Tau lepton trigger and identification at CMS in Run-2

Olivier Davignon for the CMS Collaboration

Abstract

In the context of LHC Run-2, the Compact Muon Solenoid (CMS) detector was upgraded. In particular, the CMS trigger system and particle reconstruction were improved. The CMS experiment implements a sophisticated trigger system composed of a Level-1 trigger, instrumented by custom-designed hardware boards, and software layers called High-Level-Triggers (HLT). A new Level-1 trigger architecture with improved performance has been installed and is now used to maintain the thresholds used in LHC Run-1 in the more challenging conditions experienced during Run-2. Optimized software selection techniques have also been developed at the HLT. The hadronic $\tau$ reconstruction algorithm has been modified to better account for the $\pi^0(s)$ from $\tau$ decays. In addition, improvements to discriminators against QCD-induced jets and electrons were also developed. The results of these improvements are presented and the validation of the $\tau$ identification performance is shown.

Presented at Tau2016 The 14th International Workshop on Tau Lepton Physics
Tau lepton trigger and identification at CMS in Run-2

O. Davignon* for the CMS Collaboration

*Laboratoire Leprince-Ringuet, Ecole Polytechnique, Palaiseau, France

Email: davignon@cern.ch

Abstract

In the context of LHC Run-2, the Compact Muon Solenoid (CMS) detector was upgraded. In particular, the CMS trigger system and particle reconstruction were improved. The CMS experiment implements a sophisticated trigger system composed of a Level-1 trigger, instrumented by custom-designed hardware boards, and software layers called High-Level-Triggers (HLT). A new Level-1 trigger architecture with improved performance has been installed and is now used to maintain the thresholds used in LHC Run-1 in the more challenging conditions experienced during Run-2. Optimized software selection techniques have also been developed at the HLT. The hadronic $\tau$ reconstruction algorithm has been modified to better account for the $\pi^0(s)$ from $\tau$ decays. In addition, improvements to discriminators against QCD-induced jets and electrons were also developed. The results of these improvements are presented and the validation of the $\tau$ identification performance is shown.

Keywords: trigger, tau, lepton, identification, CMS, LHC, Run-2

1. Introduction

The $\tau$-lepton$^1$ is a complex object used in the context of many analyses at CMS: study of the 125 GeV Higgs boson in the $H \rightarrow \tau\tau$ channel, hunt for extra bosons as predicted by the Minimal Supersymmetric Standard Model (MSSM), search for lepton flavor violation in Higgs boson decays ($H \rightarrow \mu\tau, e\tau$), and searches for exotic particles such as the graviton, radion, $Z'$ or $W'$. Therefore, it is necessary to develop specific and powerful techniques to efficiently select and identify the $\tau_h$ at CMS.

In 2016, during the Run-2, the LHC has exceeded its nominal performance, reaching a luminosity of $1.5 \times 10^{34} \text{cm}^{-2} \cdot \text{s}^{-1}$ and an average simultaneous number of collisions per bunch crossing (pileup) up to 52. This poses a certain number of challenges for the detector that the CMS Collaboration anticipated and addressed.

The design and implementation of an efficient $\tau$ lepton trigger has always represented a challenge in high-energy physics, especially in intense collision environments such as that of the LHC. The CMS trigger is organized in two levels: the electronics-based Level-1 (L1), which reduces the rate from 40 MHz to $\sim 100$ kHz and has to take the decision within $\sim 3.8 \mu s$, and a software-based High-Level-Trigger (HLT) which subsequently reduces the rate from $\sim 100$ kHz to $\sim 1$ kHz. The L1 trigger has been fully replaced and upgraded between 2015 and 2016, and has been successfully used for the CMS data taking in 2016. The trigger algorithms and performance, with focus on the selection of $\tau_h$, are presented in this document.

The reconstruction of $\tau_h$ is key in the success of $\tau$-related searches at CMS. Subsequent identification techniques of $\tau_h$ are implemented to reject fake $\tau_h$ candidates from QCD-induced hadronic jets, electrons, and

---

$^1$Throughout this paper, “$\tau_h$” refers to hadronically decaying $\tau$-leptons.
muons. The fundamental concepts and recent developments in the reconstruction and identification of \( \tau_h \) are described in this document.

The 2016 proton-proton data set corresponds to an integrated luminosity greater than 40 fb\(^{-1}\).

2. The \( \tau_h \)-trigger at CMS

2.1. Level-1 Trigger

The reconstruction and selection of L1 \( \tau_h \)-trigger objects is based on calorimetric information. At L1 trigger, the basic reconstruction element is the Trigger Tower (TT), which has an extension of roughly 0.087×0.087 in \( \eta \)(pseudorapidity)×\( \phi \)(azimuthal angle) and, depending on the position, sums up the energy from the electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL) or hadronic forward calorimeter (HF). With respect to the previous system, the key conceptual changes to the L1 calorimeter trigger are: all the data from a single event are streamed into one processing card, which allows a global view of the detector and pileup subtraction at L1; and the clustering of the TT energies into L1-objects \((e/\gamma, \tau_h, \text{etc.)}\) is dynamic. In addition, the flexibility of the system allows performing calibration at the TT and L1-object levels, while enabling powerful identification techniques such as isolation to be implemented.

The L1 \( \tau_h \) candidates are built from a collection of TT, which are dynamically clustered around a local energy maximum ("seed TT") to capture the \( \tau_h \)-footprint and minimize the effect of pileup. Two neighboring clusters are allowed to be merged (about 15\% of genuine \( \tau_h \) are merged) in order to specifically reconstruct \( \tau \rightarrow h^+\pi^0(s) \) and \( \tau \rightarrow h^+h^0h^0 \) decays (\( h^0 \) denotes a charged hadron). To improve the energy response and resolution, the L1-\( \tau_h \) candidates are calibrated using coefficients function of the \( E_{T}^{L1} \), \( \eta^{L1} \), merging and electromagnetic fraction of the L1 candidate. The energy response at L1 with respect to offline for \( \tau_h \) candidates of 2016 data \( ^{[3]} \) is represented in Fig. 1.

The isolation energy of a L1 \( \tau_h \) candidate is defined as the sum of the transverse energies of all TT in a 6(\( \eta \))×9(\( \phi \)) TT window around the cluster seed, to which the transverse energy of the L1 \( \tau_h \) cluster(s) is subtracted. To control the large rate of fake L1 \( \tau_h \)-candidates from QCD-induced jets, an upper cut on the L1 \( \tau_h \) isolation energy is applied. The cut depends on the \( E_{T}^{L1}, \eta^{L1} \) and an estimator of pileup called \( \eta_{PT} \), which is defined as the number of TT with non-zero energy in a central-\( \eta \) ring of 8 TT. The isolation energy cut values are built such that the efficiency to select a genuine \( \tau_h \) is constant with \( \eta \) and pileup; the cut is relaxed with \( E_{T}^{L1} \) to ensure 100\% plateau efficiency. The efficiency of the L1 \( \tau_h \)-trigger, as function of \( \rho^{ offline} \), is shown for isolated seeds in Fig. 2 and was measured in 2016 data \( ^{[3]} \).

2.2. High-Level-Trigger

The High-Level-Trigger (HLT) for \( \tau_h \) is based on the Particle Flow algorithm (PF), which performs a global event reconstruction, through an optimal combination of information from the CMS sub-detectors. A list of stable particles (\( e/\mu/\gamma/\text{charged and neutral hadrons} \)) is
issued by the PF, which are in turn combined to build higher level objects like jets, $\tau_h$ and missing transverse energy.

At HLT, $\tau_h$ are identified combining charged and neutral PF candidates included in jets reconstructed using the anti-$k_T$ algorithm with a cone of size $\Delta R = 0.4$. Up to three leading charged hadron candidates within a narrow cone (typically $\Delta R = 0.1$) around the direction of the seeding jet are assumed to be the signature of the $\tau$ decay into charged hadrons. Electromagnetic deposits in ECAL compatible with the $\pi^0$ characteristic of $\tau$ decays in $h^+h^0(s)$ are considered as part of the signature as well. To reduce the contamination from QCD-induced jets faking $\tau_h$, isolation selections are applied to HLT $\tau_h$ candidates. The isolation is defined as sum-$p_T$ of non-signal charged and neutral PF candidates that lie within the jet cone $\Delta R < 0.4$. In order to reduce the dependency of the isolation on pileup, a subtraction term reflecting the average energy deposit per unit area from soft scattering ($\rho_A A_f$) is applied. The efficiency of the HLT $\tau_h$-trigger can be observed in Fig. 3 and was measured in 2016 data [4].

![Figure 3](image-url)  
Figure 3: HLT trigger efficiency as a function of the offline $\tau_h$ transverse momentum, for an HLT $\tau_h p_T$ threshold of 20 GeV.

3. $\tau_h$-reconstruction and identification

3.1. $\tau_h$-reconstruction

The offline $\tau_h$-reconstruction at CMS is called the “Hadron Plus Strips (HPS)” algorithm [5][6]. The $\tau_h$ candidates are built from a combination of Particle Flow charged hadron(s) and $\pi^0 \rightarrow \gamma\gamma$ candidate(s), the latter reconstructed with $\eta \rightarrow \phi$ clusters of $e/\gamma$ candidates (“strips”). Using a set of constraints on the mass and transverse momenta of the decay products of the $\tau$, the HPS algorithm reconstructs the different $\tau_h$ hadronic decay modes, notably the $h^+\pi^0(s)$, and $h^+h^+h^-h^-$ signatures. The HPS algorithm used in Run-2 is similar to the one of Run-1, with a few modifications. In particular, a new feature is the dynamic strip reconstruction: the size of the window in which strips are reconstructed is taken as function of the $e/\gamma$-candidate transverse momentum. This allows for a better containment of the $\tau_h$ energy within strips and avoids the associated impact on the isolation energy which is used as a discriminating variable to identify $\tau_h$ (see Sec. 3.2). Fig. 4 shows the strip size in the $\phi$ direction, as function of the $e/\gamma$ $p_T$ for $\tau_h$ from simulation.

![Figure 4](image-url)  
Figure 4: Strip window size in the $\phi$ direction as function of the $e/\gamma$ $p_T$ for $\tau_h$ candidates from simulation. The size of the region in which strips can be reconstructed decreases with $p_T$. The dashed dots defines the region where 95% of the $\tau_h$-candidates are contained, while the red line is an exponential fit to this function.

3.2. $\tau_h$-identification

The $\tau_h$-identification at CMS is aimed at efficiently selecting the genuine $\tau_h$ while rejecting the fakes from jets, electrons, and muons. To reject the contribution from jet fakes, the main handle used is called isolation, which quantifies the activity around the $\tau_h$-candidate: on average it is less for genuine $\tau_h$ than for jets. The cut-based discriminator relies on the isolation-sum discriminating variable, which sums the $p_T$ of PF candidates (charged and neutral, excluding the ones that belong to the $\tau_h$ itself) in a cone centered around the $\tau_h$. Corrections are applied to reduce the effect of pileup on the isolation. Several working points, each targeting a specific signal efficiency, are defined by imposing cuts on the isolation-sum variable. The mis-identification probability as function of the $\tau_h$-identification efficiency is shown in Fig. 5.
An improved multivariate analysis (MVA) based discriminant against jets has also been developed. The use of an MVA approach allows the information from isolation energy, $\tau$-lifetime and shower shapes to be optimally combined. With respect to the cut-based approach to isolation, the MVA-based discriminant leads to a significant reduction of the mis-identification probability at constant signal efficiency, as demonstrated in Fig. 5.

The $\tau_h$-identification efficiency has been measured in Run-2 data, using three different data-driven techniques [6]: the Tag & Probe (T&P) method which relies on $Z \rightarrow \tau_h\tau_h$ events; the $Z \rightarrow \tau_p\tau_h/\mu\mu$ ratio method; and using $W^* \rightarrow \tau\tau$ events. Here, the results from the first method are described. The T&P sample of $Z \rightarrow \tau_p\tau_h$ candidates consists in events with at least one well identified and isolated muon (the tag) and one loosely-preselected $\tau_h$ candidate (the probe). A set of other requirements are imposed to ensure a high purity in $Z \rightarrow \tau_p\tau_h$. The events are separated into pass and fail categories, depending on if the $\tau_h$-candidate passes the isolation discriminator under study. The signal ($Z \rightarrow \tau_p\tau_h$) and background expected contributions are predicted using simulation, except the QCD multi-jet and W+jets backgrounds for which sideband techniques are used. The parameter-of-interest, the $\tau_h$-identification scale factor (defined as the ratio of observed data to simulation efficiencies to select $\tau_h$ from the $Z \rightarrow \tau_p\tau_h$ process) is then extracted from a simultaneous maximum likelihood fit in both categories, including systematic uncertainties on the signal and background normalization and shapes. In the fit, the low-purity fail category helps to constrain the backgrounds.

The results for the measurement of the $\tau_h$-identification scale factors from the T&P method, using the visible mass of the $\tau_h$ decay products as the fit variable, are summarized in Table 1. They show good consistency between the observed data and simulation. Details about anti-electron and anti-muon discriminators can be found in Ref. [6].

![Figure 5: Mis-identification probability, defined as the chance for true QCD jets from simulation to be mis-identified as $\tau_h$, as function of the $\tau_h$ identification efficiency, evaluated from $H \rightarrow \tau\tau$ simulated events. The cut-based discriminant is compared to the MVA-based one. The points correspond to fixed-efficiency working points.](image)

| Isolation discriminant | Scale Factor |
|------------------------|--------------|
| Cut-based              |              |
| Loose                  | 0.98 ± 0.06  |
| Medium                 | 0.98 ± 0.06  |
| Tight                  | 0.97 ± 0.06  |
| MVA-based              |              |
| Loose                  | 1.00 ± 0.06  |
| Medium                 | 1.00 ± 0.06  |
| Tight                  | 1.01 ± 0.06  |
| Very Tight             | 1.00 ± 0.06  |

Table 1: Data/MC scale factors for the different working points of the cut-based and MVA-based discriminator against jets. They were derived using the Tag & Probe technique on Run-2 data.

4. Conclusion

In this note, the systems and techniques used to trigger, reconstruct and identify $\tau_h$ at CMS in Run-2 have been presented. The L1 trigger system of the CMS detector has been recently fully upgraded. Excellent performance in the selection of $\tau_h$ at L1 has been achieved in 2016. The High-Level-Trigger has also demonstrated good performance, thanks to its offline-like reconstruction of $\tau_h$. The offline $\tau_h$ reconstruction and identification techniques have also been updated in the context of LHC Run-2. In particular, the development of advanced discriminators against jets has been pursued, leading to improved rejection of fakes.

Acknowledgments

The work of the author of this paper has been partly funded by the P2IO LabEx (ANR-10-LABX-0038) in the framework “Investissements d’Avenir” (ANR-11-IDEX-0003-01) managed by the French National Research Agency (ANR).

References

[1] CMS Collaboration, CERN-LHCC-2000-038.
[2] CMS Collaboration, CERN-LHCC-2013-011, CMS-TDR-12, CMS-TDR-012.
[3] CMS Collaboration, CERN-CMS-DP-2016-044.
[4] CMS Collaboration, CERN-CMS-DP-2016-056.
[5] CMS Collaboration, JINST 11, no. 01, P01019 (2016).
[6] CMS Collaboration, CMS-PAS-TAU-16-002.