Electron and photon identification in the D0 experiment

V.M. Abazov et al., Nuclear Instruments & Methods in Physics Research A (2014), http://dx.doi.org/10.1016/j.nima.2014.03.013
The D0 Collaboration 1

ARTICLE INFO

Article history:
Received 30 December 2013
Received in revised form 3 March 2014
Accepted 5 March 2014

Keywords:
Fermilab
DZero
D0
Tevatron Run II
Electron and photon identification
Electron and photon reconstruction

1. Introduction

The precise and efficient reconstruction and identification of electrons 2 and photons at the D0 experiment at the Fermilab Tevatron p p collider is essential for a broad spectrum of physics analyses, including high precision standard model (SM) measurements and searches for new phenomena. To satisfy this requirement, the D0 detector was designed to have excellent performance for the measurements of electrons and photons of energies from a few GeV up to 100 GeV. Another design requirement was to have good discrimination between jets and electrons or photons, since physics measurements often suffer from large backgrounds induced by jets being misidentified as electrons or photons. In this paper, the reconstruction of electromagnetic (EM) objects using D0 data is described. The determination of the electron energy scale and resolution is presented. Studies of the performance of the electron and photon reconstruction and identification are summarized. The results are based on measurements of Z boson decay events of Z → ee and Z → μμ (e, μ) collected in p p collisions at a center-of-mass energy of 1.96 TeV using an integrated luminosity of up to 10 fb -1.

© 2014 Published by Elsevier B.V.

2. D0 detector

The D0 detector is described elsewhere [1]. The components most relevant to electron and photon identification are the central tracking system, composed of a silicon microstrip tracker (SMT) that is located near the p p interaction point and a central fiber tracker (CFT) embedded in a 2 T solenoidal magnetic field, a central preshower (CPS), and a liquid-argon/uranium sampling calorimeter. The CPS is located before the inner layer of the calorimeter, outside the calorimeter cryostat, and is formed of one radiation length of absorber followed by three layers of scintillating strips. The D0 coordinate system is right-hand. The z-axis points in the direction of the Tevatron proton beam, and the y-axis points upwards. Pseudorapidity is defined as η = −ln[tan(θ/2)], where θ is the...
The hadronic section is composed of (by the CH layers. To eliminate possible contamination from diboson, \(Z + \) jets and \(W + \) jets processes, events with at least one isolated high-\(p_T\) muon [5], events with an invariant mass of the EM cluster and an isolated track between 60 and 120 GeV, and events with missing transverse energy [6] greater than 10 GeV are rejected. For studies of jets misidentified as photons, the \(\gamma + \) jet component containing a real photon is removed from the dijet sample by requiring that the EM cluster be non-isolated by cutting on the shower isolation fraction (see Section 4.1) of 0.07 < \(f_{iso}\) < 0.15. The \(Z \rightarrow \gamma\ell\ell\) and \(Z \rightarrow \gamma\ell\ell\) data events are collected using single electron triggers as described in Section 7.2. For \(Z \rightarrow \mu\gamma\) and dijet data events, single muon triggers [5] and jet triggers [4] are used, respectively.

The data used in physics analyses were collected by the D0 detector during Tevatron Run II between April 2002 and September 2011 and correspond to an integrated luminosity of approximately 10 fb\(^{-1}\).

The \(Z \rightarrow \gamma\ell\ell\) signal samples are generated using the AIPGEN generator [7] interfaced to PYTHIA [8] for parton showering and hadronization. The simulated transverse momentum \(p_T\) distribution of the Z boson is weighted to match the distribution observed in data [9]. Diphoton and \(Z \rightarrow \gamma\ell\ell\) signal events, and dijet background samples are generated using PYTHIA [8]. All MC samples used here are generated using the CTEQ6L1 [10] parton distribution functions, followed by a GEANT [11] simulation of the D0 detector. To accurately model the effects of multiple pp interactions in a single bunch crossing and detector noise, data from random pp bunch crossings are overlaid on the MC events. The instantaneous luminosity spectrum of these overlaid events is matched to that of the events used in the data analysis. Simulated events are processed using the same reconstruction code that is used for data.

4. EM object reconstruction and identification

EM objects – electrons and photons – are reconstructed by detecting localized energy deposits in the EM calorimeter. Confirmation of the existence of an electron track is sought from the central tracking system since an isolated high-\(p_T\) track should originate from the interaction vertex. The hadronic calorimeter, preshower, and tracking systems can be used to differentiate electrons and photons from jets.

4.1. EM cluster reconstruction

EM objects in the D0 detector are reconstructed using the nearly 55,000 calorimeter channels. Only channels with energies above noise are read out [4]. We use the same cluster reconstruction algorithm for electrons and photons, since their showers consist of collimated clusters of energy deposited mainly in the EM layers of the calorimeter. Calorimeter cells with the same \(\eta\) and \(\phi\) are grouped together to form towers. For the calculation of the energy of EM clusters, we sum the energies measured in the four EM layers and the first hadronic (FF1) layer which is included to account for leakage of energy of EM objects into the hadronic part of the calorimeter. Starting with the highest transverse energy tower (\(E_T > 500\) MeV), energies of adjacent towers in a cone of
radius \( \rho = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} \) = 0.4 around the highest \( E_T \) tower are added to form EM clusters in the CC. In the EC, EM clusters are a set of adjacent cells with a transverse distance of less than 10 cm from an initial cell with the highest energy content in the EM3 layer.

To be selected as an EM candidate, EM clusters must satisfy the following set of criteria:

- The cluster transverse energy must be \( E_T > 1.5 \text{ GeV} \).
- The fraction of energy in the EM layers is
  \[
  f_{EM} = \frac{E_{EM}}{E_{tot}}
  \]  
  (1)
  where \( E_{EM} \) is the cluster energy in the EM layers, and \( E_{tot} \) is the total energy of the cluster in all layers within the cone. At least 90% of the energy should be deposited in the EM layers of the calorimeter.
- The isolation fraction is the ratio of the energy in an isolation cone surrounding an EM cluster to the energy of the EM cluster,
  \[
  f_{iso} = \frac{E_{iso}(\rho < 0.4) - E_{EM}(\rho < 0.2)}{E_{EM}(\rho < 0.2)}
  \]  
  (2)
  where \( E_{iso}(\rho < 0.4) \) is the total energy within a cone of radius \( \rho = 0.4 \) around the cluster, summed over the entire depth of the calorimeter except the CH layers, and \( E_{EM}(\rho < 0.2) \) is the energy in the towers in a cone of radius \( \rho = 0.2 \) summed over the EM layers only. To select isolated electron or photon clusters in the calorimeter, we require an isolation fraction of less than 0.2.

For each EM candidate, the centroid of the EM cluster is computed by weighting cell coordinates with cell energies in the EM3 layer of the calorimeter. The shower centroid position together with the location of the pp collision vertex is used to calculate the direction of the EM object momentum.

Since EM objects begin to shower in the preshower detector, clusters are also formed in that detector. Single layer clusters are formed from scintillating strips for each layer. A preshower cluster is built by combining the single layer clusters from each of the three layers. These preshower clusters are extensively used to help identify the electrons and photons, and to build the multivariate identification methods, as well as to find the right interaction vertex for the photon as described in the following sections.

Electron candidates are distinguished from photon candidates by the presence of a track with \( p_T > 1.5 \text{ GeV} \) within a window of \( \Delta \eta \times \Delta \phi = 0.05 \times 0.05 \) around the coordinates of the EM cluster. The momentum of an electron candidate is recalculated using the direction of the best spatially matched track while the energy of the electron is measured by the calorimeter due to limited momentum resolution of the central tracking system. An EM cluster is considered to be a photon candidate if there is no associated track.

4.2. EM object identification

After applying the above criteria to EM clusters, there remains a considerable fraction of jets misidentified as EM objects. Further criteria must be applied to reject these misidentified jets and increase the purity of the selected electron and photon candidates. The following is a description of the quantities employed for electron and photon identification. There are a number of different selection criteria for these quantities to meet the needs of different physics analyses.

---

The cluster transverse energy must be \( E_T > 1.5 \text{ GeV} \).

The fraction of energy in the EM layers is
  \[
  f_{EM} = \frac{E_{EM}}{E_{tot}}
  \]
  (1)
  where \( E_{EM} \) is the cluster energy in the EM layers, and \( E_{tot} \) is the total energy of the cluster in all layers within the cone. At least 90% of the energy should be deposited in the EM layers of the calorimeter.

The isolation fraction is the ratio of the energy in an isolation cone surrounding an EM cluster to the energy of the EM cluster,
  \[
  f_{iso} = \frac{E_{iso}(\rho < 0.4) - E_{EM}(\rho < 0.2)}{E_{EM}(\rho < 0.2)}
  \]
  (2)
  where \( E_{iso}(\rho < 0.4) \) is the total energy within a cone of radius \( \rho = 0.4 \) around the cluster, summed over the entire depth of the calorimeter except the CH layers, and \( E_{EM}(\rho < 0.2) \) is the energy in the towers in a cone of radius \( \rho = 0.2 \) summed over the EM layers only. To select isolated electron or photon clusters in the calorimeter, we require an isolation fraction of less than 0.2.

For each EM candidate, the centroid of the EM cluster is computed by weighting cell coordinates with cell energies in the EM3 layer of the calorimeter. The shower centroid position together with the location of the pp collision vertex is used to calculate the direction of the EM object momentum.

Since EM objects begin to shower in the preshower detector, clusters are also formed in that detector. Single layer clusters are formed from scintillating strips for each layer. A preshower cluster is built by combining the single layer clusters from each of the three layers. These preshower clusters are extensively used to help identify the electrons and photons, and to build the multivariate identification methods, as well as to find the right interaction vertex for the photon as described in the following sections.

Electron candidates are distinguished from photon candidates by the presence of a track with \( p_T > 1.5 \text{ GeV} \) within a window of \( \Delta \eta \times \Delta \phi = 0.05 \times 0.05 \) around the coordinates of the EM cluster. The momentum of an electron candidate is recalculated using the direction of the best spatially matched track while the energy of the electron is measured by the calorimeter due to limited momentum resolution of the central tracking system. An EM cluster is considered to be a photon candidate if there is no associated track.
In the EC the matrix is of dimension $8 \times 8$, while in the CC $\sigma_{\pm}$ is not used and therefore the matrix has the dimension $7 \times 7$. A separate matrix is built for each ring of calorimeter cells with the same $|\eta|$ coordinate. To measure how closely the shower shape of an electron candidate matches expectations from MC simulations, a $\chi^2$ value is calculated ($\chi^2_{\text{cal}}$). Since the electron and photon candidates tend to have smaller $\chi^2_{\text{cal}}$ than jets, this variable can be used to discriminate between EM and hadronic showers.

Track isolation: For electrons and photons that are isolated, the scalar sum of the $p_T$ of all charged particle tracks with $p_T > 0.5$ GeV, excluding the associated track for the EM cluster, originating from the $pp$ collision vertex in an annular cone of $0.05 < \varphi < 0.4$ around the electron and photon candidates, $2p_T^{\text{iso}}$, is expected to be small. It is therefore a sensitive variable for discriminating between EM objects and jets.

Track match: For electron identification, to suppress photons and jets misidentified as electrons, the cluster is required to be associated with a track in the central tracking system in a road between the EM calorimeter cluster and the $pp$ collision vertex satisfying the differences $|\Delta\eta_{\text{EM} \text{trk}}| < 0.05$ and $|\Delta\phi_{\text{EM} \text{trk}}| < 0.05$ for the differences between $\eta$ and $\phi$ of the EM cluster and the associated track. To quantify the quality of the cluster-track matching, a matching probability $P(\chi^2_{\text{spatial}})$ is defined using $\chi^2_{\text{spatial}}$, which is given by

$$\chi^2_{\text{spatial}} = \left(\frac{\Delta\eta}{\delta\eta}\right)^2 + \left(\frac{\Delta\phi}{\delta\phi}\right)^2,$$

The probability is computed for each matched track. In these expressions, $\Delta\eta$ and $\Delta\phi$ are the differences between the track position and the EM cluster position in the EM3 layer of the calorimeter. The variables $\delta\eta$ and $\delta\phi$ are the resolutions of the associated quantities. The track with the highest $P(\chi^2_{\text{spatial}})$ is taken to be the track matched to the EM object. If there is no matched track, $P(\chi^2_{\text{spatial}})$ is set to $-1$.

Hits on road: Due to tracking inefficiencies, the cluster-track matching probability method is not fully efficient in separating electron from photon candidates, in particular in events with high instantaneous luminosity. To improve the separation between electron and photon candidates, a "hits on road" discriminant, $D_{\text{hor}}$, is used in the CC. For each EM object, a "road" is defined between the $pp$ collision vertex position and the CPS cluster position, if it is matched to the EM object, or else to the EM cluster position. To account for the different sign of the electric charge of electrons and positrons, two roads (positive-charge and negative-charge roads) are defined. The number of hits from CFT fibers and SMT strips along the EM cluster’s trajectory, $N_{\text{hits}}$, is counted. The discriminant $D_{\text{hor}}$ is defined by

$$D_{\text{hor}} = \frac{P_e(N_{\text{hits}})}{P_e(N_{\text{hits}}) + P_{\gamma}(N_{\text{hits}})},$$

where $P_e$ and $P_{\gamma}$ are the probabilities in the bin of $N_{\text{hits}}$, given by

$$P_e(N_{\text{hits}}) = \frac{\sum_{i=1}^{N_0} N_i}{\sum_{i=0}^{N_e} N_i},$$

$$P_{\gamma}(N_{\text{hits}}) = \frac{\sum_{i=0}^{N_{\gamma}} N_i}{\sum_{i=0}^{N_{\gamma}} N_i}$$

where $N_0$ and $N_{\gamma}$ are the number of electrons and fake electrons in the bin $N_{\text{hits}} = i$ from $Z \rightarrow ee$ and multijet data events, respectively. The maximum number of hits is $24$, as the maximum of CFT hits is $16$ and the maximum of SMT hits is $8$. Electrons tend to have $D_{\text{hor}} \approx 1$, while photons tend to have values close to $0$.

Figs. 2–4 show distributions of identification variables for EM candidates from $Z \rightarrow ee$ data and MC events, as well as from diphoton and dijet MC events. As can be inferred from the distributions, the simulation has some imperfections in modeling the shower shapes mainly caused by an insufficient description of instrumented material [2]. This is accounted for when correcting simulated electron and photon identification efficiencies utilizing data as described in Sections 7 and 8, respectively.

5. Multivariate identification methods

The variables described in Section 4.2 allow efficient identification of electron and photon candidates. However, to maximize the identification efficiencies of electrons and photons and to minimize the misidentification rate from jets in physics analyses, various multivariate analysis (MVA) techniques are explored. One MVA technique, the H-matrix method, has already been discussed in Section 4.2. Two more types of MVAs that are used in physics analyses are a Likelihood method for electrons and a neural network (NN) method for electrons and photons. H-matrix, Likelihood, and NN methods achieve an improved background rejection. However, the H-matrix is mainly based on the calorimeter information, while the Likelihood method includes the tracking information in addition, while the advantage of the NN is that it includes CPS information. The electron identification efficiency and purity are therefore found to be improved when these MVA output variables are utilized together with other electron reconstruction variables as input to a Boosted Decision Tree (BDT) [14]. All MVAs except the H-matrix are described in this section.

5.1. Electron likelihood

Likelihood-based identification of electron candidates is an efficient technique for separating electrons from background by combining information from various detector components into a single discriminant.

There are several mechanisms by which particles, either isolated or in jets, may produce electron signatures. Photon conversions may be marked by the presence of a track very close to the track matched to the EM cluster, or a large $E_T/p_T$ when the closely situated ee pair is reconstructed as a single EM cluster and only one track is identified. Here, $E_T$ is the transverse energy of the cluster measured by the calorimeter and $p_T$ is the transverse momentum of the associated track measured by the tracker. The calorimeter quantities describing the shower shape, however, are nearly identical to that of an electron, though photon calorimeter clusters may be slightly wider than an electron shower.Neutral pions ($\pi^0$) may also have nearby tracks, as they are generally produced in association with other charged hadrons. Since the $\pi^0 \rightarrow \gamma \gamma$ decay would have to overlap with a charged hadron track in order to fake an electron, the track matching quantity could be poor, and the track would not necessarily be isolated. The H-matrix $\chi^2_{\text{cal}}$ and $f_{\text{iso}}$ of the EM object may be influenced by the surrounding hadrons. The following eight variables are used to calculate the electron likelihood:

- EM energy fraction $f_{\text{EM}}$;
- EM shower isolation $f_{\text{iso}}$;
- H-matrix $\chi^2_{\text{cal}}$;
- $E_T/p_T$;
- transverse impact parameter of the selected track with respect to the $pp$ collision vertex;
- $E_T$;
- $p_T^\gamma$.

The distributions shown in this paper are generally derived from subsets of the Run II data sample. The small variations observed between different periods of Run II are treated as systematic uncertainties in physics analyses.

For definitions see Section 4.
number of tracks with $p_T > 0.5$ GeV in a cone of radius $\mathcal{R} = 0.4$ around and including the matched track;

- cluster-track matching probability $P(\chi^2_{\text{spatial}})$;

- track isolation variable $\Sigma p_T^{\text{trk}}$.

The distributions of these eight variables are normalized to unit area to generate probability density distributions for each variable from $Z \rightarrow \text{ee}$ and dijet data for signal and background, respectively. These distributions are used to assign a probability for a given EM object to be signal or background. To quantify the degree of correlation between the input variables, we calculate the correlation coefficients. We find that most of the combinations have correlation coefficients close to zero and hence are mutually uncorrelated. Others do not exceed 55% for signal or fake electrons. The product of individual probabilities from all variables is correlated with the overall probability for the EM object to be an electron. To differentiate between signal-like and background-like electron candidates, a likelihood discriminant is calculated:

$$\mathcal{L} = \frac{P_{\text{sig}}}{P_{\text{sig}} + P_{\text{bkg}}}$$

where $P_{\text{sig}}$ and $P_{\text{bkg}}$ are the overall probabilities for signal and background, respectively. Distributions of this discriminant for electron candidates in the CC and EC are presented in Fig. 5. This demonstrates the enhanced power to separate between genuine electrons, which peak at large values of the discriminant, and jets, which peak at low values.

5.2. Neural Network for electron and photon identification

To further suppress jets misidentified as electrons and photons, we train an NN [15] using a set of variables that are sensitive to differences between electrons (photons) and jets. The variables,
Fig. 3. Normalized distributions of EM object identification variables as defined in Section 4.2 for $Z \rightarrow ee$ data and MC events, and for diphoton and dijet MC events in the EC. Presented are (a) the EM energy fraction, (b) the EM shower isolation, (c) the width of the EM shower in the transverse plane, (d) the width of the EM shower in the pseudorapidity direction, (e) the track isolation, and (f) the track matching probability. The first bin of the track matching probability distribution indicates no track match.

Fig. 4. The distributions of $\chi^2_{Cal}$ for EM candidates for $Z \rightarrow ee$ data and MC events, and for diphoton and dijet MC events in the CC (a) and EC (b).
selected to explore both the tracker activity and the energy distribution in the calorimeter and CPS, are listed below.

- fraction of the EM cluster energy deposited in the first EM calorimeter layer (\(f_{\text{EM1}}\));

- number of charged particle tracks with \(p_T > 0.5\) GeV originating from the pp collision vertex within \(R < 0.05\) of the EM cluster (\(N_{\text{track}}\));

- number of charged particle tracks with \(p_T > 0.5\) GeV originating from the pp collision vertex within \(R < 0.2\) (\(N_{\text{track}}^{0.2}\));

- energy deposit in the CPS (\(E_{\text{EM}}\));

- number of cells above an EM cluster \(p_T\)-dependent threshold, given by \(0.004 \times E_{\text{EM}}\) (in GeV) \(+ 0.25\) GeV in the first EM calorimeter layer within \(R < 0.4\);

- number of hits in the \(\eta, \phi\) cells of the EM cluster; and

- track isolation variable \(\tilde{\sigma}^{pT}\).

\[ \sigma_{\text{EM1}}^2 = \sum \frac{E_i^2 \times (\phi_{\text{EM}} - \phi_i)^2}{\sum E_i^2} \]  

where \(E_i\) and \(\phi_i\) are the energy and azimuthal angle of the \(i\)th event in the \(\eta, \phi\) cells of the EM cluster; and \(\phi_{\text{EM}}\) is the azimuthal angle of the EM cluster at the EM3 layer.

\( \chi^2_{\text{Cal}} \) is calculated from the H-matrix.

Separate NNs are built for electrons and photons in the CC, whereas a single NN is used for electrons and photons in the EC. Table 1 lists the input variables utilized in each NN.

For the construction of the NN for electrons in the CC, the seven variables above are used as inputs (eNN7). Here, \(Z \rightarrow \text{ee}\) data events define the signal, and dijet data events define the background. Performance checks have been performed using \(Z \rightarrow \text{ee}\) and dijet MC events.

The NN for CC photons (\(\gamma\)NN5) is built from the same variables as eNN7 but excluding the tracker-based variable \(N_{\text{track}}\), and \(f_{\text{EM1}}\) since its distribution varies significantly with the \(E_T\) of the EM cluster. The direct diphoton MC defines the signal, and dijet MC events are used as background in training the NN. For testing, the reconstructed radiated photon from \(Z \rightarrow e^+e^-\) events in data and MC events, and dijet MC events are used.

A photon NN (\(\gamma\)NN4) is built with four input variables as listed in Table 1 for the EC region. The training is based on direct diphoton and dijet MC events. The same types of events used to test \(\gamma\)NN5 are used to test \(\gamma\)NN4. Considering the similar performance of the input variables of electrons and photons in the EC, \(\gamma\)NN4 is found to work well, and is used, for electron identification in the EC.

Fig. 6 shows the NN output distributions for reconstructed EM clusters with \(R^{(\text{spatial})} > 0.001\) (electron candidates) and without track match (photon candidates) for \(Z \rightarrow \text{ee}\) data and MC events, and for dijet background MC events. The distributions show good agreement between data and MC simulation and demonstrate good separation between signal and background.

To validate the photon NNs for jets, dijet data events in the jet-enriched calorimeter isolation region \(0.07 < f_{\text{iso}} < 0.15\) are compared to MC simulation. As shown in Fig. 7, good agreement between data and MC is observed.

### 5.3. Boosted decision trees for electron identification

To enhance the efficiency and purity of electron identification, a BDT is constructed utilizing variables that are significantly different for signal and background leading to a strong discrimination power of the BDT output distribution. The following variables are used to construct the BDT:

- EM energy fraction \(f_{\text{EM1}}\);

- EM shower isolation \(f_{\text{iso}}\);

- energy deposit in EM1, EM2, EM3, EM4 and FH1; and

- \(\sigma_{\text{EM1}}^2\) in EM1, EM2, EM3, EM4 and FH1;

- \(\sigma_{\text{EM2}}^2\) in EM1, EM2, EM3, EM4 and FH1;

- H-matrix \(\chi^2_{\text{Cal}}\);

- \(\Sigma_{\text{EM1}}\) and \(\Sigma_{\text{EM3}}\);

- cluster-track matching probability \(P(\chi^2_{\text{spatial}})\);

- “hits on road” discriminant \(D_{\text{hit}}\) in CC;

- ratio \(E_T/p_T\);

- number of charged particle tracks with \(p_T > 0.5\) GeV originating from the pp collision vertex within \(R < 0.05\) of the EM 127 cluster;

- electron likelihood discriminant \(L\); and

- output distribution of eNN7 in CC; and

- output distribution of \(\gamma\)NN4 in EC;

- squared width of the energy deposit in the CPS \(\sigma_{\text{EM1}}^2\) in CC.

---

Table 1

<table>
<thead>
<tr>
<th>Input variables</th>
<th>eNN7 in CC</th>
<th>(\gamma)NN5 in CC</th>
<th>(\gamma)NN4 in EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f_{\text{EM1}})</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>(N_{\text{track}}^{0.2})</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>(N_{\text{track}}^{0.4})</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>(\Sigma_{\text{EM1}})</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>(N_{\text{iso}}^{0.05})</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>(N_{\text{track}})</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>(\sigma_{\text{EM1}}^2)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
<tr>
<td>H-matrix (\chi^2_{\text{Cal}})</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
<td>(\checkmark)</td>
</tr>
</tbody>
</table>
The methods presented here. Typically, the signal ef-motivate the training of separate BDTs for high and low instanta-

...in the two calorimeter regions, and the CC has a better 

2 for matching the spatial positions between CPS cluster and 

material in the MC simulation, discrepancy between data and MC 

exists. This has been studied and taken into account by applying 

corrections to the simulation.

6. Energy scale and resolution calibration

After EM objects are identified as described in Sections 4 and 5, the detector response to the energy of electrons and photons is calibrated. The electron energy scale and resolution are determined from $Z \rightarrow ee$ data events. EM showers induced by electrons and photons have similar distributions in the D0 calorimeter. However, EM clusters deposit energy in the passive material such as the inner detector and solenoid before reaching the calorimeter. On average, electrons lose more energy in this material than photons [16]. To account for this difference, MC simulations tuned to reproduce the response for electrons in data are used to derive the response difference between electrons and photons. In this section, the electron energy scale and resolution, and the energy scale difference between electrons and photons, are described.
6.1. Energy scale

The amount of material in front of the calorimeter varies between 3.4 and 5 $X_0$ in the CC and between 1.8 and 4.8 $X_0$ in the EC [2]. The fraction of energy deposited in each longitudinal layer of the calorimeter depends on the amount of that passive material. The energy loss in passive material is studied taking into account the energy profile dependence on the incident angle [17]. The differences of the energy response between data and the MC simulation are determined using $Z \rightarrow \text{ee}$ events and the corrections are applied to the MC simulation.

The energy response is degraded near the module $\phi$ boundaries for the EM layers of the CC. In addition to a degradation of energy response, the centroid position of the EM cluster is shifted. To study these effects, the following variable is defined:

$$\phi_{\text{mod}} = \text{mod} \left( \frac{16 \cdot \phi_{\text{EM}}}{\pi} \right)$$

(11)

where $\phi_{\text{EM}}$ is the azimuthal angle of the EM cluster. For track-matched electrons, $\phi_{\text{mod}}$ is determined by extrapolating the associated track through the known magnetic field towards the calorimeter. For photons and non-track-matched electrons, an average correction of the $\phi_{\text{mod}}$ is applied which was determined from track-matched electrons. Regions of $0.1 < \phi_{\text{mod}} < 0.9$ are referred to as “in-fiducial”, the values outside this range are defined as “non-fiducial”. Fig. 9 shows the dielectron invariant mass ($M_{\text{ee}}$) distribution for $Z \rightarrow \text{ee}$ data events with two CC electrons. The distribution is shown separately for events with 0, 1, and 2 electrons located in fiducial regions. Electrons in or close to module boundaries suffer significant energy losses. To correct for such energy loss, the $\phi_{\text{mod}}$ dependent energy scale corrections are derived for both data and MC simulation using $Z \rightarrow \text{ee}$ events. Due to the different amount of material traversed by the electrons before reaching the calorimeter, the events are split into five $\eta$ regions to derive the correction parameters. In addition, the energy loss near $\phi$ boundaries is larger for electrons with a poorly measured shower shape corresponding to a large H-matrix $\chi^2_{\text{Cal}}$. The energy scale corrections during Run II, the uncalibrated $Z$ boson mass is shifted to lower values in data events. The cause of this effect is discussed in Ref. [2]. The MC simulation, however, predicts an increase in the average EM energy with $E_{\text{inst}}$ due to extra energy from additional pp interactions.

$$E_{\text{inst}} = \text{mod} \left( \frac{16 \cdot E_{\text{EM}}}{\pi} \right)$$

(11)

In Run II, the uncalibrated $Z$ boson mass is shifted to lower values in data events. The cause of this effect is discussed in Ref. [2]. The MC simulation, however, predicts an increase in the average EM energy with $E_{\text{inst}}$ due to extra energy from additional pp interactions. In the data, calibration of the calorimeter largely corrects for this energy scale dependence on $E_{\text{inst}}$. Residual offset corrections are derived by fitting the distribution of $E_{\text{inst}}/p_T$ for electrons in $W \rightarrow e\nu$ events, taking advantage of the fact that the $p_T$ scale is independent of $E_{\text{inst}}$.

Individual cells in the EM calorimeter are known to saturate at energies varying from about 60 to 260 GeV, depending on the cell position. As a result, an EM cluster loses on average about 0.5% of its nominal energy at 300 (500) GeV. A simple correction truncates the energy of any cells in the MC that exceed the saturation value for that cell.

Due to the different amount of energy loss between electrons and photons in the passive material, the photon energy is over-corrected by applying the electron energy scale correction. The correction is about 3\% at $p_T = 20$ GeV, and it decreases at higher energies. The correction required for forward photons is slightly smaller. The reconstructed photon energy is corrected accordingly to compensate for the over-correction.

The systematic uncertainty for the electron energy scale correction is $\approx 0.5\%$, which is mainly caused by the limited statistics of $Z \rightarrow \text{ee}$ data events. For photons, additional 0.5\% systematic uncertainty is added in quadrature from electron-to-photon energy scale correction.

6.2. Energy resolution

The energy resolution of the calorimeter as a function of the electron/photon energy, $E$, can be written as

$$\frac{\sigma_{\text{SM}}}{E} = \sqrt{\frac{C_{\text{SM}}^2 + S_{\text{SM}}^2}{E} + N_{\text{SM}}^2}$$

(12)

with $C_{\text{SM}}$, $S_{\text{SM}}$ and $N_{\text{SM}}$ as the constant, sampling and noise terms, respectively. The constant term accounts for the non-uniformity of the calorimeter response. Its effect on the fractional resolution is independent of the energy, and therefore it is the dominant effect at high energies. The sampling term is due to the fluctuations related to the physical development of the shower, especially in sampling calorimeters where the energy deposited in the active medium fluctuates event by event because the active layers are interleaved with absorber layers. The noise term comes from the electronic noise of the readout...
system, radioactivity from the Uranium, and underlying events. Since the noise contribution is proportional to 1/E it is basically negligible for high energy electrons/photons. Due to the large amount of material in front of the calorimeter, $S_{EM}$ is not a constant and is parameterized as a function of electron energy and incident angle [21]. The constant term $C_{EM}$ is derived by a fit to the measured width of the $Z\rightarrow ee$ peak [2].

These terms are measured and applied to the true energy of electron.

The electron and photon energy resolution predicted by the GEANT-based [11] simulation of the D0 detector is better than observed in data. Furthermore, there are non-Gaussian tails in the resolution distribution that are poorly modeled by the fully simulated MC described in Section 3, partly because the finite charge collection time of the readout system of the calorimeter is neglected in the simulation. To account for both effects, an ad-hoc smearing is applied to the reconstructed energy of EM clusters following the GEANT simulation according to the following function, which was introduced by the Crystal Ball Collaboration [18]:

$$f(x; \alpha, n, \sigma, \xi) = \begin{cases} \exp\left(\frac{(x-x)^2}{2\sigma^2}\right) & \text{for } x-x > -\alpha \\ \left(\frac{n}{\tau}\right) \exp\left(-\frac{\tau - x}{\tau}\right) \left(\frac{\tau - x}{\tau}\right)^{-n} & \text{for } x-x < -\alpha \end{cases}$$

(13)

Here, the $\sigma$ parameter determines the width of the Gaussian core part of the resolution. The $\alpha$ parameter controls the energy below which the power law is used, and the $n$ parameter governs the exponent of the power law. The $\xi$ parameter is the mean of the Gaussian core part of the resolution. Typically, an increase in the width of the non-Gaussian tail needs to be compensated by an increase in the mean. The mean of $f(x)$ is around 0, and the simulated energy is scaled by $1+\tau$, where $x$ is sampled from the probability distribution function according to Eq. (13).

To determine the parameters of Eq. (13), a fit is performed by varying parameters applied to the MC, and minimizing the $\chi^2$ between the data and fully simulated MC in the $M_{ee}$ distribution.

The $n$ parameter is fixed since there is enough freedom in the other three parameters to adequately describe the data. A value of $n=7$ is found to be appropriate.

The parameters are fitted separately for the following three categories of EM clusters [19]:

- **Category 1**: CC in-fiducial CC in-fiducial clusters are defined as $|\eta| < 1.1$ and $0.1 < \phi_{mod} < 0.9$. The parameters are fitted using events in which both electrons are CC in-fiducial.
- **Category 2**: CC non-fiducial CC non-fiducial clusters are defined as $|\eta| < 1.1$ and $\phi_{mod} < 0.1$ or $\phi_{mod} > 0.9$. The parameters are fitted using events containing two CC electrons, where at least one is non-fiducial. Any CC in-fiducial electrons are smeared using their already tuned parameters.
- **Category 3**: EC EC clusters are defined as having $|\eta| > 1.5$. The parameters are fitted using events containing two EC electrons, or one CC in-fiducial or non-fiducial plus one EC electron. For EC clusters, a simple Gaussian smearing is used where the fit has only two parameters ($\xi, \sigma$).

![Fig. 10](image-url) shows a comparison of $M_{ee}$ distributions for $Z\rightarrow ee$ data and MC after applying the energy scale and smearing corrections. Good agreement between data and MC simulation is observed.

7. **Efficiencies of electron identification**

Electron trigger, preselection and identification efficiencies are measured in $Z\rightarrow ee$ data and MC events by selecting two high-$E_T$ electron candidates that have an invariant mass close to the $Z$ boson mass peak. To obtain an improved simulation, differences between the efficiencies measured in data and MC simulation are used to derive correction factors to be applied to MC events taking into account kinematic dependences.
7.1. Tag-and-probe method

To measure the efficiencies, a “tag-and-probe” method is used. In $Z \rightarrow ee$ decays, a $E_T > 30$ GeV electron candidate in CC fiducial is selected as the “tag” with the following requirements:

- $f_{EM} > 0.96$;
- $f_{iso} < 0.07$;
- $\Sigma p_T^b < 2$ GeV;
- associated track $p_T > 15$ GeV;
- $L > 0.8$;
- $eN7 > 0.7$.

The “probe” – used to perform the measurement of the identification efficiency – is either an EM cluster or a track. The invariant mass of the tag and probe electrons, $M_{pp}$, is required to be close to the $Z$ boson mass. If the probe is an EM cluster, $M_{pp}$ is required to be greater than 80 GeV but less than 100 GeV. The energy resolution for high-$p_T$ tracks is worse, and the $M_{pp}$ is required to be greater than 70 GeV but less than 110 GeV when the probe is a track. If the probe passes the tag selection criteria, it will also be used as a tag, resulting in the event being counted twice. To avoid bias, the same tag-and-probe method is used for both $Z \rightarrow ee$ data and MC events.

To remove the residual background from jet production in data events, a template fit is applied to the $M_{pp}$ distributions. The signal shape is obtained from $Z \rightarrow ee$ MC simulation, and the background shape is derived from dijet data. The fit is performed in various $\eta$ and $\phi$ regions. The systematic uncertainty for the tag-and-probe method is dominated by the statistics of $Z \rightarrow ee$ data events. It is $\approx 10\%$ for low probe $E_T (< 20$ GeV) region, and $\approx 3\%$ for high probe $E_T$ region.

7.2. Trigger efficiencies

There are two types of single electron triggers [2, 20]. One class of triggers is solely based on calorimeter information and the other class includes tracking information. Calorimeter-based triggers are used for both electrons and photons. To have higher trigger efficiencies for electrons, we combine both types of triggers by taking their logical OR. The tag-and-probe method is used to measure the trigger efficiencies in data. To be consistent with offline electron identification requirements (described in Section 7.4), the trigger efficiencies are measured with respect to each set of electron identification requirements. To account for dependencies on the EM cluster position in the detector, the trigger efficiencies are parametrized as a function of $E_T$ and $\eta$ of the electron candidate. Single triggers are used with an efficiency $\approx 100\%$ for transverse momenta above 30 GeV in the fiducial regions of the calorimeter up to $|\eta| < 2.5$.

7.3. Preselection efficiencies

Preselected electrons and photons are EM clusters that satisfy the criteria described in Section 4.1. The preselection efficiency is given by the fraction of tracks that match an EM cluster passing the preselection requirements for the probe electron candidate. In Fig. 11a the preselection efficiencies are presented for probe tracks in the CC as a function of $\phi_{mod}$ for data and the MC simulation. The average efficiency is $\approx 98\%$. Data and MC simulation show good agreement, except in non-fiducial regions. Therefore, the $\phi_{mod}$-dependent correction factors as shown in Fig. 11a are applied to MC to improve the simulation. Fig. 11b shows the preselection efficiencies as a function of $\eta$ for EC electrons. Efficiency losses are observed in the region $|\eta| > 2.5$ due to partial detector coverage for increasing $\eta$. To correct for data versus MC agreement, except in non-

### Table 2

Sets of requirements to identify electrons with loose and tight quality. (%):

<table>
<thead>
<tr>
<th>Variable</th>
<th>Loose CC</th>
<th>Loose EC</th>
<th>Tight CC</th>
<th>Tight EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_{EM}$</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>$f_{iso}$</td>
<td>0.09</td>
<td>0.08</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>$2\theta_{p}\times 4$ GeV</td>
<td>4 GeV</td>
<td>2.5 GeV</td>
<td>2.5 GeV</td>
<td>2.5 GeV</td>
</tr>
<tr>
<td>$H$-matrix $\eta_{MC}$</td>
<td>$\leq 40$</td>
<td>$\leq 35$</td>
<td>$\leq 40$</td>
<td>$\leq 40$</td>
</tr>
<tr>
<td>$\sigma_{p}\times 4$</td>
<td>$\leq (+$)</td>
<td>$\leq (+$)</td>
<td>$\leq (+$)</td>
<td>$\leq (+$)</td>
</tr>
<tr>
<td>$eN7(CC),\gamma_{NN}(EC)$</td>
<td>0.4</td>
<td>0.05</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>$P(U)$</td>
<td>$\leq 1$</td>
<td>$\leq 1$</td>
<td>$\leq 1$</td>
<td>$\leq 1$</td>
</tr>
<tr>
<td>$\eta_{stat}$</td>
<td>$\geq 0$</td>
<td>$\geq 0$</td>
<td>$\geq 0$</td>
<td>$\geq 0$</td>
</tr>
<tr>
<td>$L &gt;$</td>
<td>$\geq 3$</td>
<td>$\geq 6$</td>
<td>$\geq 6$</td>
<td>$\geq 6$</td>
</tr>
<tr>
<td>$E_T/p_T &lt;$</td>
<td>$\leq 1$</td>
<td>$\leq 1$</td>
<td>$\leq 1$</td>
<td>$\leq 1$</td>
</tr>
</tbody>
</table>

Fig. 11. Preselection efficiencies of probe tracks as a function of $\phi_{mod}$ and $\phi$ for electrons in the CC (a) and EC (b). $Z \rightarrow ee$ data is compared to the MC prediction, and the ratio data/MC is presented.
differences in the EC region, $\eta$-dependent factors are applied to the simulation. No significant differences between data and MC in other variables are observed for either electrons or photons in the CC and EC regions.

### 7.4. Electron identification efficiencies

Many sets of requirements for electron identification are provided for use in physics analyses, each with different electron selection efficiencies and misidentification rates. As examples, the electron identification efficiencies for two sets of requirements are presented here. These sets are called “loose” and “tight”. Table 2 lists the specific requirements of these two operating points.

The tag-and-probe method described in Section 7.1 is used here with the exception that now the probe electron is required to fulfill the preselection criteria. The identification efficiencies are measured in $\eta-\phi$ phase space. The resulting efficiencies for electrons in data and MC events and the ratio of efficiencies in

![Fig. 12. Electron identification efficiencies as a function of (a,b) $E_T$, (c,d) $\phi$ and (e,f) $\eta$ for loose electron requirements in CC and EC. Efficiencies for data and MC simulated $Z\rightarrow ee$ events are shown, as is the ratio of the data and MC efficiencies.](image-url)
the data and the MC simulation are shown in Figs. 12 and 13. In CC, the efficiencies in the $\eta \approx 0$ region are lower than in other regions since the light yield of the CFT is lower due to a shorter path length through the scintillating fiber. In EC, the efficiencies decrease in high $\eta$ region due to the partial coverage of the tracking system. The dependence of the efficiencies on $\phi$ are mainly caused by the azimuthal variations of the CFT waveguide length not taken into account in simulation.

To account for deficiencies of the simulation, the simulation is corrected by applying $\eta$ and $\phi$ dependent correction factors. The dependence on instantaneous luminosity for the electron reconstruction efficiencies is studied and derived following $(\eta - \phi)$-dependent correction. Relative to the efficiency at low instantaneous luminosity ($\mathcal{L}_{\text{inst}} < 1.5 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$) the efficiency decreases with increasing $\mathcal{L}_{\text{inst}}$, declining by $\approx 10\%$ when $\mathcal{L}_{\text{inst}} = 2.5 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. The ratio of those efficiencies in data.

Please cite this article as: V.M. Abazov, et al., Nuclear Instruments & Methods in Physics Research A (2014), http://dx.doi.org/10.1016/j.nima.2014.03.013
and MC simulation has no dependence on the instantaneous luminosity.

For transverse momenta of 40 GeV after preselection, loose electrons have a total identification efficiency of 85% (95%) with a fake rate from misidentified jets of 5% (3%) in the CC (EC). Tight electrons at the same transverse momentum have an identification efficiency of 72% (53%) with a misidentification rate of 0.2% (0.1%) in the CC (EC).

8. Efficiencies of photon identification

8.1. Photon identification efficiencies

There are two categories of variables for photon identification. Variables based mainly on shower information are used to reject misidentified jets. Tracking-based variables are used to separate electrons from photons. There are two main mechanisms by which

Fig. 14. Photon identification efficiencies for identification variables mainly based on calorimeter information as derived from Z→ee decays. Displayed are data and MC predictions and their ratio as a function of $E_T$ (a) (b), $\eta$ (c) (d) and $\phi$ (e) (f) for CC and EC photons.
photons can appear as electrons. First, if the photon has converted into an electron–positron pair in the inner tracking system, creating a reconstructed track. The probability for conversion is ≈6%, and we do not reconstruct these converted photons explicitly. Second, if a track from particles of the underlying event is pointing to the EM cluster. In both cases, the matched track information for a photon will tend to be different from a real electron.

Because a large sample of pure photons is not available in data, $Z\rightarrow\text{ee}$ events are used to derive efficiencies for variables based mainly on the calorimeter information. For tracking-based variables, the efficiencies are measured from reconstructed radiated photons in $Z\rightarrow\gamma\ell\ell (\ell = e, \mu)$ events in data and MC. In both cases, differences between data and MC event samples are analyzed to correct the efficiency in simulation.

Due to different needs in various physics analyses, various sets of photon identification requirements are developed. We provide here photon identification efficiencies for two different sets of photon identification requirements.

The first set of photon identification requirements considered is used in the search for $H\rightarrow\gamma\gamma$ decays [21,22]. The signal is dominated by high-$p_T$ CC photons, and the analysis maximizes the photon signal acceptance. Photon candidates in the CC are required to fulfill the preselection requirements as described in Section 7.3. In addition, it is required that

- $\sum p_T^{trk} < 2 \text{ GeV}$;
- $\sigma_{\rho}^2 < 18 \text{ cm}^2$;
- Output of $\gamma \text{ NN}4 > 0.1$.

In addition the following requirements are placed on track-based variables:

- $P_{\chi^{2} \text{ spatial}} = -1$;
- $B_{\text{hor}} < 0.9$.

The measured identification efficiencies using the non-track-based variables in this selection are presented in Fig. 14 (left column) as a function of $E_T$, $\eta$ and $\phi$. The differences between data and MC are at the percent level and are constant in the presented distributions. Therefore, a single correction factor is applied to MC photon simulation.

The second set of photon identification requirements presented here is used for measurements of electroweak cross-sections, such as the measurement of the $W\gamma$ production cross-section [23]. Here, the photons tend to have low $E_T$ and a high background rejection is required. The EC photons used are required to fulfill the preselection criteria of Section 7.3 and to satisfy the following requirements:

- $\sum p_T^{trk} < 1.5 \text{ GeV}$
- $\sigma_{\rho}^2 \lesssim (7.3 \cdot \eta^2 - 35.9 \cdot |\eta| + 45.7) \text{ cm}^2$
- Output of $\gamma \text{ NN}4 > 0.05$

In addition, a track-based requirement $P_{\chi^{2} \text{ spatial}} < 0.001$ is applied.

Fig. 14 (right column) shows the identification efficiencies using the nontrack-based variables in this selection for data and MC. The difference between data and simulation depends on $\eta$. To take this into account, the correction to MC simulation is parametrized as a function of $\eta$.

For both CC and EC photons, exploring the track-based variables as presented in this section, the efficiencies to identify a photon candidate are measured. The $Z\rightarrow\gamma\ell\ell (\ell = e, \mu)$ data and MC comparison justifies that no further corrections to the photon simulation are required. The photon identification efficiency for these track-based variables is 92% (95%) in CC (EC) for an electron-positron pair in the inner tracking system, mainly due to increasing amount of material in front of the calorimeter (from about 3.4 to 4.5 cm). It becomes larger towards high $p_T$ events.

8.2. Vertex pointing

In most physics applications, it is important to know from which $p\bar{p}$ collision vertex the photon originated. Since unconverted photons leave no track, the default reconstruction vertexing algorithm does not provide high probability to find the correct photon origin if there is no high-$p_T$ track in the event. For events without leptons and with energetic photons, the most probable photon production vertex can be reconstructed due to the presence of the underlying event coming from interactions of spectator quarks, and corresponds to the vertex with highest track multiplicity [21,22,24]. In such cases, verifying that the true production vertex is found in data is important, especially in the high-instantaneous luminosity regime with many $p\bar{p}$ collision vertices.

To find the position of the photon origin along the beam line ($z$-axis) between −60 cm and 60 cm in the CC, the (x, y, z)-coordinates of the EM cluster in the EM1–EM4 layers and the position of the CPS cluster are used. Therefore, 5 points are used with radii from about 73 cm to 99 cm. Using a linear extrapolation to the $z$-axis, the most probable position of the photon origin vertex is obtained. Typical resolution of the algorithm varies between 3 and 4.5 cm. It becomes larger towards high $p_T$ mainly due to increasing amount of material in front of the calorimeter (from about 3.4 to 5 X0). The resolution has been tested in data using $Z\rightarrow\ell\ell'\gamma$ events, in which the “true” vertex ($z_{\text{true}}$) is reconstructed using the

![Fig. 15. Vertex pointing resolution in two rapidity bins: $|\eta| < 0.4$ (a), and $0.8 < |\eta| < 1.1$ (b).](image-url)
two lepton (e or μ) tracks and the photon vertex (z_{point}) is obtained using the procedure described above. The distribution of events for Δz = z_{true} - z_{point} is shown in Fig. 15. The resolution is 2.4 cm for |η| < 0.4, and 4.3 cm for 0.8 < |η| < 1.1.

The resolution in MC simulation is a factor of 1.4–1.5 better than in data events. To calibrate the pointing resolution, a smearing procedure as a function of photon pseudorapidity is applied. The Δz resolution is almost independent of photon p_T.

9. Conclusions

The precise and efficient reconstruction and identification of electrons and photons by the D0 experiment at the Tevatron pp collider at Fermilab is essential for a broad spectrum of physics analyses, including high precision SM measurements and searches for new phenomena.

In this paper, the electron and photon reconstruction and identification algorithms have been presented using data collected by the D0 detector in pp collisions at a center-of-mass energy of 1.96 TeV. The separation between electron or photon signal and multijet background is considerably improved using multivariate analysis techniques. A likelihood method for electron identification, a neural network method for electrons and photons, and a Boosted Decision Tree for electrons have been developed. An energy calibration dependent on the azimuthal angle of the EM cluster, the shower shape and the pseudorapidity has been performed separately for data and MC, leading to significant improvements in resolution and uniformity and resulting in a good agreement between data and MC.

Single electrons are triggered with an efficiency ≈ 100% for transverse momenta above 30 GeV in the fiducial regions of the calorimeter up to |η| < 2.5. For transverse momenta of E_T = 40 GeV, in general at D0 electrons can be identified with a total identification efficiency of 90% (95%) with the rate at which jets are misidentified as electrons being 5% (3%) in the CC (EC). Photons in the CC and EC regions can typically be identified with efficiencies varying between 69% and 84% with the rate at which electrons or jets are misidentified as photons being 2%–10%.

The agreement of electron and photon identification efficiencies between data and MC in fiducial regions of the detector is reasonable, with deviations only at the percent level. Larger correction factors are necessary in non-fiducial regions close to the boundaries of the calorimeter modules. These correction factors have been applied to MC events as a function of kinematic variables resulting in considerable improvements of the simulation.

Acknowledgments

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); MON, NRC Ki and RFBR (Russia); CNPq, FAPEM, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); NRF (Korea); FOM (The Netherlands); STFC and the Royal Society (United Kingdom); MSMT and GACR (Czech Republic); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); and CAS and CNSF (China).

References

[14] L. Breiman, et al., Classification and Regression Trees (Wadsworth, Stanford, 1984). (Version v04-01-00 of TMVA is used.)