PROCEEDINGS OF SCIENCE

Hadronic tau reconstruction and identification performance at ATLAS & CMS

Izaak Neutelings[1,*]

University of Zurich, Physik-Institut
Winterthurerstrasse 190, Zurich, Switzerland

E-mail: izaak.neutelings@uzh.ch

The τ lepton is an important part of the final states of many standard model processes in proton-proton collisions at the LHC. It also appears in many signatures of predicted new physics beyond the standard model. As it decays predominantly to hadrons, it is essential for the physics program at ATLAS and CMS to have a faithful reconstruction and identification of hadronic τ lepton decays, and a strong suppression of backgrounds. In the following, the ATLAS and CMS algorithms for the reconstruction and identification of hadronic τ lepton decays are presented, and their performance in proton-proton collision data collected during Run 2 of the LHC is discussed.

The Eighth Annual Conference on Large Hadron Collider Physics-LHCP2020
25-30 May, 2020
online

[1]For the ATLAS and CMS Collaborations collaboration.
[*]Speaker

PoS(LHCP2020)045

1. Introduction

Many standard model (SM) processes at the LHC have a τ lepton in their final state. This includes the observation and study of the couplings and CP properties of the Higgs boson [1, 2]. Besides SM measurements, τ leptons can appear in many signatures of predicted new physics beyond the standard model (BSM). With a mass of 1.777 GeV, the τ lepton is the only lepton that can decay to hadrons [3], and with a branching fraction of about 65%, the decay to hadrons is its dominant final state. Therefore, to faithfully reconstruct and identify hadronic τ lepton decays (τ_h) among vast amounts of backgrounds is essential for the physics program at ATLAS [4] and CMS [5]. For optimal sensitivity, a high efficiency with low misidentification rate is needed, as well as a good energy calibration and momentum resolution.

2. τ_h reconstruction and identification

The τ lepton decays about 11.5% of the time to one charged hadron, 35.5% to one charged hadron and neutral hadrons, and 15% to three charged hadrons and neutral hadrons. Most hadronic τ lepton decays under study at ATLAS and CMS have an energy above 10 GeV, and their products therefore tend to form a collimated and well isolated jet. The τ_h reconstruction algorithms typically start from jets, before classifying them in different decay modes (DM) according to the number of charged and neutral pions. The main backgrounds are quark- and gluon-initiated jets that are misidentified as τ_h candidates. In addition to the isolation and kinematic information, the algorithms typically take advantage of the fact that the τ lepton has a lifetime of about 2.9×10^{-13} s, such that it can travel a distance of about a 1 mm before decaying.

CMS employs the so-called Particle Flow (PF) algorithm [6], which reconstructs individual electrons, muons, photons, and charged hadrons using information from all parts of the detector. The τ_h reconstruction algorithm at CMS uses as input the jets that are clustered from these PF candidates using the anti-k_T algorithm with distance parameter $R = 0.4$, and have $p_\mathrm{T} > 14$ GeV and $|\eta| < 2.3$ [7]. The τ_h candidate is assigned to different decay modes by counting the number of charged hadrons and ECAL clusters called strips. The strips are defined by merging electron and photon clusters in the η-ϕ space. This is done to identify π^0s, which decay promptly to two photons that in turn can convert to an electron-positron pair. The size of the strip is adjusted dynamically as a function of the p_T of the electron and photon candidates. The algorithm considers all possible combinations of charged candidates and strips that are consistent with a τ_h decay. They are required to have a total charge of ± 1 and be within a signal cone of radius in a range of 0.05–0.1 and a mass window determined by the DM.

Recently CMS has developed a convolutional deep neural network (DNN) to suppress backgrounds. Similar to the previous boosted decision tree (BDT) discriminator, the DNN takes as input information on high-level variables, such as quantities related to lifetime of the τ_h candidate, isolation and kinematics of its electron and photon constituents, but also information on the PF hadron, muon, electron and photon in the jet [8], see Fig. 1 (left). The PF candidates are split in $\eta \times \phi$ cells, which are processed by several convolutional layers, after which they are combined with the high-level information in five dense layers. From the DNN output three discriminants are constructed that very efficiently reject misidentified muons, electrons and jets. For an efficiency

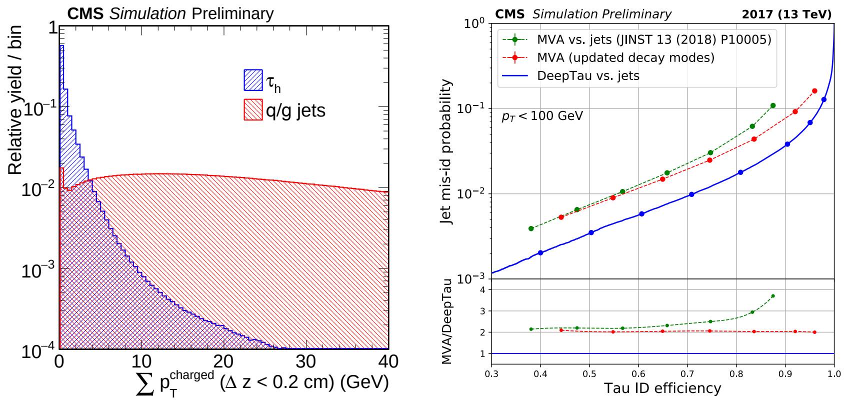

Figure 1: Left: example of a variable with strong discriminating power, the p_T-sum of charged tracks [7]. Right: jet-misidentification probability versus true τ_h efficiency, comparing the BDT- and DNN-based τ_h identification algorithms of CMS [8].

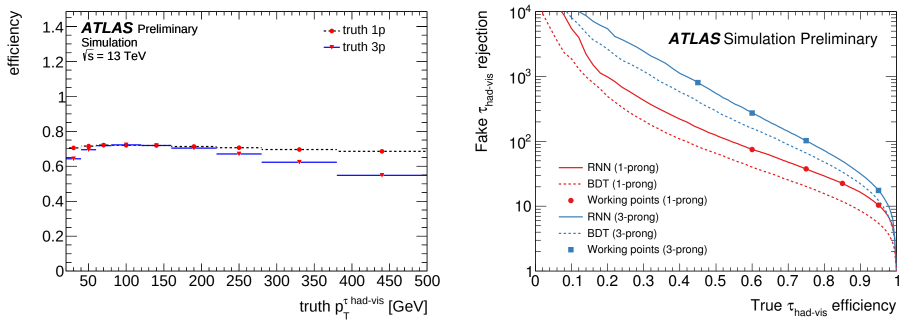

Figure 2: Left: efficiency of the baseline τ_h reconstruction at ATLAS [12]. Right: rejection of jets versus true τ_h efficiency, comparing the BDT- and RNN-based τ_h identification algorithms at ATLAS [11].

of 60%, the misidentification rate of jets is about 0.6%. As Fig. 1 (right) shows, this is a twofold improvement with respect to the previous BDT discriminator. The identification efficiency was measured in $\mathrm{Z}/\gamma^* \to \tau^+\tau^-$ events, where one τ lepton decays to a well measured muon, and the other to hadrons. By fitting the invariant mass distribution between the muon and τ_h candidate, a data-to-simulation scale factor of around 0.9 with an uncertainty of about 6% has been found, indicating a good description of the detector.

To reconstruct τ_h candidates, ATLAS starts from jets consisting of calorimeter clusters, using the anti-k_T algorithm with distance parameter $R = 0.4$ [9, 10] and requiring $p_\mathrm{T} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$. A signal cone with radius $\Delta R = 0.2$, as well as an isolation cone of $\Delta R = 0.4$ is defined around the jet's barycenter. A dedicated set of BDTs is used to identify charged tracks with $p_\mathrm{T} > 1\,\mathrm{GeV}$ and further classify them as either one or three-prong DM, using only track information. As is shown in the left panel of Fig. 2, the efficiency to reconstruct τ_h is roughly 70% for $p_\mathrm{T} < 200\,\mathrm{GeV}$ for both the one- and three-prong decays.

Most recently, ATLAS has deployed a recurrent neural network (RNN) that uses similar high-level information as mentioned for CMS, but it also takes as input low-level information on the track and cluster inputs [11]. The high-level, tracks and cluster inputs are processed separately by dense layers, before they are merged into three final dense layers. Figure 2 (right) compares the rejection of misidentified τ_h candidates of the RNN and the previously used BDT. It indicates that for a typical efficiency of 60%, the RNN has a jet misidentification rate of about 0.8% (3%) for

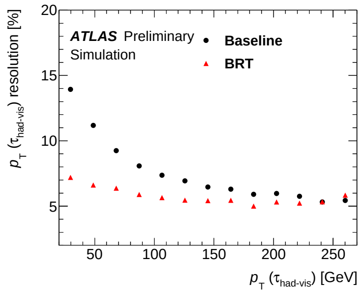

Figure 3: Resolution of the p_T comparing the calorimeter- (black) and BRT-based calibration (red) at ATLAS [10].

one-prong (three-prong) DMs. The efficiency has been measured in $Z/\gamma^* \to \tau^+\tau^-$ events, fitting the distribution of the number of tracks. A data-to-simulation scale factor close to unity with an uncertainty of about 3% has been determined, indicating a good description of the ATLAS detector.

ATLAS has developed a second algorithm to significantly improve the momentum resolution below 100 GeV. This algorithm, called the Tau Particle Flow (TPF) [9], takes advantage of tracking in order to identify neutral hadron candidates as well. After the charged hadron candidates are subtracted, the remaining clusters are identified by a specialized BDT. Then, a second BDT takes all the charged and neutral pion candidates and assigns the most likely τ_h DM.

3. τ_h energy calibration

As the τ_h reconstruction at CMS is based on the already well-calibrated PF objects, the τ_h energy response is close to unity and well modeled by the simulation with a resolution around 10% [13]. The energy scale in simulation is measured in $Z/\gamma^* \to \tau^+\tau^-$ events by fitting the τ_h mass distribution with signal templates of varying energy scale. The energy scale correction is of the order of 1% and depends on the DM [14].

A boosted regression tree (BRT) with the true τ_h p_T as target uses an interpolation between the calorimeter-based momentum and TPF-based momentum, to take advantage of the TPF's improved resolution below 100 GeV. The calibration using the BRT approach improves the resolution significantly up to about 200 GeV, as is presented in Fig. 3 (right). At momenta below 30 GeV, the resolution becomes 7% compared to 14% for the calorimeter-based calibration alone.

4. Conclusions

The reconstruction and identification of hadronic τ lepton decay in ATLAS and CMS has been presented. While each detector has a unique approach to τ_h reconstruction, they have comparable performance. Recent development of algorithms based on neural networks have significantly improved background rejection, down to less than 1% misidentification for 60% efficiency. Measurements show a good understanding of the detector needed for a faithful description of τ_h decays.

References

[1] CMS Collaboration, *Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector*, *Phys. Lett. B* **779** (2018), 283-316 doi:10.1016/j.physletb.2018.02.004 [arXiv:1708.00373 [hep-ex]].

[2] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, *Phys. Rev. D* **99** (2019), 072001 doi:10.1103/PhysRevD.99.072001 [arXiv:1811.08856 [hep-ex]].

[3] Particle Data Group, *Review of Particle Physics*, *PTEP* **2020** (2020) no.8, 083C01 doi:10.1093/ptep/ptaa104

[4] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008), S08003 doi:10.1088/1748-0221/3/08/S08003

[5] CMS Collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008), S08004 doi:10.1088/1748-0221/3/08/S08004

[6] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017) no.10, P10003 doi:10.1088/1748-0221/12/10/P10003 [arXiv:1706.04965 [physics.ins-det]].

[7] CMS Collaboration, *Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s}$ = 13 TeV*, *JINST* **13** (2018) no.10, P10005 doi:10.1088/1748-0221/13/10/P10005 [arXiv:1809.02816 [hep-ex]].

[8] CMS Collaboration, *Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons*, CMS-DP-2019-033, http://cds.cern.ch/record/2694158.

[9] ATLAS Collaboration, *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, *Eur. Phys. J. C* **76** (2016) no.5, 295 doi:10.1140/epjc/s10052-016-4110-0 [arXiv:1512.05955 [hep-ex]].

[10] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s}$ = 13 TeV*, ATLAS-CONF-2017-029, http://cds.cern.ch/record/2261772.

[11] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, http://cds.cern.ch/record/2688062.

[12] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, http://cds.cern.ch/record/2064383.

[13] CMS Collaboration, *Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS*, *JINST* **11** (2016) no.01, P01019 doi:10.1088/1748-0221/11/01/P01019 [arXiv:1510.07488 [physics.ins-det]].

[14] CMS Collaboration, *Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s}$ = 13 TeV*, *JINST* **13** (2018) no.10, P10005 doi:10.1088/1748-0221/13/10/P10005 [arXiv:1809.02816 [hep-ex]].

PoS(LHCP2020)045