The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

13 December 2011 (v4, 22 December 2011)

Tau Physics at CMS

Vladimir Cherepanov for the CMS Collaboration

Abstract

At the LHC, τ -leptons are frequently produced in the decays of W and Z bosons. The development of an efficient and accurate reconstruction and identification of τ -leptons is therefore an important item in the CMS physics program. However, this is challenging due to neutrino in the final state. We report on the development of algorithms to identify τ -lepton decays and discuss the first CMS measurements of W and Z decays involving τ -leptons using the 7 TeV dataset from the pp-run in 2010.

Presented at *PHPSI11: International Workshop on $e+e^-$ collisions from Phi to Psi*

Tau physics at CMS

Vladimir Cherepanov (on behalf of the CMS Collaboration)

III. Physikalisches Institut B, RWTH Aachen University, D-52056 Aachen, Germany

Abstract

At the LHC, τ -leptons are frequently produced in the decays of W and Z bosons. The development of an efficient and accurate reconstruction and identification of τ -leptons is therefore an important item in the CMS physics program. However, this is challenging due to neutrino in the final state. We report on the development of algorithms to identify τ -lepton decays and discuss the first CMS measurements of W and Z decays involving τ -leptons using the 7 TeV dataset from the pp-run in 2010.

1. Introduction

The τ -lepton is an important item in the CMS physics programme. This is mainly due to the fact that the decay into τ -pairs is the second favorite decay mode of the light Higgs boson. Moreover, a significant amount of τ -leptons are produced in decays of gauge bosons (W and Z). This makes the τ -lepton a very suitable tool for studying many processes at the LHC. For example, testing lepton universality using $W \rightarrow \tau\nu$ decays and measuring the electroweak coupling constants, which can be extracted from charged forward-backward and polarization asymmetries of the τ -pairs, produced from decays of Z bosons. Since the tau is the only lepton heavy enough to decay into hadrons, it can be used to study light-quark interactions, to test the Conserved Vector Current (CVC) hypothesis and to search for phenomena beyond the Standard Model (SM), like lepton flavour violating τ decays. Besides this, τ -leptons can be used in detector related studies, including those related to the reconstruction of the missing transverse energy. The key features of the τ -lepton reconstruction and identification at CMS and the first results, obtained using the developed algorithms, are described in the next sections.

2. Particle Flow Algorithm

The Particle Flow algorithm (PF)[1] is developed to provide a detailed event description based on the signals

coming from all CMS sub-detectors. The reconstruction starts by linking together the fundamental elements, namely charged particle tracks and calorimeter clusters, in the η - ϕ plane and combining them to PF blocks to fully reconstruct and classify each single particle in the event. The algorithm begins with the track being linked to any given cluster, if the extrapolated position in the corresponding calorimeter is within the cluster boundaries. Then, if the calibrated calorimetric energy is compatible with the track momentum within the measurements uncertainties, the block gives rise to a PF charged hadron and the corresponding track and clusters are removed from further processing of the event. All unlinked clusters or clusters with significantly larger energy than associated charged-particle momentum give rise to PF photons or PF neutral hadrons, depending on the link between the signals from the electromagnetic and hadron calorimeters. Finally, the PF provides the list of individually reconstructed and classified particles that are called PF candidates. Electrons and muons have a specific reconstruction and are added to the list of PF candidates, more details about the electrons and muons reconstruction can be found in the PF paper [1].

All composite objects like jets, missing energy or τ -leptons are built from the particle flow candidates.

Jets are reconstructed using the anti-KT iterative-cone algorithm [2] with a cone size of 0.5 in the η - ϕ plane. In order to estimate the performance of the PF-jet

algorithm and the alternative calo-jet algorithm, which is based only on the signals from calorimeters, a sample of QCD multi jet events was used.

Figure 1: The jet-energy resolution as a function of the transverse momentum of calo-jets (open markers) and PF jets (triangles) in the barrel region ($|\eta| < 1.5$).

Figure 1 shows the jet energy resolution as a function of the transverse momentum of the jets for both algorithms. The PF-based reconstruction demonstrates a superior performance in the region of low p_T , where the PF resolution is three times smaller than the resolution provided by the calo-jet algorithm. This is due to including the information from the tracker in the PF approach.

Figure 2: The photon-pair invariant mass distribution in the barrel ($|\eta| < 1.0$) for the data and a fit of the π^0 peak by a Gaussian.

The performance of the PF algorithm in the reconstruction of photons can be seen in Figure 2. It shows the photon-pair invariant mass obtained from a small amount of data. The agreement for the π^0 mass is within $\pm 2\%$ of the PDG value [3], which demonstrates the suitability of the electromagnetic calorimeter calibration.

3. τ reconstruction and identification

The main challenge in τ -lepton reconstruction at the CMS detector is to discriminate hadronically decaying

τ -leptons from QCD jets. The experimental signatures that are used to indicate genuine τ -lepton decays are: (i) a collimated jet, as τ -leptons produced from decays of gauge bosons are highly boosted; (ii) a low-multiplicity jet with either one or three charged mesons, mainly pions, with the possible presence of up to two neutral pions; and (iii) the decay products are expected to be isolated in the detector. Applying all of these requirements provides a powerful separation against jets from QCD events. The best performance in terms of efficiency and fake-rate is achieved by analyzing constituents of PF jets and building particular τ -lepton decay modes. Currently, two main algorithms of τ -lepton reconstruction are employed: the Hadron Plus Strips algorithm (HPS); and the Tau Neural Classifier (TaNC). In this paper only HPS [4] is covered. Special attention in the HPS algorithm is given to the reconstruction of neutral pions that immediately decay into pairs of photons. In many cases, photons convert into e^+e^- pairs in the CMS tracker material before they reach the electromagnetic calorimeter. This results in a broader signal in the ECAL due to the bending of electrons and positrons in the magnetic field. To account for possible conversions, the HPS reconstructs “strips” which are centered on the most energetic electromagnetic particle within a given jet. Then, the algorithm searches for other electromagnetic particles within the region of size 0.05-0.2 in the $\eta - \phi$ plane. In the case that a more energetic particle is found within this region, a new strip is created around this particle. The procedure stops when there are no particles that can be associated to the strip.

Finally, the charged particles in the given jet are combined with reconstructed strips, that satisfy the requirement $p_T^{strip} > 1$ GeV/c, in order to build individual hadronic τ -lepton decay modes. Currently, three main τ -decay modes can be reconstructed: $\pi^\pm \nu_\tau$, $\pi^\pm \pi^0 \nu_\tau$, and $\pi^\pm \pi^\mp \pi^\pm \nu_\tau$. This corresponds to approximately 50% of τ -lepton branching ratio. All the decay products, the charged tracks, and the strips, are required to be contained within a cone of size $2.8/p_T$.

Additionally, the isolation requirements are applied to the reconstructed τ -lepton candidate. This implies counting the number of charged tracks and deposits in the ECAL which are not associated with the reconstructed τ -lepton candidate, with transverse momentum and transverse energy above a certain thresholds within the isolation cone of size $\Delta R = 0.5$. Three working points are defined for different tightness of isolation criteria:

- The loose working point: no charged hadrons with $p_T > 1.0$ GeV/c and no photon with $E_T > 1.5$ GeV

within the isolation cone.

- The medium working point: no charged hadrons with $p_T > 0.8$ GeV/c and no photon with $E_T > 0.8$ GeV within the isolation cone.
- The tight working point: no charged hadrons with $p_T > 0.5$ GeV/c and no photon with $E_T > 0.5$ GeV within the isolation cone.

Figure 3: The expected efficiency of the τ algorithm as a function of generated visible tau p_T .

The expected efficiency for the three working points is estimated using a sample of simulated $Z \rightarrow \tau\tau$ events and is illustrated in Figure 3. For the loose working point the efficiency is estimated to be approximately 50% for a wide range of transverse momentum.

The probability of the non-tau jets to be misidentified as a τ candidate is called fake-rate and is estimated from data. The selection of QCD-type jets requires at least one jet with transverse momentum $p_T > 15$ GeV/c within $|\eta| < 2.5$. To select W-type events at least one isolated muon with transverse momentum $p_T > 15$ and $|\eta| < 2.1$ and a jet with $p_T > 10$ GeV/c and $|\eta| < 2.5$ are required. The muon enriched QCD sample is selected, requiring the event to contain one muon and one jet and, in order to suppress the contribution from W-type events, the total transverse mass is required to be $M_T < 40$ GeV/c².

The fake-rate estimated from different control samples is shown in Figure 4. The probability of the non- τ jets to pass the τ reconstruction and identification is at the level of a few percents, and is in good agreement with those expected from simulation.

Figure 4: The probability of quark and gluon jets to pass the “loose” working points of the HPS algorithm as a function of jet p_T estimated using different MC samples.

4. $Z \rightarrow \tau\tau$ cross-section measurement and τ -ID efficiency estimation

4.1. τ -ID efficiency

The efficiency of the τ reconstruction and identification was estimated from data using a sample of $Z \rightarrow \tau_\mu \tau_{had}$ events, where τ_μ, τ_{had} indicates the τ -leptons decaying to muon and hadrons correspondingly. The pre-selection requires the event to be triggered by a single muon trigger, to contain only one isolated muon with $p_T^\mu > 15$ GeV/c and $|\eta| < 2.1$, to have an isolated τ -jet candidate with $p_T^\tau > 20$ GeV/c and $|\eta| < 2.3$, which is reconstructed using the anti- k_T algorithm in a cone of size 0.5 [5]; and a highest p_T track within the jet with $p_T^\tau > 5$ GeV/c. Additionally, the requirement on the transverse mass, $M_T = \sqrt{2p_T^\mu E_T^{miss}(1 - \cos\Delta\phi)} < 40$ GeV/c, is applied, where p_T^μ is the muon transverse momentum, E_T^{miss} is the PF missing transverse energy, which is calculated as the modulus of the vector sum of the transverse momentum of all reconstructed particles in the event, and $\Delta\phi$ is the azimuthal angle between the missing transverse energy vector, and the muon transverse momentum vector.

After this, the HPS identification algorithm is run on preselected events and the sample is divided into two parts, based on whether the events pass or failed the identification criteria. The visible invariant mass of the $\mu\tau$ system of both types of the events is fitted using the signal and background templates taken from Monte-Carlo simulation. The efficiency is then calculated as: $\epsilon = N_{pass}^{Z \rightarrow \tau\tau} / (N_{pass}^{Z \rightarrow \tau\tau} + N_{fail}^{Z \rightarrow \tau\tau})$, where $N_{pass, fail}^{Z \rightarrow \tau\tau}$ is the

number of events after background subtraction. The resulting visible invariant mass for “pass” and “fail” are shown on Figure 5 and Figure 6 respectively. The fit results are summarized in Table 1. The numbers for efficiencies do not match those in Figure 3. This is due to the preselection applied to make the tau ID efficiency measurement.

Figure 5: The $\mu\tau_{jet}$ visible invariant mass (PASS).

Figure 6: The $\mu\tau_{jet}$ visible invariant mass (FAIL).

The values estimated from fitting data are compatible with the values obtained by applying the identification algorithm to simulated events. With the statistics used in this paper, the most interesting working point is the “loose” working point, since it has the highest efficiency. The data-to-MC ratio for this working point is determined with an accuracy of 24%, where the statistical uncertainty is dominant. Today, with more statistics available this ratio is known with an accuracy of 6%.

4.2. $Z \rightarrow \tau\tau$ production cross section

The $Z \rightarrow \tau\tau$ production cross section was measured in the four different final states: $\tau_\mu\tau_{had}$, $\tau_\mu\tau_e$, $\tau_e\tau_{had}$, and

$\tau_\mu\tau_\mu$. The data sample corresponds to an integrated luminosity of 36 pb^{-1} . A detailed description of this analysis can be found in ref. [6]. The online selection for final states that contain a muon requires the event to be triggered by a single-muon High Level Trigger (HLT) with a p_T threshold varying from 9 GeV/c to 15 GeV/c, depending on the instantaneous luminosity. The $\tau_\mu\tau_e$ final state was selected using the single electron HLT with a threshold of 12 GeV/c.

Furthermore, the selection for $\tau_\mu\tau_{had}$ and $\tau_e\tau_{had}$ final states requires that the muon and electron are isolated. The relative isolation variable is defined as $I_{rel}^{PF} = \Sigma(p_T^{charged} + p_T^{photon} + p_T^{neutral})/p_T^l$, where $p_T^{charged}$, p_T^{photon} , and $p_T^{neutral}$ are the momenta of the charged hadrons, photons, and neutral hadrons, that are contained in the cone of size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$ around the lepton flight direction. In the $\tau_\mu\tau_\mu$ channel, a cone of size 0.3 is used. The lepton isolation cut requires the I_{rel}^{PF} to be less than 0.1 for muons and 0.08 for electrons in the barrel and 0.04 in the endcap region.

The associated τ_{had} must have an opposite charge with $p_T > 20 \text{ GeV/c}$ and $|\eta| < 2.3$. In order to reject W+jets events, the transverse mass of the lepton and the missing transverse mass (M_T) is required to be less than 40 GeV/c^2 .

For the $\tau_e\tau_\mu$ final state, one isolated muon with $|\eta| < 2.1$ and one oppositely charged and isolated electron with $|\eta| < 2.4$ are required. Both the electron and the muon are required to have a $p_T > 15 \text{ GeV/c}$. Additional cuts on transverse masses are applied, $M(\mu, E^{miss})_T < 50 \text{ GeV/c}^2$, $M(e, E^{miss})_T < 50 \text{ GeV/c}^2$.

For the final states with τ_{had} , the main background contribution is coming from QCD and W+jets events, where both are estimated from data. The QCD multi jet background is estimated using the sample of Same Sign (SS) and Opposite Sign (OS) events, for which the lepton isolation requirement is inverted. In the so called ABCD technique the cut phase space is divided in four parts:

- Region A: OS and lepton isolated
- Region B: SS and lepton isolated
- Region C: OS and lepton non-isolated
- Region D: SS and lepton non-isolated

Here the regions B,C and D are dominated by QCD, while region A is expected to be primarily signal. Assuming that there is a low correlation between lepton isolation and SS/OS requirements the background contribution in region A can be estimated as $N_A^{QCD} =$

Table 1: The efficiency for hadronic τ decays to pass the HPS identification criteria measured by fitting the $Z \rightarrow \tau\tau$ signal in the samples of the “passed” and “failed” preselected events. The uncertainties are statistical only. The last column shows the data to MC correction factors and its full uncertainty, statistical and systematic.

Algorithm	Fit data	Expected MC	Data/MC
HPS “loose”	0.70 ± 0.15	0.70	1.00 ± 0.24
HPS “medium”	0.53 ± 0.13	0.53	1.01 ± 0.26
HPS “tight”	0.33 ± 0.08	0.36	0.93 ± 0.25

$N_B^{QCD} (\frac{N_C^{QCD}}{N_D^{QCD}})$. The contribution of W+jets events is subtracted from the high transverse mass region where it dominates the sample.

For the decay channels where both taus decay leptonically, the selection requires two oppositely charged and isolated leptons. Both leptons were required to have a high p_T and to be within the detector acceptance. The total missing transverse energy was required to be less than 50 GeV. For the $\tau_\mu\tau_\mu$ decay, which suffers from the Drell-Yan background, a multi-variant likelihood was employed to prevent the contribution from $Z \rightarrow \mu\mu$.

The resulting visible invariant mass for the $\tau_\mu\tau_{had}$ final state is shown in Figure 7. The signal yield is extracted from the template fitting of the visible invariant mass shape. The normalization for MC background shapes correspond to the estimated values.

Figure 7: The $\mu\tau_{jet}$ visible invariant mass for $\tau_\mu\tau_{had}$ final states.

The total cross section is defined as $\sigma(pp \rightarrow ZX) \times \mathcal{B}(Z \rightarrow \tau\tau) = \frac{N}{\mathcal{A}\mathcal{B}'\epsilon\mathcal{L}}$, where $N = N_{fit}(1 - f_{out})$ is the corrected fraction of signal events outside the generator level mass window, $\mathcal{A}$ is a geometrical acceptance, ϵ is the selection efficiency, and $\mathcal{B}'$ is the branching ratio.

The measured values for all four final states are presented in Table 2. They are consistent with each other

as well as with their theoretical prediction. The main systematic uncertainty on the cross section is coming from the reconstruction and identification of hadronically decaying τ -leptons and is shown separately in the last column. For fully leptonic decays, the statistical uncertainty is dominant. To obtain a combined value of the cross section a simultaneous fit is performed. The extracted value and the summary on the $Z \rightarrow \tau\tau$ cross section measurements are shown in Figure 8. The CMS results on the production cross-section of $\tau\tau$ are consistent with the results from previous experiments and with the recently reported value by the ATLAS collaboration, $0.97 \pm 0.07(\text{stat}) \pm 0.06(\text{syst}) \pm 0.03(\text{lumi}) \text{ nb}$ [7].

Figure 8: Summary of the measurements of the $Z \rightarrow \tau\tau$ production cross section.

5. Observation of $W \rightarrow \tau\nu$ at CMS

Another important contribution to τ -lepton studies at CMS is the $W \rightarrow \tau\nu_\tau$ production. This decay is important for testing charged lepton universality and for searches of charged Higgs bosons decaying to the $\tau\nu_\tau$ final state.

The study of $W \rightarrow \tau\nu$ has been performed using $18.4 \pm 0.7 \text{ pb}^{-1}$. The selection of $W \rightarrow \tau\nu_\tau$ decays requires the event to contain at least one high p_T HPS-tau candidate ($p_T^{\tau_h} > 30 \text{ GeV}$) large missing transverse energy ($E_T^{\text{miss}} > 35 \text{ GeV}$) and to satisfy the electron and muon

Table 2: The results for the cross section measured in four decay channels

Final state	$\sigma(pp \rightarrow ZX) \times \mathcal{B}(Z \rightarrow \tau\tau)$	stat.	syst.	lumi.	tau ID
$\tau_\mu\tau_{had}$	0.83	0.07	0.04	0.03	0.19
$\tau_e\tau_{had}$	0.94	0.11	0.03	0.04	0.22
$\tau_e\tau_\mu$	0.99	0.12	0.06	0.04	
$\tau_\mu\tau_\mu$	1.14	0.27	0.04	0.05	

veto. The PF- τ candidate must be isolated in the detector. This implies that there are no PF-charge-hadron and PF-photon candidates associated to the τ -lepton candidate within an isolation cone of size $\Delta R = 0.5$ around the τ -lepton flight direction. In order to reduce the muon contamination, the τ_{had} is required to have no hits in the muon chambers. To further suppress the QCD background, a cut on the ratio of the τ p_T to the sum of p_T of all PFjets in the event was used, ($R_{HT} = p_T^\tau / \Sigma p_T^{jet} > 0.65$).

The electroweak (EWK) background is estimated from the simulation, while the QCD contribution is estimated using the data-driven ABCD method. The four regions are defined:

- Region A: $R_{HT} > 0.65$ and $E_T^{miss} > 35$ GeV
- Region B: $R_{HT} > 0.65$ and $E_T^{miss} < 35$ GeV
- Region C: $R_{HT} < 0.65$ and $E_T^{miss} > 35$ GeV
- Region D: $R_{HT} < 0.65$ and $E_T^{miss} < 35$ GeV

The number of QCD events in region A is estimated as: $N_A = N_D \times N_B / N_C = 109 \pm 6$, where $N_B = 174 \pm 3$ is the number of signal events from simulation, $N_C = 46 \pm 2$ is the number of electroweak events estimated from simulation, and $N_D = 372$ is the number of observed events in data. The uncertainties are statistical only.

Figure 9 shows the resulting transverse mass distribution with the QCD background estimated using data-driven methods. More details can be found in the dedicated CMS paper [8].

6. Conclusion

CMS has developed an advanced and robust τ -reconstruction algorithm, which had been successfully commissioned with the first data collected. The efficiency of the HPS- τ algorithm is nearly 50% for a wide range of τ transverse momenta and with a few percents of fake-rate. The uncertainty of the τ -ID efficiency was measured to be 24% using a data sample of 36 pb^{-1} . Higher statistic-based estimation gives an uncertainty

Figure 9: The transverse mass of τ_{had} and E_T^{miss} after all cuts applied.

on the efficiency of 6%. The $Z \rightarrow \tau\tau$ production cross section was measured in four different τ -decay channels. The cross-section obtained using 36 pb^{-1} of 7 TeV data is 1.00 ± 0.05 (stat) ± 0.08 (syst) ± 0.04 (lumi) nb. This is consistent with the SM expectation, the $Z \rightarrow ll$ cross section, and the result on the $Z \rightarrow \tau\tau$ production obtained by the ATLAS Collaboration. A clear signal of $W \rightarrow \tau\nu$ based on 18.6 pb^{-1} is established.

References

- [1] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", **CMS PAS PFT-10-001**
- [2] CMS Collaboration, "The CMS experimet at the CERN LHC", *Journal of Instrumentation* **3 S08004** (2008)
- [3] K. Nakamura et al., [Particle Data Group Collaboration], *J. Phys. G* **37**, 075021 (2010).
- [4] CMS Collaboration, "Performance of τ reconstruction algorithm in 2010 data collected in CMS", to appear in *JINST* (2011) **arXiv:hep-ex/1109.6034**.
- [5] M.Cacciari and G.Soyez, "The anti-kt jetclustering algorithm", *JHEP* **04** (2008) 063, arxiv:0802.1189
- [6] CMS Collaboration, "Measurement of the Inclusive Z Cross Section via Decays to Tau Pairs in pp Collisions at $\sqrt{s} = 7\text{TeV}$ ", **CMS-EWK-10-013** (2011)
- [7] ATLAS Collaboration, "Measurement of the $Z \rightarrow \tau\tau$ Cross Section with the ATLAS Detector", CERN-PH-EP-2011-097(2011)
- [8] CMS Collaboration, "Observation of $W \rightarrow \tau\nu_\tau$ Production in pp Collisions at $\sqrt{s}=7 \text{ TeV}$ ", **CMS PAS EWK-11-002** (2011)