Tau reconstruction and identification in the ATLAS experiment

Andrew Leister*
On behalf of the ATLAS Collaboration
Yale University
E-mail: andrew.gerard.leister@cern.ch

In the ATLAS detector, semi-hadronic decays of tau leptons are reconstructed as collimated jets with low track multiplicity. Due to the background from multijet processes, efficient tau identification techniques with large jet rejection are essential. Since single variable criteria are not enough to efficiently distinguish taus from jets and electrons, modern multivariate techniques are used. In ATLAS, several advanced algorithms are applied to identify taus, including a projective likelihood estimator and boosted decision trees. All multivariate methods use several variables exploiting detailed information from calorimeter and tracking detectors. The algorithms and their performance are presented in detail, using $\sqrt{s} = 8$ TeV data collected at the ATLAS experiment during the 2012 LHC run and simulation data.

---

*Speaker.

1. Introduction

These proceedings summarize the techniques and algorithms used to reconstruct and identify semi-hadronically decaying taus in the ATLAS experiment [1, 2]. Taus are the only lepton to decay semi-hadronically, which occurs in 64.8% of tau decays. In this decay mode (referred to as τ_{had}), the tau decays to a tau neutrino plus one or more hadrons (predominantly pions).

The tau proper lifetime is too short for them to be detected directly by ATLAS. Instead, they are reconstructed from their decay products. After τ_{had} reconstruction, various identification algorithms, described in these proceedings, are used to discriminate τ_{had} from various backgrounds. The performance of these algorithms is tested by measuring the tau identification (ID) efficiency. Leptonic tau decays are not considered in this study.

2. Reconstruction and Identification of Hadronic Tau Candidates

τ_{had} candidates are reconstructed using an anti-k_t algorithm (R=0.4). They are seeded from jet objects with $E_{\text{T}} \geq 10$ GeV and $|\eta| \leq 2.5$. Tracks are then associated to the candidate if they fall within the core region, or $\Delta R < 0.2$, of the jet barycentre, where $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Tracks must also pass quality criteria: cuts on minimum p_{T}, silicon detector hits, and impact parameters. Tracks within an isolation annulus ($0.2 < \Delta R < 0.4$ of the barycenter) are also used to calculate ID variables and must pass quality criteria, though they are not associated to the tau candidate's core region. τ_{had} candidates are defined separately as 1-prong (1 track) or multi-prong (mostly 3 tracks).

To ensure performance in high-pileup conditions, an algorithm called the Tau Jet Vertex Association is used to calculate the primary vertex of the τ_{had} [3]. This minimizes the impact of additional interactions, which can cause tau tracks to fail the z_0 impact parameter requirement.

The tau reconstruction provides little rejection against various multi-jet backgrounds, which can be very difficult to distinguish from τ_{had}. Features of τ_{had}, such as their narrow calorimeter clusters, isolation, and distinct 1- or 3-prong track signatures, can be used to discriminate between the two. During reconstruction, tracks and calorimeter clusters are used to define several ID variables, which are used to discriminate taus from quark- or gluon-initiated jets and other leptons. Variables that provide the strongest discriminating power are ones that describe the shower shape in the calorimeter or tracks (such as energy fraction within $\Delta R < 0.1$ and average p_{T}-weighted track distance from tau axis) and ones based on the number of tracks.

Several multi-variate algorithms, collectively referred to as the tau ID, have been developed to reject quark- or gluon-initiated backgrounds from τ_{had} samples using these ID variables. Two multivariate techniques are used to reject jet backgrounds: a projective likelihood method (LLH) and method using boosted decision trees (BDT). To train these algorithms, MC samples of $Z \rightarrow \tau\tau$, $W \rightarrow \tau\nu_{\tau}$, and $Z' \rightarrow \tau\tau$ are used for signal τ_{had}, while jet-enriched data is used for backgrounds. The LLH and BDT are trained separately for 1-prong and multi-prong candidates. For each tauID method, three thresholds are established based on target signal efficiency: loose, medium, and tight. For 1-prong (multi-prong) taus, these efficiencies are 70% (65%), 60% (55%), and 40% (35%) respectively.

The signature of a 1-prong τ_{had} can be also mimicked by electrons, which create significant background. There are still characteristics that distinguish the two, such as transition radiation

properties and calorimeter shower shape. A BDT algorithm for electron discrimination utilizes variables describing these properties to reject electrons faking taus. This BDT is optimized using simulated $Z \to \tau\tau$ for signal and $Z \to ee$ for background. A muon can also be misidentified as a τ_{had} if the muon track is associated with a sufficiently energetic calorimeter cluster. To counter this, a cut-based algorithm is used for muon discrimination. Removing τ_{had} that overlap with reconstructed muons in analyses, supplemented by a cut on the energy fraction from the EM calorimeter, is most effective in distinguishing τ_{had} signal from the muon background.

3. Tau Identification Efficiency

To measure the performance of the tau ID, the identification efficiency is measured directly in data. To do this, a tag-and-probe method is used for 3 processes involving decays to 1 τ_{had}: $Z \to \tau_{\text{lep}}\tau_{\text{had}}$, $W \to \tau_{\text{had}}\nu_\tau$, and $t\bar{t} \to \tau_{\text{had}}$ + jets.

The number of τ_{had} (determined in data by a fit to number of tracks) is counted before and after the tau ID is implemented. This is used to obtain the identification efficiency in data ($\varepsilon_{\text{data}}$) and MC ($\varepsilon_{\text{MC}}$). Scale factors, defined as $\varepsilon_{\text{data}}/\varepsilon_{\text{MC}}$, quantify the tau ID performance and are used in analyses to account for discrepancies between data and MC. The $Z \to \tau_{\text{lep}}\tau_{\text{had}}$ results (Table 1) are used as the main measurement since it offers the highest precision due to low associated backgrounds. The $W \to \tau_{\text{had}}\nu_\tau$ results are used as a cross check, while the $t\bar{t} \to \tau_{\text{had}}$ + jets offer a measurement for a higher kinematic (higher p_{T}) regime.

	scale factor $\pm$ sys. unc. $\pm$ stat. unc.	
	BDT	LLH
Loose	$1.033 \pm 2.0\% \pm 1.0\%$	$1.044 \pm 1.7\% \pm 1.0\%$
Medium	$0.979 \pm 2.1\% \pm 1.1\%$	$0.985 \pm 2.1\% \pm 1.1\%$
Tight	$0.907 \pm 2.6\% \pm 1.5\%$	$0.941 \pm 2.4\% \pm 1.5\%$

Table 1: 2012 8-TeV (8.5-10.5 fb^{-1}) Z$\to \tau_{\text{lep}}\tau_{\text{had}}$: overall scale factors and uncertainties

References

[1] ATLAS Collaboration, *Identification of the Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector*, ATLAS-CONF-2013-064, https://cds.cern.ch/record/1562839.

[2] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, 2008 JINST 3 S08003.

[3] ATLAS Collaboration, *Performance of the Reconstruction and Identification of Hadronic tau Decays in ATLAS with 2011 Data*, ATLAS-CONF-2012-142, https://cds.cern.ch/record/1485531.