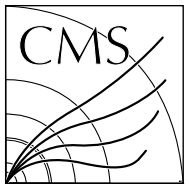
Available on CMS information server

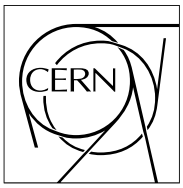
CMS AN -2010/034



The Compact Muon Solenoid Experiment

Analysis Note

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



09 February 2010 (v2, 15 March 2010)

Electron Reconstruction within the Particle Flow Algorithm

F. Beaudette [a)], D. Benedetti [a)], P. Janot [a)], and M. Pioppi [b)]

a) CERN

b) Imperial College London

Abstract

The electron reconstruction within the Particle Flow algorithm is described. Each step of the algorithm is detailed and studied for isolated and non-isolated electrons. The performance of the identification and reconstruction of electrons in jets is investigated and finally, the impact on the particle flow jet response and resolution is evaluated.

1 Introduction

The particle-flow event reconstruction [1] aims at reconstructing and identifying all stable particles in the event, i.e., electrons, muons, photons, charged hadrons and neutral hadrons, with a thorough combination of all CMS sub-detectors, aiming at an optimal determination of their direction, energy and type. This list of individual particles is then used, as if it came from a Monte-Carlo event generator, to build jets (from which the quark and gluon energies and directions are inferred), to determine the missing transverse energy (which gives an estimate of the direction and energy of the neutrinos and other invisible particles), to reconstruct and identify taus from their decay products, to quantify charged lepton isolation with respect to other particles, to tag b jets, etc.

The CMS detector [2] is equipped with a very accurate electromagnetic calorimeter (ECAL) and silicon tracker. The identification and reconstruction of electrons is nevertheless a challenging task. Indeed, the large amount of tracker material (two radiation lengths at a $\eta = \pm 1.5$ pseudo-rapidity) induces a significant Bremsstrahlung photon emission, which, because of the track curvature in the 3.8 T magnetic field, can result in a calorimeter energy deposit widely spread in the azimuthal direction.

Therefore, a special procedure, hereafter called Bremsstrahlung recovery, is needed to collect the Bremsstrahlung photon energy deposits in the calorimeter, as to reconstruct the total energy of the electron. In addition, a dedicated track reconstruction algorithm, able to accommodate for the kinks in the electron trajectory is required. This track reconstruction algorithm is unfortunately computing-time intensive and can only be run on a limited number of track seeds, resulting from a selection procedure.
In the standard CMS electron reconstruction [3], the previous issues have been addressed in the context of energetic and isolated electrons, while in the case of the particle flow, the low-energy and/or non-isolated electrons should be reconstructed as well. Therefore, the particle-flow electron reconstruction has been designed to minimise the sensitivity to the isolation, firstly in the seeding procedure, and then in the Bremsstrahlung recovery where the clustering and the linking procedure described in [1] are heavily exploited.

In this note the electron reconstruction and identification within the particle-flow algorithm is presented. The note organisation follows closely the logic of the electron reconstruction algorithm itself. The seed selection procedure for the electron track reconstruction, is presented in Sec. 2. Sometimes, Bremsstrahlung photon conversions lead to a duplication of the reconstructed electron tracks and a cleaning, described in Sec. 3, is applied. Then, the core of the algorithm, i.e., the Bremsstrahlung recovery is described in Sec. 4. The energy thus collected is corrected, as explained in Sec. 5, before being combined with the track measurement in Sec. 6. Once the association between the electron track and the energy deposits in the calorimeters has been performed, a final identification, described in Sec. 7, is carried out to disentangle electrons from hadrons. Finally, the impact of the electron reconstruction and identification on particle-flow jets is studied in Sec. 8.

2 The electron track reconstruction and seeding

The tracks in CMS are reconstructed with a Kalman Filter (KF) approach, assuming that the charged particle is a muon. For muons, the multiple Coulomb scattering is the dominant effect on the particle while crossing material, and its impact is modelled by Gaussian fluctuations. This approach is not suitable for electrons where the dominant effect is the highly non-Gaussian Bremsstrahlung emission.

Indeed, in the case of energetic Bremsstrahlung photon emission, causing a kink in the electron trajectory, the pattern recognition is often not able to follow the electron path. In contrast, when the photon energy is moderate, the pattern recognition can succeed in collecting all the hits, but the quality of the track fit can be poor.

Therefore, a dedicated electron track reconstruction had to be developed [4]. It firstly relies on a KF-based pattern recognition with relaxed criteria, allowing the trajectory to be reconstructed often up to the last tracker layer. Secondly, a specific fit is carried out, using a Gaussian-Sum Filter (GSF). In this approach, the Bremsstrahlung energy loss is modelled by a Gaussian mixture. The large number of components (currently 11) in the GSF track fit allows the sudden curvature radius change, caused by the Bremsstrahlung photon emission, to be properly taken into account. The GSF track thus provides more precise estimates of the inner and outer momentum than the KF [4]. The price to pay for this large number of components is the CPU-time consumption, at the order of a few hundred milliseconds per track. The GSF track reconstruction can thus be run on a limited number of seeds.

2.1 The ECAL-driven seeding strategy

The original seeding strategy was designed and optimised for isolated electrons with a large transverse momentum (p_{T}). It starts from the ECAL clusters, and is hereafter called the ECAL-driven approach. It relies on the property that the position of the barycentre of all the ECAL energy deposits, resulting from an electron, does not depend on the Bremsstrahlung photon emission. From the barycentre, the position of the hits in the pixel detector and in the most inner strips can be inferred, modulo the ambiguity on the charge. The general track seeds found to match the predicted hit positions are selected. The performance of this method depends on the ability to gather into one "super-cluster" all the Bremsstrahlung photon and electron energy deposits. To do so, all the energy deposits in the ECAL in a large region along ϕ are collected around the crystal with the local energy maximum. In most of the cases, this algorithm allows the electrons from converted Bremsstrahlung photons to be recovered. Only the super-clusters with a transverse energy exceeding 4 GeV are considered.

This procedure is well suited for isolated and high p_{T} electrons but not for non-isolated or low-p_{T} electrons. Indeed for electrons within a jet, the super-cluster position and energy can be biased by the contribution of overlapping particles. Moreover, because of the high track multiplicity, the backward propagation from the super-cluster can be compatible with several track seeds originating from other charged particles. To limit the number of fake seeds inside jets, the ECAL-driven seeding requires that the ratio between the HCAL and ECAL energy deposit is smaller than 0.15 [3]. The fake seed rate is thus kept under control but the efficiency on non-isolated electrons is severely limited. As far as the low-p_{T} electrons are concerned, the spread in ϕ, due to the Bremsstrahlung photons, can be so wide that the super-cluster cannot include all the deposits, which biases its barycentre position and prevents the super-cluster from being matched with a track seed.

2.2 The tracker-driven seeding strategy

In the particle-flow, to increase the seeding efficiency, in particular for non-isolated and low-p_{T} electrons, a pre-identification based on high-purity KF tracks has been developed to create "tracker-driven" seeds.
The electron pre-identification [5] starts from the tracks. A high tracking efficiency is thus required, even for tracks with a small number of hits. An iterative-tracking strategy [6, 7] was adopted to achieve both high efficiency and low fake rate. First, tracks are seeded and reconstructed with very tight criteria, leading to a moderate tracking efficiency, but a negligibly

small fake rate. The next steps proceed by removing hits unambiguously assigned to the tracks found in the previous iteration, and by progressively loosening track seeding criteria. This approach which has primarily been introduced in the context of the electron pre-identification, has become the default method for the track reconstruction in CMS.

The concept of the pre-identification is the following: when the Bremsstrahlung emission is negligible, the electron trajectory in the tracker is, with a good approximation, a helix. In such cases, even though it has been reconstructed with a muon hypothesis, the momentum is determined with good precision by the KF, and the track can be reconstructed up to the ECAL internal surface where it can be matched with the closest cluster. The momentum of the track can then be compared with the corresponding cluster energy, forming the E/p observable. If this ratio is close to unity, the track is selected.
On the contrary, if the electron loses a non-negligible fraction of its energy by Bremsstrahlung emission, the characteristics of the tracks are exploited. As mentioned earlier, in the case of energetic Bremsstrahlung photon emission, two cases are possible: either the pattern recognition stops following the electron path, resulting in a track with a small number of hits; or if the pattern recognition succeeds in collecting all the hits, the fitted track χ^2 is usually large. A selection based on these two criteria, #hits and χ^2_{KF} is firstly applied, then, a light GSF refit, with only five components, is carried out. The GSF refit χ^2_{GSF}, the $\chi^2_{\mathrm{KF}}/\chi^2_{\mathrm{GSF}}$ ratio, the number of hits, the energy loss as measured by the track, as well as the quality of the ECAL-cluster track matching are fed into a multivariate analysis using a Boosted Decision Trees estimator. In the endcaps, the matching with the preshower can be used to improve the pion rejection, but this possibility is not activated at the moment.
The work-flow of the pre-identification strategy is summarised in Fig. 1.

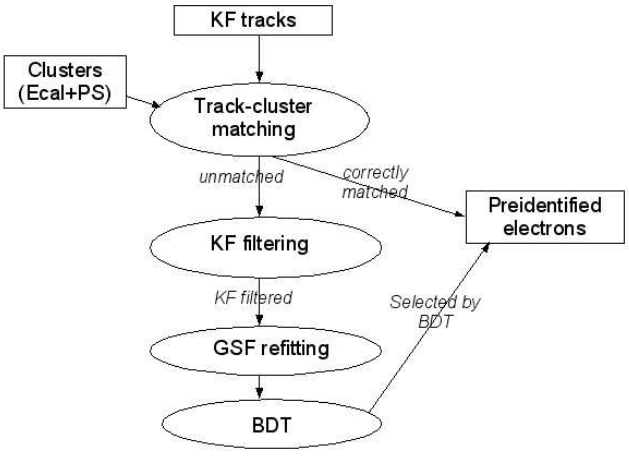
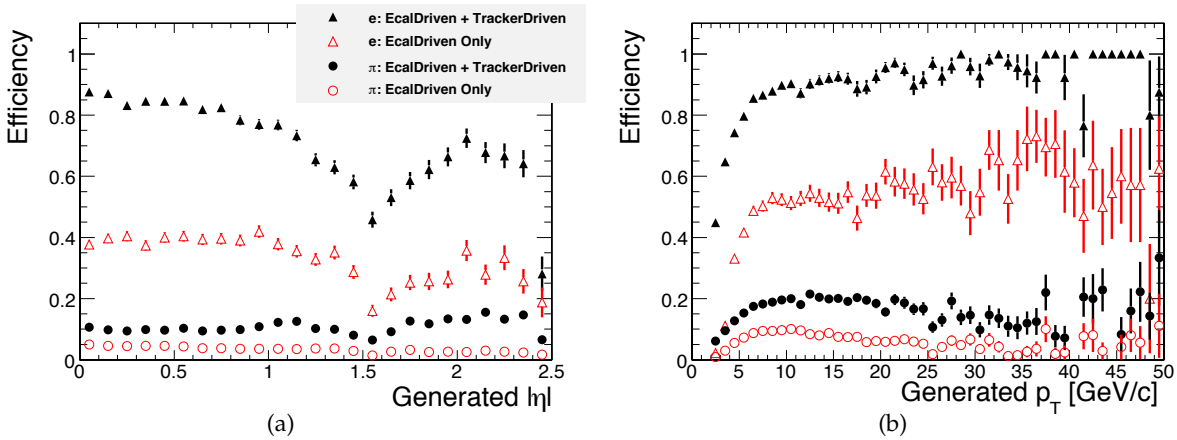


Figure 1: Workflow of the pre-identification strategy.

2.3 Overall seeding performance

The seed finding efficiency for electrons and pions has been studied in a sample of b jets with $20 < \hat{p_{\mathrm{T}}} < 120$ GeV/c, and the results are presented in Fig. 2. The efficiency is computed for electrons and pions with $p_{\mathrm{T}} > 2$ GeV/c in the $|\eta| < 2.5$ tracker acceptance.

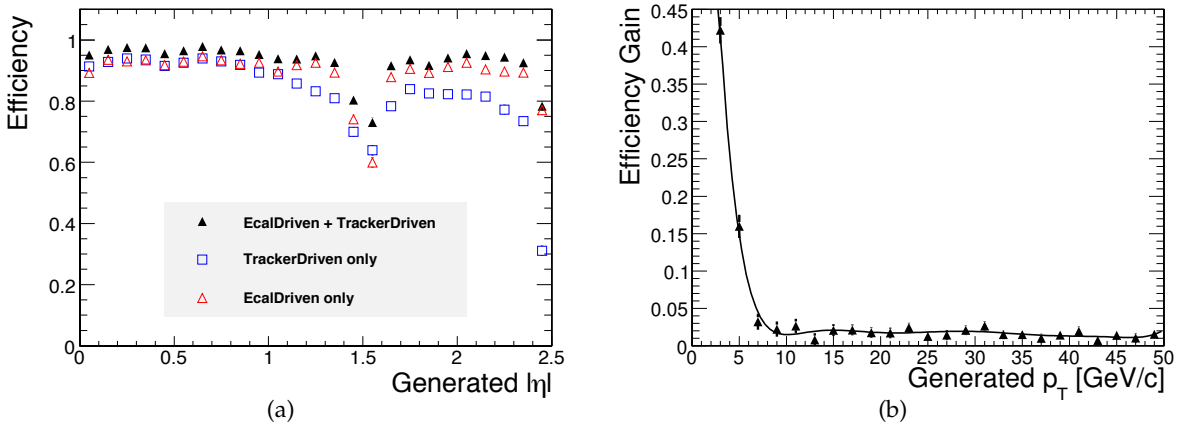
For electron within b jets, the seed efficiency is increased from 36% (ECAL-driven only) to 77% when the tracker-driven seeding is added. The overall pion mis-identification is kept low, around 10%.



(a) (b)

Figure 2: Seeding efficiency for electrons (triangles) and pions (circles) as a function of the generated η (a) and p_{T} (b) in a b-jet sample with $20 < \hat{p_{\mathrm{T}}} < 120$ GeV/c. Full circles and triangles indicate seeds found by the tracker-driven and ECAL-driven seeds, empty circles and triangles indicate seeds found by the ECAL-driven method only.

The electron seeding efficiency, for single electrons with $2 < p_{\mathrm{T}} < 70\,\mathrm{GeV}/c$, is shown in Fig. 3. In particular, in Fig. 3b, the efficiency gain, when the tracker-driven seeding is added, is shown as a function of the generated p_{T}. For electrons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}/c$, the gain is of order of 1-2%, at $p_{\mathrm{T}} = 5\,\mathrm{GeV}/c$ the gain is about 15%.



(a) (b)

Figure 3: (a) Seeding efficiency for single electrons as a function of the generated η for the ECAL-driven only, tracker-driven only and the merged list. (b) Efficiency gain as a function of the generated p_{T}, when the tracker-driven seeds are added.

3 The GSF track collection

The seeds obtained with the tracker-driven and ECAL-driven procedures are merged in order to provide a unique collection of seeds, while keeping memory of their origin. For the tracker-driven seeds the reference to the KF track, out of which the seed has been obtained, is kept; similarly the reference to the standard super-cluster is saved for the ECAL-driven seeds. When

an ECAL-driven seed, with N hits, and a tracker-driven seed share at least N-1 tracker hits, only the former is kept and the provenance is assigned to both algorithms.

The GSF tracking runs on the merged collection of seeds and the resulting GSF tracks are used by both the standard and the particle-flow electron reconstruction algorithms [3]. As explained in Sec. 2, electrons traversing the tracker detector emit a non-negligible amount of Bremsstrahlung photons which afterwards can convert into electron-positron pairs. Bremsstrahlung photons converting in the pixel detector can generate additional electron seeds both for the ECAL-driven and tracker-driven algorithms, while only the tracker-driven seeding is sensitive to the Bremsstrahlung photons converting further in the tracker. The tracker-driven seeding indeed uses all high-purity tracks, included those produced in the fourth and fifth iterations. These tracks are not required to have their first hit close to the interaction point. Those additional seeds imply that more than one GSF track per electron can be reconstructed occasionally.

In a sample of single electrons ($2 < p_{\mathrm{T}} < 70$ GeV/c), in about 14% of the cases, more than one GSF track is reconstructed (Fig. 4a). As expected, the additional GSF tracks are concentrated in the $|\eta| > 1.5$ region where there is more material in the tracker (Fig. 4b). A cleaning procedure is then necessary to identify the GSF track generated by conversion legs and select the track corresponding to the primary electron.

3.1 Cleaning strategy

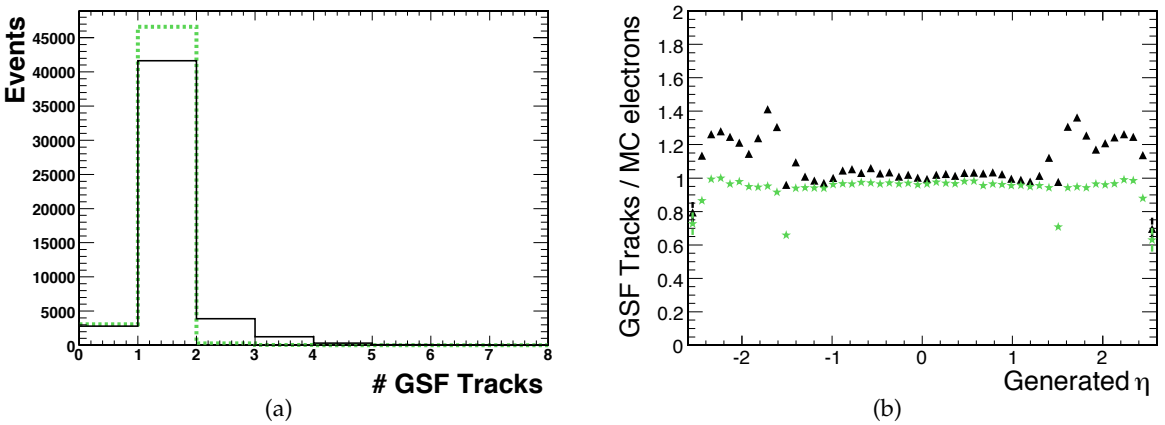
Since the particle-flow electrons are merged with the ECAL-driven electrons [3], a cleaning of the GSF tracks has to be applied before starting the two electron reconstructions. At this level, only intrinsic properties of the GSF track, including the super-cluster in case of ECAL-driven tracks, can be used.

The additional GSF tracks resulting from conversion legs have generally a reconstructed η-direction close to the one of the primary GSF track, and are not too far in ϕ. Therefore, the cleaning strategy is only applied to GSF track pairs whose momenta satisfy $|\Delta\eta| < 0.05$ and $|\Delta\phi| < 0.3$ rad, which happens in 95% of the cases. If two GSF tracks satisfying these criteria are both ECAL-driven seeded, the cleaning procedure is applied only if they are associated to the same super-cluster. Indeed, if they point to different super-clusters, it means that, even though the directions of the two tracks are similar at the vertex, the two tracks are pointing to distinct regions in the ECAL.

The first selection criterion mostly addresses the case of Bremsstrahlung conversions giving rise to displaced tracks, and is based on the radial distance of the first GSF track hit to the beam line (R_{hit}). For all the preselected GSF track pairs having $\Delta R_{\mathrm{hit}} > 5$ cm, the GSF track closest to the beam line, i.e., with the smallest R_{hit}, is considered to be the primary.
When instead both GSF tracks fulfil $\Delta R_{\mathrm{hit}} < 5$ cm, two cases are considered. If both the GSF tracks are ECAL-driven seeded, the corresponding unique standard super-cluster is available and the GSF track momentum can be compared with the super-cluster energy. The GSF track having the best E/p is considered to be the primary.
If, instead, at least one of the GSF tracks is only tracker-driven, since at least one of the tracks does not have a super-cluster attached, the selection can solely be based on tracking observables and is done in two steps. The first step of this selection consists in comparing the reconstructed charge of a track at the vertex and at the outermost state. Observation of a different charge generally means that different segments of a track have been gathered into one trajectory by the pattern recognition. Such tracks are therefore discarded. The second step is a relaxed version of the trajectory cleaning applied during the pattern recognition. Indeed, in the trajectory cleaning, if two track candidates are found to share more than 50% of their hits, the track candidate

with the largest number of hits or, if they are equal, the one with the best χ^2 is kept. Here, the same selection is applied if two GSF tracks share at least 2 hits and 50% of the tracker sensors.

The number of reconstructed GSF tracks and the GSF track and generated electrons ratio before and after the cleaning is displayed in Fig. 4. After the cleaning, the electron percentage with more than one GSF track reconstructed is 0.7%. This fraction is further reduced to 0.1% when the whole electron reconstruction process is applied (see Sec. 4).



(a) (b)

Figure 4: (a) Number of reconstructed GSF tracks for a sample of single electrons before (continuous-black line) and after (dotted-green line) the cleaning. (b) Pseudo-rapidity (η) profile of the number of reconstructed GSF track per electron before (black triangles) and after the cleaning (green stars).

4 The electron reconstruction algorithm

In order to reconstruct all particles individually, a clustering algorithm able to reconstruct individual electromagnetic showers has been developed for the particle flow, making optimal use of the ECAL granularity. As a consequence, in the case of an electron, the electron cluster and each of the Bremsstrahlung photon clusters are often reconstructed separated. Starting from the GSF track, it is then necessary to set up a strategy to link not only the electron energy deposit, but also the Bremsstrahlung energy deposits to the GSF track. Therefore, as in the whole particle-flow algorithm [1], the link procedure is crucial.

4.1 Linking procedure

In the electron reconstruction, the link algorithm has the role of connecting to each other the elements of the electron: KF and GSF tracks, ECAL, HCAL, and preshower clusters. Because the electron reconstruction within the particle-flow starts from a GSF track, the track-calorimeter cluster link plays a fundamental role. A brief summary of this link method is given below.

A GSF/KF track is linked to a given cluster if the extrapolated position from the outermost tracker measurement in the calorimeter is within the boundaries of one of the cells constituting the cluster. In the ECAL, this check is carried out at the expected depth corresponding to the electron shower maximum, while in the HCAL, it is done on the inner surface of the detector. In addition, the boundaries of the cells are appropriately enlarged to account for the presence of gaps between calorimeter cells, cracks between calorimeter modules, for the uncertainty of the position of the shower maximum, and for the effect of multiple scattering for low-momentum

charged particles.
The preshower-ECAL cluster link is also used; a preshower cluster is linked to an ECAL cluster if the preshower cluster position is within the boundaries of one of the cells of the ECAL cluster. Similarly to the ECAL-track linking, the boundaries of the cells are also appropriately enlarged.

4.2 The Bremsstrahlung recovery strategy

The ability to identify all the energy deposits of an electron, i.e., the electron cluster itself and the associated Bremsstrahlung cluster(s), in the calorimeters is important. Firstly because, if a cluster is missed, it is reconstructed as a primary photon; secondly because the comparison of the energy deposit with the track momentum plays an important role in the electron identification, and finally because the calorimeter energy measurement enters in the final momentum determination of the electron, and it has therefore to be as precise as possible.

The procedure to collect all the calorimeter energy deposits, i.e., the Bremsstrahlung recovery, is driven by the GSF track. For each tracker layer, where the material is mainly localised, a Bremsstrahlung photon emission is sought by computing a straight-line extrapolation, tangent to the track, up to the ECAL, preshower and HCAL. The track-cluster link described previously is used to define a cluster-Bremsstrahlung-photon association. So as to reduce the charged hadron contamination, the ECAL clusters already matched with a KF track are discarded from this procedure. Moreover, to limit the background due to neutral particles, the distance in η between the cluster and the extrapolation should be smaller than 0.015.

The identification of the cluster associated to the electron, and in general the association of the outermost state of the GSF track with the ECAL clusters requires a more sophisticated treatment. Indeed, the GSF track may not be sensitive to a late Bremsstrahlung emission, or the conversion of a late Bremsstrahlung photon in front of the ECAL can produce additional clusters. Therefore, several clusters can be associated with the outermost state of the track. These clusters are required to belong to the same topological cluster [1] as the one directly associated to the GSF track by the linking procedure, they should fulfil $|\Delta\eta(\text{cluster} - \text{track})| < 0.05$ and, as to minimise the charged-hadron contamination, they should not be linked to any other KF track. The cluster closest to the track extrapolation is considered the electron cluster; while the other clusters, if any, are just added in the list of clusters connected to the GSF track.

The procedure in the case of an electron emitting a single Bremsstrahlung photon, is illustrated in Fig. 5 where the GSF track, the extrapolated track tangents, and the ECAL clusters are visible.

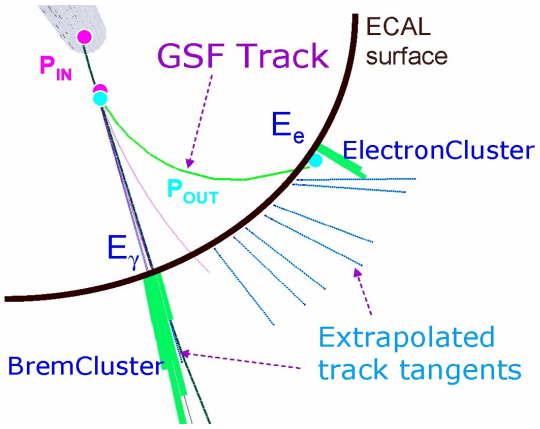


Figure 5: Electron representation using the IGUANA [8] visualisation system.

When the list of the ECAL clusters associated to the GSF track is defined, the preshower clusters associated to these ECAL clusters, are added to the list of the elements constituting the electrons. Finally, if an HCAL cluster linked to the GSF track is found, it is stored in the list of the electron elements. Since an HCAL cluster matching a GSF track can be an indication of a hadron faking an electron, the presence of the cluster is one of the ingredients of the final identification algorithm (Sec. 7), but it is not used to compute the overall electron energy deposit. In the end, all the ECAL and preshower clusters connected to the GSF track, or to one of its tangents, form one particle-flow super-cluster.

4.3 The Bremsstrahlung photon momentum measurement

After the procedure described in the previous section, each Bremsstrahlung photon is connected to a tracker layer. If the point of emission of the photon is not too close neither to the interaction point nor to the ECAL inner surface, the change of curvature of the trajectory at the level of this layer allows the momentum of the photon to be estimated. In the track reconstruction process, for each layer, a forward and a backward fit are computed. For a given tracker layer, the difference between the momenta obtained from the two fits, Δp, is an estimate of the momentum of the hypothetical emitted photon, and can be compared to the associated ECAL cluster energy.

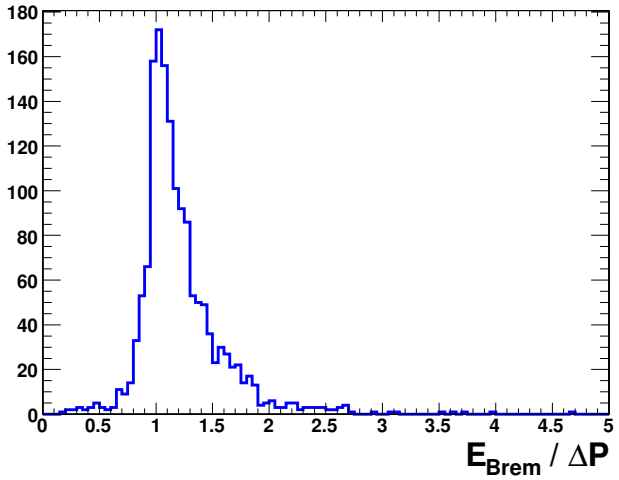


Figure 6: Ratio of the Bremsstrahlung photon energy and Δp as measured from the GSF track for single electrons.

The resulting $E/\Delta p$ ratio is presented in Fig. 6 for photons whose momentum significance, $\Delta p/\sigma_{\Delta p}$, is larger than 5 in a single electron sample.

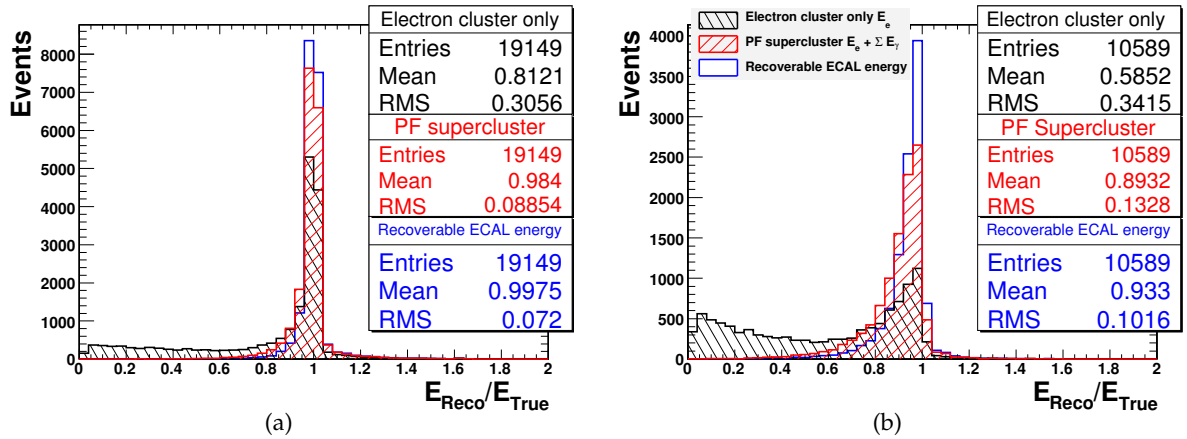
This information is however not used at the moment in the electron reconstruction. Indeed, the Δp is reliably estimated for a small fraction of the electrons, for which at least three tracker measurements are available before and after the emission point, while most of the wrongly associated clusters are linked to the first or last tangent of the GSF track.

4.4 Bremsstrahlung recovery performance

Isolated electrons

To evaluate the Bremsstrahlung recovery performance, the ratio of the energy collected in the particle-flow super-cluster and the true energy of the electron is computed and visible in Fig. 7

for a sample of single electrons with $2 < p_T < 70$ GeV/c. . In addition, the ratio of the maximum recoverable energy, i.e., the sum of the energies of all the particle-flow clusters contained in a $|\eta_{\text{cluster}} - \eta_{\text{GSF-track}}| < 0.2$ and $|\phi_{\text{cluster}} - \phi_{\text{GSF-track}}| < 1.5$ region and the electron true energy is also presented in Fig. 7.



(a) (b)

Figure 7: Fraction of the electron initial energy collected at the level of the calorimeter with different strategies: taking only the cluster immediately associated to the track, with the Bremsstrahlung recovery procedure, taking all the clusters in a wide region around the track. (a) low material budget region $|\eta| < 1$, (b) high material budget region $1 < |\eta| < 1.6$

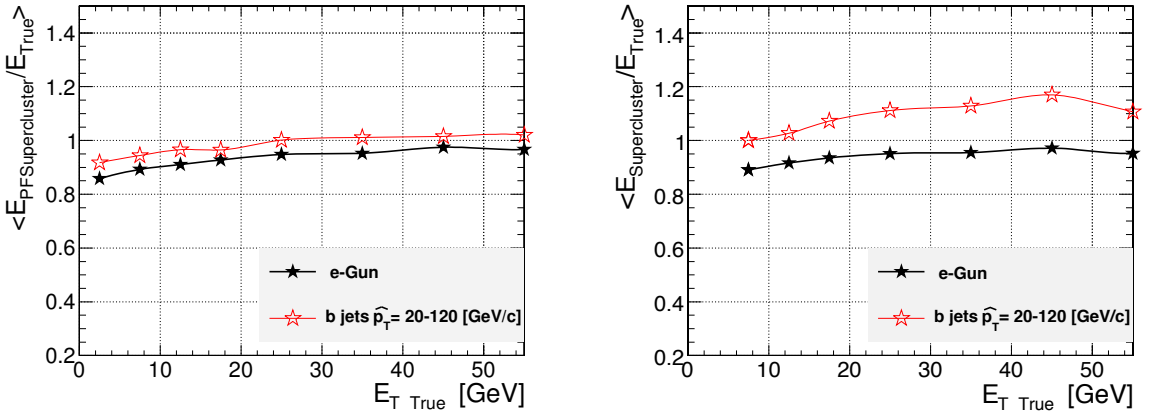
In the region with low tracker material budget ($|\eta| < 1.0$) the particle-flow super-clusters contain on average 99% of the recoverable ECAL energy while in the region with high tracker material budget ($1.0 < |\eta| < 1.6$) the fraction of energy recovered is 96%. The inability of the track tangent method to recover the total energy essentially comes from Bremsstrahlung photons which afterwards convert into an electron-positron pair. Indeed the ECAL energy deposits from the converted Bremsstrahlung photons, because of the high magnetic field of CMS and the typical low momentum of the e^+e^- pair, can fall outside of the Bremsstrahlung recovery region. In extreme cases, the electrons from conversions may not have a transverse momentum large enough to reach the ECAL, and loop inside the tracker. This behaviour is the reason for the recoverable energy being lower than the initial electron energy, and is important in the region of high material budget.

Electrons in jets

The energy collected by the particle-flow super-cluster should be ideally identical for isolated and non-isolated electrons. The mean energy of the particle-flow super-cluster divided by the true energy of the electron ($< E_{\text{Reco}}/E_{\text{True}} >$) is shown in Fig. 8a as a function of the true transverse energy ($E_{\text{T True}}$) both for a sample of isolated and non isolated electrons. A sample of electrons in b jets with $20 < \hat{p_T} < 120$ GeV/c has been considered for the non-isolated case, while a single electron gun reproducing the kinematics of the electrons in b jets has been used to get a reference sample. In Fig. 8b, the same plot is displayed for the standard super-cluster (Sec. 2).

For non-isolated electrons the standard super-clusters collect up to 20% of additional energy with respect to isolated electrons. This effect is caused by the contribution of charged hadrons and photons from the jets absorbed by the super-clustering algorithm. The particle-flow super-clusters get on average only 5% of additional energy. They are less affected by the jet contribution for two reasons: there is a specific veto on the clusters associated to other KF tracks

(charged hadrons) and the particle-flow super-clusters are delimited in space by the first track tangent and the GSF track impact points on the ECAL surface.



(a) (b)

Figure 8: Recovered raw energy normalised by the true energy in the isolated and non-isolated case, using the particle-flow super-clusters (a) and the standard super-clusters (b).

5 Electron cluster energy correction

The non-linearity at low energy due to the thresholds applied in the clustering as well as containment effects are firstly corrected for individual particle-flow clusters using energy- and $|\eta|$-dependent corrections which have been determined using unconverted photons [1]. The lack of containment of the particle-flow super-clustering, mainly due to converted Bremsstrahlung photons as described in Sec. 4, is corrected following a procedure similar to the one set up for the standard super-cluster corrections [3, 9].

Two million double-electron events generated with a flat p_T between $2 - 300$ GeV/c and a flat η distribution have been simulated using the CMS Fast Simulation program [10]. The $E_{\text{Reco}}/E_{\text{True}}$ distributions have been fitted with a Crystal Ball function in order to extract the most probable value (MPV) in various bins in (E_T, η), where E_T is the reconstructed super-cluster transverse energy. In the barrel region ($0 < |\eta| < 1.5$), the obtained MPV's are then fitted in each slice in E_T with a smoothed step function :

$$f(E_T, \eta) = p_0(E_T) - a \times |\eta| + p_1(E_T) \times \text{atan}[b \times (c - |\eta|)]. \quad (1)$$

where $a = 0.00181$, $b = 15$ and $c = 1.08$. It has been checked that a and c, taken from [3] are suitable for particle-flow clusters, while the b coefficient has been slightly adjusted. The dependency on η and E_T of the $E_{\text{Reco}}/E_{\text{True}}$ ratio MPV is well visible in Fig. 9.

Afterwards $p_0(E_T)$ and $p_1(E_T)$ are fitted (Fig. 10) with the functions:

$$p_0(E_T) = c_0 + c_1/(c_2 + E_T) + c_3/E_T \quad (2)$$

and

$$p_1(E_T) = c_4 + c_5/(c_6 + E_T). \quad (3)$$

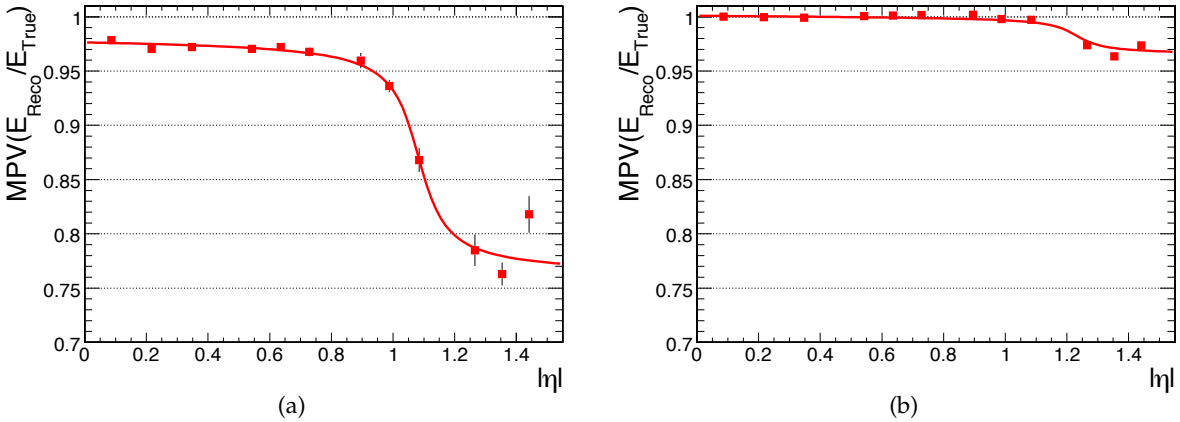


Figure 9: Most probable value of the $E_{\text{Reco}}/E_{\text{True}}$ ratio as a function of the pseudo-rapidity in two E_{T} slices $4-6$ GeV (a) and $46-48$ GeV (b). The results have been fitted with a smoothed step function.

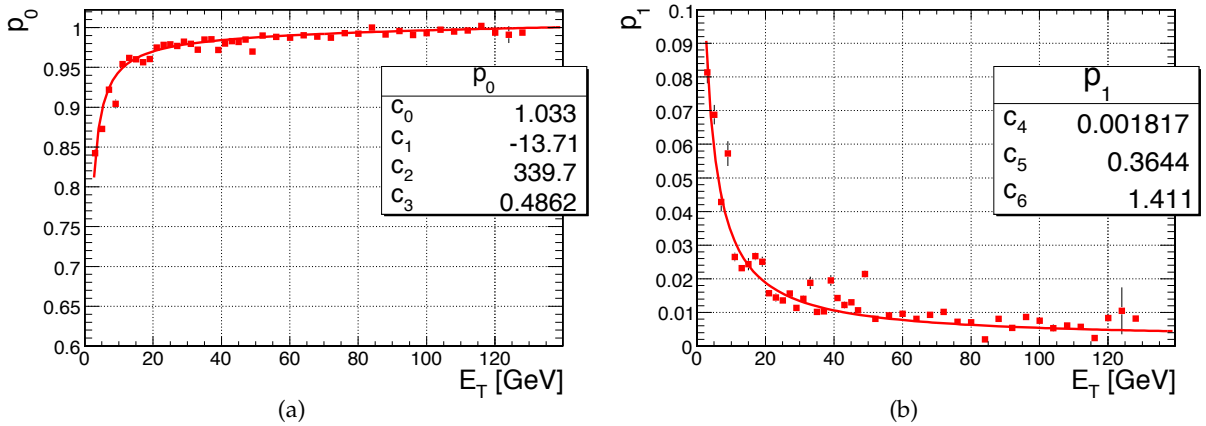


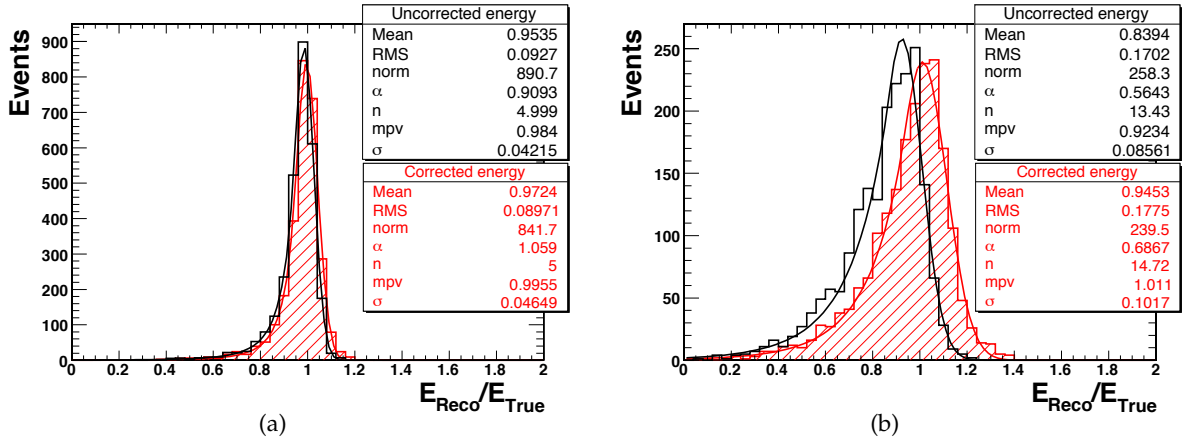
Figure 10: The p_0 (a) and p_1 (b) coefficients of the correction functions for the barrel region as a function of E_{T}.

In the endcaps the same procedure is applied with the difference that each slice in E_T is fitted with a second order polynomial:

$$f(E_T, \eta) = p_2(E_T) + p_3(E_T) \times |\eta| + p_4(E_T) \times |\eta|^2 \quad (4)$$

and each of the $p_i(E_T)$ is then fitted with the same functional form of eq. 3.
Finally, the corrected super-cluster energy is obtained as $E_{Reco}/f(E_T, \eta)$.

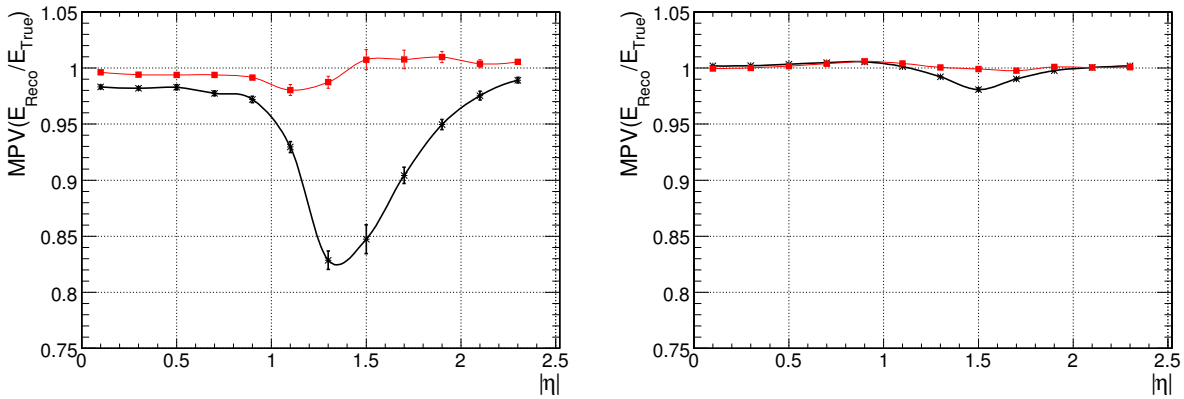
For low-E_T electrons ($E_T < 10$ GeV), where the corrections are larger, the distributions of the ratio E_{Reco}/E_{True} for the corrected and uncorrected reconstructed energy are exemplified in Fig. 11 for two different η regions.



(a) (b)

Figure 11: E_{Reco}/E_{True} distributions for the uncorrected (black, empty) and corrected (red, dashed) reconstructed energy for low $E_T < 10$ GeV electrons in two different η regions: $0 < |\eta| < 0.5$ (a) and $1.5 < |\eta| < 2.0$ (b).

The MPV of E_{Reco}/E_{True} before and after corrections is shown in Fig. 12 as a function of $|\eta|$ for two different E_T regions, $E_T < 10$ GeV and $10 < E_T < 100$ GeV. A 0.5% maximum deviation



(a) (b)

Figure 12: MPV for the E_{Reco}/E_{True} as a function of $|\eta|$ for $E_T < 10$ GeV (a) and for $10 < E_T < 100$ GeV (b) before (black curve with stars) and after (red curve with squares) corrections.

of the MPV is observed for electrons with $E_T > 10$ GeV. Below this E_T the maximum deviation increases up to 2%.

After these corrections, even though the MPV of $E_{\text{Reco}}/E_{\text{True}}$ is close to one, the mean value is still lower than unity (Fig. 11) because of the asymmetric shape of the distribution and the relative resolutions are not improved. Two directions to make the shape symmetric are envisioned. The first one consists in improving the Bremsstrahlung recovery, in particular, the recognition of converted Bremsstrahlung photons. The second one would consist in using cluster-shape-based corrections.

6 Electron momentum determination

The combination of the tracker momentum with the ECAL energy measurements leads to an improvement of the electron momentum estimation, especially for low-p_{T} electrons. In the GSF algorithm, as described in Sec. 2, the Bremsstrahlung emission is modelled by a Gaussian mixture for both the trajectory states and the energy losses. The track parameters can then be determined either from a weighted mean of all the components, or from the most probable value (or mode), as to give more importance to the component with the largest weight [9].

For electrons whose GSF track has at least eight reconstructed hits, the electron momentum determination is performed using a weighted mean between the corrected ECAL energy measurement, E_{corr}, and the momentum of the GSF track at the vertex estimated with the mode calculation, similarly to the method described in [9]. The weight associated to the track is the error on the momentum measurement, obtained from the error matrix, while the weight of the ECAL energy corresponds to the expected resolution on the total super-cluster energy. There are however two exceptions to this rule.

- The track fit is not always sensitive to an early Bremsstrahlung emission, while the corresponding photon can be recovered in the ECAL. Therefore, if both the KF and GSF momenta (with the mode calculation) lack more than 5% of energy with respect to the ECAL measurement, the electron energy is only determined from the ECAL.
- As explained earlier, in spite of the careful clustering and Bremsstrahlung recovery, some clusters might have been missed. Therefore, if the particle-flow super-cluster lacks more than 10% of energy with the respect to the track measurement, and if the track is well measured, the momentum determination is based only on the GSF track measurement. More precisely, the conditions to use only the GSF track momentum are:
  - GSF $p_{\text{mode}} > 1.1 \times E_{\text{corr}}$
  - GSF $p_{\text{mean}} > 1.1 \times E_{\text{corr}}$
  - KF $p > E_{\text{corr}}$
  - GSF $p_{\text{T mode}} < 30$ GeV/c
  - GSF $p_{\text{mode}}/\sigma_{p_{\text{mode}}} > 5$

When instead the GSF track is short (#hits < 8), the momentum of the KF track is more precise than the GSF, and the track momentum estimate is performed using a weighted mean between the ECAL energy measurement and the KF track momentum at vertex.

The performance of the combined measurement is illustrated in Fig.13 where the invariant mass of di-electron pairs from a J/Ψ or an Υ decay are shown using the corrected ECAL energy only or the electron momentum estimation described above. In these plots, the electron identification criterion described in the next section is applied.

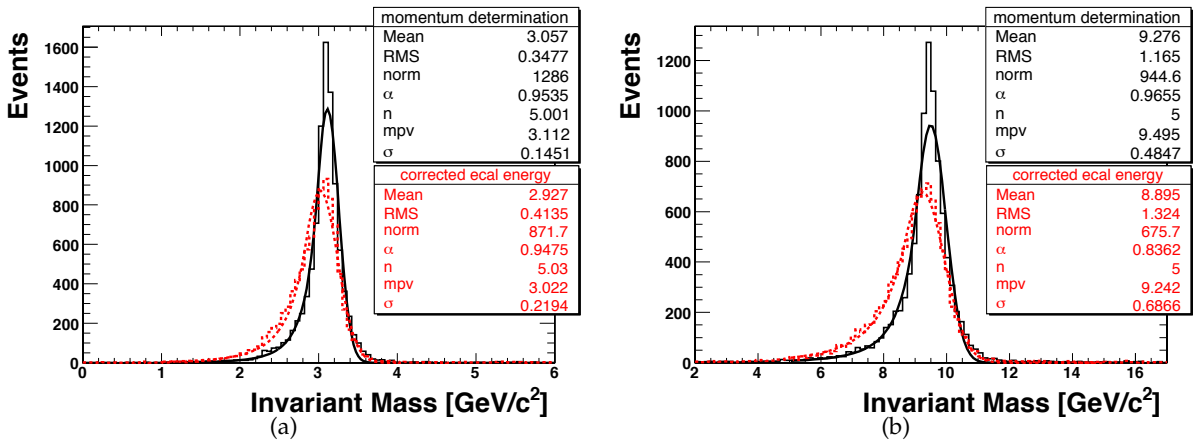


Figure 13: Invariant mass of identified electron pairs in $J/\Psi \rightarrow e^+e^-$ (a) and $\Upsilon \rightarrow e^+e^-$ (b) samples computed with the electron momentum determination algorithm (solid line) or the corrected ECAL energy (dotted line).

7 The electron identification algorithm

The information of all the CMS sub-detectors, obtained during the reconstruction step (Sec. 4) are used for building observables for the electron identification.

7.1 Electron identification observables and multivariate analysis

The most important observable for electron-charged-hadron discrimination is the energy matching between the tracker and ECAL. Therefore, if no ECAL cluster could be associated to the GSF track, no electron identification is performed. The particle-flow super-cluster contains, by construction, the information of the ECAL cluster associated to the GSF track and the ECAL cluster(s) associated to the Bremsstrahlung track tangent(s). Using this information three discriminating variables based on the energy matching can be built for each electron candidate.

As illustrated on Fig. 5, these variables are E_e/p_{out}, the ratio between the electron cluster energy and the track outer-momentum, $\sum E_\gamma/(p_{in} - p_{out})$ the ratio between the Bremsstrahlung photon energy as measured by the ECAL and by the track, and $(E_e + \sum E_\gamma)/p_{in}$, the total ECAL energy matching with the inner track momentum.
The Bremsstrahlung recovery procedure also brings additional information on the ECAL-track energy matching. If a Bremsstrahlung emission occurs before the electron has crossed at least three tracker layers, the GSF track p_{in} is often underestimated with respect to the initial electron momentum and the super-cluster energy does not match anymore the initial momentum measured by the GSF track, $(E_e + \sum E_\gamma) > p_{in}$. In such case, an ECAL cluster is however linked to one of the first three track tangents. This bias is flagged with the *EarlyBrem* variable. Alternatively, in the case of a late Bremsstrahlung emission, the ECAL particle-flow clustering might not be able to disentangle the overlapping showers, while the tracking is still able to determine correctly the momentum at the outermost state, and it leads to $E_e > p_{out}$. In this case the azimuthal position of the merged particle-flow cluster (electron and the late Bremsstrahlung contribution) is no longer centred with respect to the GSF track extrapolation in the ECAL surface but displaced towards one of the last track tangents. This situation is flagged with the *LateBrem* variable. The tracker and ECAL information are also used to match in pseudo-rapidity the ECAL cluster and the GSF track ($|\eta_{GSF} - \eta_{cluster}|$).

Then, calorimeter-related variables are used. The lateral shower shape of the ECAL cluster associated to the electron is expected to be narrow as is the case for genuine electromagnetic showers. Because the Bremsstrahlung contribution enlarges the lateral shape in the azimuthal direction, only the shower width along η, measured by the $\sigma_{\eta\eta}$ variable [9] is used. The hadron fraction of the shower, $H/(H+E_e)$, is also computed where H is the energy of the hadron cluster linked to the GSF track.

The reconstructed observables mentioned so far are shown in Fig. 14 for electrons and pions. The signal sample consists in electrons, both isolated and non-isolated, extracted from $Z \rightarrow e^+e^-$ events and from b jets (with $\hat{p_T} = [50-3000]$ GeV/c). The background consists in pions in the same b jet sample. The signal and background samples have been re-weighted as to obtain a flat distribution in $ln(p_T)$ and η.

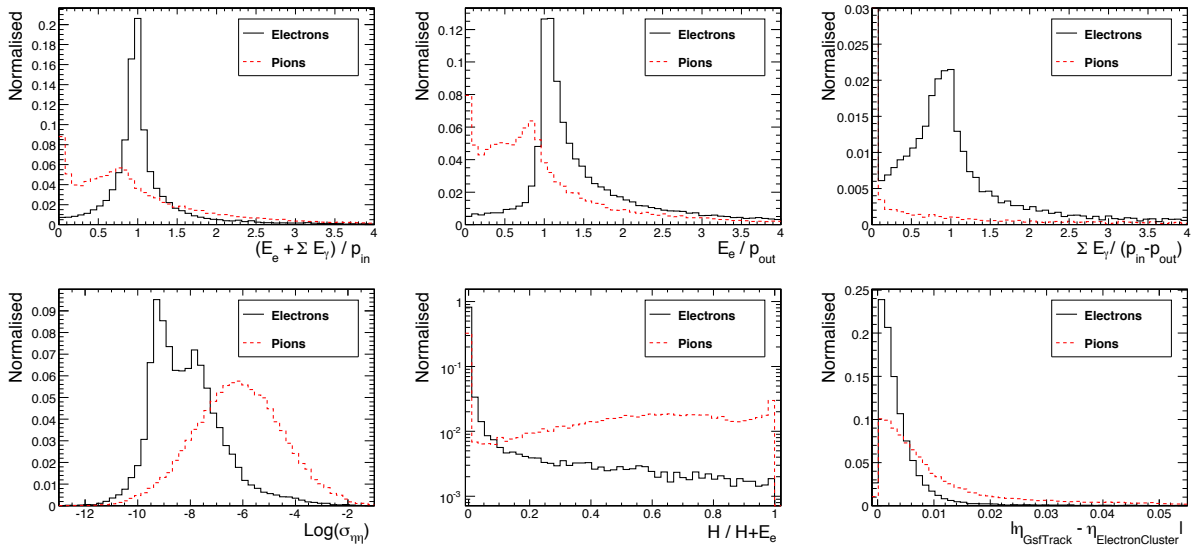


Figure 14: Electron identification variables for electrons (black continuous line) and pions (red dashed line). On the upper plots, the calorimeter-track matching variables, from left to right: E_{tot}/p_{in}, E_e/p_{out} and $\sum E_\gamma/(p_{in}-p_{out})$. On the bottom and from left to right: $\log(\sigma_{\eta\eta})$, hadronic fraction and $|\eta_{GSF}-\eta_{cluster}|$. All the histograms are normalised to unity.

For pions, the E/p selection applied in the pre-identification (Sec. 2) has a visible effect on the $(E_e+\sum E_\gamma)/p_{in}$ and E_e/p_{out} distributions. The double-peak structure in the $\sigma_{\eta\eta}$ distributions is due to the different crystal-size in the barrel and in the endcaps.

Finally, to further improve the separation power between electron and charged hadrons, pure tracking observables, both for the GSF and KF track, are considered. When the electron radiate Bremsstrahlung photon(s), its GSF track momentum at the outermost state is smaller than the momentum at the innermost state. The so-called $f_{brem} = (p_{in}-p_{out})/p_{in}$ [9] variable, measured with the GSF track, has then a flat distributions for electrons while for charged hadrons it peaks at 0 (Fig 15a). Moreover, as mentioned in the Sec. 2, the KF track from electrons tends to have a larger χ^2_{KF} (Fig 15b) and a smaller number of reconstructed tracker hits, $\#hits_{KF}$ (Fig 15c). The value 0 for the χ^2_{KF} and $\#hits_{KF}$ is due to the GSF tracks that have not a KF track associated. The resolution σ_{p_T}/p_T and the χ^2 of the GSF track are also included in the list of pure tracking observables. They do not show a good discrimination power between electrons and pions, but they are used to quantify the reliability of the parameters measured by the GSF track.

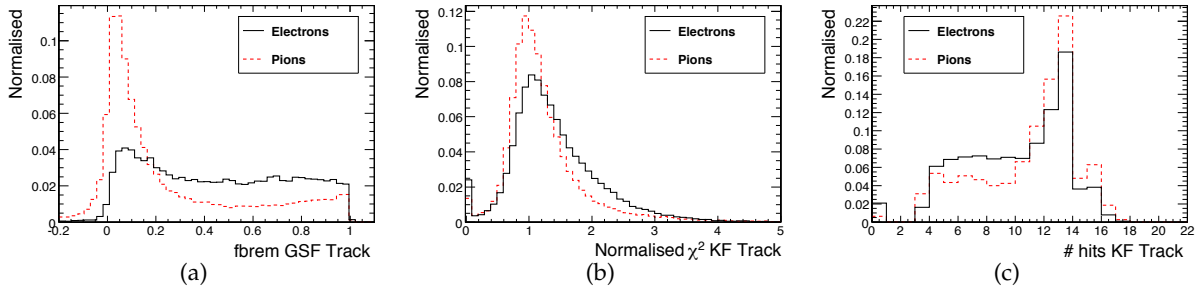


Figure 15: Tracking variables used for the electron identification: Bremsstrahlung fraction (a), χ^2 of the KF track (b), number of hits of the KF track (c). The histograms are normalised to unity.

The final list of the observables for the electron identification is given below:

- $(E_e + \sum E_\gamma)/p_{in}$
- E_e/p_{out}
- $\sum E_\gamma/(p_{in} - p_{out})$
- $EarlyBrem$
- $LateBrem$
- $\log(\sigma_{\eta\eta})$ (only for the ECAL cluster linked to the GSF track)
- $|\eta_{GSF} - \eta_{cluster}|$ (only for the ECAL cluster linked to the GSF track)
- $H/(H + E_e)$
- $f_{brem} = (p_{in} - p_{out})/p_{in}$
- χ^2_{KF} and χ^2_{GSF}
- $\#hits_{KF}$
- GSF σ_{p_T}/p_T
- GSF $ln(p_T)$
- GSF η

These observables are combined into a single discriminator using a multivariate Boosted Decision Trees (BDT) method. The training has been done using the signal and background samples previously described.

The BDT output for isolated electrons ($Z \rightarrow ee$) and non-isolated electrons and pions (b jets $\hat{p_T}[20-120]$ GeV/c) is shown in Fig. 16 for all the GSF tracks with $p_T > 2$ GeV/c. Even though the pions which enter the electron identification have been selected by the seeding procedure, and are therefore electron-like, a very good separation is achieved when the electrons are isolated. The pion-electron discrimination remains good in jets.

7.2 Efficiency on non-isolated electrons

The efficiency for non-isolated electron from B hadron decay and pions, in a b jet sample with $\hat{p_T}[20-120]$ and $t\bar{t}$ events, is shown in Fig. 17 for different values of the BDT output. In the same figure the particle-flow working point (BDT output > -0.1) is also displayed, as well as the results of the cut-based identification algorithm [11] developed for isolated electrons. The efficiencies are relative to the number of generated electron or pions with $p_T > 2$ GeV/c in

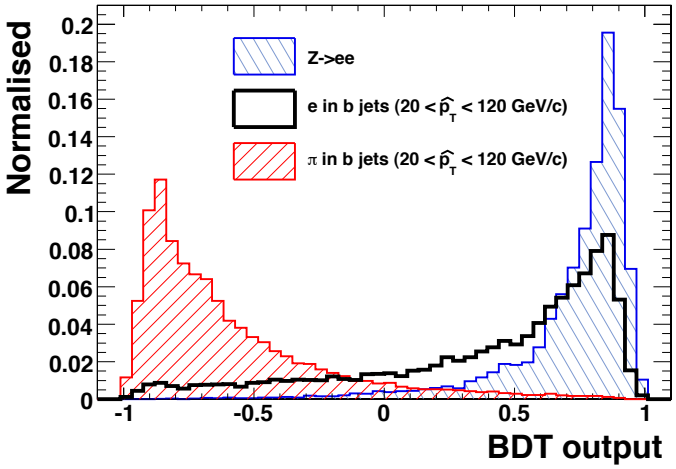


Figure 16: Output of the multivariate analysis for isolated electrons in a $Z \rightarrow ee$ sample, for non-isolated electrons in b jets and for pions. The histograms are normalised to unity.

the tracker acceptance $|\eta| < 2.5$. In b jet events, a 65% efficiency on electrons can be reached with less than 1% of pion contamination. The slightly worse results in $t\bar{t}$ events are due to the larger boost of the jets. For jets with high momentum the ECAL clustering is not always able to separate each particle contribution. The particle-flow super-clustering absorbs energy from photons and other neutral hadrons and consequently the observables based on the track and ECAL energy matching start to be degraded. In addition, the overlap of charged hadrons with photons increases the probability to have fake ECAL-track links, thus worsening the charged hadron rejection.

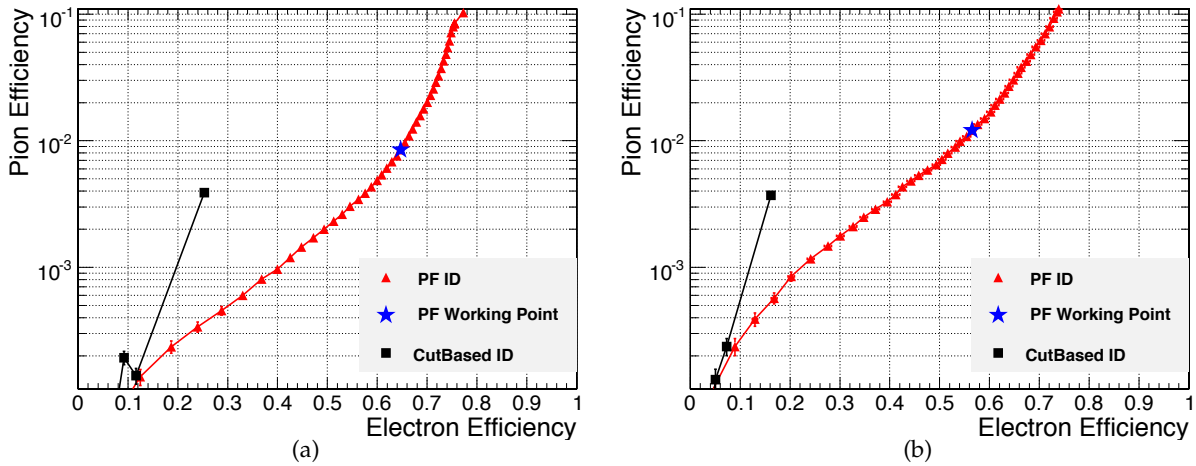


Figure 17: Efficiency for non-isolated electrons from B hadron decay and pions in a b jet sample $\hat{p_T}[20-120]$ (a) and in $t\bar{t}$ events (b). The red triangle markers represent different values of the electron identification cut (BDT) applied, the default value (BDT output > -0.1) used in the particle flow is indicated with a blue star while different cut-based optimisations developed for isolated electrons are represented with black squares.

8 Impact on the Particle Flow jet resolution

The impact of the electron identification and reconstruction in the particle-flow performance with jets has been evaluated with a sample of di-jets events and a sample of electron-enriched b jet events; both samples have been generated with a quasi-flat $\hat{p_T}$ spectrum.
The electron-enriched b jet sample has been obtained requiring at least an electron from a B decay with a generated p_T greater than 2 GeV/c. Jets have been reconstructed with the iterative-cone algorithm [12] with a cone size of 0.5 in the (η, ϕ) plane, from two types of inputs: all generated stable particles, except neutrinos, give rise to so-called "generated jets"; all particles reconstructed with the particle-flow algorithm are clustered into "particle-flow jets". The particle-flow jets are then matched to the closest generated jet in the (η, ϕ) plane.
Ideally, the jet response defined as the mean value of a Gaussian fit obtained from the distribution of $((p_T^{PFlow} - p_T^{Gen})/p_T^{Gen})$ should be identical between light jets, which contain little genuine electrons, and heavy-flavour jets, which have an electron-enriched content.

The jet response for QCD multijets events and the electromagnetic-enriched b jet sample, with and without the particle-flow electrons, is shown in Fig. 18 for the barrel region. The endcap region shows similar results. As intended, the jet response does not change for the QCD mul-

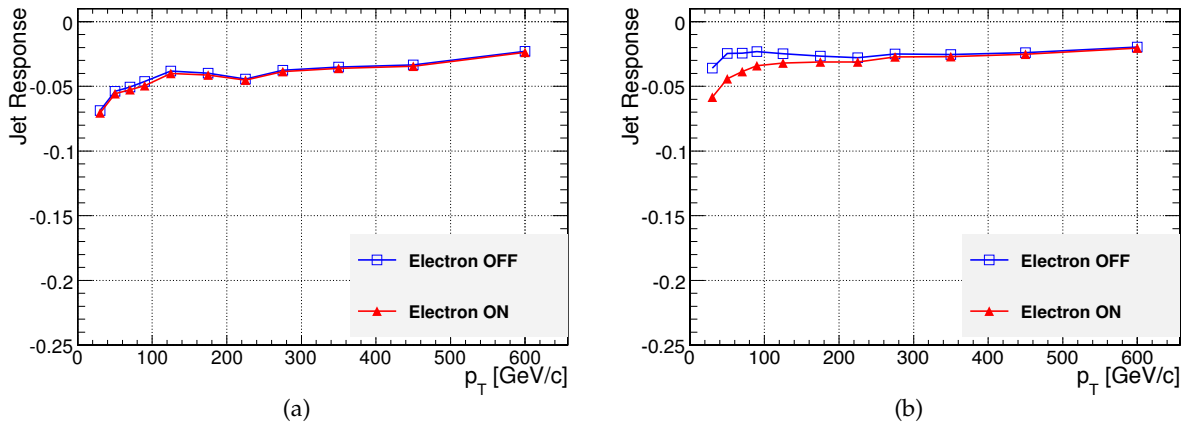


Figure 18: Jet Response in the barrel as a function of p_T for QCD multijets (a) and electromagnetic-enriched b jet (b) with/without the electron reconstruction.

tijets events when the particle-flow electron reconstruction is turned on.
In b jets, when the particle-flow electron reconstruction is not activated, an artificial increase of the jet response is observed. This is explained by two effects. When the reconstructed electrons are not included in the particle-flow algorithm, electrons are reconstructed as charged hadron candidates plus photon candidates in the case of Bremsstrahlung emissions. Because the momentum of the charged hadron candidate is mainly measured at the vertex by the reconstructed track, the momentum carried by the photon candidates results in a double counting. Moreover, when reconstructed electrons are not included, the energy corrections provided for the charged hadrons [1] are wrongly applied also to electrons. When instead the electron reconstruction is performed the energy deposit(s) due to the Bremsstrahlung emission(s) are connected to the electron energy deposit forming the particle-flow super-cluster and the momentum of the electron is given by the combination of the super-cluster energy and the GSF track momentum at the vertex.
Finally, the jet-energy resolutions, defined as the fitted width of the $((p_T^{PFlow} - p_T^{Gen})/p_T^{Gen})$ distribution, are unchanged in QCD di-jets and improved by 10% at $p_T = 20$ GeV/c in the

electromagnetic-enriched b-jet sample when the electron reconstruction is activated (Fig. 19).

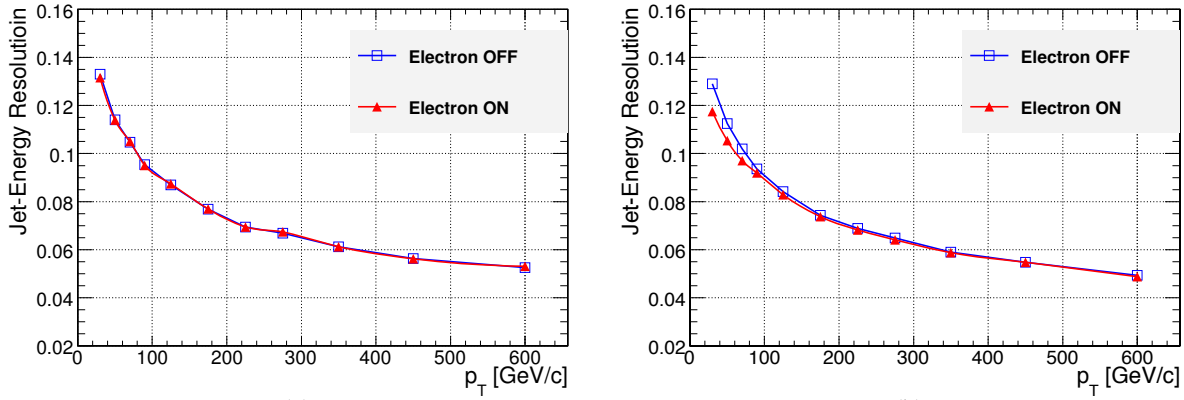


Figure 19: Jet-energy Resolution in the barrel as a function of p_T for QCD multijets (a) and electromagnetic-enriched b jet (b) with/without the electron reconstruction.

Conclusion

The electron identification and reconstruction within the particle-flow algorithm has been described. It consists in two main parts: a tracker-driven seeding for GSF tracks focused on low p_T or non-isolated electrons, complementing the ECAL-driven seeding; and an electron reconstruction and identification included in the core of the particle flow algorithm. The key step of the reconstruction, i.e., the Bremsstrahlung recovery has been explained and its performance studied. The super-cluster thus obtained allows calorimeter-track matching variables to be easily built. The resulting observables are combined with the electron track properties in a multivariate analysis.

The reconstructed electron momentum results from a combination of the ECAL and track measurements. The performance of the electron identification in jets has been presented. In a b jet sample with $\hat{p_T}[20-120]$ GeV/c, the efficiency reaches 65% for less than 1% of pion misidentification. The impact on particle flow jet resolution and response has finally been evaluated. The performance on light jets are unaffected, and a 10% improvement in the energy resolution is achieved on electron-enriched b jets.

9 Acknowledgements

We wish to thank the CMS particle-flow and e/γ groups for their support. We also would like to thank Yurii Maravin for the kind help on the usage of the super-cluster energy-correction algorithm and Roberto Covarelli, Stephen Reucroft, Claude Charlot for the reading of the first draft.

References

[1] **CMS** Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus and E_T^{miss}", *CMS Note* **PFT 2009/001** (2009).

[2] **CMS** Collaboration, "The CMS experiment at the CERN LHC", *JINST* **0803** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[3] W. Adam et al., "Electron Reconstruction in CMS", *CMS Analysis Note* **AN 2009/164** (2009).

[4] W. Adam, R. Fruhwirth, A. Strandlie, T. Todorov, "Reconstructions of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", *Journal of Physics G: Nuclear and Particle Physics* **31** (2005) N9. doi:10.1088/0954-3899/31/9/N01.

[5] M. Pioppi, "Electron Pre-identification in the Particle Flow framework", *CMS Analysis Note* **AN 2008/032** (2009).

[6] M. Pioppi, "Iterative tracking", *CMS Internal Note* **IN 2007/065** (2007).

[7] **CMS** Collaboration, "Track Reconstruction in the CMS tracker (in preparation)", *CMS Note* **TRK 2009/001** (2009).

[8] G. Alverson, G. Eulisse, S. Muzaffar, I. Osborne, L. Taylor, L. A. Tuura, "IGUANA: A high-performance 2D and 3D visualisation system", *Nucl. Instr. and Meth.* **A 534** (2004) 143–146.

[9] S. Baffioni et al., "Electron Reconstruction in CMS", *Eur. Phys. J. C* **49** (2007) 1099. doi:10.1140/epjc/s10052-006-0175-5.

[10] **CMS** Collaboration, "CMS physics : Technical Design Report", *CERN/LHCC* **CERN-LHCC-2006-001** (2006).

[11] S. Baffioni et al., "Electron Identification in CMS", *CMS Analysis Note* **AN 2009/178** (2009).

[12] A. Heister et al., "Measurement of Jets with the CMS Detector at the LHC", *CMS Note* **2006/36** (2006).