PROCEEDINGS OF SCIENCE

Performance of reconstruction and identification of τ leptons in their hadronic decays in pp collisions at $\sqrt{s} = 13$ TeV at CMS

Olena Hlushchenko[*†]
RWTH Aachen university
E-mail: olena.hlushchenko@cern.ch

Since the beginning of LHC Run 2, many improvements have been made to the tau lepton reconstruction algorithm at CMS. The hadrons-plus-strips (HPS) tau leptons reconstruction algorithm now benefits from a dynamic strip reconstruction, and has been extended to a version intended for highly Lorentz-boosted topologies. In addition, multivariate discriminators used for tau lepton identification now combine isolation with lifetime variables. The methodology and performance of the newly developed algorithms and discriminators are presented using 35.9 fb^{-1} and 41.4 fb^{-1} of proton-proton data collected with CMS in the 2016 and 2017 LHC runs.

The 39th International Conference on High Energy Physics (ICHEP2018)
4-11 July, 2018
Seoul, Korea

[*]Speaker.
[†]On behalf of CMS collaboration.

PoS(ICHEP2018)692

1. Overview

Proper reconstruction and identification of tau leptons is crucial for various analyses both in the standard model (SM) and beyond standard model (BSM) sectors. Since Higgs boson Yukawa couplings are expected to be proportional to the mass, the tau lepton with its highest mass among other leptons of $m_\tau = 1.777$ GeV becomes interesting for measuring the Higgs boson couplings to fermions.

At CMS most of the high-level analyses are based on the particles identified and reconstructed by the particle-flow algorithm [1] which combines information from all CMS sub-detectors on single proton-proton collision basis (per event). The particles are classified as photons, muons, electrons, charged or neutral hadrons.

Tau lepton's hadronic decays require a dedicated approach in order to discriminate jets originating from a genuine tau lepton from any other occurring jet, the so-called hadrons-plus-strips (HPS) algorithm. It was developed for use in the LHC Run-1 [2] and further optimized for the the Run-2 [3] campaign. Seeded by the particle-flow reconstructed objects, the algorithm attempts to reconstruct individual charged (mainly $\pi^\pm$) and neutral (mainly π^0) hadrons originating from the tau lepton decay.

The neutral hadrons are reconstructed as "strips" in the scope of the HPS algorithm referring to their detector response. Photons from the prompt π^0 decays have a high probability to convert to an e^+e^- pair in the tracker material. Their trajectories are then bent by the magnetic field along the ϕ direction before reaching the electromagnetic calorimeter, where they are grouped into a cluster. At CMS the decay modes are defined with respect to the number of reconstructed charged hadrons (prongs) and strips (π^0). This leads to considering also the non-physical decays that arise when individual components of hadronically decaying tau leptons (τ_h) can not be fully resolved. Among considered decay modes are: 1 prong 0 π^0, 1 prong 1 π^0, 1 prong 2 π^0, 3 prongs 0 π^0 and 3 prongs N π^0.

2. Discrimination against jets and tau ID scale factors

The identification of hadronically decaying tau leptons is challenging due to their similarities with quark- or gluon-initiated jets. However, the tau-induced jets are more isolated with respect to other particles in the event, deposit energy in narrower cones, and have lower multiplicity with respect to other jets. Those features are providing a handle to reduce the jet $\rightarrow \tau_h$ misidentification probability.

At CMS two main anti-jet discriminators are used: cut-based and MVA-based. The cut values are defined in a way that assures a fixed efficiency of the tau lepton to be reconstructed and identified as such. The choice of the cut value is then referred to as a working point (WP).

The cut-based discriminator, the isolation, is constructed as a p_T sum of charged and neutral components of the tau lepton candidate jet after subtraction of the respective pileup contributions in an angular $(\eta;\phi)$ plane that lies inside the $\Delta R < 0.5$ isolation cone, but outside the signal cone of a size that depends on the candidate's p_T.

The MVA-based discriminator, the MVA discriminant, is constructed by training a boosted decision tree classifier (BDT) on 23 input variables that are sufficient to cover a wide range of τ_h

signatures (such as shape of the jet, its energy distributions and multiplicity) or properties (such as τ-lifetime).

For the 2017 BDT training additionally to other input variables the Gottfried-Jackson angle [4] was included in case of a 3 prong decay mode. The performance of reconstruction and identification of the tau leptons as a function of p_T and discrimination WP covering the range of efficiency between 40% (VVTight) and 95% (VVLoose) for 2017 data taking campaign is shown in Fig. 1.

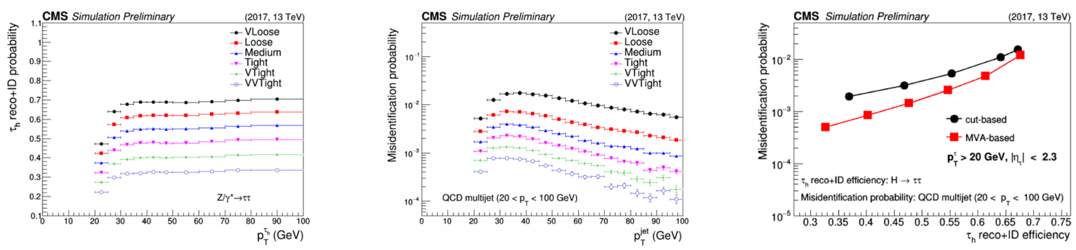

Figure 1: Expected reconstruction and identification efficiency (left) and misidentification probability (center) as a function of the tau lepton p_T for various WP of MVA discriminant, comparison of the performance of MVA and cut-based discriminants (right) [5].

To account for differences in the reconstruction and identification efficiencies in data and MC the tau ID scale factors (SF) are estimated using the tag-and-probe technique for each of the WPs independently. This is done by exploiting the Z, W' and $t\bar{t}$ decays with τ_h in the final state. Additionally the SF are extrapolated to the not accessible p_T range.

The simulated τ_h reconstructed energy is additionally corrected for differences between data and simulated events. The correction is estimated by selecting $Z \to \tau_h\tau_\mu$ events separately for each possible decay mode of τ_h and performing a maximum likelihood fit leaving the energy scale as a floating parameter [6]. The tau ID SF are estimated with a precision of less than 5% uncertainty for data collected during 2016 and 2017 data taking periods [5].

3. Discrimination against light leptons and misidentification rate measurements

In final states involving τ_h, background events can also arise from $Z \to ee$ or $Z \to \mu\mu$ decays, where an electron or a muon is misidentified as τ_h.

To suppress these backgrounds dedicated discriminants are developed [2] and the corresponding misidentification rates are estimated using tag-and-probe technique in $Z \to ee$ or $Z \to \mu\mu$ selected events.

To suppress the $e \to \tau$ misidentification probability, a dedicated BDT classifier is trained. Variables such as photon multiplicity per strip, shower shape related and variables that quantify the contribution of strips to the final τ_h observables are used as input for the BDT. The classifier is shown to have a stable,

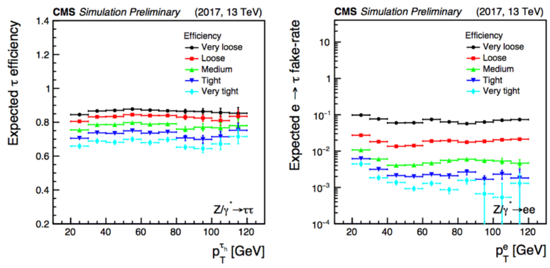

Figure 2: Expected τ_h identification (left) and $e \to \tau_h$ misidentification (right) probabilities as a function of the transverse momentum of a τ_h candidate [5].

over wide p_T range, efficiency of approximately 75% and misidentification probability smaller than 1% as is shown in Fig. 2. The uncertainty is strongly dependent on the chosen WP and ranges from 5% to 40% [5].

To reduce the $\mu \to \tau$ misidentification, candidates matching fired segments in the muon detector are discarded. This approach is shown to have close to 100% efficiency. The corresponding misidentification scale-factors are measured in η-bins and vary between 1 in central and 2 in forward regions with uncertainties up to 10%. The results are cross-checked with an analytical approach where the normalizations of the signal ($Z \to \mu\mu$) and background ($Z \to \tau\tau$) components are obtained with an analytical fit. Both methods are found to be in agreement in range of the estimated uncertainties [5].

4. Summary

Concluding, after the reconstruction, identification and application of estimated scale factors, we observe a very good agreement between the data and simulation as shown in Fig. 3. Various decay modes of the tau leptons are well resolved and identified proving an excellent understanding of the detector and reconstruction.

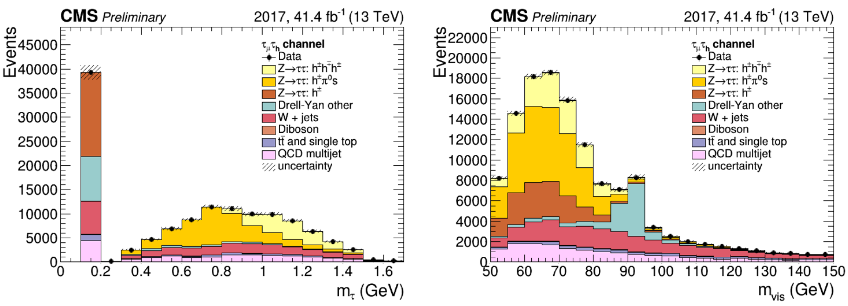

Figure 3: The mass of the visible tau (left) and di-tau (right) decay products in $Z \to \tau_\mu\tau_h$ selected events using MVA tight tau-ID WP [5].

References

[1] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017) no.10, P10003, [arXiv:1706.04965]

[2] CMS Collaboration, *Performance of tau-lepton reconstruction and identification in CMS*, *JINST* **7** (2012) P01001, [arXiv:1510.07488]

[3] CMS Collaboration, *Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV*, CMS-TAU-16-003, [arXiv:1809.02816]

[4] K. Gottfried and J. D. Jackson, *On the connection between production mechanism and decay of resonances at high energies*, *Nuovo Cim.* **33** (1964) 309

[5] CMS Collaboration, *Tau Identification Performance in 2017 Data at $\sqrt{s} = 13$ TeV*, CMS-DP-2018-026

[6] CMS Collaboration, *Performance of tau identification with 2016 data at $\sqrt{s}$ = 13 TeV*, CMS-DP-2017-006