PROCEEDINGS OF SCIENCE

Hadronic tau reconstruction and identification performance in ATLAS and CMS

Camilla Galloni* **on behalf of the ATLAS and CMS Collaborations**

E-mail: cgalloni@cern.ch

The reconstruction and identification of hadronic tau lepton decays is an essential requisite of the ATLAS and CMS physics program, for standard model measurements, such as the ones involving the Higgs boson decays, and for searches of new physics beyond the standard model. In the following, the ATLAS and CMS algorithms for the reconstruction and identification of hadronic tau lepton decays are presented, together with the measurements of the most important performance with *pp* collision data collected during the LHC Run 2.

Sixth Annual Conference on Large Hadron Collider Physics (LHCP2018)
4-9 June 2018
Bologna, Italy

*Speaker.

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

PoS(LHCP2018)228

1. Introduction

A solid and well-understood reconstruction of hadronic tau lepton decays (τ_{h}) is an important requirement for many standard model (SM) measurements and searches for physics beyond the standard model (BSM) conducted at the ATLAS and CMS experiments. Since the Yukawa couplings are proportional to the fermion masses, the Higgs boson decay to τ leptons has the highest branching fraction among leptons. In fact, the Higgs boson decay to a pair of τ leptons has recently been observed by both the experiments [1, 2]. The identification of hadronic τ lepton decays is vital to measure the coupling strength of the Higgs boson to τ leptons, as well as its charge and parity properties. Particles appearing in BSM models such as leptoquarks or extended gauge bosons can have large branching fractions to heavy fermions, thus they couple preferably to τ leptons. In order to optimize the sensitivity of these analyses, the most important aspects of the identification of hadronic τ lepton decays are a high efficiency of reconstructing these decays (with a sustainable misidentification rate) and an energy scale calibration close to the true scale with a good momentum resolution. In the following, the dedicated algorithms in use at ATLAS and CMS are presented.

2. Tau lepton properties

The τ lepton is the heaviest lepton with a mass of 1.777 GeV [3]. In 35% of the cases it decays leptonically into a lighter lepton, either an electron (e) or a muon (μ), and neutrinos (ν), while it decays to hadrons and a neutrino in the remaining 65%: 11.5% to one charged hadron, 35.5% to one charged hadron and neutral hadrons and 15% to three charged hadrons and any number of neutral hadrons. The mean life time of a τ lepton is $2.9 \cdot 10^{-13}$ s, thus, for a momentum of few tens of GeV, it can travel few millimeters before decaying. Most of the times, it decays before reaching the innermost sensitive detector layer, therefore only its decay products are observed.

3. Tau lepton reconstruction and identification

Hadronic τ lepton decay products are detected as one or three tracks in the tracking systems, and energy deposits in the electromagnetic and hadronic calorimeters, consistent with charged and neutral hadrons, most commonly, pions (π). Although also electrons and muons can mimic a τ_{h} decay with just one charged hadron, the main source of background for τ_{h} identification are quark- or gluon-initiated jets, due to their large production cross section. With respect to the latter, τ_{h}-originated jets have a lower track multiplicity and a larger electromagnetic energy component since one or more π^0s are present in 60% of τ_{h} decays. Differently from gluons and light quarks, τ leptons can decay relatively far from the interaction point, thus have a track with an impact parameter incompatible with the primary vertex or a displaced secondary vertex.

The ATLAS τ_{h} algorithm constists of two steps [4]. First the τ_{h} are reconstructed using as input jets with $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$, reconstructed with the anti-k_{T} algorithm with a distance parameter value of 0.4 applied to clusters of calorimeter cells, and calibrated using a local hadronic calibration. Tracks are associated to the candidate if they are within a cone of radius 0.2 from the candidate direction and have $p_{\mathrm{T}} > 10$ GeV. Candidates with 1 or 3 tracks are considered. The candidate momentum is built using only clusters of calorimeter cells within a cone of radius 0.2.

PoS(LHCP2018)228

As shown in Fig.1(left), the efficiency to reconstruct τ_h is roughly 70% for $p_T(\tau_h) < 250$ GeV, for both the one- and three-prong decays. For the latter, at high-p_T the efficiency slightly decreases due to the increased collimation of the decay products that results in an increased probability to miss a track because of overlapping trajectories [4].

The second identification step is based on a boosted decision tree (BDT), trained separately for the one- and three-prong decays to discriminate τ_h from jets, using variables related to calorimeter deposits and the reconstructed tracks. The variables exploited are related to the collimation of tracks and calorimeter cells, the fraction of energy deposited in the electromagnetic calorimeter, the balance of track momentum and calorimeter energy measurements, the presence of a significant impact parameter or secondary vertex and the invariant mass of the hadronic τ decay products. The input variables are corrected for the average effect of additional collisions in the same or adjacent bunch crossings (pileup). In Fig.1(right) multiple working points are defined for the τ_h selection and jet misidentification probability.

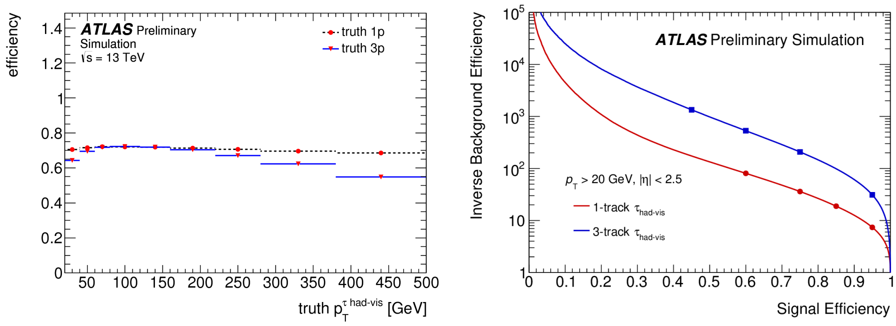

Figure 1: ATLAS reconstruction efficiency for τ_h with one or three prongs as a function of the $p_T(\tau_h)$ (left) [4] and efficiency and misidentification rate of jets as τ_h for different requirements on the BDT (right).

The CMS τ_h algorithm starts with objects reconstructed by the particle-flow (PF) algorithm [5]. The PF objects used to reconstruct individual hadronic decay products are electrons, muons, photons, and charged and neutral hadrons. The reconstructions is seeded by jets of PF candidates clustered with the anti-k_T algorithm with a distance parameter of 0.4, with $p_T > 14$ GeV and $|\eta| < 2.3$. Combinations of charged and neutral hadrons compatible with a hadronic τ decay are identified inside the jet. Electrons and photons are clustered in the electromagnetic calorimeter in strips in the $\eta - \phi$ space, in order to identify π^0s that decay to photons which can convert to pairs of electrons and positrons that are bent by the magnetic field, leading to an elongation of the strips in the ϕ dimension. Since 2015, the size of the strip is adjusted dynamically as a function of the p_T of the electron and photon candidates. Charged candidates, photons, and strips are required to have $p_T > 0.5$ GeV, $p_T > 1$ GeV, and $p_T > 2.5$ GeV, respectively. All possible combinations of charged candidates and π^0 candidates are considered if the the constituents are within a cone of radius $\Delta R = 0.3/p_T$ from the τ_h direction, within the range 0.05–0.1. Each combination is required to have unity electric charge and an invariant mass compatible with one of the intermediate resonance characteristic of each τ_h decay mode (DM). The combination with the highest $p_T(\tau_h)$ is then chosen as the reconstructed τ_h.

To reduce the number of jets misidentified as τ_h two kinds of isolation are used. A cut-based isolation is defined, with the charged particles and the photons within a cone of radius 0.5 from the

PoS(LHCP2018)228

τ_h direction, as:

$$I_{\tau_h} = \sum p_T^{ch}(d_z < 0.2\text{cm}) + \max(0, \sum p_T^{\gamma} - \Delta\beta \sum p_T^{ch}(d_z > 0.2\text{cm})) \quad (1)$$

where the d_z is the longitudinal distance from the τ_h vertex and the correction $\Delta\beta = 0.2$ is used to correct the neutral isolation component for pileup, based on its charged component. The CMS MVA isolation is based on a BDT trained on simulated samples with the following variables: the τ DM, the isolation terms of eq. (1), the number of photons and electrons and their distributions with respect to the τ_h direction, the distributions of the energy deposits, and quantities related to the displaced impact parameter or the secondary vertex (for the three-prong decay), as depicted in Fig.2(left). The cut-based and MVA isolation expected performance are shown in Fig.2(right).

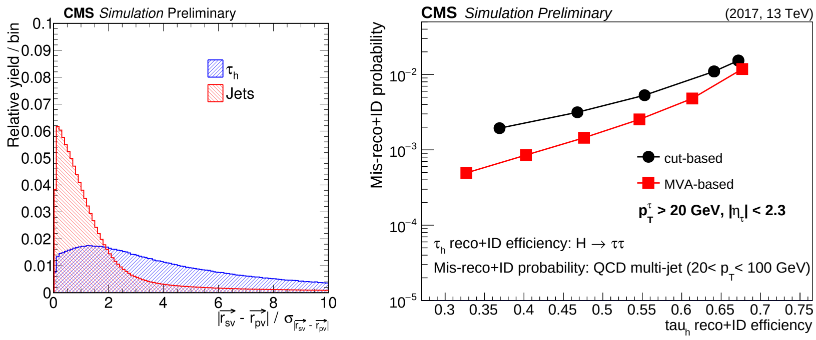

Figure 2: Flight distance significance for 3 prongs τ_h [6] and performance of isolation discriminators [7].

4. Tau lepton energy calibration

In ATLAS, the τ_h are reconstructed from jets corrected for local hadronic calibration, that account for the non-compensating nature of the calorimeter and for energy depositions outside the non-instrumented regions of the detector. These corrections, however, are not optimized for the τ_h jet smaller radius of 0.2 and hadron content, which is different from the one of a quark or gluon jets. Corrective functions are derived using simulations to remove the differences between the reconstructed and true energy (on average), depending on the number of tracks and the $\eta(\tau_h)$. Pileup corrections are also applied. The resolution of this baseline energy calibration is shown in Fig.3(left). Recently, an alternative energy calibration has been developed, based on reconstructing the hadronic τ decay products. Calorimeter and tracker information are combined with charged and neutral hadrons, identified by the BDT Tau Particle Flow (TPF), in a boosted regression tree (BRT) [8]. As shown in Fig.3(left) the TPF-BRT energy resolution is around 7% compared to 14% for the baseline $p_T = 30$ GeV. For $p_T > 200$ GeV both approaches give similar results. Then the assigned $p_T(\tau_h)$ is calibrated with a combination of the baseline and the BRT methods.

Since the CMS τ_h reconstruction deploys already calibrated PF objects, the τ_h energy response is close to unity and well modeled by the simulations with a resolution around 10% [9]. Corrections for small residual differences in the energy scale between data and simulations are measured in $Z/\gamma* \to \tau\tau \to \mu\tau_h + 3\nu$s data with maximum likelihood (ML) fits of two distributions: the mass of the visible τ decay products $m(\tau_h)$ and the invariant mass of the τ_h and the μ $(m_{vis}(\mu, \tau_h))$.

PoS(LHCP2018)228

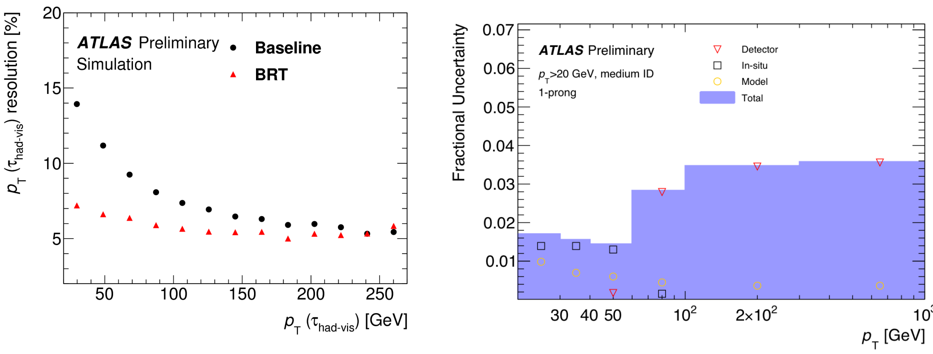

Figure 3: ATLAS $p_T(\tau_h)$ resolution comparing the baseline calibration to the TPF-BRT approach [8] (left). Fractional uncertainty on the τ energy calibration (right).

Templates with energy scales varying within a range of $\pm 6\%$ are used to extract the best fit values for the 2 distributions and the τ DMs, independently. The derived corrections are close to 1% and in agreement for the two distributions within the uncertainties.

5. Light lepton misidentification suppression

Electron can mimic the one-prong τ_h decay, since they have a charged track and can emit additional bremsstrahlung radiation that can be misidentified as π^0s. In ATLAS, this background is suppressed by vetoing object that satisfy the regular electron identification criteria. In CMS, a BDT-based discriminator is used: typically used requirements correspond to a 75% efficiency, with a misidentification rate of 10^{-2}–10^{-3}. Discriminators for muon misidentification have an efficiency of 95%–100%. In ATLAS, the τ_h candidate that overlaps with an identified muons is discarded. In CMS, τ_h candidates with matching segments in the muon detectors are rejected, with a misidentification rate of 10^{-3}–10^{-4}.

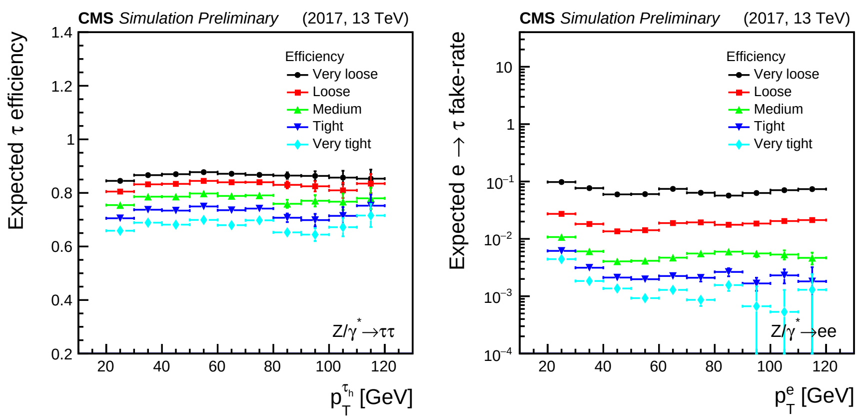

Figure 4: CMS e$\to$ τ_h misidentification suppression: discriminator efficiency on true τ_h and misidentification rate evaluated on simulated $Z/\gamma* \to \tau\tau$ and $Z/\gamma* \to ee$, respectively [7].

PoS(LHCP2018)228

6. Tau lepton performance validation

Both experiments use for the validation $Z/\gamma* \to \tau\tau \to \mu\tau_{\mathrm{h}} + 3\nu$ events. For ATLAS, the identification efficiency is measured with a fit to the distribution of the number of tracks before and after requiring identification criteria (Fig.5 (left)). In CMS, the fitted distribution is the $m_{vis}(\mu, \tau_{\mathrm{h}})$ spectrum after requiring the identification criteria (Fig.5 (right)), where the misidentified jet component is estimated from data, and the $Z/\gamma*$ production cross section is simultaneously fitted for in di-muon events. The expectation from simulations is found in agreement with data. In ATLAS, the uncertainties vary between 4% and 10% depending on the number of tracks and p_{T}. In CMS, the uncertainty is estimated from the fit to data inclusively in the DM. An extrapolation is done at high p_{T}, using efficiency measurements in $\mathrm{t\bar{t}}$ and $\mathrm{W}^* \to \tau\nu$ events, and the uncertainty is estimated conservatively to be 30% at $p_{\mathrm{T}} = 1$ TeV.

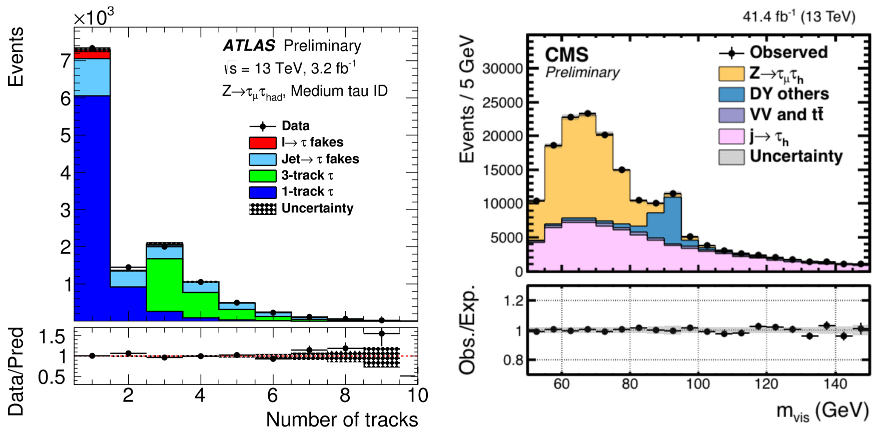

Figure 5: Left: number of tracks after passing the τ_{h} identification criteria, ATLAS [8]. Right: distribution of the $m_{vis}(\mu, \tau_{\mathrm{h}})$ spectrum after requiring the identification criteria, CMS [7].

The performance of anti-lepton discriminators is evaluated in data with tag-and-probe methods, where data and simulation agreement is checked for different η regions to reflect a possible dependency on the modeling of different parts of the detectors. Data-to-simulation corrective factors for the efficiency of the discriminators are generally close to unity in the central part of the detectors and increase slightly in the forward part.

The τ_{h} energy scale and resolution are validated in the same final state using $m(\tau_{\mathrm{h}})$ and $m_{vis}(\mu, \tau_{\mathrm{h}})$. The resolution and scale uncertainties are comparable for both experiments, the resolution is roughly 10% depending on p_{T}, η and decay mode of the τ_{h} candidate. At CMS, the scale uncertainty is evaluated from the ML fit to data inclusively and amounts to 1–3% [6]. In ATLAS, the energy scale uncertainty takes into account uncertainties evaluated from in-situ data, uncertainties for high p_{T} derived in single hadrons data and additional modeling uncertainties from simulations as shown in Fig.3. Total uncertainties vary between 2–4% depending on the p_{T}, η and number of tracks. Figure 6 shows the $m(\tau_{\mathrm{h}})$ with ATLAS 2012 data (left) and CMS 2017 data (right).

PoS(LHCP2018)228

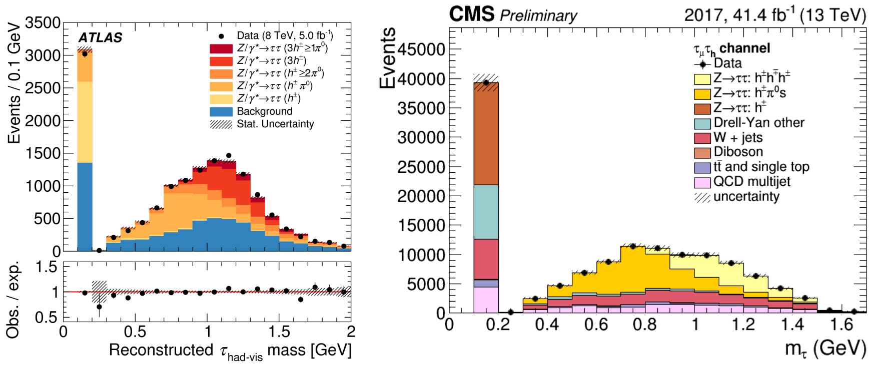

Figure 6: Distributions of the $m(\tau_h)$ for ATLAS 2012 data (left) [10]; and CMS 2017 data (right) [7].

7. Conclusions

The reconstruction and identification of hadronic tau leptons decays in ATLAS and CMS have been presented. The two experiments use conceptually different approaches that recently started to converge (the ATLAS adoption of BDT-based particle flow and the CMS usage of MVA-based isolation). Overall, both experiments deliver an excellent performance of tau lepton identification.

References

[1] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying to a pair of τ leptons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2018-021, [http://cds.cern.ch/record/2621794].

[2] CMS Collaboration, *Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector*, Phys.Lett. B779 (2018) 283, [arXiv:1708.00373]

[3] M. Tanabashi et al. (Particle Data Group), *Review of Particle Physics*, Phys. Rev. **D** 98, 030001 (2018)

[4] 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].

[5] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, JINST *12* (2017) P10003 [arXiv:1706.04965]

[6] CMS Collaboration, *Performance of reconstruction and identification of τ leptons in their decays to hadrons and τ neutrino in LHC Run-2*, CMS-PAS-TAU-16-002 [https://cds.cern.ch/record/2196972].

[7] CMS Collaboration, *Tau identification performance in 2017 data at $\sqrt{s}$=13 TeV*, CMS-DP-2018/026 [https://cds.cern.ch/record/2622155].

[8] 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].

[9] CMS Collaboration, *Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS*, JINST *11* (2016) P01019 [arXiv:1510.07488]

[10] ATLAS Collaboration, *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, Eur. Phys. J C **76** (2016) 295 [arXiv:1512.05955].

PoS(LHCP2018)228