A First Look at Tau Fake Rates from QCD Dijets in 2011 Data

Didier Alexandre

Summer Students Programme 2011

ATLAS

September 8, 2011

Abstract

In this report, an account of an 8-week project at the DESY-ATLAS group is given. Measurements of the tau fake rates from QCD dijets have been performed using 2011 ATLAS data. Special focus is given to the observed differences in the Fake Rates when using different tau identification algorithms. Restricting the identification to tau-candidates with one single prong or 3 prongs also proved to influence the Fake Rates. Several sources of systematic uncertainties are discussed. Determinig the values of these requires a detailed analysis, which was only broadly approached during this project.

Contents

1 Introduction **3**
1.1 The Mis-Identification Probability of τ-leptons - The Tau Fake Rate . . 3
1.2 Definition 4
1.3 A brief introduction to the ATLAS detector 4
1.4 Trigger 4

2 Analysis **5**
2.1 The Tau-ID algorithms 5
2.1.1 The Cut-based identification 6
2.1.2 The logarithmic likelihood 6
2.1.3 The Boosted Decision Trees 7
2.2 Sample Selection 7
2.2.1 Tag and Probe Method 8

3 Results **8**
3.1 Comparison between the 3 ID-methods 9
3.2 Dependence of the Fake Rates on the multiplicity of prongs 9
3.3 Systematics 9
3.4 Systematic uncertainties plots 11

4 Conclusion **11**

5 Acknowledgments **14**

1 Introduction

1.1 The Mis-Identification Probability of τ-leptons - The Tau Fake Rate

The τ-lepton is the heaviest particle among the leptons (m_τ = 1.8GeV). Because it is so massive, it decays very rapidly, with a lifetime of 2.9×10^{-13} seconds. Taus (τ) can decay leptonically, producing lighter leptons - muons or electrons. However, this only happens 35.3% of the time, as taus more often decay hadronically (branching ratio, BR_h=64.7%), leaving hadronic showers - also known as jets - in the calorimeters of the ATLAS detector [1].

Muons are the easiest particles to identify in the detector. With their high penetration power, they travel through all subdetectors until they reach the outermost shell of ATLAS - the muon detector - and deposit the majority of their energy in there.

Electrons don't cross as much of the detector unimpeded and leave distinct electromagnetic showers in the electromagnetic calorimeter. As they interact with the atomic nuclei of the calorimeter, they scatter in a process called 'Bremsstrahlung', thereby emitting a photon. The latter produces an electron-positron pair (pair-production) soon after. These charged particles will scatter again and this chain of processes will repeat itself several times. The result is an avalanche of photons and electrons[1]. The amount of energy they produce in the calorimeter is proportional to the energy of the originating electron.

In contrast to muons and electrons, taus don't leave such specific signatures in the detector. Most of the time, they mimick the signature of jets caused by the hadronisation of quarks and gluons (QCD jets), and there is no easy way to tell them apart from the latter. Three different methods have been developed in order to discriminate taus from QCD jets: Cut-based Identification, Projective Likelihood and Boosted Decision Trees. No matter how efficient these methods are, a complete disentanglement cannot be achieved with them and some jets are always mistakenly identified as taus. The *tau fake rate* is a measure of how probable it is for this mis-identification to occur[2]. It therefore provides information about the background of physics processes analyses involving taus.

[1] electrons stands for both electrons and positrons

[2] *tau fake rate* is in fact a misnomer, as the quantity measured is not a rate. A more correct name for it is *the Mis-identification Probability of taus.* I will continue referring to the former name in this report

1.2 Definition

The tau fake rate is defined as the ratio of the number of identified taus to the number of reconstructed taus in a dataset that only contains QCD jets:

$$f_I D = \frac{number of identified taus}{number of reconstructed taus}$$

Reconstruction algorithms are unable to distinguish between taus and QCD jets and almost every particle reconstructed as the former is also reconstructed as the latter. The tau identification algorithms will be dealt with in more detail in Section 2.1.

1.3 A brief introduction to the ATLAS detector

The ATLAS detector is formed of four major components, ranging from innermost to outermost parts as follows: the central tracker (or inner detector), the electromagnetic calorimeter (EMCAL), the hadronic calorimeter (HCAL) and the muon detector.

A magnetic field produced by a solenoid at the centre and toroid magnets in the outer regions of the Atlas detector causes charged particles to follow curved paths. The silicon pixels and strips and the Transition Radiation Tracker (TRT) of the inner detector are responsible for tracking the particles created after the pp- (proton-proton) collisions. The curvature of the tracks allows the momentum of a given particle to be measured. The calorimeter systems are made of subsequent layers of passive absorbers (lead in the EMCAL , and steel, copper and tungsten in the HCAL) and active media (liquid argon - LAr - in the EMCAL and scintillator tiles in the HCAL).[7] As a particle interacts with the absorber material, it decays and produces new charged particles and photons. These are then converted into electrical signals by the active media from which information on the energy of the decaying particle is obtained.

A muon detector is placed at the outermost shell of the detector. There is sufficient material between it and the beam axis so that only muons manage to interact with it. In Figure 1, the Atlas detector viewed from a cut-away perspective is shown. [3] [5] [6]

1.4 Trigger

The ATLAS experiment collects data at an event rate of 40 MHz. Such huge amounts of data are impossible to store. However an overwhelming majority of it is not needed, as interesting physics processes only occur in a minute fraction of all events (usually less than 1%). The trigger system reduces the event rate by selecting the most promising events. This is done in three successive stages: Level-1 (L1), Level-2 (L2) and Event-Filter (EF) trigger. The event rates resulting from each of the levels, as well as their average processing time per event are listed below in Table 1. As can be seen the L1-triggers make very quick decisions in comparison to L2 and EF triggers. Unlike the other two stages, the former does not use the detector information at full granular precision.

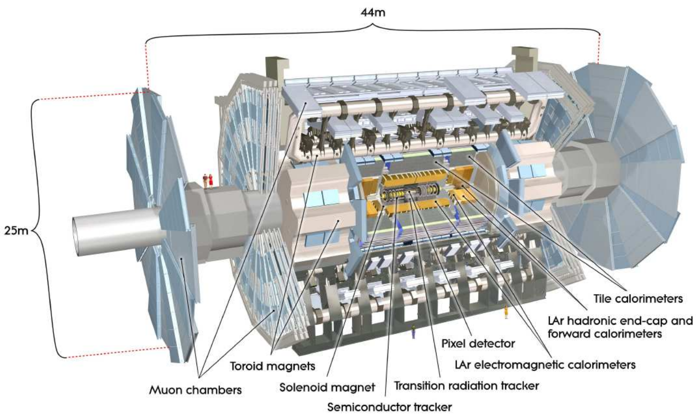


Figure 1: Cut-away view of the ATLAS detector.[5]

L1 triggers are responsible for the coarse triggering, elimintating most of the events. Finer triggering is carried out by L2 and EF triggers. The EF uses full granularity information and is able to reconstruct each event it processes.

In order to achieve the event rates listed in Table 1, certain triggers may need to reduce the number of selected events by a *prescaling* factor.[3]

Trigger Level	Event Rate (kHz)	average decision time
Level 1	75	2.5 μs
Level 2	3.5	40 ms
Event-Filter	0.2	4 s

Table 1: The data collection rate after reduced by the given trigger levels and the average processing time per event.[3]

2 Analysis

2.1 The Tau-ID algorithms

The following section outlines the 3 different methods used to distinguish hadronically decaying taus from QCD jets: the Safe-cuts, the Projective likelihood and the Boosted Decision Trees. They are trained by using simulated Monte Carlo data, in which the nature (tau or QCD jet) of the tau-candidate is known. The identification of taus is performed by applying cuts on a set of discrimination variables. These names of these variables are listed below [1]:

- the transverse energy wighted shower width in the electromagnetic calorimeter: the electromagnetic radius R_{EM}
- the p_T-weighted track width: the track radius R_{track}
- the leading track momentum fraction: f_{track}
- the core energy fraction f_{core}
- the fraction of transverse energy of the tau candidate deposited in the electromagnetic calorimeter: the electromagnetic fraction f_{EM}
- the cluster mass $m_{clusters}$
- the track mass m_{tracks}
- the transverse flight path significance S_T^{flight}

2.1.1 The Cut-based identification

This algorithm sets cuts on the following three discrimnation variables: R_{EM}, R_{track} and f_{track}. The cuts are based on the fact that one expects the hadronic shower originating from a tau to have a narrower width in comparison to one produced by a QCD jet.

2.1.2 The logarithmic likelihood

The logarithmic likelihood (Llh) discriminant is defined as follows:

$$d_{Llh} = \ln\left(\frac{L_S}{L_B}\right)$$

where L_S and L_B are the likelihood functions of the signal and the background respectively. $L_S(L_B)$ describes the probability of a given jet to belong to the signal (background) - i.e. the probability of it being a tau (QCD jet) - by taking into account all of the identification variables.

2.1.3 The Boosted Decision Trees

This discriminant is trained to produce a decision tree, with a variable cut at each node. A tau-candidate will start at the central node and continue along a certain direction if it passes that cut. If it fails to pass it, it will go the opposite direction. As it is subjected to subsequent cuts at each node, it will take one specific path along the tree until it reaches a leaf node. Every leaf node has an associated signal purity coefficient, with which the nature of the tau-candidate can be determined with the reliability corresponding to that coefficient. A schematic representation of such a decision tree is shown in Figure 2.

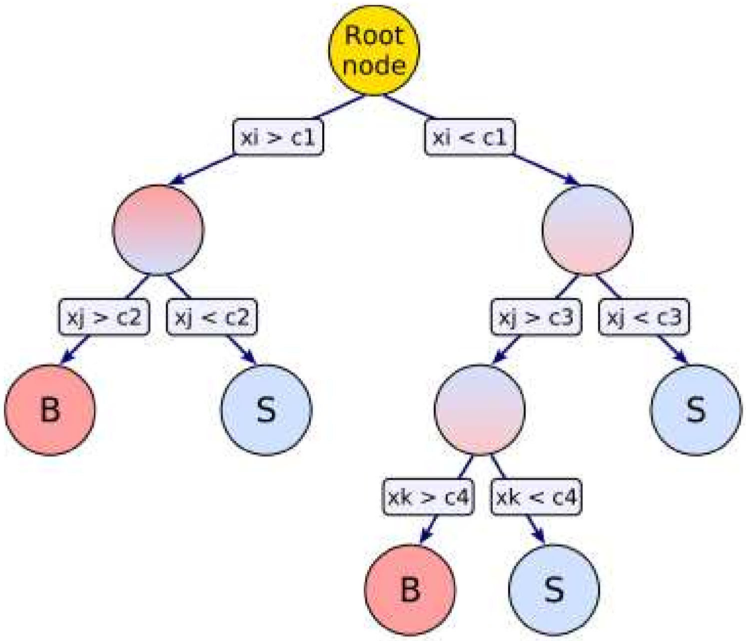


Figure 2: Schematic view of a decision tree. Cuts (c) are applied on the variables (x) at the nodes. “S” and “B” stand for signal and background respectively. [4]

2.2 Sample Selection

The data used for the calculations of the tau fake rates was 2011 Atlas data from period D to G. The data is first subjected to 2 selection cuts:

- GRL cut
  The Good-Runs List cut eliminates any event whose data quality does not satisfy the requirements for the analysis to be performed. The information on the quality of the events is obtained on-line.

- Trigger cut
  Only data selected by the triggers (as explained in section 1.4) is available offline for analysis. In most analyses, the data used is also restricted to events passing specific trigger chains to ensure the sample contains events with physics of interest. In this analysis, a combination of 4 EF triggers was used. Only events passing at least one of these 4 filters were allowed to pass the cut for further analysis.

2.2.1 Tag and Probe Method

In order to calculate fake rates, it is advantageous to have at disposal a QCD jets sample which is as unbiased by selection criteria as possible. To create such a pure sample is the purpose of the Tag-and-Probe method, which processes back-to-back objects that are balanced in transverse momentum[3] and reconstructed as QCD jets. These objects are paired if they satisfy the following conditions:

- $\eta \leq 2.5$
- $p_T \geq 15$GeV;

The leading-p_T jets of all selected pairs are taken to be the tag jets, while the ones with the lowest p_T are called the probe jets. The former are subject to several cuts, chosen such that only true QCD jets pass. The most relevant one is the cut on the track multiplicity:

- the number of tracks associated to the jet is: $N_{track}(\text{tag}) \geq 4$

This cut effectively gets rid of the genuine hadronic tau decays Those that do manage to pass it, contribute to a negligible extent to the background. It is then assumed with a high degree of confidence that any probe jet paired to a QCD tag jet is also, itself, a QCD jet. The sum of all such probe jets constitute the sample needed for the fake rate calculations.

3 Results

In this section several Fake Rate distributions are shown. In section 3.1, graphs of the tau fake rate as a function of transverse momentum and η obtained using all of the identification methods are discussed. Section 3.2 deals with the way the fake rates distributions change depending on the number of prongs (number of charged pions) associated to the tau candidates. For this purpose, it is also interesting to make a comparison between the results obtained by the 3 tau-ID techniques. The systematic

[3] the back to back criterion is defined as two jets having a difference in their ϕ position within π-0.3 $\leq \Delta\phi \leq \pi$, while the p_T-balance criterion requires that $\Delta p_T \geq \frac{p_T^{max}}{2}$, where p_T^{max} is the transverse momentum of the jet with highest p_T (the leading-p_T jet)

uncertainties due to variations in p_T-balance and alignment of the tag and probe jets are shown in section 3.4.

3.1 Comparison between the 3 ID-methods

In Figure 3 the 3 tau-identification methods are compared. The tau Fake Rate distributions for each of them are plotted in the same graphs as a function of transverse momentum. It can be observed that the multivariate techniques (LLh and BDT) are less likely to mis-identify jets as taus. This holds for almost the entire p_T-range shown: only at low p_T does the BDT method give a higher fake rate than the cuts-based algorithm. Using medium selection criteria, the multivariate techniques produce fake rates lower than the Safe-cuts method by a factor of the order of 10 (Figure 3 b)). The lower mis-identification probability of the latter can also be seen as a function of η in Figure 4.

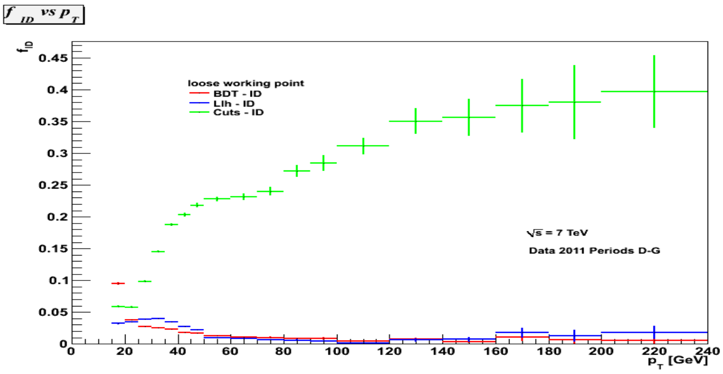
3.2 Dependence of the Fake Rates on the multiplicity of prongs

The number of prongs associated to a jet, indicate the number of charged pions that were present in the hadronic decay. Taus are usually categorised as 1 or 3 prong. Figure 5 shows that for Llh and BDT the fake rates are lower for 3-prong taus than 1-prong. However, this is not the case with the cuts identification, for which 1-prong taus give a lower mis-identification rate.

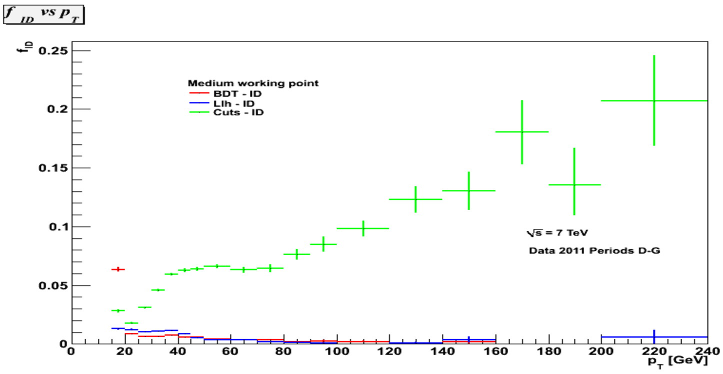
3.3 Systematics

There are several factors that can affect the performance of the identification methods, and all contribute to the total systematic uncertainties of the tau fake rates. The main sources of systematic uncertainties are:

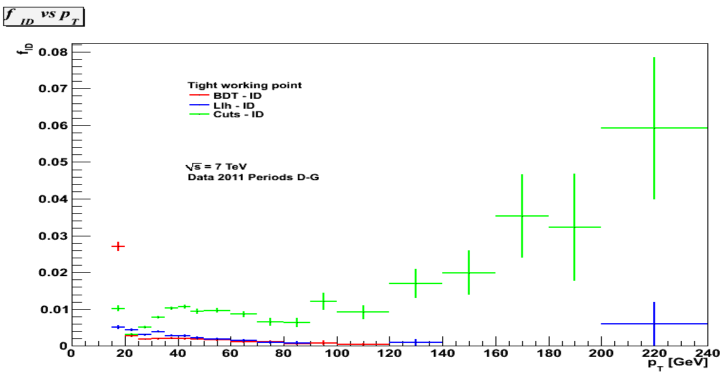
- Pile-up: the accumulation of primary vertices in an event
- quark-gluon fraction: the fraction of quark to gluon originated jets in the sample
- loose or tight p_T-balance between tag and probe jet
- the degree of alignment of tag with probe jet
- dead material (crack region, Liquid-Argon Hole): The crack region is the transition region between barrel and end-cap of the detector and the LAr hole is a defect region of the detector.
- track multiplicity: the number of tracks associated to the jets in the sample



(a)



(b)



(c)

Figure 3: Comparison between tau fake rates as a function of p_T obtained with the 3 identification methods with loose a), medium b) and tight c) tau selections.

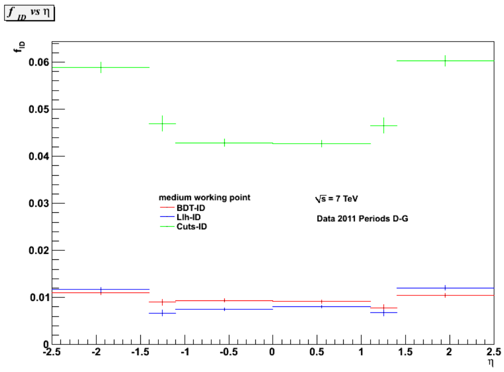


Figure 4: f_{ID} as a function of η obtained with the 3 identification methods with a medium tau selection

3.4 Systematic uncertainties plots

During my project, I considered the systematic uncertainties due to p_T-balance and tag-probe alignment (back-to-back selection) biases. To calculate theses uncertainties, the sample was split into two sub-samples by introducing a cut, which seperates them in samples of similar number of events. These are:

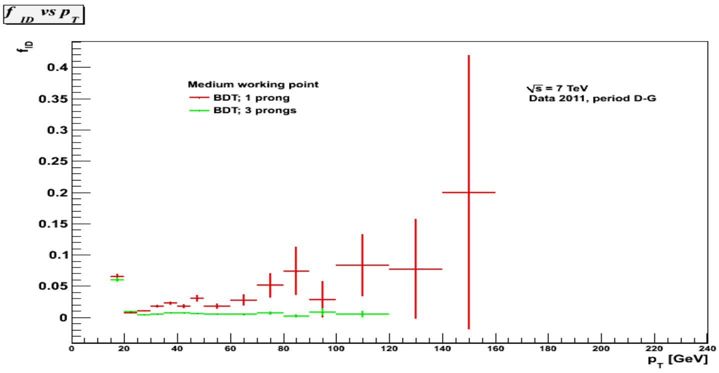
- for the back-to-back analysis
  loose cut: $\Delta\phi < \pi - \frac{1}{3} \cdot 0.3$
  tight cut: $\Delta\phi \geq \pi - \frac{1}{3} \cdot 0.3$

- for the p_T-balance analysis
  loose cut: $|\Delta p_T| \geq 0.44 \cdot \frac{p_T^{max}}{2}$
  tight cut: $|\Delta p_T| < 0.44 \cdot \frac{p_T^{max}}{2}$

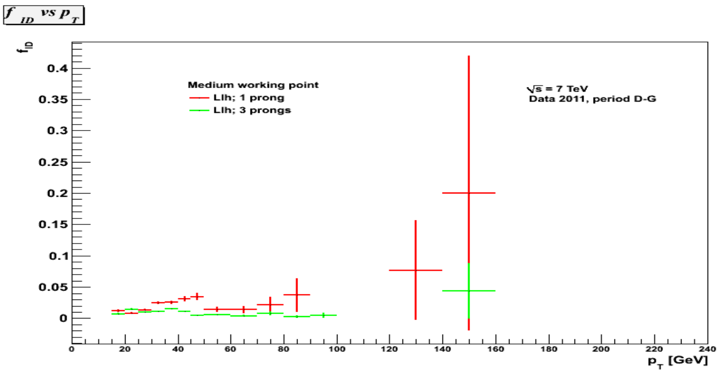
After these two sub-samples were created, the Fake Rate distributions were plotted for each, in order to see the variations between them. The systematic uncertainty values are obtained by dividing these fake rate curves by the default fake rate distribution. The graphs in Figure 6 illustrate this procedure and show the results obtained for both uncertainty sources for the loose BDT tau-identification.

4 Conclusion

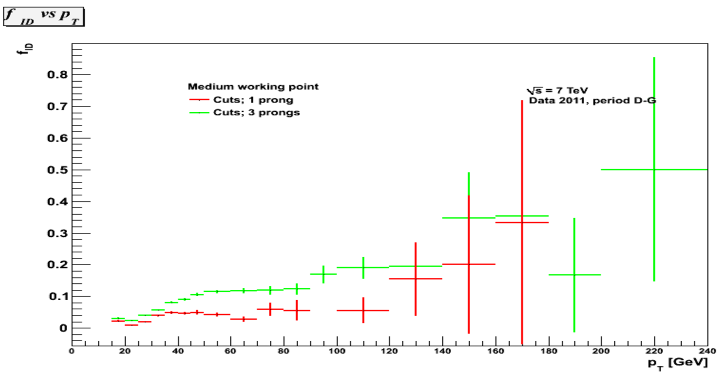
It was observed that tau fake rates decrease in the hight p_T-range with the multivariate techniques, as opposed to the cuts-identification. Another difference between the fake rates obtained by the safe cuts and the other two methods is the dependence on the number of prongs. The 3-prong BDT and 3-prong Llh identifications have a lower mis-identification probability than the respective 1-prong selections. However, for the safe cuts, lower probabilities are obtained with 1-prong, rather than 3-prong tau selection.



(a)

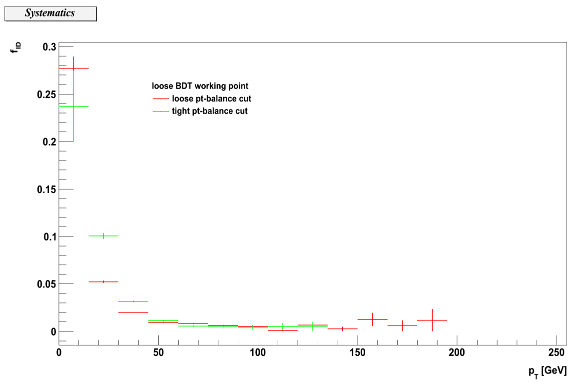


(b)

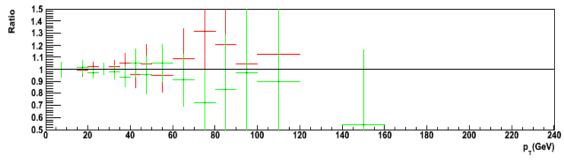


(c)

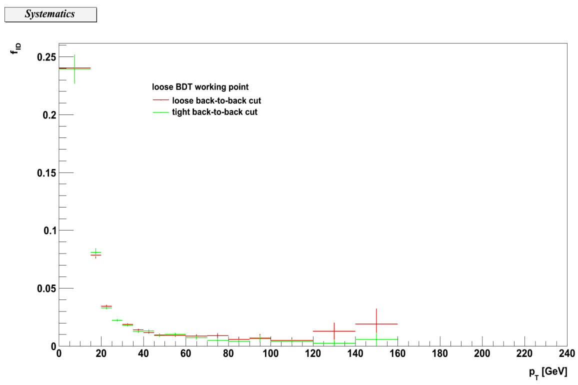
Figure 5: f_{ID} vs p_T for 1 and 3 prong medium taus for a) BDT, b) Llh and c) cuts methods.



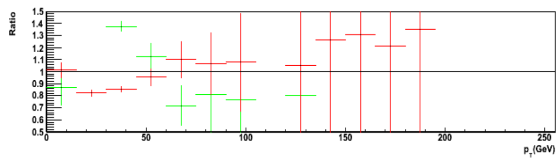
(a)



(b)



(c)



(d)

Figure 6: f_{ID} vs p_T distributions for a) loose and tight p_T-balance sub-samples and b) loose and tight back-to-back sub-samples; systematic uncertainties due to the p_T-balance c) and due to the back-to-back selection d)

5 Acknowledgments

I am specially thankful to my supervisor Nicholas Styles for his support and advice during the entire summer programme. I would also like to thank Wolfgang Ehrenfeld for his help in IT issues and my fellow summer student and office colleague Maciej Misiura for sharing some of his wide range of experience with me.

References

[1] The ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment*, ATLAS-CONF-2011-077

[2] The ATLAS Collaboration, *Measurement of the Mis-identification Probability of τ-Leptons from Hadronic Jets and from Electrons*, ATLAS-CONF-2011-113

[3] Almut Maria Pingel, *Measurement of the Tau Mis-identification Probability from QCD Jets with ATLAS data at $\sqrt{s}$=7TeV*, Diplomarbeit, Institut für Experimentalphysik, Universität Hamburg, April 2011

[4] *TMVA 4 Toolkit for Multivariate Data Analysis with ROOT, Users Guide*, version 4.0.3, CERN-OPEN-2007-007, November 2009

[5] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Institute of Physics Publishing and SISSA, August 2008

[6] Kristin Heine, *Trigger performance studies with the ATLAS detector using $J/\psi \rightarrow e^+e^-$ events*, Diplomarbeit, Institut für Experimentalphysik, Universität Hamburg, März 2011

[7] D. Pitzl, *Detectors for Particle Physics, Lecture 1*, Lectures for Summer Students in HEP, Accelerators and Computing, DESY, Hamburg, 2011 (PDF file available from: www.desy.de/summerstudents)