Tau Identification at CMS in LHC Run–2

Somnath Choudhury
(for the CMS collaboration)

21st Particles and Nuclei International Conference
Beijing (China), 1 – 5 September 2017

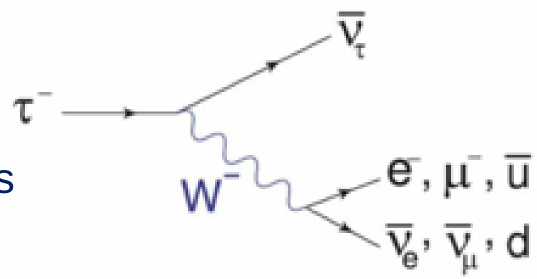
Hadronic Tau Decay

τ is the only lepton that decays to hadrons

Mass m_τ = 1.78 GeV
Lifetime τ = 290 x 10^{-15} s
$c\tau$ = 87 μm

Tau Decay Signature

- ~65% of tau decays
- 1 or 3 π^+
- 0, 1, or 2 π^0
- Via ρ or a_1 decay



Challenges

- Reject huge jet → τ_h background
- Reject e →τ_h fakes
- Reject μ →τ_h fakes(relatively easier)
- τ_h candidates are collimated:
  - A few overlapping $\pi^\pm$ and γ from π^0 decays
- Particle Flow used to resolve objects

Decay Mode	Resonance	BR [%]
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
$\tau^- \rightarrow \pi^- \nu_\tau$	$\pi(140)$	11.6
$\tau^- \rightarrow \pi^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \rightarrow \pi^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	10.8
$\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \rightarrow \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$		4.8
Other hadronic modes		1.7
All hadronic modes		64.8

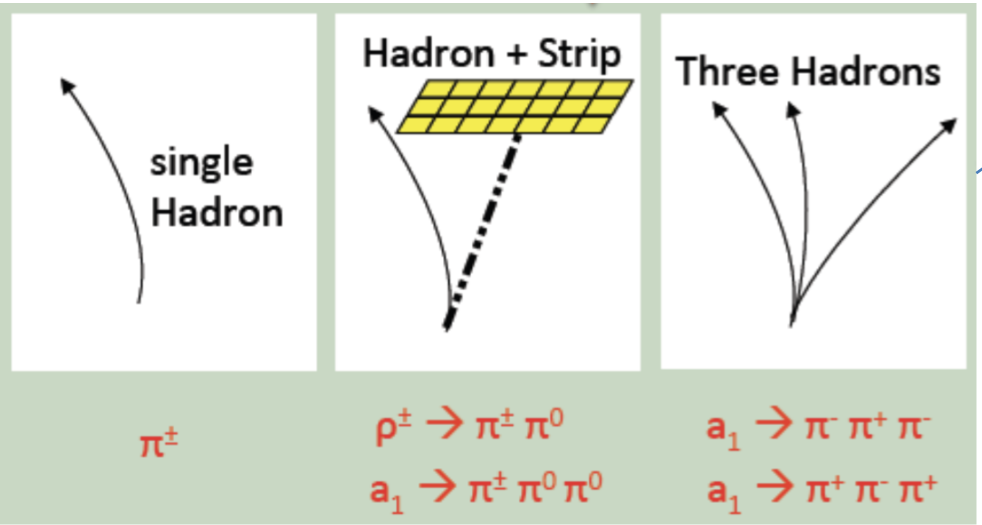
PDG

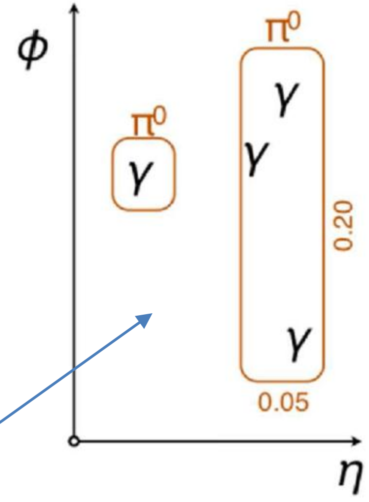
Reconstructed using standard e/μ reconstruction

Reconstruction of $\pi^\pm$, $\rho^\pm$, $a_1^\pm$ signatures

Hadron Plus Strips Algorithm (Run-1)

- Start from an anti-KT (R=0.4) PF jet
- Reconstruct decay modes with one or three charged hadrons, and one or two neutral pions
- Pions reconstructed using an elongated $\eta \times \Phi$ strips collecting energy spread from photon conversions due to magnetic field
- Charged hadrons and photons reconstructed from tracks and calorimeter energy using a particle-flow technique



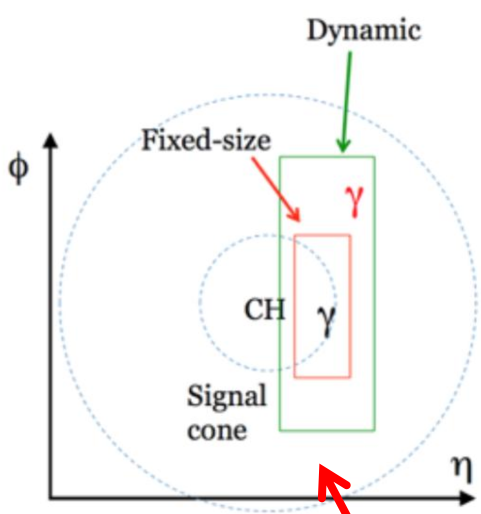


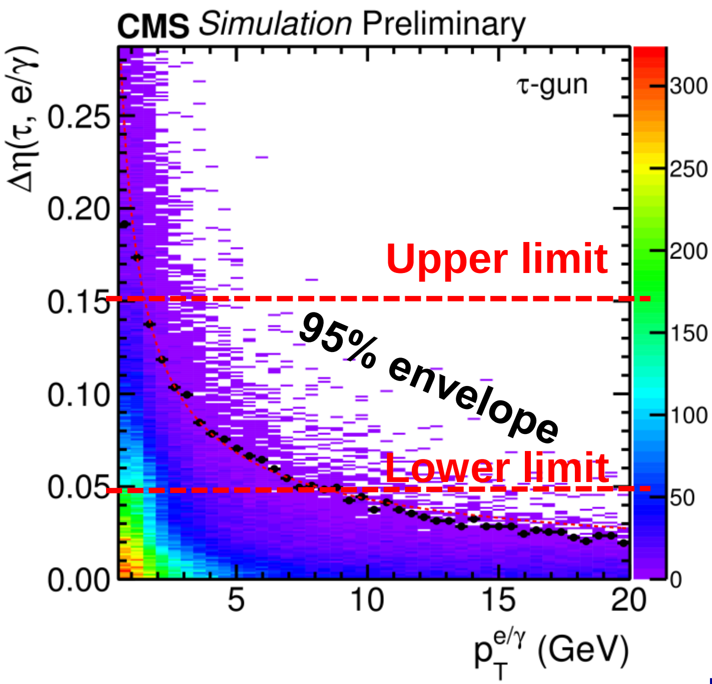
$$\eta = -\ln(\tan(\theta/2))$$

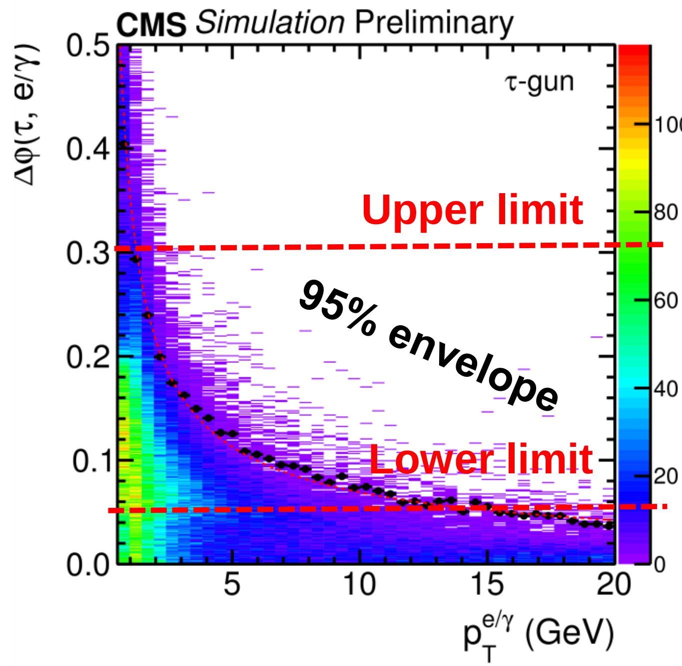
- Signal constituents are required to be in a smaller cone (p_T dependent cone)
- Mass constraints compatible with ρ and a_1 meson mass

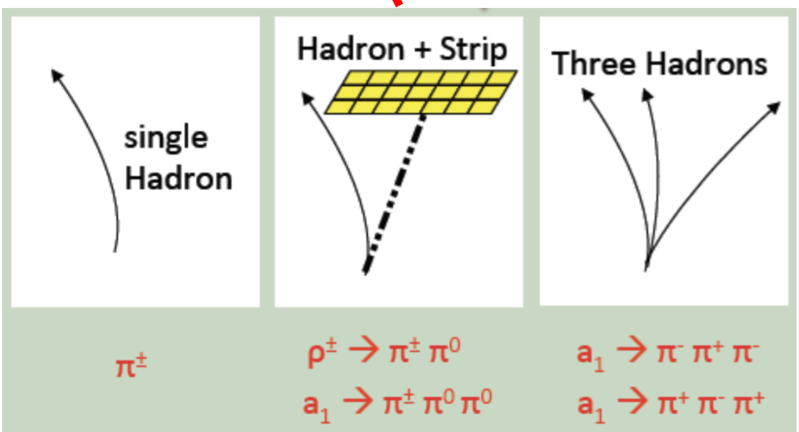
Hadronic Tau Identification

CMS-PAS-TAU-16-002









New for Run-2:

- Dynamic reconstruction of strips size to reconstruct p^0s from e/g candidates
- size of the strip depends on the e/γ p_T

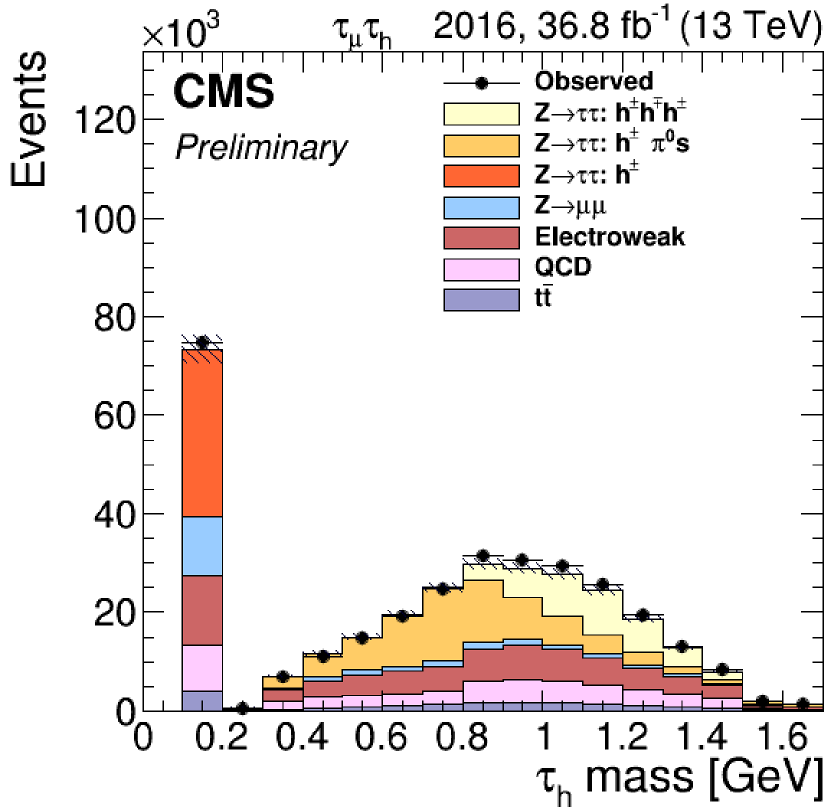
$$\Delta\eta = f(p_T^\gamma) + f(p_T^{\mathrm{strip}}) \qquad f(p_T) = 0.20 \cdot p_T^{-0.66}$$

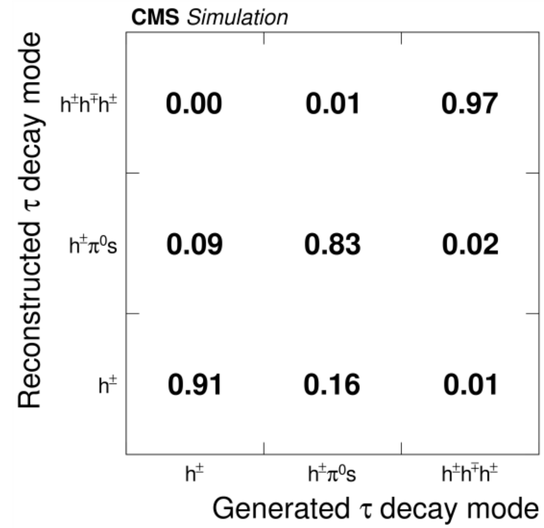
$$\Delta\phi = g(p_T^\gamma) + g(p_T^{\mathrm{strip}}). \qquad g(p_T) = 0.35 \cdot p_T^{-0.71}.$$

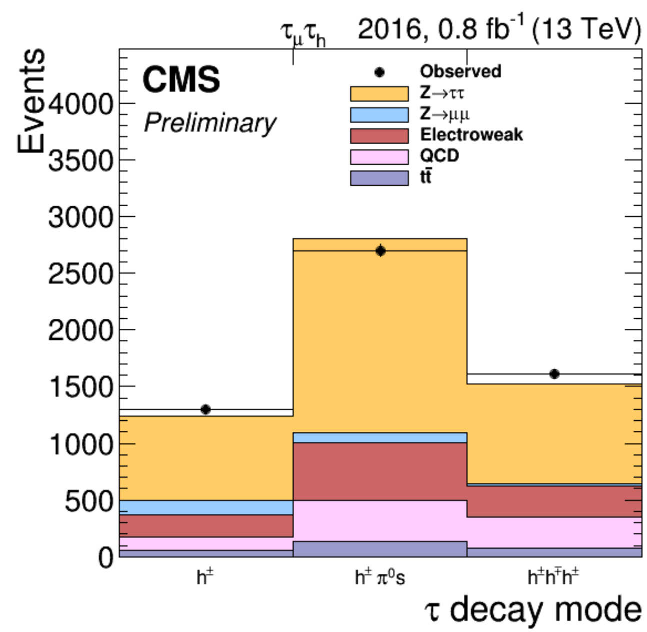
Max (Δη)= 0.15, Max(ΔΦ)= 0.3, Min(Δη)=Min(ΔΦ)=0.05

Tau Identification Performance

CMS DP -2017/002 , CMS DP -2016/015







- Purity of tau decay mode reconstruction is 80-90%
- Data are in well agreement with the expectations

Multivariate Tau Isolation

Use variables sensitive to **tau lifetime, in addition to the isolations**

BDT for MVA training (τ as signal & jets as background)

Kinematic Variables:

- $P_T(\tau)$
- $\eta(\tau)$
- Reconstructed tau decay mode

Cut-based Isolation:

- P_T (charged hadrons)
- P_T (photons)
- Pileup correction ($\Delta\beta$)
- p_T of photons in strips outside signal cone

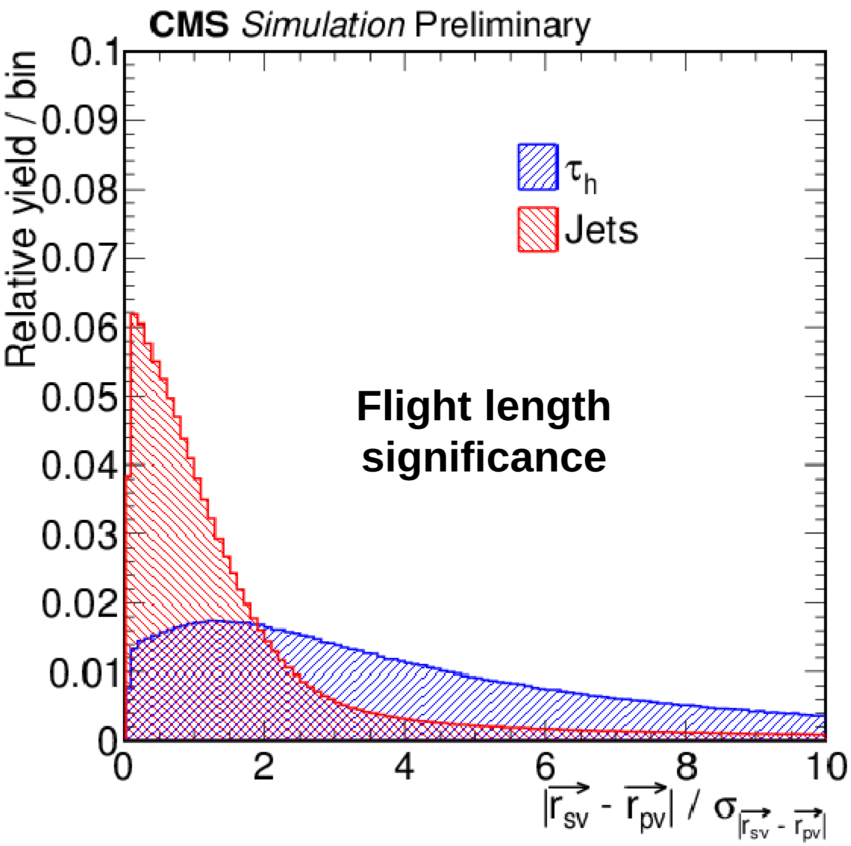
Tau lifetime variables:

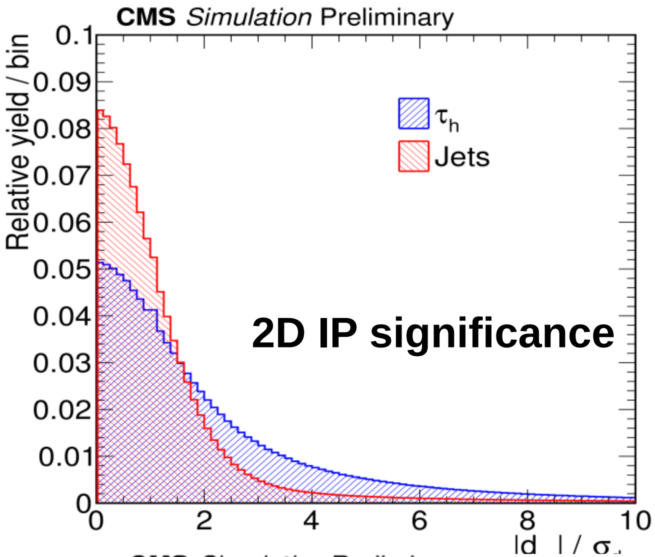
- Signed 2d and 3d impact parameter of the leading track and its significance
- Presence of secondary vertex
- τ flight length
- τ flight significance

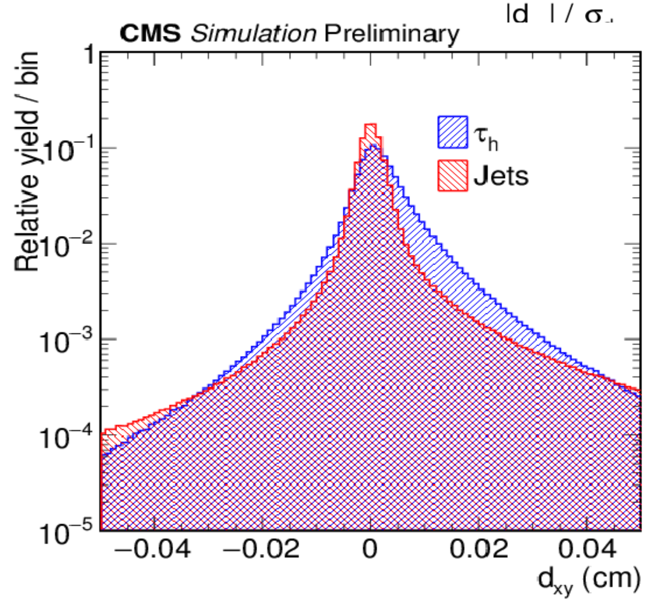
- Additional particle-flow **photon variables** within signal and isolation cones

Signal and background events re-weighted to have similar p_T and η distributions

MVA Tau ID Variables

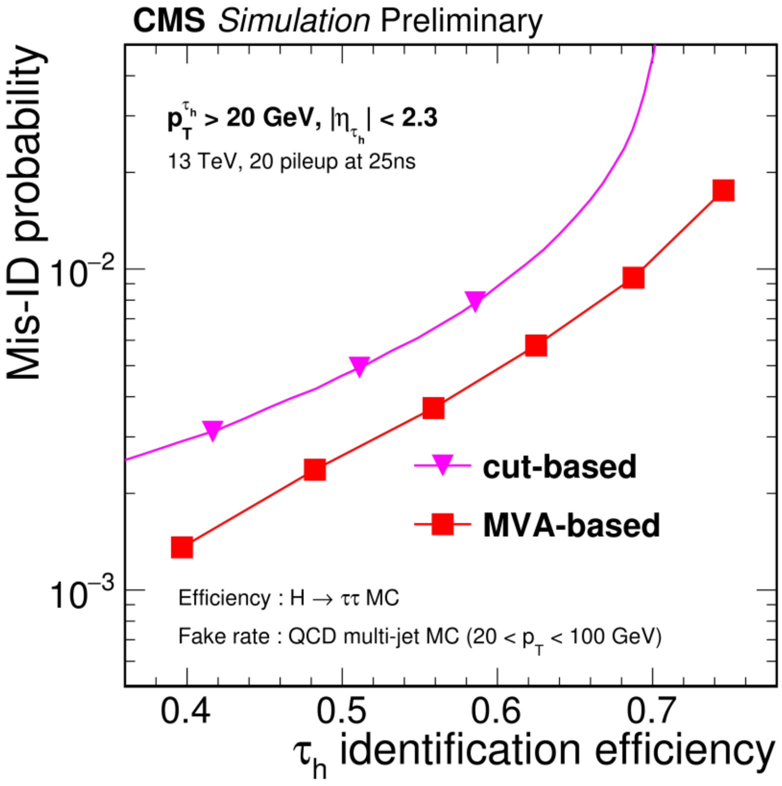


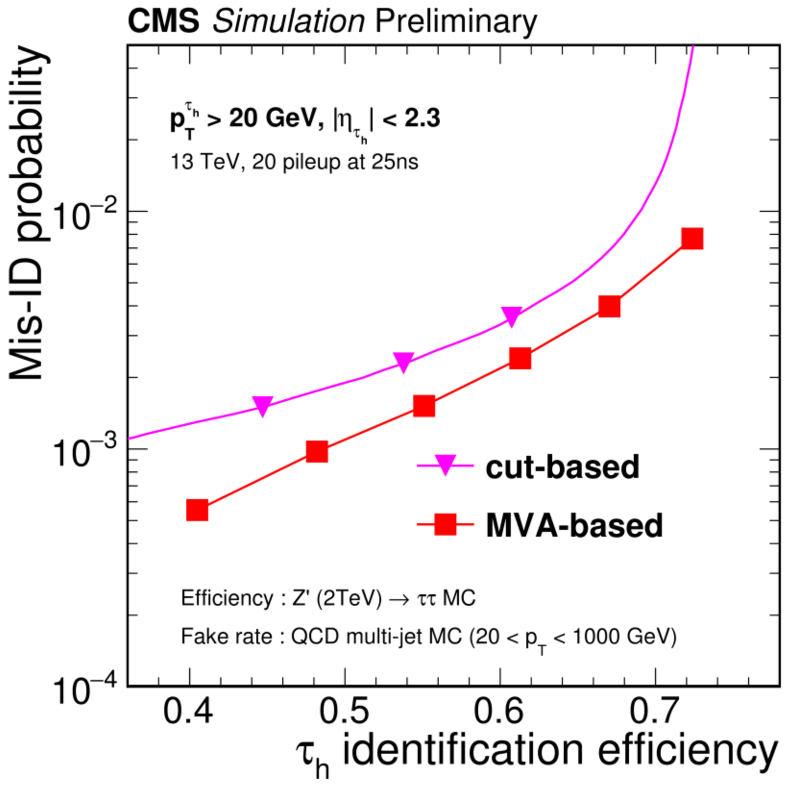




Tau MVA Isolation Performance

CMS-PAS-TAU-16-002



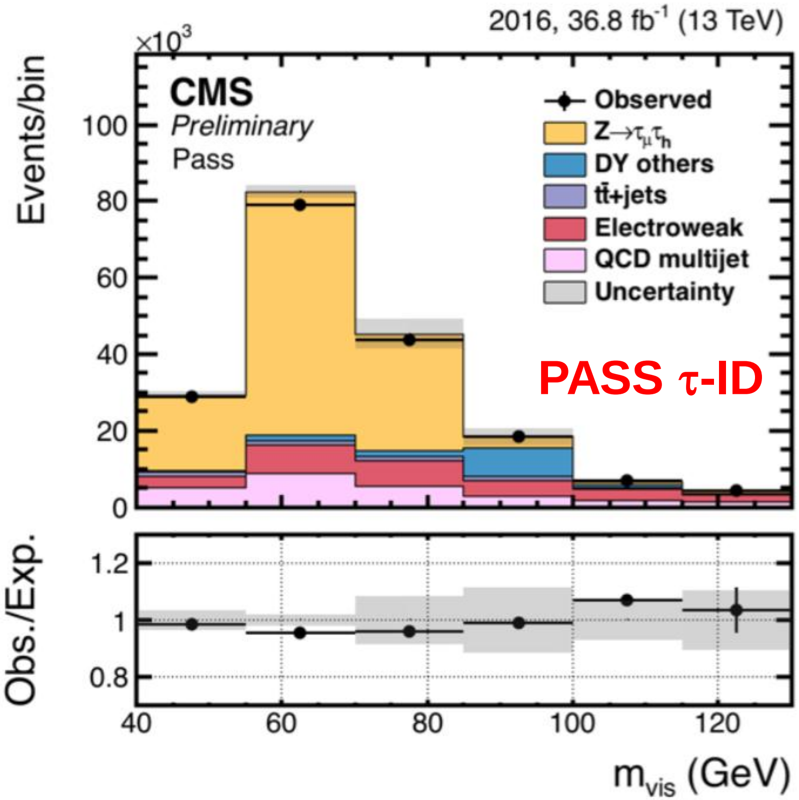


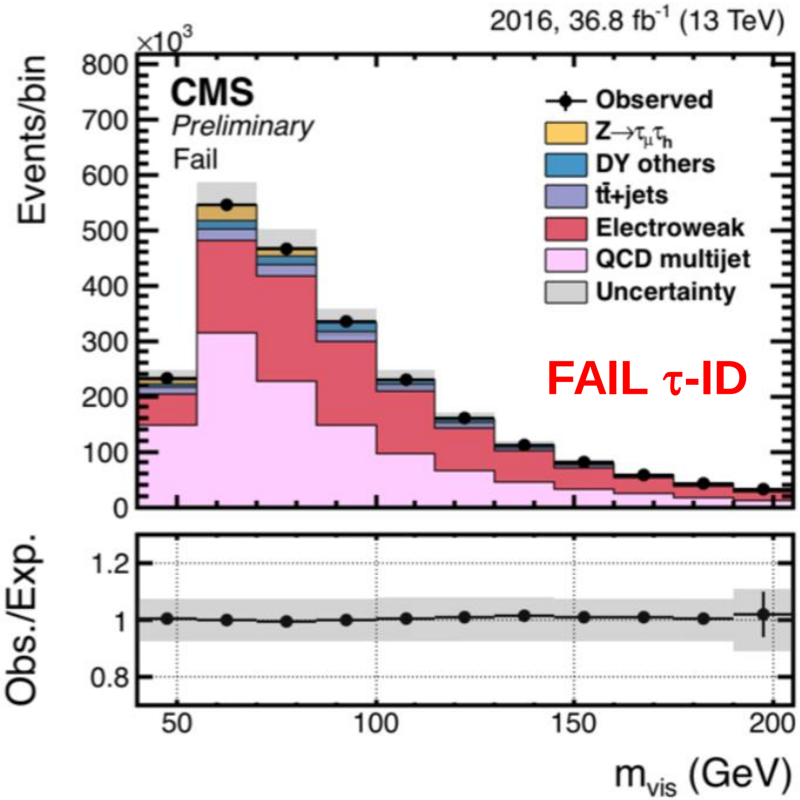
Dynamic strip algorithm with MVA isolation

Factor of ~2 reduction in fakes compared to cut-based

Visible mass of $\tau_\mu \tau_h$

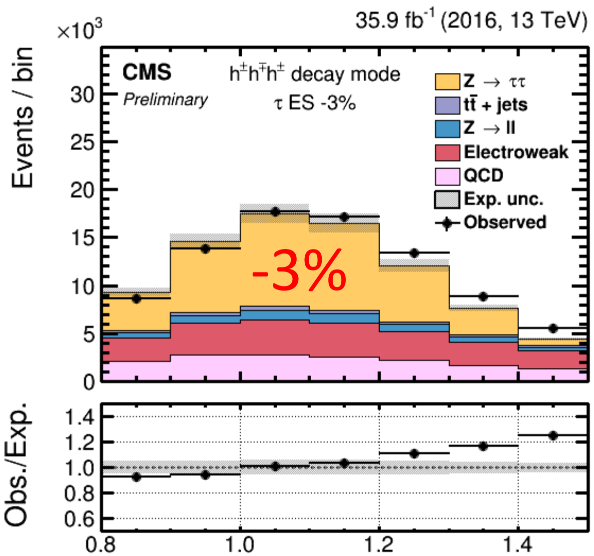
CMS DP -2017/006

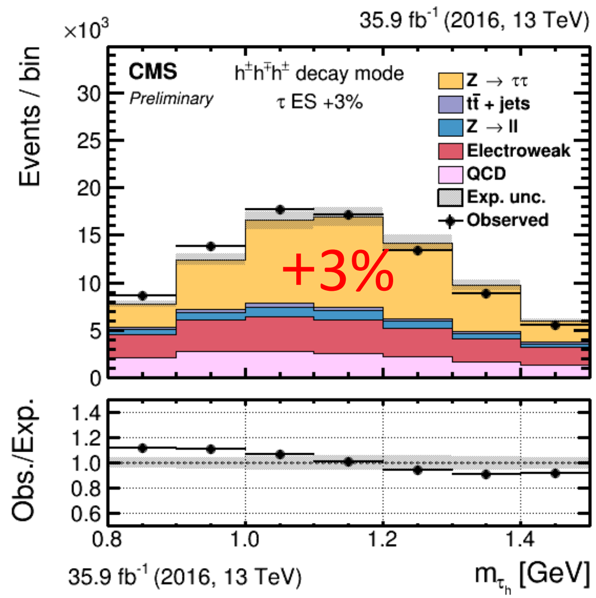


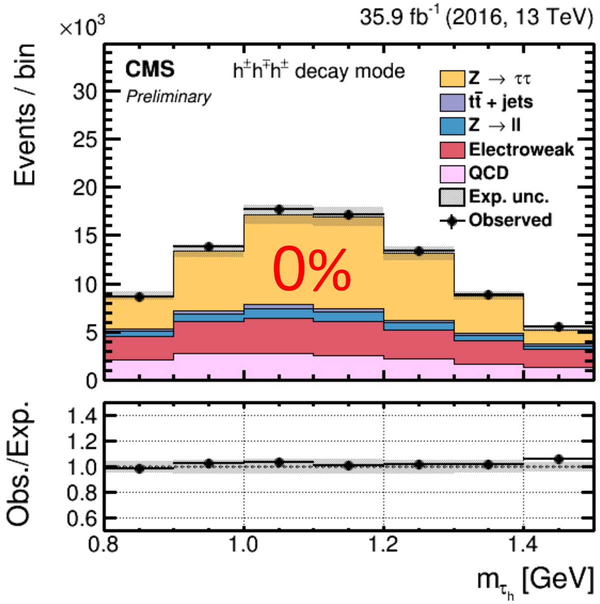


- Tau ID efficiency measured from $Z \rightarrow \tau\tau \rightarrow \tau_\mu \tau_h$ events using a **Tag** (μ) & **Probe** (τ_h) method
- Data/MC scale factor consistent with unity

τ_h Energy Scale







CMS DP -2017/006

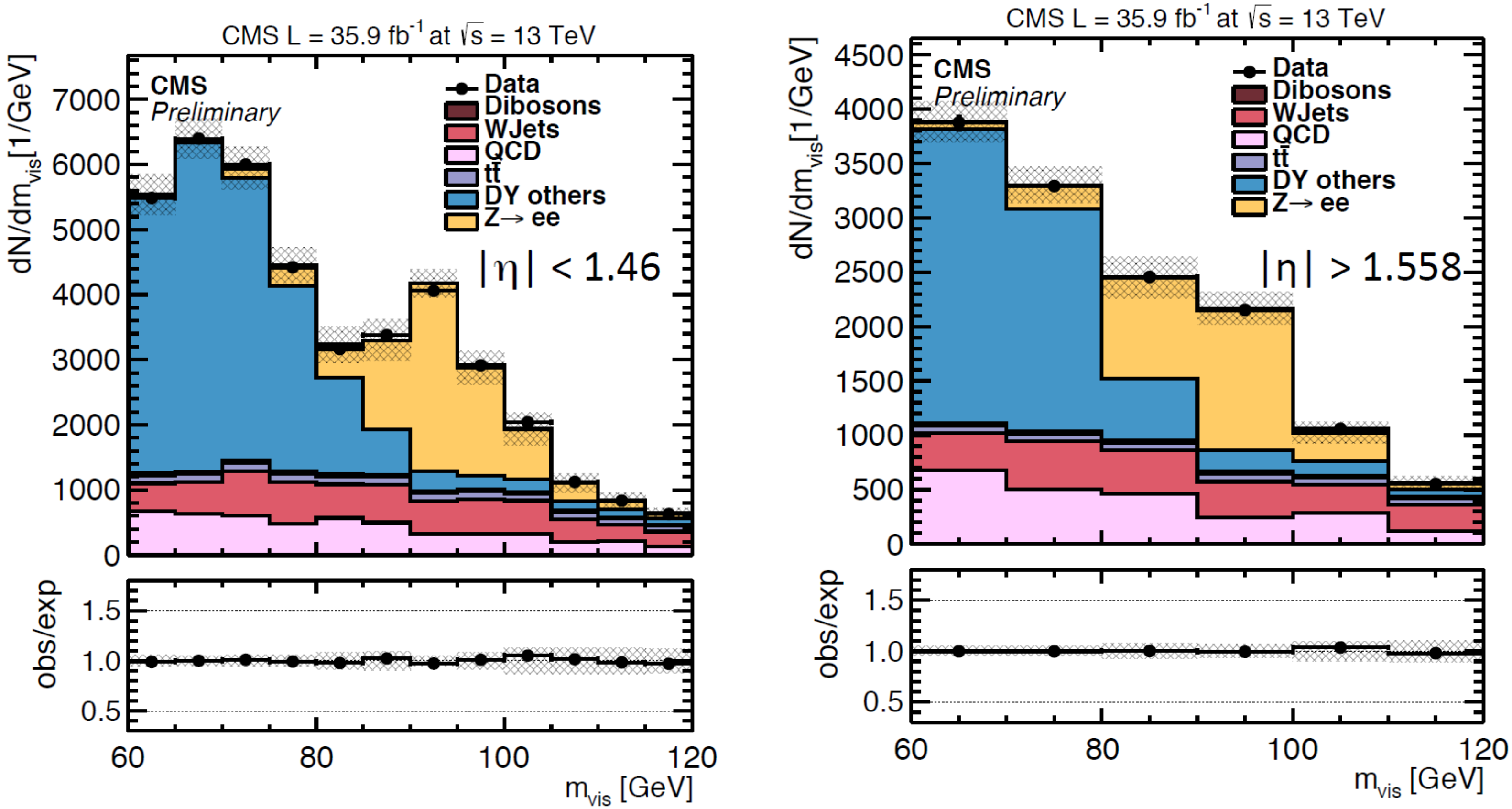
τ_h energy scale measured by fitting **$m_{vis}(\mu\tau_h)$** and **$m_{vis}(\tau_h)$** distributions

Decay mode	Tau energy scale [%]
1-prong	+0.0 ± 1.1
1-prongπ^0	+1.0 ± 0.4
3-prong	-0.1 ± 0.2

$e \to \tau_h$ Misidentification

$e \to \tau_h$ misidentification measured from Z → ee events

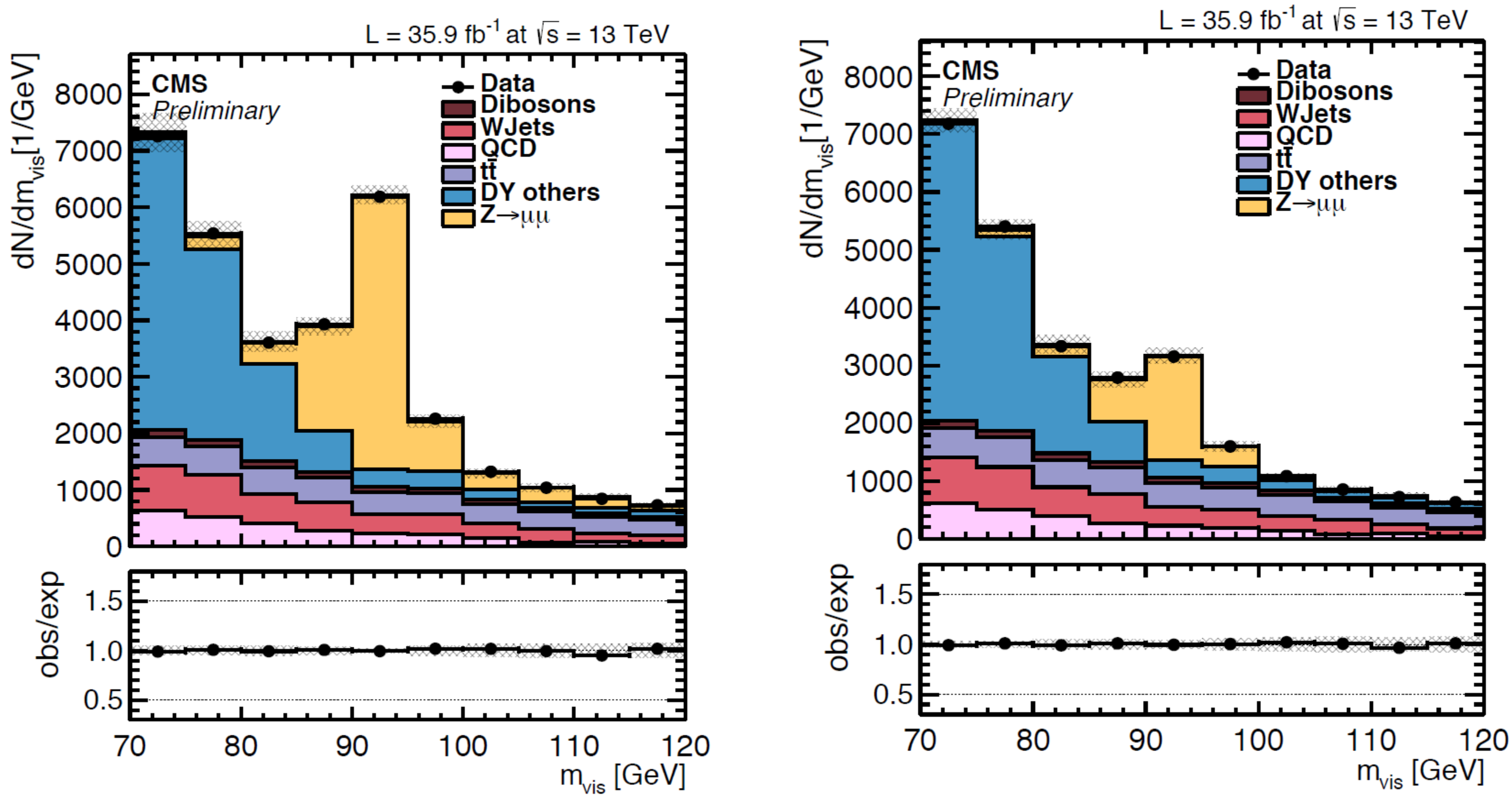
(where probe electron is reconstructed as τ_h and passes MVA anti-electron discriminator)



The Data/MC scale factors for different working points of the discriminator are about 1.3 to 1.6 with 10 to 20% uncertainty

$\mu \to \tau_h$ Misidentification

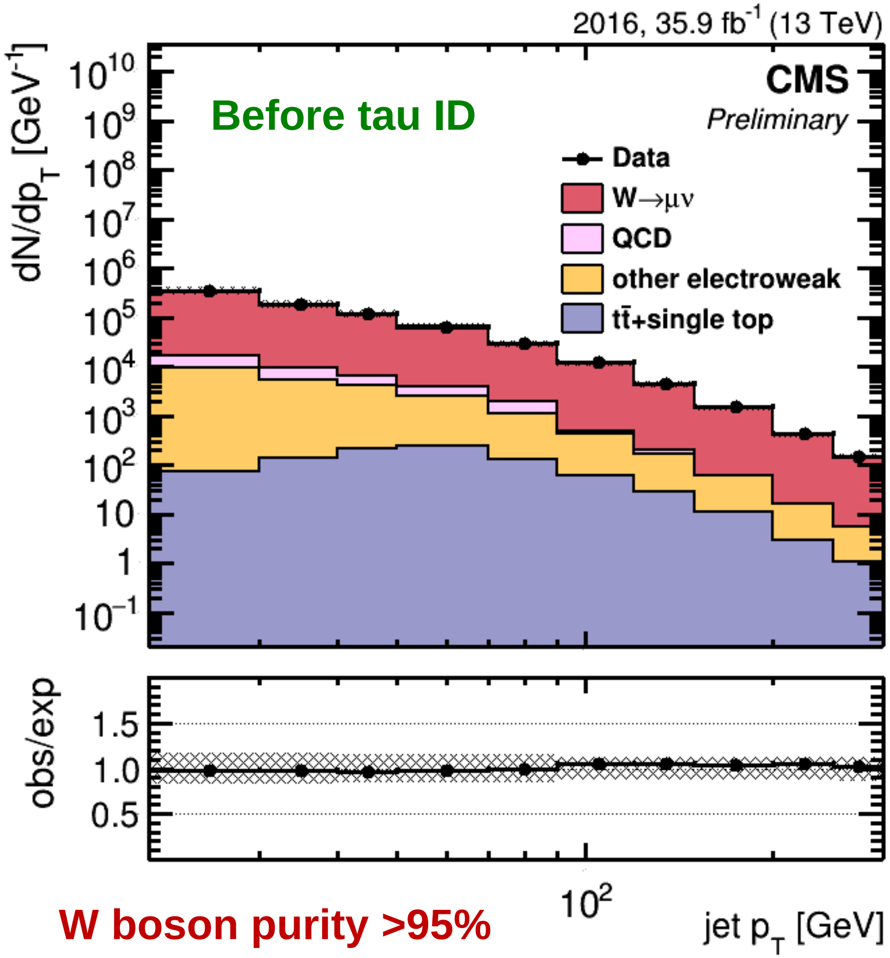
- $\mu \to \tau_h$ misidentification measured from Z →μμ events

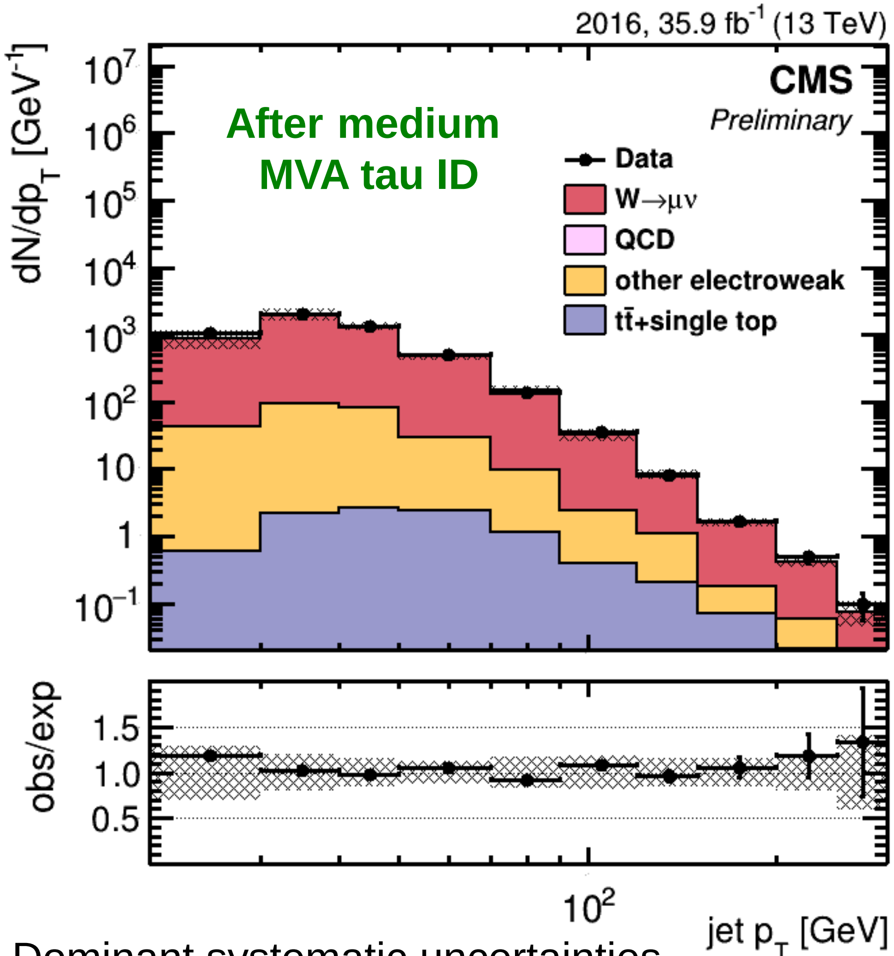


- ➢ visible mass distribution of $\mu\tau_h$ pair after maximum likelihood fit with Z→μμ event selection
- ➢ the probe muon is reconstructed as τ_h and passes the loose (left) and tight (right) working points of anti-muon discriminator.

Jet→τ_h Misidentification in W+Jet Events

CMS DP-2017/036



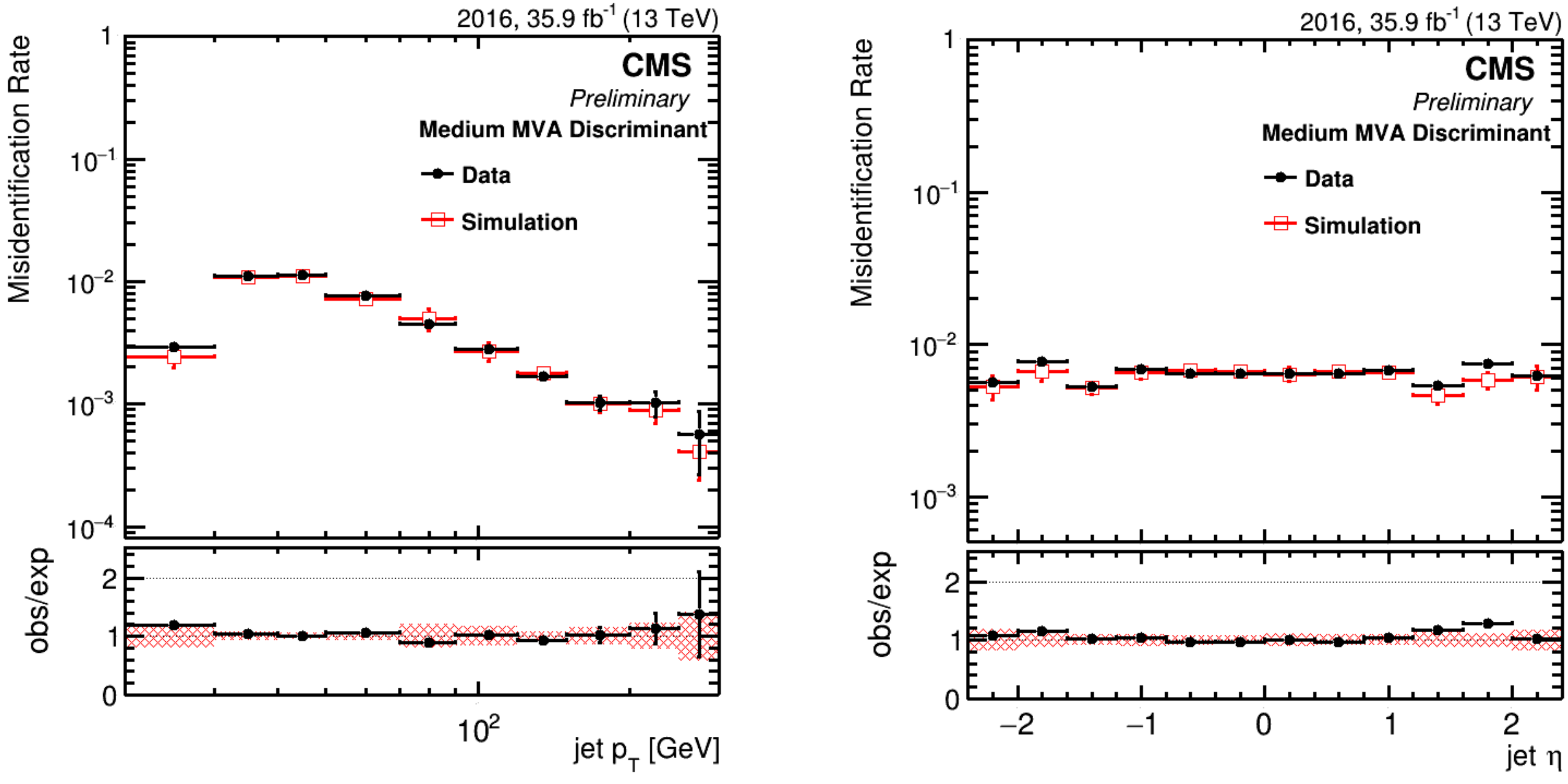


W boson purity >95%

Error band: systematic uncertainties + MC statistical uncertainties

Dominant systematic uncertainties
→ JES
→ unclustered ES

CMS DP-2017/036



Misidentification rate of jets to taus versus p_T and η for W+jet events after MVA medium tau identification discriminator: **average rate 0.7%** covering wide jet p_T range 20 GeV - 300 GeV

τ_h in Boosted Regime

- Start from a large CA8 jet (Cambridge-Aachen R=0.8)

- Use subjet(sj) finding algorithm and require 2 subjets:

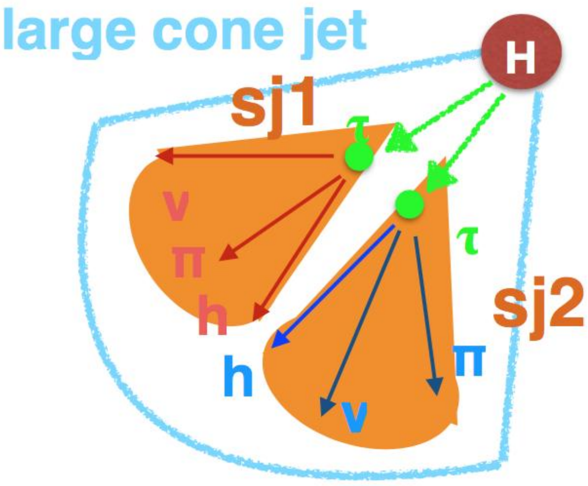
  p_T(sj1,sj2) > 10 GeV

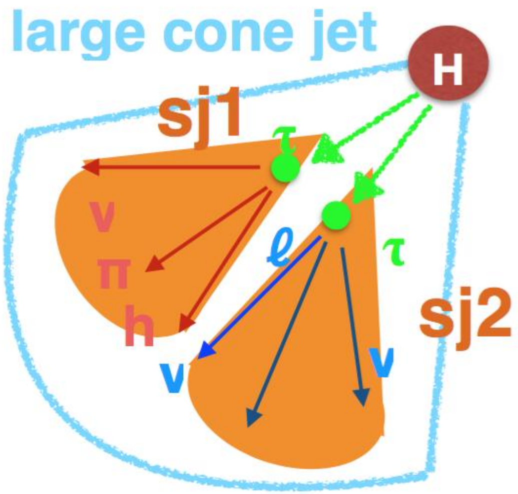
  Max(mass(sj1),mass(sj2))/mass(jet)< 0.667

- In semi-leptonic final state the lepton is considered a subjet at this stage

- Use subjets as seeds for Tau reconstruction

- Then the tau reconstruction proceeds using the standard HPS algorithm

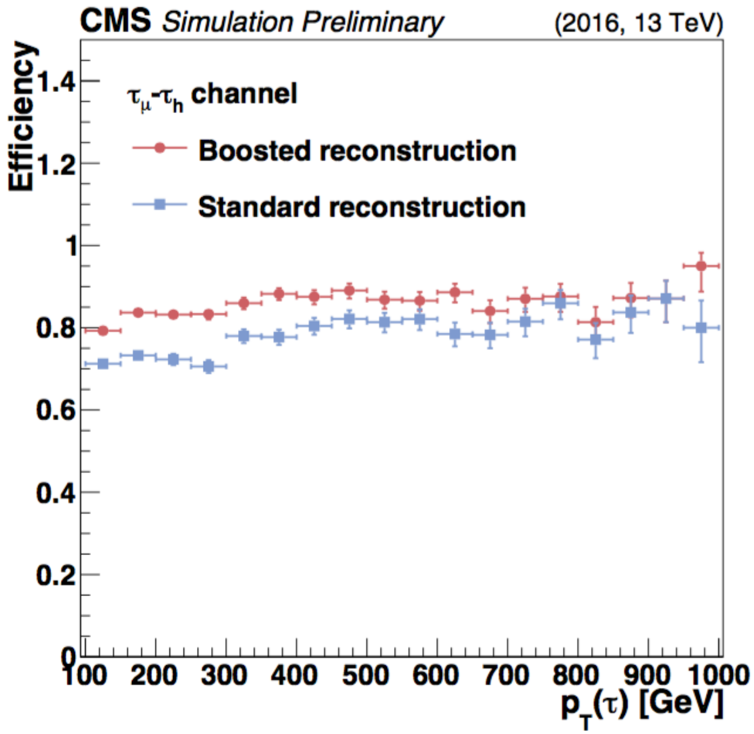


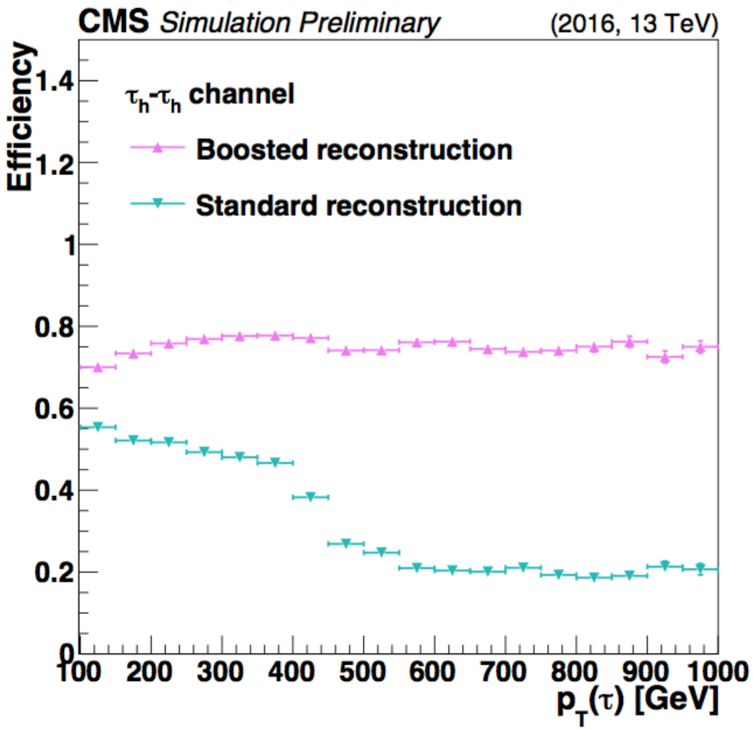


Boosted Tau ID Performance

tau reconstruction efficiency vs tau p_T

Tau $|\eta| < 2.3$ and $p_T > 20$, Loose Isolation

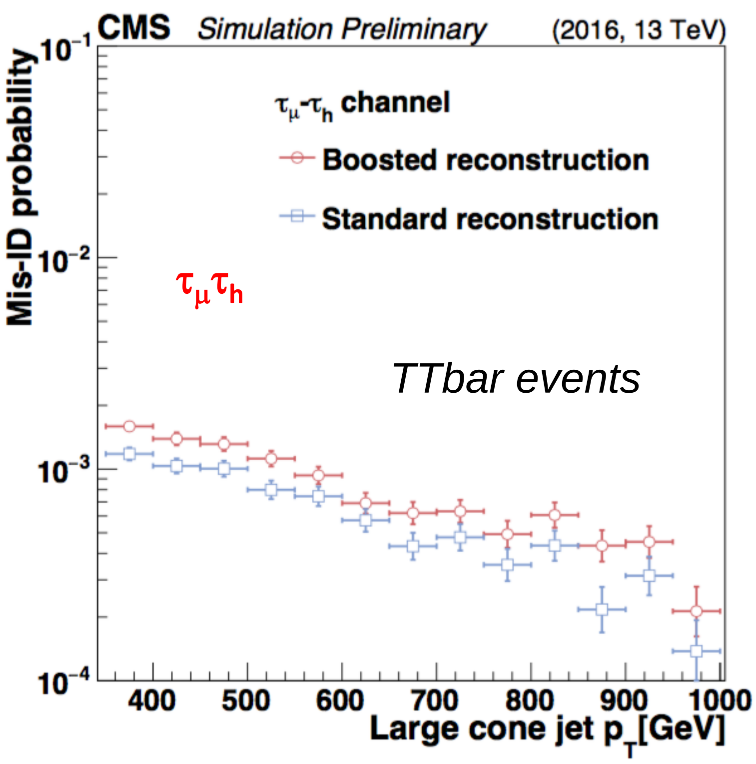


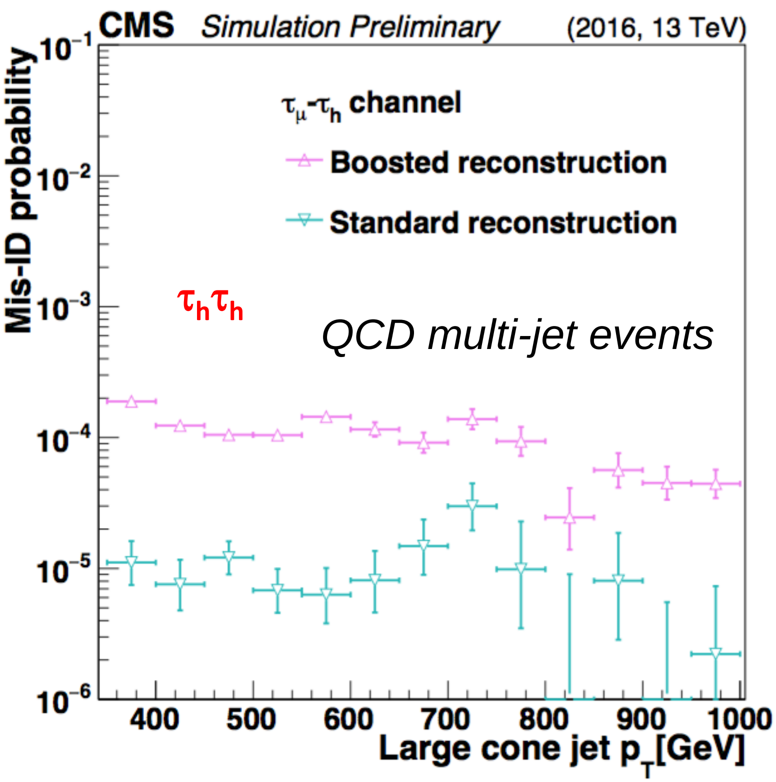


Major improvement in fully hadronic channel

Boosted Tau ID Performance

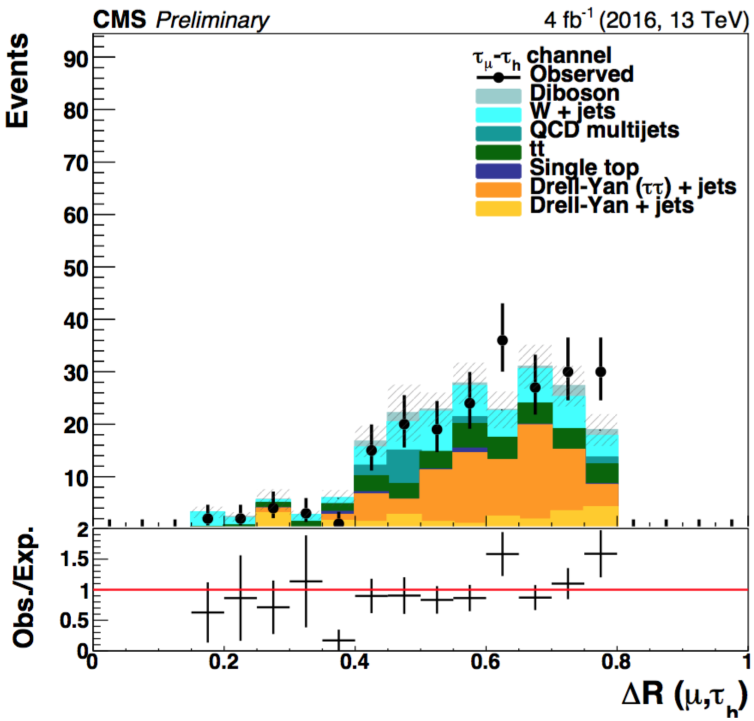
Misidentification Probability vs large Cone Jet p_T



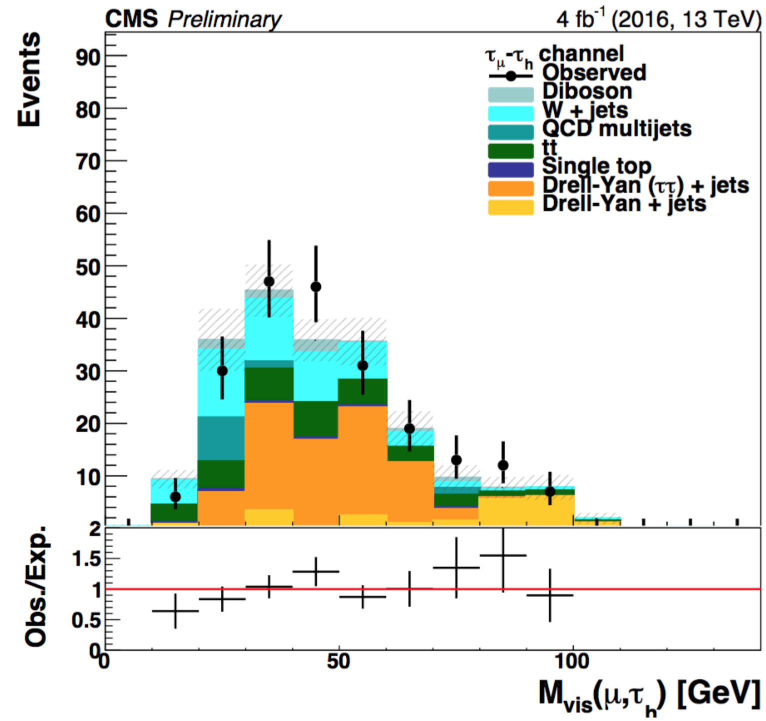


The fake probability increases significantly.
However, the background contributions at such high p_T is smaller

High p_T $Z \rightarrow \tau\tau \rightarrow \tau_\mu \tau_h$ events
(Tight muon selection and Loose MVA isolation for τ_h)



Delta R between the lepton and the hadronic τ



Mass of the visible products of the $\tau\tau$ system

Tau Identification at Trigger

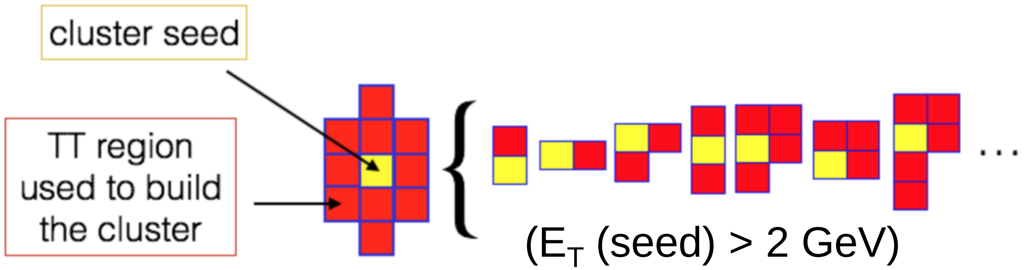
- Tau identification at the trigger level is constrained by timing as well as rates

- **Tau ID at Level-1 Trigger (Electronics)**
  - No possibility of using tracker detector
  - A simpler algorithm developed using energy deposits in the trigger towers (ECAL + HCAL towers)

- **Tau ID at High Level Trigger (Computing Farm)**
  - Use a simplified version of offline algorithms to increase efficiency and meet timing constraints
  - A simple cone based algorithm employed at HLT
    - Based on particle-flow with regional tracking

Improved algorithm in run-2 compared to run-1

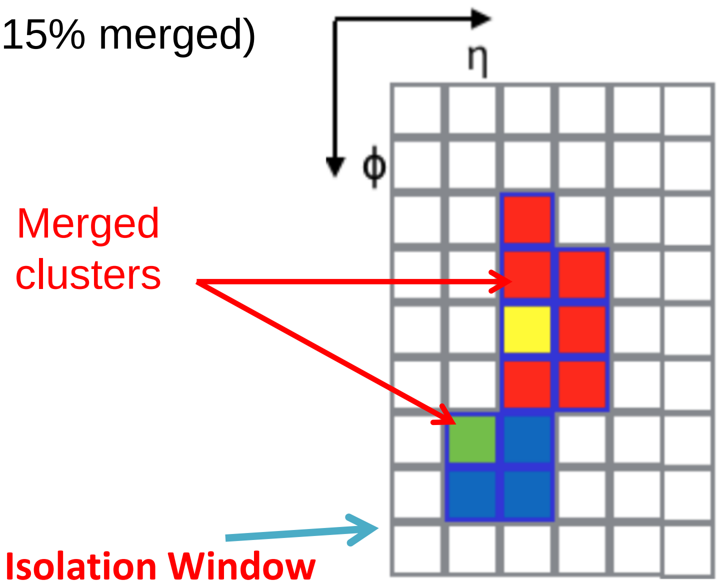
- **Clustering:** Create tau clusters from Trigger Towers



- **Merging:** Search for neighbours in a defined path (~15% merged) (tau decay products can be spread out)

- **Calibration:** As function of E_T, eta, merging, and presence of ECAL deposits, also on tower by tower basis using charged and neutral pions

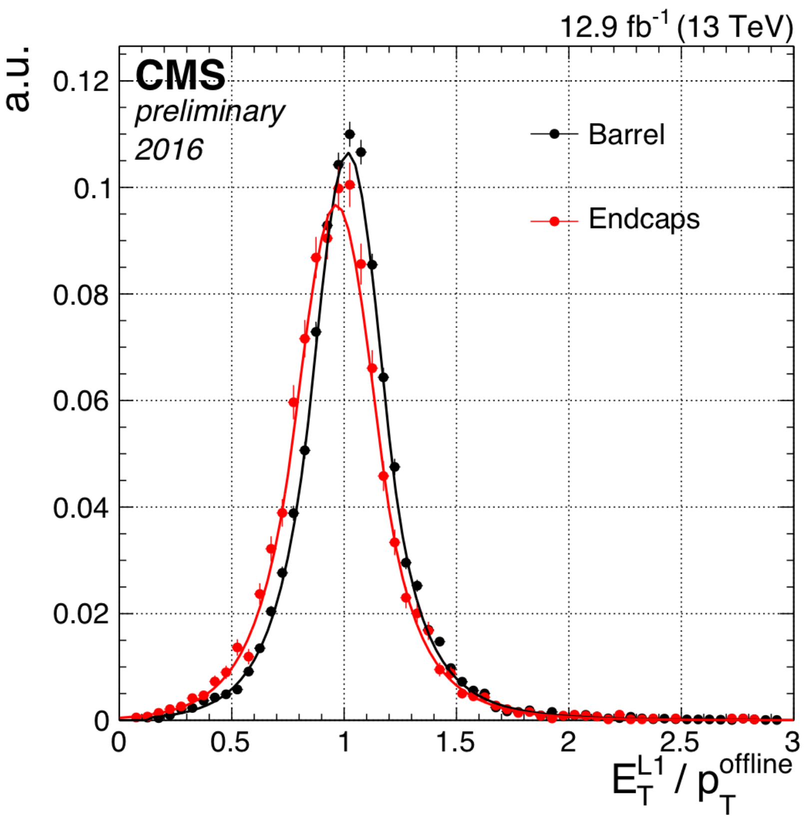
- **Isolation:** Computed as E_T(iso) = E_T(6x9) – E_T(tau) Cut on E_T(iso) depends on p_T, |η|, and pileup

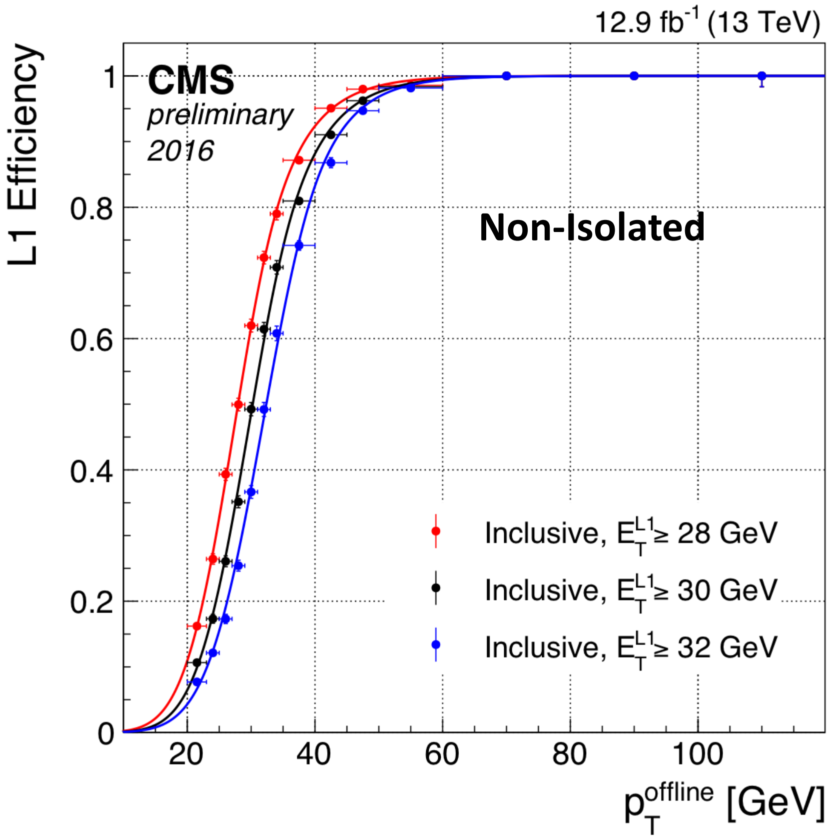


Level-1 Trigger Tau Performance

- Very good E_T response and resolution, thanks to in-situ calibration of L1 tau

re-designing of the L1 tau trigger for Run-2 helped to keep di-tau trigger thresholds at ~30 to 35 GeV

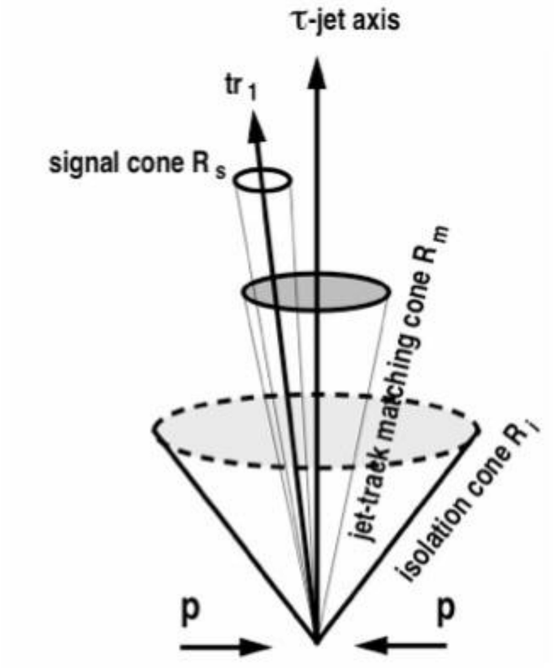




Tau ID at HLT

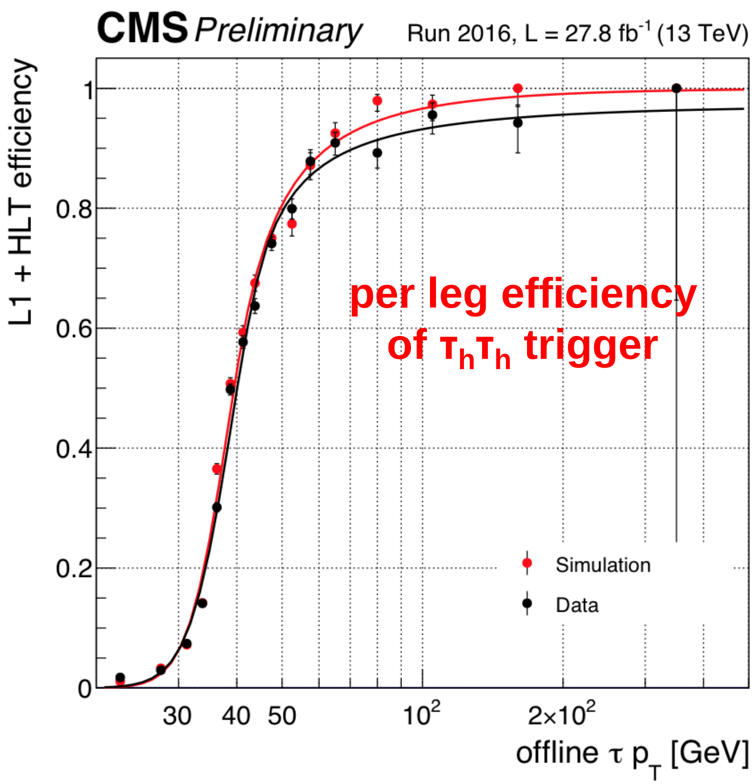
Level-1 Tau → L2 Calo Taus → L2.5 (regional pixel tracking) → L3 Tau

- L2 & L2.5 steps are needed in double-hadronic tau paths to reduce rate before PF in run at HLT
  - Needed to control timing
- Build L2 calo tau-jets seeded by L1 tau candidates
  - Require two calo tau-jets with $p_T > 26$ GeV & $|\eta| < 2.2$
- L2.5:
  - Regional pixel tracking around the calo taus
  - Use pixel tracks to reconstruct vertices
  - Candidates are required to pass pixel track based isolation
- L3:
  - Particle flow with regional pixel tracking. Regions defined around L2.5 candidates
  - Simple cone based algorithm (leading track finding)
  - Combined (track + photon) isolation

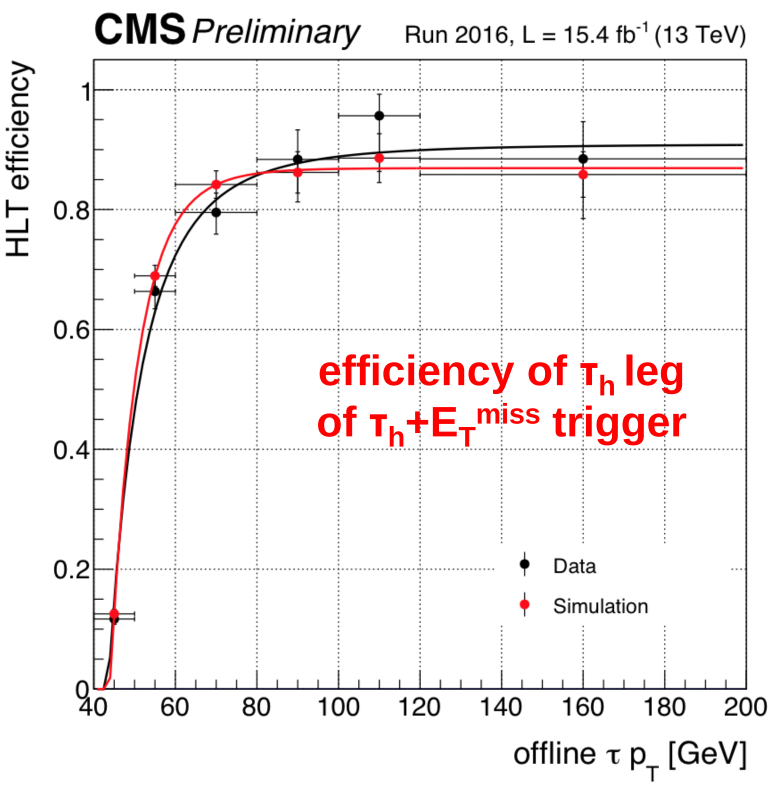


HLT Tau Performance

Per-leg combined L1 and High Level trigger efficiency of the di-τ_h (medium isolation, p_T > 35 GeV, seeded by di-τ Level-1) trigger for $H\to \tau_h\tau_h$ analysis



High Level Trigger efficiency of the τ_h leg of the τ_h + E_T^{miss} (medium isolation, p_T > 50 GeV, seeded by E_T^{miss} Level-1) trigger for the $H^\pm\to \tau_h\nu_\tau$ analysis



Summary

- CMS tau reconstruction algorithm is one of the biggest beneficiary of the particle-flow method
  - PF helps reconstruct individual decay modes => improving significantly the tau identification capability compared to leading track algorithms
  - Furthermore, the MVA based tau isolation significantly improve suppression of the jet to tau fake rate

- There is already a very good effort to identify taus in boosted regime. Efforts are made to validate the method from data (very few events with high p_T Z events)

- The tau algorithm at level-1 trigger re-designed for LHC run-2 (thanks to Phase-1 trigger upgrade) => Able to keep the trigger threshold similar or less than run-1
  - More studies ongoing for further improvement for future data taking