

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 symbol \cancel{E}_T .

However, in order to use real CMS data (as opposed to simulation) to derive predictions for the backgrounds to the search, *control samples* distinct from the *candidate* two-photon sample must be collected. 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 and control samples. Finally, the chapter concludes with a measurement of the photon

identification efficiency in Sec. 1.3.

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 time in hit, should contain no signal. This pedestal 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 1.1.

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,

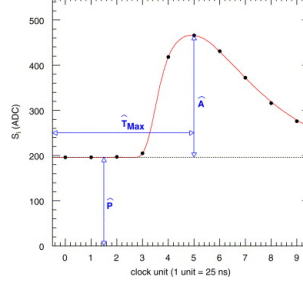


Figure 1.1: Typical ECAL channel pulse shape. \hat{P} is the pedestal value, \hat{A} is the pulse amplitude, and \hat{T}_{\max} is the hit time. The red line is the assumed pulse shape from which the weights are derived. Reprinted from ref. [?].

- they are stuck in fixed gain,
- 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).

In addition, no uncalibrated hits that are spike-like are eligible for calibration. The calibrations applied are crystal transparency loss corrections measured continuously by the laser/LED system, energy intercalibrations (relative energy calibration between crystals), absolute scale calibrations between ADC counts and GeV,¹ and 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

¹The 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 $5 \times$ cluster of crystals (plus the associated energy deposited in ES) is 50 GeV for a 50 GeV incident unconverted photon [?].

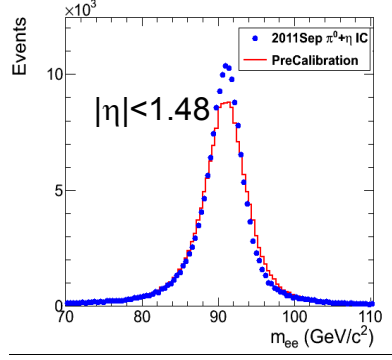


Figure 1.2: Z peak reconstructed using pre-LHC calibration constants (red) or September 2011 π^0/η -derived intercalibration constants (blue). Reprinted from ref. [?].

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:

- ϕ symmetry relative calibration between crystals, exploiting the azimuthal symmetry of CMS
- π^0 and η 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 π^0/η method [?]. Figure 1.2 shows the improvement in Z reconstruction from pre-LHC calibration constants to the latest π^0/η -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 gain-adjusted, pedestal-subtracted, and reconstructed using weights. To make a calibrated ES hit, intercalibration constants, angle correction constants, 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 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 ϕ , not η . Both algorithms work by forming

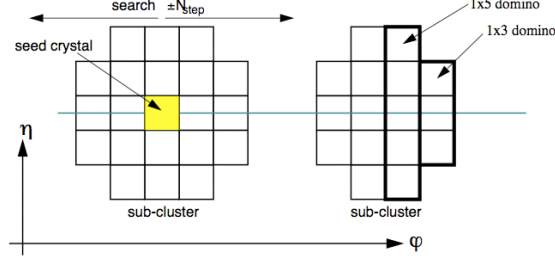


Figure 1.3: Hybrid algorithm in EB. The shower extent is essentially constant in η , but spreads out in ϕ as the two sub-clusters (or basic clusters) are grouped into the same supercluster. Reprinted from ref. [?].

basic clusters around seeds, then combining the basic clusters into *superclusters* (SC) by searching in a window extended in the ϕ direction for all basic clusters consistent with bremsstrahlung radiation from the primary electron, or with a photon conversion. Figure 1.3 illustrates the hybrid algorithm in EB. In EE, the energy deposited in ES must also be added into the total clustered energy sum.

Figure 1.4 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 SC correction factor. G and c_i were explained in Secs. 1.1.1 and 1.1.1. F is a product of three factors for hybrid SCs (two for multi5x5 SCs) [?]:

1. $C_{EB}(\eta|)$, which compensates for lateral energy leakage due to the crystal off-pointing in EB. These corrections are taken from MC simulation [?] and were confirmed in test beams [?].

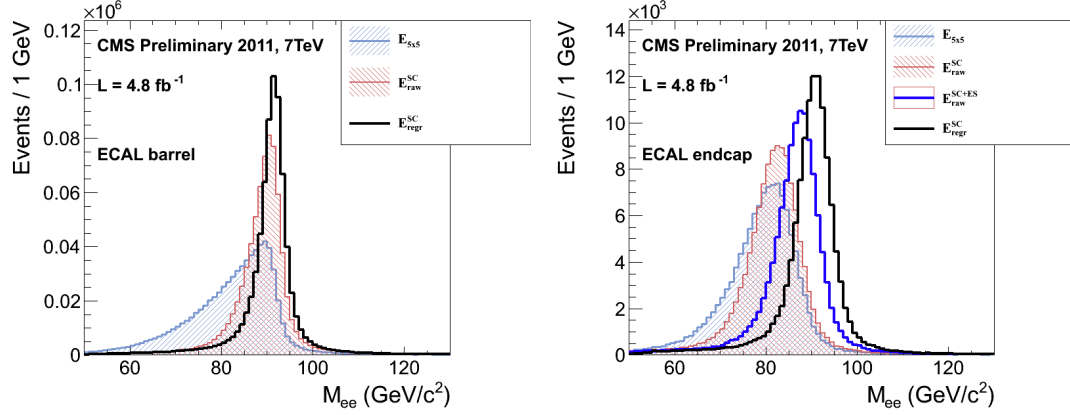


Figure 1.4: Z peak reconstructed in the dielectron channel for different kinds of clustering. The left plot is for EB and the right plot for EE. The constituent hits were calibrated with the best available intercalibrations and laser calibrations as of December 2011. The light blue histogram shows the reconstruction using a 5×5 energy sum, the red histogram shows the reconstruction using the SC energy for crystals only (the dark blue histogram on the right-hand side 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. [?].

2. $f(\text{brem})$, which corrects for the 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 η 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 $Z \rightarrow ee$ 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 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 $E_{3 \times 3}/E_{SC}$, 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 5×5 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 ?? . I_{comb} is defined as

$$I_{\text{comb}} = I_{\text{ECAL}} - 0.0792\rho + I_{\text{HCAL}} - 0.0252\rho + I_{\text{track}} \quad (1.2)$$

where I_{ECAL} , I_{HCAL} , and I_{track} are E_T sums in the annular regions defined in Figure 1.5 and ρ is the average pileup energy density in the calorimeters (per unit $\eta \cdot \phi$) as measured with the Fastjet algorithm [?, ?]. Figure 1.6 shows the ρ distribution for a sample of events with at least two 25 GeV EM objects passing the $|\eta|$, H/E , and $R9$ requirements in Table 1.1, and passing the trigger requirements in Table 1.2, representing the full 2011 dataset. Since average ρ is ~ 5 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 0.0792 and 0.0252, respectively, are calculated by fitting the average ECAL or HCAL isolation energy vs. ρ in a sample

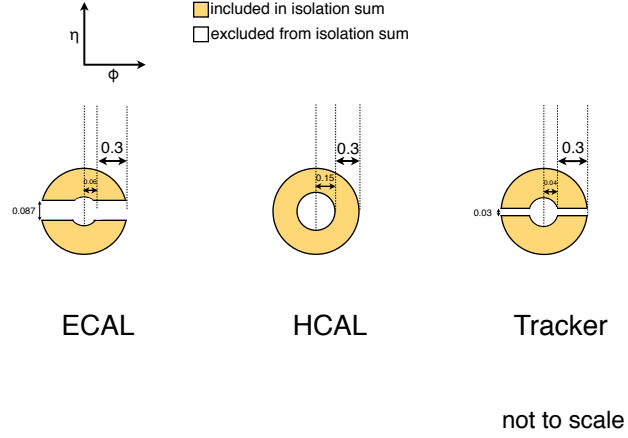


Figure 1.5: ECAL, HCAL, and track Isolation cones.

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.

A “pixel seed” is defined as a hit in the pixel detector consistent with a 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.

Table 1.1: Selection criteria for photons and fakes.

Variable	Cut (γ)	Cut (f)
SC $ \eta $	< 1.4442	< 1.4442
H/E	< 0.05	< 0.05
$R9$	< 1	< 1
Has pixel seed	No	No
I_{comb}	$< 6 \text{ GeV}$	$\geq 6 \text{ and } < 20 \text{ GeV}$

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.² Photons and electrons

²In 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,

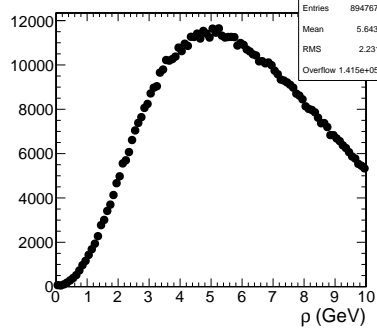


Figure 1.6: ρ distribution for a sample of events with at least two 25 GeV EM objects passing the $|\eta|$, H/E , and $R9$ requirements in Table 1.1, and passing the trigger requirements in Table 1.2. This sample covers the full 2011 dataset.

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

the electron tracking is not used.

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 χ^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
- Remove excess track-cluster links via $\Delta R_{\text{track-cluster}}$ minimization (but allow multiple tracks to be associated to one cluster)
- If the cluster energy is significantly larger than the energy of the linked track, ID as a PF photon or PF neutral hadron and remove corresponding clusters from the block
- If the cluster is not linked to a track, ID as a PF photon or PF neutral hadron and remove corresponding clusters from the block
- Remaining track-cluster links are PF charged hadrons

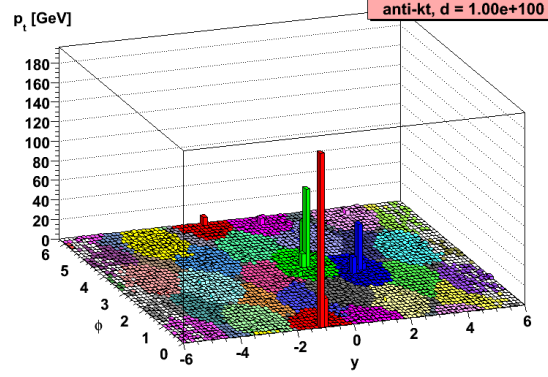


Figure 1.7: Example event display showing jets clustered via the anti- k_T algorithm. y is pseudorapidity.

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 momenta of the pair that minimize the distance variable

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2 \max(k_{Ti}^2, k_{Tj}^2)} \quad (1.3)$$

are combined, where k_{Ti} is the transverse momentum of “combined” PF candidate i . The constituent PF candidates are clustered together. The process is repeated until $d_{ij} > 1/k_{Ti}^2$ for all pairs of clustered PF momenta [?]. An illustration is given in Figure 1.7. 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 clustered, 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 geom-

etry 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_μ^{raw} [?]:

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

Figure 1.8 shows the total jet energy correction factor $C_{\text{offset}}C_{\text{MC}}C_{\text{rel}}C_{\text{abs}}$ vs. η for jets reconstructed with the anti- k_T algorithm, $R = 0.5$. The PF jet corrections are more uniform across η 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 of PF jets motivates their use in this search.

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

$$\cancel{E}_{T\text{raw}} = \left| - \sum_{i=1}^{n_{\text{PF}}} \vec{p}_{Ti} \right| \quad (1.4)$$

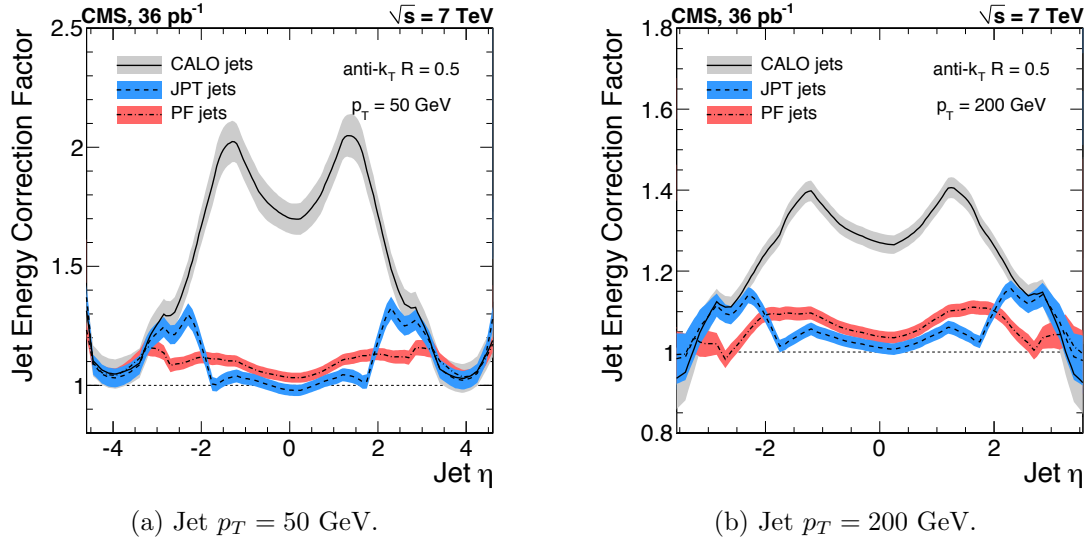


Figure 1.8: Total jet energy correction factor $C_{\text{offset}}C_{\text{MC}}C_{\text{rel}}C_{\text{abs}}$ vs. η , including uncertainty band, for jets reconstructed with the anti- k_T algorithm, $R = 0.5$. Reprinted from Fig. 26 of ref. [?].

where n_{PF} is the number of PF candidates in the event. $\cancel{E}_{T\text{raw}}$ must 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). Therefore, CMS *Type-I* \cancel{E}_T corrections simply involve 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. Since this search looks for excesses in the \cancel{E}_T distribution typically above 100 GeV of \cancel{E}_T , it is not very sensitive to the mis-measurement of low- p_T jets.

Check this statement and add illustrative plots.

Figure 1.9 shows the σ of a Gaussian fit to the x- and y-components of calibrated

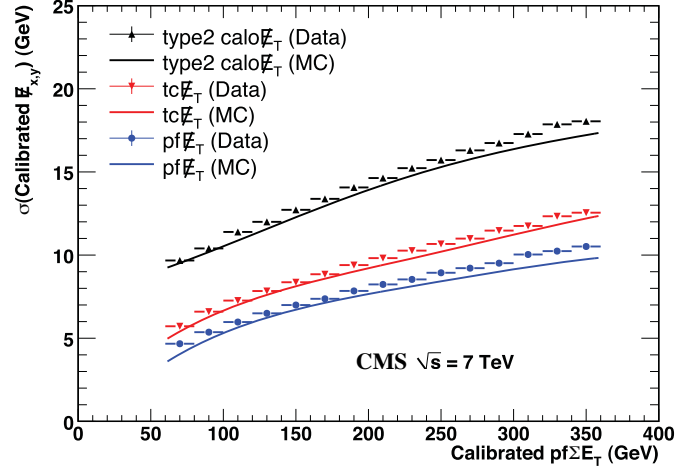


Figure 1.9: σ of a Gaussian fit to the x- and y-components of calibrated \cancel{E}_T vs. the calibrated PF E_T sum in a sample of events containing at least two jets with $p_T > 25$ GeV. σ is calibrated such that the \cancel{E}_T scale is equal for all three algorithms. PF $\sum E_T$ is corrected, on 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. [?].

\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 $E_T > 40$ GeV, while the trailing EM object is required to have $E_T > 25$ GeV. The high level triggers used to

select the four samples, by run range, are listed in Table 1.2. No trigger is prescaled.

Table 1.2: HLT paths triggered by the $\gamma\gamma$, $e\gamma$, ee , and ff samples, by run range. No triggers are prescaled.

Run range	$\gamma\gamma$	$e\gamma$	ee	ff
160404-161215	Photon26_ IsoVL_ Photon18	Photon26_ IsoVL_ Photon18	Photon26_ IsoVL_ Photon18	Photon26_ IsoVL_ Photon18
161216-166346	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_ R9Id Photon36_ R9Id_ Photon22_ CaloIdL_ IsoVL Photon36_ R9Id_ Photon22_ R9Id	Photon36_ CaloIdL_ IsoVL_ Photon22_ CaloIdL_ IsoVL Photon36_ CaloIdL_ IsoVL_ Photon22_ R9Id Photon36_ R9Id_ Photon22_ CaloIdL_ IsoVL Photon36_ R9Id_ Photon22_ R9Id

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 1.2 it is 20 GeV.

- 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 $\sigma_{i\eta i\eta} < 0.014$.
- “IsoVL”: $I_{\text{ECAL}} < 0.012E_T + 6$ GeV, $I_{\text{HICAL}} < 0.005E_T + 4$ GeV, and $I_{\text{track}} < 0.002E_T + 4$ GeV.
- “R9Id”: $R9 > 0.8$.

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_{max} is the energy in the highest energy crystal of the EM cluster and $E_{5\times 5}$ is the energy in the 5×5 crystal matrix around the seed crystal. For runs after 163268, Swiss cross spike rejection of individual crystals from HLT quantities was performed (cf. Sec. ??). 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.

1.3 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.³ Figure ?? shows distributions of photon ID variables for MC electrons and photons, where a MC electron(photon) is a `reco::Photon` object matched to a generated electron(photon) originating from a Z decay(the hard interaction) within $\Delta R = 0.3$. The MC samples used are Using a tag and probe method with $Z \rightarrow ee$ data events, the efficiency for photons is calculated as

³ $R9$ 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.