

Reconstruction and identification of tau leptons in CMS

Francesco Romeo^{1,a} on behalf of the CMS Collaboration

¹IHEP of Beijing, 19B Yuquan Lu, Shijingshan District, Beijing, 100049, China

Abstract. Tau leptons constitute an important experimental signature for analyses at the CERN LHC related to Higgs boson, Standard Model, and beyond the Standard Model measurements. We describe the algorithm used by the CMS experiment to reconstruct and identify decays of tau leptons into hadrons and a neutrino during Run 1 of the CERN LHC. The performance of the algorithm is studied in proton-proton collisions recorded at a center-of-mass energy of 8 TeV, corresponding to an integrated luminosity of 19.7 fb⁻¹. The algorithm achieves an identification efficiency of typically 50-60%, with misidentification rates for quark and gluon jets, electrons and muons that vary between per mille and percent levels.

1 Introduction

Tau leptons constitute an important experimental signature for analyses at the CERN LHC related to Higgs boson, Standard Model, and beyond the Standard Model measurements [1–6]. These analyses benefit from the high reconstruction and identification performances of the CMS tau algorithm [7], whose the description is the main subject of this proceeding.

The tau lepton has a mass of 1.777 GeV and a lifetime of 290.6 fs ($c\tau \approx 87\mu\text{m}$) [8]. The branching fractions for the main tau decay modes are given in Table 1.

Table 1. Main decay channels of the tau lepton. The symbol h denotes a charged meson (pion or kaon).

Decay channel	Resonance	BR (%)
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
$\tau^- \rightarrow h^- \nu_\tau$		11.5
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Others		3.2
All modes containing hadrons		64.8

It decays into hadrons in about two-third of the cases, typically into either one or three charged mesons (π or, less frequently, K) and up to two neutral pions. The π^0 decays instantaneously in $\gamma\gamma$. The remaining one-third of the case the τ lepton decays into an electron or muon and two neutrinos. The electrons and muons originating from tau decays are reconstructed and identified using standard

^ae-mail: fromeo@cern.ch

CMS algorithms [9, 10], while the algorithm for the tau reconstruction and identification presented in this paper focuses on "hadronic" tau decays, denoted by τ_h .

The main challenge in identifying τ_h decays is related to quark and gluon jets that can mimic τ_h decays and have a cross section for QCD jet production that exceeds by many orders of magnitude the rate at which tau leptons are produced at the LHC. To mitigate the misidentification of quark and gluon jets as τ_h , we exploit some of the tau properties: small track multiplicity, collimation and isolation of the decay products, non-negligible lifetime. The misidentification of electrons or muons as τ_h represents another possible source of background and specific identification criteria have been developed in order to reduce this possibility.

2 The CMS detector

The Compact Muon Solenoid (CMS) [11] is a general-purpose detector which operates at LHC. The natural coordinate frame used to describe the CMS detector geometry is a right-handed cartesian system with its origin at the nominal interaction point, the x -axis pointing radially inward-towards 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, η , is defined as $-\ln(\tan\frac{\theta}{2})$, the radius $r = \sqrt{x^2 + y^2}$. The CMS subdetectors, radially installed inside-out, are: tracker, electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL), muon systems. A detailed description of the CMS detector can be found in [11].

3 The hadronic tau reconstruction and identification

In this section we describe the hadrons-plus-strips (HPS) algorithm used by CMS to tag τ_h . The algorithm follows two main stages: the reconstruction step, in which charged and neutral particles are combined to look for specific τ_h decays; and the identification step, in which some discriminators are defined to separate τ_h decays from quark and gluon jets, electrons, and muons.

3.1 The hadronic tau reconstruction

CMS uses a particle-flow (PF) algorithm [12, 13] to provide a global event description. The information available in all CMS subdetectors is employed to reconstruct and identify individual particles: muons, electrons, photons, charged and neutral hadrons. These particles are then used to reconstruct higher-level objects like isolation variables for leptons or the vector imbalance in transverse momentum in the event (whose module will be referred to as E_T^{miss}). Particles reconstructed by the PF algorithm are also used as input to reconstruct jets, using the anti- k_r algorithm [14] with a distance parameter $R = 0.5$. Jets are the seed for the HPS algorithm.

In order to reconstruct the tau decay modes, the HPS algorithm considers charged particles and neutral pions. The charged particles are the constituents of the seeding jet which have a minimum transverse momentum, p_T , of 0.5 GeV. The neutral pions, given the high probability for photons from $\pi^0 \rightarrow \gamma\gamma$ to convert to e^+e^- pairs, are accounted for by clustering the photon and electron constituents of $p_T > 0.5$ GeV of the jet that seeds the tau reconstruction into $\eta - \phi$ "strips". The strip size is 0.05×0.20 , enlarged in ϕ -direction to account for the bending of e^+e^- pairs produced by photon conversions in the 3.8 T magnetic field of CMS. Only strips with $p_T > 2.5$ are considered.

For each jet, charged hadrons and strips, from which the name of the algorithm hadrons-plus-strips originates, are combined to reproduce the tau decays of Table 1, except for the decay mode $h^- h^+ h^- \pi^0$

that has a branching fraction of 4.8% and suffers huge contamination by jets. Mass window cuts are applied to test the compatibility of each hypothesis with the signatures expected for the different hadronic tau decay modes given in Table 1:

- $h^\pm h^\mp h^\pm$: Combination of three charged particles that sum to unit charge and have a mass $0.8 < m_\tau < 1.5$ GeV.
- $h^\pm \pi^0 \pi^0$: Combination of a single charged particle with two strips with $0.4 < m_\tau < 1.2 \cdot \sqrt{p_T[\text{GeV}]/100}$ GeV. The size of the mass window is enlarged for τ_h candidates of high p_T to account for resolution effects. For hadronic tau candidates of $p_T < 100$ GeV (> 1111 GeV) the upper limit of the mass window is set to 1.2 GeV (4.0 GeV).
- $h^\pm \pi^0$: Combination of one charged particle plus one strip with $0.3 < m_{\tau_h} < 1.3 \cdot \sqrt{p_T[\text{GeV}]/100}$ GeV. For τ_h candidates of $p_T < 100$ GeV (> 1044 GeV) the upper limit of the mass window is set to 1.3 GeV (4.2 GeV).
- h^\pm : A single charged particle with no strips.

The four-momentum of each τ_h candidate hypothesis (p_T , η , ϕ , and mass) is given by the momentum sum of charged particles plus strips that are included in the respective decay mode combination. The two decay modes $h^\pm \pi^0$ and $h^\pm \pi^0 \pi^0$ are analysed together and referred to as $h^\pm \pi^0 s$. The distribution of tau decay modes reconstructed in $Z/\gamma^* \rightarrow \tau\tau$ events and of the reconstructed mass of the τ_h candidates in those events are shown in Fig. 1.

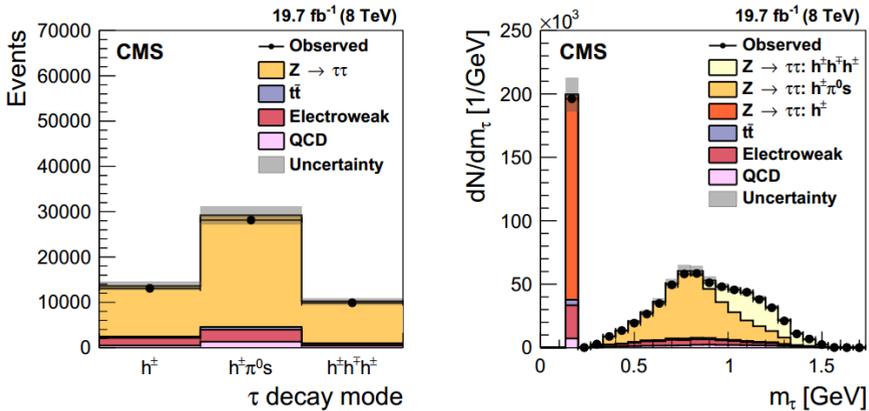


Figure 1. Distributions of the reconstructed tau decay mode (left) and of τ_h candidate mass (right) in $Z/\gamma^* \rightarrow \tau\tau$ events selected in data compared to the Monte-Carlo expectation.

3.2 The hadronic tau identification

Reconstructed τ_h candidates are required to satisfy some identification criteria in order to mitigate the background related to quark and gluon jets or due to electrons and muons.

The HPS algorithm has two isolation discriminators: cut-based and MVA-based (where MVA stands for Multivariate Analysis).

Cut-based: The isolation of τ_h candidates is computed by summing the transverse momenta of charged particles of $p_T > 0.5$ GeV plus photons of $E_T > 0.5$ GeV, reconstructed by the PF algorithm, within an isolation cone of size $\Delta R = 0.5$, centered on the τ_h direction. The effect of pileup is reduced by requiring the tracks associated to charged hadrons considered in the isolation p_T sum to be

compatible with the production vertex of the τ_h candidate within a distance of 0.2 cm in direction of the beamline and 0.03 cm in the transverse plane. Charged hadrons used to build the τ_h candidate are excluded from the isolation p_T -sum, as are photons used to build any of the strips. The effect of pileup on the photon isolation is compensated on statistical basis, by applying so-called $\Delta\beta$ corrections:

$$I_\tau = \sum p_T^{\text{charged}}(d_Z < 0.2\text{cm}) + \max(p_T^\gamma - \Delta\beta, 0), \quad \Delta\beta = 0.46 \cdot \sum p_T^{\text{charged}}(d_Z > 0.2\text{cm}) \quad (1)$$

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 defined by Eq. 1 not to exceed thresholds of 2.0, 1.0 and 0.8 GeV respectively.

MVA-based: A boosted decision tree discriminator (BDT) [15] is used to discriminate hadronic tau decays (“signal”) from quark and gluon jets (“background”). The variables used as inputs to the BDT are:

- The charged particle isolation p_T -sum $\sum p_T^{\text{charged}}(d_Z < 0.2\text{cm})$ and the neutral particle isolation p_T -sum $\sum p_T^\gamma$ defined in Eq. 1 as separate inputs.
- The reconstructed tau decay mode.
- The transverse impact parameter, d_0 , of the highest p_T track of the τ_h candidate and its significance, d_0/σ_{d_0} .
- The τ_h flight distance (for the $h^\pm h^\mp h^\pm$ decay mode)

The position of the primary event vertex is refitted after excluding the tracks associated to the τ_h candidate. The inputs are augmented by the p_T and η of the hadronic tau candidate and by the $\Delta\beta$ correction defined by Eq. 1. The purpose of the p_T and η variables is to parametrise possible dependencies of the other input variables on transverse momentum and pseudo-rapidity. The $\Delta\beta$ correction parametrises the dependency on pileup for the neutral particle isolation p_T -sum.

The BDT is trained on event samples produced using the Monte-Carlo (MC) simulation. Samples of $Z/\gamma^* \rightarrow \tau\tau$, Higgs $\rightarrow \tau\tau$, $Z' \rightarrow \tau\tau$, and $W' \rightarrow \tau\nu_\tau$ events, covering the range 20-2000 GeV in τ_h candidate p_T , are used for the “signal” category. Reconstructed τ_h candidates are required to match hadronic tau decays within $\Delta R < 0.3$ on generator level. W +jets and QCD multi-jet events are used for the “background” category. The τ_h candidates that match leptons originating from the W decays are excluded from the training. Half of the available event statistics is used for the training, the other half for evaluating the MVA performance and conducting overtraining checks.

Different working-points, corresponding to different τ_h identification efficiencies and jet $\rightarrow \tau_h$ misidentification rates, are defined by varying the cut on the MVA output.

The performance of the various isolation discriminators described so far (3-hits, 8 hits, MVA with and without lifetime information) are compared in Fig. 2, in terms of efficiency versus fake-rate (probability of misidentification). The efficiency is computed separately in simulated $Z/\gamma^* \rightarrow \tau\tau$ (left) and $Z' \rightarrow \tau\tau$ (right). The fake-rate is measured in simulated samples of multijet events. The inclusion of tau lifetime information in the MVA discriminator reduces the jet $\rightarrow \tau_h$ fake-rate by about a factor two, for the same tau identification efficiency, compared to the cut based tau isolation discriminator.

Electrons and muons may be reconstructed as τ_h in the h decay mode or, in case of electrons that radiate a bremsstrahlung photon which subsequently converts, in the $h + \pi^0$ decay mode. Specific discriminators have been developed to mitigate this contamination.

An MVA-based discriminator is used to separate hadronic taus from electrons. The discriminator is built using the energy deposits in the ECAL and HCAL calorimeters as inputs. Track-related variables as the p_T , η , or the number of hits in the tracker detector are also used.

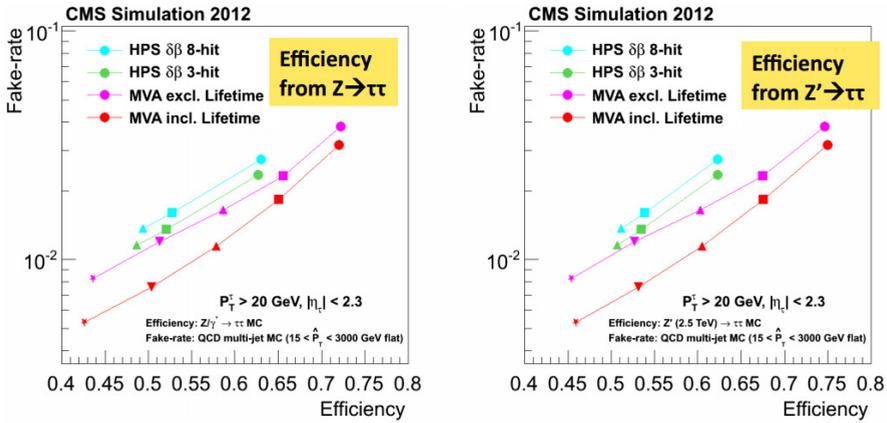


Figure 2. τ_h identification efficiency versus fake-rate for HPS cut-based isolation with 3-hits (green) and 8-hits (blue) and MVA-based isolation with (red) and without (pink) tau lifetime information. For a given discriminator, the markers correspond to the different working points.

The discriminant against muons vetoes τ_h candidates when signals in the muon system are found near the τ_h direction or when the sum of the energies in the ECAL and HCAL corresponds to < 0.2 of the momentum of the leading track of the τ_h candidate.

4 The hadronic tau performance with the LHC Run 1 at $\sqrt{s} = 8$ TeV

The performance of the τ_h algorithm is studied in proton-proton collisions recorded at a center-of-mass energy of 8 TeV, corresponding to an integrated luminosity of 19.7 fb^{-1} . Different samples of events are used to check the τ_h reconstruction and identification in data and to validate the rates with which quark and gluon jets, electrons, and muons get misidentified as τ_h decays.

4.1 Measurement of the hadronic tau identification efficiency

The τ_h identification efficiency measurement is performed on a sample of high-purity $Z/\gamma^* \rightarrow \tau\tau$ in the $\mu\tau_h$ final state. The events are selected by pairing an isolated muon with $p_T > 25$ GeV and the highest p_T jet with at least 20 GeV, $|\eta| < 2.3$ that is separated from the muon by $\Delta R(\sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}) > 0.5$. The muon and the p_T leading track of the jet are required to have opposite charge and to be compatible with originating from the same vertex. The transverse mass of the muon and E_T^{miss} , $m_T(\mu, E_T^{\text{miss}}) = \sqrt{2p_T^\mu E_T^{\text{miss}}(1 - \cos\Delta\phi)}$, lower than 40 GeV and the topological cut $p_\zeta - 1.85p_\zeta^{\text{vis}} > -15$, described in [7], are further demanded. A veto on the presence of b quark jets and extra leptons is asked. No τ_h identification criteria are directly applied during the event selection.

The τ_h identification efficiency is obtained through a simultaneous fit of the number of $Z/\gamma^* \rightarrow \tau\tau$ events, N_{pass}^τ and N_{fail}^τ , with hadronic taus passing (“pass” region) and failing (“fail” region) the τ_h identification discriminant. The τ_h identification efficiency is taken as the parameter of interest μ in the fit. The number of $Z/\gamma^* \rightarrow \tau\tau$ events in the pass and fail regions as a function of μ is given by $N_{\text{pass}}^\tau = \mu N^{Z/\gamma^* \rightarrow \tau\tau}$ and $N_{\text{fail}}^\tau = (1 - \mu)N^{Z/\gamma^* \rightarrow \tau\tau}$ respectively. The multiplicity of tracks within a cone of size $\Delta R < 0.5$ centered on the τ_h direction is used to perform the fit. For more details on the fitting

procedure, the background contaminations, and the systematic uncertainties estimations we turn to the main article [7].

Figure 3 shows the results of the efficiency measurement as a function of p_T for the cut-based (left) and MVA (right) isolation discriminators for different working points, considering both data (solid symbols) and MC simulation (open symbols). The MVA-based isolation performs better, reaching an efficiency of up to 70%. All data-to-MC ratios are compatible with unity within the estimated uncertainties of about 5%.

The efficiency to reconstruct and identify τ_h of higher p_T in more dense hadronic environments is measured using $t\bar{t}$ events. A detailed description of this measurement is reported in [7].

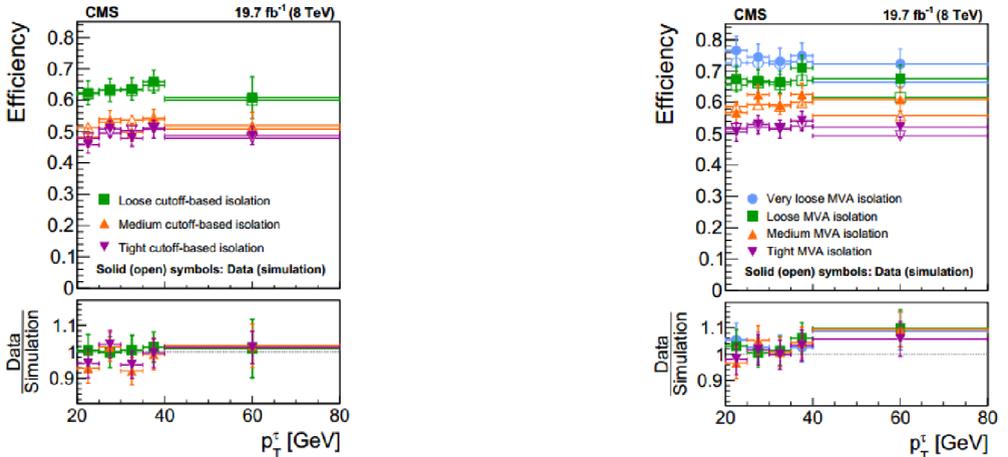


Figure 3. τ_h identification efficiency measured in $Z/\gamma^* \rightarrow \tau\tau \rightarrow \mu\tau_h$ events as function of p_T for the cut-based (left) and MVA-based (right) tau isolation discriminants, for data and MC expectation.

4.2 Measurement of the hadronic tau energy scale

The energy scale for τ_h (referred to as τ ES) is defined as the average reconstructed τ_h energy relative to the generator level energy of the visible τ decay products. MC-to-data τ ES corrections are determined in $Z/\gamma^* \rightarrow \tau\tau$ events selected in the $\mu\tau_h$ channel as described in Sec. 4.1, with the additional requirement that the selected 6jet passes the medium working point of the MVA-based τ_h isolation.

The corrections are obtained by fitting the reconstructed mass of the τ_h candidate, m_{τ_h} , and the mass of muon and τ_h candidates, m_{vis} . The m_{τ_h} and m_{vis} templates are computed by changing the τ_h four-momentum as a function of τ ES, and recomputing m_{τ_h} and m_{vis} after each such change.

The τ ES is measured separately for the reconstructed τ_h decay modes and as a function of p_T . The variable m_{τ_h} is seen to be the more sensitive observable compared to m_{vis} , as indicated by smaller uncertainties. τ ES correction is always found to be lower than 2%. There is no indication of a dependence of the measured τ ES corrections on $\tau_h p_T$.

4.3 Measurement of the rate of jets, electrons, and muons misidentified as hadronic tau

The rate for quark and gluon jets to be misidentified as τ_h decays is measured in W+jets and multijet events. W+jets events are selected requiring the presence of a muon of $p_T > 25$ GeV, $|\eta| < 2.1$, the $m_T(\mu, E_T^{miss}) > 50$ GeV, and at least one jet of $p_T > 20$ GeV, $|\eta| < 2.3$ separated from the muon by $\Delta R > 0.5$. Multijet events are selected demanding the presence of two jets with $p_T > 20$ GeV, $|\eta| < 2.3$. The misidentification rates measured in W+jets and in multijet events are shown in Fig. 4.

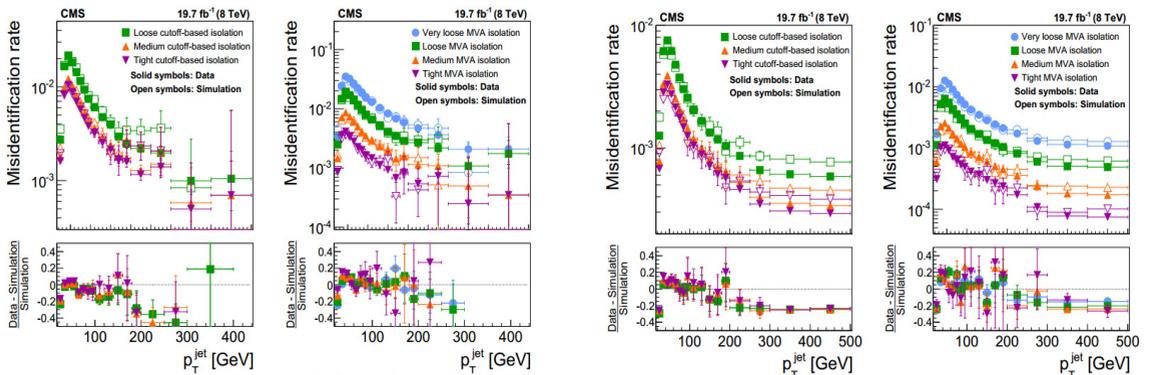


Figure 4. Fraction of times as a function of jet p_T in which quark and gluon jets in W+jets (left) and multijet (right) events pass the cutoff-based and MVA-based τ_h isolation discriminant.

In general, the misidentification rates are higher in W+jets than in multijet events. The difference is due to the higher fraction of quark jets in W+jets events that typically have a lower particle multiplicity and are more collimated than gluon jets. Overall, the jet $\rightarrow \tau_h$ misidentification rates vary between low percentages and about 10^{-4} , decreasing as a function of jet p_T , as the particle multiplicity increases for higher energy jets. MVA-based isolation, which includes τ lifetime information, performs better than cut-based isolation.

Notable differences are observed in the comparison between data and expectations. The rates for jet $\rightarrow \tau_h$ measured in data exceed the MC expectation at low p_T while the rates measured at high p_T fall short of the simulation. The magnitude of the change on data/MC is $\approx 20\%$.

The fraction of times in which electrons or muons pass the τ_h identification criteria of the dedicated discriminants against electrons or muons, are measured through the tag-and-probe technique using $Z/\gamma^* \rightarrow ee$ and $Z/\gamma^* \rightarrow \mu\mu$ events. The measured $e \rightarrow \tau_h$ ($\mu \rightarrow \tau_h$) fake-rate is on the level of per-mille for a τ_h efficiency of 90% (98%). The rates measured in data exceed the MC prediction by up to a factor 1.7 for electrons. The full details of these measurements are reported in [7].

5 First results with hadronic taus with the LHC Run 2 at $\sqrt{s} = 13$ TeV

Preliminary results related to τ_h has been obtained by analysing proton-proton collisions recorded at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 5.6 pb^{-1} . $Z/\gamma^* \rightarrow \tau\tau \rightarrow \ell\tau_h$, $\ell = e, \mu$, have been selected by requiring the presence of an isolated electron (muon) with $p_T > 20$ ($p_T > 18$) GeV, $|\eta| < 2.1$ and an isolated τ_h with $p_T > 20$ GeV, $|\eta| < 2.3$ passing the identification criteria against electrons and muons that have an opposite charge to the other lepton. $m_T(\mu, E_T^{\text{miss}}) < 40$ GeV is also required.

Figure 5 shows the visible mass distribution (left) and fully reconstructed mass using the SVFit [16] algorithm (right). The simulation is normalised to the total integrated luminosity for all contributions except from the QCD background that is estimated by same sign data. A good agreement is observed overall.

6 Conclusions

The algorithm used by the CMS experiment for the reconstruction and identification of hadronic τ decays in Run 1 data from the LHC have been presented, and their performance validated with proton-proton collision data recorded at $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 19.7 fb^{-1} .

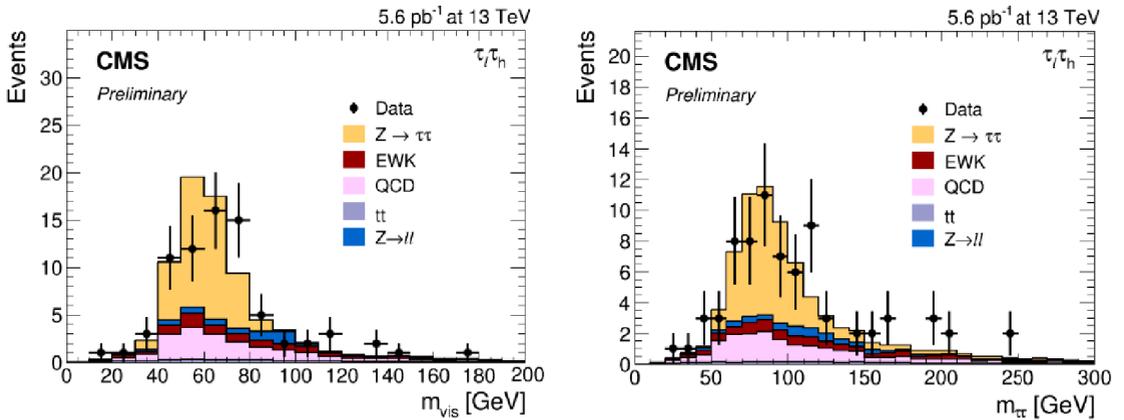


Figure 5. The visible mass distribution (left) and fully reconstructed mass using the SVFit algorithm (right) obtained by analysing proton-proton collisions recorded at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 5.6 pb^{-1} .

The algorithm achieves a τ_h identification efficiency of typically 50-60%, and misidentification rates for quark and gluon jets, electrons, and muons that vary between the per mille and per cent level. The τ_h identification efficiency measured in the data agrees with the MC expectation within the uncertainty of the measurement of about 5%. The measured jet $\rightarrow \tau_h$ misidentification rates are about 20% higher for low- p_T jets and 20% lower for high- p_T jets. The measured misidentification rate for electrons exceeds the MC expectation by up to a factor 1.7.

Preliminary results related to τ_h obtained by analysing $Z/\gamma^* \rightarrow \tau\tau \rightarrow \ell\tau_h$, $\ell = e, \mu$, events from proton-proton collisions recorded at $\sqrt{s} = 13 \text{ TeV}$ and corresponding to an integrated luminosity of 5.6 pb^{-1} show good τ_h performances in view of the LHC Run 2.

References

- [1] CMS Collaboration, JHEP **05**, 104 (2014)
- [2] CMS Collaboration, JHEP **10**, 160 (2014)
- [3] CMS Collaboration, Phys. Lett. **B 739**, 23 (2014)
- [4] CMS Collaboration, JHEP **08**, 117 (2011)
- [5] CMS Collaboration, Phys. Rev. Lett. **110**, 081801 (2013)
- [6] CMS Collaboration, Phys. Lett. **B 716**, 82 (2012)
- [7] CMS Collaboration, Submitted to JINST, arXiv:1510.07488 (2015)
- [8] Particle Data Group, Chin.Phys. **C38**, 090001 (2014)
- [9] CMS Collaboration, JINST **10**, P06005 (2015)
- [10] CMS Collaboration, JINST **7**, P10002 (2012)
- [11] CMS Collaboration, JINST **3**, S08004 (2008)
- [12] CMS Collaboration, CMS-PAS-PFT-10-001, (2009)
- [13] CMS Collaboration, CMS-PAS-PFT-10-002, (2010)
- [14] Cacciari et al., JHEP **04**, 063 (2008)
- [15] Hoecker et al., PoS **ACAT**, 040 (2007)
- [16] L. Bianchini et al., J.Phys.Conf.Ser. **513**, 022035 (2014)