3
Object Identification

At each bunch crossing at the DØ interaction point, collider data or the events recorded by the DØ sub-detectors consist mainly of digitized electronic signals resulting from the collected charge of calorimeter cells, light yield of scintillators and pattern of hits in the tracking or muon system. The DØ Online Reconstruction Program (DØ RECO) takes the trigger and detector signals from collider data or from simulations and creates the reconstructed particles that are used for the analysis. This program associates sub-detector signals with specific positions in the detector, creating energy clusters in the calorimeter, and tracks in the CFT, SMT and muon system. Different particles generated in the collisions are identified and their kinematic properties are measured. Additional quantities used for the identification of different types of particles, like the missing transverse energy or the isolation of particles relative to other objects in the calorimeter or in the tracker, are also evaluated.

The analysis presented in this thesis relies mostly on the efficient reconstruction of leptons - electrons, muons and taus. It also requires the reconstruction of missing transverse energy due to the neutrinos escaping the detector. The following sections will summarize briefly the identification of particles, commonly known as "physics objects" like leptons, jets and missing transverse energy.

3.1 The Reconstruction Program

The DØ offline reconstruction program [24] outputs data for each event to a small event thumbnail (TMB) that is converted to the Common Analysis Format (CAF) and is stored for later analysis. These reduced data files contain all physics objects and discard most of the original data from the detector facilitating rapid access to the entire data set. In the first step of this reconstruction chain, the program unpacks (or unfolds) the digitized signals from the detector, associates the electronic channels with physical detector elements and applies detector specific calibration constants. In the second step, tracks are reconstructed using the hits in the SMT and CFT detectors. These tracks are stored and used later as input to the third level in the reconstruction. In the third step, primary vertices (associated with the $p\bar{p}$ collision in each bunch crossing) and secondary vertices (from the decay of long-lived particles produced in the hard interaction) are identified and stored. The last and final step uses the information of dedicated sub-detectors to reconstruct physics objects such as electrons, photons, muons, taus and jet candidates. The object identification algorithms used in the reconstruction are explained in the following text.

3.2 Track Reconstruction

Tracks are not physics objects but they are important for measuring momentum of an object and also in object identification. When traversing through the silicon strips of the tracking system, the charge particles interact with matter by ionization and leave hits in the tracker. These hits are spatially clustered. Two algorithms are run sequentially to reconstruct helical tracks from clusters of signal hits from the central tracking detectors: the Histogram Track Finder (HTF) [25] and the Alternative Algorithm (AA) [26]. The AA algorithm shows generally a better performance and lower fake rate for low p_{T} and high impact parameters. The HTF algorithm, on the other side, is more efficient for high p_{T} tracks.

3.2.1 The AA algorithm

The Alternative Algorithm (AA) forms an initial "track hypothesis" by using combinations of hits in the SMT barrel or disk. Starting from any combination of three hits in SMT barrel or disk layers, the algorithm extrapolates the sequence of hits moving outwards to the next SMT or CFT layer. Hits found within the expected region are associated to the track hypothesis if they match certain conditions such as the axial angle between the different hits or the curvature. The quality of the fit of the track candidate has to satisfy a given pearson chi-square (χ^2) threshold. In case of multiple tracks passing, the hypothesis is split and a new track candidate is formed for each valid hit combination. A "miss" is recorded when no hit is found in a layer. The track hypothesis ends at the outer radius of the detector, or after three misses. Tracks with less than three hits in the SMT are reconstructed by using the primary vertex candidate which was found using reconstructed tracks of at least three hits in the SMT. Any three hits in the CFT are then required to fulfill the additional condition of the track hypothesis passing near a reconstructed primary vertex.

3.2.2 The HTF algorithm

The trajectory of a particle in a magnetic field can be characterized in a plane perpendicular to the direction of the field by three parameters: ρ, the radius of the curvature, d_0, the distance of closest approach (DCA) with respect to $(0,0)$, and ϕ, the azimuthal angle of the track at the point of closest approach to $(0,0)$. For track candidates with small impact parameters, every pair of hits in x and y directions that belongs to the same track corresponds to a single point in the $\rho - \phi$ plane. Filling each pair of hits into the two-dimensional $\rho - \phi$ histogram, a peak in the histogram would correspond to a track candidate. The algorithm is based on the principle that all pairs of hits on one track trajectory point to the same value of (ρ, ϕ). Points from different tracks result in a randomly distributed background in the parameter space. The HTF algorithm improves this method using the Hough transformation [27]. In this transformation each hit corresponds to a line in the parameter space. An intersection of different lines corresponds to track candidates. The final track list is sorted by the number of hits, fewest misses and lowest χ^2 value. Tracks with large values of χ^2 are removed.

3.3 Primary Vertex Reconstruction

The primary vertex (PV) represents the point of the $p\bar{p}$ collision. A precise reconstruction of the PV is essential to distinguish the objects originating from the hard interaction w.r.t overlapping events, to calculate important kinematic variables and to distinguish tracks from the primary and a possible secondary vertex (SV). Proper identification and precise measurement of secondary vertices is particularly important for identifying the flavor of jets and selecting the contribution of jets originating from heavy quarks. The DØ vertex fitting algorithm used for the vertex reconstruction can be distinguished into three basic steps.

At first, tracks identified with $p_{\mathrm{T}} > 0.5\ GeV/c$ and at least 2 hits in the fiducial SMT region are clustered along the beam axis. This allows the separation of possible additional $p\bar{p}$ interactions taking place during the bunch crossing. A two step approach is used for each of the track clusters. First an estimation of position and width of the beam is performed by fitting all the tracks in the cluster into a common vertex using a Kalman Filter based vertex fitting algorithm [28]. Then a preselection on the tracks corresponding to each cluster is performed based on their distance of closest approach to the beam spot. Subsequently the Adaptive Vertex Fitting algorithm is applied [29]. This technique is an iterative Kalman Filter that re-weights track errors according to their χ^2 contribution to the vertex by means of the Fermi-like function given by:

$$w_i = \frac{1}{1 + e^{(\chi_i^2 - \chi_{cutoff}^2)/2T}} \tag{3.1}$$

Here, w_i is the weight given to ith track, χ_i^2 is the χ^2 contribution of the i-th track to the primary vertex, χ_{cutoff}^2 is the distance where the function drops to 0.5, and T is a parameter controlling the sharpness of the function. The weight is re-computed with respect to the newly fitted vertex at each iteration until convergence is achieved.

Finally, the primary vertices are selected with the aim to identify the ones corresponding to the largest momentum transfer (high Q^2 processes). The remaining primary vertices are assigned to additional soft interactions which may take place in the same bunch crossing. Using the track p_{T} the probability for each track to have originated from a soft interaction is calculated. The minimum-bias probabilities for all tracks coming from a given vertex are multiplied to determine the soft probability for that collision vertex. The collision vertex with the lowest probability of being a soft interaction is taken as the primary vertex.

3.4 Reconstruction of Electrons

Electrons are characterized by an isolated track in the central tracking system leading to a narrow energy deposit in the calorimeter that starts in the preshower detectors and ends in the first hadronic layer. Whereas photons are characterized by a narrow energy deposit in the calorimeter only and no track in the tracking system.

The process of energy deposition for electron starts with the bremsstrahlung of a photon, while for photons it starts with the production of an e^+e^- pair. The electromagnetic (EM) particles create, at each step, more particles each carrying a fraction of the initial energy.

These particles ionize the sensitive medium (LAr) of the calorimeter, creating a signal which is proportional to the energy of the initial particle. The shower reaches its maximum when the average energy per particle becomes low enough so as no further multiplication of particles takes place. After this, the shower dies slowly through ionization for electrons or by Compton scattering for photons.

The identification of EM (either electron or photons) particles start with the reconstruction of these EM showers which in turn starts with the clustering of EM towers. The EM tower is defined as the first five layers of $0.1 \times 0.1(\eta \times \phi)$ tower of calorimeter cells (i.e, the four EM and first hadronic layer). The clustering is performed using the *Simple Cone Algorithm* [30]. The algorithm searches for EM towers exceeding a minimum "transverse" energy ($E_T = E \sin\theta$) of 0.5 GeV and uses them as seeds. Around each seed, a cone of towers with radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$ is formed. The cluster must have $E_T > 1.5$ GeV.

3.4.1 Electron Identification Variables

Although the electron identification algorithms have high efficiency of reconstructing true electron candidates (real electron), there are several sources which can lead to a fake identification of an electron candidate (fake electrons). We define several variables that are used for electron identification and are used to distinguish the real electron candidates from the fake electrons. These are discussed below:

- When summing the energy deposited in EM and hadronic layers, clusters with 90% of the energy deposited in the EM calorimeter ($EMfraction > 0.9$) are considered. These clusters are then checked for isolation in the calorimeter. The isolation of an EM cluster is defined as:

$$isolation = \frac{E_{total} - E_{em}}{E_{em}} \tag{3.2}$$

  where E_{total} is the total energy in the EM and hadronic towers in a cone of radius $\Delta R = 0.4$ and E_{em} is the total energy in the EM layers in a cone of radius $\Delta R = 0.2$ around the seed.

- Clusters are then matched with the tracks with $p_{\mathrm{T}} > 1.5$ GeV/c within a window of 0.05×0.05 in the $(\eta - \phi)$ plane, around the centroid of the EM cluster [32]. The track-matching is done by calculating χ^2_{EMtrk}, which is defined from the difference in ϕ and z in EM3 layer and the squared significance of the resolution of the transverse energy of the cluster as measured in the calorimeter, over the transverse momentum of the track (E_T/p_{T}):

$$\chi^2_{EMtrk} = \left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta z}{\sigma_z}\right)^2 + \left(\frac{E_T/p_T - 1}{\sigma_{E_T/p_T}}\right)^2 = \chi^2_{spatial} + \left(\frac{E_T/p_{\mathrm{T}} - 1}{\sigma_{E_T/p_{\mathrm{T}}}}\right)^2 \tag{3.3}$$

Here, $\sigma_{\phi/z}$ is the standard deviation of the distribution of variable ϕ/z. A probability is then calculated that the track with given χ^2_{EMtrk} matches the cluster. $\chi^2_{spatial}$ probability is also computed for the best track match using the spatial distance only.

- A 7×7 covariance matrix ($HM \times 7$) is constructed from 7 correlated shower shape observables. It provides a comparison between the shower shape of the cluster and the expected shower shape for an electron. The seven variables are:
  - Fraction of energy in each of the four EM layers,
  - Width of shower in the third EM layer,
  - Logarithm of the total EM energy,
  - z coordinate of the PV.

  The shower shape is determined using a sample (N) of MC electrons to construct the covariance matrix. Matrix elements of the covariance matrix are defined as:

$$M_{ij} = 1/N \sum_{n=1}^{N} (x_i^n - \bar{x}_i)(x_j^n - \bar{x}_j) \quad i,j = 1...7, \tag{3.4}$$

  here $(x_i, x_j)^n$ is the value of the variable i/j for particle n and $(\bar{x}_i, \bar{x}_j)$ are their corresponding averages. The inverse of the 7×7 covariance matrix is called H-Matrix: $H = M^{-1}$. The H matrix is then used to construct the χ^2, defined as:

$$\chi^2_{hmx7} = \sum_{ij}^{N} (x_i - \bar{x}_i) H_{ij} (x_j - \bar{x}_j) \tag{3.5}$$

  The H-Matrix thus constructed is called Central region H-Matrix. For real electrons this χ^2_{hmx7} variable should have small values. Similarly, a 8×8 covariance matrix ($HM \times 8$) is also constructed with an addition of one more variable (transverse shower shape) to the seven correlated shower shape variables listed above. This is also called as Forward region H-Matrix.

- An electron likelihood (*Lhood*8) variable is constructed to discriminate between EM clusters from real electrons (signal) and clusters from fake objects, such as jet faking an electron (background) [31]. Eight variables are used as input for the electron likelihood:
  - The electromagnetic fraction $EMfraction$,
  - Two covariance matrices using a different number of variables,
  - E_T/p_{T},
  - The spatial track match probability.
  - The z position of the closest approach of the matched track to the primary vertex,
  - The number of tracks inside a cone of radius $\Delta R = 0.05$,

– The sum of all track momenta within a cone of radius $\Delta R = 0.4$, excluding the candidate track itself.

Fig 3.1 shows the performance of the electron likelihood for the endcaps and the central part of the calorimeter. Although the efficiencies in both parts are comparable the contribution of instrumental background in the endcaps of the detector is higher than in the central part. Fig 3.2 shows the signal efficiency vs. the background rejection of both, the endcap and the central part of the detector. All these criteria described above are grouped together to define an operating point for electron and photon identification.

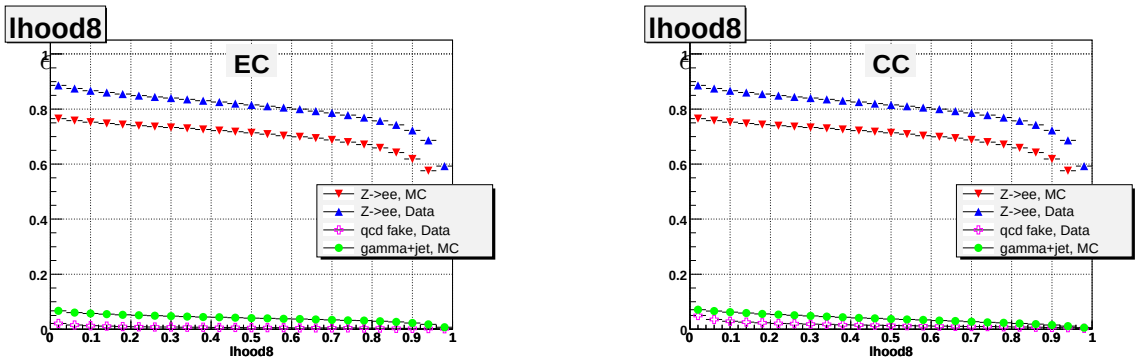

Figure 3.1: The performance of the electron likelihood for $Z \rightarrow ee$ events in collision data and MC and for background processes of multijet and $\gamma + jet$ production [62]. The left plot shows the efficiency for the endcaps of the calorimeter and the right plot shows the efficiency for the central part.

- An Artificial Neural Network (ANN) that combines variables for electron and photon selections are used to select high p_{T} isolated electrons and photons. The ANN is trained against multijet background. Four variants of ANN are used for electron and photon identifications.

  – $NNout7\_cc$ - A 7 input variable based ANN used for identifying central electrons.

  – $NNout3\_ec$ - A 3 input variable based ANN used for identifying end-cap electrons.

  – $NNout5\_gamCC$ - An ANN with 5 input variables used for identifying central photons and trained against multijet diphoton background.

  – $NNout4\_gamEC()$ - An ANN with 4 input variables used for isolating forward photons.

The measured electron energy is corrected by selecting $Z \rightarrow ee$ events and comparing the dielectron mass peak with the experimental results. The dielectron mass peak obtained using uncorrected electron energies is then scaled up to obtain corrected energies that reproduce the expected $Z \rightarrow ee$ resonance.

The efficiency for an electron candidate to pass a given criteria is measured in different parts. Firstly there is an efficiency to be reconstructed as an EM cluster and pass a given identification criteria, secondly the efficiency of the electron to be track-matched is referred

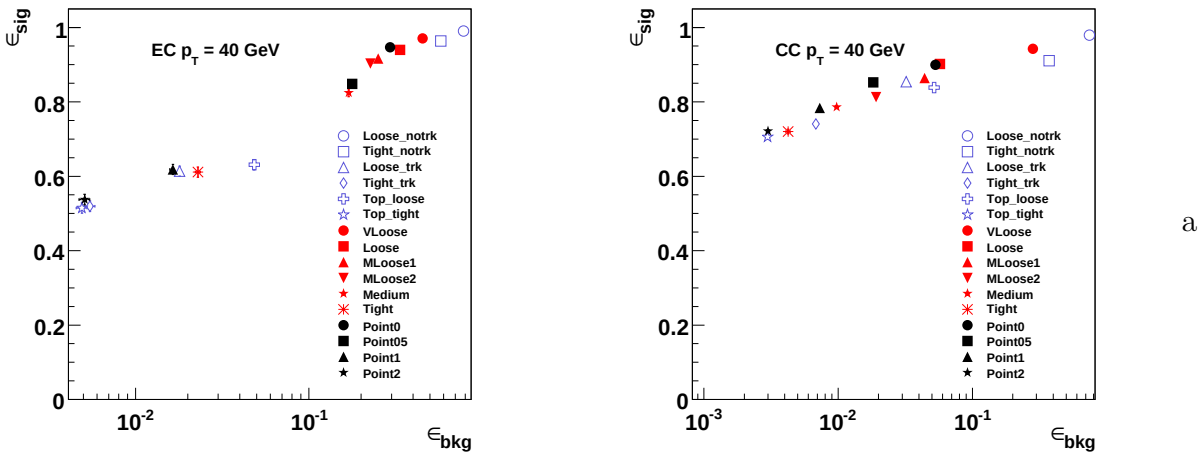

Figure 3.2: The rejection of various e-ID definitions for a 40 GeV/c electron. The comparison is done using Run IIb data. The electron ID used in the present analysis is `Point05` for CC electrons and `Point1` for EC electrons [61]. The left plot shows the rejection for the endcaps of the calorimeter and the right plot the rejection for the central part.

to as the "track match efficiency", and finally the probability for a track matched electron to pass the electron likelihood is called the "likelihood efficiency". Electron reconstruction efficiencies are measured in $Z \rightarrow ee$ collision data and MC events, whereas Electron likelihood and track match efficiencies are measured in $Z \rightarrow ee$ collision data and MC events with respect to reconstructed electrons. The efficiencies are parametrized as a function of η, ϕ and Luminosity.

3.5 Muon Reconstruction

Muons (μ) deposit only a small amount of ionization energy in the calorimeter and travel through most of the inner detector components up to the muon system, typically the outer most radial layers of a detector. The reconstruction of muons uses information from three independent sub-systems:

- hits in the three layers of the muon system.
- tracks in the central tracking system.
- signature as a minimum ionizing particle (MIP) in the calorimeter that relies on the "Muon Tracking in the Calorimeter (MTC)" algorithm [33].

Reconstructed muon candidates are classified using two parameters: muon type and muon quality. The type of muon is given by the parameter N_{seg}. A positive value of N_{seg} indicates that the muon reconstructed in the muon system ("local muon") was matched to a track in the central tracking system. A negative value of N_{seg} tells that the local muon could not be matched to a central track. The absolute value $|N_{seg}| = 1$, 2 or 3 respectively indicates that

the local muon is made up of A-layer only hits, B or C-layer only hits (outside the toroid), or both A- and B- or C-layers hits. The second parameter used to classify muons is the quality. The muon quality can be "Loose", "Medium" or "Tight". The muon quality used in the analysis 5.5.1 is based on:

- The number of hits in the muon drift and scintillation chambers.
- Number of segments reconstructed in the muon system N_{seg}.
- Timing of the scintillator hits.
- Distance of closest approach (DCA) in the $x - y$ plane w.r.t PV.
- χ^2_{trk} per degrees of freedom of the central track.
- χ^2_{global} per degrees of freedom for the match between central track and local muon track.
- Calorimeter isolation I_{cal}, computed as the energy deposited in a hollow cone of radius R centered around the muon in the calorimeter i.e, $\Sigma^{R>0.4}_{R>0.1} E^i_T$.
- Track isolation I_{trk}, computed as the sum of the tracks transverse momenta in a hollow cone of radius R centered around the muon i.e. $\Sigma^{R>0.4}_{R>0.1} p^i_T$.

In order to control the quality of central tracks matched to muons, four track quality definitions, *trackloose*, *tracknewmedium*, *trackmedium* and *tracktight* which rely on:

- The number of hits in SMT or CFT system.
- χ^2 per degrees of freedom of the central track fit.
- Distance of closest approach (DCA) with respect to the beam spot location in the (x, y) plane.

Timing requirements are placed on hits in scintillator pads in the muon spectrometer to reject background due to cosmic rays. As the arrival of cosmics is uncorrelated with the $p\bar{p}$ collisions, they typically produce "out of time" hits, while muons from the $p\bar{p}$ collisions produce hit-times close to the crossing time expected from these collisions. The distance of closest approach (DCA) of the muon to the PV is also effective in rejecting cosmic-ray events.

Isolation variables $TrackHalo$ and $CalorimterHalo$ are used to separate signal muons from those produced in heavy flavor quark decays. Since the signal muons will result in an isolated track originating from the PV whereas the muons from heavy flavor quark decays tend to be embedded inside a jet and hence are non-isolated. These isolation variables are defined in terms of any tracks that accompany the muon track or calorimeter energy deposited along the direction of the muon momentum.

- $TrackHalo$ is the scalar sum of the p_{T} of all tracks within a cone of radius $\Delta R < 0.5$ excluding the p_{T} of the muon.

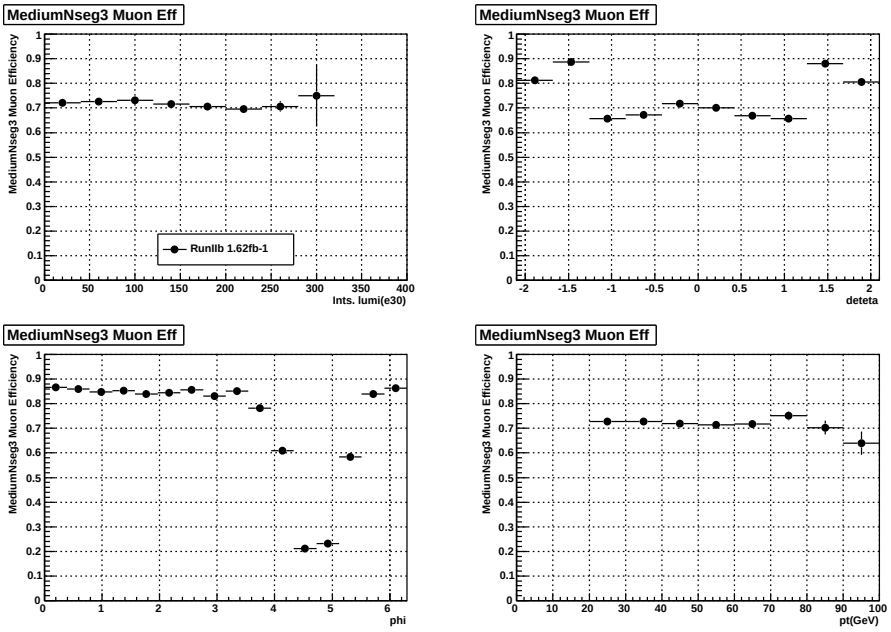

Figure 3.3: The reconstruction efficiency of muons in data for, "medium" muon quality in Run IIb data as a function of *luminosity*, ϕ_{det}, ϕ and p_T.

- *CalorimeterHalo* is the scalar sum of transverse calorimeter energy clusters in hollow cone around the muon with $0.1 < \Delta R < 0.4$ using all cells in the EM calorimeter and first layer of the Hadronic calorimeter.

The efficiency for a muon candidate to pass the muon reconstruction criteria is quite different in collision data and MC. The efficiency to reconstruct muons in collision data as a function of *Luminosity*, ϕ_{Det}, η_{Det} and p_{T} are shown for a *medium* quality muon in Fig. 3.3. To correct for this difference, scale factors are derived and applied to each MC event. High p_{T} muons from $Z \to \mu\mu$ events are used to parametrize the efficiencies in track and calorimeter isolation as a function of the instantaneous luminosity, η, ϕ and p_{T} of the muon. Additionally the p_{T} of the muons is smeared in the MC events using a correction to track curvature. This correction is derived by comparing $Z \to \mu\mu$ and $J/\psi \to \mu\mu$ events in data and MC .

3.6 Tau Reconstruction

In contrast to electrons and muons, tau-leptons (τ) decay very rapidly. Their lifetime is about 290×10^{-15} s. Therefore the decay takes place within the beam pipe before reaching any DØ subdetector. Hence they have to be identified using their decay products. The total branching ratio of leptonic and hadronic τ decays are about 35% and 65%, respectively. The products of hadronic τ decays contain mostly neutral and charged pions, and therefore the decay products mostly appear to be narrow jets with low track multiplicities, which require a dedicated τ identification algorithm. Electron and muon identification methods as described above are used to detect the leptonic τ decay products. The neutral pions are detected

with a significant energy deposited in the EM layers of the calorimeter. The hadronic τ reconstruction uses the following information:

- The algorithm begins by finding calorimeter clusters, using a simple cone algorithm with a cone size of $\Delta R = 0.3$.
- EM subclusters are then found by using a nearest neighbor algorithm in the third layer of the EM calorimeter. If a sub-cluster is found, EM cells in the other layers and preshower hits are added to it. The subclusters are designed to identify π^0 particles originating, for example, from $\tau^{\pm} \rightarrow \rho^{\pm}\nu \rightarrow \pi^{\pm}\pi^0\nu$ decays.
- All tracks with $p_{\mathrm{T}} > 1.5\ GeV/c$ within a cone size of radius $\Delta R = 0.5$ around the cluster center are ordered in terms of their p_{T}. The highest p_{T} track is associated with the cluster and up to two more tracks are associated if they are within 2 cm of the first track at the production vertex. A second track is added if the mass of the first and second track is less than 1.1 GeV/c^2 and a third is added if the mass of the first two tracks is less than 1.7 GeV/c^2. A detailed description of the τ reconstruction algorithm can be found in Ref. [36].

The τ candidates can be split into three types, defined by their detector signature:

τ-type 1: Calorimeter cluster with one track associated and no EM subcluster, corresponding mainly to the decay $\tau^{\pm} \rightarrow \pi^{\pm}\nu$.

τ-type 2: Calorimeter cluster with one track associated and at least one EM subcluster, corresponding mostly to the decay $\tau^{\pm} \rightarrow \pi^{\pm}\pi^0\nu$.

τ-type 3: Calorimeter Cluster, with two or three associated tracks, with or without EM subcluster. This corresponds mostly to the decay $\tau^{\pm} \rightarrow \pi^{\pm}\pi^{\pm}\pi^{\mp}(\pi^0)\nu$.

Jets can be easily misidentified as tau candidates due to similar detector signatures. To efficiently separate such events from the real tau events, neural networks (ANN) are used [36]. These ANNs are trained using a set of variables discriminating between jets and taus. The output of the ANN will be a value in the range of $[0, 1]$. Lower values correspond to background-like or fake tau events and higher values of the ANN output correspond to events likely to be a real τ candidate event. The ANN combines the power of each single variable and hence has a better background rejection for a given signal efficiency. The variables used in the ANN are [36]:

- $(E^{EM1} + E^{EM2})/E^{\tau}$, where E^{EM1} and E^{EM2} correspond to the energies deposited in the first and second layers of the EM calorimeter and E^{τ} is the energy of the cluster, within a cone of radius $\Delta R = 0.5$. Since the τ-type 1 deposits no energy in the EM calorimeter, hence this variable offers the best discrimination power between τ-type 1 and other τ types.
- $\Sigma p_{\mathrm{T}}^{trk}/\Sigma p_{\mathrm{T}}^{\tau_{trk}}$, where p_{T}^{trk} is the p_{T} of a track within a cone of radius $\Delta R < 0.5$. $p_{\mathrm{T}}^{\tau_{trk}}$ corresponds to the sum of the transverse momenta of the tracks, associated with the τ candidate.

  - Fine hadronic fraction is defined as the fraction of total energy deposited in the hadronic calorimeter. It is used for all τ *types*.

- $E_T^\tau / \left(E_T^\tau + \Sigma p_{\mathrm{T}}^{trk}\right)$, energy of the cluster divided by the sum of the cluster energy and the total track momentum.
- $\sqrt{(\Delta\phi/sin\theta)^2 + (\Delta\eta)^2}/\pi$, where the differences are between the vector sum of τ-track directions and the vector sum of the EM cluster. Used for τ-types 2 and 3.
- Transverse energy of the leading EM subcluster divided by the transverse energy in the layer 3 of the calorimeter in a cone of radius $\Delta R < 0.5$. Only used for τ-type 2.
- Profile, defined as: $(E_T^1 + E_T^2)/E_T^\tau$ where E_T^1 and E_T^2 are the transverse energies of the two most energetic calorimeter towers. It is used for all τ types.
- E_T^{em}/E_T^τ, where E_T^{em} corresponds to the transverse energy of the EM cluster. This variable is used for τ type I and II only.
- Transverse energy of the leading τ track divided by the transverse energy of the τ. This variable is considered in all three τ types.
- Calorimeter isolation, $\mathcal{I}_{cal} = (E_{Tot}^\tau - E_{core}^\tau)/(E_{core}^\tau)$, here E_{Tot}^τ is the cluster energy in the $\Delta R < 0.5$ cone and E_{core}^τ the energy in the $\Delta R < 0.3$ cone. A modified isolation parameter is defined: $\mathcal{I}_{cal}^2 = \mathcal{I}_{cal}/(1.5 \cdot |\eta_{det}| - 0.5)$ for the region $\eta_{det} > 1$. This variable is used for all three τ types.
- RMS of the shower, $\sqrt{\Sigma_{i=1}^n \left[(\Delta\phi_i)^2 + (\Delta\eta_i)^2\right] E_{T_i}/E_T}$, represents the width of the calorimeter cluster of the τ. Again the variable is used for all three τ types.

Fig 3.4 shows the efficiency of the τ reconstruction algorithm as a function of the visible transverse momentum of a τ lepton p_{T}^{vis}, which excludes neutrinos, as a function of η_{det}^{vis}. For a tau efficiency of 95% the typical rejection efficiency is about $\sim 85\%$ to $\sim 90\%$, depending on the τ type.

3.7 Jet Reconstruction

Hadrons deposit large amounts of energy in small angular regions in the hadronic calorimeter. These deposits of energy are collimated in the direction of the parent parton originating from the hard interaction and are identified as "Jets". These partonic jets are a result of the initial state radiation (ISR), final state radiation (FSR) or beam remnants (beam jets). Jet algorithms enable jet identification and categorizes them according to their quality. These algorithms also study the jet energy, their resolution and corrections to the jet energies such that one obtains the actual initial parton energy. A schematic of the jet development is shown in Fig. 3.5. The details of these procedures are described below.

The iterative cone algorithm [37] used for jet identification at DØ is based on the idea of associating all particles within a cone of radius ΔR in $\eta \times \phi$ plane. The cone algorithm consists of three stages as described below:

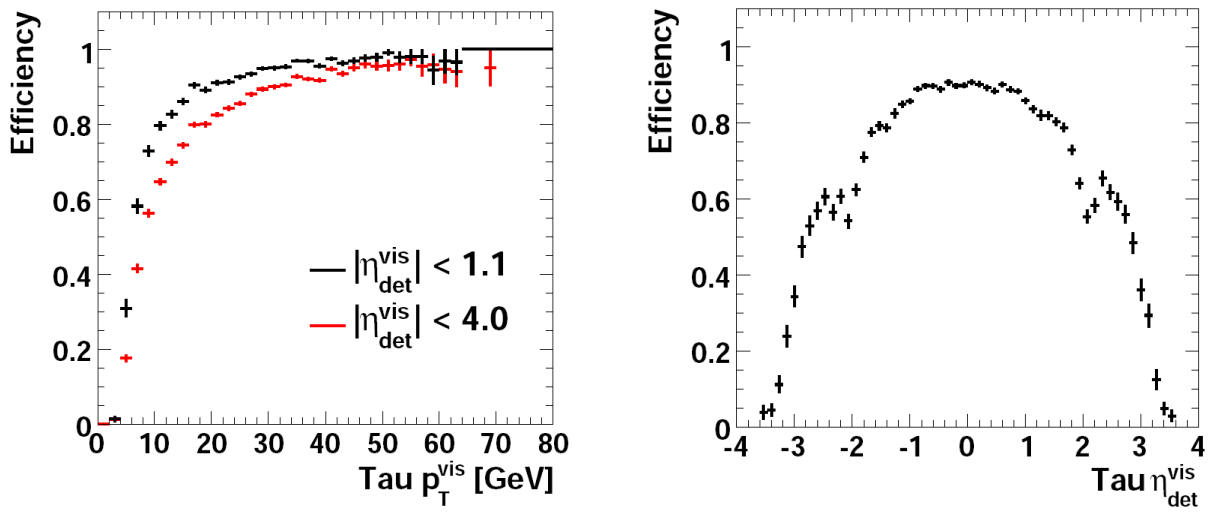

Figure 3.4: Reconstruction efficiency of hadronically decaying tau leptons as a function of τ p_T^{vis} and η_{det}.

- Pre-Clustering: In the first step "Seeds" are formed using the calorimeter towers with $E_T > 0.5\ GeV$. After being sorted in E_T these towers are clustered to higher-order E_T towers using a cone of radius $\Delta R = 0.3$ to form a "pre-cluster".

- Cone Clustering: These pre-clusters with $E_T > 1\ GeV$ and more than one tower are considered as input to the cone algorithm. The pre-cluster with the highest E_T is used as seed for the formation of "proto-jets". All pre-clusters within a cone of radius $\Delta R = 0.5$ are assigned iteratively to the "proto-jet" until the E_T-weighted cone center is found. In order to avoid sensitivity to soft radiation, stable cones around midpoints of any combination of two "proto-jets" are searched for. Finally a list of "proto-jets" from pre-clusters and midpoints is considered in the last step of merging and splitting. So far the condition that each pre-cluster is only present in one jet is not fulfilled. Therefore in the next step all pairs of "proto-jets" within a distance larger than the cone size but smaller than twice the cone size are considered. These cones get either split or merged, according to their overlapping energies. If the overlap exceeds 50% of the lower-energy cone of the two jets the jets are merged. If the overlap is smaller the pre-clusters get assigned to the cone with the cone axis closer in $\eta \times \phi$. All jets with $E_T > 6\ GeV$ are kept for the next step of jet identification.

- To reduce the calorimeter noise, an algorithm called "T42" is used [38]. This algorithm performs a finer treatment of calorimeter noise to improve the reconstruction of the calorimeter objects. It considers an isolated cell as a noise cell and discards if it is not "signal-like". T42 algorithm has been found to reduce the number of fake jets clustered on noise by about a factor of two.

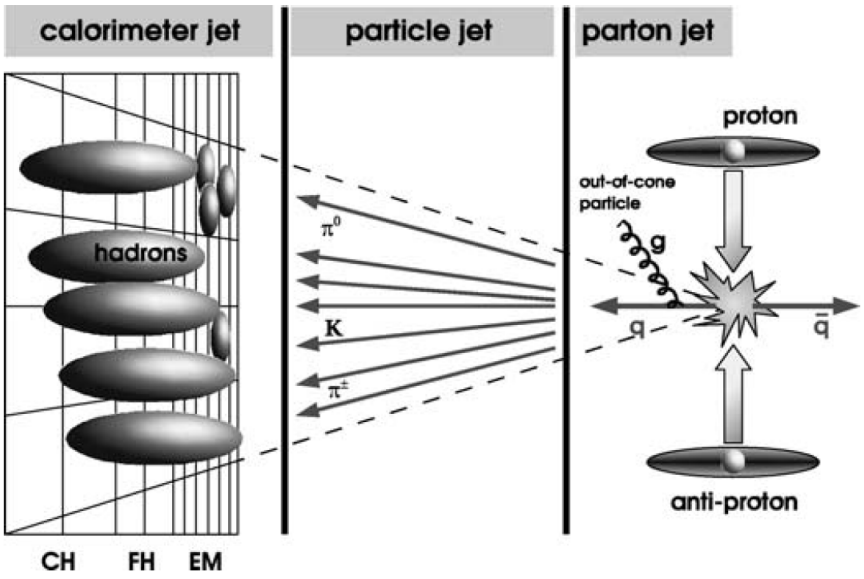

Figure 3.5: Schematic view of the jet development illustrating the idea of jet energy scale corrections, taking energy measurements on the calorimeter level to the particle or the parton level.

3.7.1 Jet Identification

Once the jets are reconstructed, some quality cuts are applied to distinguish the real jet candidates from fake jets.

- $0.5 < EMfraction < 0.9$: To ensure that they are well separated from isolated electromagnetic objects, it is required that the fraction of energy deposited in the EM calorimeter is greater than 0.5 but smaller than 0.9.
- $CoarseHadronicFraction < 0.4$: To remove jets which are faked by noise appearing in the coarse hadronic calorimeter, the fraction of energy in the hadronic calorimeter has to be smaller than 0.4.
- $L1Confirmation$: Another quality criterion is the L1 confirmation which compares the energy of the reconstructed jet to the energy reconstructed by the L1 trigger readout. Jets are accepted only if the ratio of the energy measurements is above a certain threshold, which depends on $|\eta_{det}|$ and p_{T} of jets. More details about reconstruction and identification of jets can be found in Ref. [39].

Jets passing above criteria are referred to as "Good Jets". A further criteria "Vertex Confirmation" is also applied on the Good Jets to ensure that at least two tracks in the jet are associated with the PV.

The efficiencies for reconstructed jets to pass the L1 trigger, EM fraction and CH fraction are quite different in collision data and MC. Studies using dijet events are performed to parametrize these differences in p_{T} and η_{det} of the jet [40].

3.7.2 B-Tagging

A heavy-flavor likelihood, commonly referred to as b-tag output, can be associated to each jet. A calorimeter jet is "taggable" if its cone direction matches the one of a track-jet within a distance of $\Delta R = 0.4$. A b-tagged jet is a taggable jet in which the tracks are associated to a displaced vertex. The jets are not labeled with a binary value indicating whether or not the jet originated from a b quark; instead, a continuos b-tag value is used. This value is computed combining several pieces of information that includes: the significance of the tracks displacements, the secondary vertex position, and the presence of muons. These pieces of information are combined within a multivariate discriminant, whose continuous output is a measure of the b-tag likelihood. The efficiency and the fake rate of the b-tag algorithms vary significantly as a function of the jet p_{T} and η_{det}. Nevertheless, efficiencies above 50% with fake rates below 1% are obtained for energetic central jets. A detailed description of the b-tag techniques used at DØ , along with the resulting performance, can be found in Ref [44].

3.7.3 Jet Energy Scale

Due to various reasons such as detector effects, dead material in the calorimeter, sampling rate and the fixed cone radius, the energy of a jet as measured in calorimeter can differ significantly from the initial energy of the partons forming the jet. Therefore the Jet Energy Scale (JES) correction are applied to the energy of the reconstructed jets. This correction is defined [41] as:

$$E_{jet} = \frac{E_{jet}^{raw} - O}{F_\eta \times R \times S} \tag{3.6}$$

with

E_{jet} : corrected jet energy at particle level,
E_{jet}^{raw} : uncorrected jet energy,
O : offset energy correction,
F_η : relative response correction,
R : absolute response correction,
S : showering correction.

The individual components of the jet energy scale correction are derived and applied sequentially in the order as given in Eq. 3.6. The estimation of the corrections is done separately for collision data and MC . Starting from the raw jet energy, the offset correction O is subtracted. The offset energy arises from multiple $p\bar{p}$ interactions, beam remnants, electronics and noise in the calorimeter or energy from previous collisions referred to as "pile-up". The offset correction is measured from data using minimum bias events which are triggered by the luminosity monitor. The energy density per tower is measured depending on the number of reconstructed primary vertices in order to include the luminosity dependence in the offset energy calculation. The energy deposited within a jet cone is then defined as the offset energy.

After correcting for the offset, jet energy is divided by the relative response correction F_η. This relative response correction calibrates the jet energy for the fact that the response of the calorimeter is not uniform in rapidity. Particularly the ICR and the massless gaps show a different response compared to the other calorimeter cells in the central and endcap parts of the calorimeter. For the measurement of F_η the *Missing Transverse Energy Projection Fraction* (MPF) method is applied. In this method dijet or photon+jet events are used, with the tag object always in the central detector region and the probe object in the region to be considered. The difference in transverse energy between tag and probe object can then be used to extract the relative response. Due to differences between the physical detector and its simulation this response determination is performed separately in collision data and MC.

After the determination of the relative response, the absolute response correction $\mathcal{R}$ can be measured and applied to the jet energy. $\mathcal{R}$ corrects for the difference in calorimeter response of hadrons and electrons and the energy loss in non-instrumented detector regions.

The last step is the determination and application of the shower correction $\mathcal{S}$. Due to showering in the calorimeter, the energy of the jet measured belonging inside or outside of the jet cone, could be missing from or added to the jet energy. The correction does not account for physical showering, for example due to gluon emission. $\mathcal{S}$ is measured in photon+jet events with exactly one primary vertex. The ratio between the jet energies at the particle level and the reconstruction level yields the showering correction.

Muons identified within a jet cone usually originate from semi-leptonic decays and therefore are indicative of a neutrino carrying off some of the jets energy. A separate correction has been derived on jets containing muons to account for this loss of information.

3.7.4 Jet Shifting, Smearing and Removal

Due to limitations of the modeling of the DØ detector, there are differences in the simulation of data for response of the calorimeter for single particles which can result in biases when modeling the jets. Hence it is necessary to modify the MC simulation and match it to the performance observed in data. For jet resolution, jet reconstruction efficiencies and identification efficiencies, a method called *"Jet shifting, smearing and removal"* (JSSR) is introduced [42]. The corrections is derived using the transverse momentum imbalance in photon+jets and $Z/\gamma^* + \text{jet}$ events, where the p_T imbalance is defined as:

$$\Delta S = \frac{p_\mathrm{T}^{Jet} - p_\mathrm{T}^{Z/\gamma}}{p_\mathrm{T}^{Z/\gamma}} \tag{3.7}$$

ΔS distributions are plotted for various $p_\mathrm{T}^{Z/\gamma}$ bins. This method is used for both MC and data to get the measure of the relative performance of the calorimeter between simulation and collision data. Any difference in response for collision data and MC is corrected in the MC. In deriving this correction, ΔS distributions are fitted with a function composed of product of an error function and a Gaussian. The central value of the Gaussian gives the information on the jet energy scale while the width of the Gaussian gives the jet energy resolution (JER). The ratio of the areas under the error function$\times$ Gaussian and Gaussian-only gives the jet reconstruction and identification efficiency.

Smearing corrections to the p_T of the simulated jets from JER are multiplicative factors randomly drawn from a Gaussian width equal to quadratic difference in the Gaussian widths of the data ($(\sigma^{data}_{Gauss})^2$) and MC ($(\sigma^{MC}_{Gauss})^2$) fits:

$$\sigma_{JER} = \sqrt{(\sigma^{data}_{Gauss})^2 - (\sigma^{MC}_{Gauss})^2} \tag{3.8}$$

Corrections to the simulated jet p_{T} (jet shifting) are evaluated from the difference in Gaussian means for the data and MC as a function of jet p_{T}.

3.8 Missing Transverse Energy

The missing transverse energy ($\not\!\!E_{\mathrm{T}}$) is an important tool for the present analysis. $\not\!\!E_{\mathrm{T}}$ is the signature of particles escaping the detector undetected such as neutrinos or weakly interacting particles predicted by non-SM theories. Partons involved in the original hard scattering process have large longitudinal momenta but their total transverse momentum at the interaction point is approximately zero. Therefore almost no net transverse momentum of the particles produced in the $p\bar{p}$ interaction should be produced. Thus the magnitude of the vectorial sum of the p_x and p_y components of all the particles produced are expected to be zero within the detector resolution. However, when a particle escapes the detector without interacting it produces an imbalance of transverse momentum. $\not\!\!E_{\mathrm{T}}$ is the amount of energy needed to restore the balance. The location of each cell in the detector is used and combined with the primary vertex position information to yield a direction for the energy $\vec{\eta}^{\mathrm{cell}}$. The uncorrected missing energy in an event is measured as:

$$\vec{E}_T^{\mathrm{miss}} = -\sum_{N_{cells}} \vec{E}_T^{\mathrm{cell}} \tag{3.9}$$

where $\vec{E}_T^{cell}$ is a three-momentum with a magnitude equal to the energy of a cell in the EM and fine hadronic layers of the calorimeter, and pointing from the primary vertex to the center of the cell. The CH cells are not included because they can contain a substantial amount of noise.

The missing energy can be split into its transverse components as:

$$\left(E_T^{\mathrm{miss}}\right)^2 = \left(E_{T_x}^{\mathrm{miss}}\right)^2 + \left(E_{T_y}^{\mathrm{miss}}\right)^2 \tag{3.10}$$

3.8.1 Corrections to the $\not\!\!E_{\mathrm{T}}$

To estimate real $\not\!\!E_{\mathrm{T}}$, various corrections are applied to it:

- $\not\!\!E_{\mathrm{T}}$ is reconstructed with all the calorimeter cells except the CH cells, but it is considered that the CH fraction of jets used in the analysis are signal and not noise and it must be taken into account. Hence it is subtracted from $\not\!\!E_{\mathrm{T}}$. This is called "CH correction".

- The measured EM and jet energies are re-calibrated to reflect their true energies in the detector. This change in energy impacts the $\not\!\!E_{\rm T}$ as well. Hence the $\not\!\!E_{\rm T}$ must be corrected in order to account for this change. The jet energy scale (JES) correction has a considerably larger impact on the $\not\!\!E_{\rm T}$ than the EM energy corrections. The $\not\!\!E_{\rm T}$ after these corrections is referred to as the "calorimeter $\not\!\!E_{\rm T}$".

- Muons, which usually behave as minimum-ionizing particles, deposit very little of their energy (typically about 2 GeV) in the calorimeter. The calorimeter $\not\!\!E_{\rm T}$ does not account for the presence of such particles. Therefore, some of the energy imbalance in the transverse plane is due to presence of muons and not neutrinos. Hence the momentum of all the muons is subtracted vectorially from the calorimeter $\not\!\!E_{\rm T}$.

After applying all the above corrections $\not\!\!E_{\rm T}$ is used in the analysis.

3.8.2 $\not\!\!E_{\rm T}$ Based Variables

In order to discriminate against $\not\!\!E_{\rm T}$ generated by mis-measurement of reconstructed jet energies, the variable $\not\!\!E_T^{\rm Scaled}$ is calculated. To determine it, the fluctuation in the measurement of jet energy in the transverse plane can be approximated by $\Delta E^{\rm jet} \cdot \sin\theta^{\rm jet}$ where $\Delta E^{\rm jet}$ is proportional to $\sqrt{E^{\rm jet}}$. The opening angle $\Delta\phi\,({\rm jet}, \not\!\!E_{\rm T})$ between this projected energy fluctuation and the missing transverse energy provides a measure of the contribution of the jet to the missing transverse energy. The scaled missing transverse energy is defined as:

$$\not\!\!E_T^{\rm Scaled} = \frac{\not\!\!E_{\rm T}}{\sqrt{\sum_{\rm jets} \left(\Delta E^{\rm jet} \cdot \sin\theta^{\rm jet} \cdot \cos\Delta\phi\,({\rm jet}, \not\!\!E_{\rm T})\right)^2}} \tag{3.11}$$

The distribution of $\not\!\!E_T^{\rm Scaled}$ in the RunIIb data is shown in Fig 3.6.

To act against the Z/γ^* background, we use another modified $\not\!\!E_{\rm T}$ distribution, $\not\!\!E_T^{\rm special}$. $\not\!\!E_T^{\rm special}$ assigns less significance to $\not\!\!E_{\rm T}$ in an event with small $\delta\phi$ between MET and any other object (lepton or jet) as the small opening angle may indicate a potential source of mismeasurement. It is defined as:

$$\begin{aligned} \not\!\!E_T^{\rm special} &= \not\!\!E_{\rm T} \text{ if } \Delta\phi\,(\not\!\!E_{\rm T}, \text{nearest lepton or jet}) > \pi/2 \\ \not\!\!E_T^{\rm special} &= \not\!\!E_{\rm T} \times \sin\Delta\phi\,(\not\!\!E_{\rm T}, \text{nearest lepton or jet}) \texttt{ otherwise} \end{aligned} \tag{3.12}$$

A representative distribution for this variable in the RunIIb one jet sample is shown in Fig 3.6.

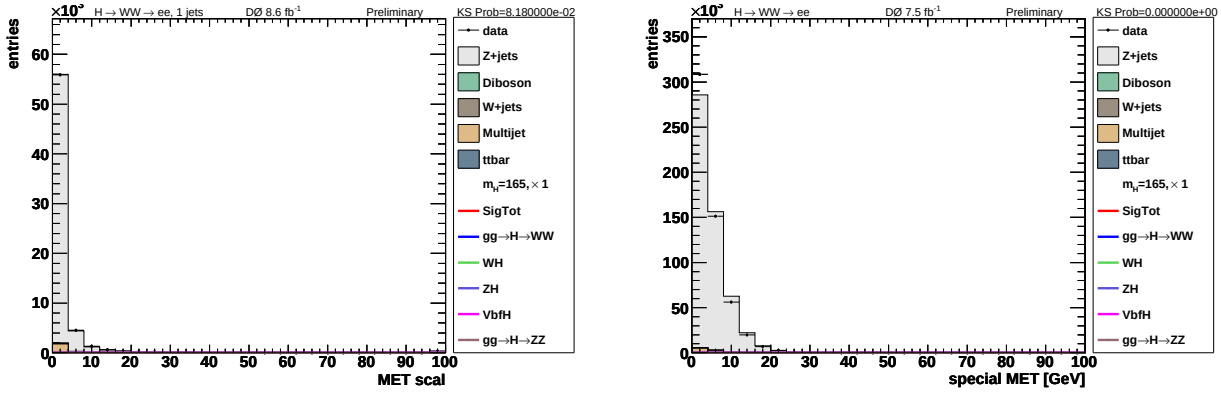

Figure 3.6: The distribution of $\not\!\!E_T^{\text{Scaled}}$ (left) and $\not\!\!E_T^{\text{special}}$ (right) in linear scale at the final selection stage for the entire RunII dataset for the one jet bin. The signal shown here corresponds to the SM Higgs boson of m_H=165 GeV/c^2.