Evidence for a Narrow Higgs-like Diphoton Resonance with a Mass of 125 GeV in pp Collisions at $\sqrt{s} = 7 - 8$ TeV

ARCHIVES

by

Joshua Lorne Bendavid

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2013

© Massachusetts Institute of Technology 2013. All rights reserved.

Author .....
Department of Physics
January 21, 2013

Certified by .....
Christoph M.E. Paus
Professor
Thesis Supervisor

Accepted by .....
Professor John Belcher
Associate Department Head for Education

Evidence for a Narrow Higgs-like Diphoton Resonance with a Mass of 125 GeV in pp Collisions at $\sqrt{s} = 7 - 8$ TeV

by

Joshua Lorne Bendavid

Submitted to the Department of Physics
on January 21, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

We have performed a search for the production of the standard model Higgs boson decaying to diphotons in pp collisions at the LHC at $\sqrt{s}$ = 7-8 TeV with the CMS detector. Having analyzed data corresponding to 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV (2011) and 5.3 fb^{-1} at $\sqrt{s} = 8$ TeV (2012), a statistically significant excess of events is observed with respect to the background prediction. Interpreted as a standard model Higgs, this excess has a local significance of 4.1 standard deviations, with the maximum significance occurring for a Higgs mass of 125 GeV. Taking into account the trials factor given the search range of 110 GeV to 150 GeV in Higgs mass, this excess has a global significance of 3.2 standard deviations. This constitutes evidence for a new particle decaying to diphotons with a mass of around 125 GeV. The rate of observed events is consistent with predictions for the standard model Higgs boson.

Thesis Supervisor: Christoph M.E. Paus
Title: Professor

Acknowledgments

I would like to thank firstly my thesis supervisor Christoph Paus for the appropriate mixture of sound advice, encouragement and independence over these past years. I would like to thank as well the other faculty members in our group, Markus Klute and Steve Nahn, as well as our heavy ion counterpart Bolek Wyslouch for their advice and support over the years. I am grateful to the entire MIT group for the vast amount of physics experience from which I have benefited through countless hours of group meetings and corridor conversations, which I will always remember fondly. I would especially like to thank my fellow diphoton contributers Fabian Stoeckli and Mingming Yang for their collaboration, dedication, and support during this remarkable period of scientific work.

I thank the entire CMS Higgs to two photon group, in which the mix of collaboration and constructive competition has enabled a truly remarkable set of results over the last two years. I would like to thank as well the entire CMS collaboration and the CERN accelerator division for all of the work over many years which has enabled the collection of the data for this analysis. I would especially like to thank the CMS electromagnetic calorimeter group for building a wonderful detector, and for the hard work in the last years to deliver the best performance possible with it.

I would like to thank my parents Janice and Alex for their patience and support since the very beginning and for their understanding on my lack of geographic and social proximity in the last several years. I would especially like to thank them for nurturing the curiosity and intellectual independence of a young boy, without which I could not be where I am today. I thank also my grandparents Mina, Sam, Ella and Zeev, whose inspiring lives have always given me perspective on the wonderful opportunities I have had in my own life.

The opportunity to work with these people at this moment, on this result has been an incredible experience and a high point in my life thus far. I can only hope that the future brings equally engaging challenges and exciting results.

Contents

1 Introduction **11**

1.1 The Large Hadron Collider 12

1.2 Higgs Boson Production and Decay at the LHC 13

1.3 The CMS Experiment 18

1.3.1 Inner Tracking Detectors 20

1.3.2 Electromagnetic Calorimeter 23

1.3.3 Hadronic Calorimeter 28

1.3.4 Muon Detectors 29

1.3.5 Data Acquisition and Trigger 30

1.4 Analysis Overview 32

1.4.1 Multivariate Analysis Techniques 33

1.5 Data Sample 40

1.6 Monte Carlo Simulation 41

2 Event Reconstruction and Selection **45**

2.1 Photon Reconstruction 45

2.1.1 Clustering 46

2.2 Photon Energy Regression 47

2.2.1 Regression Algorithm Details 54

2.3 Electron Reconstruction 59

2.3.1 Track Seed Reconstruction and Matching 60

2.3.2 Electron Track Reconstruction 60

2.3.3 Electron Track-Cluster Matching 61

2.4 Residual Photon Energy Scale and Resolution Corrections 61
2.4.1 $Z \to ee$ Selection and Reconstruction 62
2.4.2 Correction Granularity 62
2.4.3 Time-dependent Energy Scale Corrections to Data 63
2.4.4 Residual Energy Scale Corrections to Data 64
2.4.5 Energy Resolution Corrections to Simulation 64
2.4.6 Systematic Uncertainties 66
2.5 Conversion Reconstruction 66
2.6 Primary Vertex Reconstruction and Selection 73
2.6.1 Primary Vertex Reconstruction 74
2.6.2 Primary Vertex Selection 74
2.7 Photon Selection and Identification 78
2.7.1 Definition of Photon Variables 79
2.7.2 Photon Pre-selection 83
2.7.3 Photon Identification MVA 92
2.8 Diphoton mass resolution and kinematics MVA 98
2.8.1 Inputs 100
2.8.2 Per-Event Resolution Weighting 102
2.8.3 Validation 104
2.8.4 Systematic Uncertainties on di-photon MVA Output 105
2.9 Jet Reconstruction and Identification 112
2.10 Vector Boson Fusion Selection 115
2.11 Event Categorization 117

3 Interpretation and Results **119**
3.1 Signal Model 119
3.1.1 Final Signal Models 121
3.1.2 Systematic Uncertainties 122
3.2 Background Model 122
3.3 Statistical Procedure 128

3.4 Results 132

4 Conclusions **139**

4.1 Additional CMS Higgs Results and Combination 140

4.1.1 $H \to ZZ \to 4\ell$ 140

4.1.2 $H \to WW \to 2\ell 2\nu$ 141

4.1.3 $H \to \tau\tau$ 142

4.1.4 $H \to bb$ 143

4.1.5 Combined Results 144

4.2 ATLAS Results 146

4.2.1 Diphoton Search 146

4.2.2 Combined Results 150

4.3 Final Remarks 151

Chapter 1

Introduction

The standard model (SM) of particle physics has been enormously successful in describing a wide range of experimental results up to currently accessible energies. Providing a fundamental theory of the electroweak[1, 2, 3] and strong [4, 5] interactions and the associated fundamental particles, the SM contains nearly all known physics aside from gravity. The massive vector bosons mediating the weak interactions versus the massless photon mediating the electromagnetic interactions is empirical evidence for the breaking of the electroweak symmetry. In the SM this occurs as a consequence of the Higgs mechanism [6, 7, 8, 9, 10, 11]. This predicts, in addition to the previously observed fundamental particles and interactions, an additional scalar field and the associated scalar Boson. The prediction of additional Yukawa couplings to the fundamental fermions provides gauge-invariant terms in the Lagrangian corresponding to the fermion masses, and leads to the key prediction in the SM that the coupling of the Higgs boson to fermion pairs is proportional to the mass of each fermion type. Although the SM makes detailed and relatively precise predictions on the couplings of the Higgs to the other fundamental particles, it makes no precise prediction as to the mass of the Higgs boson. In order to preserve unitarity in the WW-scattering interaction, its mass must however be smaller than about 1 TeV [12, 13, 14, 15]. Previous direct searches at the LEP collider in e^+e^- collisions up to $\sqrt{s} = 209$ GeV have excluded at 95% confidence level (C.L.) the existence of a SM Higgs boson with a mass less than 114.4 GeV [16]. Direct searches at the Tevatron collider in $p\bar{p}$ collisions

at $\sqrt{s} = 1.96$ TeV have excluded a SM Higgs in the mass range 162-166 GeV at 95% C.L. [17]. Furthermore, indirect constraints from precision electroweak measurements at LEP, SLC and Tevatron favour, at 95% CL a Higgs mass less than 152 GeV[18].

The Large Hadron Collider (LHC) at CERN is a proton proton collider with a design center of mass energy of $\sqrt{s} = 14$ TeV [19]. The Compact Muon Solenoid (CMS) experiment[20] is one of two (together with ATLAS[21]) general purpose experiments at the LHC, capable of probing a wide range of physics. The observation or exclusion of the SM Higgs boson has been one of the main physics goals of the LHC programme and a major focus of the two collaborations.

1.1 The Large Hadron Collider

The LHC [19] is a proton-proton collider 26.7 km in circumference, located at the European Organization for Nuclear Research (CERN). The accelerator is located in a tunnel 45-170 m beneath the surface, lying between Lake Geneva and the Jura mountains, straddling the border between Switzerland and France. The LHC is designed to collide two beams of protons with a center of mass energy $\sqrt{s}$ of up to 14 TeV. This is accomplished using a synchrotron design, with 1232 dipole magnets with a nominal magnetic field of 8.33 T. Because both beams consist of positively charged protons, oppositely directed dipole magnetic fields are required in order to maintain their orbits in opposite directions. This is accomplished by two separate beam pipes with independent vacuum, except for smaller regions near the interaction points, and two sets of field coils within the dipole magnets. In order to achieve the necessary magnetic field, superconducting electromagnets are required, with field coils made of NbTi and cooled to 1.9 K using superfluid helium. To save space in the tunnel, the two beam pipe and magnetic coil assemblies are mounted in a common cryostat and mechanical structure. Proton beams are injected into the LHC with an energy of 450 GeV, having been accelerated up to this energy through a sequence of smaller accelerators including the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). Beams are injected into the LHC in discrete bunches. At the design bunch

spacing of 25 ns, the LHC can accommodate 2808 proton bunches per beam. Nominal design bunch intensities and beam parameters would correspond to a total instantaneous luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$, with an average of 20 proton-proton interactions per bunch crossing and a total interaction rate of about 1 billion per second.

1.2 Higgs Boson Production and Decay at the LHC

The SM makes definite predictions about the coupling of the Higgs boson to the already observed particles and so the production rates and decay modes of the Higgs at the LHC can be calculated relatively precisely as a function of the unknown Higgs mass. Although the LHC is a proton-proton collider, production of the Higgs, as all high energy processes, results from the interactions between the underlying quarks and gluons in the protons, referred to as partons[22, 23]. The relationship between the colliding proton beams and the type and energy of the underlying parton-parton interactions is given by the Parton Density Functions (PDFs) of the proton [24]. Because the LHC is a proton-proton collider, any quark-antiquark interactions must involve antiquarks from the lower momentum sea distribution, as compared to the higher momentum fraction valence quarks. As a result, at energies relevant for Higgs production, the equivalent gluon-gluon luminosity is significantly higher than for quark-antiquark interactions. The dominant production mode for the Higgs at the LHC is therefore gluon-gluon fusion, in which two initial state gluons produce a Higgs boson through an intermediate fermion loop[25]. Because the Higgs couplings to fermions are proportional to the fermion mass, this process is dominated by production of the Higgs through an intermediate top quark loop. A Higgs boson produced in this mode may be produced in association with additional high transverse momentum jets, resulting from the fragmentation of additional partons radiated from the fermion loop or the initial state gluons, as well as any additional soft hadronic activity. The next most prevalent production mode is the Vector Boson Fusion process [26], in which the Higgs is produced by W or Z boson pairs which have been radiated from initial state quarks. Since the initial state quarks are typically valence quarks carrying a

high fraction of the proton momenta, this typically results in a Higgs produced in association with two high energy forward jets in opposite sides of the detector. Because there is no colour flow between the quarks in this process, there is typically little hadronic activity in the central region for such events. Nevertheless the presence of the two forward jets recoiling against the Higgs boson tends to impart higher transverse momentum than for gluon fusion production. After VBF, the next largest Higgs production mode is associated production with a W or Z boson [27], in which a W or Z boson is produced, at leading order by a quark-antiquark pair, and radiates a Higgs boson. The recoiling W or Z boson tends to impart as well higher transverse momentum to the Higgs than in the gluon fusion production. Finally, the next largest production mode is associated production with a top anti-top pair [28, 29, 30, 31]. Other possible Higgs production modes have very small cross sections and are relevant only at very high integrated luminosity. Leading order Feynman diagrams from the four production mechanisms discussed above are shown in Figure 1-1. Graphs showing the production cross section per mode as a function of Higgs mass for center of mass energies of 7 TeV and 8 TeV are shown in Figure 1-2.

Once a Higgs has been produced, the branching ratios for the various possible decay modes depend on the couplings, fully determined in the SM, and the available phase space for the decays, determined by the unknown Higgs mass. In addition to direct decays to the SM fermions and vector bosons, additional decay modes are possible through loops, allowing the Higgs to decay to a pair of gluons, coupling through a fermion loop, and to a pair of photons, or a Z boson plus a photon, proceeding through an interfering combination of fermion and W loops. Leading order Feynman diagrams for the two contributions to the Higgs to di-photon decay are shown in Figure 1-3. The SM Higgs branching ratios, along with the total width are shown for the full range of possible Higgs masses in Figure 1-4.

At low Higgs masses, below about twice the W mass, the decay width is dominated by decays to pairs of b quarks, τ leptons, and c quarks. At these masses, decays to pairs of W and Z bosons are suppressed by the phase space requirement that the W or Z bosons be produced far off mass shell. At these masses there is no phase space for

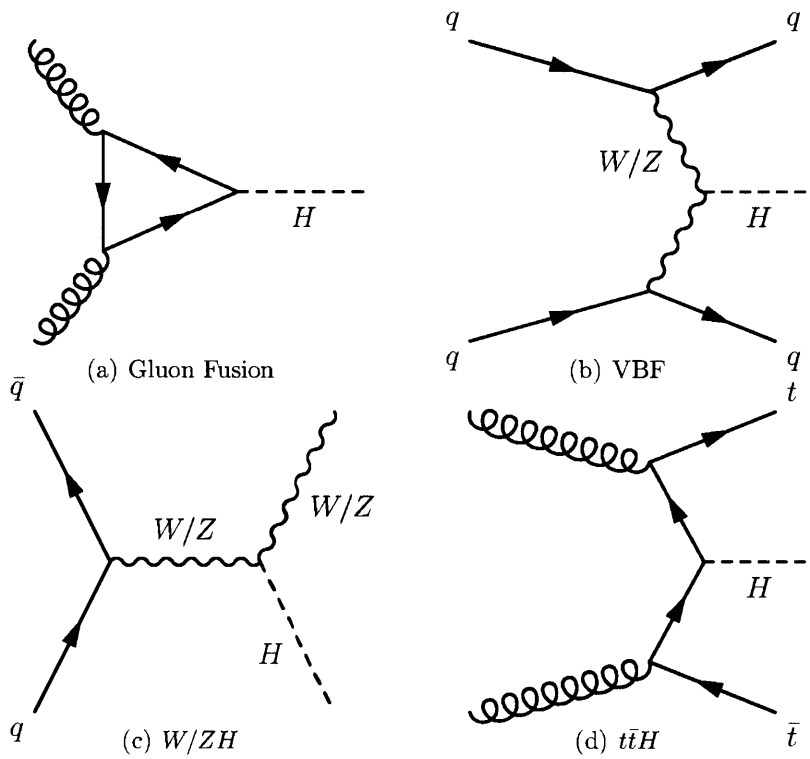


Figure 1-1: Leading order Feynman diagrams for the four leading Higgs production modes at the LHC.

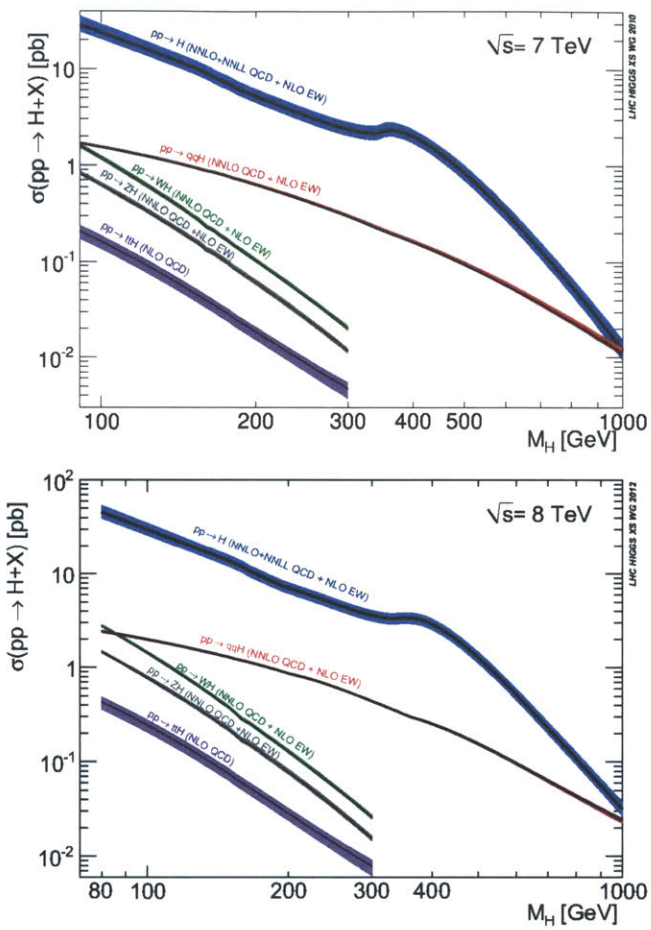
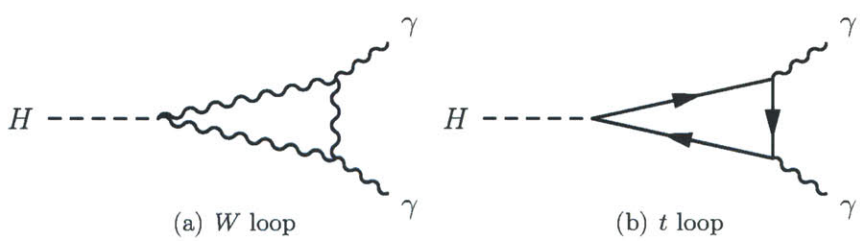


Figure 1-2: Higgs production cross sections broken down by production mode for proton proton collisions with center of mass energy 7 TeV (top) and 8 TeV (bottom) [32, 33, 34].



(a) W loop (b) t loop

Figure 1-3: Leading order Feynman diagrams for the Higgs decay to di-photon pairs, showing both the W-loop contribution (left) and the fermion, predominantly top quark, loop contribution (right).

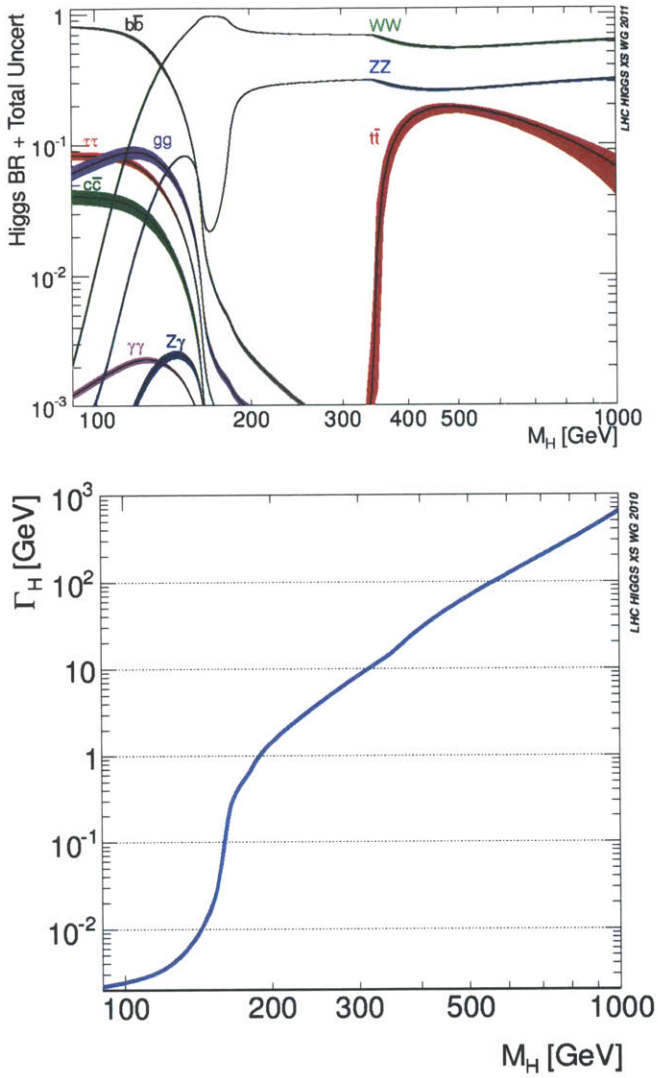


Figure 1-4: Higgs branching ratios for most possible decays (top) total width (bottom) [32, 33, 34]. Branching ratios for Higgs decays to pairs of light quarks, muons, and electrons are not shown, and are significantly smaller than those indicated, due to the small fermion masses in such decays.

decays to top quarks, although decays through predominantly top loops do contribute a substantial decay width to pairs of gluons. In addition, there are the W/fermion loop-induced decays to pairs of photons, or a Z boson plus a photon as discussed. For searches at hadron colliders, all-hadronic final states suffer from large backgrounds from QCD multijet production. The rate of useful events from W and Z pairs is therefore further reduced by the branching ratios of the W and Z bosons to charged leptons. In addition, searches involving $W \to \ell\nu$ final states are made challenging by the presence of missing energy carried by the neutrinos. The two photon decay offers a unique opportunity for a fully reconstructed final state, with relatively high total cross section compared to Higgs decays to Z bosons with 4 leptons in the final state, and much smaller backgrounds than partly or fully hadronic final states. In addition, in the low Higgs mass range, its total width is very narrow, less than about 10 MeV. This allows the reconstruction of a narrow mass peak, depending on the detector resolution. Because the boson and fermion loop contributions to the Higgs to di-photon decay interfere destructively, the precise value of the decay rate to two photons is sensitive to small deviations in the couplings from the SM predictions, making the measured rate of Higgs to di-photon events sensitive to new physics.

1.3 The CMS Experiment

Although the Higgs search is a major physics goal of the CMS experiment, the experiment is designed to be general purpose, and capable of measuring a wide range of SM and possible new physics processes produced at the LHC. General considerations include excellent momentum and energy resolution, and identification capabilities for muons, electrons and photons, as well as good jet and missing energy resolution. The close bunch spacing of as little as 25 ns requires detectors with good time resolution, at least enough to resolve the individual bunch crossings. Furthermore, the large number of charged particles per interaction, as well as the large number of interactions per crossing, requires a detector with high spatial granularity in order to be able to individually reconstruct charged particles from the interactions.

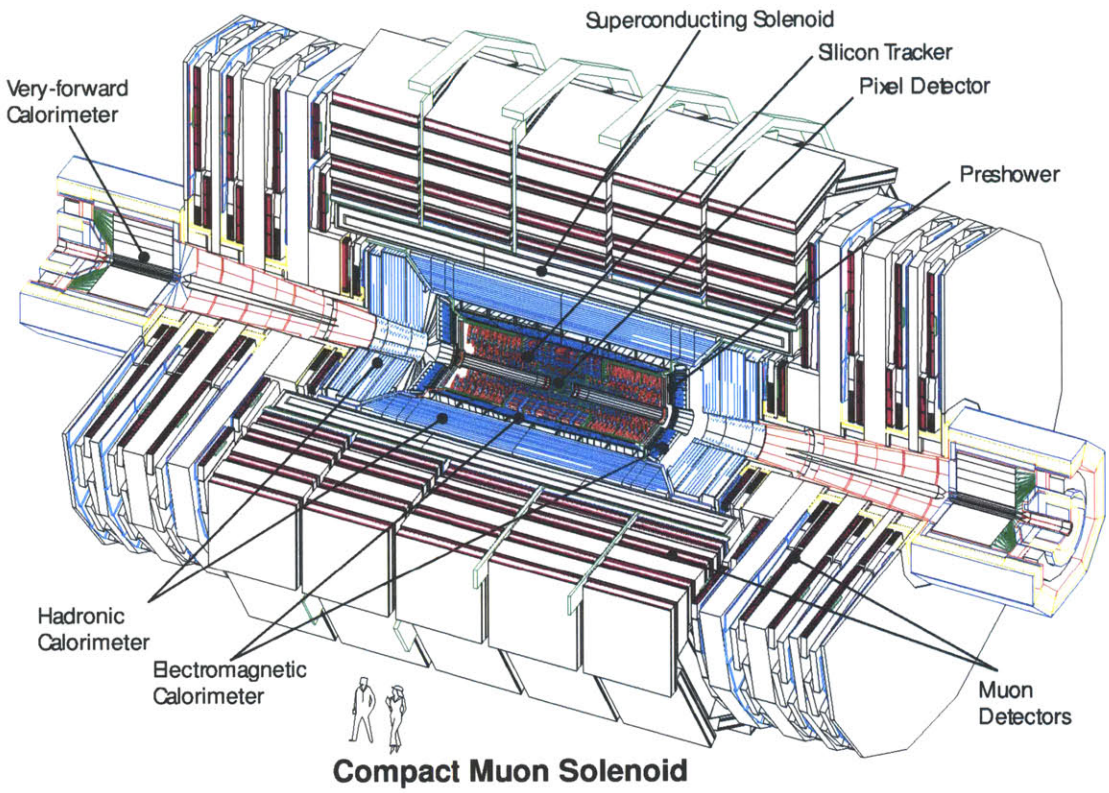


Figure 1-5: A diagram of the CMS detector, showing a cutaway of the interior, with the solenoid and various subdetectors labelled.

The overall design of CMS is centered around a 3.8 T superconducting solenoid magnet with an inner diameter of 6 m. Placed inside the solenoid are the inner tracking detectors, as well as electromagnetic and hadronic calorimeters. Outside the solenoid are muon tracking detectors as well as iron return yokes for the magnetic field, in order to ensure a strong enough field in this region to be able to measure the curvature of muons in the outer detector region. A diagram of the full detector is shown in Figure 1-5.

The overall geometry of the detector is cylindrical in nature, with subdetectors in general divided into concentric cylindrical barrel-shaped detectors covering the central region, and disc-shaped endcap detectors covering the forward region. The coordinate system of CMS consists of an origin located at the nominal interaction point, an $x-$axis in the horizontal plane, pointing radially inwards towards the centre of the

LHC, a $y-$axis pointing vertically upwards, and a z-axis running along the nominal beamline, with the $+z$ direction towards the Jura. Polar coordinates are frequently used, with the azimuthal angle ϕ in the x-y plane measured counterclockwise from the positive x axis, and the polar angle θ measured with respect to the positive z axis. The polar angle is usually expressed in terms of the pseudorapdity, defined as $\eta = -\ln\tan\theta/2$

1.3.1 Inner Tracking Detectors

The purpose of the inner tracking system[35, 36] is to measure the trajectory of charged particles, including a precise determination of the their momentum as well as their position, extrapolated either to the beamline or to the calorimeter. The inner tracking detectors in CMS consist entirely of solid state silicon based detector. In order to provide the most precise position measurement and separation of charged particles near the interaction point, the inner part of the tracking detector consists of a silicon pixel detector which provides precise precision measurements in three-dimensions. The pixel detector consists of three barrel layers, referred to as the pixel barrel region, located at radii of 4.4, 7.3, and 10.2 cm and extending from $z = -26.5$ cm to $z = +26.5$ cm. In addition, there are two endcap pixel layers, referred to as the pixel forward region. These are located at $\pm z$ =34.5, 46.5 cm, and cover a region between approximately 6 cm and 15 cm in radius from the beam.

The outer part of the tracking detectors consists of a silicon strip tracker, which provides measurements precisely localized in only two-dimensions, with most strips oriented perpendicular to the ϕ direction. In the barrel, this consists of the Tracker Inner Barrel (TIB) region, comprised of four layers between 20 and 55 cm in radius, as well as the Tracker Outer Barrel (TOB) region, consisting of an additional 6 layers between 50 and 116 cm in radius. In the endcap region, the strip tracker consists of the Tracker Inner Disk (TID) region of three layers between $|z|$ of 80 and 90 cm, plus a Tracker EndCap (TEC) region of nine layers located between $|z|$ of 124 and 280 cm. A fraction of the layers include double layered modules, with a second set of strips oriented at an angle of 100 mrad with respect to the first. The combination with these

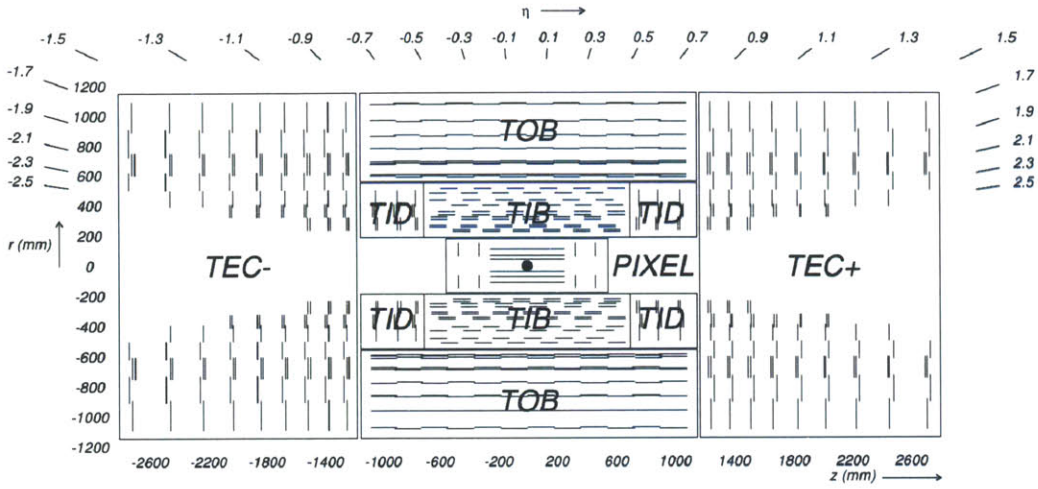


Figure 1-6: A diagram of the CMS inner tracking detectors, showing the layers of the silicon pixel and strip tracking detectors.

stereo measurements can give a position measurement in the third dimension with a precision ranging from 230 to 530 μm. A schematic view of the tracking detectors, labeled by region, is shown in Figure 1-6. The combined tracking detector system provides coverage up to $|\eta| = 2.5$, with an average of 13-17 measurements per charged particle, depending on the pseudorapidity region.

The pixel detector consists of 66 million pixel elements, each $100 \times 150\mu\text{m}^2$ in dimension, spread across 1440 modules. Each pixel consists of a $p-n$ semiconductor junction. When a charged particle crosses the junction, it excites electron-hole pairs, and the charge is collected by the readout electronics connected to the junction. In order to keep the data volume reasonable given the very large number of channels, zero suppression is performed by electronics on the sensor modules, in which only pixels with signal above a set threshold are read out. A charged particle crossing the module will generally deposit charge in at least two adjacent pixels, with the amount of charge deposited in each pixel inversely related to the distance between the particle position and the pixel. A measurement of the charge sharing between adjacent pixels therefore allows a single hit position resolution substantially smaller than the dimensions of a single pixel. In order to exploit the sharing of charge among

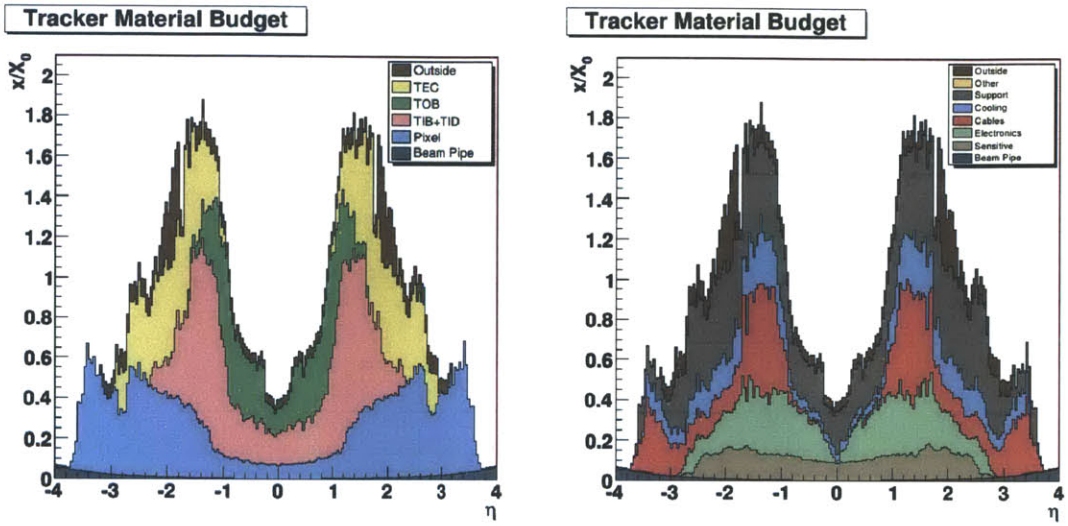


Figure 1-7: The amount of material in the inner tracking detectors, measured in units of radiation lengths, and broken down by detector region (left) and category (right).

adjacent pixels, the signal amplitude is digitized with 5 to 8 bits of information, allowing a single hit position resolution of $15 - 20$ μm.

The silicon strip detector consists of about 9.3 million strips across 15148 modules, with strips as well consisting of p-n junctions across which charge carriers are ionized by charged particles as they cross the strip. Depending on the region of the detector, the strip pitch varies between 80 and 184 μm. By exploiting charge sharing between strips, analogous to charge sharing between adjacent pixels, the single hit resolution along the ϕ direction ranges from 23 to 53 μm, smaller than the strip pitch.

The large amount of silicon in the inner tracking detectors, combined with the sophisticated electronics required leads to a substantial requirement for cabling and cooling services. This leads to a relatively large amount of material in the detector. The estimated material budget, as a function of pseudorapidity, is shown in Figure 1-7. The estimated total material budget ranges from about 0.4 radiation lengths in the very central barrel, to a peak of about 1.8 radiation lengths in the vicinity of $|\eta| = 1.5$, near the barrel-endcap transition region.

1.3.2 Electromagnetic Calorimeter

The role of the electromagnetic calorimeter (Ecal)[37, 38] is to measure the energy of incident electrons and photons. In order to achieve the best possible energy resolution, the CMS Ecal is a homogeneous and nearly hermetic calorimeter. It is made of lead tungstate ($PbWO_4$) crystals, coupled to photodetectors. There are 68524 crystals in total, divided between the Ecal barrel detector (EB) and the Ecal endcap detector (EE). Incident electrons and photons initiate showers inside the crystals, and the showering particles produce scintillation light as they interact the the crystal. This scintillation light is then measured by the photodetectors, and the amount of scintillation light is used to determine the energy deposited in each crystal. In the endcap region, a sampling preshower detector made of lead sampling and silicon active layers is installed in front of the Ecal in order mainly to improve angular resolution for photon/π^0 separation.

Lead tungstate has been chosen for the crystal material on account of its high density, and corresponding short radiation length and small Molière radius, while maintaining fast scintillation response with reasonable light output and transparency at optical wavelengths. The short radiation length and small Molière radius allows for a compact calorimeter design, while still containing the electromagnetic showers both longitudinally and transversely.

The EB detector consists of 61200 crystals arranged in a 90x360 $\eta - \phi$ grid, with coverage up to $|\eta| = 1.479$. In order to maintain hermetic coverage of the detector, the crystals must be tightly packed. In addition, to minimize the leakage of the electromagnetic shower in the small gaps between crystals, the EB geometry is such that crystals are tilted at an angle of 3 degrees with respect to the trajectory of particles incident from the nominal interaction point. In order to accommodate this tilt, together with the tight packing requirement, EB crystals are shaped as truncated pyramids, with a number of different particular variations needed depending on the precise location in EB. The crystals are on average roughly 22x22 mm^2 at their front face, and 26x26 mm^2 at their back face, comparable to the Molière radius, such that a

large fraction of the energy from an electromagnetic shower is expected to be contained within a radius of a few crystals with respect to the crystal on which the electron or photon was incident. Each crystal is 23 cm long, corresponding to 25.8 radiation lengths of material, such that longitudinal leakage of the electromagnetic showers are negligible. The back of each crystal is attached to an Avalanche Photodiode (APD), that detects the scintillation light from the crystals. The EB crystals are arranged into 36 supermodules, 18 for each of the $+z$ and $-z$ sides of the detector, such that each supermodule is 20 crystals wide in ϕ. Each supermodule consists of 4 modules, spanning from $\eta = 0$ to $\eta = \pm 1.479$. These modules are enumerated 1-4 from the center of the detector outwards. Module 1 in each supermodule consists of 25x20 crystals in the ηxϕ directions, whereas modules 2,3 and 4 in each supermodule consist of 20x20 crystals.

The EE detector consists of an additional 7324 crystals. The geometry in the endcap is much simpler than in the barrel, with crystals arranged in an $x - y$ grid in groups of 5x5 crystals, such that all EE crystals share the same geometry. This naturally produces an angle between the crystal axes and the trajectory of particles from the interaction point of between 2 and 8 degrees. EE crystals are 28.62x28.62 mm^2 at the front and 30x30 mm^2 at the back, with a length of 22 cm, corresponding to 24.7 radiation lengths. Because of the higher radiation dose in the endcap region, Vacuum Photo Triodes (VPT's) are used as photodetectors for the EE crystals in place of APD's. The EE coverage extend from $1.479 < |\eta| < 3.0$.

The preshower detector sits in front of the EE in the region from $1.653 < |\eta| < 2.6$. It consists of two alternating layers of passive lead and active silicon, acting as a sampling calorimeter. The first lead layer corresponds to two radiation lengths of material, where the second layer corresponds to one additional radiation length. The silicon layers consist of active silicon strips with a pitch of 1.9 mm. The additional spatial resolution provided by the preshower is designed in principle to improve the separation between prompt photons and neutral mesons in the endcap region. The overall layout of the Ecal subdetectors are shown in Figure 1-8

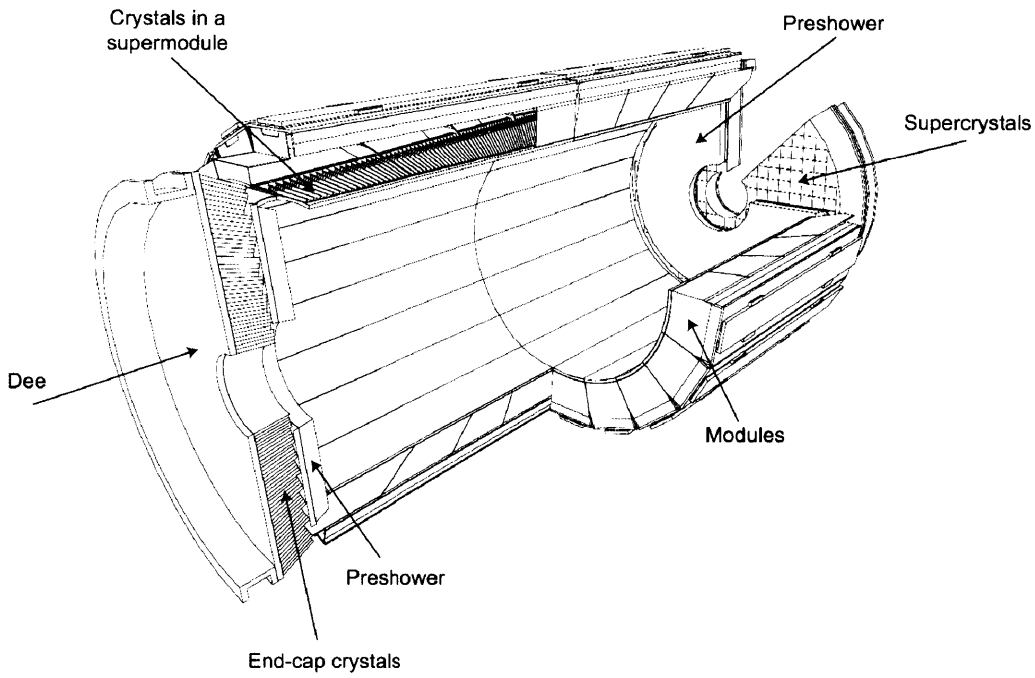


Figure 1-8: The layout of the barrel and endcap Ecal subdetectors, along with the preshower detector.

Readout and Energy Reconstruction

For each event accepted by the level one trigger, and after appropriate zero suppression, the signal from each crystal is read out in Analog to Digital Converter (ADC) counts for ten consecutive 25 ns bunch crossings. The uncalibrated energy in a single crystal is reconstructed in units of ADC counts by a weighted fit to these ten digitized samples. The readout pipeline is aligned such that for particles originating from the triggered bunch crossing, the beginning of the pulse starts in the fourth sample. The first three samples are therefore signal-free, aside from possible contamination from particles from collisions in preceeding bunch crossings. This allows the first three samples to be used for an unbiased determination of the pedestal, which may vary event by event. The final 7 samples are then used to reconstruct the amplitude of the pulse. The per-sample weights used to reconstruct the pulse amplitude are corrected at reconstruction time by per-crystal time calibration constants to account for timing differences in the electronics chains. The reconstructed energy in a single crystal is referred to as a Reconstruction Hit (RecHit). The RecHit amplitude is converted from ADC counts to GeV using a global ADC to GeV constant averaged over the entire detector, giving an uncalibrated RecHit energy.

Ecal information for use by the Level 1 trigger is read out along a parallel path, in blocks of 25 crystals and for every LHC bunch crossing. A rough version of the pulse shape fitting, as well as the summation over the 25 crystals is done in hardware, such that an estimate of the total energy in each 25 crystal trigger tower is output for use in the final trigger decision for each event. There is one additional bit associated with each trigger tower, related to the compatibility of the shower profile with an electron or photon deposit. This fine grained veto bit is also computed in hardware, based on the ratio of the highest energy 5x2 crystal region in $\eta - \phi$ compared to the total energy of the trigger tower.

Single Crystal Response and Calibration

There are a number of effects which lead to differences in response from crystal to crystal in the detector. These are broadly divided into time-independent and time-dependent effects. Time-independent effects include differences in response due to crystal manufacture, positioning or orientation in the detector, electronics, or other effects. The main time-dependent effect results from the dynamic behaviour of the crystal transparency under exposure to ionizing radiation. In general the crystals lose transparency under exposure to ionizing radiation from the collisions. This transparency loss will also dynamically recover however, during periods in which there are no collisions, or collisions at reduced instantaneous luminosity. The radiation dose varies with location in the detector and especially with pseudorapidity, but the precise transparency loss dynamics are also sensitive to variations in crystal manufacture, and therefore also vary in general from crystal to crystal. Other effects, such as radiation-induced response changes in the endcap VPT's may also contribute to time dependent response changes. The corrections are factorized into a time-independent per crystal intercalibration constant, and a per-crystal time dependent correction.

The time-independent intercalibration constants are determined using a combination of π^0 and η mass reconstruction, ϕ-symmetry in minimum bias events, and E/p calibrations with electrons from W and Z decays. The additional per-crystal and time-dependent corrections are derived instead from a dedicated laser monitoring system. This system injects laser and LED light of known intensity into the crystals during the LHC abort gap. The measured changes in output are used to derive the time-dependent corrections. The calibration sequence is such that the system is able to cycle through all crystals in the detector approximately every 30 minutes, allowing time-dependent corrections to be derived and updated on this time scale [39].

In order to have a consistent factorization between calibration constants and other corrections which are applied later in the reconstruction, it is necessary to have a precise definition for the normalization of the calibration constants. The normalization for the intercalibration constants in the barrel is defined such that the calibrated en-

ergy sum of a 5x5 grid of crystals gives on average the correct energy for a 50 GeV photon incident in module 1 in trigger towers not bordering the module gaps. The corresponding normalization in the endcap is defined such that the calibrated sum of a 5x5 grid of crystals plus associated preshower depositions gives on average the correct energy for an unconverted 50 GeV photon incident on the preshower + Ecal endcap [40].

Once the calibration constants have been applied, the result is a calibrated RecHit energy which is used for higher-level reconstruction steps.

1.3.3 Hadronic Calorimeter

The Hadronic Calorimeter (Hcal)[41] measures the energy of charged and neutral hadrons, critical for the reconstruction and measurement of jets and missing energy. In the forward region beyond the coverage of the Ecal, the Hcal system is also responsible for the measurement of electromagnetic energy. The Hcal consists of a barrel detector (HB) covering the region up to $|\eta| = 1.3$, an endcap detector (HE) covering the region from $1.3 < |\eta| < 3.0$, and a forward detector (HF) covering the region from $3.0 < |\eta| < 5.2$. In the region covered by the HB, there is an additional outer detector (HO) which is placed outside of the magnet solenoid in order to ensure measurement and containment of the tails of hadronic showers, which may extend beyond the HB and the solenoid material. Including the HO and the solenoid, the total amount of material in the region covered by the HB and HE subdetectors corresponds to at least 11.8 hadronic interaction lengths, except for a small region at the transition between the HB and HE. The overall layout of the Hcal subdetectors are shown in Figure 1-9.

The HB and HE detectors are both sampling calorimeters consisting of a combination of brass and steel absorber layers, with plastic scintillator active layers. Wavelength-shifting fibres carry light from the scintillators to hybrid photo-diodes (HPDs) from which the signal is digitized and read out. The HO detector similarly uses plastic scintillators with wavelength-shifting fibres and HPDs, but uses iron absorber plates, as well as the solenoid itself, which is effectively used as an additional absorber layer.

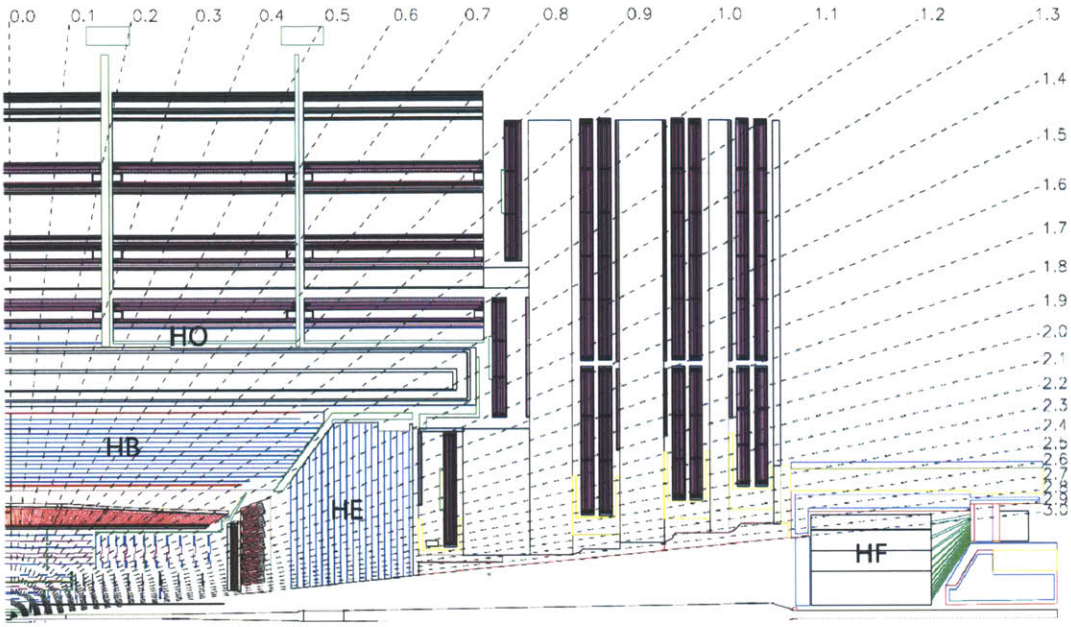


Figure 1-9: The layout of the various Hcal subdetectors showing their respective coverage in pseudorapidity.

The much harsher radiation environment in the forward region necessitates a somewhat different design for the HF detector, where scintillating quartz fibres, read out by photomultiplier tubes, are used together with steel absorber layers. The quartz scintillators are sensitive to both electrons/photons as well as hadrons, where longitudinal segmentation into two depths provides limited discrimination between electromagnetic and hadronic deposits.

1.3.4 Muon Detectors

Although not used in this analysis, the muon detectors[42] are an important part of CMS, providing triggering and identification, as well as augmenting momentum determination for muons. Since the Ecal and Hcal detectors stop nearly all electrons, photons and hadrons, the detection of a charged particle in the muon detectors is already a strong indicator that it is a muon. The primary muon detectors comprise of the Drift Tube (DT) system in the barrel ($|\eta| < 1.2$) region, as well as the Cath-

ode Strip Chamber (CSC) detector covering the region $0.9 < |\eta| < 2.4$. There is an additional Resistive Plate Chamber (RPC) system covering the region $|\eta| < 1.6$ which provides better time resolution than the DT or CSC system, and augments the performance of the trigger for muons.

1.3.5 Data Acquisition and Trigger

The triggering of events in CMS is divided into two steps, a Level One (L1) trigger[43], implemented in hardware and firmware, and a High Level Trigger (HLT)[44] implemented in software, using the same framework as the offline reconstruction, but with optimized algorithms and configuration. The overall logic is that the full detector is read out and events fully assembled only for events which pass the L1 trigger. Such events are fully assembled in the data acquisition system and then sent to the HLT, where they are either discarded, or accepted and passed on to the offline computing system for storage and offline reconstruction. With 25 ns spacing, the LHC bunch crossing rate is 40 MHz. Because there are more than one proton-proton interaction per crossing, the total inelastic collision rate is much higher, on the order of 1 billion interactions per second at design luminosity. The nominal L1 accept rate for CMS is 100 kHz, limited by the speed at which the detector electronics can be read out, and the bandwidth with which events can be assembled and passed tot he HLT. The nominal HLT accept rate is 300-400 Hz, limited by available offline storage and processing resources. In exceptional circumstances, HLT accept rates of up to several kHz are possible for short periods of time, or if not all of the data is intended to be processed offline immediately.

The L1 trigger is implemented primarily in hardware and firmware, using a combination of Application Specific Integrated Circuits (ASICs) and Field Programmable Gate Arrays (FPGAs). The L1 system processes information from every single LHC bunch crossing, at a rate of 40 MHz, and makes a decision whether to accept or reject each event. Information for use by the L1 trigger is available from the Ecal, Hcal and Muon detectors. Because this data has to be read out and processed at such a high rate, the precision and granularity of the information used by the L1 is reduced

with respect to the full readout. The full detector information is stored instead in ring buffers or similar within the readout electronics of each subdetector, and is only read out for events which are accepted by the L1. The number of channels and the speed of the readout electronics in the tracking detectors does not currently allow information from those detectors to be used as part of the L1 trigger. The trigger is in general capable of selecting events based on clusters of Ecal and Hcal deposits (such as electrons/photons, or jets), muons, or quantities constructed from global energy sums in the Ecal and Hcal, such as summed transverse energy, or missing transverse energy.

In the L1 trigger, electrons and photons are essentially indistinguishable on account of the absence of tracking information. Electron/photon triggers at L1 are based on the summed transverse energy over two Ecal trigger towers (50 crystals in total), with requirements on the fine grained veto bit as described in Section 1.3.2 for the two towers and with a configurable cut on the total transverse energy. In order to suppress spurious electron/photon triggers from electromagnetic energy in jets, an additional requirement is imposed in the L1 on the fraction of energy present in the Hcal trigger towers immediately behind these two Ecal towers. The L1 trigger in general allows to combine trigger decisions based on a combination of different objects. The analysis relies on single electron/photon and double electron/photon triggers at L1 only.

The High Level Trigger is implemented using a farm of several hundred commercially available computer systems, consisting of several thousand CPU cores. The software framework used for the HLT is identical to that used for offline processing, but optimized algorithms and configurations are used in order to allow processing at the required 100 kHz event rate. The general strategy in the HLT is to run specific processing for each event depending on which L1 trigger algorithms it was accepted by. Regional unpacking and reconstruction of the data, in regions of interest defined by the L1 trigger decision, is exploited in order to further reduce the required processing time. The HLT algorithms used for this analysis are based on regional reconstruction of photons and isolation sums around one or two triggered L1 pho-

tons. For events accepted by a single photon L1 trigger, the second photon candidate is located by globally unpacking and reconstructing the Ecal data, with subsequent isolation quantities using regional reconstruction around that candidate.

1.4 Analysis Overview

In this analysis we search for the production of the standard model Higgs in the di-photon decay channel. Given the existing exclusion, and considering the range of Higgs masses for which the branching ratio to photons is appreciable, the search is conducted for Higgs masses between 110 GeV and 150 GeV. The general analysis strategy is to search for a peak in the di-photon mass distribution on top of a large, but smoothly falling background. Given the small natural width of the SM Higgs in this mass range, the width of the mass peak is expected to be driven entirely by the detector resolution. A large irreducible background is present from QCD di-photon production in both quark and gluon initial states. Leading order diagrams for the quark-initiated Born di-photon production and the gluon initiated box di-photon production are shown in Figure 1-10. Despite the fact that the gluon-induced process does not exist at tree-level, the contribution is comparable to the quark-induced production, given the higher gluon-gluon luminosity at the LHC as compared to quark-antiquark. For an analysis carried out inclusively with respect to additional jet production, an additional source of irreducible background are QCD di-jet or photon + jets production with additional photons produced by Initial State/Final State Radiation (ISR/FSR), example diagrams for such processes are shown in Figure 1-11. In addition to the irreducible background containing two prompt photons, there is a reducible background from QCD di-jet and photon-jet production where one or more of the reconstructed photons arise from a quark or gluon jet. The primary mechanism for a jet to fake a photon involves the fragmentation of the majority of the jet energy into a leading π^0 or η meson, which subsequently decay promptly with large branching fractions to two photons. For the energy range of interest, the π^0 or η are significantly boosted, such that the two photons from the decay are nearly collinear

and are difficult to distinguish from a single photon incident on the calorimeter. Suppressing these backgrounds relies in general on applying isolation requirements against additional energy from the jet fragmentation, as well as exploiting differences in shower profile in the calorimeter to distinguish between a single incident photon and a pair of incident photons from a neutral meson decay.

In addition to the di-photon mass peak, and photon-jet discrimination, the di-photon system from signal and background have different kinematics due to differing initial states and spin correlations imposed by the scalar nature of the SM Higgs. These variations in signal to background ratio, combined with variations in the detector resolution as a function of detector region and the interaction of photons with material upstream of the Ecal are exploited to optimize the sensitivity of the search. Finally the forward jet topology of the VBF Higgs production is exploited to select a VBF-enriched set of signal events, which are also accompanied by less background than the inclusive search, albeit with a lower Higgs production cross section.

The main elements of the analysis are the primary vertex reconstruction and selection, photon reconstruction and selection, the use of multivariate techniques to optimally select and categorize the events based on expected signal purity and mass resolution, the reconstruction and tagging of forward jets for the VBF event categories, and finally the extraction of the signal and statistical analysis of the final di-photon mass distributions in the final event categories.

Because electrons behave similarly to photons in terms of their interaction with the detector, and because $Z \rightarrow ee$ events provide a clean sample of two-electron events with a similar topology and a mass just below the search region, these events are used extensively to measure, calibrate, and validate the properties of photons and di-photon events in the data.

1.4.1 Multivariate Analysis Techniques

The complex environment of the LHC and the corresponding sophisticated nature of the CMS detector produce a large amount of information associated with each collision event. Searches such as this one are statistically limited, and therefore profit

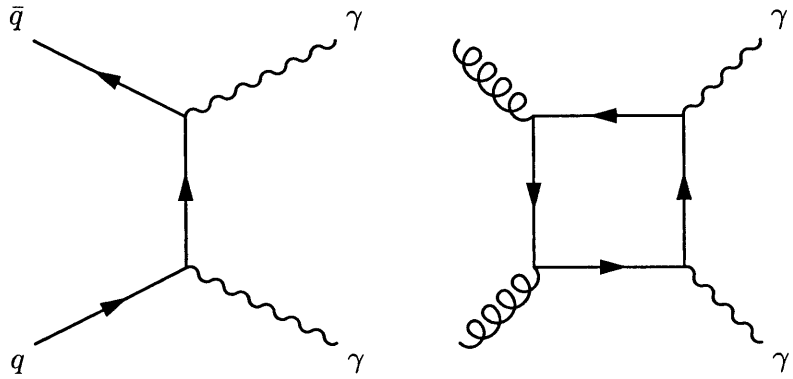


Figure 1-10: Leading order Feynmann diagrams for QCD diphoton production from quark (left) and gluon (right) initial states.

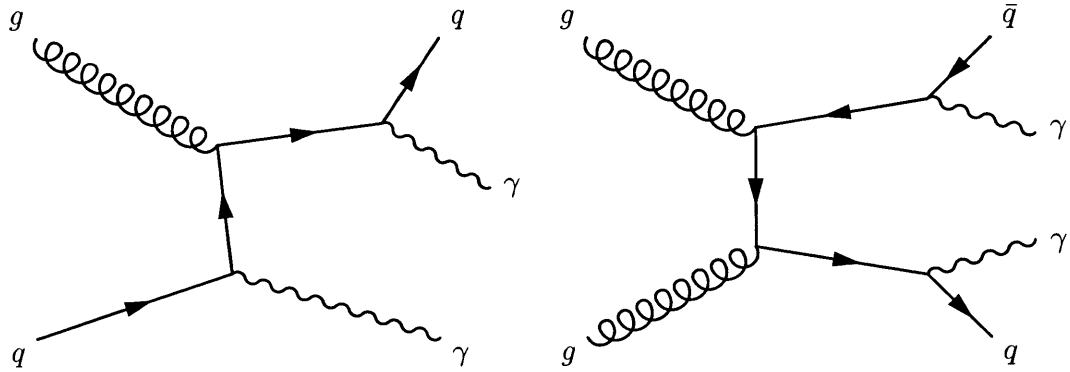


Figure 1-11: Example tree-level Feynmann diagrams for QCD diphoton production in association with one or two additional jets.

from maximally exploiting the available information. The general problem can be reduced to distinguishing as well as possible between signal events and background events. A very simple approach to this problem is to encode information about the event into a set of one-dimensional variables, and then apply simple cuts on those variables, using in the analysis events which pass the cuts, and discarding those which fail. The simplest variation on this is a "cut and count" analysis, in which the analysis result is determined purely from the number of events passing a set of cuts. This approach is easy to understand and to describe and model, but possesses a few important shortcomings:

1. Events are either retained or discarded. In principle events can be used in a more fine-grained way, according to how signal-like or background-like their features.

2. Correlations between variables are neglected. In general it is difficult to encode all of the relevant information into fully uncorrelated variables, and therefore neglecting correlations between one dimensional variables leads to a loss of information.

3. The selection of cut values which optimize the performance of an analysis is an extremely difficult problem, especially with a large number of variables. Most cut-based analyses therefore rely on a relatively small number of variables and/or use suboptimal cut values, again neglecting potentially useful information.

It is generically true that the optimal separation between two classes of events given a set of variables $\bar{x}$, is given by the likelihood ratio[45]

$$L_R = \frac{\mathcal{L}_s(\bar{x})}{\mathcal{L}_s(\bar{x}) + \mathcal{L}_b(\bar{x})} \tag{1.1}$$

where $\mathcal{L}_s(\bar{x})$ and $\mathcal{L}_b(\bar{x})$ represent the full multidimensional likelihood functions for signal and background events respectively. If the set of variables $\bar{x}$ encode all of the relevant information, then this likelihood ratio contains all of the relevant information

for distinguishing the two classes of events. This ratio quantifies the probability that a given event with features $\bar{x}$ is a signal event as opposed to a background event. This function could be used in conjunction with a maximum likelihood fit for a signal yield for example, in order to have the optimal expected discovery significance in a search.

Decision Trees for Classification

It is generally the case in high energy physics that the input variables $\bar{x}$ are the output of a complicated detector response, and therefore there is no known analytic form for the multidimensional likelihood functions above. It is a common situation however, to have access to a finite sample of events representing each class, either from Monte Carlo simulation, or from carefully defined control regions in data. There are a number of techniques designed to estimate the multidimensional likelihood ratio from a finite set of training events in this case. One such technique is the Boosted Decision Tree [46].

A Decision Tree is a simple construct consisting of a set of "nodes". Each node is either an intermediate or a terminal node. Intermediate nodes are defined by a split value associated with a particular variable, and links to two subsequent nodes. Terminal nodes are defined by a response. To determine the response for a particular event, the event is first passed to the initial "root" node of the tree. If the variable corresponding to the splitting value is less than or equal to the split value, the event is passed on to the "left" daughter node, otherwise it is passed on to the "right" daughter node. When the event reaches a terminal node, it is assigned the response value corresponding to that node. Decision trees therefore provide an efficient structure for defining hypercube-shaped regions in a multidimensional space, in order to classify or assign a response value to events based on which region they occupy.

A single decision tree can be used to approximate the likelihood ratio above by minimizing the function

$$L = -\sum_{s} \ln \frac{\mathcal{L}_s(\bar{x})}{\mathcal{L}_s(\bar{x}) + \mathcal{L}_b(\bar{x})} - \sum_{b} \ln \frac{\mathcal{L}_b(\bar{x})}{\mathcal{L}_s(\bar{x}) + \mathcal{L}_b(\bar{x})} \tag{1.2}$$

over the set of training events, where each event is labeled as signal s or background b. Each event is assigned a response value F, defined such that

$$e^{2F} \equiv \frac{\mathcal{L}_s(\bar{x})}{\mathcal{L}_b(\bar{x})} \tag{1.3}$$

with each event assigned an initial response value of zero, corresponding to no discrimination between signal and background. The loss function can then be expressed as

$$L = -\sum_s \ln \frac{1}{1+e^{-2F}} - \sum_b \ln \frac{1}{1+e^{2F}} \tag{1.4}$$

Minimization proceeds by a steepest descent algorithm, in which events are grouped according to their gradient $\frac{-\partial L}{\partial F}$. This is practically accomplished by an optimal split search at each node of the tree, choosing the splitting variable and value in order to minimize average variance of the gradient. Tree growth is terminated either by a cutoff on the number of terminal nodes, depth of the tree, or minimum number of events on each terminal node. Each terminal node is assigned a response in order to minimize the loss function L for the events on that node. The end result is that each event is assigned a response value F, which best approximates the likelihood ratio in 1.1, given for each event by the expression $\frac{1}{1+e^{-2F}}$. Once the tree has been constructed, this likelihood ratio is estimated for additional events simply be determining their response from the decision tree.

Gradient Boosting

In practice, single decision trees are sensitive to statistical fluctuations in the training data, and are not able to fully explore the multidimensional phase space without their response being driven by such fluctuations, an occurrence referred to as overtraining. The practical solution to this is "Boosting". In the gradient search formulation above, the boosting procedure is simply to form an additive series of decision trees, in which the training of each tree begins using the response determined from the existing series of trees. Since the final response is determined by a possibly large number of trees, it is not necessary for each individual tree to fully explore the multidimensional phase

space. The problem of overtraining can therefore be mitigated by limiting the depth of each of the individual trees in the boosting series, referred to as the forest. An additional handle for reducing overtraining is the addition of a shrinkage factor, in which the optimal response for each tree is multiplied by a factor between zero and one, in order to de-weight the contribution of each individual tree. In order to compensate for this, typically a larger number of trees are needed in the forest. Empirically for classification problems it has been found that optimal performance is obtained with a large number of small trees, and using shrinkage factors of order 0.1. Boosted decision trees constructed in this way are extremely robust to the addition of variables which are either redundant, or only contribute additional information in a limited region of the input variable space. In general this allows the selection of variables based on the physical information which they contain, with the final performance being relatively insensitive to the precise definition or particular combination of variables.

Classifier Usage

In a narrow technical sense, a classifier based on a Boosted Decision Tree trained on a set of variables containing all of the relevant information itself contains all of the information needed to discriminate between signal and background. The simplest possible usage of such a BDT would be to place a cut on the output. Cutting on the BDT output rather than on the input variables directly addresses two of the three limitations listed for cut and count analyses, namely that it properly exploits correlations among the input variables and is relatively simply to optimize. This cut-based usage still retains the drawback of completely accepting or discarding events. As a proxy for the likelihood ratio, the BDT contains event-by-event information on the signal probability which can be further exploited, for example by performing a maximum likelihood fit for the signal yield. In fact, if the shape of the BDT output is perfectly known for both signal and background events, then such a likelihood fit approaches the sensitivity of the full multidimensional likelihood ratio which the BDT has been trained to approximate.

This leads directly to the main challenge of multidimensional analysis techniques.

Since multivariate classifiers such as a BDT are able to train on fine details in the multidimensional distribution of input variables, they are sensitive to proper modeling of those distributions by the Monte Carlo simulation. In addition, since the mapping of the input variable space onto the classifier output is non-trivial and sometimes difficult to understand or visualize, interpreting features in a BDT distribution can be difficult, particularly in the presence of backgrounds with imprecisely-known features or composition.

The Higgs to di-photon search possesses the advantage of a search for a mass peak over a smoothly falling background. The observation of an excess in such a distribution has a straightforward and robust physical interpretation. First and foremost the strategy for the use of multivariate techniques in this analysis is to preserve, and avoid obscuring this feature. Furthermore, in order to ensure that all elements of the analysis are sufficiently easy to interpret and can be properly validated, individual multivariate classifiers are trained on factorizable and well defined subsets of the available information, such that each classifier has a well-defined physical interpretation which can be validated with the relevant control samples.

Multivariate Regression

In order to apply multivariate techniques to the photon energy and mass reconstruction without obscuring their physical interpretation it is necessary to introduce the concept of multivariate regression in addition to classifiers. In multivariate regression, rather than constructing an optimal classifier between two classes of events, the goal is to construct a prediction for an unknown dependent variable, as a function of a multidimensional input space of independent variables. This can be represented by the unknown multidimensional function

$$y = f(\bar{x}) \tag{1.5}$$

where y is the best estimate of the dependent variable as a function of the independent variables $\bar{x}$. Again due to complex detector effects, the analytic form of this function is

generally not known, but a training sample of data or Monte Carlo simulation events may be available in which the true value of the dependent variable is known on an event-by-event basis. In this case, analogous to the likelihood ratio in the classification case, a series of Boosted Decision Trees are used [47] to construct an approximation for $f(\bar{x})$. In this case each event is assigned a response F, representing the best estimate of the target variable y. An appropriate loss function to be minimized by the gradient-boosted series of trees is then the Huber loss function.

$$L = \begin{cases} \frac{1}{2}(F-y)^2 & |F-y| \leq \delta \\ \delta\left(|F-y| - \delta/2\right) & |F-y| > \delta \end{cases} \tag{1.6}$$

This minimizes the square deviation $(F-y)^2$ between the predicted and true value of the target up to some cutoff δ, beyond which the loss function is proportional to the linear deviation $|F-y|$. The multiplicative coefficients are chosen in order to match the value of L and its derivatives at the transition point. The use of this cutoff helps in general to reduce the sensitivity of the training to events in the tails of the distribution. Given this loss function and a set of training events, a series of boosted decision trees is constructed as for the classification case, where the response F now represents the best estimate of the target variable.

1.5 Data Sample

The data sample used for this analysis corresponds to 5.1 fb^{-1} of data collected with a center of mass energy of 7 TeV during the 2011 LHC run, as well as 5.3 fb^{-1} with a center of mass energy of 8 TeV, collected during the first few months of the 2012 LHC run. Due to different cross sections for Higgs production and various background processes, as well as different reconstruction versions of the available data, the 7 TeV and 8 TeV data have been analyzed separately, and combined at the statistical analysis stage. The overall analysis strategy and elements are the same between the two datasets, but there are some small differences in the details, and

these are specified where relevant.

1.6 Monte Carlo Simulation

The underlying physical interactions as well as the response of the CMS detector represent a very complex system. An effective technique to model such a system are Monte Carlo simulations, in which complete events are simulated, randomizing over the expected distributions for both the underlying physical process as well as the detector response. Monte Carlo simulation is used for the Higgs to di-photon signal, after appropriate validation and corrections, in order to model the expected properties of the signal if it were to appear in the data, necessary in order to quantify the results of the search in terms of either discovery significance/exclusion or a measured signal rate. Monte Carlo simulation of the background processes are used only to optimize the selection requirements and to train various multivariate discriminators, but is not used for the final analysis, and the correctness of the results do not depend on a correct description of the background processes by the corresponding Monte Carlo simulation.

Producing Monte Carlo events for the CMS experiment consists in general of three distinct steps:

1. **Matrix Element and Phase Space Integration**: Events are sampled according to the kinematics of the outgoing particles, up to the level of a fixed number of hard partons in the case of jets. This sampling is done taking into account the energy of the incoming protons as well as the corresponding parton density functions, as well as the amplitude of the matrix element corresponding to the underlying physics process at either Leading Order (LO) or Next to Leading Order (NLO), depending on the process.

2. **Parton Shower and Hadronization**: In order to approximate the effect of higher order QCD corrections not present at the matrix element level, additional outgoing partons are added to the event, and kinematics from the matrix

element modified accordingly. Additional photons may be added as Initial State Radiation (ISR) or Final State Radiation (FSR) in order to include the effect of higher order QED corrections. Once this is done, outgoing quarks and gluons are hadronized into appropriate collections of charged and neutral hadrons, according to the initial parton kinematics and colour flow in the event, and using semi-empirical models which have been tuned to collider data. Finally any unstable hadrons with expected decay lengths less than a few millimeters are decayed according to the relevant branching ratios. For all of the Monte Carlo samples used in this analysis, the parton shower and hadronization step is performed using Pythia 6[48].

3. **Detector Simulation**: In order to represent the response of the CMS detector with respect to the final state particles, a full detector simulation is performed, including the propagation of each particle in the magnetic field and the interactions of each particle with the passive and active elements of the detector. This simulation is performed using Geant 4[49, 50], and a detailed implementation of the CMS geometry. Interactions with the material are simulated for each particle, including energy loss, Bremsstrahlung and photon conversions. For active detector elements, the simulated energy deposits are processed through a simulation of the readout electronics for each subdetector, including effects such as simulated noise as appropriate. The digital representation of the output from this step is packed into the same format as raw detector data, such that reconstruction and subsequent processing steps can be performed using as much as possible the identical software and configuration as for real data.

The software used for the matrix element step varies according to the availability of implementations for each desired process, using a mixture of POWHEG[51, 52, 53], Madgraph[54], and Pythia 6[48], where POWHEG uses NLO matrix elements, while Madgraph and Pythia use LO matrix elements. A complete list of simulated processes as well as the corresponding matrix element generators are given in Table 1.1. Monte Carlo samples have been generated separately for the 7 and 8 TeV analyses.

Process	Matrix Element
Gluon Fusion Higgs	POWHEG[55]
Vector Boson Fusion Higgs	POWHEG[56]
W/Z Associated Production Higgs	Pythia
$t\bar{t}$ Associated Production Higgs	Pythia
Drell-Yan di-lepton + 0-2 jets	Madgraph
QCD di-photon (gluon-gluon box diagram)	Pythia
QCD di-photon + 0-2 jets	Madgraph
QCD Photon + jet	Pythia
QCD Di-jet	Pythia

Table 1.1: List of simulated processes and the corresponding matrix element generator.

Because the parton shower step may add additional photons to the event with respect to the matrix element, some care must be taken to avoid double counting of background processes. In particular, QCD photon + jet events with one additional photon added by the parton shower are already included in the madgraph di-photon + jets sample. Since the matrix element is expected to describe the kinematics of such events better than the parton shower, these events are removed from the pythia sample. Similarily, QCD di-jet events with *two* photons added by the parton shower are already included in the madgraph di-photon + jets sample, and are removed from the pythia sample.

There is an additional interference term between the gluon-induced QCD di-photon production and gluon fusion Higgs production[57] which is not included in the Monte Carlo generation. This effect changes the rate, kinematics, and lineshape of the Higgs signal. Because the observed lineshape of any Higgs signal is expected to be dominated by the detector response, the latter effect is neglected. The effect of the interference term on the shape of kinematic distributions has been found to be negligible after the application of kinematic cuts on the photons. This effect is therefore taken into account by an overall correction to the expected signal yield from gluon fusion Higgs production of -2.5%.

The cross sections for the signal processes have been computed up to NNLO+NNLL

and are documented in [32, 33, 34]. The transverse momentum spectrum of the 7 TeV gluon fusion Higgs sample has been reweighted to match to the NNLO+NNLL prediction[58]. For the 8 TeV gluon fusion sample, parameters of the POWHEG generator have instead been tuned in order to match this prediction. The cross-sections for background processes, where used for optimization, are computed from the LO matrix element generators, and corrected with approximate k-factors from NLO calculations, ranging between 1.0 and 1.3.

Chapter 2

Event Reconstruction and Selection

2.1 Photon Reconstruction

Photons are reconstructed in the detector pseudorapidity range $|\eta| < 1.4442$ in the EB and $1.566 < |\eta| < 2.5$ in the EE. The reconstruction of the photon impact position in the calorimeter and the corresponding energy occurs in three steps. First the raw detector output is reconstructed into calibrated Ecal RecHits, containing the best estimate of the energy deposited in each individual crystal of the calorimeter as described in Section 1.3.2. Second, Ecal RecHits are combined into SuperClusters (SCs), themselves consisting of one or more BasicClusters (BC's). A single BasicCluster loosely corresponds to a single photon/electron or positron incident on the Ecal face, whereas a SuperCluster is intended to contain as much as possible Bremsstrahlung photons and conversion electrons and positrons from interactions of the primary photon with the tracking detector material. Third, the final corrections to the SuperCluster energy for local and global containment of the electromagnetic shower are applied using a multivariate regression.

2.1.1 Clustering

Once the calibrated RecHits have been reconstructed, they are combined into SuperClusters and their constituent BasicClusters. In the barrel, these are constructed in a single step. RecHits are clustered iteratively in descending order of transverse energy. Once a RecHit has been included in a cluster, it is not available for use in further clusters. Since the crystals in the barrel are positioned according to an $\eta-\phi$ geometry, the clustering algorithm can naturally exploit the limited width of the electromagnetic showers in η, combined with the possibly broad deposition of energy in ϕ due to the bending of Bremsstrahlung /conversion legs in the magnetic field. The initial BasicCluster of a SuperCluster is seeded by a single crystal with transverse energy $E_T > 1$ GeV. A 5x1 strip is then constructed symmetrically around this crystal, extended in the η direction. Next, all 5x1 strips with energy above 0.1 GeV and having their central crystal at the same η as the seed crystal and within $|\Delta\phi| < 17$crystals are clustered around local maxima to form BasicClusters. BasicClusters whose highest energy 5x1 strip is less than 0.35 GeV are dropped and the remainder are included in the SuperCluster. The final resulting SuperCluster is therefore 5 crystals wide in η and includes additional clusters within $|\Delta\phi| < 17$ crystals from the seed.

In the endcap, since the crystals do not follow an $\eta - \phi$ geometry, BasicClusters are first built up independently according to crystal-based patterns, and the SuperClusters are then subsequently built up geometrically. Basic Clusters are seeded by crystals with at least 0.18 GeV of transverse energy, with the initial cluster consisting of the corresponding 5x5 grid of crystals. Additional BasicClusters may be seeded by the crystals on the boundaries of this 5x5 region which are local maxima and above the seeding threshold. Since the subsequent 5x5 region is overlapping with the initial Basic Cluster, only new crystals are included in the subsequent Basic Cluster. Once the Basic Clusters have been reconstructed, SuperClusters are seeded by Basic Clusters containing at least 1.0 GeV of transverse energy, and include additional basic clusters in a geometric window of $|\Delta\eta| < 0.14$ and $|\Delta\phi| < 0.6$. For SuperClusters within the preshower acceptance, preshower energy deposits directly in front of the

crystals included in the SuperCluster are added a posteriori.

The raw energy of each BasicCluster is computed simply as the sum of calibrated RecHit energies for all crystals associated to the BasicCluster. The position of each BasicCluster is computed by a logarithmically energy weighted sum of RecHit positions. The individual RecHit positions are calculated for this purpose by projecting from the center of the crystal face, along the crystal axis to the nominal shower depth, computed based on the total energy of the Basic Cluster. The raw SuperCluster energy is computed from the sum of Basic Cluster energies, with preshower energy deposits being separately summed and stored for the endcap SuperClusters. The position of the SuperCluster is derived from a simple energy-weighted average of the BasicCluster positions.

2.2 Photon Energy Regression

The photon energy is computed starting from the raw SuperCluster energy, adding also the preshower energy in the endcap. This raw energy represents the calibrated sum of energy deposits for all crystals and preshower deposits which have been included in the SuperCluster. The measurement of this energy is of course subject to the underlying detector resolution, driven by fluctuations in the electromagnetic shower, noise in the electronics, and residual mis-calibration. Beyond this, however, there is a fraction of the energy which is simply not included in the measurement. This is due to two conceptually distinct groups of effects. First, some fraction of the shower energy is lost in gaps between the crystals as well as larger cracks between modules and sub-modules. The size of these local containment losses vary according to the impact position of the incident photon with respect to the crystal edge, and to module or sub-module boundaries. These effects are in principle simulated by the Monte Carlo, such that photons in the Monte Carlo samples are subject to similar local containment losses and variations as in the data, although there may be additional crystal to crystal and module to module variations in gap and crack size and orientation which are not included in the simulation.

Aside from local containment losses, for photons which convert in the material before they reach the calorimeter, soft conversion legs from either the primary conversion or secondary showers may not reach the Ecal at all on account of the magnetic field. Additionally, soft conversion legs or secondary Bremsstrahlung photons may impact the Ecal far enough from the primary shower that they are not included in the SuperCluster. Collectively these effects are referred to as a loss of global containment for the cluster. These effects are also simulated by the Monte Carlo, to the extent that the material description accurately reflects the detector. In addition to containment losses, the raw energy is also subject to contamination by energy deposits from additional pileup interactions, where the number of pileup interactions and their average energy contribution also vary significantly from event to event.

In order to obtain the optimal resolution, the raw energy must be corrected for both the local containment of the shower in the calorimeter, as well as the global containment of the shower for photons which convert and shower in material upstream of the calorimeter. Such corrections are needed for both the Monte Carlo samples as well as the data. To the extent that the Monte Carlo simulation accurately represents the degree and variation of local and global containment losses, corrections derived on the Monte Carlo samples are also applicable to the data. The traditional strategy for these kinds of corrections in CMS has been to derive factorized corrections on the Monte Carlo simulation, using simple parameterized corrections in terms of a few variables. The default reconstruction includes only corrections for global containment losses, parameterized as a function of SuperCluster energy and pseudorapidity, as well as a single variable representing the elongation of the cluster in ϕ, used as an in situ measure of the degree of showering in the material. In reality, the detailed shape and properties of the cluster are sensitive to the details of the shower in the material, as well as the impact position in the Ecal and subsequent shower. There are therefore many variables which are sensitive, and indicate to some extent on a photon-by-photon basis, the degree of local and/or global containment losses. Although these effects are conceptually separate, their manifestation in the details of the cluster shape are correlated.

The non-trivial correlations and large number of potential variables motivate the use of multivariate techniques for the determination of these energy corrections. The multivariate regression procedure using Boosted Decision Trees is adopted as described in Section 1.4.1. The regression is trained on prompt photons in Monte Carlo (from the photon + jets sample) using the ratio of generator level photon energy E_{gen} to the raw SuperCluster energy E_{SCRaw} (+ preshower energy E_{PS} in the endcap) as the target variable T_E.

$$T_E = \begin{cases} \frac{E_{gen}}{E_{SCRaw}} & \text{barrel} \\ \frac{E_{gen}}{E_{SCRaw}+E_{PS}} & \text{endcap} \end{cases} \tag{2.1}$$

The regression is trained separately for photons in the barrel and endcap.

The input variables are the global η and ϕ coordinates of the SuperCluster, a collection of shower shape variables, and a set of local cluster coordinates. The shower shape variables included are the R_9 of the SuperCluster, the ratio of the 5x5 crystal energy to the raw SuperCluster energy, the energy weighted η-width and ϕ-width of the SuperCluster, the number of BasicClusters, and the ratio of hadronic energy behind the SuperCluster to the electromagnetic energy of the cluster. In the endcap, the ratio of preshower energy to raw supercluster energy is also included. Additional information is included for the seed basic cluster, consisting of the relative energy and position of the seed cluster, the local covariance matrix, and a number of crystal energy ratios. These variables provide information on the likelihood and location of a photon conversion and the degree of showering in the material, and together with their correlation with the global η and ϕ position of the SuperCluster, drive the degree of global containment correction predicted by the regression. In the barrel, the η/ϕ index of the seed crystal as well as the position of the seed cluster with respect to the crystal center are also included. These variables, together with the seed cluster energy ratios provide information on the amount of energy which is likely to be lost in crystal and module gaps and cracks, and drive the level of local containment corrections predicted by the regression. Although the global and local

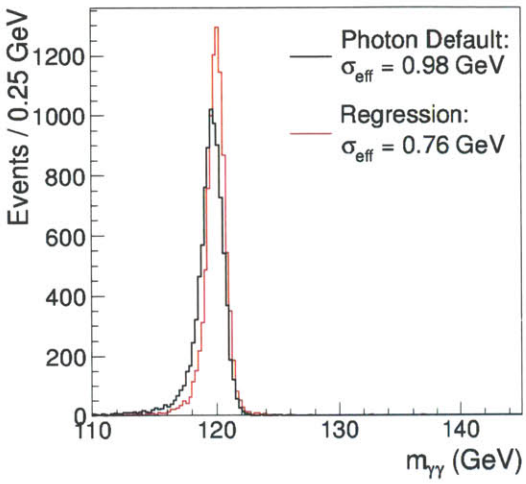
containment are conceptually different effects, the required corrections are allowed to be correlated in the regression in order to account for the fact that a showering photon is not incident at a single point on the calorimeter face, and is therefore relatively less affected by the local containment. Finally the number of primary vertices and median energy density ρ in the event are included in order to correct residual energy scale effects from pileup.

For the 7 TeV analysis in the barrel, these additional BasicCluster variables are also included for the 2nd highest energy BasicCluster (if present), and a subset of information for the lowest two energy BasicClusters, providing an additional handle for pileup suppression. Detailed information beyond the seed BasicCluster was removed for the 8 TeV analysis since it does not significantly improve the performance in data.

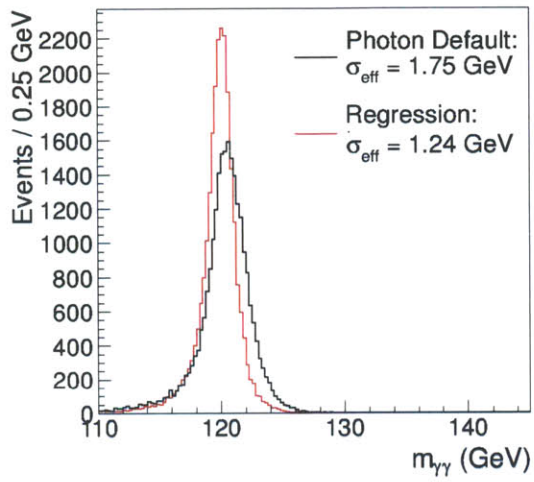
The performance of the regression for photons in Monte Carlo is shown in terms of the Higgs$\rightarrow \gamma\gamma$ mass resolution in four event classes based on pseudorapidity and R_9 in Figure 2-1. The line-shape is compared between the regression energy and the default photon energy in the reconstruction (using E5x5 for high R_9 photons and electron-tuned parameterized global containment corrections for low R_9).

The primary validation tool for the regression is the use of $Z \rightarrow ee$ events, where the for electrons in the Monte Carlo simulation can be compared to those in data. There are some differences between electrons and photons, mainly in terms of interaction with the material, but also in terms of incidence angle on the crystals which affects the local containment. Nevertheless, the photon-trained regression can still be applied to electrons and give sufficient performance for meaningful Monte Carlo to data comparisons. Additionally, a boosted decision tree with identical training settings and input variables is trained on a sample of electrons in $Z \rightarrow ee$ Monte Carlo. This regression provides more optimal performance for electrons and allows for more precise Monte Carlo simulation to data comparisons. The final validation and determination of data and Monte Carlo resolution and scale correction factors is described in the section below, and includes also possible effects of data vs Monte Carlo differences affecting the regression performance.

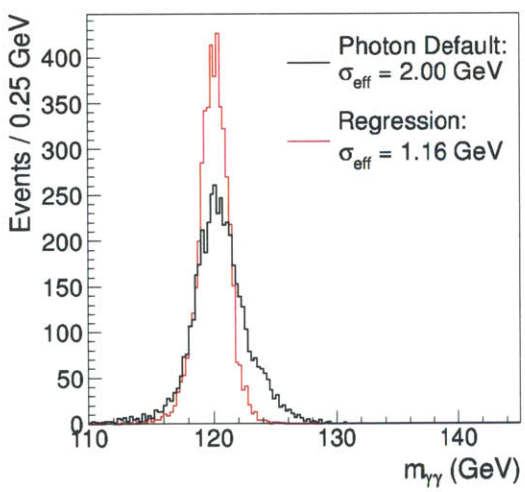
In order to make optimal use of each photon in the analysis, and also to aid in



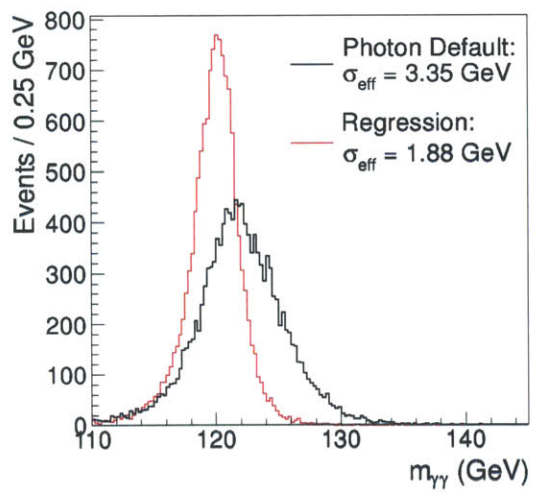
(a) Both EB, High R_9



(b) Both EB, One or more low R_9



(c) One or more EE, High R_9



(d) One or more EE, One or more low R_9

Figure 2-1: Comparison of the diphoton mass resolution in Higgs→ $\gamma\gamma$ 120 GeV Monte Carlo (8 TeV with no pileup re-weighting, using the generator level vertex and no additional energy smearing.). The default reconstructed photon energy is shown in black, and the full regression-corrected energy is shown in red.

validation of the regression, it is desirable to have, along with the optimal energy correction, a per-photon estimate of the energy resolution σ_E, allowing a per-event estimate of the di-photon mass resolution to be constructed. This is accomplished by training a second boosted decision tree on an independent set of Monte Carlo events with a target variable $T_{\sigma(E)}$ as the absolute deviation between the correction predicted by the first regression, E_{reg}/E_{SCRaw} ($E_{reg}/(E_{SCRaw} + E_{PS})$), and the true correction to generator level-energy:

$$T_{\sigma(E)} = \begin{cases} \left| \frac{E_{reg}}{E_{SCRaw}} - \frac{E_{gen}}{E_{SCRaw}} \right| & \text{barrel} \\ \left| \frac{E_{reg}}{E_{SCRaw}+E_{PS}} - \frac{E_{gen}}{E_{SCRaw}+E_{PS}} \right| & \text{endcap} \end{cases} \tag{2.2}$$

In order to form an equivalent Gaussian uncertainty, the target distribution T_E is approximated in each region of phase space as a Gaussian distribution with mean $< T_E >$ ($\bar{x}$) and resolution $\sigma_E(\bar{x})$. The relationship between σ_E and $T_{\sigma(E)}$ is then given by

$$\begin{aligned} \langle T_{\sigma(E)} \rangle &= \langle |\langle T_E \rangle - T_E| \rangle && (2.3) \\ &= \int_{-\infty}^{\infty} |w| \frac{1}{\sigma_E \sqrt{2\pi}} e^{-\frac{w^2}{2\sigma_E^2}} dw && (2.4) \\ &= 2 \int_0^{\infty} w \frac{1}{\sigma_E \sqrt{2\pi}} e^{-\frac{w^2}{2\sigma_E^2}} dw && (2.5) \\ &= \sigma_E \sqrt{\frac{2}{\pi}} && (2.6) \end{aligned}$$

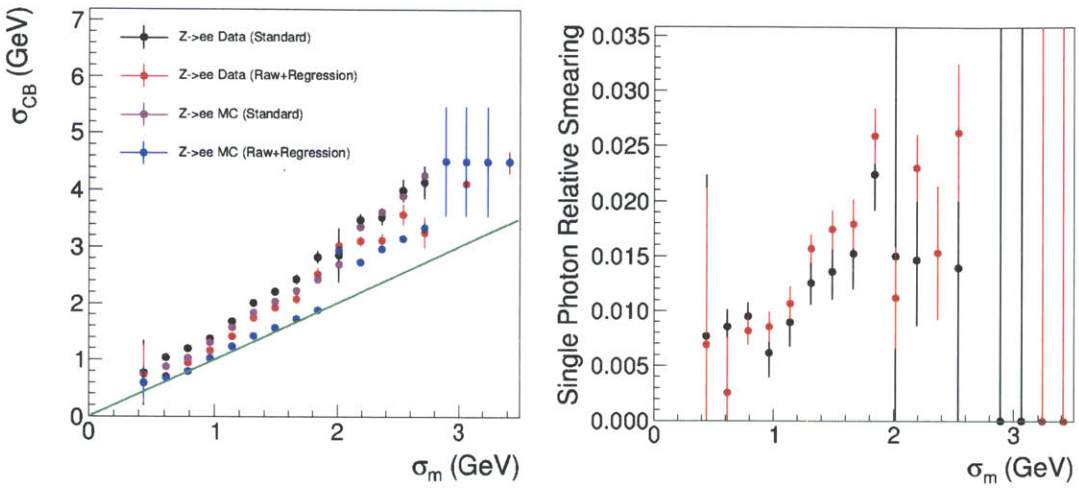
where $w \equiv T_E - < T_E >$.

The equivalent Gaussian uncertainty on the energy prediction is therefore $\sigma_E = \sqrt{\pi/2} \langle T_{\sigma(E)} \rangle \sim 1.253 \langle T_{\sigma(E)} \rangle$, taking into account the same set of input variables and correlations as the correction itself. In addition to allowing a per-event estimate of the mass resolution, this serves as a useful validation tool by providing a fine-grained division of the validation sample in bins of expected resolution. The di-electron mass resolution is measured in bins of the predicted mass resolution, and the two values are compared. The resolution is measured in each bin by means of an unbinned likelihood fit to the $Z \rightarrow ee$ mass. The signal shape for the fit is a crystal-ball

convolved with a Breit-Wigner, where the Breit-Wigner parameters are fixed to the PDG mass and width of the Z. The Gaussian width parameter of the crystal ball is taken as the detector contribution to the mass resolution. Data and Monte Carlo are compared for the default SuperCluster corrections as well as the mass computed using the regression-corrected electron energies in Figures 2-2 and 2-3. The default SuperCluster corrections here are the parameterized global containment corrections.

The regression provides a significant improvement to the resolution for both Data and Monte Carlo simulation in the barrel. Nevertheless, the data resolution is worse than the simulation resolution when comparing both the regression and the default SuperCluster corrections. The degradation of performance between simulation and data is quantified as the difference in quadrature of the σ_{CB} parameter in data versus simulation. This difference in quadrature is multiplied by $\sqrt{2}$ in order to represent the degradation of resolution at the single electron level, referred to as the single electron smearing, assuming the two electrons contribute in an uncorrelated way to the di-electron mass resolution. The smearing factor between data and simulation in the barrel is consistent whether measured with the regression or with the default SuperCluster corrections. This implies that the effects which lead to worse resolution in the data are largely uncorrelated with the containment or pileup corrections. In the endcap the regression provides substantial resolution improvements in Monte Carlo, but the required smearing is large, such that the relative improvement is smaller in data. In both barrel and endcap the measured mass resolution in Monte Carlo matches well the regression prediction, and even in the data, where the absolute performance is degraded with respect to the Monte Carlo, the measured resolution shows the expected correlation with the prediction.

A comparison of the energy scale and resolution as a function of the number of reconstructed primary vertices is shown in Figures 2-4 and 2-5. The default cluster corrections have a significant dependence on pileup in both energy scale and resolution, particularly in the endcap, which is almost entirely cured by the regression.



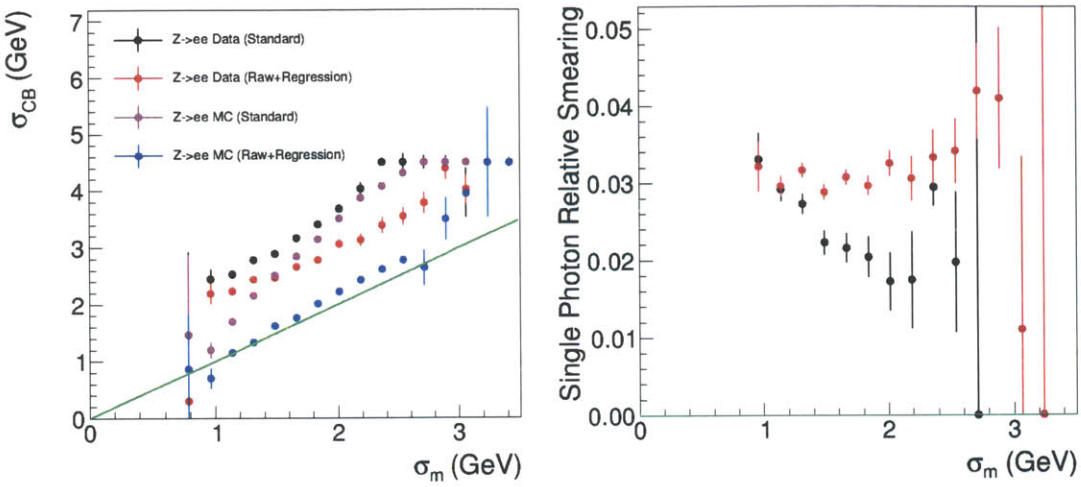
(a) Resolution

(b) Single Electron MC→ data smearing

Figure 2-2: Comparison of the measured versus predicted diphoton mass resolution from the electron-trained regression (left) as well as the implied single electron smearing vs predicted mass resolution (right) for $Z \to ee$ events in data and Monte Carlo with both electrons in the barrel. The green diagonal line shows the distribution expected if the measured and predicted mass resolution are in agreement.

2.2.1 Regression Algorithm Details

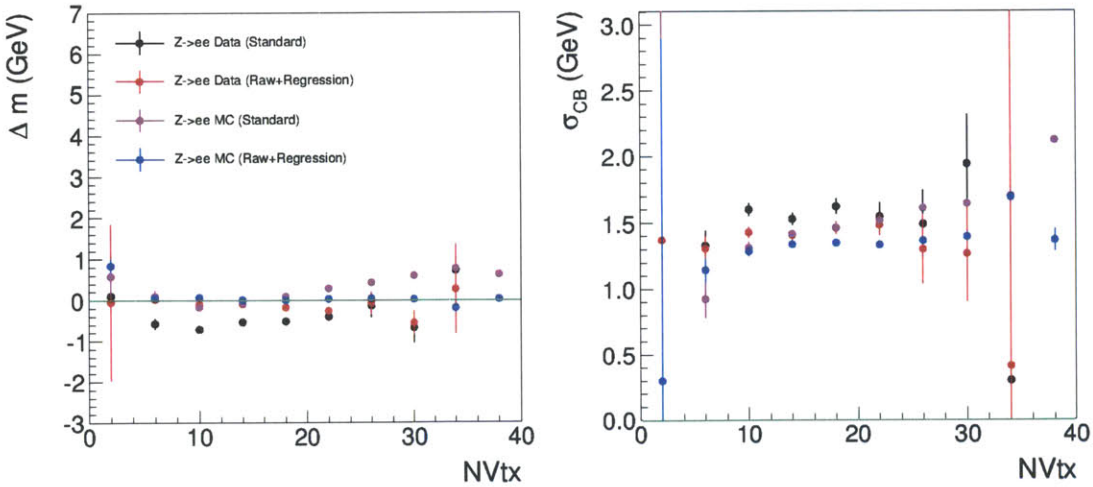
The photon energy regression used for the analysis of the 2011 data has been trained directly using the BDT regression functionality in TMVA [59]. The implementation of gradient boosted decision trees for regression in TMVA follows closely the original description of the TreeBoost algorithm [46], [47] as summarized in Section 1.4.1. The training parameters have been tuned so that tree growth is terminated only by the minimum number of events per node, set at 200. With the training samples of several million events, this leads to trees with O(10,000) nodes. The number of trees was set to 200. This corresponds to a significantly larger set of trees than is typically used for classification problems, where a larger number much smaller trees only a few levels deep commonly gives the optimal performance. The storage of TMVA BDT weights in xml files is not optimized for storage or memory footprint and therefore produces prohibitively large weights for the above trained regression, requiring several GB both on disk and in memory.



(a) Resolution

(b) Single Electron MC→ data smearing

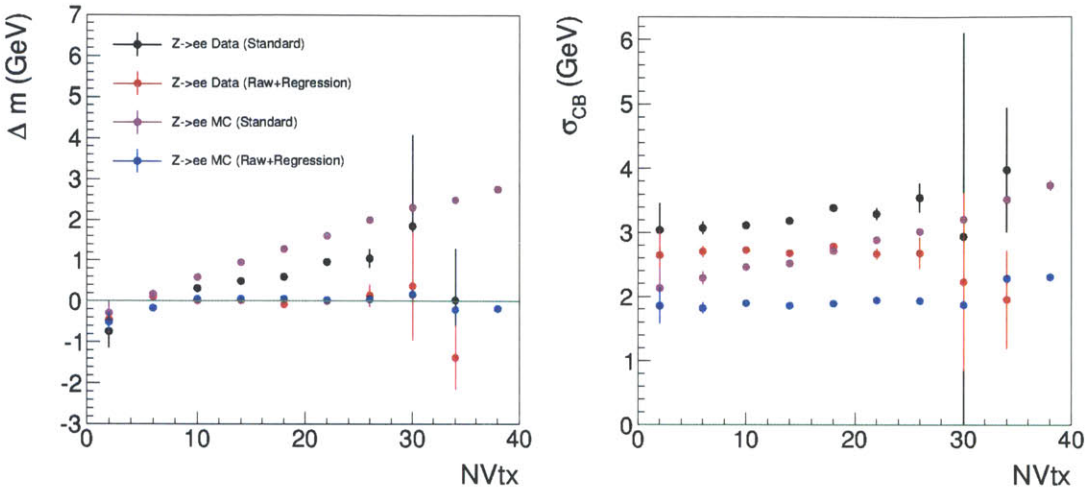
Figure 2-3: Comparison of the measured versus predicted diphoton mass resolution from the electron-trained regression (left) as well as the implied single electron smearing vs predicted mass resolution (right) for $Z \to ee$ events in data and simulation with one or more electrons in the endcap. The green diagonal line shows the distribution expected if the measured and predicted mass resolution are in agreement.



(a) Energy Scale

(b) Resolution

Figure 2-4: Comparison of the energy scale (left) and resolution (right) from the electron-trained regression as a function of the number of reconstructed primary vertices for $Z \to ee$ events in data and Monte Carlo with both electrons in the barrel.



(a) Energy Scale

(b) Resolution

Figure 2-5: Comparison of the energy scale (left) and resolution (right) from the electron-trained regression as a function of the number of reconstructed primary vertices for $Z \to ee$ events in data and Monte Carlo with with one or more electrons in the endcap.

Optimization of BDT Storage

In order to run the analysis efficiently on standard batch systems, a dedicated set of BDT storage and application utilities were implemented as persistable ROOT classes. In these classes, the tree structure is represented by flat arrays stored in std::vector containers. Each tree contains two sets of arrays, one representing intermediate nodes, and another representing the terminal nodes of the tree. The arrays representing the intermediate nodes store the index of the splitting variable as an unsigned 8 bit integer, the value of the splitting cut stored as a single precision floating point number, as well as the indices of the "left" and "right" nodes corresponding to the subsequent nodes in the tree structure for the two splitting outcomes, stored as signed 32 bit integers. Positive indices represent further intermediate nodes, where zero or negative indices represent terminal nodes. The single array representing the terminal nodes stores only the corresponding response as a single precision floating point. The choice of data types limits the regression to 256 input variables, with practically no limit on

the tree size (up to 2^{31} intermediate nodes plus 2^{31} terminal nodes). Standard ROOT IO is able to efficiently store this structure, including the use of standard compression algorithms. These classes implement functionality allowing conversion from existing TMVA-trained weights, which was used for the TMVA-trained regression for the 2011 portion of the analysis. The net result is a reduction by approximately a factor of 50 in both disk and memory requirements compared to the standard TMVA xml weights and application classes.

Optimization of BDT Training

For the 2012 portion of the analysis, the regression was trained using a reimplementation of the training algorithm, which has been optimized in a number of ways both for computing and physics performance. This implementation is closely based on the TMVA implementation but with a number of important improvements.

1. Floating point values of input variables are converted to integer quantiles (by default with 2^{16} possible values) such that each quantile contains an equal subset of the training events. This occurs at the very beginning of the training, and sorted vectors of the training events are produced for each input variable. At the beginning of the split search for a given variable on a given node, these quantiles are down-converted to a coarser set of quantiles (by default 128) covering the sub-range of the input variable present in the subset of training events appearing on the particular node. This approach has several advantages compared to the fixed binning used in TMVA. First the quantile-based binning is robust against outlier values as well as any monotonic transformation of the input variables, thereby requiring less care in the preparation of the training sample and selection. Second, large computing performance gains are realised without sacrificing physics performance by limiting the number of split values tested in each individual search without compromising the overall granularity considered over the full tree. If the full granularity contains statistically significant information, it can be utilized in deeper levels of the tree, but if not, the computing time spent searching for splits in this variable at higher levels of the

tree are kept to a minimum.

2. A new tree-growth cutoff mechanism has been implemented based on the statistical significance of the best split. This is computed simply based on the difference in mean target values on the two sides of the split, divided by the statistical uncertainty on the mean, computed from the variance approximating the target distributions as Gaussian distributed,

$$\delta(\bar{x}_R - \bar{x}_L)^2 = \frac{\sigma_R^2}{N_R^2} + \frac{\sigma_L^2}{N_L^2} \tag{2.7}$$

where $\bar{x}_L$ and $\bar{x}_R$ are the mean target values on the left, L and right, R sides of the split, σ_L^2 and σ_R^2 are the target variances on the two sides of the split, and N_L and N_R are the number of training events on the two sides of the split. The significance of the split S is then given by

$$S = \frac{\bar{x}_R - \bar{x}_L}{\delta(\bar{x}_R - \bar{x}_L)} \tag{2.8}$$

The significance cutoff also provides for automated termination of forest growth, by stopping the addition of trees once the previous tree in the forest has been limited to only one node. For the training of the regression here, parameters were tuned such that tree and forest growth was terminated primarily by the significance cutoff, combined with the existing minimum of 200 training events per node. The resulting forests contained between 150 and 300 trees, such that the total number of trees was similar to the value set by hand in TMVA for the 2011 training, though the resulting trees are on average smaller.

3. The structure of frequently-executed loops with many iterations inside the split search step set of the algorithm have been tuned to allow auto-vectorization by GCC, allowing multiple possible split values to be tested in parallel using the vector integer and floating point instructions of modern processors.

4. The loops over input variables, both for the initial sorting of training events, and

most importantly during the split search on each node has been multithreaded using the OpenMP library/implemenation in GCC. For a regression with a number of input variables equal to or larger than the number of execution threads on the computer, this allows for close to 100% utilization of all processor cores and execution threads on the machine.

The net effect of these changes is an approximately twenty-fold reduction in real time needed for regression training on a typical 8-core/16-thread machine, with possible small improvements in physics performance depending on the training settings. The optimization of CPU and memory footprint also enables in principle the use of this code for much larger regression problems. Using a slightly modified version of the code (to increase the 256 input variable limit), the algorithm has been successfully tested on a standard 8 core batch worker with 24 GB of memory with up to 600 input variables and 4 million training events.

2.3 Electron Reconstruction

Electrons are reconstructed by appropriately combining information from the Ecal and tracker. Reconstructed electrons are used both to exclude electrons from the reconstruction of photons, as well as to measure and validate various corrections to the Monte Carlo simulation. The reconstruction of electrons begins with the identical set of SuperClusters which have been used for the photon reconstruction. The SuperClusters used for electron seeding are required to have transverse energy of at least 4 GeV, as well as to pass a loose cut on the ratio of hadronic to electromagnetic energy H/E as defined in 2.7.1, and strictly looser than the photon and electron identification criteria. These SuperClusters are then used to seed the reconstruction of the electron trajectory in the tracking detectors. Candidates passing a loose preselection on the compatibility between the reconstructed track and electromagnetic cluster are promoted to electron candidates.

2.3.1 Track Seed Reconstruction and Matching

Track seeds for the electrons consist of pairs or triplets of pixel hits in the barrel region, and pairs or triplets of pixel and TID hits in the endcap region. The first two hits of the seed are required to match geometric compatibility requirements with the Ecal Supercluster. These requirements are based on backwards propagation of a helix from the reconstructed SuperCluster position, using the SuperCluster transverse energy as an initial estimate for the primary electron transverse momentum. This compatibility test is made under both charge hypotheses. If all Bremsstrahlung photons and conversion legs are included in the SuperCluster and well measured, then a radiating electron will have its SuperCluster position reconstructed along the helix of the primary electron trajectory, irrespective of the amount of radiated energy. In practice, due to the loss of some radiated energy from the cluster, as well as ambiguity related to the possibility of significant Bremsstrahlung on or before the first two layers of the trajectory, the compatibility windows need to be somewhat wider than those implied by the raw energy and position resolution. The size of the matching windows in z are also made large enough in order to accommodate variations in the position of the primary vertex along the length of the luminous region.

2.3.2 Electron Track Reconstruction

The reconstruction of electron tracks is made challenging by the large amount of material in the tracking detectors, leading to large amounts of Bremsstrahlung, as well as subsequent conversions of Bremsstrahlung photons. As a result, the standard Gaussian assumptions of the track energy loss in the material and their propagation to the final track fit are not sufficient for efficient track reconstruction or precise charge and momentum determination. The pattern recognition step, in which compatible hits are added to the trajectory in addition to the seed hits, is performed with very loose matching windows, in order to accommodate large Bremsstrahlung. The final track fit is performed using a Gaussian Sum Filter (GSF) [60] procedure, in which the non-Gaussian distribution for the expected energy loss at each layer is represented

as a weighted sum of Gaussian energy loss components. Each of these components is propagated into a Gaussian track state, for which the helix parameters and covariance matrix are determined using the Combinatorial Kalman Filter (CKF) used for general track reconstruction [61]. The final charge and momentum determination of the GSF track is determined by taking the mode of the charge and each momentum component from the weighted mixture of fitted states.

2.3.3 Electron Track-Cluster Matching

The compatibility of the reconstructed electron track with the electromagnetic cluster is determined by comparing the position of the inner trajectory extrapolated to the Ecal face, with the position of the reconstructed SuperCluster. A loose preselection is applied, requiring $|\Delta\eta| < 0.02$ and $|\Delta\phi| < 0.15$ between these two positions.

2.4 Residual Photon Energy Scale and Resolution Corrections

In order to correctly model a potential Higgs signal using the Monte Carlo, the photon energy scale and resolution must be consistent with the data. The photon-trained regression employed for the cluster corrections brings the energy scale close to unity in the Monte Carlo. The Monte Carlo simulation incorporates the bulk of the known detector effects, including a realistic estimate of the intercalibration precision. Nevertheless, residual imperfections in the calibration scheme, as well as residual data-MC discrepancies in the effect of the cluster corrections, and additional effects which are not present or underestimated in the simulation lead to a different energy scale and resolution in data with respect to the Monte Carlo simulation, with possible variations as a function of time, detector location, and shower properties. The strategy employed is to correct the photon energy scale in data events in order to match that of the Monte Carlo, and to correct the photon energy resolution in Monte Carlo events in order to match that of the data.

2.4.1 $Z \to ee$ Selection and Reconstruction

These corrections are determined using reconstructed $Z \to ee$. In order to have the closest possible correspondence to reconstructed photons, the energy of the electrons are reconstructed exclusively using information from the Ecal, where the reconstruction and clustering algorithms are identical for electrons and photons. In order to use the same energy reconstruction as the photons, the electron energy is determined using the photon-trained energy regression for the cluster corrections. This leads to a systematic under-correction of the electron energy in both data and simulation, particularly for electrons which shower significantly in the material. The absolute scale is not critical however, because corrections are derived from the ratio between the data and simulation energy scale. Nevertheless, the corresponding electron-trained regression corrections are used as part of the study of systematic uncertainties, and these do have an energy scale which close to unity for simulated electrons. In order to factorize out effects related to the primary vertex reconstruction, the electron direction is taken from the reconstructed GSF track, and has negligible contribution to the mass resolution. Events are selected with two oppositely charged electrons with reconstructed di-electron mass between 75 and 105 GeV.

2.4.2 Correction Granularity

The energy scale corrections for the data are measured in four detector regions, according to the pseudorapidity of the electron SuperCluster: $|\eta| < 1.0, 1.0 < |\eta| < 1.4442, 1.566 < |\eta| < 2.0, 2.0 < |\eta| < 2.5$. This reflects the division in barrel and endcap, as well as differences in material budget and radiation level as a function of pseudorapidity. The scale corrections are further differentiated between electrons/photons which interact significantly with the tracker material, referred to as "showering" electrons/photons and those which do not, referred to as "non-showering". This division is made experimentally according to the shower profile variable R_9, with electrons and photons with $R_9 \geq 0.94$ corresponding to the non-showering category, and those with $R_9 < 0.94$ corresponding to the showering category. The scale corrections are

factorized into a time-dependent part designed to cover residual and possibly time-dependent imperfections in the calibration, and a time-independent part, designed to cover residual differences in the material interactions and cluster corrections. The time-dependent component of the corrections are therefore not sub-divided according to R_9, and are derived only for the four pseudorapidity regions.

2.4.3 Time-dependent Energy Scale Corrections to Data

In order to derive the time-dependent corrections, the data has been divided into 16 time-contiguous periods (6 in the 7 TeV data and 10 in the 8 TeV data). The divisions have been chosen such that the number of $Z \rightarrow ee$ events are sufficient to derive scale corrections with negligible statistical uncertainties, and have been grouped according to changes in data-taking conditions or longer breaks in data-taking, where changes in instantaneous luminosity as well as extended periods with no or little luminosity can effect larger changes in crystal transparency. For each time period, the selected $Z \rightarrow ee$ events are divided into 4 subsets consisting of events where both electrons have been reconstructed in the same pseudorapidity region among the four. Although this subdivision uses only a fraction of the total events, the energy scale corrections can be determined independently in each pseudorapidity region, greatly simplifying the procedure. The $Z \rightarrow ee$ Monte Carlo is likewise divided into rapidity regions. A maximum likelihood fit is first performed for each of the four event pseudorapidity regions on the Monte Carlo, fitting the $Z \rightarrow ee$ peak using the convolution of a Breit-Wigner with a Crystal Ball function [62]. The Breit-Wigner represents the underlying line shape of the Z resonance, and it's mean and width are fixed to the PDG[63] values for the Z mass and width. The Crystal Ball represents the detector response. The Crystal Ball mean Δm, resolution and tail parameters are all freely floating in the fit to the Monte Carlo. For each time period, the corresponding data-sample is similarly fit, with the Crystal Ball tail parameters fixed to the values from the Monte Carlo fit and the Crystal Ball mean and resolution freely floating. The time-dependent component of the correction for time period i in each pseudorapidity range j is given

by the relative Monte Carlo/Data energy scale.

$$s_{ij}^{time} = \frac{m_Z + \Delta m_j^{MC}}{m_Z + \Delta m_{ij}^{data}} \tag{2.9}$$

2.4.4 Residual Energy Scale Corrections to Data

In order to derive the residual time-independent component of the energy scale corrections, the time-dependent component derived above has been applied to the data by multiplying the energy of each electron by the correction factor s_{ij}^{time} as defined above. This partially corrected data sample is then divided into 8 sub-samples where both electrons are reconstructed in the same pseudorapidity region among the four and in the same R_9 category among the two. The Monte Carlo is likewise sub-divided. Maximum likelihood fits are performed as for the time-dependent corrections, and the time-independent component of the correction per pseudorapidity $\times$ R_9 category j is likewise defined as

$$s_j^{residual} = \frac{m_Z + \Delta m_j^{MC}}{m_Z + \Delta m_j^{data}} \tag{2.10}$$

The full energy scale correction is then given by

$$s_{ij} = s_{ij}^{time} s_j^{residual} \tag{2.11}$$

2.4.5 Energy Resolution Corrections to Simulation

In order to derive the energy resolution correction to the Monte Carlo, the data has been corrected by the full energy scale correction above, by multiplying the energy of each electron by a total factor s_{ij}. At this stage the energy scale of the data has been corrected to match that of the Monte Carlo, but the energy resolution is not necessarily equivalent, and in practice the data has somewhat worse energy resolution than the Monte Carlo. Although the likelihood fits used to derive the scale corrections provide an estimate of the resolution through the Crystal Ball Gaussian resolution parameter, this estimate suffers from large statistical uncertainties in some categories. This is the case because the natural width of the Z is comparable or larger than the

mass resolution, such that there are significant statistical uncertainties in deriving the deconvolved width of the Crystal Ball function. This problem is mitigated by using the full set of $Z \rightarrow ee$ events, rather than just those with both electrons in the same $\eta \times R_9$ categories. In order to fully exploit the available events, both the data and Monte Carlo are sub-divided into the $N(N+1)/2$ sets representing all possible combinations of the $N = 9$ $\eta \times R_9$ categories for the two electrons (non-showering electrons in the $|\eta| < 1$ region are further subdivided according to whether their cluster position is in proximity to one of the module boundaries, since such electrons and photons are expected to have somewhat worse resolution than those far from the boundaries). Each subset of the Monte Carlo sample is used to compute a corresponding binned probability density function for the di-electron mass. This probability density is computed in terms of $2N$ free parameters ΔE_i and $\Delta\sigma_i$, representing the energy scale and resolution difference between data and Monte Carlo for electrons in each category. The energy of each Monte Carlo electron has a random number added to it, thrown from a Gaussian distribution with mean $E_{nom}(1+\Delta E)$ and width $E_{nom}\Delta\sigma$. The probability density function therefore corresponds to the di-electron mass distribution of the Monte Carlo in each subcategory after a relative energy scale correction of $(1+\Delta E)$ and a relative degradation of resolution (smearing) by $\Delta\sigma$. These probability density functions, together with the data in each of the 45 sub-categories is used to compute a binned likelihood value. This likelihood is minimized as a function of the 18 free parameters corresponding to the energy scale and resolution corrections for each electron category. Since the data which has been used for this procedure has already been corrected to match the simulation energy scale, the ΔE parameters in this procedure are in fact consistent with 0 within the statistical uncertainty. The final correction to the energy resolution in the Monte Carlo simulation is applied by adding a random number to the energy of each electron or photon in the Monte Carlo sample, drawn from a Gaussian distribution with mean of zero and width of $E\Delta\sigma_i$, effectively smearing the energy of each electron or photon by a relative fraction $\Delta\sigma_i$ depending on the $\eta \times R_9$ category.

2.4.6 Systematic Uncertainties

The statistical uncertainty associated with the energy scale corrections in particular are small, less than 0.1%. Nevertheless, there are systematic uncertainties associated with extrapolating the measurement from the $Z \rightarrow ee$ system to the Higgs di-photon system, both in extrapolating from the energy scale for electrons to that for photons, and extrapolating from the mean electron energy associated with the Z resonance at 91 GeV, to the somewhat higher average photon energies associated with the Higgs search range. Uncertainties associated with electron to photon extrapolation are evaluated both by re-weighting the $Z \rightarrow ee$ events according to match the shower profile and pseudo-rapidity distribution expected for photons from a Higgs decay, as well as by studying the size of the electron-photon energy scale difference in the Monte Carlo simulation, resulting in an uncertainty of 0.25% for non-showering photons in EB. Uncertainty associated with extrapolation in energy has been evaluated by studying variations in the Z mass peak under variations of the electron p_T or scalar sum p_T cuts, and correspond to 0.4% for non-showering photons in EB. These uncertainties have a negligible effect on search sensitivity, but are relevant for any measurement of the Higgs mass.

2.5 Conversion Reconstruction

The explicit reconstruction of photon conversion vertices formed from electron-positron track pairs is used to aid both the primary vertex identification as well as to help distinguish between electrons and photons. The basic strategy is to loosely pre-select oppositely charged track pairs and perform a constrained vertex fit including the constraint that the tracks are parallel at the vertex, intrinsic to the conversion topology from the massless photon and small momentum transfer. The track and vertex reconstruction is complicated by the fact that electron and positron tracks themselves radiate substantially and by the fact that a converted photon may asymmetrically share its energy among the electron and positron such that the sub-leading particle is of low transverse momentum and difficult to reconstruct efficiently.

In order to maximize tracking efficiency, the reconstruction of conversions uses all available tracks, consisting of the standard CMS iterative tracking, Gaussian Sum Filter (GSF) tracking from electron reconstruction, as well as dedicated conversion seeding steps where additional tracks are seeded either from Ecal clusters, or from existing tracks with a loose conversion constraint in the seeding. The vertex fitting of track pairs uses a least squares minimisation with Lagrange multiplier constraints to implement the common vertex and parallel track constraints. For GSF tracks, which contain a weighted mixture of Gaussian components, the vertex fit is performed using the weighted average for the parameter values and covariance matrix [1].

The vertex fit is implemented within the KinematicConstrainedVertexFitter in CMSSW, with the general mathematical framework and the details of the common vertex constraint based on [64]. The vertex fit uses a rectilinear coordinate system to express the track helices, defining each track in terms of a reference point on the helix $\boldsymbol{x}_i$ with corresponding momentum $\boldsymbol{p}_i$. The track reference point and momentum used to initalize the fit corresponds to the track state propagated to the innermost hit position, using the full material and magnetic field model of the track reconstruction. For a track with charge q_i in a constant magnetic field $\boldsymbol{B}$, the momentum at the fitted vertex $\boldsymbol{x}$ with $\boldsymbol{\Delta x_i} \equiv \boldsymbol{x} - \boldsymbol{x}_i$ is expressed as

$$p_{T,i} = p_{x,i}^2 + p_{y,i}^2 \tag{2.12}$$

$$\boldsymbol{p_{vtx,i}} = \boldsymbol{p_i} + q_i \left(\boldsymbol{\Delta x_i} \times \boldsymbol{B}\right) \tag{2.13}$$

$$\tag{2.14}$$

For a solenoidal magnetic field B in the $+z$ direction, and defining $a_i \equiv -q_i B$, the

[1] Preliminary studies suggest the possibility for performance gains in both resolution and efficiency with the use of the full state mixture, but at a large cost in computing time

momentum components at the fitted vertex position become,

$$p_{x,vtx,i} = p_{x,i} - a_i \Delta y_i \tag{2.15}$$

$$p_{y,vtx,i} = p_{y,i} + a_i \Delta x_i \tag{2.16}$$

$$p_{z,vtx,i} = p_{z,i} \tag{2.17}$$

In the standard treatment, the two constraint equations are written as

$$H^{std}_{s\phi,i}(\boldsymbol{\alpha_i}) = \Delta y_i p_{xi} - \Delta x_i p_{yi} - \frac{a_i}{2}\left(\Delta x_i^2 + \Delta y_i^2\right) \tag{2.18}$$

$$H^{std}_{sz,i}(\boldsymbol{\alpha_i}) = \Delta z_i - \frac{p_{zi}}{a_i}\mathrm{asin}\left(\frac{n_i}{p_T^2}\right) \tag{2.19}$$

with the defined quantity,

$$n_i \equiv a_i\left(\Delta x_i p_{xi} + \Delta y_i p_{yi}\right) \tag{2.20}$$

In cases where the tracks have very low p_T and have large uncertainties, the initial vertex position may be far from the track helices, or the helices may bend through very large angles between the track reference point and the vertex. In such cases the argument of the arcsin in 2.19 may become outside of the range $[-1, 1]$, such that the constraint equation cannot be evaluated and the fit is forced to abort. The common vertex constraint has therefore been re-derived in order to improve the robustness of the fit. For each track, the modified constraint equations are written as

$$H_{s\phi,i}(\boldsymbol{\alpha_i}) = \Delta y_i p_{xi} - \Delta x_i p_{yi} - \frac{a_i}{2}\left(\Delta x_i^2 + \Delta y_i^2\right) \tag{2.21}$$

$$H_{sz,i}(\boldsymbol{\alpha_i}) = \Delta z_i - \frac{p_{zi}}{a_i}\mathrm{atan2}\left(n_i,\ m_i\right) \tag{2.22}$$

with the additional defined quantity,

$$m_i \equiv p_{xi}\left(p_{xi} - a_i \Delta y_i\right) + p_{yi}\left(p_{yi} + a_i \Delta x_i\right) \tag{2.23}$$

$$\tag{2.24}$$

This differs from the standard vertex constraint equations mainly in the use of the two-branch arctangent function, in place of arcsin in the second constraint. The two-branch arctangent function is as defined in the standard C libraries, such that $\text{atan2}(y, x) \in (-\pi/2, \pi/2)$ is the angle between the positive x-axis, and the point defined by (x, y). This change increases the robustness of the fit in the case of large bending angles and/or initial parameters where the vertex is far from the helix. The corresponding non-zero partial derivatives needed to define the constraint are then

$$\frac{\partial H_{s\phi,i}}{\partial x_i} = p_{yi} + a_i \Delta x_i \tag{2.25}$$

$$\frac{\partial H_{s\phi,i}}{\partial y_i} = -p_{xi} + a_i \Delta y_i \tag{2.26}$$

$$\frac{\partial H_{s\phi,i}}{\partial p_{xi}} = \Delta y_i \tag{2.27}$$

$$\frac{\partial H_{s\phi,i}}{\partial p_{yi}} = -\Delta x_i \tag{2.28}$$

$$\frac{\partial H_{s\phi,i}}{\partial x} = -\left(p_{yi} + a_i \Delta x_i\right) \tag{2.29}$$

$$\frac{\partial H_{s\phi,i}}{\partial y} = \left(p_{xi} - a_i \Delta y_i\right) \tag{2.30}$$

$$\frac{\partial H_{sz,i}}{\partial x_i} = -\ell_i \left(m_i p_{xi} - n_i p_{yi}\right) \tag{2.31}$$

$$\frac{\partial H_{sz,i}}{\partial y_i} = -\ell_i \left(m_i p_{yi} + n_i p_{xi}\right) \tag{2.32}$$

$$\frac{\partial H_{sz,i}}{\partial z_i} = -1 \tag{2.33}$$

$$\frac{\partial H_{sz,i}}{\partial p_{xi}} = \ell_i \left[m_i \Delta x_i - \frac{n_i}{a_i}\left(2p_{xi} - a_i \Delta y_i\right)\right] \tag{2.34}$$

$$\frac{\partial H_{sz,i}}{\partial p_{yi}} = \ell_i \left[m_i \Delta y_i - \frac{n_i}{a_i}\left(2p_{yi} + a_i \Delta x_i\right)\right] \tag{2.35}$$

$$\frac{\partial H_{sz,i}}{\partial p_{zi}} = -\frac{1}{a_i}\text{atan2}\left(n,\ m\right) \tag{2.36}$$

$$\frac{\partial H_{sz,i}}{\partial x} = \ell_i \left(m p_{xi} - n_i p_{yi}\right) \tag{2.37}$$

$$\frac{\partial H_{sz,i}}{\partial y} = \ell_i \left(m p_{yi} + n_i p_{xi}\right) \tag{2.38}$$

$$\frac{\partial H_{sz,i}}{\partial z} = 1 \tag{2.39}$$

with the final defined quantity,

$$\ell_i \equiv -\frac{p_{zi}}{\left[(p_{xi} - a_i\Delta y_i)^2 + (p_{yi} + a_i\Delta x_i)^2\right] p_{Ti}^2} \tag{2.40}$$

In addition to the common vertex constraint, we impose a collinearity constraint in three-dimensional space, requiring that the momenta of the two tracks are parallel at the vertex. This corresponds to the characteristic topology of the conversion, following from the massless photon and small momentum transfer. This can be expressed by the two constraint equations

$$H_1(\boldsymbol{\alpha}) = \phi_{1,vtx} - \phi_{2,vtx}\ ,\ -\pi < H_1 \leq \pi \tag{2.41}$$

$$H_2(\boldsymbol{\alpha}) = \theta_{1,vtx} - \theta_{2,vtx}\ ,\ -\pi < H_2 \leq \pi \tag{2.42}$$

where $\phi_{i,vtx}$ and $\theta_{i,vtx}$ denote the momentum of the ith track at the fitted vertex position.

Expressing the constraints in terms of the standard track parameters in rectilinear coordinates,

$$H_1(\boldsymbol{\alpha}) = \text{atan2}(p_{y,1} + a_1\Delta x_1,\ p_{x,1} - a_1\Delta y_1) - \text{atan2}(p_{y,2} + a_2\Delta x_2,\ p_{x,2} - a_2\Delta y_2)) \tag{2.43}$$

$$H_2(\boldsymbol{\alpha}) = \text{atan2}(\sqrt{p_{x,1}^2 + p_{y,1}^2},\ p_{z,1}) - \text{atan2}(\sqrt{p_{x,2}^2 + p_{y,2}^2},\ p_{z,2}) \tag{2.44}$$

with $a_i = -q_i\ B$ for the track charge q_i and solenoidal magnetic field B (in GeV^{-1}) and $\boldsymbol{\Delta x_i} = \boldsymbol{x} - \boldsymbol{x}_i$ for fitted vertex position $\boldsymbol{x}$ and track reference point $\boldsymbol{x}_i$. Defining for the ith track,

$$k_i \equiv \frac{1}{(p_{x,i} - a_i\Delta y_i)^2 + (p_{y,i} + a_i\Delta x_i)^2} \tag{2.45}$$

then the non-zero partial derivatives required to fully define the constraint are

$$\frac{\partial H_1}{\partial p_{x,1}} = -k_1 \left(p_{y,1} + a_1 \Delta x_1\right) \tag{2.46}$$

$$\frac{\partial H_1}{\partial p_{x,2}} = +k_2 \left(p_{y,2} + a_2 \Delta x_2\right) \tag{2.47}$$

$$\frac{\partial H_1}{\partial p_{y,1}} = +k_1 \left(p_{x,1} - a_1 \Delta y_1\right) \tag{2.48}$$

$$\frac{\partial H_1}{\partial p_{y,2}} = -k_2 \left(p_{x,2} - a_2 \Delta y_2\right) \tag{2.49}$$

$$\frac{\partial H_1}{\partial x_1} = -k_1 a_1 \left(p_{x,1} - a_1 \Delta y_1\right) \tag{2.50}$$

$$\frac{\partial H_1}{\partial x_2} = +k_2 a_2 \left(p_{x,2} - a_2 \Delta y_2\right) \tag{2.51}$$

$$\frac{\partial H_1}{\partial y_1} = -k_1 a_1 \left(p_{y,1} + a_1 \Delta x_1\right) \tag{2.52}$$

$$\frac{\partial H_1}{\partial y_2} = +k_2 a_2 \left(p_{y,2} + a_2 \Delta x_2\right) \tag{2.53}$$

$$\frac{\partial H_1}{\partial x} = +k_1 a_1 \left(p_{x,1} - a_1 \Delta y_1\right) - k_2 a_2 \left(p_{x,2} - a_2 \Delta y_2\right) \tag{2.54}$$

$$\frac{\partial H_1}{\partial y} = +k_1 a_1 \left(p_{y,1} + a_1 \Delta x_1\right) - k_2 a_2 \left(p_{y,2} + a_2 \Delta x_2\right) \tag{2.55}$$

$$\frac{\partial H_2}{\partial p_{x,1}} = +\frac{p_{z,1} p_{x,1}}{|\boldsymbol{p_1}| p_{T,1}} \tag{2.56}$$

$$\frac{\partial H_2}{\partial p_{x,2}} = -\frac{p_{z,2} p_{x,2}}{|\boldsymbol{p_2}| p_{T,2}} \tag{2.57}$$

$$\frac{\partial H_2}{\partial p_{y,1}} = +\frac{p_{z,1} p_{y,1}}{|\boldsymbol{p_1}| p_{T,1}} \tag{2.58}$$

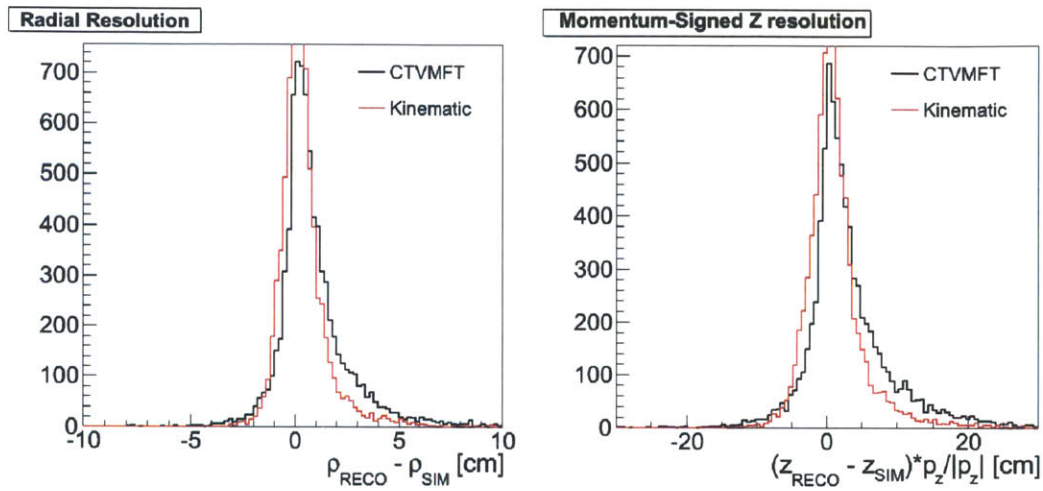
$$\frac{\partial H_2}{\partial p_{y,2}} = -\frac{p_{z,2} p_{y,2}}{|\boldsymbol{p_2}| p_{T,2}} \tag{2.59}$$

$$\frac{\partial H_2}{\partial p_{z,1}} = -\frac{p_{T,1}}{|\boldsymbol{p_1}|^2} \tag{2.60}$$

$$\frac{\partial H_2}{\partial p_{z,2}} = +\frac{p_{T,2}}{|\boldsymbol{p_2}|^2} \tag{2.61}$$

Together with the required partial derivatives, this forms the collinearity constraint in the vertex fit.

The physics performance of this vertex fitter and the constraint have been validated against a comparable vertex fit using the CTVMFT package [65], which has been used in the CDF reconstruction software. The two vertex fitters were run on

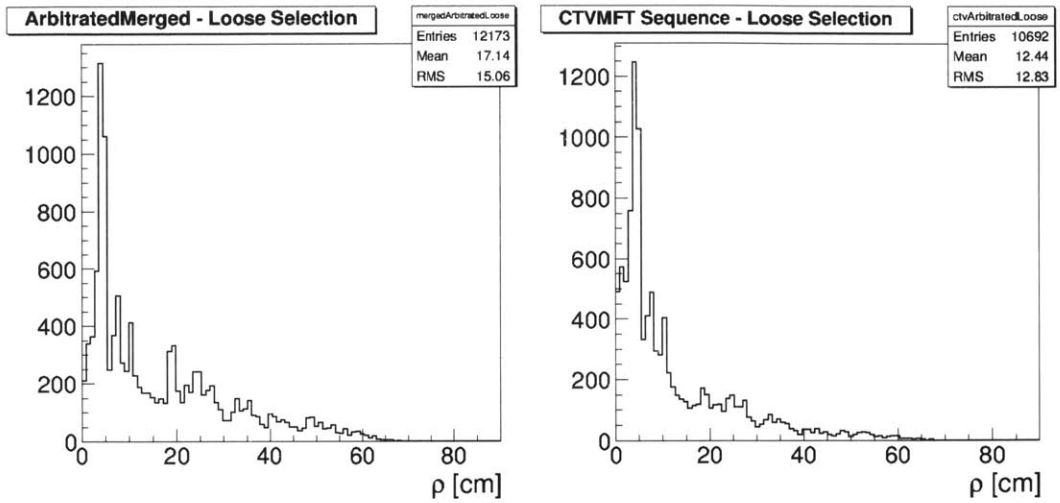


(a) Radial Resolution

(b) Z Resolution

Figure 2-6: Comparison of the resolution of the reconstructed conversion position for the KinematicConstrainedVertexFitter and CTVMFT in QCD Monte Carlo both in the radial direction (left) and along the z axis (right).

track pairs in a sample of 9000 QCD Monte Carlo events (80 GeV$< \hat{p}_T <$ 120 GeV) and conversion candidates were selected using only a minimum cut ($> 10^{-6}$) on the track and vertex fit probabilities. Figure 2-6 shows the radial and z resolution for the reconstructed conversion position. The sharper peak and reduced upper tail indicate superior position reconstruction performance as compared to CTVMFT. Figure 2-7 shows the radial distributions of the reconstructed conversions, where the peak structure corresponds to the layer structure of the barrel pixel and strip tracking detectors. The sharper and more distinctly resolved structures compared to CTVMFT further indicate the improved position reconstruction, and the increased number of candidates in the material peaks also indicate an improvement in reconstruction efficiency of approximately 10%. The sensitivity to the above optimizations arises due to the much larger material budget in CMS as compared to CDF.



(a) KinematicConstrainedVertexFitter

(b) CTVMFT

Figure 2-7: Comparison of the radial distribution of reconstructed conversion vertices for the KinematicConstrainedVertexFitter (left) and CTVMFT (right) in QCD Monte Carlo.

2.6 Primary Vertex Reconstruction and Selection

Since two-photon events do not necessarily have any prompt charged particles associated with the hard interaction, it is not possible to determine unambiguously the location of the primary interaction vertex along the beam-line. Nevertheless, even in the dominant gluon-fusion production mode, the Higgs may be produced in association with additional jets, or at least soft hadronic activity from the underlying event. This makes it possible to reconstruct the primary vertex associated with the Higgs in nearly all cases, however in events with a significant number of pileup interactions, the selection of the primary vertex corresponding to the hard interaction producing the photons may still be ambiguous. The strategy for the selection of the primary vertex is to inclusively reconstruct all of the primary interaction vertices in the event, and then select the one which is most compatible with the di-photons under consideration, exploiting the magnitude and direction of the hadronic activity, as well as the possible presence of reconstructed conversions in the tracker.

2.6.1 Primary Vertex Reconstruction

Primary vertices are reconstructed starting from an inclusive set of charged particles reconstructed in the tracking detectors. Some preselection cuts on track quality and compatibility with the beamline are applied, in order to reduce the number of fake tracks or those from secondary vertices. The tracks are first divided into subsets corresponding to distinct primary vertices, by clustering the tracks along the z-coordinate of the beamline. This is done by finding the division of tracks along the z coordinate into N clusters, which globally minimizes the χ^2 compatibility for the z coordinate and uncertainties of the tracks with N common vertices along the beamline. Each cluster of tracks is then fit to a common vertex in three-dimensional space, using an adaptive vertex fitting procedure, in which tracks with poor compatibility to the vertex can be iteratively down-weighted during the fit. The effective spatial resolution of the fit is $\mathcal{O}(50\ \mu\text{m})$ in the transverse plane and $\mathcal{O}(100\ \mu\text{m})$ in the longitudinal direction. Since the width of the LHC beams in the transverse plane is comparable to the vertex position resolution, the position and width of the luminous region is added as a constraint to the fit, thereby improving the final position resolution in the transverse plane. The output of the primary vertex reconstruction is a list of vertices, including the reconstructed position and uncertainties, as well as the list of tracks associated to each vertex.

2.6.2 Primary Vertex Selection

Hadronic Activity

In order to select the primary vertex among those which have been reconstructed which is most compatible with the di-photon pair under consideration, the properties of the associated charged hadronic activity are exploited. Due to larger momentum transfer in the hard interaction, there is in general more hadronic activity for Higgs or di-photon vertices as compared to minimum bias vertices from pileup. In addition, although not all of the momentum in the event is measured from the charged tracks, the magnitude and direction of the hadronic activity from the signal vertex

should approximately balance the momentum of the di-photon system in the transverse plane. In order to quantify the compatibility, three variables are constructed for each primary vertex reconstructed in the event:

1. sumpt2 $= \sum_i |\vec{p}_T^{\,i}|^2$: The sum of the squared transverse momentum of all of the charged tracks associated to the vertex. This variable is a measure of the overall amount of hadronic activity.

2. ptbal $= -\sum \left(\vec{p}_T^{\,i} \cdot \frac{\vec{p}_T^{\,\gamma\gamma}}{|\vec{p}_T^{\,\gamma\gamma}|} \right)$: The sum of the transverse momentum of all the charged tracks associated to the vertex, projected onto the axis of the diphoton momentum in the transverse plane. This is a measure of the degree to which the charged hadronic activity balances the transverse momentum of the di-photon system.

3. ptasym $= \frac{|\sum_i \vec{p}_T^{\,i}| - \vec{p}_T^{\,\gamma\gamma}}{|\sum_i \vec{p}_T^{\,i}| + \vec{p}_T^{\,\gamma\gamma}}$: An asymmetry quantity formed from the total transverse momentum of all the charged tracks associated to the vertex with the transverse momentum of the di-photon system. This is a further measure of the balancing between the charged hadronic activity and the di-photon system

Use of Photon Conversions

For events where one or more photons is associated to a reconstructed conversion in the tracking detectors, the pointing information from the conversions is additionally used. There are two methods used to reconstruct the z position at the beamline based on the reconstructed conversions, as illustrated in Figure 2-8. The first uses the projection of the reconstructed conversion momentum from the reconstructed conversion vertex back to the beamline. Since the position and momentum resolution of the conversion reconstruction varies significantly as a function of location in the detector, and in particular conversions which are reconstructed far from the interaction region have poorer resolution, on account of having fewer measurements in the tracking detectors. In this case, the z position at the beamline is computed instead by projecting a line joining the reconstructed conversion vertex to the reconstructed

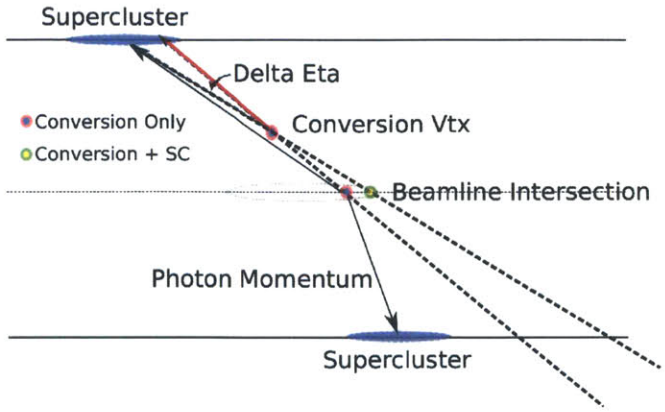


Figure 2-8: A diagram illustrating the two methods used to calculate the z position at the beamline using reconstructed photon conversions. The first method uses the projection of the conversion momentum from the conversion vertex back to the beamline, whereas the second uses a projection of the line joiniing the reconstructed SuperCluster position in the Ecal to the reconstructed conversion vertex.

SuperCluster position. The use of the position information from the Ecal mitigates the poor resolution of the reconstructed conversion in this case.

Depending on the region of the detector in which each conversion has been reconstructed, an uncertainty is assigned to the estimate of the z position at the beamline. For events in which both photons are associated with reconstructed conversions, the final estimate is taken from the weighted average of the two positions, taking into account these uncertainties. Finally for each primary vertex, a variable is constructed quantifying the compatibility of the vertex with the conversion pointing information:

- Conversion Pull $= \frac{|z_{conv} - z_{vtx}|}{\sigma_z^{conv}}$

where z_{vtx} is the z position of the reconstructed primary vertex, z_{conv} is the z position at the beamline determined from the conversion pointing, and σ_z is the uncertainty on z_{conv}. As the resolution on z_{vtx} is negligible compared to that of the conversion reconstruction, it is neglected.

Multivariate Selection

The final primary vertex selection is performed using a BDT classifier which has been trained to discriminate between primary vertices from Higgs Monte Carlo events vs

additional minimum bias vertices in the same sample. The input variables are the three hadronic recoil variables plus the one conversion variable discussed above. For events in which no reconstructed conversion is present, the conversion pull variable is set to a dummy value for both the training and evaluation, such that it is effectively ignored by the BDT. The primary vertex with the highest BDT score is selected for subsequent computation of the photon kinematics as well as use in the photon identification.

Vertex Probability Estimate

Because the mass resolution is significantly degraded for events in which the incorrect primary vertex has been selected, it is desirable to classify events taking into account the certainty with which the correct primary vertex has been selected. Since the mass resolution only depends on the proximity in z between the selected primary vertex and true location, a correct primary vertex selection is defined for this purpose as being within 1.0 cm of the true z position. The per-event probability that the correct primary vertex has been selected is estimated using an additional BDT classifier, based on the properties of the diphoton system and the reconstructed primary vertices. This classifier is trained to discriminate between events in which the correct primary vertex is chosen within 1.0 cm, and those in which it is not. The input variables are:

- $p_T^{\gamma}\gamma$ The di-photon transverse momentum. In general, events with larger di-photon transverse momentum have more hadronic activity and a higher chance of selecting the correct vertex.

- Number of reconstructed primary vertices in the event: Events with many primary vertices are less likely to have the correct vertex selected, since the chances are higher that at least one minimum bias vertex randomly has a better compatibility with the di-photon system.

- Vertex selection BDT score for the three highest ranked vertices: This contains information on how signal like is the selected vertex, as well as the next two

most likely candidates.

- Distance in z between the selected vertex and the next two vertices
- Number of photons associated to reconstructed conversions (0,1,2): The pointing information increases the vertex selection efficiency for events with reconstructed conversions

The final per-event estimate of the correct vertex selection probability is constructed from a simple linear transformation of the above BDT classifier response, using the fraction of events with the correct primary vertex selected in Higgs Monte Carlo.

2.7 Photon Selection and Identification

Because of the very large production cross-section for photon-jet events at the LHC, it is necessary to maintain a relatively low rate for jets to be reconstructed as photons in the final analysis. There are two handles in general for discrimination between jets and prompt photons. As discussed, the most photon-like jets result from fragmentation into a leading π^0 or η, which subsequently decay into a pair of photons with small opening angle. Even if the two photons cannot be cleanly distinguished, such objects nevertheless have wider shower profiles on average than a single photon incident upon the Ecal. This is particularly true along the η axis of the cluster, since the discriminating power resulting from the ϕ profile of the shower is washed out by the effect of the magnetic field, which can elongate the electromagnetic cluster in the ϕ direction for both converted single photons as well as pairs of photons from a neutral meson decay where at least one of the photons has converted. The second handle is the use of isolation energy sums. Because the π^0 or η results from jet fragmentation, there are in general additional charged and neutral particles produced in association. Since jets tend to be collimated objects, these additional particles tend to be produced close to the reconstructed photon in the detector. This leads to a generic class of discriminating variables in the form of isolation energy sums, which

generally consist of the sum of some particular type of reconstructed energy in a cone around the reconstructed object. For this purpose, it is convenient to define cones in terms of an $\eta - \phi$ metric, with the radius with respect to the reconstructed photon defined by $R^2 = \Delta\eta^2 + \Delta\phi^2$. In order to ensure that the energy from the photon itself is not included in this some, it is necessary to define a smaller veto region inside the cone, inside of which energy is excluded from the isolation sum.

In general, to make the optimal use of each event, it is desirable to use fine-grained information on a per-photon basis about the likelihood that a given reconstructed photon arises from a prompt photon as opposed to a fragmenting jet. Since the production cross sections for photon-jet and di-jet events varies substantially with the jet kinematics, and the distributions of shower profile and isolation variables vary as well with the jet kinematics as well as the possible evolution of the showers from converted photons in the upstream material, the optimal strategy is a to use a multivariate classifier. In order to have a well defined event selection for the MVA training as well as for the analysis, it is necessary to nevertheless apply a loose cut-based preselection, including fiducial and transverse momentum cuts and loose isolation and identification requirements, as well as an electron veto. The preselection is designed to be nominally tighter than the requirements imposed in the trigger, as well as the EM-enrichment filter used in some of the Monte Carlo samples, but otherwise to maintain as high efficiency as possible. The cut-based preselection is followed by the application of a multivariate discriminator based on shower-profile and isolation quantities designed to distinguish between photons and fakes from jets. This per-photon multivariate discriminator is not cut on, but rather fed forward to the per-event multivariate discriminator described in Section 2.8.

2.7.1 Definition of Photon Variables

Both the preselection and the multivariate photon-jet discriminator rely on a combination of photon-level variables characterizing the shape of the electromagnetic shower in the calorimeter, as well as the amount of energy deposited in the vicinity of the photon. Event-level variables characterizing the amount of additional energy

from pileup interactions are used either to correct some of the isolation quantities, or as MVA inputs. The exact set of variables differs slightly between the 7 TeV and 8 TeV data analysis, mainly but not exclusively in the isolation variables, but the inclusive list and definition of variables are given below.

- Fiducial and kinematic:
  - η_{SC}: The pseudorapidity of the supercluster in absolute detector coordinates, computed from the pseudorapidity of the vector joining the point (0,0,0) to the reconstructed supercluster position
  - p_T: The transverse momentum of the photon, computed using the regression-corrected Supercluster energy, with the direction taken as the vector joining the selected primary vertex to the reconstructed SuperCluster position.
  - $m_{\gamma\gamma}$: The mass of the diphoton system, using four vectors computed from the regression-corrected Supercluster energy, with the direction taken as the vector joining the selected primary vertex to the reconstructed SuperCluster position.
- Detector-based isolation:
  - EcalIso03: The scalar sum of ECal rechit transverse energy in a cone of $\Delta R < 0.3$ around the photon supercluster, excluding an inner veto cone 3.5 crystals in radius, as well as an inner veto strip 2.5 crystals in width along the η direction. The veto regions defined in terms of number of crystals are converted into an equivalent region in $\eta - \phi$ space on a per-photon basis, using the detailed detector geometry.
  - EcalIso04: As EcalIso03 but with the outer cone size increased to $\Delta R < 0.4$.
  - HcalIso03: The scalar sum of transverse energy from Hcal deposits in a cone of $\Delta R < 0.3$ around the photon supercluster, excluding an inner-veto cone of $\Delta R < 0.15$.

- HcalIso04: As HcalIso03 but with the outer cone size increased to $\Delta R < 0.4$.
- TrkIso03: The scalar sum of track transverse momentum in a cone of $\Delta R < 0.3$ around the photon direction, considering only tracks compatible with the highest $\sum p_T^2$ primary vertex. While this is not the primary vertex used for the rest of the analysis, this choice is the most closely corresponding offline choice to what is used in the HLT.
- TrkIso03Selected: The scalar sum of track transverse momentum in a cone of $\Delta R < 0.3$ around the photon direction, and excluding an inner veto cone of $\Delta R < 0.02$, considering only tracks compatible with the selected primary vertex, requiring the longitudinal impact parameter between the track and the selected primary vertex $|\Delta z| < 1.0$ cm and the transverse impact parameter $|d_{xy}| < 0.1$ cm.
- TrkIso04Worst: As TrkIso03Selected, except that the outer cone size is increased to $\Delta R < 0.4$ and the primary vertex used for track compatibility is the one which yields the largest isolation sum. The use of this variable ensures that a jet will not be mistakenly classified as isolated in the case that the selected primary vertex is different from the interaction which produced the jet.
- H/E: The ratio of Hcal energy deposits in a cone of $\Delta R < 0.15$ around the supercluster, divided by the supercluster energy.

• Particle flow-based isolation:

- ChargedPFIso02Selected: The scalar sum transverse momentum of charged hadron particle flow candidates in a cone of $\Delta R < 0.2$ around the photon direction, excluding an inner veto cone of $\Delta R < 0.02$, computed with respect to the selected primary vertex. Particle flow candidates are restricted to those having a longitudinal impact parameter with respect to the selected primary vertex $|\Delta z| < 0.2$ cm and a transverse impact param-

eter $|d_{xy}| < 0.1$ cm. The choice of cone size corresponds roughly to the EM-enrichment filter in some of the Monte Carlo samples.

  - ChargedPFIso03Selected: As ChargedPFIso02Selected but with the outer cone size increased to $\Delta R < 0.3$.
  - ChargedPFIso03Worst: As ChargedPFIso03Selected, except that the primary vertex used for track compatibility is the one which yields the largest isolation sum.
  - PhotonPFIso03: The scalar sum transverse energy of photon-type particle flow candidates in a cone of $\Delta R < 0.3$ around the photon direction, computed with respect to the beamline reference point of the particle flow candidate. For photons in the barrel, an inner veto strip of $|\Delta\eta| < 0.015$ is excluded from the isolation sum, and for photons in the endcap instead an inner veto cone of $\Delta R < 0.07$ is excluded

- Shower profile:
  - R_9: The sum of ecal rechit energy in the 3x3 grid of crystals centred on the seed crystal, divided by the raw supercluster energy.
  - $\sigma_{i\eta i\eta}$: The width of the seed BasicCluster along the η-direction, computed in terms of crystal indices, and using weighted RecHit positions in the same way as for the cluster position reconstruction.
  - $Cov(i\eta i\phi)$: The $\eta - \phi$ covariance of the seed BasicCluster, computed in terms of crystal indices, and using a log-energy weighting of the crystal positions taken at the nominal shower depth.
  - Supercluster η Width: The width of the supercluster along the η direction, computed using the energy weighted distribution of crystal positions taken at the nominal shower depth.
  - Supercluster ϕ Width: The width of the supercluster along the ϕ direction, computed using the energy weighted distribution of crystal positions taken at the nominal shower depth.

- E_{2x2}/E_{5x5}: The ratio of the maximum energy 2x2 crystal grid and the energy in a 5x5 crystal grid centered on the seed crystal.
- σ_{RR}: The sum in quadrature of the width of the preshower energy distribution in the two preshower planes.

- Event level variables sensitive to pileup:
  - N_{vtx}: The number of reconstructed primary vertices in the event.
  - ρ: A per-event measure of the amount of transverse energy from pileup interactions. This is constructed from the median transverse energy density of all kT-reconstructed jets in the event, where the kT algorithm tends to produce a large number of soft jets such that the median of this distribution is relatively insensitive to the hard interaction [66].

The shower profile variables in general suffer from relatively poor modelling in the Monte Carlo simulation. One cause of this is imprecise simulation of the electromagnetic shower evolution, such that the width of the shower is sensitive to fine tuning of the Bremstrahlung model in Geant 4. Another source is the presence of out of time pileup, which changes the relative contribution of crystals with a small amount of energy at the edges of the shower. This effect is not fully included in the simulation, since only one preceding bunch crossing at -50ns is simulated, whereas the RecHit reconstruction may still be sensitive to interactions from earlier crossings.

To correct for this, linear scalings are applied to the shower profile variables in the Monte Carlo prior to the application of preselection cuts or computation of the photon identication MVA or the computation of scale factors. The coefficients for these linear transformations are tuned to match the data distributions in $Z \to ee$ events and for high mass di-photon events.

2.7.2 Photon Pre-selection

The pre-selection of photons entering the analysis is based on a simple set of rapidity and relative transverse momentum cuts as well as a very loose set of identification and

isolation cuts intended to maximize signal efficiency, but to be nominally tighter than both the cuts employed by the trigger as well as those implicit in the EM-enrichment filter employed in a subset of the background Monte Carlo samples. Photons are required to satisfy the fiducial requirement on the Supercluster pseudorapidity, computed in absolute detector coordinates with respect to (0,0,0), of $|\eta_{SC}| < 1.4442$ or $1.566 < |\eta_{SC}| < 2.5$. The first pseudorapidity range corresponds to the Ecal barrel, excluding the gap/transition region, and the second range corresponds to the overlap of the Ecal endcap, excluding the gap/transition region and the tracking acceptance. Although the Ecal endcap coverage extends to $|\eta| < 3$, the analysis is nevertheless restricted to the tracking acceptance in order to allow for the use of track-based isolation, and validation of photons using $Z \to ee$ events. Detector performance and signal to background ratio are also degraded in this far forward region. The photons are also required to pass the relative transverse momentum requirements for the larger and smaller transverse momentum photons $p_T^{lead} > m_{\gamma\gamma}/3$ and $p_T^{sub-lead} > m_{\gamma\gamma}/4$, where the four-vectors of the photons are computed using the regression-corrected Supercluster energy, with the direction taken as the vector joining the selected primary vertex to the corresponding SuperCluster position. The use of relative transverse momentum cuts scaled by the mass reduces variations in acceptance as a function of Higgs mass, and reduces kinematic threshold effects on the shape of the background di-photon mass distribution. The further pre-selection cuts differ slightly between the 2011 and 2012 analyses due to the introduction of Particle Flow isolation in the 2012 analysis.

The cut values for the preselection in 2011 and 2012 data are given in Tables 2.1 and 2.2. In both cases cut values differ according to the detector region (barrel/endcap) and R_9 of the photon in order to follow differences in detector characteristics and purity, and corresponding differences in cut values at the HLT level.

Electron Veto

The reconstruction and selection variables used for the identification of photons are for the most part insensitive to the differences between electrons and photons, and

	Barrel		Endcap	
	$R_9 > 0.9$	$R_9 \leq 0.9$	$R_9 > 0.9$	$R_9 \leq 0.9$
H/E	< 0.082	< 0.075	< 0.075	
$\sigma_{i\eta i\eta}$	< 0.014		< 0.034	
EcalIso03 $-0.012\ p_T$	< 50 GeV	< 4 GeV	< 50 GeV	< 4 GeV
HcalIso03 $-0.005\ p_T$	< 50 GeV	< 4 GeV	< 50 GeV	< 4 GeV
TrkIso03 $-0.002\ p_T$	< 50 GeV	< 4 GeV	< 50 GeV	< 4 GeV
EcalIso03 + HcalIso03 - 0.17 ρ	< 3 GeV			
TrkIso03Selected	< 2.8 GeV			
TrkIso03	< 4 GeV			

Table 2.1: Preselection isolation and identification cuts for 2011 data.

	Barrel		Endcap	
	$R_9 > 0.9$	$R_9 \leq 0.9$	$R_9 > 0.9$	$R_9 \leq 0.9$
H/E	< 0.082	< 0.075	< 0.075	
$\sigma_{i\eta i\eta}$	< 0.014		< 0.034	
EcalIso03 $-0.012\ p_T$	< 50 GeV	< 4 GeV	< 50 GeV	< 4 GeV
HcalIso03 $-0.005\ p_T$	< 50 GeV	< 4 GeV	< 50 GeV	< 4 GeV
TrkIso03 $-0.002\ p_T$	< 50 GeV	< 4 GeV	< 50 GeV	< 4 GeV
ChargedPFIso02	< 4 GeV			

Table 2.2: Preselection isolation and identification cuts for 2012 data.

attempt to maintain efficiency also for converted photons. In the absence of an explicit electron veto, isolated electrons from electroweak processes such as Drell-Yan dielectron production would pass the photon selection with high efficiency. In this case, events from Drell-Yan production, as well as smaller cross section processes such as $W + \gamma \rightarrow e\nu$ would dominate the di-photon selection. Photon candidates are therefore rejected if they are also reconstructed as a prompt electron. In order to maintain efficiency for converted photons, these electrons have some additional requirements imposed in order to ensure they did not arise from a converted photon. Photon candidates are rejected if they share a Supercluster with an (ecal-seeded) electron candidate passing the appropriate conversion rejection requirements.

The first conversion rejection requirement is that the electron GSF track has zero missing hits prior the first hit on the track. This ensures that the electron is consistent with having been produced at the interaction point, rather than from a conversion vertex further out in the detector. To maintain robustness against inactive channels in the pixel and tracking detectors, modules which have been flagged as inactive in the reconstruction are excluded from this missing hit count. Since some photons may also convert early in the detector, within the beampipe material or within the first active layer itself, electrons from converted photons may nevertheless have zero missing hits. It is therefore also necessary to apply an explicit conversion rejection criteria by checking whether the electron shares a track with a reconstructed and selected conversion in the tracker. The conversion reconstruction for this purpose is as discussed in Section 2.5. The conversion selection in this context includes a cut on the fit probability $> 10^{-6}$ and a cut on the decay length > 2.0 cm to exclude randomly paired tracks from the interaction point. In order to remove mis-reconstructed conversions from Bremsstrahlung photons near the initial electron track, an additional cut requires that neither track in the conversion has any hits on the trajectory earlier than the conversion vertex. The effect of this requirement for electrons from $Z \rightarrow ee$ events in Monte Carlo matched to reconstructed conversions is shown in Figure 2-9.

The effect of the electron veto on converted photons is illustrated by Figure 2-10 showing the radial distribution of reconstructed conversions matched to converted

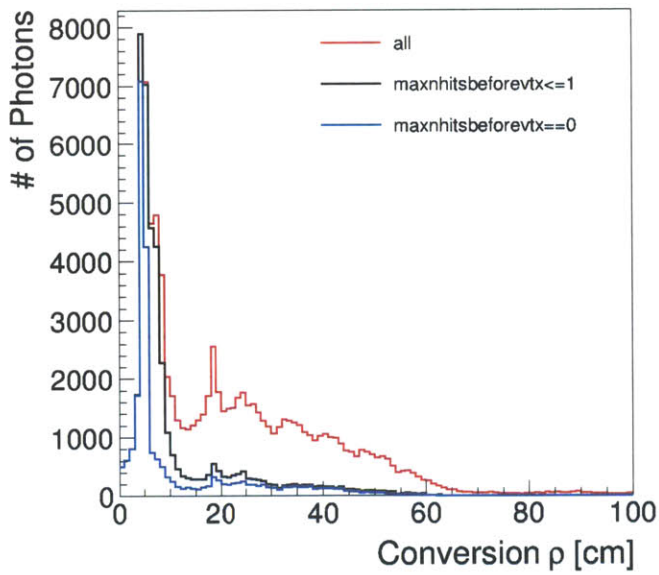


Figure 2-9: The radial distribution of reconstructed conversions matched to electrons from $Z \rightarrow ee$ Monte Carlo events. The distribution is shown with no cut on the number of hits before the conversion vertex (red), requiring 1 or fewer such hits on each track (blue) and requiring no such hits (black). The contribution and suppression of mispaired Bremsstrahlung conversions are clearly visible.

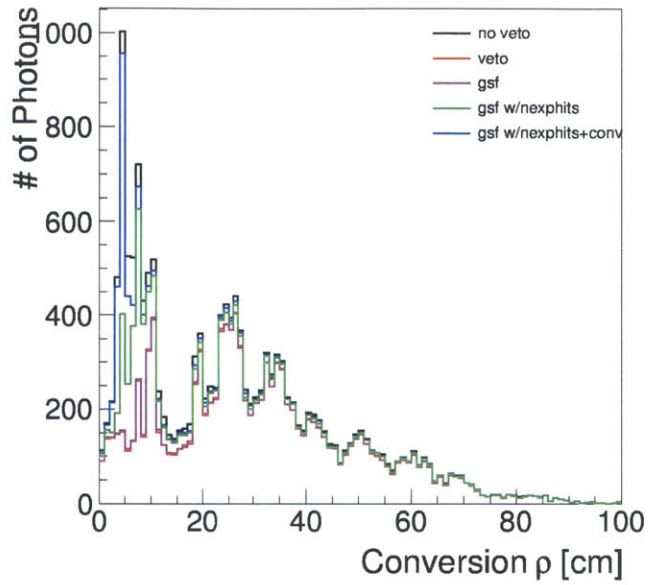


Figure 2-10: The radial distribution of reconstructed conversions matched to converted photons from $H \to \gamma\gamma$ Monte Carlo events. The distribution is shown with no electron veto (black), the final electron veto used for the analysis (blue), a version of the electron veto dropping the conversion matching (green), a version dropping additionally the missing hits requirement for the electron (magenta), and finally a veto on all electron pixel seeds (red). The main effect is the substantial recovery of efficiency for photons which convert early in the tracker compared to the tighter electron vetos.

photons from $H \to \gamma\gamma$ Monte Carlo, showing the effect of the electron veto, as well as comparing to several tighter variations of electron veto. A very severe variation of the electron veto included in this comparison is the pixel seed veto, in which photon candidates are rejected as long as their SuperCluster has a matched pixel seed for the electron track reconstruction, as described in 2.3.1. This tight pixel seed veto would reject essentially all photons which convert in the first two layers of the pixel detector, whereas the more sophisticated electron veto described above, maintains a high efficiency also for such photons.

Preselection Efficiency Measurement and Scale Factors

The efficiency for the photon identification pre-selection is measured using electrons from $Z \rightarrow ee$ events, for all requirements aside from the electron veto. This is done using the tag and probe method, starting with a trigger requiring an electron with $p_T > 32$ GeV and some id requirements on one leg, and simply a SuperCluster with $E_T > 17$ GeV on the second leg. The efficiency which is measured is therefore the efficiency from the SuperCluster to preselection level. The efficiency for a photon passing the acceptance cuts to be reconstructed as a SuperCluster is very close to 1, and assumed to be well modelled by the Monte Carlo.

Events are selected with at least two SuperClusters with $E_T > 20$ GeV, both passing the loose leg of the trigger. One SuperCluster is chosen randomly as the tag, and is required to pass an offline electron selection as well as the tight leg of the trigger, otherwise the event is discarded. The remaining events are divided into passing and failing samples according to whether the probe SuperCluster passes the photon preselection requirements. The events are further categorised as for the preselection requirements, according to whether the probe is in the barrel or endcap, and according to $R_9 >(<)0.9$. The preselection efficiency in each category is extracted from a simultaneous maximum likelihood fit to the passing and failing samples, with floating signal and background components. The signal Drell-Yan events are modelled by a Breit-Wigner convoluted with a crystal ball, with the mean and width of the Breit-Wigner constrained to the PDG mass and width of the Z, and the mean, resolution and tail parameters of the crystall ball freely floating. The background component, mainly from QCD and W+jet events with two or one fake electrons from jets, is modelled by an exponential function with floating slope parameter. Both the signal and background are additionally multiplied by an error function to model possible kinematic turn-on, with mean and width freely floating in the fits.

The mass distribution for the tag-probe pair for 7 TeV Monte Carlo events is shown in Fig. 2-11, on the left (right) for the failing (passing) events, and in the top (bottom) plots for the Barrel (Endcap) probes together with the fitted distributions.

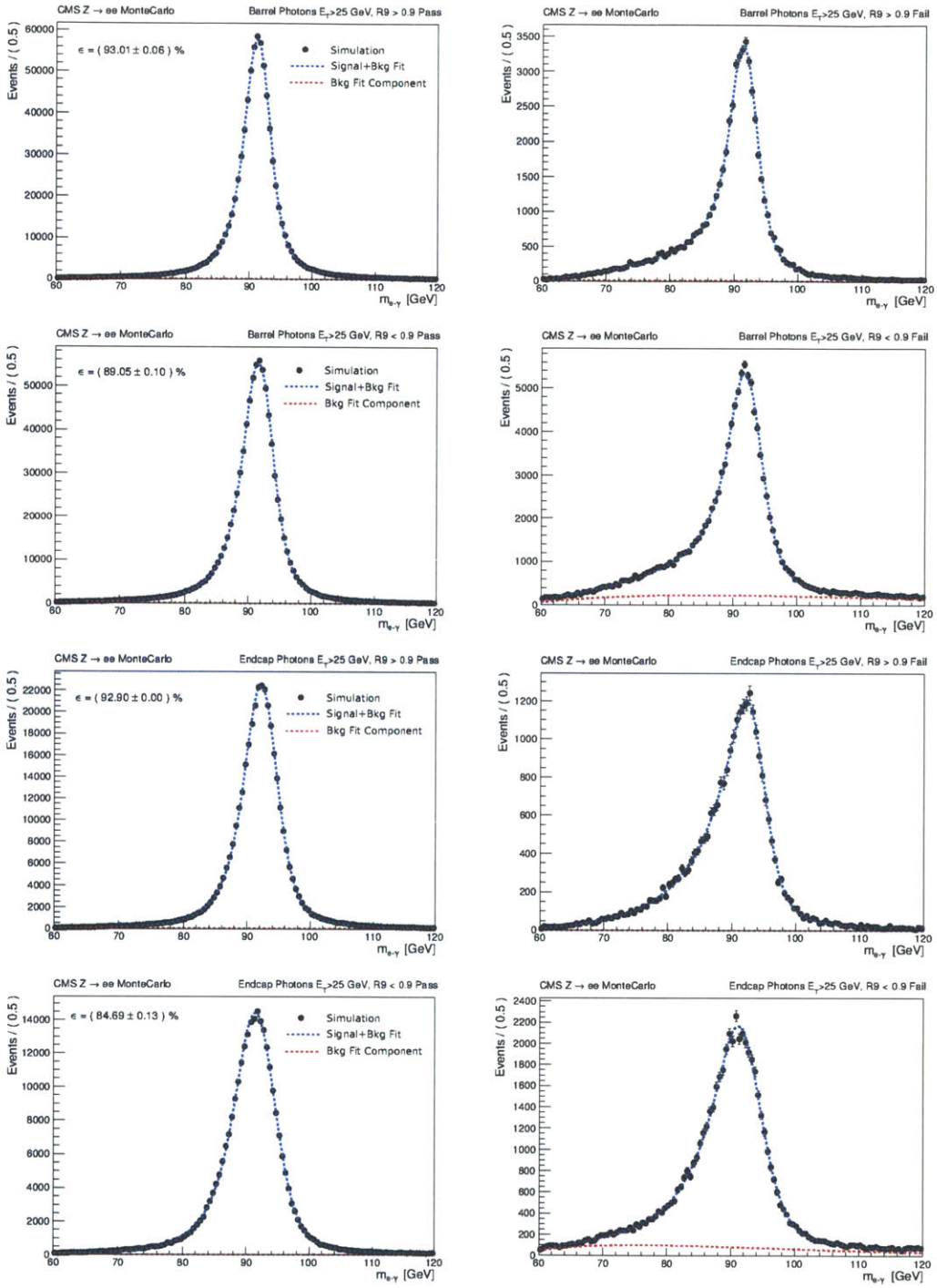


Figure 2-11: Monte Carlo: Mass of the tag-electron plus probe-photon pair computed in the tag-and-probe procedure for measuring the Monte Carlo to data photon efficiency scale factors. On the left the passing probes, on the right the failing ones, ordered in Photon categories.

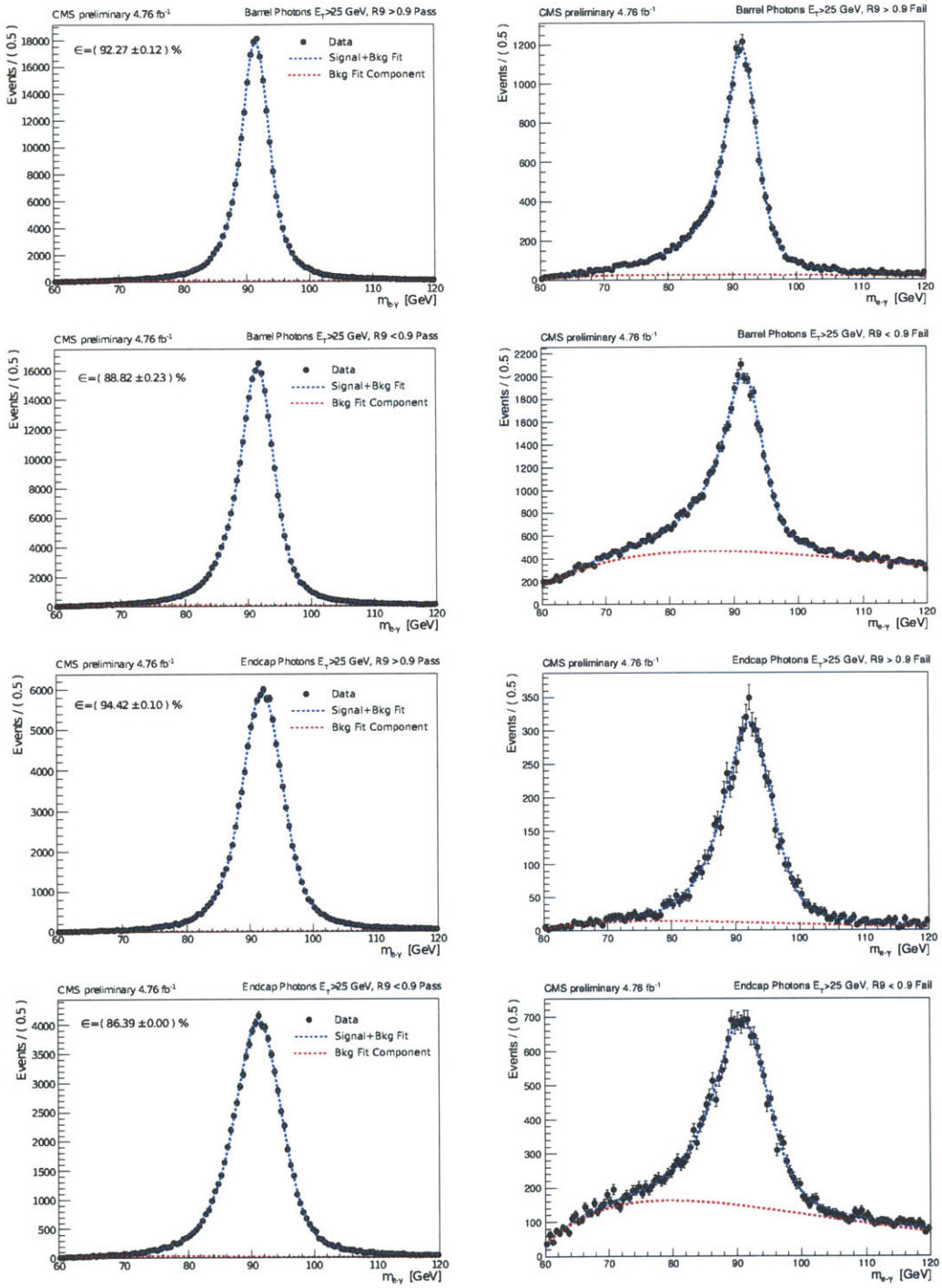


Figure 2-12: 7 TeV CMS data: Mass of the tag-electron plus probe-photon pair computed in the tag-and-probe procedure for measuring the Monte Carlo to data photon efficiency scale factors. On the left the passing probes, on the right the failing ones, ordered in Photon categories.

Table 2.3: Photon preselection efficiencies in Monte Carlo and data, as well as the Monte Carlo to data scale-factors for four categories of photons seperately for 7 TeV and 8 TeV data, Monte Carlo, and preselection. Quoted uncertainties are statistical only.

Category	Data	MC	Data/MC
EB R9>0.9 2011	0.9227 ± 0.0012	0.9301 ± 0.0006	0.992
EB R9<0.9 2011	0.8882 ± 0.0023	0.8905 ± 0.0010	0.998
EE R9>0.9 2011	0.9442 ± 0.0010	0.9290 ± 0.0003	1.016
EE R9<0.9 2011	0.8639 ± 0.0001	0.8469 ± 0.0013	1.020
EB R9>0.9 2012	0.9894 ± 0.0002	0.9916 ± 0.0001	0.998
EB R9<0.9 2012	0.9327 ± 0.0006	0.9369 ± 0.0003	0.996
EE R9>0.9 2012	0.9832 ± 0.0007	0.9771 ± 0.0002	1.006
EE R9<0.9 2012	0.9298 ± 0.0014	0.9298 ± 0.0003	0.999

The distributions for 7 TeV data are shown in Figure 2-12. The resulting efficiencies and scale-factors (Including also equivalently derived results for 8 TeV data and Monte Carlo simulation), together with their uncertainty are shown in Table 2.3. Absolute efficiencies for 8 TeV are higher due to the looser isolation cuts in the 8 TeV preselection, although data/simulation scale factors are similar and close to 1 in all cases. The scale-factors are directly applied to the Monte Carlo modelling of the expected signal yields.

Systematic uncertainties on the extrapolation from electrons to photons are considered by reweighting the probe E_T and R_9 distributions to the expected distribution for photons in a 120 GeV $H \to \gamma\gamma$ Monte Carlo sample, and redetermining scale factors. The maximum variation in the resulting scale factors which is observed with the reweighting of E_T and R_9 is 0.6% in the barrel and 1.8% in the Endcap. These variations are propagated as systematic uncertainties on the per-photon efficiency in the signal Monte Carlo samples.

2.7.3 Photon Identification MVA

The suppression of fake photons from neutral mesons in jets comes mainly from use of a multivariate photon-jet discriminator based on EM shower profile and isolation quantities, all of which have been described in 2.7.1. This is implemented with a

BDT trained on prompt photons vs non-prompt photons in Monte Carlo, passing the 7 TeV preselection. For the analysis of the 7 TeV data, the prompt photons used in the training are those from a $m_H = 121$ GeV $H \to \gamma\gamma$ sample, and the fakes are from the (EM-enriched) photon + jets sample. Because the photon+jets sample contains both prompt photons as well as fakes from jets, the object selected for the background training sample is explicitly required not to match to the prompt photon at generator-level. The 121 GeV signal sample is selected in this case because this Monte Carlo sample is not used for the construction of the final signal model, thereby avoiding any possible bias on the estimated signal efficiency due to overtraining of the photon identification classifier. Because Supercluster η is included as a an input variable to the BDT, for the purposes of the training, the Supercluster η distribution of the signal photons have been reweighted to match the distribution of the background, in order that this variable is only exploited by the BDT through its correlations with other inputs. Trainings are performed separately for barrel and endcap. The full list of input variables for the 2011 version of the photon identification MVA is given in Table 2.4. The selection of the isolation quantities is predicated on the fact that the most sensitive isolation quantity comes from the summed track transverse momentum. This occurs due to the superior resolution of the tracker compared to the calorimeters for additional particles in jets, as well as the ability to associated tracks to a particular primary vertex. In order to mitigate the effect of selecting the incorrect primary vertex in the analysis, isolation quantities are included incorporating both association to the selected primary vertex, as well as an alternate primary vertex which gives the largest isolation sum for a particular photon. This reduces the selection efficiency for fake photons from jets originating from a primary vertex other than the one selected for the analysis. The subtraction of ρ-proportional terms from the isolation sums including calorimeter based quantities corrects on average for the contribution of pileup energy to those sums on an event by event basis. The average correction cannot fully mitigate the loss of discrimination which comes from the additional partly random pileup energy collected by the isolation sums, and therefore N_{vtx} is also included in the MVA in order to modulate the weighting of the different isolation

Fiducial and Kinematic
η_{SC}
Detector-based isolation
TrkIso03Selected + EcalIso03 + HcalIso04 - 0.17ρ
TrkIso04Worst + EcalIso04 + HcalIso04 - 0.52ρ
TrkIso03Selected
H/E
Shower profile
R_9
$\sigma_{i\eta i\eta}$
Supercluster η Width
Supercluster ϕ Width
Event level variables sensitive to pileup
N_{vtx}

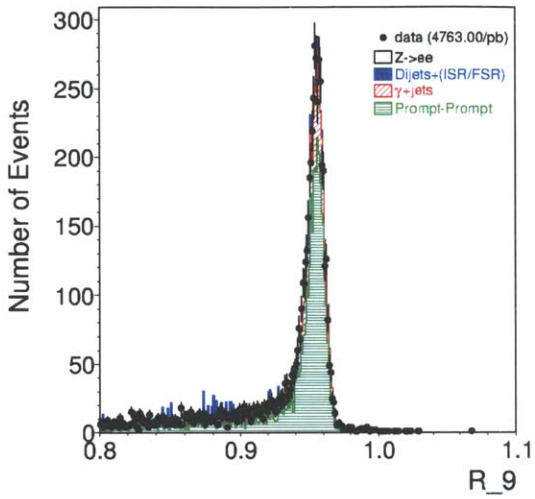
Table 2.4: Identification variables used for training of 2011 photon identification MVA.

quantities as a function of pileup.

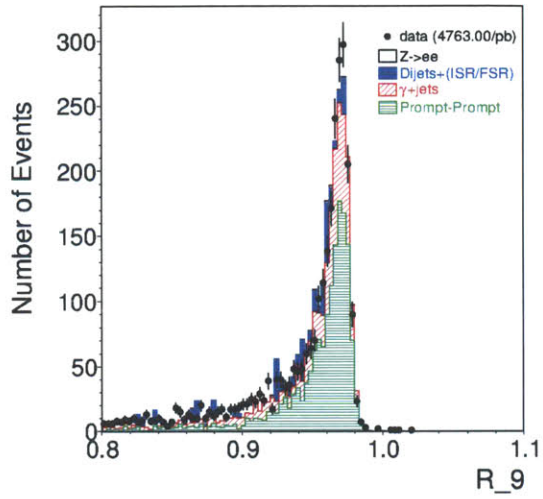
In order to correct imperfect modeling of the EM shower in the Geant 4 simulation, shower profile variables in the Monte Carlo simulation samples have been linearly rescaled in order to match the data distributions for preselected events with $m_{\gamma\gamma}$ >160 GeV in 2011 data. This high mass region is used because the fraction of prompt photons compared to fakes from jets is much higher in this region, ensuring that the scalings are appropriate for prompt photons, in case there is a significant difference in behaviour with respect to fakes from jets. This rescaling is applied for both the training and application of the identification MVA (and also for the preselection as applied to the final analysis). The distributions for shower profile variables input to the 2011 identification MVA in the high mass control region are shown in Figures 2-13 and 2-14.

The distribution of the photon ID MVA output for the leading photon in events passing the 2011 preselection in the high mass control region after rescaling of the Monte Carlo shower profile variables is shown in Figure 2-15.

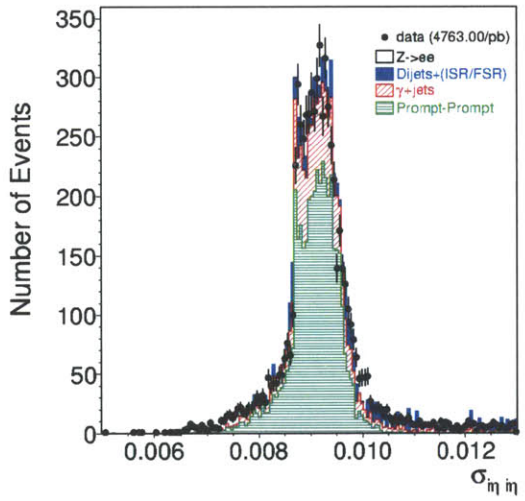
For the 2012 version of the identification MVA, an updated set of shower profile



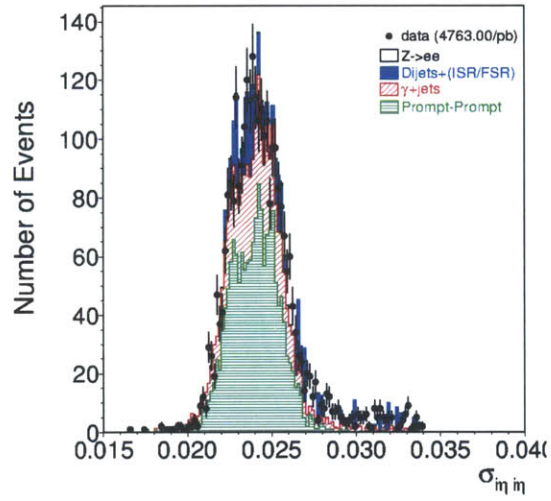
(a) R_9, Barrel



(b) R_9, Endcap

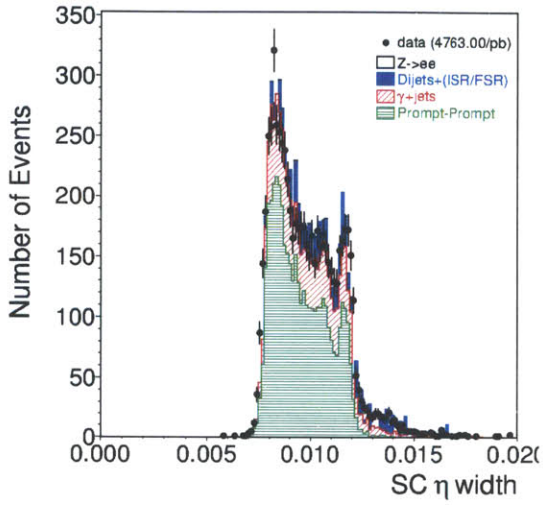


(c) $\sigma_{i\eta i\eta}$, Barrel

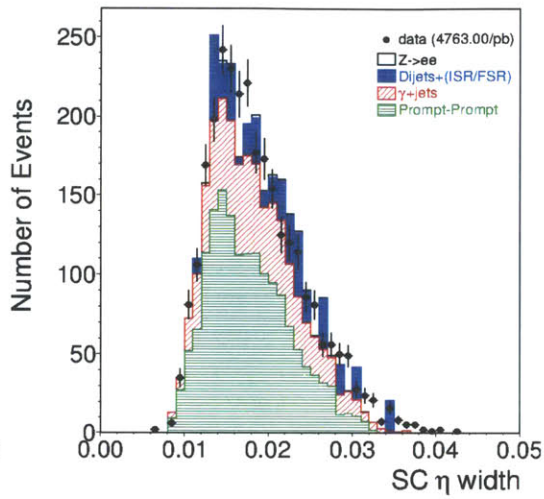


(d) $\sigma_{i\eta i\eta}$, Endcap

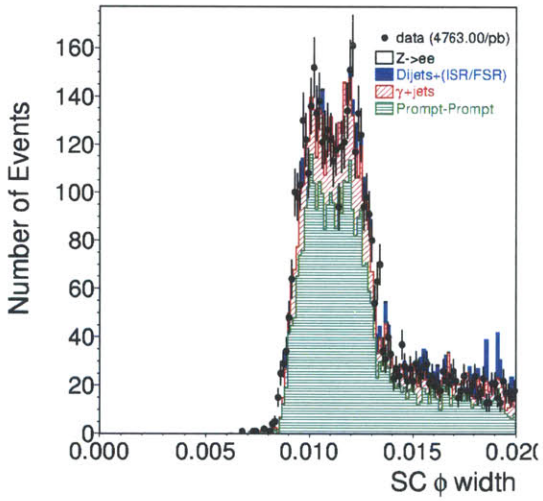
Figure 2-13: Distribution of shower profile input variables to the photon ID MVA for the leading photon in preselected diphoton events with $m_{\gamma\gamma}$ >160 GeV in 2011 data. Shower profile variables in the Monte Carlo have been linearly scaled to better match the data distributions.



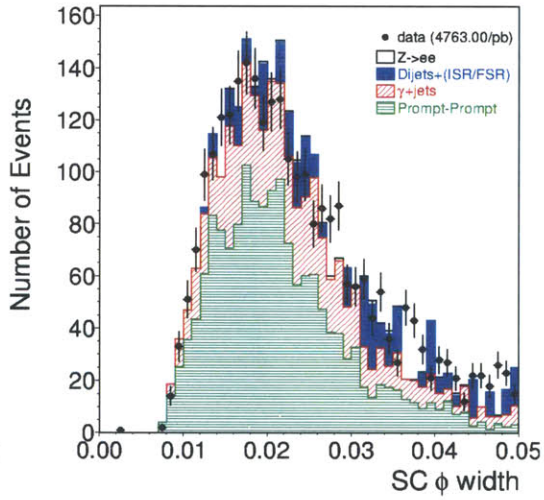
(a) SuperCluster η Width, Barrel



(b) SuperCluster η Width, Endcap

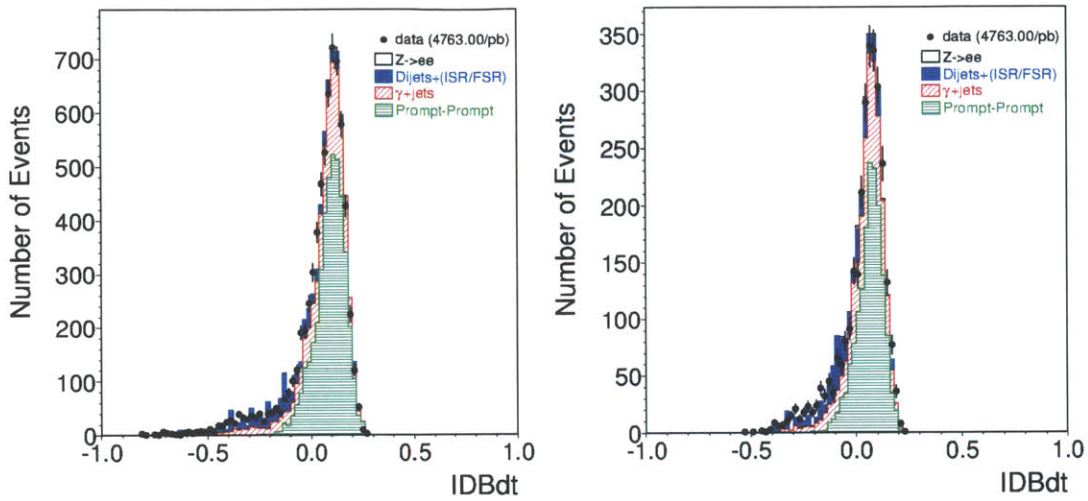


(c) SuperCluster ϕ Width, Barrel



(d) SuperCluster ϕ Width, Endcap

Figure 2-14: Distribution of shower profile input variables to the photon ID MVA for the leading photon in preselected diphoton events with $m_{\gamma\gamma}$ >160 GeV in 2011 data. Shower profile variables in the Monte Carlo have been linearly scaled to better match the data distributions.



(a) Barrel (b) Endcap

Figure 2-15: Distribution of photon ID MVA output for the leading photon in preselected diphoton events with $m_{\gamma\gamma}$ >160 GeV in 2011 data.

variables are used (including also the use of the preshower detector in the endcap), and detector based isolation quantities are replaced with those computed from particle flow objects. The use of ρ corrections in the isolation sums is replaced by the direct inclusion of ρ as an input variable to the MVA, such that the effective correction or de-weighting of isolation quantities in the MVA is moved entirely to the BDT training. The training configuration has also been modified slightly. The prompt and fake photons used for training are both taken from the photon + jet Monte Carlo sample, where events generally contain one prompt photon and one electromagnetically enriched object originating from a jet, using only objects passing the 2012 preselection. The η_{SC} and p_T distribution of the signal photons are reweighted 2-dimensionally to match that of the background. Training is again performed seperately for barrel and endcap The full list of input variables for the 2012 version of the photon identification MVA is given in Table 2.5.

Fiducial and Kinematic
η_{SC}
Detector-based isolation
ChargedPFIso03Selected
ChargedPFIso03Worst
PhotonPFIso03
Shower profile
R_9
$\sigma_{i\eta i\eta}$
$Cov(i\eta i\phi)$
Supercluster η Width
Supercluster ϕ Width
E_{2x2}/E_{5x5}
σ_{RR} (endcap-only)
Event level variables sensitive to pileup
ρ

Table 2.5: Identification variables used for training of 2012 photon identification MVA.

2.8 Diphoton mass resolution and kinematics MVA

Events are selected with two photons satisfying the preselection described above, including an additional loose preselection cut on the photon ID MVA output of $MVA > -0.3$ for the 2011 analysis and $MVA > -0.2$ for the 2012 analysis. These cuts retain more than 99% of the signal events passing the other preselection requirements, while removing around 20% of the data events with $100 < m_{\gamma\gamma} < 180\,\text{GeV}$.

The diphoton mass distribution for these selected events is shown in Fig. 2-16. Since the photon selection at this stage is very loose, the inclusive sample contains a large contribution of fake photons.

The sensitivity of the analysis is optimized taking into consideration several factors:

- Due to significant variations in detector and calibration performance, material budget, differences between converted and unconverted photons, the contribution to the di-photon mass resolution arising from the photon energy mea-

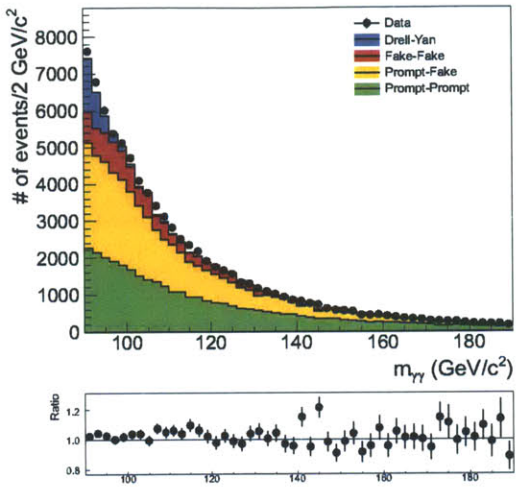


Figure 2-16: diphoton mass distribution for events with photons passing the loose preselection and MVA ID cut as well as mass-dependent p_T thresholds, comparing 2011 data to corresponding background Monte Carlo.

surements varies significantly as a function of photon rapidity and material interactions.

- The probability of correctly identifying the hard interaction among the reconstructed primary vertices in the event varies significantly as a function of the di-photon kinematics, fluctuations in the underlying event and/or fragmentation of recoiling jets from the hard interaction, as well as variations in the number and properties of pileup interactions. Selection of an incorrect primary vertex additionally contributes to the di-photon mass resolution through additional smearing in the angular resolution.

- The expected signal to background ratio against both reducible and irreducible background varies as a function of the photon kinematics both as a result of spin correlations in the production and decay of a scalar boson, as well as differences in production initial state and kinematics.

- The optimal working point for photon-jet discrimination depends on the relative amount of reducible and irreducible background and therefore varies as a

function of the photon kinematics.

Starting from this relatively loose selection, the analysis strategy is to categorize the events in order to optimize the sensitivity for extracting a signal on top of the background. In order to preserve the robustness and transparency of the results, this will be done by fitting a narrow peak on top of a smoothly falling background in each category. This requires that the categorization of events preserves the relevant physical features of the di-photon mass distributions for both signal and background. The classification of events therefore does not make any use of the di-photon mass, nor any variables which are strongly correlated with it.

The relevant properties for the optimal classification of events in the context of fitting a mass peak are the expected signal-to-background ratio, and the expected resolution of the signal peak. These depend on a number of factors as discussed above, which have been combined through the use of multivariate techniques. A BDT classifier has been trained incorporating a nominally mass-independent subset of the photon kinematics, a per-event estimate of the relative di-photon mass resolution under the hypotheses of both the correct and incorrect primary vertex selection, a per-event estimate of the probability that the correct primary vertex has been selected, and the value of the photon-jet discriminator for each photon.

2.8.1 Inputs

The kinematic variables used are the relative photon transverse momenta $p_T^{lead}/m_{\gamma\gamma}$ and $p_T^{sub-lead}/m_{\gamma\gamma}$, the photon pseudorapidities η_{lead} and $\eta_{sublead}$, and the angle between the photons in the transverse plane $\cos\Delta\phi$, which is strongly correlated with the transverse momentum of the di-photon system. These quantities are all computed from the final four-vectors of the reconstructed photons.

The two-body di-photon mass is computed simply as

$$m_{\gamma\gamma} = \sqrt{2E^{lead}E^{sub-lead}\cos\theta} \tag{2.62}$$

where E^{lead} and $E^{sub-lead}$ are the photon energies, and θ is the angle between their

momenta in three dimensions. As the position resolution of the primary vertex reconstruction in CMS is $O(100\mu\text{m})$, the mass resolution in the case of correct primary vertex selection is dominated by the resolution on the photon energy measurement. The per-event relative resolution estimate under the hypothesis of correct primary vertex selection can therefore be expressed purely in terms of the per-photon estimates of the energy resolution, σ_{Elead} and $\sigma_{Esub-lead}$, obtained from the energy regression,

$$\frac{\sigma_m^{right}}{m_{\gamma\gamma}} = \frac{1}{2}\sqrt{\frac{\sigma_{Elead}^2}{E_{lead}^2} + \frac{\sigma_{Esub-lead}^2}{E_{sub-lead}^2}} \tag{2.63}$$

In order to correct for the data-MC discrepancy in σ_E due to the imperfect shower shape modelling in the simulation, σ_E in the simulation is linearly corrected as for the shower shape variables for the 7 TeV analysis. For the 2012 analysis, the shower profile inputs to the regression are rescaled for the computation of σ_E such that rescaling the output σ_E is not necessary. A systematic uncertainty for the σ_E modeling is applied to the Monte Carlo by scaling the value of σ_E in both barrel and endcap up and down by an additional 10.0% on the nominal value after scaling. Since the energy estimate from the regression is based on Monte Carlo, the σ_E of each photon (in both data and Monte Carlo) is increased to match the estimate of the resolution in data by adding in quadrature the additional smearing applied to the single photon energy in the Monte Carlo.

In case the selected primary vertex does not correspond to the true hard-interaction of the event, the $\cos\theta$ term can contribute significantly to the di-photon mass resolution. Although the LHC beams are narrow in the transverse plane, on the order of 10's of μm, the interacting bunches are elongated along the z-direction, such that the interaction region has an average Gaussian width σ_{BS} of around 5.8 cm in 2011. Since the position of both the hard interaction vertex, and an additional pileup interaction are both randomly distributed along this distribution, the distance between the hard interaction and the selected vertex in the case of the incorrect vertex selection follows a Gaussian centred at 0 with a width of $\sigma_Z = \sqrt{2}\sigma_{BS}$.

The angular contribution to the di-photon mass resolution is then given by

$$\delta m_{vtx} = \frac{\partial m}{\partial z} \left(\vec{x}_{lead}, \vec{x}_{sub-lead}, \vec{x}_{BS}\right) \sigma_z \tag{2.64}$$

where $\vec{x}_{lead}$ and $\vec{x}_{sub-lead}$ are the reconstructed SuperCluster positions and $\vec{x}_{BS}$ is the position of the centre of the interaction region. The per-event resolution estimate on the di-photon mass under the hypothesis of incorrect primary vertex selection is then given by

$$\sigma_m^{wrong} = \sqrt{\left(\sigma_m^{right}\right)^2 + (\delta m_{vtx})^2} \tag{2.65}$$

Since the Monte Carlo does not model the variation of the beamspot length over the course of a fill, we use the average beamspot length (5.8 cm) to compute σ_m^{wrong} for both data and Monte Carlo in order to construct consistent inputs for the MVA. Finally, in order to be able to combine appropriately the relative resolution under the correct and incorrect primary vertex hypothesis, we input to the MVA as well the per-event probability that the correct primary vertex was selected, computed from a linear fit to the event-level vertex selection MVA as described in Section 2.6.2.

Likewise, the photon-jet discriminator for the lead and sub-lead photons are taken directly from the photon identification MVA described in 2.7.3. The full list of input variables for the di-photon MVA training is then given in Table 2.6.

2.8.2 Per-Event Resolution Weighting

The di-photon MVA is trained on the cross-section weighted mixture of background Monte Carlo samples, and the cross-section weighted mixture of Higgs production mechanisms for $m_h = 123$ GeV. The training sample consists of events passing the preselection, with the additional photon identification MVA cuts > -0.3 (> -0.2) for the 7 TeV (8 TeV) analysis, and with 100 GeV$< m_{\gamma\gamma} <$ 180 GeV. A standard multivariate classifier trained on a mass-independent set of inputs in this way would produce a discriminator which is independent of the di-photon mass. Unfortunately this would also produce a discriminator which is independent of the di-photon mass

Kinematic
p_T^{lead}
$p_T^{sub-lead}$
η_{lead}
$\eta_{sub-lead}$
$\cos\Delta\phi$
Per-Event Resolution Estimate
σ_m^{right}
σ_m^{wrong}
p_{vtx}
Photon-Jet Discrimination
IDMVA_{lead}
$\mathrm{IDMVA}_{sub-lead}$

Table 2.6: Variables used as input to the di-photon MVA.

resolution. Since the training makes use by default only of the inclusive signal to background ratio over the full mass range included in the training. In order to make the classifier sensitive to the resolution, the signal events have been weighted during the training by the inverse of the relative resolution estimate. The correct and incorrect primary vertex selection hypotheses are both included, weighted by the vertex selection probability. The weight applied to the signal events during training is then,

$$w_{sig} = \frac{p_{vtx}}{\sigma_m^{right}/m_{\gamma\gamma}} + \frac{1 - p_{vtx}}{\sigma_m^{wrong}/m_{\gamma\gamma}} \tag{2.66}$$

The result of this weighting is that the signal to background ratio visible to the BDT training corresponds to the signal to background ratio in a simple cut and count analysis which approximates the full analysis of fitting a mass peak.

Consider a simple cut and count analysis performed in a mass window of $\pm n\sigma$ around the hypothesis mass, where the signal is assumed to be a Gaussian mass peak containing N_s events, and the background is assumed to have a flat distribution in mass with normalization of a [GeV^{-1}]. In this case the signal to background ratio in

the mass window is given by

$$s/b = \frac{\mathrm{erf}\frac{n}{\sqrt{2}}}{2n\sigma a} \tag{2.67}$$

This means that for a cut and count analysis which is defined to contain a constant fraction of signal events, if the resolution becomes narrower by a factor of two, then the mass window becomes narrower by a factor of two, the number of background events decreases by a factor of two, and the signal to background ratio increases by a factor of two. This effect can be emulated in the BDT training either by weighting the background events by a factor proportional to σ, or by weighting the signal events by a factor inversely proportional to σ. The latter is chosen because the background training sample is more statistically limited than the signal sample, and we avoid exacerbating this with additional weights. This hypothetical cut and count analysis represents a crude approximation to the final mass fit, however training the BDT against the signal to background ratio in this configuration is expected to yield close to the optimal usage of the per-event resolution.

There is a subtle point which has been glossed over in this discussion. The above argument assumes that the per-event resolution for each signal event is equal to the *average* resolution of the signal peak in the final mass fit. This is true if and only if the final analysis categories contain very little variation in per event resolution within each category. As this condition is not enforced by the BDT training, this is not strictly true, and the simple weighting scheme used is not completely optimal. An improved treatment of this issue is the subject of ongoing study.

2.8.3 Validation

In order to check the level of agreement between data and Monte Carlo, and also to see the level of mass dependence in the di-photon MVA inputs and output, we compare di-photon data and Monte Carlo simulation in two mass regions. First in the high mass (signal-free) control region $m_{\gamma\gamma} > 160$ GeV, and second in the signal region 100 GeV$< m_{\gamma\gamma} <$ 160 GeV, where the inclusive signal to background ratio over the full mass range is nevertheless expected to be very small. Distributions in

these plots have been normalized to area to facilitate comparison of the shapes.

First, simulation describes the data well in the control region. Second, the mass dependence of the variables is rather mild, comparing the MC in the two different mass regions. The most mass sensitive variables are the relative resolution and the photon id output, where most of the mass dependence comes from the changing composition of the background sample, with a decreasing number of fake photons from jets at higher di-photon masses.

In Fig. 2-19 the BDT classifier distribution is shown. On the left the MC background (black histogram) and the signal for $m_H = 125\,\text{GeV}$ (stacked coloured histograms), compared to the data, all in the inclusive mass range 100-180 GeV. On the right the MC background in the signal region (orange histogram) is again compare to both the MC (hashed histogram) and the data (black markers) in the control region. All histograms have been normalized to the same area in order to compare the shapes. The MC and data agreement in the control region is again excellent, and the mass dependence comes again mainly from the changing background composition. The BDT output in the two mass regions is shown for the prompt di-photon Monte Carlo sample only in Fig. 2-20 and shows a much smaller difference in shape between the two mass regions, demonstrating that most of the mass dependence indeed arises indirectly from changing background composition.

The output for the di-photon MVA output for 2012 Monte Carlo and data are shown in Figure 2-21 with comparisons between signal and background Monte Carlo, and between data and background Monte Carlo. The excess of data in very low score region (not used in the analysis) suggests an underestimation of the number of fake-fake events in the Monte Carlo. This is likely caused by a combination of the EM enrichment filters applied to the Monte Carlo, as well as imperfect kinematic description and normalisation from the leading order Monte Carlo sample.

2.8.4 Systematic Uncertainties on di-photon MVA Output

Because the Monte Carlo simulation, with relevant Monte Carlo to data correction factors applied, is used to model the predicted Higgs signal, uncertainties on the

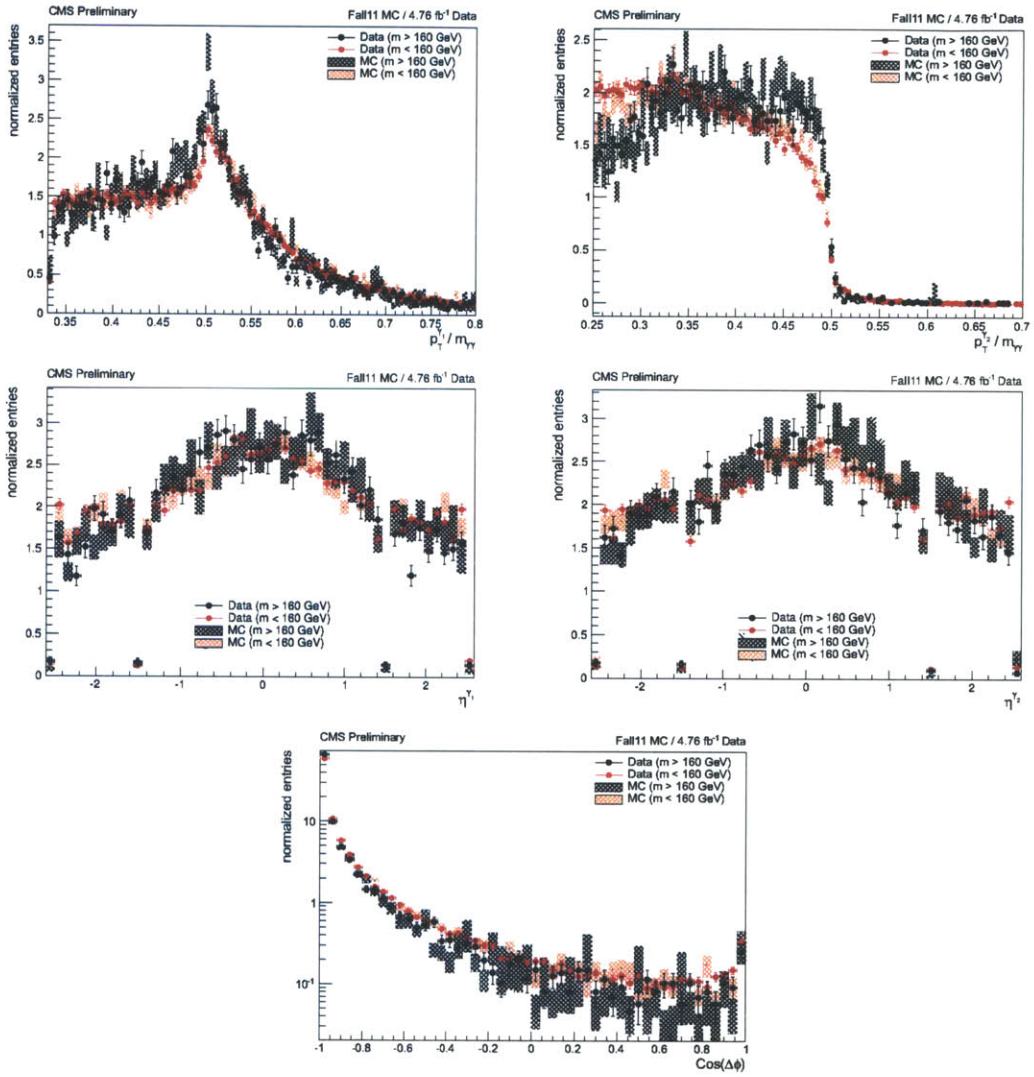


Figure 2-17: Comparison of the BDT input variables between Data (markers) and MC (hashed histogram) in the control region (black markers/histogram) and in the signal region (orange markers/histogram).

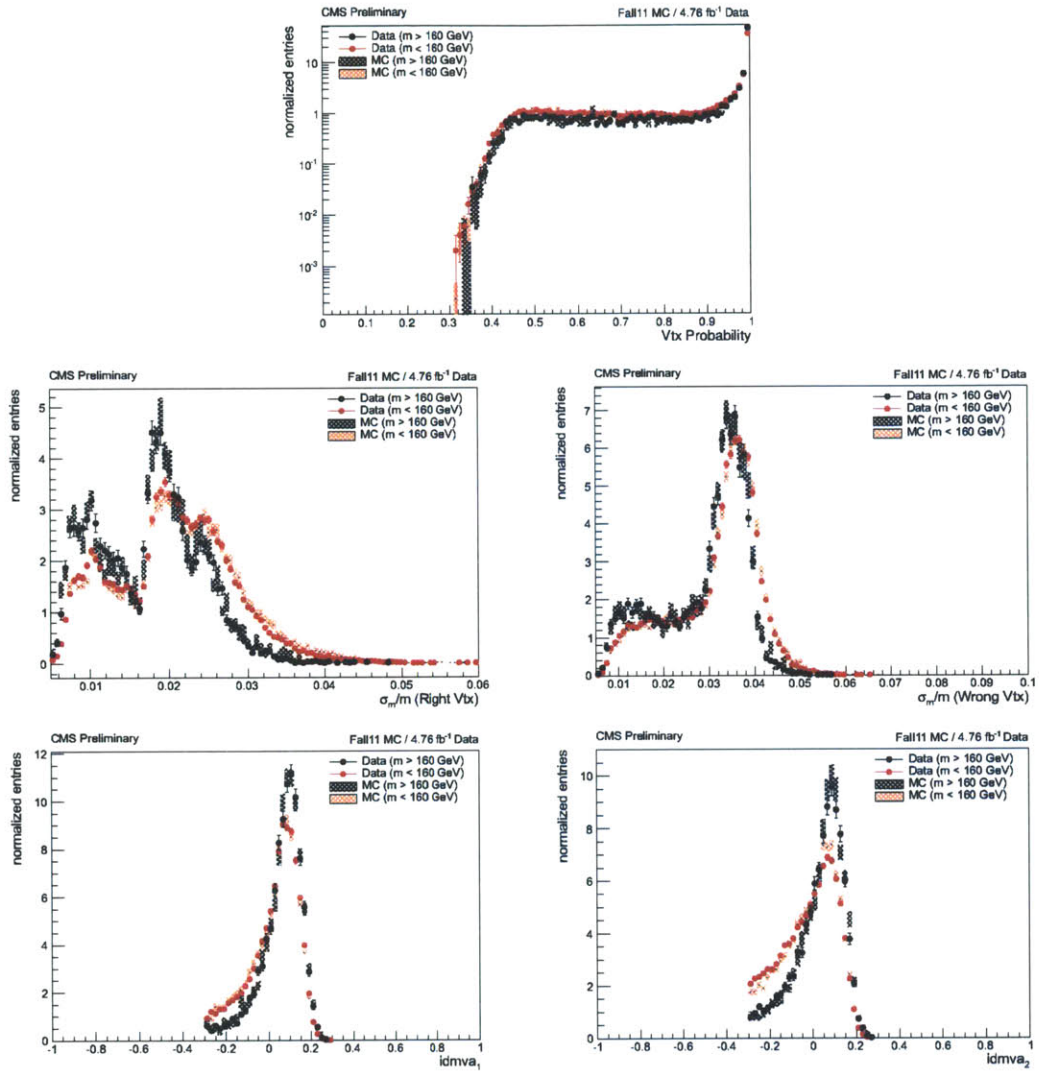


Figure 2-18: Comparison of the BDT input variables between Data (markers) and MC (hashed histogram) in the control region (black markers/histogram) and in the signal region (orange markers/histogram).

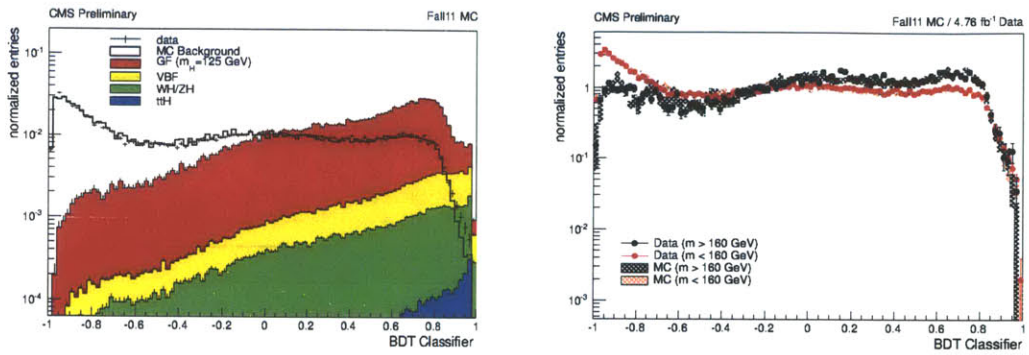


Figure 2-19: Left: Comparison of the di-photon MVA output between background Monte Carlo (black histogram) and signal Monte Carlo for $m_h = 125$ GeV for the four production mechanisms (filled histograms). Right: Comparison of the diphoton MVA output between Data (markers) and MC (hashed histogram) in the control region (black markers/histogram) and in the signal region (orange markers/histogram).

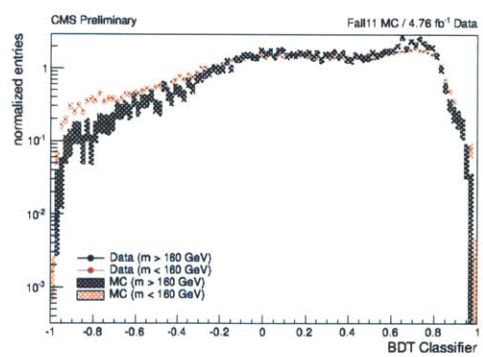


Figure 2-20: BDT output in the two mass regions for the prompt diphoton Monte Carlo only..

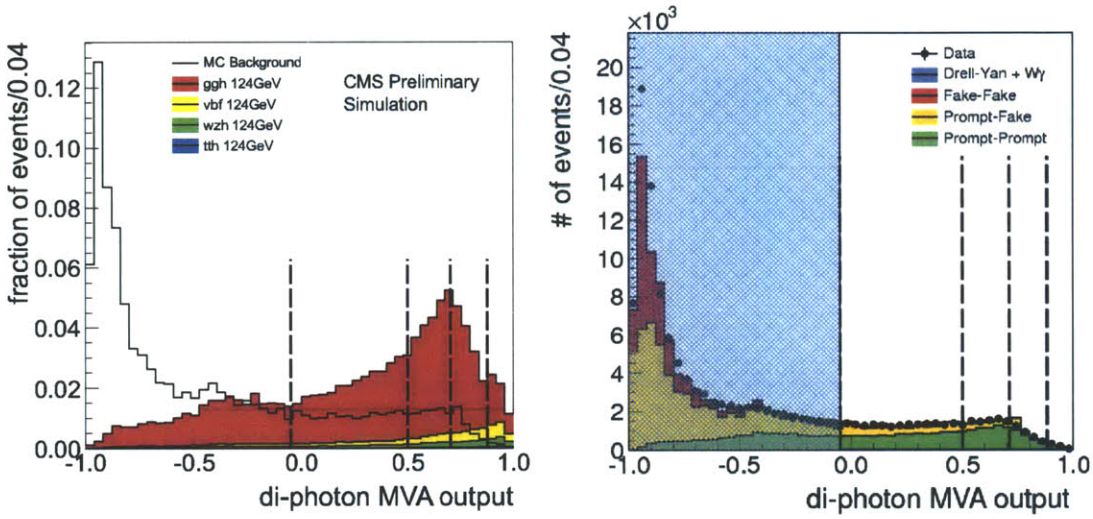


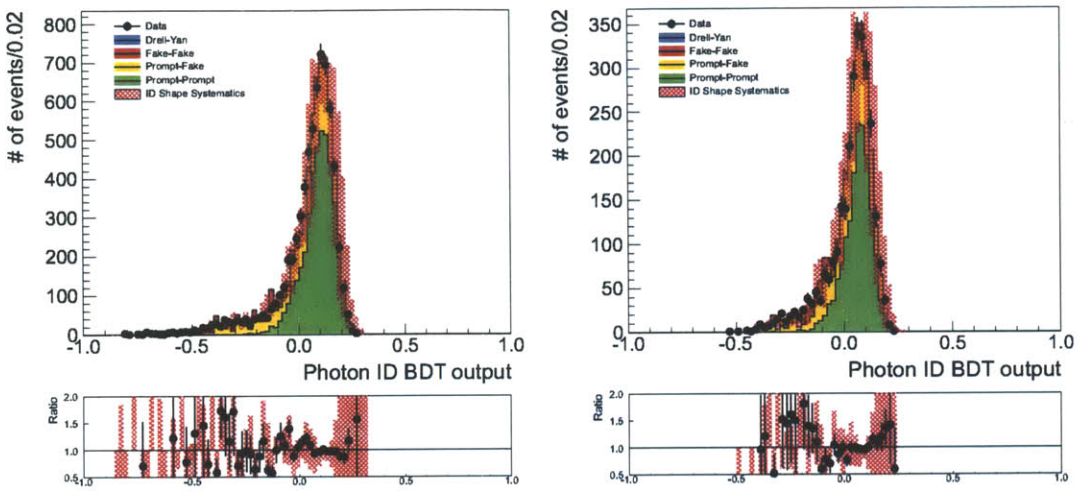
Figure 2-21: Left: Comparison of the di-photon MVA output between background Monte Carlo (black histogram) and signal Monte Carlo for $m_h = 125$ GeV at 8 TeV for the four production mechanisms (filled histograms). Right: Comparison of the diphoton MVA output between Data and background Monte Carlo for $100 < m_{\gamma\gamma} < 180$ GeV for 8 TeV. The excess of data in very low score region (not used in the analysis) suggests an underestimation of the number of fake-fake events in the Monte Carlo.

simulation prediction for the di-photon MVA input variables must be propagated as systematic uncertainties on the MVA output for the signal. Because the background will be modelled in a fully data-driven manner, the result does not rely on the Monte Carlo prediction for the di-photon MVA output shape of the various background components, and therefore no systematic uncertainty of this nature is needed for the background.

The two main sources of systematic uncertainty on the diphoton MVA output for the signal are the photon ID MVA, and the per-photon resolution estimate from the regression. Both are corrected (at the input level and at the output level respectively) for a nominal data$\rightarrow$ MC shift, but remaining imperfect modelling of the correlations between shower shape variables lead to discrepancies which must be covered by shape uncertainties.

The Photon ID MVA is a continuous-valued discriminator with output between -1.0 and +1.0, with the vast majority of prompt photons having an output between -0.2 and +0.3. The exact scale is arbitrary and depends on the details of the classifier training. The prescription which is followed to assign a systematic uncertainty from this output is to take every photon in the signal Monte Carlo and translate it's Photon ID output by ± 0.02 (± 0.01) fully correlated across all photons in the 7 TeV (8 TeV) sample. Since the diphoton MVA is in general monotonic with the Photon ID MVA output, this monotonic transformation of the ID MVA leads to a maximal possible migration of events in the diphoton MVA output, which is propagated as a migration of the signal yield among the final event classes. The Photon ID MVA output for the lead photon in diphoton events with $m_{\gamma\gamma} > 160$ GeV is shown in Fig. 2-22, along with the uncertainty bands representing the ± 0.02 shift of the output. The value of 0.02 (0.01) is chosen in order to cover the maximum observed discrepancies between data and simulation in high mass-diphoton as well as $z \rightarrow ee$ control samples. The search region lies somewhere in between in terms of transverse momentum, and therefore this variation should cover any possible event migration for the signal as well.

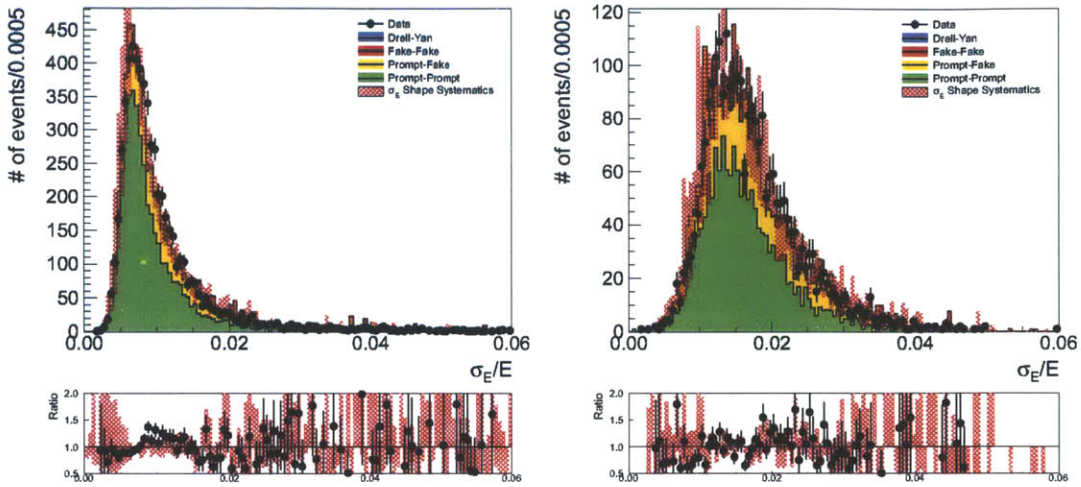
The per-photon resolution estimate is also affected by imperfect modelling of the electromagnetic shower shape in the Monte Carlo, even after correction with



(a) Barrel (b) Endcap

Figure 2-22: Distribution of photon ID MVA output for the leading photon in preselected diphoton events with $m_{\gamma\gamma}$ >160 GeV/c^2

the nominal Monte Carlo → data rescaling. This quantity enters into both right and wrong vertex hypotheses for the per-event mass resolution estimate. Data-MC discrepancies in this variable are diluted in the final mass resolution estimates by the sum in quadrature with the MC to data resolution smearing term. Nevertheless, the systematic uncertainty from modelling of σ_E is assigned by varying the MC to data rescaling factor by ± 0.10, again fully correlated for all photons in the sample, keeping in mind that this rescaling occurs prior to the sum in quadrature of the smearing term. The diphoton MVA output is again expected to be monotonic with the resolution estimate, such that this monotonic transformation produces a possible migration of events in the diphoton MVA output. The per-photon resolution estimate σ_E (again prior to the sum in quadrature of the smearing term) for the lead photon in diphoton events with $m_{\gamma\gamma} > 160$ GeV is shown in Fig. 2-23, along with the uncertainty bands representing the ± 0.10 uncertainty on the scaling factor. The effect of this discrepancy, as well as the shift corresponding to the systematic uncertainty are both reduced by the addition of the smearing term in quadrature, as well as the sum in quadrature of the resolution estimate from the two photons in the event.



(a) Barrel (b) Endcap

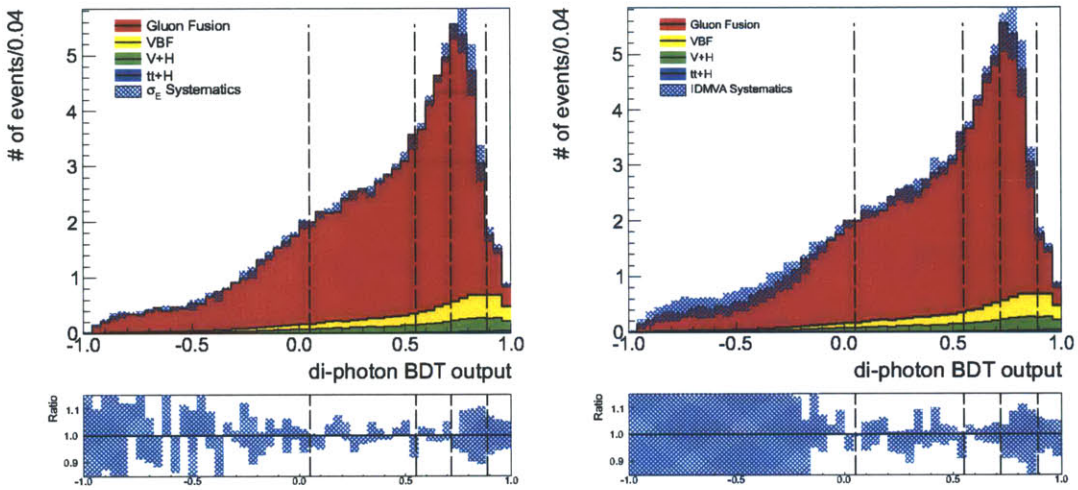
Figure 2-23: Distribution of per-photon resolution estimate σ_E/E (prior to the sum in quadrature with the smearing factor) for the leading photon in preselected diphoton events with $m_{\gamma\gamma}$ >160 GeV/c^2

The effect of these two shape systematics on the di-photon MVA output for a 125 GeV is shown in Fig. 2-24.

Although the Monte Carlo background is not used for the final result, the effect of applying the same shape systematics is shown in Fig. 2-25 for illustrative purposes. These plots show only the shape systematics which are in common applicable to both signal and background, there are additional significant uncertainties on the k-factors and background composition which are not shown here, and which are in part responsible for the lower simulation prediction in the very low score region.

2.9 Jet Reconstruction and Identification

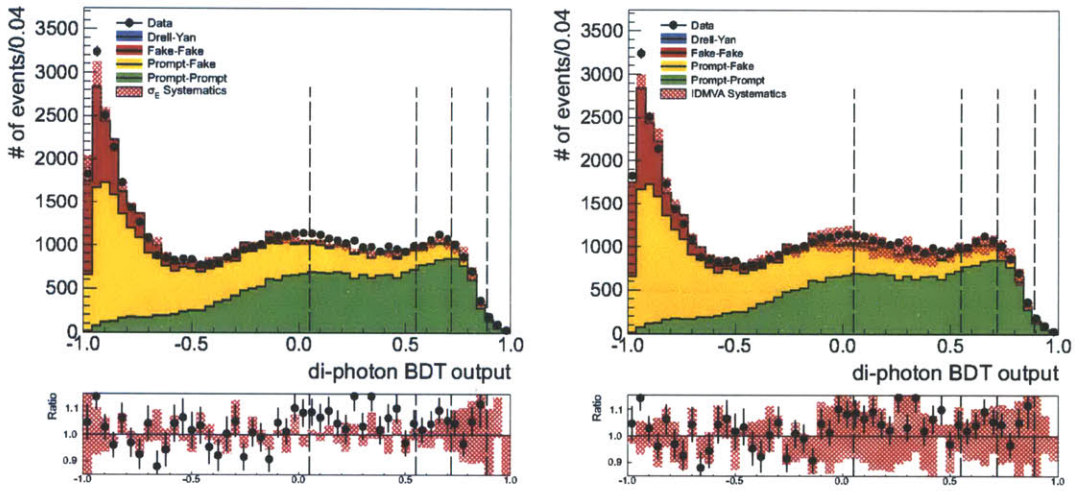
Jets are reconstructed by clustering the PF Candidates using the anti-k_T algorithm [67] with a clustering parameter of 0.5. The effect of additional energy from pileup interactions is corrected on average jet by jet and event by event using the median energy density ρ as computed by FastJet [66], with η-dependent scale factors for the



(a) σ_E Shape Systematic

(b) Photon ID MVA Shape Systematic

Figure 2-24: The effect of the shape systematics arising from the per-photon resolution estimate σ_E and the Photon ID MVA output on the final diphoton MVA output for 125 GeV Standard Model Higgs in Monte Carlo. The nominal MVA output is shown as the stacked histogram, and the variation due to the respective shape uncertainties are shown as the hashed band. The dotted lines correspond to the final event categories which are described in Section 2.11.



(a) σ_E Shape Systematic

(b) Photon ID MVA Shape Systematic

Figure 2-25: The effect of the shape systematics arising from the per-photon resolution estimate σ_E and the Photon ID MVA output on the background Monte Carlo in 100 GeV$< m_{\gamma\gamma} <$ 180 GeV, with selected data events overlayed. The nominal MVA output is shown as the stacked histogram, and the variation due to the respective shape uncertainties are shown as the hashed bands. These plots show only the shape systematics which are in common applicable to both signal and background, there are additional significant uncertainties on the k-factors and background composition which are not shown here. The dotted lines correspond to the final event categories which are described in Section 2.11.

correction calibrated in minimum bias events separately for data and simulation. The average jet energy scale as a function of η and p_T has been calibrated using Monte Carlo truth, and residual jet energy scale corrections for data have been derived using transverse momentum balancing in γ-jet and di-jet events. The four-vectors of the jets are computed as the four-vector sum of the constituent PF Candidates.

Due to the higher pileup conditions in 2012 data, an additional set of identification requirements are applied to the jets in order to distinguish between jets originating from the hard interaction and those originating from from pileup interactions as a result of random clustering of energy or the overlap of two or more softer jets. These requirements are based on a set of cuts on the jet shape and, within the tracking acceptance, the compatibility of charged jet constituents with the primary vertex.

2.10 Vector Boson Fusion Selection

The production of an SM Higgs through VBF production produces a subset of Higgs events with a unique topology consisting of two high energy jets in addition to the Higgs, produced with high pseudorapidity and in detected in opposite sides of the detector. By exploiting explicitly the jet kinematics in addition to the di-photon selection, a selection of events is defined with enhanced expected signal to background ratio, and which provides as well some discrimination between gluon fusion and VBF production of a SM Higgs. Requirements on the jets for this additional selection follow a simple set of kinematic cuts based on the two highest p_T jets with $|\eta| < 4.7$, based on the following variables:

- $p_T^{lead-jet}, p_T^{sub-lead-jet}$: The transverse momentum of the two leading jets.
- m_{jj}: Di-jet mass
- $|\Delta\eta(j,j)|$ Pseudorapidity gap between the two leading jets
- $Z = \eta_{\gamma\gamma} - \frac{1}{2}(\eta^{j1} + \eta^{j2})$: Zeppenfeld variable between the di-photon system and the jets

$p_T^{\gamma 1}/m_{\gamma\gamma}$	$> 55/120$
$p_T^{\gamma 2}/m_{\gamma\gamma}$	$> 1/4$
p_T^{j1}	> 30 GeV
p_T^{j2}	> 20 GeV
$\lvert\Delta\eta(j,j)\rvert$	> 3.5
$\lvert Z\rvert$	< 2.5
m_{jj}	> 350 GeV
$\lvert\Delta\phi(jj,\gamma\gamma)\rvert$	> 2.6

Table 2.7: VBF-tag selection for the 7 TeV analysis.

	Loose	Tight
$p_T^{\gamma 1}/m_{\gamma\gamma}$	$> 1/2$	$> 1/2$
$p_T^{\gamma 2}/m_{\gamma\gamma}$	$> 1/4$	$> 1/4$
p_T^{j1}	> 30 GeV	> 30 GeV
p_T^{j2}	> 20 GeV	> 30 GeV
$\lvert\Delta\eta(j,j)\rvert$	> 3.0	> 3.0
$\lvert Z\rvert$	< 2.5	< 2.5
m_{jj}	> 250 GeV	> 500 GeV
$\lvert\Delta\phi(jj,\gamma\gamma)\rvert$	> 2.6	> 2.6

Table 2.8: VBF-tag selection for the 8 TeV analysis.

- $|\Delta\phi(jj,\gamma\gamma)|$: The angle in the transverse plane between the di-jet system and the di-photon system

The final VBF-tag selection includes also additional cuts on the relative photon transverse momenta. For the 7 TeV analysis, a single VBF-tag selection was defined, based on the selection in Table 2.7.

For the 8 TeV analysis, in order to further optimize the sensitivity and discrimination between gluon fusion and VBF Higgs production, the VBF-tag selection was divided into two exclusive categories with different signal to background and gluon fusion to VBF composition. The selection details for the tight and loose VBF-tag categories in the 8 TeV analysis are given in Table 2.8.

2.11 Event Categorization

The events entering the analysis have been split into mutually exclusive categories according to the di-photon MVA output and the VBF-tag selection. VBF-tag events form their own event categories (a single category for 7 TeV and two categories for 8 TeV according to the loose and tight VBF-tag selections), the remaining untagged events are further divided into categories using the di-photon MVA output. The MVA values corresponding to the untagged category boundaries have been optimized using the signal ($m_H = 124$ GeV) and background Monte Carlo samples. The figures of merit for the optimization were the expected 95% exclusion limit as computed using the Asymptotic CLs procedure described in Section 3.3, as well as the expected discovery significance for a SM Higgs, with the two criteria having been verified to yield the same optimal points. The optimization was done iteratively by sub-dividing the untagged events into categories, with the split point in the di-photon MVA output scanned across the full range from -1.0 to 1.0. This procedure was terminated after splitting into 5 categories, as further splitting into 6 categories was found to lead to negligible ($< 1\%$) additional improvement in the expected limit or significance. Finally, the fifth category was dropped, with negligible loss of estimated sensitivity. This region of phase space corresponds to the highest background region with the lowest signal to background ratio, consisting of events with photons in the endcap, badly showering in the tracker, or with larger amounts of energy in their isolation cones, as well as actual fake photons from jets. The VBF-tag categories are not further subdivided according to the di-photon MVA, however a lower cut is placed on the di-photon MVA for these categories, in order to match the region of phase space which is discarded from the untagged categories. The definition of categories for the 7 TeV and 8 TeV analyses are given in Tables 2.9 and 2.10.

This represents a relatively simple scheme for combining information about the di-jet system with that of the di-photon system. Some gains in sensitivity as well as separation of production mechanisms might be possible with a more sophisticated treatment, especially with more data.

Category	Di-photon MVA Requirements	VBF-tag Requirements
Untagged 0	MVA$\geq$ 0.89	!VBF-tag
Untagged 1	0.74 $\leq$MVA$<$ 0.89	!VBF-tag
Untagged 2	0.55 $\leq$MVA$<$ 0.74	!VBF-tag
Untagged 3	0.05 $\leq$MVA$<$ 0.55	!VBF-tag
VBF-tag	MVA $\geq$ 0.05	VBF-tag

Table 2.9: Event catogory definitions for the 7 TeV analysis. In the above, "MVA" refers to the value of the di-photon MVA output, and VBF-tag refers to whether or not the event passes the VBF-tag selection for the 7 TeV analysis described in 2.10

Category	Di-photon MVA Requirements	VBF-tag Requirements
Untagged 0	MVA$\geq$ 0.88	!VBF-loose
Untagged 1	0.71 $\leq$MVA$<$ 0.88	!VBF-loose
Untagged 2	0.50 $\leq$MVA$<$ 0.71	!VBF-loose
Untagged 3	-0.05 $\leq$MVA$<$ 0.50	!VBF-loose
Tight VBF-tag	MVA ≥ -0.05	VBF-tight
Loose VBF-tag	MVA ≥ -0.05	VBF-loose && !VBF-tight

Table 2.10: Event category definitions for the 2012 analysis. In the above, "MVA" refers to the value of the di-photon MVA output, and VBF-loose/tight refer to whether or not the event passes the VBF-tag selections for the 8 TeV analysis described in 2.10

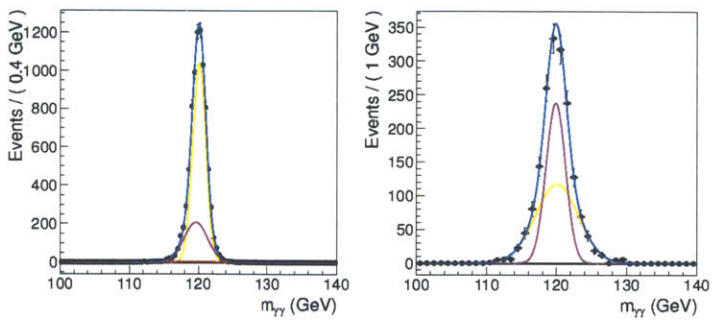
Chapter 3

Interpretation and Results

The overall strategy for interpretation of the data and to extract a result is to perform maximum likelihood fits to the di-photon mass distribution simultaneously across all of the event categories. This requires models for the probability density functions (PDFs) of both the expected signal, as well as the background in each category. A statistical procedure must also be defined in order to interpret the results of these fits, including the treatment of the systematic uncertainties.

3.1 Signal Model

In order to statistically interpret the observed data, it is necessary to have a description of the signal which specifies the overall efficiency times acceptance, as well as the shape of the di-photon mass distribution in each of the event categories. The Monte Carlo simulation is used, after the smearing of the resolution, and the application of all efficiency corrections and scale factors, to build a parameterized model for the signal which is defined continuously for any value of the Higgs mass between 110 GeV and 150 GeV. The shape model is to derived from the signal Monte Carlo using an analytic function. The parameters of this function are determined by fitting the Monte Carlo samples for each available mass point. Finally, the full signal model is defined by a linear interpolation of each fit parameter between the fitted mass values. As input to the fits, we use individually the signal Monte Carlo samples for each mass



(a) Correct Vertex Selection (b) Incorrect Vertex Selection

Figure 3-1: Fit results for signal shape model with correct and incorrect primary vertex selection for a single class in 120 GeV Gluon-Fusion Higgs Monte Carlo. The black points are the weighted Monte Carlo events and the blue lines are the corresponding fit results. Individual Gaussian components of the fits are also shown.

point for each of the four production mechanisms gluon-fusion, vector boson fusion, W/Z associated production, and $t\bar{t}$+Higgs associated production. The analytic functions for each production mechanism are added together at the end according to their relative cross-sections in the standard model.

The signal shape for events with the correct primary vertex selection is dominated by the detector and reconstruction response in the Ecal. The signal shape for these events is modeled empirically in each class by a sum of 2 or 3 Gaussians, depending on the class. The means, widths, and relative fractions of the Gaussians are left free in the fit to the Monte Carlo.

The signal peak for events with incorrect primary vertex selection is smeared significantly by the variation in the z position of the selected primary vertex with respect to the true Higgs production point. The signal shape for these events is modelled in each class by a sum of 2 Gaussians, or a single Gaussian, depending on the class. The means, widths, and relative fraction for classes with two Gaussians, are left free in the fits to the Monte Carlo. A representative set of fits for events with correct and incorrect primary vertex selection, in one class for gluon-fusion production are shown in Figure 3-1.

The combined shape in each class for correct and incorrect vertex selection is

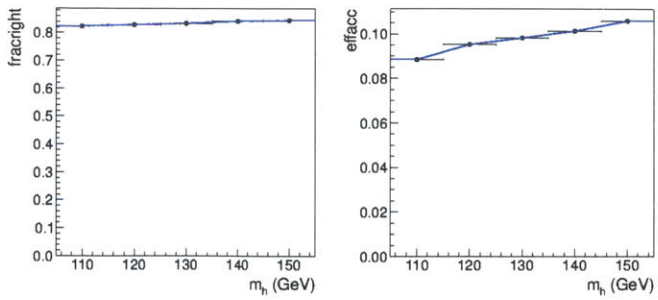
constructed by adding the shapes for the two sub-components together, according to the correct vertex selection efficiency determined from each Monte Carlo sample. This efficiency is treated as another model parameter for the purposes of interpolation between mass points, although in practice the vertex selection efficiency does not vary much as a function of Higgs mass.

In order to facilitate the interpretation of the signal model in terms of a standard model Higgs production cross-section, and in order to facilitate the use of the signal model simultaneously across all of the event classes, we parametrise the signal yield in terms of a per class acceptance times efficiency, computed from each Monte Carlo sample. The evolution of the correct vertex fraction, as well as the per-category acceptance times efficiency with Higgs mass are shown for one category, for gluon-fusion production in Figure 3-2.

3.1.1 Final Signal Models

The final parametrised shapes for each class for a Higgs mass of 120 GeV for the Standard Model cross-section weighted mixture of all production mechanisms is shown in Figures 3-3 and 3-4 for 7 TeV and 8 TeV respectively. The parametrised shape corresponding to the combination of all event categories for 7 TeV and 8 TeV is shown in Figure 3-5. The better mass resolution for 7 TeV as compared to 8 TeV is a consequence of the near-final Ecal calibration constants used for the 7 TeV data, as compared to the Prompt Reconstruction constants used for the 8 TeV data, leading to smaller required Monte Carlo to data energy resolution smearing. There is a smaller contribution as well from the higher pileup in the 8 TeV sample, leading to a few percent reduction in the correct primary vertex selection efficiency.

The determination of the full set of signal model parameters at each Monte Carlo mass point is used to construct a signal model continuous in Higgs mass by performing a linear interpolation of each fit parameter, which gives rise to a smooth evolution of the signal shape. A closure test of this procedure is shown in Figure 3-6 where the 115 GeV Monte Carlo is shown along with the interpolated shape in one of the classes for for gluon-fusion production. The agreement between the interpolated



(a) Vertex Selection Efficiency (b) Acceptance×Efficiency

Figure 3-2: Correct vertex selection efficiency and Acceptance×Efficiency for one class for the fitted masses of 110, 120, 130, 140 GeV along with the linear interpolations between mass points.

model and the direct Monte Carlo prediction demonstrates that the interpolation procedure yields a reliable signal prediction for intermediate masses.

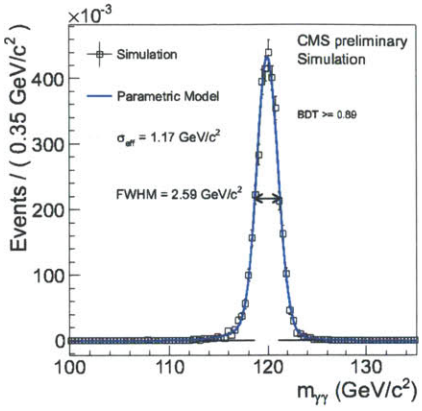
3.1.2 Systematic Uncertainties

Systematic uncertainties affecting the shape are incorporated as parametric variations of the model. Uncertainty in the vertex selection efficiency are treated by varying the relative additive fraction of the right and wrong vertex shapes. Uncertainty in the energy scale is incorporated as a shift in the mean of each Gaussian, and uncertainty on the resolution is incorporated by the analytic convolution or deconvolution of an additional width with each of the Gaussians.

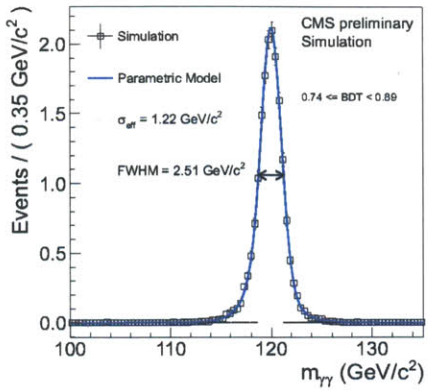
Systematic uncertainties which affect the di-photon BDT output are incorporated as category migration systematics, through correlated log-normal uncertainties on the category yields, as described in more detail in Section 3.3.

3.2 Background Model

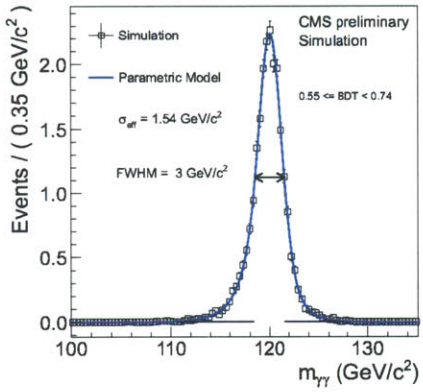
Since the level of background after selection is large, and a comprehensive Monte Carlo description is not available at NLO or beyond, the background is instead modelled in



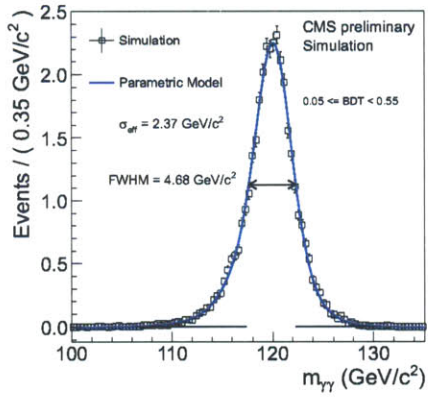
(a) Untagged 0



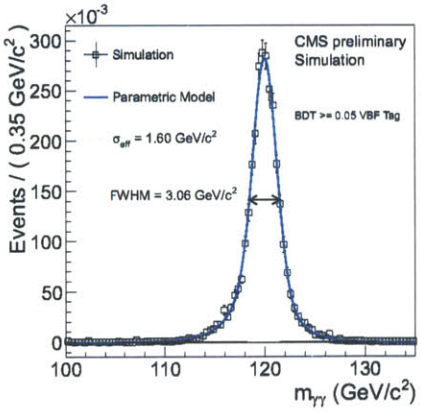
(b) Untagged 1



(c) Untagged 2

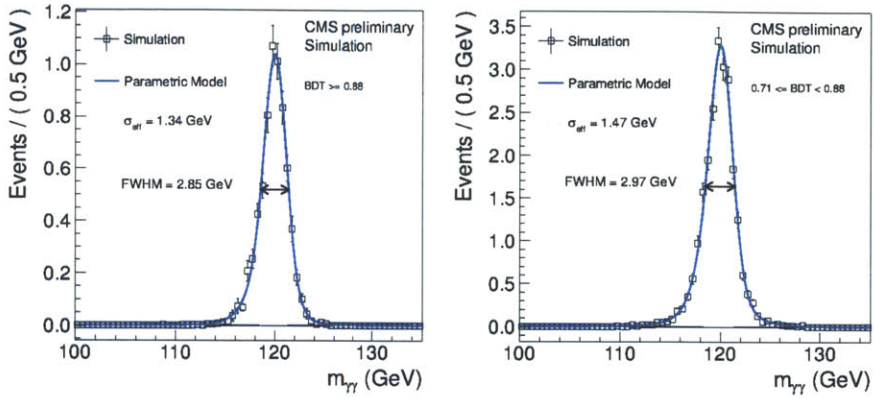


(d) Untagged 3



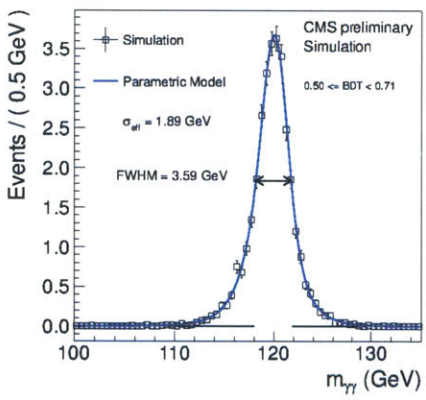
(e) VBF-tag

Figure 3-3: Full parameterized signal shape for the 5 event categories for the 7 TeV analysis in 120 GeV Higgs Monte Carlo. The black points are the weighted Monte Carlo events and the blue lines are the corresponding models. Also shown are the effective σ values and corresponding intervals.

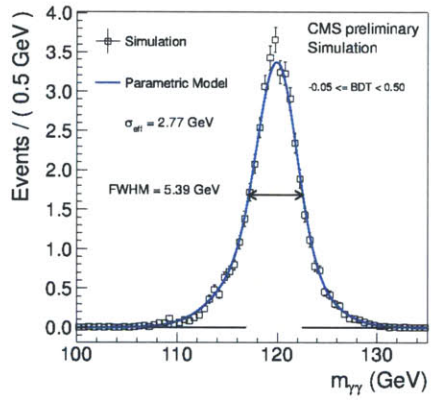


(a) Untagged 0

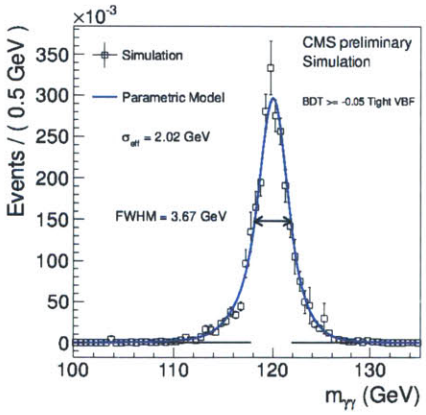
(b) Untagged 1



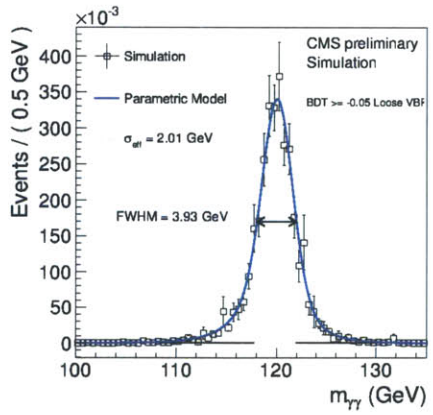
(c) Untagged 2



(d) Untagged 3

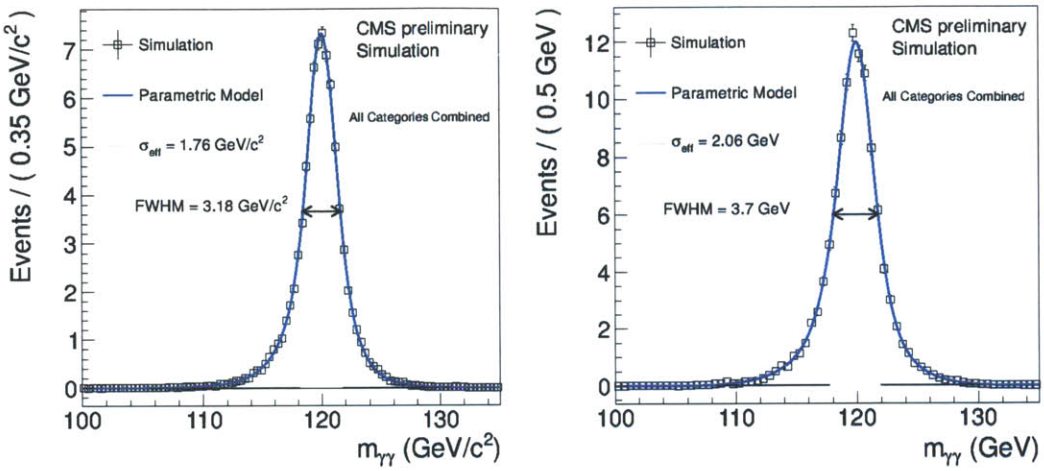


(e) Tight VBF-tag



(f) Loose VBF-tag

Figure 3-4: Full parameterized signal shape for the 6 event categories for the 8 TeV analysis in 120 GeV Higgs Monte Carlo. The black points are the weighted Monte Carlo events and the blue lines are the corresponding models. Also shown are the effective σ values and corresponding intervals.



(a) 7 TeV (b) 8 TeV

Figure 3-5: Full parameterized signal shape integrated over all event classes in 120 GeV Higgs Monte Carlo. The black points are the weighted Monte Carlo events and the blue lines are the corresponding model. Also shown is the effective σ value and corresponding interval.

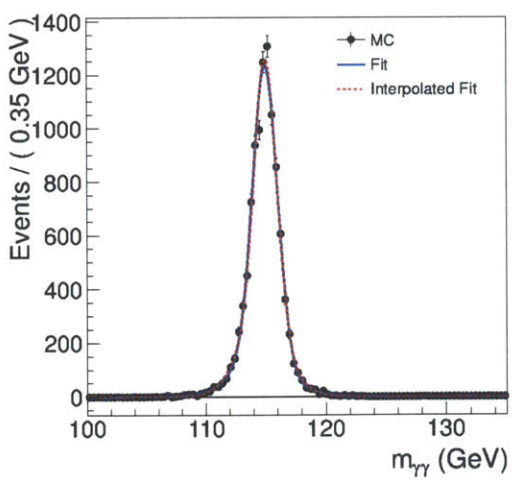


Figure 3-6: Combined signal shape model for correct and incorrect primary vertex selection for one class in 115 GeV gluon fusion Higgs Monte Carlo. The black points are the weighted Monte Carlo events and the blue lines are the corresponding fit models. The red dashed lines are the signal models interpolated from the 110 GeV and 120 GeV fits.

an entirely data driven manner. The general physical constraint that the background be smooth and continuous in the di-photon mass distribution, combined with the narrow width of the di-photon mass peak for the expected signal allows for in-situ modelling of the background, exploiting the signal-free region farther from the hypothesised Higgs mass. The background shape is modelled parametrically with freely floating shape and normalisation parameters. Since there is no specific prediction for the functional form of the background, and since selection efficiencies and subdivision into categories can introduce subtle and non-trivial turn-on effects, we choose a functional form which is flexible enough to cover a wide range of possible underlying shapes. In order to allow for sufficient sidebands across the entire search range, and taking into account the possibility for larger background uncertainties or un-physical behaviour near the boundaries of the fit range, likelihood fits, and consequently the background model are defined for the mass range $100 < m_{\gamma\gamma} < 180$ GeV.

For each event category, a number of possible functional forms have been fit to the data and used as the background pdf for toy Monte Carlo studies. These include:

- Sum of N exponential terms: $\exp_N(m) = \sum_{i=1}^{N} f_i e^{k_i m}$
- Sum of N power terms: $\mathrm{pow}_N(m) = \sum_{i=1}^{N} f_i m^{-k_i}$
- Nth order Bernstein polynomial: $\mathrm{B}_N(m) = \sum_{i=1}^{N} f_i^2 \binom{N}{i} x^i (1-x)^{N-i}$, $x \equiv \frac{m-100}{80}$
- three-term Laurent Series: $L_3 = f_1 m^{-3} + f_2 m^{-4} + f_3 m^{-5}$
- four-term Laurent Series: $L_4 = f_1 m^{-3} + f_2 m^{-4} + f_3 m^{-5} + f_3 m^{-6}$

The Bernstein basis is chosen for the polynomial form here because the function can be constrained to be positive over the entire fit range simply by rescaling the input variable to lie between 0 and 1, and by constraining all of the coefficients to be non-negative. This avoids stability problems in fits to data and toy Monte Carlo by avoiding likelihood evaluation errors due to negative PDF values. Rather than directly constrain the polynomial coefficients, the fit parameters are instead squared in order to ensure they are non-negative. This leads to better numerical behaviour in MINUIT[68].

These functional forms cover a range of possible shapes for the background. For each event category, each functional form is fit to the data, with the sufficient number of terms being determined by an f-test, where terms are only included if they significantly improve the χ^2 probability of the fit. The resulting background PDFs are used to generate toy Monte Carlo. For each toy Monte Carlo experiment, the signal yield is extracted with a maximum likelihood fit using a variety of different background PDFs for the fit, and the bias with respect to the injected signal is assessed. Given that possible signals are only considered for 110 GeV$< m_H <$ 160 GeV, the bias is only evaluated for injected signals in this range. Because it would not be straightforward to introduce additional systematic uncertainties to account for any remaining bias, we require that this residual bias is at least five times smaller compared to the statistical uncertainty on the background estimate, such that it can be neglected. The background estimate may be biased or possess un-physical behaviour near the edges of the fit range, but this does not have any effect on the results, because it is sufficiently far from any signal under consideration. The only functional form for the background fit among those tested that reproduces a sufficiently small bias for the full range of generating PDFs is the polynomial. The order of Bernstein polynomial chosen for the fits is the lowest order resulting in a sufficiently small bias against all of the functional forms, where empirically the limiting factor is in all cases the toy Monte Carlo in which a power series was used as the generating PDF. This full set of toy Monte Carlo studies was carried out for the 7 TeV analysis in order to establish this qualitative behaviour, such that for the 8 TeV analysis only the toy Monte Carlo with a power series as the generating function and polynomials as the fitting function were repeated, in order to determine the necessary order of polynomial for the fits. The resulting Bernstein polynomial orders used for the background fits in each event category are listed in Table 3.1.

Category	Bernstein Polynomial Order
Untagged 0 (2011)	4
Untagged 1 (2011)	5
Untagged 2 (2011)	5
Untagged 3 (2011)	5
VBF-tag (2011)	4
Untagged 0 (2012)	4
Untagged 1 (2012)	5
Untagged 2 (2012)	5
Untagged 3 (2012)	5
Tight VBF-tag (2012)	3
Loose VBF-tag (2012)	4

Table 3.1: The order of Bernstein polynomial used for the background model in each event category.

3.3 Statistical Procedure

The statistical procedure used to interpret the data are based on the jointly agreed details of the ATLAS and CMS collaborations described in [69]. Results are extracted based on unbinned maximum likelihood fits to the data under various background-only and signal plus background hypotheses. For results combining multiple event categories, likelihood fits are performed simultaneously across all event categories, such that the overall log likelihood is the simple sum over all categories.

The signal PDF is that described above, where the shape and normalization for each production method are separately tracked and defined. Penalty terms are added to the likelihood for each uncorrelated source of systematic uncertainty and the corresponding nuisance parameter x_i. Nuisance parameters which affect the signal yield for various production mechanisms and categories parametrize variations of the yield following log-normal distributions. The value of the nuisance parameter itself is distributed internally as a Gaussian with mean of zero and sigma of one. The effect on the signal yield for each production mechanism and category is either null, defined by a symmetric log-normal distribution with parameter κ_{ijk}, or defined by an asymmetric log-normal distribution with parameters κ_{ijk}^{up}, κ_{ijk}^{down}. The first subscripts

on these parameters enumerate the nuisance parameters, the second, the production mechanism, and the third the event category. For a symmetric log-normal variation, the nominal signal yield for a given mechanism and event category is multiplied by a factor

$$f_{ijk}(x_i) = e^{\left(\ln \kappa_{ijk}\right)x_i} \tag{3.1}$$

. For x_i distributed as a Gaussian with mean zero and sigma one, then $f_{ijk}(x_i)$ is distributed as the corresponding log normal distribution with mean of zero and sigma of $|\ln \kappa_{ijk}|$, with probability density

$$p\left[f_{ijk}(x_i)\right] = \frac{1}{f_{ijk}(x_i)|\ln \kappa_{ijk}|\sqrt{2\pi}} e^{-\frac{\left(\ln f_{ijk}(x_i)\right)^2}{2|\ln \kappa_{ijk}|^2}} \tag{3.2}$$

.

For an asymmetric log-normal variation, the nominal signal yield for a given mechanism and event category is multiplied by a factor

$$f_{ijk}(x_i) = \begin{cases} e^{\left(\ln \kappa_{ijk}^{down}\right)x^i} & , x \leq -0.5 \\ e^{\left\{\frac{1}{2}(\ln \kappa_{ijk}^{up} - \ln \kappa_{ijk}^{down}) + \frac{1}{8}x_i\left[4x_i^2(12x_i^2-10)+15\right](\ln \kappa_{ijk}^{up} + \ln \kappa_{ijk}^{down})\right\}x_i} & , -0.5 < x < 0.5 \\ e^{\left(\ln \kappa_{ijk}^{up}\right)x^i} & , x \geq 0.5 \end{cases} \tag{3.3}$$

such that $f_{ijk}(x_i)$ is log-normally distributed with sigma of $|\ln \kappa_{ijk}^{up}|$ for $x \geq 0.5$, log-normally distributed with sigma of $|\ln \kappa_{ijk}^{down}|$ for $x \leq -0.5$ and a smooth and continuous interpolation of the two distributions in the intermediate region $-0.5 < x < 0.5$.

Nuisance parameters which affect the signal shape have a Gaussian constraint term added to the likelihood with mean μ_i and width σ_i. The nuisance parameters associated with the background normalization and shape are constrained entirely from the data and therefore have no additional penalty terms added to the likelihood.

The complete negative log likelihood is then

$$-\ln\mathcal{L} = \sum_{k=1}^{N_{\text{categories}}} \left\{ -\sum_{ievt=1}^{N_{\text{events}}} \ln p_k(m_{\gamma\gamma}|\bar{y},\bar{x}) + \mathcal{N}_k(\bar{y},\bar{x}) - N_k \ln \mathcal{N}_k(\bar{y},\bar{x}) \right\} - \sum_{i=1}^{N_{\text{sig. nuisances}}} \ln \text{Gaussian}(x_i|\mu_i,\sigma_i) \tag{3.4}$$

, where $\bar{x}$ are the nuisance parameters associated with systematic uncertainties for the signal and background normalization and shape, $\bar{y}$ are the physical parameters being tested (such as mass and coupling of the Higgs boson), $\mathcal{N}_k(\bar{y},\bar{x})$ and N_k are the total number of fitted and observed events in each category, and $p_k(m_{\gamma\gamma}|\bar{y},\bar{x})$ is the normalized signal plus background probability density for each category. The fitted number of events in each category is

$$\mathcal{N}_k(\bar{y},\bar{x}) = \sum_{j=1}^{N_{\text{prod. mech.}}} \mathcal{N}_{jk}^{sig}(\bar{y},\bar{x}) + \mathcal{N}_k^{bkg}(\bar{x}) \tag{3.5}$$

with $\mathcal{N}_{jk}^{sig}$ representing the number of signal events for each of the four Higgs production mechanisms in each category, and $\mathcal{N}_k^{bkg}$ representing the number of background events in each category. Each $\mathcal{N}_k^{bkg}$ is associated with its own corresponding freely floating nuisance parameter, such that the number of background events in each category is constrained only by the data. The number of signal events is

$$\mathcal{N}_{jk}^{sig} = \prod_{i=1}^{N_{\text{sig. nuisances}}} [f_{ijk}(x_i)]\, \epsilon_{jk} A_{jk} L \sigma_j(\bar{y}) \tag{3.6}$$

where $f_{ijk}(x_i)$ are the variations of the signal yield due to systematic uncertainties described in 3.1 and 3.3, $\epsilon_{jk}A_{jk}$ is the nominal category-exclusive acceptance times efficiency for the given production mechanism and category, and $\sigma_j(\bar{y})$ is the nominal signal cross section for the given category, given physical model parameters $\bar{y}$. The

per-category probability density is

$$p_k(m_{\gamma\gamma}|\bar{y},\bar{x}) = \frac{1}{\mathcal{N}_k(\bar{y},\bar{x})}\left[\sum_{j=1}^{N_{\text{prod. mech.}}} \mathcal{N}_{jk}^{sig}(\bar{y},\bar{x})p_{jk}^{sig}(m_{\gamma\gamma}|\bar{y},\bar{x}) + \mathcal{N}_k^{bkg}(\bar{x})p_k^{bkg}(m_{\gamma\gamma}|\bar{x})\right] \tag{3.7}$$

For most of the results discussed here, the physical parameters $\bar{y}$ consist of a single overall scaling factor for the standard model Higgs production cross sections, μ, and the Higgs mass m_H.

Best-fit values for quantities such as cross section ratios and mass are determined by the value which maximizes the likelihood, as determined using MINUIT [68]. Uncertainties are determined using the MINOS procedure in MINUIT, with 68% confidence interval bounds defined by the parameter values, with all other parameters profiled, which give $\Delta \log \mathcal{L} = 1/2$ with respect to the global minimum.

The statistical significance for the observation of an excess to be due to a signal rather than an upward fluctuation of the background is determined using the profile likelihood method. Two maximum likelihood fits are performed to the data, one for the background-only hypothesis, such that the signal strength μ is set to zero, and the second for the signal plus background hypothesis, with μ freely floating ($\mu \geq 0$). The signal plus background fit is carried out for each value of hypothesis mass m_H in 0.5 GeV steps for 110 GeV$\leq m_H \leq$ 150 GeV. For each hypothesis mass, the local statistical significance is computed from the profiled log likelihood ratio

$$q = -\ln \mathcal{L}(\hat{\mu},\hat{x},m_H) + \ln \mathcal{L}(\mu = 0,\hat{x}) \tag{3.8}$$

Since the signal plus background fit contains, at a given hypothesis mass, exactly one degree of freedom more than the background-only fit, the test statistic q is asymptotically expected to be distributed as a χ^2 distribution with one degree of freedom in case the underlying data is the result of background only. In this case, the local significance is given simply by the asymptotic approximation $s = \sqrt{2q}$.

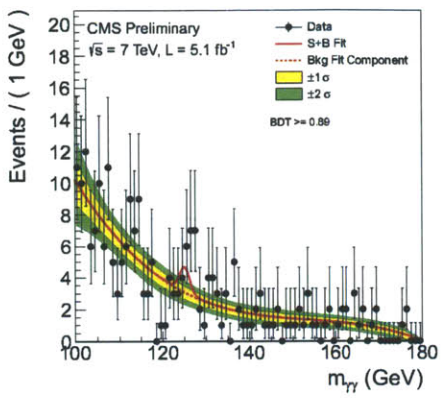
3.4 Results

The di-photon mass distributions for each event category after the final selection are shown in Figures 3-7 and 3-8 for 7 TeV and 8 TeV data respectively. Also shown are the results of the full simultaneous signal plus background fit over all 11 event categories with fully floating overall cross section multiplier μ and hypothesis mass m_H. The best fit cross section multiplier is $1.56^{+0.45}_{-0.41}$, with the global minimum occurring for a hypothesis mass $m_H = 125.1$ GeV. Since these uncertainties are obtained including profiling over the signal nuisances parameters, they include both statistical and systematic contributions, with the statistical contribution being the dominant for the present data set.

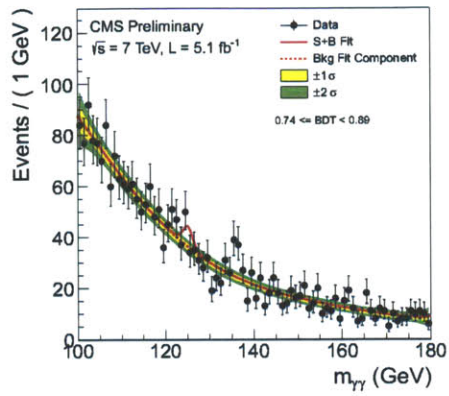
The di-photon mass distribution for the sum of all 11 event categories is shown in Figure 3-9, along with the total signal plus background and background components from the combined fit. Although this distribution includes all of the data entering the analysis, it is not optimal for interpreting the results, because categories with high signal to background ratio and good mass resolution, but fewer events, are overwhelmed by the inclusion of a large number of events from categories with worse signal purity and resolution.

The 95% confidence level exclusion limit, expressed in terms of the cross section multiplier μ as a function of hypothesis mass m_H is shown in Figure 3-10. Exclusion limits computed separately for the 7 TeV and 8 TeV data are shown in Figure 3-11. The median expected exclusion limit for all data combined is below the standard model cross section over most of the search range. The observed results exclude a standard model Higgs in several disconnected regions of Higgs boson mass, but most notably not in the vicinity of the large excess around 125 GeV.

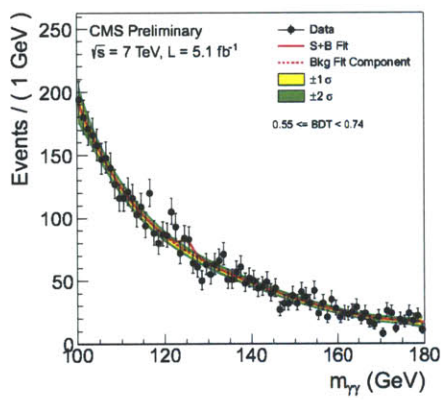
The search results, expressed in terms of the probability for the background alone to fluctuate to give a result as signal-like as that observed in data is shown in Figure 3-12, where the probability corresponds to the local p-value in the asymptotic limit. The most significant excess, evaluated in 0.5 GeV steps, occurs for $m_H = 125$ GeV, with a local significance of 4.1 standard deviations. The expected local significance



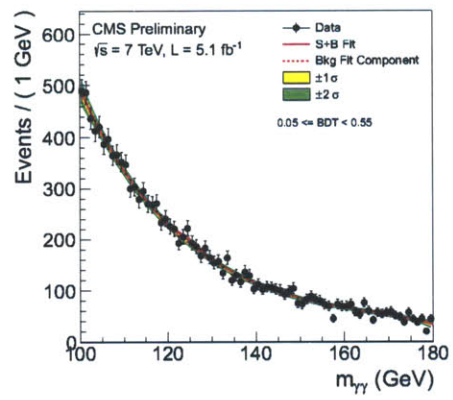
(a) Untagged 0



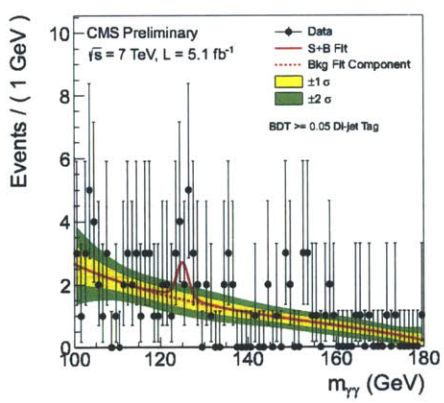
(b) Untagged 1



(c) Untagged 2

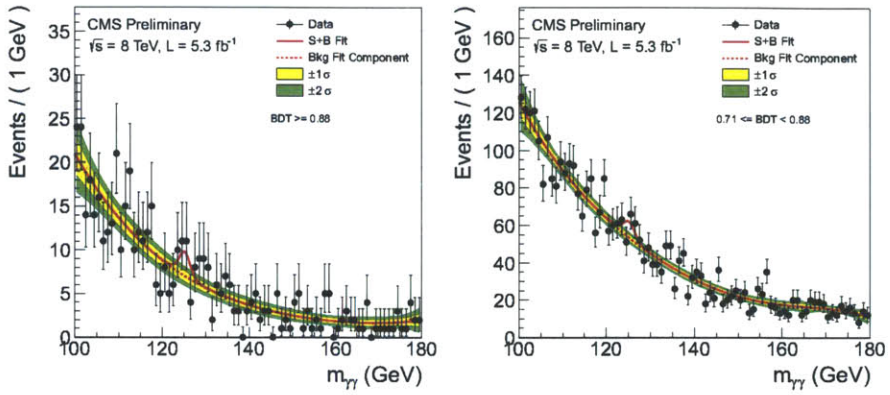


(d) Untagged 3



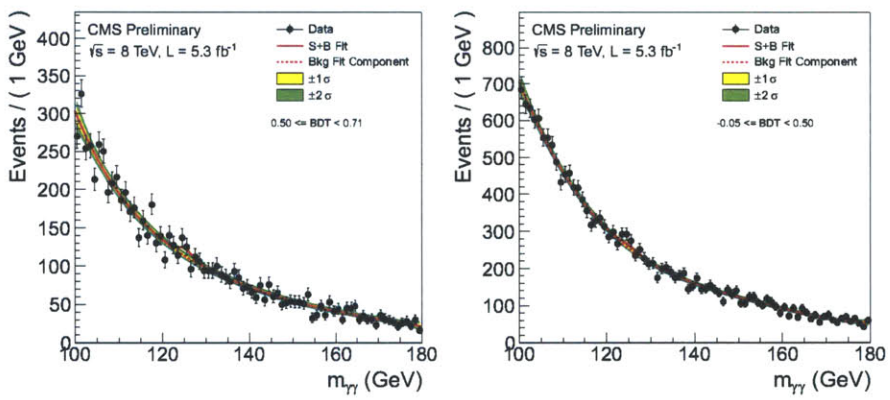
(e) VBF-tag

Figure 3-7: Final di-photon mass distributions for the 5 event categories for the 2011 analysis. The data is shown as the black points with error bars, and the signal plus background model from the full signal plus background fit is shown as the red solid line. The red dashed line shows the background component of the signal plus background fit.



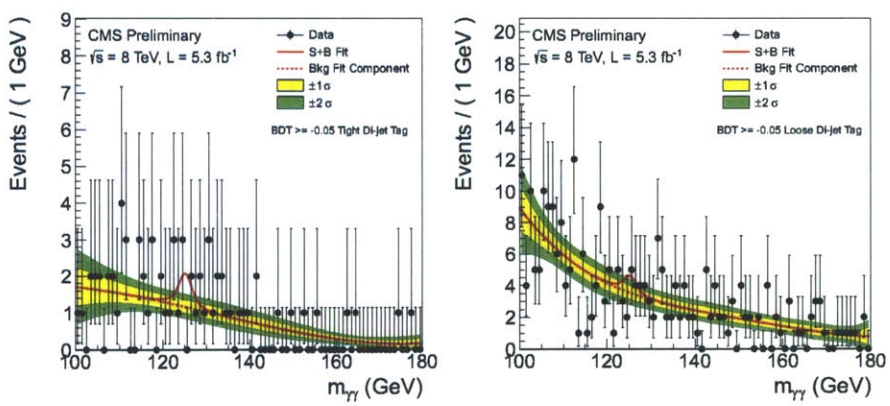
(a) Untagged 0

(b) Untagged 1



(c) Untagged 2

(d) Untagged 3



(e) Tight VBF-tag

(f) Loose VBF-tag

Figure 3-8: Final di-photon mass distributions for the 6 event categories for the 2012 analysis. The data is shown as the black points with error bars, and the signal plus background model from the full signal plus background fit is shown as the red solid line. The red dashed line shows the background component of the signal plus background fit.

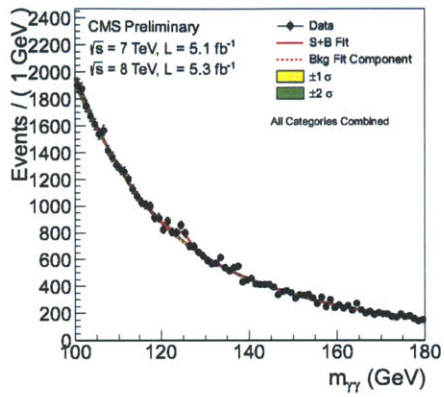


Figure 3-9: Final di-photon mass distributions for the sum of the 11 event categories. The data is shown as the black points with error bars, and the signal plus background model from the full signal plus background fit is shown as the red solid line. The red dashed line shows the background component of the signal plus background fit. The fit components shown are not from a dedicated fit to the combined distribution, but rather from the summed contributions from the 11 event categories.

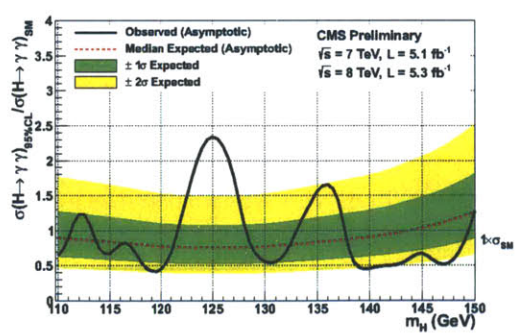
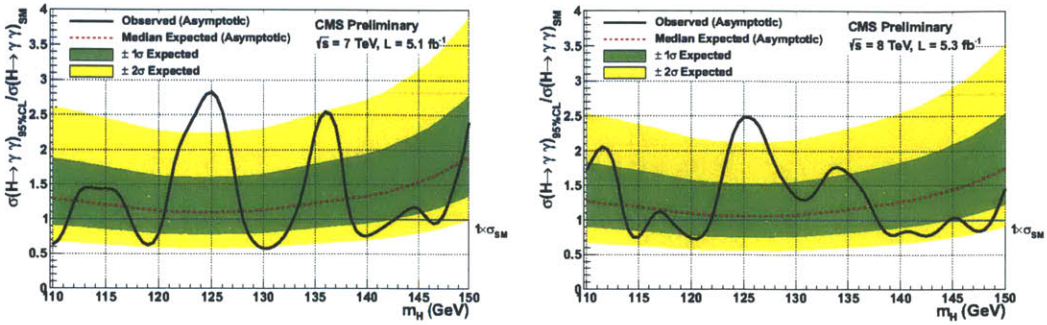


Figure 3-10: Exclusion limit at 95% confidence level for a Standard Model Higgs, expressed in terms of ratio to standard model cross section as a function of Higgs mass, and incorporating both 2011 and 2012 datasets.



(a) 7 TeV (b) 8 TeV

Figure 3-11: Exclusion limit at 95% confidence level for a Standard Model Higgs, shown separately for the 2011 and 2012 datasets.

for a standard model Higgs with mass 125 GeV, evaluated using the Asimov dataset, is 2.8 standard deviations. Since the search is conducted over the full mass range from 110 GeV$< m_H <$ 150 GeV, the trials factor to account for the possibility that a background fluctuation could appear anywhere in the mass range can be considered. Inferring this trials factor using the zero-crossing method [70], the global significance for this excess over the entire search range is 3.2 standard deviations. The global significance has also been explicitly computed, using toy Monte Carlo to assess the probability, with no asymptotic approximations, yielding a global significance of $3.2 \pm$ 0.1 standard deviations. The uncertainty on this number is statistical, and due to the finite number of toy Monte Carlo datasets.

The simultaneous likelihood fit across the eleven event categories makes optimal use of the information contained in the mass distributions for the purpose of extracting the final results, however the distributions presented as such are difficult to interpret visually, given that the excess at 125 GeV appears with contributions spread over most of the eleven categories. In order to aid in visualization of the result, a weighted combination of the 11 event categories is shown in Figure 3-13, with events weighted by a single factor for each category, proportional to $S/(S+B)$, computed in a $\pm 2\sigma_{\text{eff}}$ window around the fitted signal mass in each category, as computed from the simultaneous signal plus background fit. Since only the overall cross section multiplier

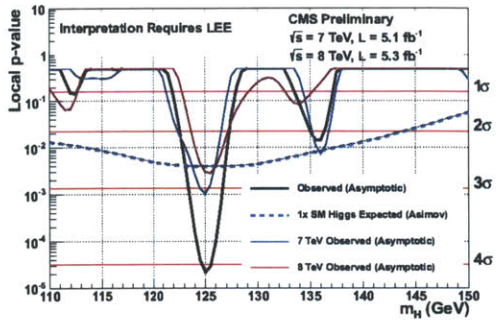


Figure 3-12: Local p-values for the background to fluctuate to give a result as signal-like as the observed, as a function of hypothesis mass. The Solid black line shows the observed local p-value for the combined dataset, with blue and red solid lines showing the observed p-value for the 7 TeV and 8 TeV datasets. The blue dashed line corresponds to the expected significance for a Standard Model Higgs at each mass, obtained by computing the significance for an Asimov toy dataset, thrown with statistical fluctuations suppressed.

μ is floated in the fit, the relative signal yield across the event categories is primarily constrained by the standard model expectation, rather than by the data, with variations allowed only within the systematic uncertainties. The weights are normalized in order to conserve the total number of fitted signal events in the peak. The $S/(S+B)$ weighting is an approximate representation of the importance of each category in the simultaneous likelihood fit. This weighting also corresponds to the optimal weighting if one were to extract results from a fit to the weighted mass distribution [71]. All results are obtained instead from the simultaneous fit, but this optimized weighting is still relevant for visual interpretation of the results. The fit components which are shown correspond to the weighted sum of the individual category contributions rather than a dedicated fit to the weighted combination.

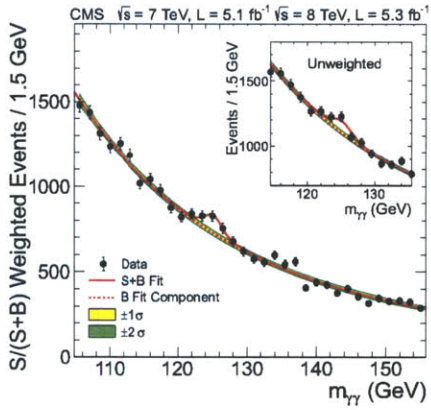


Figure 3-13: Final di-photon mass distributions for the weighted sum of the 11 event categories, with weights proportion to $S/(S+B)$ in each category. The data is shown as the black points with error bars, and the signal plus background model from the full signal plus background fit is shown as the red solid line. The red dashed line shows the background component of the signal plus background fit. The fit components shown are not from a dedicated fit to the combined distribution, but rather from the weighted sum of contributions from the 11 event categories. The data and fit components from the unweighted combination of the 11 categories are shown in the inset.

Chapter 4

Conclusions

We have performed a search for the production of the standard model Higgs boson decaying to diphotons in pp collisions at the LHC at $\sqrt{s}$ = 7-8 TeV. Having analyzed data corresponding to 5.1 fb^{-1} at $\sqrt{s}$ = 7 TeV (2011) and 5.3 fb^{-1} at $\sqrt{s}$ = 8 TeV (2012), a statistically significant excess of events is observed with respect to the background prediction. Interpreted as a standard model Higgs, this excess has a local significance of 4.1 standard deviations, with the maximum significance occurring for a Higgs mass of 125 GeV. Taking into account the trials factor given the search range of 110 GeV to 150 GeV in Higgs mass, this excess has a global significance of 3.2 standard deviations. This constitutes evidence for a new particle decaying to diphotons with a mass of around 125 GeV. The rate of observed events is consistent with predictions for the standard model Higgs boson.

In addition to the diphoton results, CMS has carried out the standard model Higgs search for a range of additional final states in the same dataset[72]. The ATLAS collaboration has also analyzed data collected over the same period, for the diphoton final state as well as several others [73]. These results are briefly summarized below.

4.1 Additional CMS Higgs Results and Combination

In addition to the diphoton results, CMS has conducted a search for the standard model Higgs in an overlapping mass region including Higgs decays to four charged leptons through a pair of Z bosons, decays to two charged leptons and two neutrinos through a pair of W bosons, decays to pairs of τ leptons, and decays to pairs of b quarks, as well as combined results for the five decay channels.

4.1.1 $H \to ZZ \to 4\ell$

A search has been conducted for a standard model Higgs decaying into four leptons through a pair of Z bosons. The search includes 4μ, $2e2\mu$ and $4e$ final states. The four lepton system for a Higgs decay forms a narrow mass peak, limited by the detector resolution, with muon final states having somewhat better mass resolution as compared to the electron final states. This channel has much smaller backgrounds than the diphoton search, but also much lower expected signal yield. The main background is irreducible non-resonant ZZ production, with a smaller reducible component arising mainly from events with two isolated charged leptons, plus additional leptons from heavy quark decays, mainly from $Z + b\bar{b}$ and $t\bar{t}$ events. The search is conducted using a two-dimensional maximum likelihood fit to the four-lepton mass, as well as a kinematic discriminator based on angular distributions of the four lepton system in its rest frame as well as the masses of the two di-lepton pairs corresponding to the Z bosons in the decay. This discriminator is a likelihood ratio for Higgs versus continuum ZZ discrimination, constructed from the analytic leading-order matrix element with parameterized detector response for the signal, and Monte Carlo simulation for the background. The mass and kinematic discriminator shapes for the signal and for the irreducible background are modeled using Monte Carlo simulation with appropriate correction factors, while the irreducible background is normalized using data control regions.

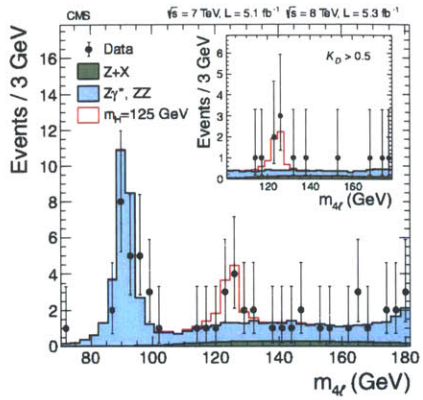


Figure 4-1: Four lepton mass distribution for the $H \to ZZ \to 4\ell$ search in CMS. The main plot shows the inclusive distribution, while the inset shows the four-lepton mass distribution after requiring the value of the kinematic discriminator to be greater than 0.5.

The four lepton mass distributions for all final states combined is shown in Figure 4-1 inclusively and after a cut on the kinematic discriminator in the inset.

Interpreted as a standard model Higgs boson, the excess of events between 120 GeV and 135 GeV has a local significance of 3.2 standard deviations, with the highest significance for a Higgs mass of 125.6 GeV. The expected sigificance for a standard model Higgs with this mass is 3.8 standard deviations, and this excess corresponds to a best fit cross section ratio compared to the standard model of $\mu = 0.7^{+0.4}_{-0.3}$, and is therefore also compatible with a standard model Higgs.

4.1.2 $H \to WW \to 2\ell 2\nu$

The standard model Higgs search in the $2\ell 2\nu$ final state differs significantly from the diphoton and four lepton searches on account of the missing energy carried by the neutrinos, preventing the reconstruction of a narrow mass peak. The search is conducted on the basis of two reconstructed charged leptons, as well as a moderate amount of missing transverse energy in the event. The main backgrounds include an irreducible contribution from non-resonant WW production, as well as a reducible contribution from W+jets, with one reconstructed lepton arising from mis-reconstruction of a

quark or gluon jet. For the same-flavour ee and $\mu\mu$ final states, there is an additional reducible background from Z+jets events with two charged leptons and missing transverse energy arising from detector mis-measurement. There are a significant number of events produced with $t\bar{t}$ dilepton decays containing two charged leptons and missing transverse energy. Because these events are usually produced with additional b or light quark/gluon jets, the search is carried out in exclusive jet bins and with anti-selection for jets originating from b quarks. In addition to zero and one jet event categories, a two-jet category is included in the search with kinematic requirements on the jets optimized to select VBF Higgs events.

The 7 TeV search is performed, in the zero and one jet event categories, using a binned likelihood fit to a BDT classifier trained to discriminate between Higgs events and the irreducible WW background, combined with a cut and count search in the two-jet category. The 8 TeV search for this dataset is performed using a cut and count approach in all three jet bins. The combined results of the search are shown, in the form of a 95% confidence level exclusion limits, in Figure 4-2.

The search results are characterized by a small excess over the background prediction, which, due to the low resolution of the search with respect to the mass of the Higgs boson, appears across a wide range of Higgs masses. The excess has a significance of 1.4 standard deviations, where the expected significance for a 125 GeV standard model Higgs in this search is 2.4 standard deviations. The search results are compatible with both a standard model Higgs with a mass around 125 GeV as well as the background only hypothesis.

4.1.3 $H \to \tau\tau$

The standard model Higgs search in the $\tau\tau$ decay mode includes final states where the taus decay to electron-muon pairs, dimuon pairs, an electron plus a hadronic tau decay, as well as a muon plus hadronic tau decay. Events are categorized according to jet multiplicity and the transverse momentum of the decay products. The search is carried out by means of a binned maximum likelihood fit to the ditau mass distribution, where the ditau mass is reconstructed by means of a likelihood fit combining

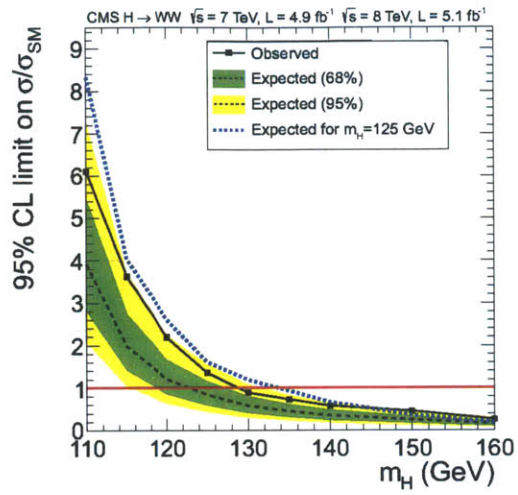


Figure 4-2: Exclusion limit at 95% confidence level for a Standard Model Higgs in the $WW \to 2\ell 2\nu$ final state, expressed in terms of ratio to standard model cross section as a function of Higgs mass, and incorporating both 7 TeV and 8 TeV datasets. The dashed black line shows the median expected limit in the case of background only whereas the dotted blue line shows the expected limit in the case of a standard model Higgs boson with a mass of 125 GeV.

the visible decay products and the missing transverse energy in the event. The results of the search are summarized by the exclusion limit in Figure 4-3.

No significant excess is observed, and the results are compatible with the background-only hypothesis as well as a 125 GeV Higgs boson.

4.1.4 $H \to bb$

The search for the Higgs decaying to $b\bar{b}$ pairs is carried out, for Higgs events produced in association with a W or Z boson of relatively large transverse momentum. This helps suppress the very large background from QCD multijet events. Events are selected with two jets satisfying b-tagging requirements based on the presence of displaced tracks or vertices from the b hadron decay, as well as high transverse momentum charged leptons and/or missing transverse energy corresponding to the W/Z decay. Events are categorized according to the vector boson decay mode and transverse momentum. Search results are extracted using a binned maximum likeli-

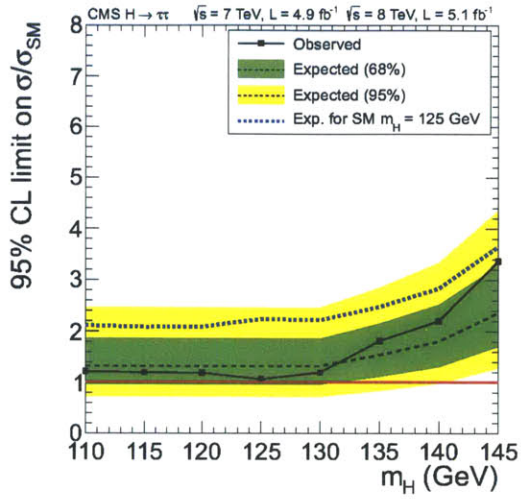


Figure 4-3: Exclusion limit at 95% confidence level for a Standard Model Higgs in the $\tau\tau$ final states, expressed in terms of ratio to standard model cross section as a function of Higgs mass, and incorporating both 7 TeV and 8 TeV datasets.

hood fit in each category to a BDT classifier. Search results are summarized by the exclusion limits in Figure 4-4.

The small excess which is observed has a statistical significance of 0.7 standard deviations, where the expected significance for a 125 GeV Higgs is 1.9 standard deviations. The results are as well compatible with either a standard model Higgs or the background only hypothesis.

4.1.5 Combined Results

The combined search results of the five decay channels are summarized in Figure 4-5, showing the local p-value as a function of Higgs mass for each decay channel as well as the combination.

The maximum combined local significance is 5.0 standard deviations, occurring at a Higgs mass of 125.5 GeV, with the main contributions to the excess coming from the diphoton and four-lepton search results. The expected significance for a standard model Higgs of this mass is 5.8 standard deviations. Taking into account the trials factor for a search range of 110-145 GeV gives a global significance of 4.5 standard

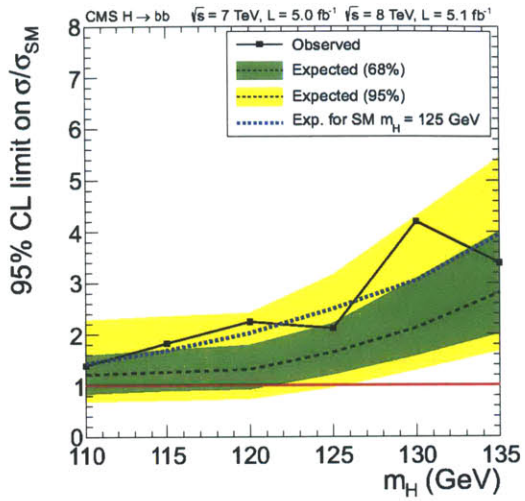


Figure 4-4: Exclusion limit at 95% confidence level for a Standard Model Higgs in the *bb* final states, expressed in terms of ratio to standard model cross section as a function of Higgs mass, and incorporating both 7 TeV and 8 TeV datasets..

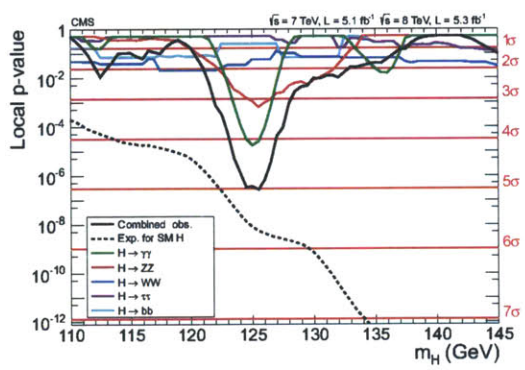


Figure 4-5: Local p-values for the background to fluctuate to give a result as signal-like as the observed, as a function of Higgs mass. The solid black line shows the observed local p-value for the combination of the five search channels. The black dashed line corresponds to the expected significance for a standard model Higgs at each mass, obtained by computing the significance for an Asimov toy dataset, thrown with statistical fluctuations suppressed. The observed local p-values for the individual search channels are shown by the colored solid lines.

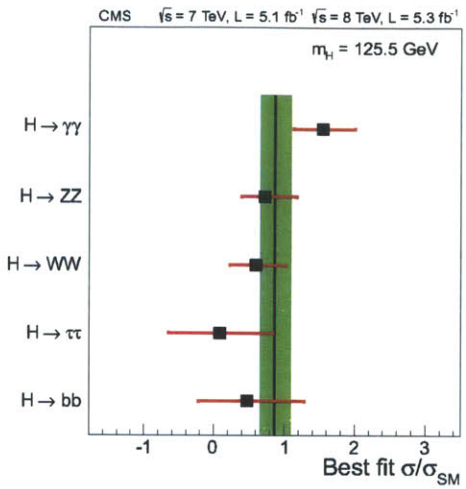


Figure 4-6: Best fit ratio to standard model cross section for each of the five Higgs decay channels, shown as the data points with error bars, and for the combination, indicated by the solid line and error band.

deviations for the combined search. A summary of the best fit ratio to the standard model cross section for a Higgs mass of 125.5 GeV for each search channel is shown in Figure 4-6, indicating that the results are consistent with a standard model Higgs given the present uncertainties.

For the diphoton and four lepton final states, the two channels with high mass resolution, the two-dimensional likelihood contours for Higgs mass versus cross section ratio are shown in Figure 4-7. The best fit Higgs mass, profiling independently over the signal strength for the diphoton inclusive and dijet tag categories as well as the four lepton signal strength, is $m_H = 125.3 \pm 0.4(\text{stat.}) \pm 0.5(\text{syst.})$ GeV.

4.2 ATLAS Results

4.2.1 Diphoton Search

The ATLAS Higgs to diphoton search [74] incorporating data collected over the same time period includes data corresponding to 4.8 fb^{-1} at $\sqrt{s} = 7$ TeV as well as 5.9 fb^{-1} at $\sqrt{s} = 8$ TeV. The basic search strategy is similar to that of CMS, a search for a

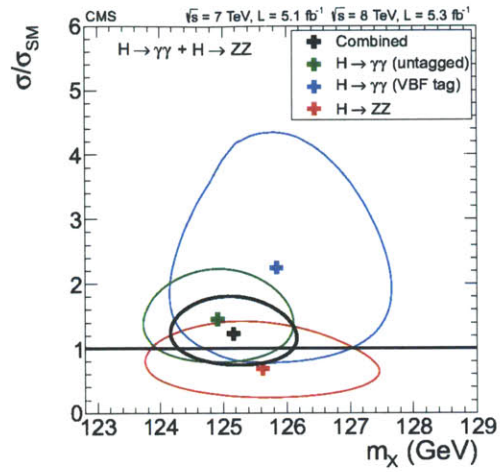


Figure 4-7: Two-dimensional maximum likelihood values and 68% confidence level contours for the diphoton and four lepton channels in Higgs mass versus cross section ratio. Diphoton results, divided into dijet-tagged and untagged event categories, and four lepton results are shown individually by the coloured lines, and the combined fit is shown by the black line.

narrow diphoton mass peak on top of a large smoothly falling background, dominated by irreducible QCD diphoton production, but still with an appreciable contribution from photon+jet event with a jet misidentified as a photon.

There are some important differences in the event reconstruction compared to CMS, driven in part by differences in the detector design. The design strength of the CMS Ecal is its excellent intrinsic energy resolution, due in large part to the homogeneous nature of the calorimeter. One trade-off necessitated by the lack of longitudinal segmentation in the CMS Ecal is the absence of any angular information about the incident photons, such that the CMS analysis relies entirely on the tracking information for the reconstruction of the primary vertex location. The ATLAS electromagnetic calorimeter on the other hand is a sampling design with lead absorbers and three active layers of liquid argon in the central region. Although the energy resolution for unconverted photons in the central region is somewhat worse, the longitudinal segmentation, combined with fine-grained transverse segmentation in the first layer provides sufficient angular resolution that the angular contribution

to the diphoton mass resolution in the ATLAS search is negligible.

Reconstruction and identification of photons is broadly similar to CMS, though with less usage of multivariate techniques. Energy corrections for photon clusters are derived from simulation as in CMS, but using a simpler parameterized correction as compared to the multivariate regression in CMS. Cluster corrections in ATLAS, in addition to the local and global containment effects which are similar as for CMS, must also correct for the energy deposited in the passive material of the sampling layers. Photon identification is based on shower profile and isolation quantities as for CMS. ATLAS uses a multivariate classifier based on a neural network for photon-jet discrimination in the 7 TeV data. This is conceptually similar to a BDT classifier in that it is trained to approximate the relevant likelihood ratio. The 8 TeV analysis instead uses a simpler cut-based photon-jet discrimination at present.

Events are categorized, as in the CMS analysis, in order to exploit different signal to background ratios and resolution. The categorization relies on a simple set of cut based selections, depending on the location of the photons in the detector, whether or not the photons are matched to conversions in the tracking detector, and the transverse momentum of the diphoton system. There is one additional event category for events with two additional jets passing a set of VBF topology cuts. There are a total of ten event categories each for 7 TeV and 8 TeV data, which are kept in separate event categories as for CMS.

The final signal extraction is based on a simultaneous likelihood fit to the 20 event categories, using parametric functions with freely floating parameters for the background, and parametric functions fixed to the Monte Carlo samples within systematic uncertainties for the signal, again very similar to the CMS analysis. The choice of parametric forms for the background model differs somewhat from CMS however, with a mix of fourth order Bernstein polynomials, simple exponentials, and exponentials of second order polynomials being used, depending on the category. The simple exponential and exponential of a second order polynomial are in particular more restrictive functions with fewer degrees of freedom than the third to fifth order polynomials which are used in the CMS analysis, implying that the contribution

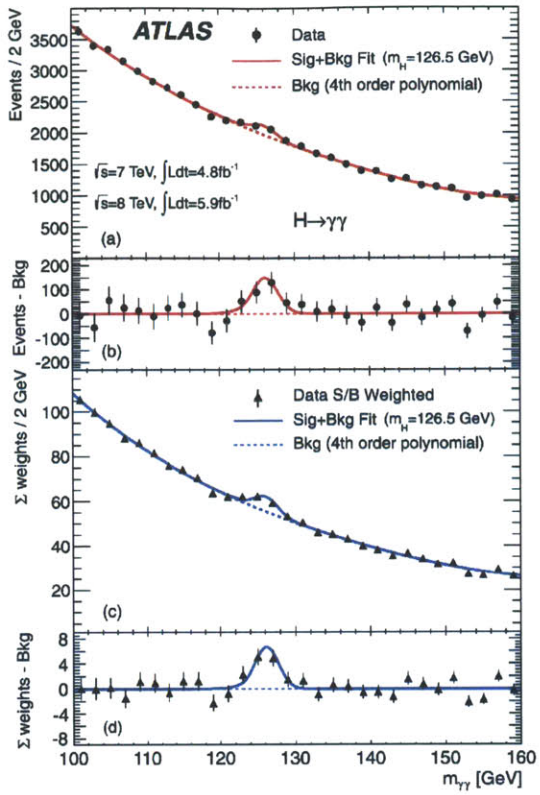
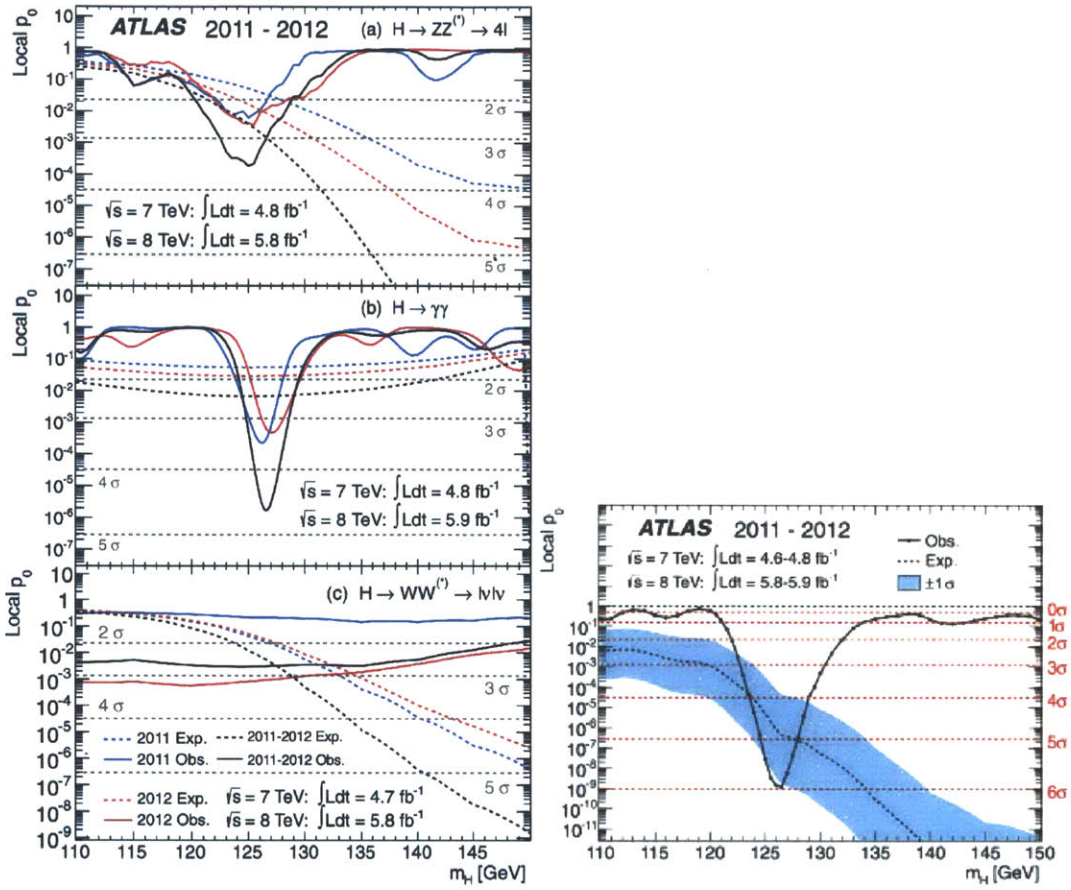


Figure 4-8: ATLAS diphoton mass distribution across 20 event categories in 7 TeV and 8 TeV data in both unweighted form (top) and with event weights proportional to $\ln(1 + S/B)$ (bottom).

of the background shape uncertainty to the total statistical uncertainty of the fit is somewhat less for the ATLAS analysis than for CMS.

The diphoton mass distribution for the combined ATLAS search across all 20 event categories is shown in Figure 4-8, both in unweighted form, and with event weights proportional to $\ln(1 + S/B)$.

The most significant excess occurs for a Higgs mass of 126.5 GeV with a local significance of 4.5 standard deviations. The expected significance for a standard model Higgs of this mass in the ATLAS search is 2.5 standard deviations (compared to 2.8 standard deviations for the CMS search.). The best fit ratio to standard model cross section is $\mu = 1.8 \pm 0.5$, compatible with the CMS result as well as a standard model Higgs at this mass.



(a) Individual Channels (b) Combination

Figure 4-9: ATLAS local p-values as a function of Higgs mass for individual channels (a) and for the combined results (b).

4.2.2 Combined Results

In addition to the diphoton results, ATLAS has also carried out the standard model Higgs search in the $ZZ \to 4\ell$ and $WW \to 2\ell 2\nu$ channels for this dataset, as well as combined results [73]. Local p-values as a function of Higgs mass are shown for the three channels individually and for the combination in Figure 4-9.

The combined local significance is 6.0 standard deviations, occurring for a Higgs mass of 126.5 GeV. The expected significance for a standard model Higgs of this mass is 4.9 standard deviations, compared to about 5.8 standard deviations for the combined CMS search. The combined best fit mass value for ATLAS diphoton and

four lepton results is $m_H = 126.0 \pm 0.4(\text{stat.}) \pm 0.4(\text{syst.})$ GeV, well compatible with the CMS result.

4.3 Final Remarks

The evidence for a new particle decaying to diphotons is augmented by compatible excesses of events in the CMS four lepton search, as well as both the diphoton and four lepton searches in ATLAS, in both 2011 and 2012 datasets, and with compatible mass and cross sections between experiments. The combined local significance for each experiment individually reaches 5.0 standard deviations, sufficient to claim the observation of a new particle in each of the two experiments independently. The observed event yields are broadly consistent with a standard model Higgs with a mass near 125 GeV. The observation of decays to diphotons and to pairs of Z bosons implies that the new particle is a boson, and decays to diphotons imply that the spin of the new particle must be different from 1 [75, 76].

A definitive statement about the nature of the new particle awaits future precision measurements of its couplings, spin, and parity, with this observation marking the beginning of a long campaign to measure those properties at the LHC and future experiments.

Bibliography

[1] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961.

[2] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[3] Abdus Salam. Weak and Electromagnetic Interactions. *Conf.Proc.*, C680519:367–377, 1968.

[4] H. David Politzer. Reliable Perturbative Results for Strong Interactions? *Phys.Rev.Lett.*, 30:1346–1349, 1973.

[5] D.J. Gross and Frank Wilczek. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys.Rev.Lett.*, 30:1343–1346, 1973.

[6] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys.Rev.Lett.*, 13:321–323, 1964.

[7] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys.Lett.*, 12:132–133, 1964.

[8] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys.Rev.Lett.*, 13:508–509, 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, 1964.

[10] Peter W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys.Rev.*, 145:1156–1163, 1966.

[11] T.W.B. Kibble. Symmetry breaking in non-Abelian gauge theories. *Phys.Rev.*, 155:1554–1561, 1967.

[12] John M. Cornwall, David N. Levin, and George Tiktopoulos. Uniqueness of spontaneously broken gauge theories. *Phys.Rev.Lett.*, 30:1268–1270, 1973.

[13] John M. Cornwall, David N. Levin, and George Tiktopoulos. Derivation of Gauge Invariance from High-Energy Unitarity Bounds on the s Matrix. *Phys.Rev.*, D10:1145, 1974.

[14] C.H. Llewellyn Smith. High-Energy Behavior and Gauge Symmetry. *Phys.Lett.*, B46:233–236, 1973.

[15] Benjamin W. Lee, C. Quigg, and H.B. Thacker. Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass. *Phys.Rev.*, D16:1519, 1977.

[16] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys.Lett.*, B565:61–75, 2003.

[17] T. Aaltonen et al. Combination of Tevatron searches for the standard model Higgs boson in the W+W- decay mode. *Phys.Rev.Lett.*, 104:061802, 2010.

[18] Precision Electroweak Measurements and Constraints on the Standard Model. 2010.

[19] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[20] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[21] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[22] Elliott D. Bloom, D.H. Coward, H.C. DeStaebler, J. Drees, Guthrie Miller, et al. High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees. *Phys.Rev.Lett.*, 23:930–934, 1969.

[23] Martin Breidenbach, Jerome I. Friedman, Henry W. Kendall, Elliott D. Bloom, D.H. Coward, et al. Observed Behavior of Highly Inelastic electron-Proton Scattering. *Phys.Rev.Lett.*, 23:935–939, 1969.

[24] John C. Collins and Davison E. Soper. Parton Distribution and Decay Functions. *Nucl.Phys.*, B194:445, 1982.

[25] H.M. Georgi, S.L. Glashow, M.E. Machacek, and Dimitri V. Nanopoulos. Higgs Bosons from Two Gluon Annihilation in Proton Proton Collisions. *Phys.Rev.Lett.*, 40:692, 1978.

[26] R.N. Cahn and Sally Dawson. Production of Very Massive Higgs Bosons. *Phys.Lett.*, B136:196, 1984.

[27] S.L. Glashow, Dimitri V. Nanopoulos, and A. Yildiz. Associated Production of Higgs Bosons and Z Particles. *Phys.Rev.*, D18:1724–1727, 1978.

[28] Risto Raitio and Walter W. Wada. Higgs Boson Production at Large Transverse Momentum in QCD. *Phys.Rev.*, D19:941, 1979.

[29] John N. Ng and Pierre Zakarauskas. A QCD Parton Calculation of Conjoined Production of Higgs Bosons and Heavy Flavors in p anti-p Collision. *Phys.Rev.*, D29:876, 1984.

[30] Z. Kunszt. Associated Production of Heavy Higgs Boson with Top Quarks. *Nucl.Phys.*, B247:339, 1984.

[31] William J. Marciano and Frank E. Paige. Associated production of Higgs bosons with t anti-t pairs. *Phys.Rev.Lett.*, 66:2433–2435, 1991.

[32] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*, CERN, Geneva, 2011.

[33] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. *CERN-2012-002*, CERN, Geneva, 2012.

[34] LHC Higgs Cross Section Working Group. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections`.

[35] V Karimki. *The CMS tracker system project: Technical Design Report. oai:cds.cern.ch:368412.* Technical Design Report CMS. CERN, Geneva, 1997.

[36] *The CMS tracker: addendum to the Technical Design Report. oai:cds.cern.ch:490194.* Technical Design Report CMS. CERN, Geneva, 2000.

[37] *The CMS electromagnetic calorimeter project: Technical Design Report. oai:cds.cern.ch:349375.* Technical Design Report CMS. CERN, Geneva, 1997.

[38] Philippe Bloch, Robert Brown, Paul Lecoq, and Hans Rykaczewski. *Changes to CMS ECAL electronics: addendum to the Technical Design Report. oai:cds.cern.ch:581342.* Technical Design Report CMS. CERN, Geneva, 2002.

[39] The CMS Collaboration. CMS PAS EGM-11-001, Performance of the CMS ECAL with 5 fb^{-1} of data at $\sqrt{s} = 7$ TeV, 2012.

[40] Paolo Meridiani and Chris Seez. CMS IN -2011/002, Definition of calibrated ECAL RecHits and the ECAL calibration and correction scheme, 2011.

[41] *The CMS hadron calorimeter project: Technical Design Report. oai:cds.cern.ch:357153.* Technical Design Report CMS. CERN, Geneva, 1997.

[42] *The CMS muon project: Technical Design Report. oai:cds.cern.ch:343814.* Technical Design Report CMS. CERN, Geneva, 1997.

[43] *CMS TriDAS project: Technical Design Report; 1, the trigger systems. oai:cds.cern.ch:706847.* Technical Design Report CMS.

[44] Sergio Cittolin, Attila Rcz, and Paris Sphicas. *CMS trigger and data-acquisition project: Technical Design Report. oai:cds.cern.ch:578006.* Technical Design Report CMS. CERN, Geneva, 2002.

[45] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, 1933.

[46] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28:2000, 1998.

[47] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.

[48] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.

[49] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl.Instrum.Meth.*, A506:250–303, 2003.

[50] John Allison, K. Amako, J. Apostolakis, H. Araujo, P.A. Dubois, et al. Geant4 developments and applications. *IEEE Trans.Nucl.Sci.*, 53:270, 2006.

[51] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 0411:040, 2004.

[52] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 0711:070, 2007.

[53] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 1006:043, 2010.

[54] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 1106:128, 2011.

[55] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO Higgs boson production via gluon fusion matched with shower in POWHEG. *JHEP*, 0904:002, 2009.

[56] Paolo Nason and Carlo Oleari. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *JHEP*, 1002:037, 2010.

[57] Lance J. Dixon and M. Stewart Siu. Resonance continuum interference in the diphoton Higgs signal at the LHC. *Phys.Rev.Lett.*, 90:252001, 2003.

[58] Daniel de Florian, Giancarlo Ferrera, Massimiliano Grazzini, and Damiano Tommasini. Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC. *JHEP*, 1111:064, 2011.

[59] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[60] Wolfgang Adam, R. Frhwirth, Are Strandlie, and T. Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. 2005.

[61] W. Adam, B. Mangano, T. Speer, and T. Todorov. Track reconstruction in the CMS tracker. 2005.

[62] Tomasz Skwarnicki. A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances. 1986.

[63] J. Beringer et al. Review of Particle Physics (RPP). *Phys.Rev.*, D86:010001, 2012.

[64] Paul Avery. Applied Fitting Theory VI: Formulas for Kinematic Fitting. `http://www.phys.ufl.edu/~avery/fitting/kinematic.pdf`, 1999.

[65] J.M. Marriner. Secondary vertex fit with mass and pointing constraints (ctvmft). 1996.

[66] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys.Lett.*, B659:119–126, 2008.

[67] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 0804:063, 2008.

[68] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput.Phys.Commun.*, 10:343–367, 1975.

[69] ATL-PHYS-PUB-2011-011, Procedure for the LHC Higgs boson search combination in summer 2011. 2011.

[70] Eilam Gross and Ofer Vitells. Trial factors or the look elsewhere effect in high energy physics. *Eur.Phys.J.*, C70:525–530, 2010.

[71] Roger J. Barlow. Event Classifications Using Weighted Methods. *J.Comput.Phys.*, 72:202–219, 1987.

[72] Serguei 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.

[73] Georges 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.

[74] Observation of an excess of events in the search for the standard model higgs boson in the gamma-gamma channel with the atlas detector. Technical Report ATLAS-CONF-2012-091, CERN, Geneva, Jul 2012.

[75] L.D. Landau. On the angular momentum of a two-photon system. *Dokl.Akad.Nauk Ser.Fiz.*, 60:207–209, 1948.

[76] Chen-Ning Yang. Selection Rules for the Dematerialization of a Particle Into Two Photons. *Phys.Rev.*, 77:242–245, 1950.