Available on CMS information server **CMS IN -2010/008**

CMS Internal Note

The content of this note is intended for CMS internal use and distribution only

12 March 2010 (v3, 26 March 2010)

Review of clustering algorithms and energy corrections in ECAL

M. Anderson, A. Askew, A.F. Barfuss, D. Evans, F. Ferri, K. Kaadze, Y. Maravin, P. Meridiani, C. Seez

Abstract

Many experimental signatures involve electrons and photons in the final state over a large energy range. The first step in the identification and measurement of electrons and photons in the CMS detector is performed in the Electromagnetic Calorimeter (ECAL). The purpose of this note is to explain how the energy and position of electrons and photons are reconstructed. This description is based upon the procedures employed in the software release CMSSW 3 X.

CMS Internal Note

The content of this note is intended for CMS internal use and distribution only

2010/03/26- Revision - checkout

A Review of clustering algoritms and energy corrections in the Electromagnetic Calorimeter

M. Anderson, A. Askew, A.F. Barfuss, D. Evans, F. Ferri, K. Kaadze, Y. Maravin, P. Meridiani, C. Seez

Abstract

Many experimental signatures involve electrons and photons in the final state over a large energy range. The first step in the identification and measurement of electrons and photons in the CMS detector is performed in the Electromagnetic Calorimeter (ECAL). The purpose of this note is to explain how the energy and position of electrons and photons are reconstructed. This description is based upon the procedures employed in the software release CMSSW_3_X.

Contents

Introduction

Many experimental signatures involve electrons and photons in the final state over a large energy range. The first step in the identification and measurement of electrons and photons in the CMS detector is performed in the Electromagnetic Calorimeter (ECAL). The purpose of this note is to explain how the energy and position of electrons and photons are reconstructed. This description is based upon the procedures employed in the software release CMSSW_3_X.

The bulk of the ECAL is a homogeneous scintillating calorimeter, made of trapezoidal lead tungstate ($PbWO_4$) crystals as the active medium. The Moliere radius of $PbWO_4$ is 2.19 cm. The detector is divided into a barrel region (EB) transverse to the beam direction up to $|\eta| < 1.48$ and two endcaps (EE) that cover the region from $1.48 < |\eta| < 3.0$ perpendicular to the beam direction. Crystals in the barrel region have a depth of 25.8 radiation lengths X_0 and are 2.2×2.2 cm at the front face. The endcap crystals are shorter than the barrel crystals, with a depth of 24.7 X_0 and front face of 2.9×2.9 cm. Preshower detectors (ES) are situated in front of the inner face of each endcap covering $1.65 < |\eta| < 2.6$. Each preshower comprises silicon sensors (divided into 2mm strips) and Pb absorbers presenting an additional 3 X_0 of material in front of the EE (at $|\eta| = 1.7$). The physical layout of the ECAL is illustrated in Figure 1. For a more detailed description of the ECAL geometry, see [1]

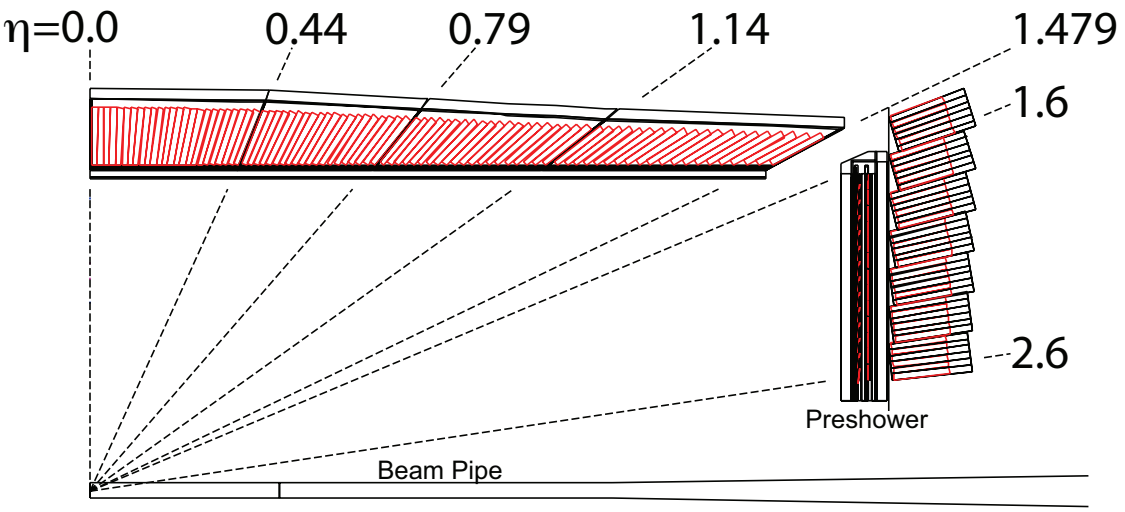

Figure 1: An illustration of the ECAL layout.

The ECAL is immersed in the field of the CMS solenoid coil, with $\vec{B} = 3.8$ T.

Particles interacting with the ECAL deposit a fraction or their energy which is converted into a measurable light signal by the crystals. Groups of crystals with deposited energy above a certain threshold are collected together into basic clusters (BCs), which are themselves grouped using a bremsstrahlung hypothesis into super clusters (SCs). The reconstructed energy of a SC is assumed to be comparable to the energy of the incident particle.

Existing energy corrections for energy losses due to the presence of tracker material in front of ECAL were tuned for underestimated material budget (a), low energy ($<$ 100 GeV) electrons and photons (b) and causing overcorrection of energy for high energetic showers, Hybrid (for EB) and Island (for EB and EE) [2] clustering algorithms (c). Since then, a more realistic description of the tracker material budget has been achieved. In addition, the Hybrid algorithm

has been updated and a new clustering algorithm (Multi5x5) has been implemented for energy reconstruction in EE.

We address all these items in our studies in order to follow current energy correction scheme [3] [4] and to get an optimal correction function.

1 Clustering

Individual electromagnetic showers are spread over some number of neighbouring crystals. Bremsstrahlung emitted as the electron traverses the silicon tracking detector is spread in the ϕ-direction in the plane perpendicular to the beam due to the bending of the electron trajectory by the solenoidal magnetic field.

To give the best possible energy measurement, neighboring crystals are "clustered" together. Because of the differences between the geometric arrangement of the crystals in the barrel and endcap regions, a different clustering algorithm is used in each region. Both the barrel and endcap algorithms perform the same task of grouping crystals associated with individual electromagnetic showers and collecting showers which are close in the η direction but within some larger window in ϕ.

The input to the clustering processes in the barrel region is the set of hits in the barrel region only, and likewise the endcap clustering uses only the endcap hits. This means that a cluster reconstructed close to the divide between the barrel and endcap may contain a lower than expected fraction of the true electron energy.

To recover some energy loss, clusters in both the EB and EE, which are within some maximum separation, are themselves grouped. This process is referred to as superclustering and the resulting groups of associated basic clusters are referred to as superclusters.

The algorithms by which the basic clustering and superclustering are performed, both in the barrel and endcap regions, are described thereafter in more detail. Instructions to reconstruct basic clusters and superclusters from the EcalRecHits are also given in the twiki page [2].

1.1 Clustering in the barrel region

Clustering in the barrel region is performed using the Hybrid algorithm, which takes advantage of the $\eta - \phi$ geometry of the crystals to perform the collection of both the energy in individual showers and of the set of showers compatible with a bremsstrahlung hypothesis. This is done by collecting energy within a rectangular window extended in the ϕ direction. The algorithm operates on the set of crystals sorted in descending order of E_T as follows:

- Test the E_T of the next crystal that does not already belong to a cluster. To avoid obvious noise contamination and low energy backgrounds, the crystal E_T has to be above a threshold; then if $E_T > E_T^{hybseed}$, this crystal can seed the clustering process. If not, the clustering process terminates.
- Construct a 3×1 domino of crystals in $\eta - \phi$. If $E_{domino} > E_{wing}$ then extend the domino to 5×1 crystals symmetrically about the seed crystal. The default value of $E_{wing} = 0.0$ ensures that 5×1 dominoes are always built.
- Repeat the second step for all crystals with the same η as the seed crystal and within $\phi < \phi_{road}$. To be included in the cluster, dominoes must satisfy $E_{domino} > E_{thresh}$.
- Group the remaining dominoes into local energy maxima, which are connected by side in the ϕ direction and then remove all dominoes belonging to any local maxima

with a highest energy domino below E_{seed}.

The result of the procedure is illustrated in Figure 2. The hybrid supercluster is made up of a series of showers at constant η but spread in the ϕ-direction. Each energy deposit can be well contained in 5×5 crystals.

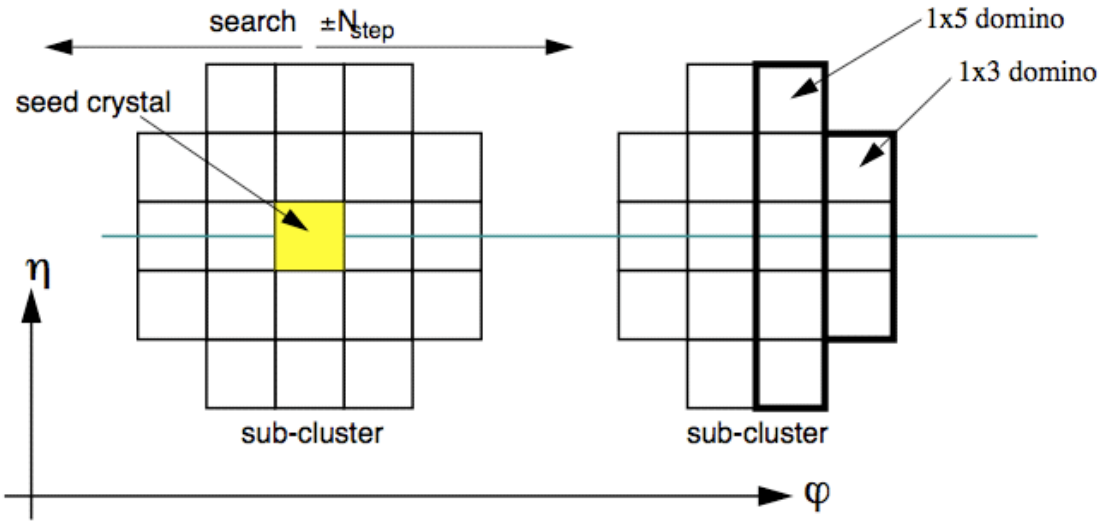

Figure 2: An illustration of the Hybrid clustering algorithm used in the ECAL barrel region.

The behaviour of the algorithm is controlled by the CMSSW configuration,

`RecoEcal/EgammaClusterProducers/python/hybridSuperClusters_cfi.py`.

The default values of the various thresholds, and their names in the cfi_py file are given in Table 1.1.

Parameter	in hybridSuperClusters_cfi.py	Default
ϕ_{road}	`step`	17 crystals
$E_T^{hybseed}$	`HybridBarrelSeedThr`	1 GeV
E_{wing}	`ewing`	0 GeV
E_{thresh}	`ethresh`	0.1 GeV
E_{seed}	`eseed`	0.35 GeV

Table 1: Usual values of the Hybrid clustering algorithm parameters.

1.2 Clustering in the endcaps

Since the crystals in the endcap are not arranged in an $\eta - \phi$ geometry as in the barrel, the hybrid algorithm cannot be applied there. The same idea of collecting energy deposits within a window in η and ϕ must be implemented differently. This is achieved by the Multi5×5 algorithm, which operates as follows on the set of crystals sorted in descending order of E_T, as follows:

- Test the E_T of the next crystal that does not already belong to a cluster. If $E_T > E_T^{seed}$ then this crystal can seed the clustering process. If not, the clustering process terminates.
- Continue if the crystal is a local maxima in energy by comparing its energy to its four neighbours by side in a Swiss Cross pattern. If the crystal is not a local maxima, return to the previous step.
- Construct a 5×5 matrix of crystals about the seed, including only crystals that do not already belong to a cluster.

Due to pedestal subtraction the energy in a crystal may be calculated to be negative. Such crystals are still included in the 5x5 matrix because the effect of the pedestal subtraction should average out over the complete matrix. To allow closely overlapping showers due to bremsstrahlung to be collected, the outer 16 crystals of the 5x5 matrix may seed a new matrix, thus the matrices can overlap. However, no crystal that is already included in a cluster may belong to another.

To produce the final set of clusters by recovering bremsstrahlung, a rectangular window in η and ϕ, which is extended in the ϕ-direction is created around energy deposits with a transverse energy above $E_T^{seed\ bc}$. Other energy deposits falling within the window are added to form the cluster. This procedure is performed in descending order of E_T of the energy deposits and an energy deposit may only belong to one cluster.

The behaviour of the algorithm is controlled by two CMSSW configuration files, one for the grouping of crystals into discrete energy deposits and a second for the clustering of energy deposits. These are

`RecoEcal/EgammaClusterProducers/python/multi5x5BasicClusters_cfi.py`.
`RecoEcal/EgammaClusterProducers/python/multi5x5SuperClusters_cfi.py`.

The default values of the various thresholds, and their names in the cfi_py file are given in Table 1.2.

Parameter	in multi5x5BasicClusters_cfi.py	Default
E_T^{seed}	`IslandEndcapSeedThr`	0.18 GeV

Table 2: Usual values of the Multi5×5 basic clustering algorithm parameters.

Parameter	in multi5x5SuperClusters_cfi.py	Default
$E_T^{seed\ bc}$	`seedTransverseEnergyThreshold`	1.0 GeV
η^{road}	`endcapEtaSearchRoad`	0.14 GeV
ϕ^{road}	`endcapPhiSearchRoad`	0.6 GeV

Table 3: Usual values of the Multi5×5 super clustering algorithm parameters.

An example of the result of the processes is shown in Figure 3. This illustrates the collection of two overlapping energy deposits that have been identified by the algorithm and grouped. The highest E_T crystal, which is a local maximum, is clustered first resulting in a full 5x5 matrix. The second local maxima has already been clustered, however it is still eligible to seed a new 5x5 matrix. This is constructed using the remaining free crystals.

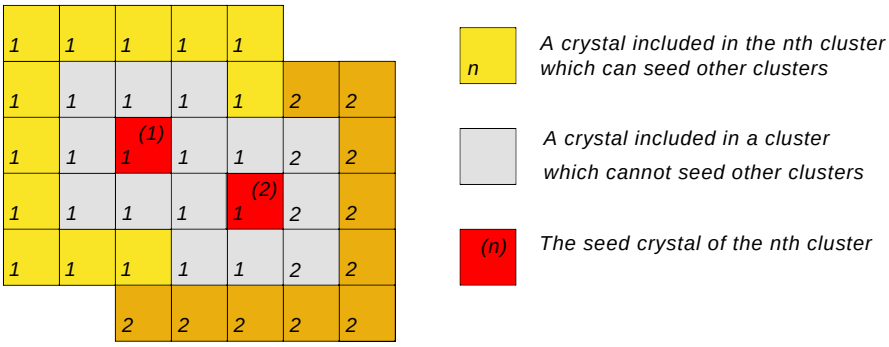

Figure 3: An illustration of two overlapping multi5×5 clusters. Crystals indicated in yellow are eligible to seed further multi5×5 clusters provided they are local maxima in energy.

1.2.1 Recovery of energy deposited in the preshower

In the endcap, the region $1.6 < |\eta| < 2.6$ is covered by the preshower detector. Electrons and photons reconstructed in this region will typically deposit some fraction of their energy in the preshower, so this energy must be measured and added to each cluster. This is done by summing the energy from the preshower strips that intersect at the position extrapolated between each energy deposit in the calorimeter and the primary vertex. This energy sum is computed for and added to each endcap supercluster before any energy scale corrections are applied. The parameters for this procedure are controlled by the configuration

`RecoEcal/EgammaClusterProducers/python/multi5x5SuperClustersWithPreshower_cfi.py`

2 Energy Scale Corrections

A high resolution of the ECAL is crucial in every search which expects electrons and/or photons in the final state. A precise measurement of energy deposit in the calorimeter also improves the resolution of missing transverse momentum (MET), present in the signature of number of new physics searches.

There are different sources of damaging the ECAL energy resolution. One of them is the interaction of particles with matter (here, the tracker material), which results in bremsstrahlung and photon conversions and causes the energy of electromagnetic particles measured in ECAL to be underestimated. The function of SC energy corrections is to compensate this energy loss, or in other words to control energy scales in the ECAL.

2.1 General procedure

Energies in the ECAL are estimated as follows:

$$E = F \times \sum_{\text{cluster}} G \times c_i \times A_i$$

where A_i is the signal amplitude in ADC counts, G is a global scale calibration term, c_i is the intercalibration term (i.e. the uniform individual channel response to a reference).

As described in the energy correction scheme [4], F is the factorization of three individual energy corrections applied sequentially to Hybrid superclusters (EB) and two for SCs reconstructed with the Multi5x5 algorithm and taking the preshower energy into account (EE):

a) first, the $C_{EB}(\eta)$ factor is used to compensate for lateral energy leakage from the exposed faces of the EB crystals, and which then only applies to EB clusters;

b) secondly, a correction called $f(brem)$ aims at correcting for the response of the clustering algorithm to the shower, where *brem* describes the spread of the electromagnetic cascade, i.e. the dimensions of the cluster;

c) finally, due the non-linear distribution of matter in the detector and of the energy dependence, a residual correction $f(E_T, \eta)$ is applied to all reconstructed superclusters.

Because of bremsstrahlung, one mostly deals with electron deposits. The $Z \rightarrow ee$ process is thus a rather good "standard candle" process for this study: the Z mass peak can be reconstructed from a small dataset ($\approx 10/\text{pb}$) and allows to control the electron energy scale. The procedure to derive the energy corrections leads to a description of the energy corrections in two categories: corrections for clustering effects (safely derived from simulation), and corrections for residual effects to be derived in situ, from data.

2.2 Specific correction for barrel clusters

Most of the lead tungstate crystals in the ECAL endcap are oriented with front edge towards the point (0,0,0) or tilted with very small angle. Hence, the first "correction", denoted $C_{EB}(\eta)$, and due to lateral leakage of shower is not needed for energy reconstruction in EE and only applies to SCs with $|\eta| < 1.479$. This function is obtained from Monte Carlo simulation and is in very good agreement with test beam data. The effect of this correction, designed for electrons, is of the order of 1%, while for photons it appears to be slightly smaller.

After correcting EB superclusters with $C_{EB}(\eta)$, the distribution of SC energy E^{reco} is compared to the generated particle energy E^{MC}. In an ideal case, the ratio of the two distributions should

be a gaussian centered on 1. Though because of underestimation of reconstructed particle energy due to radiations in the tracker, the distribution expands to low values of the energy, and this is even more visible in forward regions ($|\eta| > 0.8$).

2.3 Corrections applied to all superclusters

The energy loss due to electron bremsstrahlung is characterized by the variable *brem* defined as follows:

$$brem = \frac{\sigma_\phi}{\sigma_\eta} = \frac{\sum_{Xtal} \sqrt{\frac{E_{Xtal}}{E_{SC}}(\phi_{Xtal} - \phi_{SC})^2}}{\sum_{Xtal} \sqrt{\frac{E_{Xtal}}{E_{SC}}(\eta_{Xtal} - \eta_{SC})^2}}$$

Due to the solenoidal field, the energy deposits in ECAL spread in ϕ direction, while no spread occurs in η direction and the SC width describes the shower size: this is the natural width of a SC (σ_η). Normalization of σ_ϕ with σ_η allows to treat the showers in various energy ranges, all in a single approach. The distribution of the *brem* variable is much narrower for high energetic showers.

It is assumed that $f(brem)$ correction depends less on tracker material and, up to some precision, we can estimate it from Monte Carlo simulation. On the contrary, $f(E_T, \eta)$ through η heavily depends on the amount of material and needs to be re-computed *in situ*.

In order to parameterize the energy loss due to bremsstrahlung as a function of σ_ϕ/σ_η, we consider short ranges of the distribution in the barrel and fit the ratio E^{reco}/E^{MC} with a 'Crystal Ball' function. The mean value of the distribution is extracted from the fit for different ranges of *brem*. In an ideal case, this correction should recover all energy losses in the tracker. However, because of a non-linear distribution of matter in the detector, and a visible dependence on the energy, some additional correction is needed, and has to be parameterized by the transverse energy E_T and the pseudorapidity η.

A similar procedure is used to obtain $f(E_T, \eta)$, considering small η ranges of 0.1 and 2 GeV-ranges of E_T. For each energy range, a correction factor is calculated from the mean value of the Crystal Ball fit of E_T^{reco}/E_T^{MC} and parameterized as a function of η.

The procedure to compute the corrections $f(brem)$ and $f(E_T, \eta)$ in EE is analog to the one used in EB, except that the coverage of the pre-shower calorimeter (ES) [5] is different ($1.653 < |\eta| < 2.6$) from the coverage of the endcap ($1.479 < |\eta| < 3.0$).

The energy corrections are currently implemented in the `EcalReco/EgammaCoreTools/plugins` package (by Federico Ferri), allowing the user to access each individual correction. An example of its usage can be found on the following TWiki page

`https://twiki.cern.ch/twiki/bin/view/CMS/ECALDPGClusterization`

Note that a possibility of applying "data-driven" $f(E_T, \eta)$ correction is implemented. This tool makes it also possible for anyone to derive custom energy corrections for electromagnetic superclusters from the existing correction scheme.

References

[1] CMS Collaboration, "The Electromagnetic Calorimeter Project - Technical Design Report", CERN/LHCC 97-23, December 1998.

[2] https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideEcalRecoClustering

[3] C. Seez, P. Meridiani, "ECAL calibration and the definition of calibrated RecHits", v. 2.1, June 2007.

[4] C. Seez, P. Meridiani, "Addendum to 'ECAL calibration and the definition of calibrated RecHits': details, B-field, and the cluster correction scheme", v. 1.1, November 2007.

[5] CMS Collaboration, CMS ECAL Preshower and Endcap Engineering Design Review. v. 2 - Preshower, CERN-ECAL-EDR-4, November 2000.