

# Title TBC

Louie Corpe

Supervisors: Paul Dauncey, Chris Seez

June 5, 2014

## Abstract

An overview of the study of the Higgs boson via its decay to two photons at CMS is presented. A brief introduction into the history of the Standard Model and the motivation for Higgs searches is given, as well as a detailed description of the CMS experiment at CERN. The analysis of the  $H \rightarrow \gamma\gamma$  decay mode is described in some detail, with reference to recent work. A discussion on the outlook and future work to be done in this decay channel is also given.

## Contents

1	Introduction	2
2	The SM and $H \rightarrow \gamma\gamma$	2
2.1	The Standard Model . . . . .	2
2.2	$H \rightarrow \gamma\gamma$ . . . . .	3
3	The CMS detector	4
3.1	Overview . . . . .	4
3.2	The tracker and the Electromagnetic Calorimeter . . . . .	4
3.3	Other subdetectors . . . . .	5
4	The $H \rightarrow \gamma\gamma$ analysis	5
4.1	Overview of the legacy analysis . . . . .	5
4.2	Photon resolution studies . . . . .	6
4.2.1	Motivation . . . . .	6
4.2.2	Method and Result . . . . .	7
4.2.3	Estimation of the error . . . . .	8
4.2.4	Checks . . . . .	9
5	ECAL calibration and service work	11
5.1	Overview of the task . . . . .	11
5.2	The ECALELF tool . . . . .	11
5.3	ECALELF with CMSSW_7_0_X . . . . .	12
5.4	CMSSW_7_0_X regression validation . . . . .	12
6	Future work	13
7	Conclusion	13
	References	13

# 1 Introduction

-Introduction to CERN and the LHC. -Overview of my PhD: purpose of this report -Will present my work on CMSSW validation and Hgg photon energy resolution. -

The  $H \rightarrow \gamma\gamma$  analysis played an important role in the discovery of the Higgs boson, and has recently published an analysis using all of the data from run I (known as the “legacy analysis”). This legacy analysis has demonstrated the sensitivity of this decay mode: an observation of the Higgs boson signal with over  $5\sigma$  allowed a standalone discovery to be claimed CITE LEGACY. My Ph.D. will involve working within this group. In run II, the LHC will ramp up the center of mass energy  $\sqrt{s}$  to  $13\text{TeV}$  and eventually  $14\text{TeV}$ . This increase, coupled with a rise in the luminosity of the beam (CHECKME), will allow even more data to be taken, and should usher in the era of precision Higgs physics. The tradeoff to pay, however, is that further strain will be put on the trigger, reconstruction and event processing frameworks, both during data acquisition and after. In order to make the most of the opportunities offered by run II, much work needs to be done in these domains to ensure that the CMS collaboration is ready for the switch on on 2015. One example of where such work is required is in the calibration of the ECAL. The  $H \rightarrow \gamma\gamma$  analysis relied heavily on the excellent performance of the ECAL in run I to achieve its results. The analysis also required software to convert this performance into resolution. In run II, new constraints mean that the method used to reconstruct electrons and photons within the ECAL have changed (INSERT Mustache method??). This report will present work I have done in the validation of this update in the software, as well as a review of the resolution achieved by the  $H \rightarrow \gamma\gamma$  specific software (‘known as *h2gGlobe*) in simulated data in the legacy analysis.

## 2 The SM and $H \rightarrow \gamma\gamma$

### 2.1 The Standard Model

The standard model (SM) of particle physics came into being in the mid-1970’s, when the discovery of the  $J/\psi$  particle [1, 2] and results from deep inelastic scattering experiments at SLAC confirmed the predictions of quark models. The SM has been an immensely successful theory, accurately describing many processes in high energy physics. Among other features, the model placed all fundamental forces apart from gravity into one framework, and united the electromagnetic and weak forces into the electroweak force [3–5]. In order to do this, a mechanism was required to permit the  $W^\pm$  and  $Z$  vector bosons to have mass while allowing the photon to remain massless. Crucially, this process was required to preserve gauge invariance. Such a theory was independently proposed by several theorists [6–11], and is commonly referred to as the Higgs mechanism. In 1964, Higgs postulated that one outcome of this mechanism was that it should yield an observable particle, the Higgs boson [8]. Over the decades, the particles postulated by the SM were discovered: the  $\tau$  lepton in 1975 [12], the  $b$  quark in 1977 [13], the gluon in 1979 [14–16], the  $Z$  and  $W^\pm$  bosons in 1983 [17, 18], the top quark in 1995 [19, 20] and the  $\nu_\tau$  in 2000 [21]. Each new discovery cemented the SM as one of the most successful theories of modern times. By the turn of the millennium, all but one particle postulated by the SM had been observed: the Higgs boson, which had proved elusive despite decades of searches. The Higgs search prompted the construction of the Large Hadron Collider (LHC) at CERN, and two multi-purpose detectors, ATLAS and CMS, were designed with the Higgs observation as one of their main physics goals. In 2012, the two experiments jointly announced the observation of a Higgs-like particle of mass  $\sim 125$  GeV, ending a 50-year lull between postulation and discovery [22, 23].

Although the SM has been very successful as a theory, it falls short of being a “theory of everything”, and is clearly incomplete. For instance, it does not accommodate mass terms for neutrinos, which are required to explain the origin of neutrino oscillations observed in many experiments [24–26]. Furthermore, it does not contain a viable candidate for dark matter, which is needed to explain the mass deficit of the universe [27]. Other issues such as the hierarchy problem [28] and the origin of matter-antimatter asymmetry in our universe [29] also persist. Clearly, the SM is incomplete or approximate, and many efforts in modern high energy physics are being made to discover “Beyond the Standard Model” (BSM) physics. Supersymmetry proposes one such extension to the SM, but many other models exist, and the search for new physics will be one of the primary objectives of experimentalists during run II of the LHC in 2015.

Even in the Higgs sector, more work remains to be done. Precision measurements of the properties of this particle are needed: does it behave like the SM predicts? Is there only one such particle? Detailed

studies are required to ascertain couplings, the differential cross section, width and other attributes of this Higgs particle. Deviations from the expected values of such properties could provide valuable insight into the nature or indeed existence of BSM physics.

## 2.2 $H \rightarrow \gamma\gamma$

According to SM, the Higgs boson's coupling with particles is proportional to their mass. As such, its production modes in the environment of the LHC are dominated by interactions involving the heavier particles of the SM. Typically, the Higgs boson is produced by one of the following mechanisms, as illustrated in figure 2: a) gluon-gluon fusion, via a loop of top quarks, b) vector boson fusion (so-called “VBF”), with associated quark production, c) associated vector boson production (also known as “Higgsstrahlung”) and d) top quark fusion with associated top quark production.

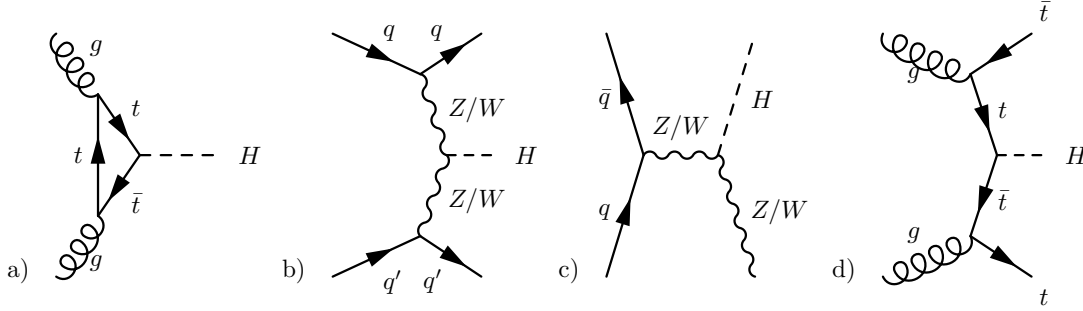


Figure 1: Higgs production mechanisms at the LHC.

By the same token, the SM Higgs boson decays to pairs of particles with branching ratios proportional to the square of their mass. The production of a pair of  $t$  quarks is not kinematically allowed because their mass is high, so the most likely decay modes are  $H \rightarrow ZZ, W^\pm W^\mp, b\bar{b}$  and  $\tau^+\tau^-$ . In addition, a small fraction of decays of the Higgs boson ( $< 1\%$ ) can occur via a loop diagram to a pair of high-energy photons. The branching fractions and cross sections of these production and decay modes are available in the Handbook of LHC Higgs Cross Sections [30,31].

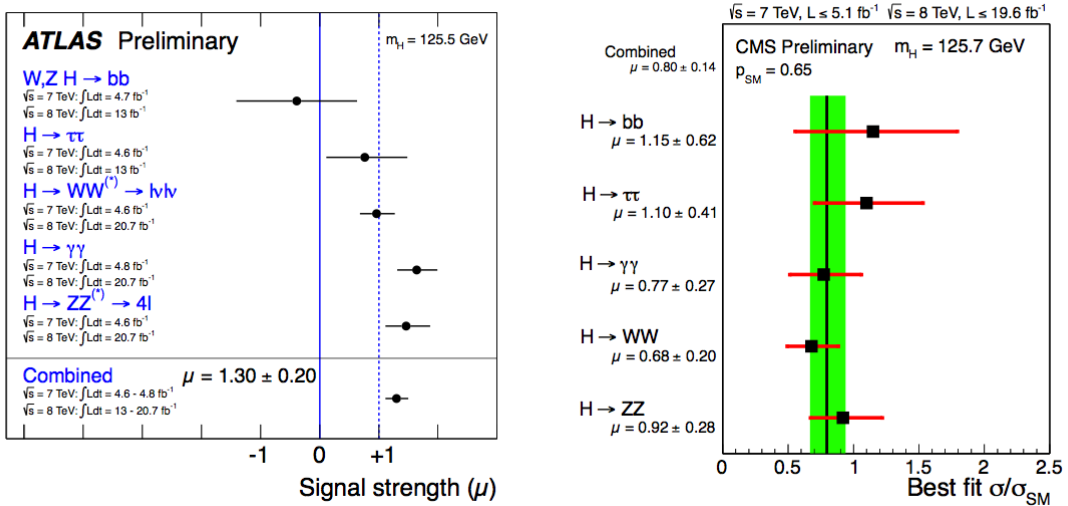


Figure 2: The signal strengths of the Higgs boson observation by decay mode and in combination at ATLAS (left) and CMS (right). The results are consistent with the observation of a particle of  $m_H \sim 125 \text{ GeV}$  [32].

The Higgs boson was discovered in 2012 using data from runs at  $\sqrt{s} = 7$  and  $8 \text{ TeV}$ , and was found to have a mass  $m_H \sim 125 \text{ GeV}$ . The signal strengths of the various decay modes at the CMS and ATLAS

experiments can be seen in figure 3. Despite fewer than 1% of Higgs boson decays occurring via  $H \rightarrow \gamma\gamma$ , this channel played a crucial role in the discovery, and remains one of the two most sensitive methods of studying the Higgs boson. This is in part thanks to the excellent performance of the CMS ECAL.

### 3 The CMS detector

#### 3.1 Overview

The LHC is a synchrotron that was built in the tunnel which previously contained the Large Electron Positron (LEP) collider at CERN, near Geneva. It has a circumference of 27km and was designed to collide beams of protons (or lead ions) head on. In its latest run, it achieved a centre of mass energy of  $\sqrt{s} = 8$  TeV. The LHC shut down for a planned upgrade period in 2012, and is due to restart collisions in March 2015. At this stage, it should be able to deliver a centre of mass energy close to its design value of  $\sqrt{s} = 14$  TeV. As mentioned previously, the LHC is equipped with two multi-purpose detectors, CMS and ATLAS, which simultaneously discovered the Higgs boson using the data from run I. This report will focus on the CMS detector. The overall layout can be seen in the figure 1 below [33]. CMS is a layered

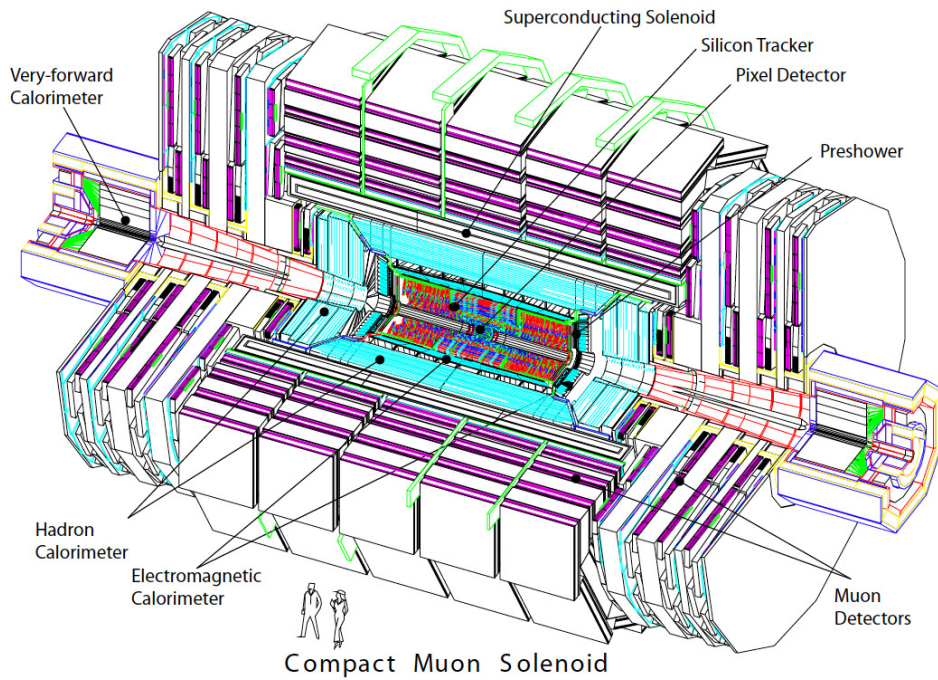


Figure 3: The CMS detector perspective view.

detector over 21m long and 14m in diameter, weighing over 12,500 tons. It consists of a superconducting solenoid magnet 13m long and 5.9m in diameter, generating a 3.8T magnetic field. On the outside of this, 4 layers of iron act as a return yoke, and house muon detector chambers (drift tubes in the barrel region and cathode strip chambers in the end cap region). The high magnetic field allows for good energy resolution in the compact space of the detector. The calorimeters and inner tracker are housed within the solenoid.

#### 3.2 The tracker and the Electromagnetic Calorimeter

The tracker and the electromagnetic calorimeter are the two subdetectors which the  $H \rightarrow \gamma\gamma$  uses most (and indeed, almost exclusively). This report will therefore deal with them in a more detailed manner than it will for the other subdetectors. The inner tracker consists of 4 layers of silicon pixel detectors close to the interaction region, surrounded by 10 layers of silicon micro-strip detectors. (INSERT MORE TRACKER INFO) These allow reconstruction of tracks and secondary vertices in the high track multiplicity environment of the LHC. The electromagnetic calorimeter (ECAL) is made up of an array

of 61,200 lead  $\text{PbWO}_4$  (lead tungstate) crystals in the barrel section and 14,648 crystals in the end caps [33]. This is a central feature of the CMS detector as it allows for excellent energy resolution of incoming photons. The resolution of the ECAL is modelled with the equation

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (1)$$

where  $S$  represents the stochastic term,  $N$  represents the noise terms and  $C$  represents a constant term [34]. The design values of these parameters are approximately  $S = 2.8\% \text{ GeV}^{\frac{1}{2}}$ ,  $N = 0.12 \text{ GeV}$  and  $C = 0.3\%$ . The performance of the ECAL matched these design values during the 2012 run [35].  
INSERT FIGURE of ecal performance?

### 3.3 Other subdetectors

The ECAL is then surrounded by a sampling hadronic calorimeter composed of brass/scintillator in the body and iron/quartz-fibre in the forward regions. The muon chambers  
INSERT INFORMATION  
ipsum lorem Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed pharetra dapibus lorem sed mattis. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Etiam sollicitudin nisi in convallis interdum. Vivamus nec libero at nunc molestie elementum a et arcu. Curabitur facilisis felis eget condimentum mattis. Fusce tellus erat, porttitor aliquet porta at, rhoncus in justo. Aenean pretium nulla vel mi molestie imperdiet. Praesent aliquet velit nec ligula congue, eu lacinia orci ultrices. Mauris commodo cursus massa, ut cursus tortor consectetur eu. Suspendisse eu dictum odio. Morbi tincidunt ultrices erat in ultrices.

## 4 The $H \rightarrow \gamma\gamma$ analysis

### 4.1 Overview of the legacy analysis

The  $H \rightarrow \gamma\gamma$  analysis relies on the fact that there is a small but non-negligible branching fraction of Higgs bosons which decay to two highly energetic photons via loops of particles, generally virtual top quarks or  $W^\pm$  vector bosons. Although around the mass  $m_H \sim 125 \text{ GeV}$  there is a large irreducible QCD background, the excellent diphoton energy resolution ( $\sim 1\%$ ) of the CMS ECAL allows the reconstruction of a narrow peak above the background. The resolution available via this decay channel meant that it was identified early on as “one of the most promising channels in the search for a SM Higgs boson in the low mass range” [36]. The following description is based on the analysis of the  $\sqrt{s} = 7$  and  $8 \text{ TeV}$  data samples citeHDisc.

The analysis depends heavily on boosted decision trees (BDTs) to identify and classify events. A BDT is an algorithm which categorises items based on multiple input variables [37]. Different weights are applied depending on whether an item passes or fails cuts based on these input variables. The weights are summed at the end and the final value is used to determine which category the item falls in. The algorithm is trained (i.e. the aforementioned weights and cut thresholds are determined) with simulations or data with similar properties.

The first step in the analysis is to identify candidate diphoton events. There is a loose trigger, identifying any pair of photons in an event based on ECAL isolation, shower shape and energy. A more stringent selection takes place later to choose only diphoton events likely to have originated from a Higgs decay.

The next step is to locate the photon production vertex. This is an important step because an accurate determination of the origin of the photon tracks can be combined with the measured ECAL energies and spatial location of the photon showers to accurately reconstruct the Higgs 4-momentum, and thus its invariant mass. A simple analysis of conservation of 4-momentum in the laboratory frame yields the required formula:

$$m_H = m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - \cos\alpha)} \quad (2)$$

where  $m_{\gamma\gamma}$  is the invariant mass of the diphoton system,  $E_{1,2}$  are the energies of the reconstructed photons and  $\alpha$  is the opening angle between the two photons. It is clear that since the ECAL resolution is excellent,  $E_{1,2}$  are well known. The only other contribution to the mass resolution comes from the opening angle  $\alpha$ . In order to accurately measure this angle, the photon creation vertex must be identified to within 10 mm of its true position. If this is the case, the error contribution from  $\alpha$  is negligible.

Vertex identification can be achieved by matching the kinematic properties of the tracks coming from these vertices with the diphoton system’s transverse momentum. Another option is to use information from the tracker if the photon converted into an  $e^+e^-$  pair: in this case the track directions can be combined with the impact position in the ECAL to extrapolate back to the photon production vertex. A BDT trained on simulated data is used to identify the vertex, using quantities such as those mentioned above as input variables. The vertex identification efficiency is measured to be of the order of 80%, and the associated systematic uncertainty is well understood.

Since the photon pairs expected from Higgs decay should be highly energetic, offline cuts are applied to select candidate diphoton events likely to have originated from a Higgs boson decay. These cuts are defined as  $E_T > m_{\gamma\gamma}/3$  for one candidate photon and  $E_T > m_{\gamma\gamma}/4$  for the other [38], where  $E_T$  is the transverse energy of the photon and  $m_{\gamma\gamma}$  is the invariant mass of the diphoton system. It is possible that a photon may have pair converted ( $\gamma \rightarrow e^+e^-$ ) before reaching the ECAL. Such occurrences are identified using the variable  $R_9$ , defined as the sum of the energy of the  $3 \times 3$  array of lead tungstate crystals centred around the most energetic crystal in the supercluster, divided by the total energy of said supercluster. A value of  $R_9$  smaller than 0.94 is indicative of a photon having undergone pair conversion.  $R_9$  is used as an input variable in a BDT which classifies events based on their sensitivity, as described below. A further BDT is also used to remove “non-prompt” photons (i.e. photons not created at the primary vertex) and other particles misidentified as photons, such as pions.

The events are then segmented into categories based on the expected signal to background ratio and diphoton invariant mass resolution. Studying these separately increases the overall sensitivity of the analysis. An additional categorisation where candidate events also pass a dijet trigger corresponds to VBF Higgs production (See figure 2.b: the quarks are scattered at a sufficiently large opening angle for them to undergo hadronisation and form jets rather than be lost down the beam-pipe). This mode is analysed separately because its signal to background ratio is over an order of magnitude better than for the other categories [38]. The categorisation is achieved using another BDT, where the input variables are chosen to be dimensionless to avoid any bias in the mass distribution. For example, the transverse momenta of the photons are used amongst other input variables, but they are scaled by the calculated diphoton invariant mass to yield a dimensionless criterion.

The background model is obtained using a data-driven technique rather than a MC prediction, as this avoids any systematic uncertainties based on mis-modelling. For each of the categories, the diphoton mass distribution outside of the region of interest is fitted. The function used to fit the data is chosen based on minimisation of the bias introduced in each case. Candidate functions involve exponentials, power laws, polynomials and Laurent series. The bias is found to be negligible when Bernsteins of order 3 to 5 (depending on the category) are used.

Finally, the observed invariant mass distribution is compared to that predicted by the background fit. An observed excess of events around  $\sim 125.0$  GeV with a local significance of  $4.1\sigma$  in the  $\gamma\gamma$  decay channel signalled the existence of the Higgs boson in 2012. Combined with measurements from other decay modes, CMS announced a measurement of  $m_H = 125.3 \pm 0.4(\text{stat.}) \pm 0.5(\text{syst.})$  GeV with a local observed significance of  $5.0\sigma$  [39].

## 4.2 Photon resolution studies

### 4.2.1 Motivation

An important task after the publication of the legacy analysis is to learn lessons from the performance of the previous run. This allows better preparation for run II. As such, a paper is being prepared to mark the performance of the  $H \rightarrow \gamma\gamma$  software during the later part of the first run. One aspect of this was to run a study to look at the resolution of photon energies as reconstructed using the globe software package. I was tasked with repapring this plot.

The study consisted of looking at the reconstruction of Higgs boson decay photons in simulated collision at 8TeV, assuming  $m_H = 125\text{GeV}$ . The resulting plot was intended to give the effective energy resolution  $\sigma_{\text{eff}}/E$  for individual photons, given in small  $|\eta|$  bins. Monte Carlo samples for the 8TeV LHC conditions were used in this study. More precisely, a sample of simulated  $H \rightarrow \gamma\gamma$  events corresponding to each Higgs production mode (Vector Boson Fusion, Gluon-Gluon Fusion, Associated Vector Boson Production and Associated Top Quark Production), weighted by cross section, were used as the dataset for the study. Since the intention is to provide an account of the sensitivity of the  $H \rightarrow \gamma\gamma$  analysis, the MC samples were passed through the globe MVA framework, and therefore were processed as they would

have been in the legacy analysis, with full smearing and corrections.

#### 4.2.2 Method and Result

In order to produce the required plots, the globe MVA framework code was modified so as to produce an additional root tree, containing only the information needed for this study. The writing of this additional tree took place after all the actual analysis within the framework was complete. In order to measure the effective energy resolution of the photons, it was necessary to compare the “true” energy of the photons  $E_{true}$  (as generated in the simulation) with the “reco” energy of the photons  $E_{reco}$  (as provided after reconstruction and application of corrections in the the globe framework). To make this comparison in a meaningful way, the generated photons and the reconstructed needed to be matched correctly. This was done using an unambiguous geometrical method: for each generated photon in an event, the value of  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$  (where  $\Delta$  refers to the difference between generated and reconstructed values) was calculated for the two reconstructed photons. The reconstructed photon minimising the value of  $\Delta R$  was chosen as the match for that generated photon. (DELTA R and DELTA ETA PLOTS??)

For each event, each individual photon was categorised first by  $R_9 < 0.94$  or  $R_9 \geq 0.94$ , then by its value of  $|\eta|$ . The width of the  $|\eta|$  categories was 0.1, apart from the region near the transition between the ECAL barrel and ECAL endcap, where the data are excluded (resulting in two slightly larger categories on either side, namely  $1.3 \leq |\eta| < 1.4442$  and  $1.566 \leq |\eta| < 1.7$ ), thus 23 categories. Taking into account the two  $R_9$  options, this left 46 photon categories. For each one, a histogram of the values of  $E_{reco}/E_{true}$  was created. An algorithm is used to calculate  $\sigma_{eff}$  (defined as half the width of the smallest window containing 68.3% of events in a distribution). This algorithm recursively scans potential windows in the histogram until the smallest matching the required condition is found. A different algorithm also calculates the “Full Width Half Maximum” (FWHM), which is the width of the peak when measured at half of its maximum height. In this case, the algorithm fits each histogram to a combination of a Gaussian and a Crystal Ball function (AVAILABLE IN APPENDIX), then creates a very finely binned histogram of the best fit, and finds the bin with the largest number of events. Then it finds the first bin with half that number of events, and the last bin with half the number of events. For a true Gaussian distribution,  $\sigma_{eff} = \sigma = \text{FWHM}/2\sqrt{2\ln 2} \simeq \text{FWHM}/2.35$ . In other cases,  $\sigma_{eff} \geq \text{FWHM}/2.35$ . Fig. 4 shows examples of these histograms.

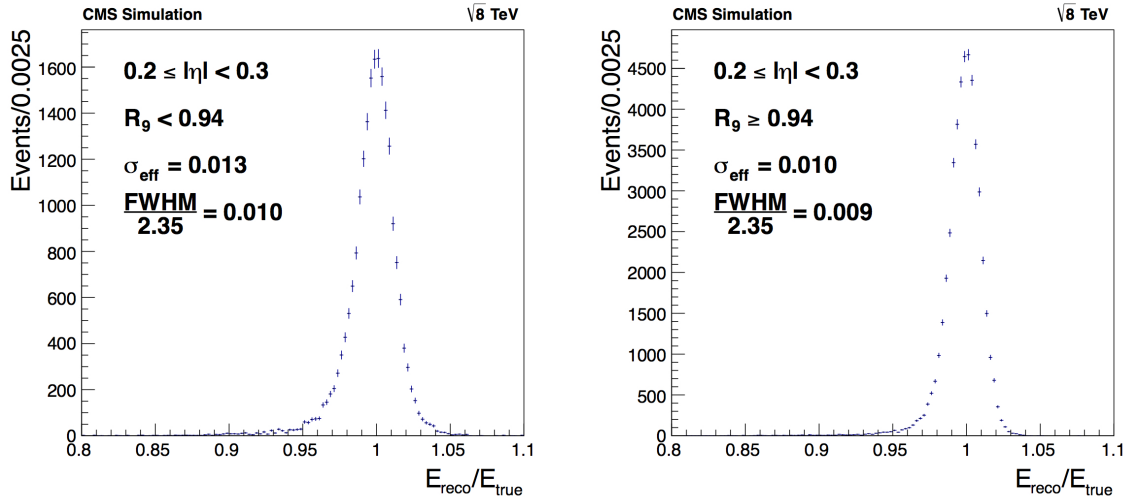


Figure 4: Example  $E_{reco}/E_{true}$  histograms for two  $|\eta|$  and  $R_9$  categories. The vertical axis indicates the number of events if each event were counted with the weight associated to a GGH event. Overlaid on each plot is the calculated value of  $\sigma_{eff}$  and  $\text{FWHM}/2.35$ . For a distribution with a Gaussian peak but non-Gaussian tails, we expect  $\sigma_{eff} \geq \text{FWHM}/2.35$ , which is what is observed.

The values of FWHM are not used in the rest of the study, however they provide a good sense check that the distributions make sense. On the other hand,  $\sigma_{eff}$  is used as the input for the main result. The values of  $\sigma_{eff}$  for the distributions of  $E_{reco}/E_{true}$  are equal to  $\sigma_{eff}/E$  (FIXME why??), and are used as



inputs to the effective energy resolution plot. Each of the 46  $|\eta|$  and  $R_9$  categories provide a value, which is then plotted versus  $|\eta|$  to give the main plot (Fig.5).

As expected, the effective resolution is consistently worse for low  $R_9$  categories than for their high  $R_9$  equivalent. This is because low  $R_9$  photons are likely to have undergone pair conversion, and the resulting electrons and positrons may have emitted energy via bremsstrahlung before hitting the ECAL. Furthermore, the categories where  $|\eta|$  lies over a supermodule boundary (vertical dashed lines) have worse resolution than neighbouring categories. This is also expected, as the interfaces between the supermodules are less hermetic than the body of the supermodules, leading to energy lost between the gaps. Finally, the resolution in the endcaps (ie categories where  $1.5 \leq |\eta| < 2.5$ ) is substantially worse than the resolution in the barrel. This is also expected (FIXME why?). In conclusion, this study set out to estimate the individual photon resolution in small  $|\eta|$  bins in the CMS ECAL for simulated  $H \rightarrow \gamma\gamma$  events at  $m_H = 125\text{GeV}$ , after the full and final corrections applied with the globe framework. The result is that for high  $R_9$  photons in the barrel, the effective energy resolution is of the order of  $\sim 1\%$  in the barrel and  $\sim 2/4\%$  in the endcaps. For low  $R_9$  photons, the resolution is slightly worse.

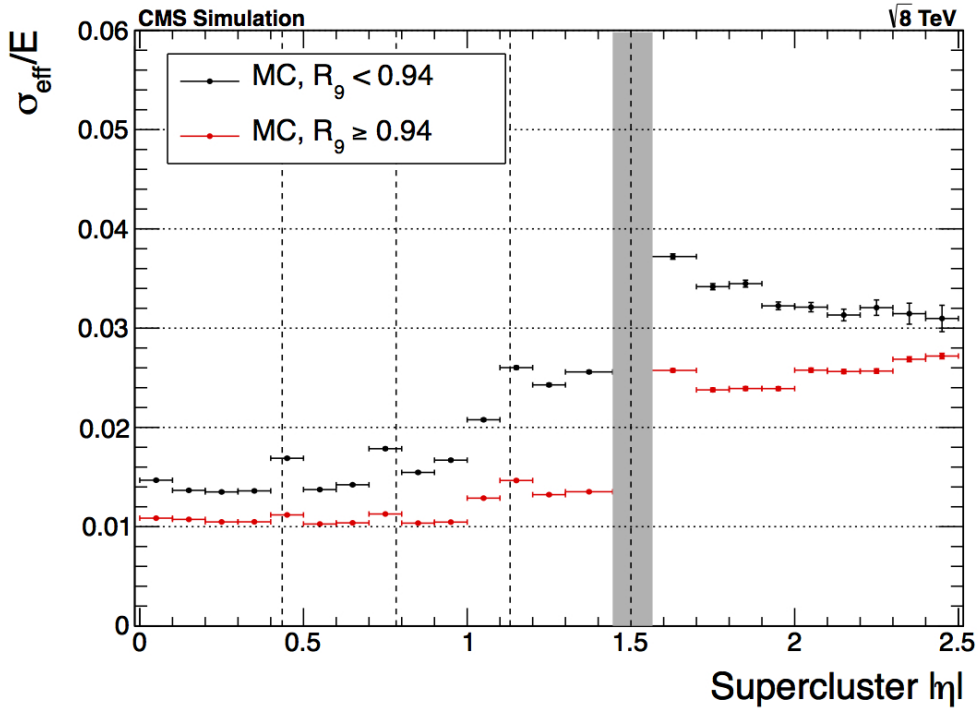


Figure 5:  $\sigma_{\text{eff}}/E$  plotted against  $|\eta|$ . Horizontal dotted lines shown as guides only. Vertical dashed lines show boundaries between supermodules. The grey band indicates the interface between the ECAL barrel and endcap regions, which we do not study. Horizontal error bars indicate width of  $|\eta|$  bin, while vertical error bars (visible chiefly in the high  $|\eta|$ , low  $R_9$  categories), indicates the estimated uncertainty on the measurement of the resolution.

#### 4.2.3 Estimation of the error

The vertical error bars in Fig. 5 provide an estimate of the uncertainty associated with the measurement of the resolution in each  $|\eta|$  and  $R_9$  category. The error for a given  $|\eta|$  and  $R_9$  category was assumed to be at least as large as the standard statistical error, given by  $\text{Err}_{\text{stat}} = \sigma_{\text{eff}}/\sqrt{2N}$  (where  $N$  is the number of photons in the category). However, additional uncertainty may have been added by the algorithm. To check this, the sample was divided into ten equivalent subsamples. For each subsample, the entire study was rerun, and the resulting ten values of  $\sigma_{\text{eff}}/E$  were recorded for each of the 46  $|\eta|$  and  $R_9$  categories. The RMS of these ten values was calculated in each case. If the error really were equal to  $\text{Err}_{\text{stat}}$ , we would expect that the subsample RMS error in each category would be  $\sim \sqrt{10} \simeq 3.2$  times larger than



the error for the full sample. However, when comparing the calculated subsample RMS error to the minimum statistical error for each category, the factor was on average 4.4 rather than 3.2, indicating that the true error is (very roughly) 1.5 times larger than  $\text{Err}_{\text{stat}}$ . As a result, the error added into Fig. 5 was  $\text{Err}_{\text{stat}} \times 1.5$ , thus providing a rough estimate of the uncertainty of the resolution. For most bins, the fact that  $N$  is very large means that the vertical error bars are completely invisible, however for low  $R_9$ , high  $|\eta|$  categories, which have fewer photons, the error bars are no longer negligible.

#### 4.2.4 Checks

Several checks were performed during the course of the study in order to verify the consistency of the results. The main ones are summarised here. Firstly, the photon matching mechanism was checked thoroughly. The good performance of the matching mechanism can be seen in Fig. 6.

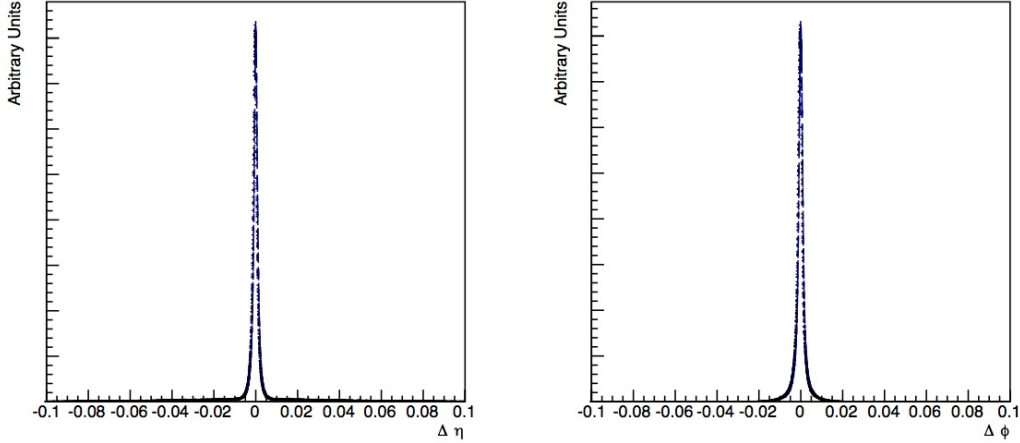


Figure 6: The plot on the left (right) shows  $\Delta\eta$  ( $\Delta\phi$ ) for photons used in this study, where  $\Delta X$  refers to the difference between  $X_{\text{reco}}$  and  $X_{\text{true}}$ . The resulting distributions are very narrowly centered about 0, showing that the matching algorithm is performing very well. In the rare cases where  $\Delta\eta$  and  $\Delta\phi$  are different from 0, this is due to poorly reconstruction of photons rather than mismatching.

Secondly, the study was run using the CiC (REFER EARLIER FIXME) globe framework, rather than the MVA globe framework. Since the CiC framework was used as a crosscheck for the MVA framework in the actual legacy analysis, it provided an excellent opportunity to check consistency in this study. As can be seen in Fig. 7, the CiC version leads to slightly worse resolution in the endcaps, which is as expected since the MVA framework was designed to be more sensitive in this region.

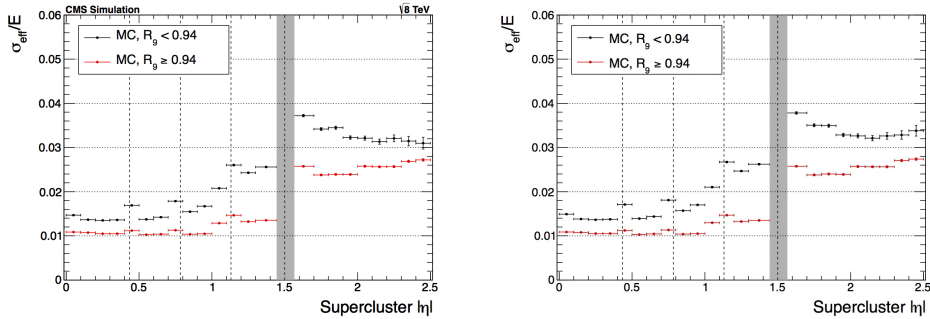


Figure 7: The plot on the right (left) shows the final result when the initial data are run through the globe CiC (MVA) framework. The results are almost identical, with the biggest differences observable in the low  $R_9$ , high  $|\eta|$  categories. This is as expected, as the MVA analysis was designed to be more sensitive, especially in this region.

A Final check was run to check whether modifying the binning in of the  $E_{\text{reco}}/E_{\text{true}}$  plots had an effect on the result of the  $\sigma_{\text{eff}}$  algorithm. Various bin sizes were considered, however, the change in the resulting values of  $\sigma_{\text{eff}}$  were less than negligible (at most 1.7% in the most extreme cases).

## 5 ECAL calibration and service work

### 5.1 Overview of the task

The overall collection of CMS Software is called CMSSW. “*The overall collection of software, referred to as CMSSW, is built around Framework, an Event Data Model, and Services needed by the simulation, calibration and alignment, and reconstruction modules that process event data so that physicists can perform analysis.*” [40]

Effectively, CMSSW provides a way to process and select events from the raw data and allow physicists in the CMS collaboration to analyze the outcome and produce physics results. One way in which CMSSW does this is in the reconstruction of particles. Of particular interest to the  $H \rightarrow \gamma\gamma$  group is the way in which photons are reconstructed within the ECAL. The latest version of CMSSW is known as CMSSW\_7\_0\_X, where the X can represent a number of different sub-versions. A key distinction between this and the version used during run I (known as CMSSW\_5\_3\_X) is indeed the way in which photons are reconstructed. In the new version, a “Particle Flow” (PF) (FIXME explain??) approach has been implemented. Using the new PF clusters, CMSSW\_7\_0\_X applies the so-called “moustache clustering” method, so called because of the shape of the energy distributions with respect to the seed crystal (effectively, the new method uses more complex shapes than simple boxes around crystals, as used in the previous iteration). (FIXME picture? more explanation needed). In this method, all ECAL objects (electrons and photons) are reconstructed with the algorithm and undergo particle identification as a different time. This method was implemented because it is able to take advantage of the ECAL granularity to improve containment of electrons and photons while rejecting pileup. Advantages of the method also include better low- $p_T$  electron resolution, particle-based isolation...

However, although in principle the new clustering method should provide many advantages, it needs to be validated. Any significant loss in the energy resolution of ECAL objects could have a potentially disastrous effect on the  $H \rightarrow \gamma\gamma$  analysis in particular, but in general CMS’s capability to make important measurements in run II. My service work therefore is to work with the ECAL detector performance group (DPG) to help check the performance of CMSSW\_7\_0\_X against the performance of CMSSW\_5\_3\_X.

### 5.2 The ECALELF tool

The **ECAL Calibration with Electron Framework** is a generic tool which uses  $Z \rightarrow e^+e^-$  events to monitor and calibrate the ECAL. ECALELF works within CMSSW to provide aligned and calibrated data upon which tests and studies can be run. In this sense, it is an ideal tool to test the new clustering method for electrons, and can be adapted to validate the reconstruction of photons also. The general workflow for calibration with ECALELF is to take *RECO* data and MC (ie, raw data which has been reconstructed using the native CMSSW version) for a given set of conditions, and apply a selection to retrieve only candidate  $Z \rightarrow e^+e^-$  events. A further selection is made via BDT to reduce to a pool of candidate events. Once this is complete a map between MC and data events based on finely segmented  $\eta$  and  $R_9$  categories. In general, the two electrons in a candidate event will not belong to the same category. In this case, it is difficult to know what correction factor needs to be applied to the energy of each electron in order to obtain the correct invariant mass distribution. However when both electrons in the candidate event belong to the same  $\eta$  and  $R_9$  category, one can look at the MC invariant mass distribution in that category, compare it to the invariant mass distribution of the candidate data events in that category, and infer that the correction to be applied to electrons (as a function of energy) in order to retrieve the required distribution. In this way,  $Z \rightarrow e^+e^-$  events with both leptons in the same category allow the calculation of the required correction (as a function of energy) to be applied to *individual* electrons in that category. Once this has been done for all categories, the events where the electrons/positrons belong to different categories can also be corrected (since each one will belong to a category where the required correction for an individual electron is known). The result will be a set of corrections to apply in any  $\eta, R_9$  scenario, thus providing the calibration. This is known as the “smearing method”.

Apart from actual calibration, the ECALELF tool can be used to run general studies on Ecal performance using the output data from the setup before the application of the smearing method. Namely, one can theoretically compare the ECALELF ntuples resulting from different CMSSW releases, since the ntuples will be in the same format.

### 5.3 ECALELF with CMSSW\_7\_0\_X

Although ECALELF provides an opportunity to test the performance of the new clustering method by comparing it to the previous release's ntuples, work needed to be done in order to make ECALELF work in the new CMSSW environment. Various variables had changed from one version to the other, it an update of the software was needed. However, in order to run a proper comparison, a version of ECALELF which functioned within both 53X and 70X was needed. In other words, we needed to develop code which could actively detect which CMSSW version was being used, and make the required substitutions at compile time. I carried out this work and completed it in the first quarter of 2014.

### 5.4 CMSSW\_7\_0\_X regression validation

## 6 Future work

The  $H \rightarrow \gamma\gamma$  working group are in the process of finalising the legacy analysis for the  $\sqrt{s} = 7$  and 8 TeV data samples from run I. This will include a mass measurement for the Higgs boson based on a similar analysis to the one described above. An analysis of the spin properties of this particle, which I was involved in, will also be included. This publication will reflect the final word on the older data samples from the perspective of the CMS  $H \rightarrow \gamma\gamma$  analysis. The LHC should begin colliding again in 2015, and new data will then become available. Various new measurements will need to be made to further understand the properties of the Higgs boson. Further data will be able to shed light on the existence of a potential second Higgs boson (multiple such bosons are predicted in various BSM frameworks), and will allow precision measurements of differential cross sections, couplings and spin/parity. Deviations from the expected properties of the SM Higgs boson could signal the potential existence of BSM physics.  $H \rightarrow \gamma\gamma$  will remain one of the most sensitive channels with which to analyse the properties of the Higgs during run II. On a more personal note, my work within the CMS experiment will initially be centred around the ECAL, preparing for calibration in run II. The way that photons are reconstructed is being modified, and I will help determine what the effect of this change will be on the sensitivity of the  $H \rightarrow \gamma\gamma$  analysis. As 2015 approaches, I will become involved in planning and implementing the next round of analysis, to help further pin down the properties of the Higgs boson.

## 7 Conclusion

To conclude, it is clear that even though we have now discovered the elusive Higgs boson, many questions remain unanswered. In 2015, some BSM theories which might help to answer these questions will be tested. If nothing is found in direct searches, the precision analysis of the properties of the Higgs boson will be one of key avenues with which to search for deviations from the SM. As such, further understanding of the Higgs boson's properties are not only desirable, but indeed imperative if we are to understand more about the fundamental structure of the universe.

## References

- [1] J. E. Augustin et al. Discovery of a Narrow Resonance in  $e^+e^-$  Annihilation. *Phys. Rev. Lett.*, 33:1406–1408, Dec 1974.
- [2] J. J. Aubert et al. Experimental Observation of a Heavy Particle *J. Phys. Rev. Lett.*, 33:1404–1406, Dec 1974.
- [3] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev.*, D2:1285–1292, Oct 1970.
- [4] A. Salam. Weak and Electromagnetic Interactions. *Conf. Proc.*, C680519:367–377, 1968.
- [5] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.
- [6] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [7] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, September 1964.
- [8] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.
- [9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.
- [10] P. W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966.
- [11] T. W. B. Kibble. Symmetry breaking in non-abelian gauge theories. *Phys. Rev.*, 155:1554–1561, Mar 1967.

- [12] M. L. Perl et al. Evidence for anomalous lepton production in  $e^+ e^-$  annihilation. *Phys. Rev. Lett.*, 35:1489–1492, 1975.
- [13] S. W. Herb et al. Observation of a dimuon resonance at 9.5-GeV in 400-GeV proton - nucleus collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.
- [14] R. Brandelik et al. Evidence for planar events in  $e^+ e^-$  annihilation at high energies. *Phys. Lett. B*, 86:243–249, September 1979.
- [15] D. P. Barber et al. Discovery of Three-Jet Events and a Test of Quantum Chromodynamics at PETRA. *Phys. Rev. Lett.*, 43:830–833, September 1979.
- [16] C. Berger et al. Evidence for gluon bremsstrahlung in  $e^+ e^-$  annihilations at high energies. *Phys. Lett. B*, 86:418–425, October 1979.
- [17] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 GeV at the CERN SPS collider. *Phys. Lett.*, B126:398–410, 1983.
- [18] M. Banner et al. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN  $\bar{p} - p$  collider. *Phys. Lett.*, B122:476–485, 1983.
- [19] F. Abe et al. Observation of top quark production in anti-p p collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.
- [20] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.
- [21] K. Kodama et al. Observation of tau neutrino interactions. *Phys. Lett.*, B504:218–224, 2001.
- [22] S. Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.
- [23] G. Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [24] Y. Fukuda, T. Hayakawa, E. Ichihara, et al. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, August 1998.
- [25] Q. R. Ahmad, R. C. Allen, T. C. Andersen, et al. Measurement of the Rate of  $\nu_e + d \rightarrow p + p + e^-$  Interactions Produced by  $^8\text{B}$  Solar Neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87(7):071301, August 2001.
- [26] F. P. An, J. Z. Bai, A. B. Balantekin, et al. Observation of Electron-Antineutrino Disappearance at Daya Bay. *Phys. Rev. Lett.*, 108(17):171803, April 2012.
- [27] P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, et al. Planck 2013 results. I. Overview of products and scientific results. (arXiv:1303.5062), March 2013.
- [28] S. P. Martin. A Supersymmetry primer. *Kane, G.L. (ed.): Perspectives on supersymmetry II*, (arXiv:hep-ph/9709356), 1997.
- [29] U. Sarkar. Particle and astroparticle physics. *Taylor and Francis*, 2008.
- [30] S. Dittmaier et al. Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. (arXiv:1101.0593. CERN-2011-002), 2011.
- [31] S. Dittmaier et al. Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. (arXiv:1201.3084. CERN-2012-002), 2012.
- [32] S. Heinemeyer, C. Mariotti, G. Passarino, et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. (arXiv:1307.1347. CERN-2013-004), 2013. Comments: 404 pages, 139 figures, to be submitted to CERN Report. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>.
- [33] G. L. Bayatian et al. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. 2006.

- [34] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.
- [35] Federico De Guio. Performance of the CMS electromagnetic calorimeter and its role in the hunt for the Higgs boson in the two-photon channel. *J.Phys.Conf.Ser.*, 455:012028, 2013.
- [36] C. J. Seez et al. Photon decay modes of the intermediate mass Higgs. *Conf. Proc.*, C901004:474–487, 1990.
- [37] Byron P. Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 543(2–3):577 – 584, 2005.
- [38] S Chatrchyan et al. Search for the standard model Higgs boson decaying into two photons in  $pp$  collisions at  $\sqrt{s} = 7$  TeV. *Phys. Lett.*, B710:403–425, 2012.
- [39] S. Chatrchyan et al. Observation of a new boson with mass near 125 GeV in  $pp$  collisions at  $\sqrt{s} = 7$  and 8 TeV. *JHEP*, 06:081, 2013.
- [40] G. L. Bayatian, S. Chatrchyan, G. Hmayakyan, et al. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items.