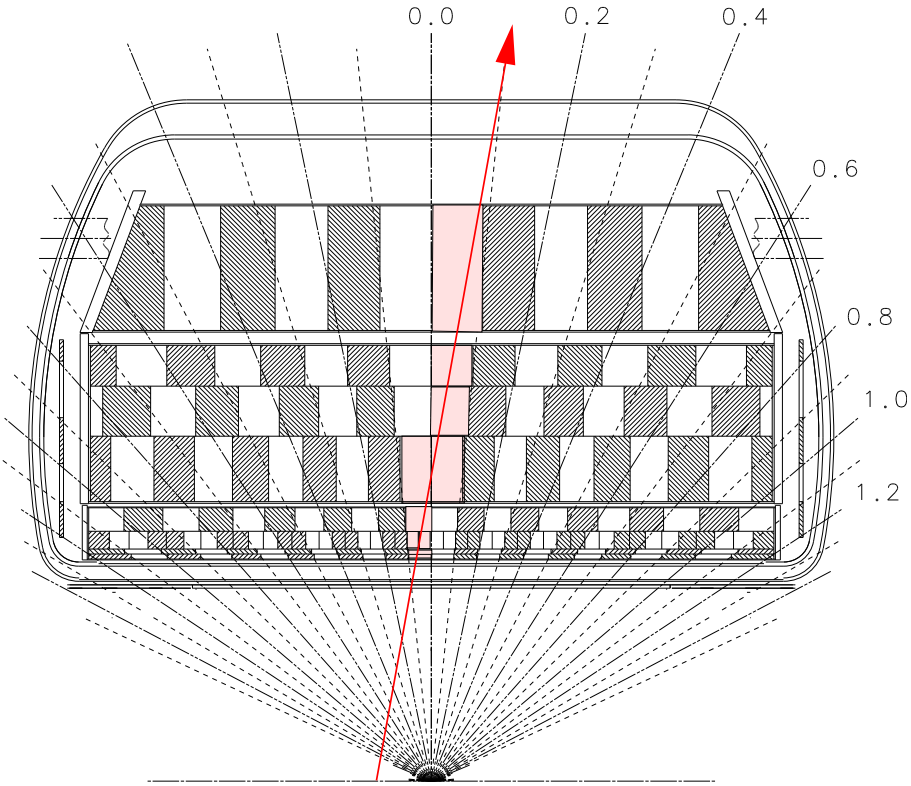
The Road Method

(an algorithm for the identification of electrons in jets)

F. Beaudette and J.-F. Grivaz
Laboratoire de l'Accélérateur Linéaire – Orsay

D0 note 3976

April 11, 2002



Abstract

The road method is an algoritm for the identification of electrons in jets which is based on the extrapolation of charged tracks into the calorimeter. Calorimeter cells are collected around the track extrapolations, and the transverse energies within the roads thus defined are computed in each floor. Electron identification estimators are constructed. The performance of this method for the tagging of heavy-flavour jets is evaluated in terms of efficiency and purity.

1 Introduction

The standard electron reconstruction and identification procedure in D0 is based on calorimeter clusters. It relies essentially on the characteristic transverse and longitudinal shapes of electromagnetic showers [1]. While this method is efficient for high energy isolated electrons, such as those arising from W or Z decays, it fails for electrons belonging to hadron jets, such as those from semileptonic heavy-flavour decays. The reason is that hadron and electron showers tend to overlap in collimated jets so that the electron cluster characteristics are obscured (Fig. 1–left).

An alternative electron identification procedure, the *road method*, has therefore been designed to cope with these configurations, which takes advantage of the calorimeter granularity and of the improved tracking capabilities of the Run-II D0 detector. The road method is based on the extrapolation of reconstructed charged particle trajectories into the calorimeter. For a given track, only the energy contained in a narrow tube, or *road*, along the track extrapolation is considered. The contribution from neighbouring hadronic showers (or even from extraneous photons) is therefore reduced (Fig. 1–right).

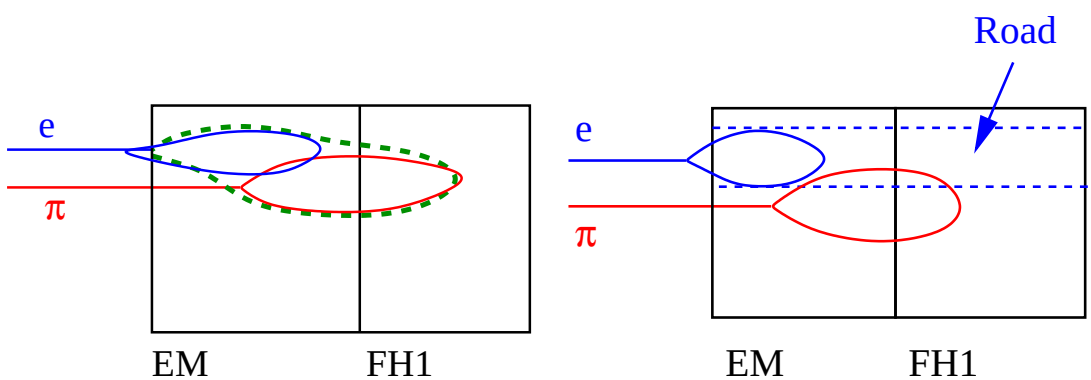


Figure 1: An electron shower and a pion shower overlap in the calorimeter. The two clusters are merged (left), while the pion contribution to the road is reduced (right).

The road method was first developed for the central calorimeter, and its performance for the tagging of b quarks was reported in Ref. [2]. Since then, it has been extended to the end calorimeters, and the code has been integrated in the standard D0 soft electron reconstruction package `SEMReco` [3].

A description of the road algorithm is given in the next section, followed by an example of application to the tagging of heavy flavours.

2 The road algorithm

2.1 Algorithm overview

The road algorithm consists of the following steps:

- for each reconstructed charged particle track selected, extrapolate its trajectory through the calorimeter;
- in every floor of the calorimeter, establish the list of cells belonging to the road;
- in every floor, calculate the transverse energy contained in the road.

The initial selection of charged tracks to be delivered to the road algorithm is performed in `SEMReco` as explained in Ref. [3]. (Presently, a p_T cut of 1.5 GeV/c is applied, and a matching calorimeter tower with a transverse energy of at least 0.5 GeV and an electromagnetic fraction larger than 70% is required.)

The identification criteria using the energies within the road in the various floors do not belong to the algorithm as such. They depend on the physics process considered, but general guidelines and a detailed example will be given in this note.

The current version of the algorithm does not make use of the preshowers, nor of the Inter Cryostat Detector (ICD), nor of the Massless Gaps (MGs).

2.2 Track extrapolation

The parameters of the selected tracks are taken at the point of closest approach with respect to the z-axis. A uniform magnetic field equal to the value at the center of the detector, taken from the database, is assumed throughout a cylinder of length 260 cm and radius 58.7 cm, corresponding to the inner radius of the coil. The residual magnetic field outside this cylinder is ignored. The extrapolated trajectory is therefore approximated by an helix inside the cylinder and by a straight line outside. The intersections with the various calorimeter floors listed in Appendix A are calculated at the level of their front, central and back planes.

2.3 Road construction

If the track extrapolation intersects the central plane of the third floor at an $\eta_{\rm det}$ value such that $|\eta_{\rm det}| < 1$, a *Central Road* is constructed by the the `RoadFinder` program. Otherwise, in the forward region where the geometry is more intricate, the `ECRoadFinder` program is used instead.

The road construction algorithm was designed using a simulated sample of single electrons, uniformly distributed in η between 0 and 3.2 and in energy between 5 and 25 GeV; the vertex distribution along the beam axis is a Gaussian with width $\sigma_z = 28$ cm. The ratio of the total energy collected in the calorimeter to the generated electron energy is displayed in Fig. 2 for electrons in the central ($\eta_{\rm det} < 1$) and in the end ($1.5 < \eta_{\rm det} < 3.2$) calorimeters. (For historical reasons, no *calweights* were applied for these distributions nor for the similar ones in the following.) The distribution of the fraction of calorimeter energy

contained in the first three calorimeter floors is shown in Fig. 3. Average fractions of 90 and 95% are obtained in the central and forward regions, respectively, the difference being largely due to the cracks in the central calorimeter. Since most of the energy deposited by electrons in the calorimeter is contained in the first three floors, the construction of the road has been optimized using the features of the energy deposits in those floors.

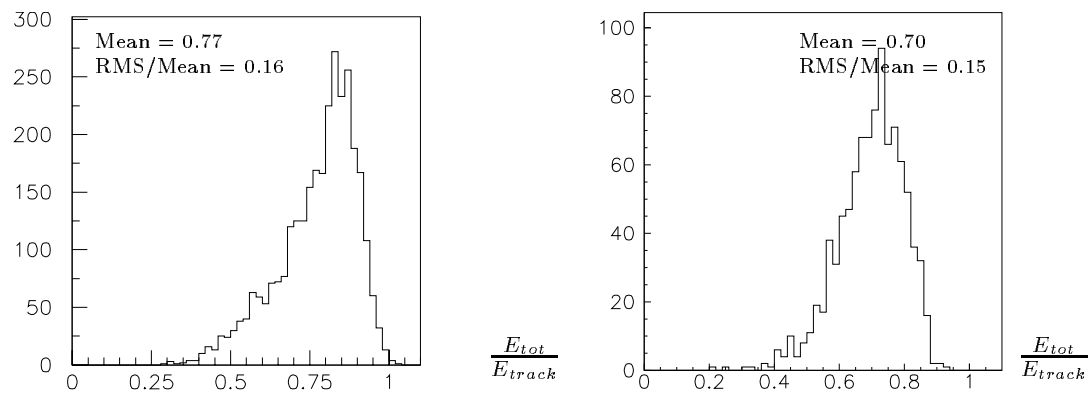


Figure 2: Ratio of the total energy in the calorimeter to the generated electron energy for the central (left) and end (right) calorimeters.

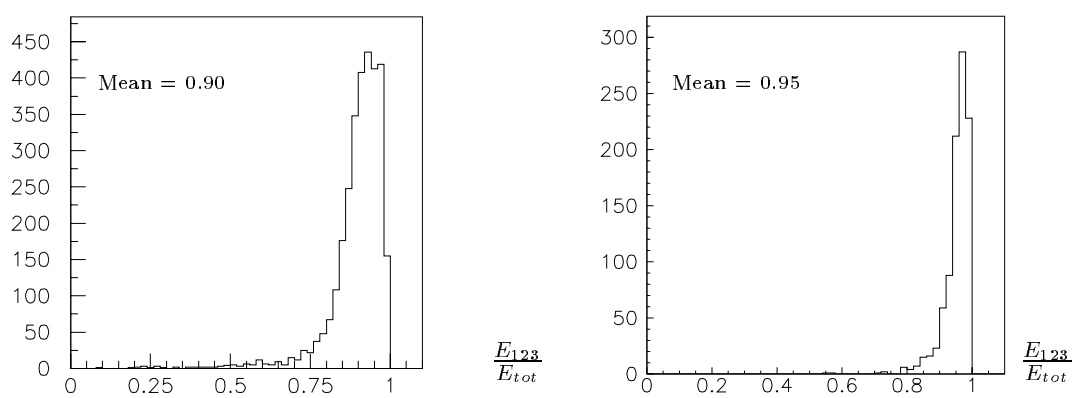


Figure 3: Fraction of the calorimeter energy contained in the first three floors for the central (left) and end (right) calorimeters.

2.3.1 Central Road

The geometry of the central calorimeter (CC) is sufficiently simple, so that the central road can be calculated without making use of the accurate but rather slow `calorimeter_geometry` package. (This package is however called once during the initialisation phase in order to get and store the inner and outer radii of each floor.) There are four electromagnetic floors and four hadronic floors (three fine and one coarse). All CC cells have a trapezoidal shape. They span azimuthal angles and pseudorapidity ranges of $2\pi/64$ (~ 0.1) and 0.1, respectively, except in the third floor where the granularity is four times

finer: $2\pi/128$ (~ 0.05) in azimuth and 0.05 in pseudorapidity. The cell size in the third floor is similar to the transverse extension of an electromagnetic shower. Each cell is labeled by its layer number (as given in Appendix A), and by its iη and iφ indices. In the construction of the road, the concepts of *η-rings* and *φ-slices* are used: η-rings are made of all cells with the same iη, and φ-slices of all cells with the same iφ. In the third floor, rings and slices are twice thinner than in the other floors.

For each track, the road is calculated floor by floor. In a given floor, the coordinates of the entrance and exit points of the track extrapolation are calculated, and the η-rings and φ-slices to which they belong are determined. The numbers of rings and slices crossed in the first three floors are shown in Fig. 4 for the single electron sample. It can be seen that, most of the time, only one η-ring is crossed in the first two floors. In the third floor, both because of the finer granularity and of the larger thickness of the floor, two η-rings are often crossed by the track extrapolation. In the vast majority of cases, on the other hand, only one φ-slice is crossed, even in the third floor. This is because, for the electron momentum range considered, the track extrapolations are almost parallel to the edges of the φ-slices in spite of the deflection of the trajectories in the magnetic field.

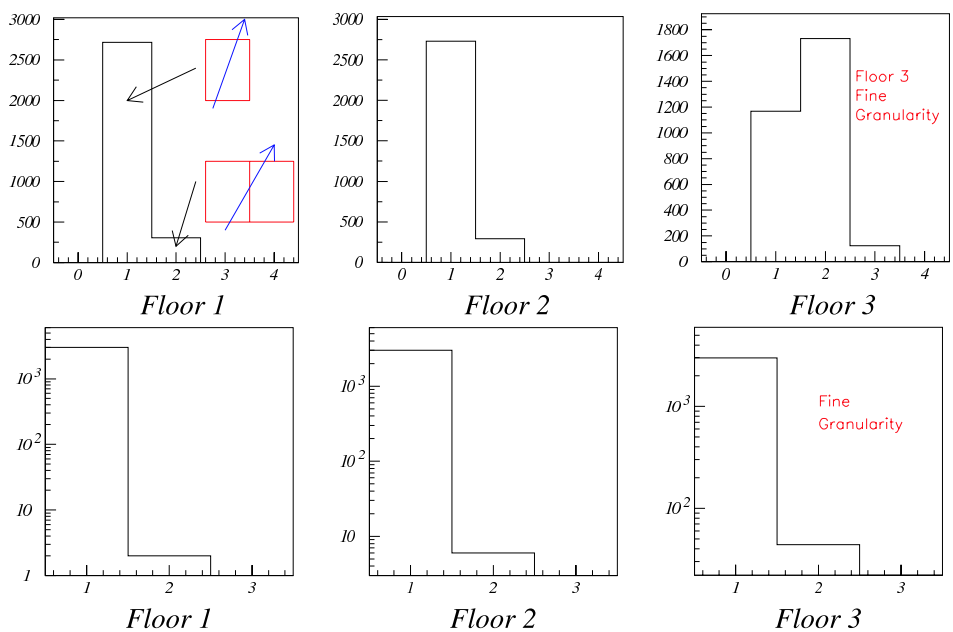


Figure 4: Numbers of η-rings (top) and of φ-slices (bottom, in logarithmic scale) crossed by the extrapolation of the track in the first three floors.

In first approximation, the road in a given floor consists of the cells located at the intersections of the crossed η-rings and φ-slices. There are however exceptions to this general rule.

Firstly, it is required that both the entrance and exit points belong to physically existing cells in the CC. If it is not the case, *i.e.,* if the track extrapolation leaves the CC in the floor considered, neither that floor nor any of the following ones are included in the central road; they are nevertheless taken into account at a later stage, together with the end calorimeters.

Secondly, since the aim of the road method is to discriminate electrons against hadrons, the road should satisfy the contradictory requirements of being as narrow as possible while at the same time containing as much of the electron energy as possible. Because of the non-vanishing transverse extension of electromagnetic showers, it may be worthwhile including neighbouring η-rings or φ-slices in addition to those crossed by the track extrapolation. On the other hand, it may be preferable not to include a crossed η-ring or φ-slice if it is expected to contain only a small fraction of the energy of an electron. These issues were investigated with the sample of single electrons.

The case where, in a given floor, only one η-ring is crossed by the track extrapolation, is considered first. The energy fraction in the ring crossed, calculated with respect to the total energy deposited in the floor considered, is shown in Fig. 5 as well as the corresponding fractions in the neighbouring rings. As expected, most of the energy is contained in the ring which is crossed, but in some cases the neighbouring rings do contain a significant fraction of the energy.

When two η-rings are crossed, most of the energy is, as expected, contained in those two rings, as can be seen in Fig. 6. The energy imbalance among the two rings reflects the longitudinal profile of the shower. In the first two floors, the shower is developing; the second ring crossed therefore contains more energy than the first one. In the third floor, on the contrary, the shower starts fading away, and the first ring thus contains more energy.

In the third floor it occasionnally happens, because of the finer granularity, that three η-rings are crossed. As expected, and as shown in Fig. 7, the fraction of deposited energy contained in the third ring rarely exceeds 20%. It turns out that better advantage of the third floor granularity is taken, and as a result the discrimination between electrons and hadrons is improved, if this third ring is not included in the road.

It has been seen that, in not unfrequent instances, neighbouring η-rings contain a substantial fraction of the deposited energy. A similar observation is made for φ-slices, but in that case the effect of the transverse extension of the shower can be more easily quantified since the track extrapolation runs almost parallel to the slice edges, and thus usually crosses only one slice. The fraction of the energy deposited in the floor which is contained in the slice neigbouring the crossed one is shown in Fig. 8 as a function of the distance of the track extrapolation to the edge of the slice; this distance, calculated at the entrance point into the floor, is normalized to the slice width.

The decision to include or not a neighbouring slice in the road is controlled by the *dismerge* parameter (Fig. 9): the neigbouring slice is included if the above defined distance is smaller than *dismerge* ($2 \times dismerge$ in the third floor because of the twice finer granularity). The maximum value of *dismerge* ($2 \times dismerge$ in the third floor) is 0.5, so that at most one slice is added (a twice thinner slice in the third floor). The same procedure is applied to the η-rings except that, in order to cope with the angle between the track extrapolation and the ring edges, both the entrance and the exit points must lie in the *dismerge* zone for the neighbouring cell to be added. No improvement of the performance has been observed when two different *dismerge* parameters are used for η-rings and φ-slices.

In the end the road is built in every floor from the intersections of up to two η-rings and up to two φ-slices, crossed by the track extrapolation or added as neighbours when

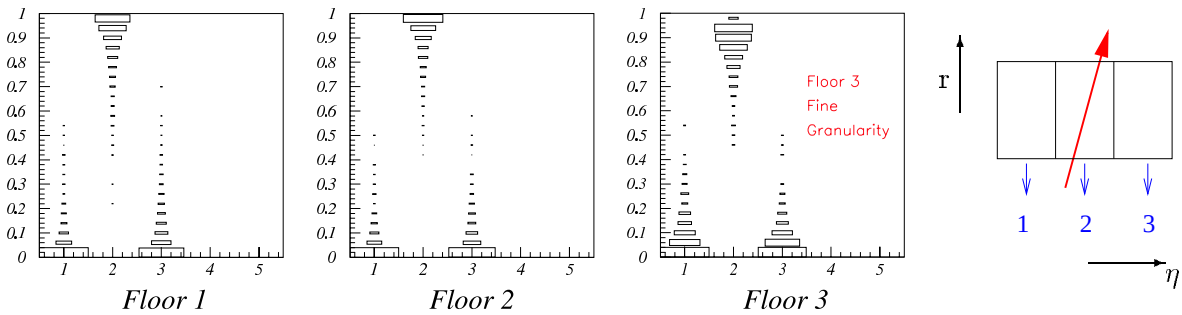


Figure 5: For the first three floors of the calorimeter, fractions of the energy deposited in the floor contained in the crossed ring and in the neighbouring ones, when only one ring is crossed by the track extrapolation. Ring number 1 is the first neighbour along the track, ring number 2 is the crossed ring, and ring number 3 is the last neighbour. In the electron sample used here, η is always positive.

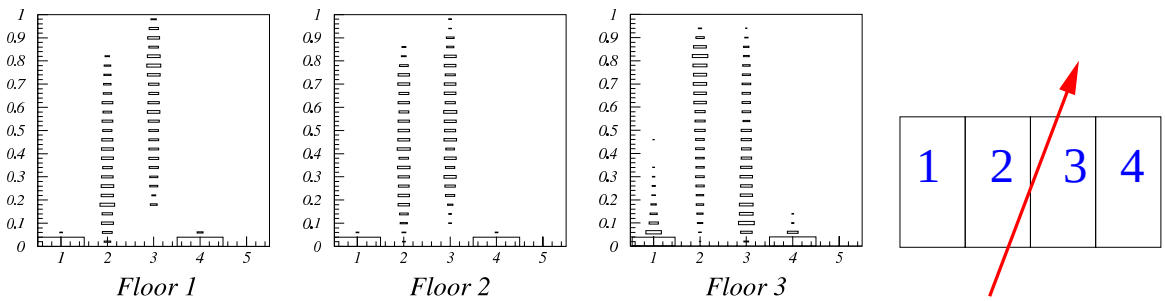


Figure 6: For the first three floors of the calorimeter, fractions of the energy deposited in the floor contained in the crossed rings and in the neighbouring ones, when two rings are crossed by the track extrapolation. Ring number 1 is the first neighbour along the track, rings number 2 and 3 are the crossed rings, and ring number 4 is the last neighbour. In the electron sample used here, η is always positive.

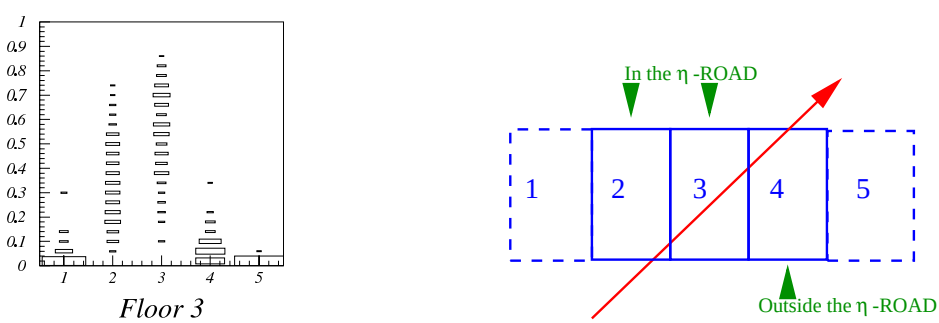


Figure 7: For the third floor of the calorimeter, fractions of the energy deposited in the floor contained in the crossed rings and in the neighbouring ones, when three rings are crossed by the track extrapolation. Ring number 1 is the first neighbour along the track, rings number 2, 3 and 4 are the crossed rings, and ring number 5 is the last neighbour. In the electron sample used here, η is always positive.

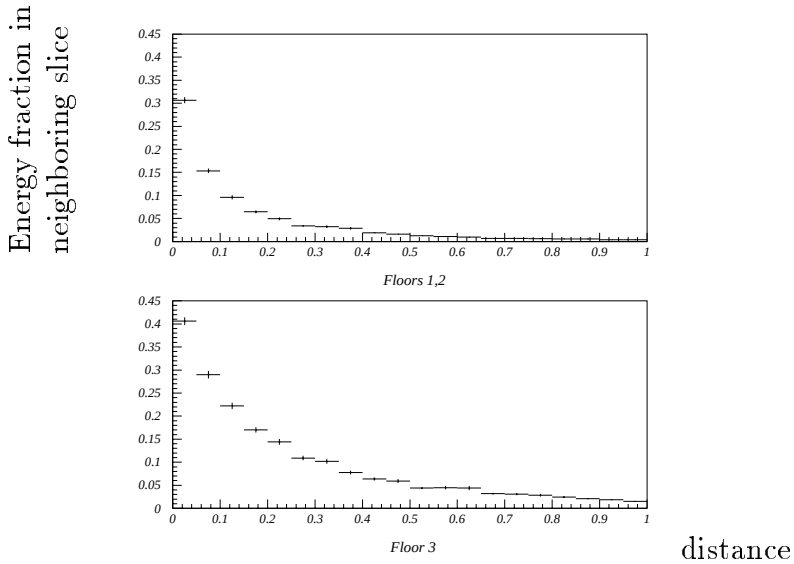


Figure 8: Fraction of the energy deposited in the floor contained in the neighbouring φ-slice, as a function of the distance to the slice edge normalized to the slice width.

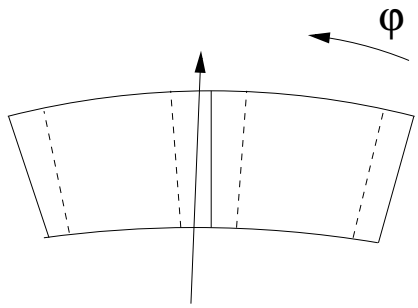


Figure 9: Illustration of the use of the *dismerge* parameter: the neighbouring φ-slice is added if the track extrapolation enters the floor between the dashed line and the slice edge.

the track extrapolation is contained in the *dismerge* zone. As a result, the road consists in each floor of one, two or four cells. An example is given in Fig. 10. The distributions of the number of cells within the road in floors 1 and 3 are shown in Fig. 11.

The ratio of the energy contained in the road to the total energy deposited in the first three floors is shown in Fig. 12 for the single electron sample, with a *dismerge* value of 0.15. This fraction is 89% on average, corresponding to 80% of the total energy deposited in the calorimeter. With respect to the generated electron energies, the resolution is 21%, to be compared to a raw calorimeter resolution of 16%.

The current implementation of the algorithm in the reconstruction program is such that three roads are calculated, for three *dismerge* values. The default *dismerge* values are presently 0.25, 0.125 and 0. The choice of which of those roads to use depends on the specific application, but the intermediate value turns out to be adequate for most purposes. Technical details are given in Appendix B

2.3.2 Road in the End Calorimeters

The geometry in the end calorimeters (EC) is much more intricate than in the CC, especially in the hadronic section which is split into an inner, a middle and an outer part. As in the CC, the electromagnetic section contains four floors, but the third floor granularity is twice thinner in both η and φ only for $|\eta| < 2.6$. In view of these complications, the `calorimeter_geometry` package is used instead of a fast recalculated geometry as for the CC road construction.

In contrast to the situation in the CC, the physical cell size in the EC decreases rapidly with increasing $|\eta|$, as can be seen in Fig. 13, and becomes smaller than the typical size of an electron shower. As a result, while the goal of the road construction in the EC remains to define an as small as possible tube along the track extrapolation which still contains a large fraction of an electron shower, the implementation needs to be somewhat different from that developed for the CC: defining the road as the crossed cells, with the possible addition of direct neighbours, is inadequate. The simplest alternative algorithm is to use a cylinder around the extrapolation of the track, with a fixed radius R similar to that of an electron shower.

The track extrapolations are performed by the `caltrack_extra` package [4] which provides a list of crossed cells as well as the coordinates of the entrance and exit points in each floor. This procedure is rather slow, but it is applied only once per event, and the information is shared with other algorithms, such as the jet finders.

In each floor, the cell crossed by the track extrapolation at the level of the central plane is included in the road, and all cells with center within a cylinder of radius R around the track extrapolation are added to the road. Since the cell size varies with η, the number of cells in the road also depends on it, as shown in Fig. 14 for $R = 3\,\mathrm{cm}$. The change in the granularity of the third floor at $|\eta| = 2.6$ is clearly visible.

The ratio of the energy contained in the road to the total energy deposited in the first three floors is shown in Fig. 15 for the single electron sample, for $R = 3\,\mathrm{cm}$. This fraction is 89% on average, corresponding to 85% of the total energy deposited in the calorimeter. With respect to the generated electron energies, the resolution is 21%, to be compared to a raw calorimeter resolution of 15%.

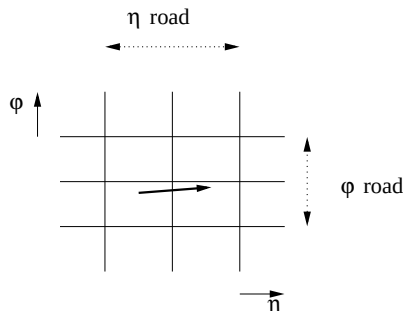


Figure 10: Example of cells included in the road in a given floor: the arrow represents the extrapolation of the track, from the entrance to the exit point; here, only one φ-slice is crossed, but the neighbouring slice has been added; as a result, the road contains four cells.

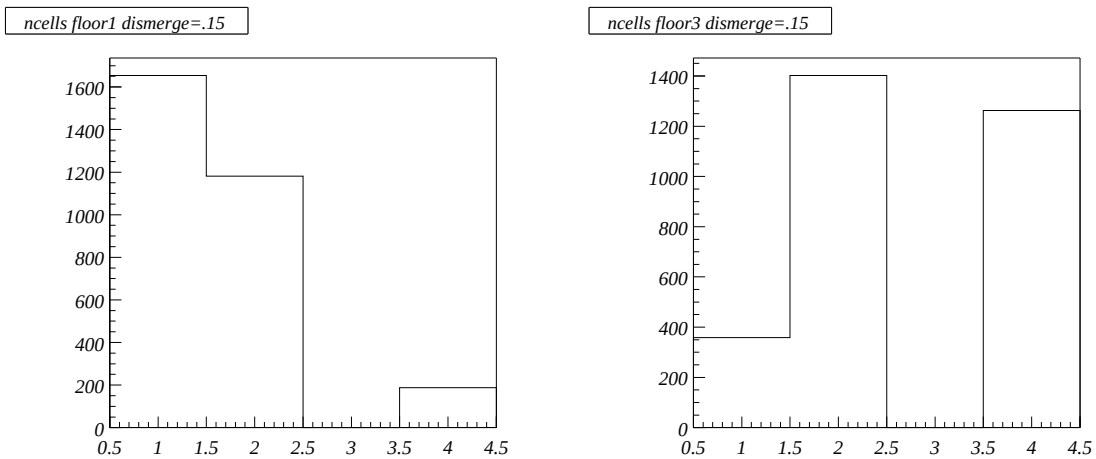


Figure 11: Number of celles included in the road in floors 1 (left) and 3 (right), for $dismerge = 0.15$.

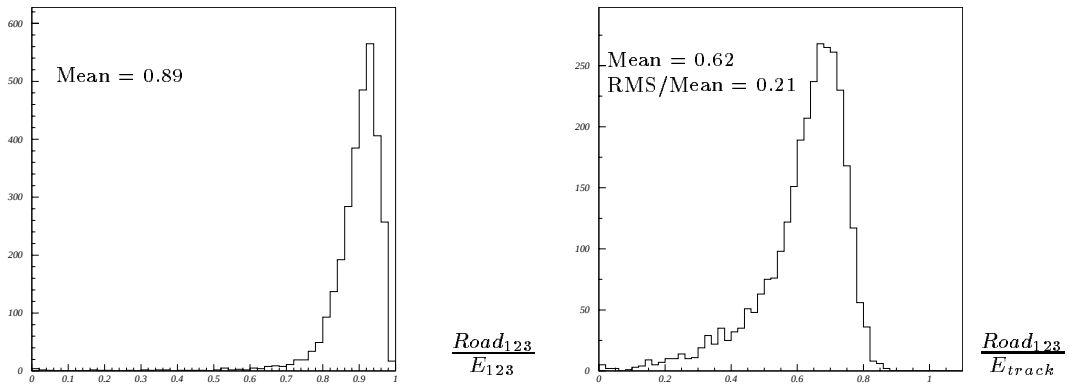


Figure 12: Fraction of the energy deposited in the first three floors (left) and of the generated electron energy (right) contained in the road.

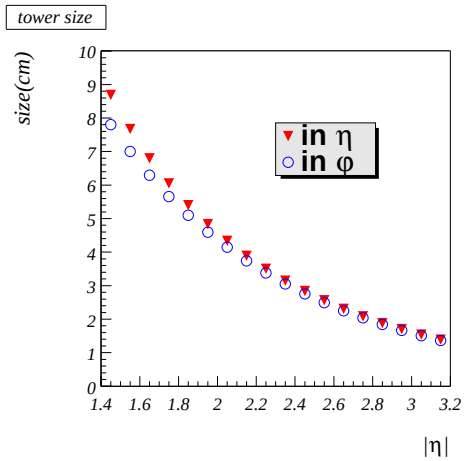


Figure 13: Cell size in the EC as a function of $|\eta|$, at the level of the second floor.

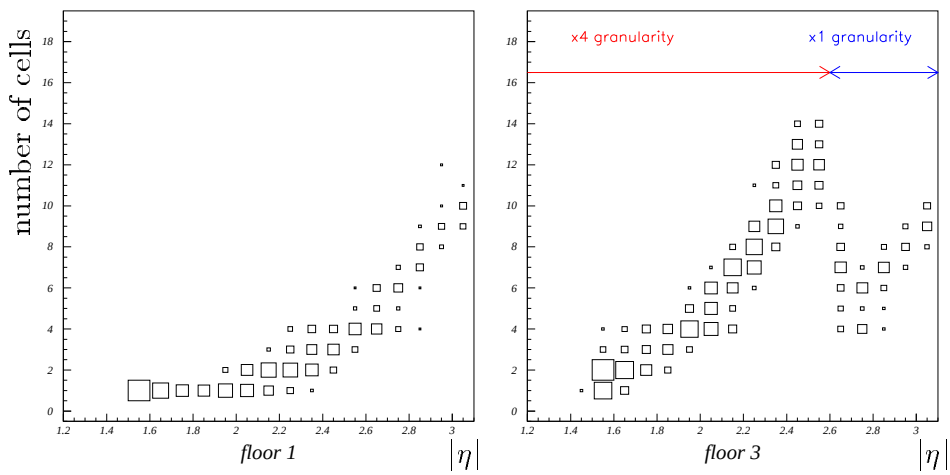


Figure 14: Number of cells included in the road in floors 1 (left) and 3 (right), for $R = 3\,\mathrm{cm}$.

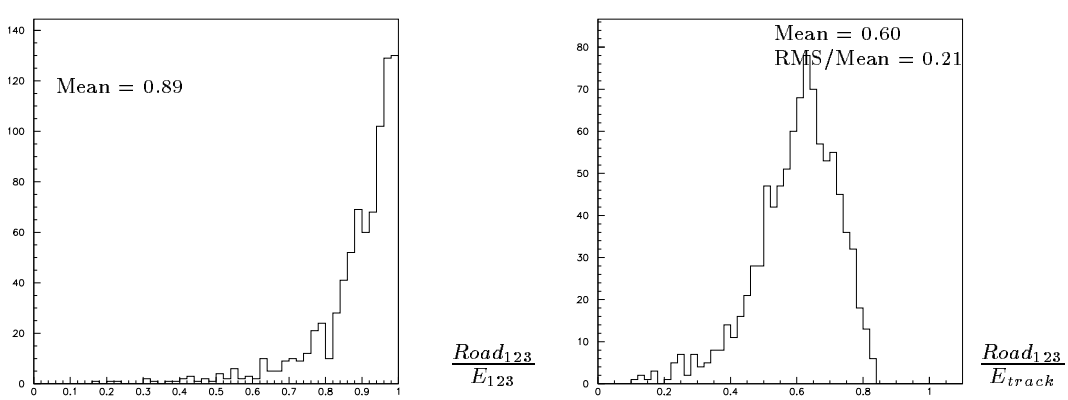


Figure 15: Fraction of the energy deposited in the first three floors (left) and of the generated electron energy (right) contained in the road.

In a similar fashion as for the CC, three roads are computed for different radii in the current implementation in the reconstruction program. The default radii are presently 4, 3 and 2.5 cm. The intermediate value is adequate for most purposes. Technical details are given in Appendix B.

To minimize geometrical acceptance gaps, the CC and EC road constructions are combined. If, for a given track, a road has been constructed in the CC, it is extended by the EC algorithm to the unused CC floors, if any, and to the EC. Since it uses the standard `calorimeter_geometry` package, the "EC" algorithm can also be applied to the whole CC.

2.4 Algorithm output and identification estimators

As explained above, three roads are constructed for the three values of the *dismerge* (in the CC) and radius R (in the EC) parameters. For a given road, the output consists of the set of transverse energies that it contains in the various calorimeter floors. (In fact, depending on the version of the reconstruction program, the energies in the hadronic sections may or may not be merged in a single quantity, as detailed in Appendix B.)

For each given road and given floor, the transverse energy is computed as the sum of the transverse energies of the cells belonging to the road. Only the cells with energy exceeding a given threshold are taken into account; presently, this threshold is set at 10 MeV to avoid including negative energy cells. The cell transverse energies are calculated as $E_T = E_{\text{cell}} \sin(\theta_{\text{track}})$. The θ angle deduced from the vertex and cell positions is not used in this calculation because this angle is erroneous in case of primary vertex misidentification.

In addition, two standard electron identification estimators are provided for each road:

- $EMF = \dfrac{\sum_{\text{floor number } i=1,2,3} E_T(i)}{\sum_{\text{all floors}} E_T(i)}$
- $E/p = \dfrac{\sum_{\text{floor number } i=1,2,3} E_T(i)}{p_T(\text{track})}$

where $E_T(i)$ is the transverse energy within the road in floor i. In contrast to the usual practice in D0, the above defined *electromagnetic fraction* EMF does not include the energy in the fourth floor.

3 Application to the tagging of heavy flavours

In this section, the performance of the road method for the tagging of high p_T b and c jets by electrons is studied in terms of efficiency and background rejection, along the lines of Ref. [2]. For that purpose, the following simulated samples have been used:

- 2000 $Z \to b\bar{b}$ with at least one $b \to eX$ *direct* decay,
- 2000 $Z \to b\bar{b}$ with at least one $b \to c \to eX$ *cascade* decay,
- 2000 $Z \to c\bar{c}$ with at least one $c \to eX$ decay, and
- 2000 QCD events with a $\hat{p}_T$ in the hard process in excess of 40 GeV/c.

All these events were generated with `Pythia`. The simulation was performed with `mcp05` in *mixture geometry*, and with an average pile-up of 2.5 events. The version of the reconstruction program was `p10.13.00`. The energy deposits in the road have been modified to emulate the *calweights* as implemented in `p10`. Finally, the `SEMReco` requirements on matching calorimeter towers were deactivated for the present study.

For the time being, only the above defined standard estimators have been considered in the electron identification procedure. Unless otherwise specified, the values chosen for the *dismerge* parameter and for the radius R are 0.125 and 3 cm, respectively.

3.1 Electron identification in $Z \to b\bar{b} \to eX$

3.1.1 Track selection

In the sample of 2095 electrons from direct b decays, only those with a generated p_T larger than 2 GeV/c are considered; the generated p_T spectrum is shown in Fig. 16. The generated electron tracks are extrapolated to the central plane of the third calorimeter floor, where $\eta_{\rm det}$ is defined; the distribution of $\eta_{\rm det}$ is shown in Fig. 16. In the central region ($|\eta_{\rm det}| < 1$) 788 direct electrons are selected, and 540 in the forward region ($1.4 < |\eta_{\rm det}| < 3.2$). In the central region, a fiducial cut is also performed in some cases in order to avoid the calorimeter cracks; only electrons such that $0.014 < \varphi_{\rm det}[\pi/16] < \pi/16 - 0.014$ are kept, where $\varphi_{\rm det}$ is calculated from the generated track extrapolation to the third floor. The $\varphi_{\rm det}$ cut retains 665 central electrons.

An angular matching between generated and reconstructed tracks is performed for the selected electrons. The normalized distance $\Delta r = ((\Delta\varphi/\delta_\varphi)^2 + (\Delta\eta/\delta_\eta)^2)^{1/2}$ is used for that purpose. As in Ref. [2], $\delta_\varphi = 0.0015$ and $\delta_\eta = 0.004$ are chosen, and the closest reconstructed track is associated to the generated electron provided that $\Delta r < 10$. The distribution of Δr is shown in Fig. 17.

The same p_T, $\eta_{\rm det}$ and, when required, $\varphi_{\rm det}$ cuts as for the generated electrons are applied to the reconstructed tracks. In the central region, 658 (552) such matching reconstructed tracks are found without (with) the $\varphi_{\rm det}$ cut. In the forward region, the number of matching reconstructed tracks is 329. The correlation of the reconstructed and generated p_T is diplayed in Fig. 18.

For the QCD background, the same p_T, $\eta_{\rm det}$ and $\varphi_{\rm det}$ criteria are applied to the reconstructed tracks, leading to 7865 (6705) tracks in the central region without (with) $\varphi_{\rm det}$

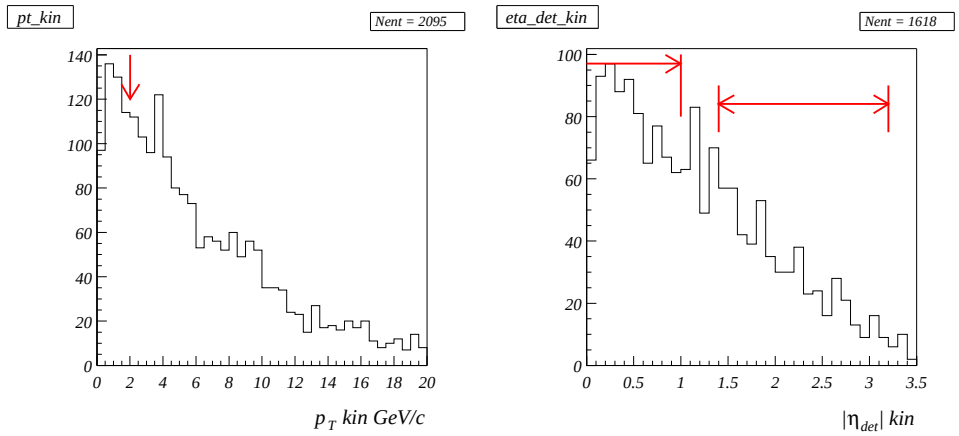


Figure 16: Distributions of the generated p_T for direct electrons (left), and of $|\eta_{\rm det}|$ for direct electrons with $p_T > 2\,{\rm GeV}/c$ (right).

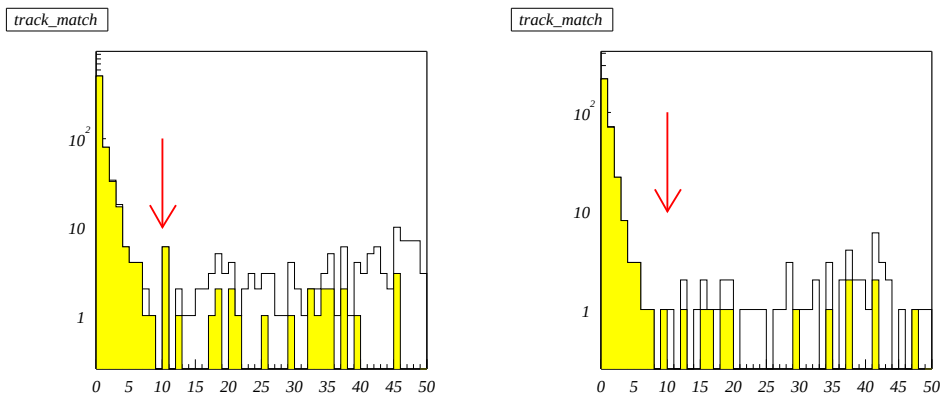


Figure 17: Normalized angular distance Δr between generated and reconstructed electron tracks in the central (left) and forward (right) regions. The minimum distance is shown shaded.

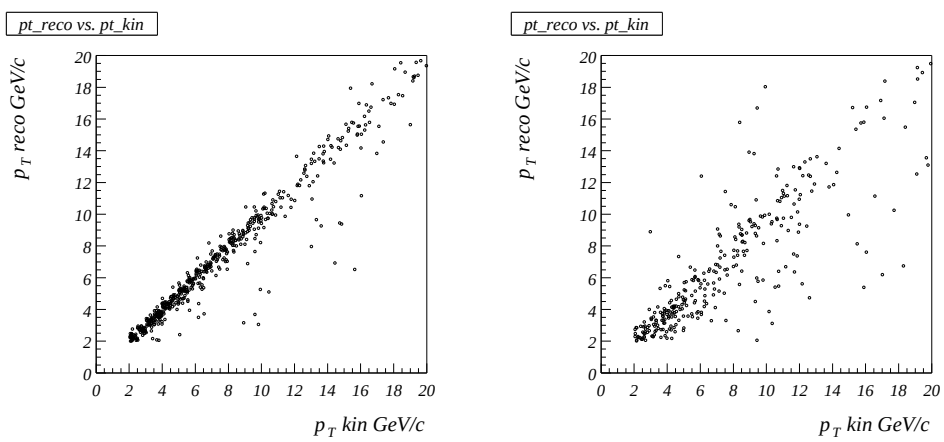


Figure 18: Reconstructed *vs* generated p_T for matched electrons in the central (left) and forward (right) regions.

cut, and to 4710 tracks in the forward region. The same matching criteria as above are used to recognize and reject electrons from b- or c-hadron decays. This leaves 7848 (6691) and 4705 tracks in the central and forward regions.

In the following, the $\varphi_{\rm det}$ cut is applied for the determination of the identification criteria based on the standard estimators. It is removed for the evaluation of the overall performance of these criteria and, unless explicitely specified otherwise, in all subsequent discussions.

3.1.2 Identification criteria

The EMF and E/p distributions in the central region are shown in Fig. 19 for both the direct electron signal and the QCD background. The corresponding distributions in the forward region are shown in Fig. 20. As expected, the electromagnetic fraction EMF clusters at values close to unity for signal electrons. Except for a small peak at $EMF = 1$ which is due to low energy hadrons, the background EMF distribution is flatter and extends to much lower values. The signal E/p distribution shows that 82% (77%) of the shower energy is contained in the first three floors of the road, with a relative resolution of 20% (35%) in the central (forward) region. The broader E/p distribution in the forward region is largely due to the degradation of the resolution on the track momentum. In contrast, the background E/p distribution is peaked at low values.

The two-dimensional distributions of E/p and EMF are shown in Fig. 21 for the signal and for the background in the central region. The corresponding distributions for the forward region are shown in Fig. 22. The boxes represent cuts optimized to retain as fake electrons 1% of the background tracks

- $(EMF > 0.81)$.AND.$(0.68 < E/p < 1.07)$ in the central region, and
- $(EMF > 0.88)$.AND.$(0.48 < E/p < 1.20)$ in the forward region.

These cuts, hereafter called *standard identification cuts*, select 390 and 196 electrons in the central and forward regions, respectively. The corresponding efficiencies are

- 71% in the central region, and
- 60% in the forward region.

The generated electron p_T spectrum before and after the standard identification cuts is shown in Fig. 23. The optimization procedure is described in Appendix C.

Since electrons originating from heavy-flavour decays have been removed from the QCD sample, the electron candidates selected in that sample, hereafter designated as *fake electrons*, are dominated by hadron misidentification, with only 5% associated with actual electrons.

These results were obtained for roads constructed with *dismerge* $= 0.125$ and $R = 3\,$cm. The performance for other choices, after appropriate reoptimization of the identification cuts, is displayed in Table 1, which confirms that the chosen values are close to optimal for the signal considered here.

As already stated, the EC algorithm can also be applied in the CC. The results obtained in that way, with cut reoptimization, are shown in Table 3.1.2. The performance

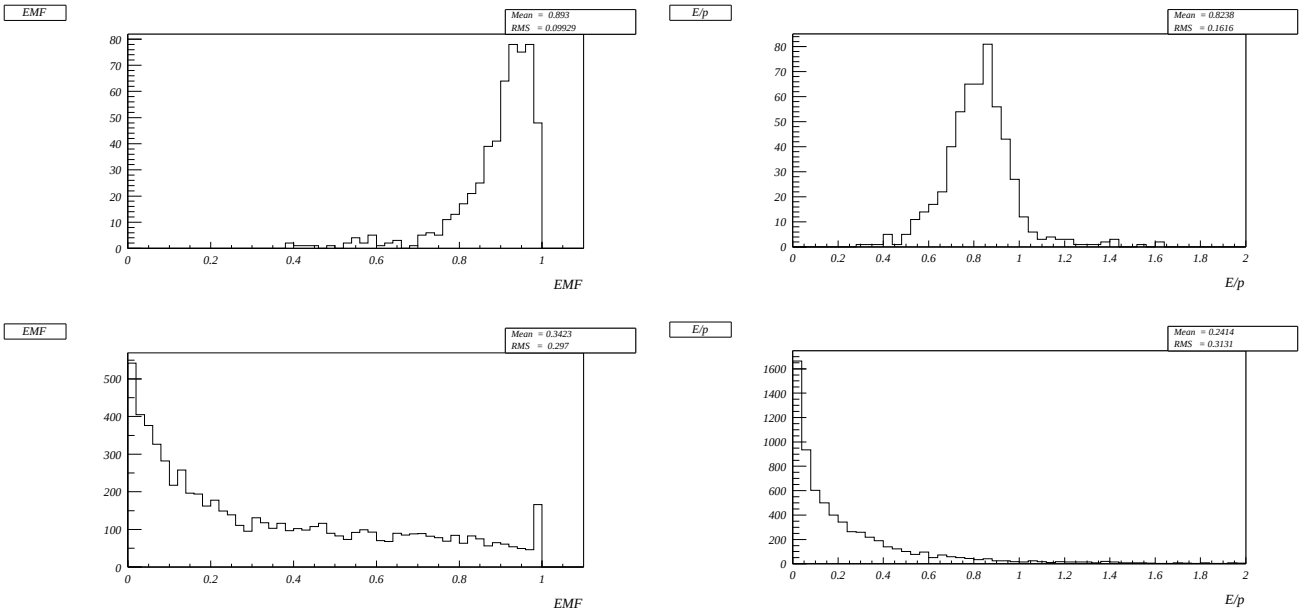


Figure 19: Distributions of the electromagnetic fraction EMF (left) and of E/p (right) for signal electrons (top) and for QCD events (bottom) in the central region.

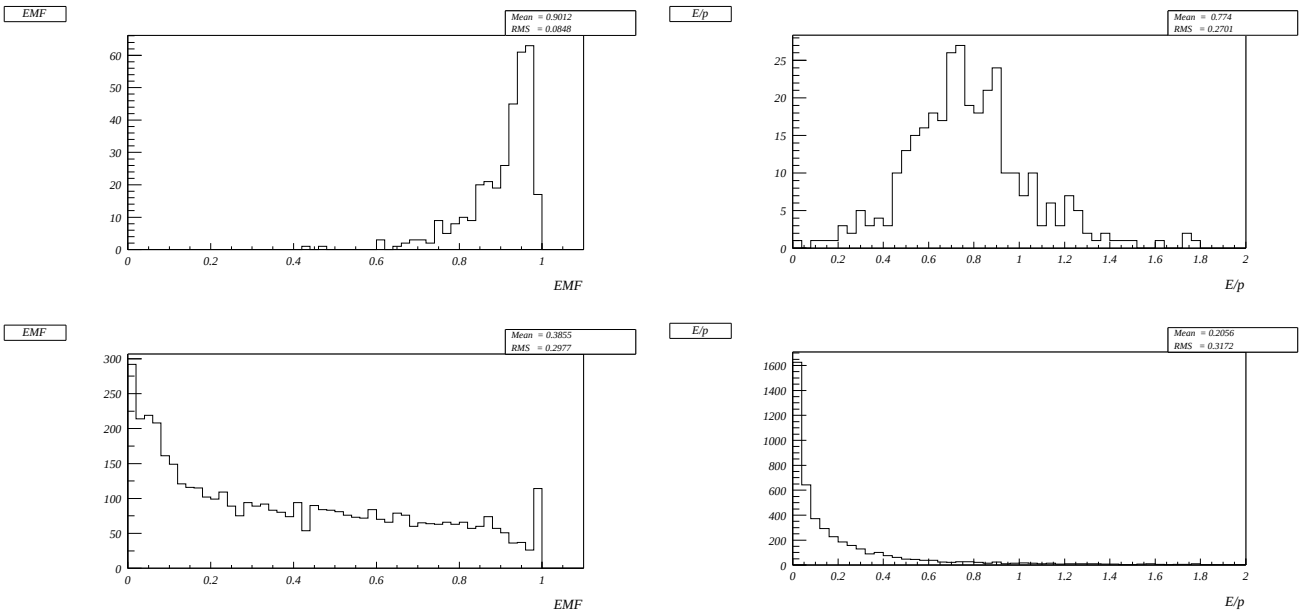


Figure 20: Distributions of the electromagnetic fraction EMF (left) and of E/p (right) for signal electrons (top) and for QCD events (bottom) in the forward region.

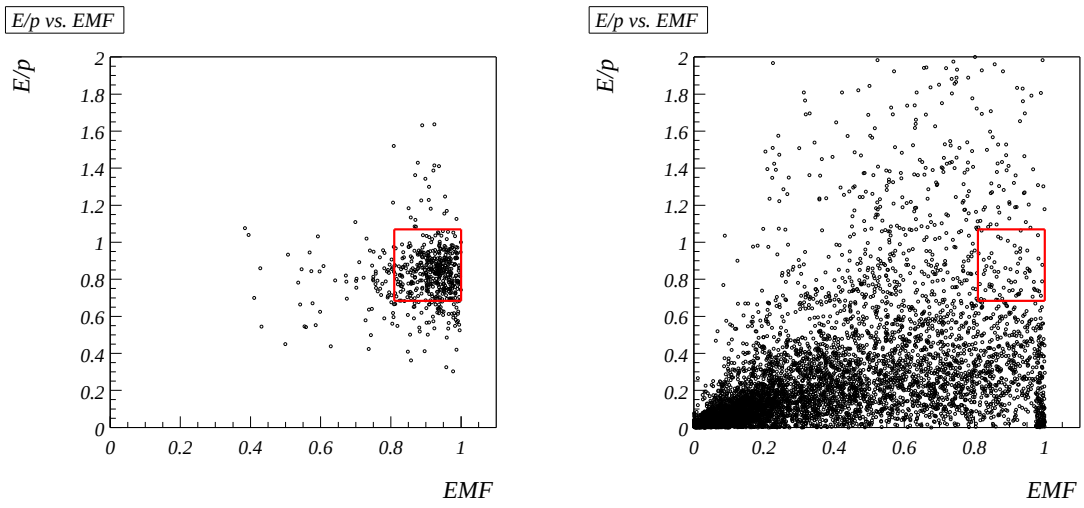


Figure 21: Electromagnetic fraction *vs* E/p for signal (left) and background (right) in the central region.

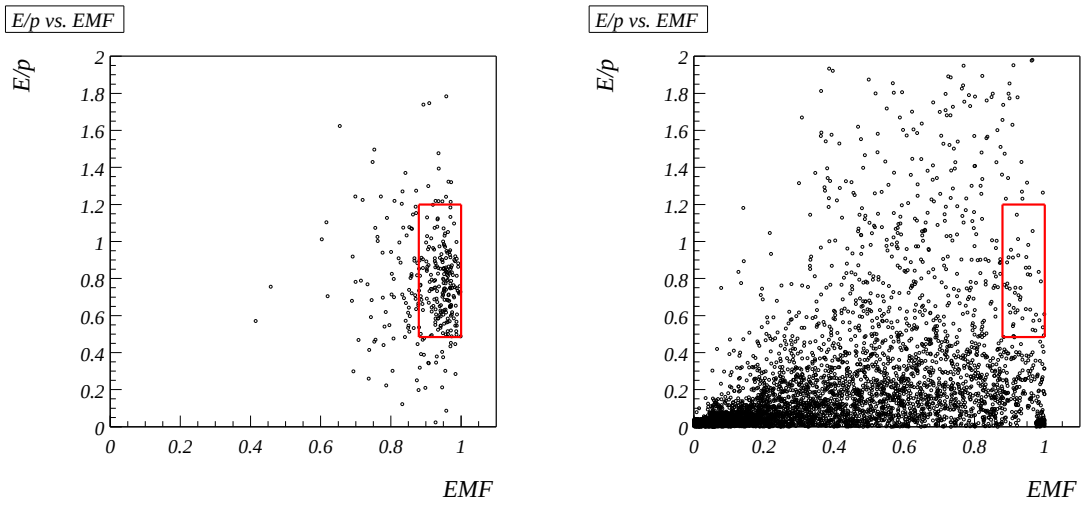


Figure 22: Electromagnetic fraction *vs* E/p for signal (left) and background (right) in the forward region.

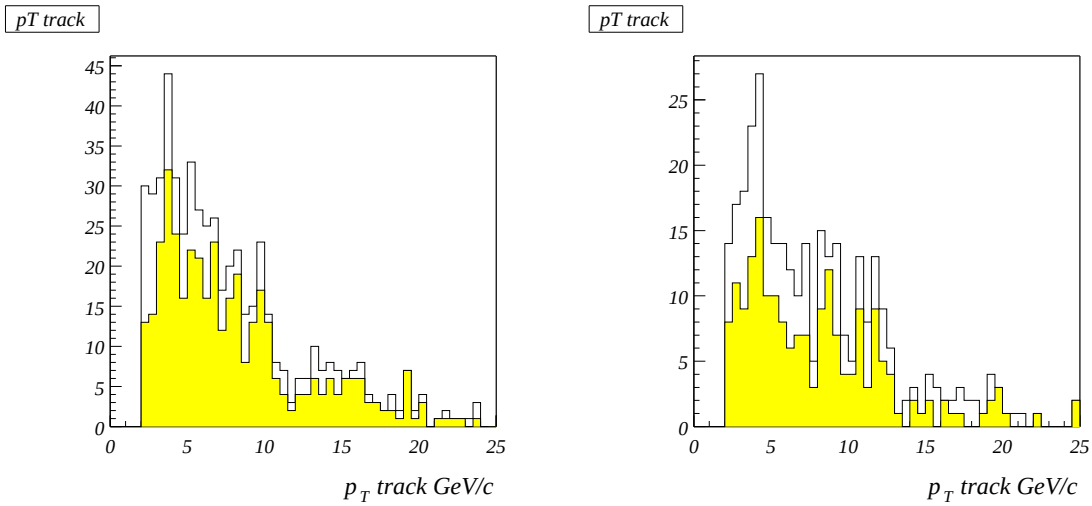


Figure 23: In the central (left) and forward (right) regions, generated p_T distributions of the electrons before and after (shaded) the standard identification cuts.

is reduced but, as could be expected, still reasonable. For $R = 2.5\,\mathrm{cm}$, it is similar to that obtained for $dismerge = 0$, which can be understood since in that case the EC road width is similar to the cell size in the third floor of the CC.

Table 1: Identification efficiencies for various *dismerge* parameter and road radius values.

	CC : *dismerge*			EC : radius(cm)		
parameter value	0	0.125	0.25	2.5	3	4
efficiency for 1% mis-id	65%	71%	69%	57%	60%	55%

Table 2: Identification efficiencies for various radii, when the EC algorithm is applied in the central calorimeter.

radius (cm) value	2.5	3.	4.
efficiency for 1% mis-id	65%	66%	62%

3.1.3 Overall performance of the identification criteria

Fiducial cuts were applied for the optimization of the standard electron identification criteria in order to disentangle the specific differences between electron and hadronic showers from local detector effects such as those encountered in the intercryostat region or near the calorimeter cracks. Keeping the standard identification cut values obtained above for a misidentification rate of 1%, the variation as a function of φ_{det} of the efficiency

for direct electrons and of the misidentification rate is shown in Fig. 24 for the central region, still for $p_T > 2\,\mathrm{GeV}/c$; the performance clearly degrades near the cracks, but it nevertheless remains appreciable.

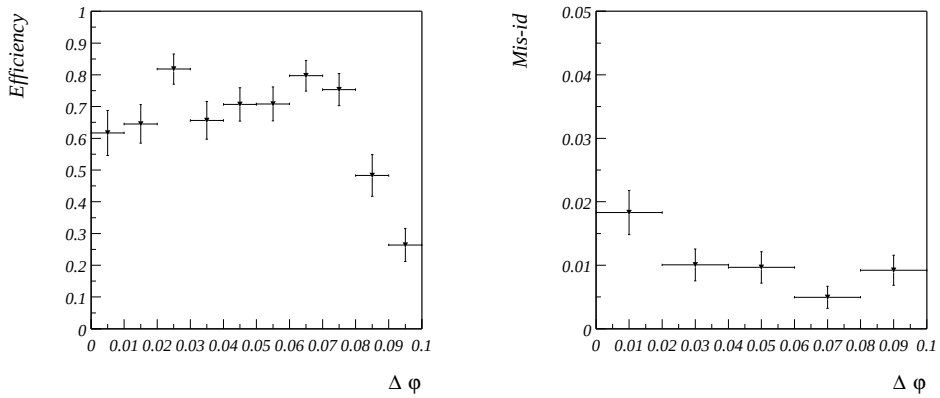


Figure 24: Identification efficiency (left) and misidentification rate (right) in the central region, as a function of $\Delta\varphi = |\varphi_{\mathrm{det}}[\pi/16] - \pi/32|$, for $p_T > 2\,\mathrm{GeV}/c$.

The performance of the standard identification cuts as a function of p_T is shown in Fig. 25 both in terms of efficiency and of misidentification rate. The generated p_T is used for the signal electrons, and the reconstructed p_T for the background. In the central region, the efficiency improves with increasing p_T; such a feature is not visible in the forward region because it is due to a threshold effect in energy, rather than in transverse energy. The misidentification-rate distributions are structureless.

The dependance on η_{det} of the efficiency and of the misidentification rate is shown in Fig. 26 for $p_T > 2\,\mathrm{GeV}/c$. As expected, the efficiency is reduced and the background rate is increased in the intercryostat region. The efficiency vanishes for $1.2 < |\eta_{\mathrm{det}}| < 1.4$, where the coverage by the electromagnetic calorimeter is interrupted.

The optimization of the standard identification cuts was performed for an arbitrarily fixed misidentification rate of 1%. When this rate is varied, the same optimization procedure leads to the efficiency *vs.* purity curve shown in Fig. 27.

3.1.4 Kinematic cuts

Among the electron candidates selected by the standard identification criteria, those which originate from direct b-hadron decays exhibit characteristic kinematic features: they tend to carry a large fraction of the energy of the primary b quark, and to be emitted with a large transverse momentum relative to the direction of their parent b hadron. To take advantage of these features, the electron candidate must first be associated to the hadron jet to which it belongs. The distribution of the angular distance $\Delta R_{(\mathrm{e-jet})}$ between the electron candidate and any jet in the event is shown in Fig. 28, as well as the distance $\Delta R_{(\mathrm{e-jet,min})}$ to the closest jet. The jet algorithm used here is *RunII-cone* with $R = 0.7$. The electron candidate and the closest jet are associated if $\Delta R_{(\mathrm{e-jet,min})} < 1$, which leaves 635 and 309 electrons in the central and forward regions, out of the 658 and 329 reconstructed ones.

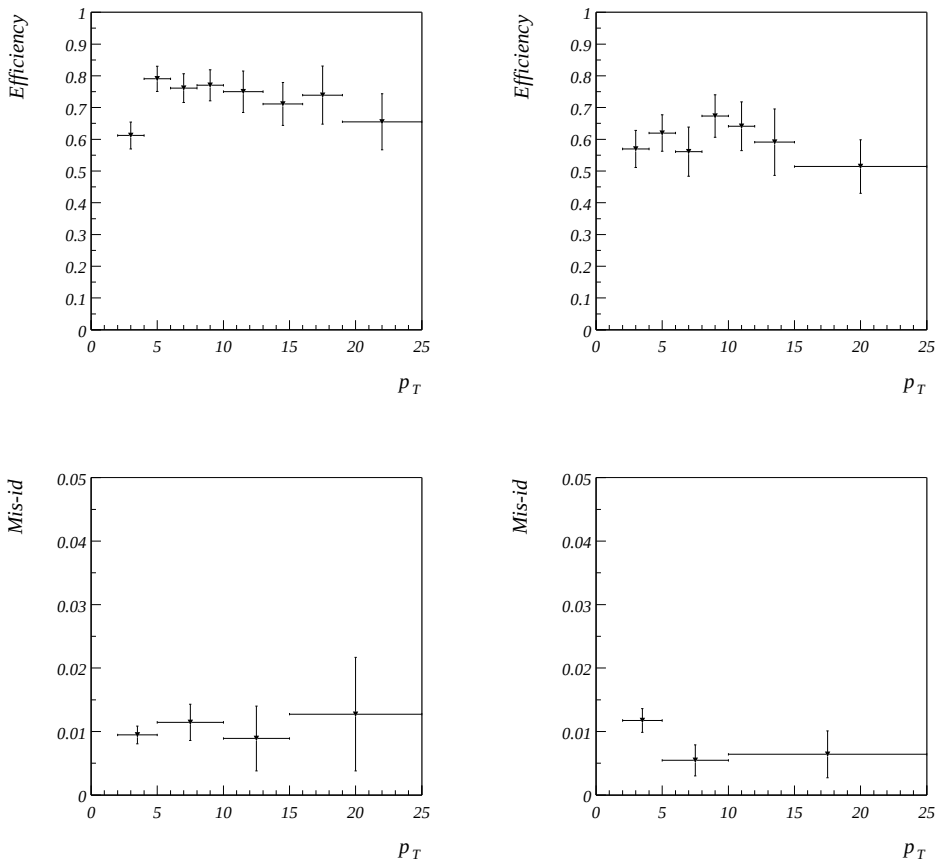


Figure 25: Identification efficiency (top) and misidentification rate (bottom) in the central (left) and forward (right) regions, as a function of p_T. The overflows are included in the last bins.

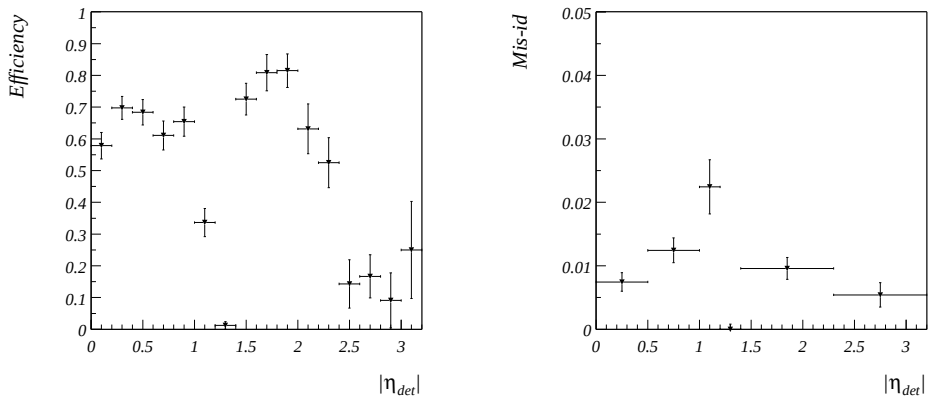


Figure 26: Identification efficiency (left) and misidentification rate (right), as a function of $|\eta_{\rm det}|$, for $p_T > 2\,{\rm GeV}/c$.

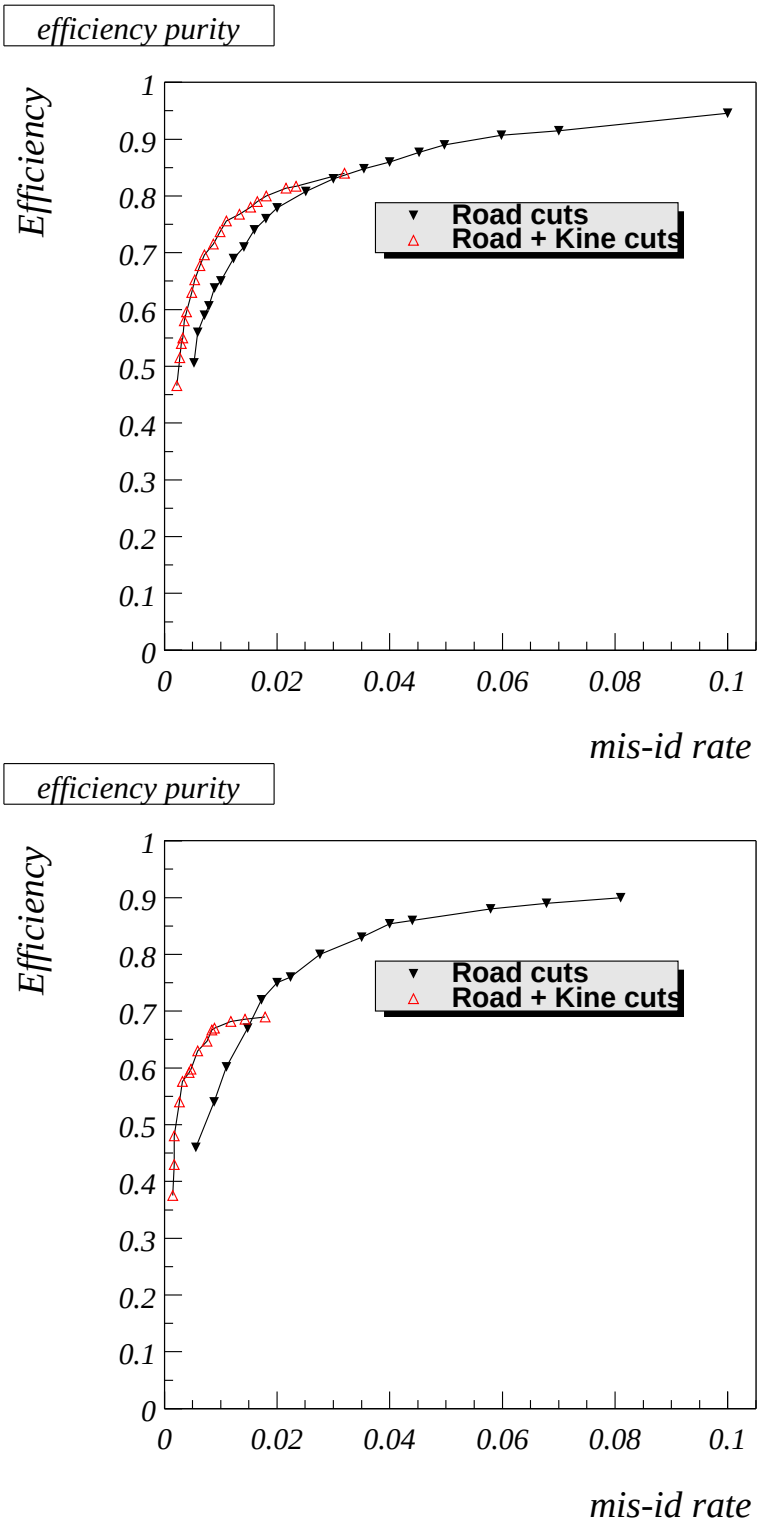


Figure 27: Efficiency *vs* purity in the central (top) and forward (bottom) regions before and after kinematic cuts.

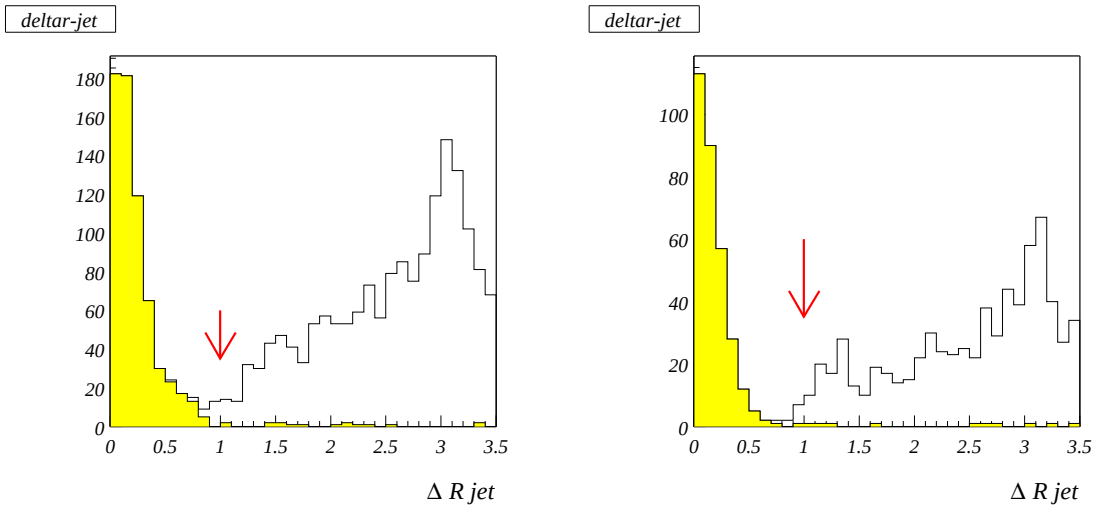


Figure 28: Angular distance $\Delta R_{(\mathrm{e-jet})}$ between the electron candidate and all jets in the event, and distance $\Delta R_{(\mathrm{e-jet,min})}$ to the closest jet (shaded), in the central (left) and forward (right) regions.

The fraction of the jet energy carried by the electron is estimated as $x_E = p_T/E_T$, where p_T is the track transverse momentum and E_T is the jet transverse energy, and the transverse momentum of the electron relative to the jet axis as $p_t = p_T \times \Delta R_{(\mathrm{e-jet,min})}$. The two-dimensional distribution of x_E and p_t is shown in Fig. 29 for the 413 electrons selected by the standard identification criteria in the central region. Of those, the cuts

$$(x_E > 0.35).\mathrm{OR}.(p_t > 0.75\,\mathrm{GeV}/c)$$

select 378, while 30 fake electrons are retained out of 78; the signal efficiency is 60%, and the misidentification rate is reduced to 0.4%. In the forward region (Fig. 30), the same cuts select 148 electrons out of 185, and keep 9 fake electrons out of 46; the signal efficiency is 48%, and the misidentification rate 0.2%. The generated electron p_T spectrum before and after kinematic cuts is shown in Fig. 31.

A summary of the performance of the selection with standard identification and kinematic cuts is given in Table 3 for the direct electron sample and in Table 4 for the QCD background.

The effect of the kinematic cuts on the efficiency *vs.* purity curve is shown in Fig. 27. It can be seen that the kinematic cuts are beneficial in the high purity regime. For misidentification rates larger than a few percent, on the contrary, it is preferable to apply only the identification cuts.

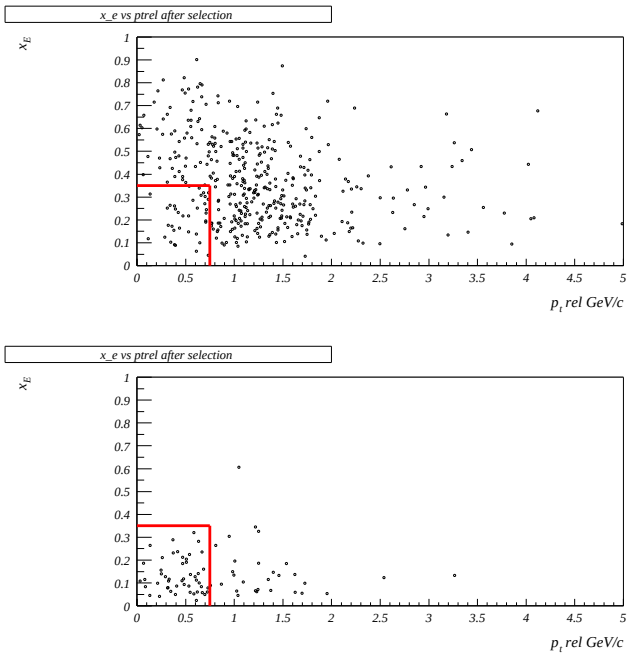


Figure 29: Distribution of x_E *vs* p_t for selected direct electrons (top) and for the fake electron candidates in the QCD sample (bottom), in the central region.

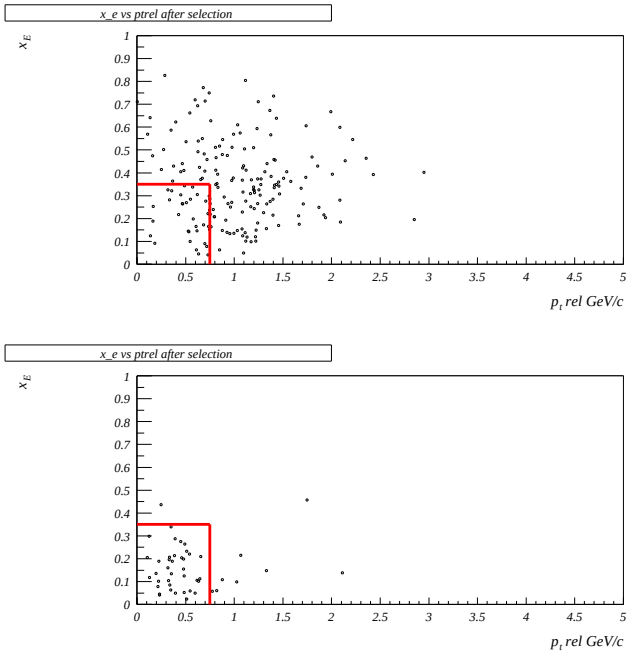


Figure 30: Distribution of x_E *vs* p_t for selected direct electrons (top) and for the fake electron candidates in the QCD sample (bottom), in the forward region.

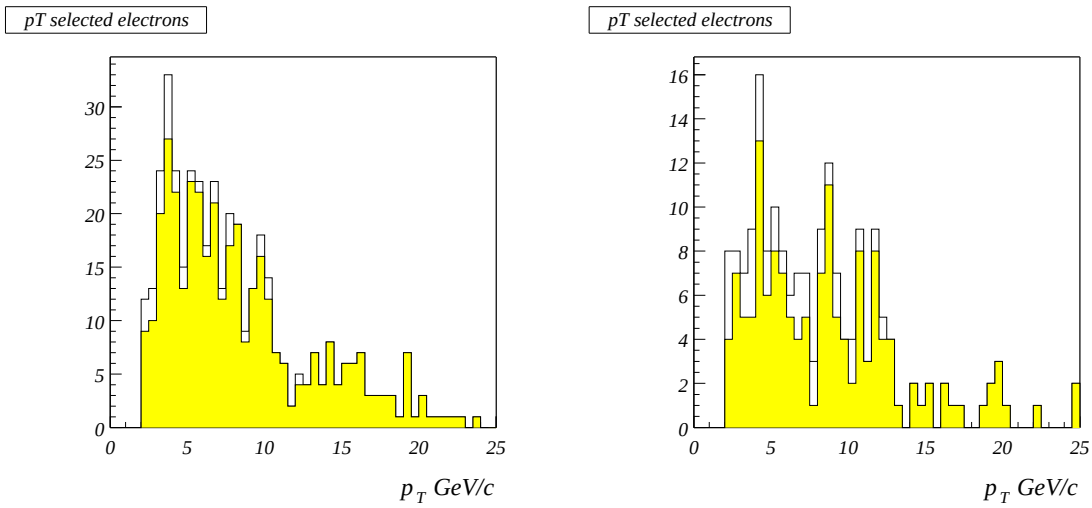


Figure 31: In the central (left) and forward (right) regions, p_T distributions of the electrons before and after (shaded) kinematic cuts.

Table 3: Summary of the selection of direct electrons in the $Z \to b\bar{b} \to eX$ sample. The identification efficiencies without and with kinematic cuts are calculated with respect to the numbers of electron tracks associated with a jet.

Number of electrons	CC	EC
generated	2095	
in η acceptance with $p_T > 2\,\mathrm{GeV}/c$	788	540
with matched reconstructed track	658	329
with an associated jet	635	309
identified	413 (65%)	185 (60%)
after kinematic cuts	378 (60%)	148 (48%)

Table 4: Summary of the selection of fake electron candidates in the QCD sample. Electrons from semileptonic heavy-flavour decays have been removed. The misidentification rates without and with kinematic cuts are calculated with respect to the numbers of tracks associated with a jet.

Number of tracks	CC	EC
in η acceptance with $p_T > 2\,\mathrm{GeV}/c$	7848	4705
with an associated jet	7571	4058
identified	78 (1.0%)	46 (1.1%)
after kinematic cuts	30 (0.4%)	9 (0.2%)

3.2 b-quark tagging with electrons

Out of 2095 direct electrons from b-hadron decays in $Z \to b\bar{b}$ events, a total of 598 are selected by the standard identification criteria, of which 526 are retained by the proposed kinematic cuts. Given the value of 10.6% for the $b \to e$ branching ratio [5], the efficiency for the tagging of a b quark by a direct electron is therefore 3.0% (2.7%) without (with) kinematic cuts. To these efficiencies should be added the contribution from electrons produced in cascade b-hadron decays.

3.2.1 Contribution of cascade decays

The generated cascade electron p_T spectrum and their $\eta_{\rm det}$ distribution after the $p_T > 2\,{\rm GeV}/c$ cut are shown in Fig. 32. As expected, the p_T spectrum of cascade electrons is softer than that of direct electrons.

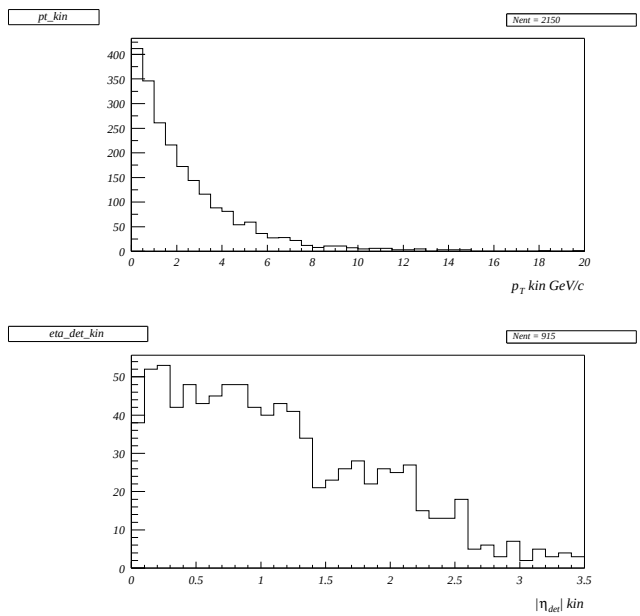


Figure 32: Distributions of the generated p_T for cascade electrons (top), and of $|\eta_{\rm det}|$ for cascade electrons with $p_T > 2\,{\rm GeV}/c$ (bottom).

Electrons from cascade decays are selected in the same manner as direct electrons, with results shown in Table 5. In particular, the effect of the identification cuts is visible in Fig. 33. Electron identification is less efficient than for direct decays because cascade electrons are emitted closer to the core of the jet, *i.e.,* with a smaller transverse momentum with respect to the jet axis. The effect is even more dramatic when the kinematic cuts are applied, as can be seen in Fig. 34.

The totals of 212 and 81 electrons selected without and with kinematic cuts out of a sample of 2150 cascade electrons correspond to efficiencies of 0.8% and 0.3% for the tagging of a b quark by a cascade electron, given the 8.0% value of the $b \to c \to e$ branching ratio [5].

Table 5: Summary of the selection of cascade electrons in the $Z \to b\bar{b} \to c \to eX$ sample.The identification efficiencies without and with kinematic cuts are calculated with respect to the numbers of electron tracks associated with a jet.

Number of electrons	CC	EC
generated	2150	
in η acceptance with $p_T > 2\,\mathrm{GeV}/c$	459	285
with matched reconstructed track	333	140
with an associated jet	324	136
identified	141 (44%)	71 (52%)
remain after kinematic cuts	60 (19%)	21 (15%)

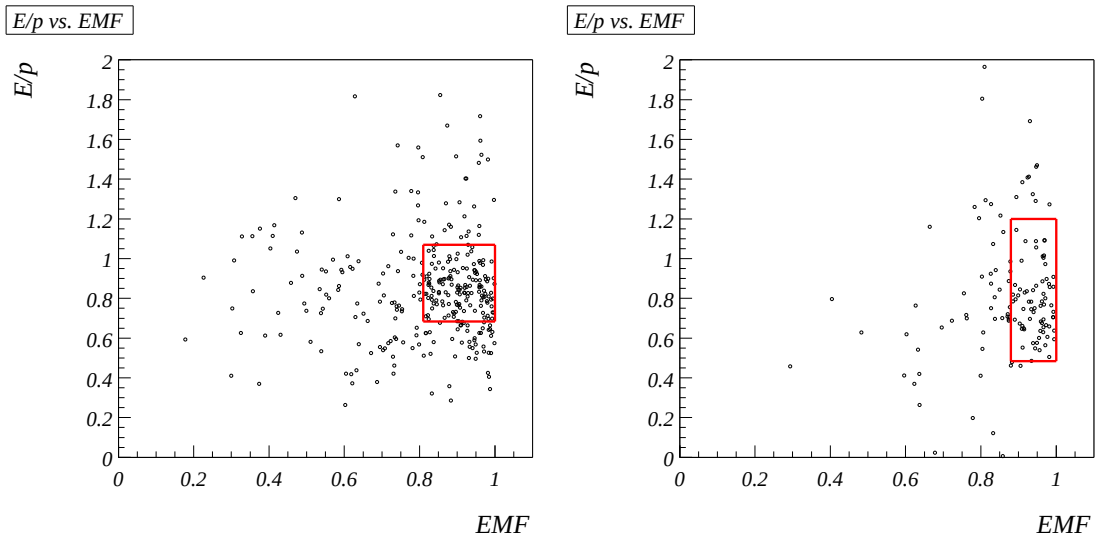


Figure 33: Electromagnetic fraction EMF *vs* E/p for cascade electrons in the central (left) and forward (right) regions.

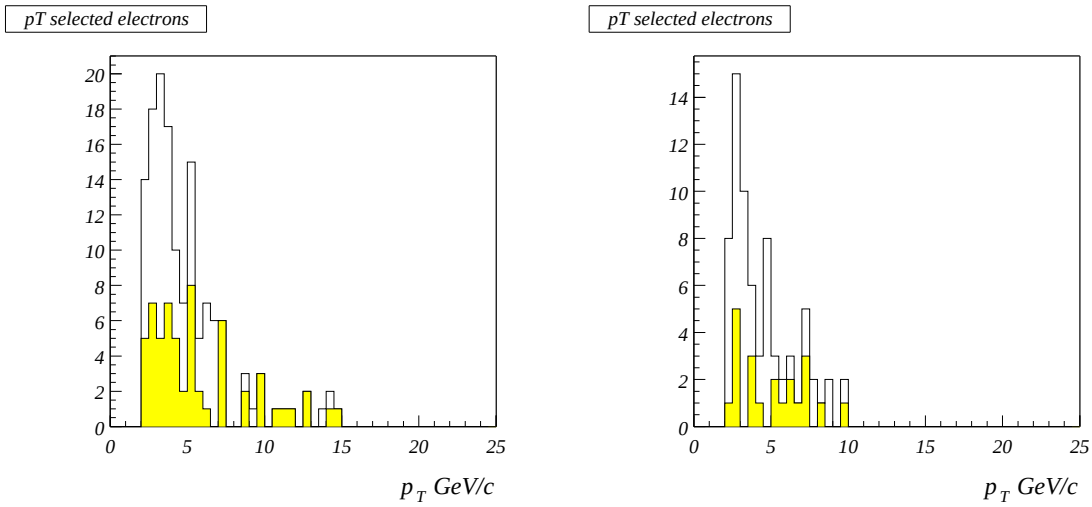


Figure 34: In the central (left) and forward (right) regions, generated p_T distributions of cascade electrons before and after (shaded) kinematic cuts.

3.2.2 Overall b-tagging performance

With the standard identification criteria, 124 (39) fake electrons are selected without (with) kinematic cuts in the sample of 2000 QCD events generated with $\hat{p}_T > 40\,\mathrm{GeV}/c$, where *fake* electron is to be understood as "electron candidate not associated with an electron from a semileptonic heavy-flavour decay". Under the simplifying assumption that two high p_T jets arise from the hard process in every event, fake b-tagging rates of 3.1% and 1.0% result, without and with kinematic cuts, respectively. If a mistag rate at the percent level is required, the kinematic cuts should therefore be applied. Under those conditions, the following probabilities are obtained for high p_T jets:

- 25.1% to tag a b $\rightarrow$ e decay,
- 2.7% to tag a generic b quark by a direct electron,
- 3.8% to tag a b $\rightarrow$ c $\rightarrow$ e decay,
- 0.3% to tag a generic b quark by a cascade electron,
- 1.0% to tag a generic quark by a fake electron, and
- 4.0% to tag a generic b quark by a direct, cascade, or fake electron.

The signal to background ratio for b jets is increased by a factor of 4.0.

3.3 Electron identification in $Z \to c\bar{c} \to eX$

The generated electron p_T spectrum and their $\eta_{\rm det}$ distribution after the $p_T > 2\,{\rm GeV}/c$ cut are shown in Fig. 35. The p_T spectrum is softer than the one of direct electrons from $Z \to b\bar{b}$. The final performance is therefore expected to be somewhat degraded. This is indeed observed when the standard identification criteria are applied, as can be seen in Table. 6.

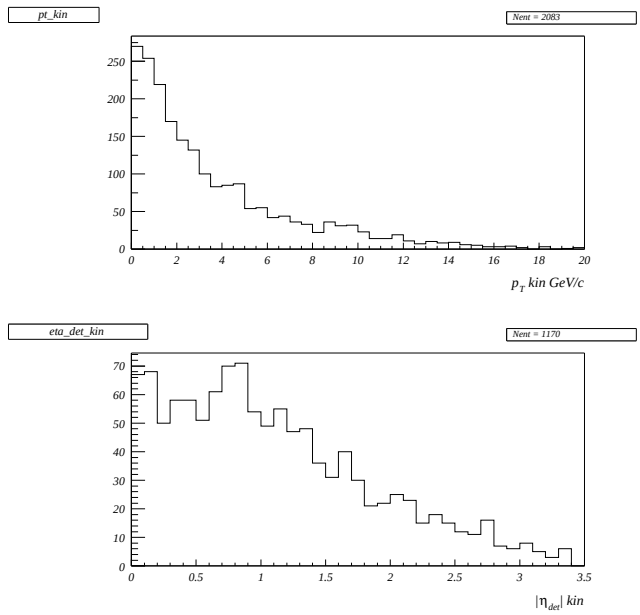


Figure 35: For $Z \to c\bar{c} \to eX$ events, distributions of the generated electron p_T (top), and of $|\eta_{\rm det}|$ for electrons with $p_T > 2\,{\rm GeV}/c$ (bottom).

Table 6: Summary of the selection of electrons in the $Z \to c\bar{c} \to eX$ sample.The identification efficiencies are calculated with respect to the numbers of electron tracks associated with a jet.

Number of electrons	CC	EC
generated	2083	
in η acceptance with $p_T > 2\,{\rm GeV}/c$	608	341
with matched reconstructed track	461	202
with an associated jet	453	198
identified with standard cuts	192 (42%)	71 (36%)
identified with optimized cuts (standard road)	200 (44%)	86 (43%)
identified with optimized cuts (narrow road)	208 (46%)	91 (46%)

The degradation in the identification performance can be understood by inspection of the distributions of EMF and E/p shown in Fig. 36. It can be seen, in particular, that the EMF distribution has a longer tail than in the case of direct $b \to e$ decays from $Z \to b\bar{b}$ events. This is due to an increased overlap of electron and hadronic showers in

the narrower c jets. A reoptimization of the identification cuts for c $\rightarrow$ e decays from Z $\rightarrow$ c$\bar{\text{c}}$ events, still for a 1% misidentification rate in the QCD sample, leads to slightly improved efficiencies, displayed in Table 6, particularly in the forward region. For the same reason, it is also to be expected that some improvement may result from the use of a narrower road. This can indeed be observed in Table 6, where the results for a road constructed with $dismerge = 0$ and a radius R of 2.5 cm are shown.

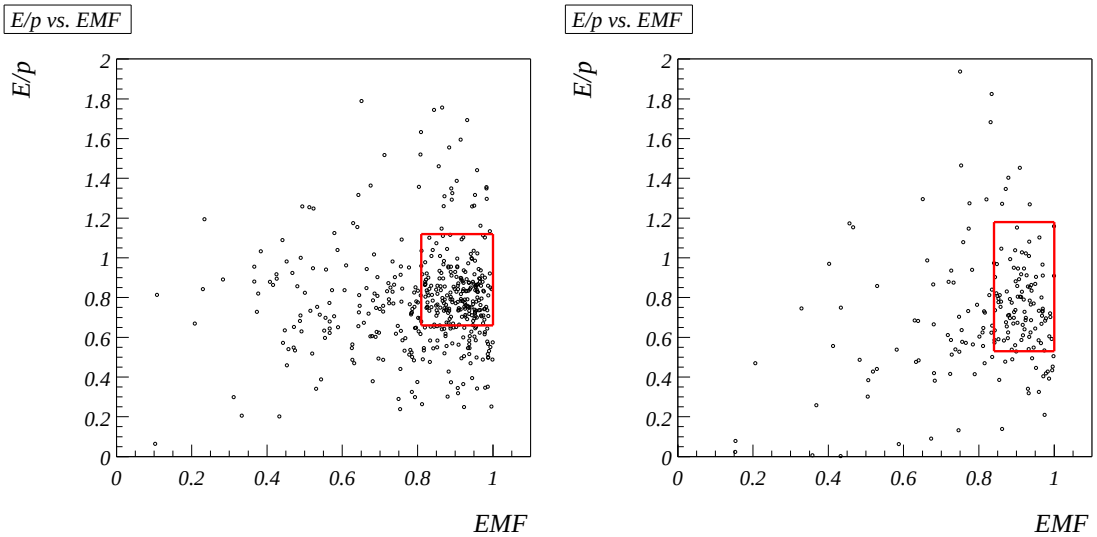


Figure 36: Electromagnetic fraction EMF *vs.* E/p in the central (left) and forward (right) regions.

The kinematic cuts should also be modified to account for the differences between c and b decays. For

$$(x_E > 0.125).\text{OR}.(p_t > 1\,\text{GeV}/c),$$

as shown in Figs. 37 and 38, the results displayed in Table. 7 are obtained. The effect of the kinematic cuts on the p_T distribution of identified electrons is shown in Fig. 39.

Table 7: Summary of the selection performance for electrons from Z $\rightarrow$ c$\bar{\text{c}}$. In the QCD sample, electrons from semileptonic heavy-flavour decays have been removed. The identification efficiencies (misidentification rates) without and with kinematic cuts are calculated with respect to the numbers of electron tracks (of tracks) associated with a jet.

	Number of identified electrons	Number of fake electrons
In the CC	208 (46%)	76 (1.0%)
After kinematic cuts	184 (41%)	41 (0.5%)
In the EC	91 (46%)	46 (1.1%)
After kinematic cuts	77 (39%)	22 (0.5%)

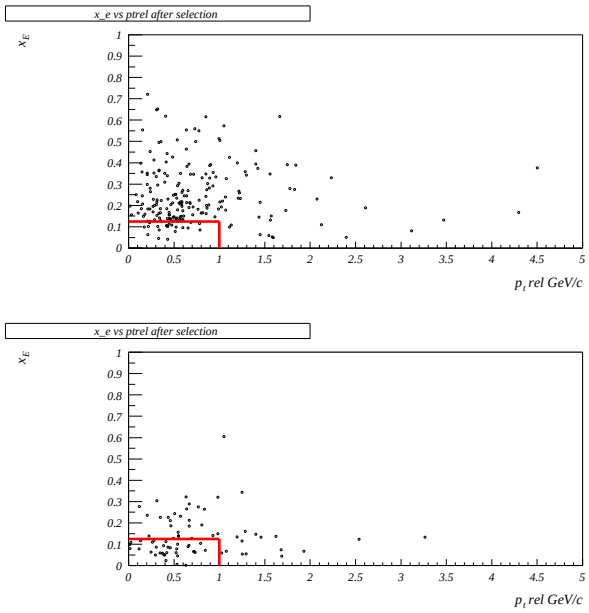


Figure 37: Distribution of x_E *vs* p_t for selected electrons from c $\rightarrow$ e decays (top) and for the fake electron candidates in the QCD sample (bottom), in the central region.

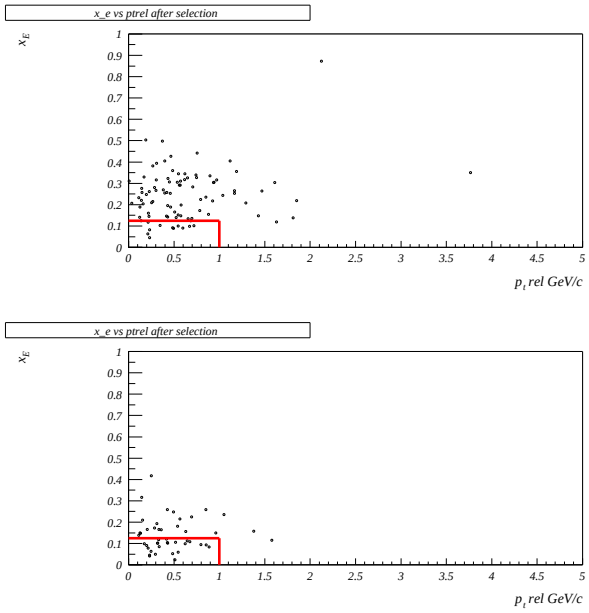


Figure 38: Distribution of x_E *vs* p_t for selected electrons from c $\rightarrow$ e decays (top) and for the fake electron candidates in the QCD sample (bottom), in the forward region.

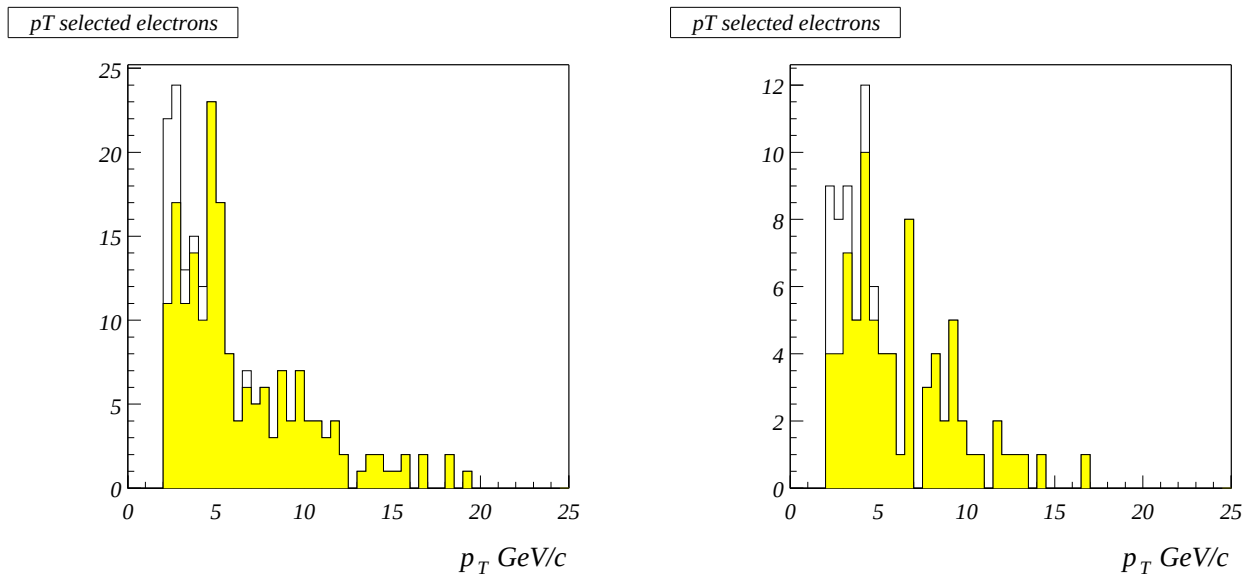


Figure 39: For identified electrons from $Z \to c\bar{c}$, generated p_T distributions before and after (shaded) kinematic cuts, in the central (left) and forward (right) regions.

In the end, 261 electrons are selected out of a sample of 2083 electrons from c-hadron decays in $Z \to c\bar{c}$ events. The signal selection efficiency is 12.5%, with a misidentification rate of 0.5%. The mistag probability is 1.6% for a generic quark. Taking into account the value of 9.8% for the $c \to e$ branching ratio [5], the probability to tag a generic c quark by a real or fake electron is 2.8%. The signal to background ratio for c jets is moderately increased by a factor of 1.8.

4 Conclusion

An algorithm for the identification of electrons in jets, the *road method*, has been developed and incorporated in the D0 reconstruction program. For each reconstructed charged particle track, the heart of the algorithm consists in extrapolating it into the calorimeter, in establishing the list of cells within some restricted volume, or *road*, around the extrapolation, and in constructing identification estimators based on the track momentum and on the energy contents in the various calorimeter floors within the road. For the time being, the preshowers, the intercryostat detectors and the massless gaps are not utilized. Due to the different geometrical structures of the central and end calorimeters, two slightly different algorithms are used, and their results are combined for track extrapolations intersecting both regions.

On average, 90% of the energy deposited by electrons in the central calorimeter is contained in the first three floors, and 90% of that energy is contained within a road which is typically built out of one or two cells in the first two floors, and of two or four cells in the third floor where the granularity is four times finer. Because of the calorimeter structure, the typical numbers of cells are larger in the forward region. Standard electron

identification estimators are proposed: EMF, the fraction of energy in the road contained in the first three floors; and E/p, the ratio of the energy contained in the first three floors of the road to the track momentum measured in the inner track detector.

To design identification criteria for electrons originating from heavy-flavour hadron decays in high p_T jets, samples of $\mathrm{Z} \to \mathrm{b\bar{b}}$ and $\mathrm{Z} \to \mathrm{c\bar{c}}$ events have been analysed, in conjunction with a sample of QCD events generated with a $\hat{p}_T$ in the hard process in excess of $40\,\mathrm{GeV}/c$. A procedure for the optimization of the identification criteria, depending on the background and efficiency levels required and on the physics process considered, has been developed. In addition to the selection based on those identification criteria, kinematic cuts have also been considered. They make use of the fraction of jet energy carried by the electron, and of the electron transverse momentum relative to the jet axis.

Electrons produced by direct $\mathrm{b} \to \mathrm{eX}$ decays in $\mathrm{Z} \to \mathrm{b\bar{b}}$ events are selected with an efficiency of 56% for a hadron misindentification rate of 0.3%. Here, the normalization is made with respect to tracks reconstructed with a $p_T > 2\,\mathrm{GeV}/c$, in the geometrical acceptance $|\eta_{\mathrm{det}}| < 1$ or $1.4 < |\eta_{\mathrm{det}}| < 3.2$, and within $\Delta R < 1$ from a reconstructed jet. The corresponding overall efficiency to tag a direct $\mathrm{b} \to \mathrm{eX}$ decay is 25%, with a 1% mistag probability for a generic jet. At present, the main improvement in performance is likely to come from an increased charged track reconstruction efficiency, currently at the level of 75% in the angular range considered.

To be applied to the data, the road method requires efficient track reconstruction, good relative alignment of the calorimeter and of the tracking system, and calorimeter calibration for low energy electrons. Results are to be expected within the coming months.

Appendices

A Floor and layer numbering

According to the `caladdress` package, the numbering of the floors and layers is the following.

Table 8: Floor names, index, and layer numbering.

Name	Floor Number	Layer number
EM1	1	1
EM2	2	2
EM3	3	3,4,5,6
EM4	4	7
FH1	5	11
FH2	9	12
FH3	10	13

Floor	Floor Number	Layer number
FH4	11	14
CH1	12	15
CH2	13	16
CH3	14	17
CCMG	6	8
ICD	7	9
ECMG	8	10

B Implementation details

The road method is computed by three classes which belong to the `SEMReco` package [3]. `SEMReco` reads the tracks from the `ChargedParticleChunk`. In the present version of `ChPartReco.rcp`, this is equivalent to using `Gtr401`.

`RoadFinder` determines the list of cells within the road in the $|\eta_{\rm det}| < 1$ region, and `ECRoadFinder` elsewhere. The parameters which drive the road construction are included in `SEMReco.rcp`. Three roads with different widths are computed. The current CC algorithm default values are:

```
float dismerge_container  = 0.25
float dismerge_road1  = 0.125
float dismerge_road2  = 0.
```

From the p11 version on, the followings default values are used by the EC algorithm:

```
float radius_container = 4. // cm
float radius_road1 = 3.     // cm
float radius_road2 = 2.5    // cm
```

There is a one to one correspondance between the values used in the CC and in the EC. The largest road contains the others; therefore *dismerge_container* (*radius_container*) has to be greater than *dismerge_road*1 (*radius_road*1) and than *dismerge_road*2 (*radius_road*2).

When the road is filled with the transverse energy deposits, only cells with $E >$ *E_threshold* are included. The default threshold is:

```
float E_threshold = 0.01  //in GeV
```

Finally, a selection can be made at the reconstruction level using EMF and E/p. By default, this selection is not activated:

```
bool requireRoadselection = false
int road_number = 1   // 0 = container
float road_emfraction_cut = 0.8
float road_eopmin_cut = 0.55
float road_eopmax_cut = 1.1
```

The road information is stored in a `SEMparticle` object, which is defined in the `em_evt` package. At the end of the reconstruction, it is stored in the `SEMparticleChunk`. Since the t01.73.00 release, and consequently in p11, the transverse energy deposits can be obtained with:

```
/// Get Road transverse energy deposits EM1,2,3,4,FH,CH,others
void getRoadEt(std::vector<float>& ret,int num=1) const;
```

where *num* is the road label, *i.e.*, *num* = 0 means the largest road, built with the parameters *dismerge_container* and *radius_container*, *num* = 1 and *num* = 2 mean road 1 and road 2. If *num* is not given, then *num* = 1 is the default.

It is also possible to get directly E/p and EMF:

```
/// road Electromagnetic fraction
float road_emf(int num=1) const;
/// Road E/p
float road_eop(int num=1) const;
```

The LinkIndex to the `ChargedParticle` is:

```
/// Return ChargedParticle LinkIndex.
const ChpIndex& chpindex() const;
```

The number of the last floor crossed by the central road is:

```
/// Get the last floor crossed by CC-Road
int out_floor() const ;
```

A vector containing the addresses of the cells within the road can be obtained with:

```
/// Get the Roads_cell vector.
void getRoads_cellvector(std::vector<Roads_cell>& rcellvec) const ;
```

where `Roads_cell` is a structure defined in `em_evt`. One `Roads_cell` represents one cell belonging to the road_container. It contains a CellAddress and two bits indicating whether the cell belongs to road1, road2 or both.

The `SEMparticleChunk` does not go into the thumbnail. In order to save space, only the transverse energy deposits are stored, and the transverse energies in the fine and coarse hadronic sections are added together. The three roads are available in the thumbnail. The methods to get the energy deposits, the electromagnetic fraction and E/p have been recoded in the `EMparticle` class, but have exactly the same syntax as before.

The `em_analyze` package [6] fills the `EMSEM` block of the root-tuple, which contains the road method output. In all versions up to now, the transverse energies in all floors and for the three road widths can be found in the variables: rRC, rR1, rR2 which contain the container-road, road 1 and road 2, respectively. They are two-dimensional arrays, with the first index for the particle number and the second for the floor number(between 0 and 13). The TRKindx makes it possible to link the `SEMparticle` to the track by matching it with the chidx variable in the `CH` block.

Some additional information can be found in the root-tuple:

- rccfl : number of the last central floor crossed by the track
- rcrossfl : bit pattern indicating the floors crossed, with the conventions of Appendix A.

In the p10-branch, for versions higher than p10.11.00, the information for the road built with dismerge=0.125 and radius=4 cm has been made easier to access:

- reop : E/p of the candidate
- emf : electromagnetic fraction EMF
- rem1f, rem2f, rem3f, rem4f : total road energy fractions in floors EM1 to 4.

For versions higher than t01.73.00 and for the p11-branch, almost the same has been done for all three roads. The names are:

- rCeop, r1eop, r2eop : E/p of the candidate
- rCem1f, r1em1f, r2em1f : total road energy fractions in the first floor. Similar quantities are available for the three other electromagnetic floors.

C Optimization procedure

The optimization procedure is carried out requiring a given misidentification rate. Three cuts have to be adjusted:

- the minimum value $EMF_{\rm low}$ for the electromagnetic fraction EMF,
- the minimum value $(E/p)_{\rm low}$ for E/p, and
- the maximum value $(E/p)_{\rm high}$ for E/p.

For a given value of $EMF_{\rm low}$, both E/p cuts are adjusted in order to maintain the required background level. This is done as follows.

- For a given value of $(E/p)_{\rm high}$, $(E/p)_{\rm low}$ is varied until the required background level is reached, and the signal efficiency is calculated.
- This procedure is repeated for various values of $(E/p)_{\rm high}$. As expected, $(E/p)_{\rm low}$ increases with $(E/p)_{\rm high}$, while the efficiency exhibits a maximum, as shown in Fig. 40. A local fit is performed around this maximum to obtain the optimal (E/p) cut values and the associated signal efficiency.

The same procedure is repeated for various $EMF_{\rm low}$ values. An example of the simultaneous dependence on $EMF_{\rm low}$ of the efficiency and of the cuts on E/p is shown in Fig. 41. The maximum maximorum for the signal efficiency is determined, thus fixing the optimal value for $EMF_{\rm low}$, from which the optimal values of $(E/p)_{\rm low}$ and $(E/p)_{\rm high}$ are deduced. (Due to the limited statistics of fake electron candidates in the QCD sample, some smoothing of $(E/p)_{\rm high}$ is performed for large $EMF_{\rm low}$ values.)

When the required misidentification rate is allowed to vary, the optimized efficiency changes, as shown in Fig. 27 for direct electrons from $\mathrm{Z} \to \mathrm{b}\bar{\mathrm{b}} \to \mathrm{eX}$. The corresponding variations of the optimal cuts are shown in Figs. 42 and 43 for the central and forward regions, respectively.

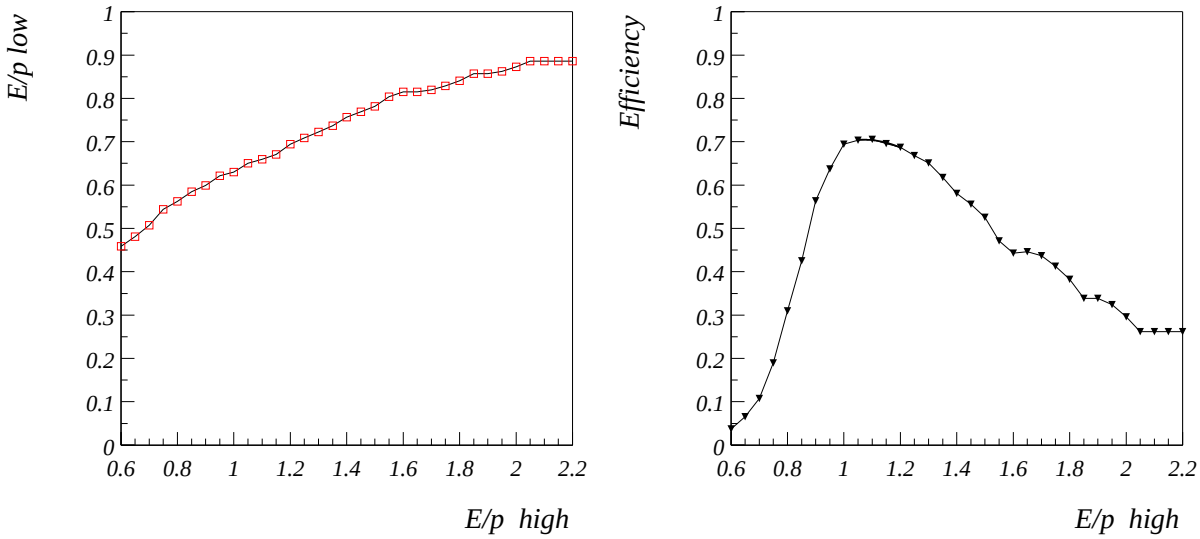


Figure 40: For $EMF_{\text{low}} = 0.83$, optimal $(E/p)_{\text{low}}$ value (left) and associated efficiency (right), as a function of $(E/p)_{\text{high}}$. The signal consists of direct electrons from $Z \to b\bar{b} \to eX$ in the central region.

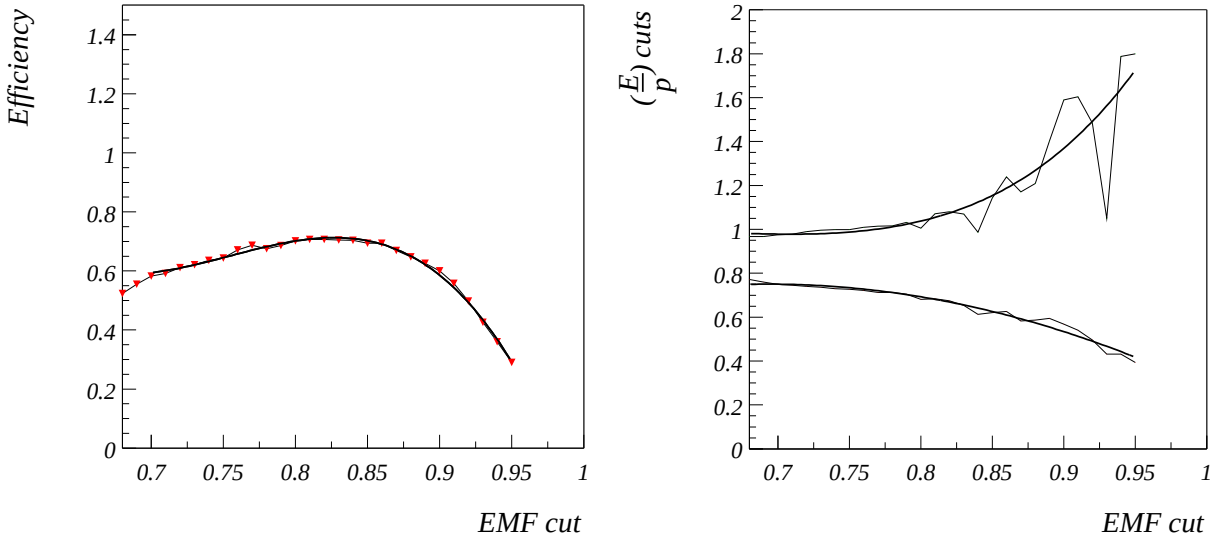


Figure 41: Signal efficiency (left) and E/p cut values (right), as a function of the EMF_{low} cut value. The signal consists of direct electrons from $Z \to b\bar{b} \to eX$ in the central region.

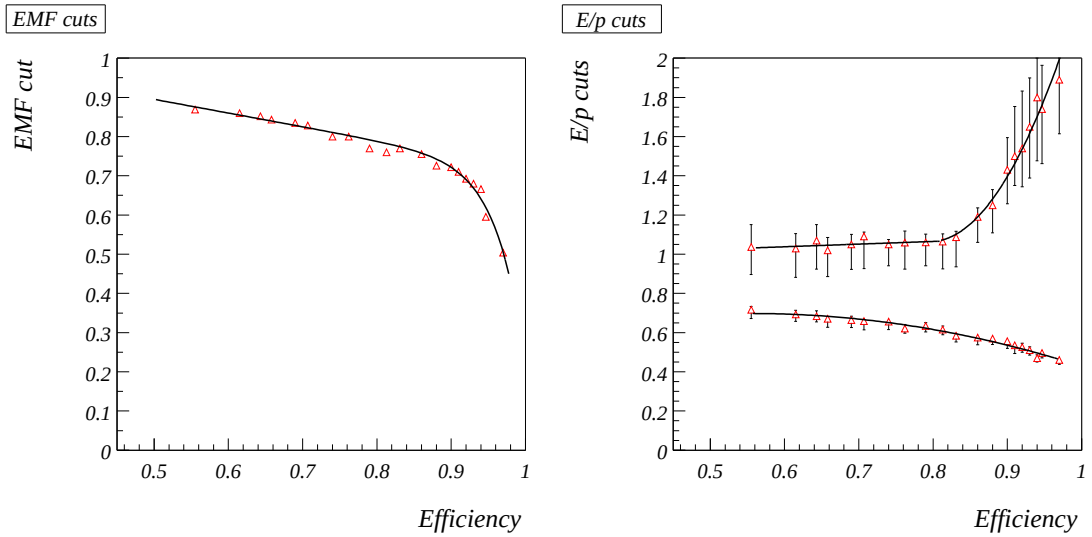


Figure 42: Optimized cuts as a function of the efficiency for direct electrons from $Z \to b\bar{b} \to eX$ in the central region.

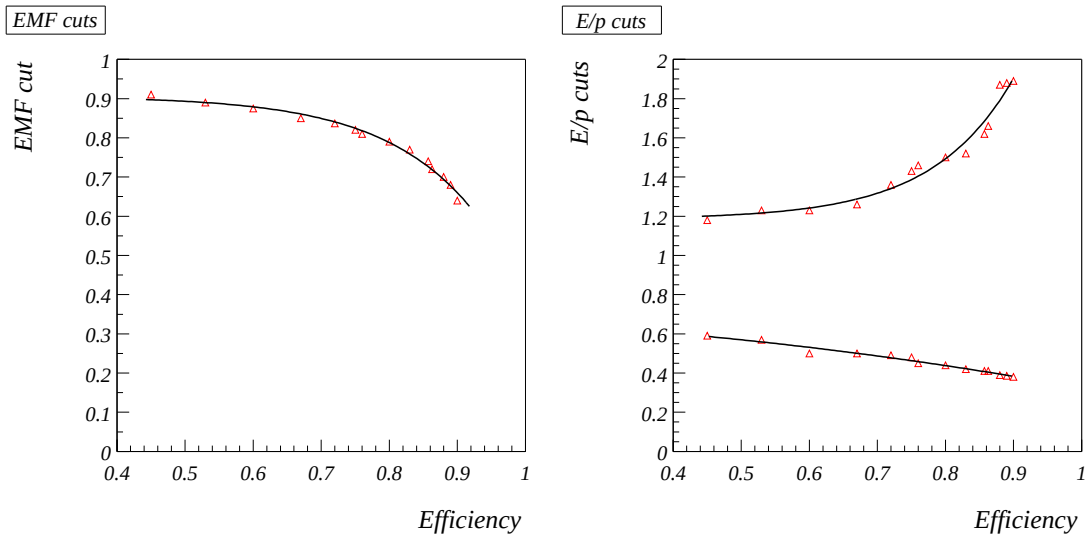


Figure 43: Optimized cuts as a function of the efficiency for direct electrons from $Z \to b\bar{b} \to eX$ in the forward region.

References

[1] EMRECO *"EM-ID group web-page :"*
`http://www-d0.fnal.gov/phys_id/emid/d0_private/emid_vertical_review.html`

[2] F. Beaudette and J.-F. Grivaz, *"b-tagging with electrons using the road method"*, D0 Note 3828.

[3] F. Fleuret, F. Beaudette and S. Protopopescu, *"The D0 Soft Electron Reconstruction Package SEMReco"*, D0 Note 3872.

[4] L. Duflot *"caltrack_extra documentation"*
`http://www-d0.fnal.gov/d0dist/dist/releases/current/\`
`caltrack_extra/doc/caltrack_extra.html`

[5] D. Abbaneo *et al.*, *"A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model"*, CERN-EP/2001-021.

[6] F. Fleuret. *"The D0 Electron/Photon Analysis Package EMAnalyze"*, D0 Note 3888.