Study of charge misidentification probability for $Z \to e^+e^-$

Summer student: Sabin Roman, The University of Edinburgh

Supervisors: Alexander Glazov, Mikhail Karnevskiy

September 8, 2011

Abstract

This report presents an analysis of charge misidentification for the ATLAS detector in the Central-Central region based on the process $Z \to e^+e^-$. The main goal of the analysis is to estimate the charge misidentification probability: for all same charges and separately for $++$ and $--$ electron pairs.

Contents

1 Introduction 3

2 Selection of $Z \to e^+e^-$ events 5

3 Background estimation 6

4 Global charge misidentification 7

5 Charge misidentification in bins 9

5.1 Data and results 9

5.2 Final results and consistency check 10

6 Conclusion 11

1 Introduction

The ATLAS detector comprises a superconducting solenoid surrounding the inner detector (ID) and a large superconducting toroid magnet system around the calorimeters. The ID system is immersed in a 2 T axial magnetic field and provides tracking information for charged particles in a pseudorapidity range matched by the precision measurements of the electromagnetic calorimeter. The silicon pixel and strip tracking detectors (SCT) cover the pseudorapidity range $\eta < 2.5$. The liquid argon (LAr) based electromagnetic (EM) calorimeter is divided into one barrel ($\eta < 1.3$) and two end-cap components (abbreviated EMEC; we consider only $1.3 < \eta < 2.5$).

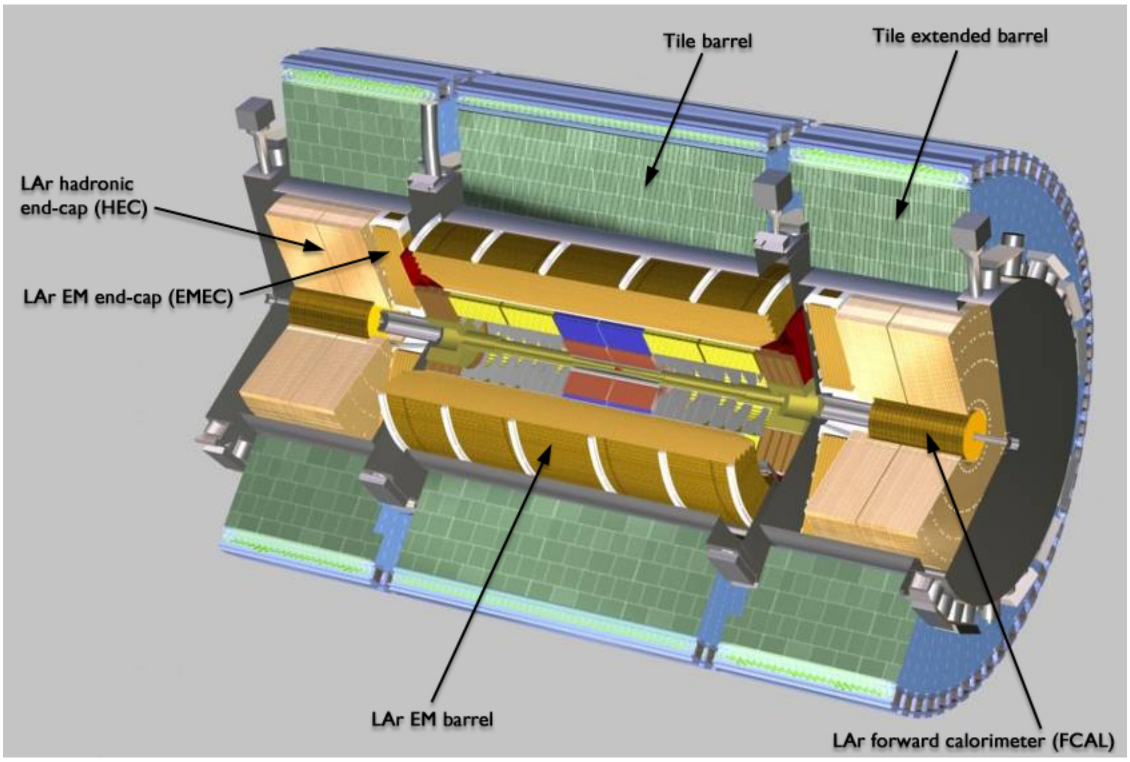


Figure 1: The ATLAS detector calorimeters

The ATLAS detector has a three-level trigger system consisting of Level-1 (L1), Level-2 (L2) and the Event Filter (EF). The L1 trigger rate at design luminosity is approximately 75 kHz. The L2 and EF triggers reduce the event rate to approximately 200 Hz before data transfer to mass storage. After the final stage the data is broken into streams. The stream we are working with is the Egamma stream, which means the data passed the electron and photon trigger. The analysis is based on EF_E20_Medium trigger which requires an electromagnetic object with transverse energy E_T >20GeV. The luminosity of the data is around 1.4 fb^{-1}.

The charge of the particles is determined from the bending of the patricle trajectories in the magnetic field. A correct charge determination is very important for particle identification and reduction of background. For example, searches for the standard model Higgs production and decay into ZZ, $pp \rightarrow H \rightarrow ZZ$ require exactly two pairs of oppositely charged leptons. Knowing charge identification efficiency is thus very important for determination of Standard Model physics. In this note a method of charge identification efficiency based on $Z \rightarrow e^+e^-$ events is described. The method uses the high statistics sample of low background $Z \rightarrow ee$ events[†] selected in 2011.

The selection of $Z \rightarrow ee$ events is described in Section 2. An estimation of background is given in Section 3. The global charge misidentification results for data and MC are presented in Section 4. Charge misidentification in bins is discussed in 5. Finally we conclude with Section 6.

[†]We refer to both electrons and positrons as "electrons". The signs are occasionally dropped.

2 Selection of $Z \to e^+e^-$ events

In working with data it is important to select relevant samples for analysis. The criteria applied in this selection are called "cuts". The event selection is based on selecting two electrons, satisfying kinematic cuts for $Z \to ee$ decays and electron identification cuts for the two scattered electrons. The electron identification cuts are encoded in so called ISEm bit mask with 32 bits. Each bit corresponds to a cut. If a electron passed the cut, the bit is 1, otherwise 0. Depending on the numbers of cuts passed, electrons are classified in: loose (first few cuts passed), medium, "M" (first 16) and tight, "T" (all cuts passed). Applying the tight cut leaves mostly signal with few background events, but because the cut is too severe real signal is also lost. A pre-selection was done and we selected events with 2 electrons of which one of the electrons is medium.
The kinematics cuts that were applied are:

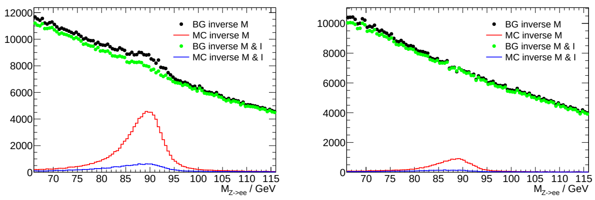
- For the transverse energy (to satisfy online trigger condition): E_T >20 GeV
- For the invariant mass: 66 GeV $< M_{ee} <$ 116 GeV
- For the pseudo-rapidity: $|\eta| < 2.47$, crack $1.37 < \eta < 1.52$

We start off with these cuts and add, invert or ignore additional electron identification cuts to obtain clean signal or background samples. In addition to the tight and medium identification cuts we consider electron isolation, "I", and opposite charge requirement. The isolation cut removes background from electrons associated with jets, e.g. heavy quark decays. The opposite charge cut is inverted for the study of the charge misidentification.

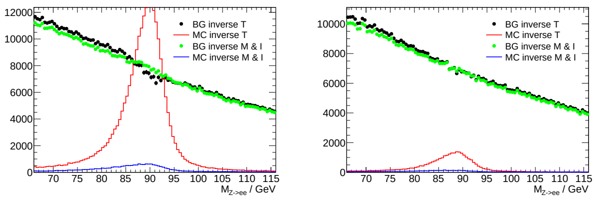
We use MC10b Monte Carlo sample which uses the Pythia event generator. The luminosity of the Monte Carlo is about 4.9 fb^{-1}. The MC was scaled by the ratio of luminosity of the data to the luminosity of the MC.

3 Background estimation

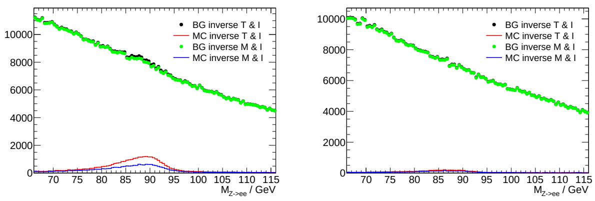
To get a reliable estimate of the charge misidentification rate we need accurate values for the number of same charge and opposite charge signal events, and for this we need a good estimate of the corresponding background. To estimate the background is our sample we first have to establish the background shape. We can get this background by appliying certain cuts (meant to reduce signal as much as possible), namely inverting the medium, tight or isolation cuts. We need to compare the backgrounds obtained using these different cuts so we choose the one with less signal. Here we see the different possibilities compared to the one we opted for: inverting the Medium and Isolation cuts. To get the background we apply these cuts to the data sample and then subtract the Monte Carlo with the same cuts. Figure 2 compares M_{ee} distributions for opposite charged (left) and same charge (right) backgrounds. For the opposite charge backgrounds the different possibilities have roughly the same shape and size, so it doesn't matter that much which is chosen. The same is true for the same charge backgrounds.



(a) inv M & inv M&I



(b) inv T & inv M&I



(c) inv T&I & inv M&I

Figure 2: Opposite & same charge background estimation by various cuts. The red and blue graphs represent the Monte Carlo with the same cuts which was substracted from the data to get the background.

4 Global charge misidentification

The charge misidentification probability for a single track is given by

$$\epsilon_{track} = \frac{\text{Tracks with wrongly reconstructed charge}}{\text{All tracks}}$$

We use events with two tracks. Each of the tracks can have charge misidentified. We define an event based charge misidentification ϵ as the ratio of the number of same charge signal events, N_{sg}^{sc} divided by the total number of signal events, N_{sg}. These numbers can be determined using the total number of same charge (all) events, N_{sg}^{sc} (N_d) and the total same change (all) background, N_{bg}^{sc} (N_{bg}). The misidentification probability is then:

$$\epsilon = \frac{N_{sg}^{sc}}{N_{sg}} = \frac{N_d^{sc} - N_{bg}^{sc}}{N_d - N_{bg}}$$

$$= \frac{N_{sg}^{sc}}{N_{sg}^{sc} + N_{sg}^{op}} = \frac{N_d^{sc} - N_{bg}^{sc}}{(N_d^{sc} - N_{bg}^{sc}) + (N_d^{op} - N_{bg}^{op})}$$

where N_d^{sc} is the number of events with electrons of the same charge, N_{bg}^{sc} is the number of background events with electrons of the same charge; so $N_d^{sc} - N_{bg}^{sc}$ gives the number of signal events with electrons of the same charge. N_d is the total number of events and N_{bg} is the number of events in the background; so $N_d - N_{bg}$ gives the total number of signal events (with opposite and same charge ee pairs).
We have

$$N_d = N_d^{op} + N_d^{sc}$$

where N_d^{op} is the opposite charge data. Also

$$N_{bg} = N_{bg}^{op} + N_{bg}^{sc}$$

and N_{bg}^{op} the number of opposite charged events in the background; it is estimated by comparing data with Monte Carlo and fitting the background level. N_{bg}^{sc} is obtained in a similar way but the opposite charge cut is inversed. To get the data we start off from the pre-selection, then apply some cuts to get real and misidentified signal. To get the opposite charge data we apply the: medium, isolation and opposite charge cuts. This will mostly consist of real signal events but still has some background events which have to be estimated and substracted. The invariant mass distribution of the opposite charge data, along with the corresponding MC and the fitted background level, is shown in Figure 3. To get the same charge data we invert the opposite charge cut, so that only events that fail this cut are saved. Again, this data has some background events. Figure 4 shows the invariant mass distribution for same charge events.

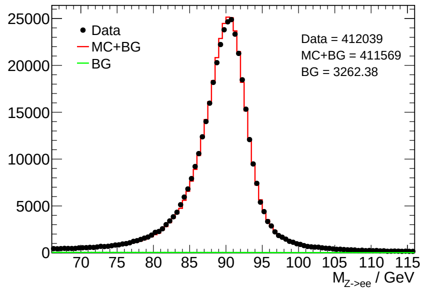


Figure 3: Invariant mass distribution for opp. charge events. The MC is scaled by the ratio: luminosity of data/luminosity of MC. The background is normalized using invariant mass side-bounds 66 GeV $< M_{Z\rightarrow ee} <$ 75 GeV and 100 GeV $< M_{Z\rightarrow ee} <$ 115 GeV; the background is small. The data and Monte Carlo agree well.

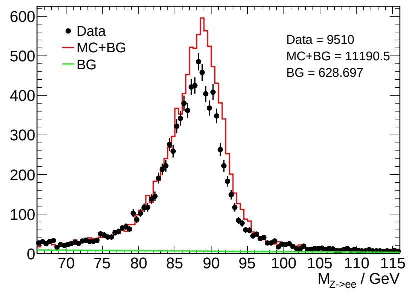


Figure 4: Invariant mass distribution for same charge events. The MC is scaled and the background is normalized as above. The background is again samll. MC overestimates the rate of same charge events.

Because $N_{sg}^{sc} << N_{sg}$, the statistical uncertainty for ϵ is given by $\pm\frac{\sqrt{N_{sg}^{sc}}}{N_{sg}}$. We can do the same analysis for ++ and - - pairs separately. Finally, we get these results:

Charge / $\epsilon(\%)$	**SC**	++	- -
Global (Data)	2.13 ± 0.02	1.08 ± 0.02	1.05 ± 0.02
Global (MC)	2.50 ± 0.03	1.28 ± 0.02	1.24 ± 0.02

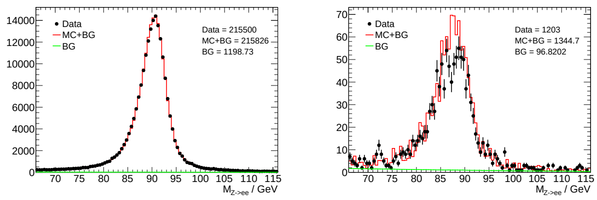
Table 1: Same charge, ++ and - - charge misidentification probabilities for data and Monte carlo

The charge misidentification probability is about 2 per cent. The ++ and − probabilities add up to the same charge misidentification probability; also, they are roughly equal. This is to be expected, the misidentification should be symmetric for ++ and - - pairs; the failure to identify charge properly shouldn't be biased. The Monte Carlo has a greater misidentification probability. As we saw the same charge histogram for data and MC did not fit well and this explains the difference.

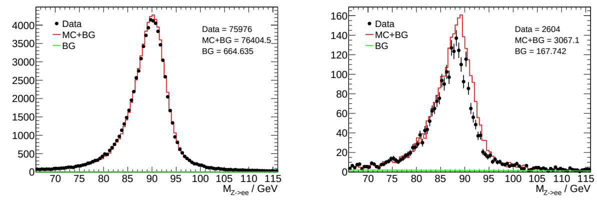
5 Charge misidentification in bins

5.1 Data and results

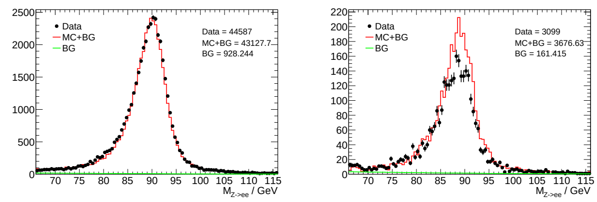
Charge misidentification depends on direction of the electron track since different tracking detectors are involved. Therefore it is interesting to analyses events depending on topology of the scattered electrons. The events can be classifed into three categories: both electrons fall in the barrel region (EMB) with $|\eta_{1,2}| < 1.37$, both electrons are reconstructed in EMEC with $1.52 < |\eta_{1,2}| < 2.47$ and one electron reconstruct in EMB, another in EMEC. These classes are labeled as Barrel-Barrel, EMEC-EMEC and Barrel-EMEC. The invariant mass distributions of the electrons for the three classes are shown in Figure 5.
We can do the above analysis for each bin. The graphs on the left are the invariant mass distribution for opposite charge events; the ones on the right are for the same charge events.



(a) Barrel-Barrel



(b) Barrel-EMEC



(c) EMEC-EMEC

Figure 5: Invariant mass distribution for opposite and same charge events in bins. The MC is scaled by the ratio: luminosity of data/luminosity of MC. The background is normalized using invariant mass side-bounds 66 GeV $< M_{Z\rightarrow ee} <$ 75 GeV and 100 GeV $< M_{Z\rightarrow ee} <$ 115 GeV; the background is small. Good agreement between data and MC for opp. charge; MC overestimates rate for same charge events.

The results are:

Charge / $\epsilon(\%)$	**SC**	++	**- -**
BB (Data)	0.5 ± 0.02	0.26 ± 0.01	0.25 ± 0.01
BB (MC)	0.6 ± 0.02	0.29 ± 0.01	0.29 ± 0.01
BE (Data)	3.1 ± 0.06	1.6 ± 0.05	1.5 ± 0.04
BE (MC)	3.7 ± 0.07	1.9 ± 0.05	1.8 ± 0.05
EE (Data)	6.3 ± 0.1	3.1 ± 0.08	3.2 ± 0.08
EE (MC)	7.7 ± 0.1	3.9 ± 0.09	3.8 ± 0.09

Table 2: Same charge, ++ and - - charge misidentification probabilities in bins for data and Monte carlo

As expected, the ++ and – probabilities add up to the same charge misidentification probability and we see that, again, the charge misidentification for ++ and - - pairs are roughly the same. We compute the same quantities for the Monte Carlo and we see that the probabilities are again higher than for the data. The results show the following trend: the misidentification get worse in the EMEC region. This is because there are different trackers in the Barrel and EMEC regions.

5.2 Final results and consistency check

We can deduce the probability that an electron is misidentified in the Barrel (p_B)and EMEC(p_E) regions. The equations satisfied by p_B and p_E are:

$$p_B(1-p_B) + p_B(1-p_B) = (0.5 \pm 0.02)\%$$
$$p_B(1-p_E) + p_E(1-p_B) = (3.1 \pm 0.06)\%$$
$$p_E(1-p_E) + p_E(1-p_E) = (6.3 \pm 0.1)\%$$

We have $p_B = (0.26 \pm 0.08^{\dagger})\%$ and $p_E = (3.26 \pm 0.06)\%$. We now compare $p_B(1-p_E) + p_E(1-p_B) = (3.5 \pm 0.06)\%$ with $(3.1 \pm 0.06)\%$. The consistency is not well satisfied but this could be due to a over-estimation of the charge misidentification probability in the Barrel-Barrel bin. The difference could also be due to a variation of the misidentification probability inside EMEC bin and different event topology for EE versus BE samples. For the Monte Carlo we have:

$$p_B(1-p_B) + p_B(1-p_B) = (0.6 \pm 0.02)\%$$
$$p_B(1-p_E) + p_E(1-p_B) = (3.7 \pm 0.06)\%$$
$$p_E(1-p_E) + p_E(1-p_E) = (7.7 \pm 0.1)\%$$

We get $p_B = (0.29 \pm 0.08)\%$, $p_E = (4 \pm 0.07)\%$ and need to check $p_B(1-p_E)+p_E(1-p_B) = (4.3 \pm 0.06)\%$ by comparing with $(3.7 \pm 0.07)\%$. We see that we have a larger discrepancy that for the data. This is to be expected because the Monte Carlo overestimates the charge misidentification probability in the Barrel bin a lot less than in the EMEC bin.

† Uncertainty computed using error propagation formulas.

6 Conclusion

A study of the charge misidentification probability using $Z \rightarrow ee$ events from a 2011 sample with integrated luminosity $L = 1.4\ \text{fb}^{-1}$ has been performed. The misidentification probability is studied in the barrel and end cap regions separately. A low $(0.26 \pm 0.08)\%$ misidentification probability is found for EMB region which is in agreement with MC estimate of $(0.29 \pm 0.08)\%$. A higher misidentification probability of $(3.26 \pm 0.06)\%$ is found for the EMEC region. For EMEC, the Monte Carlo simulation overestimate the misidentification probability giving a value of $(4 \pm 0.07)\%$.

References

[1] ATLAS Computing Group. ATLAS Computing. Technical Design Report. ATLAS TDR-017, CERN-LHCC-2005-022. 4 July 2005.

[2] The ATLAS Collaboration. Performance of the Electron and Photon Trigger in p-p Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector at the LHC. ATLAS-CONF-2011-114.

[3] ATLAS NOTE. An Analysis of Z, W Cross Section Determination in the Electron Channels with ATLAS.

[4] Authors: The ROOT team. Editors: Ilka Antcheva, Olivier Couet. ROOT. An object oriented data analysis framework. User Guide 5.26 December, 2009. http://root.cern.ch/.

[5] An Introduction to the Standard Model of Particle Physics 2nd Edition, *W.N. Cottingham and D.A. Greenwood*