

HIGGS  $\rightarrow \tau\tau$  ANALYSIS IN THE CMS EXPERIMENT\*

MICHAŁ OLSZEWSKI

on behalf of the CMS Collaboration

Faculty of Physics, University of Warsaw  
Pasteura 5, 02-093 Warszawa, Poland*(Received May 9, 2016)*

In July 2012, the CMS and ATLAS collaborations announced the discovery of a Higgs boson with a mass of about 125 GeV and properties in agreement with expected from the Standard Model. In this note, we review the analysis performed for the  $H \rightarrow \tau\tau$  decay mode during the first stage of the LHC operation in the CMS. Further, we present the work done during the First Long Shutdown and the first stage of Run 2 of the LHC on the field of hadronic tau lepton offline reconstruction.

DOI:10.5506/APhysPolB.47.1697

**1. Introduction**

In a frame of the Standard Model (SM), the Higgs boson can be produced in the hadron collider like LHC mainly via gluon–gluon fusion, vector boson fusion (VBF) and with association of weak bosons. The observation of the particle is performed in a statistical way throughout identification of its decay products, their kinematics, and is strictly correlated with our ability of the background rejection. The Higgs boson was first observed in the decay channels with diboson and diphotons which are well-reconstructed and indicate non-QCD process. Sensitivity in the double tau lepton is suppressed in a hadron collider although its branching fraction of about 6% is second highest for 125GeV SM Higgs boson. It is caused by the large QCD-background which hinder the identification of hadronically decaying taus, and by neutrinos in the final state worsening the resolution of mass reconstruction. Nevertheless, this particular channel is a valuable source of data for Higgs physics studies for several reasons:

---

\* Presented at the Cracow Epiphany Conference on the Physics in LHC Run 2, Kraków, Poland, January 7–9, 2016.

- decay into taus give a chance for direct measurement of Yukawa coupling of Higgs into fermions;
- the angle distributions of decay products of tau leptons can be used as a probe of CP nature of the Higgs boson;
- some beyond SM scenarios, in particular large-tan  $\beta$  MSSM, can significantly raise branching ratio of Higgs-like resonances into taus.

## 2. Higgs $\rightarrow \tau\tau$ analysis overview

In the analysis, the signal is extracted from the tau pair mass distribution [1]. Due to neutrinos, in the final state, the simple Higgs boson mass estimate taken as a sum of masses of tau visible decay products is not sufficient to effectively differentiate peak of a resonance. Therefore, the statistical maximum likelihood fit method is utilized [2], where we assume true ditau mass  $m_{\tau\tau}$  for every event being the one that maximize the following probability function

$$P(m_{\tau\tau}^i) = \int \delta(m_{\tau\tau}^i - m_{\tau\tau}(\vec{y}, \vec{a})) f(\vec{z}, \vec{y}, \vec{a}) d\vec{a},$$

where  $\vec{z}$  is the missing transverse energy (MET) vector,  $\vec{y}$  is the four-momenta of the visible decay products and  $\vec{a}$  are the unknown parameters specifying the kinematics of the two  $\tau$ -leptons decays. In the above formula, the  $f(\vec{z}, \vec{y}, \vec{a})$  term is responsible for delivering quantitative measure of the compatibility of kinematic variables for the observed and theoretical distributions. The relative  $m_{\tau\tau}$  resolution obtained with this method varies between 10% and 20% depending on considered taus decay modes, and by itself improves the final expected significance of the order of 40% when compare to visible mass usage.

### 2.1. Event selection and categories

The starting point of the analysis consist of creating independent samples with different decay modes of tau lepton ( $\tau_h\tau_h$ ,  $e\tau_h$ ,  $\mu\tau_h$ ,  $e\mu$ ,  $\mu\mu$  and  $ee$ ), and performing remaining steps separately. Using simulated event samples, the trigger and offline selection criteria have been optimized for each channel to maximize the sensitivity to a SM Higgs boson signal [1]. All reconstructed leptons are required to lie within the detector acceptance, originate from the primary vertex, meet minimum transverse momentum condition, be isolated and have opposite sign in pairs. In addition, the  $\ell$ -MET system need to pass the low transverse mass ( $m_T$ ) cut

$$m_T \equiv \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos(\Delta\phi))} < 30 \text{ GeV},$$

where  $p_{\text{T}}^{\ell}$  is the  $\ell$  transverse momentum and  $\Delta\phi$  is the difference in azimuthal angle between the  $\ell$  direction and the  $\vec{E}_{\text{T}}^{\text{miss}}$ . This, in practice, effectively reduces  $W + \text{jets}$  background (see below).

We split the analysis into categories, so it is possible to tune the final cuts and further increase the significance. The three main categories are defined by choosing events with two, one and zero jets, which reflects the anticipated Higgs production process. The two-jet category is intended to be enriched in VBF production and delivers a clean Higgs boson signal, on the price of a reduced yield. The remaining categories serve mainly to constrain the  $Z \rightarrow \tau\tau$  background. Additionally, the jet-based division is extended by selecting events with the high transverse momentum of the Higgs boson candidate, for which  $m_{\tau\tau}$  resolution is improved.

## 2.2. Background

The main background in all  $H \rightarrow \tau\tau$  channels comes from Drell–Yan production of  $Z$  boson. We can distinguish three decay modes contributing differently to the background:

- $Z \rightarrow \tau\tau$  decay is the main irreducible background. It is modeled with the  $Z \rightarrow \mu\mu$  data sample recorded in each data-taking period and for which the muons are replaced with particles resulting from the reconstruction of simulated  $\tau$  lepton decays. The complex detector signature of an event is, therefore, taken directly from data, allowing for the significant reduction of jet energy scale, the missing transverse energy, and the luminosity measurement uncertainties for estimation of this background [1].
- $Z \rightarrow \mu\mu$  and  $Z \rightarrow ee$  decay. In such a case,  $Z$  boson gives the contribution to the signal in  $ll$  and  $\ell\tau_h$  decay channels. The reason for the latter is twofold: (1) the electron (2–3%) and muon ( $\sim 1\%$ ) can be misidentified as a  $\tau_h$  and (2) one lepton can be lost and simultaneously a jet can be misidentified as a  $\tau_h$ .

The  $W + \text{jets}$  background plays important role in  $e\tau_h$  and  $\mu\tau_h$  channels, if  $W$  decays leptonically and jet is misidentified as a  $\tau_h$ . The shape of this background is taken from the simulation, whereas the total signal yield is normalized with enriched  $W + \text{jets}$  sample of high- $m_{\text{T}}$  data.

The similar approach is used for the important in the  $\mu e$  channel  $t\bar{t}$  background. The simulated shape of the  $m_{\tau\tau}$  distribution is normalized with enriched sample created by selecting events with minimum one  $b$ -tagged jet.

The contamination from QCD multijet events is a main reducible background in the  $\tau_h\tau_h$  channel and also relevant in the  $e\tau_h$  and  $\mu\tau_h$  channels. It is entirely estimated from data. In the  $\tau_h\tau_h$  channel, the multijet background shape and yield are obtained from the control region with a relaxed  $\tau_h$  isolation requirement, for which contributions from Drell–Yan,  $t\bar{t}$ , and  $W$  + jets processes are subtracted. The QCD multijet background yield in the signal region is obtained by multiplying the yield in the control region by an extrapolation factor, obtained using identical signal region and control region definitions applied to a sample of the same-charge  $\tau_h\tau_h$  events [1]. In  $e\tau_h$  and  $\mu\tau_h$  channels, the control region with the same-sign leptons is used. The non-QCD backgrounds in the QCD-sideband are normalized to the background prediction and subtracted from the observed control region mass distribution. The obtained signal is rescaled by measured ratio between the opposite-sign and the same-sign events.

The other backgrounds, like diboson and single-top-quark, are small and their contributions are estimated with MC simulation.

### 3. Tau reconstruction

The branching ratio for the tau decaying into hadrons is about 65%, with final state predominantly containing either one or three charged mesons and up to two neutral pions. The electrons and muons coming from the leptonically decaying taus are reconstructed by standard techniques [3–6]. The reconstruction of hadronically decaying taus starts with anti- $k_T$  jets, which are used as seeds for dedicated Hadron-Plus-Strip (HPS) algorithm [7]. As  $\pi^0$  decay virtually instantaneously into  $\gamma$  particles, which, in turn, can convert into  $e^+e^-$  pair, HPS algorithm tries to reconstruct  $\pi^0$ s from photons and electrons constituents of the jet first. All energy deposits in electromagnetic calorimeter within a  $0.05 \times 0.20$   $\eta$ - $\phi$  window are clustered into a strip. The strip is enlarged in  $\phi$ -direction to account for the  $e^+e^-$  pair bending in the detector magnetic field. For each tau candidate, charged hadrons and strips are combined to reproduce the tau decay mode. During Run 1 of the LHC, three topologies were considered: (1) single hadron (relevant to  $\tau \rightarrow h\nu_\tau$  decays); (2) hadron plus one or two strips (relevant to  $\tau \rightarrow h\pi^0\nu_\tau$  and  $\tau \rightarrow h\pi^0\pi^0\nu_\tau$  decays); (3) three hadrons (relevant to  $\tau \rightarrow hhh\nu_\tau$  decays). The tau reconstruction is followed by an identification. In this step, it is required for the tau to pass some isolation criteria in order to mitigate the jet to tau misidentification. The cut-based isolation is defined via the formula

$$I_\tau = \sum P_T^{\text{charged}}(d_Z < 0.2 \text{ cm}) + \max(P_T^\gamma - \Delta\beta, 0) ,$$

where two right-hand side terms are the transverse momentum sum of charged particles (excluding hadrons used for reconstruction) and photons, respectively. The  $\Delta\beta$  correction factor accounts for overestimation of neutral energy deposits due to pile-up. Loose, medium and tight working-points are defined as  $I_\tau$  cuts of 2.0, 1.0 and 0.8 GeV. Finally, a specific discriminants are applied to suppress muons and electrons misidentified as a  $\tau_h$ .

### 3.1. The Run 2 improvements

The described above default tau reconstruction and identification algorithm came by some improvements for the Run 2, from which the most important being:

- *Dynamic strip reconstruction.* The  $e^+e^-$  pair can go outside the strip window for low- $p_T$  electrons. This would produce energy deposit in isolation region of the tau jet, causing isolation cuts ( $< 2$  GeV) to fail. This effect is more pronounced at higher- $p_T$  taus, as the decay product usually have higher transverse momentum. In the new implementation of the algorithm, the original  $0.05 \times 0.20$   $\eta$ - $\phi$  window size is dynamically changed (enlarged) with electron  $p_T$ .
- *New decay modes.* During Run 1, 3-prong decays of tau with additional neutral pions were not considered. In some cases, the occurrence of this process was interpreted as fake tau due to violation of isolation cut by  $\pi^0$  energy deposit. For this reason, the  $\tau \rightarrow hhh\pi^0\nu_\tau$  topology was added. Moreover, to take into account tracking inefficiencies of the detector, the 2-prong topology (with or without  $\pi^0$ s) is now also considered. The new decay modes improve reconstruction efficiency by 3–4%, what is especially important for high energy taus for which the jet-to-tau fake rate is low.
- *MVA-based isolation.* Cut-based isolation was replaced by a multivariate discriminant combining isolation and lifetime and tau shape information. Figure 1  $\tau_h$  shows identification efficiency *versus* jet-to-tau fake rate for HPS cut-based and MVA-based isolation.
- *Improved boosted  $\tau_h$ -ID.* The massive resonances produced in the  $pp$  collisions that decay into diboson possibly lead to topology with two boosted taus passing the detector matter alongside each other. With at least one tau decaying hadronically, the efficient reconstruction of such event require separate treatment. The boosted technique is used, which essentially isolate two subjects from given fat jet and uses subjects as seeds to HPS algorithm.

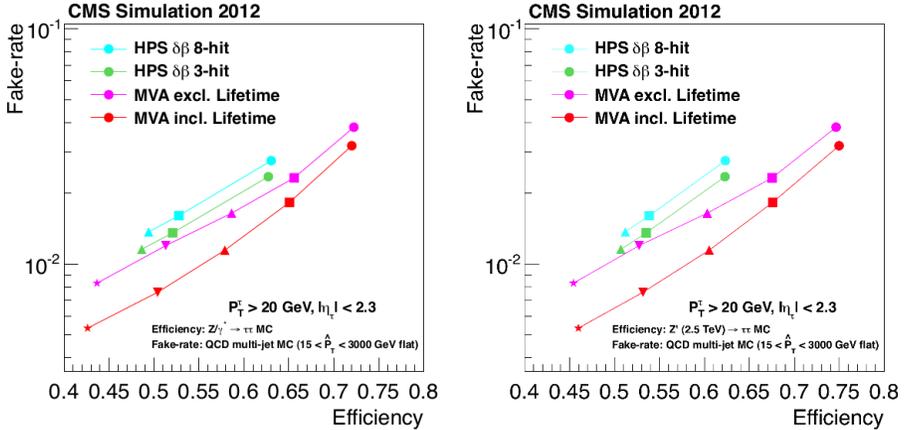


Fig. 1. (Colour on-line)  $\tau_h$  identification efficiency *versus* jet-to-tau fake rate for HPS cut-based isolation with 3-hits (grey/green) and 8-hits (pale grey/blue), and MVA-based isolation with (black/red) and without (dark grey/pink) tau lifetime information. For a given discriminator, the markers correspond to the different working points. The simulated results for SM  $Z$  boson (left) and heavy  $Z'$  (2.5 TeV) resonance (right) are shown.

#### 4. Results and summary

The data collected by the CMS detector during the first stage of LHC working allowed the collaboration to publish the evidence for the Higgs boson decaying into a pair of  $\tau$  leptons. The observed significance was equal

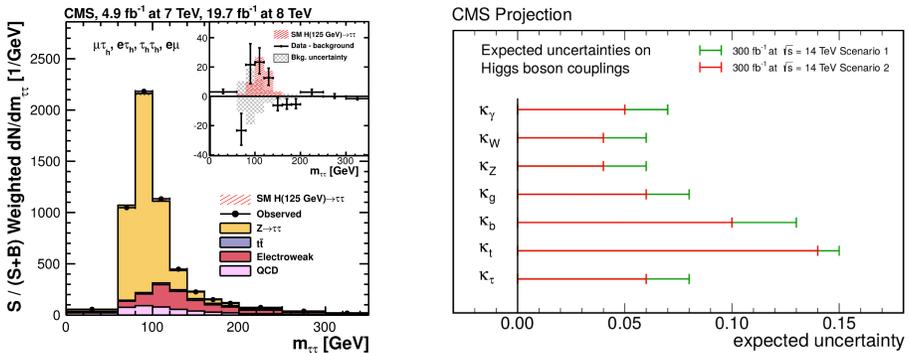


Fig. 2. Plot of combined observed and predicted  $m_{\tau\tau}$  distributions for the  $H \rightarrow \tau\tau$  (left), and estimated precision on the measurements of different Higgs boson coupling parameters (right) [8]. The projections assume  $\sqrt{s} = 14$  TeV and an integrated dataset of 300 fb<sup>-1</sup>. The calculated uncertainties reduction *versus* Run 1 measurements is a factor about two [9].

to 3.2 standard deviations for  $m_H = 125$  GeV. Figure 2 shows the combined observed and predicted  $m_{\tau\tau}$  distributions for the  $\mu\tau_h$ ,  $e\tau_h$ ,  $\tau_h\tau_h$ ,  $e\mu$  channels. The CMS  $H \rightarrow \tau\tau$  search performed greatly during Run 1 and Run 2 of the LHC brings beneficial increase in the cross sections for the signal processes. Together with increased luminosity, improved reconstruction and identification algorithms and upgraded hardware of the detector the upcoming year will deliver new results in the Higgs measurements.

This work was supported by the National Science Centre (Poland) grants 2014/13/B/ST2/02543 and 2014/14/M/ST2/00428.

## REFERENCES

- [1] CMS Collaboration, *J. High Energy Phys.* **1405**, 104 (2014).
- [2] L. Bianchini, J. Conway, E.K. Friis, C. Veelken, *J. Phys.: Conf. Ser.* **513**, 022035 (2014).
- [3] CMS Collaboration, *JINST* **7**, P10002 (2012) [[arXiv:1206.4071](#) [[physics.ins-det](#)]].
- [4] CMS Collaboration, *JINST* **10**, P06005 (2015) [[arXiv:1502.02701](#) [[physics.ins-det](#)]].
- [5] CMS Collaboration, Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and  $E_T^{\text{miss}}$ , CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.
- [6] CMS Collaboration, Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from  $pp$  Collisions at 7 TeV, CMS-PAS-PFT-10-002, 2010.
- [7] CMS Collaboration, *JINST* **7**, P01001 (2012).
- [8] CMS Collaboration, [arXiv:1307.7135v2](#) [[hep-ex](#)].
- [9] CMS Collaboration, *Eur. Phys. J. C* **75**, 212 (2015).