https://doi.org/10.1142/9789811280184_0003

Chapter 3

The Higgs Boson Discovery

Christoph Paus* and Stefano Rosati†

*Massachusetts Institute of Technology (MIT)

†Istituto Nazionale di Fisica Nucleare (INFN) Sezione di Roma

The Higgs boson discovery has been the highlight of particle physics in the last few decades and the crowning achievement of the LHC to date. It was accomplished in its initial running period, Run 1 (2010–2012). In this chapter we will review the status of the standard model Higgs boson searches at the start of LHC, and how the ATLAS and CMS experiments prepared. Analyses of the five main decay channels that lead to the simultaneous discovery, announced in the CERN seminar on July 4th, 2012, will be discussed. We will give an overview of the preparation for the data taking, the physics organization put in place, the analyses of the discovery channels, and finally discuss the results that lead to the discovery.

1. Introduction

The discovery of the Higgs boson, announced by the ATLAS and CMS experiments on July 4th 2012, marks an important milestone in particle physics. The scalar Higgs field and the spontaneous symmetry breaking mechanism[1–6] and the related particle were suggested to exist since the mid 1960's as a way to allow elementary particles to acquire their masses in a natural way by ensuring the standard model (SM) Lagrangian remain locally gauge invariant. At its discovery the Higgs boson was the last SM particle that had not yet been experimentally confirmed and its central role is further underlined by the fact that it is the only elementary *scalar* known to date.

The Higgs boson discovery had long been anticipated even before the start of Run 1, in fact the ATLAS and CMS experiments had detailed

analyses performed on Monte Carlo simulations before their detectors were even completed. The sensitivity of the LHC experiments required to detect the Higgs boson served as the benchmark for the detector design. A number of adjustments to detectors and analyses were made when in the 2000's the electroweak precision data more and more clearly indicated that the expected Higgs boson mass should be rather small, below about 200 GeV.[7]

2. Getting ready for the analysis

At the start of Run 1, ATLAS and CMS carefully reviewed the portfolio of analyses to be performed urgently and decided on a set of high priority analyses. Apart from obvious new physics searches, the Higgs boson search with five main channels were at the top of the priority list. The most relevant decay channels for the Higgs boson are the decays: $H \to ZZ^* \to 4\ell$, $H \to \gamma\gamma$, $H \to WW \to \ell\nu\ell\nu$, $H \to \tau\tau$, and $H \to b\bar{b}$. The decay channels are in descending order of their expected significance at 125 GeV, but the sensitivity depends on the mass value of the Higgs boson and the level of sophistication of the various analyses implementations. The expected significance of these analyses is depicted in Fig. 1 for the CMS experiment and ATLAS had very similar expected sensitivities.

In preparation for the Higgs boson search SM processes like J/ψ, Z boson and W boson decays to leptons were measured and then used to tune the lepton selections and efficiencies. It follows a short description of triggers, object reconstructions and the related key performances.

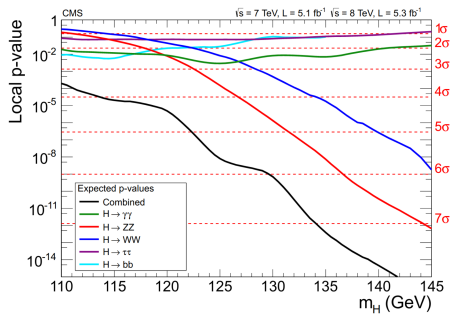


Fig. 1. Expected sensitivity for the five main Higgs boson decay channels for the CMS experiment.

2.1. *Trigger*

The online trigger selection was based on the identification of candidate muons, electrons, and photons with the lowest possible p_{T} threshold given the Run 1 LHC instantaneous luminosity. Either single lepton or dilepton triggers were used for the $H \to ZZ^* \to 4\ell$ and $H \to WW \to \ell\nu\ell\nu$ channels, while diphoton calorimetric triggers were used for the $H \to \gamma\gamma$ channel.

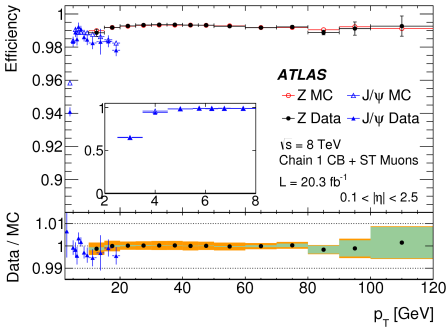
2.2. *Reconstruction of the input objects*

The basic reconstructed objects used in the Higgs discovery channels were leptons (electrons or muons), photons, and missing energy $E_{\mathrm{T}}^{\mathrm{miss}}$. The Higgs boson decay channels to tau leptons and bottom quarks were also published by the CMS collaboration but did not significantly contribute to the observation.

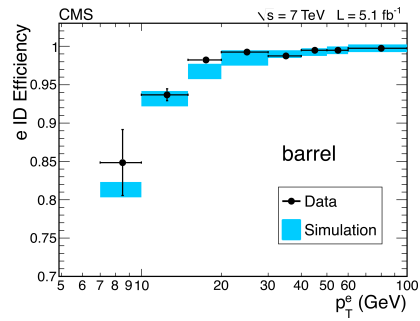
In ATLAS muons were identified and reconstructed combining a muon spectrometer track or segment with a matching track in the inner detector (ID).[8,9] Photon and electron candidates were reconstructed from EM calorimeter energy clusters. Electrons were reconstructed by matching ID tracks to clusters satisfying a set of criteria aiming at the identification of electromagnetic showers.[10] Photons identification was based on shower shapes in the EM calorimeter and on energy leakage into the hadronic calorimeter.[11,12] A cut-based and a neural network selection was used for photon identification in 7 TeV and 8 TeV data, respectively. In CMS, all physics objects were reconstructed with the "particle-flow" event description algorithm,[13,14] which uses an optimized combination of all subdetector information to best reconstruct each particle (muons, electrons, photons, charged and neutral hadrons). Multivariate approaches were used to refine the initial loose selections of muons, electrons and photons. Both experiments were reconstructing jets using the anti-k_t algorithm[15] to cluster the reconstructed objects. The missing energy was defined as the negative vector sum of the transverse momenta of the reconstructed objects, including muons, electrons, photons, jets and clusters of calorimeter cells not associated to these objects. Energy depositions and tracks from overlapping proton-proton collisions ("pileup") and the underlying event were carefully accounted for to ensure optimal selection efficiencies. In particular, lepton or photon isolation requirements are very sensitive to pileup if not designed properly.

For both ATLAS and CMS, data-driven methods were used to assess the reconstruction performances of all the objects used in the analyses.

Scaling factors were applied to the MC to achieve good representation of the actual performance in the data. For leptons, the reference channels were the Z and J/ψ decays to pair of muons or electrons. Energy scales and resolutions were derived from fits of the Z peak.[9,16] The so-called "tag and probe" method was applied to Z bosons to determine reconstruction and identification efficiencies for electrons and muons. An example of the measured reconstruction and identification efficiencies is shown in Fig. 2(a) for muons in the ATLAS experiment and in Fig. 2(b) for electrons in the CMS experiment.



(a) ATLAS muon reconstruction efficiency, as a function of the muon p_{T}. The inset shows the efficiency in the low p_{T} region. The bottom panel shows the ratio between the efficiencies in data and those expected from the MC simulation.



(b) CMS reconstruction efficiencies for electrons in the barrel, at 7 TeV center-of-mass energy. The points with error bars represent the measurements from data, while the histogram shows the efficiency obtained from MC. The shaded region represents the combined statistical and systematic uncertainties.

Fig. 2. An example of the lepton reconstruction efficiencies obtained from data-driven methods, compared to those expected from the MC simulation.

3. Analysis Organization

The Higgs search program was performed in a range consistent with the sensitivity of the individual decay channels but covering the entire region from about 100 GeV up to about 1 TeV, where the Higgs boson mass was not yet excluded by previous experiments. The analyses in ATLAS and CMS were optimized in an unbiased fashion, i.e., not looking at the signal region but only using MC simulation samples. The MC predictions were carefully normalized and constrained using background control regions, to

estimate the background contamination in the signal region of each selection and to validate the MC description and relevant detector effects of the background processes. Data in the signal regions were only analyzed once the analysis optimization was completed and widely discussed within the working groups, and frozen after an approval procedure. In both experiments, the analysis was organized in working groups, each focusing on one of the highest sensitivity decay channels. Working group meetings were very frequent as well as general meetings for common discussions. Both experiments worked on the analysis optimization and validation until the last few weeks before the discovery. Signal regions remained hidden until the analysis procedures were finalized and approved.

In the following we begin with the two most significant channels for the observation. These channels also allow the full reconstruction of the final state and of the mass of the Higgs boson: the $H \to ZZ^* \to 4\ell$ and the $H \to \gamma\gamma$ channel. We finish with the $H \to WW \to \ell\nu\ell\nu$ channel which cannot reconstruct the full Higgs boson due the undetected neutrinos, but due to its large rate improves the combined sensitivity in the low mass region.

3.1. *The $H \to ZZ^* \to 4\ell$ channel*

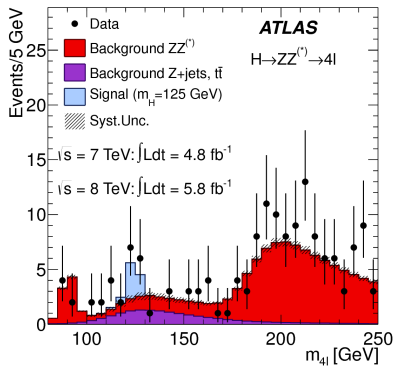
In the $H \to ZZ^* \to 4\ell$ channel the experimental signature is a narrow four-lepton mass peak on top of a small background. The main background is the irreducible $Z\ Z^*$ contribution from direct production via $q\ \overline{q}$ and gluon-gluon interaction. Other backgrounds, relevant in particular in the low mass region, are $Z+$jets and $t\overline{t}$ production, where charged lepton candidates arise from leptonic decays of hadrons with b- and c-quarks, and from jets misidentified as leptons. The analysis selection starts from events with two same-flavor opposite-charge lepton pairs, with all four lepton tracks associated to the same interaction vertex.

Four independent sub-channels, $4e$, $2e\,2\mu$, $2\mu\,2e$, 4μ, with different mass resolutions and background compositions, were considered for the discovery analyses. In ATLAS, each electron (muon) had to satisfy $p_{\mathrm{T}} > 7$ GeV ($p_{\mathrm{T}} > 6$ GeV) and be within a fiducial region $|\eta| < 2.47$ ($|\eta| < 2.7$). The corresponding requirements in CMS are $p_{\mathrm{T}} > 7$ GeV and $p_{\mathrm{T}} > 5$ GeV with $|\eta| < 2.5$ ($|\eta| < 2.4$) for electrons and muons, respectively. The two lepton pairs masses corresponding the two Z bosons are important variables to separate signal from background. To reject reducible backgrounds, the lepton isolation and the impact parameter significance were used by both

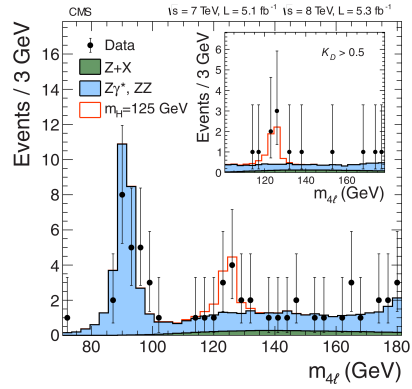
experiments. The irreducible ZZ^* background was determined from the MC simulation, and normalized to the theoretical cross section. Reducible backgrounds were determined from data control regions built by relaxing or reverting some of the identification, isolation or impact parameter requirements.

In ATLAS, a Z boson mass constrained kinematic fit was applied to the lepton pair with the mass closest to the Z mass to improve the four lepton mass resolution. In CMS, a kinematic discriminant was built based on five angles and the masses of the two leptons pairs, which fully describe the kinematics of the final state in its center-of-mass frame. In the hypothesis of a SM-like scalar Higgs boson, the discriminant was defined as the likelihood ratio $K_{\rm D} = P_{\rm sig}/(P_{\rm sig} + P_{\rm bkg})$.[17]

The distribution of the four-lepton mass in data and in MC, is shown in Fig. 3 for ATLAS (Fig. 3(a)) and CMS (Fig. 3(b)). The peak at the Z boson mass is clearly visible for both experiments, it is however more pronounced in the case of CMS due to the looser requirements on the lepton momenta and on the subleading lepton pair mass.



(a) Distribution of the four-lepton mass for the candidates selected by the ATLAS experiment. Points are the data, compared to the histograms that represent the background expectation. The expected signal for a SM Higgs with $m_H = 125$ GeV is also shown.



(b) Distribution of the four-lepton mass for the CMS experiment. The inset shows the mass distribution satisfying the requirement on the kinematic discriminant $K_D > 0.5$.

Fig. 3. Distribution of the four-lepton mass for the (a) ATLAS and (b) CMS experiment. The points represent the data, the filled histograms show the backgrounds and the expected signal for $m_H = 125$ GeV.

3.2. *The $H \to \gamma\gamma$ channel*

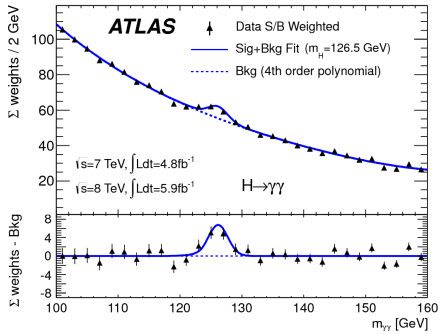
In the $H \to \gamma\gamma$ search channel, the signature is a narrow peak in the diphoton mass distribution. The main background is the irreducible background from SM diphoton production; additional contributions come from gluon plus jet and di-jet production with one or two jets misidentified as photons. In the analyses of both experiments, the events are separated into mutually exclusive categories based on the characteristics of the reconstructed photons and on the additional presence of two jets. In particular, a two-jets category aims to identify events in which the Higgs production has happened through the vector boson fusion process.

The identification of the interaction vertex is critical to keep an optimal resolution for the two photons invariant mass. In the dominant gluon fusion production process of the Higgs boson, it is hard to identify the correct vertex because photons have no tracks. To address this issue, the ATLAS analysis identifies the primary vertex by combining the flight directions of the two photons as reconstructed exploiting the longitudinal segmentation of the electromagnetic calorimeter and its pointing direction measurement, the parameters of the beam spot and the Σp_{T}^2 of the tracks associated to each reconstructed vertex. In the CMS analysis, the primary vertex is identified using a multivariate discriminant which uses, as input, the kinematic properties of the tracks associated to each vertex and the properties of the diphoton kinematics.[18]

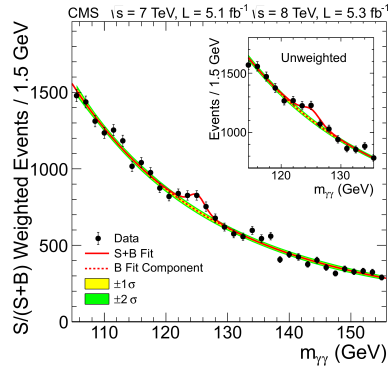
The background in each category was estimated from data, by fitting the diphoton mass spectrum with a model selected for each category. Models were chosen to have a good statistical power while minimizing potential biases. The distribution of the diphoton mass for the ATLAS and the CMS experiment are shown in Figs. 4(a) and 4(b), respectively.

3.3. *The $H \to WW \to \ell\nu\ell\nu$ channel*

This channel is very sensitive in the Higgs mass region around 160 GeV, just above the threshold for the production of a pair of W bosons, but its sensitivity extends downwards to the lower mass region. Only leptonic W decays are considered because hadronic W decays have a large background. Therefore the final state reconstructed in the detector is characterized by two opposite-charged leptons with high p_{T} and large $E_{\mathrm{T}}^{\mathrm{miss}}$ due to the presence of the two neutrinos, which cannot be seen in the detector. The signal topology and the background composition depend on the number of jets present in the final state. In order to optimize the signal sensitivity, the



(a) Mass distribution of diphoton candidates of the ATLAS experiment. Events are weighted as a function of the signal over background ratios of their corresponding categories, as described in the text. The bottom inset shows the residual of the data with respect to the fitted background.

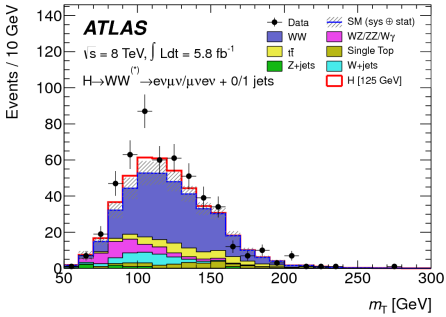


(b) Mass distribution of diphoton candidates of the CMS experiment. Each event is weighted by the signal over signal plus background value of its corresponding category as described in the text. The inset shows the unweighted mass distribution in the region around 125 GeV.

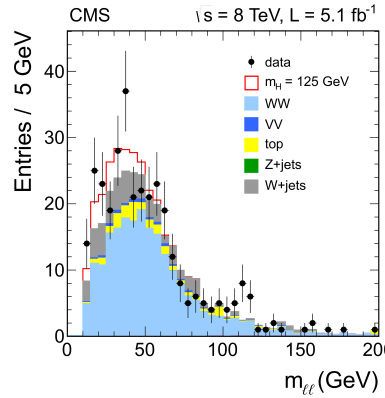
Fig. 4. Distributions of the diphoton mass for the (a) ATLAS and (b) CMS experiment.

event selection criteria were optimized separately for the zero-jet, one-jet and two-jet categories, where the two-jet category includes all events with two or more reconstructed jets.

The main backgrounds are the non-resonant $W\ W$, $t\bar{t}$ and $W\ t$. Drell-Yan lepton pair production, with $E_{\mathrm{T}}^{\mathrm{miss}}$ arising from mis-reconstructed leptons and jets, constitutes a significant and dominant background, for the same-flavor channel. This channel is included in the CMS analysis, while only the different-flavor channels $e\nu\mu\nu$ have been considered in ATLAS, due to the larger Drell-Yan contribution to the same-flavor decay. The main backgrounds were estimated using partially data-driven methods, i.e., normalizing the MC predictions to the data in control regions dominated by each background source. For all jet multiplicities, the distributions considered to test the presence of a signal were the transverse mass m_T defined as: $m_T = \sqrt{(E_{\mathrm{T}}^{\ell\ell} + E_{\mathrm{T}}^{\mathrm{miss}})^2 - |\mathbf{p}_{\mathrm{T}}^{\ell\ell} + \mathbf{E}_{\mathrm{T}}^{\ell\ell}|^2}$ for ATLAS and the dilepton mass $m_{\ell\ell}$ for CMS. Figure 5(a) shows the transverse mass distribution for the zero-jet and one-jet channels together, for the ATLAS experiment. The distribution of $m_{\ell\ell}$ for the CMS experiment is shown in Fig. 5(b) for data and MC.



(a) ATLAS distribution of the transverse mass m_{T} for the $e\mu$ and μe events selected by the zero- and one-jet analyses. The signal prediction for a SM Higgs boson with mass 125 GeV is shown as the red histogram stacked on top of the backgrounds. The hashed area indicates the total uncertainty on the background.



(b) Dilepton mass $m_{\ell\ell}$ distribution for the CMS experiment. The expected signal for a SM Higgs boson with mass 125 GeV is shown as the red histogram stacked on top of the backgrounds.

Fig. 5. Mass distributions for the $H \to WW \to \ell\nu\ell\nu$ Higgs decay channels for the (a) ATLAS and (b) CMS experiments.

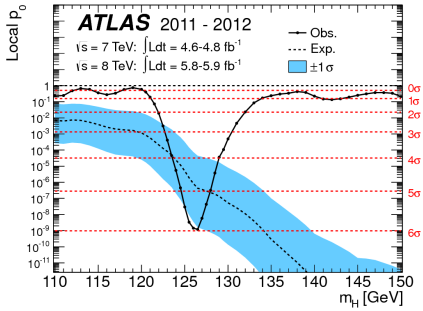
4. Combination and results

4.1. *Statistical procedure*

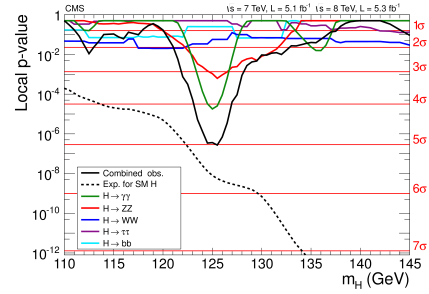
The statistical procedure used to interpret the analysis results was developed by the ATLAS and CMS collaborations within the LHC Higgs Combination Group.[19–22] The parameter of interest is the cross section times the relevant branching fraction, denoted as signal strength $\mu = \sigma/\sigma_{\mathrm{SM}}$. This means that $\mu = 0$ corresponds to no Higgs boson signal, the background-only hypothesis, while $\mu = 1$ corresponds to the SM Higgs boson signal on top of the background. Exclusion limits are derived based on the CL_s criterion.[23,24]

4.2. *Observing a narrow resonance*

The local p-values obtained from the combination of all search channels are shown in Figs. 6(a) and 6(b) for the ATLAS and CMS experiment, respectively. The ATLAS experiment combined the most sensitive channels $H \to ZZ^* \to 4\ell$, $H \to \gamma\gamma$ and $H \to WW \to \ell\nu\ell\nu$. The CMS experiment



(a) The local p-value observed (solid line) and expected in the hypothesis of a SM Higgs boson signal (dashed line) versus m_H.



(b) The observed (expected) local p-values versus m_H are shown as the black solid (dashed) line. The solid colored lines also show the observed local p-values per decay channel.

Fig. 6. Observed and expected local p-values in the low m_H region. The horizontal lines indicate the p-values corresponding to significances of one to seven standard deviations.

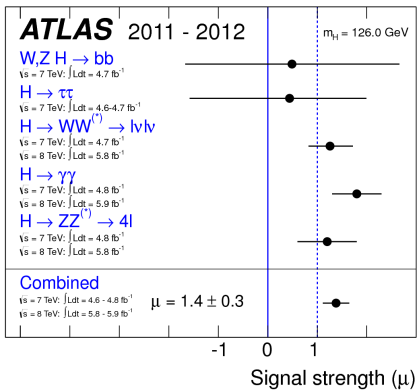
added to those three channels also the $H \to b\overline{b}$ and $H \to \tau\tau$ decay channels for the final combination.

Both experiments observed a clear excess in the region close to $m_H = 125$ GeV. The excess was dominated by the two high sensitivity and high mass resolution channels and was confirmed by the low mass resolution channels, in particular the $H \to WW \to \ell\nu\ell\nu$ channel. The leading contributions to the discovery came from the $H \to ZZ^* \to 4\ell$ channel for the ATLAS analysis, while the $H \to \gamma\gamma$ channel was most significant for the CMS analysis. From the combination of the $H \to ZZ^* \to 4\ell$, $H \to \gamma\gamma$ and $H \to WW \to \ell\nu\ell\nu$ channels, the ATLAS observed (expected) significance was 6.0 (4.9) standard deviations, while the CMS experiment's corresponding values were 5.0 (5.8) standard deviations for the combination of all five channels. The observed and expected local p-values are shown in Fig. 6(a) for the ATLAS experiment and in Fig. 6(b) for the CMS experiment.

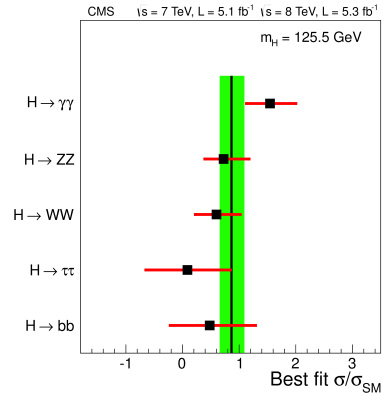
While the excess was observed within the context of the SM Higgs boson search, the experiments were rather careful to state what they had found and not jump to conclusions. What was clear was that there was a new particle materializing as a narrow — consistent with detector resolution — resonance. Its observed decay to two photons excluded it from being a spin one particle and thus, it is most probably a spin zero boson. In addition more detailed tests were immediately performed. The compatibility of the observed excess with the expectation from the SM was evaluated by

measuring the signal strength σ/σ_{SM} in each decay channel. The results are shown in Figs. 7(a) and 7(b) for the two experiments. The signal strength depends on the Higgs mass considered at the time, which was $m_H = 126$ GeV for the ATLAS experiment and $m_H = 125.3$ GeV for the CMS experiment. The signal strength values measured by both experiments were compatible with the SM expectation.

These findings lead both experiments to announce the discovery of a new particle, a boson compatible with the expectations of a SM Higgs boson at around 125 GeV.



(a) Best-fit signal strengths for $m_H = 126$ GeV observed by the ATLAS experiment for each of the decay channels analyzed, and combined.



(b) Best-fit signal strengths for $m_H = 125$ GeV observed by the CMS experiment for each of the decay channels analyzed, and combined.

Fig. 7. Summary of the signal strengths for the various channels and the combined analyses per experiment. Both CMS and ATLAS observe a clear signal at five or more standard deviations consistent with the expectations of a SM Higgs boson at a mass around 125 GeV.

5. Conclusions

On July 4th 2012, the Large Hadron Collider experiments ATLAS and CMS announced the discovery of a new boson within the context of their standard model Higgs searches, a particle that behaved much like the Higgs bosons which had been hypothesized almost fifty years before. They did it within the first data taking period of the LHC, called Run 1, which operated at 7 and 8 TeV, about half the final planned center-of-mass energy. This new

boson for all we know now is the Higgs boson and completed the standard model because at that time it was the last missing particle that had not yet been observed.

Completing the SM might seem like a final step for the outside observer, but explaining it and also addressing all fundamental questions it leaves unexplained are more like a new beginning. Questions at the top of the priority list for particle physicists are: is the Higgs boson really the Higgs boson and could it be a portal to another world of physics? What is the nature of dark matter, which we observe to exist? Why is the matter-antimatter asymmetry in the universe so large? How can gravity be included into our model of the universe? And there are many more questions which particle physicists hope to answer with the extension of the LHC program and future colliders at the energy frontier.

References

1. F. Englert and R. Brout, *Phys. Rev. Lett.* **13**, 321 (1964).
2. P. Higgs, *Phys. Lett.* **12**, 132 (1964).
3. P. Higgs, *Phys. Rev. Lett.* **13**, 508 (1964).
4. C. H. G.S. Guralnik and T. Kibble, *Phys. Rev. Lett.* **13**, 585 (1964).
5. P. Higgs, *Phys. Rev.* **145**, 1156 (1966).
6. T. Kibble, *Phys. Rev.* **155**, 1554 (1967).
7. t. T. E. W. G. ALEPH CDF D0 DELPHI L3 OPAL SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak, and H. F. Groups, *CERN-PH-EP-2010-095.* **arXiv:1012.2367 [hep-ex]** (2010).
8. ATLAS Collaboration, *Phys. Lett. B.* **710**, 383 (2012).
9. ATLAS Collaboration, *Eur. Phys. J. C.* **74**, 3130 (2014).
10. ATLAS Collaboration, *Eur. Phys. J. C.* **72**, 1909 (2012).
11. ATLAS Collaboration, *Phys. Rev. D.* **83**, 052005 (2011).
12. ATLAS Collaboration, *Phys. Rev. Lett.* **108**, 111803 (2012).
13. CMS Collaboration, *CMS Physics Analysis Summary.* **CMS-PAS-PFT-09-001** (2009). `http://cdsweb.cern.ch/record/1194487`.
14. CMS Collaboration, *CMS Physics Analysis Summary.* **CMS-PAS-PFT-10-001** (2010). `http://cdsweb.cern.ch/record/1247373`.
15. M. Cacciari, G.P. Salam and G. Soyez, *JHEP.* **0804**, 063 (2008).
16. CMS Collaboration, *JHEP.* **06**, 081 (2013).
17. CMS Collaboration, *JHEP.* **1204**, 036 (2012).
18. CMS Collaboration, *Phys. Lett. B.* **710**, 403 (2012).
19. ATLAS Collaboration and CMS Collaboration, *ATL-PHYS-PUB-2011-011, CERN-CMS-NOTE-2011-005* (2011).
20. L. Moneta, K. Belasco, K.S. Cranmer, S. Kreiss, A. Lazzaro et al., *PoS ACAT2010.* p. 057 (2010). arxiv:1009.1003 [physics.data-an].

21. K. Cranmer, G. Lewis, L. Moneta, A. Shibata, W. Verkerke, *CERN-OPEN-2012-016* (2012). http://cdsweb.cern.ch/record/1456844.
22. W. Verkerke, D. Kirby, *Tech. Rep., SLAC, Stanford, CA* (2003). arXiv:physics/0306116 [physics.data-an].
23. T. Junk, *Nucl. Instrum. Meth. A*. **434**, 435 (1999).
24. A.L. Read, *J. Phys. G*. **28**, 2693 (2002).