Searches for the SM Higgs boson decaying into two photons, ZZ and two taus in CMS

Marco Meneghelli
Bologna University and INFN
E-mail: marco.meneghelli@cern.ch

Abstract. Results are presented on the search for the Standard Model (SM) Higgs bosons decaying into two photons, ZZ and two taus. Multivariate techniques are used to enhance the sensitivity. The full data sample of 5.1 fb$^{-1}$ of pp collisions collected in 2011 at a CM energy of 7 TeV with the CMS experiment has been analyzed, as well as a significant fraction of the 2012 luminosity delivered at the new CM energy of 8 TeV.

1. Introduction
In July 2012 the CMS and ATLAS collaborations announced the discovery of a new particle, in the search for the SM Higgs boson, at a mass of about 125-126 GeV [1][2]: a boson whose characteristics are compatible with the SM Higgs ones. The local significances of data excesses, over the backgrounds, were more than 5 standard deviations in both experiments. After having observed a new particle, its properties have to be measured: particularly the mass, the cross section, the spin-parity and the couplings to other particles (or the Branching Ratios BR). CMS results on the SM Higgs boson searches are presented in the following, for three of the most important channels: $H \rightarrow \tau\tau, H \rightarrow \gamma\gamma, H \rightarrow ZZ \rightarrow 4l$.

The CMS experiment apparatus is described elsewhere [3]. Its main features are a compact cylindrical structure, built around the solenoidal magnet that creates a field of 3.8 T in the inner part of the detector, where are a silicon tracker, an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). Surrounding the magnet is the muon system. The CMS performances are characterized by a high efficiency and resolution in physical object reconstruction.

2. Search for the SM Higgs boson decaying to tau pair
The search for the SM Higgs boson decaying to a tau pair [4] is, de facto, the combination of five searches (called sub-channels), depending on the decay of the tau leptons in the final state (leptonic decay to electron or muon plus neutrinos, or hadronic decay): $H \rightarrow \tau\tau \rightarrow \mu\tau_h, H \rightarrow \tau\tau \rightarrow e\tau_h, H \rightarrow \tau\tau \rightarrow \mu\nu_\tau, H \rightarrow \tau\tau \rightarrow \mu\mu, H \rightarrow \tau\tau \rightarrow \tau_h\tau_h$. The experimental signature of the events is, in the sub-channels with leptons, the presence of isolated leptons and missing transverse energy. In the hadronic and semi-leptonic sub-channels, hadronic taus are required to be identified.

The search is sensitive to the Higgs boson presence in the low mass range ($110 < m_H < 145$ GeV), where the branching ratio for the Higgs to decay in a tau pair is large. The background rate is about one hundred times larger than the signal one, the main backgrounds being:
the Drell-Yan process with taus in the final state: $Z \rightarrow \tau\tau$

- the top quark pair production

- other Drell-Yan processes with leptons in the final state: $W + jets$, $Z \rightarrow ll$ and di-boson production

- the QCD background

The event reconstruction relies on the “Particle Flow” algorithm (PF), to identify physical objects. Starting from sub-detector information the algorithm builds, for each event, an output of individual particles (charged/neutral hadrons, electrons, muons, photons) which are used to reconstruct hadronic jets, taus and estimate the amount of missing transverse energy. Among the various techniques available at CMS, the PF has shown to provide the best jet reconstruction: jets are, in addition, corrected with calibration factors. The b-tagging allows to discriminate between heavy-flavor and light flavor jets. The missing transverse energy from PF is further corrected with a “Boosted Decision Tree” (BDT) multivariate analysis, to account for the effect of different pile-up conditions. Hadronic taus are typically constituted of a charged hadron plus two neutral pions or by 3 charged hadrons. They are, in both cases, isolated from other particles: tau isolation is evaluated summing transverse momenta/energies of particles in a geometric cone around the main tracks. Leptonic taus are reconstructed as leptons, whose isolation is computed in a similar way.

The event selection requires, for each event, a trigger containing muons, electrons and or tau objects. In the case of semi-leptonic channels an isolated lepton and an opposite charge hadronic tau are required. A cut on invariant mass $m_T$ is applied to discriminate against $W + jets$. In the case of full-leptonic channels, two opposite-sign same-flavor isolated leptons and missing transverse energy are required. A cut on a $m_T$-like variable is applied to discriminate against $W + jets$. For the full-hadronic channel, two reconstructed $\tau_h$ with high $p_T$ are requested. The mass $m_{\tau\tau}$ is reconstructed with a likelihood approach. The selected events, in each sub-channel, are divided into categories relying on jet multiplicity and tau $p_T$, in order to increase the sensitivity of the analysis.

The $Z \rightarrow \tau\tau$ background is estimated from data using the $Z \rightarrow \mu\mu$ events. The $Z \rightarrow \tau\tau$ yield is estimated with an embedding procedure, substituting the muons in $Z \rightarrow \mu\mu$, with simulated taus. Some QCD processes can yield fake leptons/taus: this rate is controlled using a signal-free control region, with same-sign taus. The control region approach is used also to estimate the contribution of the $W+jets$ process: fake tau rate is controlled using events with large $m_T$, while fake lepton rate is controlled using other data-driven techniques, typically defining a control region in a signal-free phase space, and extrapolating the event yields here to the signal region. Concerning the $t\bar{t}$, $Z \rightarrow ll$ and di-boson processes, their rate is estimated from background enriched sidebands.

Systematics uncertainties are assumed on the integrated luminosity, on the jet energy scale, on the background normalization, on the $pp \rightarrow Z$ cross section, on lepton efficiencies, trigger efficiency and tau identification efficiency and on the Higgs production cross section.

The final mass distributions, divided in categories in each sub-channel, are dominated by the huge backgrounds (see e.g., Fig. 1): just the combination of all sub-channels and categories can provide a weakly-discriminating variable. As shown in Fig. 1, that displays the mass distribution for the 0-jet category, there is a very good control of backgrounds.

The results are expressed using 95% CL upper limits on the signal cross section, normalized to the SM nominal one. The cross section modifier (or signal strenght) is defined as $\mu = \frac{\sigma}{\sigma_{SM}}$. Upper limits on this quantity are calculated with the CLs method. A little excess with respect to background-only expectations is observed, broadly distributed (see Fig. 2). The significance of this excess, for the hypothesis $m_H = 125$ GeV, is 1.5 standard deviations.

The best fit to $\mu$ is $\hat{\mu} = 0.7 \pm 0.5$ at $m_H = 125$ GeV.
We can conclude that what is observed is compatible with the SM expectations, but more data are needed to measure any signal property.

**Figure 1.** $H \rightarrow \tau\tau$: mass distribution for the $\mu\tau_h$ sub-channel, VBF category (left) and $e\tau_h$ sub-channel, 0-jet category (right). The 0-jet category is for background-control only.

**Figure 2.** $H \rightarrow \tau\tau$: results of statistical analysis of data. Upper limits on signal cross section compared to what expected in case of background only (left) and the same compared to what expected in case of signal (right).

3. **Search for the SM Higgs boson decaying to photon pair**

   The search for the SM Higgs boson decaying to a photon pair [5] is one of the most sensitive at low mass. The experimental signature of the events is the presence of two isolated photons coming from a primary interaction vertex.
The search is sensitive to the Higgs boson presence in the low mass range \(110 < m_H < 145\) GeV: the signal should be highlighted by the presence of a narrow resonance peak over the broader smooth background. The background rate is about ten times larger than the signal one, the main backgrounds being:

- the irreducible di-photon production
- the instrumental QCD background, where one or two hadronic jets are misidentified as photons

**Photon reconstruction** relies on the presence of significant energy deposits (superclusters) in the calorimeters. The energy is assigned to photon objects and corrected with the usage of a BDT. The photon identification relies on another BDT, trained on a \(pp \rightarrow \gamma + jet\) simulated sample and taking as input shower topology variables and isolation variables. The single photon efficiencies are above 95%. Jets are reconstructed with the Particle Flow algorithm.

The **Event selection** requires di-photon triggers, with asymmetric \(E_T\) thresholds and complementary photon selection: a loose calorimetric identification, a high R9 variable (sum of energy in 3x3 calorimeter crystals). The primary interaction vertex is assigned using the kinematic properties of the tracks associated to it (transverse momenta), and their correlation with di-photon kinematics, using a BDT. Events are further classified using another BDT (trained on simulated background and signal processes) that is designed to sort the events bases on their signal-like characteristics: four classes are individuated, relying on the BDT output, and another (two) class (es) is (are) individuated by the di-jet tag for 7 (8) TeV data, to discriminate “Vector Boson Fusion” (VBF) Higgs production.

For the signal, the “Next to Leading Order” (NLO) contribution is re-weighted to “Next to NLO” (NNLO), taking into account the interference with the \(pp \rightarrow \gamma \gamma\) process. The \(m_{\gamma\gamma}\) distribution is a resonance peak, dominated by the instrumental resolution, modeled with a sum of Gaussians. The backgrounds are entirely modeled on data. There is a 28% of reducible background contamination, while the rest is irreducible di-photon production. The \(m_{\gamma\gamma}\) shape is fitted on data in the \(100 < m_{\gamma\gamma} < 180\) GeV range, polynomial functions are used.

Systematics uncertainties are assumed on the single-photon efficiency, the energy scale and resolution and ID BDT. Other systematics uncertainties are assumed on the integrated luminosity, on the vertex-finding and trigger efficiency, on the di-jet efficiency, on the Higgs cross section.

The statistical analysis is performed with the profile likelihood method, fitting simultaneously the \(m_{\gamma\gamma}\) distribution in the sub-classes. Upper limits on the cross section modifier are computed (see Fig. 3), the largest excess with respect to background-only expectations (see Fig. 3) is observed at \(m_H = 125\) GeV: the local significance is 4.1 standard deviations. The global significance, in the mass range \(110 < m_H < 150\) GeV, is 3.2 standard deviations.

The fit to the signal strength gives \(\hat{\mu} = 1.56 \pm 0.43\) at \(m_H = 125\) GeV.

A visualization of the results, showing the signal peak spiring over the background, is presented in Fig. 4.

**4. Search for the SM Higgs boson decaying to \(ZZ \rightarrow 4l\)**

The search for the SM Higgs boson decaying to a Z pair, each Z decaying leptonically [6], is one of the most sensitive and clean channels, despite the relatively small signal rate. The experimental signature of the events is the presence of two isolated lepton-pairs (muons or electrons) coming from the primary interaction vertex (primary leptons). The search is sensitive to the Higgs boson presence in the a wide mass range \((115 < m_H < 800\) GeV): the signal should be highlighted by the presence of a narrow resonance peak over the broader background. The ratio between signal and background rates is locally of the order one in all the mass range, the main backgrounds being:
Figure 3. $H \rightarrow \gamma\gamma$: results of statistical analysis of data. Upper limits on signal cross section (left) and local p-values for the background-only hypothesis (right).

Figure 4. $H \rightarrow \gamma\gamma$: visualization of results. Weighted sum of events in the sub-classes (left) and the histogram with the background subtracted (right).

- the irreducible double $Z$ production
- the instrumental backgrounds $Z + jets$ and $t\bar{t}$, where jets of secondary leptons are misidentified as primary leptons

Muon reconstruction matches tracker and muon system information to build complete objects and assign them the kinematics. The PF algorithm is used to identify muons, enhancing the purity of the sample. Electron reconstruction matches tracker and ECAL information to build complete objects and assign them the kinematics. A multivariate BDT technique (trained on $H \rightarrow 4e$ as signal and $W + fake$ as background) is used to identify electrons: the input variables are the bremsstrahlung, a tracker-ECAL matching variable and a set of variables describing the ECAL shower shape. Lepton isolation is evaluated by summing the $p_T$ of tracks or energy deposits in a geometrical cone around the track of each lepton, and normalizing the sum to the lepton $p_T$. A cut on the significance of lepton impact parameter is applied, to select primary leptons. The muon identification efficiency is above 95%, while electron identification efficiency
ranges between 60% and 95%, depending on the $p_T$.

The event Selection requires triggers with a pair of electrons or muons, a triplet of electrons or a muon and an electron. From the leptons passing the ID criteria, the $Z_1$ and $Z_2$ candidates are built, with a “Final State Radiation” (FSR) recovery, done by associating FSR photons to $Z$ candidates. A kinematic discriminant variable $k_D$ is built using all kinematic observables of leptons and $Z$ candidates.

The ZZ irreducible background yield is directly estimated from theoretical predictions and simulation. The reducible background yield is estimated using a signal-free control region $Z+X$, the data rate here is extrapolated to the signal region with a single lepton fake rate, estimated from data. Systematic uncertainties are assumed on trigger and lepton reco-ID efficiencies, on the energy-momentum calibration, on the reducible background yield and on the theoretical cross sections.

The event distribution is in good agreement with the SM expectations. A peak around 126 GeV, over the almost flat background, is observed (see Fig. 5).

![Figure 5. $H \to ZZ \to 4l$: $m_4l$ distribution in the low-mass range. Data are superimposed to the expected backgrounds and the signal for $m_H = 126$ GeV.](image)

The statistical analysis of data is performed with the profile likelihood method, using as observables $m_{4l}$ and $k_D$. The largest excess of data with respect to background-only expectations is observed at $m_H = 126$ GeV, with a local significance of 4.5 standard deviations (see Fig. 6). On the other hand, the excluded mass range at 95% CL, for the Standard Model Higgs boson is $[113, 116] \cup [129, 720]$ GeV (see Fig. 6).

The simultaneous fit to the cross section modifier and the mass (see Fig. 7) gives:

$$\hat{\mu} = 0.8^{+0.35}_{-0.28}$$
$$\hat{m}_H = 126.2 \pm 0.6(stat) \pm 0.2(syst) \text{ GeV}$$

The number of events observed around the peak is in good agreement with what expected from the signal+background hypothesis (see table 1).
Figure 6. $H \rightarrow ZZ \rightarrow 4l$: results of statistical analysis of data. Upper limits on signal cross section (left) and local p-values for the background-only hypothesis (right).

Table 1. 4-lepton events for $m_{4l} \in [121.5, 130.5]$. The number of events observed in data is compared to what expected in case of signal + background.

| $m_{4l}$ $\in [121.5, 130.5]$ | expected 7 TeV | expected 8 TeV | observed |
|-------------------------------|----------------|----------------|----------|
| 4e                            | 1.25           | 2.20           | 3        |
| 4$\mu$                        | 2.09           | 4.26           | 6        |
| 2e2$\mu$                      | 3.14           | 5.97           | 8        |
| Total                         | 6.48           | 12.43          | 17       |

A spin-parity hypothesis is tested with dedicated discriminants, which use all the kinematic observables: the $0^+$ hypothesis is compatible with data, while the $0^-$ one is disfavored with a $CLs = 2.4\%$ (see Fig. 7).

5. Conclusions
The results for the SM Higgs search at CMS in three of the most important channels have been presented. They all exploit the excellent detector performances in terms of efficiency and resolution in reconstructing physical objects.

All the searches see excesses in data, with respect to background-only expectations, that suggest the presence of a new boson, compatible with being the Standard Model Higgs. The $H \rightarrow ZZ$ and $H \rightarrow \gamma\gamma$ analyses measure the boson mass at about $125 - 126$ GeV. The ZZ analysis provides a first measurement of the spin-parity quantum numbers, showing that data suggest a $0^+$ scenario.

More precise results require more data, we look forward for the analysis of the complete 2012 dataset, with almost 22 fb$^{-1}$ at 8 TeV collected by CMS.
Figure 7. $H \rightarrow ZZ \rightarrow 4l$: estimation of boson observables. Simultaneous fit to mass and cross section (left) and first results on $J^P$ measurement (right).

References
[1] The ATLAS Collaboration 2012 Phys. Lett. B 716 p. 1-29
[2] The CMS Collaboration 2012 Phys. Lett. B 716 p. 30-61
[3] S. Chatrchyan et al. 2008 JINST 3 S08004
[4] The CMS collaboration CMS PAS HIG-12-015 Search for the standard model Higgs boson decaying to tau pairs
[5] The CMS collaboration CMS PAS HIG-12-043 Evidence for a new state decaying into two photons in the search for the standard model Higgs boson in pp collisions
[6] The CMS collaboration CMS PAS HIG-12-041 Updated results on the new boson discovered in the search for the standard model Higgs boson in the $H \rightarrow ZZ \rightarrow 4l$ channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV