

Search for the SM Higgs Boson in Di- τ Final States at CMS

Roger Wolf, for the CMS Collaboration^{1,a}

¹Massachusetts Institute of Technology

Abstract. A search for the standard model Higgs boson decaying to tau pairs is performed using events recorded by the CMS experiment at the LHC in 2011 and 2012 at a center-of-mass energy of 7 and 8 TeV respectively. The dataset corresponds to an integrated luminosity of 17 fb⁻¹, split in 4.9 fb⁻¹ of data taken at 7 TeV center-of-mass energy and 12.1 fb⁻¹ at 8 TeV. The tau-pair invariant-mass spectrum is studied in five different final states, $\mu\tau_h + X$, $e\tau_h + X$, $e\mu + X$, $\tau_h\tau_h + X$, and $\mu\mu + X$. Upper limits with respect to the standard model prediction in the mass range of 110-145 GeV are determined. An observed (expected) 95% confidence level exclusion limit for $m_H = 125$ GeV is found to be 1.63 (1.00) times the standard model cross section.

1 Introduction

On July 4, 2012, the discovery of a new boson, with mass around 125 GeV and with properties compatible with those of a standard-model Higgs boson, was announced at CERN by the ATLAS and CMS collaboration [1, 2]. The reported excess is most significant in the SM Higgs searches using the decay modes into $\gamma\gamma$ and ZZ . The results in the $\tau\tau$ decay mode showed no excess of observed events in the mass range near 125 GeV, still compatible with both, a downward fluctuation from a background-only or background plus SM Higgs boson hypothesis. In this document a search for the SM Higgs boson is reported using final states with pairs of τ leptons in proton-proton collisions at $\sqrt{s}=7$ and 8 TeV at the LHC using the data that have been collected in 2011 and 2012 corresponding to an integrated luminosity of 17 fb⁻¹ recorded by the CMS experiment. This luminosity splits in 4.9 fb⁻¹ of data taken at 7 TeV center-of-mass energy and 12.1 fb⁻¹ at 8 TeV.

Five independent τ pair final states are studied: $\mu\tau_h + X$, $e\tau_h + X$, $e\mu + X$, $\mu\mu + X$, and $\tau_h\tau_h + X$, where τ_h denotes a reconstructed hadronic decay of a τ lepton. In each channel, the signal is separated from the background using the fully reconstructed invariant mass of the selected pair of τ -leptons as described in section 5. The events are classified by the number of additional jets in the final state to improve the ratio of signal to background events and to enhance the contribution of different Higgs boson production mechanisms. A more detailed description of this analysis can be found in [3]. The analysis that will be described in this document has been combined with an complementary SM Higgs boson search in the exclusive production channel in association with W or Z bosons, which can be found documented in [4]. The combined limit shown in section 6 includes this dedicated complementary search.

^ae-mail: roger.wolf@cern.ch

2 Event reconstruction and selection

The information of all sub-detectors of CMS is reconstructed according to a particle-flow (PF) algorithm [5–7], resulting in a complete list of reconstructed particles emerging from each collision. These particles are used to identify the decay products of the di- τ pair, isolated muons, electrons and τ_h -leptons and to define particle jets and the missing transverse energy in the event.

The events for this analysis have been online recorded requiring a combination of electron, muon and tau trigger objects [8–10] depending on the exact decay channel. The identification and isolation criteria and the transverse momentum thresholds for these objects have progressively been tightened as the LHC instantaneous luminosity increased over the data-taking periods. In the $\tau_h\tau_h + X$ channel, the trigger fake rate has been kept under control by requiring two τ_h trigger objects of $p_T > 30$ GeV, together with a jet trigger object of $p_T > 30$ GeV and $|\eta| < 3.1$, reconstructed using the particle flow algorithm on trigger level.

Electrons, muons and τ_h -leptons are required to be isolated, to have a p_T that exceeds the corresponding trigger thresholds and to be well contained in the fiducial volume of the detector: in the $e\tau_h + X$ and $\mu\tau_h + X$ channels, the p_T of the selected electron (muon) is required to be above 20 (17) GeV. The absolute value of η is required to be smaller than 2.1. The electron (muon) is required to be accompanied by an oppositely charged τ_h of $p_T > 20$ GeV with $|\eta| < 2.3$. In the 2012 dataset analysis, the electron and muon p_T thresholds have been increased to 24 GeV and 20 GeV, respectively, to account for higher trigger thresholds. In the $e\mu + X$ channel, events are required to contain an electron with $|\eta| < 2.3$ and an oppositely charged muon within $|\eta| < 2.1$, requiring $p_T > 20$ GeV for the leading and $p_T > 10$ GeV for the sub-leading lepton. In the $\tau_h\tau_h + X$

channel, both τ_h are required to have $p_T > 45$ GeV and $|\eta| < 2.1$.

Hadronically-decaying τ -leptons (τ_h) are reconstructed and identified using the hadron-plus-strips algorithm [11], selecting candidates with three charged hadrons or with one charged hadron and up to two neutral pions. The neutral pions are identified by clustering reconstructed photons in strips along the ϕ direction, taking possible photon conversions in the tracker material into account.

The overlap between the various channels, is eliminated by rejecting from the $e\tau_h + X$ ($\mu\tau_h + X$) channel events with an additional muon (electron), and by rejecting leptonic events from the $\tau_h\tau_h + X$ channel.

For the characterization of the rest of the event jets are used. These are reconstructed from all reconstructed particles using the anti- k_T jet algorithm [12, 13] with a distance parameter of $D = 0.5$. A residual calibration factor is applied to the jet energy to account for imperfections in the neutral hadron calibration, the jet energy containment, and an estimation of the contribution of pile-up and underlying event particles. Jets from pile-up are identified and removed using a TMVA BDT taking into account input variables like the momentum and spatial distribution of the jet particles, charged and neutral particle multiplicities and the compatibility of the charged hadrons in the jet with the reconstructed primary vertex of the hard interaction. For the event categorization jets with $|\eta| < 4.7$ are taken into account.

In the $e\mu + X$ channel events with jets originating from b -quarks are vetoed in the further event classification. These jets are identified using a Combined Secondary Vertex (CSV) b -tagging algorithm that combines track impact parameter and secondary vertex information in a single likelihood discriminant [14]. Correction factors on the b -tagging efficiency and the misidentification probability in the simulation and systematic uncertainties on these quantities have been obtained, as a function of the p_T and η of the jet, from the comparison of the simulation with data.

The missing transverse energy is used for the topological event selection and in the reconstruction of the di- τ invariant mass. The resolution of the PF missing transverse energy is reconstructed as the negative vectorial sum of the transverse momenta of all particles. In this analysis, a multivariate regression is used to provide a more precise measurement of the MET in the presence of pile-up, taking as input the PF MET itself, as well as different definitions of MET computed from:

- charged hadrons from the primary vertex
- charged hadrons from the primary vertex, and neutral particles in jets passing the pile-up jet identification described above.
- charged hadrons from pile-up vertices and neutral particles in jets that fail the pile-up jet identification.
- charged hadrons from the primary vertex and all neutral particles in the event. The vectorial sum of the transverse momenta of neutral particles within jets failing the

pile-up jet identification is then added to this PF MET definition.

For the average number of 20 reconstructed primary vertices, corresponding to the average pile-up situation in 2012, the resolution of this estimator of the MET is a factor of two better than the one obtained from the naive calculation of MET.

In the $e\tau_h + X$ and $\mu\tau_h + X$ channel the transverse mass

$$m_T = \sqrt{2p_T \cdot \text{MET} \cdot (1 - \cos(\Delta\phi))}$$

is required to be less than 20 GeV, where p_T is the lepton transverse momentum and $\Delta\phi$ is the difference in ϕ between the lepton and the MET vector. In the $e\mu + X$ and the $\mu\mu + X$ channel, instead of a requirement on m_T a requirement on the variable $(p_\zeta^{\text{miss}} - 0.85p_\zeta^{\text{vis}}) > -20$ GeV is imposed, where p_ζ^{miss} and p_ζ^{vis} are the projections of the MET vectors and of the vectorial sum of the two leptons onto the bisector (ζ) of the directions of the leptons in the transverse plane.

3 Event categorization

To further enhance the sensitivity of the search, the selected events are split into exclusive categories based on the jet multiplicity, and on the p_T of the reconstructed τ decay products as listed below:

- **2-Jet (VBF):** In this category, two jets with $p_T > 30$ GeV are required to tag the vector-boson fusion Higgs production process. The two jets are required to have an invariant mass $m_{jj} > 500$ GeV and to be separated in pseudo-rapidity by $\Delta\eta > 3.5$. A rapidity gap is defined by requiring no reconstructed jet with $p_T > 30$ GeV between the two tagging jets. In the $e\mu + X$ channel, the large background contribution from $t\bar{t}$ events is suppressed by rejecting events with a b -tagged jet with $p_T > 20$ GeV.
- **1-Jet:** In this category at least one jet is required with $p_T > 30$ GeV. It is required not to be part of the VBF event category, and not to contain any b -tagged jet with $p_T > 20$ GeV. In the $e\tau_h + X$ channel, the large background from $Z \rightarrow ee + \text{jets}$ events in which an electron is misidentified as τ_h is further suppressed by requiring an additional requirement of $\text{MET} > 30$ GeV.
- **0-Jet:** This category contains all events with no jet with $p_T > 30$ GeV, and no b -tagged jet with $p_T > 20$ GeV.

The 1-Jet and 0-Jet categories are split in two bins of p_T of the τ -lepton. In the $e\tau_{+X}$ and $\mu\tau_h + X$ channel the threshold is at $p_T > 40$ GeV of the τ_h . In the $e\mu + X$ channel the threshold is 35 GeV on the muon and in the $\mu\mu + X$ channel the threshold is 30 GeV on the leading muon in the event. The 0-Jet categories are only used to constrain background normalizations, identification efficiencies, and energy scales.

In the $\tau_h\tau_h + X$ channel, a leading jet with $p_T > 50$ GeV and $|\eta| < 2.1$ is required to match the trigger requirement, and two categories are defined:

- **2-Jet (VBF):** This category is required to contain a sub-leading jet with $p_T > 30$ GeV in addition to the leading jet. The two-jet system is required to have an invariant di-jet mass of $m_{jj} > 250$ GeV and the two leading jets are required to be separated by $\Delta\eta > 2.5$ with no other jet with $p_T > 30$ GeV in between. Multijet background is reduced by requiring the p_T of the system formed by the two τ_h -leptons and of the MET in the event, $p_{T,H}$, to be larger than 110 GeV.
- **1-Jet:** In this category the jet requirement is fulfilled already by the inclusive selection in the $\tau_h\tau_h+X$ channel. Multijet background is reduced by requiring $p_{T,H} > 140$ GeV.

4 Estimation of backgrounds

The signal is compromised by various sources of known background processes. The largest and irreducible source of background events originates from $Z \rightarrow \tau\tau$ events, which is estimated directly from data using $Z \rightarrow \mu\mu$ events, where the reconstructed muons are replaced by the reconstructed particles from simulated τ decays (embedding). The normalization for this process is determined from the measurement of the $Z \rightarrow \mu\mu$ yield in data. Other significant sources of background originate from QCD multijet events in which there is one jet misidentified as an isolated electron or muon, and a second jet is misidentified as τ_h and from W +jets events in which there is a jet misidentified as a τ_h . The rates for these processes are estimated using the number of observed same-charge τ -pair events, and from events with large transverse mass, respectively. For the $\tau_h\tau_h + X$ channel, the shape of the multijet background is estimated from a complementary sample of events with opposite-sign charge where the isolation requirement on one of the two τ_h -leptons has been relaxed. The contributions from non-multijet backgrounds are estimated from simulation and subtracted from this control region. The resulting distribution in this control region is scaled to the signal region using scale factors from the same signal and control region in an independent control sample where the τ_h -leptons have the same charge.

Other background processes include $t\bar{t}$ production and $Z \rightarrow ee/\mu\mu$ events and di-boson production.

To model the SM Higgs boson signals the event generators PYTHIA and POWHEG [15] are used. The TAUOLA [?] software package is used to simulate the decay of τ -leptons in all cases. Additional next-to-next-to-leading order (NNLO) k -factors from FEHIPRO [17, 18] are applied to the Higgs boson p_T spectrum of Higgs boson events produced via gluon gluon fusion for samples produced at $\sqrt{s}=7$ TeV. Samples produced at $\sqrt{s}=8$ TeV use an improved version of POWHEG which shows good agreement in the Higgs boson p_T spectrum with NNLO calculations.

The presence of pile-up is incorporated into the simulation by additional interactions, reweighting the simulated events to match the pile-up situation observed in data. The background samples, which are obtained from data by

definition contain the correct distribution of pile-up interactions. For the simulated samples the missing transverse energy response is corrected using a prescription, based on data, where Z bosons are reconstructed in the di- μ channel, and the missing transverse energy scale and resolution are calibrated as a function of the p_T of the Z boson. All background yields are input to a maximum likelihood fit for the signal extraction, as described below.

5 Di- τ mass reconstruction

The $H \rightarrow \tau\tau$ signal is extracted and separated from background processes with the help of a full reconstruction of the invariant mass of the di- τ system $m_{\tau\tau}$ utilizing a maximum likelihood estimate based on the expected phase space of the hadronic (leptonic) decay of the two corresponding τ -leptons. Input variables of this likelihood are the four momenta of the visible parts of the two τ -leptons, the MET and the covariance matrix of the MET resolution, that is extracted on an event by event basis. Unconstrained parameters are the decay angle θ_i^* in the rest frame of the correspondingly decaying τ -lepton and the invariant mass $m_{\nu\nu,i}$ of the neutrino pair in the case of a leptonic decay.

6 Results

An estimate on the SM Higgs boson signal strength as a function of the Higgs boson mass is extracted from a simultaneous maximum likelihood fit to the binned distributions of the fully reconstructed invariant mass of the di- τ system, $m_{\tau\tau}$ in all channels and all event categories, including correlated and un-correlated shape altering and normalization uncertainties on individual background processes, reconstruction efficiencies and τ_h , electron and jet energy scale as additional nuisance parameters. All nuisance parameters are profiled in the fitting procedure. Shape altering uncertainties are introduced by a vertical template morphing technique. Including the statistical uncertainties for weakly populated bins in individual background templates more than 550 individual nuisance parameters are profiled during the fitting procedure. The most important uncertainties correspond to the τ_h identification efficiency (6%), the τ_h energy scale (3% introduced as a shape altering uncertainty) and the uncertainty on the normalization of the background from multijet W +jet events in the $e\tau+X$ and $\mu\tau+X$ channel (10-30% depending on the event category). The 0-Jet event categories, where no signal is expected are used only for the in situ calibration of the correlated uncertainties, like the τ_h identification efficiency and the τ_h energy scale. The fit is checked for a good quality and the pulls of all nuisance parameters are checked to follow the statistical expectation. The $m_{\tau\tau}$ distributions in the most sensitive event categories, 1-Jet, high p_T and 2-Jet (VBF), in the $\mu\mu + X$, $e\mu + X$, $e\tau + X$ and $\mu\tau + X$ channel, after the fit of the hypothesis of all considered background processes and a SM Higgs boson at 125 GeV, are shown in figures 1 and 2. The expected yield for a SM Higgs boson of 125 GeV is also shown.

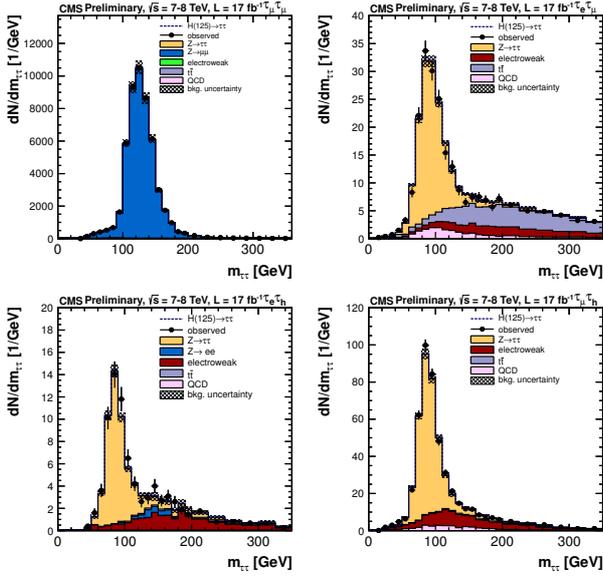


Figure 1. Fully reconstructed invariant mass of the di- τ system in the 1-Jet high p_T event category. Shown are invariant mass distributions of the (upper left) $\mu\mu+X$, (upper right) $e\mu+X$, (lower left) $e\tau_h+X$, and (lower right) $\mu\tau_h+X$ decay channel. The fully reconstructed invariant mass of the di- τ system in the 1-Jet event category in the $\tau_h\tau_h+X$ decay channel is shown in figure 3 (left).

The shaded band corresponds to the remaining uncertainties on the background estimate after the fit. To reduce the number of figures the individual datasets for 7 TeV and 8 TeV have been combined into one histogram. The $m_{\tau\tau}$ distributions in the two event categories in the $\tau_h\tau_h+X$ channel are shown in figure 3.

The shown data are combined with a dedicated search for the SM Higgs boson in the production mode in association with W or Z bosons, which is documented in [4]. The fit results in a signal strength of 0.7 ± 0.5 for a SM Higgs boson with a mass of 125 GeV. This result is translated into a 95% CL upper limit in the signal strength for a SM Higgs boson as function of the Higgs boson mass using the modified asymptotic CL_S method. The limit of all channels and all event categories combined is shown in figure 4. The red line indicates the expected limit for a background-only hypothesis, the shaded bands indicate the one and two σ uncertainties on the expected limit. The black lines corresponds to the observed limit, revealing a slight excess over the expectation, that is still compatible with the background-only hypothesis.

The sensitivity of each contributing channel and event category, in terms of the expected 95% CL upper limit on the signal strength of the SM Higgs boson as function of the Higgs mass, are shown in figure 5. For the breakup in channels all event categories in each corresponding decay channel have been combined. For the breakup in event categories all contributing decay channels, as well as the low and high p_T sub-categories have been combined and the constraint in the 0-Jet event categories has been taken into account.

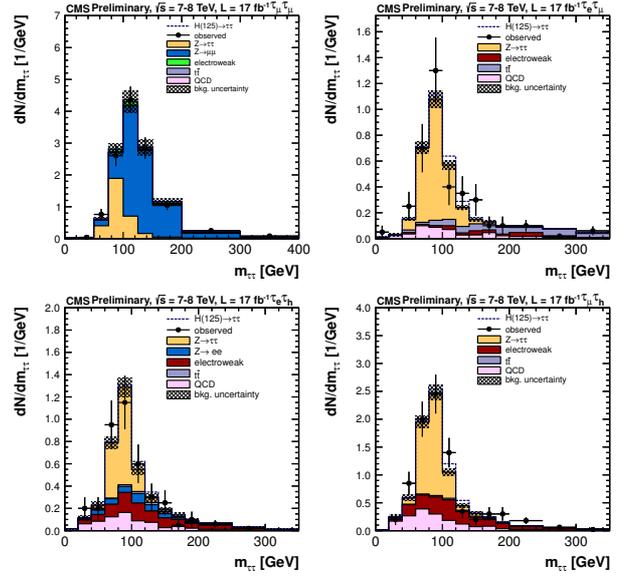


Figure 2. Fully reconstructed invariant mass of the di- τ system in the 2-Jet (VBF) event category. Shown are invariant mass distributions of the (upper left) $\mu\mu$, (upper right) $e\mu$, (lower left) $e\tau_h$, and (lower right) $\mu\tau_h$ decay channel. The fully reconstructed invariant mass of the di- τ system in the 2-Jet (VBF) event category in the $\tau_h\tau_h$ decay channel is shown in figure 3 (right).

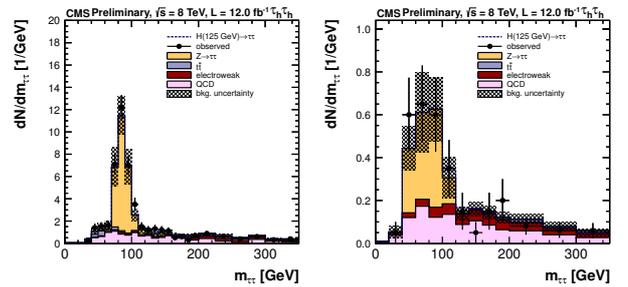


Figure 3. Fully reconstructed invariant mass of the di- τ system in the $\tau_h\tau_h$ decay channel. Shown are the invariant mass distributions in the (left) 1-Jet and (right) the 2-Jet (VBF) event category.

Also shown in figure 5 are the (lower left) the observed limit together with the unbiased expectation for a SM Higgs boson with a mass of 125 GeV. The expectation is shown by the red line. The shaded bands correspond to the one and two σ uncertainty on the expectation. This expectation and the uncertainties have been obtained from 5000 independent toy experiments based on the Asimov dataset for a SM Higgs boson with signal strength one and a mass of 125 GeV at each mass point. The interpretation of this figure should take into account that the expectation is unbiased, i.e. the nuisance parameters in the modified asymptotic CL_S method have not been fitted to the data, but to the corresponding pseudo-dataset for each toy. The expectation has been obtained from the median and the uncertainties from the corresponding quantiles of the corresponding toy samples.

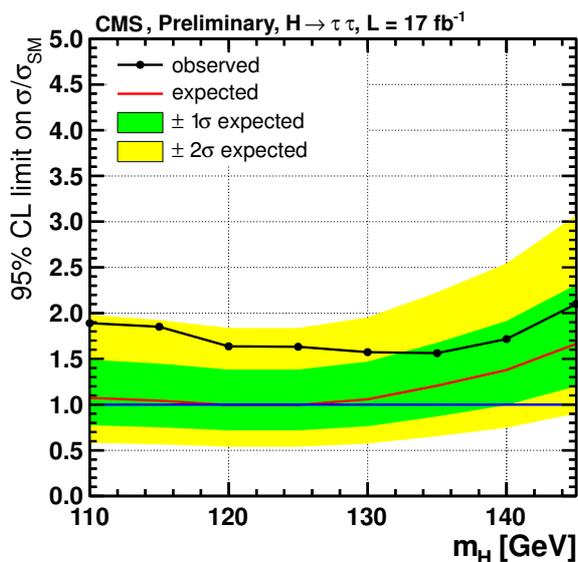


Figure 4. Expected and observed 95% CL upper limit for the observation of a SM Higgs boson for masses between 110 and 145 GeV. The red line indicates the expected limit, the yellow and green band indicate the one and two σ uncertainty on the expected limit. The black points indicate the observed limit.

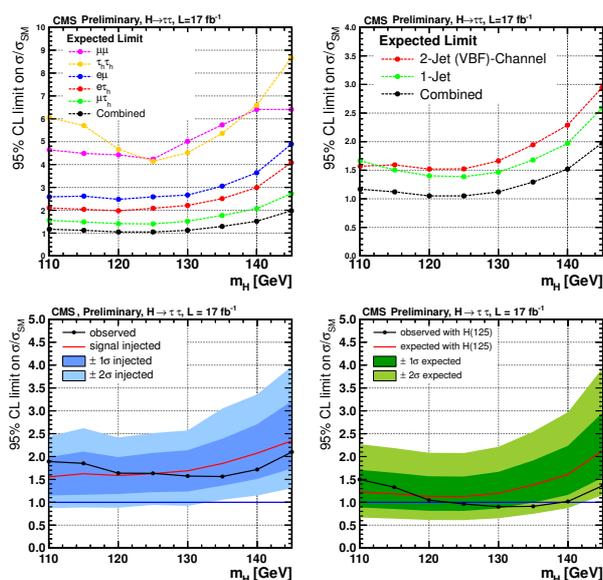


Figure 5. The sensitivity of the presented analysis expressed in terms of the expected limit. Shown are the sensitivity (upper left) split in decay channels and (upper right) split in event categories. The sensitivity per category is shown including the constraint of the background contributions in the 0-Jet event category. The sensitivity of the 1-Jet event category corresponds to the combination of the sub-event categories for high and low p_T of the τ lepton. Also shown are two different representations of the observed and expected limit. Shown is (lower left) a comparison of the observed limit together with the expectation for a SM Higgs boson at $m_H=125$ GeV and (lower right) the expected and observed 95% CL upper limit for the search of a second Higgs boson, considering the SM Higgs boson with $m_H=125$ GeV as background.

In figure 5 (lower right) the 95% CL upper limit for the search for an additional Higgs boson is shown introducing a SM Higgs boson with a mass of 125 GeV as background. Both, figures 5 (lower left) and (lower right) show a very good agreement with the expectation of a SM Higgs boson with a mass of 125 GeV and no excess beyond that. The observed limit at 125 GeV is found to be 1.63 for an expected limit of 1.00 in combination with the search in the exclusive production channel in association with a W or Z boson. The slight excess corresponds to a Bayesian significance of 1.50σ for an expected significance of 2.45σ . It should be noted that this significance has been obtained from a Bayesian approach suppressing negative values of the significance and that especially for low significances this approach can deviate from the significance obtained from a frequentist approach. The observed limit and significant is compatible with a SM Higgs boson at 125 GeV. It is also compatible with the zero signal hypothesis.

References

- [1] ATLAS Collaboration Collaboration, Phys.Lett. **B**, 710, (2012).
- [2] CMS Collaboration Collaboration, Phys.Lett. **B**, 716, (2012).
- [3] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-HIG-12-043, (2012).
- [4] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-HIG-12-032, (2012).
- [5] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, (2009).
- [6] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-PFT-10-002, (2010).
- [7] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-PFT-10-003, (2010).
- [8] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-EGM-10-004, (2010).
- [9] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-MUO-10-002, (2010).
- [10] CMS Collaboration, JHEP **8**, 117, (2011).
- [11] CMS Collaboration, JINST **7** P01001, (2012).
- [12] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04** 063, (2008).
- [13] CMS Collaboration, JINST **6** 11002, (2011).
- [14] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-BTV-12-001 (2012).
- [15] P. Nason and C. Oleari, JHEP **02** 037, (2010).
- [16] S. Jadach, J. H. Kuhn, and Z. Was, Computer Physics Communications **64** 275, (1991).
- [17] C. Anastasiou, K. Melnikov, and F. Petriello, Nucl. Phys. **B 724**, 197, (2005).
- [18] C. Anastasiou, S. Bucherer, and Z. Kunszt, JHEP **10** 068, (2009).