# Chapter 4

# ELECTRON RECONSTRUCTION AND IDENTIFICATION AT ATLAS

The central theme of this thesis is the improvement of the selection of signal electrons in SUSY searches. In particular, Chapter 5 deals with the estimation of electron charge mis-identification, Chapter 6 discusses a supersymmetry seach that involves leptons (electrons or muons) in the final state, and Chapter 7 measures the identification efficiencies for in-jet electrons[1]. This chapter presents a more extended discussion of electron reconstruction and identification at ATLAS [37].

At ATLAS, a signal electron that has been selected passes through two major steps, reconstruction and identification. Electron reconstruction, discussed in Section 4.1, is the selection, using information from the inner detector and the electromagnetic calorimeter, of a set of objects which are called electron candidates. Electron identification, on the other hand, refers to the selection from a pool of electron candidates; it is discussed in Section 4.2.

Figure 4.1 [37] shows the hypothetical path of an electron, in red trajectory, through the ATLAS detector. The electron emerges near the collision point, passes the tracking system (made up of the pixel detectors, the silicon-strip detectors, and the TRT) before entering the electromagnetic calorimeter. Also shown in the figure is the path of a photon, in dashed trajectory, produced by the interaction of the electron with the material in the tracking system.

## 4.1 Electron Reconstruction

At ATLAS, we expect most electrons passing through the detector to interact with the material of the latter and lose a significant amount of energy through bremsstrahlung. These interactions, which may happen along the path of the electrons, typically cause radiated photons and consequently electron-position pairs, all of which tend to collimate. Thus, electron reconstruction at ATLAS consists of three fundamental components:

[1]These refer to electrons that are found within $\Delta R = 0.4$ of high $p_T$ jets.

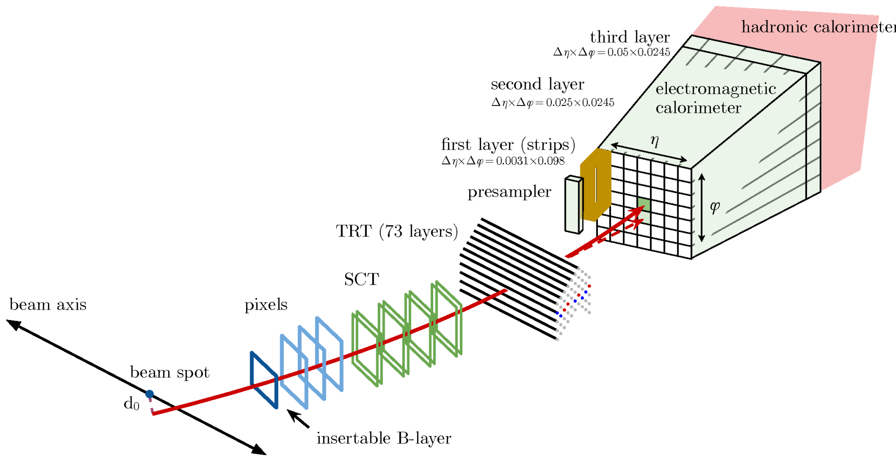



Figure 4.1: The hypothetical path of an electron through the detector [37] is shown in red in the figure. The electron moves through the tracking system (pixel detectors, silicon-strip detectors, and the TRT) before entering the electromagnetic calorimeter. The dashed red line shows the path of a photon that comes from the interaction of the electron with the material in the tracking system.

- Localized clusters of energy deposits in the electromagnetic calorimeter;
- Tracks in the inner detector;
- Matching of tracks to the clusters.

These components will be discussed in some detail in the following. The discussion will focus on what was done before a new electron and photon reconstruction method, dynamic and topological cell clustering-based, was introduced at ATLAS [38].

### 4.1.1 Seed-cluster reconstruction

Electromagnetic energy cluster candidates are reconstructed from localized energy deposits in the electromagnetic calorimeter (Section 3.3.2.2) using an algorithm known as the sliding-window algorithm [39]. To this end, the $\eta \times \phi$ plane of the electromagnetic calorimeter is divided into a grid of $200 \times 256$ elements, also called towers, of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. The algorithm starts from localized energy deposits of size $3 \times 5$ towers in $\eta \times \phi$ where the total transverse energy exceeeds 2.5 GeV, moving in steps of 0.025 in either the $\eta$ or the $\phi$ direction and amassing neighboring localized energy deposits. These accumulated clusters of energy deposits are referred to as seed-cluster candidates. In the case where two candidates overlap within an area of $\Delta\eta \times \Delta\phi = 5 \times 9$ units of $0.025 \times 0.025$, both will undergo a selection process in which

- Only the one that has transverse energy at least 10% higher than the other is kept; or otherwise
- Only the one that contains the highest transverse momentum in the central tower is kept.

The reconstruction efficiency is found to depend on $\eta$ and on the transverse energy. Figure 4.2 shows the dependency on $E_T$. The efficiency ranges from 96% at $E_T = 7$ GeV to more than 99% above $E_T = 15$ GeV.

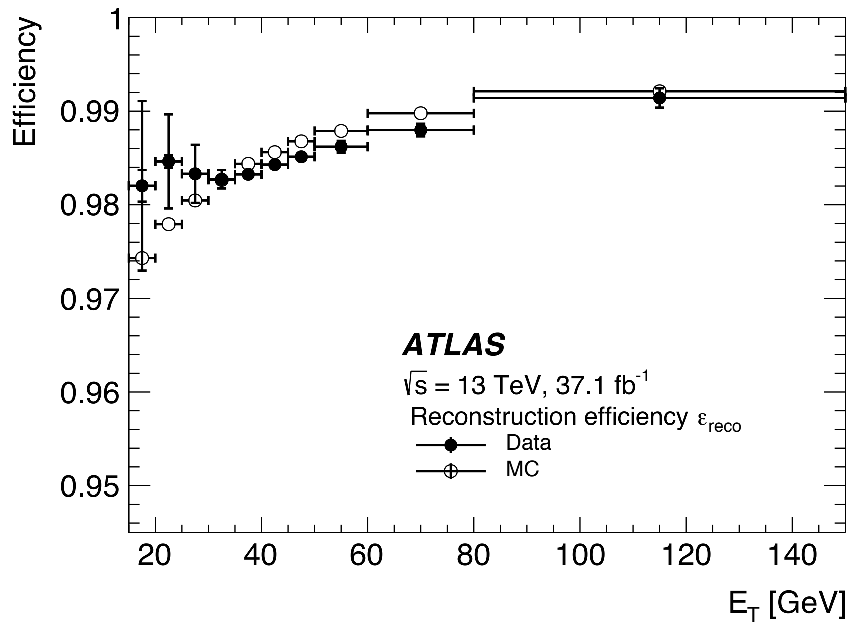



Figure 4.2: The reconstruction efficiency relative to reconstructed clusters as a function of $E_T$ for $Z \to ee$ events.

### 4.1.2 Track reconstruction

The interactions of charged particles with the inner detector material created hits [40] in the latter. In track reconstruction, these hits are assembled into clusters in the pixel and SCT detectors, from which three-dimensional measurements called space-points are built. In the silicon-detector layers, sets of three space-points are used to form track seeds. Then a pattern-recognition algorithm proceeds to build track candidates, in which energy loss of a particle due to its interactions with the detector material is modelled assuming the particle is a pion. A modified pattern-recognition will be used in the case where a track seed having $p_T > 1$ GeV cannot be extended to a full track of at least seven hits per track candidate and the associated electromagnetic calorimeter cluster satisfies shower width and depth requirements. The modified algorithm allows up to 30% energy loss for bremsstrahlung at each intersection of the track with the detector material.

Track candidates with $p_T > 400$ MeV are fit using the ATLAS Global $\chi^2$ Track Fitter [41], taking into account which pattern-recognition algorithm was used. Ambiguities arising from track candidates sharing hits are also resolved in this step. Figure 4.3 shows that the reconstruction efficiency ranges from 80% at $E_T = 1$ GeV to more than 98% above $E_T = 10$ GeV.

An additional fit, using the Gaussian-sum filter (GSF) [42] method to better model energy loss of the particle, is applied on tracks having at least four silicon hits and that are loosely matched to electromagnetic clusters. The method takes into account non-linear effects related to bremsstrahlung and models experimental noise by a sum of Gaussian functions.

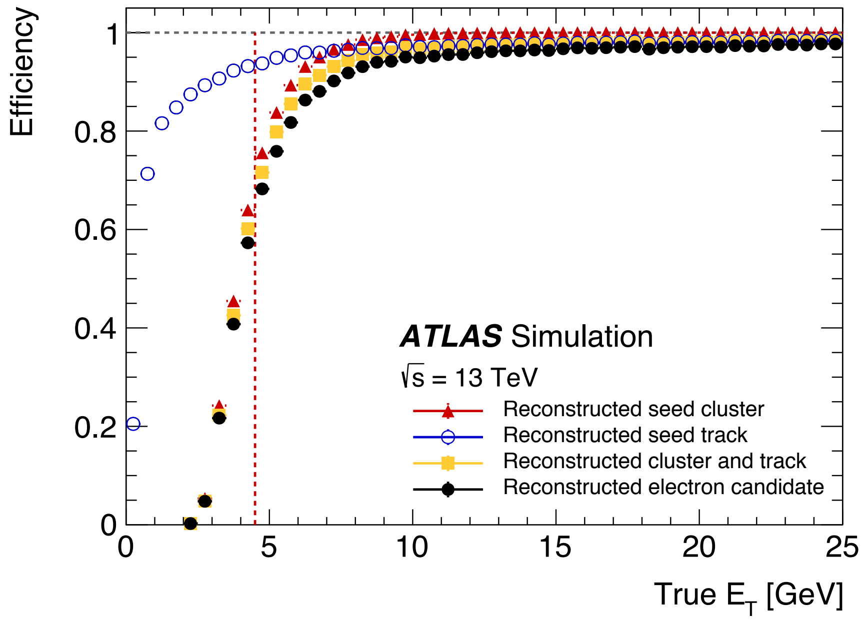



Figure 4.3: The total reconstruction efficiency for simulated electrons in a single-electron sample as a function of the true generator $E_T$ for each step in the reconstruction process: $\Delta\eta \times \Delta\phi = 3 \times 5$ seed-cluster reconstruction (red triangles), seed-track reconstruction using the Global $\chi^2$ Track Fitter (blue open circles), both steps but using GST tracking (yellow squares), and the final reconstructed electron candidate (black closed circles).

### 4.1.3 Electron-candidate reconstruction

In this final step, the GSF-track candidate is matched to the candidate calorimeter seed cluster and the final cluster size is determined. If during the matching procedure several tracks may be matched to a same cluster then an algorithm using such information as the number of hits in the silicon detectors, the number of hits in the innermost sillicon layer, and others, is applied to select out the primary track. The resulting object is called an electron candidate if it has an associated track with at least four hits in the silicon layers and no association with a vertex from photon conversion. If on the other hand its primary track can be matched to a secondary vertex and has no pixel hits, the object is classified as a photon candidate instead.

Subsequently the candidate electron undergoes an additional classification — mainly to keep a high photon-reconstruction efficiency — to determine if it still should be considered as a potential photon candidate. The classification uses the candidate electron's $E/p$ and $p_T$, the presence of a pixel hit, and the secondary-vertex information.

The energy of the final electron candidate is computed from the calibrated energy of the extended-window cluster, which is formed from the original seed cluster by expanding the size of the latter of $\phi$ or $\eta$. The calibration uses multivariate techniques [44, 45].

Figure 4.3 and 4.4 show the reconstruction efficiency as a function of $E_T$ and as a function of $\eta$ in bins of $E_T$, respectively, from $Z \rightarrow e^+e^-$ events. They show that for $E_T > 15$ GeV, the reconstruction efficiency varies from approximately 97% to 99%. Moreover, simulation efficiency is lower than data efficiency in the low $E_T$ region ($E_T < 30$ GeV) but is higher in the higher $E_T$ region ($E_T > 30$ GeV).

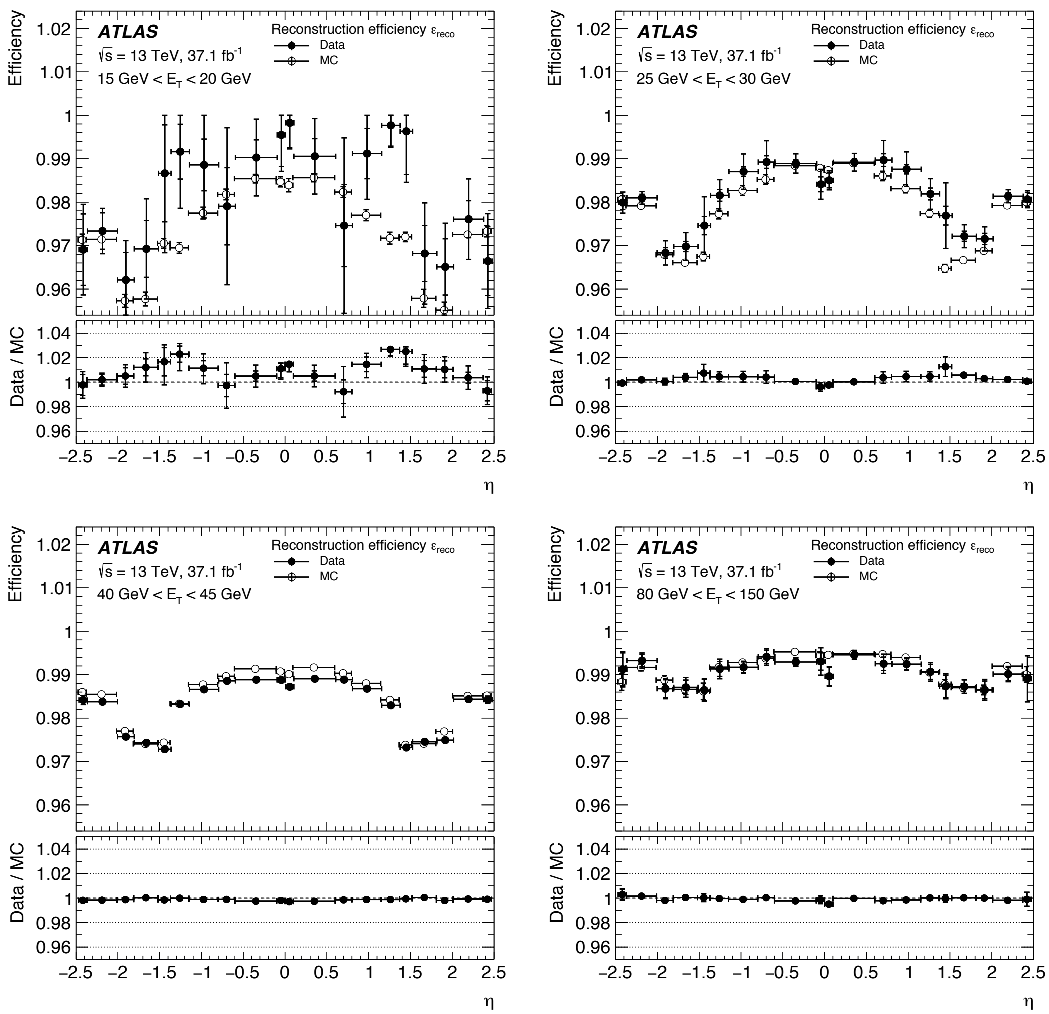



Figure 4.4: Reconstruction efficiencies relative to reconstructed clusters evalued in the 2015-2016 dataset (closed points) and in simulation (open points) and the ratios between the two in $Z \rightarrow ee$ events. The efficiencies are shown as a function of $\eta$ in four $E_T$ bins: 15-20 GeV (top left), 25-30 GeV (top right), 40-45 GeV (bottom left), and 80-150 GeV (bottom right).

## 4.2 Electron Identification

Electron candidates coming out of reconstruction consist mostly of background electrons that are made up of hadrons, electrons from photon conversions, and electrons from heavy-flavor decays. Electron identification is the step whereby these background electrons are reduced.

### 4.2.1 Likelihood Identification

Prompt electrons[2] that enter the central region ($|\eta| < 2.47$) are selected using a method called the likelihood identification. In this method, the differences in shower shapes, in track conditions, penetration depth, and others between prompt electrons and background electrons are analyzed in detail. Specifically, the following quantities [37], which are classified into seven types, are used:

- Hadronic leakage:
  - $R_{\text{had1}}$: Ratio of the transverse momentum in the first layer of the Hadronic Calorimeter to that of the Electromagnetic Calorimeter.
  - $R_{\text{had}}$: Ratio of the transverse momentum in the Hadronic Calorimeter to that of the Electromagnetic Calorimeter cluster (used in range $0.8 < |\eta| < 1.37$).
- Third layer of EM calorimeter
  - $f_3$: Ratio of the energy in the third layer to the total energy in the Electromagnetic Calorimeter (used only for $E_T < 80$ GeV).
- Second layer of EM calorimeter
  - $\omega_{\eta^2}$: Lateral shower width.
  - $R_\phi$: Ratio of the energy in $3 \times 3$ cells over the energy in $3 \times 7$ cells centered at the electron cluster position.
  - $R_\eta$: Ratio of the energy in $3 \times 7$ cells over the energy in $7 \times 7$ cells centered at the electron cluster position.
- First layer of EM calorimeter
  - $\omega_{\text{stot}}$
  - $E_{\text{ratio}}$
  - $f_1$
- Track conditions
  - $n_{\text{Blayer}}$: the number of hits in the innermost pixel layer.
  - $n_{\text{Pixel}}$: the number of hits in the Pixel detector.
  - $n_{\text{Si}}$: the total number of hits in the pixel and SCT detectors.
  - $d_0$
  - $|d_0/\sigma(d_0)|$
  - $\Delta p/p$
- TRT

[2] These refer to electrons that originate from the prompt decays of particles such as $W$, $Z$, and other beyond the Standard Model particles.

  - eProbabilityHT

- Track-cluster matching

  - $\Delta\eta_1$: $\Delta\eta$ between the cluster position in the first layer and the extrapolated track
  - $\Delta\Phi_{\text{res}}$
  - $E/p$: ratio of the cluster energy to the track momentum (for $E_T > 150$ GeV)

These are used as inputs to two likelihood functions, one for signal electrons and one for background electrons, which take the forms

$$L_S(\mathbf{x}) = \prod_{i=1}^{n} P_{S,i}(x_i), \qquad L_B(\mathbf{x}) = \prod_{i=1}^{n} P_{B,i}(x_i)$$

respectively. Here $\mathbf{x}$ is a vector of entries $x_i$ which are the inputs that correspond to the quantities listed above, each of which has a signal probability distribution function (pdf) and a background pdf. $P_{S,i}(x_i)$ is the value of the signal pdf of the quantity $i$ at the value $x_i$, and likewise $P_{B,i}(x_i)$ is the value of the background pdf. The pdfs are derived using simulation samples, with corrections applied when discrepancies with the corresponding data are found. The correlations between the inputs are neglected.

Then, from each electron candidate discriminant value $d_L$ is computed according to the formula

$$d_L = \frac{L_S}{L_S + L_B}$$

This discriminant is actually transformed into

$$d'_L = -\tau^{-1} \ln(d_L^{-1} - 1) \tag{4.1}$$

which then serves as a quantity to assess if an electron candidate should be considered a prompt electron. The parameter $\tau$ is set to 15 [43]. Figure 4.5 shows a comparison of $d_L$ and $d'_L$ for prompt electrons from $Z$-boson decays and for background, illustrating the effective separation between the two.

### 4.2.2 Operating Points

In general, background rejection and identification efficiency are inversely related i.e. the higher is one, the lower is the other, and vice versa. In order to cover various signal efficiencies and background rejection factors as needed by physics analyses, ATLAS has defined four so-called identification operating points. They are, in order of increasing background rejection power, VeryLoose, Loose, Medium, and Tight. All operating points have fixed requirements on tracking criteria:

- Loose, Medium, and Tight: at least two hits in the Pixel detector and a total of seven hits in the pixel and silicon strip detectors combined. To reduce background from photo conversions, Medium and Tight require one of these pixel hits to be in the innermost pixel layer or, if this layer is out of order, the layer immediately after it.

There is a variation of the Loose operating point, called LooseAndBLayer, that is the same as Loose except with the addition of the requirement of a hit in the innermost layer also.

- VeryLoose: one hit in the pixel detector, regardless of the layer.

A particular value of $d'_L$ (Equation 4.1) is defined for each of them, and in the context of an operating point, electron candidates with computed $d'_L$ larger than the defined value are considered prompt electrons. In the likelihood method, candidates that pass an operating point having a higher background rejection power also pass the operating points having lower background rejection powers.

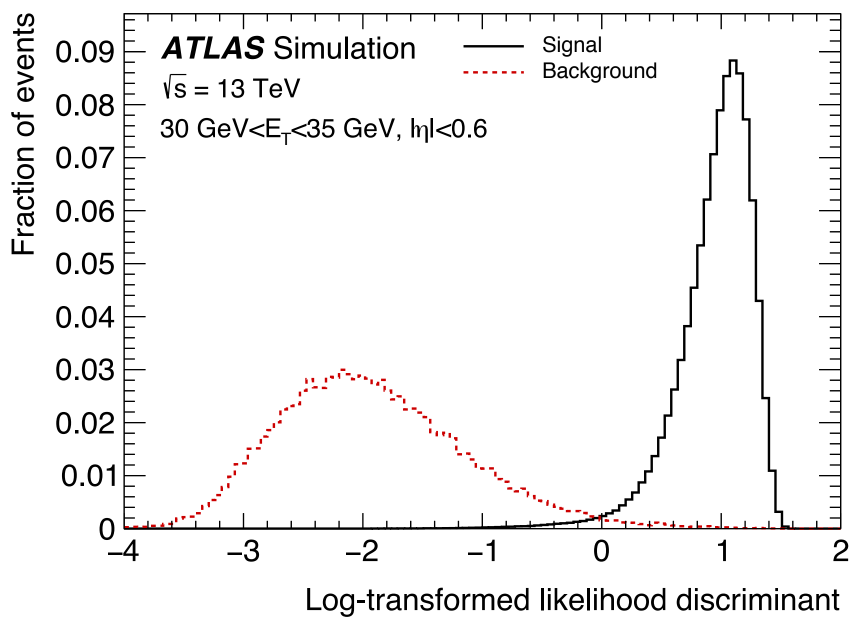
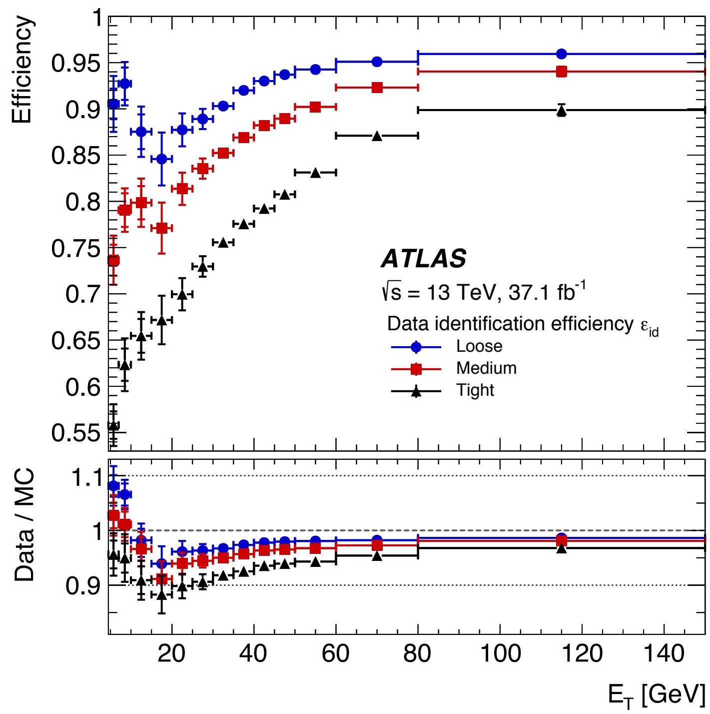



Figure 4.5: The discriminant $d'$ (Formula 4.1) for reconstructed electron candidates with good quality tracks with 30 GeV $< E_T <$ 35 GeV. The black distribution shows prompt electrons in a $Z \to ee$ simulation sample, and the red distribution shows background electrons in a generic two-to-two process simulation sample.

Figure 4.6 shows the efficiencies measured in $J/\psi \to ee$ and $Z \to ee$ events for data and the corresponding data-to-simulation ratios. Specifically, the variations of the efficiencies in $E_T$ and $\eta$ for the Loose, Medium, and Tight operating points are displayed. The efficiency ranges from 55% at $E_T = 4.5$ GeV to 90% at $E_T = 100$ GeV for the Tight point, and from 85% at $E_T = 20$ GeV to 96% at $E_T = 100$ GeV for the Loose point.



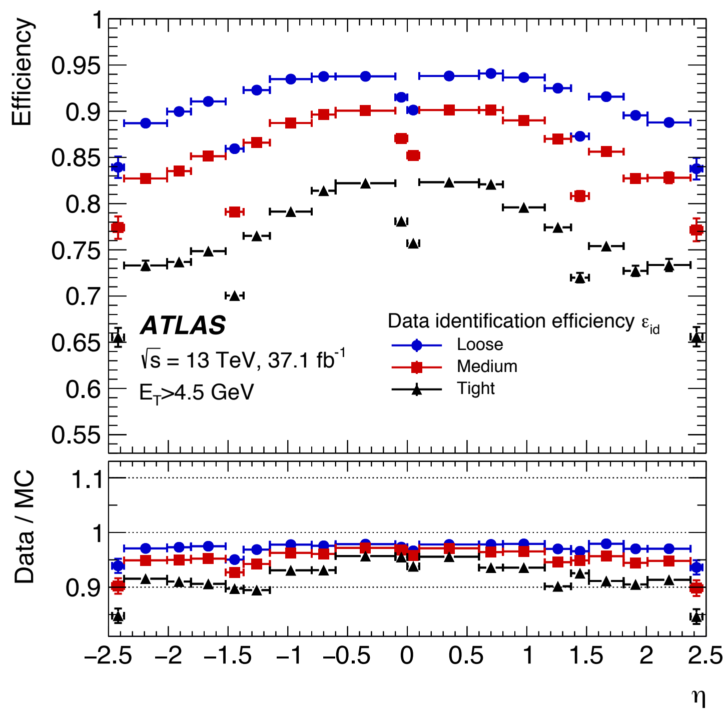




Figure 4.6: The likelihood identification efficiencies as functions of $E_T$ and $\eta$ in $Z \to ee$ events for Loose, Medium, and Tight (shown in blue, red, and black respectively). The data efficiencies are obtained by applying data-to-simulation efficiency ratios measured in $J/\psi \to ee$ and $Z \to ee$ events to $Z \to ee$ simulation.