each shape which results in an additional systematic uncertainty which is taken into account during the limit calculation.

5. – Results

After all selections a total of 36 events are observed in both channels in the whole analyzed statistics. The prediction from simulation indicates that the background enriched region is dominated by $W+$jets, $Z\to\tau\tau$, and $t\bar{t}$ events, and agrees well with the observed data. In fig. 2 is shown the distribution of the visible mass of the di-tau system in the two channels after the best fit to data that takes into account all the systematic uncertainties listed in the previous section. The observed yields and background estimates are presented in table II.

The exclusion limits for the $W^{\pm}H^{0}$ associated production are calculated using the signal shape of the SM Higgs extracted from MC and the shapes corresponding to various background production processes.

The 95% confidence level (CL) upper limit and its $\pm1\sigma$ and $\pm2\sigma$ uncertainty regions on the rate of the signal with respect to the SM cross section are calculated with the

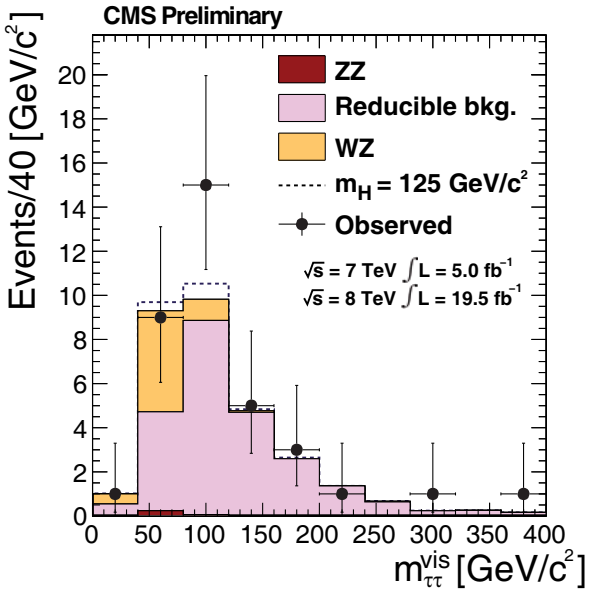

Fig. 2. – Observed and expected Higgs boson candidate mass spectra. The expected contribution from the associated production of a SM Higgs boson with mass $m_H = 125\,\mathrm{GeV}/c^2$ is shown by the dashed line. The yields of each process are determined using the same maximum-likelihood fit to a signal-plus-background hypothesis used in the limit setting procedure.

Asymptotic CL_s algorithm [5] using the event count only (for 7 TeV data) or the visible mass of the selected di-tau pair (for 8 TeV data). The reason for using a counting experiment approach for 7 TeV is that there are so few data events observed that no di-tau mass shape can sensibly be extracted from the data events. The algorithm uses a frequentist statistical test where a hypothesis with only background processes is tested against the model with background plus signal. In this algorithm the shape of the di-tau pair visible mass and its uncertainty is considered together with the systematic uncertainties on normalization, introduced as nuisance parameters. The exclusion limits obtained for the two channels are shown in fig. 3. These results have been combined with

TABLE II. – *Yields for the 2011 and 2012 analyzed data and the estimated and measured background contributing processes in the* $\mu\tau\tau$ *and* $e\tau\tau$ *channels. Signal events contain all contributions from* WH, ZH, $t\bar{t}H$, *but also from VBF and gluon-gluon fusion. The systematic uncertainty is not included*

Process	$l\tau\tau$
Fakes	20.4 ± 4.3
WZ	6.2 ± 1
ZZ	0.38 ± 0.06
Total Background	27.1 ± 4.5
WH, $H \to \tau\tau$ ($m_H = 125\,\mathrm{GeV}$)	1.2 ± 0.2
Observed	36

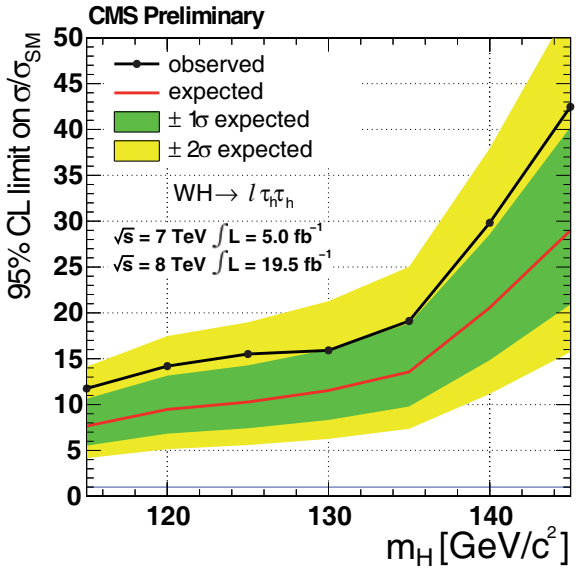

Fig. 3. – Observed and median expected 95% CL upper limits on SM Higgs production set by $l\tau\tau$ channels.

the ones obtained in the other channels of associated production process (ZH, $WH \to ll\tau$) [1]. The final combination is shown in fig. 4 in the background only hypothesis and with Higgs signal injected. The data are compatible with both the background-only prediction and the presence of a SM Higgs boson. Upper limits of 2.9 to 4.6 times the predicted SM value are set at 95% CL for the product of the SM Higgs boson production cross section and decay branching fraction in the mass range $110 \leq m_H \leq 145\,\mathrm{GeV}/c^2$.

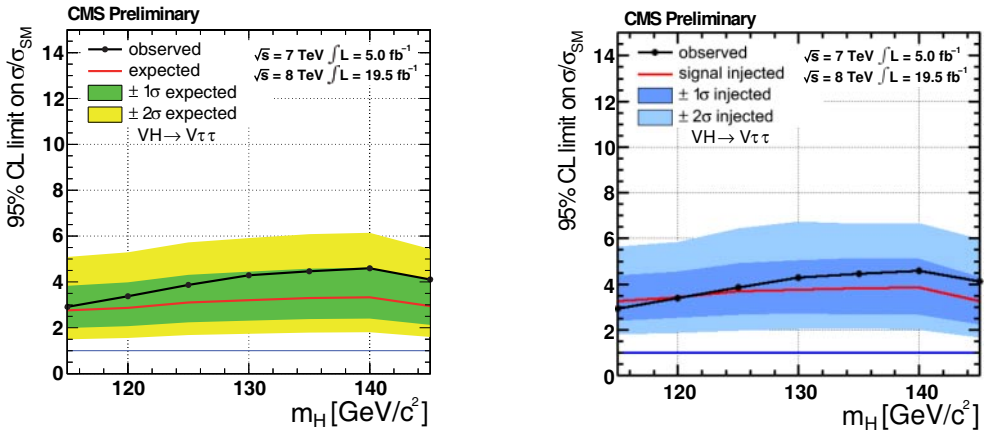

Fig. 4. – Observed and median expected 95% CL upper limits on SM Higgs production set by $ZH \to ll\tau\tau$ and $W \to l\tau\tau$ channels in the background-only hypothesis (left) and with Higgs signal injected (right).

REFERENCES

[1] CMS Collaboration, "Search for the standard model Higgs boson decaying to tau pairs produced in association with a W or Z boson", *CMS-PAS-HIG-12-053*
[2] CMS Collaboration, *JINST*, **3** (2008) S08004.
[3] CMS Collaboration, "Tau identification in CMS", *CMS-PAS-TAU-11-001*, 2-3.
[4] CMS Collaboration, "Search for the standard model Higgs boson decaying to tau pairs", *CMS-PAS-HIG-12-043*, 5.
[5] Cowan G. *et al.*, *Eur. Phys. J. C*, **71** (2011) 1.