The Compact Muon Solenoid Experiment
Conference Report

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

27 November 2008 (v2, 01 December 2008)

Tau reconstruction and identification with particle-flow techniques using the CMS detector at LHC

Michele Pioppi

Abstract

New Physics beyond the Standard Model could well preferentially show up at the LHC in final states with taus. The development of efficient and accurate reconstruction and identification of taus is therefore an important item in the CMS physics programme. The potentially superior performance of a particle-flow approach can help to achieve this goal with the CMS detector.

Presented at *TAU 08*,
10th International Workshop On Tau Lepton Physics,
22-25 /09/2008,
Novosibirsk, Russian Federation

Tau reconstruction and identification with particle-flow techniques using the CMS detector at LHC.

M. Pioppi ^{a *}

on behalf of the CMS collaboration

^aCERN, CH-1211, Genève Switzerland

New Physics beyond the Standard Model could well preferentially show up at the LHC in final states with taus. The development of efficient and accurate reconstruction and identification of taus is therefore an important item in the CMS physics programme. The potentially superior performance of a particle-flow approach can help to achieve this goal with the CMS detector.

1. Introduction

Because the tau is the heaviest of the three leptons, specific final states involving taus are expected to show up in the Standard Model (SM), and would appear abundantly in many processes arising from new physics beyond the SM. Because taus decay to hadronic final states in about 64% , the reconstruction and the identification of these hadronic decays is an important element of the CMS physics programme. In this summary, the strategy used by CMS to reconstruct and identify hadronic decays of taus with particle-flow techniques is outlined. Several and substantial improvements to the tau reconstruction, through the particle-flow algorithm developments, and identification are still expected and are actively being worked on.

2. Experimental Challenges For Taus at Hadron Colliders

From a calorimeter point-of-view, hadronic taus resemble ordinary quark and gluon jets arising from QCD multijet production, with electromagnetic and hadronic energy deposits, from the neutral and charged pions, respectively. The outstanding challenge at hadron colliders is therefore to reduce this enormous QCD background, the cross section of which is many orders of magnitude larger than any new physics signa-

tures. Another complication arises from the fact that a significant fraction of the tau momentum escapes undetected with the ν_τ , which renders tau jets softer and further reduces the experimental discrimination of signals involving taus.

3. Particle Flow Overview

The particle-flow reconstruction algorithm aims at providing a global (*i.e.*, complete and unique) event description at the level of individually reconstructed particles, with an optimal combination of the information coming from all CMS subdetectors. The reconstructed and identified individual particle list includes muons, electrons (with individual reconstruction and identification of all Bremsstrahlung photons), photons (either unconverted or converted), charged hadrons (without or with a nuclear interaction in the tracker material), as well as stable and unstable neutral hadrons. These particles can be non isolated, and in some cases they produce a intricate overlap of reconstructed charged tracks, ECAL and HCAL energy clusters, and signal in the muon chambers. The complete list of reconstructed particles may then be used to derive composite physics objects, such as clustering into jets with standard jet algorithms.

The algorithms discussed in this summary use this list of particles both for reconstruction (jet clustering) and identification of taus. Specifically, all reconstructed particles in the event, includ-

*Now at Imperial College, London, United Kingdom

ing charged pions and photons from any possible hadronic tau decay products, are clustered into jets using a Cone algorithm with radius of 0.5 [1]. The tau algorithms benefits from both the improved energy and angular resolution with respect to the calorimeter-based algorithm (Fig. 1) and the depth of information available describing each individual particle in the jet [2]. The limited energy and angular resolution of the calorimeter-based jets is dominated by the hadron calorimeter resolution and granularity. As the tau decay products are mostly photons and charged pions, the particle-flow-based jets benefit fully from the tracker and electromagnetic calorimeter superior resolutions. In addition, the bias in the calorimeter-based jet energy cannot be corrected by the standard jet-energy calibration. The latter is indeed primarily aimed at QCD jets, and would overestimates by a large factor the tau true energy. Instead, particle-flow-based jets are calibrated by construction, with the use of the accurate charged-particle-momentum and photon-energy determination. These refinements allow to significantly improve the QCD jet rejection.

Figure 1. Comparison between particle-flow-based (red) and calorimeter-based (black) reconstruction of single τ^- 's with $p_T = 50 \text{ GeV}/c$ (typical of a Z decay). Difference, in GeV, between the measured and the true visible transverse momentum (ΔE_T).

4. Base Tau Reconstruction and Identification

The tau reconstruction and identification proceeds in two distinct stages: (1) a common pre-selection, which serves as a basis for all final states with taus: it employs relatively simple and robust methods similar to those used for trigger conditions, and is aimed at strongly suppressing backgrounds while still preserving a large fraction of the genuine taus; (2) sophisticated tau identification algorithms, described in Section 5, suitable and tunable for each individual physics analyses, towards achieving the desired high purity.

Several reasons justify this split, including the difficulty of using so called “tag-and-probe” methods (where one “tags” an object and then attempts to measure the identification efficiency by “probing” an associated target object) to estimate the tau identification efficiency from data. Unlike in the case of electrons and muons, hadronic taus require a generic pre-selection to reduce the huge QCD background, keeping a large efficiency for all tau decay modes. The pre-selection results in relatively pure and unbiased $Z \rightarrow \tau\tau \rightarrow \mu(e) + \tau\text{-jet}$ samples, enabling the use of tag-and-probe techniques to estimate the efficiency of sophisticated tau identification algorithms from the pre-selection point onwards. Possible loss of efficiency in the pre-selection due to the presence of pileup or underlying particles can also be determined from the data via $Z \rightarrow ee$ and $\mu\mu$ samples.

The essential features of the common pre-selection are as follows. First, a transverse-momentum threshold is applied to each jet and only those satisfying this threshold are further considered as a possible tau candidate. Next, at least one charged hadron with p_T in excess of $5 \text{ GeV}/c$ is required to be found at a distance from the jet direction smaller than 0.1 in the (η, φ) plane. The highest-momentum charged hadron satisfying this cut is called the “leading track”.

A narrow “signal cone”, expected to contain all tau decay products, is then defined around the direction of the leading track, and an “isolation annulus”, expected to contain little activity if the tau is indeed isolated, is defined as a cone

larger than but excluding the signal cone. It is important to note that the cone particle contents are determined with the particle directions measured at the primary event vertex, as delivered by the particle-flow algorithm, and are thus unaffected by sweeping effects caused by the strong magnetic field. To enforce the tau isolation, no reconstructed charged hadrons with p_T above 1 GeV/ c and no photons with p_T larger than 1.5 GeV/ c are allowed in the isolation annulus.

The efficiency for reconstructing a jet with $p_T > 15$ GeV/ c matched to a true tau-jet or a generated QCD jet is shown by the black circles in Fig. 3. (Here QCD events are generated with a $\hat{p}_T$ between 5 and 120 GeV/ c .) A “generated jet” is a cluster of generated and detectable particles into a jet with the same jet-clustering algorithm as that used on reconstructed particles. The seemingly larger efficiency for the background is therefore an artifact of the substantially harder spectrum of the simulated QCD jets than the simulated taus in $Z \rightarrow \tau\tau$ events.

The leading-track finding marginal efficiency (*i.e.*, determined with respect to the tau candidates satisfying the previous cut) reflects the probability of a tau (or a QCD jet) to actually contain a charged particle with $p_T > 5$ GeV/ c , folded with the track reconstruction efficiency. The latter dominates the asymptotic behavior of the efficiency curve for high-momentum taus and jets. The probability of finding a track with $p_T > 5$ GeV/ c for taus is larger than the corresponding probability in QCD jets with the same transverse momentum, because of the larger average particle multiplicity in QCD jets, hence the democratic energy sharing among more particles. The leading-track requirement therefore provides a significant suppression of QCD backgrounds (vastly dominated by low- p_T QCD jets).

Up to now, many CMS analyses [3] have used fixed-sized signal and isolation cones of typical sizes $\Delta R = 0.07$ and 0.45, respectively, defined in the (η, φ) plane. An alternative approach used here is to utilize the fact that high energy taus are Lorentz boosted and hence become more collimated at higher energy: the “signal” cone size is defined to shrink inversely proportional to the jet transverse energy, $5/E_T$, with a minimum limit

of 0.07 and a maximum limit of 0.15.

A comparison of the performance of the (historical) $\Delta R_{\text{sig}} = 0.07$ and the (new) $\Delta R_{\text{sig}} = 5/E_T$ signal cones in terms of the marginal efficiency of the charged hadron isolation requirements is shown in Fig. 2 for taus as a function of p_T . An increase of approximately 20% is observed in the signal efficiency for the low- p_T region, with an approximate doubling of the background rate. The better efficiency of the shrinking-cone algorithm is due to a better acceptance for the three-prong taus in the low-to-intermediate p_T range, due to the larger signal cone in which all three tracks can fit. The recovery of the three-prong decays is essential to make the base selection independent of the decay mode. An eventually better rejection of the background is expected with the higher-level identification algorithms, at no cost for the signal, from a detailed analysis of the jet shape and particle content. Photon isolation is another pow-

Figure 2. Comparison of the marginal efficiencies for the charged-hadron isolation requirement efficiency as a function of the true visible p_T for taus with the shrinking ($\Delta R_{\text{sig}} < 5/E_T$) and fixed ($\Delta R_{\text{sig}} = 0.07$) signal-cone definitions.

erful discriminator against QCD jet backgrounds. Since, the substantial amount of tracker material leads to high rate of photon conversions, another

development in the context of the particle-flow algorithm is to reconstruct secondary tracks originating from photon conversions (which is not yet completed). As a consequence, low-energy electrons from photon conversions may appear as photons in the isolation annulus (being strongly bent by the magnetic field). When conversion reconstruction becomes available, the signal cone definition will be re-optimized to provide additional background rejection. A first study has already demonstrated encouraging improvement.

Neutral hadron isolation was shown to have some rejection power, but it is much more dependent on: *(i)* possible double counting in the particle-flow algorithm due to the hadron calorimeter resolution; *(ii)* the higher noise in the hadronic calorimeter with respect to the other subsystems. For these reasons it is not used for the time being.

The global efficiencies of the successive pre-selection cuts as a function of p_T and η are shown for the shrinking-cone algorithm in Fig. 3. The efficiencies are determined with respect to the taus (or QCD jets) with a true visible p_T in excess of 5 GeV/c and a true visible η between -2.5 and 2.5 .

5. Higher Level Identification Criteria

Following the base tau reconstruction, which emphasizes robustness, high efficiency, and dataset size reduction, a high level identification stage is designed to achieve higher purity samples suitable for individual physics analyses. While it is important to develop tools and algorithms as early as possible, there is no doubt that vigorous re-optimization and extensions of the high level algorithms will be necessary when data become available. So far, reasonably well understood algorithms include criteria to reject electrons and muons.

5.1. Electron Rejection

After suppressing QCD jet backgrounds, isolated electrons produced in the electroweak processes, e.g. $Z \rightarrow ee$, become an important source of misidentified taus in many physics analyses. Such electrons are not efficiently rejected by the

Figure 3. Global efficiencies of the successive pre-selection cuts for taus (top) and QCD jets (bottom) as a function of the true visible p_T for the $5/E_T$ shrinking-cone algorithm.

isolation algorithms in the base reconstruction, requiring a special treatment to reduce their contamination. Because of the large amount of material in CMS tracker, electrons often emit a large fraction of their energy as Bremsstrahlung photons.

Hence, a particle-flow electron pre-identification algorithm has been developed, based on a fast multivariate analysis of tracker and calorimeter information, which provides efficient seeds for full electron reconstruction (which captures individual Bremsstrahlung photons) within jets and at low momenta. The electron pre-identification achieves 90-95% efficiency

across the entire tracker acceptance, with about 5% pion efficiency. In order to optimize the electron rejection efficiency beyond 95%, two additional variables are formed. The first variable, E/P , is defined as the summed energy of all ECAL clusters in a narrow strip $|\Delta\eta| < 0.04$ with respect to the extrapolated impact point of the leading track on the ECAL surface, divided by the momentum of the leading charged hadron inside the jet (the strip extends in $\Delta\phi$ for up to 0.5 in the direction of the expected Bremsstrahlung photon deposition). This variable is expected to cluster around unity for electrons, and to be scattered around smaller values for charged pions from tau decays. The second variable, $H_{3\times 3}/P$, is defined as the summed energy of all HCAL clusters within a $\Delta R < 0.184$ around the extrapolated impact point of the leading track on the ECAL surface, divided by the momentum of the leading charged hadron inside the jet. This variable is expected to cluster around zero for electrons, and to be somewhat randomly distributed for charged pions from tau decays.

Taus pre-identified as electrons have a behaviour similar to electrons regarding these two variables. Tight cuts have therefore to be applied to reject the true electrons without losing tau efficiency. On the other hand, taus that are not pre-identified as electrons can be cut with looser criteria to reject as many electrons as possible. The optimized electron rejection cuts are found to be

1. $E/P < 0.8$ or $H_{3\times 3}/P > 0.15$ for the candidates not pre-identified as electrons;
2. $E/P < 0.95$ or $H_{3\times 3}/P > 0.05$ for the candidates pre-identified as electrons;

which lead to an efficiency of 92.5% for true taus and 1.5% for true electrons. A summary of all results is shown in Fig. 4, where the quantity $H_{\max}/P$ is defined as the energy of the leading HCAL cluster divided by the momentum of the leading charged hadron inside the jet; the label E_{id} represents the electron pre-identification cut. The optimized electron rejection cuts described above are labeled as “Optimized Electron Veto”.

Figure 4. Electron vs. tau selection efficiency for several typical electron rejection approaches compared to the Optimized Electron Veto. The different efficiency curves are obtained by varying a cut on each of the variables listed in the legend. The efficiency of optimized veto is displayed as a blue cross. The quantities used are defined in the text.

5.2. Muon rejection

As in the case of isolated electrons, without additional rejection criteria, isolated muons could contaminate the identified tau candidates with an unacceptable rate. The very high efficiency of standard muon reconstruction and identification in CMS provides nearly optimal rejection of muons otherwise identified as tau candidates. Default reconstructed muons include (i) tracks matched with muon chamber segments; and (ii) tracks that do not match any signal in the muon system, *e.g.*, because of gaps between muon chambers, but have calorimeter energy deposits consistent with a minimum-ionizing particle hypothesis. Variables evaluating the compatibility of the calorimeter and segment measurements with originating from a muon are derived for each track with a likelihood technique, as described in Refs. [4,5]. Hence in defining the muon rejection criteria, two distinct options are considered: the tau candidate is rejected either if (1) the leading track matches any identified muon

(including the sole calorimetry compatibility), or (2) the leading track matches an identified muon with the presence of at least one segment in the muon chambers. The resulting muon rejection efficiency is above 99%, and the selection efficiency for hadronic taus remains at greater than 99%.

5.3. Future improvements

Finally, several and substantial improvements to the tau reconstruction (through the particle-flow algorithm developments) and identification (through high-level analysis tools, like multivariate analysis techniques) are still expected and are actively being worked on. For example, the inclusion of photon conversion tagging will allow a better tuning of the photon isolation requirement, further suppressing the QCD jet background at no cost for the signal efficiency.

6. Conclusion

This paper describes tau reconstruction and identification using particle flow with the CMS detector. There are three major components: a general particle flow reconstruction, a common tau reconstruction using reconstructed particles, and a higher level identification. Since the common reconstruction selection will be used to define the CMS tau secondary datasets, it therefore has to satisfy several requirements: robustness with respect to unexpected detector effects, high efficiency for selecting true hadronically decaying taus and sufficient rejection of QCD jet backgrounds to ensure manageable size of secondary datasets. The proposed schema satisfies all of these requirements. While further significant improvements are still being pursued, existing methods already provide a strong rejection of electron and muon backgrounds, preserving high efficiency for selecting hadronic taus.

REFERENCES

1. The CMS collaboration, Performance of Jet Algorithms in CMS, CMS-PAS JME-2007-003.
2. The CMS collaboration, CMS Strategies for tau reconstruction and identification using particle-flow technique, CMS-PAS PFT-08-001.
3. The CMS collaboration, Physics Technical Design Report Volume II: Physics Performance and Software, J. Phys. G (Nucl. Part. Phys.), 34, 2007
4. The CMS collaboration, Physics Technical Design Report Volume I: Detector Performance and Software, CERN/LHCC 2006-001
5. James, E. and others, Muon Identification in CMS, CMS Note 2006-010