Tau Identification at CMS

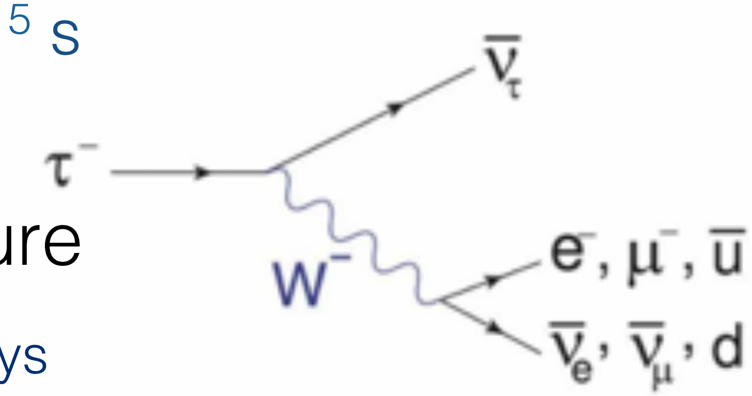
Arun Nayak
Institute of Physics
Bhubaneswar, India

Jets @ LHC Meeting
ICTS - Bengaluru

Hadronic Decay of Tau Lepton

τ is the only lepton that decays to hadrons

Mass m_τ = 1.78 GeV
Lifetime $\tau = 290 \times 10^{-15}$ 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 γs from π^0 decays
- PFlow is used to resolve objects

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

PDG

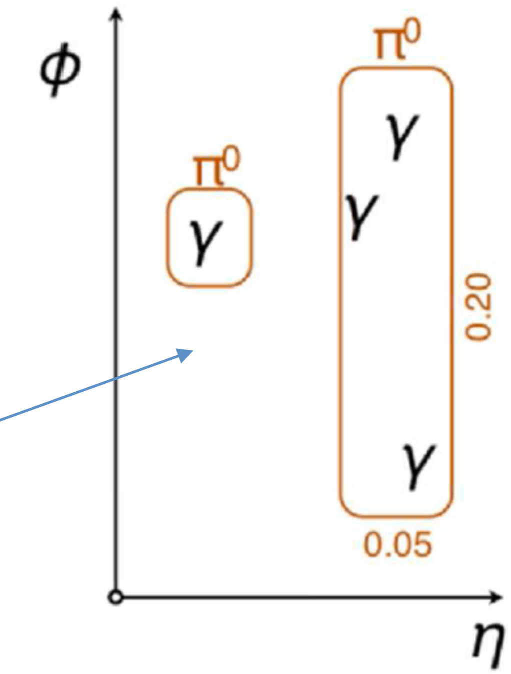
Reconstructed using standard e/μ reconstruction

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

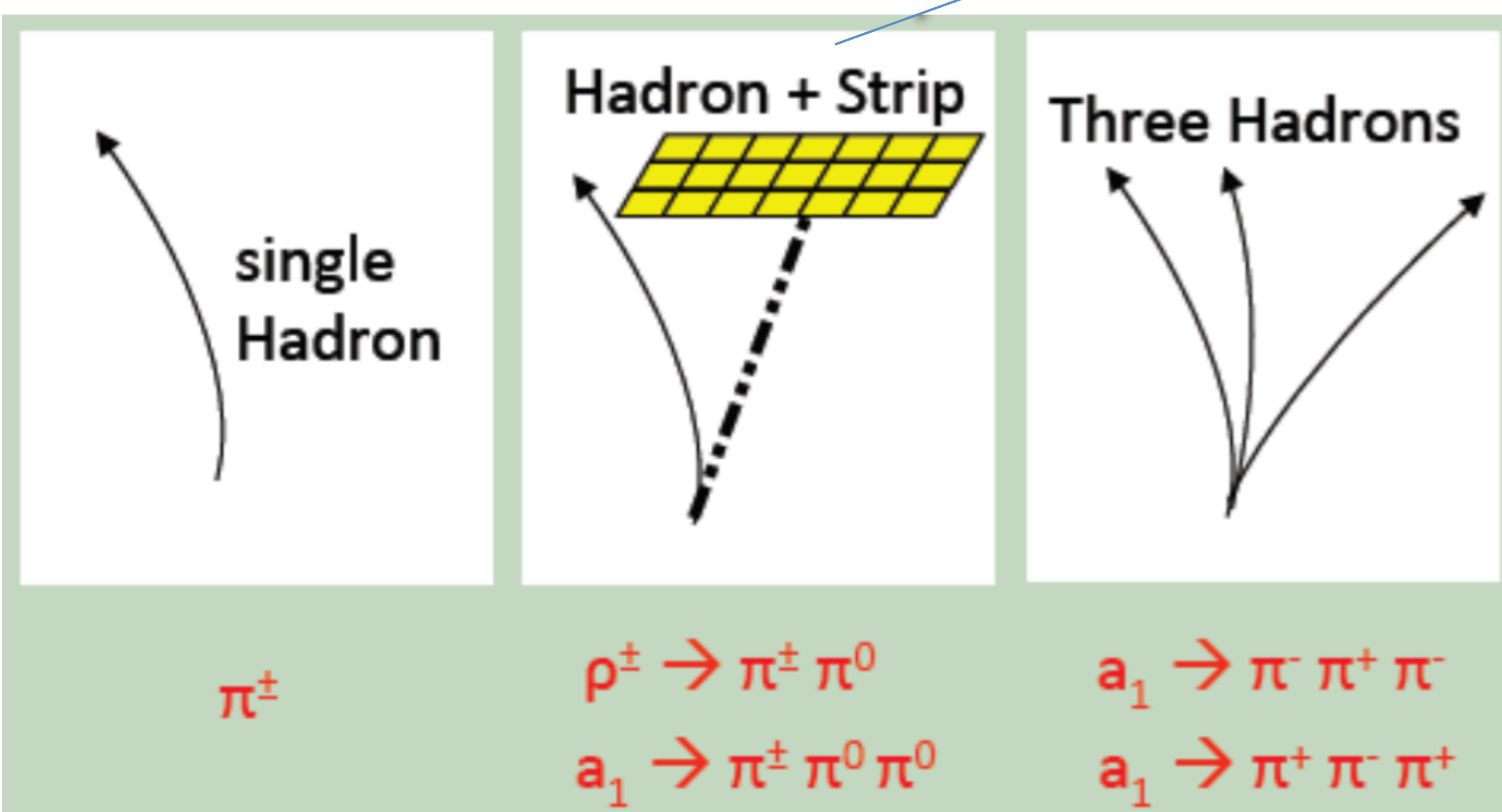
Identification of Hadronic τ decay

Hadron Plus Strips Algorithm (run-1)

- Start from an AntiK$_T$ (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



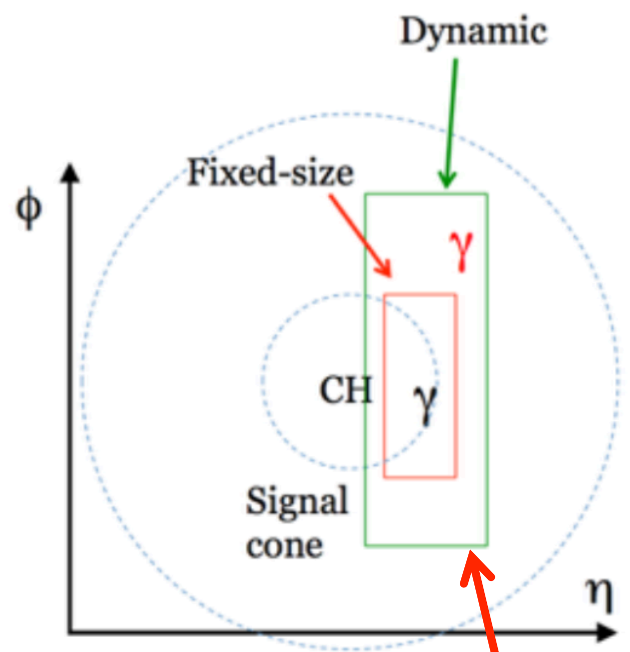
$\eta = -\ln(\tan(\theta/2))$

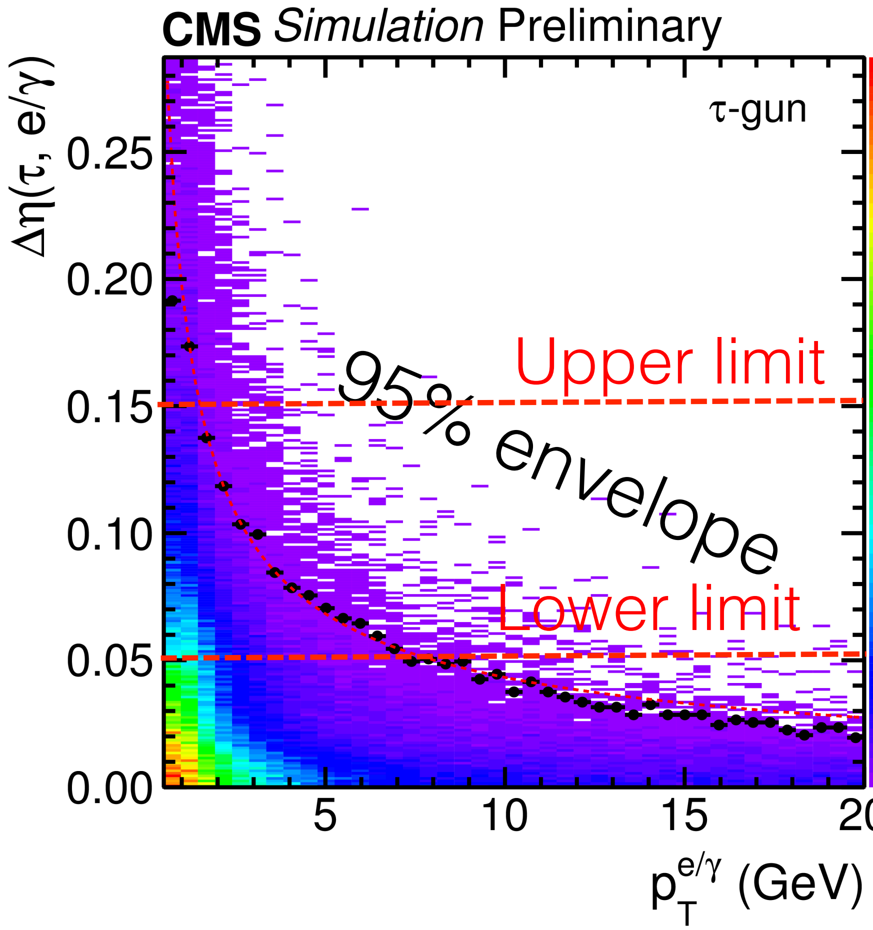


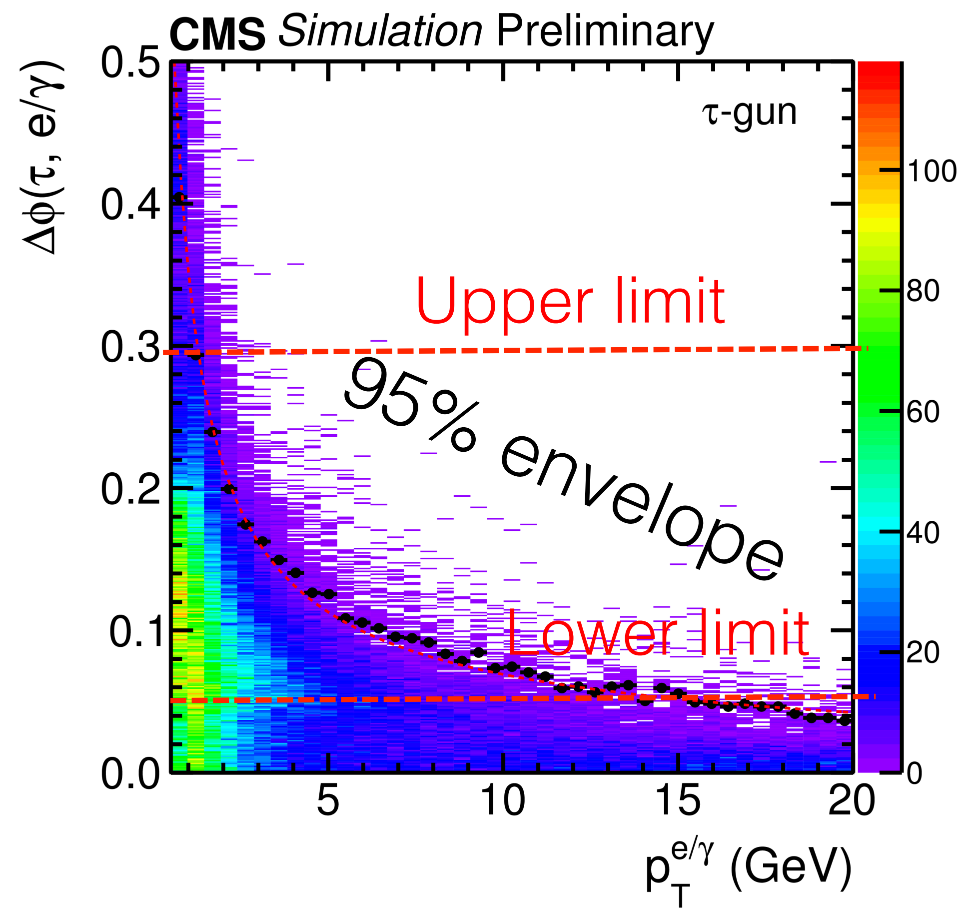
- Signal constituents are required to be in a smaller cone (p_T dependent cone)
- Mass constraints compatible with ρ and a_1 meson mass

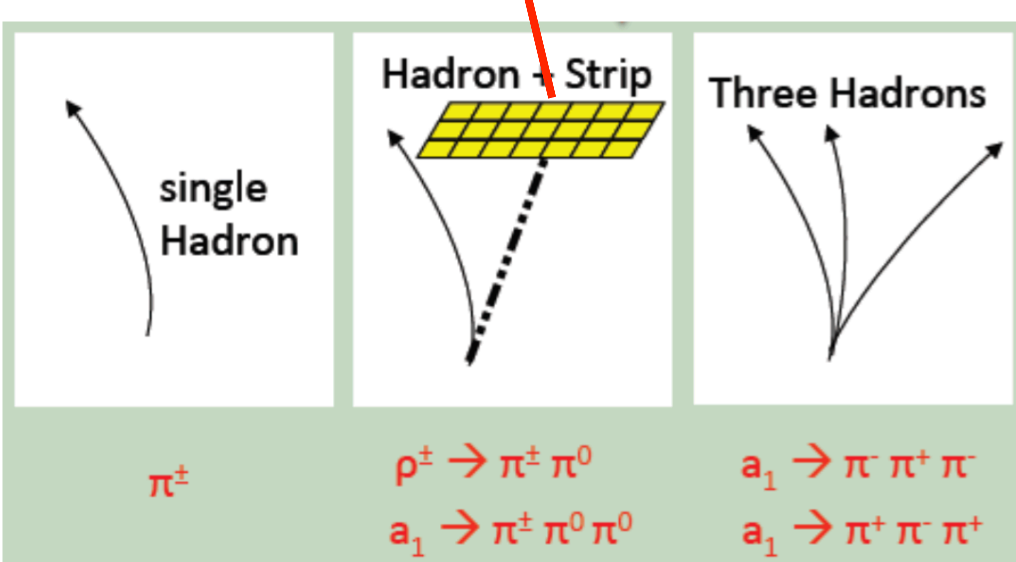
Identification of Hadronic τ decay

CMS-PAS-TAU-16-002









New for Run-II:

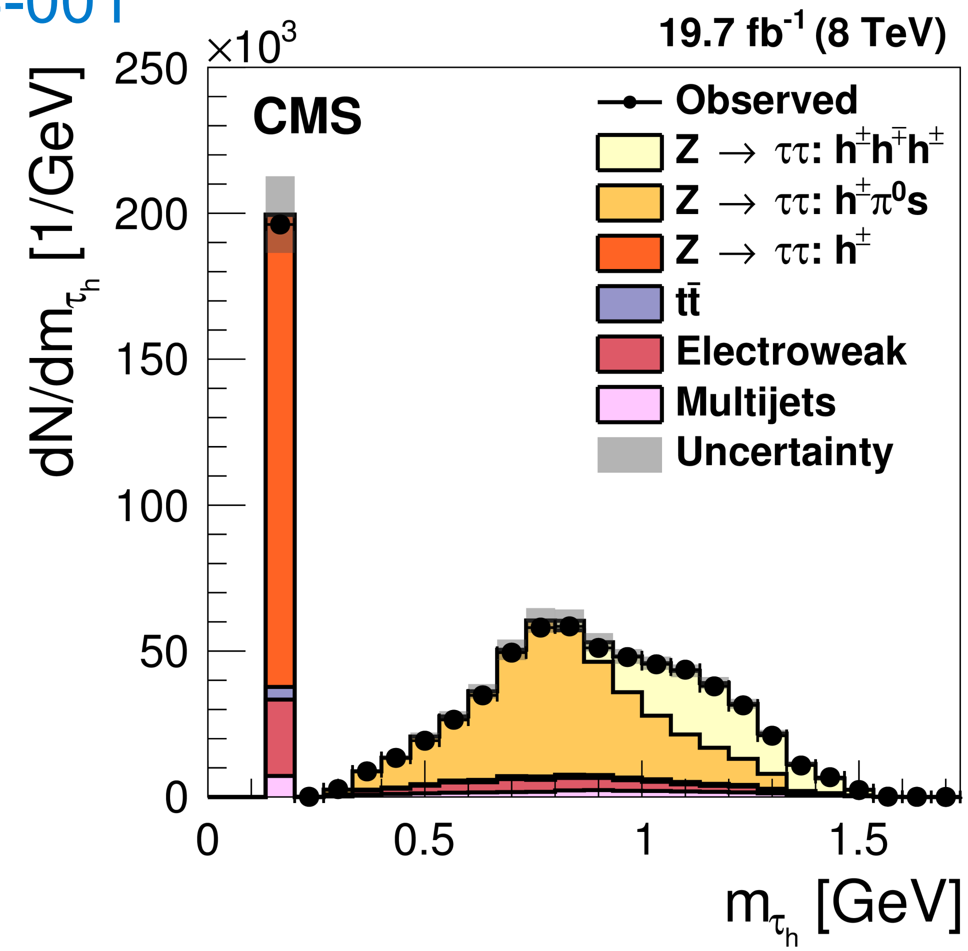
- Dynamic reconstruction of strips size to reconstruct π^0s from e/γ candidates
- size of the strip depends on the e/γ p_T

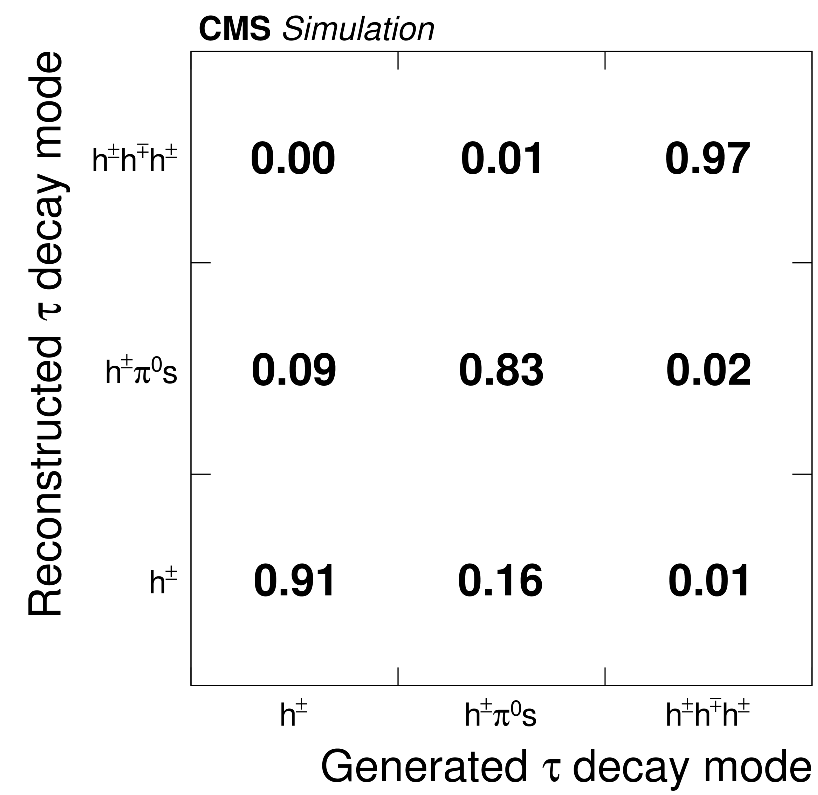
$$
\begin{aligned}
\Delta\eta &= f(p_T^{\gamma}) + f(p_T^{\text{strip}}) & f(p_T) &= 0.20 \cdot p_T^{-0.66} \\
\Delta\phi &= g(p_T^{\gamma}) + g(p_T^{\text{strip}}). & g(p_T) &= 0.35 \cdot p_T^{-0.71}.
\end{aligned}
$$

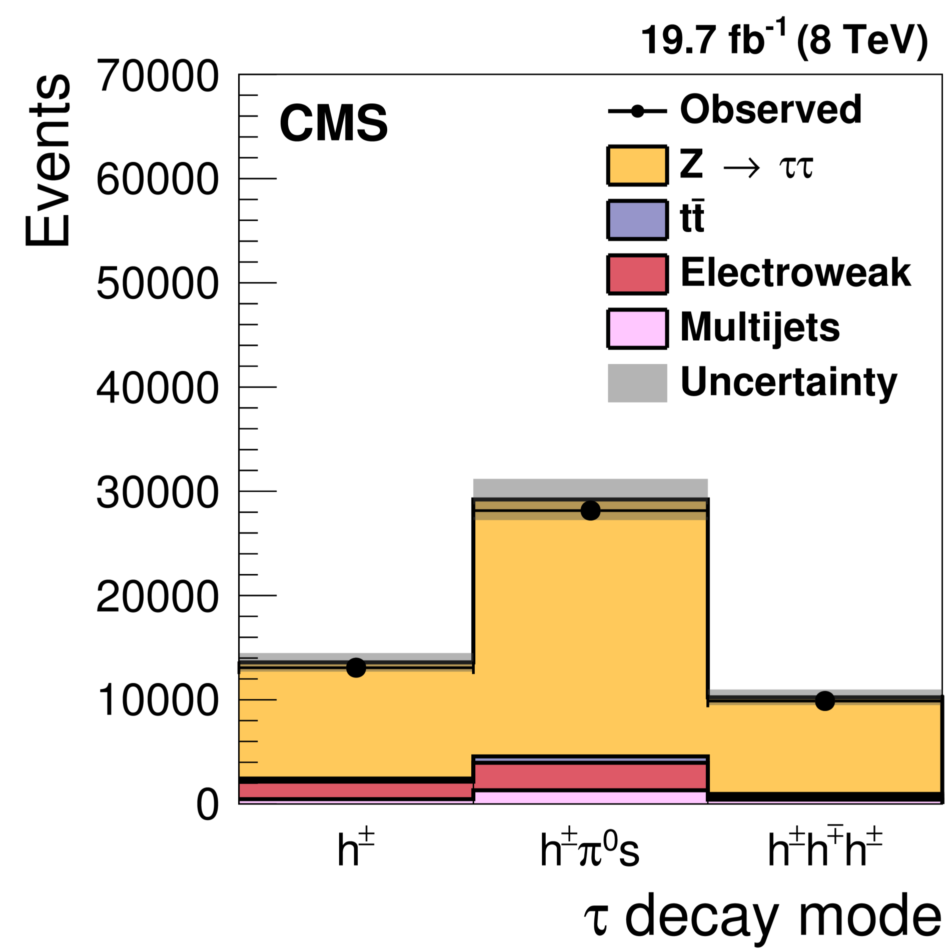
- Max ($\Delta\eta$) = 0.15, Max($\Delta\phi$) = 0.3, Min($\Delta\eta$) = Min($\Delta\phi$) = 0.05

τ-identification Performance

CMS-TAU-14-001







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

Isolation

Handle to discriminate against hadron jets

Tau isolation computed by summing momenta of particles within cone of size dR = 0.5 (or 0.3) around tau direction:

$$I_\tau = \sum P_T^{h\pm}(d_Z < 2mm) + \max(0, \sum P_T^{\gamma} - \Delta\beta \sum P_T^{h\pm}(d_Z > 2mm))$$

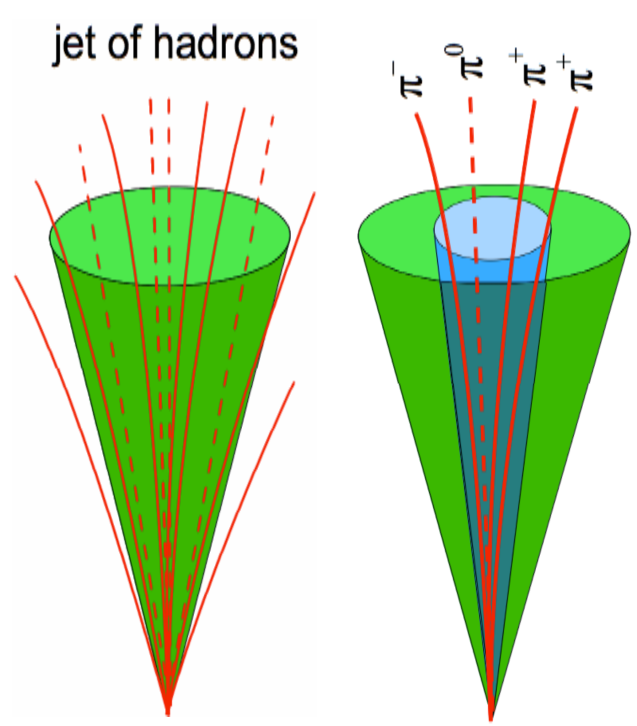
charged isolation

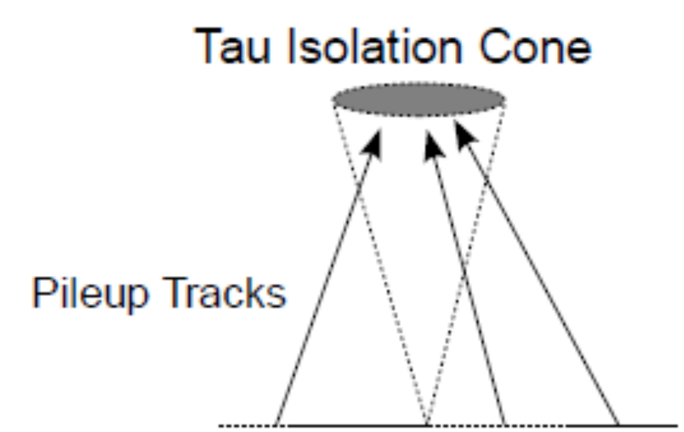
photon isolation

pileup correction

- ✓ Charged hadrons from pileup vertices suppressed by requiring dz < 2mm
- ✓ τ_h production vertex taken to be the vertex that is closest to the highest P_T ("leading") track of the τ_h

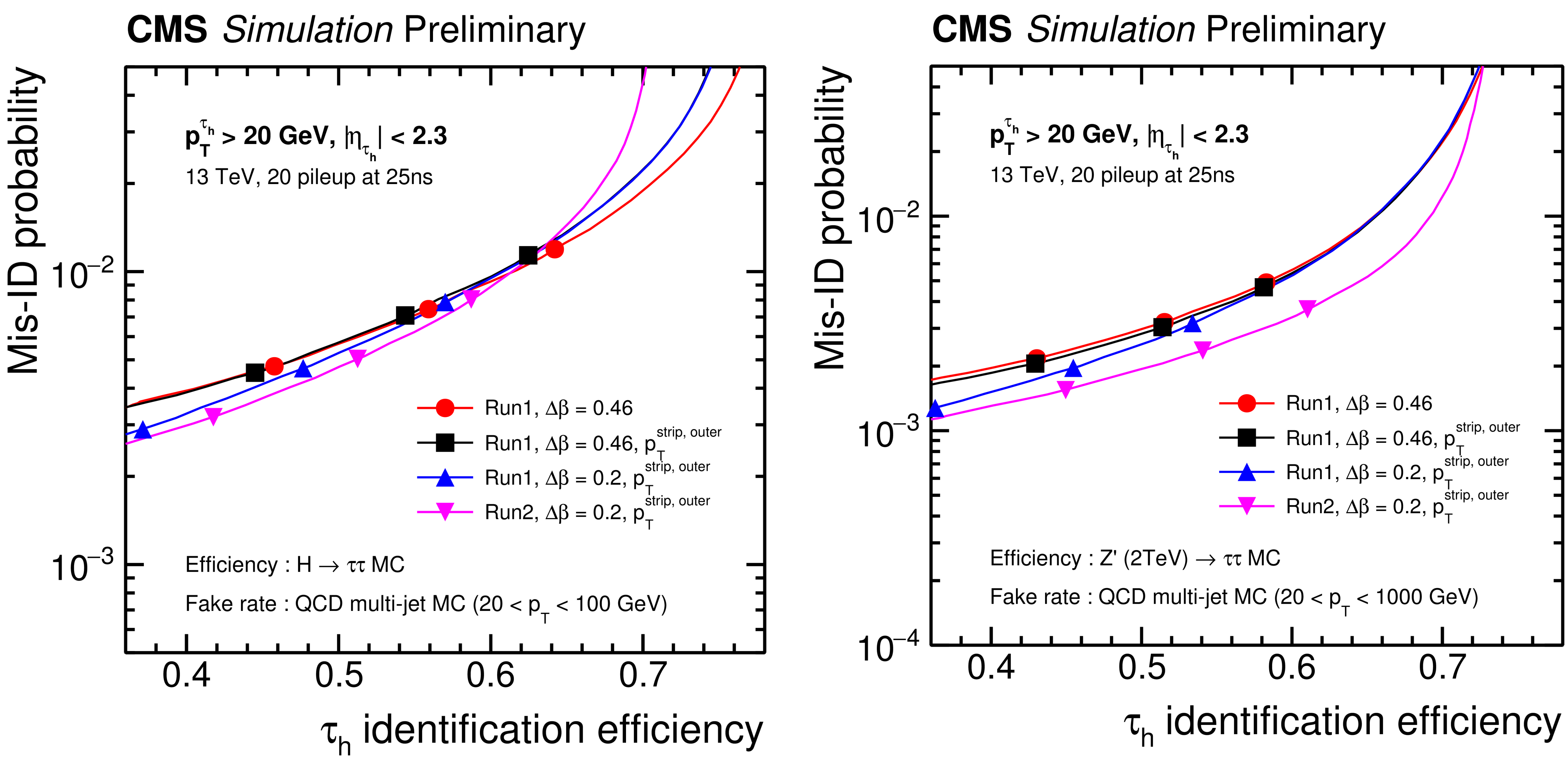
Isolation cuts and pileup correction re-optimized for run-II





τ-Isolation Performance

CMS-PAS-TAU-16-002



Dynamic strip algorithm with cut-based isolation

Improved performance compared to run-1

Multivariate τ_h Isolation

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

> BDT for MVA training (τs as signal & jets as background)

Kinematic Variables:

- $P_T(\boldsymbol{\tau})$
- $\eta(\boldsymbol{\tau})$
- Reconstructed tau decay mode

Cut-based Isolation:

- P_T (charged hadrons)
- P_T (photons)
- Pileup correction ($\delta\boldsymbol{\beta}$)
- p_T of photons in strips outside signal cone

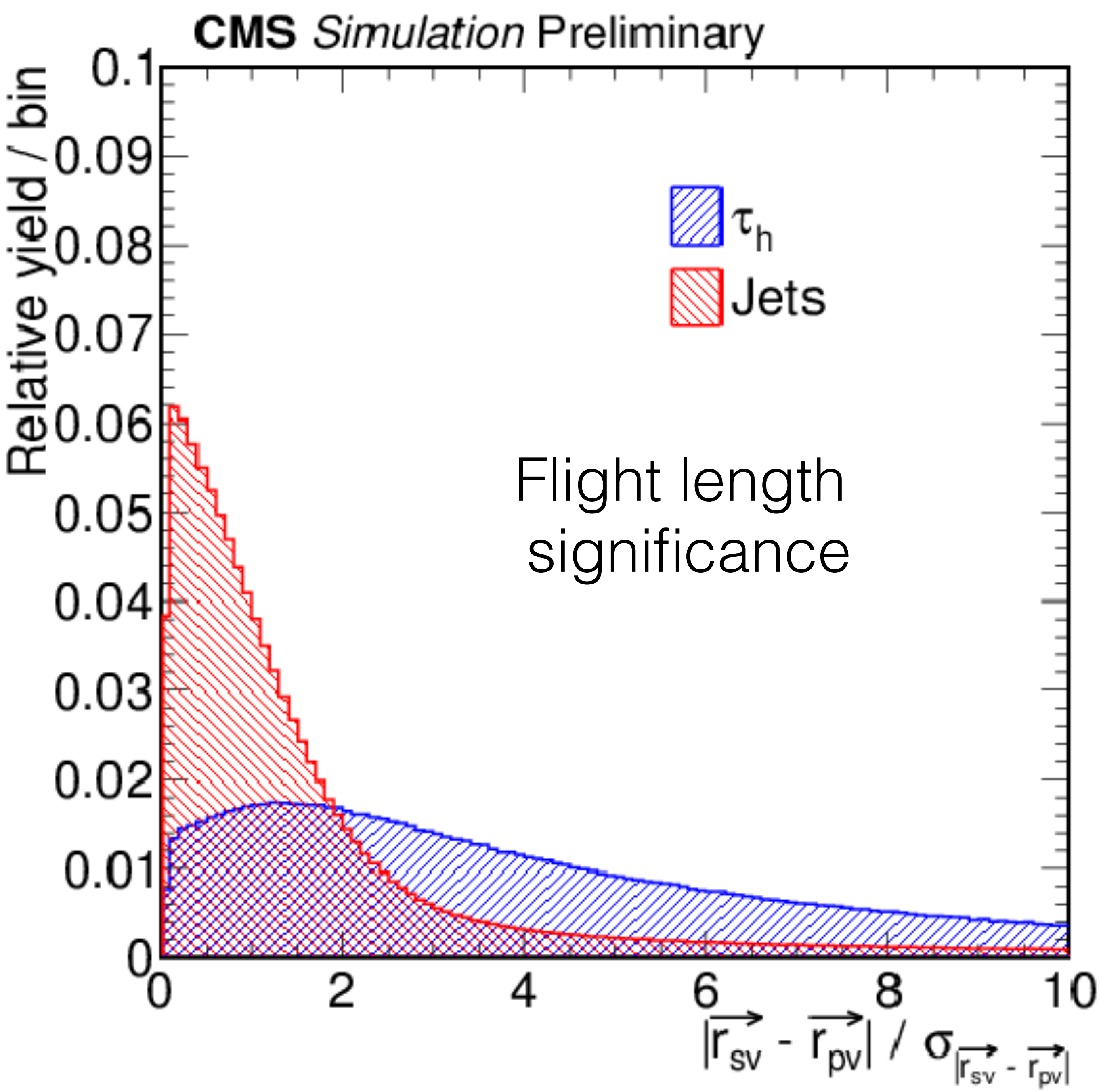
Tau lifetime variables:

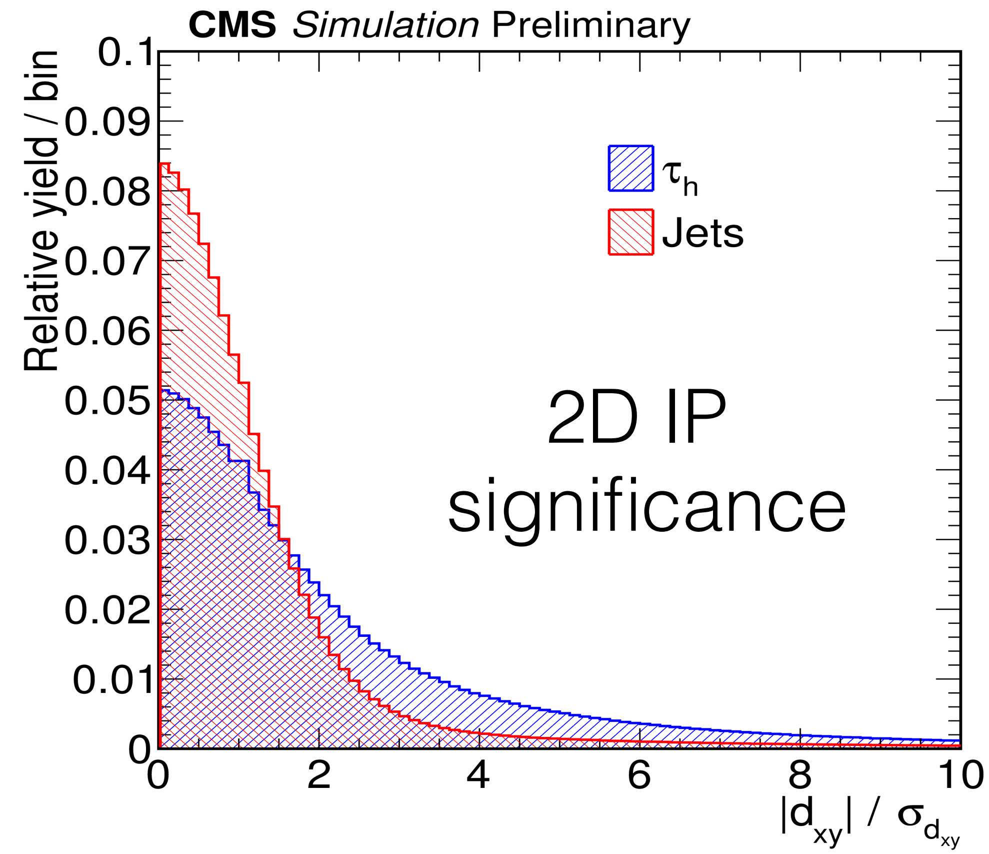
- Signed 2d and 3d impact parameter of the leading track & its significance
- Presence of secondary vertex
- $\boldsymbol{\tau}$ flight length
- $\boldsymbol{\tau}$ flight significance

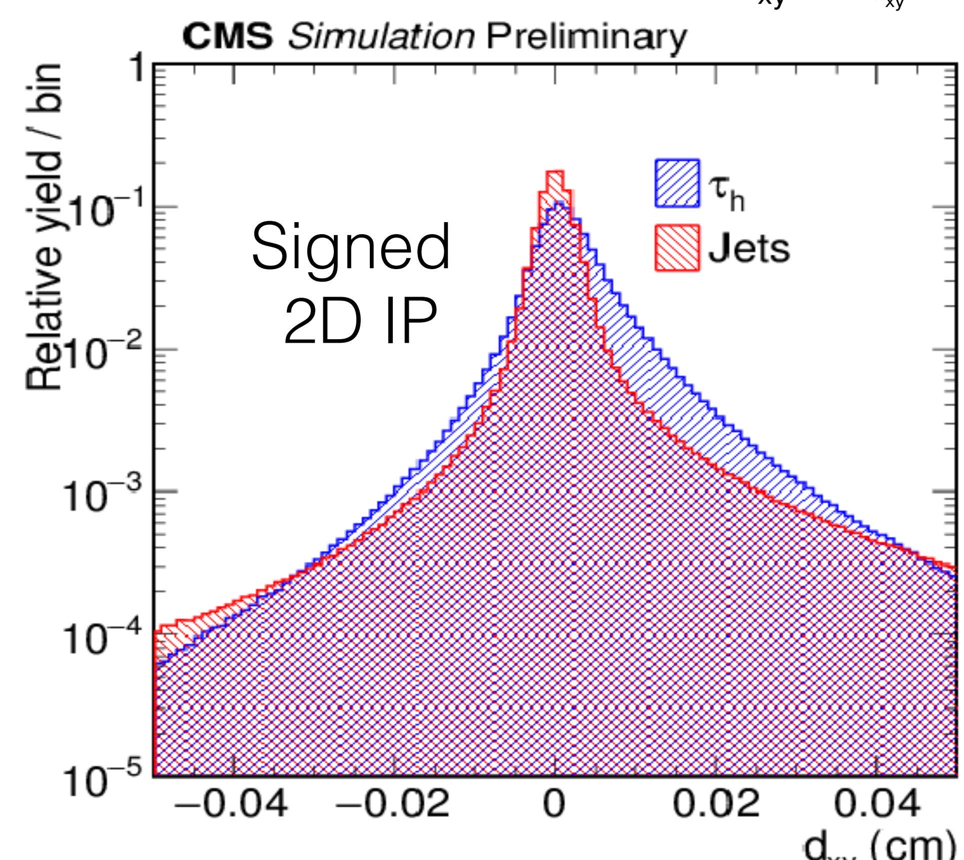
- Additional particle-flow photon variables within signal and isolation cones

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

MVA ID Variables

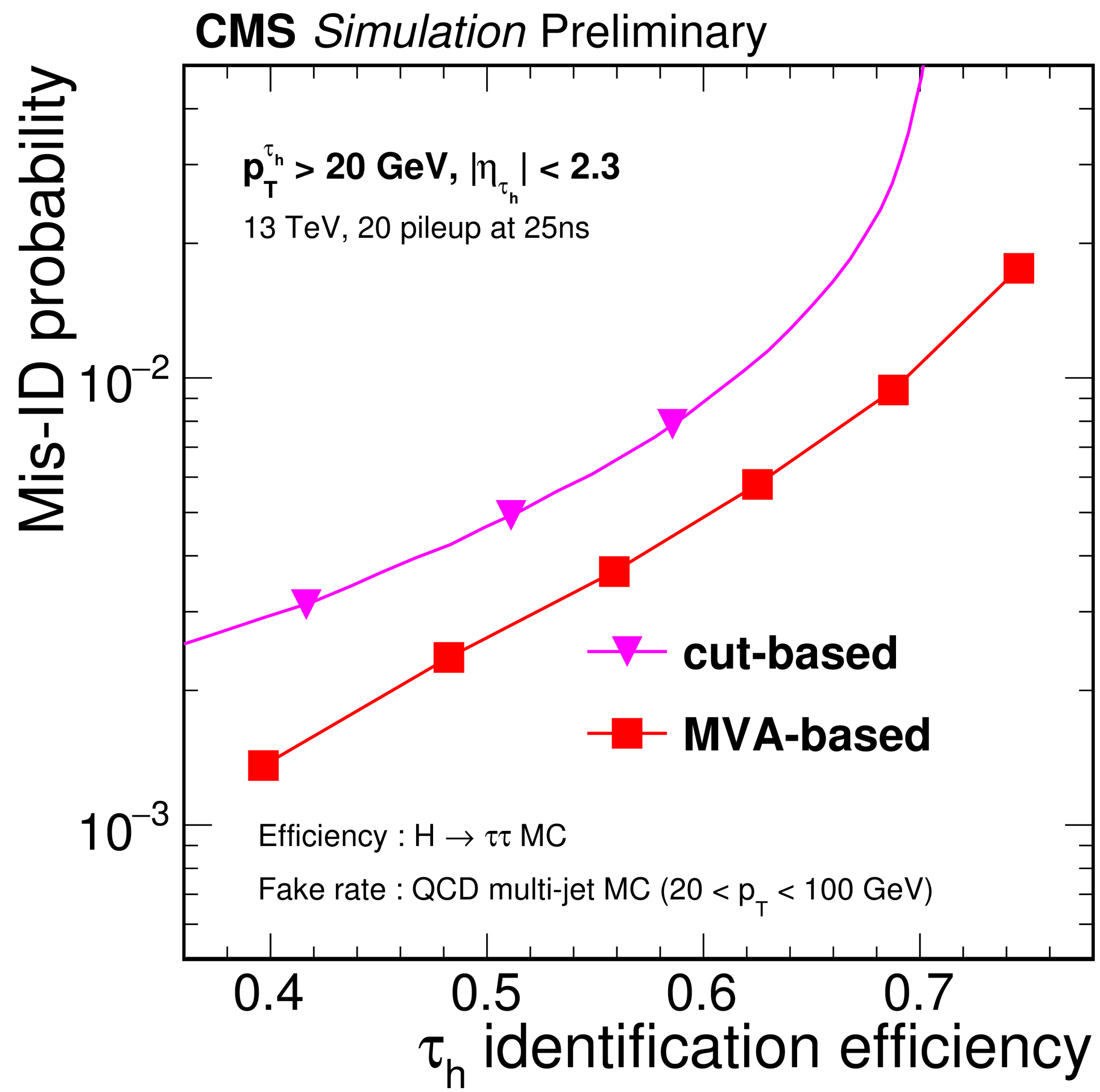


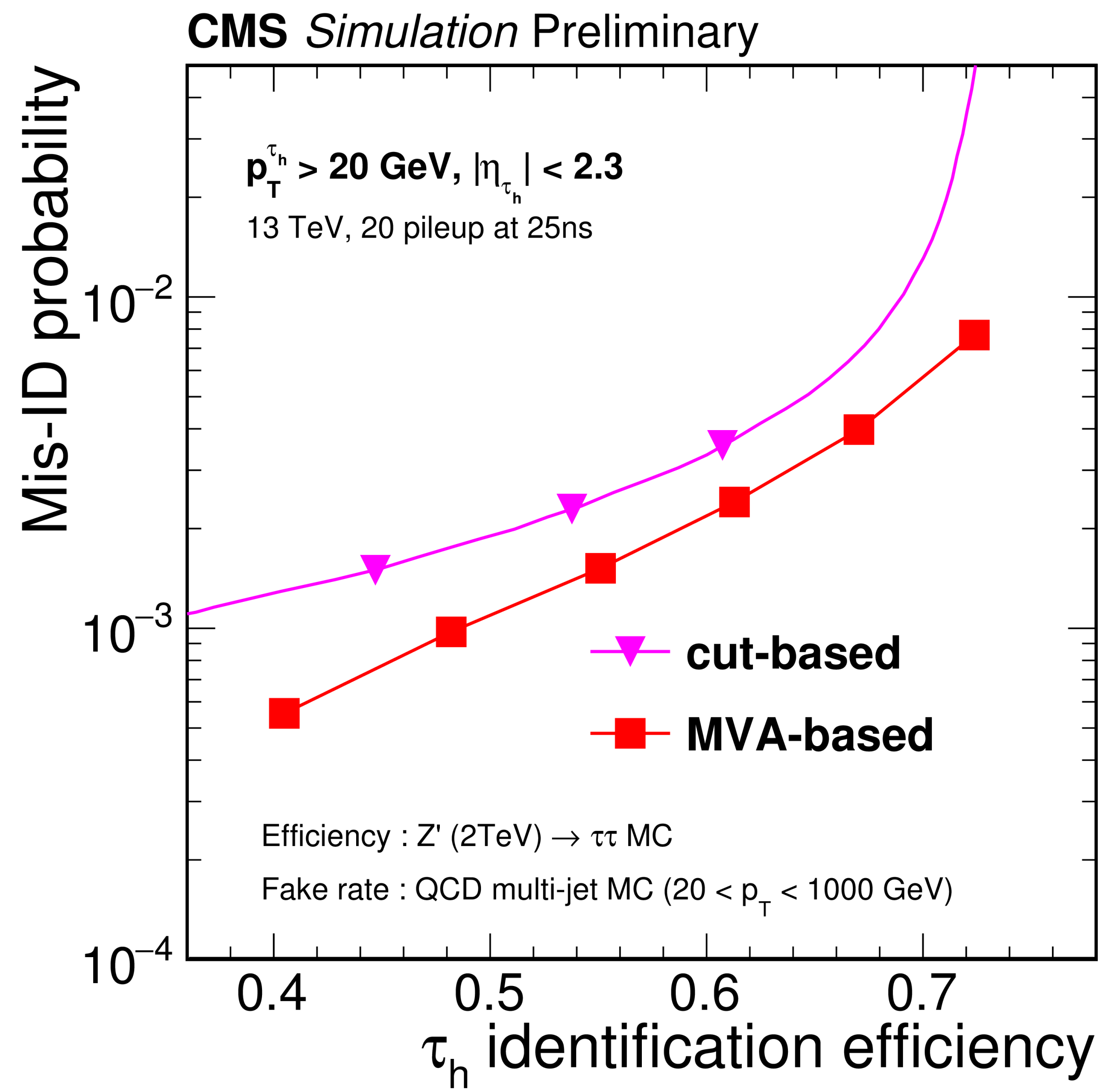




τ MVA Isolation Performance

CMS-PAS-TAU-16-002





Dynamic strip algorithm with MVA isolation

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

τ-ID measurements with 13 TeV data

CMS-PAS-TAU-16-002, CMS DP-2016/040

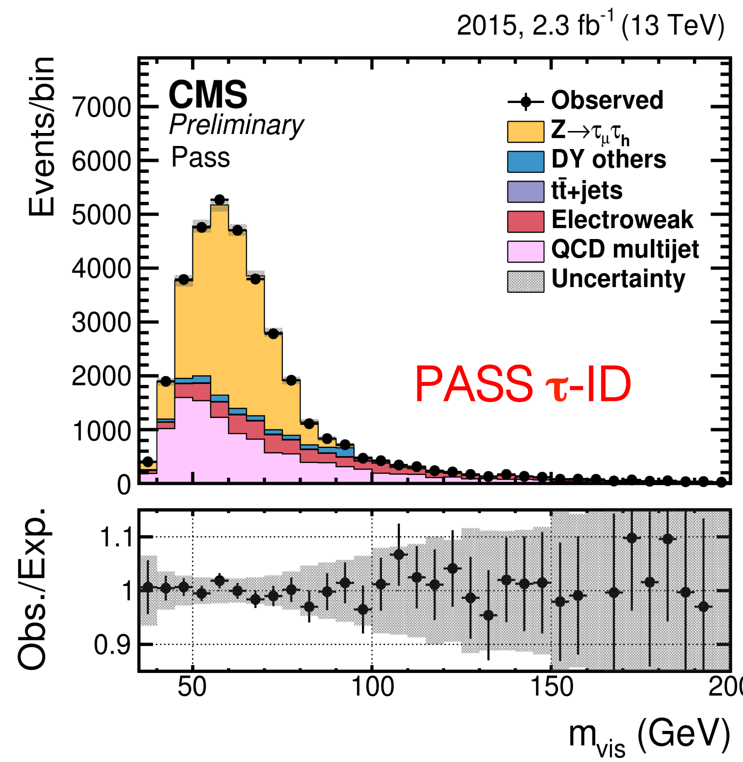
- Tau ID efficiency measured from $Z \rightarrow \tau\tau \rightarrow \tau_{\mu}\tau_{h}$ events using a **Tag** (μ) & **Probe** (τ_h) method.

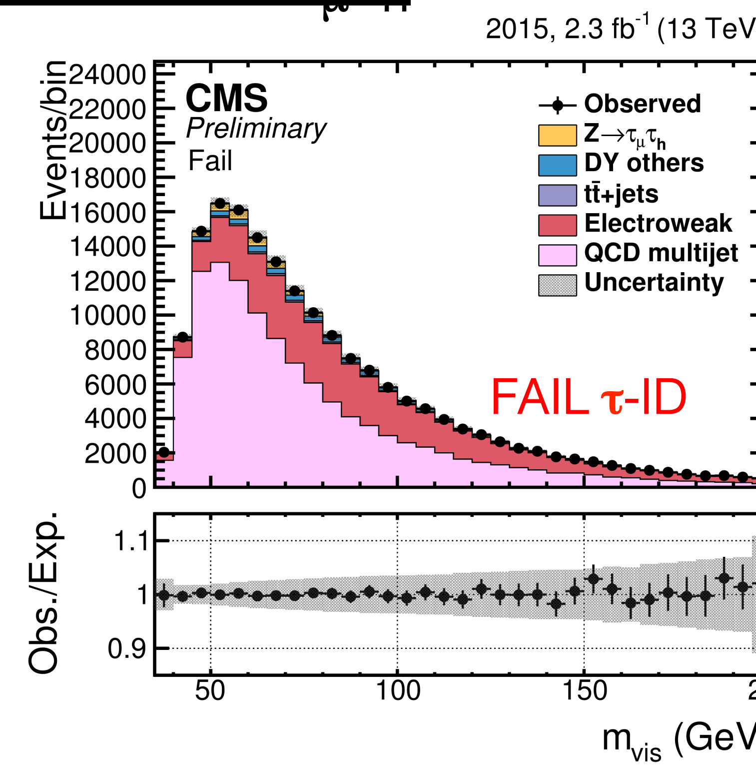
  Data/MC SF consistent with 1

- $e \rightarrow \tau_h$ & $\mu \rightarrow \tau_h$ fake rate measured from $Z \rightarrow ee$ & $Z \rightarrow \mu\mu$ events, respectively

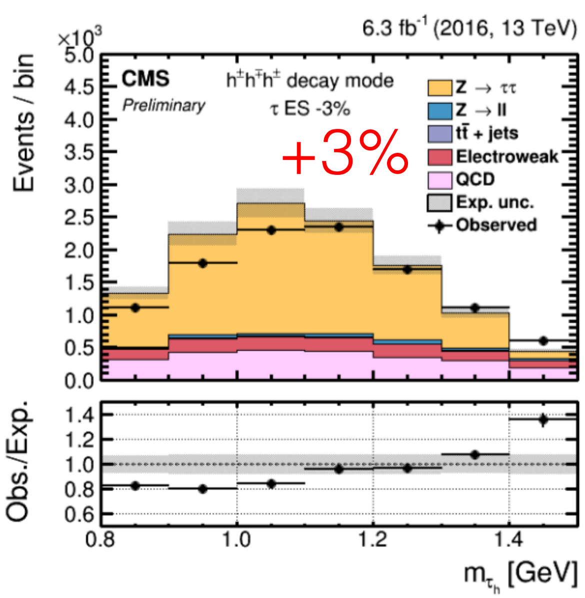
- Jet $\rightarrow \tau_h$ fakes measured from $W(\mu\nu)$+jets events

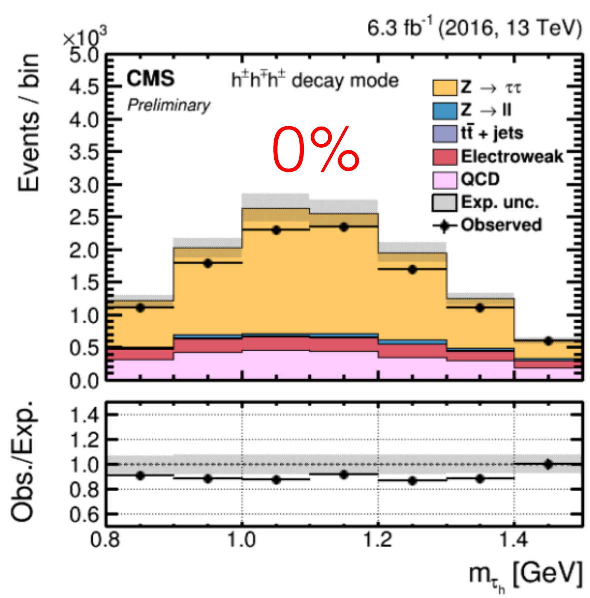
Visible mass of $\tau_{\mu}\tau_{h}$

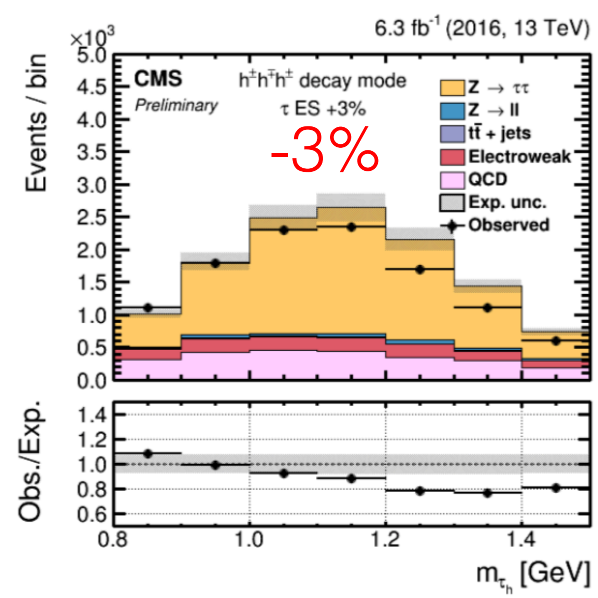




τ_h energy scale measured by fitting $m_{vis}(\mu\tau_h)$ and $m_{vis}(\tau_h)$ distributions



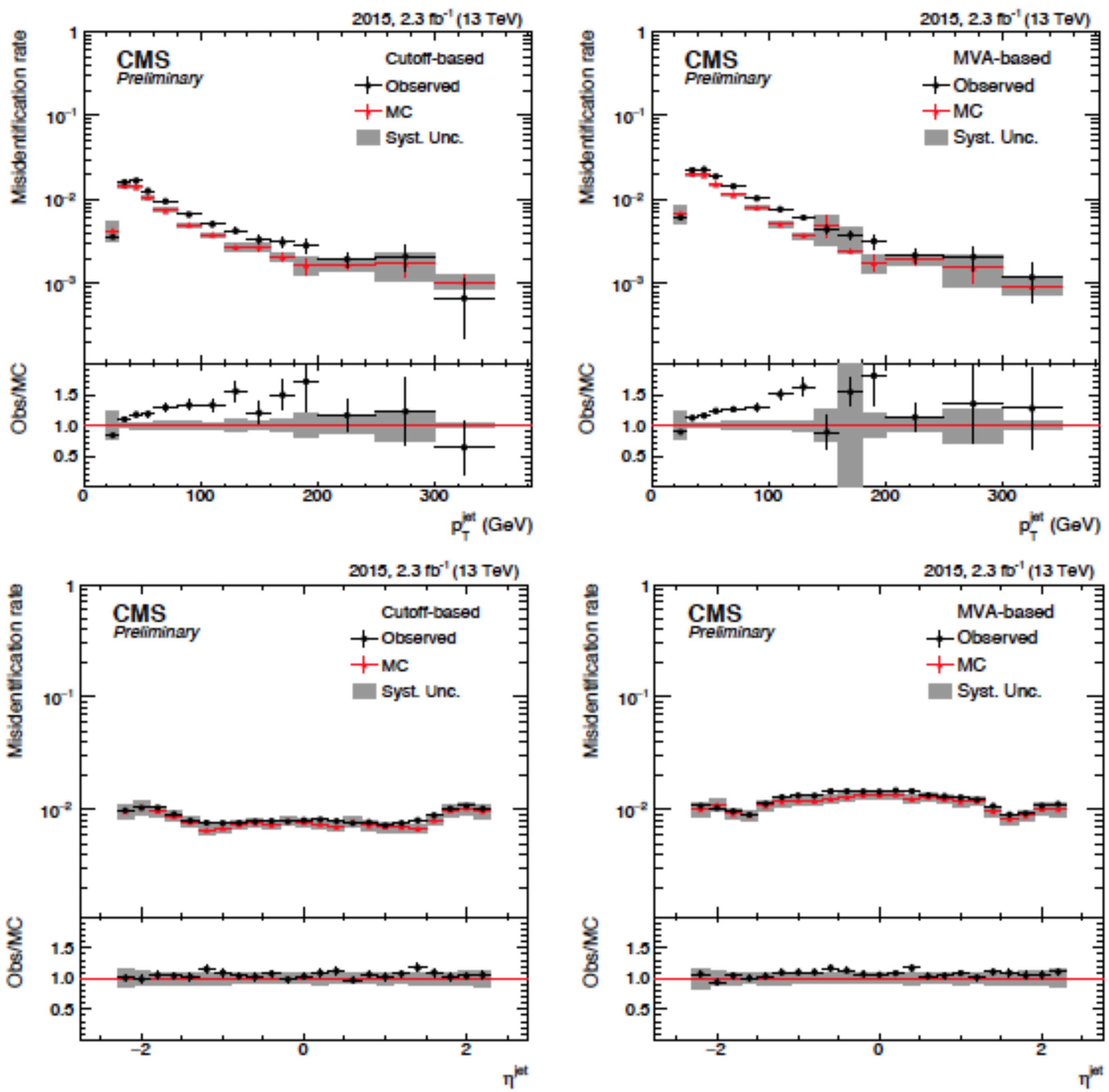




Decay mode	Tau energy scale [%]
1-prong	+0.0 ± 1.1
1-prongπ^0	+1.0 ± 0.4
3-prong	-0.1 ± 0.2

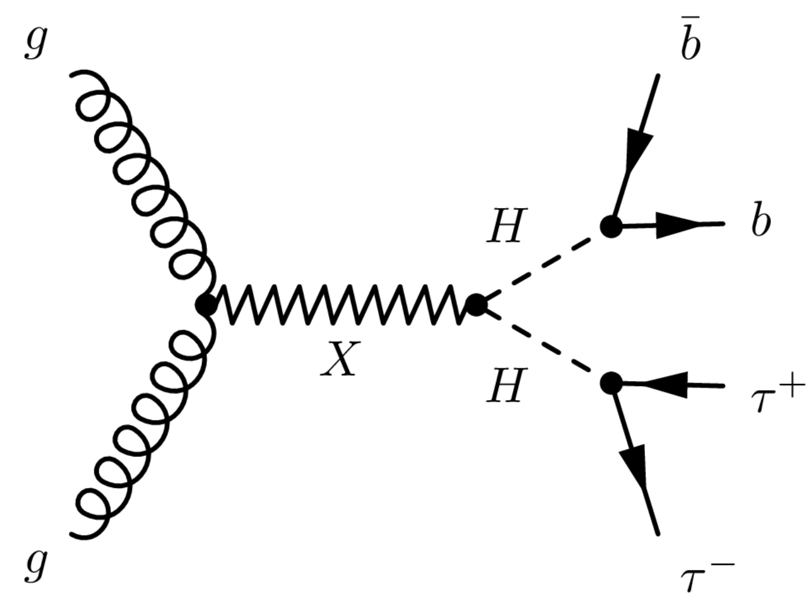
Mis-Identification Probability

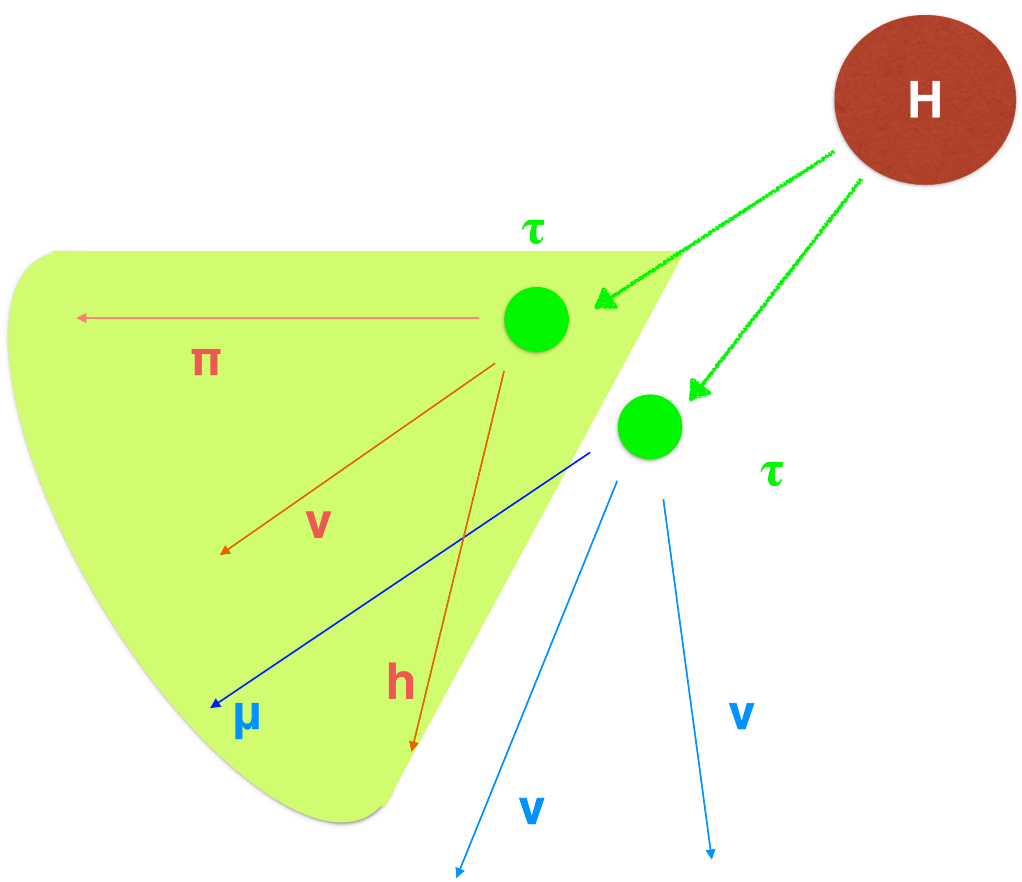
Fake rate measured in W+jets events

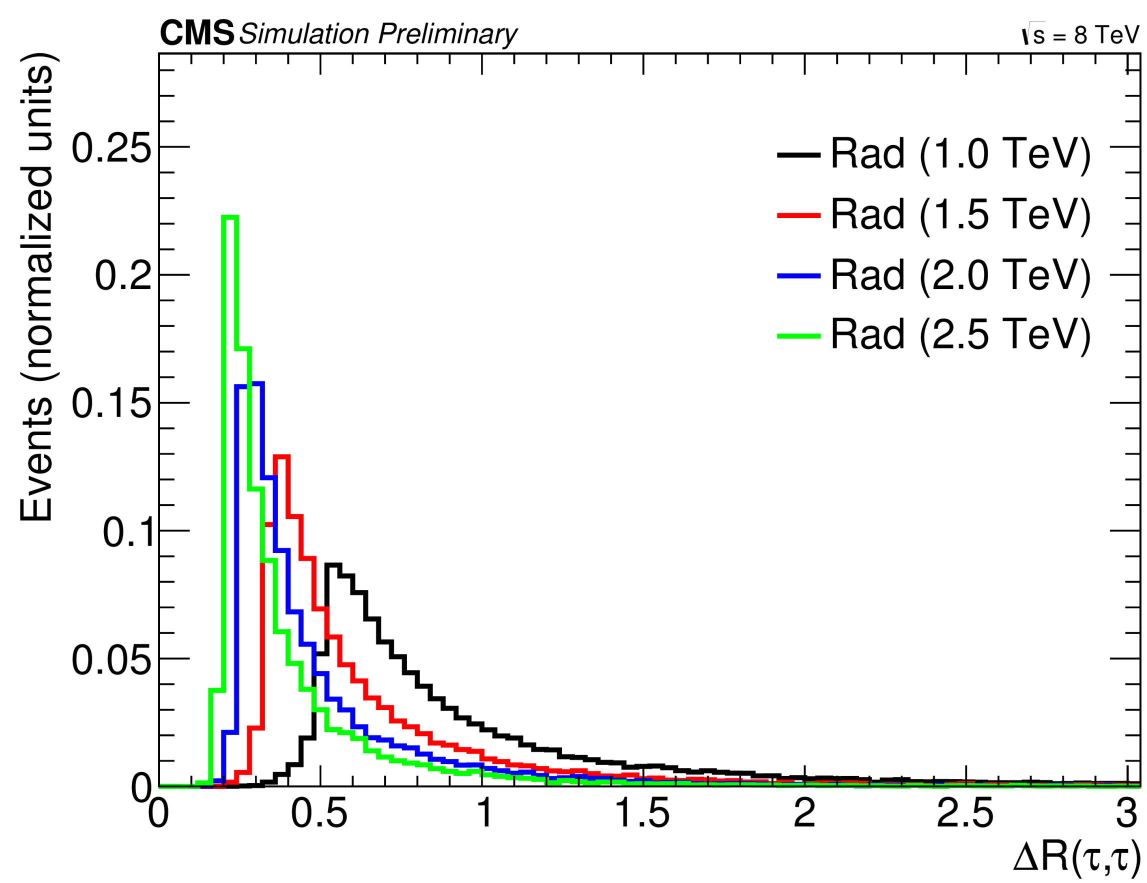


Boosted Scenario

- High mass resonances (e.g. radion) may decay to pair of high p_T Higgs bosons
- Taus from the Higgs boson can be collimated and/or the decay products can be overlapped

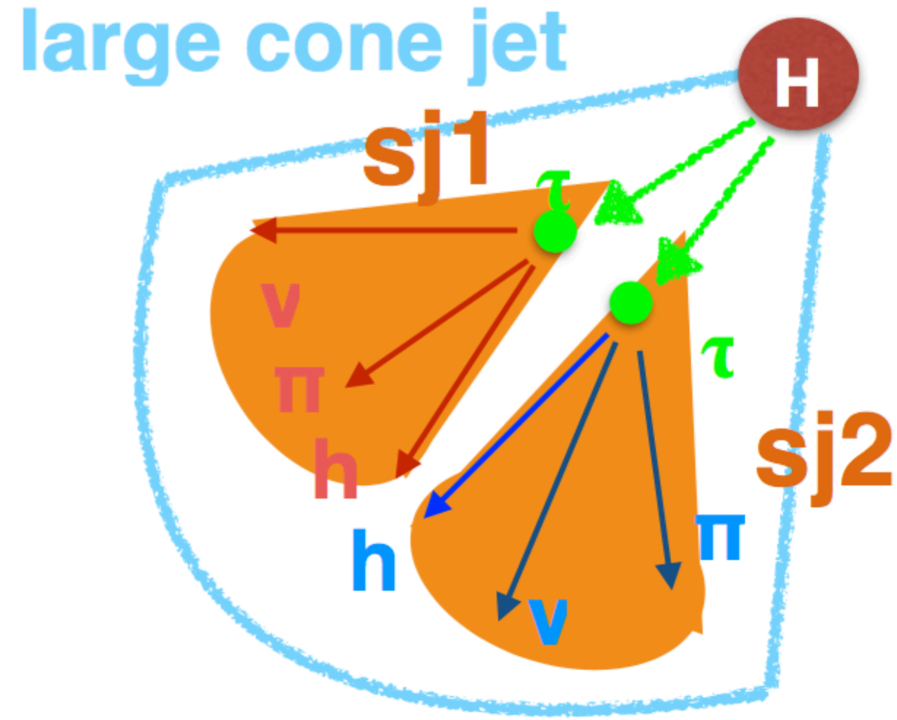


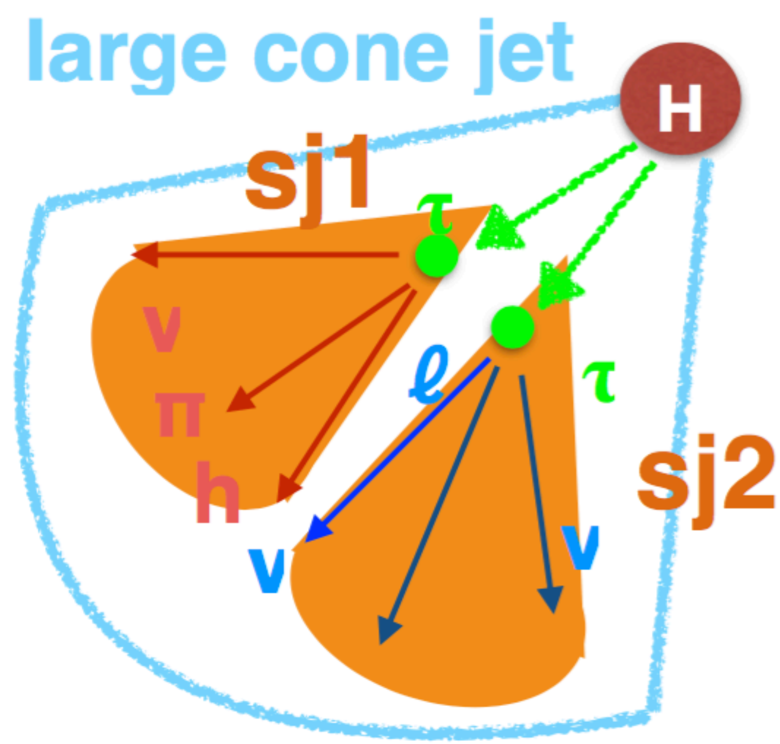




τ reconstruction in boosted regime

- 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 semileptonic 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

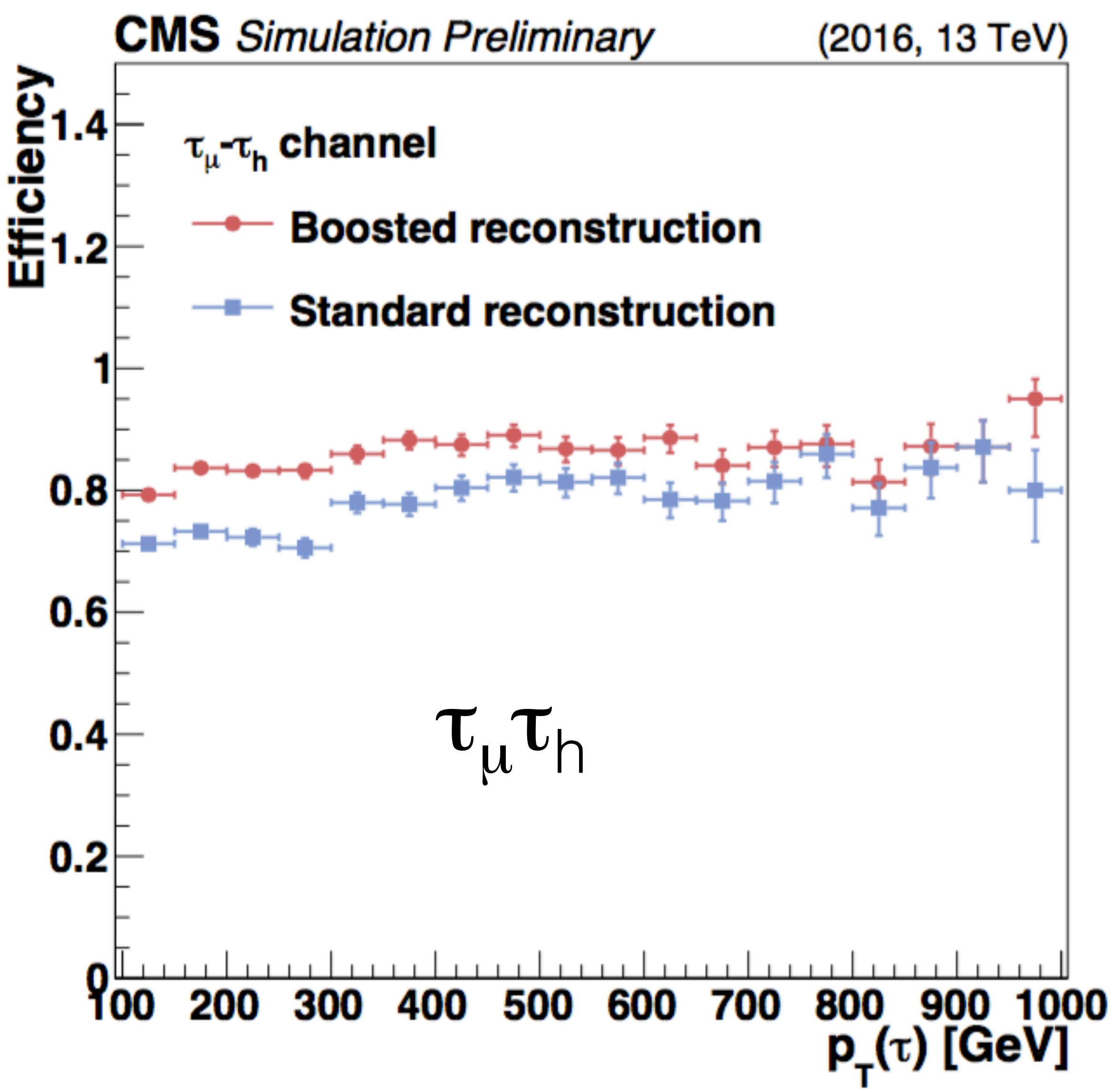


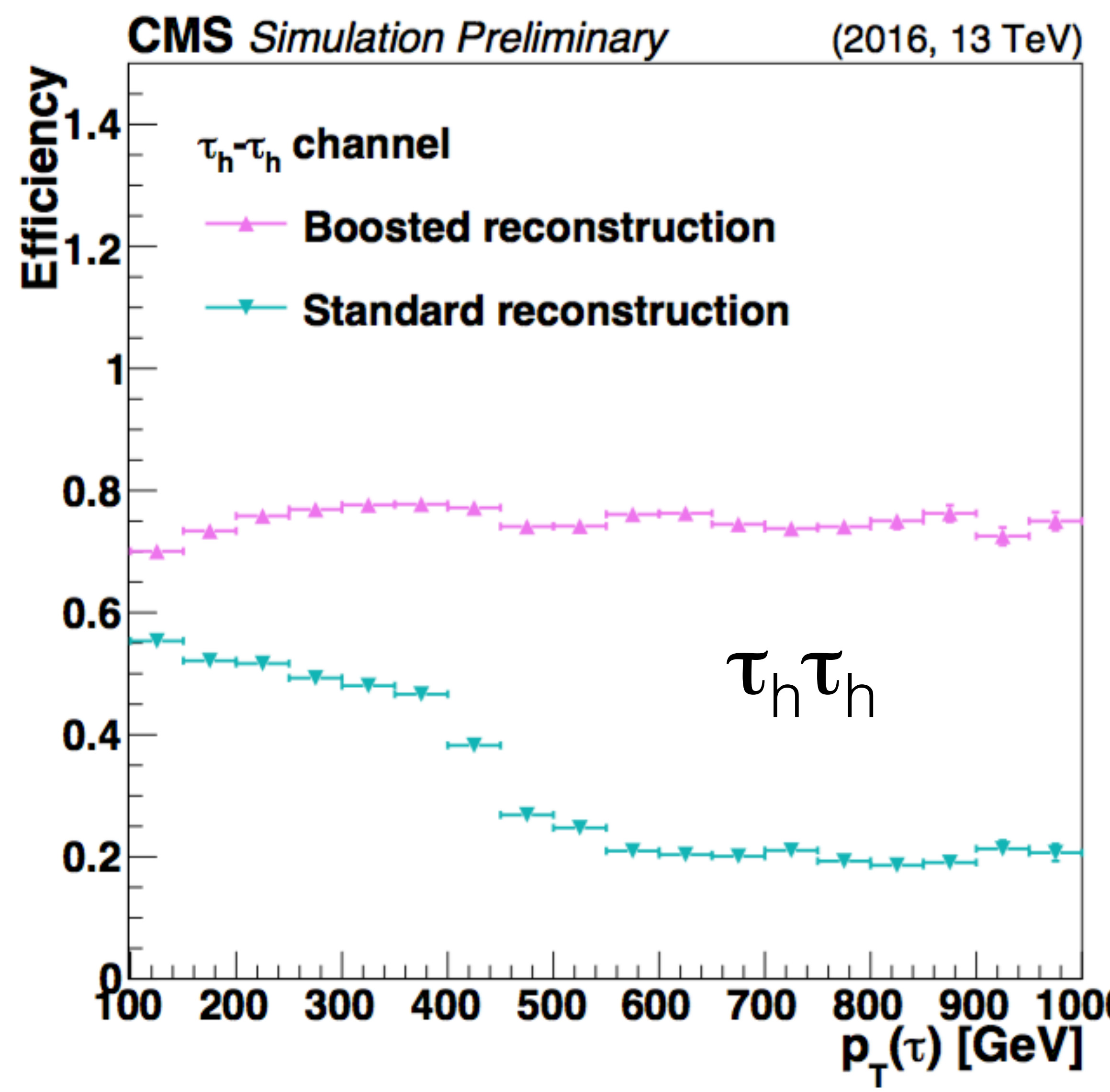


Boosted tau ID performance

tau reco efficiency vs tau p_T

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



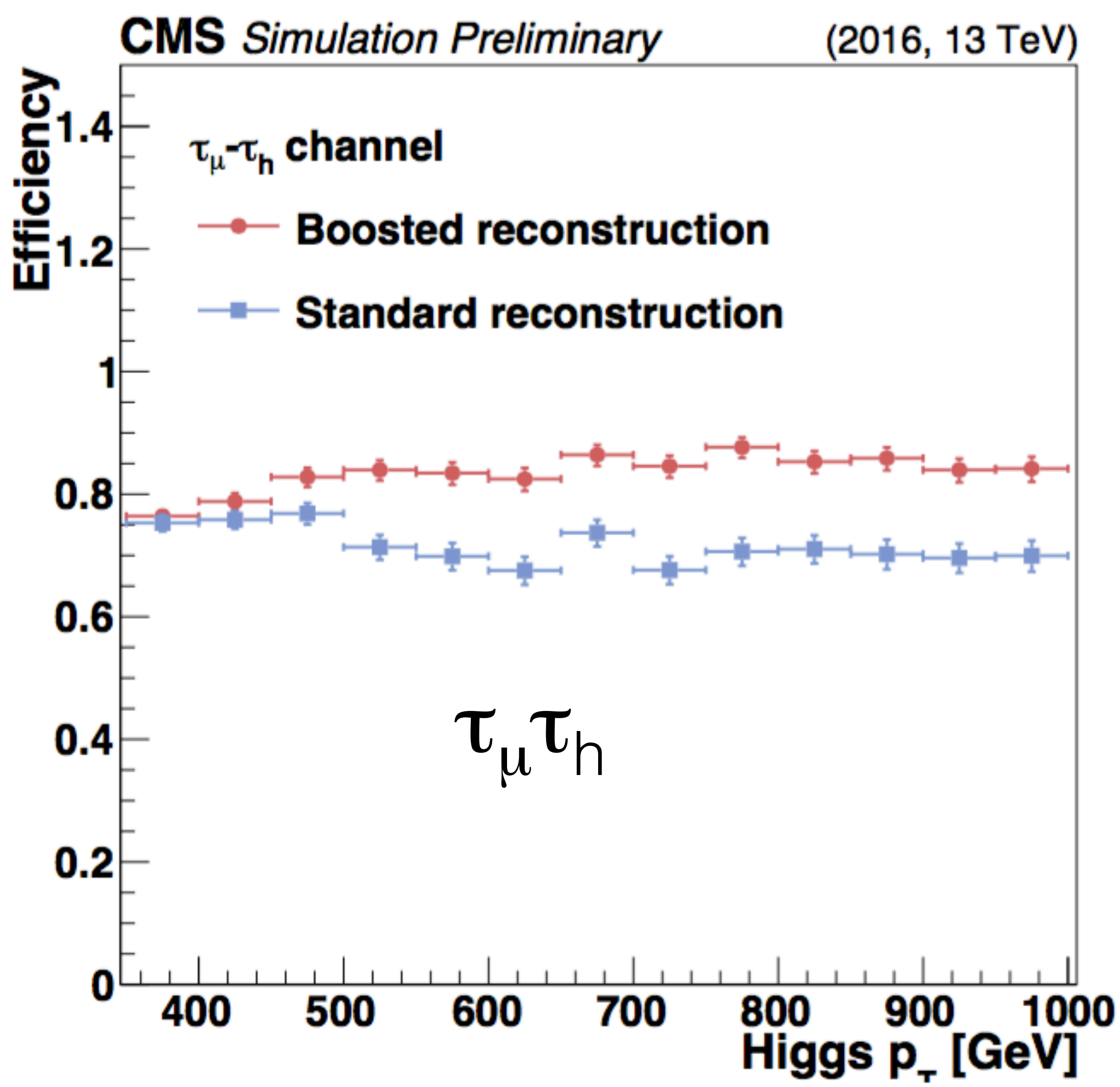


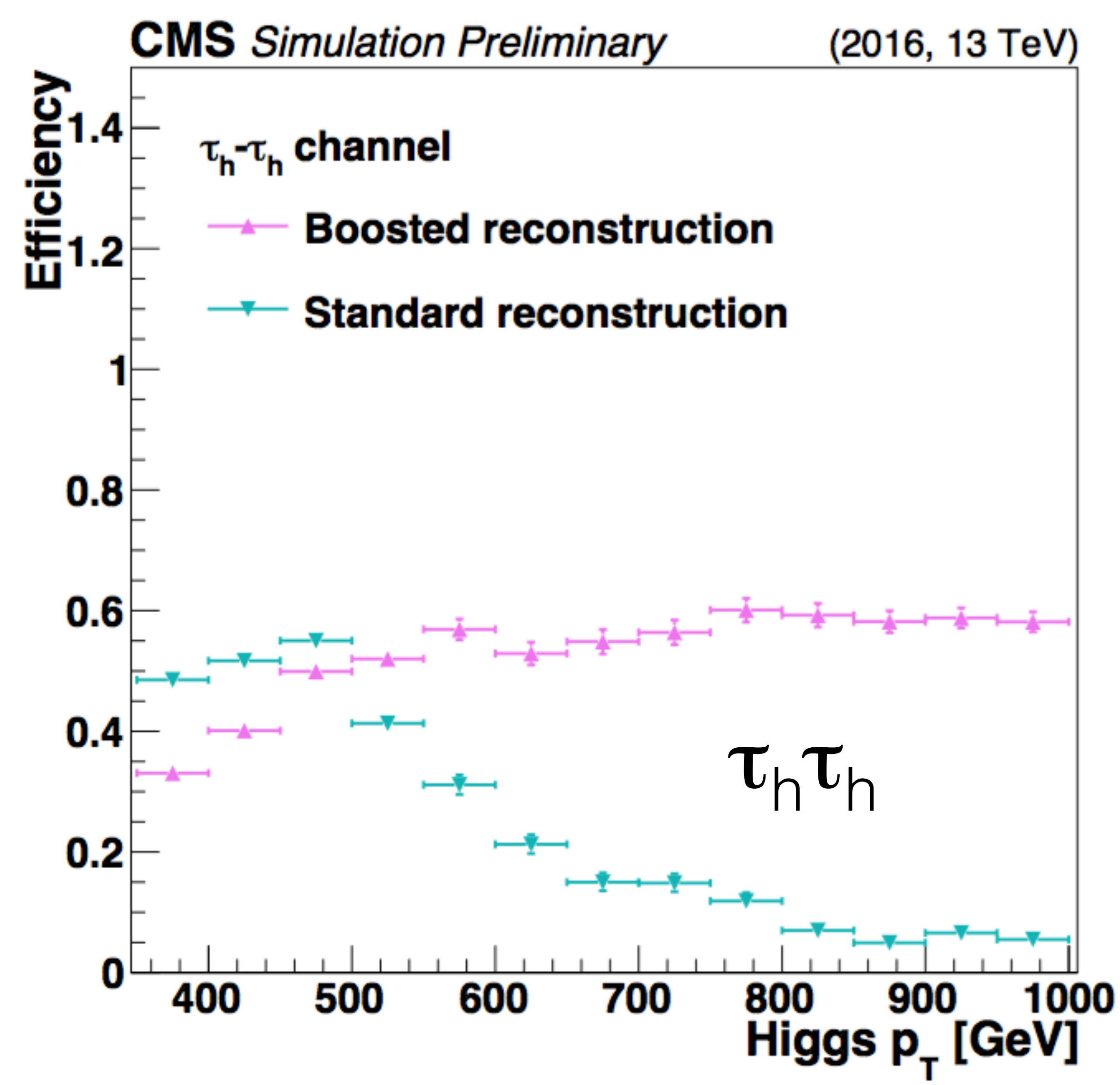
Major improvement in fully hadronic channel

Boosted tau ID performance

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

$H \rightarrow \tau\tau$ reco efficiency vs Higgs p_T

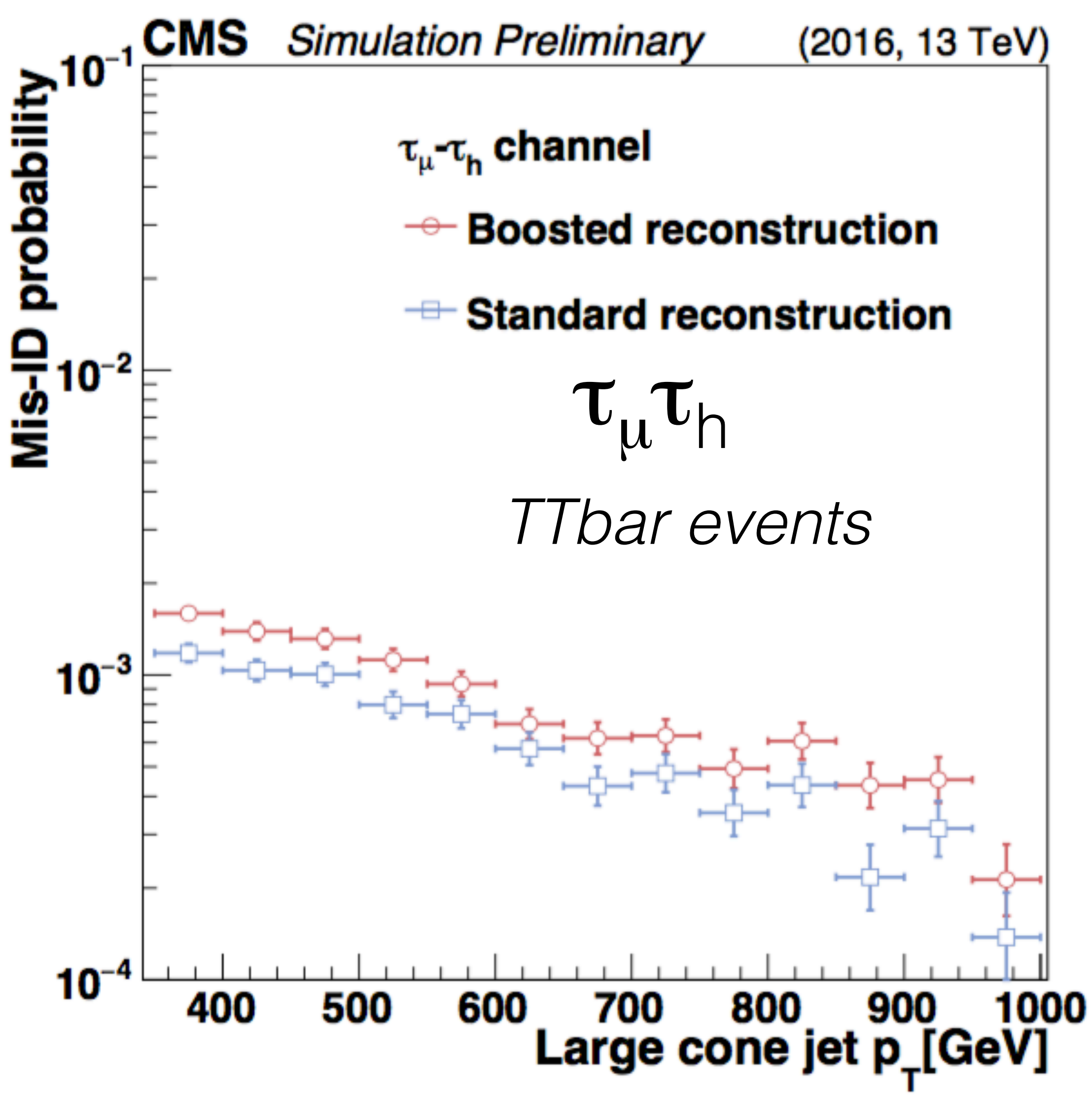


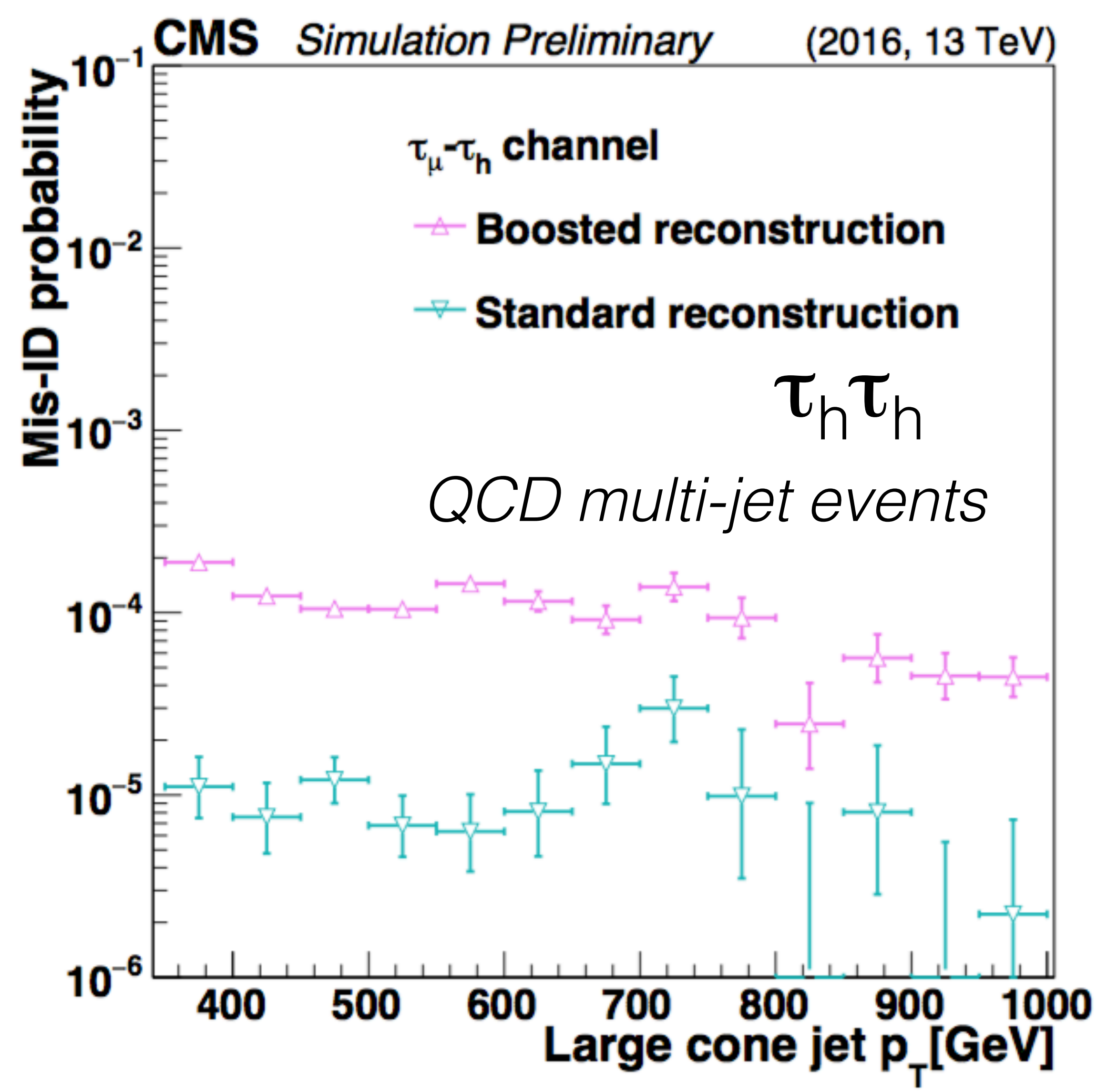


Major improvement in fully hadronic channel

Boosted tau ID performance

Mis-Identification Probability vs large Cone Jet p_T



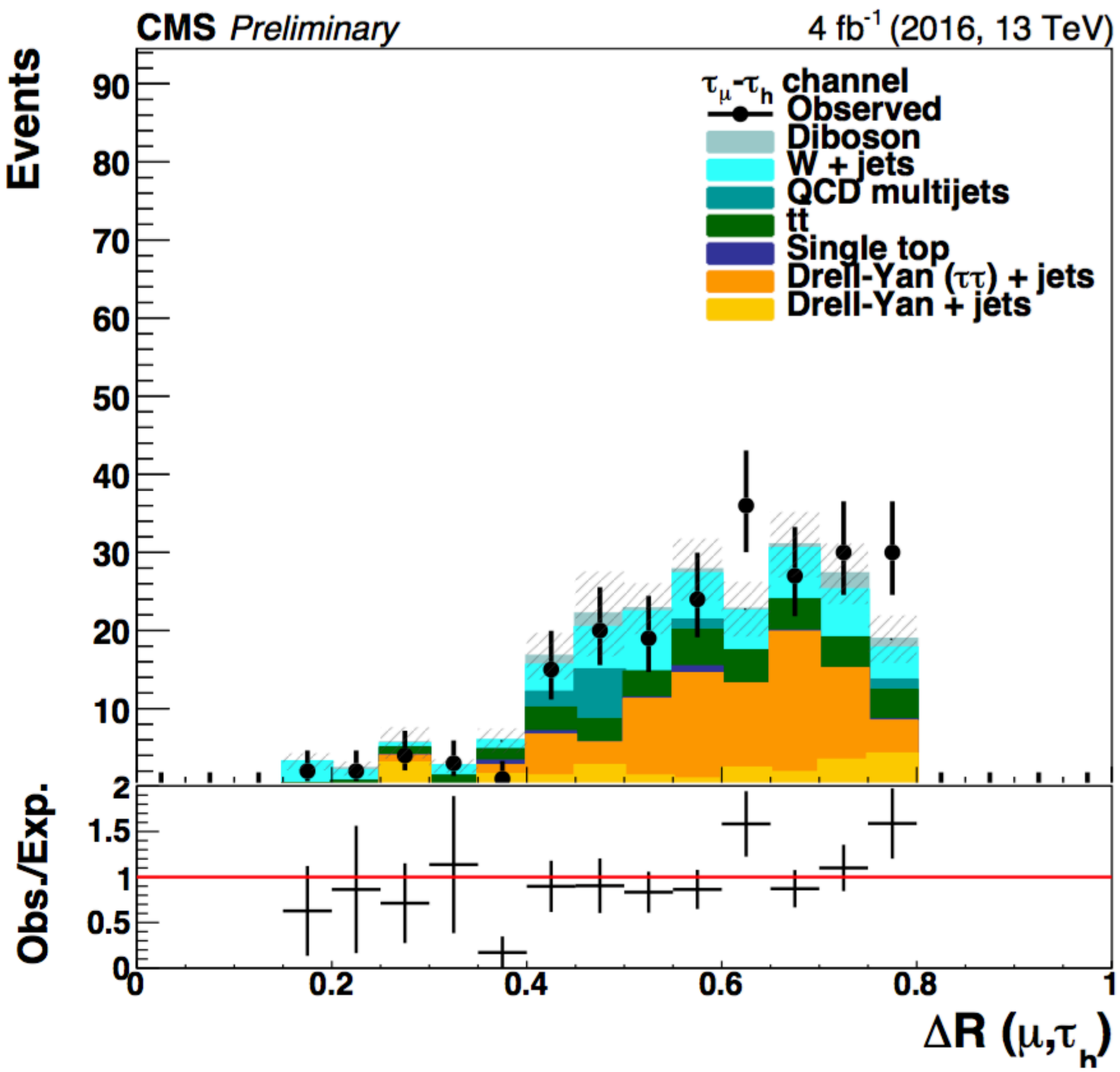


The fake probability increases significantly.
However, the background contributions at such high p_T is smaller

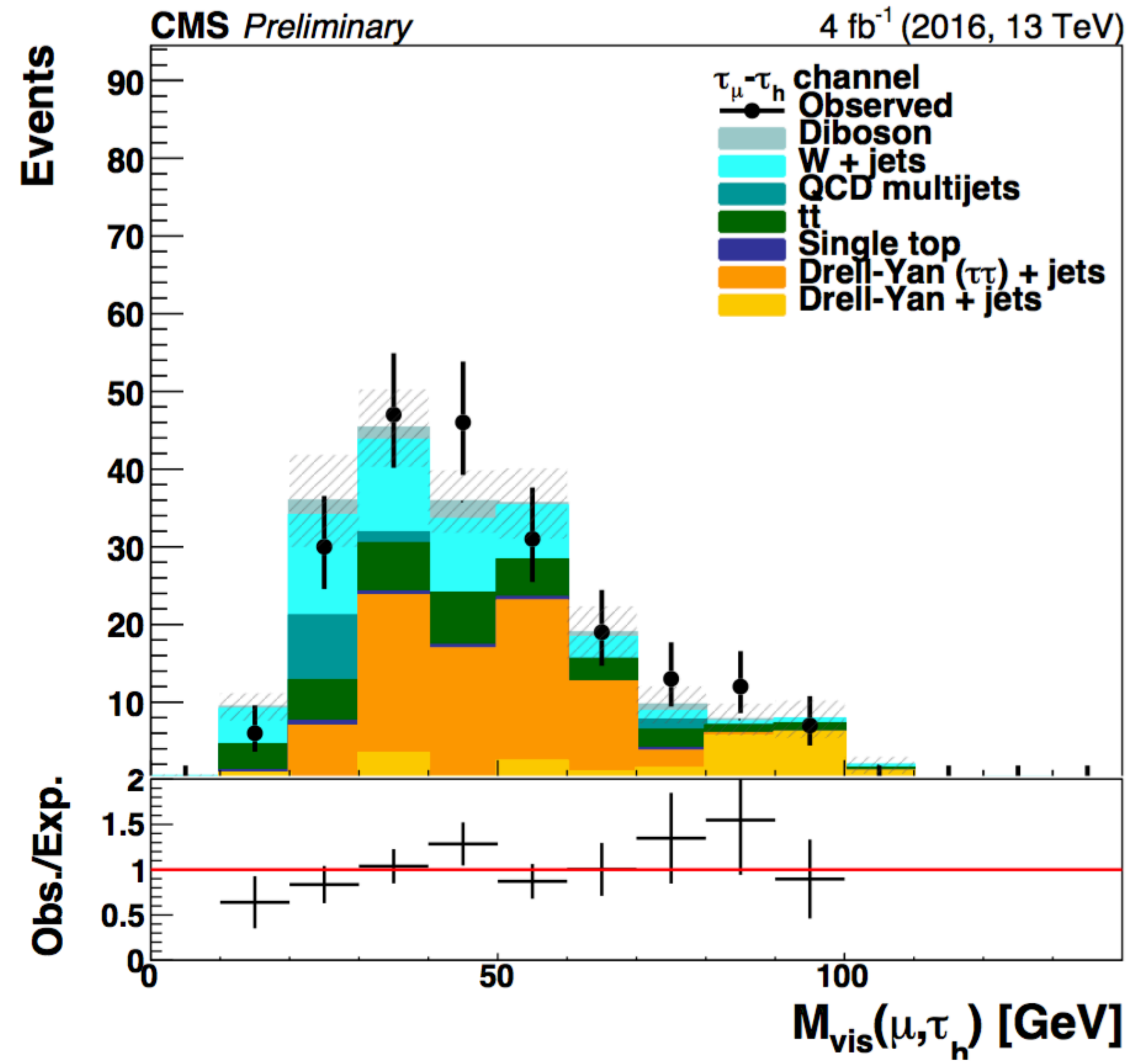
Validation with data

High p_T $Z \rightarrow \tau\tau \rightarrow \mu\tau_h$ 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 the trigger

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

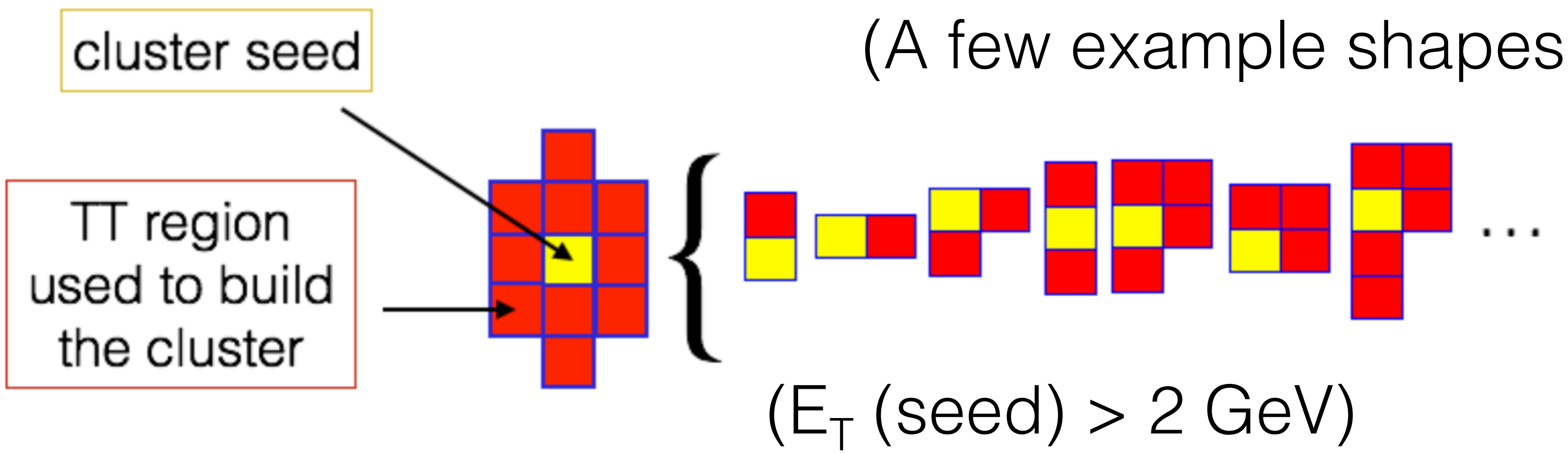
- Tau ID at Level-1 Trigger (Electronics)
  - No possibility of using tracker detector
  - A simpler algorithm developed using energy deposits in the trigger towers (ECAL + HCAL towers)

- Tau ID at High Level Trigger (Computing Farm)
  - No possibility of using offline algorithm yet due to timing constraint
  - A simple cone based algorithm employed at HLT
    - Based on particle-flow with regional tracking

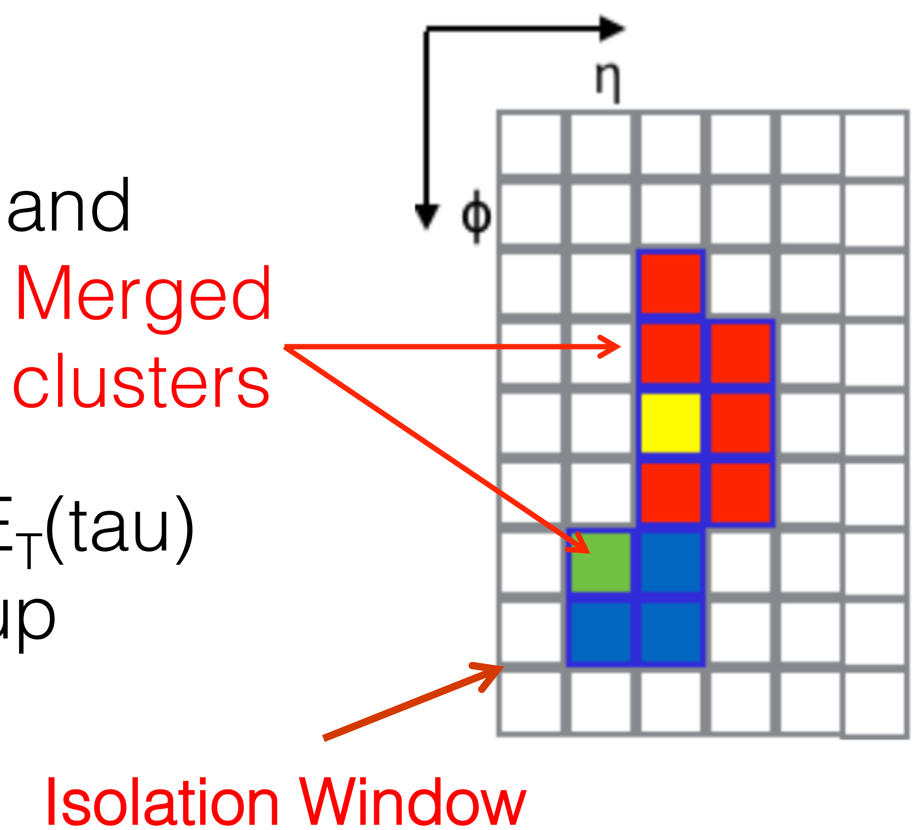
Summary of tau identification at Level-1

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)

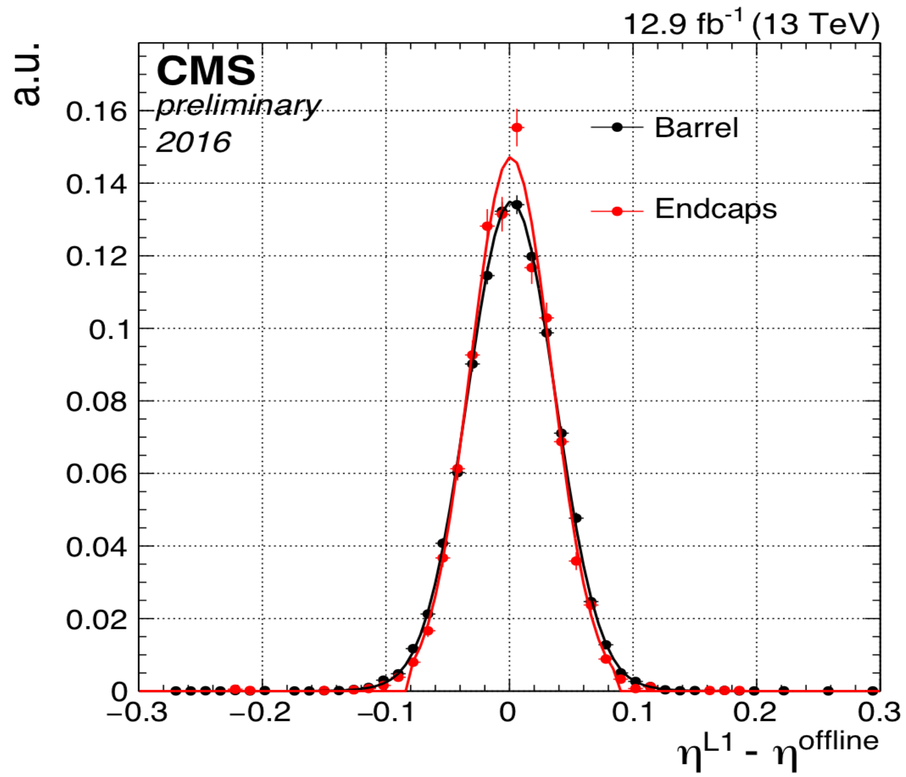


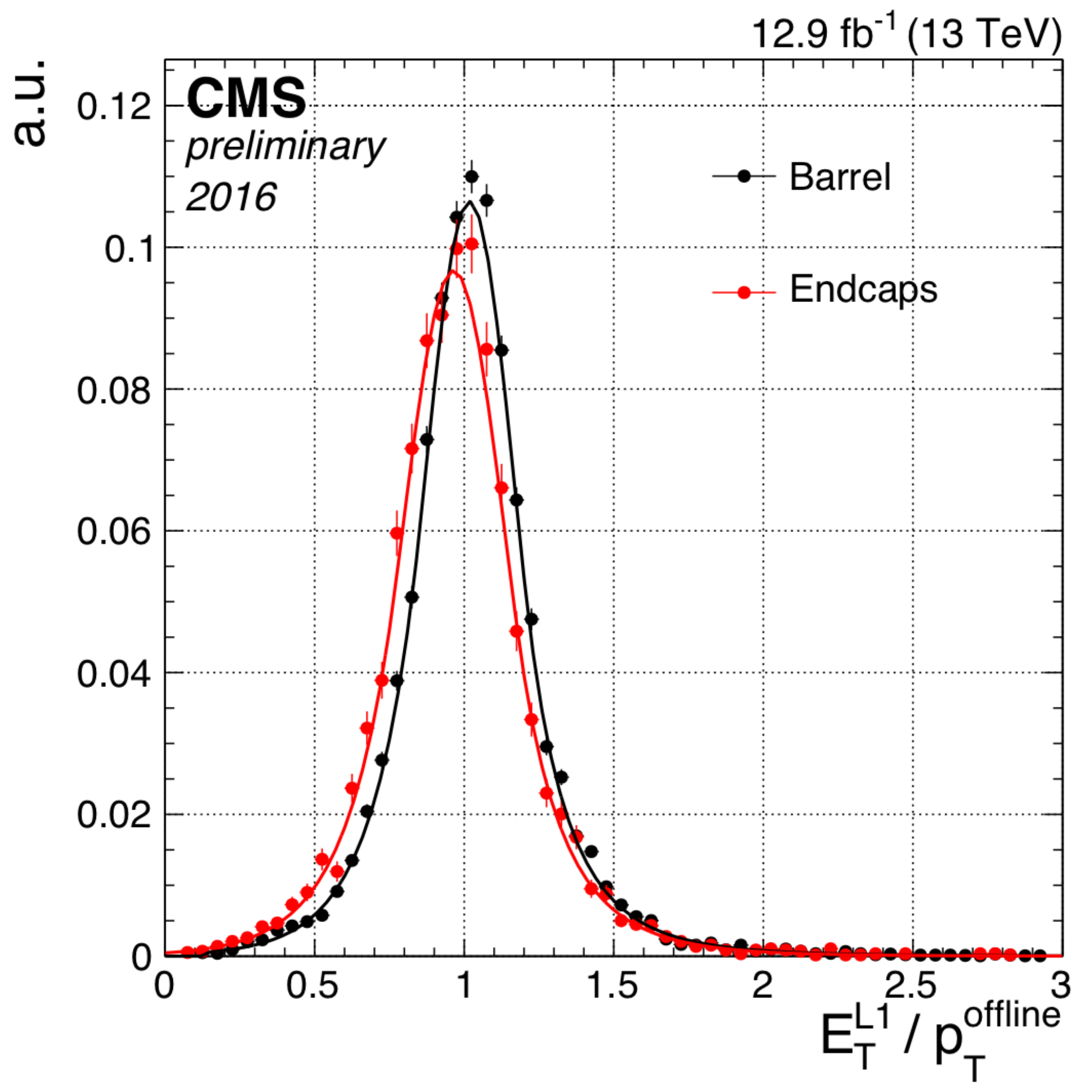
- **Calibration:** As function of E_T, eta, merging, and presence of ECAL deposits

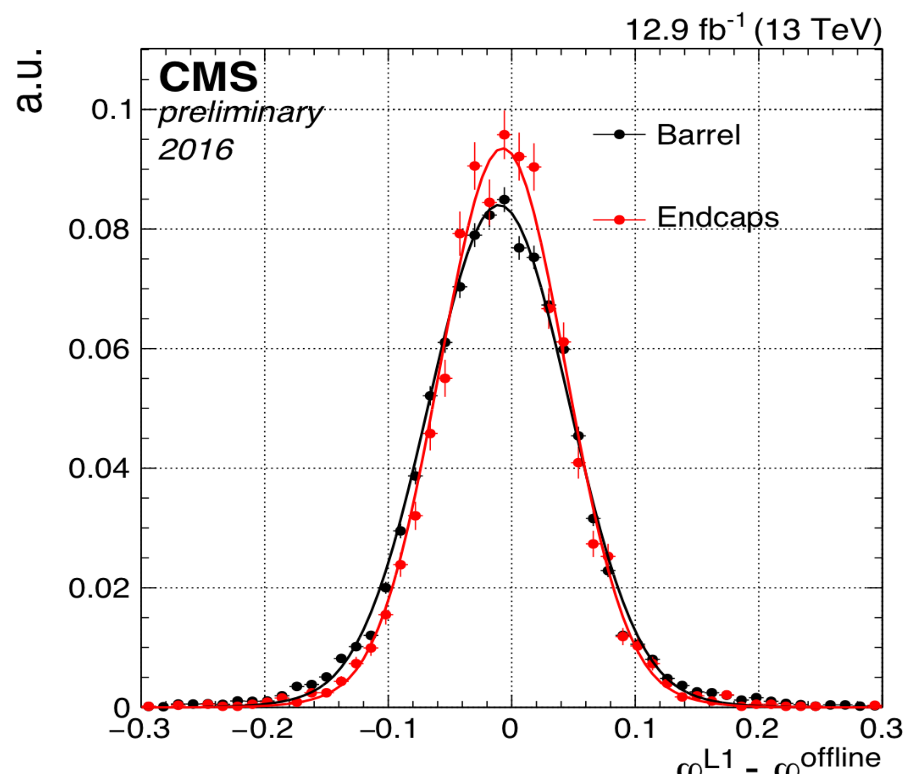
- **Isolation:** Computed as $E_T(\text{iso}) = E_T(6\text{x}9) - E_T(\text{tau})$
  Cut on E_T(iso) depends on p_T, $|\eta|$, and pileup

L1 tau-trigger performance

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

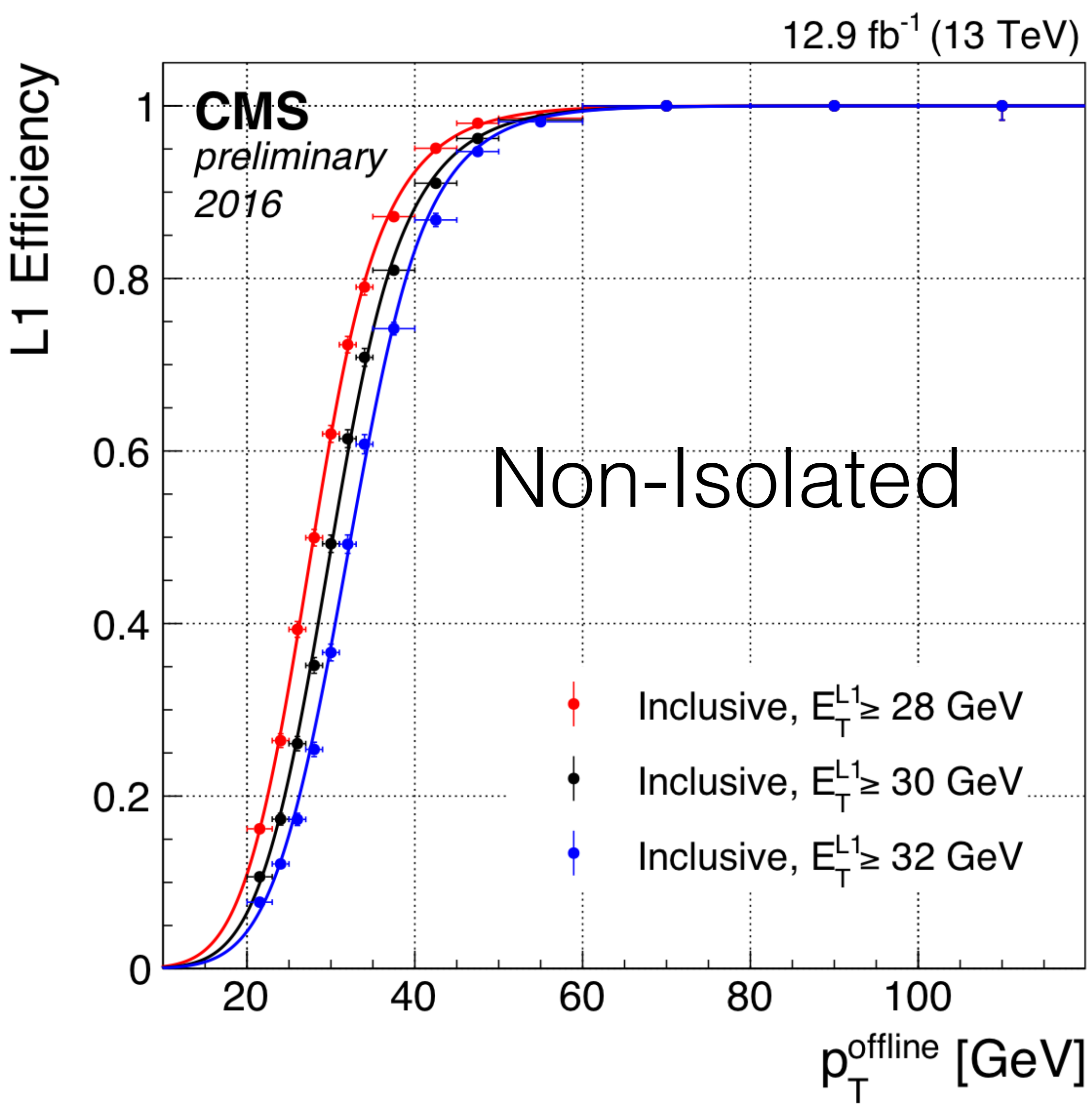


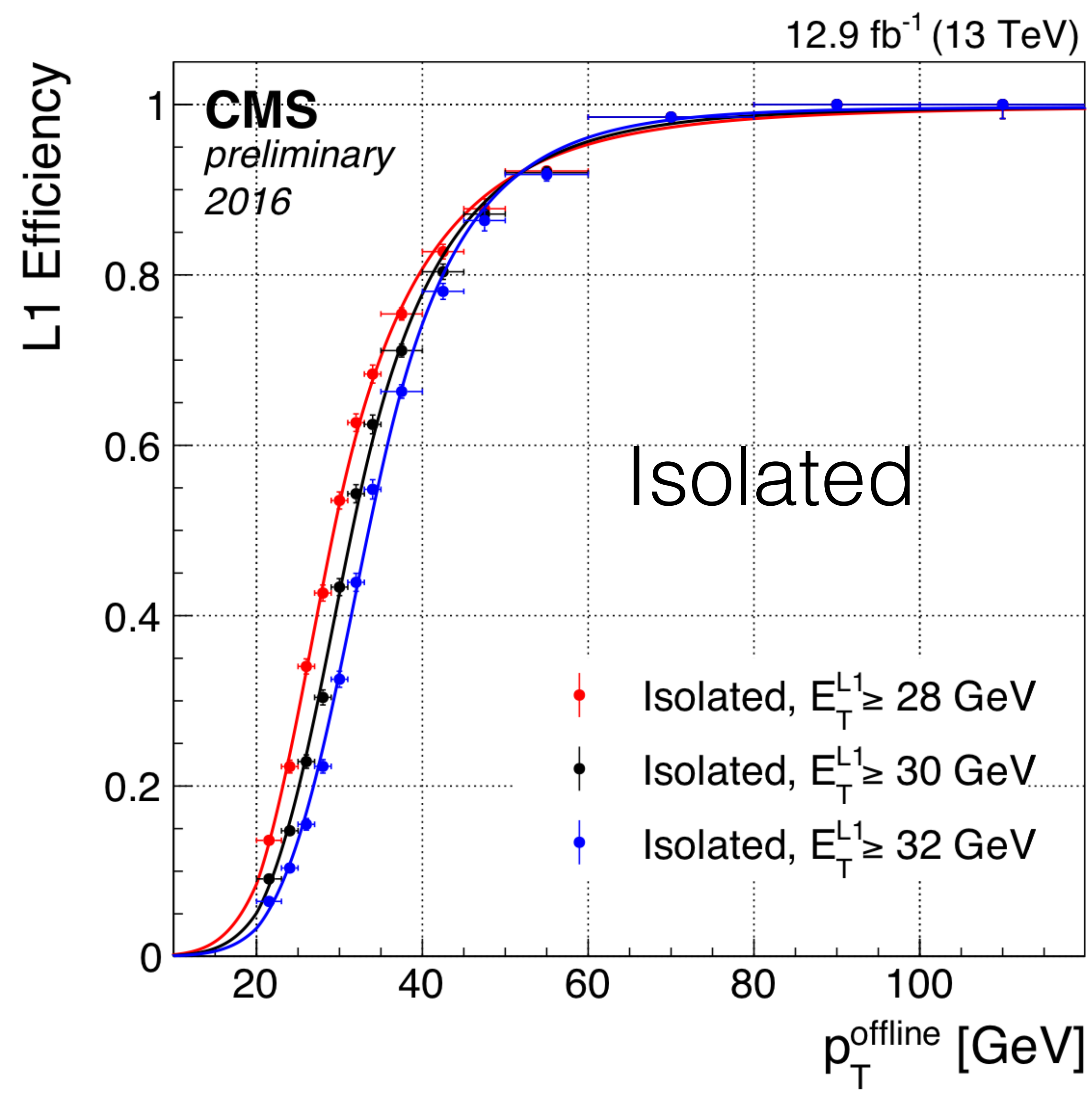




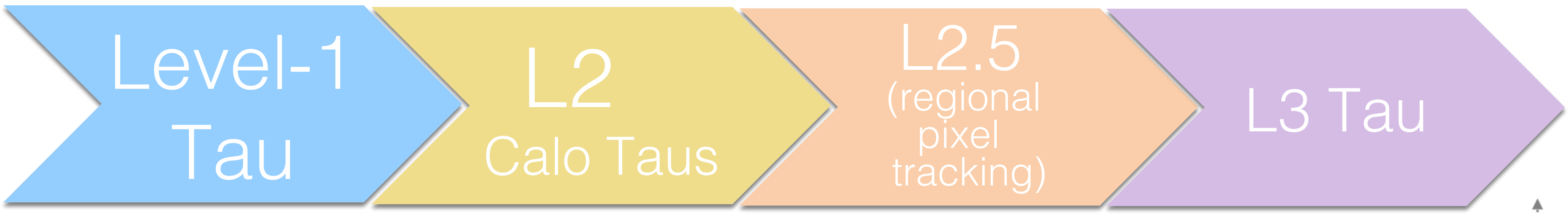
L1 tau-trigger performance

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

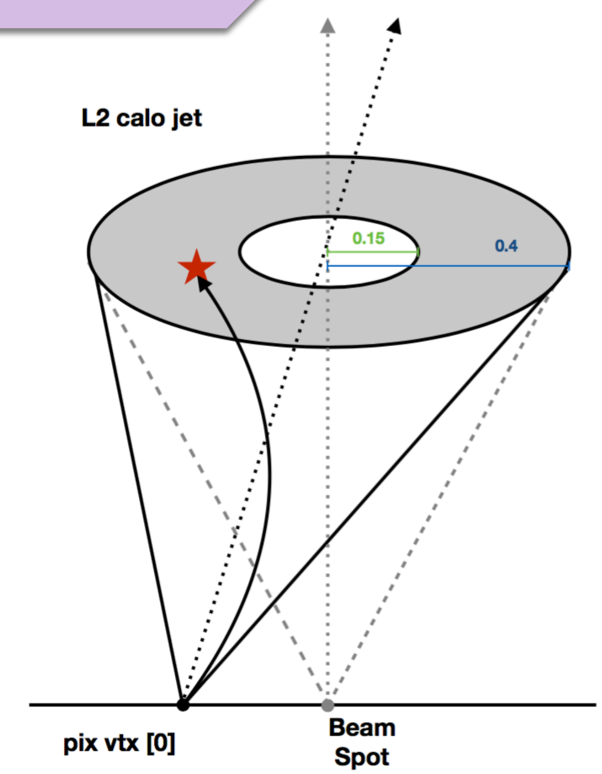


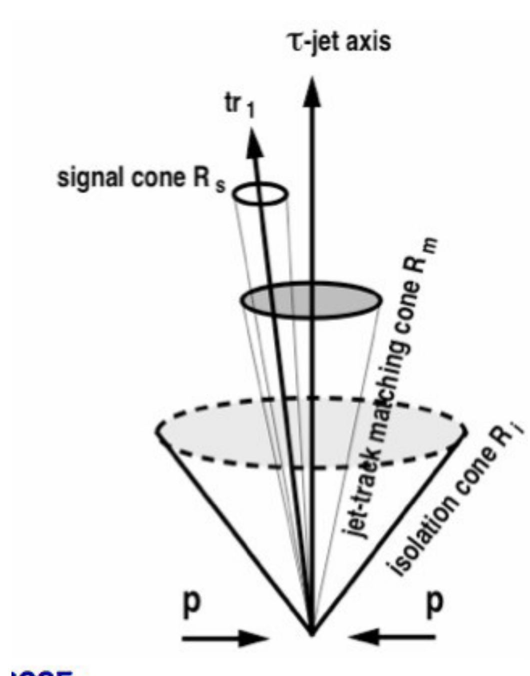


Tau identification at HLT



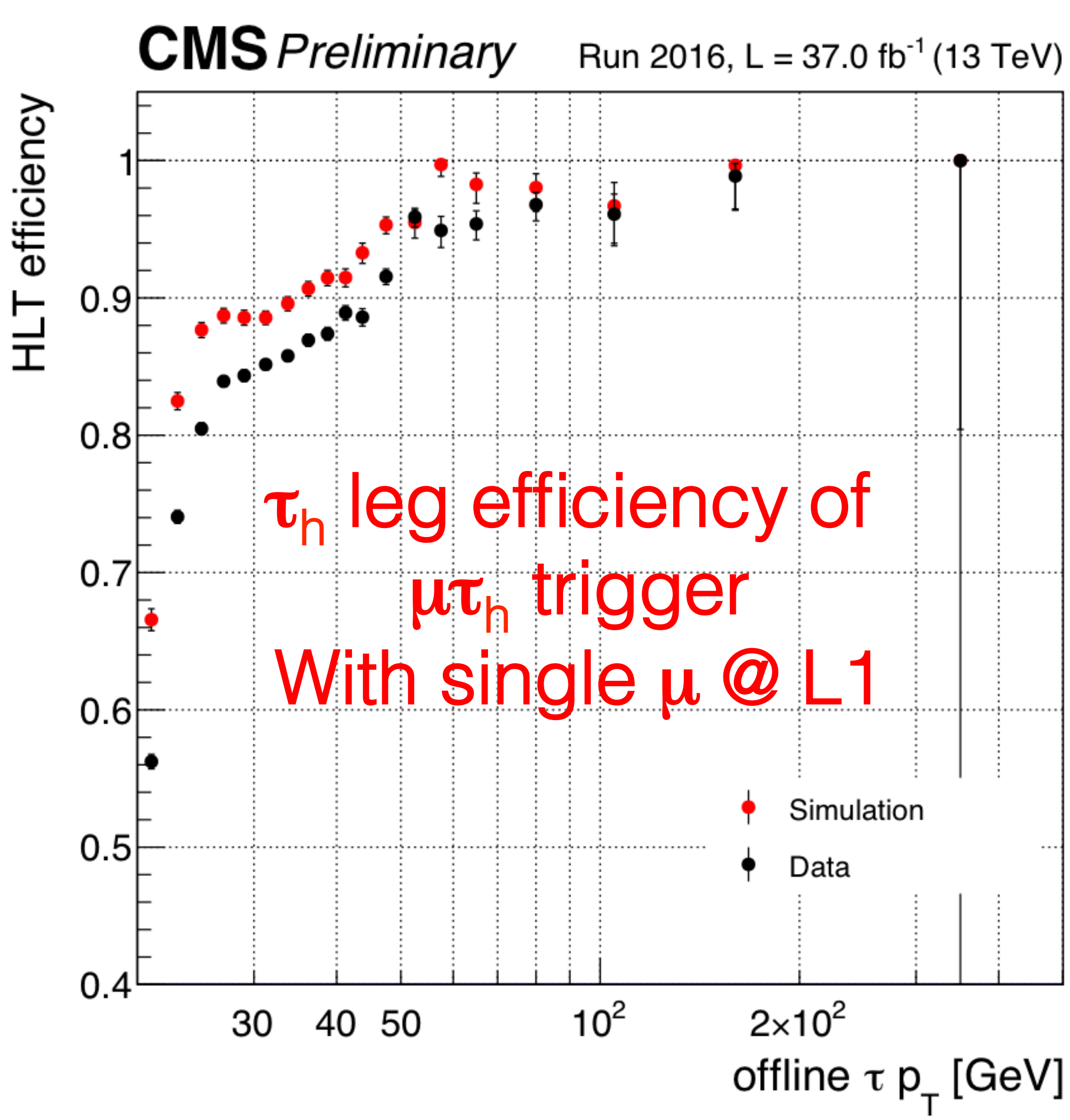
- L2 & L2.5 steps are needed in double-hadronic tau paths to reduce rate before PF in run at HLT
  - Needed to control timing
- Build L2 calo tau-jets seeded by L1 tau candidates
  - Require two calo tau-jets with $p_T > 26$ GeV & $|\eta| < 2.2$
- L2.5:
  - Regional pixel tracking around the calo taus
  - 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



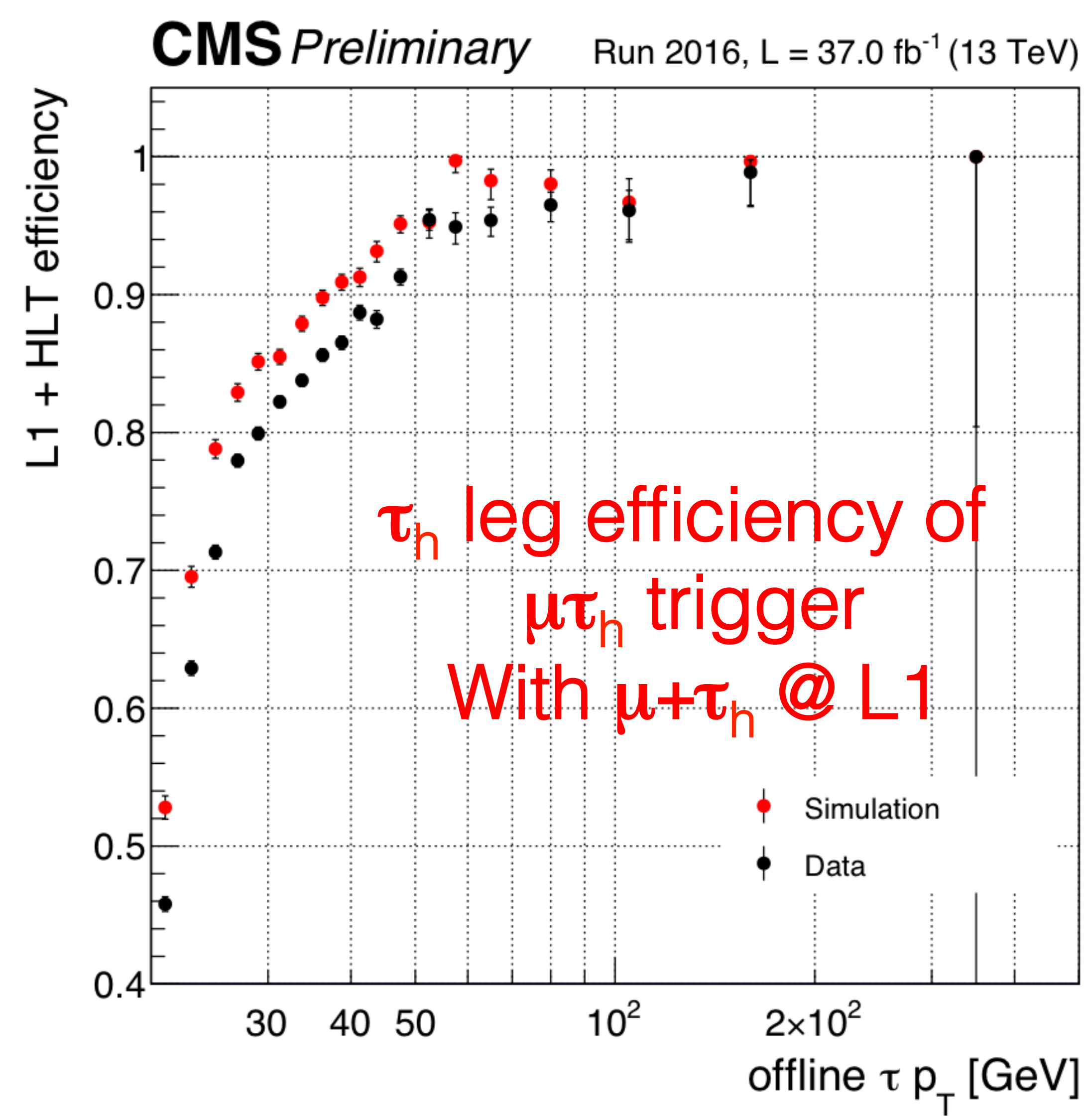


HLT tau performance

High Level Trigger efficiency of the τ leg of the $\mu\tau_h$ (loose isolation, p_T > 20 GeV, seeded by single-μ Level-1) trigger for the $H\rightarrow \tau_\mu\tau_h$ analysis

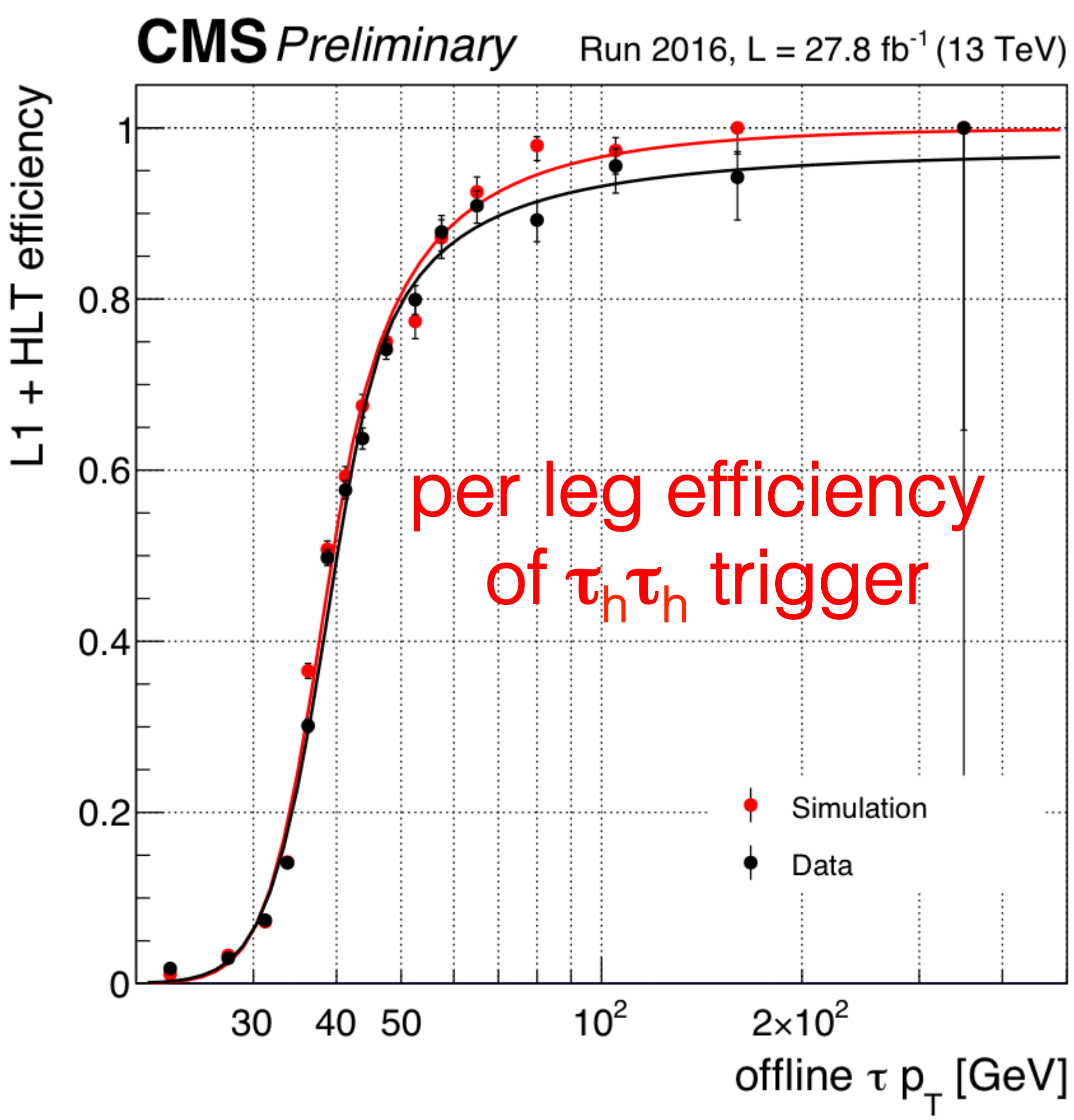


Combined L1 and High Level trigger efficiency of the τ leg of the $\mu\tau_h$ (loose isolation, p_T > 20 GeV, seeded by cross μ+τ Level-1) trigger for the $H\rightarrow \tau_\mu\tau_h$ analysis

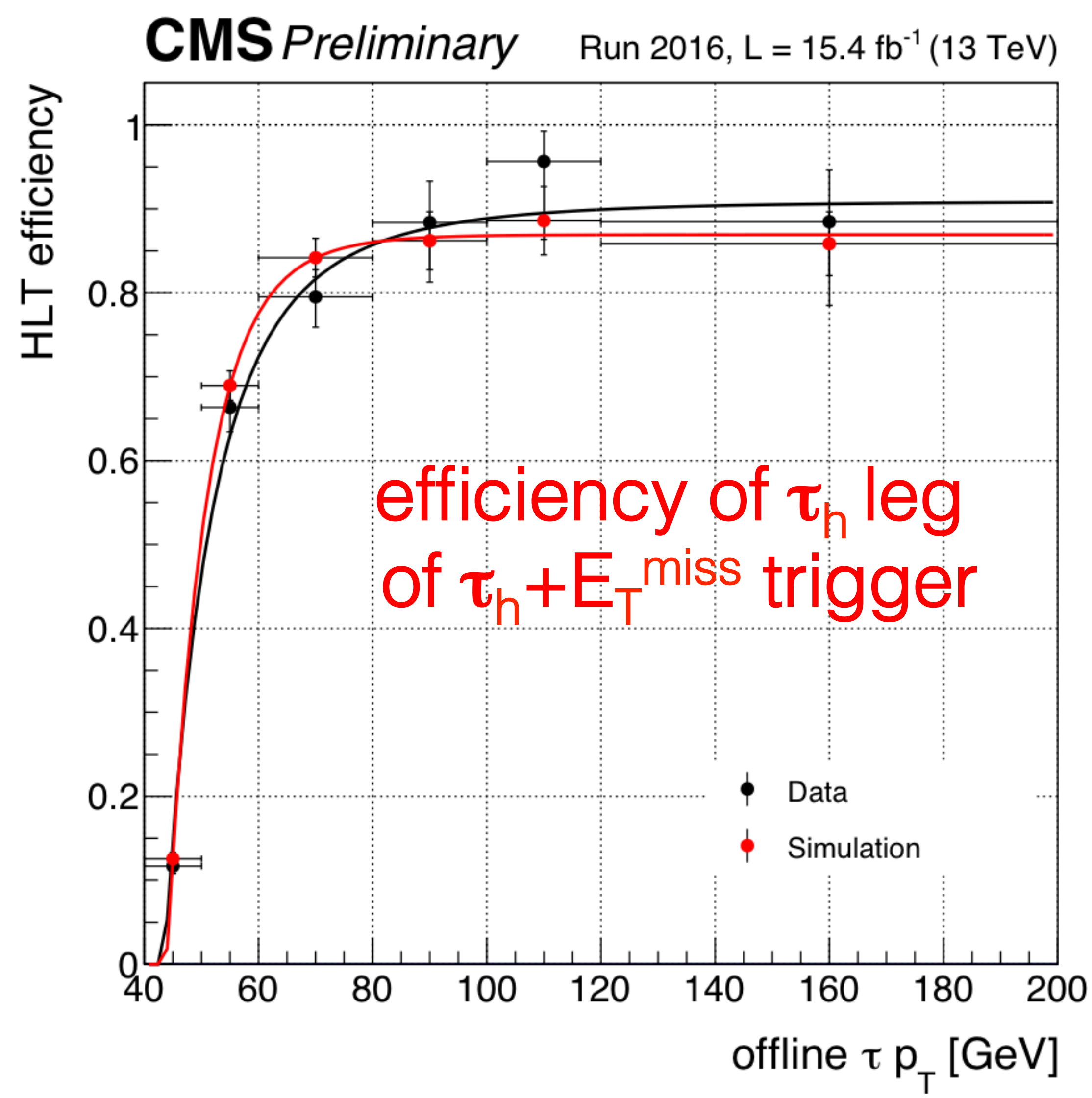


HLT tau performance

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 the $H \rightarrow \tau_h\tau_h$ 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 $H^{\pm} \rightarrow \tau_h\nu_\tau$ 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 stage-2 trigger upgrade) => Able to keep the trigger threshold similar or less than run-1
  - More studies ongoing for further improvement for 2017 data taking

BACKUP

Data/Simulation Comparison in the Semi-leptonic Final State

Data is selected with isolated single muon trigger (p_T> 22 GeV) in 2016 and correspond to an integrated luminosity of 4 /fb

Lepton Selection:

- μ: Tight identification with Isolation applied, p_T >30 GeV and |η|<2.5
- τ: Tau Identification with loose MVA based isolation, p_T >30 GeV and |η|<2.3

Pair Selection:

- ΔR(μ,τ)<0.8 and SVFit p_T(μ,τ) >200 GeV
- The lepton pair with the highest pt of the invariant visible system is selected →maximize number of events
- Veto events with additional identified leptons of the same flavor (p_T> 20 GeV)→reduce events from leptons faking taus
- Veto events with b-tagged jets medium working point→reduce events from top production
- Veto events with transverse mass (M_T (μ,τ)<50 GeV) →reduce events from W+jets background

HLTtau paths

HLT_IsoMu21_eta2p1_LooseIsoPFTau20_Single
L1 seeded by L1_SingleMu20er

HLT_IsoMu19_eta2p1_LooseIsoPFTau20
seeded by L1_Mu18er_Tau20er

HLT_DoubleMediumIsoPFTau35_Trk1_eta2p1_
Reg seeded by L1_DoubleIsoTau28er

HLT_LooseIsoPFTau50_Trk30_eta2p1_MET90
seeded by L1_ETM80

MVA ID variables

MVA ID variables