

1740 Chapter 7

1741 IN-JET ELECTRON 1742 IDENTIFICATION 1743 EFFICIENCIES

1744 In early 2015 the LHC restarted after two years of shutdown, beginning what is
1745 referred to as Run 2. The new centre-of-mass energy was 13 TeV, in place of the
1746 previous 8 TeV. The higher energy opens up unexplored parameter space and allows
1747 further probe of supersymmetry (SUSY) and other beyond-the-Standard-Model pro-
1748 cesses [79–82]. At ATLAS, many SUSY searches involve supersymmetric particles
1749 that decay into the standard model top quarks and we expect, with higher centre-of-
1750 mass energy, more massive supersymmetric particles and consequently an increase
1751 in the production of high p_T top quarks. Since the top quarks also decay, in fact
1752 most of the time into a W boson and a b quark, we in turn expect boosted decay
1753 topology, in other words the daughter particles of the top quarks, which include the
1754 daughter particles of the W boson and the b quark, tend to stay close to each other
1755 (Figure 7.1). In the case of leptonic top quark decays, the lepton tends to be very
1756 close to the produced b -jet.

1757 This chapter describes the work to measure the identification efficiencies for elec-
1758 trons that are found inside $\Delta R = 0.4$ of high p_T jets, which will also be referred to
1759 in the chapter as in-jet electrons. Prior to the work in this chapter, there was no
1760 attempt to measure the identification efficiencies for in-jet electrons. The chapter
1761 will be organized as follows. In section 7.1 we motivate the need for the measurement
1762 of the identification efficiencies for in-jet electrons. Section 7.2 describes the method
1763 used to perform the measurements. Section 7.3 presents the measured efficiencies
1764 for in-jet electrons and section 7.4 gives some conclusions.

1765 The data used for this chapter was collected in the period 2015–2016 at 13 TeV
1766 center-of-mass and corresponded to an integrated luminosity of 36.47 fb^{-1} .

1767 7.1 Motivation The entire section 7.1 has been re-written

1768 At ATLAS, prior to Run 2, electrons found outside $\Delta R = 0.4$ of jets were used by
1769 most analyses. Indeed, although there were some attempts to select signal electrons
1770 inside jets [?], at 7–8 TeV such electrons can hardly be expected; objects inside

Fixed.
1766-1769

Rephrase to say that we become sensitive to
higher mass susy particles, which in term
would produce higher pt decay products

Rephrase to mean essentially all the time

Not true, there was the one using Z to II
where the I is inside jets. Yours is the first
one in the ttbar topology

Too many significant digits wrt uncertainty

1750 Chapter 7

1751 IN-JET ELECTRON 1752 IDENTIFICATION 1753 EFFICIENCIES

1754 In early 2015 the LHC restarted after two years of shutdown, beginning what is
1755 referred to as Run 2. The new cent-of-mass energy was 13 TeV, in place of the
1756 previous 8 TeV. The higher energy opens up unexplored parameter space and allows
1757 further probe of supersymmetry (SUSY) and other beyond-the-Standard-Model pro-
1758 cesses [85–88]. At ATLAS, many SUSY searches involve supersymmetric particles
1759 that decay into the Standard Model top quarks and we expect, with higher centre-of-
1760 mass energy, higher sensitivity to more massive supersymmetric particles and
1761 consequently an increase in the production of high p_T top quarks. Since the top
1762 quarks also decay — essentially all the time — into a W boson and a b quark, we
1763 in turn expect boosted decay topology, in other words the daughter particles of the
1764 top quarks, which include the daughter particles of the W boson and the b quark,
1765 to stay close to each other (Figure 7.1).

1766 This chapter describes the work to measure the identification efficiencies for elec-
1767 trons that are found inside $\Delta R = 0.4$ of high p_T jets, which will also be called in-jet
1768 electrons, using $t\bar{t}$ events. Prior to the work in this chapter, there was no attempt
1769 to measure the identification efficiencies for in-jet electrons in a $t\bar{t}$ topology. The
1770 chapter will be organized as follows. In section 7.1 we motivate the need for the
1771 measurement of the identification efficiencies for in-jet electrons. Section 7.2 de-
1772 scribes the method used to perform the measurements and presents the measured
1773 efficiencies. Section 7.3 presents some conclusions.

1774 The data used for this chapter was collected in the period 2015–2016 at 13 TeV
1775 center-of-mass and corresponded to an integrated luminosity of 36.5 fb^{-1} .

1776 7.1 Motivation

1777 Prior to Run 2, ATLAS center-of-mass energy 7–8 GeV allows limited sensitivity to
1778 high mass resonances. Because many beyond-Standard-Model particles are predicted
1779 to decay into the Standard Model top quarks, the limited sensitivity reduces the
1780 chance in which we could expect boosted top quark decays. Such decays, however,

1781 are expected to become significant as the centre-of-mass of the LHC reached 13 TeV
 1782 starting from Run 2. In a boosted top quark decay scenario, the produced particles,
 1783 which in this case are the daughters of the W and the b quark that come from
 1784 the top quark, are found close to each other [90–92]. Figure 7.2 shows the angular
 1785 distance ΔR (Formula 3.1) between the W 's and the b -quarks as a function of the
 1786 top p_T , in the context of a hypothetical particle Z' with mass $m_{Z'} = 1.6$ TeV [92]
 1787 that decays into a $t\bar{t}$ pair. Also shown in the same figure is the separation between
 1788 the light quarks of the subsequent hadronic decay of the W boson. As can be seen,
 1789 the angular distance decreases as the top quark p_T increases, and at high top quark
 1790 p_T a non-negligible fraction of the distances becomes very small.

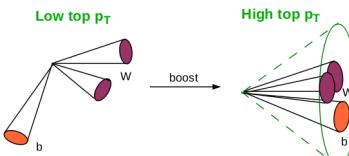


Figure 7.1: An illustration of low p_T top quark decay (left) and boosted top decay (right) of a high p_T top quark. In the case of high p_T top quark decay the daughter particles of the top quark, which include the daughter particles of the W and the b quark, are expected to be found close to each other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

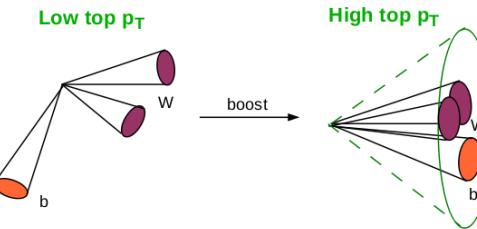


Figure 7.1: An illustration of low p_T top quark decay (left) and boosted top decay (right) of a high p_T top quark. In the case of high p_T top quark decay the daughter particles of the top quark, which include the daughter particles of the W and the b quark, are expected to be found close to each other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

1776 In Run 2, as the centre-of-mass of the LHC reached 13 TeV, high p_T top quarks
 1777 are expected to come from the decays of massive supersymmetric or other beyond-
 1778 the-Standard-Model particles. As we have already mentioned, a top quark decays
 1779 almost all of the time into a W and a b quark, and high p_T top quarks are expected to
 1780 undergo boosted decays where the produced particles, here the daughters of the W
 1781 and the b quark, are found close to each other. In particular, there are two possible
 1782 scenarios: either the W decays hadronically, in which case the produced jets stay
 1783 close to each other, or it may decay leptonically, in which case the leptons are
 1784 expected to be found near the produced b -jets. Either scenario is important and the
 1785 possibility of encountering boosted topology was assessed by several ATLAS physics
 1786 groups, albeit most works focused on hadronic boosted top quark decays [84–86].
 1787 Figure 7.2 shows the angular distance ΔR (Formula 3.2) between the W 's and the
 1788 b -quarks as a function of the top p_T , in the context of a hypothetical particle Z' at
 1789 mass $m_{Z'} = 1.6$ TeV [86] that decays into a $t\bar{t}$ pair. It also shows the separation
 1790 between the light quarks of the subsequent hadronic decay of the W boson. As can
 1791 be seen, the angular distance decreases as the top quark p_T increases, and at high
 1792 top quark p_T a non-negligible fraction of the distances becomes very small. In the
 1793 case of leptonic top quark decays, we should expect the decay products of the W 's
 1794 to be found very near the b -jets.

1791 Leptonic boosted top quark decay is also an important channel in searches for
 1792 beyond-Standard-Model particles that decay into the Standard-Model top quarks.
 1793 Table 7.1 shows a measurement of the fraction of in-jet electrons over signal elec-
 1794 trons as a function of the top quark p_T , at truth-level. The measurement used
 1795 PowhegPythia $t\bar{t}$ events simulated at 13 TeV centre-of-mass energy (Chapter 6, Sec-
 1796 tion 6.2), where dilepton events consisting of a muon and an electron were selected.
 1797 The selections made use of the p_T -dependent overlap removal

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

1798 where the overlap removal is required to keep the overlapping b -jets (Chapter 6,
 1799 Section 6.3). As is shown, more and more electrons are found inside jets as the top
 1800 quark p_T increases. The number of in-jet electrons becomes quite significant from
 1801 500 GeV, being approximately 25% there and reaching nearly 40% at 650 GeV. If
 1802 the top quark p_T is allowed to go up to 1 TeV, the figure is 64%. This result supports
 1803 the fact that different ATLAS analyses searching for heavy beyond-Standard-Model
 1804 particles decaying into lighter sparticles, such as the gluinos and stops that decay into
 1805 neutralinos, in which the final state involve the Standard Model top quarks, were able
 1806 to increase signal acceptances considerably if in-jet electrons were selected [85, 86].

Paragraph not OK, let's discuss in person

78

1771 jets that are identified as electrons are mostly background electrons that are either
 1772 hadrons faking jets or real electrons coming from heavy-flavour jet decays. As a
 1773 result, most analyses rejected electrons inside $\Delta R = 0.4$ of jets and only worked
 1774 with electrons outside jets, of which different efficiencies, among them identification
 1775 efficiencies, are measured and provided by the ATLAS Egamma group [37].

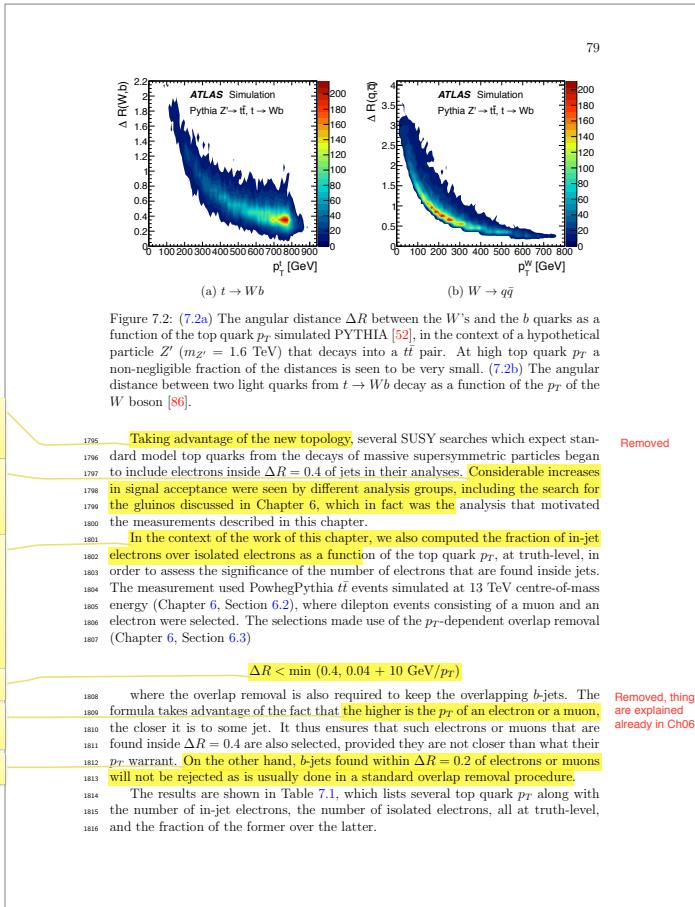
Same as above, need to be discussed

1776 In Run 2, as the centre-of-mass of the LHC reached 13 TeV, high p_T top quarks
 1777 are expected to come from the decays of massive supersymmetric or other beyond-
 1778 the-Standard-Model particles. As we have already mentioned, a top quark decays
 1779 almost all of the time into a W and a b quark, and high p_T top quarks are expected to
 1780 undergo boosted decays where the produced particles, here the daughters of the W
 1781 and the b quark, are found close to each other. In particular, there are two possible
 1782 scenarios: either the W decays hadronically, in which case the produced jets stay
 1783 close to each other, or it may decay leptonically, in which case the leptons are
 1784 expected to be found near the produced b -jets. Either scenario is important and the
 1785 possibility of encountering boosted topology was assessed by several ATLAS physics
 1786 groups, albeit most works focused on hadronic boosted top quark decays [84–86].
 1787 Figure 7.2 shows the angular distance ΔR (Formula 3.2) between the W 's and the
 1788 b -quarks as a function of the top p_T , in the context of a hypothetical particle Z' at
 1789 mass $m_{Z'} = 1.6$ TeV [86] that decays into a $t\bar{t}$ pair. It also shows the separation
 1790 between the light quarks of the subsequent hadronic decay of the W boson. As can
 1791 be seen, the angular distance decreases as the top quark p_T increases, and at high
 1792 top quark p_T a non-negligible fraction of the distances becomes very small. In the
 1793 case of leptonic top quark decays, we should expect the decay products of the W 's
 1794 to be found very near the b -jets.

fraction given
later in table
Remove

This sentence is utterly empty of new information

This sentence is utterly empty of new information



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 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 .

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 [38, 47].

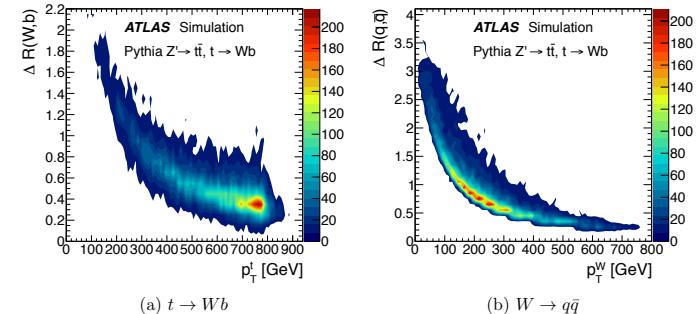


Figure 7.2: (7.2a) The angular distance ΔR between the W 's and the b quarks as a function of the top quark p_T simulated PYTHIA [57], in the context of a hypothetical particle Z' ($m_{Z'} = 1.6$ TeV) that decays into a $t\bar{t}$ pair. At high top quark p_T a non-negligible fraction of the distances 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].

1810 7.2 Method

1811 The method that is used to measure the identification efficiencies for in-jet electrons
 1812 is discussed in detail in Section 7.2.1, 7.2.2, and 7.2.3. Background estimations is
 1813 described in Section 7.2.4.

1814 7.2.1 Boosted Dilepton $e\mu$ Events

1815 In order to measure the identification efficiencies for in-jet electrons, a sample of
 1816 reconstructed electrons (Chapter 6, Section 6.3) inside high- p_T jets was obtained
 1817 by selecting boosted $t\bar{t}$ dilepton ($e\mu$) events. This is expected to result not only
 1818 in a very pure $t\bar{t}$ sample, but also in a topology close to that of many SUSY and
 1819 other beyond-Standard-Model searches. In contrast, the standard method for mea-
 1820 suring electron identification efficiencies, the tag-and-probe method supported by
 1821 the Egamma group at ATLAS [47], makes use of $Z \rightarrow e^+e^-$ events for high-energy
 1822 electrons ($E_T > 10$ GeV). Even though a clean sample of electrons may be obtained
 1823 in a relatively straightforward way by selecting events around the Z mass peak, we
 1824 expect, if electrons inside high p_T jets are required, a sample unrepresentative of
 1825 events with a boosted topology and limited in statistics.

1826 Thus, a sample of electrons was obtained by selecting a hard muon and a recon-
 1827 structed electron with only a p_T requirement applied, which will also be referred to
 1828 below simply as reconstructed electrons. The efficiency at a particular identifica-
 1829 tion operating point (Section 4.2.2) is defined by the ratio

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

1830 Both the numerator and the denominator are contaminated with background
 1831 electrons which require a careful estimate (Section 7.2.4), particularly because back-
 1832 ground electrons are expected to reside primarily inside jets.

1833 7.2.2 Data and Monte Carlo Samples

1834 The data used for this chapter was collected in the period 2015-2016 and corre-
 1835 sponds to an integrated luminosity of 36.5 fb^{-1} . The following simulation samples
 1836 (Chapter 6, Section 6.2), at 13 TeV centre-of-mass energy, are used:

- 1837 • $t\bar{t}$ events from the Powheg+Pythia generator. As a hard muon will be re-
 1838 quired in the sample in which the identification efficiencies are measured (Sec-
 1839 tion 7.2.3), these events naturally partition into either a dileptonic set ($e\mu$)
 1840 when truth-level electrons are present inside high- p_T jets, or a semileptonic set
 1841 otherwise. The latter, with jets from the fully hadronic decay of one of the
 1842 top quarks constituting a source of background electrons, is expected to be a
 1843 dominant background.

- 1844 • $W + \text{jets}$, which will be used as a background. As the W boson may produce
 1845 a hard muon, the presence of jets make these events a source of background
 1846 events to signal dilepton $e\mu t\bar{t}$ events.

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%

80

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 .

1817 It is seen that more and more electrons are found inside jets as the top quark p_T
 1818 increases. In fact, the number of in-jet electrons becomes quite significant from 500
 1819 GeV, being approximately 25% there and reaching nearly 40% at 650 GeV. If the
 1820 top quark p_T is allowed to go up to 1 TeV, the figure is 64%. This result supports the
 1821 fact that different ATLAS analyses saw considerable increases in signal acceptance
 1822 as they included in-jet electrons in their selections.

1823 This chapter develops a method and performs the initial measurements for the
 1824 identification efficiencies of electrons found inside $\Delta R = 0.4$ of jets. The measure-
 1825 ments for electrons outside jets are done by the ATLAS Egamma group [37, 43].

1826 7.2 Method

1827 In general, to measure the identification (ID) efficiencies of electrons we start from
 1828 a sample of reconstructed electrons. The efficiency at a particular identification
 1829 operating point (section 4.2.2) is defined by the ratio

This section and 7.2.1 has
been rewritten

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

1830 The number of identified electrons, the numerator, is contaminated with back-
 1831 ground electrons, and so is the denominator. As a result, part of the work to obtain a
 1832 good sample of electrons inside high p_T jets includes at the same time adequate back-
 1833 ground estimation methods, particularly because background electrons are expected
 1834 to reside primarily inside jets.

1835 In this thesis, a sample of electrons was obtained by selecting boosted H dilep-
 1836 ton ($e\mu$) events, described below in Section 7.2.1, 7.2.2, and 7.2.3. Background
 1837 estimations will be described in Section 7.2.4.

Isn't it over the total number of electrons?
That would make more sense, and only then
could the word fraction be used

Say Atlas analyses searching for heavy BSM
particles decaying to top quarks

Has reconstructed electrons been well
defined by now? It includes some cuts like
track quality variables. If that has all been
done in chapter 4, refer back to the relevant
section for the definition of reconstructed
electrons

This section currently doesn't work, it only
contains trivial information with the formula
above. I think it's better to just say the
sample which is the key and most interesting
part of the measurement. Remove that
paragraph or move earlier, and merge next
section to the current one (and rephrase by
linking the equation with the sample).

This has been attempted before, review the previous measurements and provide numbers. Please perform appropriate research and do the extra mile to write a good thesis! You are aware of the measurements I refer to

This has been attempted before, review the previous measurements and provide numbers. Please perform appropriate research and do the extra mile to write a good thesis! You are of the Add that the emu channel has the advantage of providing a very pure tbar sample

Add at the end that this sample will constitute the denominator of the previous equation. Could add also that once an identification criteria is added, this subsample will constitute the numerator of that equation.

Add at the end that this sample will constitute the denominator of the previous equation. Could add also that once an identification criteria is added, this subsample will const the numerator of

Is that the same samples as chapter 6? Say so explicitly and that they were defined there, and that here you describe how they will be used

Add a + before Pythia. Actually just remove the generator if you're not specifying it for the other samples. But since there are several options for tbar, specify it but in the text.

Specify that it will be the dominant source of background in the analysis

Rephrase to Wt is treated as a source of signal electrons, since it contains a pair of W bosons that can decay to a prompt emu pair.

Rephrase that they will contain only one W boson, and therefore cannot produce a prompt emu pair

Rephrase that they contain only one W boson, and therefore not produce a

81

1838 7.2.1 Boosted Dilepton $e\mu$ Events

1839 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 $Z \rightarrow e^+e^-$ events for high-energy electrons ($E_T > 10$ 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.

1840 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 selection 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.

1841 Compared to $Z \rightarrow ee$ 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.

1842 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.

1850 **1850 7.2.2 Data and Monte Carlo Samples**

1860 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): Change position; I think in the revised text just naming the chapter implies the same samples

1863 ○ PowhegPythia [\[4\]](#). 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

1864 ○ 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

1865 ○ 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.

For these, read corresponding text in 7.2.2

- 1847 • 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.

1852 7.2.3 Signal Region

1853 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 $e\mu$ and $t\bar{t}$ events, are applied.

1857 Pre-selection

- 1858 • One primary vertex
- 1859 • Muon trigger. The following triggers were used for the periods 2015 and 2016:
 - 1860 – 2015: HLT_mu26_imedium || HLT_mu40
 - 1861 – 2016: HLT_mu26_ivarmedium || HLT_mu50
- 1862 • p_T -dependent overlap removal, where the overlapping b -jets are kept (Formula 7.1).
- 1863 • 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.
- 1864 • Exactly one identified muon and ≥ 1 electrons inside jets are required for each event, where
 - 1865 – 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 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.
 - 1866 – 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,
- 1867 • ≥ 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.

Add: that are aimed at isolating ttbar emu events

These names are internal jargon. Are these defined in chapter 6? If yes say so, if not need to explain what they mean

Number the equation in section 7.1 and refer directly to that equation

Specify muon candidates. Only use muons if it's a truth one.

Identified

Also jargon, same comment as previously

~~Also jargon, same comment as previously. Under the way it's stated now. Before it made more sense to specify when that cut was 50 GeV. But 30 GeV is pretty standard for all purposes so leave as is~~

Add after electrons: ", which is expected to originate from the interaction point."

~~Add after electrons: which is expected to originate from the interaction point. Add candidate to electron everywhere in this chapter when you refer to the experimental object and not a truth electron. Important!~~

Add also approximate expected number for ttbar file from semileptonic like you do for W+jets and single top

~~Add also approximate expected number for ttbar file from semileptonic like you do~~

82

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: Done

Pre-selection

- One primary vertex
- Muon trigger. The following triggers were used for the periods 2015 and 2016:
 - 2015: HLT_mu26_imedium || HLT_mu40 I have not found a way to write these yet
 - 2016: HLT_mu26_ivarmedium || HLT_mu50 Later

p_T -dependent overlap removal, with the option of keeping overlapping b -jets turned on (Section 7.1) Done

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 parameter and the longitudinal impact parameter, and $\text{ptvarcone30}/\text{d}_r < 0.06$, 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.

These points have been revised

Added

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 $m_{\text{jet}}^{\text{el}}$. 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_{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.

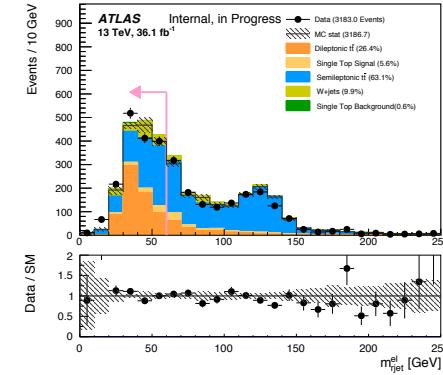
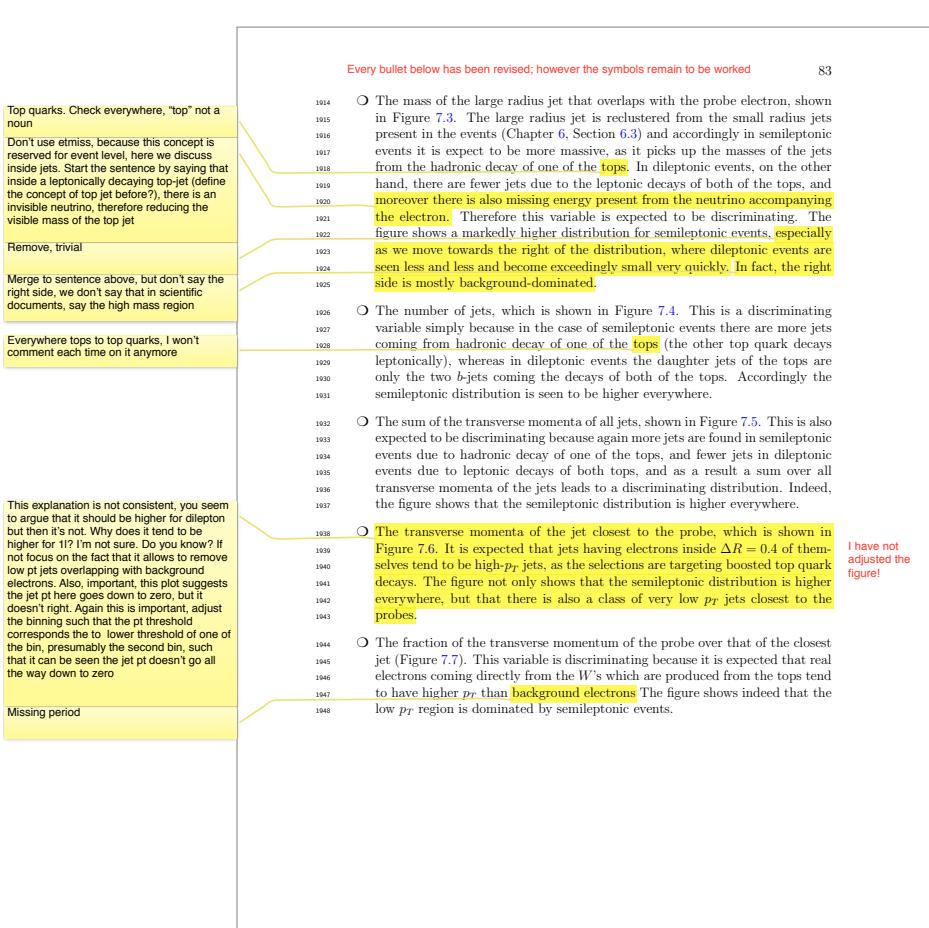


Figure 7.3: $m_{\text{jet}}^{\text{el}}$. The semileptonic contribution is higher everywhere, especially on the right side of the distribution where there is little signal contamination.

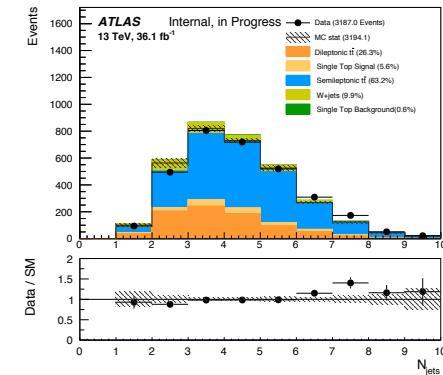


Figure 7.4: N_{jets} . The semileptonic contribution is higher because of the hadronic decay of one of the tops.

84

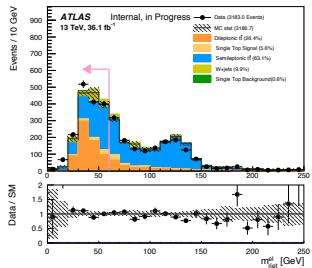


Figure 7.3: m_{jet}^{el} . The semileptonic contribution is higher everywhere, especially on the right side of the distribution where there is little signal contamination.

Figures will be fixed later

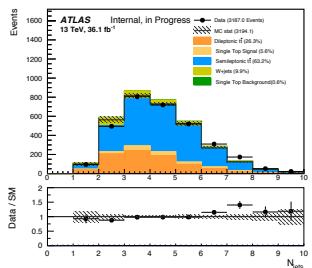


Figure 7.4: The number of jets. The semileptonic contribution is higher because of the hadronic decay of one of the tops.

Figures will be fixed later

For the first figure, describe it fully in the caption. This should include for example the ratio plot below. Check out a paper for example. The following figure captions can then say see figure 7.3 for explanation of the figure.

For the first figure, describe it fully in the caption. This should include for example the ratio plot below. Check out a paper for example. The following figure captions can then say see figure 7.4 or explanation of the

Introduce the symbol used in the plots so for example the number of jets N_{jets} . Do this for all figures

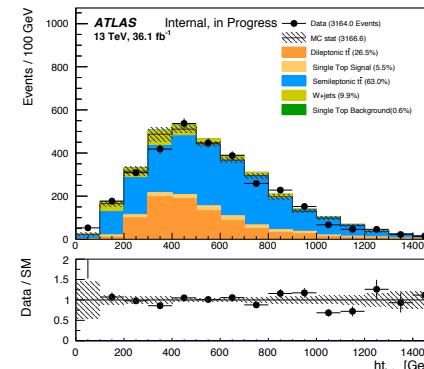


Figure 7.5: The sum of the transverse momenta of all jets.

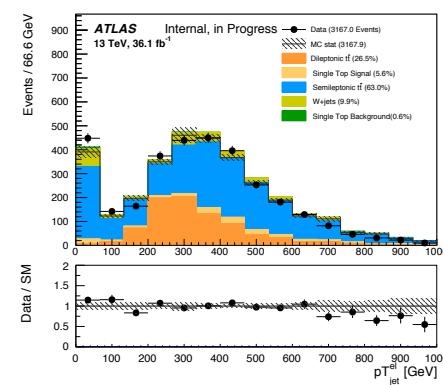
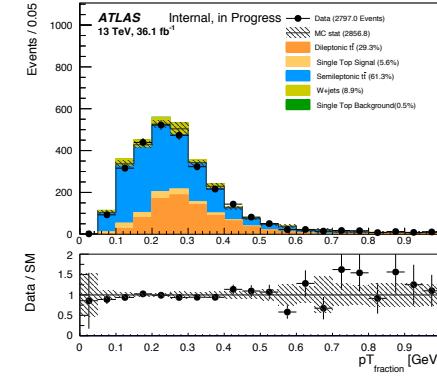
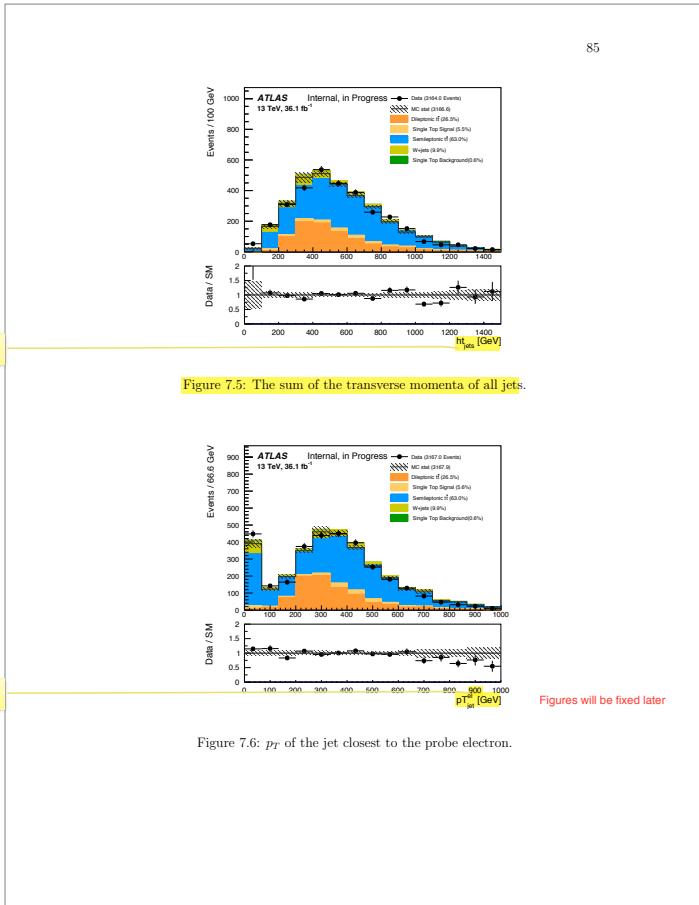


Figure 7.6: p_T of the jet closest to the probe electron.



1923 Further cuts to arrive at the signal region Of all the discriminating variables
 1924 shown above, $m_{\text{jet}}^{\text{el}}$ seems to be the most discriminating variable. In addition, its
 1925 distribution shows two distinct regions, one abundant in signal electrons and one
 1926 largely dominated by background electrons. As will be discussed later in the chapter,
 1927 the region $< 60 \text{ GeV}$ will define the signal region where the identification efficiencies
 1928 are measured, and the region $> 60 \text{ GeV}$ will define the control region for background
 1929 estimation. With this in mind, we decided to apply cuts on the other discriminating
 1930 variables to further remove the undesired background, while leaving $m_{\text{jet}}^{\text{el}}$ untouched.
 1931 The cuts are as follows:

- 1932 • Missing transverse momentum $E_T^{\text{miss}} > 25 \text{ GeV}$, to ensure that the QCD multi-
 1933 jet background is negligible.
- 1934 • The number of jets < 5 and sum of p_T of jets $< 700 \text{ GeV}$, to remove semilep-
 1935 tonic events (Figure 7.4 and 7.5).
- 1936 • p_T of jet closest to the probe is between 150 GeV and 500 GeV , to remove
 1937 semileptonic events (Figure 7.6) and at the same time make sure that boosted
 1938 $t\bar{t}$ dilepton events are selected.
- 1939 • $p_T(\text{probe})/p_T(\text{closest jet}) > 0.16$ (Figure 7.7).

1940 The resulting distribution $m_{\text{jet}}^{\text{el}}$ is shown in Figure 7.8. There are 1102 events, of
 1941 which 734 are in the signal region $< 60 \text{ GeV}$ and 368 in the background-dominated
 1942 region $\geq 60 \text{ GeV}$. In the signal region, simulation shows an expected 484.5 dilep-
 1943 ton events and 29.65 single top Wt production events, whereas for the background

¹⁹⁴⁴ semileptonic events, $W+jets$, and single top s -channel and t -channel are predicted
¹⁹⁴⁵ to be 229.6, 4.1, and 91.6 respectively.

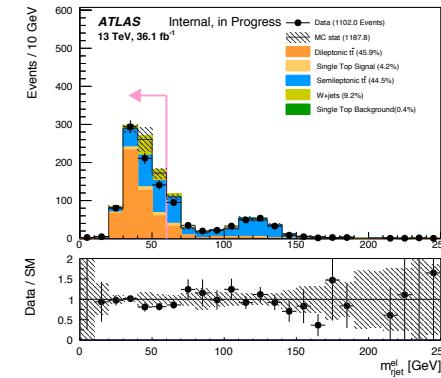
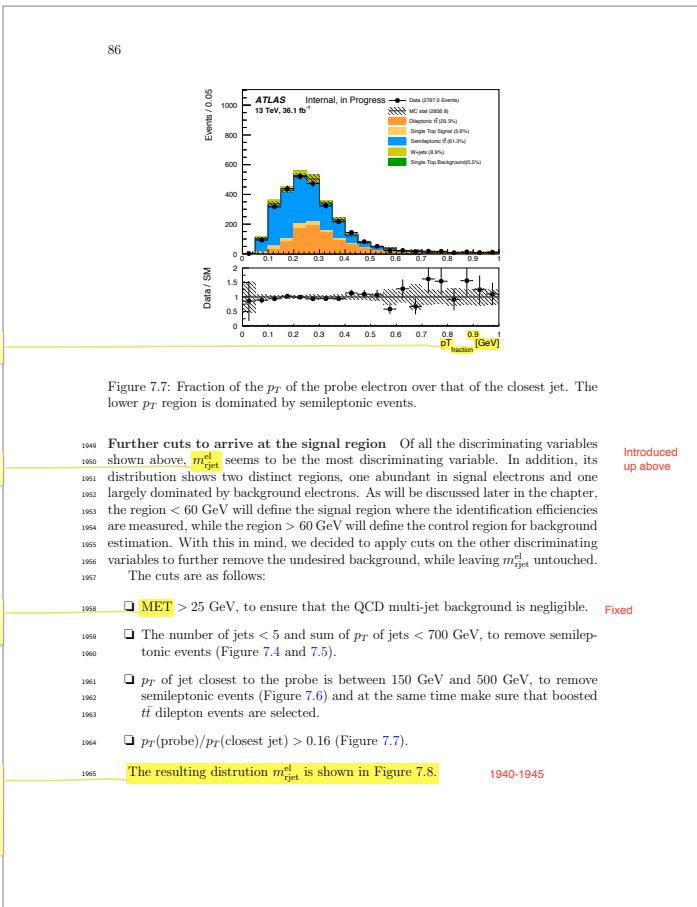


Figure 7.8: m_{jet}^{el} after the pre-selection cuts. The region < 60 GeV will define the signal region, and the region ≥ 60 GeV will define the control region for background estimation.

7.2.4 Background Estimation

¹⁹⁴⁷ The identification efficiency for electrons inside jets depends on the particular operating point (Loose, Medium, or Tight) at which the measurement is carried out. ¹⁹⁴⁸ Such an efficiency, which will be denoted ϵ , is the ratio of a numerator and a denominator (Section 7.2), both of which are expected to be contaminated with background ¹⁹⁴⁹ electrons that need to be estimated. If P denotes the number of electron candidates ¹⁹⁵⁰ passing a particular ID operating point, B_P the number of background electrons ¹⁹⁵¹ passing the operating point, N the total number of reconstructed electron candidates ¹⁹⁵² in the sample, and B_N the number of background electrons present in the ¹⁹⁵³ sample, the efficiency ϵ may be written as

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

¹⁹⁵⁶ Because analyses using in-jet electrons all use the Medium or Tight operating ¹⁹⁵⁷ 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 ¹⁹⁵⁸ the denominator, giving in each case the required numerator. Background estimations ¹⁹⁵⁹ will consist of estimating the term B_P separately for Medium and Tight in the ¹⁹⁶⁰ numerator, and estimating the common term B_N in the denominator.

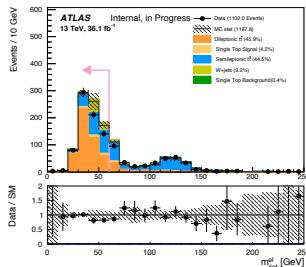


Figure 7.8: $m_{\text{jet}}^{\text{el}}$ after the pre-selection cuts. The region $< 60 \text{ GeV}$ will define the signal region, and the region $\geq 60 \text{ GeV}$ will define the control region for background estimation.

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

$$\epsilon = \frac{P - B_P}{N - B_N} \quad (7.1)$$

The identification efficiencies for the Medium and Tight ID operating points are the most 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

revised
completely

Say here already « Medium » or « Tight »
electron identification operating point

Not sure that they are the most widely used? In any case, the main reason in my mind is that nobody using electrons inside jets would use the Loose OP, it would contain too much background electrons. So say that analyses employing in-jet electrons all use the medium or tight OP.

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 $m_{\text{jet}}^{\text{el}}$ 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.

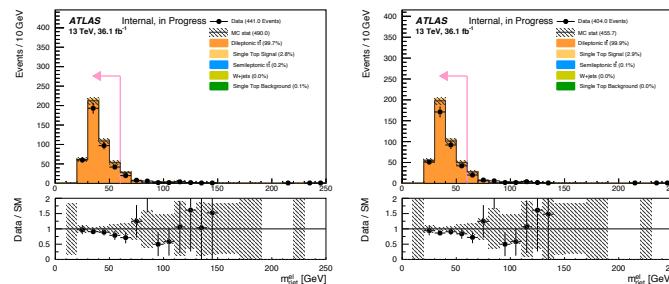


Figure 7.9: The distribution of $m_{\text{jet}}^{\text{el}}$ for electrons passing the Medium (left) and Tight (right) operating points. Background electrons figure 0.3% and 0.1% respectively.

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 of 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 ($\leq 60 \text{ GeV}$) and one in the background-dominated region ($> 60 \text{ GeV}$, 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 $m_{\text{jet}}^{\text{el}}$ 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 $t\bar{t}$.

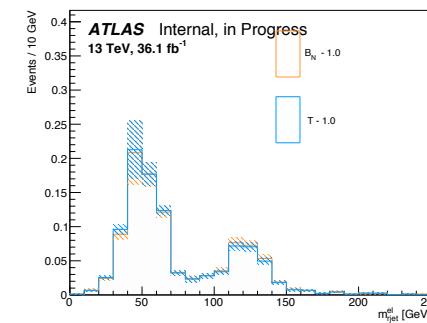
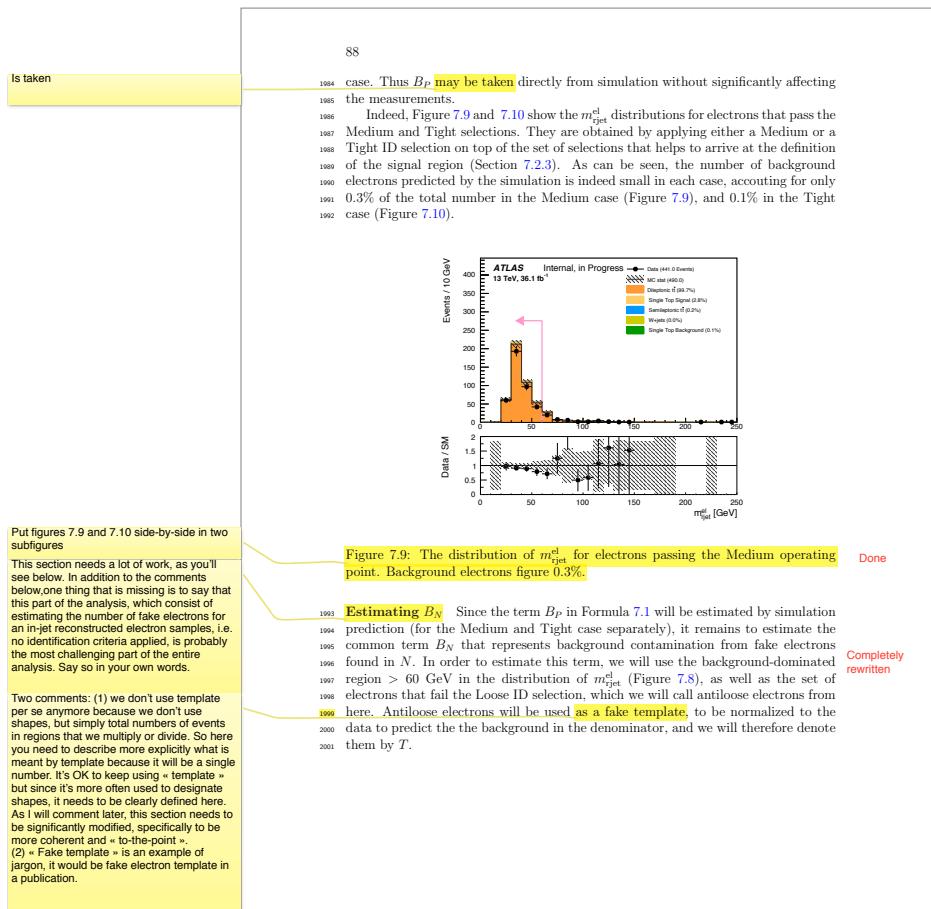


Figure 7.10: The distribution $m_{\text{jet}}^{\text{el}}$ of B_N against that of T , normalized to unity. T describes very well B_N and therefore it is reasonable to estimate B_N using T .

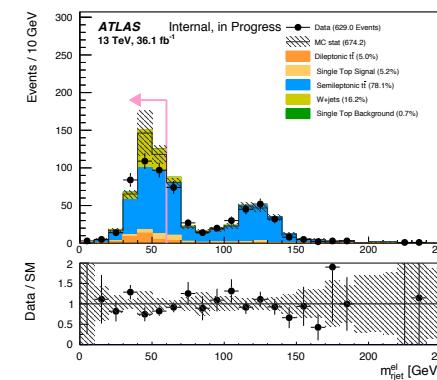
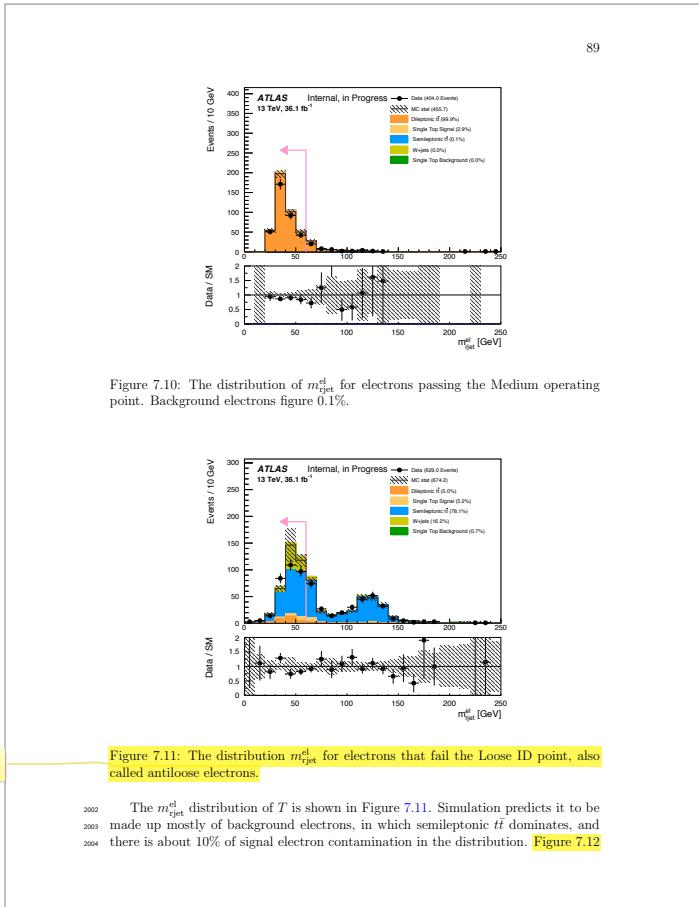


Figure 7.11: The distribution $m_{\text{jet}}^{\text{el}}$ for electrons that fail the Loose ID point, also called antiloose electrons.



Background estimation using antiloose electrons proceeds in detail as follows:

- First, T and $T_>$ are obtained by selecting antiloose electrons. Thus the method is data-driven, T and $T_>$ from simulations are not used.
- In addition to N , the set of reconstructed electron candidates in the signal region, there is also the set of reconstructed electron candidates in the background-dominated region, which will be denoted $N_>$. Signal contamination is subtracted from $T_>$, the resulting set of which is denoted $\bar{T}_>$, and signal contamination is subtracted from $N_>$, where the resulting set is denoted by $\bar{N}_>$. Then $\bar{T}_>$ is normalized to $\bar{N}_>$, to obtain a normalization factor.
- Signal contamination is subtracted from T , the resulting set of which is denoted \bar{T} , and the normalization factor is applied to \bar{T} , to obtain the number of background electrons in the signal region.

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

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

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

7.2.5 The Measurements of the Identification Efficiency

The identification efficiency ϵ shown in Formula 7.2, where B_P is taken from simulation and B_N is evaluated according to Formula 7.3, is

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

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

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

2028 signal contamination in $N_>$, provided the corresponding identification efficiency is
2029 properly taken into account. In other words,

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

2030 where the efficiency in 7.4, which is being measured, is used again. The effi-
2031 ciency will be evaluated iteratively, until the change from one iteration to the next
2032 is less than 0.5%. The value of 0.5% will be taken as the uncertainty due to signal
2033 contamination subtraction in $N_>$.

2034 The efficiencies, as well as the total statistical and systematic uncertainties (Sec-
2035 tion 7.2.6), are $0.870 \pm 0.017 \pm 0.031$ for Medium and $0.784 \pm 0.019 \pm 0.020$ for
2036 Tight. As is seen, the efficiency is higher for Medium than for Tight, consistent with
2037 expectation. The statistical uncertainties are slightly larger for Tight, also consistent
2038 with expectation, as the stats for Tight is slightly less than that for Medium. The
2039 relevant quantities in Formula 7.4 that are used to compute the efficiencies in data
2040 are listed in Table 7.2.

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

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

2041 The efficiencies and statistical uncertainties in simulation for the Medium and
2042 Tight operating points are also computed and are 0.871 ± 0.010 and 0.807 ± 0.011
2043 respectively. Thus Medium in data and in simulation agree, while there is a devia-
2044 tion of about 2% for Tight, possibly revealing the difficulty of modeling accurately
2045 electrons inside jets for the latter operating point.

2046 The next section discusses in detail the treatment of statistical and systematic
2047 uncertainties.

7.2.6 Uncertainties

2049 The measurement of the identification efficiency is accompanied by statistical and
2050 systematic uncertainties. The identification efficiencies, the statistical uncertain-
2051 ties, and the systematic uncertainties have been quoted in Section 7.2.5, they are
2052 $0.870 \pm 0.017 \pm 0.031$ and $0.784 \pm 0.019 \pm 0.020$ for Medium and Tight respec-
2053 tively. Thus the statistical uncertainty is approximately 2% for Medium and 2.4%
2054 for Tight, and the systematic uncertainty is higher, approximately 3.6% and 2.6%

Show this figure before 7.11 and use it in
your explanation above of what is T and how
it is used to estimate B_N . Right now it's not
obvious, and I think you'll need equations
also to do this property. Also need to say
clearly that 7.12 is simulated data

Simulated

Remove this part of the sentence, too
obvious/familiar

90

2005 shows T against the set of background part in N , namely B_N , both normalized to
2006 unity. It shows that T describes very well the shape of background electrons B_N ,
2007 and therefore it is reasonable to estimate B_N using T . However, instead of using
2008 the simulation distribution we will use the corresponding data distribution, so that
2009 in the following T will mean the corresponding distributions of m_{jet}^{el} in data. This
2010 is to ensure that that typical systematic uncertainties associated with simulation
2011 distributions could be avoided, because these systematic uncertainties are not small
2012 and are also often complicated to obtain.

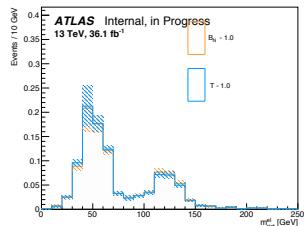


Figure 7.12: The distribution m_{jet}^{el} of B_N against that of T , normalized to unity. T
describes very well B_N and therefore it is reasonable to estimate B_N using T .

This paragraph is basically red-explaining
L1993-2001 now using equations, merge in
a single coherent and « to-the-point »
explanation that employs fig. 7.11 and 7.12
in the explanation at the appropriate places,
and also using equations like you start to do
here (but from the start)

2013 In the actual measurement of the efficiency, we need to take into account the
2014 fact that T itself is contaminated with some signal electrons, as already mentioned
2015 above. Thus let \bar{T} be T minus this signal contamination, which we will obtain by
2016 taking the distribution of T in data minus the signal contamination predicted in the
2017 simulation distribution of T . Then, the background-dominated region will be used
2018 to find a normalization factor such that

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

2019 where the normalization factor will be found as follows. To start, to any quantity
2020 in the signal region ≤ 60 GeV there corresponds a quantity in the background-
2021 dominated region > 60 GeV, which will be denoted with a subscript $>$. Thus to
2022 N there corresponds $N_>$, and to T corresponds $\bar{T}_>$. The term $N_>$ is the set of
2023 reconstructed electron candidates in the background-dominated region. In order to
2024 use it in the following we will subtract any signal contamination it may have, the
2025 method of which will be described in the next section, and denote $\bar{N}_>$ to be the
2026 resulting term, i.e. $N_>$ minus the signal contamination. Then

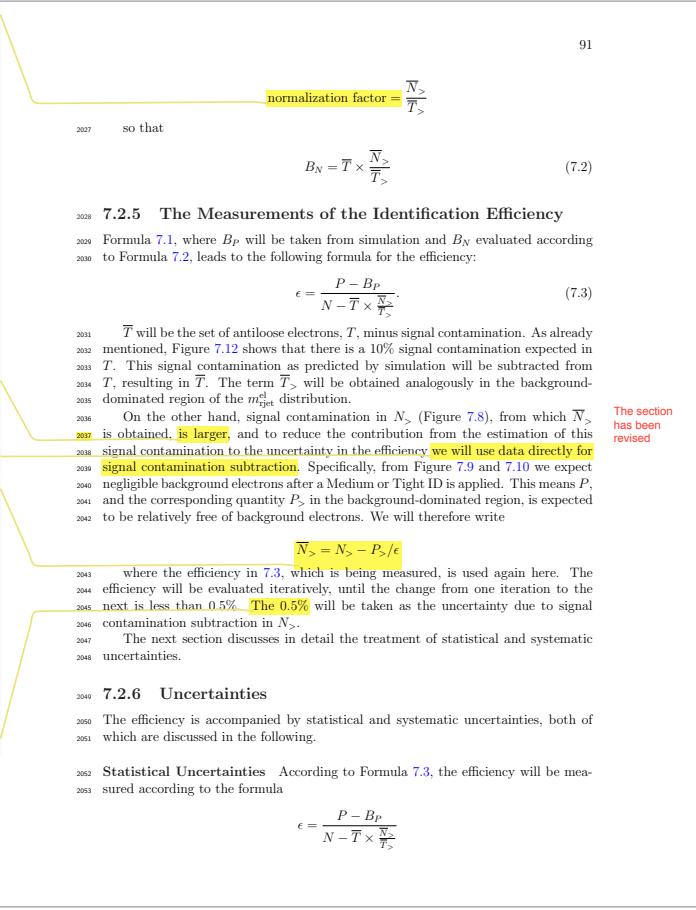
The end of that section is a bit fast, why is the normalization factor as such in words? Need to say that it is the ratio of the number of reconstructed electrons over antiloose electrons (not said anywhere in words so far). Also VERY important. This section does not describe the assumptions of the method. This is maybe the most important part of the section! You need to say that with this method, we assume that this ratio of reco to antiloose electrons is the same whether we loose or tight jet electron for the same sample. In other words, the efficiency to fail the loose selection is uncorrelated to The mass of the jet the electron belongs to. Say that. Also say that systematic uncertainties associated to the background estimation method will be described in section 7.X. Finally, background estimation section usually contain the results of the estimation. There's no reason not to provide the results here, as well as N'bar, T'bar, etc, is there? So please do so here, meaning to move part of the results tables, i.e. the part relevant to this section, here

Provide an actual approximate number from simulation to give a sense. This is probably the most recurring comment of the entire thesis: "always" avoid vague statements, we're doing physics!

Replace by «, we will use a data-drive approach »

IMPORTANT: explain this equation properly in words to complete the equation. This section is currently unclear because this is not done, and it suggests you don't fully understand what you do. So it needs to be said here that, to obtain the signal electron contamination to $N_>$, we will use directly the number of electrons passing the high mass region passing the medium or tight identification criteria. However, since these efficiencies are not 100%, the true number of signal electrons will be that number divided by the efficiency « epsilon ». which is what we want to measure. That efficiency will thus be estimated iteratively, etc. Need to say also what is the initial value of epsilon used? Also mention the number of iterations it took

That value of 0.5%



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

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

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

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 $\bar{N}_>$ the corresponding quantity in the background-dominated region minus signal contamination.
- \bar{T} is the set of antiloose electrons minus signal contamination, and $\bar{T}_>$ the corresponding quantity in the background-dominated region.

Since N contains P , and $\bar{N}_>$ contains $\bar{T}_>$, 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

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

The correlation between $\bar{N}_>$ and $\bar{T}_>$ remains, and moreover F and \bar{T} are also correlated, because in the Medium case or in the Tight case, F represents electrons failing Medium or Tight respectively, and since \bar{T} represents electrons failing Loose (minus signal contamination), in each case \bar{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 $\bar{T}_>$, to write

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

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

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

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

2054 where

- 2055 • P is the number of electrons that pass Medium or Tight.
- 2056 • B_P is background contamination due to fake electrons in P .
- 2057 • N is the set of reconstructed electron candidates, and $\bar{N}_>$ the corresponding quantity in the background-dominated region minus signal contamination.
- 2058 • \bar{T} is the set of antiloose electrons minus signal contamination, and $\bar{T}_>$ the corresponding quantity in the background-dominated region.

2061 Since N contains P , and $\bar{N}_>$ contains $\bar{T}_>$, the quantities in the formula are
2062 not all independent. We may remove the correlation between N and P by writing
2063 $N = P + F$, where F is the set of electrons that fail a particular ID point. Then

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

2064 The correlation between $\bar{N}_>$ and $\bar{T}_>$ remains, and moreover F and \bar{T} are also
2065 correlated, because in the Medium case or in the Tight case, F represents electrons
2066 failing Medium or Tight respectively, and since \bar{T} represents electrons failing Loose
2067 (minus signal contamination), in each case \bar{T} is a subset of F and there is accordingly
2068 a correlation.

2069 In order to remove all the correlations and write the efficiency completely in terms
2070 of statistically independent quantities we will first multiply both the numerator and
2071 the denominator by $\bar{T}_>$, to write

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

2072 Then we will add and subtract $\bar{T} \times \bar{T}_>$, to have

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

2073 The difference $F - \bar{T}$ represents the set of electrons that fail Medium or Tight
2074 but pass the Loose identification, and the difference $\bar{N}_> - \bar{T}_>$ represents the set of
2075 electrons that pass the Loose identification. If we treat each of the differences as
2076 a single term, and set $S = F - \bar{T}$ and $R_> = \bar{N}_> - \bar{T}_>$ respectively, the efficiency
2077 becomes

$$\epsilon = \frac{(P - B_P)\bar{T}_>}{P\bar{T}_> + S\bar{T}_> - R_> \times \bar{T}} \quad (7.4)$$

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 $R_> = \bar{N}_> - \bar{T}_>$ respectively, the efficiency
2082 becomes

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

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

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

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

$$\begin{aligned} \frac{\partial\epsilon}{\partial P} &= \frac{B\bar{T}_> - A\bar{T}_>}{B^2}, & \frac{\partial\epsilon}{\partial B_P} &= \frac{-B\bar{T}_>}{B^2}, & \frac{\partial\epsilon}{\partial \bar{T}_>} &= \frac{B(P - B_P) - A(P + S)}{B^2}, \\ \frac{\partial\epsilon}{\partial S} &= \frac{-A\bar{T}_>}{B^2}, & \frac{\partial\epsilon}{\partial R_>} &= \frac{AT}{B^2}, & \frac{\partial\epsilon}{\partial T} &= \frac{AR_>}{B^2}. \end{aligned}$$

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 ≤ 60 GeV, we may define it at ≤ 50 or ≤ 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).

93

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

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

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

$$\begin{aligned} \frac{\partial\epsilon}{\partial P} &= \frac{B\bar{T}_> - A\bar{T}_>}{B^2}, & \frac{\partial\epsilon}{\partial B_P} &= \frac{-B\bar{T}_>}{B^2}, & \frac{\partial\epsilon}{\partial \bar{T}_>} &= \frac{B(P - B_P) - A(P + S)}{B^2}, \\ \frac{\partial\epsilon}{\partial S} &= \frac{-A\bar{T}_>}{B^2}, & \frac{\partial\epsilon}{\partial \bar{R}_>} &= \frac{AT}{B^2}, & \frac{\partial\epsilon}{\partial \bar{T}} &= \frac{A\bar{R}_>}{B^2}. \end{aligned}$$

Remove or rephrase, statistical uncertainty always mean by convention the stats in data, not MC, and the current sentence mixes both concepts and is thus confusing

Add between these two words « , which are used for the background estimation, »

Avoid « we have decided », which is more likely to attract comments, and rephrase more neutrally

May be \rightarrow are

Same here, may be \rightarrow are. Also this sentence begs for an explanation. At least say an alternative template selection. How different is the mass shape of that template with respect to the default one?

Unclear, do you mean statistical uncertainty by « Contributions »?

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 $\bar{T}_>$, from which result \bar{T} and $\bar{T}_>$, 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 $\bar{T}_>$, $\bar{R}_>$, and \bar{T} in Formula 7.4 are treated as contributions to the total systematic uncertainty as well.

Removed

Has been revised completely

2108 • The simultaneous variations, either up or down, of the signal contaminations
2109 in T and $\bar{T}_>$, the subtractions of which from both terms give \bar{T} and $\bar{T}_>$. A
2110 25% variation is used for a conservative estimate of these contributions, which
2111 are **0.017** (approximately 2%) in Medium and **0.015** (approximately 1.9%) in
2112 Tight.

2113 • The change of the template T , from the distribution of antiloose electrons to
2114 the distribution of antiloose electrons in events with exactly 2 b -jets. The
2115 contributions to the total systematic uncertainty are **0.008** (approximately
2116 0.9%) for Medium and **0.007** (approximately 0.9%) for Tight.

2117 • The statistical uncertainties from the counting of $\bar{T}_>$, $\bar{R}_>$, and \bar{T} in Formu-
2118 la 7.5. They are 0.002 (approximately 0.2%), 0.008 (approximately 0.9%),
2119 and 0.002 (approximately 0.2%) respectively for Medium and 0.001 (approx-
2120 imately 0.1%), 0.006 (approximately 0.8%), and 0.001 (approximately 0.1%)
2121 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 .
- 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, 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.

2107 7.3 Identification Efficiencies

2108 In this section the integrated efficiencies as well as the binned efficiencies for the
 2109 Medium and Tight operating points are presented, along with the associated uncer-
 2110 tainties.

2111 7.3.1 Integrated Efficiencies

2112 The identification efficiencies for electrons inside jets are measured for the Medium
 2113 and Tight operating points; they are evaluated iteratively according to Formula 7.3,
 2114 which is

$$\epsilon = \frac{P - B_P}{N - \bar{T} \times \frac{N_P}{T_P}},$$

2115 where

$$\bar{N}_P = N_P - P_P / \epsilon$$

2116 The integrated efficiencies are presented in the following, along with the associ-
 2117 ated uncertainties.

2118 7.3.1.1 The efficiencies

2119 The efficiencies, as well as the total statistical and systematic uncertainties, are
 2120 listed in Table 7.2. It is seen that the efficiency is higher for Medium than for
 2121 Tight, consistent with expectation. The statistical uncertainties are slightly larger
 2122 for Tight, also consistent with expectation, as the stats for Tight is slightly less than
 2123 that for Medium.

	MEDIUM	TIGHT
Statistical Uncertainty	± 0.017	± 0.019
Systematic Uncertainty	± 0.031	± 0.020

Has been revised

2118 Table 7.2: Efficiencies, Statistical and Systematic Uncertainties in Data for the
 2119 Medium and Tight operating points.

2120 The relevant quantities in Formula 7.3 that are used to compute the efficiencies
 2121 in data are listed in Table 7.3.

Remove that table and simply replace as an
 equation of the results epsilon_medium =
 0.870 + bla + bla, etc. Put them in boxes to
 emphasize them.

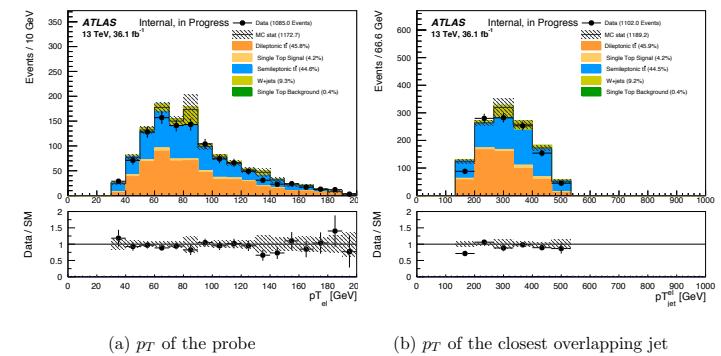


Figure 7.13: The distributions of p_T of the probe and p_T of the closest overlapping jet.

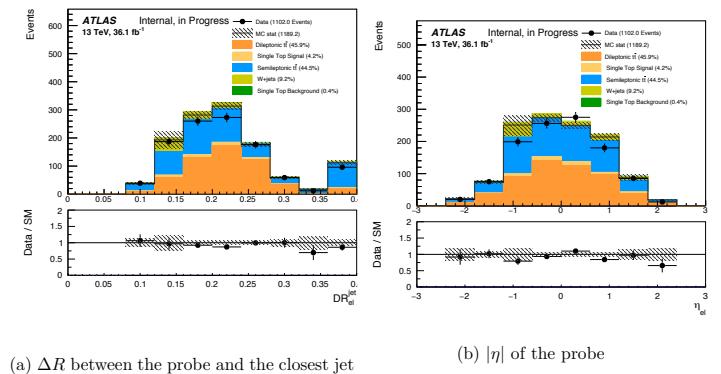


Figure 7.12: The distributions of ΔR between the probe and the closest overlapping jet and $|\eta|$ of the probe.

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 ΔR between the probe and the closest over-

Merge the cells of this line and used one in a single cell centered in the table, since these quantities are common to both medium and tight. Also use only one digit after the \times , *

	MEDIUM	TIGHT
P	392	356
B_P	1.47	0.40
N	734	734
N_S	368	368
$P_>$	49	48
T	267.35	267.35
$T_>$	292.52	292.52

95

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

Come on, you must know by now that you need to say something about the comparison between results in data and MC? SO what's the conclusion? Need to say so even if obvious (which is not necessarily the case here). Also, another open question that needs to be answered NOW is what's the results your measurement with respect to EGamma (the most vanilla result, i.e. $Z \rightarrow ee$ result)?

Same here, remove that table and provide as centered « equations »

Remove that table, overkill, only needed for data.

MEDIUM	TIGHT
0.871	0.807

HAS BEEN REVISED

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 ($< 0.5\%$).

Major comment: All of this section and the next one need to be merged with the section describing the systematics. I.e. once you describe the systematic uncertainty, also provide the quantitative result for that uncertainty. This is what is always done in papers. Since you won't have given the results already, provide the uncertainties in relative terms, i.e. fractional uncertainties (or whatever is understandable to the reader).

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 .

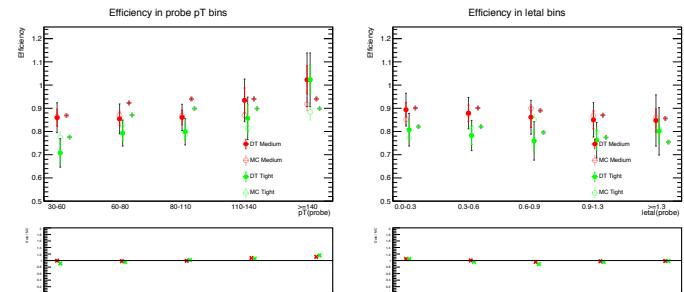


Figure 7.14: The efficiencies in p_T of the probe as well as in $|\eta|$ of the probe. Also shown are the efficiencies for standard electrons and the associated uncertainties (which are very small and therefore are barely visible).

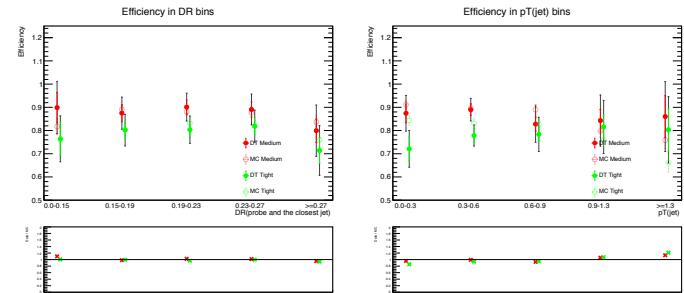


Figure 7.15: The efficiencies in ΔR between the probe and the closest jet, as well as in p_T of the closest overlapping jet.

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

in the period 2015–2016, at 13 TeV center-of-mass and totaled 36.5 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. The efficiencies as functions of the properties of the electrons and of the overlapping jets also measured. In all of the results, the efficiencies predicted by simulation agree with those obtained from the measurements on data.

96

See previous comment

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

HAS BEEN REVISED

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 .

There's too many tables in that section, what we want is a single table that describes the fractional uncertainty of each major uncertainty sources, so e.g. stat uncert.: 15%, template variation: 7%, stat uncert. from bkgd estimation: 12%, etc. this would be enough, and have only one such table in a single place. Unless otherwise noted, all tables in this section can be removed, and their results can be described in the same text that first describe what these uncertainties are.

Remove that table, one notch too detailed, table 7.3 will be enough

No need for a table if the result is trivial, this table can be removed and it's enough to describe the result in the text.

in the period 2015–2016, at 13 TeV center-of-mass and totaled 36.5 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. The efficiencies as functions of the properties of the electrons and of the overlapping jets also measured. In all of the results, the efficiencies predicted by simulation agree with those obtained from the measurements on data.

EMPTY

97

	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 uncertainty from B_P .

Contribution from signal contaminations in T and $T_>$. The contributions due to signal contamination 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	HAS BEEN REVISED
$\times 1.25$	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
$T \times 1.5$?	?
$T \div 1.5$?	?
$T_>$	317	317
Dilepton contamination	5.45	5.45
Single top signal contamination	19.03	19.03
$T_> \times 1.5$?	?
$T_> \div 1.5$?	?

Table 7.11

Haven't you estimated these uncertainties before?

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
\bar{T}	106.03	106.03
$\bar{T}_>$	114.49	114.49

Table 7.13

2159 **Contributions from re-marking the signal region** The contribution to the to-
2160 tal systematic uncertainty from re-marking the signal region is shown in Table 7.14.
2161 It is another major contribution to the total systematic uncertainty. The signal re-
2162 gion is marked, in place of at 60 GeV, at 50 GeV and 80 GeV in turn. The total
2163 contribution is computed as a sum of quadrature of the two individual contribu-
2164 tions. The relevant quantities used for the calculations of the efficiencies are listed
2165 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
$\bar{N}_>$	509	368
$P_>$	91	90
\bar{T}	182.76	182.76
$\bar{T}_>$	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.

2147 in the period 2015-2016, at 13 TeV center-of-mass and totaled 36.5 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.

EMPTY

in the period 2015-2016, at 13 TeV center-of-mass and totaled 36.5 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. The efficiencies as functions of the properties of the electrons and of the overlapping jets also measured. In all of the results, the efficiencies predicted by simulation agree with those obtained from the measurements on data.

EMPTY

99

	MEDIUM	TIGHT
P	420	384
B_P	1.56	0.43
N	734	734
$\overline{N}_>$	238	238
$P_>$	21	20
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.

This title is jargon, maybe « Efficiencies as a function of the properties of the electron and of the overlapping jet»

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,

HAS BEEN REVISED

○ $|\eta|$ 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

Don't make sub-bullets, simply list them on a line. Same applies to next three variables

2147 in the period 2015-2016, at 13 TeV center-of-mass and totaled 36.5 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.

EMPTY

- 100
- 2189 - 150-220 GeV,
- 2190 - 220-280 GeV,
- 2191 - 280-340 GeV,
- 2192 - 340-400 GeV,
- 2193 - 400-500 GeV.

2194 The distributions of these variables are shown in Figure 7.13 and 7.14. The
 2195 binned efficiencies are shown in Figure 7.15 and 7.16.

HAS BEEN REVISED

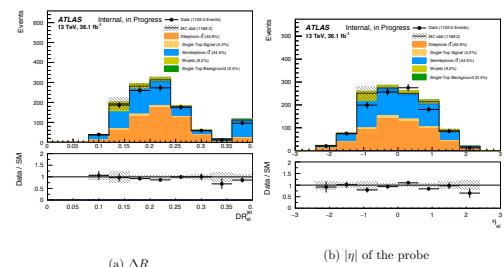


Figure 7.13: The distributions of ΔR between the probe and the closest overlapping jet and $|\eta|$ of the probe.

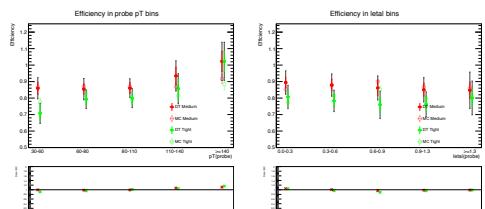


Figure 7.15: The binned efficiencies in p_T of the probe as well as in $|\eta|$ of the probe.

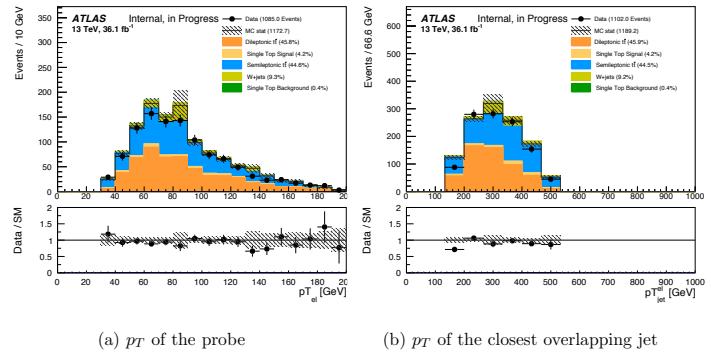


Figure 7.14: The distributions of p_T of the probe and p_T of the closest overlapping jet.

in the period 2015-2016, at 13 TeV center-of-mass and totaled 36.5 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. The efficiencies as functions of the properties of the electrons and of the overlapping jets also measured. In all of the results, the efficiencies predicted by simulation agree with those obtained from the measurements on data.

EMPTY

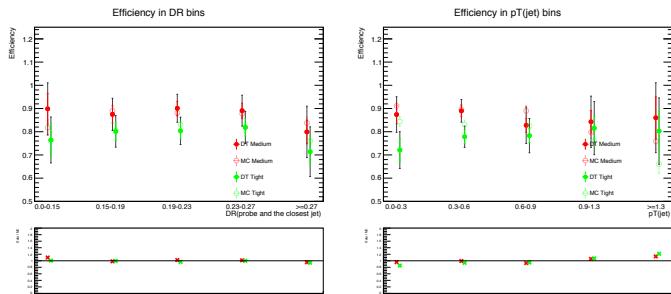


Figure 7.16: The binned efficiencies in ΔR between the probe and the closest jet, as well as in p_T of the closest overlapping jet.

2196 7.4 Conclusions

has been revised

, and the first ever using dilepton tbar events.

Only one digit after the , can you check that is the right number?

Rephrase

Quite close is too vague, either say they agree or not

2197 This chapter describes the work to measure the identification efficiencies for in-jet
 2198 electrons. It was the first attempt to perform such a measurement since Run 2
 2199 began. The measurement used the data collected in the period 2015-2016, at 13
 2200 TeV center-of-mass and totalled 36.47 fb^{-1} in integrated luminosity. A sample of
 2201 electrons for the measurements was obtained by selecting boosted $t\bar{t}$ dilepton ($e\mu$)
 2202 events. Background estimations used both simulations and data, and the efficiencies
 2203 were evaluated iteratively. The efficiencies were measured for the Medium and Tight
 2204 operating points, both on data and simulation. Efficiencies in bins were also mea-
 2205 sured. In all of the results, the efficiencies predicted by simulation are quite close to
 2206 those obtained from the measurements on data.

2147 in the period 2015-2016, at 13 TeV center-of-mass and totaled 36.5 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.

EMPTY