Chapter 7

IN-JET ELECTRON IDENTIFICATION EFFICIENCIES

In early 2015 the LHC restarted after two years of shutdown, beginning what is referred to as Run 2. The new centre-of-mass energy was 13 TeV, in place of the previous 8 TeV. The higher energy opens up unexplored parameter space and allows further probe of supersymmetry (SUSY) and other beyond-the-Standard-Model processes [79–82]. At ATLAS, many SUSY searches involve supersymmetric particles that decay into the standard model top quarks and we expect, with higher centre-of-mass energy, more massive supersymmetric particles and consequently an increase in the production of high p_T top quarks. Since the top quarks also decay, in fact most of the time into a W boson and a b quark, we in turns expect boosted decay topology, in other words the daughter particles of the top quarks, which include the daughter particles of the W boson and the b quark, tend to stay close to each other (Figure 7.1). In the case of leptonic top quark decays, the lepton tends to be very close to the produced b-jet.

This chapter describes the work to measure the identification efficiencies for electrons that are found inside $\Delta R = 0.4$ of high p_T jets, which will also be referred to in the chapter as in-jet electrons. Prior to the work in this chapter, there was no attempt to measure the identification efficiencies for in-jet electrons. The chapter will be organized as follows. In section 7.1 we motivate the need for the measurement of the identification efficiencies for in-jet electrons. Section 7.2 describes the method used to perform the measurements. Section 7.3 presents the measured efficiencies for in-jet electrons and section 7.4 gives some conclusions.

The data used for this chapter was collected in the period 2015-2016 at 13 TeV center-of-mass and correponded to an integrated luminosity of 36.47 fb^{-1}.

7.1 Motivation

At ATLAS, prior to Run 2, electrons found outside $\Delta R = 0.4$ of jets were used by most analyses. Indeed, although there were some attempts to select signal electrons inside jets [**?**], at 7-8 TeV such electrons can hardly be expected; objects inside

jets that are identified as electrons are mostly background electrons that are either hadrons faking jets or real electrons coming from heavy-flavour jet decays. As a result, most analyses rejected electrons inside $\Delta R = 0.4$ of jets and only worked with electrons outside jets, of which different efficiencies, among them identification efficiencies, are measured and provided by the ATLAS Egamma group [37].

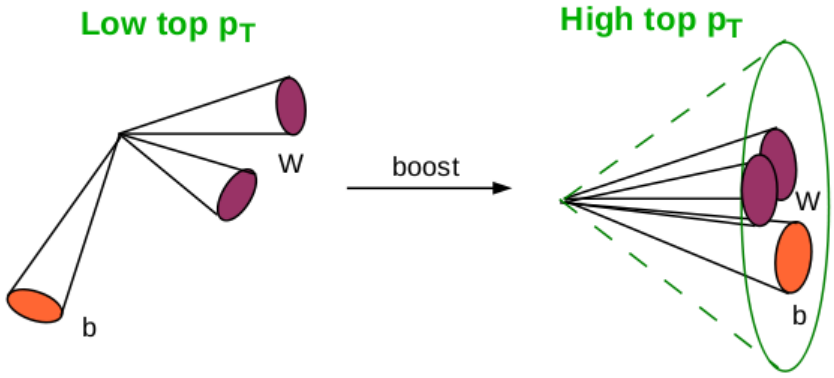
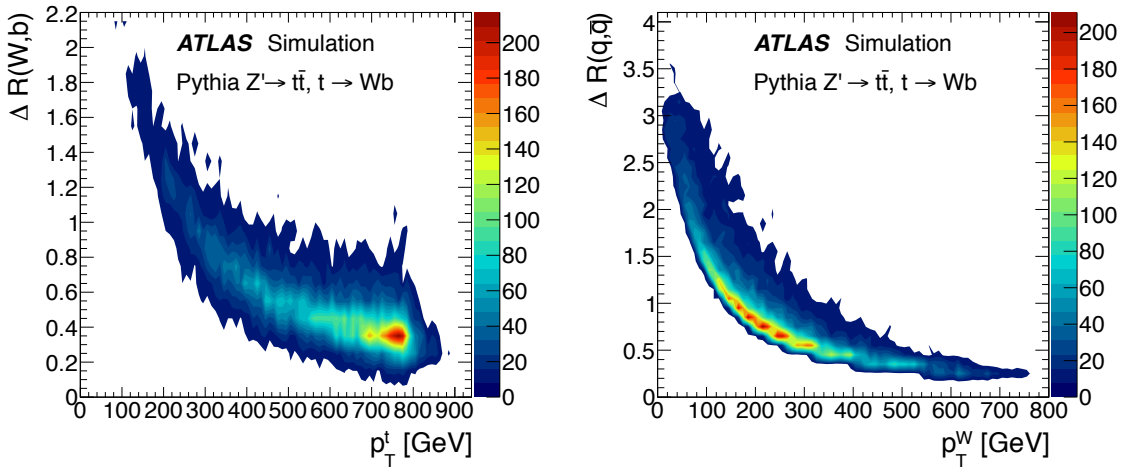


Figure 7.1: An illustration of low p_T top quark decay (left) and boosted top decay (right) of a high p_T top quark. In the case of high p_T top quark decay the daughter particles of the top quark, which include the daughter particles of the W and the b quark, are expected to be found close to each other [83].

In Run 2, as the centre-of-mass of the LHC reached 13 TeV, high p_T top quarks are expected to come from the decays of massive supersymmetric or other beyond-the-Standard-Model particles. As we have already mentioned, a top quark decays almost all of the time into a W and a b quark, and high p_T top quarks are expected to undergo boosted decays where the produced particles, here the daughters of the W and the b quark, are found close to each other. In particular, there are two possible scenarios: either the W decays hadronically, in which case the produced jets stay close to each other, or it may decay leptonically, in which case the leptons are expected to be found near the produced b-jets. Either scenario is important and the possibility of encountering boosted topology was assessed by several ATLAS physics groups, albeit most works focused on hadronic boosted top quark decays [84–86]. Figure 7.2 shows the angular distance ΔR (Formula 3.2) between the W's and the b-quarks as a function of the top p_T, in the context of a hypothetical particle Z' at mass $m_{Z'} = 1.6$ TeV [86] that decays into a $t\bar{t}$ pair. It also shows the separation between the light quarks of the subsequent hadronic decay of the W boson. As can be seen, the angular distance decreases as the top quark p_T increases, and at high top quark p_T a non-negligible fraction of the distances becomes very small. In the case of leptonic top quark decays, we should expect the decay products of the W's to be found very near the b-jets.



(a) $t \to Wb$ (b) $W \to q\bar{q}$

Figure 7.2: (7.2a) The angular distance ΔR between the W's and the b quarks as a function of the top quark p_T simulated PYTHIA [52], in the context of a hypothetical particle Z' ($m_{Z'} = 1.6$ TeV) that decays into a $t\bar{t}$ pair. At high top quark p_T a non-negligible fraction of the distances is seen to be very small. (7.2b) The angular distance between two light quarks from $t \to Wb$ decay as a function of the p_T of the W boson [86].

Taking advantage of the new topology, several SUSY searches which expect standard model top quarks from the decays of massive supersymmetric particles began to include electrons inside $\Delta R = 0.4$ of jets in their analyses. Considerable increases in signal acceptance were seen by different analysis groups, including the search for the gluinos discussed in Chapter 6, which in fact was the analysis that motivated the measurements described in this chapter.

In the context of the work of this chapter, we also computed the fraction of in-jet electrons over isolated electrons as a function of the top quark p_T, at truth-level, in order to assess the significance of the number of electrons that are found inside jets. The measurement used PowhegPythia $t\bar{t}$ events simulated at 13 TeV centre-of-mass energy (Chapter 6, Section 6.2), where dilepton events consisting of a muon and an electron were selected. The selections made use of the p_T-dependent overlap removal (Chapter 6, Section 6.3)

$$\Delta R < \min\,(0.4,\, 0.04 + 10\ \text{GeV}/p_T)$$

where the overlap removal is also required to keep the overlapping b-jets. The formula takes advantage of the fact that the higher is the p_T of an electron or a muon, the closer it is to some jet. It thus ensures that such electrons or muons that are found inside $\Delta R = 0.4$ are also selected, provided they are not closer than what their p_T warrant. On the other hand, b-jets found within $\Delta R = 0.2$ of electrons or muons will not be rejected as is usually done in a standard overlap removal procedure.

The results are shown in Table 7.1, which lists several top quark p_T along with the number of in-jet electrons, the number of isolated electrons, all at truth-level, and the fraction of the former over the latter.

Top quark p_T (GeV)	Fraction
≤ 300	8.4%
≤ 425	17.2%
≤ 500	24.0%
≤ 650	39.0%
≤ 750	49.0%
≤ 800	53.0%
≤ 900	59.0%
≤ 1000	64.0%

Table 7.1: The fraction of in-jet electrons over the number of isolated electrons, both at truth-level, as a function of the top quark p_T. The fraction increases and becomes very significant at high top quark p_T.

It is seen that more and more electrons are found inside jets as the top quark p_T increases. In fact, the number of in-jet electrons becomes quite significant from 500 GeV, being approximately 25% there and reaching nearly 40% at 650 GeV. If the top quark p_T is allowed to go up to 1 TeV, the figure is 64%. This result supports the fact that different ATLAS analyses saw considerable increases in signal acceptance as they included in-jet electrons in their selections.

This chapter develops a method and performs the initial measurements for the identification efficiencies of electrons found inside $\Delta R = 0.4$ of jets. The measurements for electrons outside jets are done by the ATLAS Egamma group [37, 43].

7.2 Method

In general, to measure the identification (ID) efficiencies of electrons we start from a sample of reconstructed electrons. The efficiency at a particular identification operating point (section 4.2.2) is defined by the ratio

$$\text{ID efficiency} = \frac{\text{The number of identified electrons}}{\text{The number of reconstructed electrons}}$$

The number of identified electrons, the numerator, is contaminated with background electrons, and so is the demoninator. As a result, part of the work to obtain a good sample of electrons inside high p_T jets includes at the same time adequate background estimation methods, particularly because background electrons are expected to reside primarily inside jets.

In this thesis, a sample of electrons was obtained by selecting boosted $t\bar{t}$ dilepton ($e\mu$) events, described below in Section 7.2.1, 7.2.2, and 7.2.3. Background estimations will be described in Section 7.2.4.

7.2.1 Boosted Dilepton $e\mu$ Events

The standard method for measuring electron identification efficiencies is the tag-and-probe method supported by the Egamma group at ATLAS [43]. Such a measurement uses $Z \to e^+e^-$ events for high-energy electrons ($E_{\mathrm{T}} > 10$ GeV). The method is well-documented, and it is straight-forward to obtain a clean sample of electrons. It is expected, however, that because of the nature of the events, requiring electrons to be inside high p_T jets would lead to a sample unrepresentative of events with a boosted topology, and moreover the resulting sample would be statistics-limited also.

In this thesis, $t\bar{t}$ events at 13 TeV centre-of-mass are used to select electrons inside high p_T jets. In particular, dilepton $e\mu$ events will be selected; they are expected to be a source of reconstructed electrons that will be used for the measurement of the identification efficiencies. The selections will be such that the muon in an event will be a hard muon, while the electron on the other hand will remain almost untouched, i.e. it will be a reconstructed electron with a p_T cut applied.

Compared to $Z \to ee$ events, such $t\bar{t}$ events are expected to provide more statistics for electrons inside high p_T jets. In addition, because they directly involve the top quarks, they will bear close topology to many SUSY and other beyond-the-Standard-Model searches.

The remaining detail of the method is described below, including a discussion of the data as well as the simulations used, the definition of the signal region, and background subtractions.

7.2.2 Data and Monte Carlo Samples

The data used for this chapter was collected in the period 2015-2016. The integrated luminosity was 36.47 fb^{-1}. The following simulation samples, all at 13 TeV centre-of-mass, are used (Chapter 6, Section 6.2):

- PowhegPythia $t\bar{t}$. This set is used for selecting electrons inside high p_T jets. As will be discussed in Section 7.2.3, $t\bar{t}$ events will be partitioned into a semileptonic set (a muon from one top quark and all jets from the other top) and a dileptonic set ($e\mu$ events). The former will constitute a background and the latter a source of signal events. Indeed, in the semileptonic case, jets from one of the tops constitute a source of background electrons

- W+jets. This is used as a background. Indeed, as a top quark almost always decays into a W and a b-jet, W+jets naturally constitutes a background where the W produces a hard muon and the jets are a source of background electrons

- Single top. This includes the Wt production as well as the s-channel and t-channel productions. The Wt production is a source of signal electrons, while the other two productions provide a source of top+jets and are therefore background, as the top quark may be a source of a hard muon and the jets in the events are a source of background electrons.

7.2.3 Signal Region

The signal region will be the kinematic region in which the measurement of the identification efficiencies is performed. Before defining the signal region, we apply the following preliminary cuts, called pre-selection cuts:

Pre-selection

- ❏ One primary vertex
- ❏ Muon trigger. The following triggers were used for the periods 2015 and 2016:
  - 2015: `HLT_mu26_imedium || HLT_mu40`
  - 2016: `HLT_mu26_ivarmedium || HLT_mu50`
- ❏ p_T-dependent overlap removal, with the option of keeping overlapping b-jets turned on (Section 7.1)
- ❏ Veto events with bad or cosmic muons. Highly energetic jets could reach the muon spectrometer and create hits in the latter, or jet tracks in the inner detector could be erroneously matched to muon spectrometer segments, both cases of which are sources of bad muons. Thus events with these muons, along with those with muons coming from cosmic rays, are removed.
- ❏ Each event is required to have exactly one tagged muon and ≥ 1 electrons inside jets, where
  - The muon is required to have $p_T > 30$ GeV, $d_0/\sigma(d_0) < 3.0$, $z_0 < 0.5$ in terms of the transverse impact paramater and the longitudinal impact parameter, and ptvarcone30$/p_T < 0.06$, and be trigger-matched.
  - The electrons have $p_T \geq 30$ GeV and must overlap with some jets, i.e. they must be found inside $\Delta R < 0.4$ of jets. There could be more than one electron in the event, however only the leading one in terms of p_T will be used as the probe. The p_T cut is applied because it is applied in most analyses where in-jet electrons are actually used.
- ❏ ≥ 1 b-tagged jet, instead of exactly 2 b-tagged jets as is usually expected in $t\bar{t}$ events, since events with electrons inside jets are being selected and consequently b-tagging efficiency may suffer because the tracks of the electrons may confuse the b-tagging algorithm.

After these cuts, we arrive at a set of 3183 events with one hard muon and at least one electron found inside some jet. Several variables that were found to be discriminating are listed and discussed below, along with plots of their distributions. It is seen in the plots that the prominent source of background events is semileptonic events, whereas W+jets and single top s-channel and t-channel constitute two small sources of background, predicted by simulations to be 315.483 and 19.120 respectively.

- The mass of the large radius jet that overlaps with the probe electron, shown in Figure 7.3. The large radius jet is reclustered from the small radius jets present in the events (Chapter 6, Section 6.3) and accordingly in semileptonic events it is expect to be more massive, as it picks up the masses of the jets from the hadronic decay of one of the tops. In dileptonic events, on the other hand, there are fewer jets due to the leptonic decays of both of the tops, and moreover there is also missing energy present from the neutrino accompanying the electron. Therefore this variable is expected to be discriminating. The figure shows a markedly higher distribution for semileptonic events, especially as we move towards the right of the distribution, where dileptonic events are seen less and less and become exceedingly small very quickly. In fact, the right side is mostly background-dominated.

- The number of jets, which is shown in Figure 7.4. This is a discriminating variable simply because in the case of semileptonic events there are more jets coming from hadronic decay of one of the tops (the other top quark decays leptonically), whereas in dileptonic events the daughter jets of the tops are only the two b-jets coming the decays of both of the tops. Accordingly the semileptonic distribution is seen to be higher everywhere.

- The sum of the transverse momenta of all jets, shown in Figure 7.5. This is also expected to be discriminating because again more jets are found in semileptonic events due to hadronic decay of one of the tops, and fewer jets in dileptonic events due to leptonic decays of both tops, and as a result a sum over all transverse momenta of the jets leads to a discriminating distribution. Indeed, the figure shows that the semileptonic distribution is higher everywhere.

- The transverse momenta of the jet closest to the probe, which is shown in Figure 7.6. It is expected that jets having electrons inside $\Delta R = 0.4$ of themselves tend to be high-p_T jets, as the selections are targeting boosted top quark decays. The figure not only shows that the semileptonic distribution is higher everywhere, but that there is also a class of very low p_T jets closest to the probes.

- The fraction of the transverse momentum of the probe over that of the closest jet (Figure 7.7). This variable is discriminating because it is expected that real electrons coming directly from the W's which are produced from the tops tend to have higher p_T than background electrons The figure shows indeed that the low p_T region is dominated by semileptonic events.

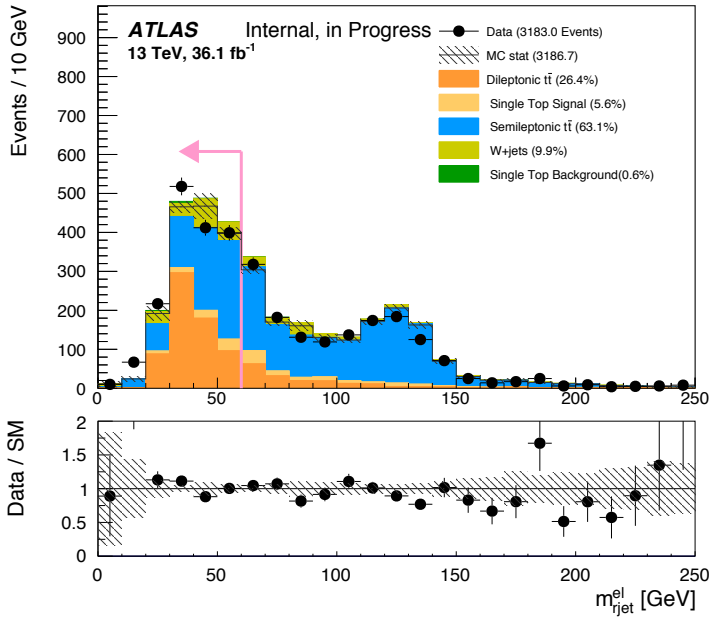


Figure 7.3: $m^{\mathrm{el}}_{\mathrm{rjet}}$. The semileptonic contribution is higher everywhere, especially on the right side of the distribution where there is little signal contamination.

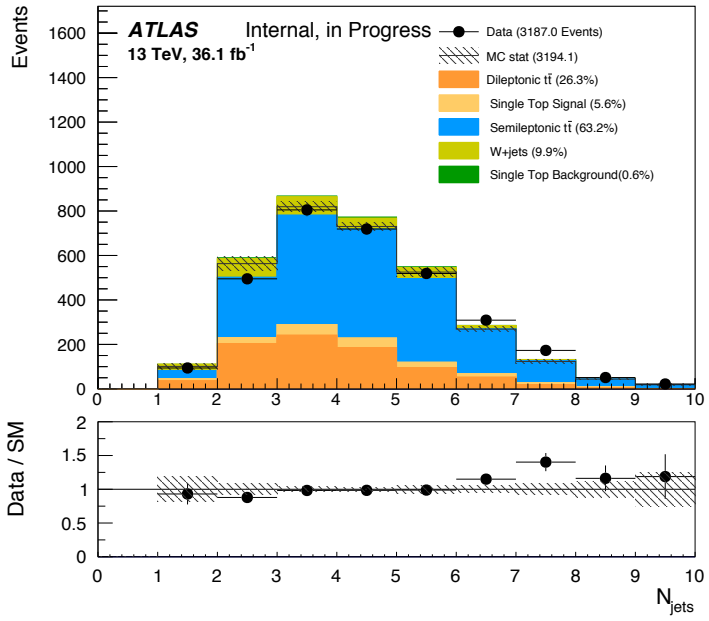


Figure 7.4: The number of jets. The semileptonic contribution is higher because of the hadronic decay of one of the tops.

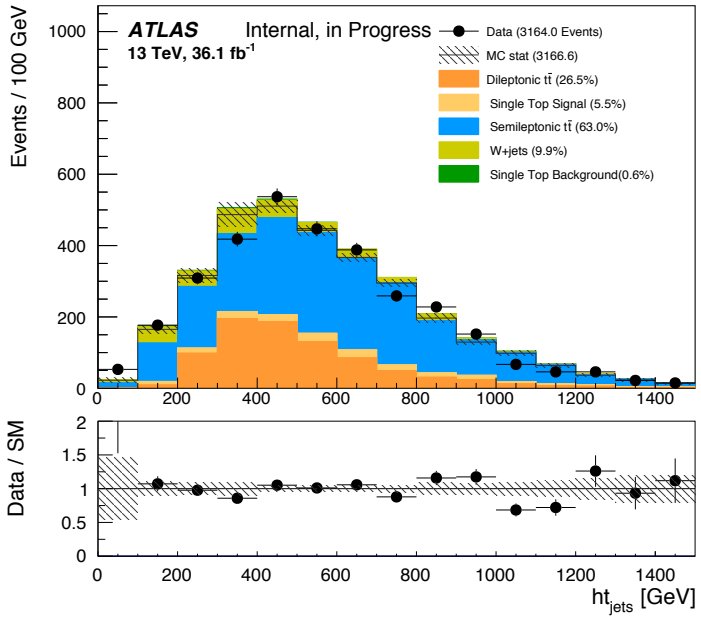


Figure 7.5: The sum of the transverse momenta of all jets.

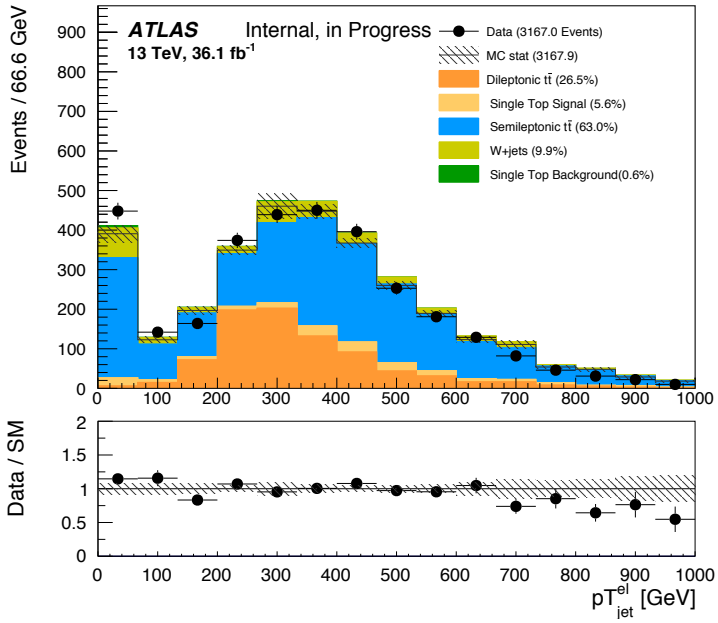


Figure 7.6: p_T of the jet closest to the probe electron.

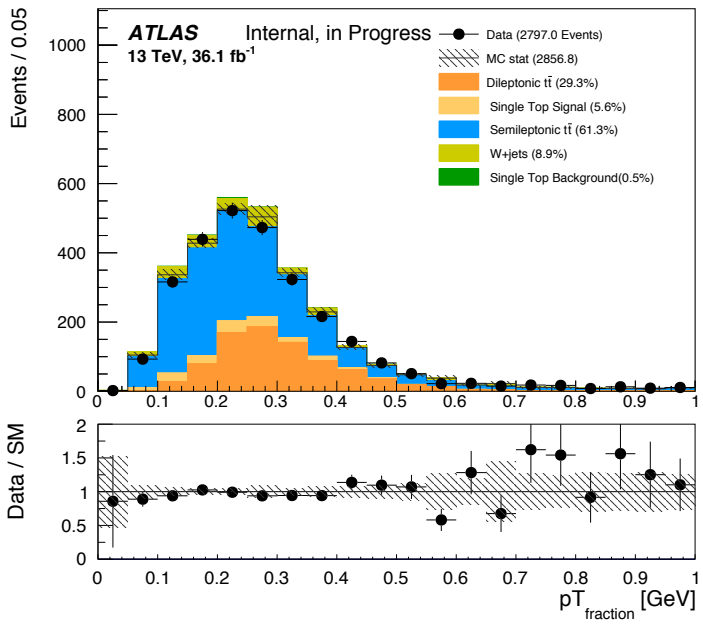


Figure 7.7: Fraction of the p_T of the probe electron over that of the closest jet. The lower p_T region is dominated by semileptonic events.

Further cuts to arrive at the signal region Of all the discriminating variables shown above, $m^{\text{el}}_{\text{rjet}}$ seems to be the most discriminating variable. In addition, its distribution shows two distinct regions, one abundant in signal electrons and one largely dominated by background electrons. As will be discussed later in the chapter, the region < 60 GeV will define the signal region where the identification efficiencies are measured, while the region > 60 GeV will define the control region for background estimation. With this in mind, we decided to apply cuts on the other discriminating variables to further remove the undesired background, while leaving $m^{\text{el}}_{\text{rjet}}$ untouched.

The cuts are as follows:

- ❑ MET > 25 GeV, to ensure that the QCD multi-jet background is negligible.
- ❑ The number of jets < 5 and sum of p_T of jets < 700 GeV, to remove semileptonic events (Figure 7.4 and 7.5).
- ❑ p_T of jet closest to the probe is between 150 GeV and 500 GeV, to remove semileptonic events (Figure 7.6) and at the same time make sure that boosted $t\bar{t}$ dilepton events are selected.
- ❑ $p_T(\text{probe})/p_T(\text{closest jet}) > 0.16$ (Figure 7.7).

The resulting distrution $m^{\text{el}}_{\text{rjet}}$ is shown in Figure 7.8.

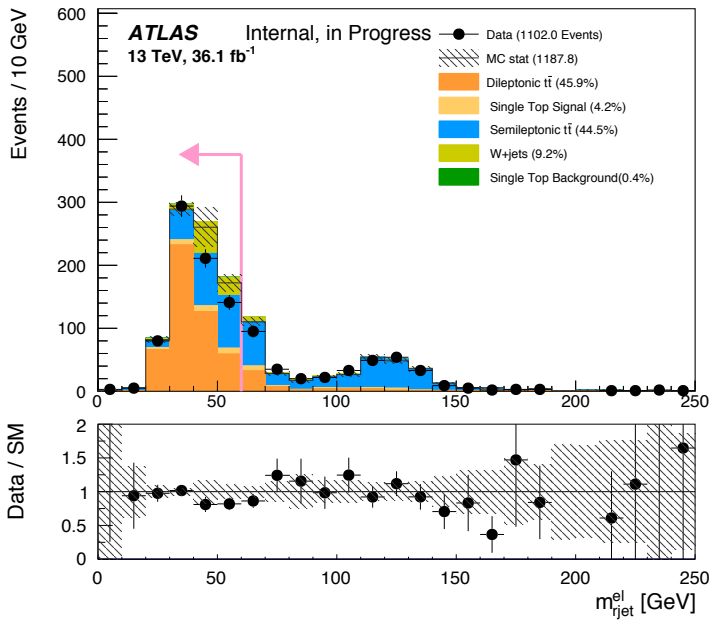


Figure 7.8: $m^{\text{el}}_{\text{rjet}}$ after the pre-selection cuts. The region < 60 GeV will define the signal region, and the region ≥ 60 GeV will define the control region for background estimation.

7.2.4 Background Estimation

We are looking to measure the identification efficiency for electrons inside jets at a particular operating point. If ϵ denotes such an efficiency, then it is the ratio of a numerator and a denominator (Section 7.2), both of which are expected to be contaminated with background electrons that will need to be estimated. We will write P as the number of electron candidates passing a particular ID operating point, B_P the number of background electrons passing the operating point, N the total number of reconstructed electron candidates in the sample, and B_N the number of background electrons present in the sample, so that the efficiency ϵ is in fact

$$\epsilon = \frac{P - B_P}{N - B_N} \tag{7.1}$$

The identification efficiencies for the Medium and Tight ID operating points are the mosted widely used points in ATLAS analyses, especially in SUSY analyses, and accordingly they are the only points that are measured in this thesis. Thus, either a Medium or a Tight ID selection will be applied on top of the sample that represents the denominator, thereby the numerator is obtained. Background estimations will consist of estimating the term B_P separately for Medium and Tight in the numerator, and estimating the common term B_N in the denominator.

Estimating B_P We expect background electrons to rarely pass the Medium or Tight ID points, and as a result we expect the term B_P to be very small in either

case. Thus B_P may be taken directly from simulation without significantly affecting the measurements.

Indeed, Figure 7.9 and 7.10 show the $m_{\text{rjet}}^{\text{el}}$ distributions for electrons that pass the Medium and Tight selections. They are obtained by applying either a Medium or a Tight ID selection on top of the set of selections that helps to arrive at the definition of the signal region (Section 7.2.3). As can be seen, the number of background electrons predicted by the simulation is indeed small in each case, accouting for only 0.3% of the total number in the Medium case (Figure 7.9), and 0.1% in the Tight case (Figure 7.10).

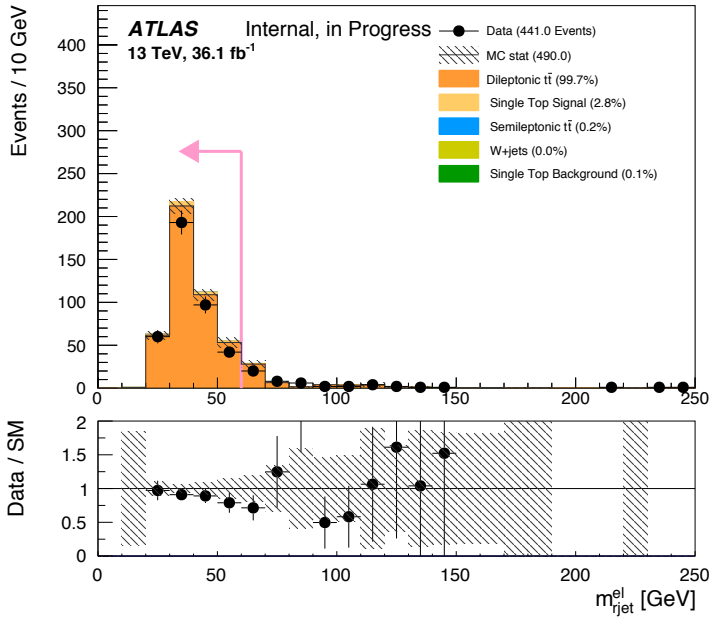


Figure 7.9: The distribution of $m_{\text{rjet}}^{\text{el}}$ for electrons passing the Medium operating point. Background electrons figure 0.3%.

Estimating B_N Since the term B_P in Formula 7.1 will be estimated by simulation prediction (for the Medium and Tight case separately), it remains to estimate the common term B_N that represents background contamination from fake electrons found in N. In order to estimate this term, we will use the background-dominated region $>$ 60 GeV in the distribution of $m_{\text{rjet}}^{\text{el}}$ (Figure 7.8), as well as the set of electrons that fail the Loose ID selection, which we will call antiloose electrons from here. Antiloose electrons will be used as a fake template, to be normalized to the data to predict the the background in the denominator, and we will therefore denote them by T.

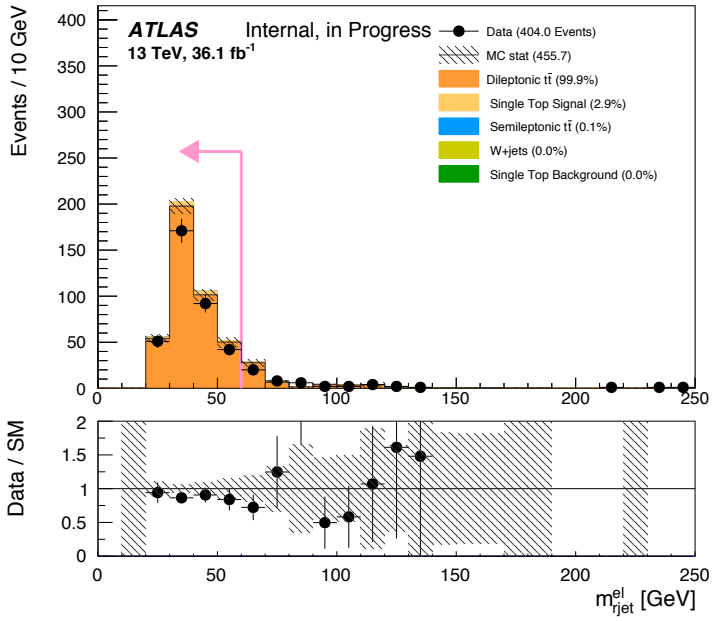


Figure 7.10: The distribution of $m^{\mathrm{el}}_{\mathrm{rjet}}$ for electrons passing the Medium operating point. Background electrons figure 0.1%.

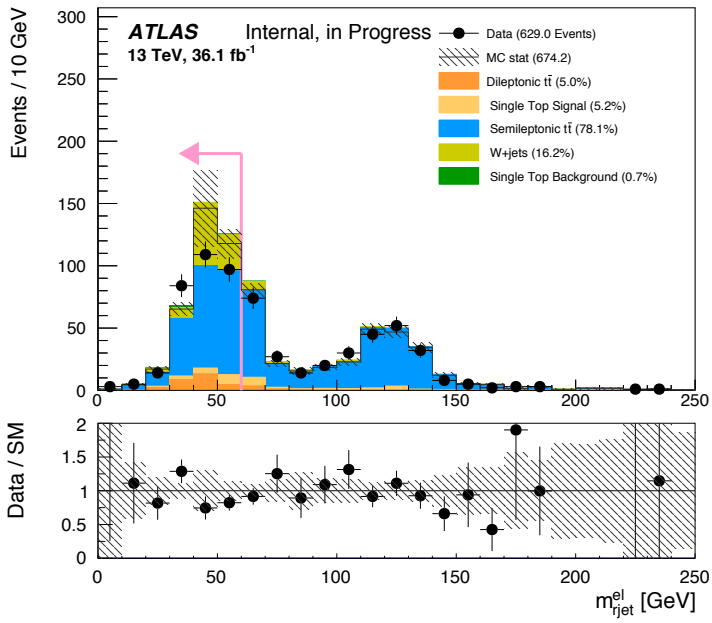


Figure 7.11: The distribution $m^{\mathrm{el}}_{\mathrm{rjet}}$ for electrons that fail the Loose ID point, also called antiloose electrons.

The $m^{\mathrm{el}}_{\mathrm{rjet}}$ distribution of T is shown in Figure 7.11. Simulation predicts it to be made up mostly of background electrons, in which semileptonic $t\bar{t}$ dominates, and there is about 10% of signal electron contamination in the distribution. Figure 7.12

shows T against the set of background part in N, namely B_N, both normalized to unity. It shows that T describes very well the shape of background electrons B_N, and therefore it is reasonable to estimate B_N using T. However, instead of using the simulation distribution we will use the corresponding data distribution, so that in the following T will mean the corresponding distributions of $m_{\text{rjet}}^{\text{el}}$ in data. This is to ensure that that typical systematic uncertainties associated with simulation distributions could be avoided, because these systematic uncertainties are not small and are also often complicated to obtain.

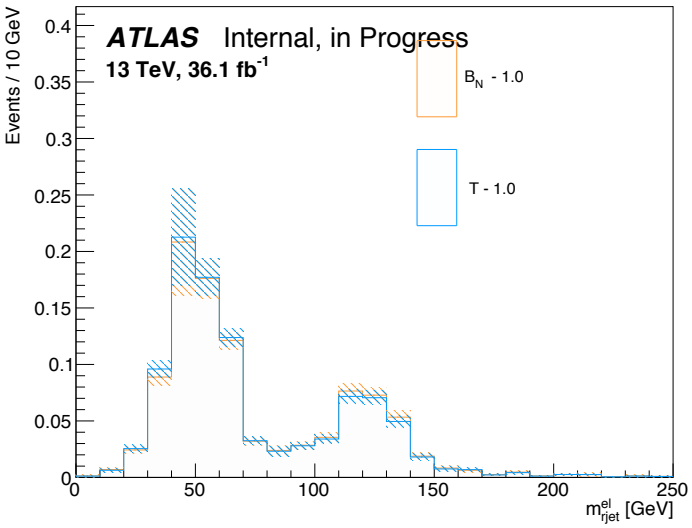


Figure 7.12: The distribution $m_{\text{rjet}}^{\text{el}}$ of B_N against that of T, normalized to unity. T describes very well B_N and therefore it is reasonable to estimate B_N using T.

In the actual measurement of the efficiency, we need to take into account the fact that T itself is contaminated with some signal electrons, as already mentioned above. Thus, let $\overline{T}$ be T minus this signal contamination, which we will obtain by taking the distribution of T in data minus the signal contamination predicted in the simulation distribution of T. Then, the background-dominated region will be used to find a normalization factor such that

$$B_N = \overline{T} \times \text{normalization factor},$$

where the normalization factor will be found as follows. To start, to any quantity in the signal region $\leq$ 60 GeV there corresponds a quantity in the background-dominated region $>$ 60 GeV, which will be denoted with a subscript $>$. Thus to N there corresponds $N_>$, and to $\overline{T}$ corresponds $\overline{T}_>$. The term $N_>$ is the set of reconstructed electron candidates in the background-dominated region. In order to use it in the following we will subtract any signal contamination it may have, the method of which will be described in the next section, and denote $\overline{N}_>$ to be the resulting term, i.e. $N_>$ minus the signal contamination. Then

$$\text{normalization factor} = \frac{\overline{N}_>}{\overline{T}_>}$$

so that

$$B_N = \overline{T} \times \frac{\overline{N}_>}{\overline{T}_>} \tag{7.2}$$

7.2.5 The Measurements of the Identification Efficiency

Formula 7.1, where B_P will be taken from simulation and B_N evaluated according to Formula 7.2, leads to the following formula for the efficiency:

$$\epsilon = \frac{P - B_P}{N - \overline{T} \times \frac{\overline{N}_>}{\overline{T}_>}}. \tag{7.3}$$

$\overline{T}$ will be the set of antiloose electrons, T, minus signal contamination. As already mentioned, Figure 7.12 shows that there is a 10% signal contamination expected in T. This signal contamination as predicted by simulation will be subtracted from T, resulting in $\overline{T}$. The term $\overline{T}_>$ will be obtained analogously in the background-dominated region of the $m^{\text{el}}_{\text{rjet}}$ distribution.

On the other hand, signal contamination in $N_>$ (Figure 7.8), from which $\overline{N}_>$ is obtained, is larger, and to reduce the contribution from the estimation of this signal contamination to the uncertainty in the efficiency we will use data directly for signal contamination subtraction. Specifically, from Figure 7.9 and 7.10 we expect negligible background electrons after a Medium or Tight ID is applied. This means P, and the corresponding quantity $P_>$ in the background-dominated region, is expected to be relatively free of background electrons. We will therefore write

$$\overline{N}_> = N_> - P_>/\epsilon$$

where the efficiency in 7.3, which is being measured, is used again here. The efficiency will be evaluated iteratively, until the change from one iteration to the next is less than 0.5%. The 0.5% will be taken as the uncertainty due to signal contamination subtraction in $N_>$.

The next section discusses in detail the treatment of statistical and systematic uncertainties.

7.2.6 Uncertainties

The efficiency is accompanied by statistical and systematic uncertainties, both of which are discussed in the following.

Statistical Uncertainties According to Formula 7.3, the efficiency will be measured according to the formula

$$\epsilon = \frac{P - B_P}{N - \overline{T} \times \frac{\overline{N}_>}{\overline{T}_>}}$$

where

- P is the number of electrons that pass Medium or Tight.
- B_P is background contamination due to fake electrons in P.
- N is the set of reconstructed electron candidates, and $\overline{N}_>$ the corresponding quantity in the background-dominated region minus signal contamination.
- $\overline{T}$ is the set of antiloose electrons minus signal contamination, and $\overline{T}_>$ the corresponding quantity in the background-dominated region.

Since N contains P, and $\overline{N}_>$ contains $\overline{T}_>$, the quantities in the formula are not all independent. We may remove the correlation between N and P by writing $N = P + F$, where F is the set of electrons that fail a particular ID point. Then

$$\epsilon = \frac{P - B_P}{P + F - \overline{T} \times \frac{\overline{N}_>}{\overline{T}_>}}$$

The correlation between $\overline{N}_>$ and $\overline{T}_>$ remains, and moreover F and $\overline{T}$ are also correlated, because in the Medium case or in the Tight case, F represents electrons failing Medium or Tight respectively, and since $\overline{T}$ represents electrons failing Loose (minus signal contamination), in each case $\overline{T}$ is a subset of F and there is accordingly a correlation.

In order to remove all the correlations and write the efficiency completely in terms of statistically independent quantities we will first multiply both the numerator and the denominator by $\overline{T}_>$, to write

$$\epsilon = \frac{(P - B_P)\overline{T}_>}{P\overline{T}_> + F\overline{T}_> - \overline{T} \times \overline{N}_>}.$$

Then we will add and subtract $\overline{T} \times \overline{T}_>$, to have

$$\begin{aligned}\epsilon &= \frac{(P - B_P)\overline{T}_>}{P\overline{T}_> + F\overline{T}_> - \overline{T} \times \overline{T}_> + \overline{T} \times \overline{T}_> - \overline{T} \times \overline{N}_>} \\ &= \frac{(P - B_P)\overline{T}_>}{P\overline{T}_> + (F - \overline{T})\overline{T}_> - (\overline{N}_> - \overline{T}_>)\overline{T}}\end{aligned}$$

The difference $F - \overline{T}$ represents the set of electrons that fail Medium or Tight but pass the Loose identification, and the difference $\overline{N}_> - \overline{T}_>$ represents the set of electrons that pass the Loose identification. If we treat each of the differences as a single term, and set $S = F - \overline{T}$ and $\overline{R}_> = \overline{N}_> - \overline{T}_>$ respectively, the efficiency becomes

$$\epsilon = \frac{(P - B_P)\overline{T}_>}{P\overline{T}_> + S\overline{T}_> - \overline{R}_> \times \overline{T}} \tag{7.4}$$

which is now a function of six independent quantities, $\epsilon = \epsilon(P, B_P, \overline{T}_>, S, \overline{R}_>, T)$. The statistical uncertainty of the efficiency then follows the standard error propagation formula,

$$\Delta\epsilon^2 = \left(\frac{\partial\epsilon}{\partial P}\right)^2 \Delta P^2 + \cdots + \left(\frac{\partial\epsilon}{\partial T}\right)^2 \Delta T^2 \tag{7.5}$$

Let A denote the numerator in Formula 7.4 and B the denominator. The terms in the formula above are then

$$\frac{\partial\epsilon}{\partial P} = \frac{B\overline{T}_> - A\overline{T}_>}{B^2}, \quad \frac{\partial\epsilon}{\partial B_P} = \frac{-B\overline{T}_>}{B^2}, \quad \frac{\partial\epsilon}{\partial\overline{T}_>} = \frac{B(P - B_P) - A(P + S)}{B^2},$$

$$\frac{\partial\epsilon}{\partial S} = \frac{-A\overline{T}_>}{B^2}, \quad \frac{\partial\epsilon}{\partial\overline{R}_>} = \frac{AT}{B^2}, \quad \frac{\partial\epsilon}{\partial\overline{T}} = \frac{A\overline{R}_>}{B^2}.$$

The statistical uncertainty contribution due to the term B_P will be treated as negligible, as this term comes from simulation. Then, among the remaining terms, only P and S are present in the signal region that are not used for background estimation, and as a result the statistical uncertainty of the efficiency will be taken from the contributions of these two terms. The contributions to the uncertainty from other terms will be taken as contributions to the total systematic uncertainty, which is discussed below.

Systematic Uncertainties Contributions to the total systematic uncertainty from different sources will be added in quadrature. The sources are listed below.

- ❏ The variation of the signal region, i.e. instead of marking the signal region at 60 GeV, we may mark it at 50 or 80 GeV, the asymmetry is due to the fact that signal distributions on both sides of the 60 GeV mark are not equal in equal intervals.
- ❏ The term B_P which represents the background contamination in P is taken from simulation and may be varied up and down. To be conservative, we have decided to make a 50% variation.
- ❏ The uncertainty due to signal contamination subtraction from T and $T_>$, from which result $\overline{T}$ and $\overline{T}_>$, may be obtained by conservatively varying the signal contamination 25% up and down.
- ❏ The template T, which is the distribution of antiloose electrons, may be replaced by the distribution of antiloose electrons in events with exactly 2 b-jets.
- ❏ In addition, the contributions from the counting of $\overline{T}_>$, $\overline{R}_>$, and $\overline{T}$ in Formula 7.4 are treated as contributions to the total systematic uncertainty as well.

7.3 Identification Efficiencies

In this section the integrated efficiencies as well as the binned efficiencies for the Medium and Tight operating points are presented, along with the associated uncertainties.

7.3.1 Integrated Efficiencies

The identification efficiencies for electrons inside jets are measured for the Medium and Tight operating points; they are evaluated iteratively according to Formula 7.3, which is

$$\epsilon = \frac{P - B_P}{N - \overline{T} \times \frac{\overline{N}_>}{\overline{T}_>}}.$$

where

$$\overline{N}_> = N_> - P_>/\epsilon$$

The integrated efficiencies are presented in the following, along with the associated uncertainties.

7.3.1.1 The efficiencies

The efficiencies, as well as the total statistical and systematic uncertainties, are listed in Table 7.2. It is seen that the efficiency is higher for Medium than for Tight, consistent with expectation. The statistical uncertainties are slightly larger for Tight, also consistent with expectation, as the stats for Tight is slightly less than that for Medium.

	MEDIUM **0.870**	TIGHT **0.784**
Statistical Uncertainty	±0.017	±0.019
Systematic Uncertainty	±0.031	±0.020

Table 7.2: Efficiencies, Statistical and Systematic Uncertainties in Data for the Medium and Tight operating points.

The relevant quantities in Formula 7.3 that are used to compute the efficiencies in data are listed in Table 7.3.

	MEDIUM	TIGHT
P	392	356
B_P	1.47	0.40
N	734	734
$\overline{N_>}$	368	368
$P_>$	49	48
$\overline{T}$	267.35	267.35
$\overline{T_>}$	292.52	292.52

Table 7.3: The relevant quantities for computing the efficiencies according to Formula 7.3.

The efficiencies in simulation for the Medium and Tight operating points are also computed and are listed in Table 7.4 along with the statistical uncertainties. The relevant quantites are displayed in Table 7.5.

MEDIUM **0.871**	TIGHT **0.807**
± 0.010	± 0.011

Table 7.4: The integrated efficiencies in simulations for the Medium and Tight operating points, along with the associated statistical uncertainties.

	MEDIUM	TIGHT
Numerator		
Dilepton	435.45	403.31
Single top signal	12.39	11.81
Denominator		
Dilepton	484.53	484.53
Single top background	29.66	29.66

Table 7.5: The relevant quantities for computing the efficiencies in simulations for the Medium and Tight operating points.

7.3.1.2 Statistical Uncertainties

As has been discussed in Section 7.2.6, when evaluating the efficiencies in data, the quantities in the signal region that are not used for background estimation are P and S, and the statistical uncertainty of the efficiency is taken from the contributions of these two terms, computed according to Formula 7.4 and shown in Table 7.2. Of the two, the contribution from S is the dominant one; the contribution from P is small ($< 0.5\%$).

7.3.1.3 Systematic Uncertainties

The total systematic uncertainty receives contributions from different sources, as discussed in Section 7.2.6. The individual contributions are shown below.

Contributions from $\overline{T}_>$, $\overline{R}_>$, and $\overline{T}$ The contributions to the total systematic uncertainty that come from the counting of $\overline{T}_>$, $\overline{R}_>$, and $\overline{T}$ in Formula 7.4 are listed in Table 7.6 below. They each contributes $< 1\%$ to the total uncertainty. The relevant quantities used for the calculations are listed in Table 7.7. The term $\overline{R}_>$ is computed as the difference $\overline{N}_> - \overline{T}_>$ (Section 7.2.6), where $\overline{N}_> = N_> - P_>/\epsilon$ as discussed in Section 7.2.5.

	MEDIUM	TIGHT
$\Delta\overline{R}_> = \sqrt{\overline{R}_>}$	± 0.008	± 0.006
$\Delta\overline{T} = \sqrt{\overline{T}}$	± 0.002	± 0.001
$\Delta\overline{T}_> = \sqrt{\overline{T}_>}$	± 0.002	± 0.001

Table 7.6: Contributions to the total systematic uncertainty from the individual sources.

	MEDIUM	TIGHT
$\overline{R}_>$	19.11	14.24
$\overline{T}$	267.35	267.35
$\overline{T}_>$	292.52	292.52

Table 7.7: The quantities $\overline{R}_>$, $\overline{T}_>$, and $\overline{T}$ that are used to compute the contributions of these terms to the total systematic uncertainty according to Formula 7.4. The term $\overline{R}_>$ is computed as the difference $\overline{N}_> - \overline{T}_>$ (Section 7.2.6).

Contribution from B_P The contribution to the total systematic uncertainty from varying the term B_P (Formula 7.4) is shown below in Table 7.8. The total contribution is taken to be the sum in quadrature of the two variations. The countings of B_P in each of the variations are shown in Table 7.9. We see that the contribution from B_P is very small.

	MEDIUM	TIGHT
$B_P \times 1.5$	0.002	0.000
$B_P \div 1.5$	0.001	0.000
B_P total contribution	0.002	0.000

Table 7.8: Contributions to the total systematic uncertainty from the term B_P.

	MEDIUM	TIGHT
Semileptonic $t\bar{t}$	1.13	0.40
Background single top	0.34	0.00
W+jets	0.00	0.00
$B_P \times 1.5$	2.2	0.61
$B_P \div 1.5$	0.98	0.27

Table 7.9: Contributions to the systematic uncertinty from B_P.

Contribution from signal contaminations in T and $T_>$ The contributions due to signal comtamination in the terms T and $T_>$ are listed in Table 7.10. These are a major sources of contribution to the total systematic uncertainty. The signal contaminations in the two terms T and $T_>$ are varied simultaneously, either up or down by 25%. The countings obtained from the variations are shown in Table 7.11.

	MEDIUM	TIGHT
$\times 1.25$	0.014	0.012
$\div 1.25$	0.010	0.010
Total contribution	0.017	0.015

Table 7.10

	MEDIUM	TIGHT
T	312	312
Dilepton contamination	28.43	28.43
Single top signal contamination	16.23	16.23
$\overline{T} \times 1.5$	?	?
$\overline{T} \div 1.5$	?	?
$T_>$	317	317
Dilepton contamination	5.45	5.45
Single top signal contamination	19.03	19.03
$\overline{T}_> \times 1.5$	?	?
$\overline{T}_> \div 1.5$	?	?

Table 7.11

Contribution from changing the template to antiloose+2b The contribution to the total systematic uncertainty from the template change is shown in Table 7.12. The template change leads to changes in quantities $\overline{T}$ and $\overline{T}_>$, which are listed in Table 7.13.

	MEDIUM	TIGHT
antiloose+2b	0.008	0.007

Table 7.12: Contributions to the total systematic uncertainty from changing the template to antiloose+2b

	MEDIUM	TIGHT
$\overline{T}$	106.03	106.03
$\overline{T}_>$	114.49	114.49

Table 7.13

Contributions from re-marking the signal region The contribution to the total systematic uncertainty from re-marking the signal region is shown in Table 7.14. It is another major contribution to the total systematic uncertainty. The signal region is marked, in place of at 60 GeV, at 50 GeV and 80 GeV in turn. The total contribution is computed as a sum of quadrature of the two individual contributions. The relevant quantities used for the calculations of the efficiencies are listed in Table 7.15 and 7.16.

	MEDIUM	TIGHT
50 GeV	0.010	0.004
80 GeV	0.020	0.001
Total contribution	0.022	0.010

Table 7.14: Contributions to the total systematic uncertainty re-marking the signal region at 50 GeV and 80 GeV.

	MEDIUM	TIGHT
P	350	314
B_P	1.47	0.40
N	593	593
$\overline{N_>}$	509	368
$P_>$	91	90
$\overline{T}$	182.76	182.76
$\overline{T_>}$	377.11	377.11

Table 7.15: The relevant quantities to compute the efficiencies for the Medium and Tight operating points when marking the signal region at 50 GeV.

	MEDIUM	TIGHT
P	420	384
B_P	1.56	0.43
N	734	734
$\overline{N_>}$	238	238
$P_>$	21	20
$\overline{T}$	355.54	355.54
$\overline{T_>}$	204.33	204.33

Table 7.16: The relevant quantities to compute the efficiencies for the Medium and Tight operating points when marking the signal region at 80 GeV.

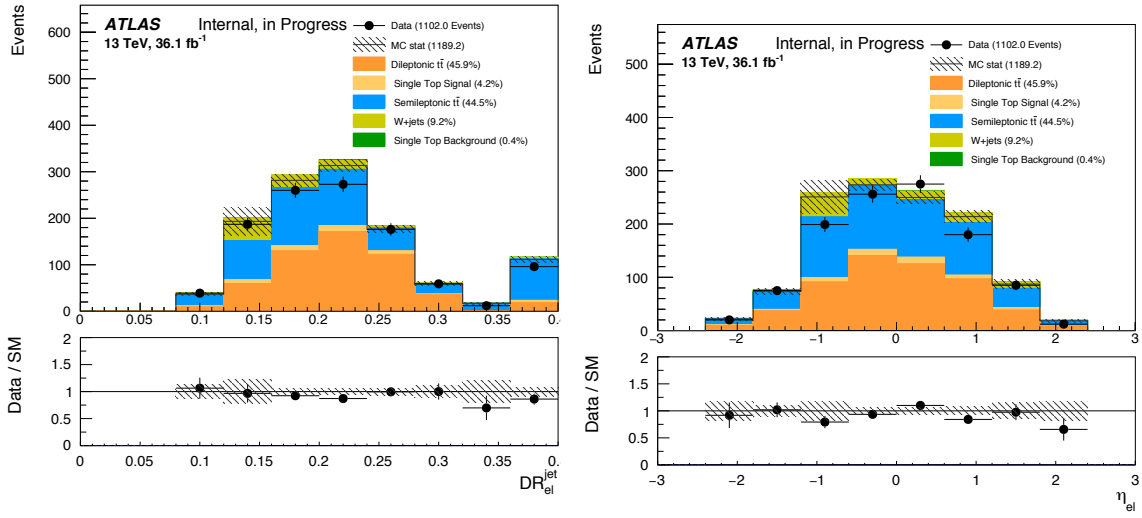
7.3.2 Efficiencies in Bins

In addition to the integrated efficiencies, binned efficiencies are also measured, to check the possible dependencies of the efficiencies on certain variables. The variables and their associated binnings are:

- p_T of the probe, in five bins
  - 30-60 GeV,
  - 60-80 GeV,
  - 80-110 GeV,
  - 110-140 GeV,
  - $>$ 140 GeV.
- $|\eta|$ of the probe, in five bins
  - 0.0-0.3,
  - 0.3-0.6,
  - 0.6-0.9,
  - 0.9-1.3,
  - $>$ 1.3.
- ΔR between the probe and the closest overlapping jet, in five bins
  - 0.0-0.15,
  - 0.15-0.19,
  - 0.19-0.23,
  - 0.23-0.27,
  - 0.27-0.4.
- p_T of the closest overlapping jet, in five bins

- 150-220 GeV,
- 220-280 GeV,
- 280-340 GeV,
- 340-400 GeV,
- 400-500 GeV.

The distributions of these variables are shown in Figure 7.13 and 7.14. The binned efficiencies are shown in Figure 7.15 and 7.16.



(a) ΔR

(b) $|\eta|$ of the probe

Figure 7.13: The distributions of ΔR between the probe and the closest overlapping jet and $|\eta|$ of the probe.

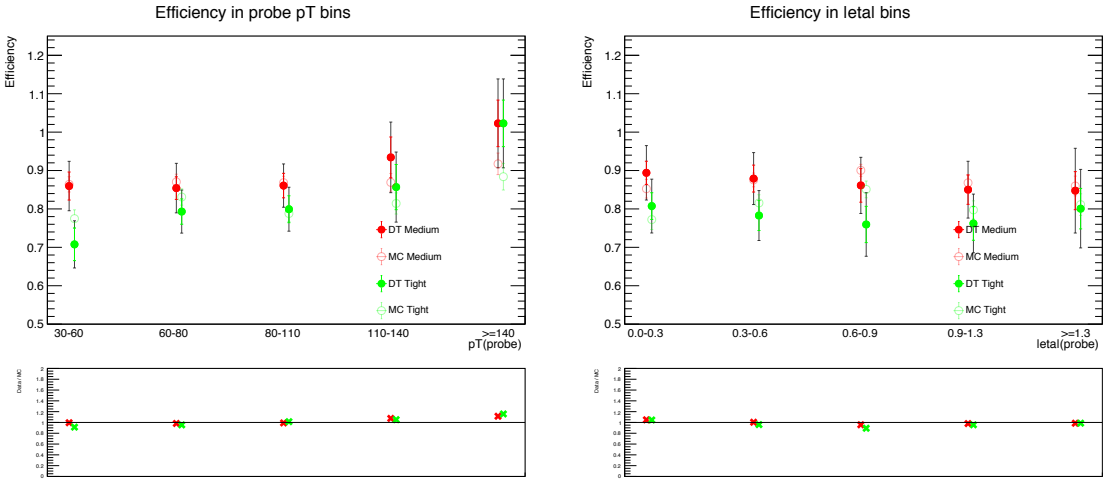
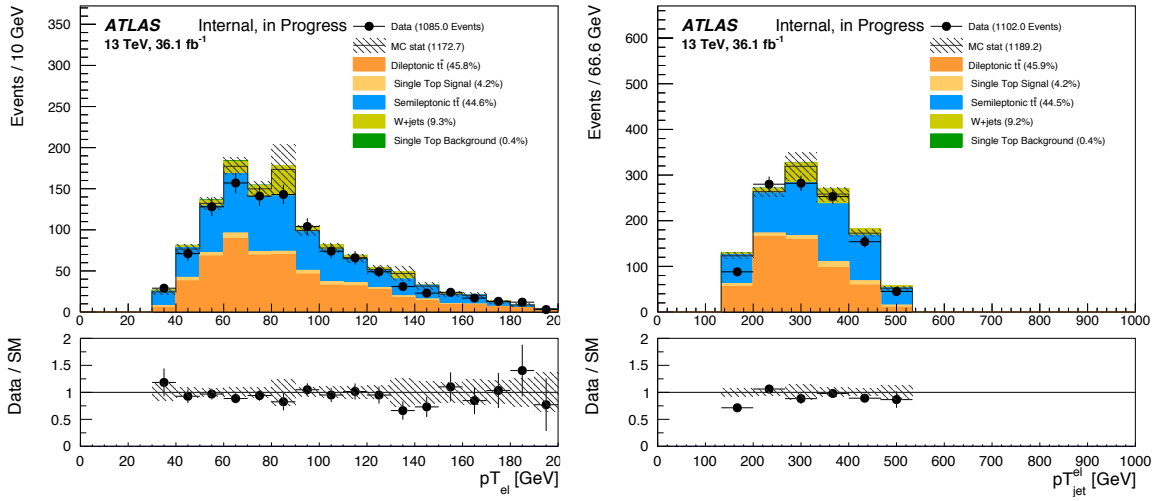


Figure 7.15: The binned efficiencies in p_T of the probe as well as in $|eta|$ of the probe.



(a) p_T of the probe (b) p_T of the closest overlapping jet

Figure 7.14: The distributions of p_T of the probe and p_T of the closest overlapping jet.

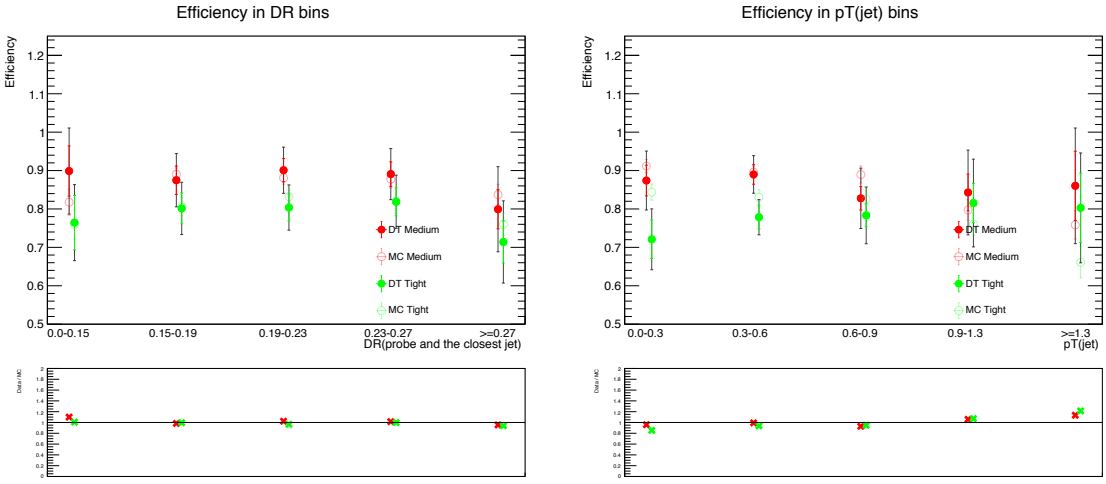


Figure 7.16: The binned efficiencies in ΔR between the probe and the closest jet, as well as in p_T of the closest overlapping jet.

7.4 Conclusions

This chapter describes the work to measure the identification efficiencies for in-jet electrons. It was the first attempt to perform such a measurement since Run 2 began. The measurement used the data collected in the period 2015-2016, at 13 TeV center-of-mass and totalled 36.47 fb^{-1} in integrated luminosity. A sample of electrons for the measurements was obtained by selecting boosted $t\bar{t}$ dilepton ($e\mu$) events. Background estimations used both simulations and data, and the efficiencies were evaluated iteratively. The efficiencies were measured for the Medium and Tight operating points, both on data and simulation. Efficiencies in bins were also measured. In all of the results, the efficiencies predicted by simulation are quite close to those obtained from the measurements on data.