

# Chapter 1

## Event Selection

In keeping with the phenomenology described in Sec. ??, the candidate GGM events selected in this search consist of two high- $E_T$  photons and a significant momentum imbalance transverse to the beam, indicating the production of an escaping gravitino. This momentum imbalance is usually referred to as *missing transverse energy* and is denoted by the symbols  $\cancel{E}_T$  or  $\text{ME}_T$ . The GGM signature is shown in Figure 1.1.

However, in order to use real CMS data (as opposed to simulation) to derive predictions for the backgrounds to the search, *control samples* that are not expected to contain any GGM signal events and are distinct from the *candidate* two-photon sample must be collected from the LHC data. These samples consist of different numerical combinations of photons, electrons, and jets, and are explained in more detail in Chapter ?. Since this search is performed in the high- $\cancel{E}_T$  tail of the  $\cancel{E}_T$  distribution, where adequate detector simulation is very difficult, it is advantageous to use *data-driven* background estimates, which capture the true detector response, over numbers derived from simulation.

In the following sections, the reconstruction of photons, electrons, jets, and  $\cancel{E}_T$  is explained. Sec. 1.1 begins with an explanation of the high level reconstruction. It is followed by Sec. 1.2, which describes the triggers used to collect the candidate

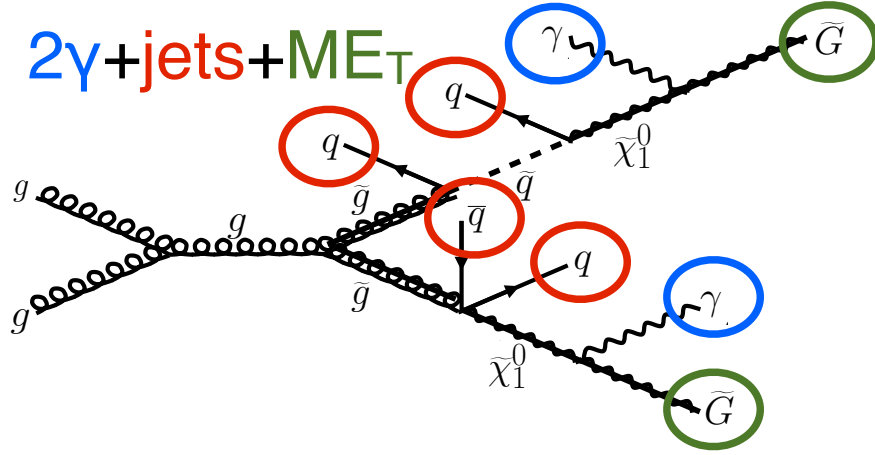


Figure 1.1: Two gluinos each decay via  $\tilde{\chi}_1^0 \rightarrow \gamma \tilde{G}$ .

and control samples. Sec. 1.3 describes event cleaning cuts that are applied to the candidate and control samples. Finally, the chapter concludes with a measurement of the photon identification efficiency in Sec. 1.4.

## 1.1 Object Reconstruction

This section describes the *offline* object reconstruction, i.e. the reconstruction of particle objects from events that have already been triggered and written to permanent storage, as opposed to the building of trigger objects explained in Secs. ?? and 1.2.

### 1.1.1 Photons

#### Uncalibrated EB/EE Hits

Photon reconstruction begins with the ADC count value for each of the 10 recorded time samples per ECAL crystal per trigger. To construct an *uncalibrated hit*, the gain (1, 6, or 12; see Sec. ??) of each sample is determined and the ADC count value scaled appropriately. The pedestal is estimated from the average of the first three samples, which, for a properly timed in hit, should contain no signal. This pedestal

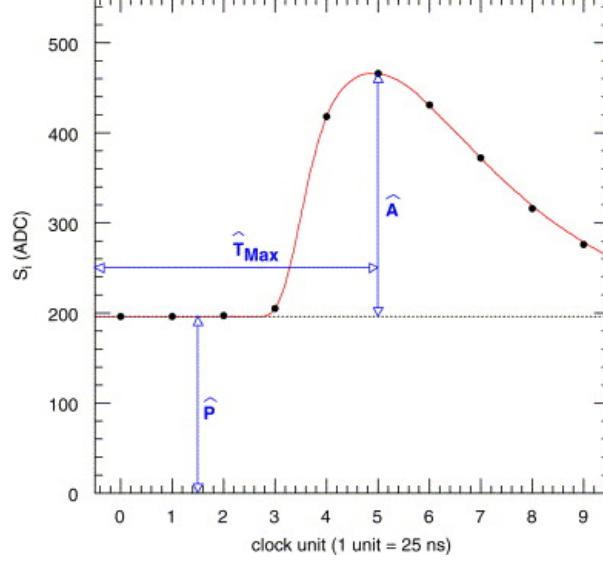


Figure 1.2: Typical ECAL channel pulse shape.  $\hat{P}$  is the pedestal value,  $\hat{A}$  is the pulse amplitude, and  $\hat{T}_{\max}$  is the hit time. The red line is the assumed pulse shape from which the weights are derived. Reprinted from ref. [? ].

value is subtracted from the rest of the samples. Finally, the amplitude of the pulse is reconstructed using a predetermined weight for each sample [? ]. The weights correspond to the pulse shape expected from the MGPA and shaping circuit response. The time of the hit is also reconstructed using the ratios between neighboring time samples [? ]. A typical ECAL channel pulse shape is shown in Figure 1.2.

### Calibrated EB/EE Hits

In the next phase of the photon reconstruction, calibrations are applied to the uncalibrated hits to form *calibrated hits* with energy measured in GeV. Channels are excluded from seeding calibrated hits if

- they are excessively noisy,
- they are stuck in fixed gain (i.e. the MGPA gain does not change properly to avoid saturation),
- they are totally dead,

- they have one or more neighboring dead channels, or
- they do not have good trigger primitives (i.e. trigger primitive is missing, saturated, or *spike-like*; cf. Secs. ?? and ??).

*ECAL spikes* are hits in which low energy protons and heavy ions from jets ionize in the sensitive volume of the EB APD (Sec. ??), causing the APD to register a fake large-amplitude hit. Because they are not the result of a real electromagnetic shower, spikes tend to be isolated. They may also appear to arrive early or late with respect to the nominal bunch crossing. Most spikes are reconstructed with a hit time  $\sim 10$  ns earlier than real EM hits because unlike real hits, whose pulse shapes include the time constant associated with crystal scintillation, the reconstructed spikes only involve the rise time of the electronics. There also is a long tail of late arriving spikes due to slow neutrons from jets [? ].

Because of their particular timing and topological characteristics, cuts have been developed to effectively identify and reject spike-like hits. This analysis utilizes both the “Swiss cross” cut  $1 - E_4/E_1 > 0.95$ , where  $E_1$  is the energy of the spike candidate crystal and  $E_4$  is the sum of the energies in the four crystals whose edges are next to the four edges of the spike candidate crystal, and a timing cut  $t \geq 3$  ns, to flag spikes. More information about these cuts can be found in ref. [? ]. A simpler algorithm using the fine grain veto bit of the ECAL L1 TPG (Sec. ??) is used to reject spikes at the trigger level.

In addition to the trigger primitives, no uncalibrated hits that are spike-like are eligible for calibration. The calibrations applied are corrections to account for crystal transparency and VPT photocathode loss (leading to signal loss), measured continuously by the laser/LED system; energy intercalibrations (relative energy calibration between crystals); absolute scale calibrations between ADC counts and GeV;<sup>1</sup> and

---

<sup>1</sup>The ADC-GeV scale factors (one for EB and one for EE) are defined such that the sum of fully calibrated and scaled hits in a particular  $5 \times 5$  cluster of crystals (plus the associated energy deposited in ES) is 50 GeV for a 50 GeV incident unconverted photon [? ].

time intercalibrations (relative time calibration between crystals).

The ECAL crystals were pre-calibrated before installation in CMS using laboratory light yield and photodetector gain measurements [? ]. In addition, some EB and EE crystals were intercalibrated using test beams [? ], and all EB crystals were intercalibrated with cosmic ray muons [? ]. EE precalibrations were validated with LHC *splash events* in 2009 [? ? ], in which the beam was dumped onto a collimator approximately 150 meters upstream of CMS, causing a spray of muons to enter CMS at one endcap and exit at the other. Splash events were also used to derive time intercalibration constants. Before colliding beam operations commenced, the intercalibration precision was estimated to be 0.5%-2.2% in EB and 1%-5% in EE [? ].

Three calibration methods were employed once colliding beam operations began:

- $\phi$  symmetry relative calibration between crystals, exploiting the azimuthal symmetry of CMS
- $\pi^0$  and  $\eta$  relative calibration between crystals, using the diphoton decays of these particles
- $E/p$  absolute calibration, comparing the momentum measured in the tracker  $p$  to the energy measured in the ECAL  $E$  of a sample of electrons from  $Z$  decay

By September 2011, the intercalibration precision in EB was measured to be between 0.3% and 1.1% using the  $\pi^0/\eta$  method [? ]. Figure 1.3 shows the improvement in  $Z$  reconstruction from pre-LHC calibration constants to the latest  $\pi^0/\eta$ -derived constants.

### Calibrated ES Hits

ES calibrated hits are formed from the three samples read out per sensor. Just as in the case of EB/EE crystals, ES uncalibrated hits are gain-adjusted, pedestal-subtracted,

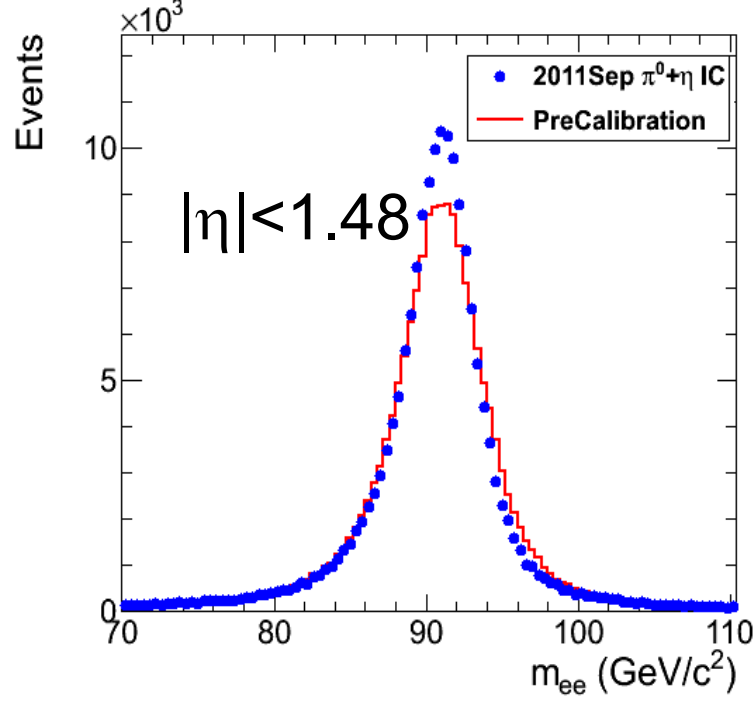


Figure 1.3:  $Z$  peak reconstructed using pre-LHC calibration constants (red) or September 2011  $\pi^0/\eta$ -derived intercalibration constants (blue). Reprinted from ref. [? ].

and reconstructed using weights. To make a calibrated ES hit, intercalibration constants, angle correction constants (for the non-uniformity of sensor angle with respect to the vertical across ES), and a MIP-GeV absolute scale factor are applied.

### Clustering

After calibrated ECAL hits are formed, they must be clustered into shapes that represent the energy deposit from a single particle. *Basic clusters* are formed around seed hits, defined as a hit that

- has calibrated  $E_T > 1(0.18)$  GeV in EB(EE),
- does not originate from a dead channel or one with faulty hardware,
- is not poorly calibrated,

- was reconstructed with the standard algorithm (i.e. not a special recovery algorithm for channels with subpar data integrity),
- is not saturated,
- is not spike-like, and
- is in time (EB).

EB basic clusters are formed around the seeds via the *hybrid* algorithm, while EE basic clusters are formed with the `multi5x5` algorithm [? ]. In addition to handling non-radiating electrons and unconverted photons, both algorithms are designed to also recover all of the energy associated with electron bremsstrahlung deposits and photon conversions. The geometry of the CMS magnetic field means that bremsstrahlung and conversions will tend to spread the shower out in  $\phi$ , not  $\eta$ . Both algorithms work by forming basic clusters around seeds, then combining the basic clusters into *superclusters* (SC) by searching in a window extended in the  $\phi$  direction for all basic clusters consistent with bremsstrahlung radiation from the primary electron, or with a photon conversion. Figure 1.4 illustrates the hybrid algorithm in EB. In EE, the energy deposited in ES must also be added into the total clustered energy sum.

Figure 1.5 shows the effect of superclustering on  $Z \rightarrow ee$  reconstruction.

## Supercluster Corrections

The total clustered ECAL energy is defined as

$$E = F \times \sum_{i=1}^{n_{\text{crystal}}} G \times c_i \times A_i \quad (1.1)$$

where  $G$  is the ADC-GeV or MIP-GeV scale factor,  $c_i$  are the intercalibration constants,  $A_i$  is the uncalibrated hit amplitude in ADC counts, and  $F$  is a SC correction

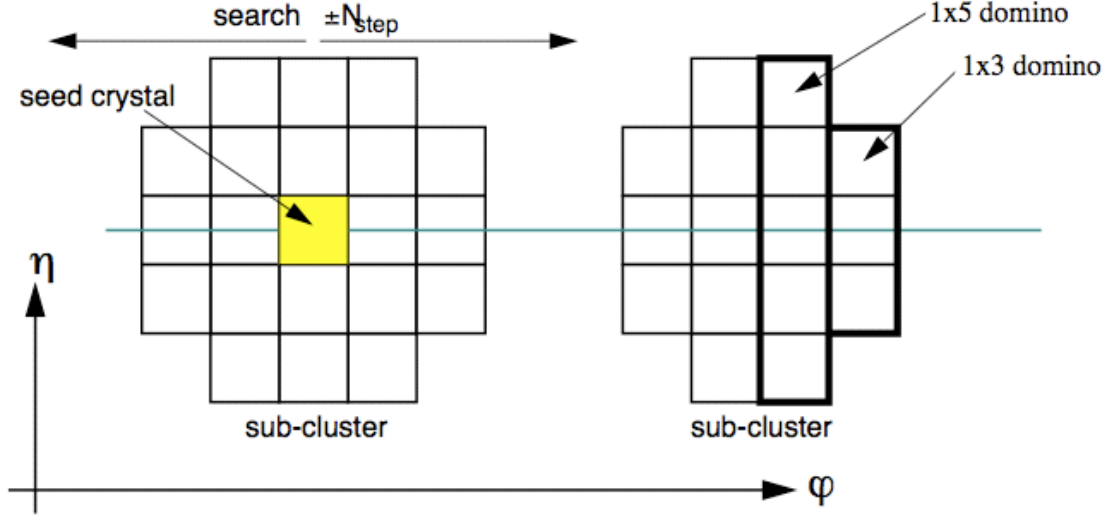


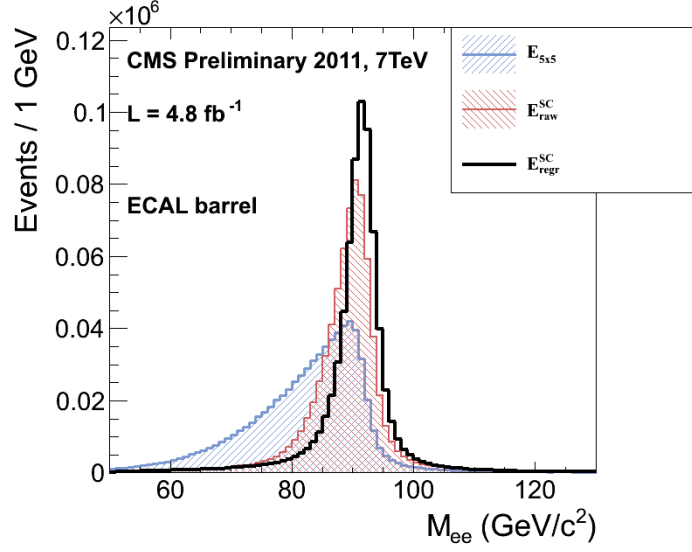
Figure 1.4: Hybrid algorithm in EB. The shower extent is constant in  $\eta$  (five crystals), but spreads out in  $\phi$  as the two sub-clusters (or basic clusters) are grouped into the same supercluster. The maximum extent in  $\phi$  is 17 crystals. Reprinted from ref. [? ].

factor.  $G$  and  $c_i$  were explained in Sec. 1.1.1.  $F$  is a product of three factors for hybrid SCs (two for multi5x5 SCs in EE, items 2 and 3 below) [? ]:

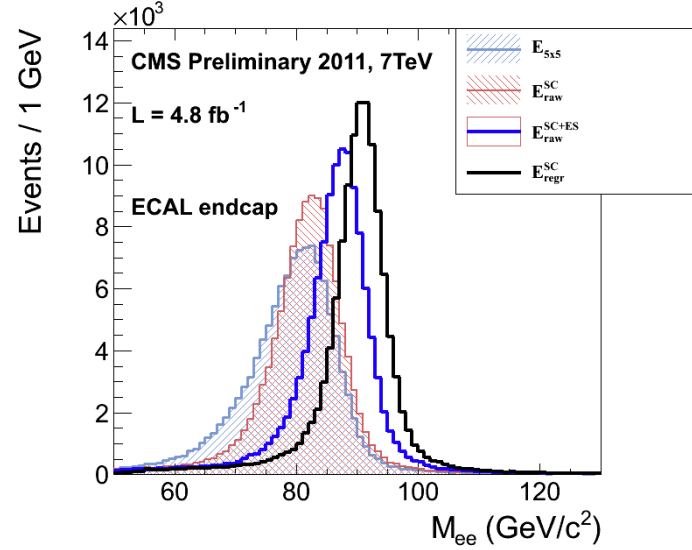
1.  $C_{EB}(\eta)$ , which compensates for lateral energy leakage due to the crystal off-pointing in EB only. These corrections are taken from MC simulation [? ] and were confirmed in test beams [? ].
2.  $f(\text{brem})$ , which corrects for biases in the clustering algorithms for showers characterized by differing amounts of bremsstrahlung. These corrections are taken from MC simulation [? ].
3. Residual correction  $f(E_T, \eta)$ , due to the variation in  $\eta$  of detector material traversed by a primary electron or photon, and to any residual  $E_T$  dependence of the reconstruction. These corrections are determined from MC and validated on  $Z \rightarrow ee$  data samples [? ].

As a benchmark of ECAL calibration performance, the extra energy smearing in MC needed to achieve data/MC agreement in the  $Z$  width was between  $\sim 0.9\%$  (in





(a) EB.



(b) EE.

Figure 1.5:  $Z$  peak reconstructed in the dielectron channel for different kinds of clustering. The constituent hits were calibrated with the best available intercalibrations and laser calibrations as of December 2011. The light blue histogram shows the reconstruction using a  $5 \times 5$  energy sum, the red histogram shows the reconstruction using the SC energy for crystals only (the dark blue histogram in the EE plot adds in the energy from ES), and the black histogram shows the reconstruction after the SCs are corrected using a multivariate method [? ]. Reprinted from Fig. 30 of ref. [? ].

the central part of EB for electrons with little bremsstrahlung) and  $\sim 3.3\%$  (in the outer part of EE for heavily radiating electrons) [? ].

### From Supercluster to Photon

The CMS photon object is any SC with  $E_T > 10$  GeV and  $H/E < 0.5$ , unless the SC  $E_T > 100$  GeV, in which case the  $H/E$  requirement is dropped.  $H/E$  is defined as the ratio of energy in the HCAL in a  $0.15$  cone around the SC centroid, directly behind the SC, to the SC energy. SCs with  $R9 > 0.94(0.95)$  in EB(EE), where  $R9$  is defined as the ratio of the energy in the central  $3 \times 3$  cluster of crystals divided by the SC energy  $E_{3 \times 3}/E_{SC}$ , are the best calibrated and most accurate type of electromagnetic shower. Therefore, for these objects, the photon energy is defined as the energy sum of the fully calibrated hits in the central  $5 \times 5$  cluster around the seed (with  $C_{EB}(\eta)$  applied for EB photons). For all other SCs, the photon energy is equal to the fully corrected SC energy (cf. Sec. 1.1.1).

In this search, candidate photons and *fake photons* ( $f$ , “fakes”) are further selected according to the criteria listed in Table 1.1. Fakes are used in the determination of the QCD background, as explained in Chapter ??.

Table 1.1: Selection criteria for photons and fakes. “Pixel seed,”  $I_{\text{comb}}$ , and  $\sigma_{i\eta i\eta}$  are defined in the text.

Variable	Cut ( $\gamma$ )	Cut ( $f$ )
SC $ \eta $	$< 1.4442$	$< 1.4442$
$H/E$	$< 0.05$	$< 0.05$
$R9$	$< 1$	$< 1$
Has pixel seed	No	No
$I_{\text{comb}}, \sigma_{i\eta i\eta}$	$< 6$ GeV AND $< 0.011$	$(\geq 6$ AND $< 20$ GeV) OR $\geq 0.011$

$I_{\text{comb}}$  is defined as

$$I_{\text{comb}} = I_{\text{ECAL}} - 0.093\rho + I_{\text{HCAL}} - 0.0281\rho + I_{\text{track}} \quad (1.2)$$

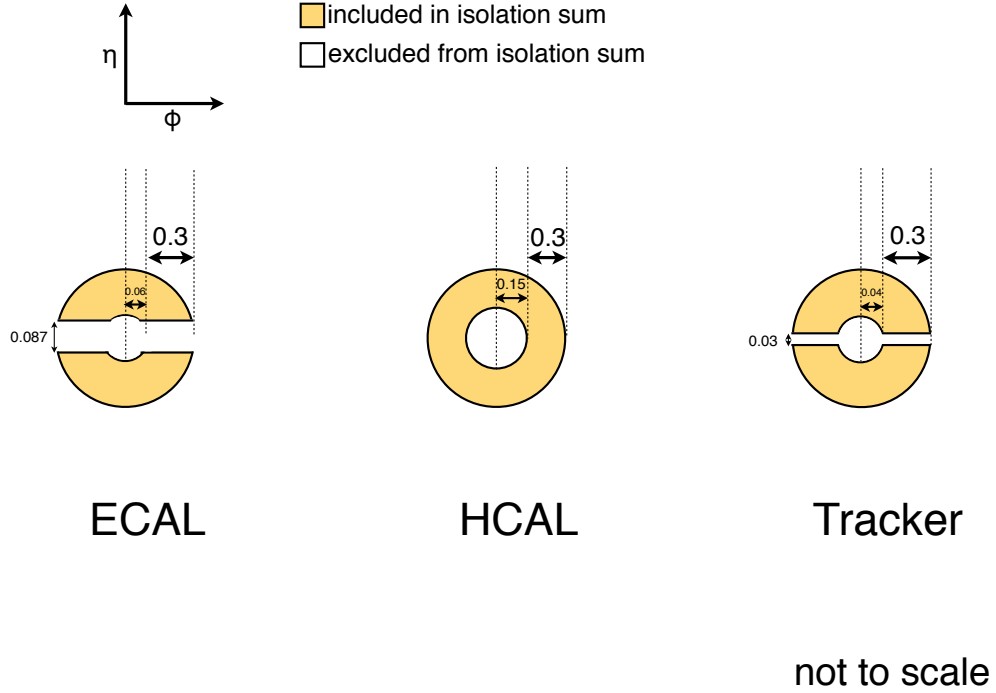


Figure 1.6: ECAL, HCAL, and track isolation cones.

where  $I_{\text{ECAL}}$ ,  $I_{\text{HCAL}}$ , and  $I_{\text{track}}$  are  $E_T$  sums in the annular regions defined in Figure 1.6 and  $\rho$  is the average pileup energy density in the calorimeters (per unit  $\eta \cdot \phi$ , or unit calorimeter surface area) as measured with the Fastjet algorithm [? ?].  $\rho$  is constant over the  $\eta$  range of the calorimeter. Note that the ECAL and track isolation veto strips at constant  $\eta$  ensure that the isolation cuts are similarly efficient for converted photons, radiating electrons, and unconverted photons.

$\sigma_{i\eta i\eta}$  is the log energy weighted extent of the shower in  $\eta$  and is defined as

$$\sigma_{i\eta i\eta} = \frac{\sum_{i=1}^{25} w_i (\eta_i - \bar{\eta})^2}{\sum_{i=1}^{25} w_i} \quad (1.3)$$

where the sums run over the  $5 \times 5$  matrix of crystals surrounding the seed,  $w_i = \max(0, 4.7 + \ln(E_i/E))$ ,  $E_i$  is the energy of the  $i^{\text{th}}$  crystal,  $E$  is the total energy in the

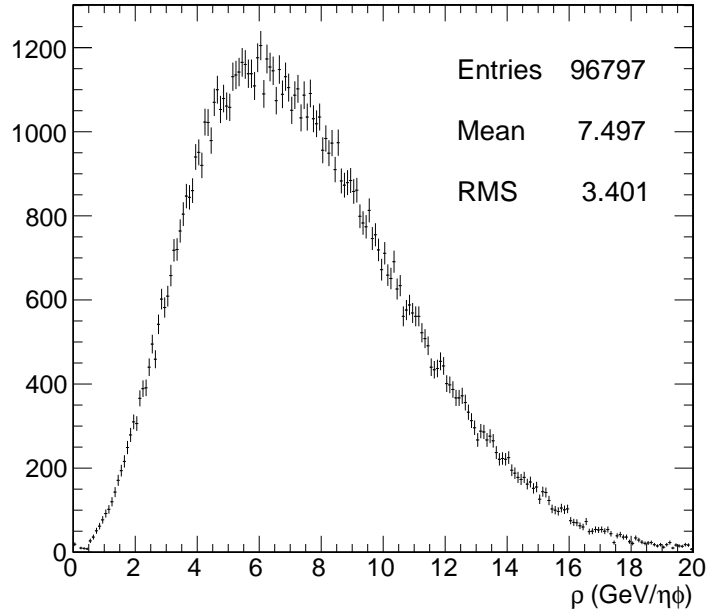


Figure 1.7:  $\rho$  distribution for a sample of two-photon events, with at least one 40 GeV and one 25 GeV photon, passing the selection requirements in Table 1.1 and the trigger requirements in Table 1.3. This sample covers the full 2011 dataset of  $4.7 \text{ fb}^{-1}$ .

25 crystals,  $\eta_i$  is the offset in  $\eta$  of the  $i^{\text{th}}$  crystal from the seed, and  $\bar{\eta}$  is the weighted average  $\eta$  of the 25 crystals (using the  $w_i$  as weights) [? ].

Figure 1.7 shows the  $\rho$  distribution for a sample of two-photon events, with at least one 40 GeV and one 25 GeV photon, passing the selection requirements in Table 1.1 and the trigger requirements in Table 1.3. This sample represents the full 2011 dataset of  $4.7 \text{ fb}^{-1}$ . Since the average  $\rho$  is  $\sim 7.5 \text{ GeV}$ , and there is a long tail above this average value, it is necessary to subtract pileup energy from the ECAL and HCAL isolation cones to recover otherwise clean photons in events with large pileup. The ECAL and HCAL *effective areas* of 0.093 and 0.0281, respectively, are calculated by fitting the average ECAL or HCAL isolation energy vs.  $\rho$  in a sample of  $Z \rightarrow ee$  events to a straight line. The slope of the line—which has the units of  $\eta \cdot \phi$ , or area—is the effective area.

The cut on combined isolation of 6 GeV (Table 1.1) is the result of an  $S/\sqrt{B}$

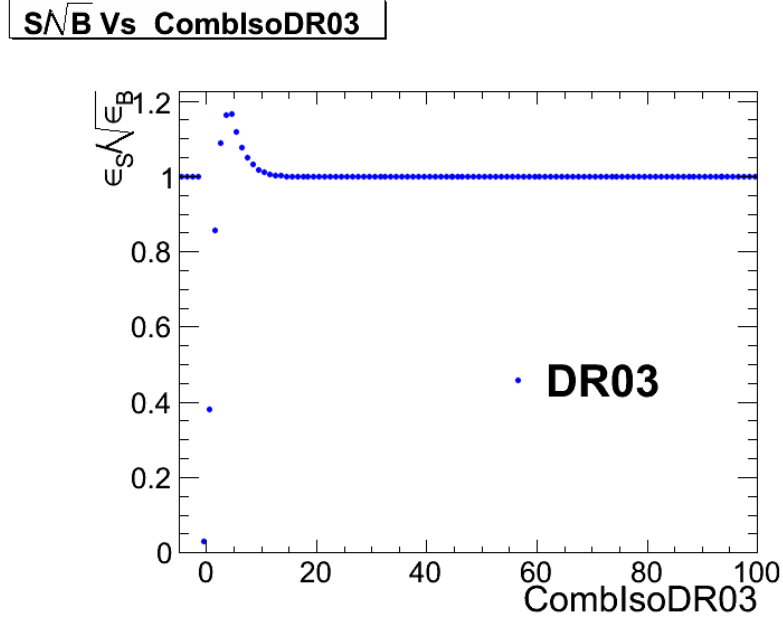


Figure 1.8:  $S/\sqrt{B}$  ( $S$  and  $B$  defined in the text) vs.  $I_{\text{comb}}$ . “DR03” in the legend indicates that this combined isolation was calculated in  $\Delta R = 0.3$  cones, as used throughout this analysis. Reprinted from Fig. 7 of ref. [? ].

optimization procedure [? ].  $S$  is a sample of photons in simulated GGM events that are products of neutralino decay, while  $B$  is a sample of photons matched to generated hadronic jets in simulated QCD events. Figure 1.8 shows the value of  $S/\sqrt{B}$  vs. combined isolation, in particular the pronounced peak around 6 GeV.

The upper bound on fake photon combined isolation guarantees that poorly isolated dijet events, with  $\cancel{E}_T$  resolution dissimilar to the candidate diphoton events, do not enter the  $\cancel{f}\cancel{f}$  sample. The exact value of 20 GeV (cf. Table 1.1) arises from a low- $\cancel{E}_T$   $\cancel{f}\cancel{f}/\gamma\gamma$   $\chi^2$  optimization procedure [? ]. Figure 1.9 shows the value of the Neyman’s  $\chi^2$  between the  $\cancel{f}\cancel{f}$  and  $\gamma\gamma$   $\cancel{E}_T$  distributions, truncated at either 25 or 50 GeV, vs. upper bound on fake combined isolation. As shown in the figure, 20 GeV keeps the  $\chi^2$  small while also being large enough that a sufficient number of  $\cancel{f}\cancel{f}$  events may be collected.

Finally, a “pixel seed” is defined as a hit in the pixel detector consistent with a

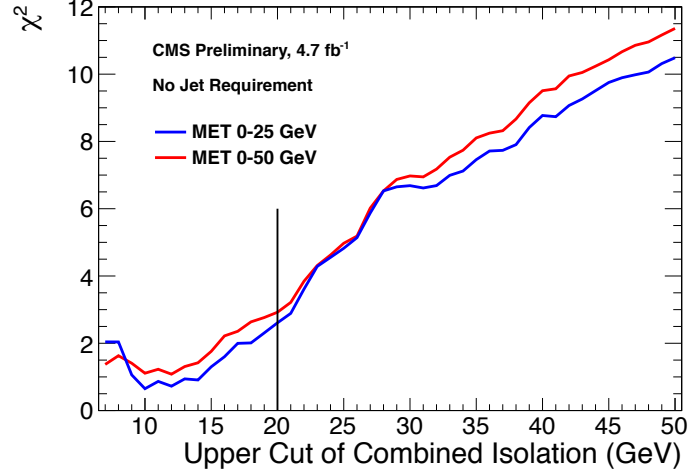
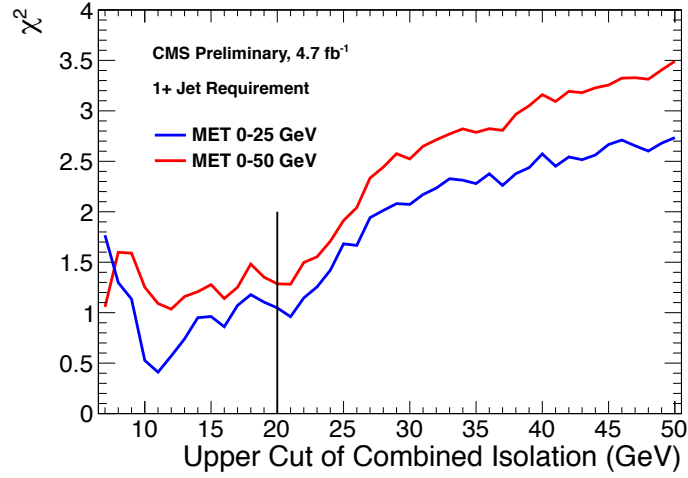
(a) Events with  $\geq 0$  jets.(b) Events with  $\geq 1$  jet.

Figure 1.9: Neyman's  $\chi^2$  between the  $ff$  and  $\gamma\gamma \cancel{E}_T$  distributions, truncated at either 25 (red) or 50 (blue) GeV, vs. upper bound on fake combined isolation. A jet is defined as in Table 1.2, but with the  $\Delta R$  cleaning criteria of Table 1.2 applied to the two primary EM objects and all additional electrons, photons, and fake photons. The full reweighting and normalization procedure is employed in the  $\cancel{E}_T$  calculation (see Sec. ??). Reprinted from Fig. 9 of ref. [? ].

track extrapolated from the position of the ECAL SC back to the primary vertex. Real photons, having no charge and therefore no bending in the magnetic field, should not have a pixel seed.

### 1.1.2 Electrons

Electrons are reconstructed identically to photons, except that in the electron case the presence of a pixel seed is enforced, rather than vetoed.<sup>2</sup> Photons and electrons are defined by very similar criteria so that  $Z \rightarrow ee$  events can be used to model the QCD background in the two-photon sample without introducing any bias in the electron energy measurement (cf. Sec. ??).

### 1.1.3 Jets and Missing Transverse Energy

#### Particle Flow

In this analysis, jets and  $\cancel{E}_T$  are formed from *particle flow* (PF) candidates. The particle flow algorithm [?] uses information from all CMS subdetectors to reconstruct as accurately as possible the positions and momenta of all visible jet constituents, exploiting the fine granularity of the tracker and ECAL to achieve a greatly improved momentum resolution over calorimeter-only jets [?]. The PF algorithm is summarized below [?].

1. Reconstruct the fundamental detector objects via iterative procedures

- Tracks in the inner silicon layers
  - High efficiency and low fake rate for charged hadrons in jets

---

<sup>2</sup>In many CMS analyses, electrons are reconstructed very differently from photons. In particular, a special tracking algorithm [?] is used to best follow a radiating electron. However, in this analysis, the electron tracking is not used.

- Relaxed primary vertex constraint allows photon conversions, particles originating from nuclear interactions in the silicon, and long-lived particles to be reconstructed

- Calorimeter clusters
- Muon tracks in the outer muon layers

## 2. Create a “block” of linked fundamental objects

- Link silicon tracks to calorimeter clusters via  $\Delta R_{\text{track-cluster}}$  (account for electron bremsstrahlung)
- Link clusters in one calorimeter layer to clusters in a separate layer via  $\Delta R_{\text{cluster-cluster}}$
- Link silicon tracks to muon tracks via global track  $\chi^2$

## 3. ID the particles in the block

- If global (silicon + muon layers) muon  $p_T$  is compatible with silicon track  $p_T$ , ID as a muon and remove corresponding tracks from block
- ID electron tracks via special algorithm and removed all corresponding tracks and cluster from block
- Remove fake tracks from the block
- Allow multiple tracks to be associated to one HCAL cluster, but not multiple HCAL clusters to be associated to one track—for each track, keep only the HCAL cluster link that is closest in  $\Delta R$  to the track
- If the cluster energy is significantly larger then the energy of the linked tracks, ID as a PF photon or PF neutral hadron
- If the cluster is not linked to a track, ID as a PF photon or PF neutral hadron



- If the cluster energy is smaller than the energy of the linked tracks, ID each track as a PF charged hadron

## Jets

PF candidates are clustered into jets by means of the anti- $k_T$  algorithm with  $R = 0.5$  [? ]. In this algorithm, all possible pairs of PF candidates  $i, j$  are looped over, and the pairs that minimize the distance variable

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2 \max(k_{Ti}^2, k_{Tj}^2)} \quad (1.4)$$

are clustered together, where  $k_{Ti}$  is the transverse momentum of PF candidate  $i$ . The process is repeated, using the pairwise-clustered PF candidates as input objects to the next round of clustering, until  $d_{ij} > 1/k_{Ti}^2$  for all pairs of clustered PF candidates [? ]. An illustration is given in Figure 1.10. The anti- $k_T$  algorithm is infrared and collinear safe, leading to well-behaved theoretical predictions and ease of comparison between data and MC simulation. It also tends to form circular jets, making it easy for experimental effects such as expected out-of-cone energy and fiducial acceptance to be measured or simulated. For these reasons, the anti- $k_T$  jet clustering algorithm was chosen for this analysis.

Once jets are found, they must be corrected for biases in the energy measurement due to non-compensation [? ], invisible energy (lost to overcoming nuclear binding energy, in neutrinos, or in unclustered muons, for example) [? ], detector geometry and cracks [? ], zero suppression and trigger inefficiencies [? ], pileup, and effects of the clustering algorithm [? ]. Four multiplicative correction factors are applied to the raw jet four-momentum  $p_\mu^{\text{raw}}$  [? ]:

- $C_{\text{offset}}(p_T^{\text{raw}})$ , which accounts for extra energy due to noise, pileup, and the un-

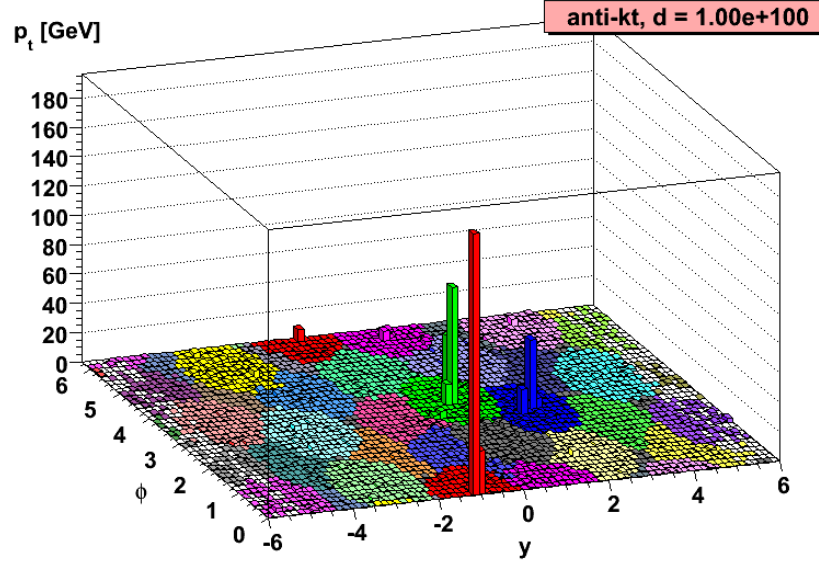


Figure 1.10: Example event display showing jets clustered via the anti- $k_T$  algorithm.  $y$  is pseudorapidity. Reprinted from slide 85 of ref. [? ].

derlying event;

- $C_{MC}(C_{offset}p_T^{raw}, \eta)$ , which is derived from MC and accounts for most of the  $p_T$  and  $\eta$  dependence;
- $C_{rel}(\eta)$ , which accounts for the remaining differences in uniformity over the entire calorimeter between data and MC; and
- $C_{abs}(C_{rel}C_{MC}C_{offset}p_T^{raw})$ , which accounts for the remaining differences in linearity over the full  $p_T$  range between data and MC.

Figure 1.11 shows the total jet energy correction factor  $C_{offset}C_{MC}C_{rel}C_{abs}$  vs.  $\eta$  for jets reconstructed with the anti- $k_T$  algorithm,  $R = 0.5$ . The PF jet corrections are more uniform across  $\eta$  than those of CALO jets (composed of simple calorimeter towers) or JPT jets (Jet Plus Tracks; composed of calorimeter energies replaced, where possible, with matching track  $p_T$ ) [? ]. In addition, for  $p_T$  in the range 30-200 GeV and  $|\eta|$  up to 2.0, the PF jet energy correction uncertainty is lower than that of the other two types of jets, and never exceeds  $\sim 3\%$  [? ]. The superior performance

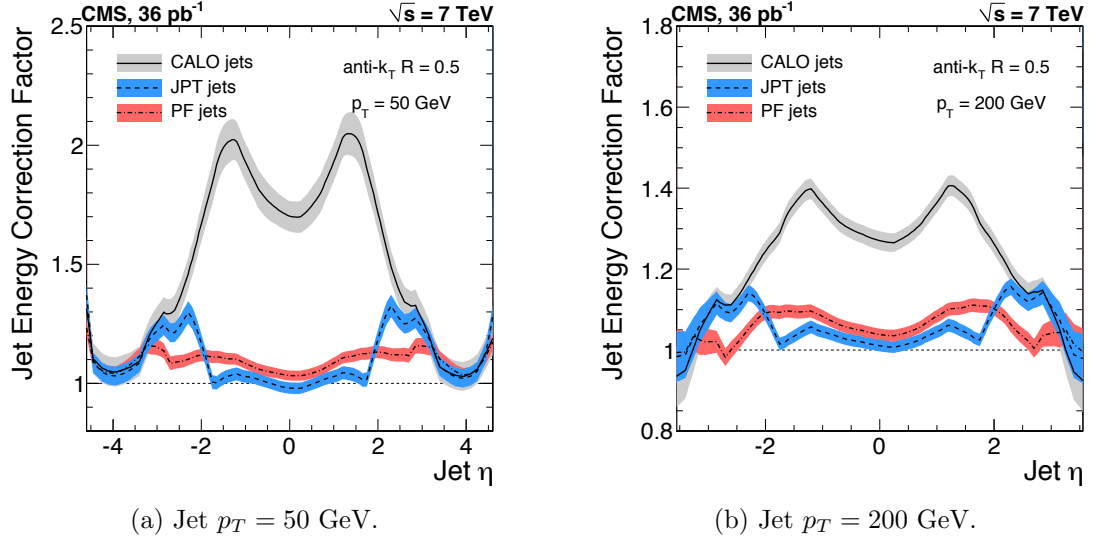


Figure 1.11: Total jet energy correction factor  $C_{\text{offset}}C_{\text{MC}}C_{\text{rel}}C_{\text{abs}}$  vs.  $\eta$ , including uncertainty band, for jets reconstructed with the anti- $k_T$  algorithm,  $R = 0.5$ . Reprinted from Fig. 26 of ref. [?].

of PF jets motivates their use in this search.

In this analysis, candidate and QCD control events are binned by number of jets satisfying the criteria in Table 1.2.

### Missing Transverse Energy

To be consistent with the jet reconstruction,  $\cancel{E}_T$  in this analysis is also reconstructed from PF candidates. Raw  $\cancel{E}_T$  is defined as

$$\cancel{E}_{T\text{raw}} = \left| - \sum_{i=1}^{n_{\text{PF}}} \vec{p}_{Ti} \right| \quad (1.5)$$

where  $n_{\text{PF}}$  is the number of PF candidates in the event.  $\cancel{E}_{T\text{raw}}$  may be corrected for the same effects that necessitate jet corrections, since  $\cancel{E}_{T\text{raw}}$  is usually the result of jet mis-measurement (except, of course, in electroweak physics processes that include an energetic neutrino, or SUSY production). CMS *Type-I*  $\cancel{E}_T$  corrections simply involve

Table 1.2: Definition of HB/HE hadronic jets.

Variable	Cut
Algorithm	L1FastL2L3Residual corrected PF
$p_T$	$> 30$ GeV
$ \eta $	$< 2.6$
Neutral hadronic energy fraction	$< 0.99$
Neutral electromagnetic energy fraction	$< 0.99$
Number of constituents	$> 1$
Charged hadronic energy	$> 0.0$ GeV if $ \eta  < 2.4$
Number of charged hadrons	$> 0$ if $ \eta  < 2.4$
Charged electromagnetic energy fraction	$< 0.99$ if $ \eta  < 2.4$
$\Delta R$ to nearest PF electron <sup>a</sup> , muon <sup>b</sup> , or one of the two primary EM objects	$> 0.5$

<sup>a</sup>A PF electron is defined as an electron reconstructed with the PF algorithm [?] with  $p_T > 15$  GeV,  $|\eta| < 2.6$ , and  $(I_{\text{charged}} + I_{\text{photon}} + I_{\text{neutral}})/p_T < 0.2$ , where  $I_{\text{charged}}(I_{\text{photon}})(I_{\text{neutral}})$  is the sum of PF charged hadron(PF photon)(PF neutral hadron) momenta in a  $\Delta R = 0.4$  cone around the PF electron.

<sup>b</sup>Muons are reconstructed [?] from a combination of muon station and inner tracker hits. Here, a muon must have track  $\chi^2 < 10$ , at least one good muon station hit, inner track transverse impact parameter  $< 0.02$  cm, inner track longitudinal impact parameter  $< 0.5$  cm,  $p_T > 15$  GeV,  $|\eta| < 2.6$ , and  $(I_{\text{ECAL}} + I_{\text{HCAL}} + I_{\text{track}})/p_T < 0.2$ , where  $I_{\text{ECAL}}(I_{\text{HCAL}})(I_{\text{track}})$  is the sum of ECAL(HCAL)(track) momenta in a  $\Delta R = 0.3$  cone around the muon.

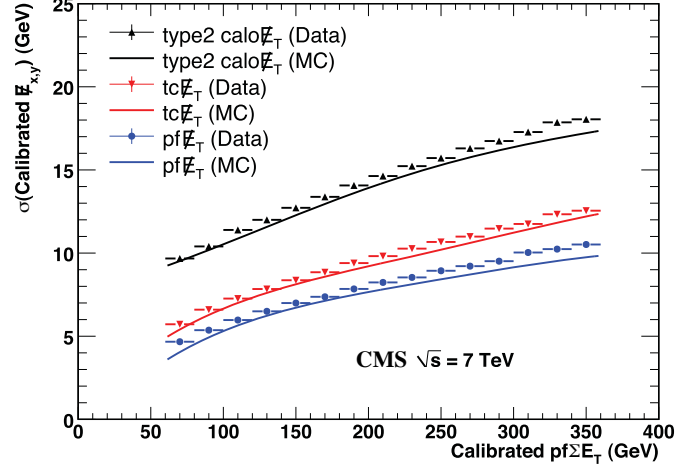


Figure 1.12:  $\sigma$  of a Gaussian fit to the x- and y-components of calibrated  $\cancel{E}_T$  vs. the calibrated PF  $E_T$  sum in a sample of events containing at least two jets with  $p_T > 25$  GeV.  $\sigma$  is calibrated such that the  $\cancel{E}_T$  scale is equal for all three algorithms. PF  $\sum E_T$  is corrected, on average, to the particle level using a Pythia v8 simulation [? ]. The blue markers (data) and line (MC) refer to PF jets. Reprinted from Fig. 13 of ref. [? ].

replacing the PF jets with their corrected energies (cf. Sec 1.1.3) and recalculating  $\cancel{E}_T$ . Only jets with electromagnetic fraction (EMF) below 90% and  $p_T > 20$  GeV are replaced. This ensures that very electromagnetic jets (as well as isolated leptons, which also receive no correction), which consist chiefly of neutral pions and are measured accurately by the ECAL, do not receive a correction derived for jets with a large fraction of their energy in charged hadrons. In addition, the  $p_T$  cut guarantees that jet corrections are only applied where they are known to within a few percent. For this search, the level of agreement between the SM background estimate and the two-photon search sample in a low- $\cancel{E}_T$  control region is the same regardless of whether the  $\cancel{E}_T$  is corrected or not, so for simplicity the Type-I  $\cancel{E}_T$  corrections are not used (see Sec. ??).

Figure 1.12 shows the  $\sigma$  of a Gaussian fit to the x- and y-components of calibrated  $\cancel{E}_T$  vs. the calibrated PF  $E_T$  sum in a sample of events containing at least two jets with  $p_T > 25$  GeV. Again, PF  $\cancel{E}_T$  outperforms  $\cancel{E}_T$  constructed of calorimeter towers or track-corrected calorimeter deposits.

## 1.2 HLT

From the objects described in Sec. 1.1, four samples of events are formed:

- $\gamma\gamma$  candidate sample, in which the two highest  $E_T$  objects are photons,
- $e\gamma$  control sample, in which the two highest  $E_T$  objects are one electron and one photon,
- $ee$  control sample, in which the two highest  $E_T$  objects are electrons, and
- $ff$  control sample, in which the two highest  $E_T$  objects are fakes.

In all samples, the leading EM object is required to have offline reconstructed  $E_T > 40$  GeV, while the trailing EM object is required to have offline reconstructed  $E_T > 25$  GeV. The high level triggers used to select the four samples, by run range, are listed in Table 1.3. No trigger is prescaled.

Each piece of the HLT path name is defined as follows.

- **Photon:** Energy deposit in the ECAL that fired an L1 trigger (cf. Sec. ??). For `Photon26_IsoVL_Photon18`, the L1 seed  $E_T$  threshold is 12 GeV, while for all other triggers in Table 1.3 it is 20 GeV (cf. Sec. ??).
- Integer following the word **Photon:**  $E_T$  threshold in GeV for offline reconstructed photon, using the full photon reconstruction of Sec. 1.1.1 minus the laser calibrations and assuming the primary vertex at (0, 0, 0).
- **CaloIdL:** For EB photons,  $H/E < 0.15$  and  $\sigma_{in\eta} < 0.014$ .
- **IsoVL:**  $I_{\text{ECAL}} < 0.012E_T + 6$  GeV,  $I_{\text{HCAL}} < 0.005E_T + 4$  GeV, and  $I_{\text{track}} < 0.002E_T + 4$  GeV.
- **R9Id:**  $R9 > 0.8$ .

Table 1.3: HLT paths triggered by the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $f\bar{f}$  samples, by run range. No triggers are prescaled.

Run range	$\gamma\gamma$	$e\gamma$	$ee$	$f\bar{f}$
160404-163261	Photon26_ IsoVL_ Photon18	Photon26_ IsoVL_ Photon18	Photon26_ IsoVL_ Photon18	Photon26_ IsoVL_ Photon18
161216-166967	Photon36_ CaloIdL_ Photon22_ CaloIdL	Photon36_ CaloIdL_ Photon22_ CaloIdL	Photon36_ CaloIdL_ Photon22_ CaloIdL	Photon36_ CaloIdL_ Photon22_ CaloIdL
166347-180252	Photon36_ CaloIdL_ IsoVL_ Photon22_ CaloIdL_ IsoVL	Photon36_ CaloIdL_ IsoVL_ Photon22_ CaloIdL_ IsoVL	Photon36_ CaloIdL_ IsoVL_ Photon22_ CaloIdL_ IsoVL	Photon36_ CaloIdL_ IsoVL_ Photon22_ CaloIdL_ IsoVL  Photon36_ CaloIdL_ IsoVL_ Photon22_ R9Id  Photon36_ R9Id_ Photon22_ CaloIdL_ IsoVL  Photon36_ R9Id_ Photon22_ R9Id

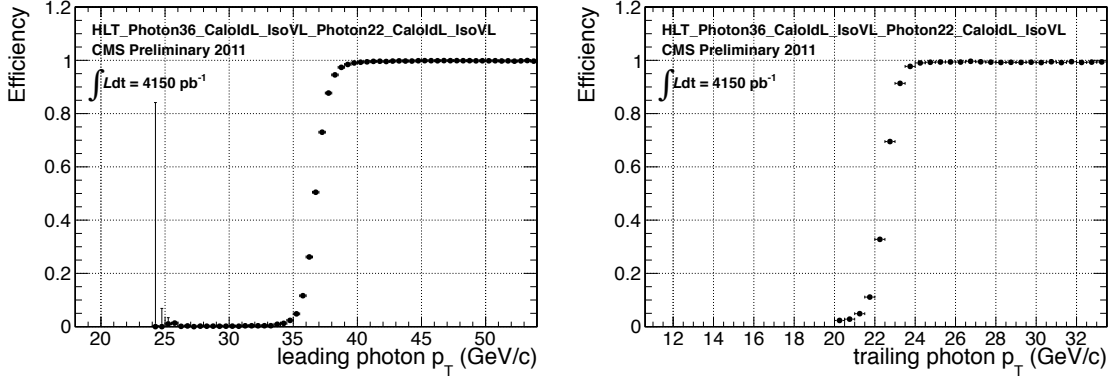


Figure 1.13: Efficiency of HLT\_Photon36\_CaloIdL\_IsoVL\_Photon22\_CaloIdL\_IsoVL for offline selected leading photon (left) and trailing photon (right) vs. photon  $p_T$ . Reprinted from Fig. 2 of ref. [? ].

In addition, the versions of HLT\_Photon26\_IsoVL\_Photon18 and Photon36\_CaloIdL\_Photon22\_CaloIdL that were active during runs 160404-163268 included a cut  $E_{\text{max}}/E_{5 \times 5} < 0.98$  for spike rejection.  $E_{\text{max}}$  is the energy in the highest energy crystal of the EM cluster and  $E_{5 \times 5}$  is the energy in the  $5 \times 5$  crystal matrix around the seed crystal. For runs after 163268, Swiss cross spike rejection of individual crystals from HLT quantities was performed (cf. Sec. 1.1.1). All information about the evolution of the CMS HLT settings can be found in the HLT configuration browser at <http://j2eeps.cern.ch/cms-project-confdb-hltdev/browser/>.

As an example of the naming convention just described, the HLT path Photon36\_CaloIdL\_IsoVL\_Photon22\_R9Id is fired if one photon is found with  $E_T > 36$  GeV passing the CaloIdL and IsoVL requirements, and another is found with  $E_T > 22$  GeV passing the R9Id requirement.

For the offline  $E_T$  cuts described in this section, the triggers are  $> 99\%$  efficient, as shown in Figure 1.13 [? ]. The efficiencies are measured with respect to triggers with lower  $E_T$  thresholds.



### 1.3 Event Quality

To suppress instrumental backgrounds, a set of event quality cuts are applied to the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  samples. First, all events are required to pass a good run selection, as determined by the CMS Physics Validation Team [?]. The good run selection excludes luminosity sections during which a sufficient part of the CMS detector was unpowered or malfunctioning. Such conditions could occur if, for example, a high voltage supply trips off in the middle of a run, or a DAQ error corrupts data quality but is not spotted until after the data have been collected. The severity of a detector problem is judged by its effect on a wide range of analyses and reconstruction algorithms. Of the  $\sim 5 \text{ fb}^{-1}$  of integrated luminosity delivered by the LHC in 2011,  $4.68 \text{ fb}^{-1}$  passed the good run selection. This analysis is performed on the entire 2011 certified dataset.

Second, all events must contain at least one good interaction vertex. The criteria for a good vertex are:

- $\chi^2 \neq 0$  OR  $\text{ndof} \neq 0$  OR  $N_{\text{tracks}} \neq 0$ , where  $\chi^2$  and  $\text{ndof}$  are calculated for the track fit to the vertex, and  $N_{\text{tracks}}$  is the number of tracks in the vertex fit
- $\text{ndof} > 4$
- $|z| < 24 \text{ cm}$ , where  $z$  is the  $z$ -coordinate of the vertex position
- $|\rho| < 2 \text{ cm}$ , where  $\rho$  is the transverse displacement of the vertex position from the beam line

The good vertex requirement eliminates non-collision backgrounds such as beam scraping, beam halo, cosmic muon interactions, and instrumental effects.

Third, the two electromagnetic objects in the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  events must be separated in  $\phi$  by at least 0.05. This requirement protects against beam halo bremsstrahlung, in which a halo muon traveling parallel to the beam line radiates an

energetic photon while itself depositing a large amount of energy in the ECAL. In this case, the two ECAL hits would likely be at the same  $\phi$  (and  $\rho$ ).

Fourth, the two EM objects must be separated in  $R$  by at least 0.6. Since the isolation cone size used is 0.3, this ensures that the isolation energy of one EM object cannot be in the veto strip (Fig. 1.6) of the other.

Finally, the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $f\bar{f}$  events must pass an HCAL noise filter and ECAL dead channel filter. The HCAL noise filter guarantees that all HCAL reconstructed hits are inconsistent with any noise source. Noise sources [?] include:

- Ion feedback in the HPDs absent any true incident photons, in which a thermal electron ionizes a molecule in the HPD acceleration gap, faking a real signal
- HPD discharge affecting nearly all channels in the same HPD [?], partially explained by the effect of the 4 T CMS magnetic field on the flashover voltage of the dielectric [?]
- Concurrent signals in nearly all 72 channels of a single RBX, as yet unexplained
- HF PMT window hits (as opposed to the usual quartz fiber hits)
- ADC saturation

Since HCAL noise may induce fake jets or  $E_T$ , events are rejected if any of the following criteria are true:

- Any HPD has  $> 17$  hits
- A single HPD has  $> 10$  hits, but every other HPD has zero hits
- An RBX has  $> 10$  zero-ADC-count hits
- Any HB/HE reconstructed hit corresponding to an RBX with  $> 50$  GeV of energy fails a two-dimensional cut defined by the variables  $(TS4 - TS5)/(TS4 +$

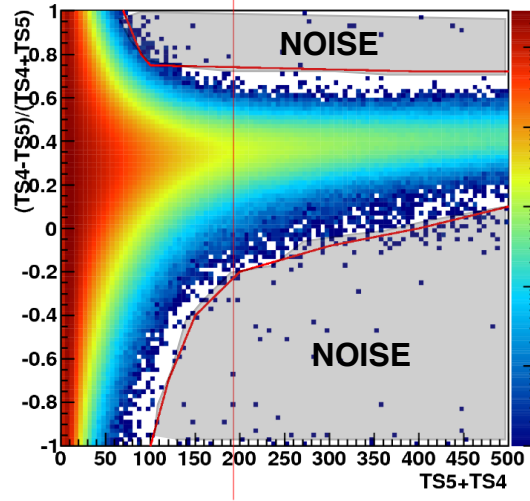


Figure 1.14:  $(TS4 - TS5)/(TS4 + TS5)$  vs.  $TS4 + TS5$  for a minimum bias sample. HB/HE hits are considered noisy if they lie in the sparsely populated gray region labeled "NOISE" defined by the curved red lines. Adapted from ref. [? ].

$TS5)$  vs.  $TS4 + TS5$ , where  $TS4(TS5)$  is the hit amplitude in the fourth(fifth) time sample read out for that hit. The cut is defined in Fig. 1.14.

The ECAL dead channel filter is designed to flag events in which significant EM energy was deposited in a masked region of the ECAL by using the trigger primitive information for the corresponding trigger tower. Energy deposited in a masked region of ECAL can cause fake  $\cancel{E}_T$ . Events are rejected if the trigger primitive  $E_T$  exceeds the maximum value of 63.75 GeV in any trigger tower that is masked in the readout.

## 1.4 Photon Identification Efficiency

In order to determine the cross section (or cross section upper limit) for a GGM signal, the photon identification efficiency is needed. Since no suitably large sample of  $Z \rightarrow \mu\mu\gamma$  events in CMS exists yet, the efficiency calculation relies on the similarity between detector response to electrons and photons. A scale factor to correct the MC photon ID efficiency to the real photon efficiency for the data is obtained from the ratio of the electron efficiency from the data to the electron efficiency from MC.

The different types of photon ID variables—calorimeter and track isolation, ratio of hadronic to electromagnetic energy of the shower, and transverse shower shape—are chosen so that their distributions for isolated electrons and photons are similar.<sup>3</sup>

The photon selection efficiency is

$$\epsilon_\gamma = \epsilon_\gamma^{\text{MC}} \times \frac{\epsilon_e^{\text{data}}}{\epsilon_e^{\text{MC}}} \quad (1.6)$$

where

- $\epsilon_\gamma$  is the photon ID efficiency in data,
- $\epsilon_\gamma^{\text{MC}}$  is the photon ID efficiency in MC,
- $\epsilon_e^{\text{data}}$  is the electron ID efficiency obtained using  $Z \rightarrow ee$  electrons in the data that satisfy the photon ID cuts, and
- $\epsilon_e^{\text{data}}$  is the electron ID efficiency obtained using  $Z \rightarrow ee$  electrons in MC that satisfy the photon ID cuts.

The ratio  $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$  is defined as the scale factor by which the GGM signal MC photon ID efficiency must be multiplied to give an estimate of the photon ID efficiency in data. The photon ID requirements of Table 1.1 plus the **IsoVL** HLT requirement described in Sec. 1.2 and Table 1.3 are repeated in Table 1.4.

### 1.4.1 Tag and Probe Method

A *tag and probe* method using  $Z$  events is utilized to measure the efficiency of the photon ID cuts in Table 1.1. The tag is a well-identified electron. The probe, by contrast, is as loosely identified as possible, and all tags must pass the probe criteria

---

<sup>3</sup> $R9$  differs between photons and radiating electrons, but the requirement  $R9 < 1$  is loose enough not to introduce problems with the use of electrons to measure the photon ID efficiency.

in addition to the stringent tag criteria. The tag and probe criteria used in this study are shown in Table 1.5.

The invariant mass of the tag and probe are required to be within a narrow window around  $Z$  mass. Assuming that the probabilities of the tag and probe legs of the  $Z$  decay to pass the photon ID cuts are uncorrelated, the efficiency can be estimated as

$$\epsilon = \frac{N_{\text{tag-pass}}}{N_{\text{tag-pass}} + N_{\text{tag-fail}}} \quad (1.7)$$

where  $N_{\text{tag-pass}}$  is the number of tag-probe pairs in which the probe leg passes the photon ID cuts under study and  $N_{\text{tag-fail}}$  is the number of tag-probe pairs in which the probe leg fails the cuts. Implicit in these definitions is a double counting of pairs in which both electrons pass the tag and probe criteria [? ]. In addition, in the rare circumstance (less than 1% in MC [? ]) that two or more probes may be matched to one tag, the pair with invariant mass closest to the  $Z$  mass is chosen.

In practice,  $N_{\text{tag-pass}}$  and  $N_{\text{tag-fail}}$  are returned by a simultaneous unbinned maximum likelihood fit to the invariant mass distributions of tag-pass and tag-fail events, with appropriate signal and background PDF assumptions. The fit form used is

Table 1.4: Candidate photon ID requirements.

Variable	Cut
$I_{\text{ECAL}}$	$< 0.012E_T + 6 \text{ GeV}$
$I_{\text{HCAL}}$	$< 0.005E_T + 4 \text{ GeV}$
$I_{\text{track}}$	$< 0.002E_T + 4 \text{ GeV}$
$H/E$	$< 0.05$
$\sigma_{i\eta i\eta}$	$< 0.011$
$I_{\text{ECAL}} - 0.0792\rho + I_{\text{HCAL}} - 0.0252\rho + I_{\text{track}}$	$< 6 \text{ GeV}$
$R9$	$< 1$

Table 1.5: Tag and probe criteria. The superscript 0.4 indicates that the isolation variable was calculated in a cone of  $\Delta R = 0.4$  around the photon candidate. The isolations without superscripts use the standard  $\Delta R = 0.3$  cones.

Variable	Cut	
	Tag	Probe
RECO object	photon	photon
HLT	HLT_Ele17_CaloIdVT_CaloIsoVT_TrkIdT_TrkIsoVT_SC8_Mass30_v* (must have fired the 17 GeV leg)	—
$H/E$	$< 0.05$	$< 0.15$
$I_{\text{ECAL}}^{0.4}$	$< 0.006E_T + 4.2 \text{ GeV}$	—
$I_{\text{HCAL}}^{0.4}$	$< 0.0025E_T + 2.2 \text{ GeV}$	—
$I_{\text{track}}^{0.4}$	$< 0.001E_T + 2.0 \text{ GeV}$	—
$E_T$	$> 25 \text{ GeV}$	—
SC $E_T$	—	$> 15 \text{ GeV}$
SC $ \eta $	$< 1.4442$	$< 1.4442$
$\sigma_{i\eta i\eta}$	$< 0.009$	—
Has pixel seed	Yes	—
Track match type	General track <sup>a</sup>	—
Track match $\Delta R$	$< 0.04$	—
Track match $p_T$	$> 15 \text{ GeV}$	—
Track match $ \eta $	$< 1.479$	—

<sup>a</sup>A general track is reconstructed with the CMS standard combinatorial track finder [?].

$$\begin{aligned}
f_{\text{tag-pass}}(m_{\text{tag-pass}}) &= \epsilon N_S f_S^{\text{pass}}(m_{\text{tag-pass}}) + N_B^{\text{pass}} f_B^{\text{pass}}(m_{\text{tag-pass}}) \\
f_{\text{tag-fail}}(m_{\text{tag-fail}}) &= (1 - \epsilon) N_S f_S^{\text{fail}}(m_{\text{tag-fail}}) + N_B^{\text{fail}} f_B^{\text{fail}}(m_{\text{tag-fail}}) \quad (1.8)
\end{aligned}$$

where  $f_{\text{tag-pass}}(m_{\text{tag-pass}})$  and  $f_{\text{tag-fail}}(m_{\text{tag-fail}})$  are the tag-pass and tag-fail PDFs, respectively;  $\epsilon$  is the efficiency;  $N_S$  is the total number of  $Z$  signal events summed over both samples;  $f_S^{\text{pass}}(m_{\text{tag-pass}})$  and  $f_S^{\text{fail}}(m_{\text{tag-fail}})$  are the tag-pass and tag-fail signal PDFs, respectively;  $N_B^{\text{pass}}$  and  $N_B^{\text{fail}}$  are the numbers of background events in the tag-pass and tag-fail samples, respectively; and  $f_B^{\text{pass}}(m_{\text{tag-pass}})$  and  $f_B^{\text{fail}}(m_{\text{tag-fail}})$  are the tag-pass and tag-fail background PDFs, respectively. This particular implementation of the tag and probe methodology is based on tag `CMSSW_4_2_5` of the CMSSW package `PhysicsTools/TagAndProbe`, and uses the MINUIT2 [?] library, as coded in RooFit [?], for the likelihood maximization. For this study, CMSSWv4.2.8 was used.

For both samples, the signal shape is assumed to be a Crystal Ball function [?] convoluted with the  $Z$  generated lineshape, while the background shape is a PDF that describes the falling background as well as the kinematic turn-on at low invariant mass. The background PDF, called `RooCMSShape` [?], is given by

$$f_{\text{RooCMSShape}}(x) = \begin{cases} 1e20 & \text{for } (x - \mu)\gamma < -70 \\ 0 & \text{for } (x - \mu)\gamma > 70 \\ \text{erfc}((\alpha - x)\beta) \exp(-(x - \mu)\gamma) & \text{otherwise} \end{cases} \quad (1.9)$$

where  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\mu$  are parameters of the fit, most of which are held fixed. In the three lowest  $E_T$  bins, all parameters of the tag-pass and tag-fail background PDFs are left floating, because the effect of the relaxed  $E_T$  cuts has a significant effect on

the background shape. More details of the signal and background PDFs are given in Table 1.6. The fixed signal and background parameter values were determined by fitting a small sample ( $0.0 \leq \eta < 0.25$ ) of Fall111 MC signal (DYJetsToLL) and background (QCD\_Pt-20to30\_BCtoE, QCD\_Pt-30to80\_BCtoE, QCD\_Pt-80to170\_BCtoE, GJet\_Pt-20\_doubleEMEnriched, WJetsToLNu, TTJets) with parameters left floating.<sup>4</sup>

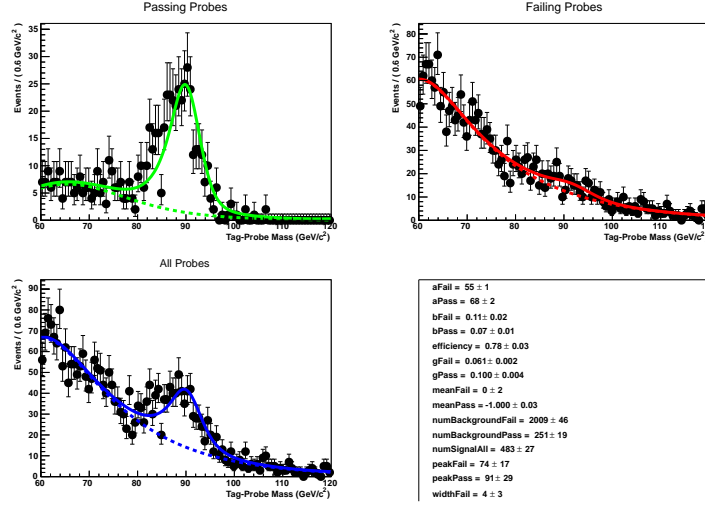
Table 1.6: Parameter values (parameter definitions are in the text) for the signal and background PDFs for the different samples. The background PDF applies to all efficiency bins except the four lowest  $E_T$  bins, which use a floating RooCMSShape background. When a bracketed range is given, the parameter is allowed to float within that range. When a constant is given, the parameter is fixed to that constant.

PDF	Crystal Ball fit parameters				RooCMSShape fit parameters			
	$\mu$	$\sigma$	$\alpha$	n	$\mu$	$\alpha$	$\beta$	$\gamma$
Tag-pass signal	[-1.0, 1.0]	[1.0, 3.0]	0.87	97.0	N/A	N/A	N/A	N/A
Tag-fail signal	[-1.0, 1.0]	[1.0, 3.0]	0.73	134.9	N/A	N/A	N/A	N/A
Tag-pass background	N/A	N/A	N/A	N/A	65.0	61.949	0.04750	0.01908
Tag-fail background	N/A	N/A	N/A	N/A	$\alpha$	[50.0, 100.0]	0.065	0.048

Some fit examples are shown in Figures 1.15 and 1.16. In Fig. 1.15, which shows fits to data and MC for  $15 \text{ GeV} \leq \text{probe } E_T < 20 \text{ GeV}$ , the kinematic turn-on is below the invariant mass range covered by the plot. The exponentially falling background is easily seen underneath the signal, and is especially pronounced in the background-dominated tag-fail sample.

<sup>4</sup>See Appendix ?? for a discussion of the MC samples.





(a) Data.

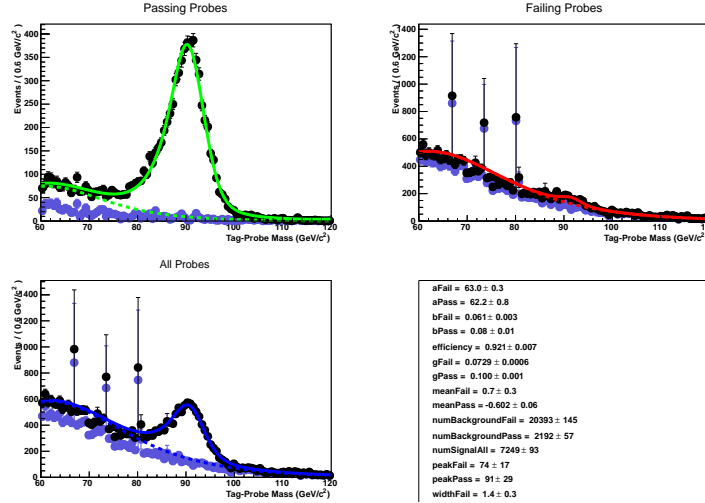
(b) MC. The purple points are the background MC (photon + jet,  $W$ , QCD, and  $t\bar{t}$ ).

Figure 1.15: Tag and probe invariant mass fits for  $15 \text{ GeV} \leq \text{probe } E_T < 20 \text{ GeV}$ . Errors are statistical only. The tag-pass fit is shown in green in the upper-left-hand plot, the tag-fail fit in red in the upper-right-hand plot, and a fit to both samples in blue in the lower-left-hand plot. Dotted lines are the background components of the fits; solid lines are signal plus background.

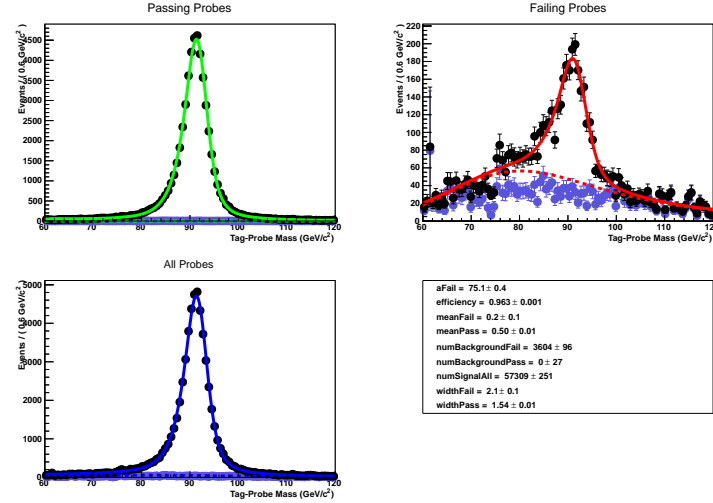
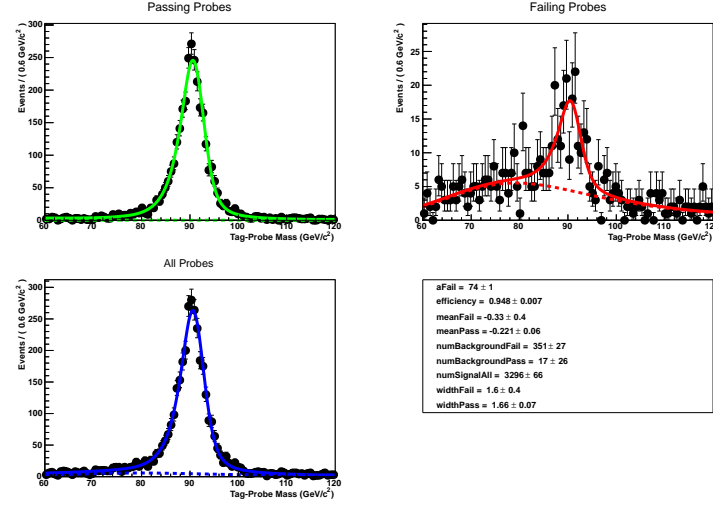


Figure 1.16: Tag and probe invariant mass fits for  $-0.25 \leq \text{probe } \eta < -0.5$ . Errors are statistical only. The tag-pass fit is shown in green in the upper-left-hand plot, the tag-fail fit in red in the upper-right-hand plot, and a fit to both samples in blue in the lower-left-hand plot. Dotted lines are the background components of the fits; solid lines are signal plus background.

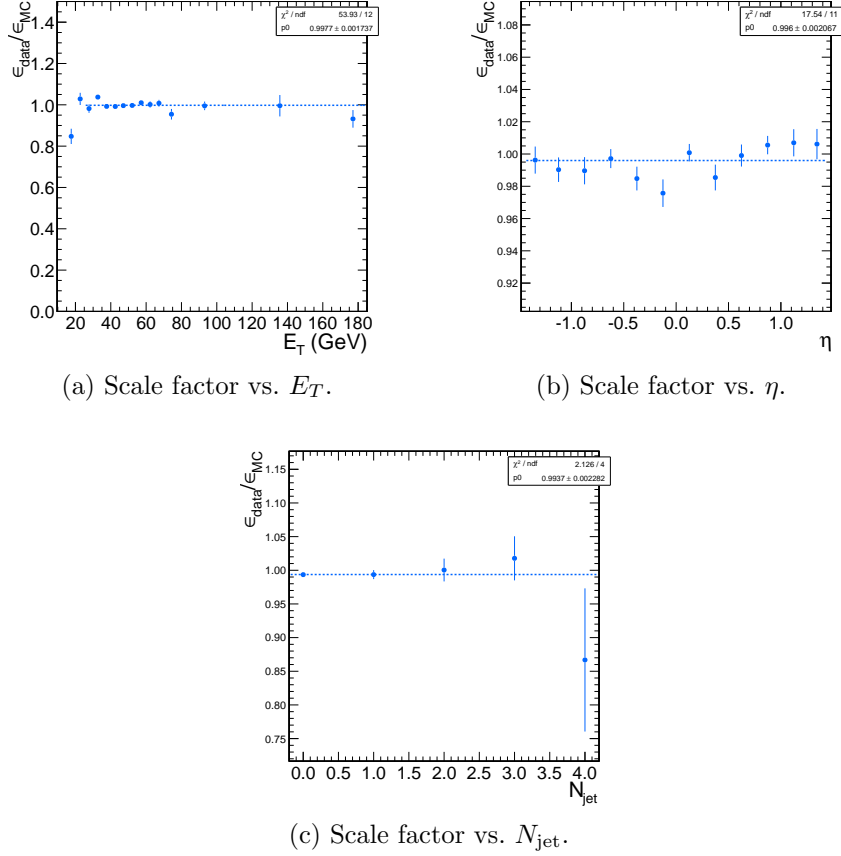


Figure 1.17: Dependence of the photon ID efficiency scale factor on some kinematic variables. Errors are statistical only.

#### 1.4.2 Photon Efficiency Scale Factor $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$

Figure 1.17 shows the dependence of the photon ID efficiency scale factor  $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$  on  $E_T$ ,  $\eta$ , and  $N_{\text{jet}}$ , where jets are defined as in Table 1.2, but with only the two  $Z$  electrons considered as candidates for overlap removal. Errors are statistical only. There no significant dependence of the scale factor on these variables, so only one scale factor is computed from the entire dataset.

The effect of pileup is studied by comparing the efficiencies  $\epsilon_e^{\text{data}}$  and  $\epsilon_e^{\text{MC}}$  vs. the number of primary vertices ( $N_{\text{PV}}$ ) in the event. The efficiency only drops a few percent for events with large  $N_{\text{PV}}$  after using pileup-corrected isolation cuts, as can be seen in Figure 1.18(a). The MC tracks the data, and the scale factor is flat vs.  $N_{\text{PV}}$ , as

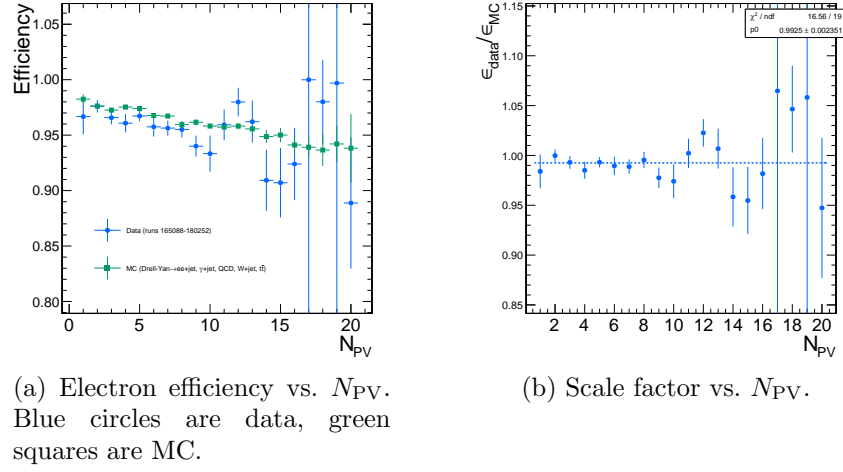


Figure 1.18: Dependence of the photon ID efficiency scale factor on the number of primary vertices per event. Errors are statistical only.

shown in Fig. 1.18(b).

The scale factor is measured to be  $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}} = 0.994 \pm 0.002(\text{stat.}) \pm 0.035(\text{syst.})$ .

Four main sources of systematic error, in addition to the statistical error of 0.2%, were studied.

#### Different behavior of electrons and photons in MC

Even though the photon ID cuts are designed to be similarly efficient for both electrons and photons, there might be a small difference in the performance between the two kinds of particles, e.g. because of electron bremsstrahlung. To check this effect, the MC electron ID efficiency was calculated using a  $Z \rightarrow ee$  sample and the MC photon ID efficiency was calculated using a  $\gamma + \text{jets}$  sample. Both samples were reconstructed in CMSSWv3.6. Half the difference between these two results, 0.5%, was taken as an error on the scale factor.

**Pileup** To account for the possibility that the MC simulation may not adequately reproduce the data in a high pileup environment, the data/MC scale factor for events with 1-4 good reconstructed primary vertices was calculated, along with the same for events with  $\geq 5$  good reconstructed primary vertices. The

difference between the scale factors from both samples, 2.4%, was taken as an error on the scale factor from pileup.

**Signal fit over/underestimation** It was found that the signal fit slightly underestimates the data in the tag-pass sample, and slightly overestimates it in the tag-fail sample. To cover this effect with a systematic error, the efficiencies in data and MC, and then the scale factor, were recalculated using the background (from fit) subtracted integrals of the tag-pass and tag-fail distributions, rather than the fitted signal yields in those distributions. The difference between the scale factor found in this way and the nominal scale factor, 1.9%, was taken as an error on the scale factor.

**Signal and background shape assumption** To assess the magnitude of the error from the signal and background shape assumptions, the tag-pass and tag-fail tail parameters (Crystal Ball  $\alpha$  and  $n$ ) were varied by  $\pm 1\sigma$ , and the background shape was varied between `RooCMSShape`, exponential, power law, and quadratic. All possible combinations of varied parameters were generated, and the data and MC were refit and new scale factors generated according to those combinations. The error was taken as the largest deviation of the scale factor from nominal, 1.8%.

Finally, the pixel veto efficiency was estimated from MC to be  $0.96 \pm 0.005(\text{syst.})$ , with error due to varying assumptions of the tracker material distribution [? ]. In general, the photon ID selection used in this analysis is very efficient for GGM photons and robust to pileup, and its efficiency is fairly well measured.