EPJ Web of Conferences 95, 04074 (2015)
DOI: 10.1051/epjconf/20159504074

Identification of hadronic tau decays in CMS

Rosamaria Venditti[1] on behalf of the CMS Collaboration

[1] *Dipartimento di Fisica M. Merlin, Bari, via Amendola 173*

Abstract. 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 of 8 TeV, corresponding to an integrated luminosity of 19.7 fb^{-1}.

1 Introduction

Tau leptons constitute an important experimental signature for many physics analyses at the LHC [1–5].

Tau lepton decays in lighter leptons with a 18% branching ratio (BR). They decay hadronically with a 65% BR, typically into either one or three charged mesons (predominantly $\pi^{\pm}$) plus up to two neutral pions. The latter decay instantaneously via $\pi^0 \to \gamma\gamma$. As the electrons and muons originating from tau decays are difficult to distinguish from those produced in the primary *pp* interaction, the algorithms for tau reconstruction and identification aim at hadronic tau decays (τ_h).

The main challenge in identifying hadronic tau decays, denoted by τ_h, is discriminating from quark and gluon jets. The cross-section for QCD jet production exceeds by many orders of magnitude the rate with which tau leptons are produced at the LHC. To distinguish genuine tau leptons from quark and gluon jets, the algorithm exploit the main futures of hadronic taus decays: low multiplicity, high collimation and the high isolation with respect to other particles in the event. Moreover taus coming from boson decay are often highly boosted, so they have a significant lifetime. In some exceptional and less probable cases, also electrons and muon candidates can mimic an hadronic tau. Dedicated algorithms have been developed to reduce this type of background.

2 CMS Experiment

The central feature of the CMS apparatus is a superconducting solenoid providing a magnetic field of 3.8 T. Within its volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass/scintillator hadron calorimeter. Muons are detected in gas-ionization chambers embedded in the steel flux return yoke outside the solenoid [6]. The CMS experiment uses a right-handed coordinate system, with the origin at the nominal interaction point, the x axis pointing to the centre of the LHC, the y axis pointing up (perpendicular to the LHC plane), and the z axis along the anticlockwise-beam direction. The polar angle θ is measured from the positive z axis and the azimuthal angle ϕ is measured in the transverse (x, y) plane. The pseudorapidity is defined as $\eta = -\log\tan(\theta/2)$. A particle-flow algorithm [7, 8] is used to combine information from all CMS subdetectors to identify and reconstruct individual particles in the event, namely muons, electrons, photons, charged hadrons, and neutral hadrons. The resulting particles are used to reconstruct jets, hadronically decaying tau leptons, and the missing transverse energy vector $\vec{E}_T^{miss}$, defined as the negative of the vector sum of the transverse momenta of all reconstructed particles, and its magnitude E_T^{miss}.

3 Tau Reconstruction in CMS

Hadronic decays of tau leptons are reconstructed and identified by the “Hadrons plus Strips” (HPS) algorithm, designed to reconstruct individual decay modes of the tau [9]. The HPS algorithm is seeded by jets reconstructed using the anti-kT algorithm with a distance parameter R=0.5 [10]. Reconstruction of the tau decay mode requires reconstruction of the neutral pions. The pions decay into pairs of photons which have a high probability to convert into electron-positron pairs when traversing the tracking detector. The bending of the e^+e^- pairs by the magnetic field is accounted for by clustering the photon constituents of the jet that seeds the tau reconstruction into “strips" that are narrow in η and wide in ϕ direction. Strips containing one or more photons are kept as π^0 candidates. τ_h candidates are built by combining the strips with the charged particle constituents of the jet . The combinations considered by the algorithm are:

- *One Prong*: one charged hadron reconstructed by the PF algorithm without any strips.
- *One Prong Plus One* π^0: combination of one charged hadron plus one strip, whose mass fulfills the requirement $0.3 < M < 1.3 \cdot \sqrt{p_T/100}$ GeV.
- *Three Prongs* : three charged hadrons originating from the same vertex whose invariant mass fulfills the requirement 0.8 < M < 1.5 GeV.

All tau decay products are required to be within a narrow cone. Finally the four-momentum of the tau is computed as the sum of charged particles plus strips included in the respective combination. Figure 1 shows the distribution of τ_h candidate mass in $Z \to \tau\tau \to \mu\tau_h$ events. The $Z \to \tau\tau \to \mu\tau_h$ contribution is split according to the tau decay mode reconstructed by the HPS algorithm. The good agreement between data and simulation demonstrates the performance of the HPS algorithm in the reconstruction of the different hadronic tau decay modes.

4 Tau Isolation

The main challenge in identifying τ_h is the discrimination from quark and gluon jets whose cross-section exceeds by many orders of magnitude the rate with which tau leptons are produced at the LHC. The HPS algorithm exploits the main features of τ_h compared to quark and gluon jets: low multiplicity, long lifetime, high collimation and the isolation with respect to other particles in the event. A cut based and a multivariate (MVA) approach have been developed.

Cut-Based Isolation HPS algorithm computes τ_h candidate isolation as the sum of the transverse momenta of the charged particles with $p_T > 1.0$ GeV plus photons of $E_T > 1.5$ GeV, within a cone of size $\Delta R = 0.5$ centered on the τ_h . To ensure that charged particles contributing to the isolation come from hard scatter, their longitudinal impact parameter is required to not exceed 0.2 cm. The charged

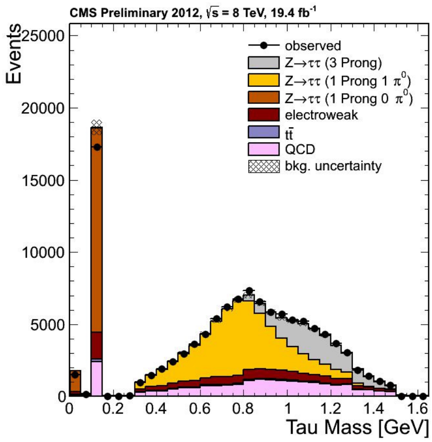

Figure 1. Observed and expected distribution of τ_h candidate visible mass in $Z \to \tau\tau \to \mu\tau_h$ events after the full selection applied. *One prong* τ_h contribution is shown in orange, *One Prong plus One* π^0 is in yellow and *Three prongs* τ_h contribution is in grey.

hadrons used to built the τ_h as well as photons used to built any of the strips are excluded from the isolation p_T sum and the final formula is given by:

$$I_\tau = \Sigma P_T^{charged}(dZ < 0.2 \text{ cm}) + max(\Sigma E_T^{\gamma} - \Delta\beta, 0). \tag{1}$$

The $\Delta\beta$ term is a correction introduced to mitigate the contribution of pile–up interaction. It consist in computing the sum of the transverse momenta of charged particles associated to pile-up interactions and scaling the sum by a factor of 0.4576, that has been chosen to make the identification efficiency independent of pile-up.

Two sets of cut-based tau isolation discriminators are computed, by requiring that at least 8 (HPS combined isolation 8-hit) and 3 (HPS combined isolation 3-hit) hits in the pixel plus silicon strip tracking detectors are associated to each track. Loose, Medium and Tight working-points are defined by requiring the isolation p_T-sum not to exceed thresholds of 2.0, 1.0 and 0.8 GeV respectively.

MVA Isolation A multivariate discriminator has been trained that combines isolation variables and τ_h lifetime information to enhance the discriminating power to separate genuine τ_h candidates from quark and gluon jets. For τ_h candidates reconstructed in the decay modes *One Prong* or *One Prong Plus One* π^0 the transverse impact parameter with respect to the primary event vertex and its significance are used as input of the MVA discriminator. For τ_h candidates reconstructed in the *Three Prongs* decay mode the distance between the primary vertex and the tau decay vertex together with its significance are used in the final discriminator computation. Kinematic information such as p_T, η and decay mode of the τ_h candidate are added as input variables as well as the isolation variables ($\Sigma P_T^{charged}(dZ <$ 0.2 cm) , ΣP^{γ}, $\Delta\beta$). The purpose of the $p_T$and η variables is to parametrize possible dependencies of the other input variables on transverse momentum and pseudo-rapidity. Events used for MVA training are reweighted such that the two-dimensional $p_T$and η distribution of the τ_h candidates for signal and background are identical, in order for the MVA to remain independent of event kinematics. A boosted decision tree (BDT) is trained using simulated events of $H \to \tau\tau$, $Z' \to \tau\tau$, $W \to \tau\nu$, $Z \to \tau\tau$ as signal. W+jets and QCD multi-jet events are used for the background category. An alternative MVA-based discriminator has been trained to assess the gain in tau identification performance that arises from the usage of tau lifetime information

The performance of the various isolation discriminators described so far (3-hits, 8 hits, mva with and without lifetime information) are compared in figure 2, in terms of efficiency versus fake rate. The

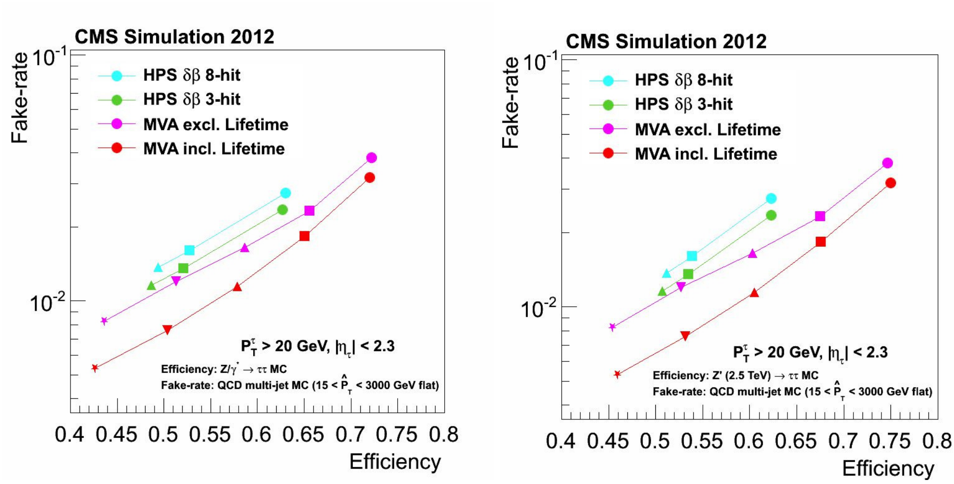

Figure 2. τ_h identification efficiency versus fake-rate for HPS combined 3-hits (green) and 8-hits (blue) isolation and MVA with (red) and without (pink) tau lifetime information. For a given discriminator, the marker correspond to the different working points.

efficiency is computed separately in simulated $Z \to \tau\tau$ (left) and $Z' \to \tau\tau$ (right). The fake rate is measured in simulated sample of QCD events with a flat $\hat{p}_T$ distribution. The inclusion of tau lifetime information in the MVA discriminator reduces the jet $\to \tau_h$ fake-rate by about a factor two, for the same tau identification efficiency, compared to the cut based tau isolation discriminator.

5 Discriminators against electrons and muons

Electrons and muons have a high probability to be reconstructed in the *One Prong* decay mode or, in case of electrons that radiate a Bremsstrahlung photon which subsequently converts, in the decay mode *One Prong Plus One* π^0. Dedicated discriminators have been developed to distinguish light leptons from genuine τ_h.

An MVA-based discriminator is used to separate hadronic taus from electrons. The discriminator is built using as inputs the shape of energy deposits in the ECAL, the sum of energy deposits in the HCAL plus the matching between track momenta and the energy deposits in the ECAL.

The discriminator against muons is based on vetoing tau candidates in case signals in the muon system are found near the τ_h direction or in case the calorimetric energy associated to the leading track of the τ_h candidate is much less then its track momentum.

The performance of these discriminators have been measured in data using the Tag & Probe technique [11] in $Z \to ee$ and $Z \to \mu\mu$ events and compared to the simulation.

The MVA–based discriminator against electrons reduces the $e \to \tau_h$ fake–rate to of 0.5% (0.1%) for an efficiency of 90% (70%). The measured $\mu \to \tau_h$ fake–rates are on the level of one permille for an efficiency of 98%.

6 Tau Identification Performance

The τ_h identification performance has been measured in terms of efficiency and fake-rate, using the data collected in 2012 at $\sqrt{s}$=8 TeV corresponding to an integrated luminosity of L=19.7 fb^{-1}. The τ_h

identification efficiency has been measured with the *Tag&Probe* technique in selected $Z \to \tau\tau \to \mu\tau_h$ events. The identification efficiency is obtained by measuring the number of $Z \to \tau\tau$ events passing and failing the tau identification discriminator under study. The yield of the $Z \to \tau\tau$ signal and the contribution of backgrounds is determined by a fit of the distribution of muon plus tau visible mass. The results are shown graphically in Figs. 3. Data and Monte Carlo simulation agree within

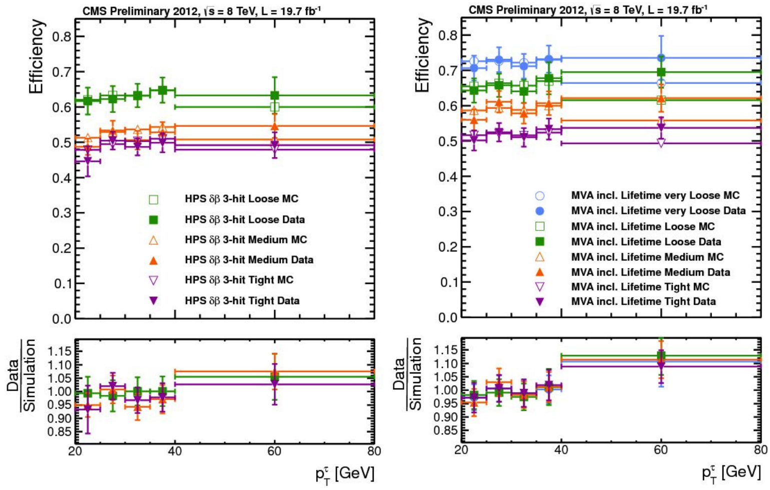

Figure 3. Tau identification efficiency in data and in Monte Carlo simulations in bins of hadronic tau p_T for the HPS cut-based isolation with $\delta\beta$ correction (left), and MVA based tau identification discriminators that include tau lifetime information (right).

uncertainties, typically amounting to 5%.

The rate with which quark or gluon jet are misidentified as hadronic taus (fake-rate) are measured in QCD multi–jet events. Events are selected by requiring at least two jets of $p_T > 20$ GeV and $|\eta| < 2.5$. The jet $\to \tau_h$ fake-rates measured as function of jet p_T are shown in Fig. 3. Measured fake–rates are compared to the Monte Carlo simulation. The probability for quark and gluon jets in QCD multi–jet events to pass the tau identification discriminators typically decreases as function of jet p_T. The maximum jet $\to \tau_h$ fake–rate amounts to 2% for jets of $P_T \approx 50$ GeV that pass the Loose working point of the cut based isolation. A minimal fake–rate on the level of 10^{-4} is achieved for jets of $p_T \approx 500$ GeV by the MVA based tau identification discriminator that includes tau lifetime information.

7 Conclusion

The algorithm devoted to τ_h reconstruction identification in CMS, HPS, has been described showing the strategies used for decay mode reconstruction and discrimination with respect to jet, muons and electrons. The algorithm performs well over a large energy range and is found to be robust against pile-up.

The algorithm has been commissioned using data collected by CMS experiment during LHC run 1. The tau identification efficiency measured in data is in agreement with the Monte Carlo simulation within the uncertainty of the measurement, amounting to 5%. The measured jet $\to \tau_h$ fake-rates are about 20% higher (20% lower) compared to the Monte Carlo prediction for low (high) p_T jets. The

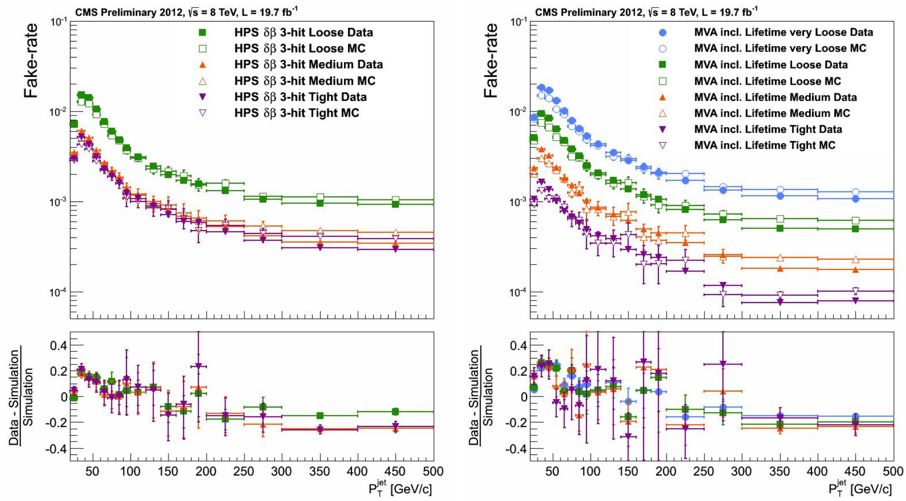

Figure 4. Probabilities for quark and gluon jets in QCD multi–jet events to pass the cut-based discriminator (left) and the MVA based tau identification discriminator that includes tau lifetime information (right), as function of jet p_T.

HPS algorithm achieves a tau identification efficiency of typically 50-60%, for a jet $\rightarrow \tau_h$, $e \rightarrow \tau_h$, $\mu \rightarrow \tau_h$ fake-rates that vary between the per mille and percent level.

References

[1] CMS collaboration, *Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons* (JHEP 05 2014, CMS-HIG-13-004)

[2] CMS collaboration, *Search for MSSM Neutral Higgs Bosons Decaying to Tau Pairs in pp Collisions* (CMS-HIG-13-021)

[3] CMS Collaboration, *Search for pair production of third generation leptoquarks and stops that decay to a tau and a b quark*, CMS PAS EXO-12-002 (2012).

[4] CMS Collaboration, *Search for high mass resonances decaying into τ lepton pairs at $\sqrt{=}7$ TeV*, Phys.Lett. B716 (2012) 82-102.

[5] CMS Collaboration, A search for a doubly-charged Higgs boson in pp collision at $\sqrt{=}7$ TeV, Eur.Phys.J. C72 (2012) 2189

[6] CMS collaboration, *The CMS experiment at the CERN LHC*, (JINST S08004, 2008), pp 3.

[7] CMS collaboration, *Particle-flow event reconstruction in CMS and performance for jets, taus and MET*, (CMS-PAS-PFT-09-001, 2009).

[8] CMS collaboration, *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*, (CMS-PAS-PFT-10-001, 2010).

[9] CMS collaboration, *Performance of τ-lepton reconstruction and identification in CMS*, 2012 JINST 7 P01001

[10] Cacciari, M. and Salam, G. P. and Soyez, G. ,*The anti-kt jet clustering algorithm*, 2008, JHEP 04 063

[11] CMS collaboration, *Measurements of Inclusive W and Z Cross Sections in pp Collisions at* $\sqrt{s}$ *= 7 TeV*, CMS PAS EWK-10-002 (2010).