Tau reconstruction at CMS with a focus on high p_T taus

Soham Bhattacharya[1]

on behalf of the CMS Collaboration

[1]TIFR, Mumbai, India

Abstract

In this poster we present the algorithm and performance of tau reconstruction at the CMS experiment, while highlighting a dedicated reconstruction algorithm that uses calorimeter hits instead of tracks to reconstruct taus with high transverse momentum. Describing the standard Hadron-Plus-Strip (HPS) algorithm and its dependence on track reconstruction and shower modelling, we present the calorimetric tau reconstruction that uses minimal track information for high p_T taus. The pros and cons of these algorithms are discussed along with their performance and potential uses.

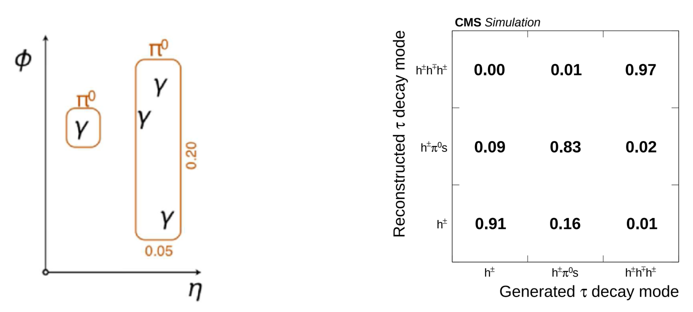
This study is work in progress, and is an attempt to tune the reconstruction for high p_T taus. *The calo-tau algorithm is not yet an official tau reconstruction algorithm for CMS.*

Tau decay modes

Decay mode (DM)	DM code $5(n_{h^\pm}-1)+n_{\pi^0}$	Resonance (mass in GeV)	$\mathcal{B}$ [%]
Leptonic decays			
$\tau^- \to e^-\bar{\nu}_e\nu_\tau$			17.4
$\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$			17.8
Total			35.2
Hadronic decays			
$\tau^- \to h^-\nu_\tau$	0		11.5
$\tau^- \to h^-\pi^0\nu_\tau$	1	ρ (0.77)	25.9
$\tau^- \to h^-\pi^0\pi^0\nu_\tau$	2	a_1 (1.26)	9.5
$\tau^- \to h^-h^+h^-\nu_\tau$	10	a_1 (1.26)	9.8
$\tau^- \to h^-h^+h^-\pi^0\nu_\tau$	11		4.8
Others			3.3
Total			64.8

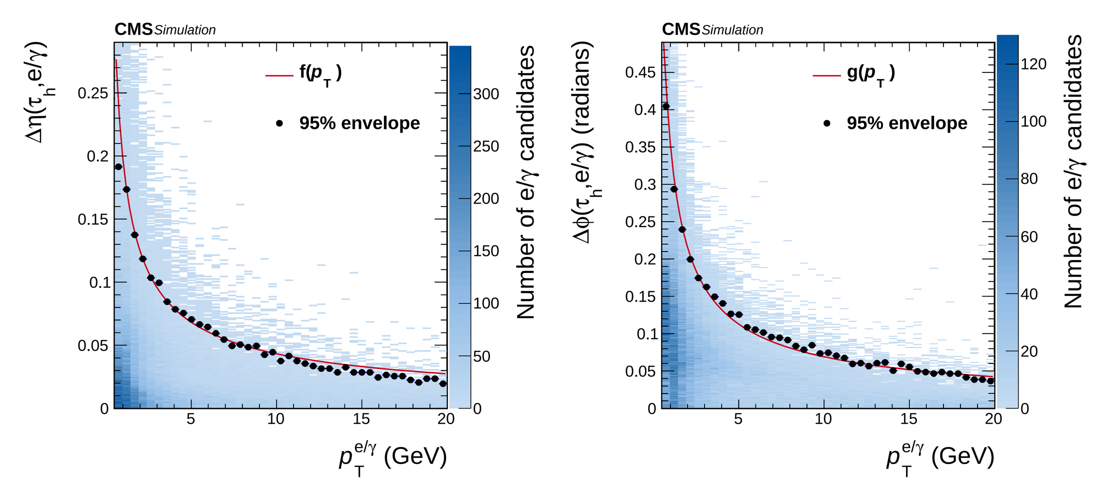
Hadron plus strips (HPS) algorithm

- The HPS algorithm is **seeded by jets.**
- The electron and photon constituents in the jet are clustered into "strips" which try to capture the neutral pion decay.
- The strip size was fixed ($\Delta\eta \times \Delta\phi = 0.05 \times 0.2$) in Run-1, and is dynamic (p_T dependent) for Run-2.

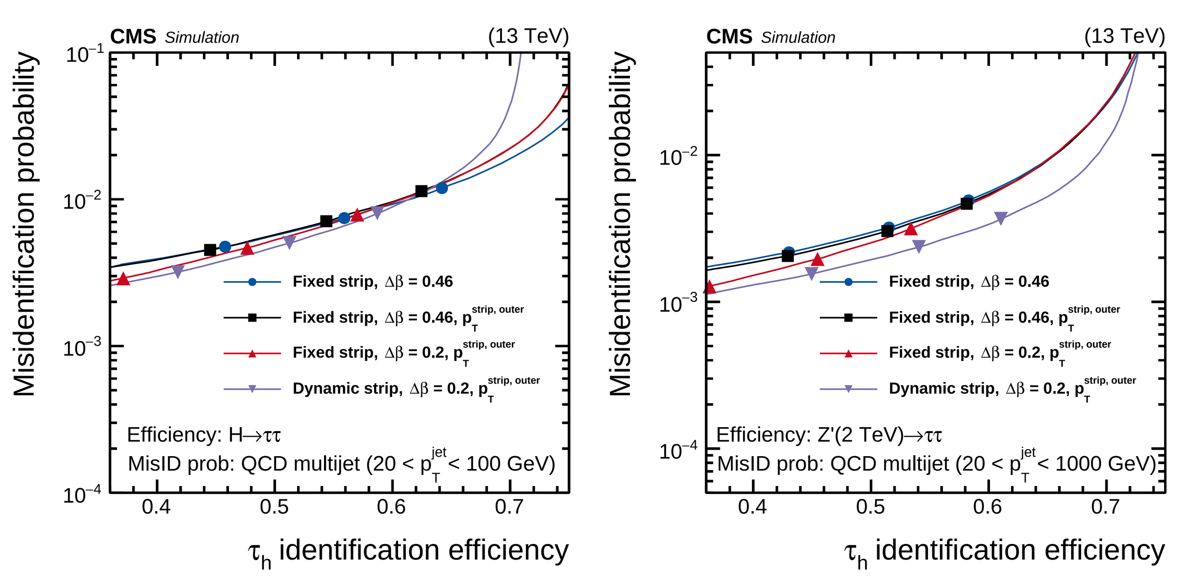


- The algorithm forms the following τ_h candidates (expected from the τ_h decay modes):
  - $h^\pm$: A single charged hadron candidate without any strips.
  - $h^\pm\pi^0$: Combination of one charged hadron and one strip.
  - $h^\pm\pi^0\pi^0$: Combination of one charged hadron and two strips.
  - $h^\pm h^\pm h^\pm$: Combination of three charged hadrons without any strips.

Dynamic strip reconstruction



- The τ_h decay products can contain low p_T components (from charged pion interaction with tracker material, or from multiple conversions and brehmsstrahlung from the e/γ from the π^0 decay). These can contribute to the τ_h isolation. Can **increase the strip size accordingly** to take these into account.
- The τ_h decay products will be more boosted (and hence collimated) with increasing p_T. Can **reduce the strip size accordingly** to decrease contamination.



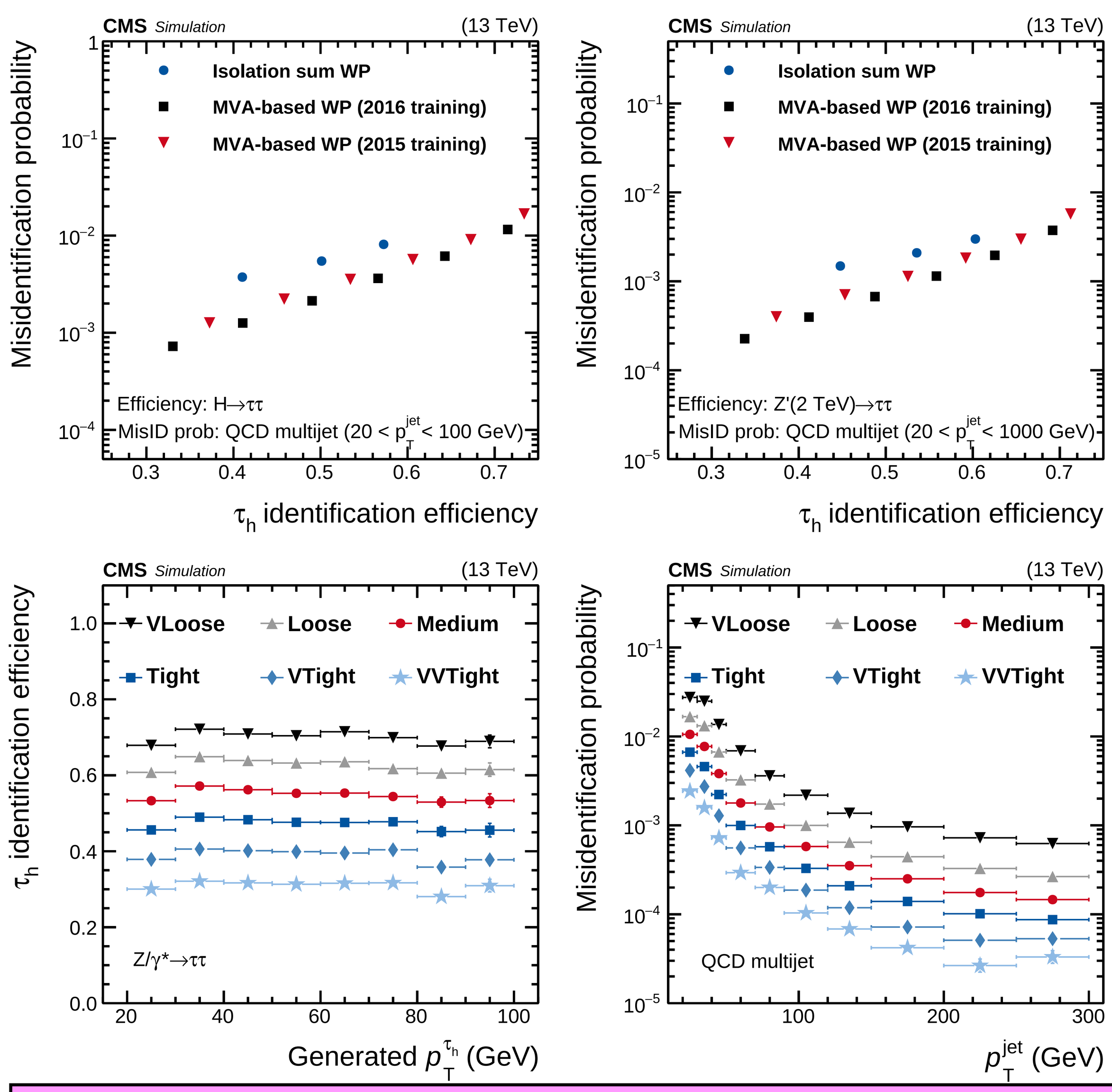
τ_h^{HPS} vs. QCD jet discrimination: Isolation-sum

- QCD jets are expected to have higher activity (tracks, calorimeter deposits) in an annular region around the signal cone.
- An isolation cone size of $\Delta R = 0.5$ around the τ_h axis is considered.
- The isolation of a τ_h candidate is computed as:

$$I_\tau = \sum_{d_z<0.2\text{ cm}} p_T^{\text{charged}} + \max\left(0, \sum p_T^\gamma - \Delta\beta \sum_{d_z>0.2\text{ cm}} p_T^{\text{charged}}\right)$$

- The $\Delta\beta$ term takes care of the **contribution from pileup to the photon isolation.** $\Delta\beta$ is 0.2 (0.46) for Run-2 (Run-1).
- In addition to this, a cut on the p_T-sum of the e/γ that are out of the signal cone but in the strips, helps to reduce the misidentification probability.

τ_h^{HPS} vs. QCD jet discrimination: MVA



The high p_T regime: Need for a more robust algorithm

- The HPS algorithm **relies heavily on track reconstruction.** A boosted three-prong τ_h may not have all its tracks well resolved, and may appear as a two-pronged object, and hence will not be treated as a τ_h (hadronically decaying τ lepton) candidate.
- A tau-reconstruction algorithm that relies primarily on calorimeter deposits only, will be free of these issues, and can be useful at very high p_T ($\sim$ TeV).
- A calorimeter based tau reconstruction algorithm is more robust against possible mismodellings in Monte-Carlo simulations, and can serve as a cross-check of whether potential high p_T τ_h signals are lost in data.

The calo-tau algorithm

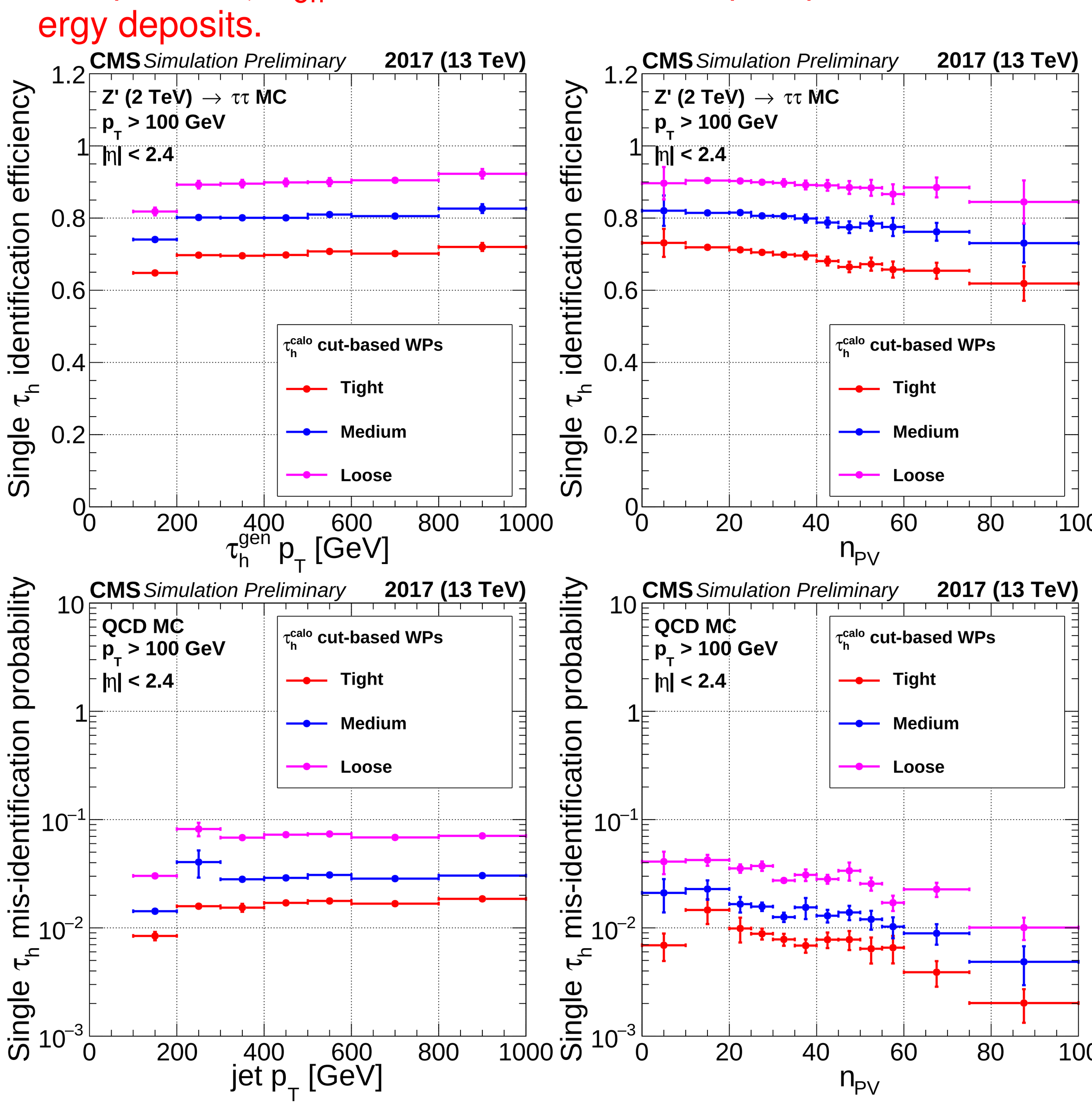
- Seed the algorithm with a calorimeter jet reconstructed with the anti-kt algorithm with a distance parameter of 0.4 (AK4 calo-jet).
- The existence of a track with the following condition is required to select the jet as a τ candidate. **Note that the p_T measurement of this track does NOT play a significant role.** The track ($p_T > 0.5$ GeV) must be within a cone of $\Delta R < 0.1$ around the jet axis. The track's transverse impact parameter (d_0) must be < 0.1 cm.
- **Set the 4-momentum of the calo-tau to that of the calo-jet.** Track p_T measurement does NOT play any role here either.

τ_h^{calo} vs. QCD jet discrimination: Isolation-sum

- Isolation variable:

$$I_{\text{iso},\rho}^{\text{comb}} = H_T^{\text{iso-trk}} + \max(0, E_T^{\text{iso-ECAL}} - \rho A_{\text{eff}})$$

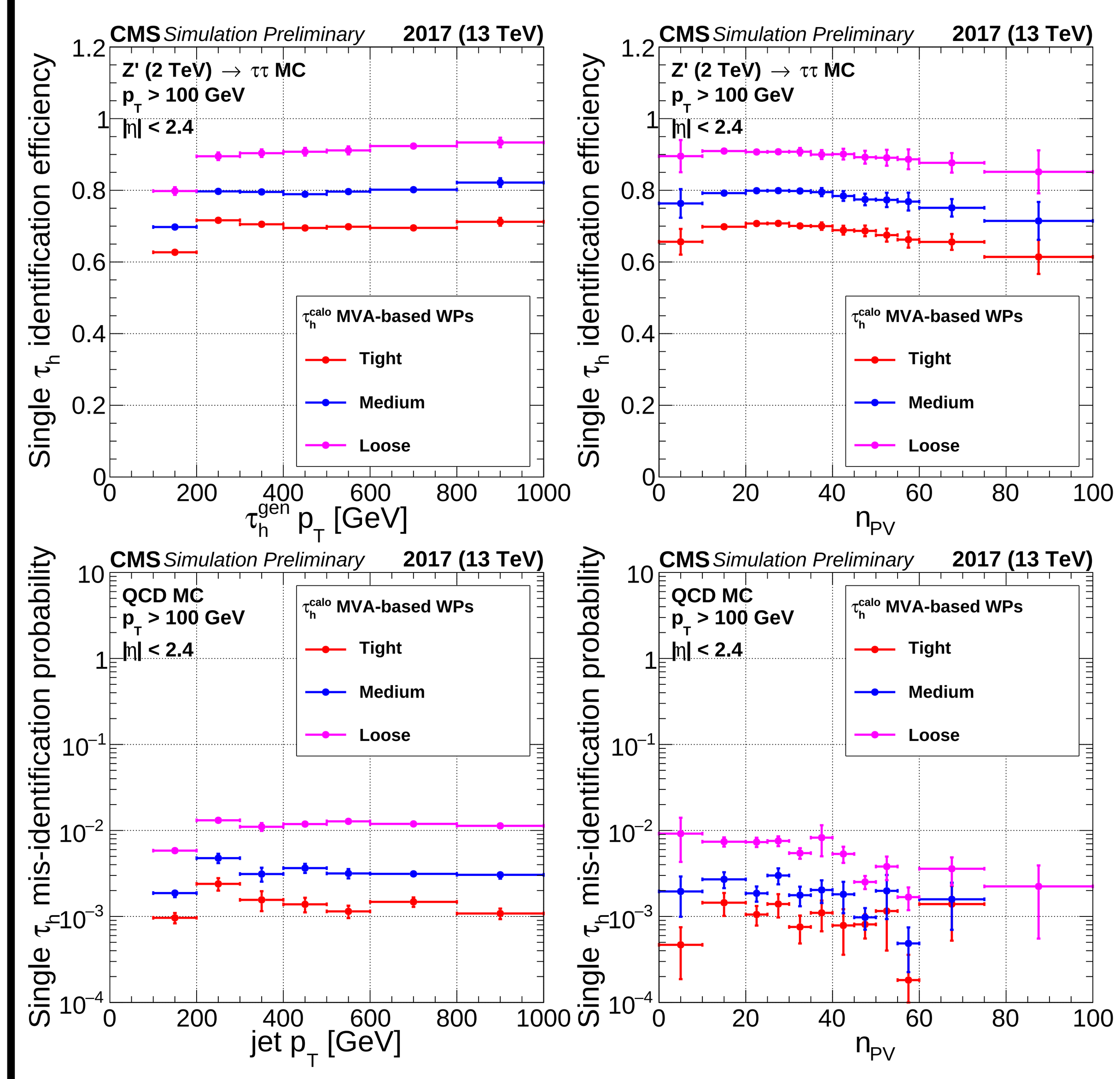
- $H_T^{\text{iso-trk}}$: Scalar sum of the p_T of racks in the annular region $0.07 < \Delta R < 0.5$ (w.r.t. the leading signal track).
- $E_T^{\text{iso-ECAL}}$: Sum of ECAL deposits with $E_T > 0.5$ GeV within an the annular region $0.15 < \Delta R < 0.5$ w.r.t. the leading signal track.
- ρ is the median of calo-jet energy divided by its area.
- A_{eff} is an effective area whose value (0.2) is chosen such that the efficiency is independent of pileup.
- The product ρA_{eff} is the contribution from pileup to the ECAL energy deposits.



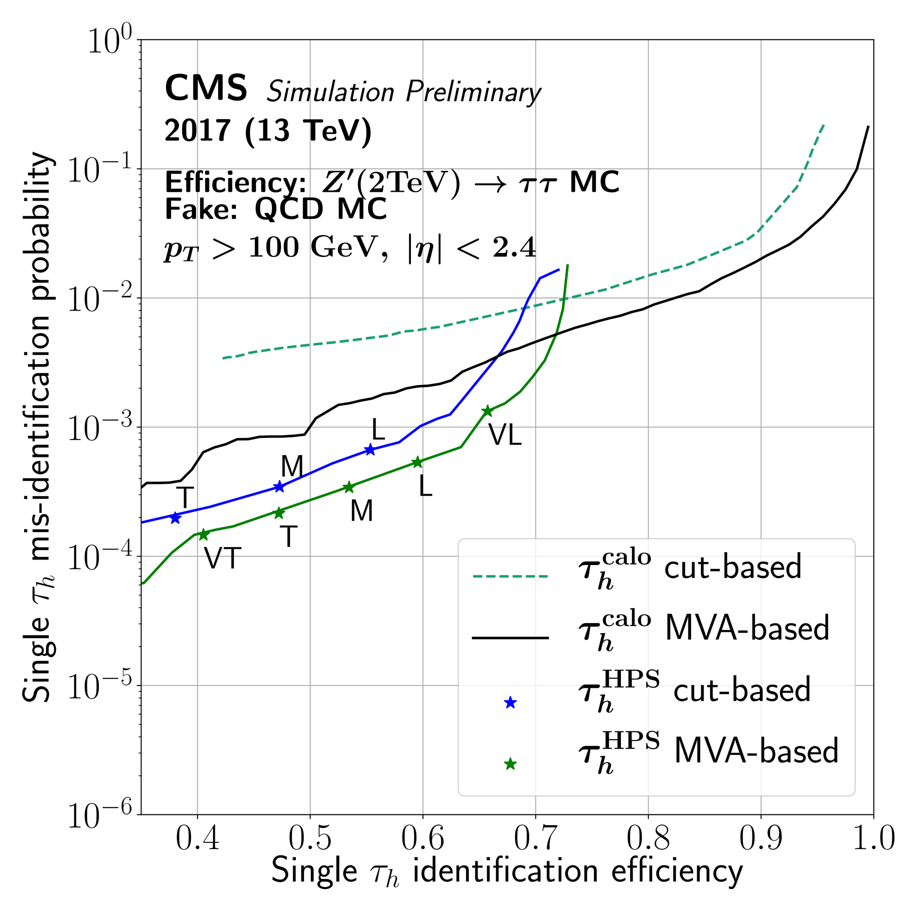
τ_h^{calo} vs. QCD jet discrimination: MVA

Some of the most useful MVA variables:

- $n_{\text{sig-trk}}$: Number of signal tracks. Tracks within $\Delta R < 0.07$ w.r.t. the leading signal track.
- $n_{\text{iso-trk}}$: Number of isolation tracks. Tracks in the annular region $0.07 < \Delta R < 0.5$ w.r.t. the leading signal track.
- m: Invariant mass of the calo-tau.
- E_T^{iso}: Sum of the ECAL energy deposits (transverse component) in the isolation annulus.
- $d_{xy}^{\text{sig-trk}_1}$: The transverse impact parameter of the leading signal track.
- $d_z^{\text{sig-trk}_1}$: The longitudinal impact parameter of the leading isolation track.
- p_T weighted average of ΔR between the τ_h^{calo} and the following:
  - The ECAL energy deposits in the signal cone ($\Delta R < 0.15$ w.r.t. the leading signal track).
  - The ECAL energy deposits in the isolation annulus ($0.15 < \Delta R < 0.5$ w.r.t. the leading signal track).



τ_h^{calo} vs. τ_h^{HPS}



Summary

- The relative performances of the HPS and calo-τ algorithms have been presented.
- The calo-tau algorithm performs better in the high efficiency region ($> 70\%$), compared to the standard HPS algorithm.
- Can be useful for high mass searches where the backgrounds are low, and hence one can use higher efficiency working points.

References

- [arXiv:1510.07488] Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS.
- [arXiv:1809.02816] Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV.