

## Tau Identification at CMS in Run II

---

**Isabel Ojalvo**<sup>\*†</sup>

*University of Wisconsin-Madison (US)*

*E-mail: [isabel.rose.ojalvo@cern.ch](mailto:isabel.rose.ojalvo@cern.ch)*

During LHC Long Shutdown 1 necessary upgrades to the CMS detector were made. CMS also took the opportunity to improve further particle reconstruction. A number of improvements were made to the Hadronic Tau reconstruction and Identification algorithms. In particular, electromagnetic strip reconstruction of the Hadron plus Strips (HPS) algorithm was improved to better model signal of  $\pi^0$  from tau decays. This modification improves energy response and removes the tau footprint from isolation area. In addition to this, improvement to discriminators combining isolation and tau life time variables, and anti-electron in MultiVariate Analysis technique was also developed. The results of these improvements are presented and validation of Tau Identification using a variety of techniques is shown.

*38th International Conference on High Energy Physics  
3-10 August 2016  
Chicago, USA*

---

<sup>\*</sup>Speaker.

<sup>†</sup>A footnote may follow.

## 1. Overview

Understanding the mechanism of electroweak symmetry breaking through which the W and Z bosons become massive is a central goal of the Large Hadron Collider (LHC) physics program. In the Standard Model, the Higgs mechanism allows for electroweak symmetry breaking and calls for the existence of a new massive particle, the Higgs boson. During Run I at the LHC the discovery of the Higgs boson was announced by the ATLAS and CMS collaborations with the most significant excesses seen in  $ZZ$ ,  $\gamma\gamma$ , and  $WW$ .

A key way to understand the Higgs is through its decay to  $\tau$  leptons. Using this fermionic final state one can determine the strength of the Yukawa coupling as well as the spin of the Higgs. One can also search for heavier resonances as are proposed in the Minimally Super Symmetric Standard Model or search for exotic particles which arise in many Beyond the Standard Model theories.

With a mass of  $m_\tau = 1.777$  GeV, the  $\tau$  lepton is the only lepton heavy enough to decay into hadrons. In about one third of the cases  $\tau$ 's decay leptonically to a muon ( $\tau_\mu$ ) or an electron ( $\tau_e$ ) with two neutrinos, and are reconstructed and identified with the usual techniques for muons and electrons. In the remaining cases,  $\tau$  leptons decay hadronically, to a combination of charged and neutral mesons with a  $\tau$  neutrino. Hadronically decaying  $\tau$ 's, denoted by  $\tau_h$ , are reconstructed and identified with the hadrons-plus-strips (HPS) algorithm, which was developed for use in the LHC Run-1. The challenge a  $\tau$  identification algorithm has to face is the distinction between genuine  $\tau_h$ , and quark and gluon jets, which are copiously produced in QCD multijet process and can be misidentified as  $\tau_h$ . The primary way to reduce the jet  $\rightarrow \tau_h$  misidentification backgrounds is to use the fact that the particles produced in  $\tau_h$  decays are of lower decay product multiplicity, deposit energy in a narrow region compared to a quark or gluon jet, and the primary constituents are typically isolated with respect to other particles in the event. In some final state selections, the misidentification of electrons or muons as  $\tau_h$  candidates may constitute a sizeable background as well.

At CMS, particle reconstruction is performed using the particle-flow (PF) algorithm which combines information from CMS subdetectors to reconstruct and identify particles from each proton-proton collision. The particles are classified as either charged hadrons, neutral hadrons, photons, muons, and electrons. The particle list from each event are then used to reconstruct the missing transverse energy, jets, the  $\tau_h$  candidates and to quantify lepton isolation.

## 2. Tau Identification in Run I

Hadronic decays of  $\tau$  leptons are reconstructed and identified with the HPS algorithm. The basic features of the algorithm are identical to those of Run-1. The HPS algorithm is seeded by the reconstructed jets. The algorithm makes use of the constituents of the jets to reconstruct the neutral pions that are present in most  $\tau_h$  decays. The high probability for photons originating from  $\pi^0 \rightarrow \gamma\gamma$  decays to convert to  $e^+e^-$  pairs is accounted for by collecting the photon and electron ( $p_T > 0.5$  GeV) constituents of the jet into clusters (*strips*). In Run I, the size of the strips was set to a fixed value of  $0.05 \times 0.20$  in the  $\eta - \phi$  direction. For Run II, the size of the strip is allowed to vary dynamically in order to account for nuclear interactions between the tracker material and low

$p_T$  secondary particles and to account for the more narrow, boosted topologies of high  $p_T$  neutral hadrons.

The  $\tau_h$  candidates are formed by combining the strips with the charged-particle constituents of the jet. Based on the observed number of strips and charged particles, it is assigned to be one of the following *decay modes*:

- a single charged particle without any strips: 1 prong 0  $\pi^0$ ;
- combination of one charged particle and one strip: 1 prong 1  $\pi^0$ ;
- combination of a single charged particle with two strips: 1 prong 2  $\pi^0$ ;
- combination of three charged particles: 3 prong.

MVA computation for isolation, anti-electron and anti-muon discriminators are performed using the  $\tau_h$  candidate and PF candidates in a cone of  $\Delta R < 0.5$  around the  $\tau_h$ .

### 3. Improvements to Tau Identification in Run II

After Run-1, additional studies were performed in order to optimize the strip size. In practice, there were cases where  $\tau_h$  decay products contributed to the isolation, such as: A charged pion from  $\tau_h$  decay experiences nuclear interaction with tracker material and produces several secondary particles with low  $p_T$ . This ends up with low  $p_T$  electrons and photons that go outside strip window. This will affect the isolation of the  $\tau_h$ , although it is part of the  $\tau_h$  decay product, or Photons from  $\pi^0 \rightarrow \gamma\gamma$  have a large probability to convert to an  $e^+e^-$  pair and, after multiple conversion and bremsstrahlung, electrons and photons may go outside the fixed size window. This will also affect the isolation.

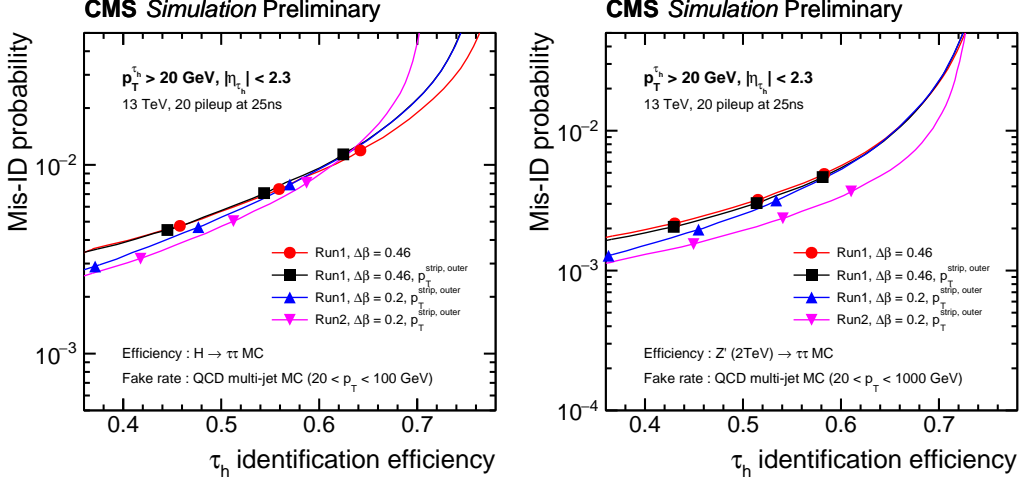
If the  $\tau_h$  has a large  $p_T$  the decay product tend to be boosted in the  $\tau_h$  flight direction. In this case, a smaller strip size than that considered in Run-1 can reduce background contributions in the strip while accounting for all  $\tau_h$  decay products. Therefore, a dynamic strip algorithm which is allowed to change in size based on the total  $p_T$  of its constituents has been developed for Run II.

For discrimination against Jets two primary methods are offered. The first sums the PF candidates in a 0.5  $\Delta R$  cone around the candidates subtracting off the contribution due to pileup and uses this to calculate the relative isolation of the  $\tau_h$  in a cut-based approach. The second method involves training a Boosted Decision Tree with genuine  $\tau_h$ 's as the target variable and jets as the background. The BDT input variables included in Run-2 MVA  $\tau_h$ -ID are the following:

- Shape variables:  $p_T^{strip,outer}$  and  $p_T$ -weighted  $\Delta R$ ,  $\Delta\eta$  and  $\Delta\phi$  (with respect to the  $\tau_h$  axis) of photons and electrons in strips inside or outside of signal cone,
- $\tau$ -lifetime information: the signed impact parameter of the leading track of the  $\tau_h$  candidate, and its significance,
- Multiplicity: the total number of photon and electron candidates ( $p_T > 0.5$  GeV) in signal and isolation cones.

## 4. Tau ID Efficiency

A summary of measurements that were presented [3] in are shown below.



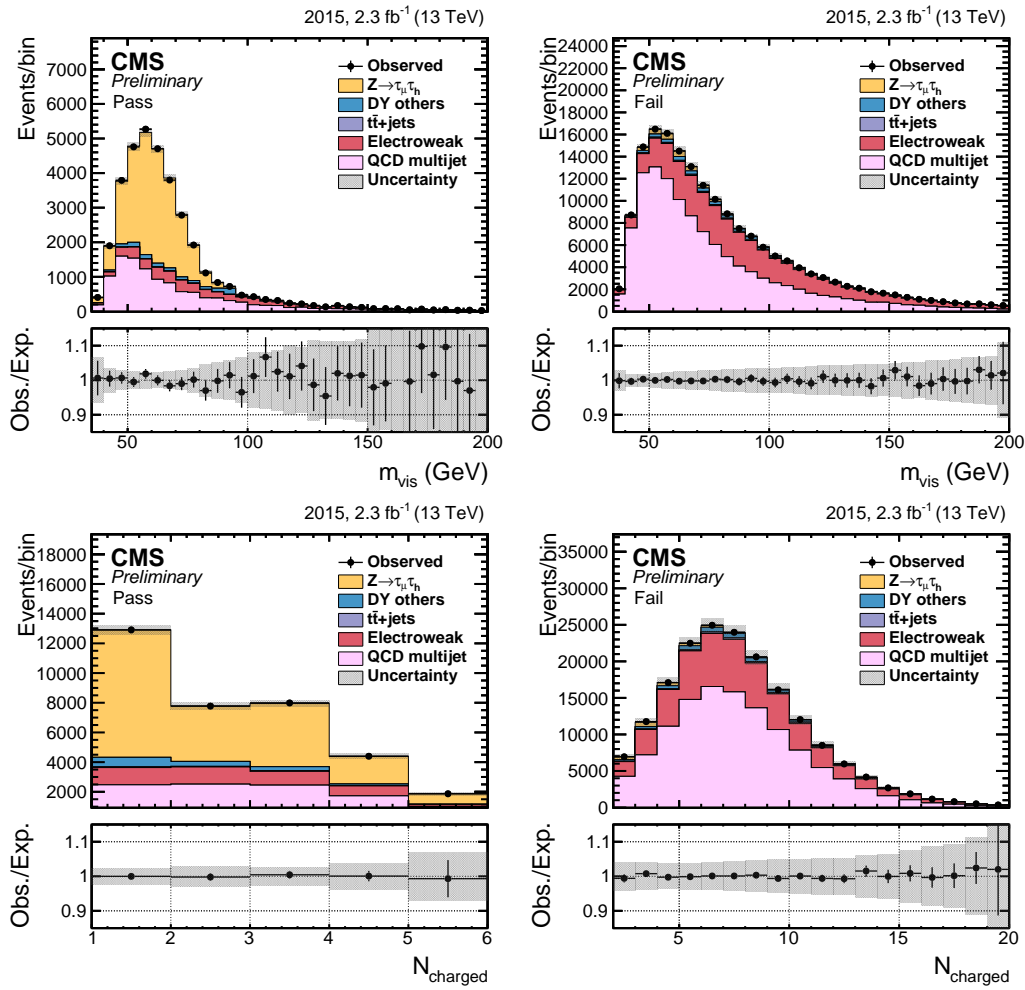
**Figure 1:** Misidentification probability as a function of  $\tau_h$  identification efficiency, evaluated using  $H \rightarrow \tau\tau$  and QCD MC samples (left), and  $Z' (2\text{ TeV}) \rightarrow \tau\tau$  and QCD MC samples (right). Four different configurations of reconstruction plus isolation method are compared (from top to bottom): Run-1 fixed size strip with  $\Delta\beta = 0.46$ , Run-1 fixed size strip with  $\Delta\beta = 0.46$  and  $p_T^{\text{strip,outer}}$  cut, Run-1 fixed size strip with  $\Delta\beta = 0.2$  and  $p_T^{\text{strip,outer}}$  cut, Run-2 dynamic strip with  $\Delta\beta = 0.2$  and  $p_T^{\text{strip,outer}}$  cut. The three points on each curve correspond to, from left to right, the Tight, Medium and Loose working point. The misidentification probability is calculated with respect to jets, which pass minimal  $\tau$  reconstruction requirements.

**Table 1:**  $\tau_h$  identification efficiency scale factor, normalization of the W cross-section,  $r$ , and correlation coefficient between the two quantities obtained from the fit. The scale factors are measured for both isolation-sum and MVA-based discriminators.

| Isolation discriminant         | $\tau$ id SF | $r$ (W norm.)   | correlation     |       |
|--------------------------------|--------------|-----------------|-----------------|-------|
| Isolation-sum $\Delta R = 0.5$ | Loose        | $0.94 \pm 0.21$ | $1.01 \pm 0.05$ | -0.18 |
|                                | Medium       | $0.91 \pm 0.19$ | $1.01 \pm 0.05$ | -0.19 |
|                                | Tight        | $0.83 \pm 0.20$ | $1.01 \pm 0.05$ | -0.17 |
| MVA-based $\Delta R = 0.5$     | Loose        | $0.96 \pm 0.17$ | $1.01 \pm 0.05$ | -0.22 |
|                                | Medium       | $0.95 \pm 0.15$ | $1.01 \pm 0.05$ | -0.25 |
|                                | Tight        | $0.94 \pm 0.15$ | $1.01 \pm 0.05$ | -0.27 |
|                                | Very Tight   | $0.94 \pm 0.14$ | $1.01 \pm 0.05$ | -0.27 |

## 5. Summary

In conclusion, the new Tau Identification has been shown to have excellent performance in Run



**Figure 2:** Postfit distributions in the *pass* (left) and *fail* (right) control regions, using  $m_{\text{vis}}$  (top) or  $N_{\text{charged}}$  (bottom) as observable, for the Loose working point of the MVA-based isolation.

II and has already been used in a number of searches and measurements. The excellent performance of  $\tau_h$  Identification is essential to the physics program at CMS.

## References

- [1] Khachatryan, Vardan and others *Reconstruction and identification of  $\tau$  lepton decays to hadrons and  $\nu_\tau$  at CMS*. CERN-PH-EP-2015-26
- [2] Chatrchyan, Serguei and others *Evidence for the 125 GeV Higgs boson decaying to a pair of  $\tau$  leptons*. CERN-PH-EP-2014-001
- [3] CMS *Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2*. CMS-PAS-TAU-16-002