Evidence for the Higgs boson in the $\tau\tau$ decay channel using the CMS detector

Rosamaria Venditti[1,a] on behalf of the CMS Collaboration

[1] *Dipartimento di Fisica M. Merlin, Bari, via Amendola 173*

Abstract. A search for the standard model Higgs boson decaying to τ pair has been performed in proton-proton collision data recorded by CMS detector at LHC at centre of mass energy 8 TeV (7 TeV) corresponding to an integrated luminosity 20 fb^{-1} (5 fb^{-1}). The production modes considered are gluon-gluon fusion, VBF and associated production with a vector boson. The analysis strategy and the resulting evidence for the Higgs boson in tau pair channel are reported.

1 Introduction

On July 4 2012 the discovery of a new boson with a mass around 125 GeV and properties compatible with those of the standard model (SM) Higgs boson, was announced at CERN by the ATLAS and CMS Collaborations [1, 2]. The excess of events was most significant in the bosonic channels but no excess was observed in fermionic ones. The measurement of Yukawa couplings between the Higgs field and the fermionic fields is instead essential to identify this boson as the SM Higgs boson. The $\tau\tau$ decay mode is the most promising because of the large event rate if compared to the $\mu\mu$ decay mode and the smaller contribution from background events with respect to the $\bar{b}b$ decay mode. The search has been performed on the data collected by CMS, corresponding to an integrated luminosity of 5 fb^{-1} (20 fb^{-1}) at $\sqrt{s}$ =7 TeV ($\sqrt{s}$ =8 TeV).

2 CMS Experiment

The central feature of the CMS apparatus is a superconducting solenoid providing a magnetic field of 3.8 T. Within its volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass/scintillator hadron calorimeter. Muons are detected in gas-ionization chambers embedded in the steel flux return yoke outside the solenoid. A more detailed description of the CMS detector can be found in ref. [3]. The CMS experiment uses a right-handed coordinate system, with the origin at the nominal interaction point, the x axis pointing to the centre of the LHC, the y axis pointing up (perpendicular to the LHC plane), and the z axis along the anticlockwise-beam direction. The polar angle θ is measured from the positive z axis and the azimuthal angle ϕ is measured in the transverse (x, y) plane. The pseudorapidity is defined as $\eta = -\log\tan(\theta/2)$. The number of inelastic proton-proton collisions occurring per LHC bunch crossing was, on average, 9(21) in 2011(2012).

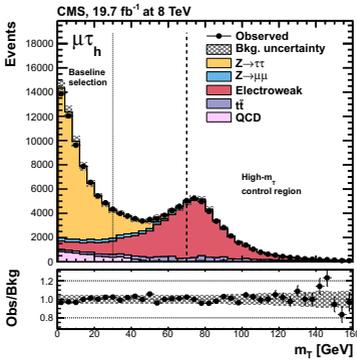[a]e-mail: rosamaria.venditti@cern.ch

Figure 1. Observed and predicted m_T distribution in the 8 TeV $\mu\tau$ analysis after the baseline selection without m_T<30 GeV requirement applied. The dashed line delimits the high-m_T control region that is used to normalize the yield of the W+jets contribution in the analysis as described in the text. The electroweak background contribution includes events from W+jets, diboson, and single-top-quark production. The "bkg. uncertainty" band represents the combined statistical and systematic uncertainty in the background yield in each bin. The bottom inset shows the ratio of the observed and predicted numbers of events.

The additional proton-proton collisions happening in the same bunch crossing are termed pileup (PU). Events are globally reconstructed with a particle-flow algorithm that links information from all of the sub detectors, reconstructing particles into mutually excluding categories [4]. Jets are reconstructed from all particles using the anti-k_T jet clustering algorithm with a 0.5 distance parameter. Jets originating from the hadronization of b quarks are identified using the combined secondary vertex algorithm [5] which exploits observables related to the long lifetime of b hadrons.

3 Baseline Selection

Higgs boson production via VBF and gluon gluon fusion processes is studied by requiring two charged leptons in the final state: $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$, $e\mu$, $\mu\mu$, ee. Here τ_h denotes the hadronic decays of the tau lepton that are reconstructed thanks to the Hadron Plus Strip [6] algorithm. Associated production with W(Z) boson is studied by requiring one (two) extra well-identified and high energetic lepton(s). Kinematic requirement on leptons transverse momentum (termed p_T in the following) and η are applied to match the trigger threshold and detector acceptance. Requirements on lepton isolation, computed by summing transverse momenta of the particles in a cone around the lepton direction after subtracting the PU contribution, are applied to discriminate genuine leptons coming from boson decay or τ decay from those contained in QCD jets. The two leptons assigned to the Higgs boson decay are required to be of opposite charge. The W+jets background is reduced by requiring that the transverse mass m_T of the light lepton and E_T^{miss} system, shown in figure 1, should not exceed 30 GeV. In the $e\mu$ channel, the $t\bar{t}$ background is reduced using a multivariate discriminant built with kinematic variables related to the $e\mu$ system and the E_T^{miss}, the distance of closest approach between the leptons and the primary vertex, and the value of the b-tagging discriminator for the leading jet if any. In WH search with τ fully hadronic final state a multivariate discriminator involving kinematic variables related to the $\tau_h\tau_h$ system and E_T^{miss}, has been trained to suppress the relevant Z+jet, W+jet and QCD background. In the semileptonic channels instead the dominant contribution from $Z \rightarrow ll$ events is reduced by requiring same sign charged light leptons. In the ZH channel the Z + jets events are reduced by requiring high sum of light lepton transverse momenta . In all channels, events containing at least one b-tagged jet with p_T> 20 GeV are rejected to reduce the $t\bar{t}$ background [7].

4 Event categories

The event sample is split into mutually exclusive categories, defined to maximize the sensitivity of the analysis to the presence of a SM Higgs boson with 110 GeV < m_H <145 GeV. In each channel events
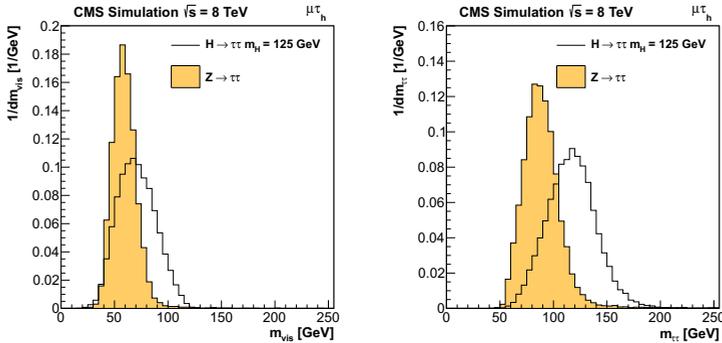
Figure 2. Distributions of (left) the τ-pair invariant mass m_{vis}, and (right) the svfit mass $m_{\tau\tau}$ in the $\mu\tau$ channel.

are classified according to the number of reconstructed jets. In all channels, events from VBF process are selected by requiring two high p_T jets with a large pseudorapidity gap and a large invariant mass. In this way events from the dominant $Z \to \tau\tau$ contribution are suppressed because the VBF tagging rejects the gluon-initiated jets from initial state radiation in the Drell-Yan process. Events failing the VBF tag requirements are collected in the 1-jet category if they contain at least one jet, and in the 0-jet category otherwise. The latter has low sensitivity to the presence of a SM Higgs boson and is mainly used to constrain the $Z \to \tau\tau$ background for the more sensitive categories. The 1-jet and VBF categories are further split according to the transverse momentum of the Higgs boson candidate because better separation between the $H \to \tau\tau$ signal and the $Z \to \tau\tau$ background is achieved thanks to the boosted τ leptons. Moreover this allows to reject the large QCD background that dominates the $\tau_h\tau_h$ channel [7].

5 The τ-pair invariant mass reconstruction

The visible mass m_{vis} of the τ-pair could be used to separate $H \to \tau\tau$ events from irreducible $Z \to \tau\tau$ background, but the large amount of undetected energy due to neutrinos from tau decay degrades this variable discriminating power. The τ-pair mass is reconstructed using a maximum likelihood technique on event-by-event basis [7]. The SVFIT algorithm computes the τ-pair mass that is most compatible with the momenta of the visible tau decay products and the E_T^{miss}. Free parameters, corresponding to the missing neutrino momenta, are subject to kinematic constrains and are eliminated by marginalization. The relative $m_{\tau\tau}$ resolution achieved by the SVFIT algorithm is estimated from simulation and found to be about 15%. The SVFIT mass reconstruction allows for a better separation between signal and background than m_{vis} alone (figure 2), yielding an improvement in the final expected significance of ~40%.

6 Background Estimation

The main background processes are estimated from data. In the $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$, $e\mu$ channels the main source of background are $Z \to \tau\tau$ events. This contribution is estimated with the "embedding" technique: $Z \to \mu\mu$ events are selected in data and the muons are replaced with simulated τ decay. The Drell-Yan event yield is rescaled to the observed yield using the inclusive sample of $Z \to \mu\mu$. The

largest systematic uncertainty is due to the τ selection efficiency (8%). The shape of W + jets events is modelled using the simulation. The yield in a high-m_T control region, dominated from W+jets event as shown in figure 1, is normalized to the observed yield. The extrapolation factor to the low-m_T signal region is obtained from the simulation and has an estimated systematic uncertainty of 10% to 25%. The $t\bar{t}$ process is one of the main backgrounds in the $e\mu$ channel. The shape is predicted by the simulation, and the yield is adjusted to the one observed using a $t\bar{t}$-enriched control sample, extracted by requiring b-tagged jets in the final state. The dominating systematic uncertainty is related to the b-tagging efficiency (1.5%-7.4%). Shape of QCD events is taken from data events where leptons from Higgs candidate decay have the same charge, after subtracting the no-QCD processes taken from simulations. The yield in the signal region is extrapolated by applying the OS/SS ratio computed in an independent sample dominated by QCD events, obtained inverting the requirement on lepton isolation. The systematic uncertainty ranges from 10% to 50% to account for the limited number of events. In the associated production channels the main background comes from Z+jets, W+jets and QCD events where one or two jets are misidentified with the leptons coming from Higgs candidate decay. The misidentification probability is measured as a function of the fake lepton p_T in an independent event sample ($Z \rightarrow \mu\mu+$ jets, W+jets, multijets). The fake lepton candidate selection criteria are relaxed and the contribution into the signal region is extrapolated after weighting each event for the misidentification probability. The systematic uncertainty ranges from 15% to 35% [7].

7 Results

The search for an excess of SM Higgs boson events over the expected background involves a global maximum likelihood fit based on final discriminating variables ($m_{\tau\tau}$ or m_{vis} in all channels except for ee and $\mu\mu$ where the output of two boosted decision trees is used). The distributions of the final discriminating variable obtained for each category and each channel at 7 and 8 TeV are combined in a binned likelihood, involving the expected and observed numbers of events in each bin. The expected number of signal events is the one predicted by the SM multiplied by a signal strength modifier μ treated as free parameter in the fit. The systematic uncertainties are represented by nuisance parameters that are varied in the fit according to their PDF [8]. The excess of events observed in the most sensitive categories is highlighted in figure 3 left, that shows the observed and expected $m_{\tau\tau}$ distributions for all categories of the $\mu\tau_h$, $e\tau_h$, $e\mu$ and $\tau_h\tau_h$ channels combined. The distributions are weighted by the S/(S + B) ratio where S is the expected signal yield for a SM Higgs boson with m_H= 125 GeV ($\mu = 1$) and B is the predicted background yield. The excess of events is quantified by calculating the corresponding local p-values using a profile-likelihood ratio test statistics [8]. Figure 3 right shows the distribution of local p-values and significances as a function of the Higgs boson mass hypothesis. The expected significance for a SM Higgs boson with $m_H = 125$ GeV is 3.6 standard deviations. The observed significance equals 3.4 standard deviations for $m_H = 125$ GeV. The corresponding best-fit value for μ is $\hat{\mu} = 0.86 \pm 0.29$ at mH = 125 GeV. The best-fit value for μ, combining all channels, is $\hat{\mu} = 0.78 \pm 0.27$ at m_H= 125 GeV. Figure 4 left shows the results of the fits performed in each decay channel for all categories. These compatibility tests shows the consistency of the various observations with the expectation for a SM Higgs boson with $m_H = 125$ GeV. Figure 4 right shows a likelihood scan in the (k_V, k_f) parameter space of the couplings of the Higgs boson to vector bosons and fermions, respectively. The observed likelihood contour is consistent with the SM expectation of $k_V=k_f=1$[7].

References

[1] 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) pp.1
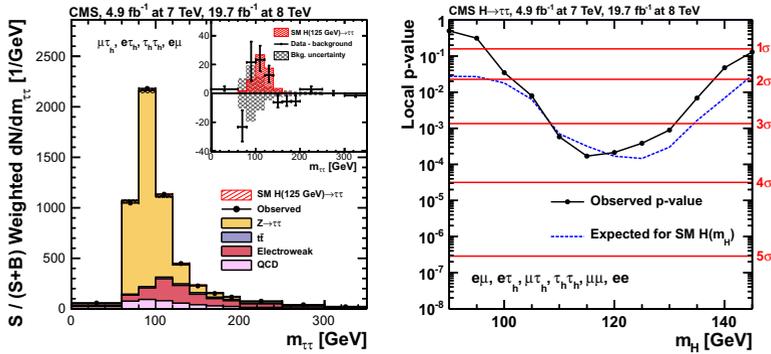
Figure 3. Combined observed and predicted $m_{\tau\tau}$ distributions (left) and local p-value and significance in number of standard deviations as a function of the SM Higgs boson mass hypothesis (right) for the $\mu\tau$, $e\tau$, $\tau\tau$, $e\mu$ channels.
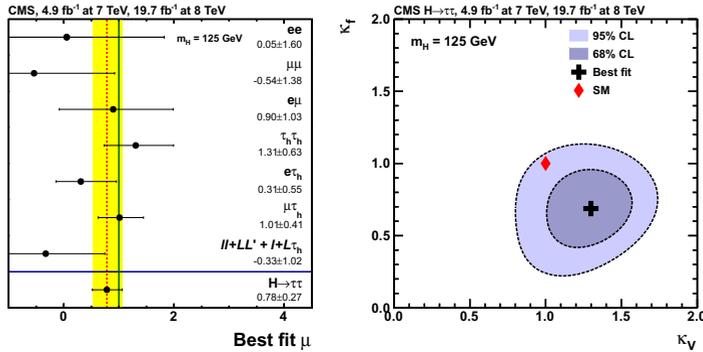

Figure 4. Left: Best-fit signal strength values for independent channels. $ll' + LL'$ and $l + L\tau_h$ stems for associated production channels results. Right: likelihood scan as a function of k_V and k_f.

[2] 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) pp 30

[3] CMS collaboration, *The CMS experiment at the CERN LHC*, (JINST S08004, 2008), pp 3.

[4] CMS collaboration, *Particle-flow event reconstruction in CMS and performance for jets, taus and MET*, (CMS-PAS-PFT-09-001, 2009).

[5] CMS collaboration, *Identification of b-quark jets with the CMS experiment*, (JINST, 8 P04013, 2013).

[6] CMS collaboration, *Performance of τ-lepton reconstruction and identification in CMS*, 2012 JINST 7 P01001

[7] CMS collaboration, *Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons* (JHEP 05 2014, CMS-HIG-13-004)

[8] ATLAS and CMS collaborations, LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, ATL-PHYS-PUB-2011-11 (2011).