Tau ID in CMS

Christian Veelken
NICPB Tallinn

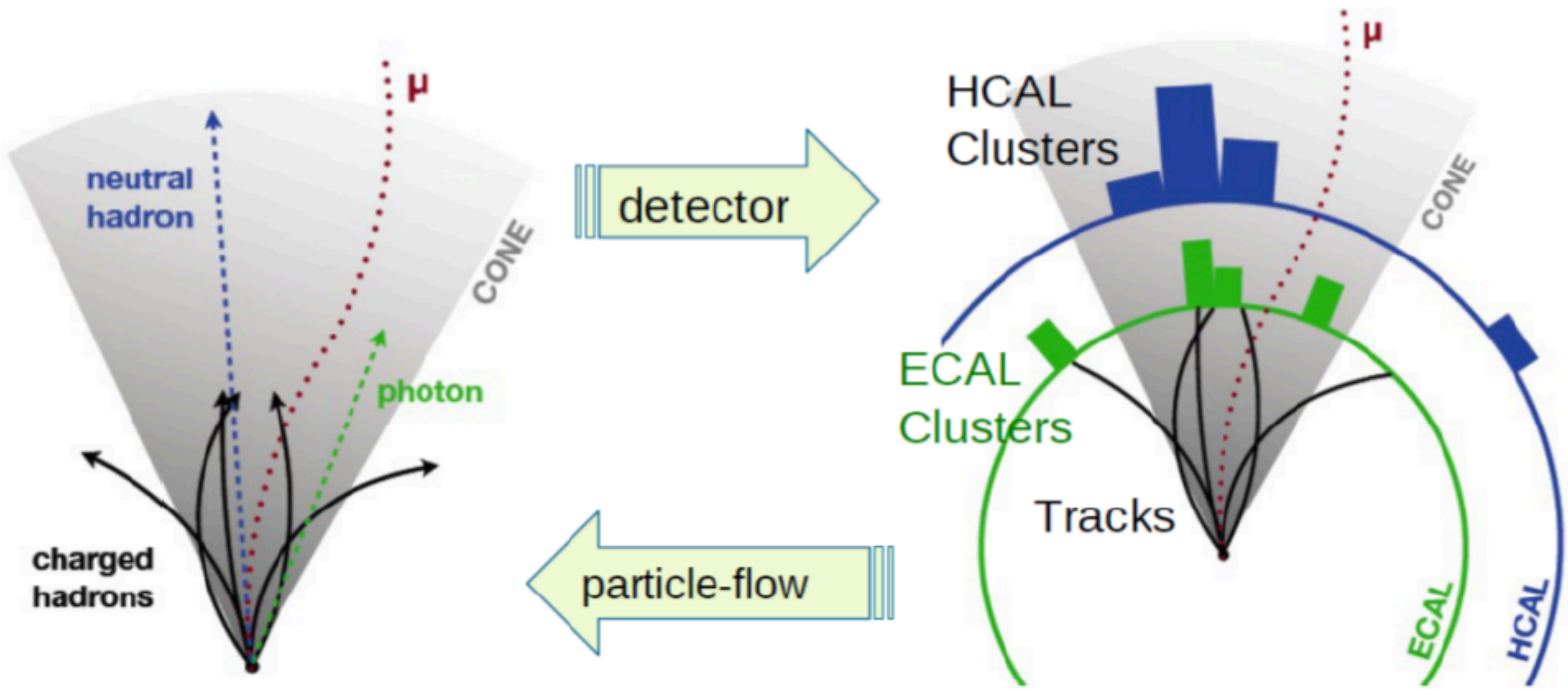
ττ Workshop DESY, November 16th 2015

Today's Presentation

- Recap: CMS tau ID
- Latest public CMS : "Reconstruction and identification of τ lepton decays to hadrons and ν_{τ} at CMS" **arXiv:1510.07488**
- Algorithm improvements and first results for run 2
- Reconstruction of boosted taus

Particle-Flow Algorithm

Consistent Interpretation of all detector signal in terms of individual particles:
e, μ, photons, charged hadrons, neutral hadrons



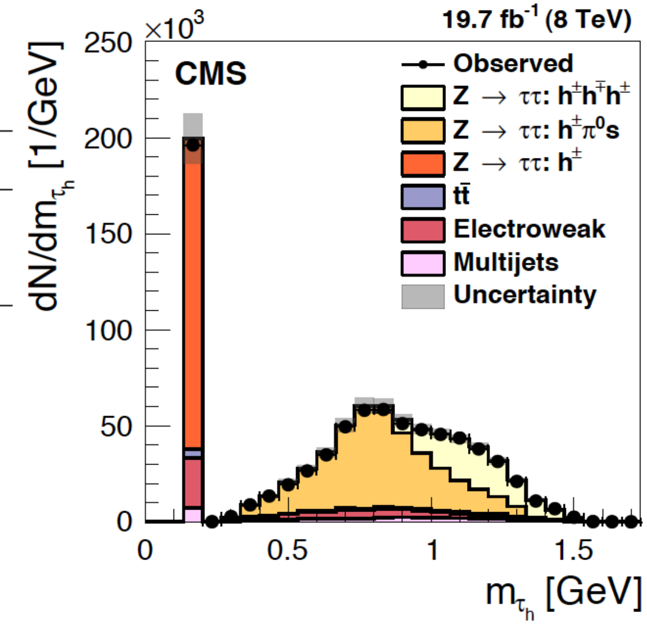
Higher level objects are reconstructed using individual particles as input:
τ_h, jets (incl. b-tagging), E_T^{miss}

Tau Decays

Mass m_τ= 1.78 GeV

Lifetime $c \bullet \tau$ = 87 μm

Decay mode	Meson resonance	$\mathcal{B}$ [%]
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
$\tau^- \to h^- \nu_\tau$		11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other modes with hadrons		3.2
All modes containing hadrons		64.8



Electrons/muons from $\tau \to e/\mu$ decays reconstructed by standard CMS electron/muon reconstruction

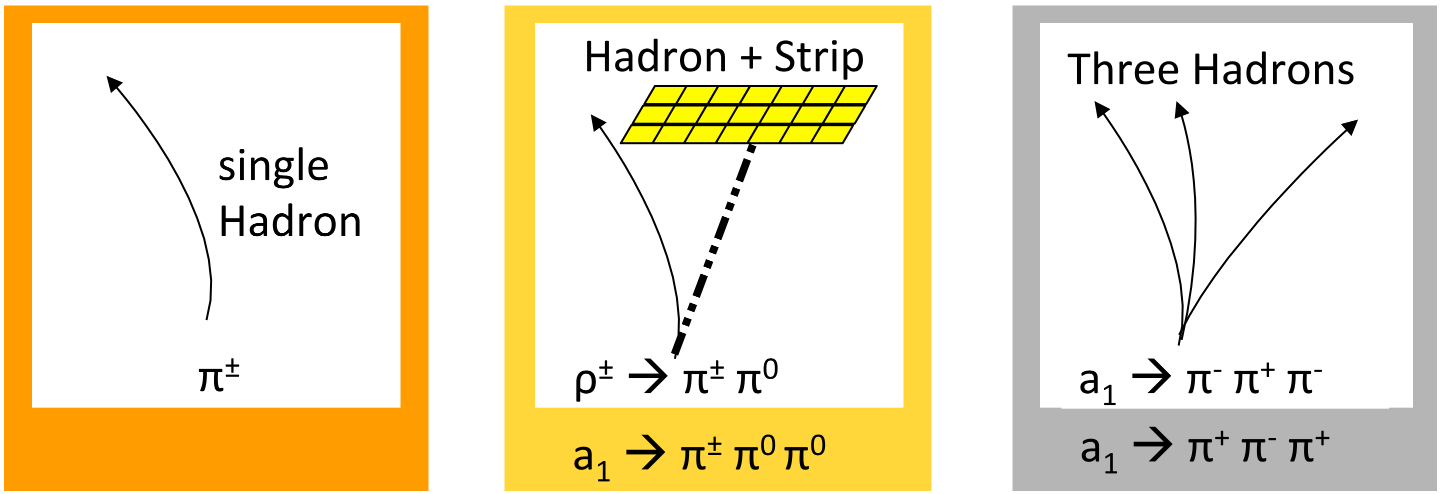
τ_h Identification ≅ reconstruction of $\pi^\pm$, $\rho^\pm$, $a_1^\pm$ signatures

Acceptance of τ_h reconstruction: $P_T^\tau \gtrsim 20$ GeV and $|\eta_\tau| < 2.3$ typically used by physics analyses in CMS

"Hadron plus Strips" (HPS) Algorithm

① Seeded by anti-k_T (R = 0.5) jets build from particle-flow output

② Build combination of charged Hadrons + Strips (= τ_h candidate)



③ Select combinations passing mass window cuts for $\pi^{\pm}$, $\rho^{\pm}$, a_1

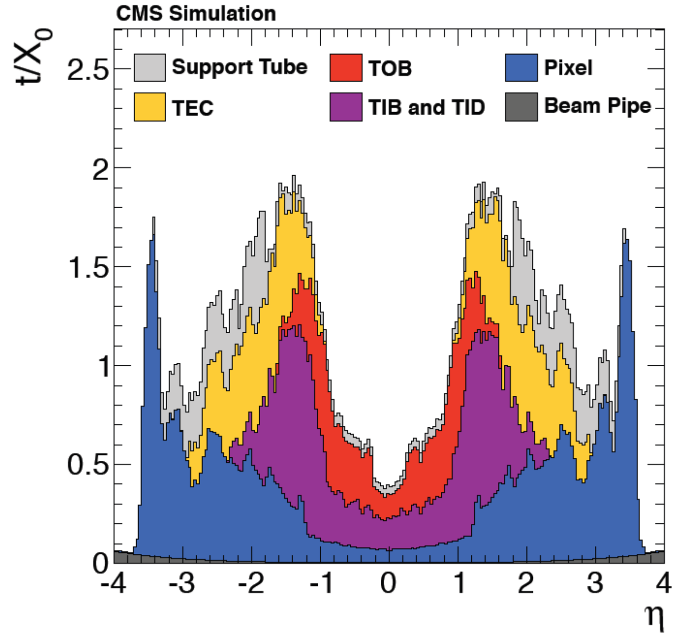
④ Most isolated τ_h kept in case multiple combinations pass mass cuts

⑤ Apply cut-based isolation or MVA based τ_h identification

⑥ Compute discriminators against e and μ

HPS π^0 Reconstruction

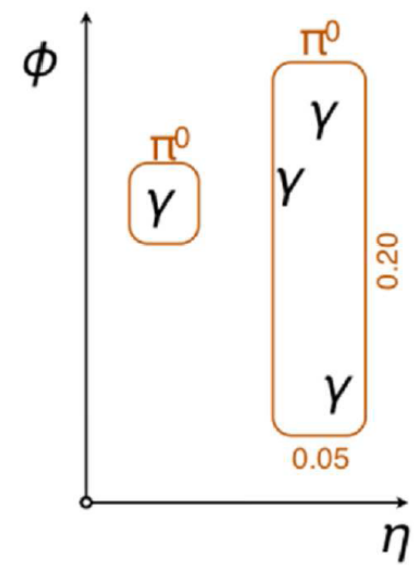
Neutral pions decay via $\pi^0 \rightarrow \gamma\gamma$ almost instantaneously



CMS has all silicon tracking detector

➔ High probability for photons to convert

➔ 3.8 T magnetic field separates e^+e^- pair in ϕ

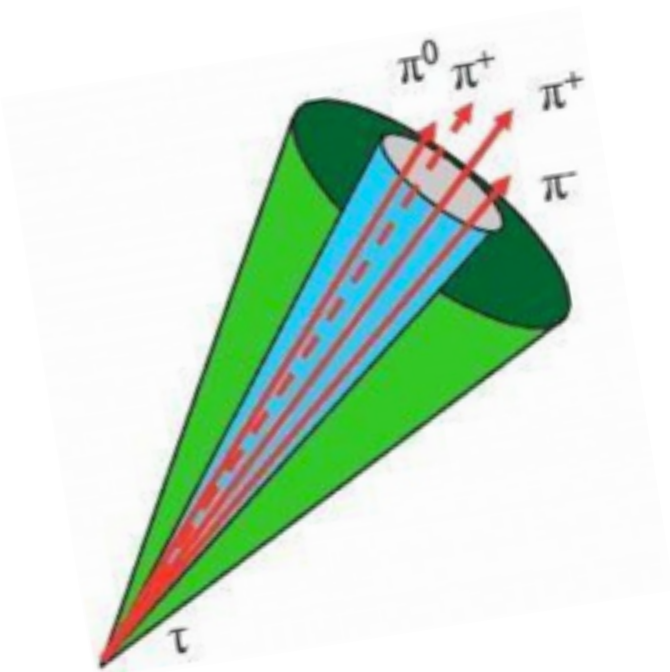


➔ Reconstruct (converted) photons in η-ϕ strips

Strips may contain either both photons from π^0 decay or just one

Cut-based Tau Isolation

$$\text{Iso} = \underbrace{\sum P_T^{h\pm}(|dz| < 2\text{mm})}_{\text{charged particles from PV}} + \max(0, \overbrace{\sum P_T^{\text{neutral}}}^{\text{neutral particles}} - \underbrace{k}_{\text{Factor chosen to make Iso flat vs. number of pileup interactions}} \bullet \underbrace{\sum P_T^{h\pm}(|dz| > 2\text{mm})}_{\text{charged particles from pileup}})$$



Different cuts on Iso define loose, medium and tight tau ID working-points:

WP	Requirement
Loose	Iso < 2.0 GeV
Medium	Iso < 1.0 GeV
Tight	Iso < 0.8 GeV

All 3 WPs are actually quite tight: necessary to reduce large QCD background

MVA Tau ID

Idea: Use tau lifetime information to improve τ_h ID performance

Input Variables:

- ΣP_T(charged contribution).
- ΣP_T(neutral contribution).
- Pile up correction.

Cut based isolation variables

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

*Used to parameterize P_T and η dependence of the MVA input variables

τ reco and kinematic variables

- abs(d_{xy}) of the τ .
- sign(d_{xy}) of the τ .
- d_{xy} significance.

Variables sensitive to the impact parameter

- Presence of a secondary vertex.
- τ flight length.
- τ flight length significance.

SV sensitive variables

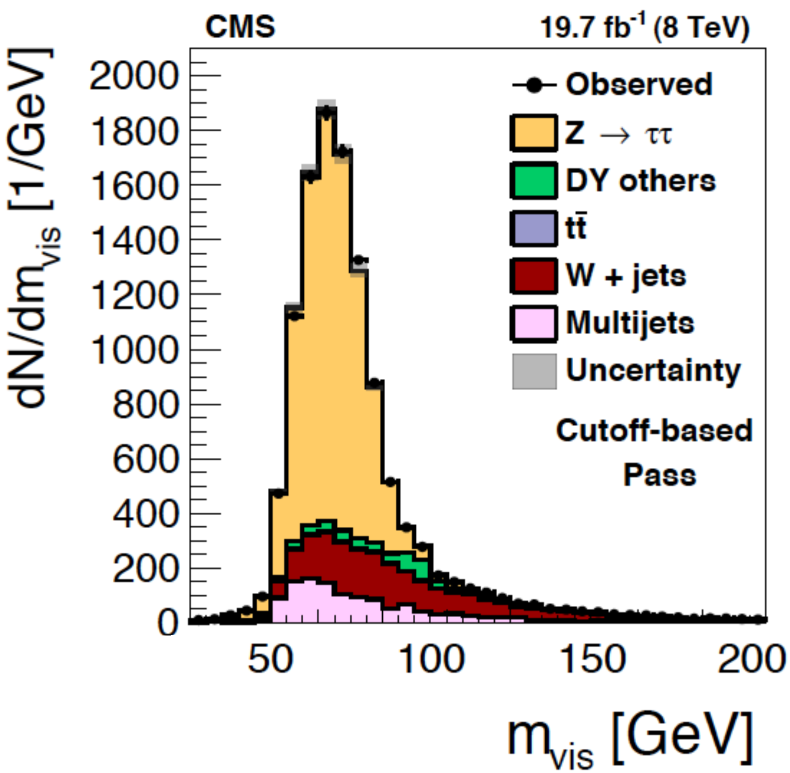
→ Loose, medium, tight WPs defined by cut on BDT output

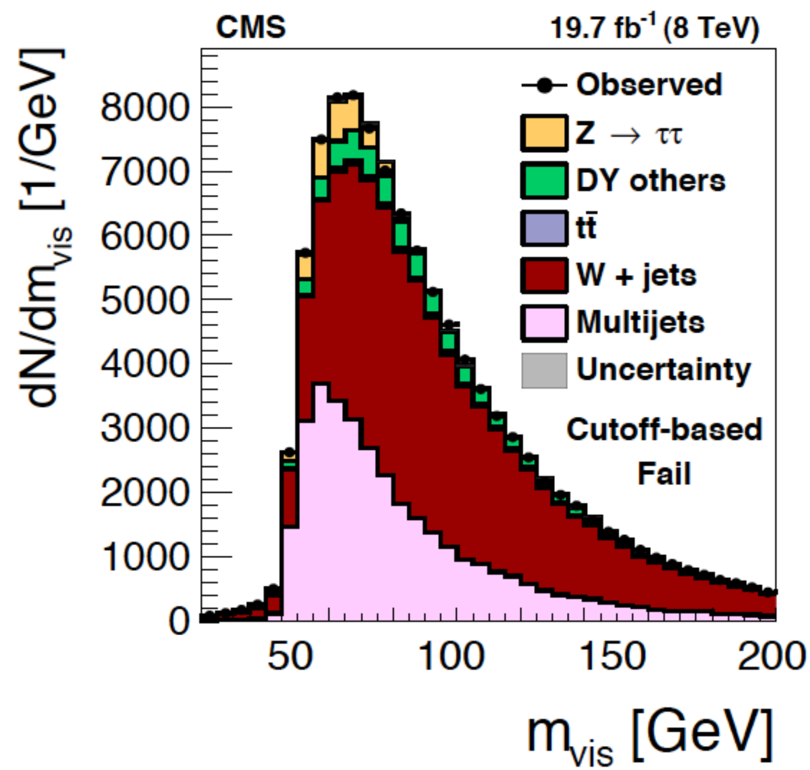
Tau ID Efficiency Measurement

Select sample of Z → ττ → μτ_h candidate events without applying tau ID

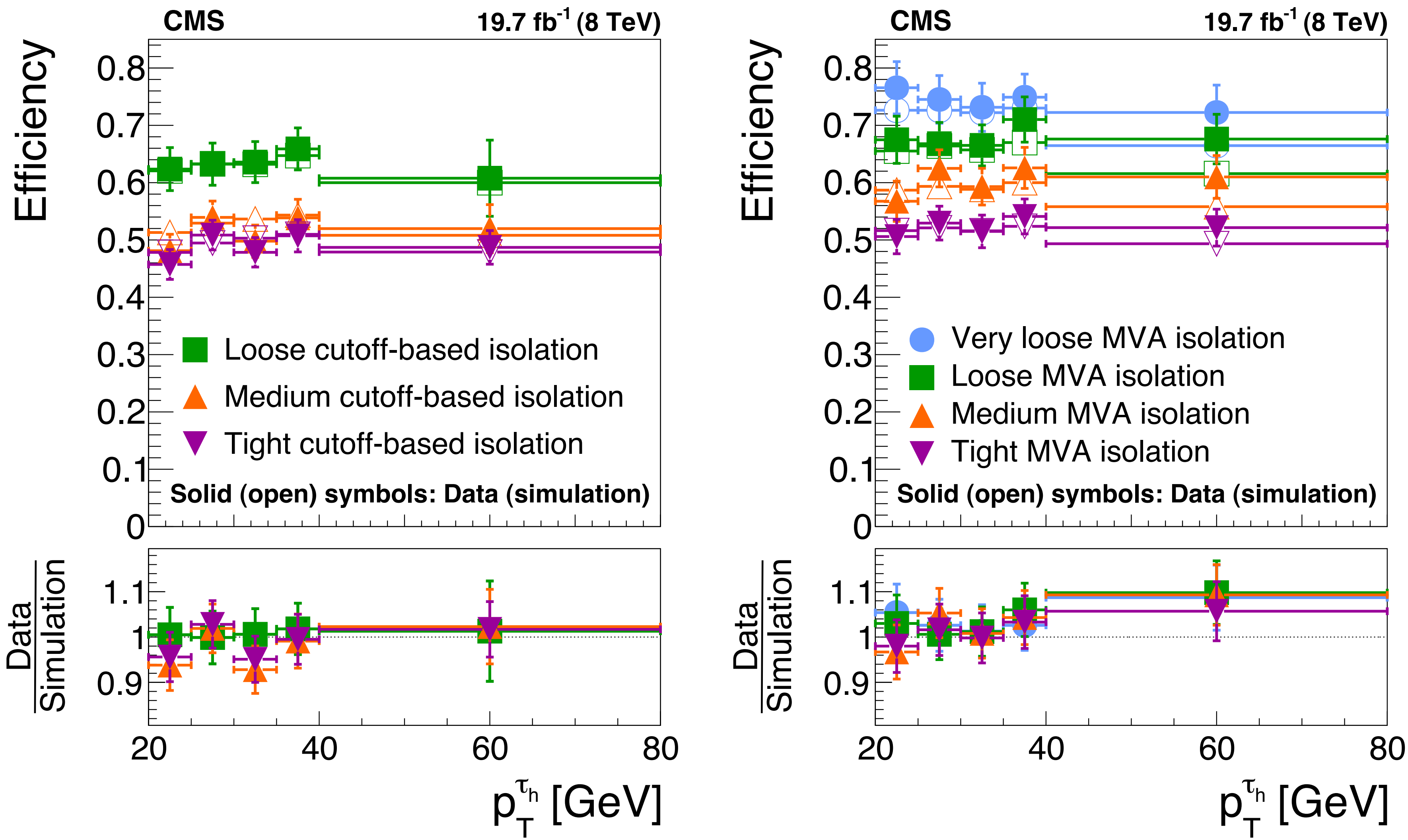
Challenge: sizeable background contributions

➔ Determine tau ID efficiency by simultaneous fit of μτ_h visible mass distribution in "Pass" and "Fail" regions with shape templates for Z → ττ → μτ_h signal and background processed





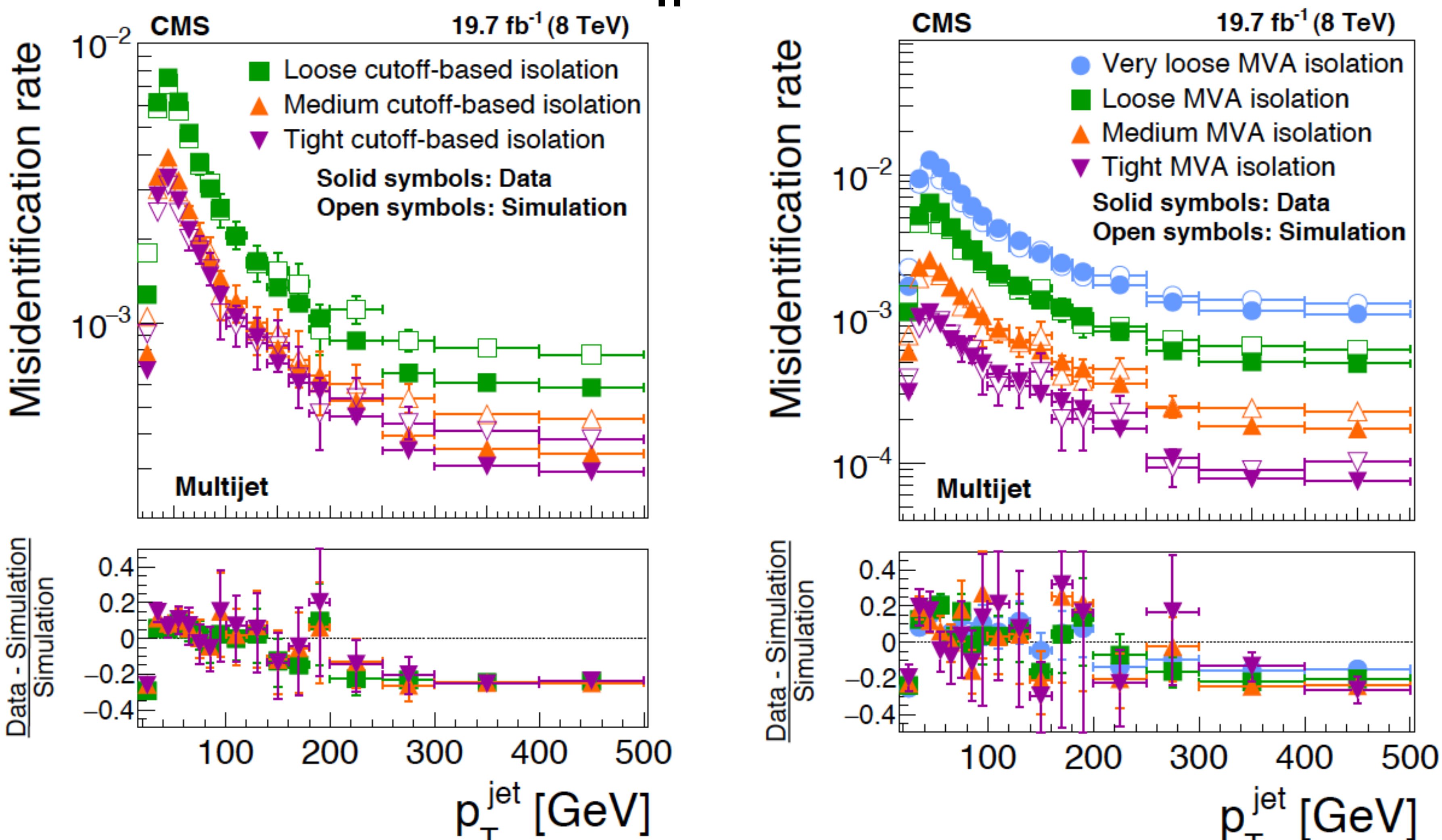
Tau ID Efficiency



➔ Tau ID efficiency 50-60%, flat vs. P_T for $P_T \geq 40$ GeV

➔ Efficiency measured in data in good agreement between MC simulation

Jet → τ_h Fake-rate

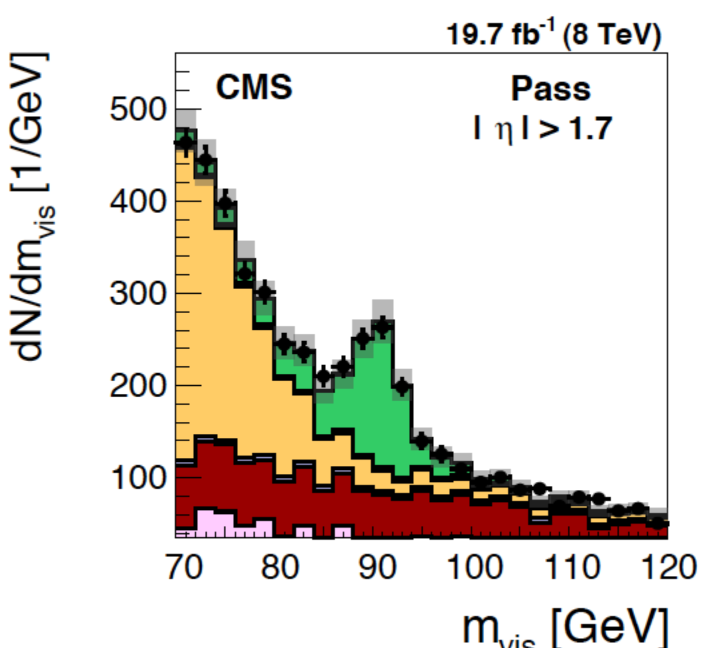


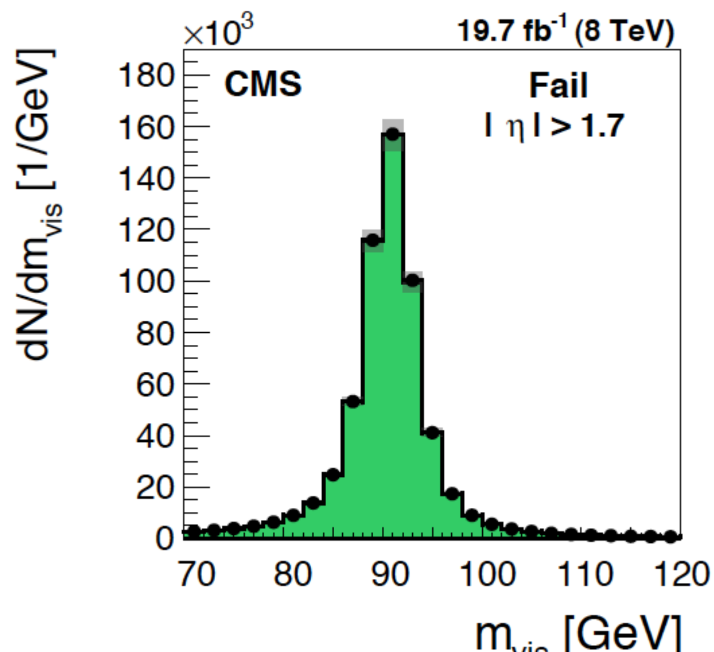
➔ Jet → τ_h fake-rate varies between 10^{-2} and 10^{-4} level, steeply falling with P_T

➔ Up to 20% difference between jet → τ_h fake-rate measured in data and MC simulation

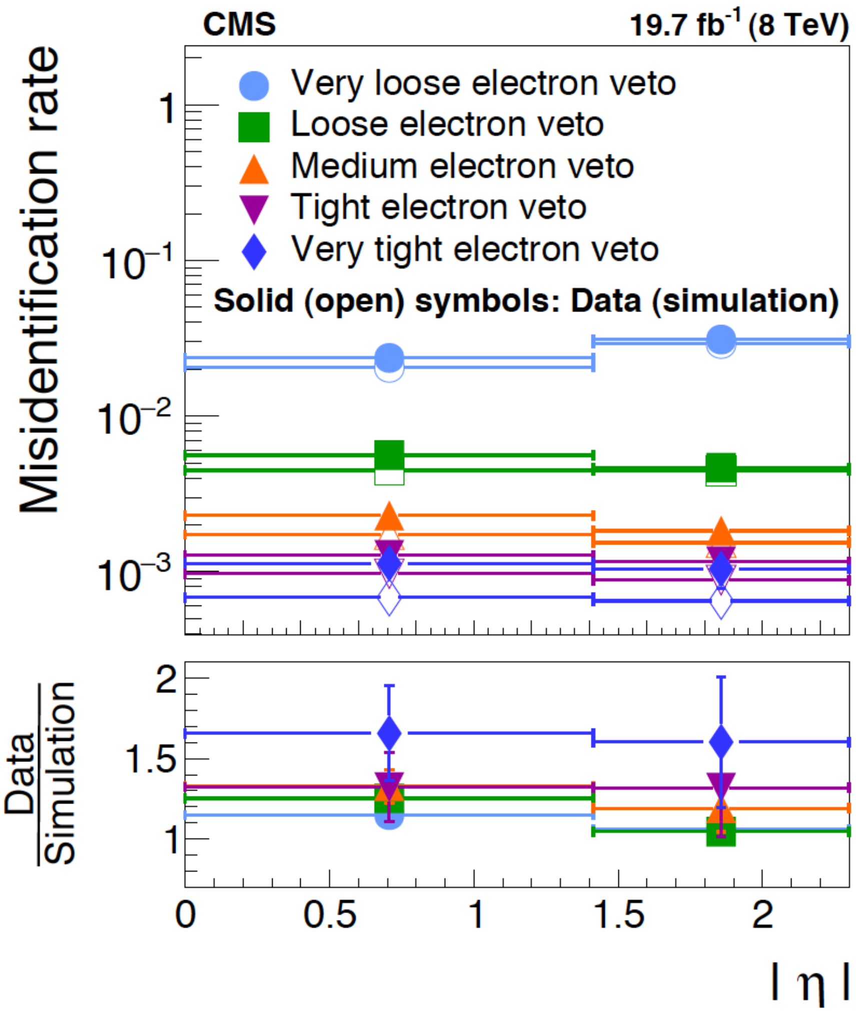
Measurement of e → τ_h and μ → τ_h Fake-rates

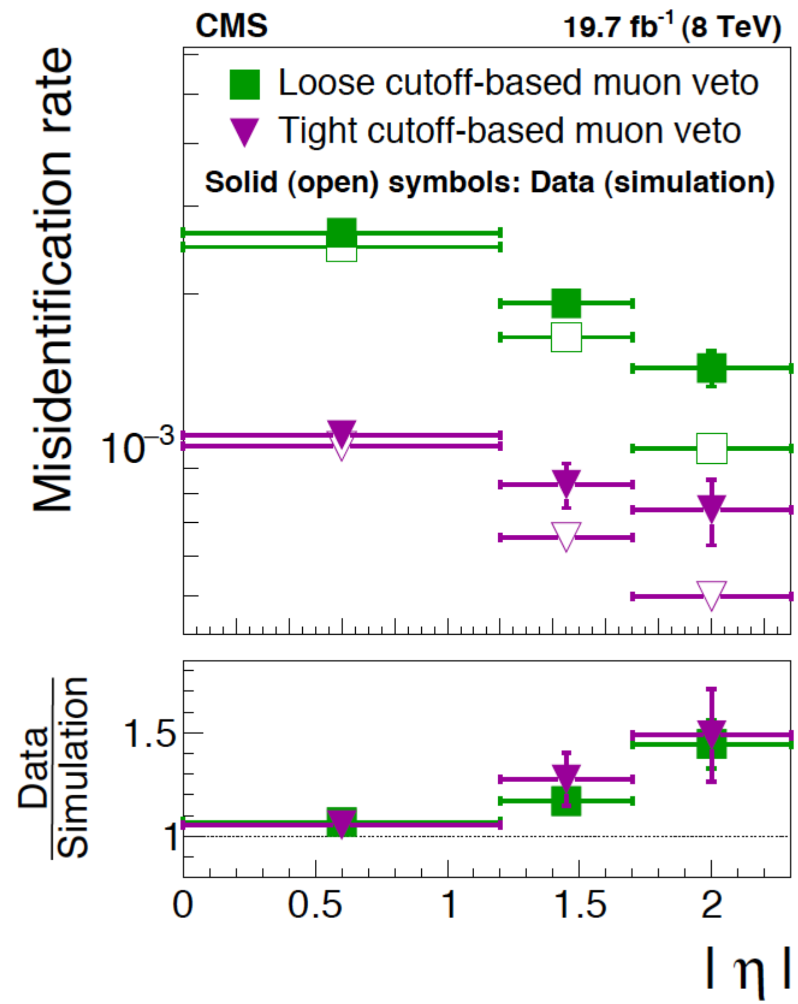
- Select Z/γ* → ee (Z/γ* → μμ) events passing single electron (single muon) trigger and containing at least one Tag and Probe pair
- Tag: well identified and isolated electron (muon) of $P_T > 30$ GeV ($P_T > 25$ GeV) && $|\eta| < 2.1$
- Probe: τ_h candidate of $P_T > 20$ GeV && $|\eta| < 2.3$, reconstructed by HPS algorithm, and of opposite charge as the Tag
- Fake-rate obtained from simultaneous fit of Tag+Probe mass distribution with shape templates for Z/γ* → ee (Z/γ* → μμ) signal and background processes





$e \rightarrow \tau_h$ and $\mu \rightarrow \tau_h$ Fake-rates





- ➔ $e \rightarrow \tau_h$ and $\mu \rightarrow \tau_h$ fake-rates reach per mille level
- ➔ Differences between data and MC simulation corrected for by applying suitable scale-factors on analysis level

Tau ID Improvements for Run 2

Tau decay mode reconstruction and charged isolation is affected little by pileup (PU) expected for run 2

➔ Preparation for run 2 concentrated on optimizing neutral isolation:

- ECAL timing cuts
- PU-weighted isolation

Backup

- Isolation based on "Pileup Per Particle Identification" (PUPPI) algorithm

JHEP 1410 (2014) 59

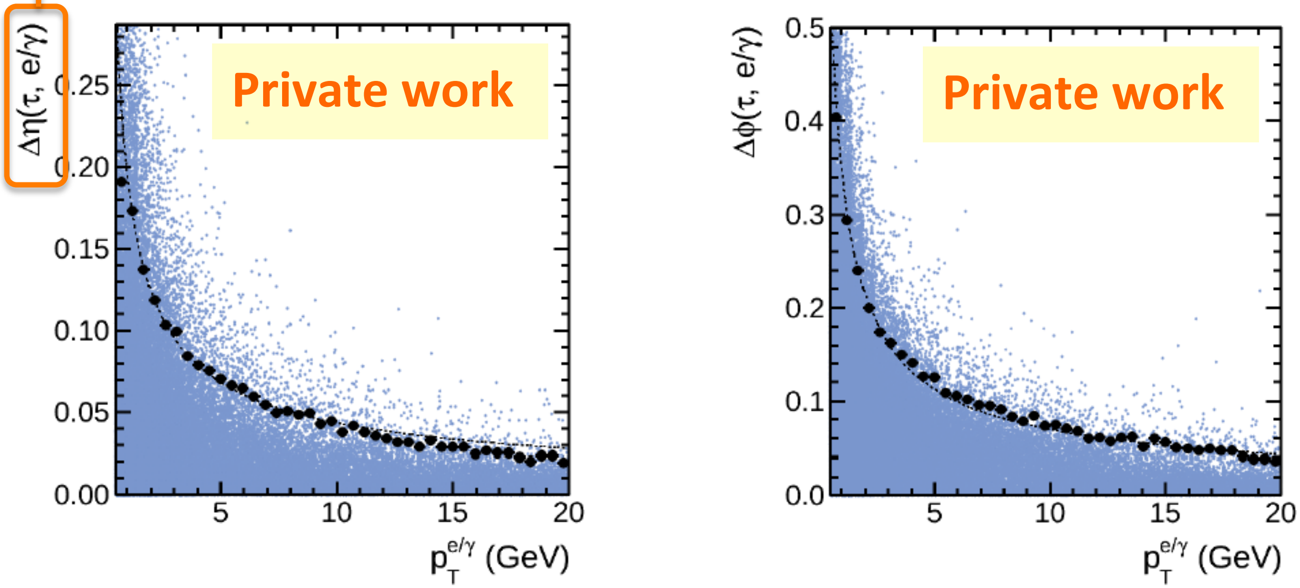
While optimizing the neutral isolation for run 2, we noticed that some ECAL energy deposits that are due to τ decay products fail to get included in the τ_h reconstruction

➔ Effect was on the level of few GeV, but caused significant decrease of tau ID efficiency in case tight neutral isolation cuts are applied

Dynamic Strip Reconstruction

Effect: Bending within 3.8 T magnetic field increases for e^+e^- pairs produced in photon conversions if electrons and positrons have low p_T

Distance between ECAL energy and generator level τ_h direction



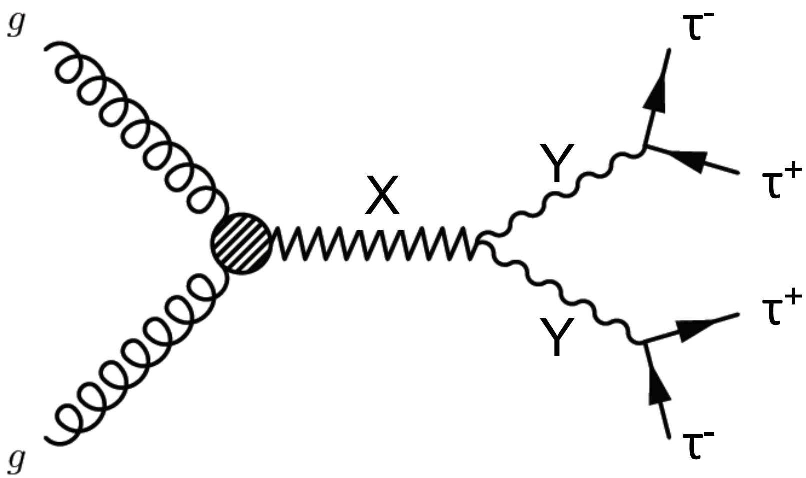
➔ Increase strip size in case photons have low p_T

Dynamic strip reconstruction is the default algorithm for run 2

➔ Allows to tighten neutral isolation cuts, yielding O(20%) reduction in jet → τ_h fake-rate for same tau ID efficiency as during run 1

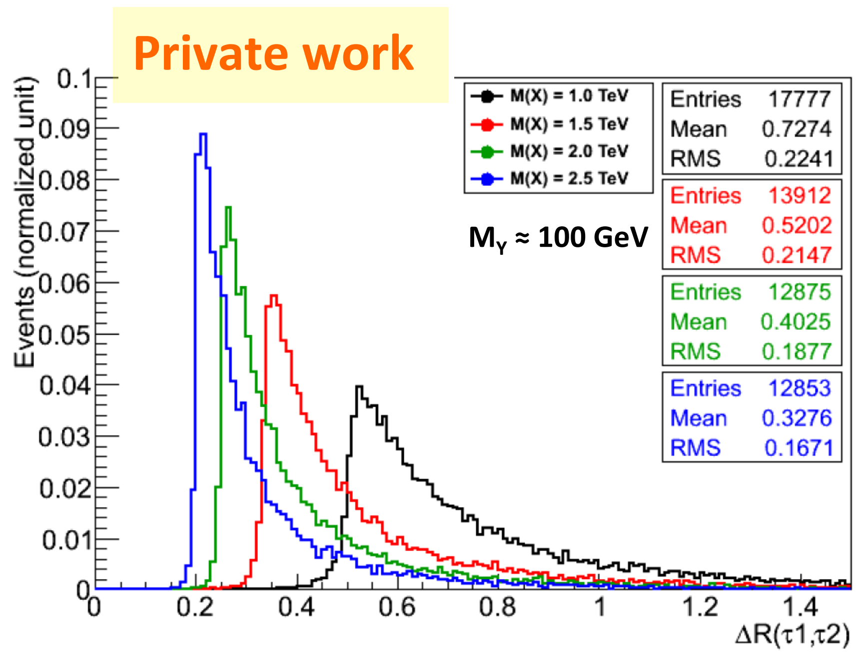
Reconstruction of "boosted" Taus

Taus produced in cascade decays of resonances of high mass($M_X \gg M_Y$) may overlap within dR < 0.5 in the detector, spoiling each other's isolation



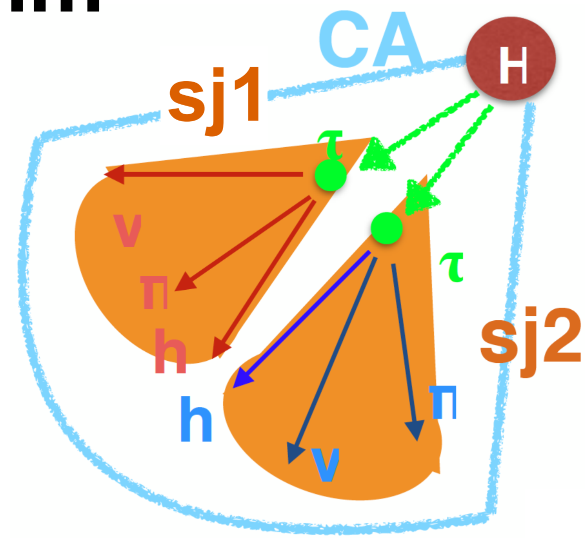
Angular separation between taus from Y decay approximately

$$\Delta R = \frac{2.0 \cdot M_Y}{P_T^Y} \quad , P_T^Y \approx \frac{M_X}{2}$$



➔ Efficiency of "standard" tau isolation decreases significantly when $M_X \geq 10 \cdot M_Y$

Boosted Tau Algorithm



Most analyses of tau lepton production include events with $\tau \rightarrow e\nu\nu$ and $\tau \rightarrow \mu\nu\nu$ decays

➔ Looking for generic approach that handles hadronic tau decays as well as decays to e/μ

Solution:

① Use subjet techniques to find the decay products of the 2 taus in a FatJet (decay products can be either e, μ or τ_h)

NB.: $M_\tau \ll M_Y$

➔ Mass-drop criteria can be very powerful

② Run standard tau reconstruction algorithm on each subjet

③ Match identified electrons and muons to subjets (e.g. by $\Delta R < 0.1$)

④ Compute e, μ and τ_h isolation ignoring particles associated to 2nd subjet within the same FatJet

Status of Boosted Taus in CMS

Algorithm has been used by search for Z' → Zh → qqττ in run 1

Phys.Lett. B748 (2015) 255

- Subject finding efficiency = 92%
  (This efficiency needs to be "paid" only once for a pair of 2 taus)

- Efficiency to pass e, μ, τ_h isolation is a few % lower than for non-boosted taus ("standard" tau reconstruction and isolation do work)

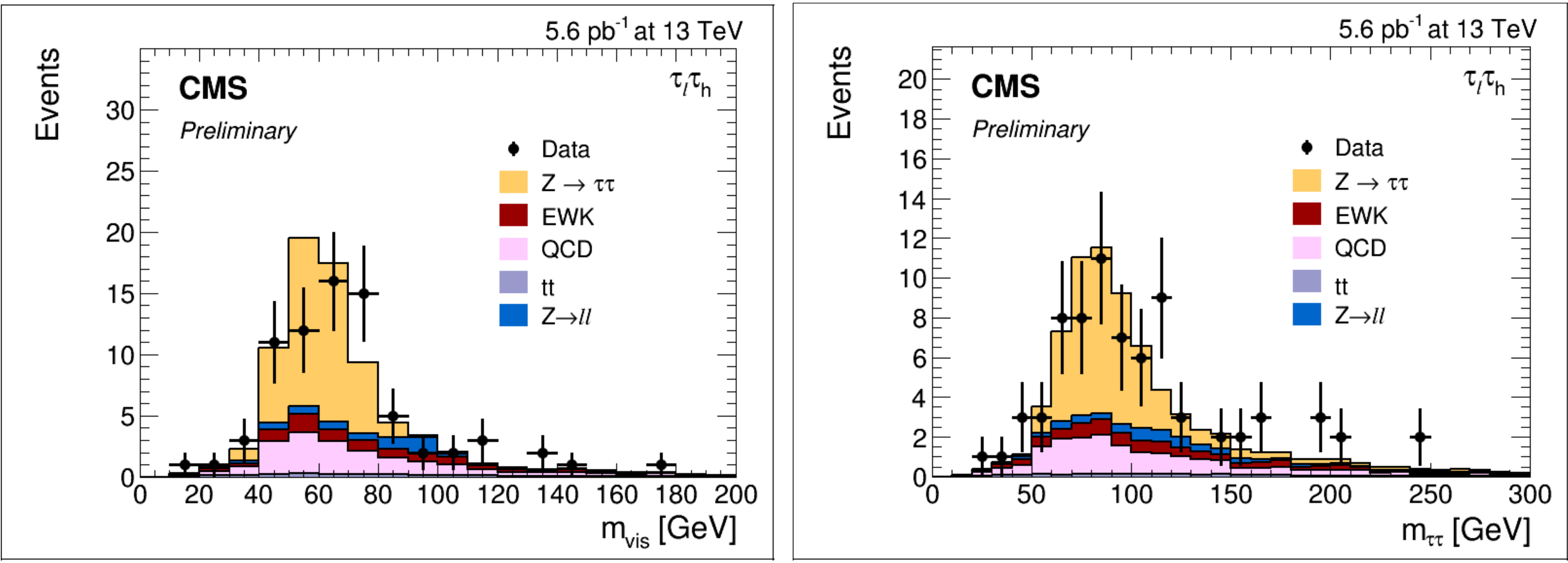
Plan: Boosted tau algorithm will be officially supported for CMS run 2 analyses

At the moment, the software is being finalized

How to validate the algorithm with data (in terms of efficiency and fake-rate) requires more though

$Z \rightarrow \tau\tau$ Signal in Run 2 Data

CMS-DP-2015/016



First $Z \rightarrow \tau\tau$ peak in CMS run 2 data, shown at summer conferences

Events selected in $e\tau_h$ and $\mu\tau_h$ channels (triggered by single lepton triggers)

QCD estimated from data, $Z \rightarrow \tau\tau$ signal and all other backgrounds taken from MC

➔ Tau ID works as expected in run 2 data (and SVfit works as expected too)

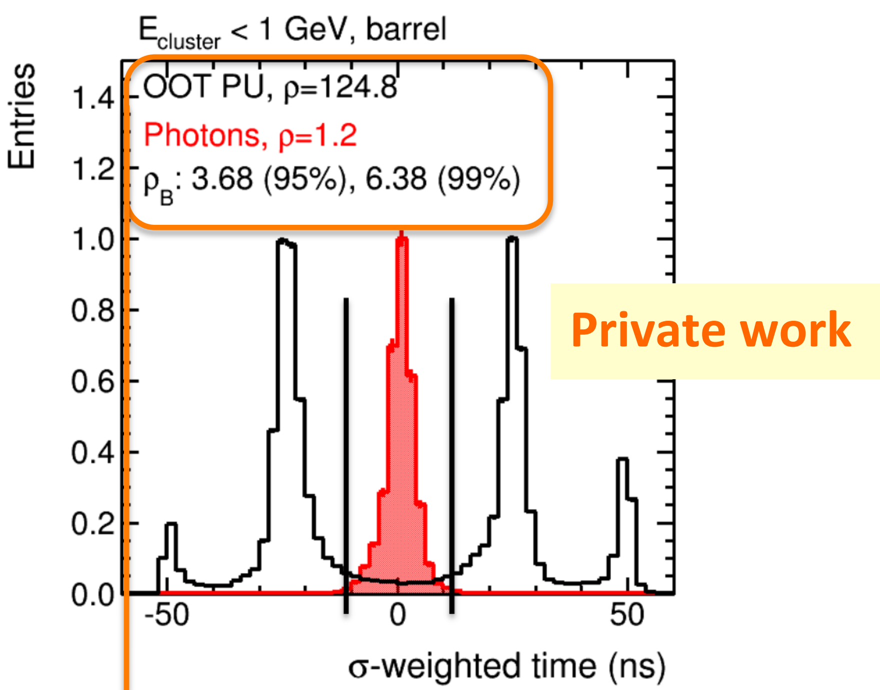
Summary

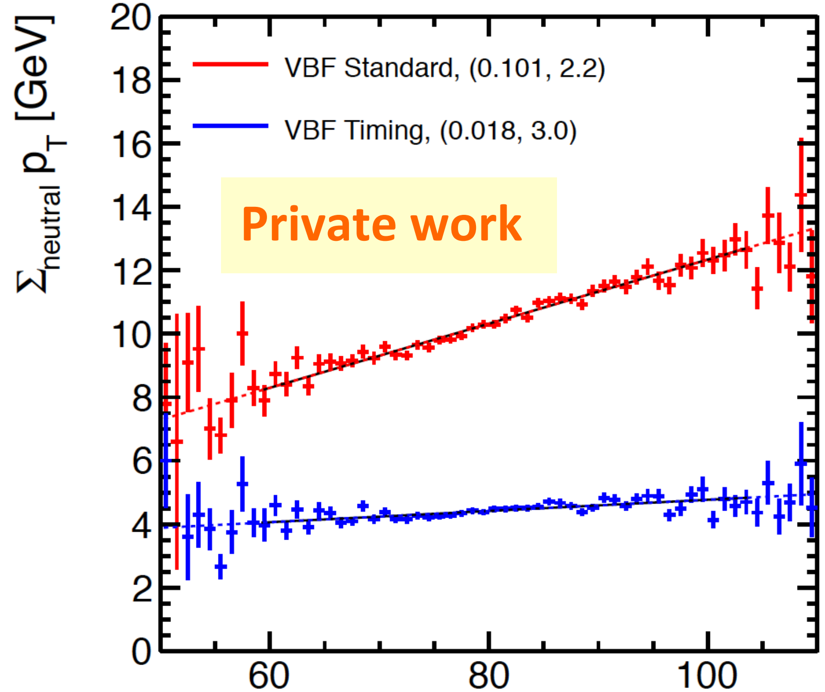
- The "Hadron plus Strips" (HPS) tau ID algorithm used during run 1 has recently been published **arXiv:1510.07488**
- The particle-flow and HPS algorithms will continue to be used for tau reconstruction and identification by CMS in run 2
- The dynamic strip reconstruction allows to tighten the neutral isolation cuts, reducing the jet → τ_h fake-rate by O(20%) for the same tau ID efficiency
- The MVA based tau identification algorithms have been retrained for run 2
- In summary, the CMS tau ID is prepared for run 2 and expected to provide a similar performance as in run 1

Backup

ECAL timing cuts

Neutral particles due to pileup from different bunch-crossings ("out-of-time" pileup) can be removed efficiently using ECAL timing information





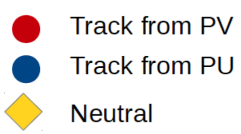
Number PU interactions in previous plus subsequent bunch-crossing

Energy due to out-of-time PU reduced by factor 20 (34) while keeping 99% (95%) of energy from current bunch-crossing

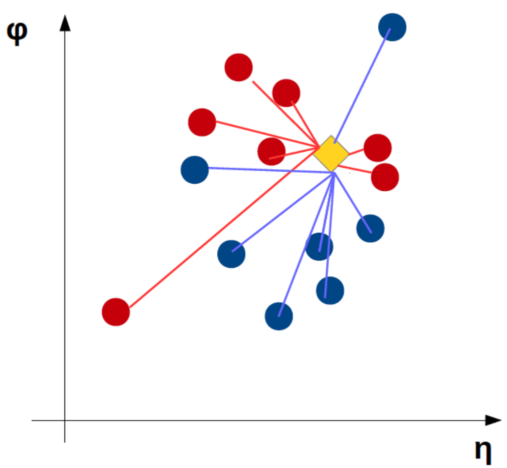
➔ Effect of out-of-time pileup reduced to small level

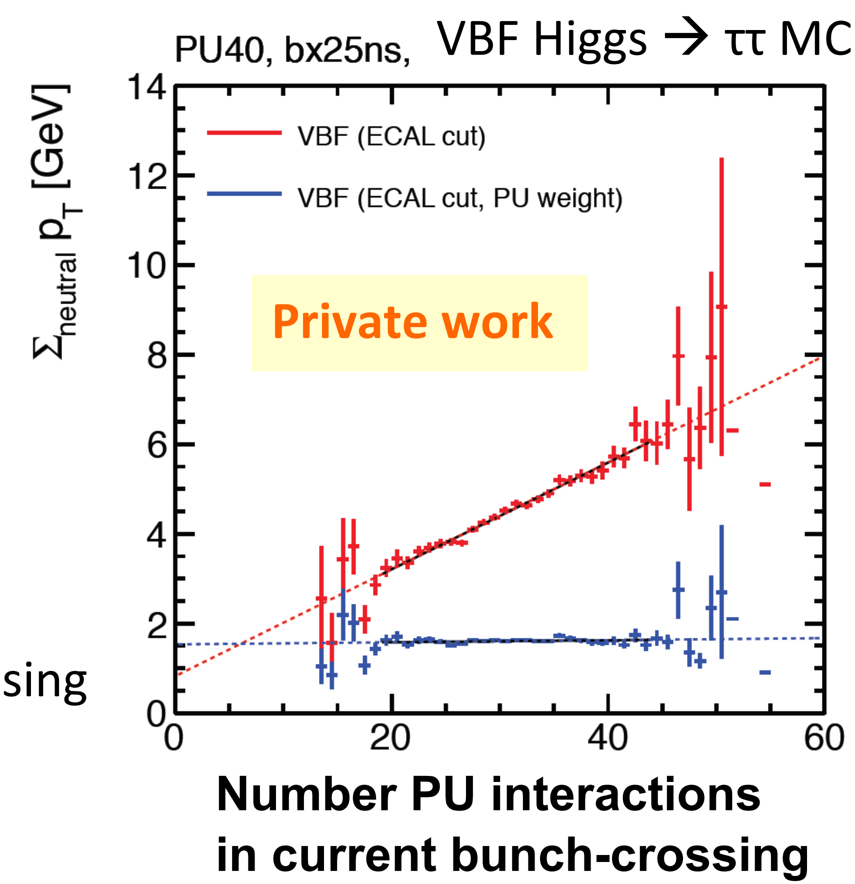
PU-weighted Isolation

Idea: weight neutral particles in isolation cone by local energy density of particles originating from PV and pileup (PU), estimated using tracks



$$w_{i,\log(p_T/\Delta R)} = \frac{\sum_{\mathrm{PV},j} \log(p_T/\Delta R_{ij})}{\sum_{\mathrm{PV},j} \log(p_T/\Delta R_{ij}) + \sum_{\mathrm{PU},k} \log(p_T/\Delta R_{ik})}$$





➔ Effect of pileup in current bunch-crossing ("in-time" PU) reduced to small level