Tau reconstruction and identification with 7 TeV collisions in ATLAS

A. Kaczmarska^{a*}

On behalf of the ATLAS Collaboration

^aInstitute of Nuclear Physics PAN, 152 Radzikowskiego, Krakow, Poland

The first LHC collisions at a centre-of-mass energy of $\sqrt{s} = 7$ TeV recorded with the ATLAS detector are used to study the reconstruction and identification algorithms for hadronic τ decays. Although the dataset used here, corresponding to an integrated luminosity of 244 nb^{-1} , contains a small number of real τ leptons, the background jets reconstructed as τ candidates can be used to assess performance aspects of these algorithms. Distributions of identification variables are compared in data and Monte Carlo samples, and the background efficiency of τ identification criteria is measured in a QCD dijet enriched data sample. Good agreement is observed between data and the fake τ candidates selected from simulated QCD dijet events, giving confidence in the performance of the algorithms on real τ leptons, to be observed soon.

1. INTRODUCTION

The τ lepton plays an important role in the LHC physics programme, for example in searches for a Higgs boson or supersymmetry [1]. Decays of Standard Model gauge bosons to τ leptons, $W \rightarrow \tau\nu$ and $Z \rightarrow \tau\tau$, are important background processes in such searches, and their cross sections need to be measured beforehand. These decays will also be used to ensure that the reconstruction and identification of τ leptons are well understood.

Since March 2010 the ATLAS [2] experiment at the LHC has been collecting proton-proton collisions events at a centre-of-mass energy of $\sqrt{s} = 7$ TeV. These data are analysed to study the performance of the τ -lepton reconstruction and identification algorithms. While the number of true τ leptons present in the dataset is expected to be negligible, the performance of the reconstruction and identification of hadronic τ decays can still be studied. Distributions of variables used in the identification of τ leptons in these events can be compared to predictions from Monte Carlo (MC) simulation. The rejection power of the identification algorithms, along with associated systema-

tic uncertainties, can be evaluated on this background enriched sample.

2. DATA SAMPLES AND EVENT SELECTION

The studies presented here are mostly based on data collected with the ATLAS detector at a centre-of-mass energy of $\sqrt{s} = 7$ TeV, corresponding to an integrated luminosity of approximately $L = 244 \text{ nb}^{-1}$ [3]. The data considered are required to have been taken with stable LHC beam conditions, and to pass data quality requirements for the inner detector (tracker) and the calorimeter.

Furthermore all events must satisfy the Level 1 trigger condition requiring a τ -trigger object passing a $p_T = 5$ GeV threshold. At least two τ candidates with $p_T > 30$ GeV and $p_T > 15$ GeV must be found in the event, which are required to be separated by at least 2.7 radians in azimuth (the angle in the plane transverse to the beam pipe). This cut aims at selecting events with back-to-back jets, enriching the sample with fake τ candidates from QCD jet processes. In order to remove any bias due to the trigger requirement, the sample of τ candidates that are studied excludes the leading τ candidate. With these requirements the selected data sample con-

*Supported in part by the Polish Government grant NN202 127937 (years 2009-2011) and Polish-DAAD collaboration 764/N-DAAD/2010/0.

tains about 2.9 million events with 3.9 million τ candidates.

QCD jet MC samples are used for comparison. The allowed range of the transverse momenta of the outgoing partons in the rest frame of the hard interaction are restricted to be between 8 and 280 GeV. These samples are generated with PYTHIA [4] using the DW tune [5] and passed through a GEANT4 simulation of the ATLAS detector [6]. When showing distributions for true τ -lepton candidates, a $Z \rightarrow \tau\tau$ MC sample with the MC09 tune [7] is used.

3. TAU RECONSTRUCTION ALGORITHM

The τ lepton decays 65% of the time to one or more hadrons and 35% of the time leptonically, in both cases with accompanying neutrinos. Because of their short lifetime, it is very difficult to separate τ leptons decaying to electrons or muons from prompt electrons and muons, and τ identification therefore focuses on reconstructing hadronically decaying τ candidates.

Hadronically decaying τ leptons are reconstructed starting from either calorimeter or track seeds [8]. *Track-seeded* candidates have a seeding track with $p_T > 6$ GeV satisfying quality criteria. *Calorimeter-seeded* candidates consist of calorimeter jets reconstructed with the anti- k_t algorithm [9] starting from topological clusters (topoclusters) [10]. The candidate is required to have $p_T > 10$ GeV. Candidates are labelled *double-seeded* when a seed track and a seed jet are within a distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.2$ of each other. The p_T of the τ candidate is calibrated by applying multiplicative factors derived from MC studies, in order to reconstruct the p_T of signal τ leptons accurately. In the studies presented here, candidates with both seeds and candidates with only a calorimeter-seed with at least one associated track are considered, as there are very few candidates without a calorimeter seed or an associated track. Figure 1 shows the number of associated tracks of the τ candidate (a real τ lepton is expected to have mostly one or three such tracks).

Figure 1. Number of associated tracks of τ candidates. The number of τ candidates in MC samples are normalised to the number of τ candidates selected in data. The data correspond to an integrated luminosity of 15.6 nb^{-1} .

4. TAU IDENTIFICATION METHODS

The reconstruction of τ candidates provides very little rejection against QCD jet backgrounds. Rejection comes from a separate identification step and is usually based on several discriminating variables.

Variables used for identification methods have shown good separation potential in MC studies. Since a hadronic τ decay is characterized by collimated energy deposits in the calorimeters and one or few collimated tracks, these properties are used to distinguish them from QCD jets. The variables that are used in the identification of τ leptons with early data include:

Cluster mass: Invariant mass computed from associated topoclusters, m_{clusters} .

Track mass: Invariant mass of the track system, m_{tracks} .

Track radius: p_T -weighted track width:

$$R_{\text{track}} = \frac{\sum_i^{\Delta R_i < 0.2} p_{T,i} \Delta R_i}{\sum_i^{\Delta R_i < 0.2} p_{T,i}},$$

where i runs over all tracks associated to the τ candidate, ΔR_i is a cone distance between the τ jet seed axis and $p_{T,i}$ is the track transverse momentum.

Leading track momentum fraction:

$$f_{\text{trk},1} = \frac{p_{\text{T},1}^{\text{track}}}{p_{\text{T},\text{vis}}^{\tau}},$$

where $p_{\text{T},1}^{\text{track}}$ is the transverse momentum of the leading track of the τ candidate and $p_{\text{T},\text{vis}}^{\tau}$ is the visible transverse momentum of the τ candidate.

Electromagnetic radius: Transverse energy weighted shower width in the electromagnetic (EM) calorimeter:

$$R_{\text{EM}} = \frac{\sum_i^{\Delta R_i < 0.4} E_{\text{T},i}^{\text{EM}} \Delta R_i}{\sum_i^{\Delta R_i < 0.4} E_{\text{T},i}^{\text{EM}}},$$

where i runs over cells in the first three layers of the EM calorimeter associated to the τ candidate, ΔR_i is a cone defined relative to the τ -jet seed axis and $E_{\text{T},i}^{\text{EM}}$ is the cell transverse energy.

Core energy fraction: Fraction of transverse energy in the core ($\Delta R < 0.1$) of the τ candidate:

$$f_{\text{core}} = \frac{\sum_i^{\Delta R_i < 0.1} E_{\text{T},i}}{\sum_i^{\Delta R_i < 0.4} E_{\text{T},i}},$$

where i runs over all cells associated to the τ candidate within ΔR_i of the τ jet seed axis.

Electromagnetic fraction: Fraction of calibrated transverse energy of the τ candidate deposited in the EM calorimeter:

$$f_{\text{EM}} = \frac{\sum_i^{\Delta R_i < 0.4} E_{\text{T},i}}{\sum_j^{\Delta R_j < 0.4} E_{\text{T},j}},$$

where $E_{\text{T},i}$ ($E_{\text{T},j}$) is the calibrated transverse energy deposited in cell i (j), and i runs over the cells in the first three layers of the EM calorimeter, while j runs over the cells in all layers of the calorimeter.

Since the instantaneous luminosities for the data used here are low, pile-up effects are expected to be small for the distributions shown here. With higher luminosity, pile-up will affect the distributions of these variables for both fake and true τ candidates, thus reducing their separation power.

The distributions for two identification variables mentioned above, R_{EM} and R_{track} , are shown in Fig. 2 for τ candidates in data and MC

Figure 2. EM radius R_{EM} (top) and track radius R_{track} (bottom) of τ candidates. The number of τ candidates in MC samples are normalized to the number of τ candidates selected in data.

samples. The agreement of the distributions for data and background MC samples is quite good for all identification variables.

The identification of τ leptons is applied after the reconstruction step. Three independent identification algorithms have been investigated: simple cuts, boosted decision trees (BDT) [11,12], and a projective likelihood (LL).

The performance for τ -identification algorithms is expressed in terms of two quantities: signal and background efficiencies. The signal efficiency is defined as $\varepsilon_{\text{sig}} = N_{\text{pass,match}}^{\tau} / N_{\text{match}}^{\tau}$, where N_{match}^{τ} is the number of reconstructed τ candidates that are matched within a cone of $\Delta R < 0.2$ with a true, hadronically decaying τ lepton with visible transverse momentum $p_{\text{T}}^{\text{vis}} > 15$ GeV and visible pseudorapidity $|\eta^{\text{vis}}| < 2.5$, reconstructed with the correct number of associated tracks; while $N_{\text{pass,match}}^{\tau}$ is the number of these

Figure 3. Background efficiencies obtained for data and MC samples as a function of the reconstructed p_T^τ (top). Signal efficiencies predicted by a $Z \rightarrow \tau\tau$ MC sample as a function of the reconstructed visible p_T^τ (bottom).

reconstructed candidates that pass the identification criteria. A $Z \rightarrow \tau\tau$ MC sample is used to evaluate the signal efficiency. The background efficiency is defined as $\varepsilon_{\text{bkgd}} = N_{\text{pass}}^{\text{bkgd}} / N_{\text{total}}^{\text{bkgd}}$, where $N_{\text{pass}}^{\text{bkgd}}$ is the number of the τ candidates that pass the identification criteria, and $N_{\text{total}}^{\text{bkgd}}$ is the number of τ candidates.

The cut-based identification uses three relatively uncorrelated variables: R_{EM} , R_{track} , and $f_{\text{trk},1}$. Cuts are optimized on signal and background MC samples for minimum background efficiency and tuned for roughly 30% (tight), 50% (medium), and 60% (loose) signal efficiency on the $Z \rightarrow \tau\tau$ MC sample for true, hadronically

Figure 4. Background efficiencies as a function of n_{vtx} .

decaying τ leptons. Different cuts are applied for τ candidates with $n_{\text{track}} = 1$ and $n_{\text{track}} \geq 2$.

Figure 3 shows the background efficiencies obtained for data and MC samples as a function of the reconstructed p_T^τ as well as the signal efficiencies from the $Z \rightarrow \tau\tau$ MC sample for loose, medium, and tight selections. The background efficiencies measured in data agree well with the MC prediction, showing the good performance with data of the cut-based identification that was optimized with MC samples. Two systematic uncertainties on the measured background efficiencies are considered from two different effects: the transverse momentum calibration for τ candidates and pile-up effects due to varying beam conditions. These uncertainties may partly account for some of the data-MC discrepancies observed in the high- p_T^τ region. The variation of the background efficiency was studied by comparing the current calibration of τ candidates using the GCW [13] scheme, with the EM+JES [14] scheme. When using the EM+JES calibration, the background efficiency for the cut-based identification decreases by 2.1%, 8.5%, and 9.6% for loose, medium, and tight selections, respectively. Over the course of the data taking period relevant for this analysis, the beam intensity increased by a factor of three. Increased beam intensities lead to different pile-up conditions that affect the dis-

Table 1

Background efficiencies for loose, medium, and tight selection cuts. The measured background efficiencies in data are compared to the MC DW tune prediction. Uncertainties for the background efficiencies in data are from transverse momentum calibration and pile-up effects.

Selection	$\varepsilon_{\text{bkgd}}$ (data)	$\varepsilon_{\text{bkgd}}$ (MC)
loose	$(3.2 \pm 0.2) \times 10^{-1}$	3.4×10^{-1}
medium	$(9.5 \pm 1.0) \times 10^{-2}$	9.9×10^{-2}
tight	$(1.6 \pm 0.3) \times 10^{-2}$	1.9×10^{-2}

Figure 5. Background efficiencies in data and MC as a function of visible p_T with the medium selection for cut-based, BDT, and LL identification (top). Signal efficiencies from MC as a function of visible p_T with the medium selection for cut-based, BDT, and LL identification (bottom).

tributions of the identification variables. Since pile-up effects are a major concern for the tau ID variables, the background efficiencies have been evaluated as a function of the number of vertices n_{vtx} , a variable that is highly correlated with pile-up activity. This is shown in Fig. 4. The robustness of the simple cut ID against the presence of pile-up is satisfactory.

The measured background efficiency for the given data sample is listed in Table 1 along with the corresponding MC DW tune prediction. The uncertainties shown in this table for the data are from the two systematic effects mentioned earlier, which are treated as independent. The statistical uncertainties from the data sample are negligible.

The BDT uses all seven variables listed before, while the LL uses all variables except f_{core} (due to correlations with other variables), to provide discrimination of τ leptons against jets. The efficiencies for signal and background obtained for the BDT and LL are also compared to those obtained for the cut-based identification for medium criteria in Fig. 5. The increased rejection power against fake τ candidates of the more sophisticated BDT and LL discriminants is evident, and the differences between MC and data are being investigated.

5. SOURCES OF REAL TAUS

The $W \rightarrow \tau\nu$ and $Z \rightarrow \tau\tau$ channels are the processes where true τ leptons can be first observed in ATLAS. Both channels provide τ leptons with relatively low visible transverse momenta and events with genuine missing transverse energy.

At NNLO, the $W \rightarrow \tau\nu$ signal is predicted to

be produced with a cross section times branching ratio of $\sigma \times \text{BR} = 10.46 \times 10^3$ pb. Those events must be triggered with the τ lepton and missing transverse energy combined trigger and should allow for the measurement of the τ -identification efficiency. Although the $Z \rightarrow \tau\tau$ process has a production cross section that is about ten times lower compared with that of $W \rightarrow \tau\nu$ the presence of an additional τ lepton gives a different event topology. By selecting $Z \rightarrow \tau(e, \mu)\tau(\text{had})$ events, one can trigger on lepton candidates, providing an unbiased sample of hadronic τ decays that can be used to study τ lepton reconstruction, measure the τ -lepton reconstruction and identification efficiency, and measure the τ -trigger efficiency. In addition, because the visible mass distribution of the $(e, \mu)\text{-}\tau(\text{had})$ system is sensitive to the energy scale of the reconstructed τ candidates, a measurement of the τ -lepton reconstruction energy scale can be made with this sample.

6. CONCLUSIONS

The variables used for τ identification have been investigated in a QCD jet enriched sample using 244 nb^{-1} of data collected by the ATLAS experiment at the LHC at a centre-of-mass energy of $\sqrt{s} = 7$ TeV. All variables are well described by MC predictions and show good separation power between τ leptons and fake τ candidates from QCD jets. The background efficiency for three cut-based selections (loose, medium, tight) was measured as a function of the reconstructed p_T^τ the τ candidate, and found to be in good agreement with MC predictions. Systematic uncertainties on the background efficiencies from transverse momentum calibration and pile-up effects were determined. Both data and MC predictions show that the BDT and LL identification algorithms increase the background rejection power significantly over the cut-based identification. These methods will be tested on $W \rightarrow \tau\nu$ and $Z \rightarrow \tau\tau$ events in data in the near future.

REFERENCES

1. See talks by Doerthe Ludwig and Ricardo Goncalo in these proceedings.
2. The ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 (2008) S08003.
3. The ATLAS Collaboration, Luminosity Determination Using the ATLAS Detector, ATLAS-CONF-2010-060 (2010).
4. T. Sjostrand, S. Mrenna, and P. Z. Skands, PYTHIA 6.4 Physics and Manual, JHEP 05 (2006) 026.
5. R. Field, Min-Bias and Underlying Event at the Tevatron and the LHC, talk at the Fermilab MC Tuning Workshop (Oct 2002). <http://www-cdf.fnal.gov/physics/conferences/cdf8547RDFTeV4LHC.pdf>.
6. The ATLAS Collaboration, The ATLAS Simulation Infrastructure, arXiv:1005.4568v1 [physics.ins-det]. Submitted to Eur. Phys. J. C.
7. The ATLAS Collaboration, ATLAS Monte Carlo tunes for MC09, ATL-PHYS-PUB-2010-002 (2010).
8. The ATLAS Collaboration, Expected Performance of the ATLAS Experiment Detector, Trigger and Physics, CERN-OPEN-2008-020 (2008).
9. M. Cacciari, G. P. Salam, and G. Soyez, The anti-kt jet clustering algorithm, JHEP 04 (2008) 063.
10. The ATLAS Collaboration, Calorimeter clustering algorithms: Description and performance, ATL-LARG-PUB-2008-002 (2008).
11. L. Breiman, J. Friedman, C. Stone, and R. Olshen, Classification and Regression Trees. Chapman& Hall, 1984.
12. Y. Freund and R. Shapire, Experiments with a New Boosting Algorithm, in Proceedings 13th International Conference on Machine Learning. 1996.
13. The ATLAS Collaboration, Properties of Jets and Inputs to Jet Reconstruction and Calibration with the ATLAS Detector Using Proton-Proton Collisions at $\sqrt{s} = 7$ TeV, ATLAS-CONF-2010-053.
14. The ATLAS Collaboration, Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at $\sqrt{s} = 7$ TeV and measured with the ATLAS detector, ATLAS-CONF-2010-056.