

# 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 center-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 [85–88]. At ATLAS, many SUSY searches involve supersymmetric particles that decay into the Standard Model top quarks and we expect, with higher center-of-mass energy, higher sensitivity to more massive supersymmetric particles and consequently an increase in the production of high  $p_T$  top quarks. Since the top quarks also decay — essentially all the time — into a  $W$  boson and a  $b$  quark, we in turn 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, to stay close to each other (Figure 7.1).

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 called in-jet electrons, using  $t\bar{t}$  events. Prior to the work in this chapter, there was no attempt to measure the identification efficiencies for in-jet electrons in a  $t\bar{t}$  topology. 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 and presents the measured efficiencies. Section 7.3 presents some conclusions.

The data used for this chapter was collected in the period 2015–2016 at 13 TeV center-of-mass and corresponded to an integrated luminosity of  $36.5 \text{ fb}^{-1}$ .

### 7.1 Motivation

Prior to Run 2, ATLAS center-of-mass energy 7–8 GeV allows limited sensitivity to high mass resonances. Because many beyond-Standard-Model particles are predicted to decay into the Standard Model top quarks, the limited sensitivity reduces the chance in which we could expect boosted top quark decays. Such decays, however,

are expected to become significant as the centre-of-mass of the LHC reached 13 TeV starting from Run 2. In a boosted top quark decay scenario, the produced particles, which in this case are the daughters of the  $W$  and the  $b$  quark that come from the top quark, are found close to each other [90–92]. Figure 7.2 shows the angular distance  $\Delta R$  (Formula 3.1) between the  $W$ 's and the  $b$ -quarks as a function of the top  $p_T$ , in the context of a hypothetical particle  $Z'$  with mass  $m_{Z'} = 1.6$  TeV [92] that decays into a  $t\bar{t}$  pair. Also shown in the same figure is 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.

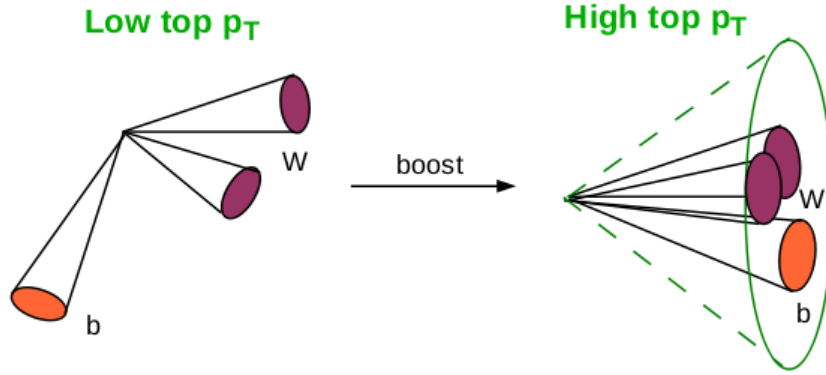


Figure 7.1: An illustration of low  $p_T$  top quark decay (left) and boosted top decay (right) of a high  $p_T$  top quark. In the case of high  $p_T$  top quark decay the daughter particles of the top quark, which include the daughter particles of the  $W$  and the  $b$  quark, are expected to be found close to each other [89].

Leptonic boosted top quark decay is also an important channel in searches for beyond-Standard-Model particles that decay into the Standard-Model top quarks. Table 7.1 shows a measurement of the fraction of in-jet electrons over signal electrons as a function of the top quark  $p_T$ , at truth-level. 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

$$\Delta R < \min(0.4, 0.04 + 10 \text{ GeV}/p_T), \quad (7.1)$$

where the overlap removal is required to keep the overlapping  $b$ -jets (Chapter 6, Section 6.3). As is shown, more and more electrons are found inside jets as the top quark  $p_T$  increases. 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 searching for heavy beyond-Standard-Model particles decaying into lighter sparticles, such as the gluinos and stops that decay into neutralinos, in which the final state involve the Standard Model top quarks, were able to increase signal acceptances considerably if in-jet electrons were selected [85, 86].

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

Table 7.1: The fraction of in-jet electrons over the number of signal 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  $p_T$ .

1807 This chapter develops a method and performs the initial measurements for the  
 1808 identification efficiencies of electrons found inside  $\Delta R = 0.4$  of jets. The measure-  
 1809 ments for electrons outside jets are done by the ATLAS Egamma group [38, 47].

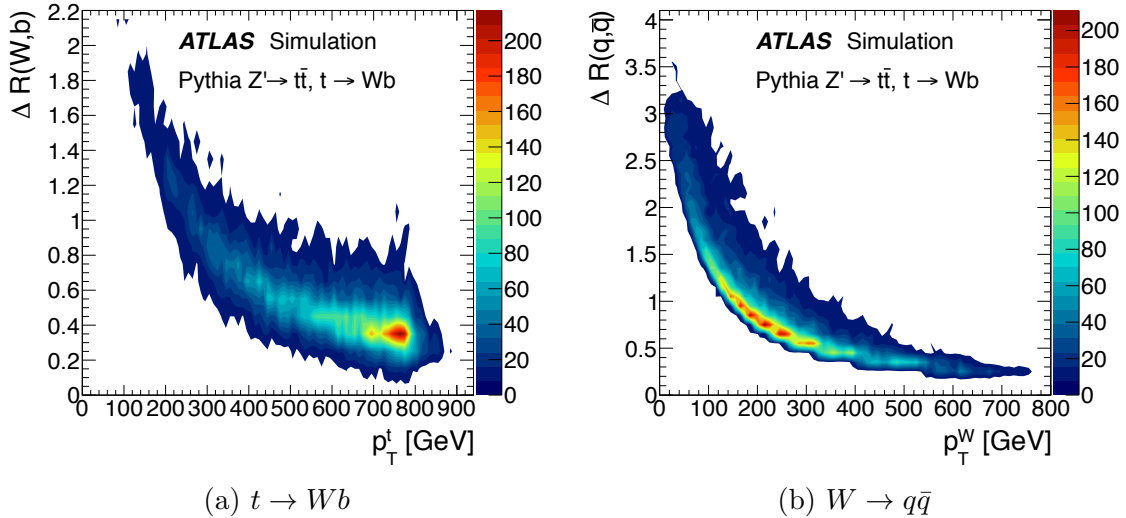


Figure 7.2: (7.2a) The angular distance  $\Delta R$  between the  $W$ 's and the  $b$  quarks as a function of the top quark  $p_T$  simulated PYTHIA [57], in the context of a hypothetical particle  $Z'$  ( $m_{Z'} = 1.6$  TeV) that decays into a  $t\bar{t}$  pair. At high top quark  $p_T$  a non-negligible fraction of the distances is seen to be very small. (7.2b) The angular distance between two light quarks from  $t \rightarrow Wb$  decay as a function of the  $p_T$  of the  $W$  boson [92].

## 7.2 Method

The method that is used to measure the identification efficiencies for in-jet electrons is discussed in detail in Section 7.2.1, 7.2.2, and 7.2.3. Background estimations is described in Section 7.2.4.

### 7.2.1 Boosted Dilepton $e\mu$ Events

In order to measure the identification efficiencies for in-jet electrons, a sample of reconstructed electrons (Chapter 6, Section 6.3) inside high- $p_T$  jets was obtained by selecting boosted  $t\bar{t}$  dilepton ( $e\mu$ ) events. This is expected to result not only in a very pure  $t\bar{t}$  sample, but also in a topology close to that of many SUSY and other beyond-Standard-Model searches. In contrast, the standard method for measuring electron identification efficiencies, the tag-and-probe method supported by the Egamma group at ATLAS [47], makes use of  $Z \rightarrow e^+e^-$  events for high-energy electrons ( $E_T > 10$  GeV). Even though a clean sample of electrons may be obtained in a relatively straightforward way by selecting events around the  $Z$  mass peak, we expect, if electrons inside high  $p_T$  jets are required, a sample unrepresentative of events with a boosted topology and limited in statistics.

Thus, a sample of electrons was obtained by selecting a hard muon and a reconstructed electron with only a  $p_T$  requirement applied, which will also be referred to below simply as reconstructed electrons. The efficiency at a particular identification operating point (Section 4.2.2) is defined by the ratio

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

Both the numerator and the denominator are contaminated with background electrons which require a careful estimate (Section 7.2.4), particularly because background electrons are expected to reside primarily inside jets.

### 7.2.2 Data and Monte Carlo Samples

The data used for this chapter was collected in the period 2015-2016 and corresponds to an integrated luminosity of  $36.5 \text{ fb}^{-1}$ . The following simulation samples (Chapter 6, Section 6.2), at 13 TeV centre-of-mass energy, are used:

- $t\bar{t}$  events from the Powheg+Pythia generator. As a hard muon will be required in the sample in which the identification efficiencies are measured (Section 7.2.3), these events naturally partition into either a dileptonic set ( $e\mu$ ) when truth-level electrons are present inside high- $p_T$  jets, or a semileptonic set otherwise. The latter, with jets from the fully hadronic decay of one of the top quarks constituting a source of background electrons, is expected to be a dominant background.
- $W$ +jets, which will be used as a background. As the  $W$  boson may produce a hard muon, the presence of jets make these events a source of background events to signal dilepton  $e\mu t\bar{t}$  events.

- Single top events, which include the  $Wt$  production as well as the  $s$ -channel and  $t$ -channel productions. The  $Wt$  production is treated as a source of signal electrons, since it contains a pair of  $W$  bosons that can decay to a prompt  $e\mu$  pair, whereas the remaining two productions each contains only one  $W$  boson and as a result cannot produce a prompt  $e\mu$  pair.

### 7.2.3 Signal Region

The kinematic region in which the measurement of the identification efficiencies is performed is called the signal region. It is defined after the following preliminary selections, which are called the pre-selection cuts and aimed at isolating dilepton  $e\mu$   $t\bar{t}$  events, are applied.

#### 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, where the overlapping  $b$ -jets are kept (Formula 7.1).
- Events with bad or cosmic muons are removed. 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 of which cases are sources of bad muons. Events with these muon candidates, along with those having muons from cosmic rays, are rejected.
- Exactly one identified muon and  $\geq 1$  electrons inside jets are required for each event, where
  - The muon is required to have  $p_T > 30$  GeV,  $d_0/\sigma(d_0) < 3.0$ , and  $z_0 < 0.5$  in terms of the transverse impact parameter and the longitudinal impact parameter. It must also have  $\text{ptvarcone30}/p_T < 0.06$ , where  $\text{ptvarcone30}$  is defined as the scalar sum of the momenta of the tracks with  $p_T > 1$  GeV in the cone with  $\Delta R < \min(10\text{GeV}/p_T, 0.3)$ , and must be a muon that has been triggered.
  - The electrons must have  $p_T \geq 30$  GeV, which is a common cut in most analyses where in-jet electrons are used, and must overlap within  $\Delta R < 0.4$  with some jets. There could be more than one electron present in the event, however the leading  $p_T$  electron will be used,
- $\geq 1$   $b$ -tagged jet, instead of exactly 2  $b$ -tagged jets as is usually expected in  $t\bar{t}$  events, since we are selecting events with electrons inside jets and the  $b$ -tagging efficiency may suffer because the tracks of the electron, which is expected to originate from the interaction point, may confuse the  $b$ -tagging algorithm.

These cuts result in a set of 3183 events with one hard muon and at least one electron candidate found inside some jet. In the following, we discuss several variables that have been found to be discriminating, along with their distribution plots. Simulation shows an expected 814.2 dilepton events and 178.5 single top  $Wt$  production events. On the other hand, the prominent source of background comes from semileptonic events, predicted to be 2010.8, whereas  $W$ +jets and single top  $s$ -channel and  $t$ -channel constitute two small sources of background, predicted to be 315.5 and 19.1 respectively.

- The mass of the large radius jet that overlaps with the probe electron, shown in Figure 7.3 and denoted  $m_{\text{rjet}}^{\text{el}}$ . 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 expected to be more massive, as it picks up the masses of the jets from the hadronic decay of one of the top quarks. In dileptonic events, on the other hand, there are fewer jets due to leptonic decays of both of the top quarks, and in addition the neutrino that accompanies the electron may reduce the visible mass of the reconstructed large radius jet. As is shown in the figure, the higher mass region is dominated by background events.
- The number of jets, which is shown in Figure 7.4 and denoted  $N_{\text{jet}}$ . Three jets are expected from a fully hadronic decaying top quark, as compared to only one jet from a semileptonic decay, and as a result semileptonic events, in which one top quark decays hadronically and one decays semileptonically, is expected to have a greater number of jets than dileptonic events, where both jets decay hadronically. In the figure, the semileptonic distribution is seen to be higher everywhere.
- The sum of the transverse momenta of all jets, shown in Figure 7.5. As above, a larger number of jets is expected in semileptonic events due to the fully hadronic decay of one of the top quarks, and in dileptonic events fewer jets are expected because of leptonic decays of both of the top quarks. Consequently a sum over all transverse momenta of the jets is expected to lead to a discriminating distribution. As is seen in the figure, the semileptonic distribution is higher everywhere.
- The transverse momenta of the jet closest to the probe, which is shown in Figure 7.6. This variable allows the removal of low  $p_T$  jets overlapping with background electrons.
- The fraction of the transverse momentum of the probe electron over that of the closest jet (Figure 7.7). We expect real electrons from the  $W$ 's produced from the top quarks to have higher  $p_T$  than background electrons. In the figure, the low  $p_T$  region can be seen to be dominated by semileptonic events.

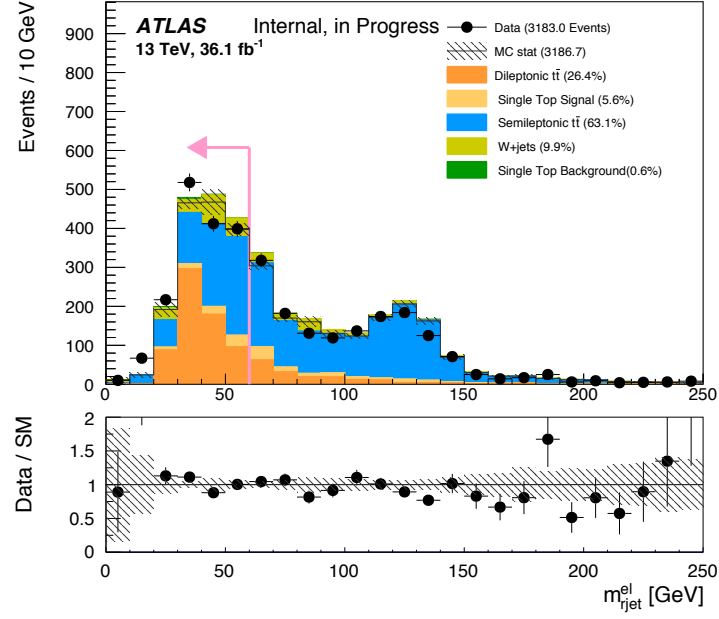


Figure 7.3:  $m_{rjet}^{el}$ . The semileptonic contribution is higher everywhere, especially on the right side of the distribution where there is little signal contamination.

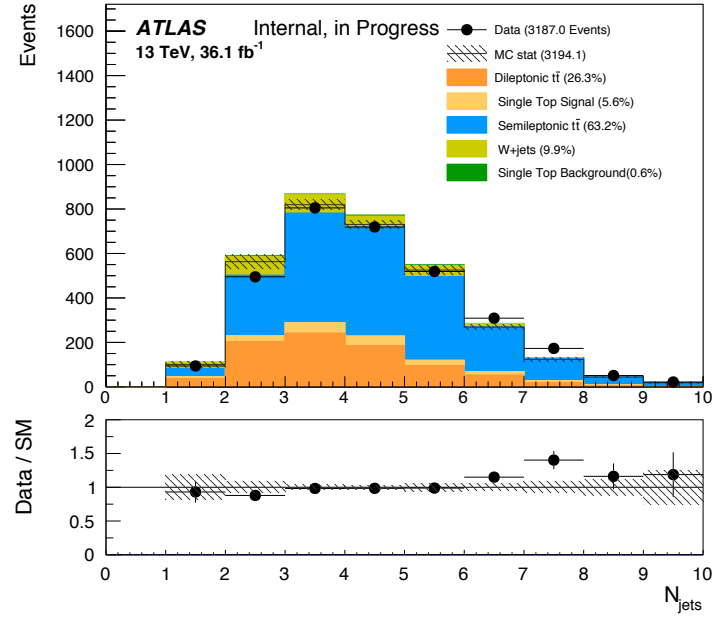


Figure 7.4:  $N_{jets}$ . The semileptonic contribution is higher because of the hadronic decay of one of the tops.

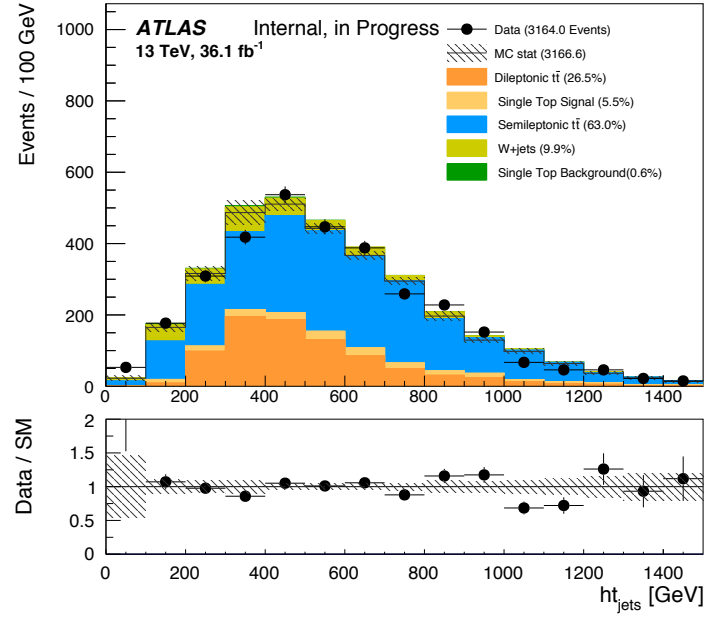
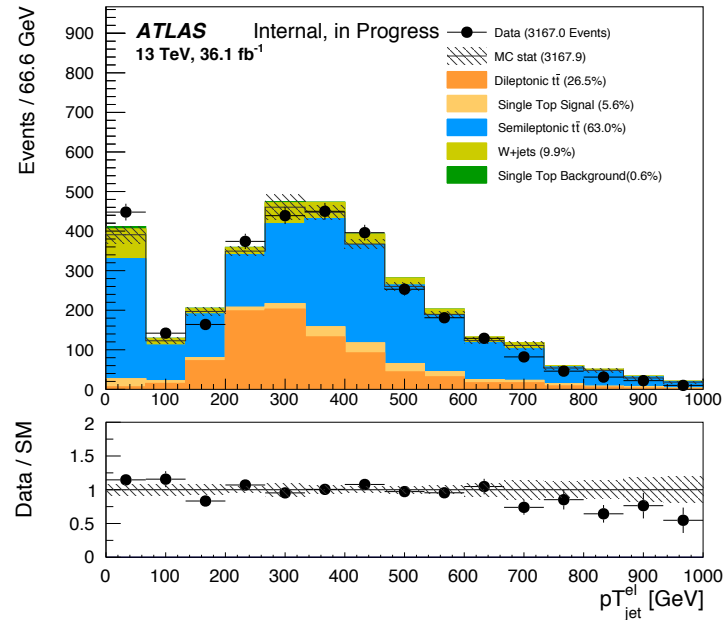


Figure 7.5: The sum of the transverse momenta of all jets.

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



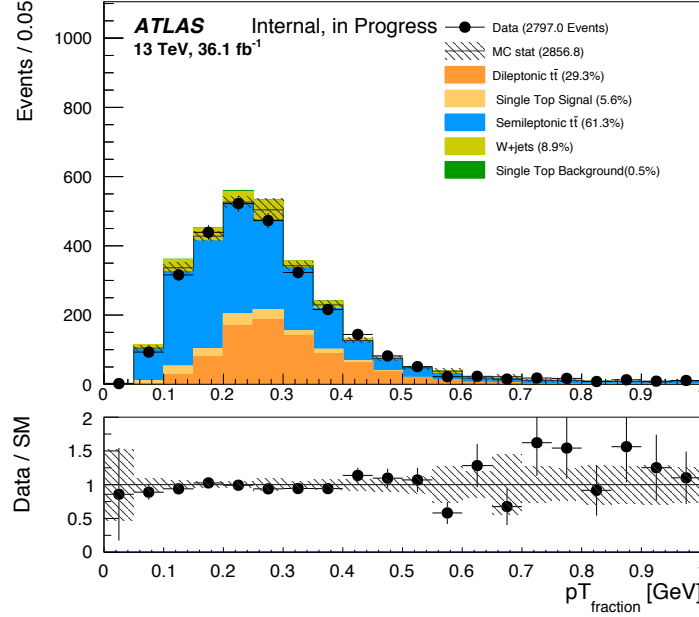


Figure 7.7: Fraction of the  $p_T$  of the probe electron over that of the closest jet. The lower  $p_T$  region is dominated by semileptonic events.

**Further cuts to arrive at the signal region** Of all the discriminating variables shown above,  $m_{\text{rjet}}^{\text{el}}$  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, and 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  $m_{\text{rjet}}^{\text{el}}$  untouched.

The cuts are as follows:

- Missing transverse momentum  $E_T^{\text{miss}} > 25$  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 distribution  $m_{\text{rjet}}^{\text{el}}$  is shown in Figure 7.8. There are 1102 events, of which 734 are in the signal region  $< 60$  GeV and 368 in the background-dominated region  $\geq 60$  GeV. In the signal region, simulation shows an expected 484.5 dilepton events and 29.65 single top  $Wt$  production events, whereas for the background

semileptonic events,  $W$ +jets, and single top  $s$ -channel and  $t$ -channel are predicted to be 229.6, 4.1, and 91.6 respectively.

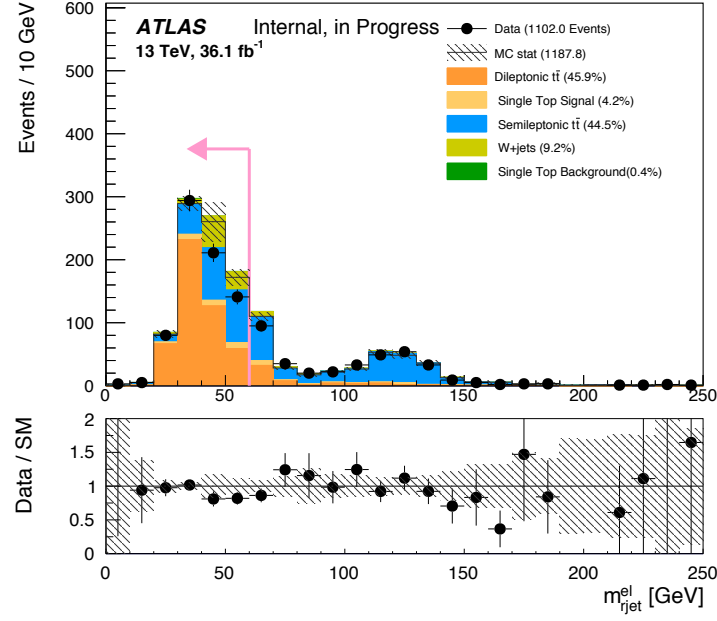


Figure 7.8:  $m_{rjet}^{el}$  after the pre-selection cuts. The region  $< 60$  GeV will define the signal region, and the region  $\geq 60$  GeV will define the control region for background estimation.

## 7.2.4 Background Estimation

The identification efficiency for electrons inside jets depends on the particular operating point (Loose, Medium, or Tight) at which the measurement is carried out. Such an efficiency, which will be denoted  $\epsilon$ , is the ratio of a numerator and a denominator (Section 7.2), both of which are expected to be contaminated with background electrons that need to be estimated. If  $P$  denotes 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, the efficiency  $\epsilon$  may be written as

$$\epsilon = \frac{P - B_P}{N - B_N} \quad (7.2)$$

Because analyses using in-jet electrons all use the Medium or Tight operating point, these are the only two points which will be measured in this chapter. Accordingly, a Medium or Tight ID selection will be applied on the sample representing the denominator, giving in each case the required numerator. 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  $B_P$**  Since we expect background electrons to rarely pass the Medium or Tight ID points, we expect in turn the term  $B_P$  to be very small in either case. Thus  $B_P$  is taken directly from simulation, and the measurements are not expected to be affected significantly.

Figure 7.9 shows the  $m_{\text{rjet}}^{\text{el}}$  distributions for electrons that pass the Medium and Tight selections. The distributions are obtained by applying a Medium or Tight ID selection in addition to the selections that define the signal region (Section 7.2.3). The number of background electrons predicted by the simulation can be seen to be indeed small in each case, accounting for only 0.3% of the total number in the Medium case and 0.1% in the Tight case.

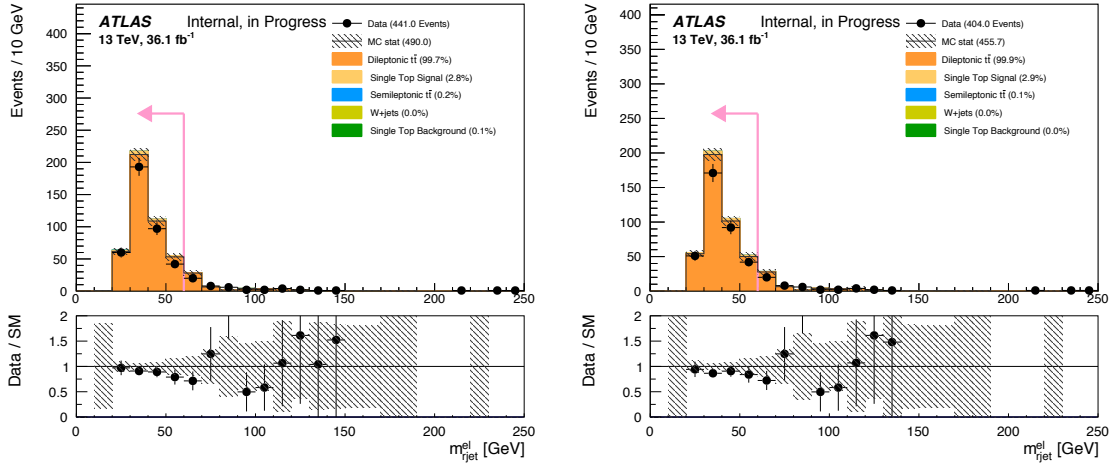


Figure 7.9: The distribution of  $m_{\text{rjet}}^{\text{el}}$  for electrons passing the Medium (left) and Tight (right) operating points. Background electrons figure 0.3% and 0.1% respectively.

**Estimating  $B_N$**  The term  $B_N$  represents background contamination from fake electrons found in  $N$  (Formula 7.2). Since  $N$  contains only reconstructed in-jet electrons with no ID applied, estimating  $B_N$  is expected to be the most challenging part of the measurements.

The method employed for estimating  $B_N$  in the following makes use the set of electrons that fail the Loose ID selection, which will be called antiloose electrons hereafter. These electrons are made up of two parts, one in the signal region ( $\leq 60$  GeV) and one in the background-dominated region ( $> 60$  GeV, Figure 7.8). The part in the background-dominated region will be used to obtain a normalization factor, which will then be applied to the part in the signal region to estimate the number of background electrons. In what follows, the set of antiloose electrons will also be referred to as the fake electron template. Its part in the signal region will be denoted by  $T$ , and that in the background-dominated region will be denoted by  $T_{>}$ .

In order to check if the set of antiloose electrons would be a suitable distribution, the set of background electrons in  $N$ , namely  $B_N$ , is plotted against the former and shown in Figure 7.10, both normalized to unity. As is seen in the figure, the antiloose selection is expected to be effective for classifying background electrons in the sample. On the other hand, Figure 7.11 shows the composition of antiloose electrons in the

1990  $m_{\text{rjet}}^{\text{el}}$  distribution. Simulation predicts about 10% of signal electron contamination,  
 1991 but otherwise the distribution is made up of mostly background electrons dominated  
 1992 by semileptonic  $t\bar{t}$ .

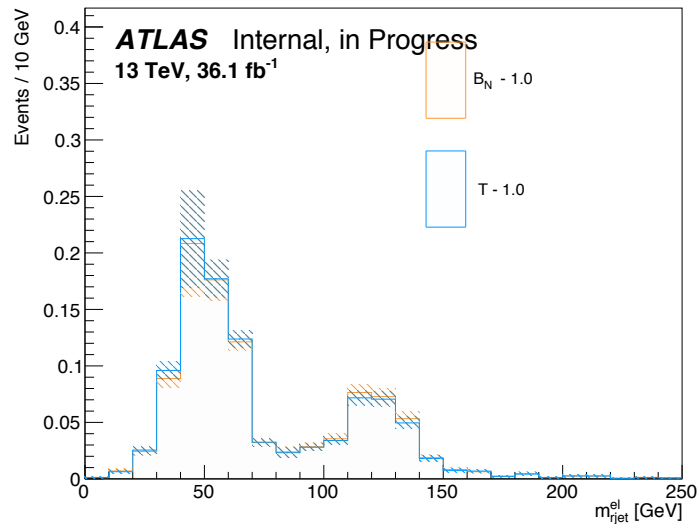


Figure 7.10: The distribution  $m_{\text{rjet}}^{\text{el}}$  of  $B_N$  against that of  $T$ , normalized to unity.  $T$  describes very well  $B_N$  and therefore it is reasonable to estimate  $B_N$  using  $T$ .

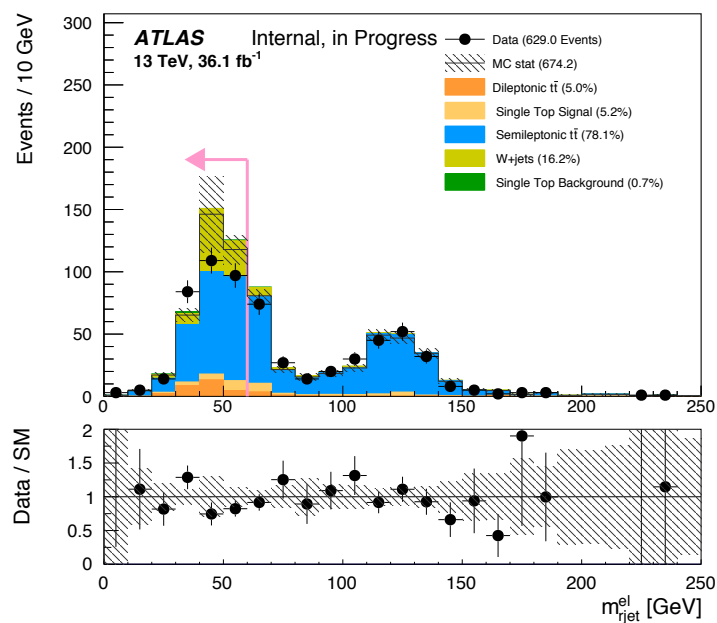


Figure 7.11: The distribution  $m_{\text{rjet}}^{\text{el}}$  for electrons that fail the Loose ID point, also called antiloose electrons.

1993 Background estimation using antiloose electrons proceeds in detail as follows:

- 1994 1. First,  $T$  and  $T_{>}$  are obtained by selecting antiloose electrons. Thus the method  
1995 is data-driven,  $T$  and  $T_{>}$  from simulations are not used.
- 1996 2. In addition to  $N$ , the set of reconstructed electron candidates in the sig-  
1997 nal region, there is also the set of reconstructed electron candidates in the  
1998 background-dominated region, which will be denoted  $N_{>}$ .  
1999 Signal contamination is subtracted from  $T_{>}$ , the resulting set of which is de-  
2000 noted  $\bar{T}_{>}$ , and signal contamination is subtracted from  $N_{>}$ , where the resulting  
2001 set is denoted by  $\bar{N}_{>}$ . Then  $\bar{T}_{>}$  is normalized to  $\bar{N}_{>}$ , to obtain a normalization  
2002 factor.
- 2003 3. Signal contamination is subtracted from  $T$ , the resulting set of which is de-  
2004 noted  $\bar{T}$ , and the normalization factor is applied to  $\bar{T}$ , to obtain the number  
2005 of background electrons in the signal region.

2006 In other words, the background to be estimated in the signal region,  $B_N$ , is  
2007 measured according to

$$B_N = \bar{T} \times \frac{\bar{N}_{>}}{\bar{T}_{>}} \quad (7.3)$$

2008 The following section discusses signal contamination subtractions in  $T$ ,  $T_{>}$ , and  
2009  $N_{>}$ , and the measurements of the identification efficiencies.

## 2010 7.2.5 The Measurements of the Identification Efficiency

2011 The identification efficiency  $\epsilon$  shown in Formula 7.2, where  $B_P$  is taken from simu-  
2012 lation and  $B_N$  is evaluated according to Formula 7.3, is

$$\epsilon = \frac{P - B_P}{N - \bar{T} \times \frac{\bar{N}_{>}}{\bar{T}_{>}}}. \quad (7.4)$$

2013  $\bar{T}$  is the set of antiloose electrons in the signal region,  $T$ , minus signal contamina-  
2014 tion, and  $\bar{T}_{>}$  is the corresponding quantity in the background-dominated region. As  
2015 there is an expected of 10% of signal contamination in the set of antiloose electrons  
2016 (Figure 7.10),  $\bar{T}$  and  $\bar{T}_{>}$  will be obtained by subtracting signal contamination as  
2017 predicted by simulations from  $T$  and  $T_{>}$  respectively.

2018 On the other hand, signal contamination in  $N_{>}$  (Figure 7.8), from which  $\bar{N}_{>}$  is  
2019 obtained, is larger. In fact, as has been mentioned at the end of Section 7.2.3, there  
2020 are 368 events in the background-dominated region, of which simulation predicts  
2021 signal contamination, made up of dilepton events and single top  $Wt$  production  
2022 events, to be  $60.5 + 20.4 = 80.9$  events. In order to reduce the contribution from the  
2023 estimation of this signal contamination to the uncertainty in the efficiency we will use  
2024 a data-driven approach. According to Figure 7.9, the number of background electrons  
2025 after a Medium or Tight ID selection is expected to be negligible. We expect as a  
2026 result  $P$ , and the corresponding quantity  $P_{>}$  in the background-dominated region,  
2027 to be relatively free of background electrons. Thus  $P_{>}$  could be used to represent

signal contamination in  $N_{>}$ , provided the corresponding identification efficiency is properly taken into account. In other words,

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

where the efficiency in 7.4, which is being measured, is used again. The efficiency will be evaluated iteratively, until the change from one iteration to the next is less than 0.5%. The value of 0.5% will be taken as the uncertainty due to signal contamination subtraction in  $N_{>}$ .

The efficiencies, as well as the total statistical and systematic uncertainties (Section 7.2.6), are  $\mathbf{0.870 \pm 0.017 \pm 0.031}$  for Medium and  $\mathbf{0.784 \pm 0.019 \pm 0.020}$  for Tight. As is seen, 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. The relevant quantities in Formula 7.4 that are used to compute the efficiencies in data are listed in Table 7.2.

	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	
$\overline{T}_{>}$	292.52	

Table 7.2: The relevant quantities for computing the efficiencies according to Formula 7.4.

The efficiencies and statistical uncertainties in simulation for the Medium and Tight operating points are also computed and are  $\mathbf{0.871 \pm 0.010}$  and  $\mathbf{0.807 \pm 0.011}$  respectively. Thus Medium in data and in simulation agree, while there is a deviation of about 2% for Tight, possibly revealing the difficulty of modeling accurately electrons inside jets for the latter operating point.

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

## 7.2.6 Uncertainties

The measurement of the identification efficiency is accompanied by statistical and systematic uncertainties. The identification efficiencies, the statistical uncertainties, and the systematic uncertainties have been quoted in Section 7.2.5, they are  $\mathbf{0.870 \pm 0.017 \pm 0.031}$  and  $\mathbf{0.784 \pm 0.019 \pm 0.020}$  for Medium and Tight respectively. Thus the statistical uncertainty is approximately 2% for Medium and 2.4% for Tight, and the systematic uncertainty is higher, approximately 3.6% and 2.6%

2055 respectively. This section discusses in detail the treatment of the statistical and  
 2056 systematic uncertainties, which are listed in Table 7.3 at the end of this section.

2057 **Statistical Uncertainties** According to Formula 7.4, the efficiency is measured  
 2058 according to the formula

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

2059 where

- 2060 •  $P$  is the number of electrons that pass Medium or Tight.
- 2061 •  $B_P$  is background contamination due to fake electrons in  $P$ .
- 2062 •  $N$  is the set of reconstructed electron candidates, and  $\bar{N}_>$  the corresponding  
 2063 quantity in the background-dominated region minus signal contamination.
- 2064 •  $\bar{T}$  is the set of antiloose electrons minus signal contamination, and  $\bar{T}_>$  the  
 2065 corresponding quantity in the background-dominated region.

2066 Since  $N$  contains  $P$ , and  $\bar{N}_>$  contains  $\bar{T}_>$ , the quantities in the formula are  
 2067 not all independent. We may remove the correlation between  $N$  and  $P$  by writing  
 2068  $N = P + F$ , where  $F$  is the set of electrons that fail a particular ID point. Then

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

2069 The correlation between  $\bar{N}_>$  and  $\bar{T}_>$  remains, and moreover  $F$  and  $\bar{T}$  are also  
 2070 correlated, because in the Medium case or in the Tight case,  $F$  represents electrons  
 2071 failing Medium or Tight respectively, and since  $\bar{T}$  represents electrons failing Loose  
 2072 (minus signal contamination), in each case  $\bar{T}$  is a subset of  $F$  and there is accordingly  
 2073 a correlation.

2074 In order to remove all the correlations and write the efficiency completely in terms  
 2075 of statistically independent quantities we will first multiply both the numerator and  
 2076 the denominator by  $\bar{T}_>$ , to write

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

2077 Then we will add and subtract  $\bar{T} \times \bar{T}_>$ , to have

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

2078 The difference  $F - \bar{T}$  represents the set of electrons that fail Medium or Tight  
 2079 but pass the Loose identification, and the difference  $\bar{N}_> - \bar{T}_>$  represents the set of

2080 electrons that pass the Loose identification. If we treat each of the differences as  
 2081 a single term, and set  $S = F - \bar{T}$  and  $\bar{R}_> = \bar{N}_> - \bar{T}_>$  respectively, the efficiency  
 2082 becomes

$$\epsilon = \frac{(P - B_P)\bar{T}_>}{P\bar{T}_> + S\bar{T}_> - \bar{R}_> \times \bar{T}} \quad (7.5)$$

2083 which is now a function of six independent quantities,  $\epsilon = \epsilon(P, B_P, \bar{T}_>, S, \bar{R}_>, T)$ .  
 2084 The statistical uncertainty of the efficiency then follows the standard error propaga-  
 2085 tion formula,

$$\Delta\epsilon^2 = \left(\frac{\partial\epsilon}{\partial P}\right)^2 \Delta P^2 + \dots + \left(\frac{\partial\epsilon}{\partial T}\right)^2 \Delta T^2 \quad (7.6)$$

2086 Let  $A$  denote the numerator in Formula 7.5 and  $B$  the denominator. The terms  
 2087 in the formula above are then

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

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

2088 Since  $P$  and  $S$  are the only terms in the signal region not used for background  
 2089 estimation, the statistical uncertainty of the efficiency is taken from the contributions  
 2090 of these two terms. For both operating points, the contribution from  $S$  is the major  
 2091 one; the contribution from  $P$  is small ( $< 0.5\%$  from the total  $2\%$  for Medium and  
 2092  $2.4\%$  for Tight).

2093 The contributions to the uncertainty from other terms, which are used for back-  
 2094 ground estimation, are taken as contributions to the total systematic uncertainty.

2095 **Systematic Uncertainties** Contributions from different sources to the total sys-  
 2096 tematic uncertainty ( $3.6\%$  for Medium and  $2.5\%$  for Tight), which are discussed  
 2097 below, are added in quadrature.

- 2098 • The variation of the signal region. In addition to defining the signal region at  
 2099  $\leq 60$  GeV, we may define it at  $\leq 50$  or  $\leq 80$  GeV, the asymmetry because  
 2100 of the fact that signal distributions on both sides of the point 60 GeV are not  
 2101 equal in equal intervals. The contribution to the total systematic uncertainty  
 2102 is **0.022** (approximately  $2.5\%$ ) for Medium and **0.010** (approximately  $1.3\%$ )  
 2103 for Tight.
- 2104 • The variation of the term  $B_P$ , which is taken from simulation and represents  
 2105 background contamination in  $P$ . A  $50\%$  variation is used for a conservative  
 2106 estimate of the contribution of this term, which has been seen to be negligible  
 2107 for both Medium and Tight ( $< 0.2\%$  in both cases).



- The simultaneous variations, either up or down, of the signal contaminations in  $T$  and  $T_{>}$ , the subtractions of which from both terms give  $\bar{T}$  and  $\bar{T}_{>}$ . A 25% variation is used for a conservative estimate of these contributions, which are **0.017** (approximately 2%) in Medium and **0.015** (approximately 1.9%) in Tight.
- The change of the template  $T$ , from the distribution of antiloose electrons to the distribution of antiloose electrons in events with exactly 2  $b$ -jets. The contributions to the total systematic uncertainty are **0.008** (approximately 0.9%) for Medium and **0.007** (approximately 0.9%) for Tight.
- The statistical uncertainties from the counting of  $\bar{T}_{>}$ ,  $\bar{R}_{>}$ , and  $\bar{T}$  in Formula 7.5. They are 0.002 (approximately 0.2%), 0.008 (approximately 0.9%), and 0.002 (approximately 0.2%) respectively for Medium and 0.001 (approximately 0.1%), 0.006 (approximately 0.8%), and 0.001 (approximately 0.1%) respectively for Tight.

	MEDIUM	TIGHT
Systematic Uncertainties		
$S$	0.022	0.010
$B_P$	0.002	0.000
$T$	0.017	0.015
Statistical Uncertainties		
	0.010	0.011

Table 7.3: The statistical and systematic uncertainties for the Medium and Tight operating point.

## 7.2.7 Efficiencies as Functions of the Properties of the Electron and of the Overlapping Jet

In addition to the integrated efficiencies, the efficiencies as functions of the properties of the electron and of the overlapping jet are also measured. The measurements include the following variables (Figure 7.12 and 7.13).

- $p_T$  of the probe, in five bins: 30-60 GeV, 60-80 GeV, 80-110 GeV, 110-140 GeV, and  $> 140$  GeV.
- $|\eta|$  of the probe, in five bins: 0.0-0.3, 0.3-0.6, 0.6-0.9, 0.9-1.3, and  $> 1.3$ .
- $\Delta R$  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, and 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, and 400-500 GeV.

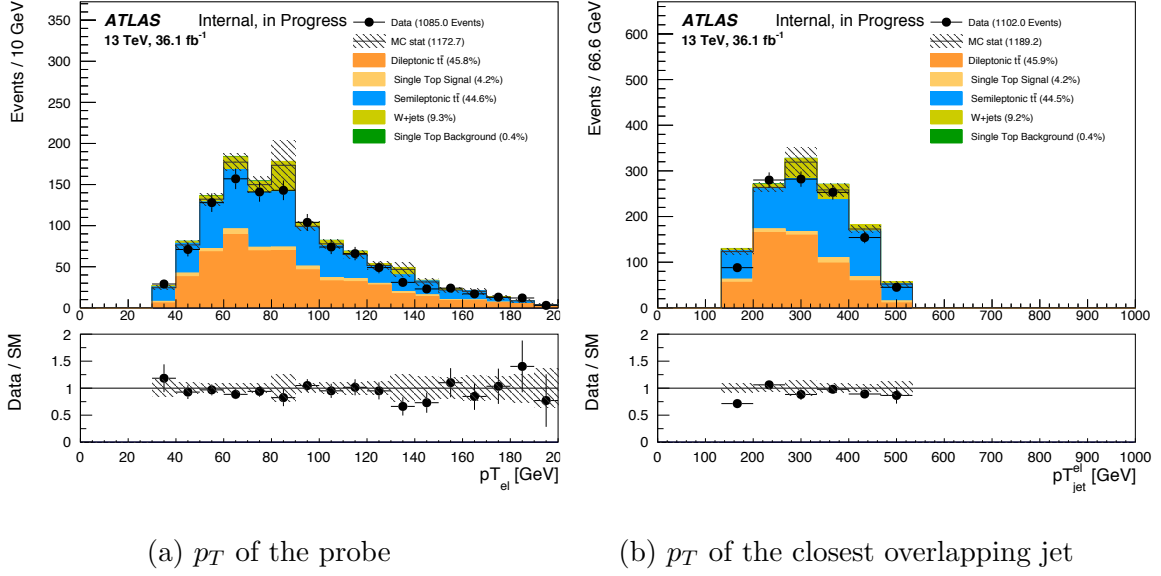


Figure 7.13: The distributions of  $p_T$  of the probe and  $p_T$  of the closest overlapping jet.

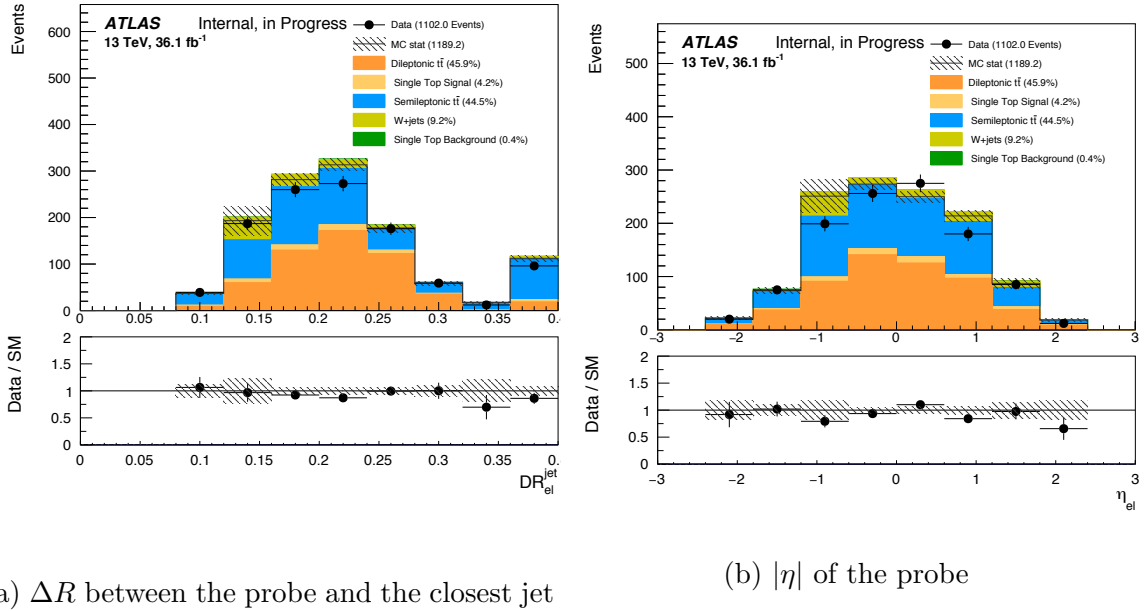


Figure 7.12: The distributions of  $\Delta R$  between the probe and the closest overlapping jet and  $|\eta|$  of the probe.

2134 The efficiencies, for the Medium and Tight operating points, as a function of the  
 2135  $p_T$  and  $|\eta|$  of the probes are shown in Figure 7.14. Also shown are the efficiencies  
 2136 for standard electrons ([38, 47]), which, as can be seen, agree with those for in-jet  
 2137 electrons within the error bars. As a function of the  $p_T$  of the probe, the efficiencies  
 2138 increase as  $p_T$  increases. On the other hand, no obvious dependency is seen in the  
 2139 case of  $|\eta|$ .

2140 The efficiencies as a function of the  $\Delta R$  between the probe and the closest over-

lapping jet, and as a function of the  $p_T$  shown in Figure 7.15. In the latter case the efficiencies are higher for lower  $p_T$ .

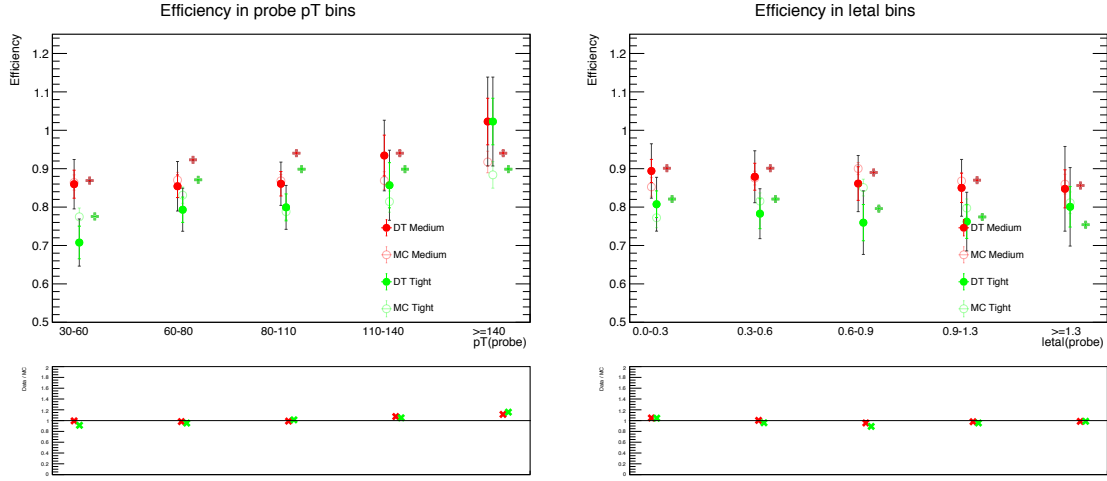


Figure 7.14: The efficiencies in  $p_T$  of the probe as well as in  $|\eta|$  of the probe. Also shown are the efficiencies for standard electrons and the associated uncertainties (which are very small and therefore are barely visible).

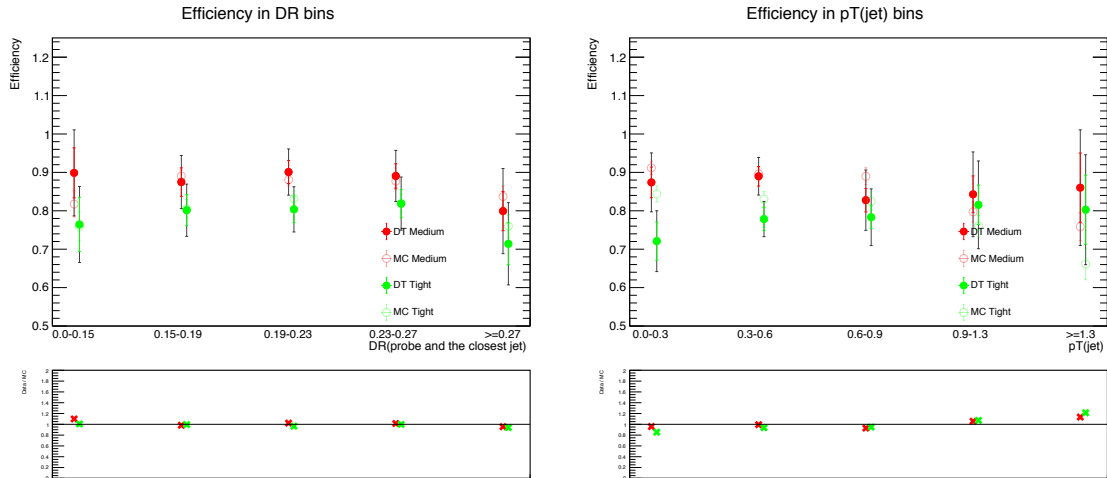


Figure 7.15: The efficiencies in  $\Delta R$  between the probe and the closest jet, as well as in  $p_T$  of the closest overlapping jet.

## 7.3 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, and the first ever using dilepton  $t\bar{t}$  events. The measurement used the data collected

2147 in the period 2015-2016, at 13 TeV center-of-mass and totaled  $36.5 \text{ fb}^{-1}$  in integrated  
2148 luminosity. A sample of electrons for the measurements was obtained by selecting  
2149 boosted  $t\bar{t}$  dilepton ( $e\mu$ ) events. Background estimations used both simulations and  
2150 data, and the efficiencies were evaluated iteratively. The efficiencies were measured  
2151 for the Medium and Tight operating points, both on data and simulation. The  
2152 efficiencies as functions of the properties of the electrons and of the overlapping jets  
2153 also measured. In all of the results, the efficiencies predicted by simulation agree  
2154 with those obtained from the measurements on data.