DESY Summer Student 2011

Measurement of identification efficiency of electrons with close-by jets using ATLAS 2011 data

ATLAS

Authors: Philip Dießner
Julia Iturbe

Supervisor: Elin Bergeaas Kuutmann
Thorsten Kuhl
Hongbo Zhu

1 Introduction

The study of top quarks produced at the Large Hadron Collider (LHC) provides validation of the Standard Model (SM) and could play an important role in the discovery of new physics. Within the SM, the production of top anti-top pairs ($t\bar{t}$) is an important background in searches for the Higgs boson since this is likely to couple to the very massive top quark. In addition, the study of $t\bar{t}$ events may provide evidence for new physics that modifies the production and/or decay of top quarks. High energy events, like the ones at LHC, may produce $t\bar{t}$ pairs with an invariant mass that is much higher than the combined mass of the quarks. This will lead to events where the decay products of a top quark are boosted in the direction of flight of this quark, being detected in areas of the experiment that are close together. To find $t\bar{t}$ events, an electron or a muon is needed as a signature. It is necessary to find methods to separate electrons from close-by jets created by boosted tops, even if they are overlapping. The identification cuts used for these electrons have to be studied and improved.

In this report we present the analysis we did to calculate the efficiency and scale factor of the tight identification cut for $\Delta R < 0.4$ using the $Z \to e^+e^-$ decay. The efficiency of an identification cut refers to the ability of a certain set of selections to accurately identify an object, in our case, an electron. This is done via the tag and probe method which consists of selecting a well identified electron and pairing it first with a candidate electron at container level and then with a candidate electron that passes the identification cut and finally, the outcome of this two processes is compared. The scale factor is the comparison of the efficiency of the identification cut when calculated with data and when calculated with a Monte Carlo simulation. The purpose of calculating the scale factor is making the Monte Carlo simulations look more like real data.

In the next section we will introduce the physics of the top quark as a motivation for our analysis. Then we briefly present the LHC and some information concerning the ATLAS experiment and the ATLAS detector. In section 4 we discuss the way in which electrons are reconstructed in the detector. After that we introduce the event on which our analysis is based and explain with more detail the tag and probe method mentioned above. In section 6 we present two different variables used to calculate efficiencies; the invariant mass of the pair of electrons and the isolation variable. We compare the results for the number of events obtained with these two variables and after observing their agreement proceed to our analysis using the isolation variable. Using it, we calculate the efficiencies and scale factor for $\Delta R < 0.4$ and estimate the statistical and systematic uncertainties.

Natural units will be used for this report, setting $\hbar = c = 1$.

2 Motivation and Goals

2.1 Top Physics

The top quark has a mass of around $m_t \approx 173$ GeV and a life time of the order of $\tau \approx 10^{-25}$ s [1]. This time is too short for the top to hadronize, but it will decay into a bottom quark and a W boson. This boson will decay in a pair of light quarks, $q\bar{q}$ or a charged lepton and neutrino pair $l^{\pm}\nu_l$. Quarks hadronize under the strong force. Their energy and momentum will be distributed among arising hadrons and form a large shower in the calorimeter, a jet.

Physics beyond the Standard Model is expected to be observed at higher energy scales than the mass of now known particles. For example, there may be a new gauge boson with a mass in the order of TeV that has otherwise similar properties to the Z boson and

is therefore called Z'. This boson could decay into the heaviest quarks, a $t\bar{t}$ pair. Such an event will be seen as a peak in the invariant mass plot of the top pair and because there is no particle in the Standard Model that is heavy enough to decay into top quarks, it is a sign for a Z' or another $t\bar{t}$ resonance. Using the decay of the W bosons, one classifies such an event as dileptonic, when both W bosons decay into a $l^{\pm}\nu_l$ pair, as semileptonic when one decays in such a pair and one into quarks, and as hadronic when there are only quarks in the final state. In our case, to select a $t\bar{t}$ event we want a semileptonic event like in Figure 1. There we have a charged lepton, a neutrino and four jets, two from b quarks and two from a W boson, in the final state. The charged lepton is used to tag the event. The four jets, one of those b-tagged[1], are needed to calculate the invariant mass of the event. Also, one can calculate the missing energy transverse to the beam pipe that the neutrino carries away, because there must be energy conservation in this transverse plane. The variable of missing transverse energy E_T^{miss} describes the amount of energy missing to reach energy conservation. Events with only jets are harder to reconstruct because they have a high QCD background. Events with two charged leptons can be reconstructed best but the branching ratio of a top in such events is not high enough for sufficient statistics and two neutrinos carrying away energy also poses a problem.

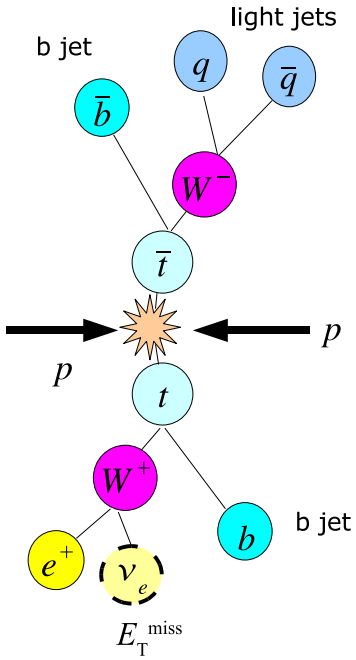

Figure 1: Sketch of a high energy collision creating a top pair that decays semi-leptonically. This leads to a charged lepton, four jets and missing transverse energy in the detector

When a Z' boson decays into a $t\bar{t}$ pair, the top quarks will gain a boost in momentum because of the excess in center-of-mass energy, if the mass of the Z' is large enough; $m_{Z'} > 1 - 1.5$ TeV. This boost will be also given to the decay products of the quarks and may result in an overlap of the created electron and the jet from the bottom quark in the detector. New algorithms are needed to discriminate those electrons because events with such overlap are not used for normal Standard Model analyzes.

[1]Using special properties and algorithms to test if a jet originates from a b quark.

3 LHC and ATLAS

3.1 LHC

The LHC (**L**arge **H**adron **C**ollider) is a storage ring collider with the world's highest center-of-mass energy colliding protons at $\sqrt{s} = 7$ TeV. It is located at CERN and has the goal to find or exclude the Higgs boson and to look for physics beyond Standard Model. There are four main experiments at the LHC: ALICE, ATLAS, CMS and LHC-b. In Figure 2 a schematic drawing of the LHC and the experiments is shown.

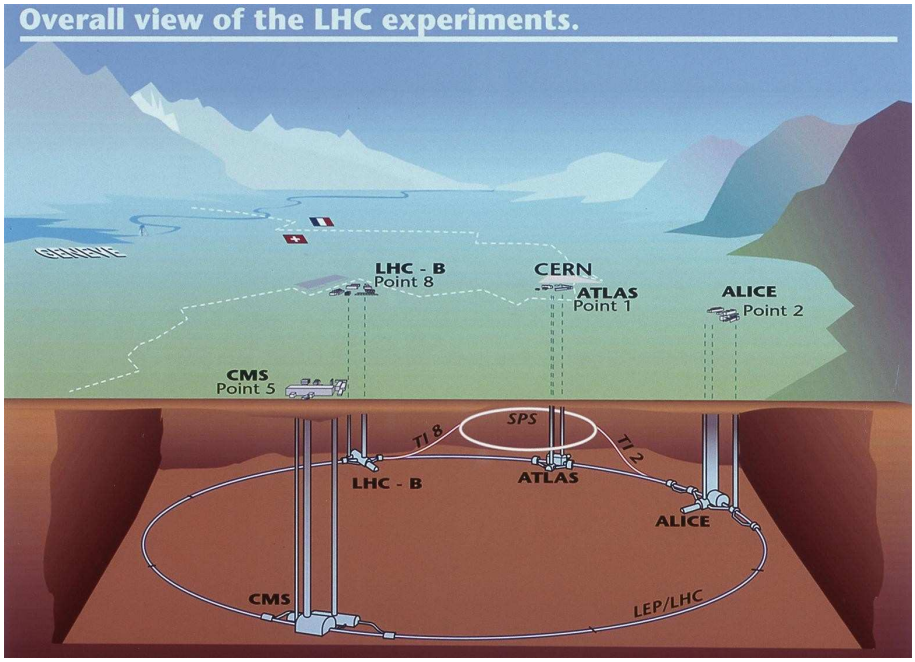

Figure 2: The layout of the LHC ring and the experiments (source: CERN)

The collider is located around 100 m underground to have some protection from cosmic radiation and it has a circumference of 26.7 km. The tunnel holds two adjacent parallel beam pipes, each containing a proton beam that travel in opposite directions and are brought to collision at the sites of the four major experiments. Superconducting magnets are used to control the beams. Dipole magnets are used to keep the beams on their path and quadrupole magnets to focus them. The protons do not travel continuously distributed in the beam pipes but in bunches. That means there is a spacing between proton packages so that the collisions happen at discrete intervals that are longer than 25 ns, providing a highest event frequency of 40 Mhz.

The instantaneous luminosity L is defined as

$$L = \frac{f_{rev} n_{bunch} N_p^2}{4\pi\sigma_x\sigma_y}. \tag{1}$$

where f_{rev} denotes the revolving frequency of a bunch, n_{bunch} the number of bunches and N_p the number of protons in a bunch. The term $4\pi\sigma_x\sigma_y$ is a description for the effective collision area of the beams, where σ_i is the spread in the dimension i. The instantaneous luminosity is used to describe the intensity of the beam. The design luminosity of the LHC is 10^{34} cm^{-2}s^{-1}. When the instantaneous luminosity is integrated over time, the

expected number of events N for a certain process can be calculated as

$$N = A\epsilon\sigma \int L \, \mathrm{d}\, t. \tag{2}$$

Here σ denotes the cross section of the process, A the acceptance of the detector and ϵ the efficiency to select the event out of the data. The integrated luminosity is used to describe the amount of data that is collected at LHC. Currently there are more than 2 fb^{-1} collected.

3.2 The ATLAS detector

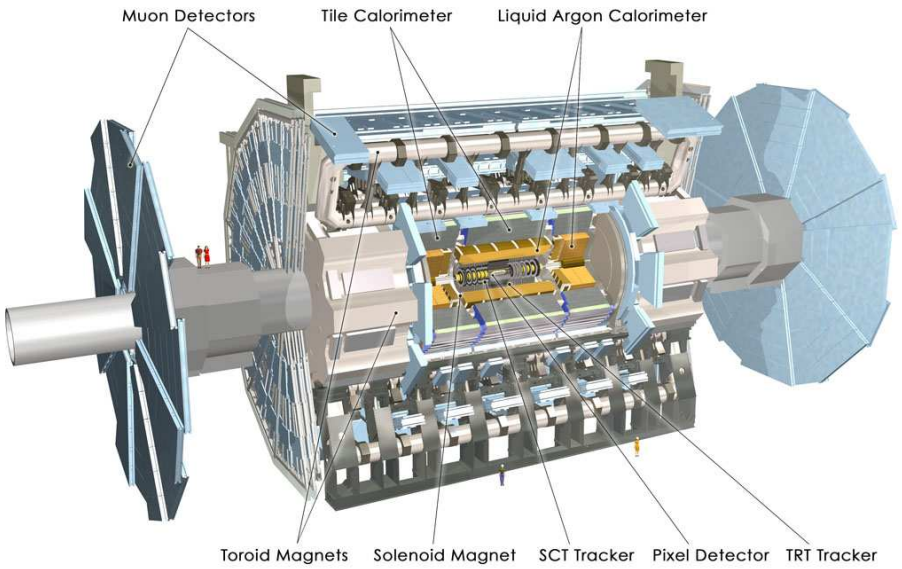

Figure 3: The ATLAS experiment (source: CERN)

ATLAS[2] (**A** **T**oroidal **L**HC **A**pparatu**S**) is used as a general-purpose detector. Its general structure can be seen in Figure 3. Closest to the beam pipe and the interaction point of the two proton beams is the inner detector surrounded by a magnetic field. With this part it is possible to measure the momenta of the created charged particles using Lorentz force and find information to identify them. The tracking detector system has pixel, SCT(Semiconductor Tracker) and TRT (Transition Radiation Tracker) sub systems. The pixel and SCT detectors consist of semiconducting pixel and micro-stripe sensors; the TRT consists of straws filled with a Xe-based gas mixture. The pixel detector is closest to the interaction point and therefore needs a high accuracy and radiation hardness. The SCT sensors use the classic principle of semiconducting detectors for accuracy and reliability. The resolution of the TRT is not as good as with semiconducting technology but it provides additional electron/pion separation.

The next part outside the magnetic field consists of the calorimeters. These are used to measure the energy of the particles through the absorption of them. The inner calorimeter is the electromagnetic calorimeter; it is layed out to be effective for electromagnetic interactions like bremsstrahlung and pair production and can so absorb electrons and photons. Both barrel and endcap are lead liquid argon systems. In the outer hadronic calorimeter mesons and baryons that passed mainly through the electromagnetic calorimeter will be absorbed by means of strong interaction and ionization. The barrel is made of steel and

scintillation tiles and the endcap, of copper and liquid argon. The outermost layer consists of the muon chambers. They are needed to identify and precisely measure the momenta of outgoing muons that are not absorbed in the calorimeters. For the selection of events of interest a three-level trigger system is used. The first level uses information from a local subset of detectors. It looks for muons, electrons, photons, jets, and tau-leptons decaying into hadrons with a high transverse momentum, as well as large missing and total transverse energy. With this information, regions of interest are built that specify certain areas in the spatial space which hold interesting features. These regions are forwarded to the second level trigger. This uses all available data in the regions to select if a certain event is interesting enough to be processed at the last trigger level. At the third trigger level, the event filter, data from the entire event is analyzed and, at the end, the event rate is reduced to $200 - 400$ Hz. These events are recorded for analysis.

The needed spatial coordinates are the angles of the spherical coordinates; the polar angle ϕ and the azimuthal angle θ. θ is measured with respect to the beam pipe and the origin at the collision point. The transverse plane lies perpendicular to the beam pipe.

θ defines the pseudorapidity η as

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{3}$$

A more detailed description of the ATLAS experiment can be found at [2].

4 Object Reconstruction

Using the tracking chambers and calorimeters one measures the direction of a outgoing particle and the total energy E.

For this report information from the tracking detectors and the electromagnetic calorimeter where taken into account. The electron candidates are required to have $|\eta| < 2.47$ for the cluster in the calorimeter but excluding the region $1.37 < |\eta| < 1.52$ with the transition region between barrel and endcap calorimeter. The wanted electrons originate from a Z boson and a cut to the transverse momentum is applied: $p_T > 20$ GeV. For jet reconstruction the anti-k$_t$ jet finding algorithm [3] was used.

4.1 Data and Monte Carlo

The integrated luminosity of the used data sample amounts to 1.0 fb^{-1} and was collected in 2011. A single electron trigger with an E_T threshold of 20 GeV called EF_e20_medium was used.

To have reliable comparison between data and theory one uses Monte Carlo (MC) simulation to generate events by computer randomly following a distribution predicted by theory. This generated events will be compared to the data from experiment using a wide range of statistical methods and analyzing tools. Doing so allows us to verify or reject theories when observing a lack or excess production of events in the data for certain areas.

There are areas where simulation can not reproduce data exactly because there is no theory reproducing reality perfectly. Because of this, in these areas. one needs to bring Monte Carlo to the level of data by hand. This is done with the help of scale factors (SF) which are extracted from variables that are easy to analyze, like later explained efficiencies, and than can be applied to more difficult analyzes to correct the MC modeling.

For this report a data set created with the PYTHIA [4] event generator and the GEANT4 [5] detector simulation was used.

5 Methodology

5.1 $Z \rightarrow e^+e^-$

The Z boson, along with the W boson, is the intermediator of the weak interactions. The Z boson is electrically neutral and it is its own antiparticle. It has a mass of (91.1876 ± 0.0021) GeV and baryon and lepton number zero. [1] Both the W and Z bosons are produced at high rates at $\sqrt{s} = 7$ TeV proton-proton collisions at the LHC and they provide the most common source of isolated, high p_T leptons. They are also important backgrounds for beyond the Standard Model searches. To calculate efficiencies in our analysis we use the decay of the Z boson into an electron and a positron. The reasons for using this event are: high statistics, since this process is produce at quite high rates at the LHC, and the fact that the Z bosons guarantee clean samples with backgrounds relatively easy to calculate.

5.2 Electron Identification

To find electrons out of the possible candidates in an event one starts at the so called *container level*. Here every electron candidate is contained that has tracks from the inner detector connecting to shower areas in the electromagnetic calorimeter. At this level, normally all electrons are included but there is also a substantial background from jets with the same signature. To reduce the background a series of *identification cuts* is introduced with rising exclusion potential. Using weak cuts it is possible to reduce some of the jet background, while keeping most electrons. At the *tight* level one has excluded almost all background, but also lost signals from real electrons that did not appear clearly in the detectors. Examples for cuts are that the shower has a certain maximum width or that the matching in η and ϕ between the track in the inner detector and the shower in the calorimeter has a certain upper limit. For an exact definition of the different identification parameters see Appendix C and [6]. In this report we will use *container level* and *tight* electrons.

5.3 Tag and Probe

To calculate the efficiency of a certain selection we use a method called "tag and probe" using $Z \rightarrow e^+e^-$ events. This method starts by searching for a good electron, "tag", which satisfies certain types of criteria. Then the efficiency of interest is measured testing the cuts on a second electron candidate called "probe" electron. For the tag, we selected a *tight* electron and pair this with every other electron candidate in the event that has opposite charge. After forming these pairs, we calculate their Lorentz vectors and we fill a histogram with the invariant mass of the sum of this vectors. What we get is a very clear peak around the mass of the Z boson. Therefore we are confident that the electrons we have chosen are in fact products of a Z boson decay. Such a histogram is shown in Figure 4. Table 1 contains information about the parameters used for this report.

The efficiency is defined as

$$\epsilon_{tight} = \frac{N_{\text{tight}}}{N_{\text{container-level}}}, \tag{4}$$

where N denotes the number of electrons at each level. It gives information about the number of electrons that appear at container level and pass the identification cuts. Different methods to approximate those efficiencies will be discussed later.

Using efficiencies one can define scale factors (see chapter 4.1) for Monte Carlo as

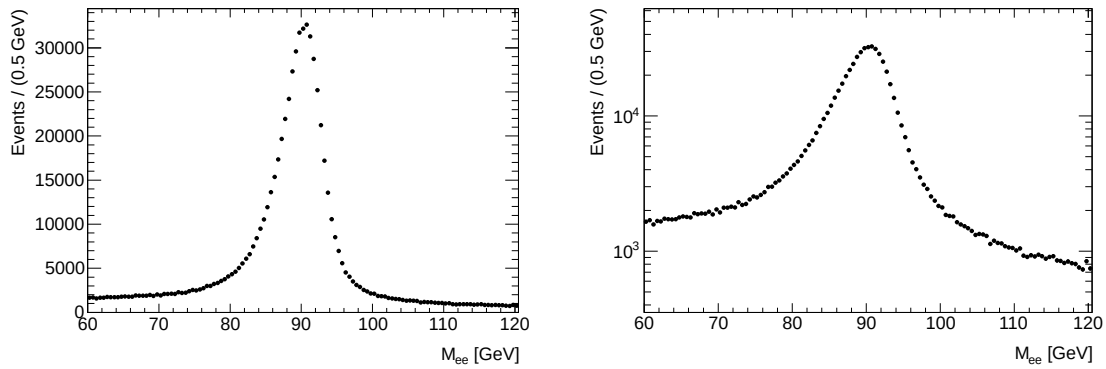

Figure 4: The invariant mass of an electron pair; consistent with a Z peak of about 91 GeV, indicating that the electrons indeed are the decay products of a Z boson, Monte Carlo created by PYTHIA [4] and GEANT4 [5]; normal and logarithmic scale.

$$SF = \frac{\epsilon_{data}}{\epsilon_{MC}}. \tag{5}$$

The scale factor can be parametrized with regard to certain variables; e.g. η, ϕ or ΔR and later be used to correct the Monte Carlo modeling.

Tag	Probe
Identification: Tight Systematic variation: $E_T^{cone}(0.2) < 4$ GeV	Identification: Tight, container level
$\|\eta\| < 2.47$, excluding $1.37 < \|\eta\| < 1.52$ $p_T > 20$ GeV	
Event selection:	trigger: EF_e20_medium invariant mass window: $(80, 100)$ GeV Systematic variations: wide mass window $(75, 105)$ GeV narrow mass window $(85, 95)$ GeV

Table 1: The identification cuts used in this report for tag and probe including the event selection parameters and variations for systematic uncertainties.

5.4 Analysis Basics

In a collision event a number of particles are produced and many variables are needed to describe it appropriately. The variables that are necessary for this report will be explained here. The variable ΔR describes the distance between electron and jet in a detector and is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \tag{6}$$

where $\Delta\eta$ and $\Delta\phi$ denote the difference in angular coordinates of the electron and jet.

Another used variable is E_T^{Cone}, which describes the isolation of an electron with respect to its surroundings. It is the sum over all the transverse energy E_T in the electromagnetic calorimeter in a cone with a certain R_0 around an electron candidate and minus the the

energy deposit of the electron candidate.

$$E_T^{Cone}(R_0) = \sum_{\Delta R < R_0} E_T - E_T(\text{electron candidate}) \tag{7}$$

In our case, $R_0 = 0.2$ is used and therefore $E_T^{Cone}(0.2)$. The shape of E_T^{Cone} is significantly different for electrons and for jets. Therefore, it can be used to discriminate between those two.

With E, p_T, η and ϕ one can calculate the four-momentum p of a particle. Adding the four-momenta of two electrons it is possible to calculate the invariant mass of a possible origin particle, like a Z boson, as

$$M_{ee} = \sqrt{(p_{e_1} + p_{e_2})^2} = \sqrt{(E_{e_1} + E_{e_2})^2 - (\vec{p}_{e_1} + \vec{p}_{e_2})^2}. \tag{8}$$

6 Measurement of Efficiencies

As next step we will discuss two methods for calculating efficiencies using the invariant mass M_{ee} or the isolation variable $E_T^{cone}(0.2)$.

6.1 Using Invariant Mass (Method 1)

Using the tag and probe method we select pairs of electrons with the characteristics described in Table 1 and we create two histograms with the invariant mass of these pairs, one for the pair of electrons in which the probe passes the tight cut and one for the pair of electrons in which the probe does not necessarily passes the cut. In both histograms we get a peak around the mass of the Z boson indicating that the pairs of electrons in this region most likely came from the Z boson decay. Also, in the histograms we get background from events that behave like electrons but are actually jets that look like electrons. We then proceed to calculate the efficiency of the cut. For this, we count the number of signals (or pairs) that we get in a mass window of 80 to 100 GeV after subtracting the background.

We do this by using predefined functions to fit the signal and the background, then we substract the background and integrate over the number of signal events in the mass window. For fitting this variable we use the Breit-Wigner function convoluted with the Crystal Ball function for the signal. For the background we use either an exponential function or the RooDecay function. This last one is a function included in the ROOT [7] library Roofit and consists of a Gaussian peak convoluted with an exponential decay. The Crystal Ball function is described by a Gaussian peak, which on one side is cut off and replaced with a power-law function. This is needed to describe the effects of the usage of a detector to the resonance peak of the Z boson. Examples of these fits are shown in figure 5 and their respective formulas can be found in Appendix D. Once we have the numbers for the container level electron and the tight electron, we apply the definition of efficiency.

6.2 Using Isolation (Method 2)

The variable which is commonly used to calculate efficiencies is the invariant mass but in our analysis we introduced another variable, the isolation variable $E_T^{cone}(0.2)$ (see equation 7). The shape of this variable is different for electrons and jets. For electrons it gives a relatively narrow peak around 0 GeV while for jets it is a broad peak with the center at some higher energy. The reason for this lies in the wide opening angle a jet shower in the calorimeter normally shows compared to an electron shower. Because some standardized

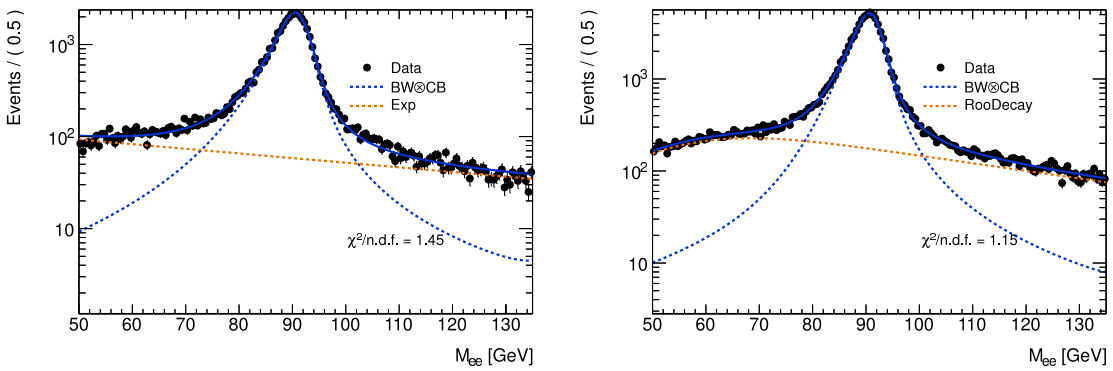

Figure 5: Fitting the invariant mass M_{ee} distribution with a Breit-Wigner function convoluted with a Crystal Ball function and both exponential (left) and RooDecay (right) background (Data with integrated luminosity of 1 fb^{-1}).

electron shower size is used to measure the E_T(electron-candidate) for the $E_T^{cone}(0.2)$, a jet will have a higher energy deposition outside this space.

To calculate the efficiency of the tight cut we again used the tag and probe method to create histograms of the isolation variable. The histograms for container level and tight cut electrons were created with pairs of electrons whose invariant mass was in the window $(80, 100)$ GeV. In the case of the tight cut we obtained practically no background and counting the entries was enough to obtain the number of signal events. For the electrons at container level we used the RooDecay function for the signal and the Landau function for the background and we proceed as in the case of the invariant mass. Such a fit is shown in Figure 6. For the functions see Appendix D.

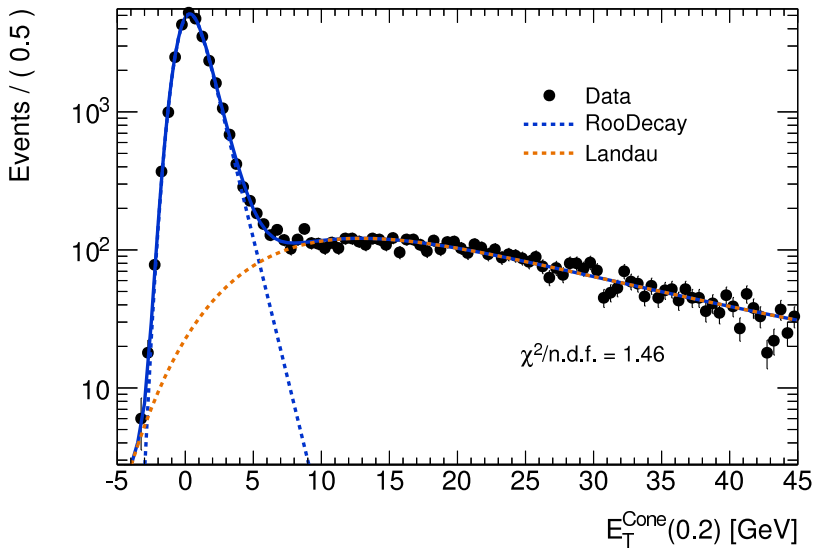

Figure 6: Fitting the isolation distribution with a RooDecay function as signal and a Landau function as background (Data with integrated luminosity of 1 fb^{-1}).

Using the isolation variable is preferable to the invariant mass for the case of electrons and jets with small separation. When observing the invariant mass of the electrons one finds a leakage of energy from the jet into the electron cluster when ΔR gets smaller. This

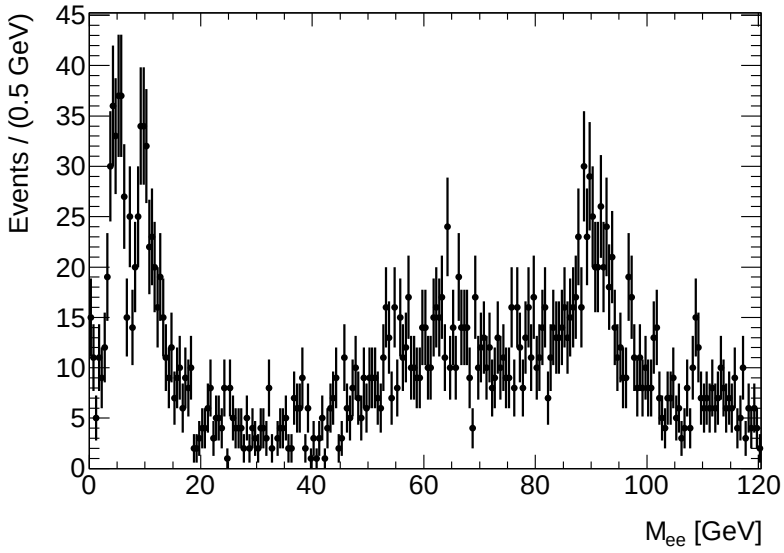

Figure 7: Distribution of the invariant mass for small ΔR. Comparing this to Figure 5, one sees the distortion of the whole range (Data with integrated luminosity of 1 fb^{-1}).

leads to a deformed peak in the mass spectrum as seen in Figure 7 which cannot give the best information about the number of signal events. Another upside of using this variable is that the method would be independent of the electron's origin and therefore could be extended to the boosted top case. In order to probe if there is any possible bias, the number of signals of the cut is plotted as a function of bins in the variables η and ϕ. In the Figures 8 and 9 we compared the number of signal events we get in the mass window $(80, 100)$ GeV using the invariant mass distribution and the isolation variable distribution and we found that these numbers differ by less than 3 percent. Because the difference in the $\eta - \phi$ space is small enough, it is reasonable to switch from the invariant mass to the isolation variable for our analysis.

6.3 Efficiency Calculation of Monte Carlo

Calculating the efficiencies of the simulation datasets is easier to accomplish and has a higher precision. There exist variables for each electron candidate in a MC dataset that tells if an electron comes from a Z boson.

These variables allow that simple counting of the wanted electrons in the *container-level* and after the *tight* cut is sufficient to get $N_{\text{container-level}}$ and N_{tight}. The efficiencies are then calculated the same way as the for the data.

6.4 Uncertainties

6.4.1 Statistical Uncertainties

The statistical uncertainty of the efficiencies ϵ_{tight} coming from data is given by

$$\Delta\epsilon_{tight} = \frac{\sqrt{(1-2\epsilon)\left(\Delta N_{\text{tight}}\right)^2 + \left(\epsilon_{tight} \cdot \Delta N_{\text{container-level}}\right)^2}}{N}. \tag{9}$$

This is the commonly accepted way to calculate this uncertainty following [8].

Using error propagation, one can derive the uncertainty of the scale factor SF (eq. 5) and obtain

$$\Delta SF = \sqrt{\left(\frac{\Delta\epsilon_{Data}}{\epsilon_{MC}}\right)^2 + \left(\frac{\epsilon_{Data}\Delta\epsilon_{MC}}{\epsilon_{MC}^2}\right)^2} = SF\sqrt{\left(\frac{\Delta\epsilon_{MC}}{\epsilon_{MC}}\right)^2 + \left(\frac{\Delta\epsilon_{Data}}{\epsilon_{Data}}\right)^2}. \quad (10)$$

6.4.2 Systematic Uncertainties

In our analysis we made some considerations that could be thought as partially arbitrary, for example the choice of the mass window we worked with. However, one can estimate how much this considerations actually affect our results. In order to do so, we altered some of these considerations. We calculated the efficiency of the tight cut broadening the mass window to $(75, 105)$ GeV and then narrowing it to $(85, 95)$ GeV. Also, we imposed the condition $E_t^{Cone}(0.2) < 4$ GeV on the tag electron. Finally, we change the fitting function for the signal from RooDecay to Crystal Ball. The first conservative estimate of the systematic uncertainties uses the spread of the central value and can be calculated as

$$\Delta SF_{sys} = \sqrt{\sum (SF_{\text{std}} - SF_{\text{var}})^2}, \quad (11)$$

where SF_{std} denotes the scale factor with the mass window $(80, 100)$ GeV and SF_{var} the scale factor with variations in different parameters. The sum goes over all the different variations taken.

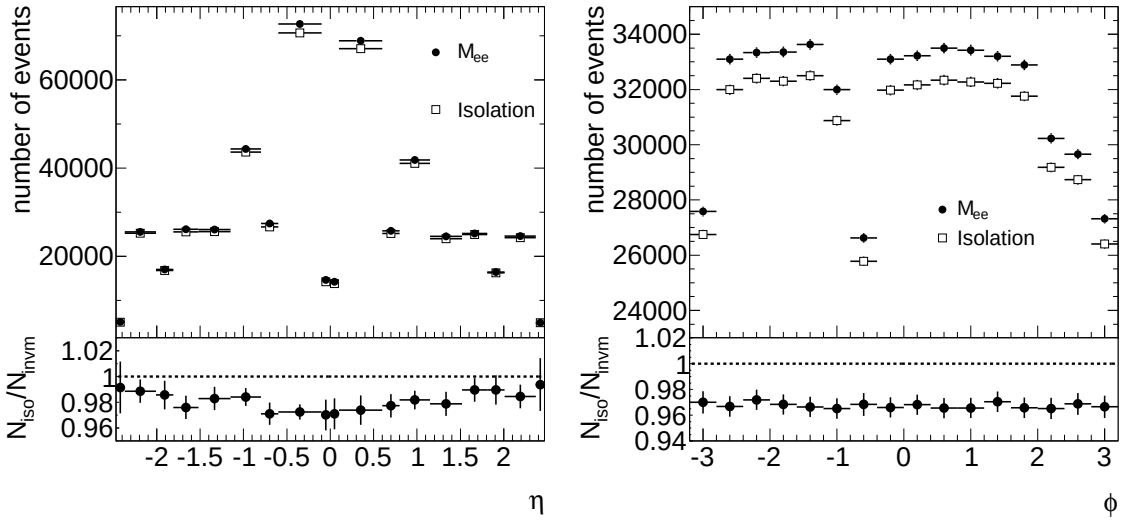

Figure 8: Comparing the number of electrons in the data at *container level* using method one (exponential background) and two as function of η and ϕ. (Data with integrated luminosity of 1 fb^{-1})

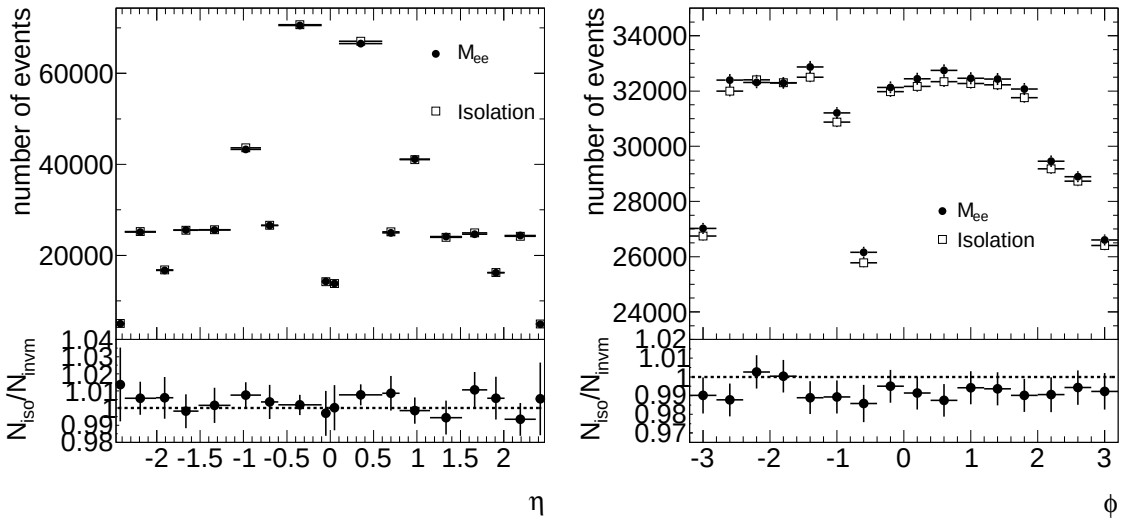

Figure 9: Comparing the number of electrons in the data at *container level* using method 1 (RooDecay background) and 2 as function of η and ϕ. (Data with integrated luminosity of 1 fb^{-1})

7 Results

	$\epsilon_{tight;Data}$	$\epsilon_{tight;MC}$	SF_{tight}
mass window $(80-100)$ GeV and RooDecay function for signal fit in data	0.52 ± 0.03	0.70 ± 0.01	0.745 ± 0.045
with $E_T^{Cone} < 4$ GeV cut on tag	0.52 ± 0.03	0.69 ± 0.01	0.75 ± 0.05
mass window $(75-105)$ GeV	0.45 ± 0.03	0.69 ± 0.01	0.65 ± 0.04
mass window $(85-95)$ GeV	0.53 ± 0.04	0.69 ± 0.01	0.76 ± 0.05
Crystal Ball function for signal fit in data	0.33 ± 0.04	0.70 ± 0.01	0.47 ± 0.05

Table 2: Efficiencies and scale factors for $\Delta R < 0.4$ using different settings. All uncertainties are statistical.

In Figure 10 one can see the efficiencies and scale factor as functions of ΔR. For all bins with $\Delta R > 0.4$ the efficiency is higher for the data than for Monte Carlo which leads to the scale factor being larger than one. The distribution in these bins is relative uniform with an scale factor of $SF \approx 1.4$. In the bin for $\Delta R < 0.4$ the efficiencies drop from around $0.76 - 0.78$ to 0.70 ± 0.01 for Monte Carlo and to 0.52 ± 0.03 for data. This can be understood since at close distances the jet obscures the electron signal and makes it look less like an electron normally looks in the detector. Because the drop in Monte Carlo is far smaller than in data, the scale factor becomes smaller than one: $SF = 0.745 \pm 0.045$. The uncertainty is statistical. To find the systematic uncertainties we use Table 2 and equation 11 but only using values for the wider mass window, the $E_T^{Cone} < 4$ GeV cut and the change in fit function for the signal from RooDecay to Crystal Ball. We found $\Delta SF = 0.29$.

This high systematic uncertainty results from the conservative guess one does using equation 11. But a deeper understanding of the influence of different parameters is necessary to estimate the systematics in a better way. It is already assessable that it is necessary to find the right function for fitting E_T^{Cone} signal of the electron because the estimations between fitting the signal using the RooDecay or Crystal Ball function have differences of almost 40 % and carry the biggest contribution to the systematic uncertainty. The difference in the fitting functions can be seen in Figure 16. One way to find the needed function could be finding a E_T^{Cone} distribution that is purely formed by jet events and use it to fit the background. This should make it easier to evaluate the form of the electron signal. The statistical uncertainty is probably underestimated because the number of events in the $\Delta R < 0.4$ bin is not high enough to have stable fitting and error approximation. This can be clearly seen in Figure 16 and the resulting χ^2/ndf. Therefore, higher statistics are necessary and they should be obtained with a longer run time of the LHC.

The efficiencies and scale factor as functions of ϕ are shown in Figure 11 and 14. And ich can be seen that they have a nearly uniform distribution. This is how it would have been expected because the ATLAS detector should be isotropic with regards to the angle ϕ. Small differences can be explained by the realistic structure of the detector so that there is more non detecting material in some directions than in others. The identification cut is more efficient for data and so the scale factor is larger than one, varying between $SF = 1.02$ and $SF = 1.04$.

The efficiencies and scale factor as function of η are shown in Figure 12 and 15. It

should be mirror-symmetric to $\eta = 0$ because the ATLAS detector setup is theoretical that way but in reality, for example, one side of the tracker is not as responsive as the other. The variations over the whole range of the variable comes from the different amount of detecting and non-detecting material a particle passes for different η due to the architecture of the detector.

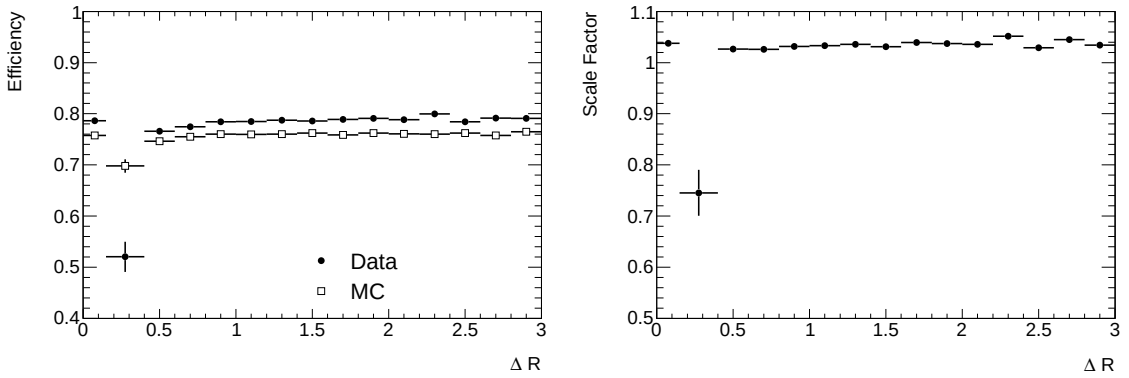

Figure 10: The efficiencies and the scale factor as functions of ΔR between electron and closest jet. The first bin shows data belonging to zero-jet-events or with $\Delta R > 3$

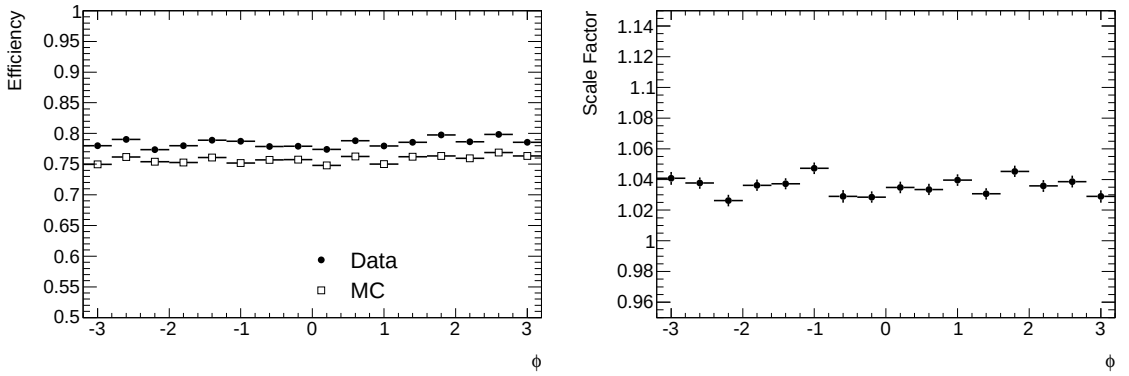

Figure 11: The efficiencies and the scale factor as functions of ϕ.

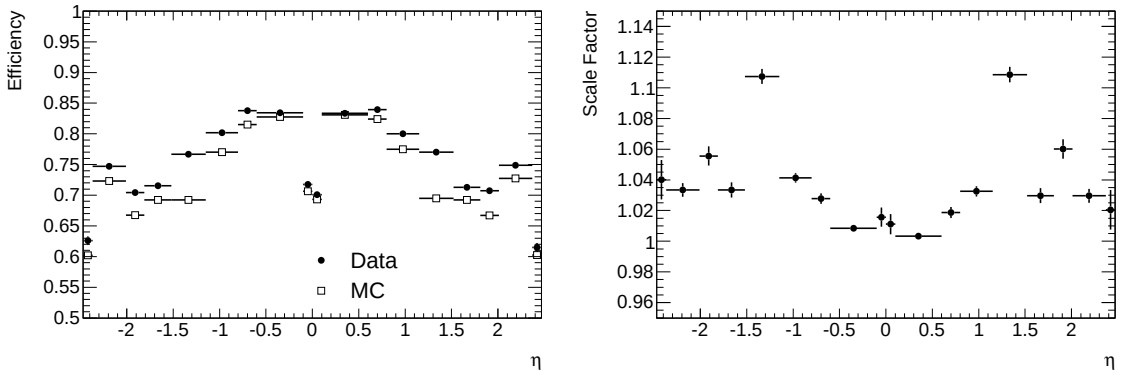

Figure 12: The efficiencies and the scale factor as functions of η.

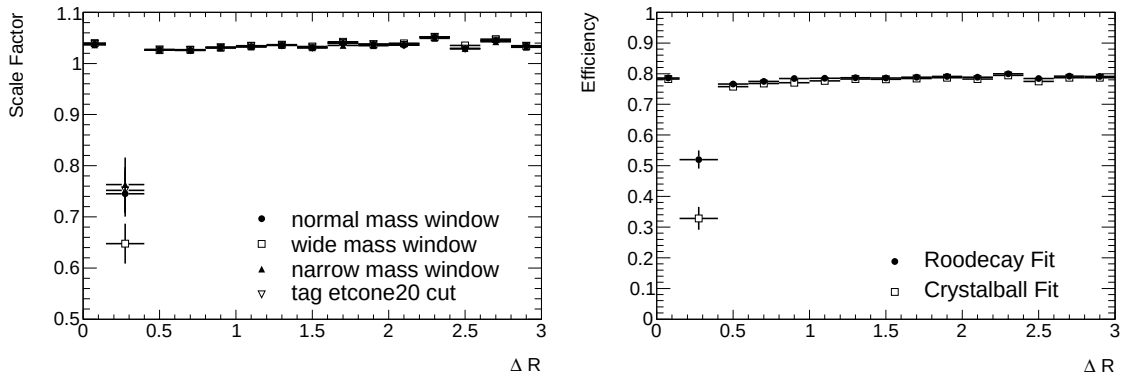

Figure 13: Varying different cuts to derive efficiencies and scale factors as functions of ΔR. The first bin shows data belonging to zero-jet-events or with $\Delta R > 3$

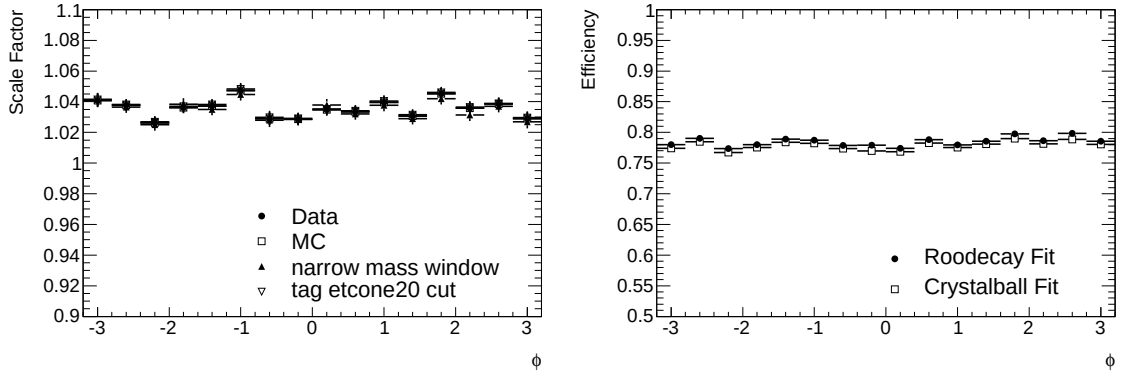

Figure 14: Varying different cuts to derive efficiencies and scale factors as functions of ϕ.

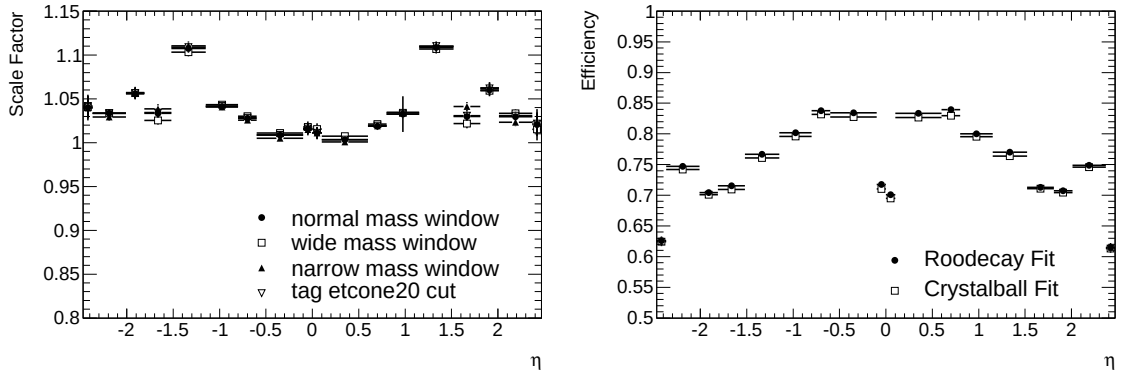

Figure 15: Varying different cuts to derive efficiencies and scale factors as functions of η.

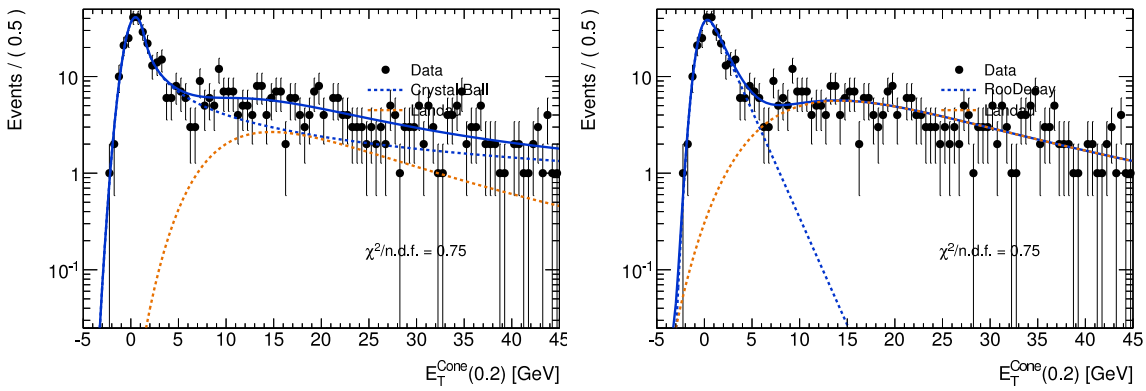

Figure 16: Fitting of $E_t^{Cone}(0.2)$ for bin $\Delta R < 0.4$ using the different functions described in the report (Data with integrated luminosity of 1 fb^{-1}).

8 Conclusion

In this report we analyzed the efficiencies and scale factors of the tight identification cut for $\Delta R < 0.4$ using the $Z \to e^+e^-$ decay. The result was obtained via the tag and probe method that we studied in this report for the case of a small ΔR. The usual approach of fitting the invariant mass distribution of the electron pair and a new method studying the distribution of the isolation variable $E_T^{Cone}(0.2)$ were used. We found that the first method is not successful for the case of small ΔR but with the method that uses $E_T^{Cone}(0.2)$ we were able to obtain efficiencies and a scale factor for $\Delta R < 0.4$.
The result $SF(\Delta R < 0.4) = 0.745 \pm 0.045$ (stat.) ± 0.290 (sys.) is dominated by systematic uncertainties and the next step should be finding better parametrization for the E_T^{Cone} distribution of electron signal and jet background. It is not clear if there is one single function good enough to describe the signal shape for all ΔR because the closer a jet is to an electron the more energy could leak into the cone distorting the distribution shape.

The statistical uncertainties will be reduced with more accumulated data. After lowering the uncertainties for the $Z \to e^+e^-$ decay it should be studied if this method can be applied to events with boosted tops.

9 Acknowledgment

We would like to thank DESY for opening its doors to the summer students and allow us to have a glance at the work a physicist does. We also want to thank Karl Jansen for organizing this program, for his support and his eagerness to make the most out of this experience for us. A special thanks goes to Klaus Mönig for selecting us to be part of his group and to our supervisors Hongbo Zhu, Thorsten Kuhl and Elin Bergeaas Kuutmann for their constant support. We particularly want to thank Luz Gomez, Valentina Ferrara and Christoph Wasicki for their willingness and patience to help us whenever we needed them, despite the fact that we came and invaded their offices. Also, we want to thank the other summer students for sharing this experience with us, we are very glad we met you and we hope we'll meet again some day in the future.

A ROOT

ROOT [7] is a framework for data processing developed at CERN aimed to support high energy physics research. ROOT provides an extremely powerful data structure that allows fast access of huge amounts of data – orders of magnitude faster than any database. Also, ROOT trees spread over several files can be chained and accessed as a unique object, allowing for loops over large amounts of data. Another very useful feature of ROOT is that it provides powerful mathematical and statistical tools to analyze data and also, the data can be generated following any statistical distribution, making it possible to simulate complex systems. In addition, one can use ROOT to present results in the form of histograms, scatter plots, fitting functions, etc. Finally, ROOT allows you to use the CINT C++ interpreter or Python to write macros or compile your program. As part of our analysis, we used some of ROOT's predefined functions to fit the histograms of the invariant mass of the pair of electrons. Afterward, we were introduced to a new library of ROOT called Refit. This library is very helpful when wanting to model an expected distribution of events in a certain physical event. One can model the distribution of observables, which are the measured quantities, in terms of physical parameters of interest and other parameters that describe detector effects (resolution, efficiency, etc.), the parameters being the degrees of freedom of the model. The model will be a probability density function (pdf) normalized over the allowed range of the observables.

B SFrame

SFrame is a general High Energy Physics analysis package based on ROOT trees. SFrame follows the cycle-based analysis that is often used in HEP by splitting an analysis into several cycles. Each cycle takes as input a number of ROOT trees and produces also ROOT trees but in a different output format and also control histograms. In this framework, users can implement their specific analysis after they have produced the ROOT trees for the data sources. The user only has to provide an ExecuteEvent() method for each cycle in which the calculation steps and the histograms filling are done. All features such as input and output of trees and histograms, loop over events and weighting are provided by the framework. Additionally, the user has to provide some description of the cycle such as the integrated luminosity the background should be weighted to and of course, the input ROOT trees and the output format. Initially, SFrame was intended to work on the ROOT trees provided by the ATLAS SusyView package. However, SFrame developed towards a SUSY independent package that can be used for any HEP analysis based on ROOT trees. Institutes involved in the development are CERN, University of Hamburg and New York University. Meanwhile it is used by several groups (DESY, CERN, Manchester, University of Hamburg, NYU and Bonn University).

C Electron identification

For container level electrons there must exist a track in the inner detector and a shower in the electromagnetic calorimeter. This two must be connected. For the tight identification cut see Table 3 and [6] Ch. 5.

Type	Description
Loose cuts	
Acceptance of the detector	$\|\eta\| < 2.47$
Hadronic leakage	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $\|\eta\| < 0.8$ and $\|\eta\| > 1.37$) Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $\|\eta\| > 0.8$ and $\|\eta\| < 1.37$)
Second layer of EM calorimeter	Ratio in η of cell energies in 3×7 versus 7×7 cells. Lateral width of the shower
Medium cuts (includes Loose)	
First layer of EM calorimeter.	Total shower width Ratio of the energy difference associated with the largest and second largest energy deposit over the sum of these energies
Track quality	Number of hits in the pixel detector (< 1) Number of hits in the pixels and SCT (< 7). Transverse impact parameter (¡5 mm).
Track matching	η between the cluster and the track (< 0.01)
Tight cuts (includes Medium)	
b-layer	Number of hits in the b-layer (≥ 1).
Track matching Track quality	$\Delta\phi$ between the cluster and the track (< 0.02). Ratio of the cluster energy to the track momentum Tighter $\Delta\eta$ cut (< 0.005) Tighter transverse impact parameter cut (< 1 mm).
TRT	Total number of hits in the TRT. Ratio of the number of high-threshold hits to the total number of hits in the TRT
Conversions	Electron candidates matching to reconstructed photon conversions are rejected

Table 3: Variables used for loose, medium and tight electron identification cuts for the central region of the detector ($|\eta| < 2.47$), see also [6] Ch. 5

D Function used for fitting

In this report several functions are used for fitting.

- The Breit-Wigner function:

$$p(E; M, \Gamma) = \frac{1}{2\pi} \frac{\Gamma}{(E-M)^2 + \Gamma^2/4} \tag{12}$$

- The Crystal Ball function:

$$f(x; \alpha, n, \bar{x}, \sigma) = N \cdot \begin{cases} \exp(-\frac{(x-\bar{x})^2}{2\sigma^2}), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-\bar{x}}{\sigma})^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leqslant -\alpha \end{cases} \tag{13}$$

  with

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right)$$

$$B = \frac{n}{|\alpha|} - |\alpha|$$

- The Landau function:

$$p(x; \mu, \sigma) = \frac{1}{\pi} \int_0^\infty e^{-t \log t - \frac{x-\mu}{\sigma} t} \sin(\pi t)\, dt \tag{14}$$

- The Roodecay function: this is a Gaussian function convoluted with an exponential function. Both functions will be assumed as known.

References

[1] Particle Data Group Collaboration, K. Nakamura et al., *Review of particle physics*, J. Phys. **G37** (2010) 075021.

[2] ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[3] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k t jet clustering algorithm*, Journal of High Energy Physics **2008** (2008) no. 04, 063. http://stacks.iop.org/1126-6708/2008/i=04/a=063.

[4] T. Sjostrand et al., *High-energy physics event generation with PYTHIA 6.1*, Comput. Phys. Commun. **135** (2001) 238–259, arXiv:hep-ph/0010017.

[5] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Methods Phys. Res., A **506** (2003) 250–303.

[6] G. Aad et al., *Expected electron performance in the ATLAS experiment*, Tech. Rep. ATL-COM-PHYS-2010-990, CERN, Geneva, Nov, 2010.

[7] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81–86. See also http://root.cern.ch/.

[8] C. Blocker, *Uncertainties on Efficiencies*, . http://www-cdf.fnal.gov/physics/statistics/statistics_recommendations.html. CDF/MEMO/STATISTICS/PUBLIC/7168.