CMS Physics Analysis Summary

Contact: cms-pog-conveners-egamma@cern.ch 2010/07/21

Photon reconstruction and identification at $\sqrt{s} = 7$ TeV

The CMS Collaboration

Abstract

We describe the commissioning of photon reconstruction and identification in the CMS experiment using 74 nb^{-1} of integrated luminosity collected at the LHC at $\sqrt{s}$ = 7 TeV. The most important observables for photon reconstruction, identification, and isolation have been measured and compared with the Monte Carlo simulation. Results are first shown for a loosely selected, background-dominated sample. A robust photon selection to be used for commissioning and early analysis is then described. Results for a signal-enriched sample using this selection are then presented. The commissioning of new observables designed for the identification of the conversion of isolated photons is then discussed. Finally, the contamination of the sample from beam halo bremsstrahlung is investigated, and an upper limit on such contamination for the current sample is established.

1 Introduction

The LHC began operation at $\sqrt{s}$ = 7 TeV at the end of March 2010. This paper presents a commissioning of the reconstruction, identification and isolation observables for photon objects based on an integrated luminosity of 74 nb^{-1} collected by the CMS experiment [1].

The key quantities which contribute to photon measurements are described in Section 2, and observables for the reconstruction, identification and isolation of photons candidates are compared with the expectation from Monte Carlo simulations. Commissioning of photon conversion reconstruction is presented in Section 3. The measurement of the corresponding key observables for isolated photons are compared with the expectation from Monte Carlo simulations. Finally, the number of fake photon candidates in data due to non-collision backgrounds is estimated with data-driven techniques in Section 4.

2 Supercluster and Photon reconstruction, corrections and observables

2.1 Supercluster reconstruction

The CMS electromagnetic crystal calorimeter (ECAL) was designed to have both excellent energy resolution and high granularity, making it a powerful instrument to measure photons with high precision at the LHC. The ECAL is organized into two pseudo-rapidity regions: the ECAL barrel or EB ($|\eta| < 1.479$) and the ECAL endcap or EE ($1.479 < |\eta| < 3.0$). In the endcap region there is a lead-silicon-strip preshower detector (ES) consisting of two orthogonal silicon strip detectors that have a strip pitch of 1.9 mm. One plane of the detector is at a depth of 2 X_0 and the other is at 3 X_0. ES coverage ranges from $1.653 < |\eta| < 2.6$ The ECAL is surrounded by a brass/scintillator sampling hadron calorimeter (HCAL) with coverage up to $|\eta| < 3.0$. The HCAL is sub-divided into towers with a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$, corresponding to 5×5 ECAL crystals in the barrel region. The CMS tracking system is composed of several layers of silicon pixel and silicon strip detectors and covers the region $|\eta| < 2.5$. In the following, we only consider candidates which fall in this η range. The tracker and both calorimeters are embedded inside a 3.8 T solenoidal magnet.

The presence of material in front of the calorimeter can cause conversion of photons and bremsstrahlung from electrons and positrons. Because of the strong magnetic field the energy flow associated to primary electrons or converted primary photons then spreads in ϕ with increasing distances from the collision vertex within the tracker volume. To collect the photon energy in the ECAL, local deposits ("basic clusters") are summed into superclusters (SCs) which are extended in ϕ. The clustering algorithm used in EB, called the 'hybrid' algorithm, is described in Ref. [2]. Superclustering in EE and ES proceeds similarly, but uses a slightly different algorithm, the 'Multi5x5', which adds together fixed 5×5 crystal basic clusters. EE cluster positions are extrapolated to the ES where ES clusters are built. The total endcap basic cluster energy is the sum of cluster energies in the EE and ES.

After applying small energy corrections (described in Section 2.3) the superclusters are used to reconstruct photons and electrons, and to seed electron track reconstruction. While the clustering algorithm is independent of the source (e.g. a primary electron or a so-called "prompt" photon), the energy measurement depends on a classification which is discussed in the case of photons in Section 2.4. To illustrate our understanding of the clustering and photon measurement process, the results in the following are presented first for a loosely selected sample of superclusters (consisting mostly of QCD-induced background) and then for a signal-enriched

sample of photon candidates.

2.2 Supercluster selection and observables

In order to compare observables important to supercluster reconstruction between the data collected in 2010 and Monte Carlo simulations, we imposed the following selection:

- We required the events to pass the high-level trigger path HLT_Photon15_L1R, seeded by the Level-1 EG5 trigger [3].
- We required that superclusters have a minimum uncorrected ('raw') transverse energy of 20 GeV.
- We considered only superclusters in regions in η also covered by the tracking detectors ($|\eta| \leq 2.5$), and we exclude the region in η between the barrel and the endcaps, which is occupied to a significant extent by services (($|\eta| > 1.4442$ and $|\eta| < 1.566$)
- We have removed superclusters containing anomalous ECAL signals by imposing the criteria on topology and signal time outlined in [4].
- We cut on the hadronic to electromagnetic energy ratio $H/E < 0.05$ to reject jets with a substantial hadronic component, and to avoid differences with the simulation of the Level-1 trigger, which did not take into account the relaxed H/E cut used online at startup.

Reconstructed superclusters in data were compared to those in a Monte Carlo sample of QCD and photon plus jet events with outgoing parton transverse momentum in the center of mass frame of $\hat{p}_T > 15$ GeV simulated with PYTHIA 6 [5], tune D6T. In the data sample analyzed, there are a total of about 1.3×10^5 events passing the above selection.

Figure 1 shows the pseudorapidity distribution of reconstructed superclusters. Here the Monte Carlo and data statistics are normalized using the integrated luminosity in data. A profile histogram of the ratio between data and Monte Carlo simulation is shown beneath each distribution in this section. The measurements are fairly well reproduced by the Monte Carlo simulation. The dips at the transition between the ECAL Barrel and Endcap is due to the fiducial cut introduced above.

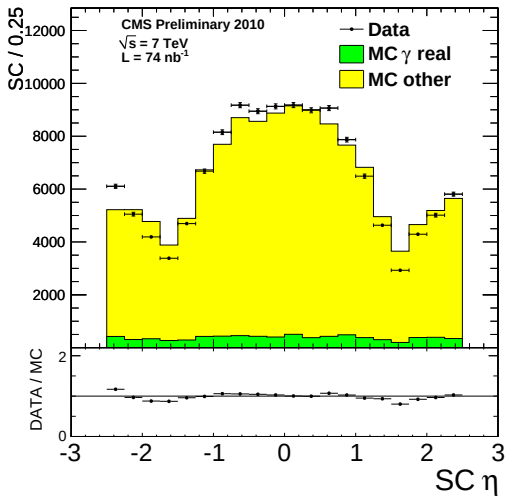


Figure 1: Detector pseudorapidity of superclusters in 7 TeV minimum bias collisions

Figure 2 shows the distribution of the uncorrected transverse energy of reconstructed superclusters. The supercluster E_T spectrum is well reproduced by the simulation over a wide range.

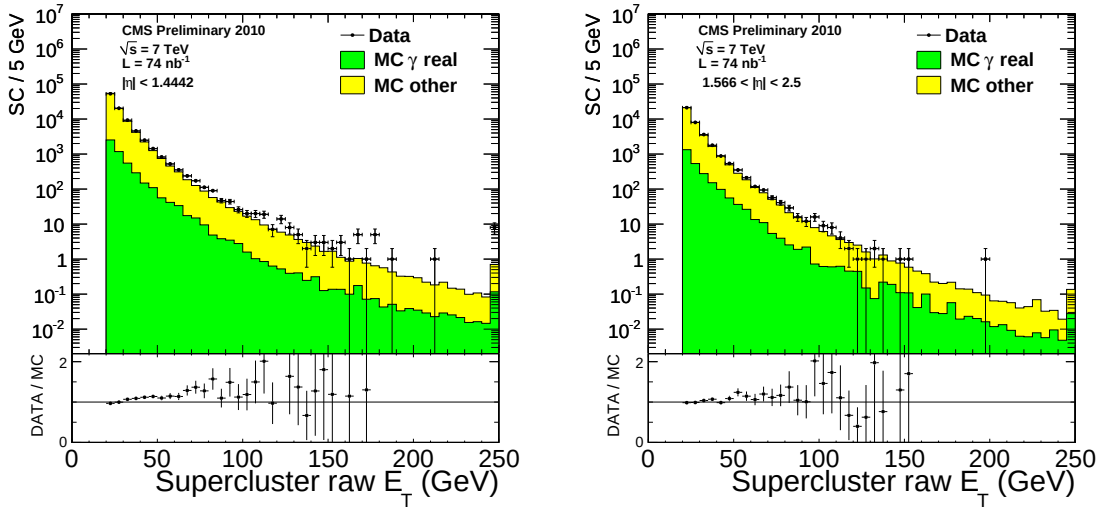


Figure 2: Uncorrected transverse energy spectra for EB (left) and EE (right) superclusters. The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

2.3 Supercluster energy corrections

Energy corrections have been developed using Monte Carlo simulation to take into account the interactions with the material in front of ECAL and shower containment:

- To compensate for the η dependence of the lateral energy leakage, arising from the 3° offpointing of the EB crystals. This correction is applied only to EB superclusters.
- A correction to compensate for interactions with material in front of ECAL. Since these interactions spread energy only in the ϕ direction, this loss can be parametrised as a function of the ratio of the supercluster size in ϕ to its size in η.
- A residual correction which is a function of the supercluster E_T and η, to compensate for variations along η in the amount of tracker material and the dependence on E_T of the bremsstrahlung and conversion processes.

For simulated signal photons, these energy corrections are typically about 1% of the uncorrected energy.

2.4 From corrected supercluster to photon candidate

Photon objects are reconstructed from the corrected superclusters, assigning the candidate momentum to the location of the reconstructed primary vertex.

The energy of each photon candidate is estimated based on an observable called $r9$ which is the ratio of the energy contained within the 3×3 array of crystals centered on the seed crystal of the photon candidate's supercluster to the total energy contained in the supercluster. This quantity is used to determine if the photon is converted or unconverted. If the $r9$ of the candidate is above 0.94 (0.95) in the barrel (endcap), the energy of the 5×5 crystals ($E_{5\times5}$) around the highest energy crystal is used. Otherwise, the supercluster energy is used. The $r9$ threshold is larger in the endcap because the crystals are larger than in the barrel.

Figure 3 shows the distribution of the $r9$ observable used for the energy determination of photon candidates. The shape of the distribution is similar in data and Monte Carlo, so the fraction of the sample of photon candidates lying above or below the typical 0.94 (0.95) cut agrees with Monte-Carlo within 10%

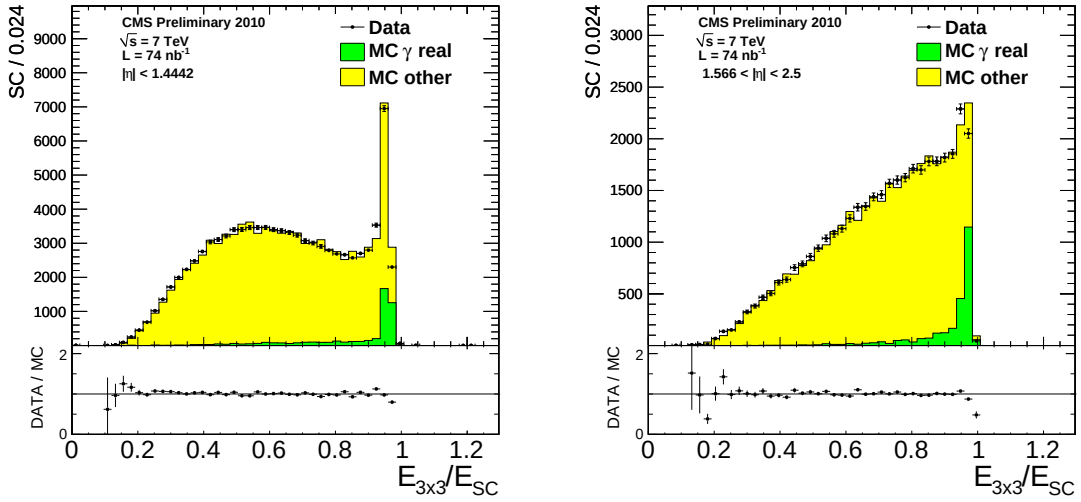


Figure 3: Ratio of the energy of the 3x3 matrix of crystals surrounding the most energetic crystal to the total energy of the supercluster, for EB (left) and EE (right) superclusters. The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

2.5 Photon identification and isolation

To increase the purity of the photon sample, we apply additional isolation and identification requirements. The selection criteria and the cut values have been adjusted on the basis of Monte Carlo simulation, in absence of a data-driven control sample, aiming to maximize the background rejection while keeping the efficiency flat as a function of η and E_T. This is a robust selection intended to be used for commissioning and early analysis. By design, we use a very similar isolation scheme to that used for electrons, so that photon ID efficiencies can be measured on an electron control sample in data when sufficient statistics are available. The photon selection is based on the following variables:

- The supercluster is required not to match pixel hits consistent with a track from the interaction region.
- ECAL isolation: the sum of ECAL E_T around the photon candidate in an annular region of inner radius $\mathcal{R} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.06$ and outer radius 0.4. A three-crystal wide strip along ϕ is excluded.
- HCAL isolation: the sum of HCAL E_T around the photon candidate in an annular region of inner radius $\mathcal{R} = 0.15$ and outer radius 0.4.
- Tracker isolation: the scalar sum of p_T of tracks consistent with the primary vertex in a hollow cone around the photon candidate in an annular region of inner radius $\mathcal{R} = 0.04$ and outer radius 0.4. The inner radius is chosen to avoid counting the momentum of photon conversion tracks in the isolation sum.
- $\sigma_{i\eta i\eta}$ is the η-η element of the η-ϕ covariance matrix, which provides another expression for the extent in η of the supercluster, similar to σ_η. (It is calulated with

logarithmic weights, and relative to $E_{5\times5}$ instead of the supercluster energy). It is given by:

$$\sigma^2_{i\eta i\eta} = \frac{\sum_i^{5\times5} w_i (i\eta_i - i\eta_{seed})^2}{\sum_i^{5\times5} w_i},\ w_i = \max(0,\ 4.7 + \ln\frac{E_i}{E_{5\times5}}), \quad (1)$$

where E_i and $i\eta_i$ are the energy and η index of the i^{th} crystal within the 5×5 electromagnetic cluster, $E_{5\times5}$ is the energy of the 5×5 crystals around the seed crystal, and $i\eta_{seed}$ is the η index of the seed crystal.

Table 1: Cut thresholds for the barrel and the endcap regions.

Variable	Barrel	Endcap
pixel seed	require none	
E_T	30 GeV	
Tracker Iso	2.0 GeV	
ECAL Iso	4.2 GeV	
HCAL Iso	2.2 GeV	
H/E	0.05	
$\sigma_{i\eta i\eta}$	0.01	0.03

The values of the cuts adopted for the selection are shown in Table 1. With this selection, the Monte-Carlo predicts a purity of about 50% signal photons with $E_T > 30$ GeV. The efficiency of the selection on isolated photons in Monte Carlo is about 90% in the EB and 80% in EE. The total number of photon candidates which pass the selection in the analyzed data sample is about 2.9×10^3.

The signal-enriched sample can be used to compare photon variables in data to the Monte Carlo. Photon candidates in 2010 data were compared to a sum (normalised to relative cross-sections) of those in Monte Carlo samples of QCD and γ + jet events. In each figure, we show the estimated contributions (based on Monte Carlo truth information) from real photons coming from both the hard interaction and from ISR/FSR, as well as that coming from the QCD background. In each case the simulation describes the overall shape of the distributions well.

Figure 5 shows the sum of the ECAL, HCAL, and track isolation before the selection, Figures 6-8 show the $N-1$ distributions for each variable after applying all the cuts in the selection except for the cut on the plotted variable. In general the agreement of the simulation with the data is satisfactory. In some cases (e.g. the track isolation in the endcap region) the observed isolation values are below what the simulation predicts. This effect is not completely understood.

Figure 4 shows the $N-1$ plot for the pixel seed variable, where the fraction of candidates with a seed is consistent with the simulation, and the enhancement in signal purity for events without the seed can be seen. Figure 9a shows the $N-1$ plot for the $\sigma_{i\eta i\eta}$ variable in EB. The signal contribution to the peak below 0.010 is clearly visible. Figure 9b shows the $N-1$ plot for the $\sigma_{i\eta i\eta}$ variable in EE. The signal contribution to the peak below 0.030 is clearly visible. Figure 12 shows the $r9$ distribution. After photon ID the distribution has on average larger values, as expected from the simulation.

Figure 10 shows the transverse energy spectrum of candidates in data. Figure 11 shows the η distribution of the selected photon candidates.

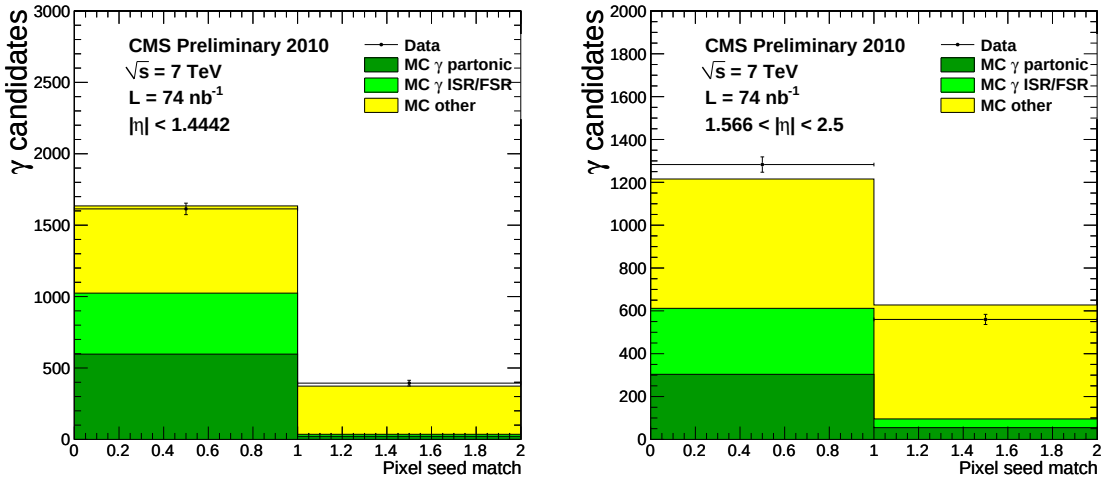


Figure 4: $N-1$ plot of the flag for presence of pixel seed in EB (left) and EE (right) photons, used in the selection in the text. The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

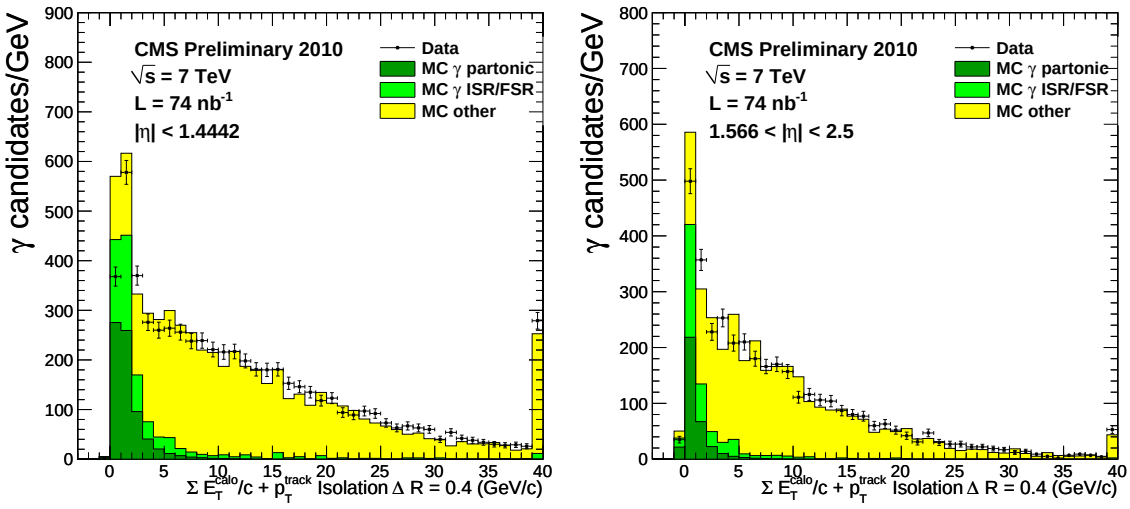


Figure 5: Sum of the isolation variables (ECAL, HCAL, and tracks) for barrel (left) and endcap (right) photon candidates, before applying photon isolation cuts. The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

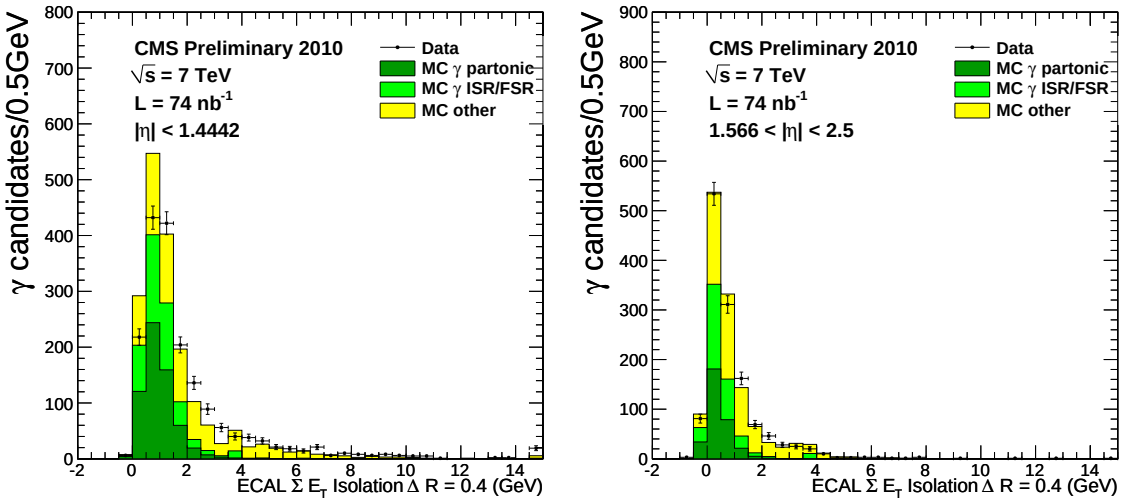


Figure 6: $N-1$ ECAL isolation distribution for data and MC, shown for barrel (right) and endcap (left). The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

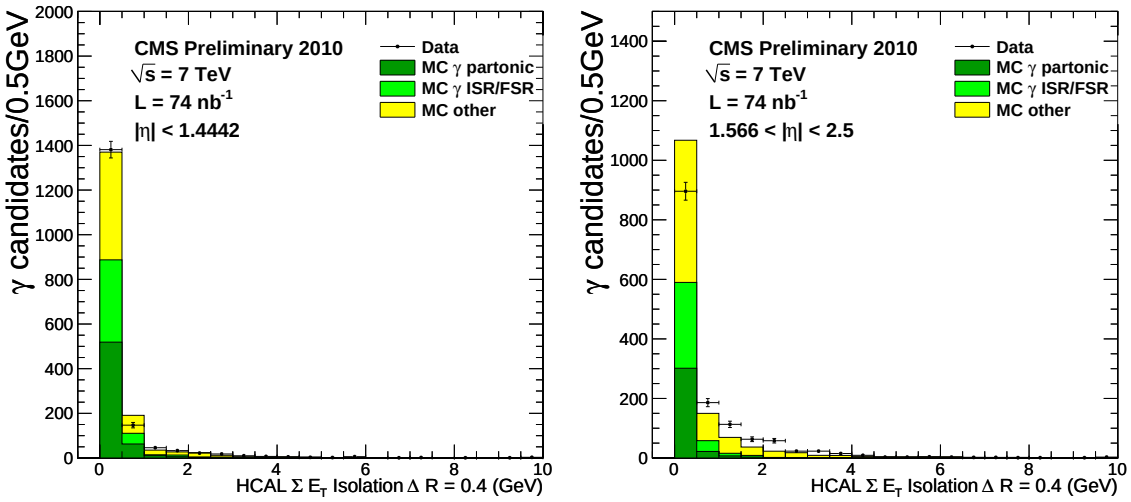


Figure 7: $N-1$ HCAL isolation distribution for data and MC, shown for barrel (right) and endcap (left). The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

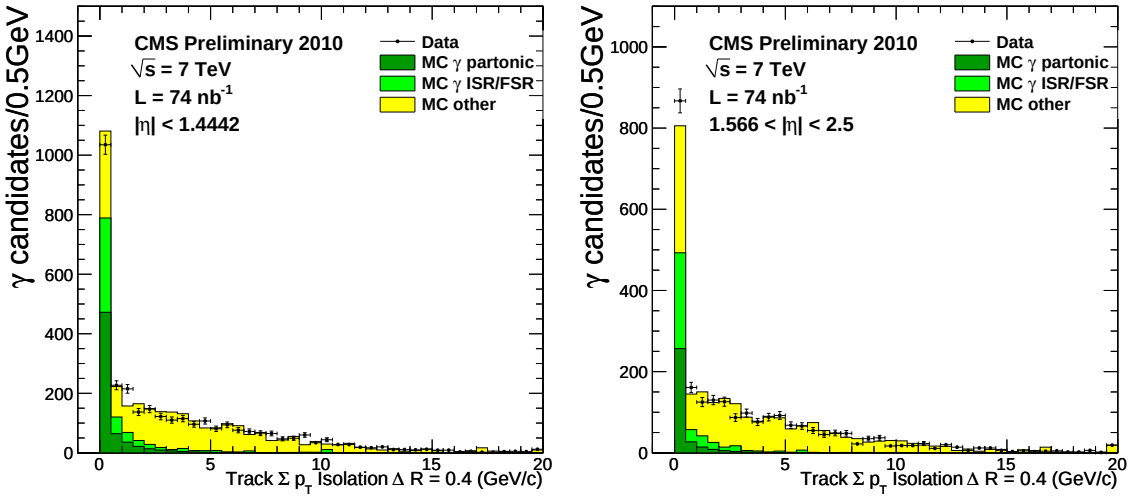


Figure 8: $N-1$ Track isolation distribution for data and MC, shown for barrel (right) and endcap (left). The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

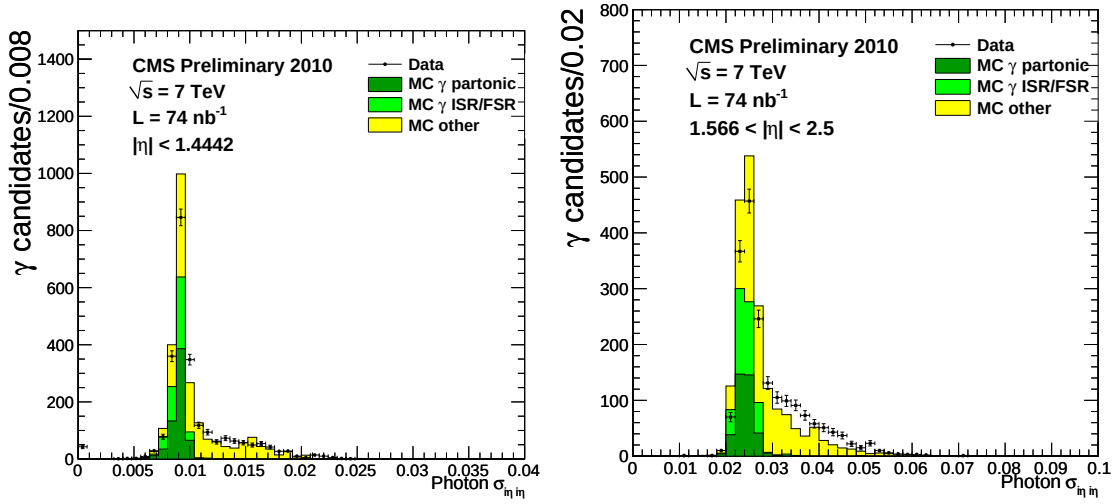


Figure 9: The $\sigma_{i\eta i\eta}$ shower shape variable for barrel and endcap photon candidates. The $N-1$ distributions are shown before cutting on the variables for photon identification. The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

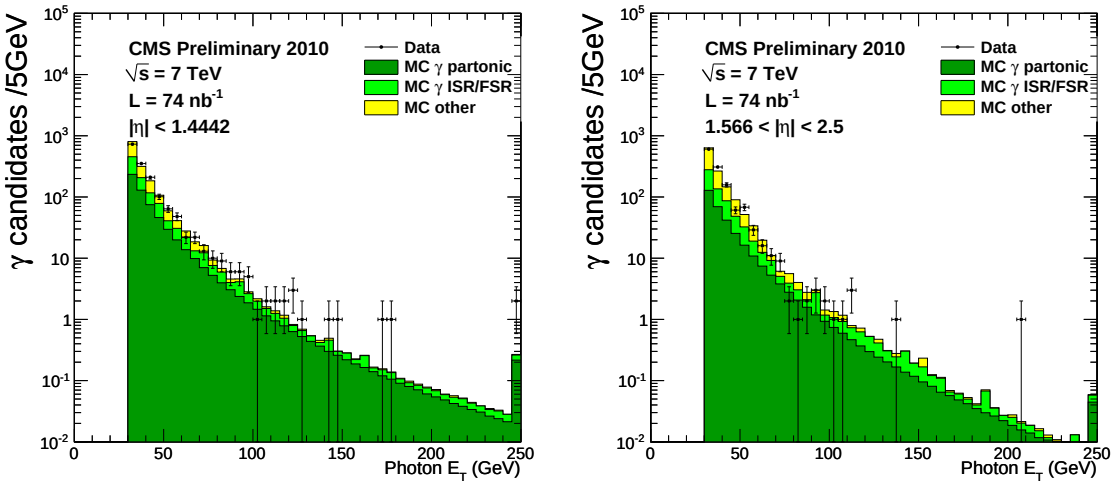


Figure 10: Transverse energies of EB (left) and EE (right) reconstructed photons in the selected sample. The purity increases with E_T. The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

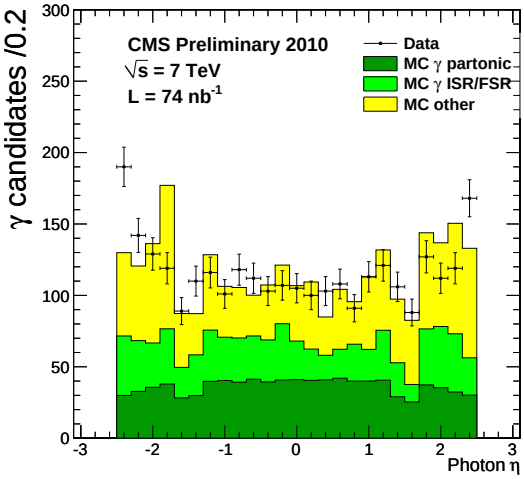


Figure 11: Pseudorapidity of reconstructed photons in the selected sample.

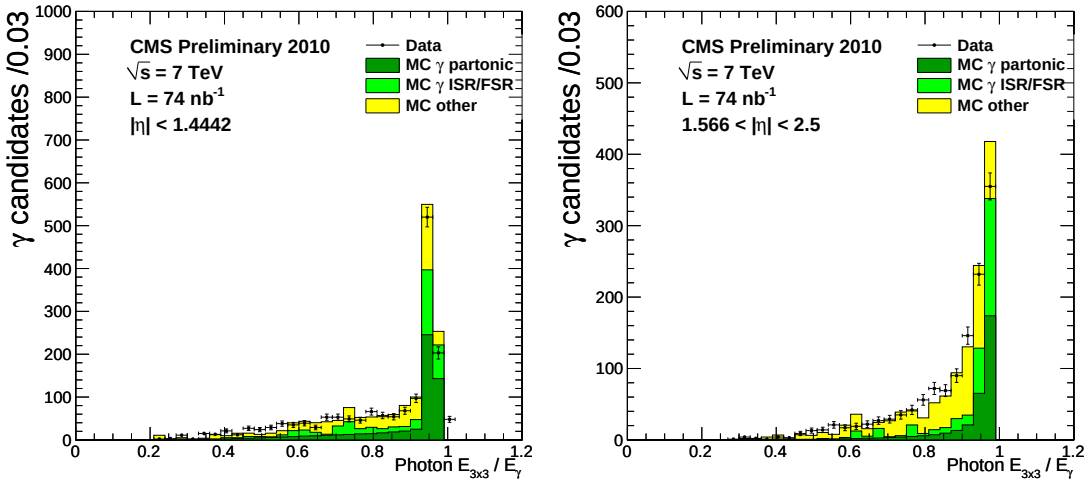


Figure 12: The $r9$ variable of reconstructed photons in the selected sample in EB and EE. The Monte Carlo results are normalized separately for each plot to the number of entries in the data histogram.

3 Photon conversions

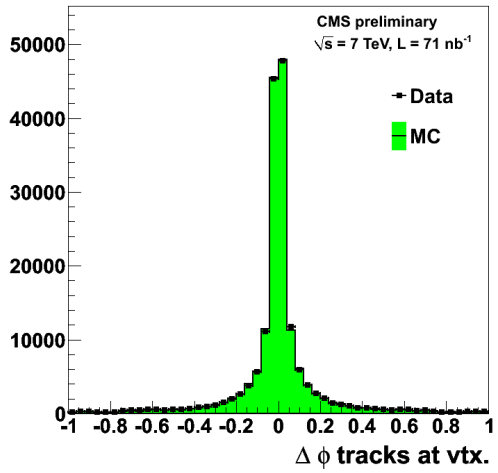
On average there is about one radiation length of material in front of the ECAL. The relatively massive CMS Tracker leads to a large probability of photon conversion and electron bremsstrahlung radiation in the tracker volume.

Conversion finding based on standard tracking ([6], [7]) optimizes the reconstruction of soft conversions which hardly reach the ECAL. Here we use an ECAL-driven conversion track finding which is mainly intended to address isolated converted photons [8]. The conversion reconstruction algorithm combines calorimetric and tracking information to achieve high efficiency throughout the CMS tracker volume. Data collected by CMS at 7 TeV, corresponding to 71 nb^{-1}, are compared to simulation. Conversion signal mostly comes from low p_T π^0s and constitutes background to all analyses involving isolated photons. Conversion candidates were required to have $p_T > 17$ GeV and to pass the HLT_Photon15_L1R trigger.

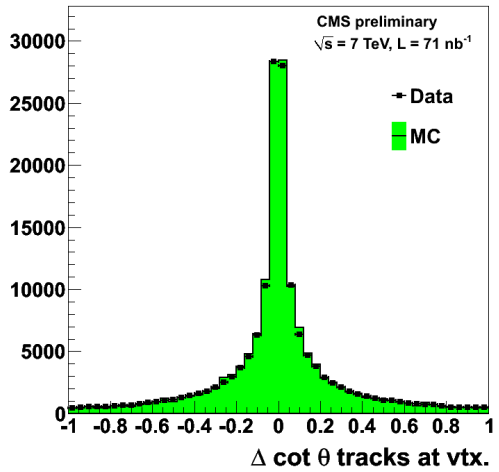
Conversion candidates are built from superclusters satisfying a loose preselection ($H/E < 0.5$ and $E_T > 2$ GeV). The constituent basic clusters are used as starting points for the inward conversion track finding, using the E_T of the basic cluster as an initial guess for the trajectory of the track. If a pair of matching hits are found in the last three layers of the tracker, they are used to re-evaluate the track parameters. Track finding then proceeds inward, considering the mean energy loss experienced by electrons in the tracker material. Once the inward ECAL-seeded tracks are built, their innermost hits are used as starting points for the other arm of the conversion. Inside-out tracks are then built outwards, again considering the mean energy loss for an electron. The possible combinations of oppositely charged tracks are selected and combined to form conversion candidates. A detailed description of the method is given in [8].

Photon conversions are characterized by a pair of oppositely charged tracks with small opening angle and consistent with zero invariant mass. The variables used to separate the conversion candidates from long-lived particles and vertices from mis-reconstructed tracks are the angular separation measured at the production vertex between the conversion electrons in the transverse plane ($\Delta\phi$) and in the longitudinal plane ($\Delta \cot\theta$) and the χ^2 of the vertex fit. The

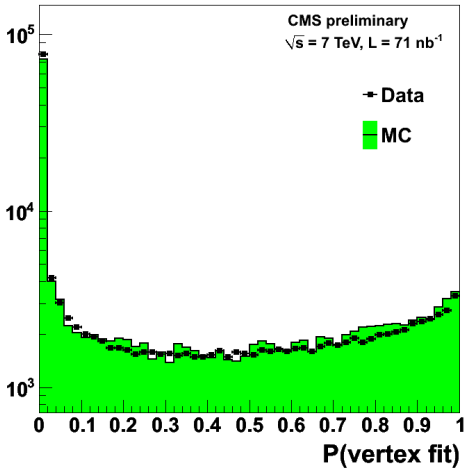
opening angle and χ^2 for the vertices before selection are shown in Fig. 13 for raw conversion candidates obtained in the reconstruction, before any additional selection is performed. The MC is normalized to the number of candidates in data.



(a)



(b)



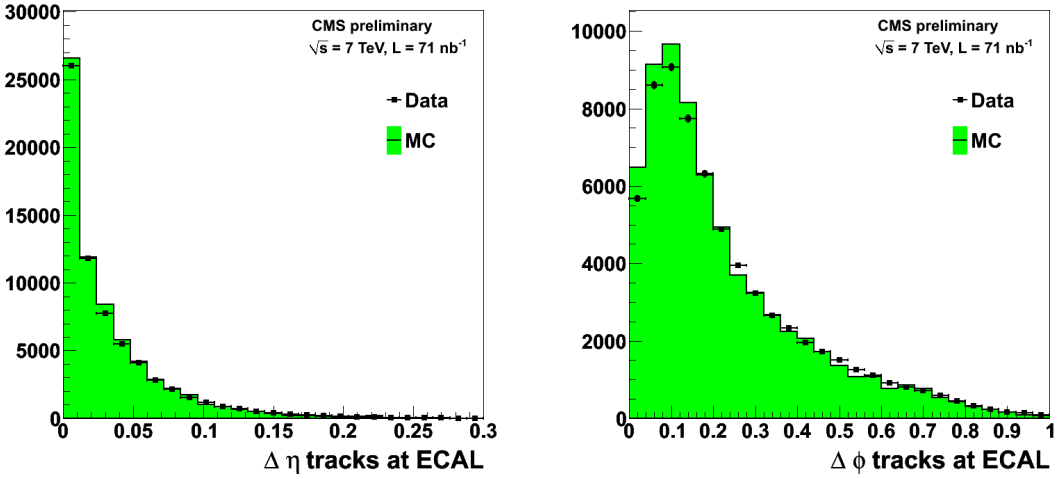
(c)

Figure 13: Angular separation between the conversion electrons measured in the longitudinal plane (a) and in the transverse plane (b); (c) χ^2 probability for the conversion vertex fit.

The selection then requires $|\Delta\phi| < 0.2$, $|\Delta \cot\theta| < 0.3$ and a valid vertex with χ^2 probability greater than 5×10^{-4}.

Distributions after the selection are shown in Fig. 14: the $|\Delta\eta|$ and $|\Delta\phi|$ of electrons is measured at ECAL. The $|\Delta\eta|$ between the electrons is small as expected from the parallel tracks while the $|\Delta\phi|$ shows the expected opening angle at the ECAL front face. The agreement between data and simulation is good.

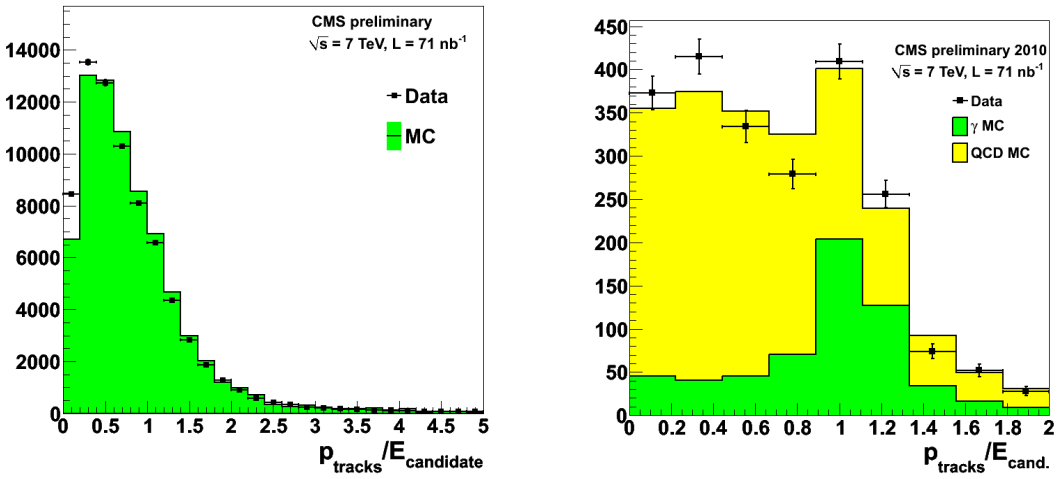
The p/E spectrum for conversions is shown in Fig. 15, both before and after applying the selection given in Table 1. This can be used to discriminate between isolated converted photons



(a) (b)

Figure 14: Angular separation between the conversion electrons at the impact with ECAL; (a) in the longitudinal plane and (b) in the transverse plane.

(where the distribution should peak at 1) and π^0s where only one of the two legs converts. The data and simulation are in good agreement.



(a) (b)

Figure 15: Conversion p/E (a) with p from the tracks and E from the super-cluster; the distribution is inclusive for barrel and endcap. Conversion p/E for candidates with $|\eta| < 1.45$, with the photon candidate selection from Table 1 applied (b). The MC distribution for QCD fakes and signal photons are shown separately.

4 Non-collision backgrounds

For signatures with photons and missing energy, another significant source of background are the bremsstrahlung photons emitted by high energy muons from cosmic rays or beam halo. The timing distribution of cosmic rays is uniform, while the beam halo is tightly correlated with the interaction time, making it more challenging to suppress once the instantaneous luminosity is high enough to produce multiple interactions per beam-crossing. Here, we will concentrate on the beam halo background.

We use two dedicated beam halo taggers to suppress this background and a data-driven technique to estimate its residual contribution. The taggers are based on hadronic calorimeter and muon system. Because halo muons traverse the detector parallel to the beam line, the halo muons' position in the hadronic endcap calorimeter (HE) and the electromagnetic barrel calorimeter (EB) are correlated. The selection for the HE halo tag is listed in Table 2(a). Similar to the HE halo tagger, the inner most positions of non-central tracks in the endcap muon system (EMU) can indicate halo bremsstrahlung events in the EB. The selection for the EMU halo tag is shown in Table 2(b).

The performance of taggers is calibrated using real data. We select events corresponding to the crossings of filled bunches, with photons with E_T above 15 GeV, passing the selection discussed in Table 1, except the cut on $\sigma_{i\eta i\eta}$. Instead, we require the ratio of the sum of the two most energetic crystals in the shower to the sum of the 5×5 crystal matrix be less then 0.95.

For "halo" sample, we require no reconstructed tracks in the event, and missing transverse energy above 25 GeV, since the beam halo events are expected to be unbalanced. The photon candidates are required to be tagged by either HE or EMU tagger.

"Prompt" sample is selected by requiring $E_T^{miss} < 15$ GeV. Photon candidates are required to have no EB-HE or EMU halo tags, and |photon seed time| < 3 ns. Finally, we require the events to have more than three reconstructed tracks in the event. This selection minimizes any possible halo contamination.

We define the "candidate" sample as events with at least three reconstructed tracks, $E_T^{miss} > 25$ GeV, containing a photon with $E_T > 30$ GeV, not tagged by wither tagger, with |photon seed time| < 3 ns.

Halo events typically have negative timing in ECAL with respect to prompt events, due to the fact that halo crosses the ECAL in time with the bunches but does not have to traverse the extra distance from the interaction point to ECAL as prompt particle does. Figure 16 shows the photon arrival time for the three samples.

Table 2: Halo Tag Definitions

(a) HE Tag

$E_{HE-hit} > 1$ GeV
115 cm $< \rho_{HE-hit} <$ 130 cm
$\Delta\phi_{photon,HE-hit} < 0.2$

(b) EMU Tag

115 cm $< \rho_{EMU-track} <$ 170 cm
$\Delta\phi_{photon,EMU-track} < 0.2$

Assuming the shape of the photon shower to be elliptical, we calculate major and minor axes of the shower in $\eta - \phi$ space. The discriminating variable is the angle A which the major axis of the shower makes with the beam direction (η-axis). Prompt showers spread dominantly along ϕ, $A \sim \pi/2$ and halo bremsstrahlung showers spread dominantly along the beam direction, $A \sim 0$. The distributions of this angle A in the "halo" and "prompt" samples form templates that are fit to the candidate sample as shown in Fig. 17. We estimate the amount of halo

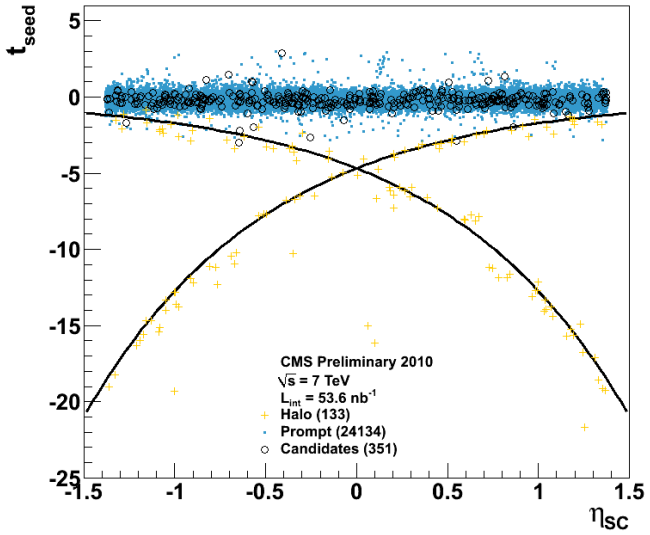


Figure 16: Data. Super cluster position vs. time of photon seed. Prompt and candidate events arrive at the EB at $t = 0$. Halo events primarily arrive out-of-time with respect to prompt events. The black curves show the expected arrival time for halo particles due to the path length difference to reach ECAL.

contribution to the "candidate" sample below 5.9 events at 95% CL.

5 Summary

Using the first 7 TeV LHC data delivered to CMS, we have carried out a series of studies on photon candidates. The basic reconstruction quantities for photons have been compared to the Monte Carlo simulation with good agreement. Using a selection which is intended to enrich the sample in signal photons and suppress the background from QCD, we have illustrated that the photon reconstruction and identification performance is similar to that expected from simulation. First results on the ECAL-seeded conversion finding also show good agreement with simulation. Finally, we have shown that the number of fakes due to non-collision backgrounds in the sample is under control.

The commissioning of the photon object at CMS is still in progress. Future results with a large sample of high p_T candidates and control samples of high-PT electrons will improve on the results shown here and lessen the dependence on simulation.

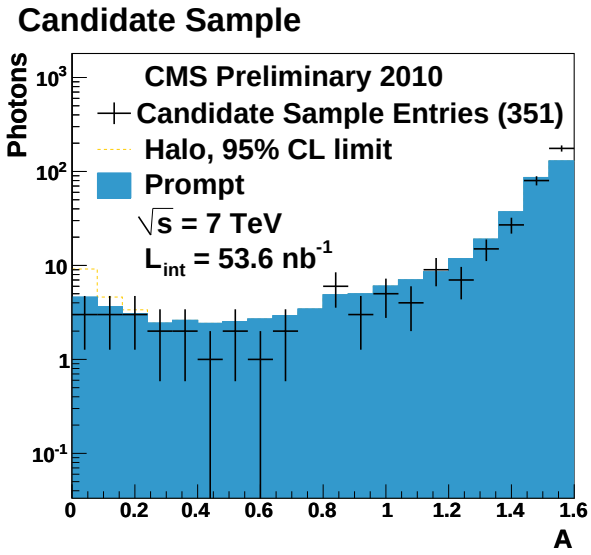


Figure 17: Black crosses are the shower's angle A of the candidate sample. The prompt template (blue) is shown stacked with the 95% CL upper limit contribution from halo (gold).

References

[1] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[2] CMS Collaboration, "Detector performance and software", *CMS Physics Technical Design Report* **Vol. 1** (2006).

[3] CMS Collaboration, "Electron reconstruction and identification at sqrt(s) = 7 TeV", *CMS Physics Analysis Summary* **EGM-10-004** (2010).

[4] CMS Collaboration, "Electromagnetic calorimeter commissioning and first results with 7 TeV data", *CMS Physics Analysis Summary* **EGM-10-002** (2010).

[5] P. T. Sjostrand, S. Mrenna, "PYTHIA 6.4 Physics and Manual; v6.420, tune D6T", *JHEP* **05** (2006).

[6] CMS Collaboration, "Tracking and Vertexing Results from First Collisions", *CMS Physics Analysis Summary* **TRK-10-001** (2010).

[7] CMS Collaboration, "Studies of Tracker Material", *CMS Physics Analysis Summary* **TRK-10-003** (2010).

[8] Marinelli N., "Track finding and identification of converted photons", *CMS Note* **2006/005** (2006).

A Additional plots

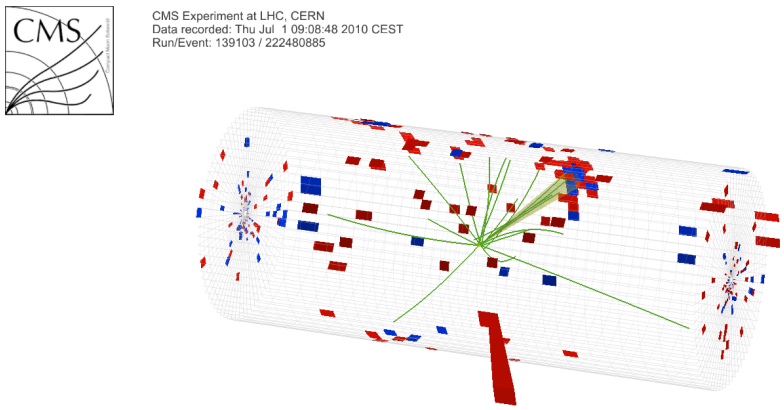


Figure 18: Event display of a photon candidate passing the selection. The photon candidate points out of the page and is isolated in ECAL (red), HCAL (blue) and the tracker (green).

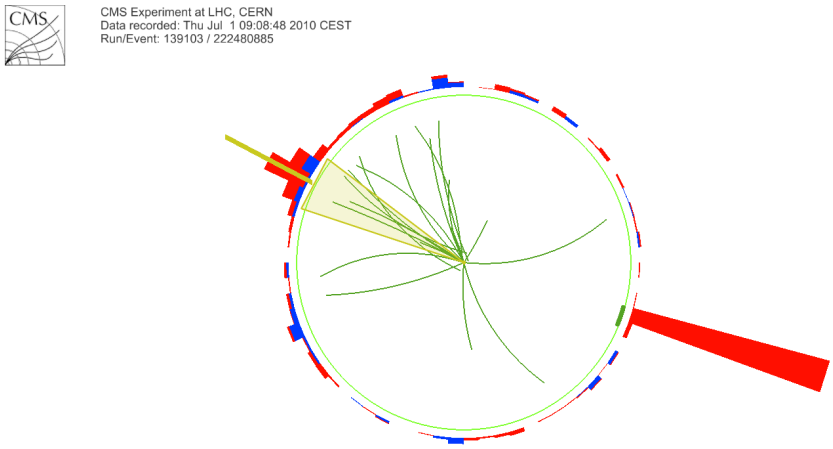


Figure 19: R-ϕ view of the same event. The photon candidate is balanced by a jet on the other side of the detector.

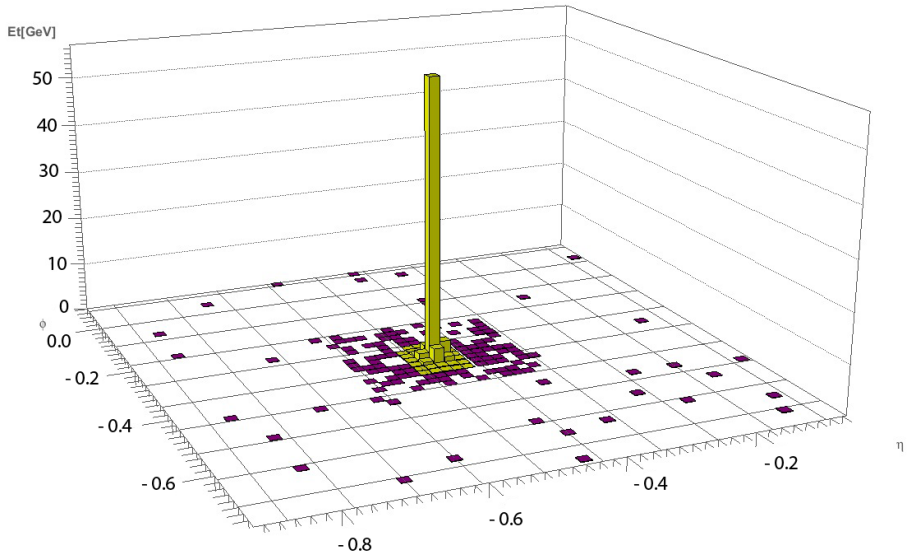


Figure 20: Detail view of the photon candidate in the same event. Clustered ECAL crystals are shown in yellow, unclustered energy in violet. The candidate has $r9 > 0.94$ and is considered to be unconverted.

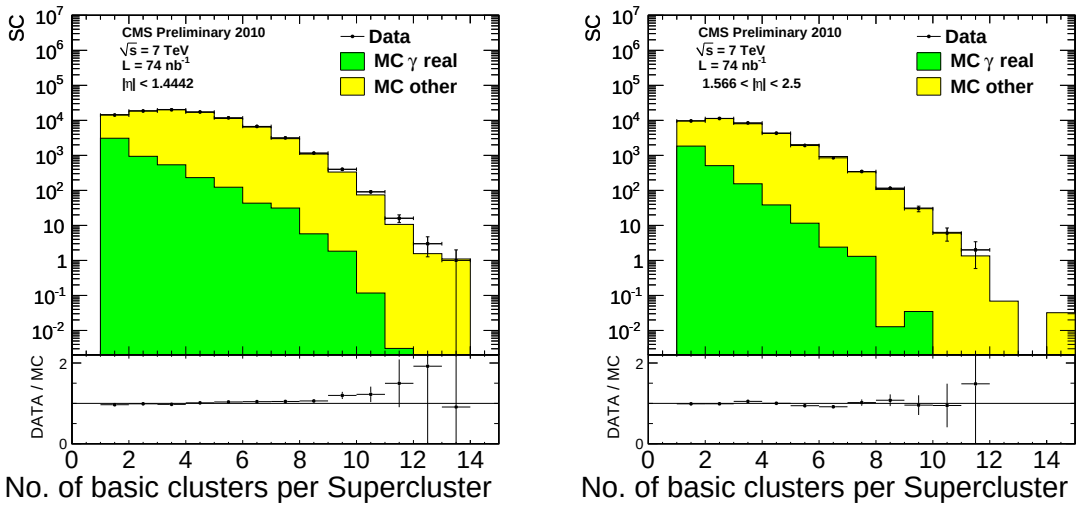


Figure 21: Number of basic clusters included in EB (left) and EE (right) superclusters

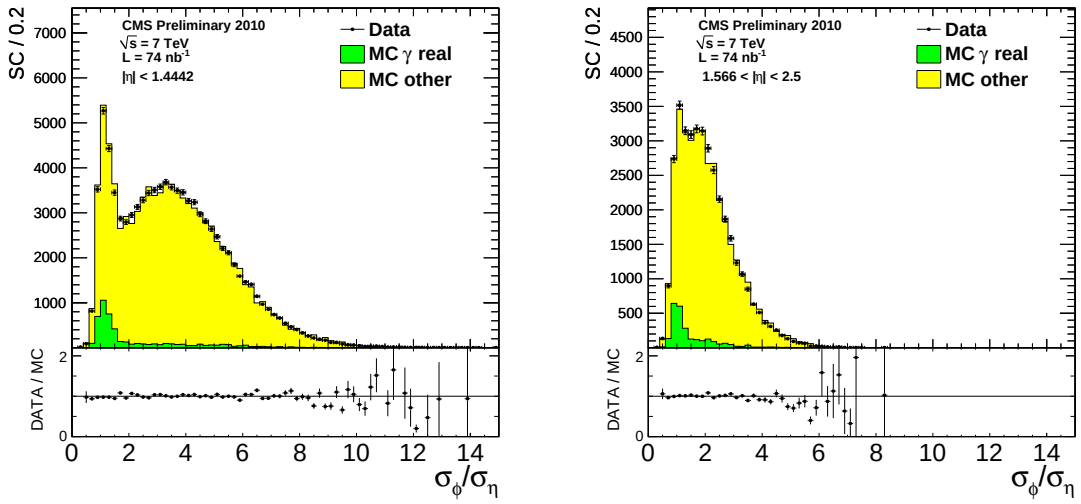


Figure 22: The ϕ/η width ratio used for energy corrections in EB (left) and EE (right).