Identification of hadronic tau decays in CMS.

Raman Khurana[1] on behalf of the CMS collaboration

[1] Saha Institute of Nuclear Physics, Kolkata-700 064, India

DOI: http://dx.doi.org/10.3204/DESY-PROC-2014-04/181

The algorithm used for reconstruction and identification of hadronic tau decays by the CMS experiment at the LHC is presented. The tau reconstruction in CMS takes advantage of the particle-flow algorithm which allows to reconstruct individual hadronic decay modes. The performance of the algorithm in terms of tau identification efficiency and rates for jets to be misidentified as hadronic tau decays is measured in *pp* collision data recorded in 2012 at a center-of-mass energy ($\sqrt{s}$) of 8 TeV, corresponding to an integrated luminosity of 19.7 fb^{-1}.

1 Tau leptons and their decay

Tau is the heaviest known lepton ($m_\tau = 1.78$ GeV/c^2) which decays into lighter leptons (BR $\sim 35\%$) or hadrons τ_h (BR $\sim 65\%$) in the presence of upto two neutrinos. CMS has developed Hadron Plus Strips (HPS) [1, 2] algorithm for reconstruction of hadronic tau decay modes. It uses decay mode identification techniques which allows to reconstruct hadronic tau decays, denoted by τ_h, with high efficiency and suppress the potentially large backgrounds from quarks and gluons that occasionally hadronize into low particle multiplicity jets.

2 Particle Flow and Hadron Plus Strips algorithm

CMS utilises the fine granularity of the calorimeters, precision tracking and muon system to identify and reconstruct final state particles in an event. This is achieved by the use of particle flow (PF) [3] algorithm to make the best use of information available from all the sub-detectors and reconstruct all the stable partiles (namely, charged hadrons, photons, neutral hadrons, muons and electrons) in the detector. The resulting list of particles is used to reconstruct hadronic decays of taus by the HPS algorithm.

The HPS algorithm is designed to reconstruct individual decay modes of the tau. This requires reconstruction of the neutral pions that are produced in the majority of hadronic tau decays. The high probability for photons originating from $\pi^0 \to \gamma\gamma$ decays to convert within the volume of the tracking detector is accounted for by clustering the photon constituents of the jet that seeds the tau reconstruction into $\eta - \phi$ *strips*. The size of the strips, 0.20×0.05, is enlarged in η direction, taking into consideration the bending of e^+e^- pairs produced by photon conversions in the 3.8 Tesla magnetic field. Strips containing one or more photons and passing a cut $p_T > 2.5$ GeV on the sum of photon transverse momenta are kept as π^0 candidates for further processing. The τ_h candidates are built by combining the the neutral objects with charged hadrons reconstructed by the PF algorithm. Figure 1 (left) shows the

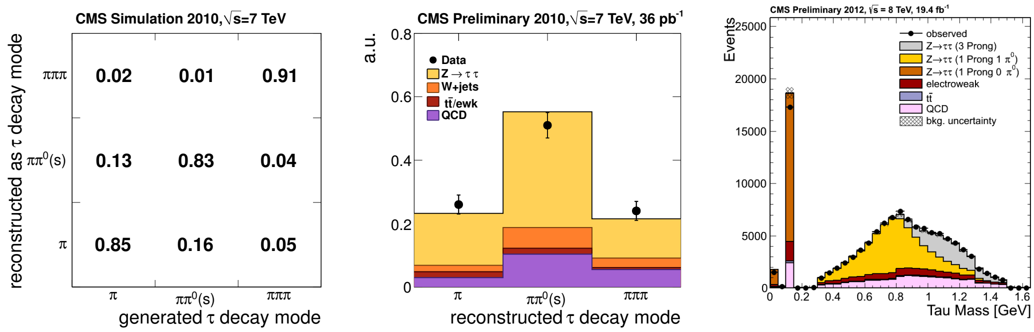

Figure 1: Performance of the decay mode reconstruction.

fraction of generated taus of a given type reconstructed in a certain decay mode for a sample of simulated $Z \to \tau\tau$ events. A data-to-Monte Carlo (MC) comparison in $Z \to \tau\tau \to \mu\tau_h$ events of the number of taus reconstructed in different tau decay modes and of the τ_h candidate mass is shown in Figure 1.

Requiring reconstructed hadronic tau candidates to pass strict isolation requirements constitutes the main handle to reduce the large jet background. Being colorless and produced in decays of colorless bosons, tau leptons are typically isolated with respect to other particles in the event, and so are their decay products, in contrast to quark and gluon jets. The isolation sum is computed using two different approaches :

- Cut based isolation : The isolation of the τ_h candidates is computed by summing the transverse momenta of charged particles of $p_T > 1$ GeV plus photons of $E_T > 1.5$ GeV within a cone of size of $\triangle R = 0.5$ centered on the τ_h direction. The contribution of pile-up to the τ_h isolation is subtracted by means of $\triangle\beta$ corrections.

$$I_\tau = \Sigma p_T^{charged}\,(d_Z < 0.2 \text{ cm}) + \max\,(p_T^\gamma - \triangle\beta, 0) \tag{1}$$

  The $\triangle\beta$ corrections are computed by summing the p_T of charged particles that have a longitudinal impact parameter $d_Z > 0.2$ cm with respect to the τ_h production vertex and are within a cone of size $\triangle R = 0.8$ around the τ_h direction. The sum is scaled by a factor 0.4576, chosen to make the τ_h identification efficiency insensitive to pile-up.

$$\triangle\beta = 0.4576 \cdot \Sigma p_T^{charged}\,(d_Z > 0.2 \text{ cm}) \tag{2}$$

- Multivariate (MVA) approach : The MVA-based tau identification discriminator utilizes the transverse impact parameter of *leading* (highest p_T) track of the τ_h candidate. In case of τ_h candidates reconstructed in the three charged hadron decay mode the distance between the tau production and decay vertex, reconstructed by fitting the three tracks to a common vertex, is used as additional handle to remove the jet $\to \tau_h$ background. A boosted decision tree discriminator (BDT) has been trained to discriminate hadronic tau decays (*signal*) from quark and gluon jets (*background*).

Several working points are provided, corresponding to looser or tighter cuts on the isolation p_T-sum I_τ and on the MVA output respectively, and yielding different tau identification efficiency and jet $\to \tau_h$ misidentification rates.

3 Efficiency and misidentification rate

The efficiency to identify hadronic tau decays is measured in Z→ $\tau\tau \rightarrow \tau_\mu\tau_h$ events selected in the 2012 data using the *tag*-and-*probe* method. The hadronic tau identification efficiency ϵ_τ is defined as the ratio of the number of genuine hadronic taus passing the tau identification discriminator under study over the total number of genuine hadronic taus:

$$\epsilon_{tau} = \frac{N^{\tau}_{pass}}{N^{\tau}_{pass} + N^{\tau}_{fail}} \tag{3}$$

The number of hadronic taus passing and failing the tau identification discriminator under study, N^{τ}_{pass} and N^{τ}_{fail}, is obtained via a simultaneous template fit. The multiplicity of tracks within a cone size $\triangle R = 0.5$ around the τ_h candidate, N_{tracks}, is used as observable to perform the fit. The efficiency measured in data and simulation is shown in Fig. 2 along with the data-to-simulation scale factors for both cut based and MVA-based tau identification discriminators. The probability for quark and gluon jets to be misidentified as hadronic tau decays is measured in QCD multijet events. The results are shown in Fig 3.

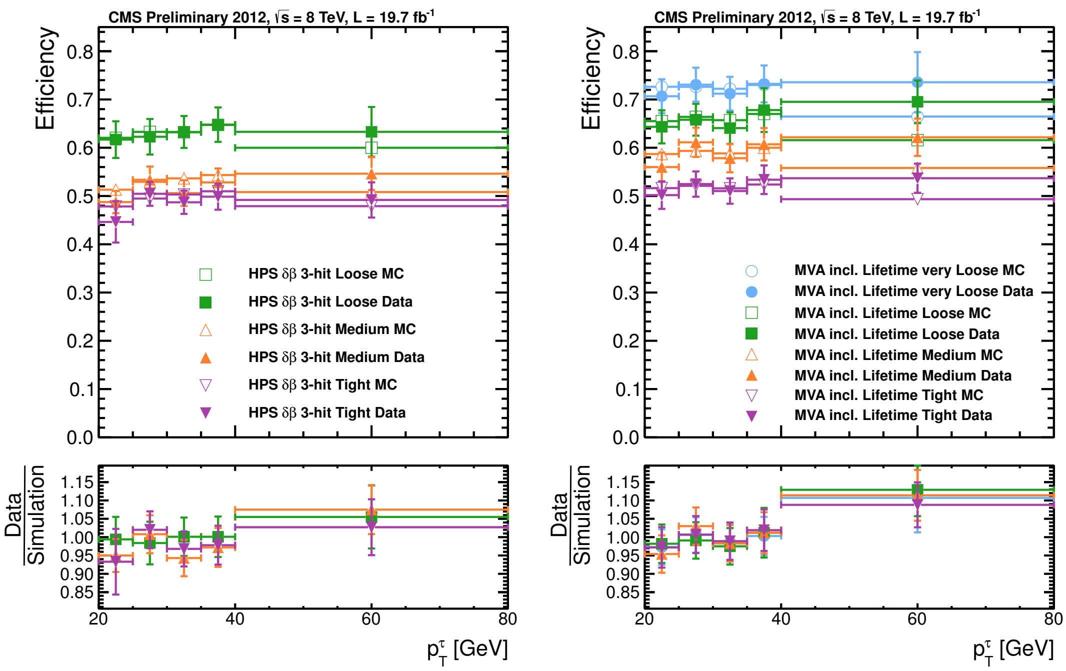

Figure 2: Tau identification efficiency measured as function of p_T in $Z \rightarrow \tau\tau \rightarrow \tau_\mu\tau_h$ events compared to Monte Carlo predictions, for the cut based HPS combined isolation (left) and for the MVA-based tau identification discriminator that includes tau lifetime information (right).

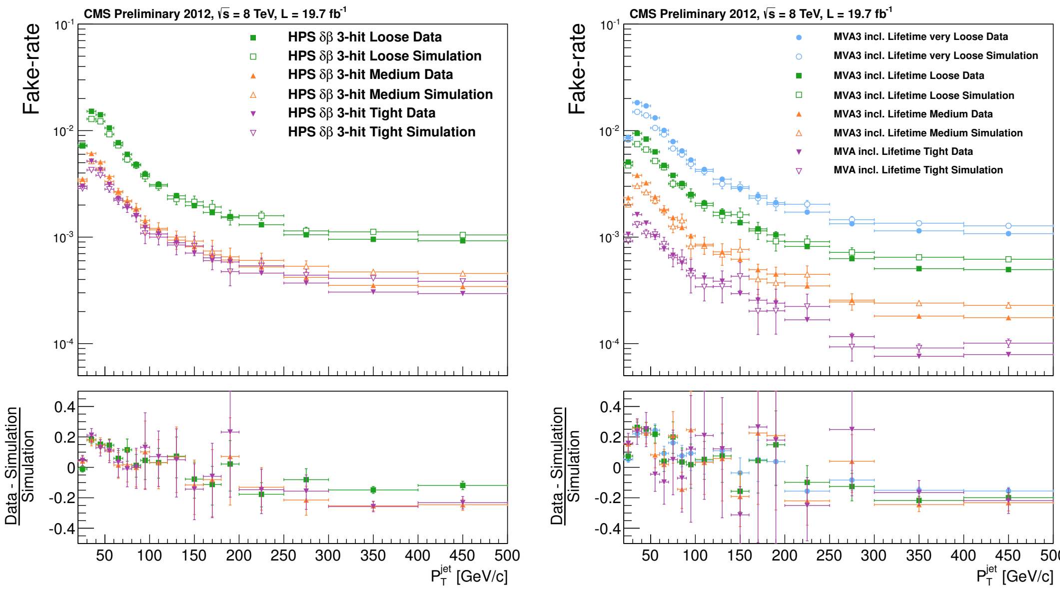

Figure 3: Jet $\rightarrow \tau_h$ misidentification rate in QCD multijet events, for cut based (left) and MVA-based (right) tau identification discriminators.

4 Conclusions

CMS has developed an advanced and robust τ_h reconstruction algorithm, which has been successfully commissioned with the data collected in LHC run 1. The decay mode of the tau is reconstructed correctly with a probability higher than 80%. The tau identification efficiency amounts to 50–60% and is almost independent as function of p_T, while the jet $\rightarrow \tau_h$ misidentification rate varies strongly with p_T, ranging from about 1% for low p_T jets and loose tau identification criteria to $\mathcal{O}(10^{-4})$ for high p_T jets and tight tau identification criteria.

References

[1] CMS Collaboration, *Performance of tau-lepton reconstruction and identification in CMS*, *JINST* **7** (2012) P01001, [`ins-det/1109.6034`].

[2] CMS Collaboration, Tau ID Performance, CMS-DP-2014-015, https://cds.cern.ch/record/1704439

[3] CMS Collaboration, Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}, CMS-PAS PFT-09-001 (2009).