Performance of reconstruction and identification of τ leptons in their hadronic decays in pp collisions at $\sqrt{s} = 13$ TeV at CMS

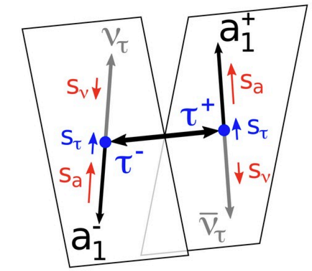
07.07.2018

Speaker:
Olena Hlushchenko
(on behalf of CMS collaboration)

Physics Institute III B
RWTH AACHEN UNIVERSITY

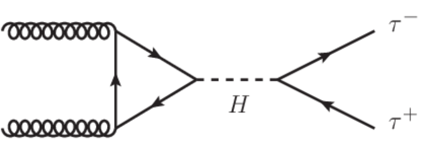
Introduction

- Features:
  - → hadronic decays (m = 1.78 GeV, Br=65%)
  - → large lifetime of $(2.9\times10^{-13}$ s)

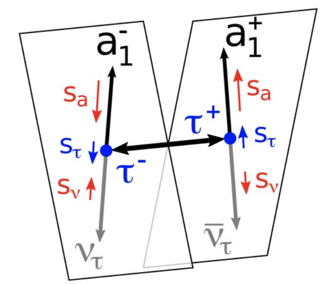


(a) Decay planes for $S = 1$ (scalar).

- Motivation:
  - → SM tests:
    - » Higgs Yukawa couplings proportional to the mass
      - » highest branching ratio due to the highest mass among leptons.

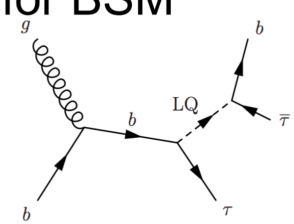


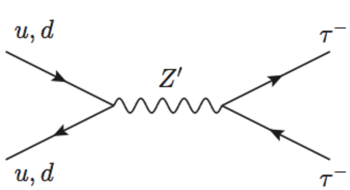
    - – Possible final state CP measurements in $H \to \tau\tau$
    - – Measurement of Higgs to fermion coupling via $H \to \tau\tau$

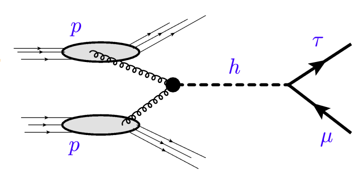


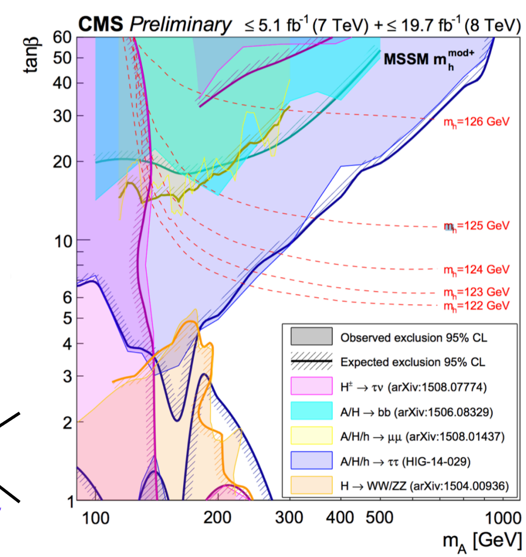
(b) Decay planes for $S = 0$ (pseudoscalar).

  - → Often pure or enhanced channel for BSM searches:
    - » Z', W' decays
    - » Leptoquarks decays
    - » Decays of MSSM Higgs boson at large $tan(\beta)$
    - » Lepton non-universality tests
    - » CP violation in B --> D** tau nu



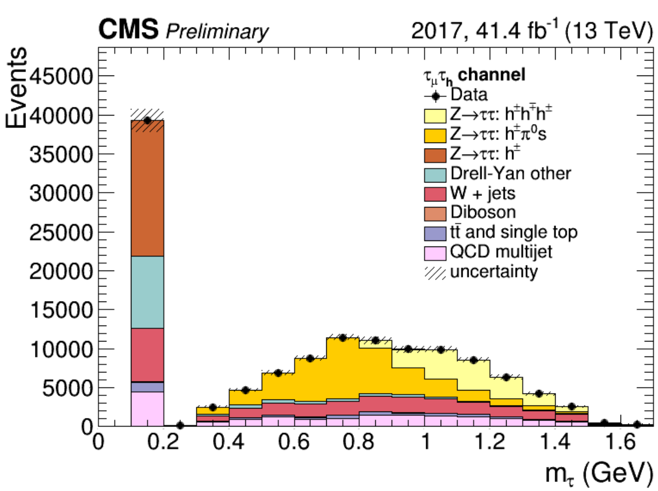


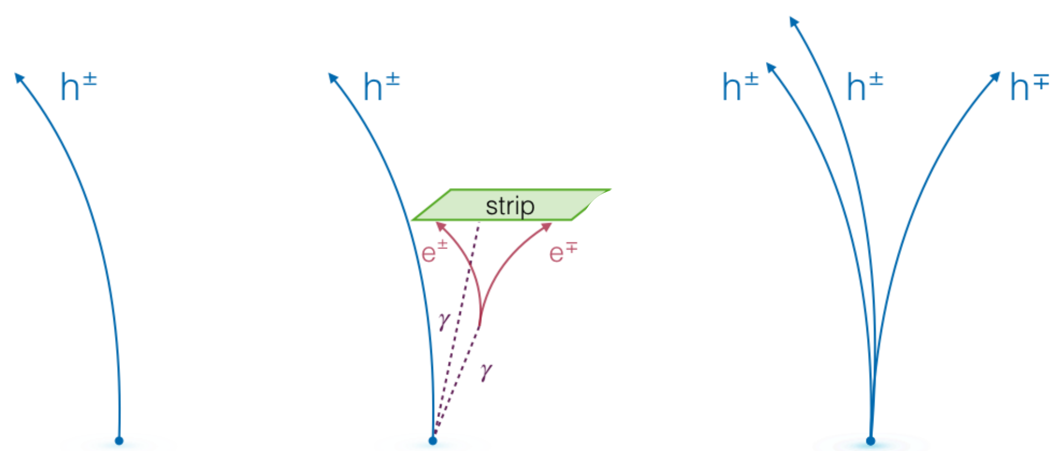




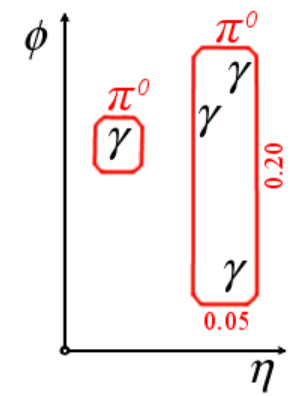
Tau reconstruction and ID

- A ”hadrons-plus-strips” (**HPS**) is based on reconstructed jets reprocessing:
  - → Reassembles **jets(anti-kt, R=0.4)** to its constituents
    - » PF: electrons, photons, muons, hadrons
  - → Reconstructs the $\boldsymbol{\pi}^0$ (as “Strips”):
    - » 2016: p_T > 0.5 GeV ; 2017: p_T >1 GeV
  - → Reconstructs all possible decay modes (**DM)**
  - → **The DM with highest p_T and charge/strip multiplicity that pass the mass cut-window is assigned to each jet**

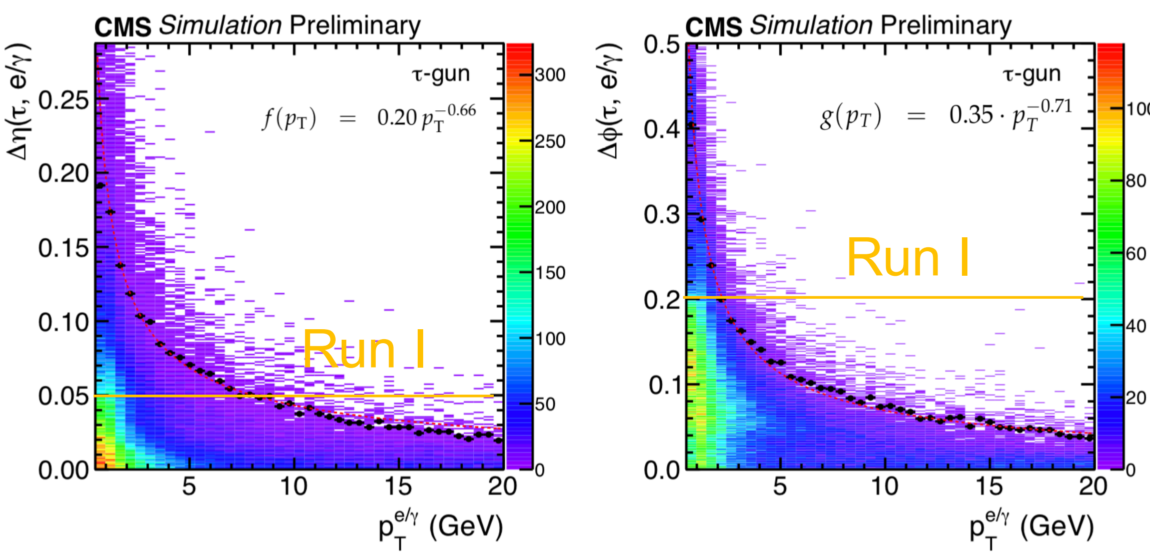




Tau reconstruction and ID: Dynamic-strip reconstruction



1. e/γ with largest p_T (seed)
2. Merge p_T of the next largest e/γ within a strip
3. Strip position is recomputed
4. Until no e/γ to merge



CMS PAS TAU-16-002

Discrimination against jets

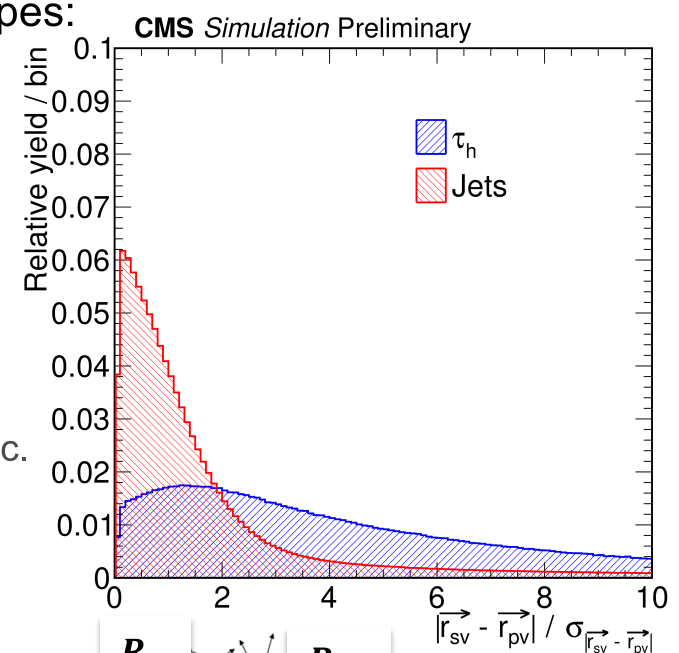
RWTH AACHEN UNIVERSITY

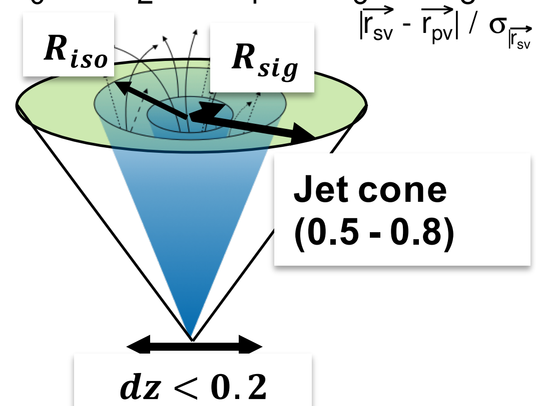
$$I_\tau = \sum p_T^{\text{charged}}(d_Z < 0.2\,\text{cm}) + \max\left(0, \sum p_T^{\gamma} - \Delta\beta \sum p_T^{\text{charged}}(d_Z > 0.2\,\text{cm})\right)$$

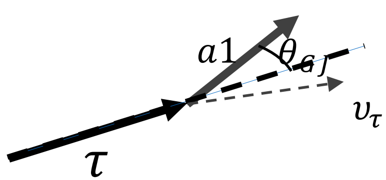
$$p_T^{\text{strip, outer}} = \sum p_T^{e/\gamma}(\Delta R > R_{\text{sig}}) < 0.10 \cdot p_T^{\tau}$$

Cut-based

- MVA: BDT is seeded with 23 input variables of types:
  - → Isolation sums
  - → Distributions and multiplicities of particles inside/outside the cone:
    - » number of photons in tau, photons P_T outside signal cone etc.
  - → Lifetime-related:
    - » flight length (significance), secondary vertex, impact parameter vector etc.
  - → Tau quantities:
    - » p_T^τ, $|\eta_T^\tau|$, decay mode(DM), energy ratios etc.
  - → 2017 training strategy improvements:
    - » added Gottfried-Jackson angle for 3 prong decay mode
    - » p_T cut-off for photons is increased

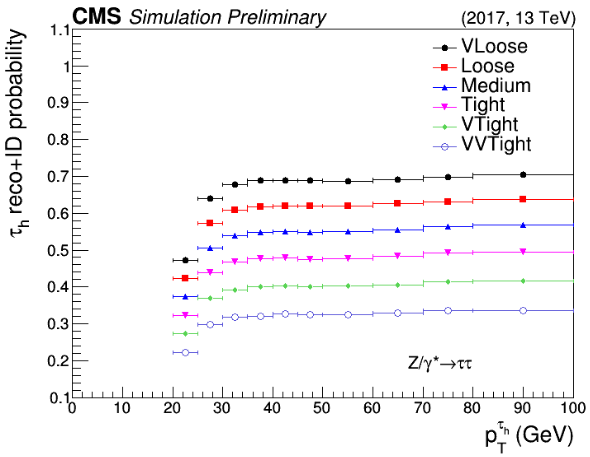


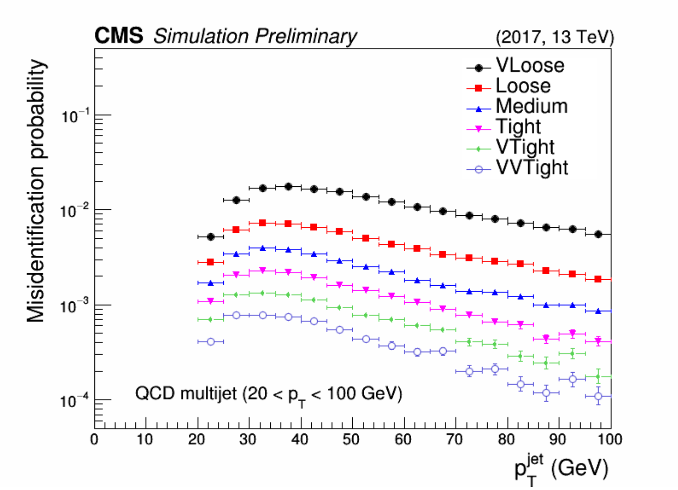


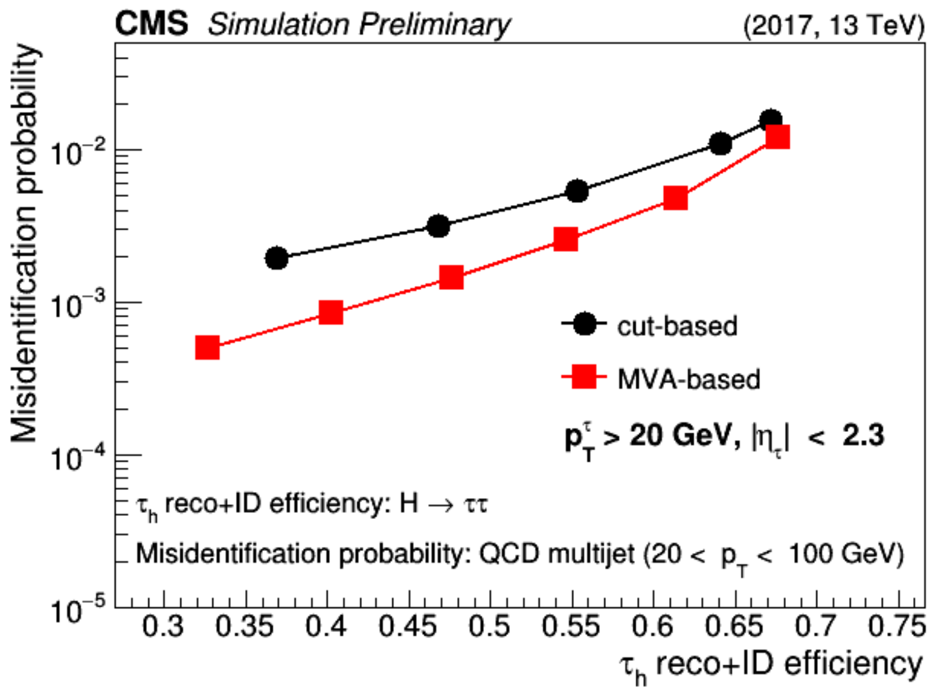


Discrimination against jets

CMS DP-2018/026





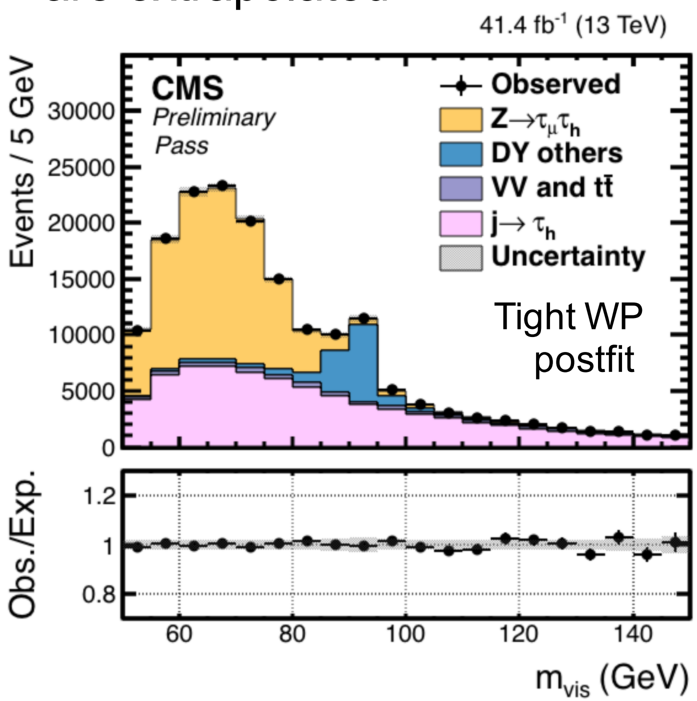


The working points (WP) are chosen to have isolation efficiencies between 40%(VVTight) and 90%(VLoose), in steps of 10% and 95%(VVLoose)

Tau ID scale factors

RWTH AACHEN UNIVERSITY

- Scale factors (SF) are measured using the Tag&Probe
- The most significant correction
  - → <5% uncertainty
- For high tau p_T^τ the SF and it's uncertainty are extrapolated

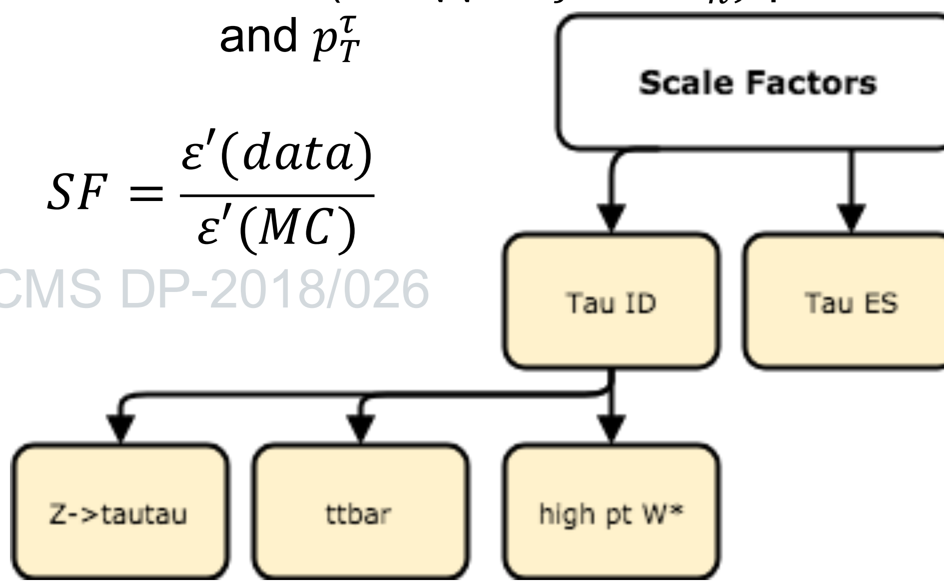


Example for $Z \rightarrow \mu\tau_h$:

- Tag: well identified μ
- Probe: reco tau with different isolation requirement
- Maximum-likelihood fit of **visible mass** performed in pass and control ($Z \rightarrow \mu\mu$) regions
- Backgrounds from MC but…
- W+Jets & QCD estimated from the data ($Z \rightarrow \mu\mu + jet \rightarrow \tau_h$) per DM and p_T^τ

$$SF = \frac{\varepsilon'(data)}{\varepsilon'(MC)}$$

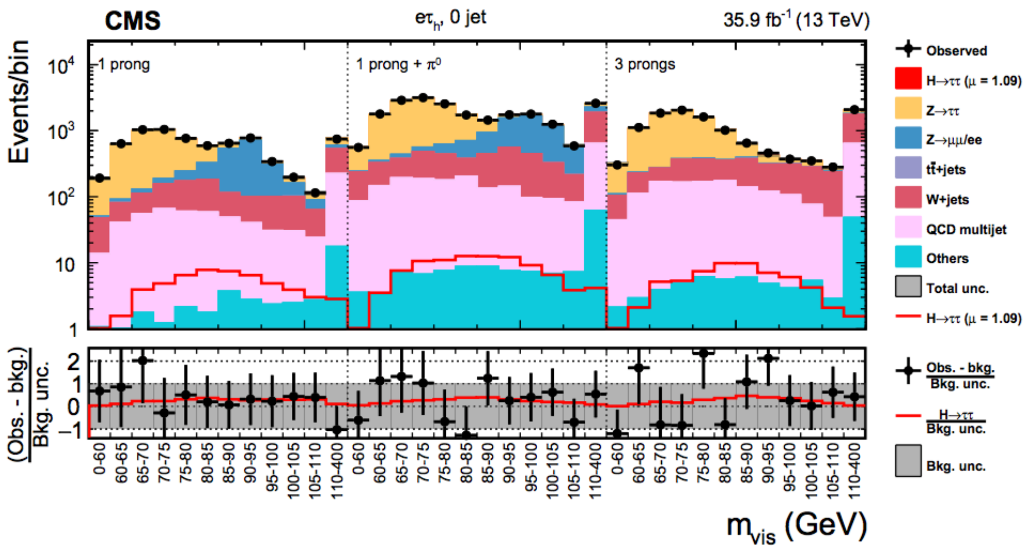
CMS DP-2018/026



20..60 GeV　60..100 GeV　>100 GeV

Discrimination against electrons

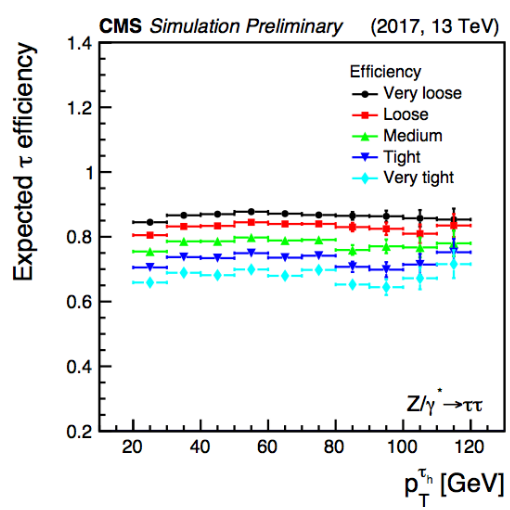
RWTHAACHEN UNIVERSITY

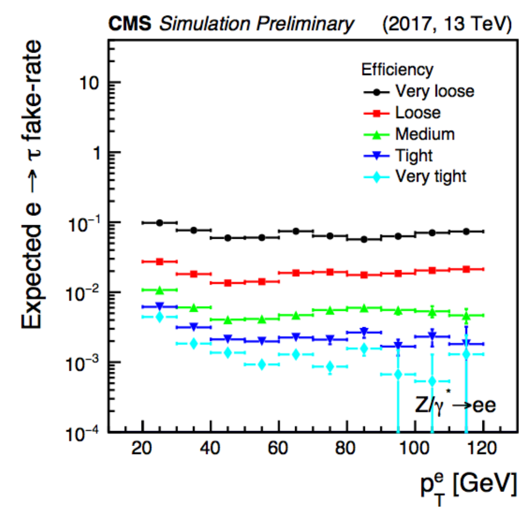


Background events in final states involving taus can arise from $Z \to ee$ decays, where the electrons can be misidentified as taus

CMS-HIG-16-043

- **MVA BDT classifier**
- **Input variables**:
  - → Photon multiplicity per strip
  - → Shower-based variables
  - → Relation of strips and τ_h quantities
- 75% Efficiency
  - → <1% misID





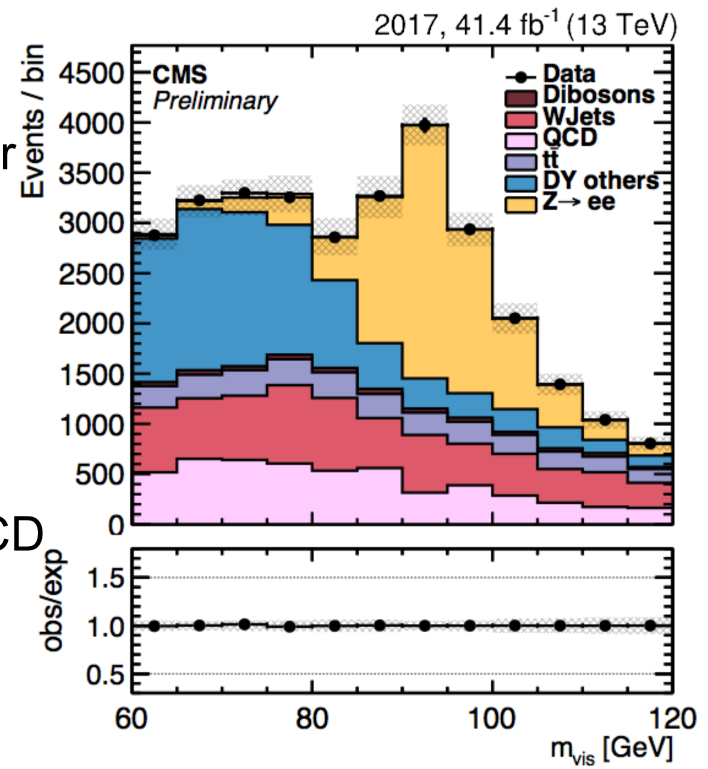
CMS DP-2018/026

Fake rates measurements: $e \rightarrow \tau_h$

- $Z \rightarrow ee$
- SFs measured separately in barrel or endcap for different WP of anti-e discriminator using Tag&Probe
- Maximum-likelihood fit of **visible mass** performed separately in pass, fail:
  - → probe(loose τ_h) is matched to generated electron
  - → Backgrounds from MC but Wjets and QCD
- The MisID fake rate was shown to be compatible in Barrel an Forward regions (within 10%) and be strongly dependent on the chosen WP
  - → 5% to 40% uncertainty

CMS DP-2018/026

Barrel region



Pass Medium WP; Postfit

ICHEP2018
Olena Hlushchenko |RWTH Aachen University|

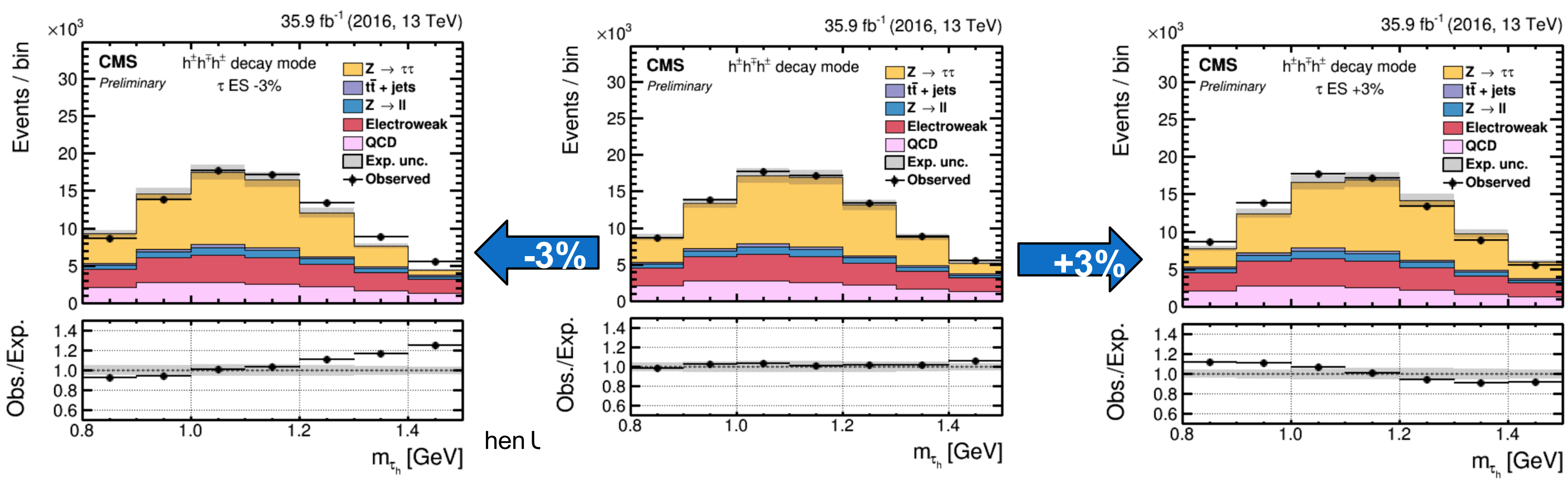
Energy scale measurements

RWTH AACHEN UNIVERSITY

CMS DP-2017/006

- Sample: $Z \rightarrow \tau\tau \rightarrow \mu\tau_h / e\tau_h$
  - → Both approaches are compatible across all decay modes (separately and combined)
- Shape templates are produced by shifting τ_h energy
- Results are compatible with unity within the uncertainty between different DM

$$ES = \frac{p_{reco}^{\tau,vis}}{p_{gen}^{\tau,vis}}$$

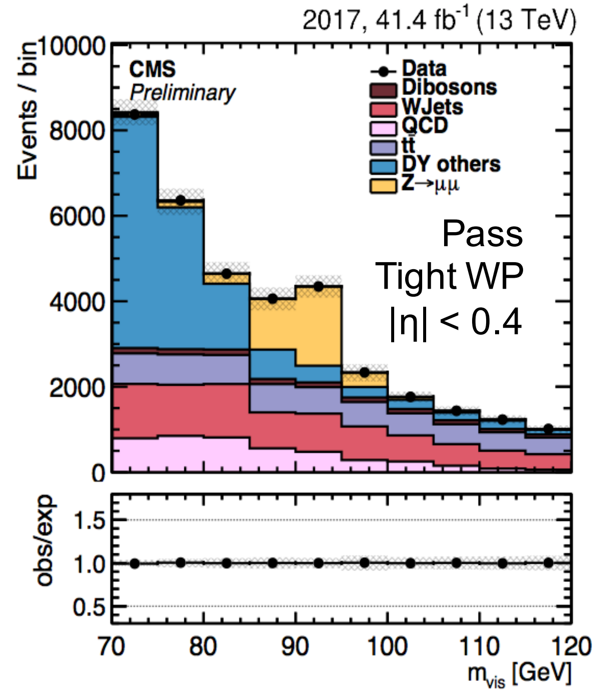


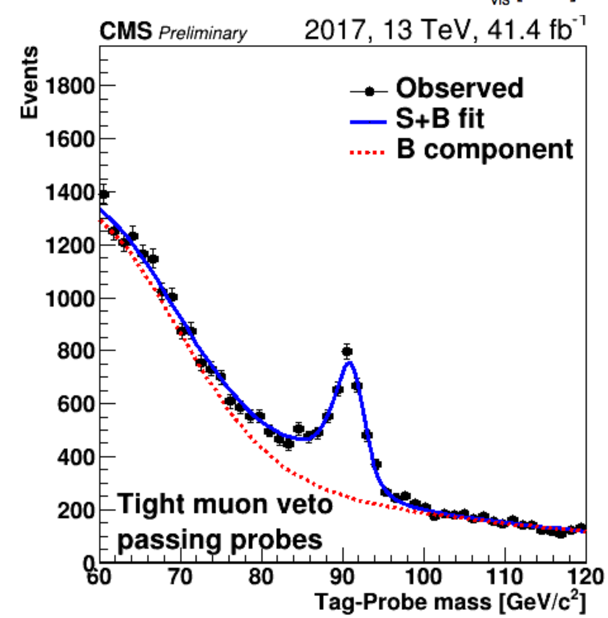
Fake rates measurements: $\mu \rightarrow \tau_h$

Cut-based anti-muon discriminator:

- Veto candidates matching segments in the muon detector
- ~100% efficiency

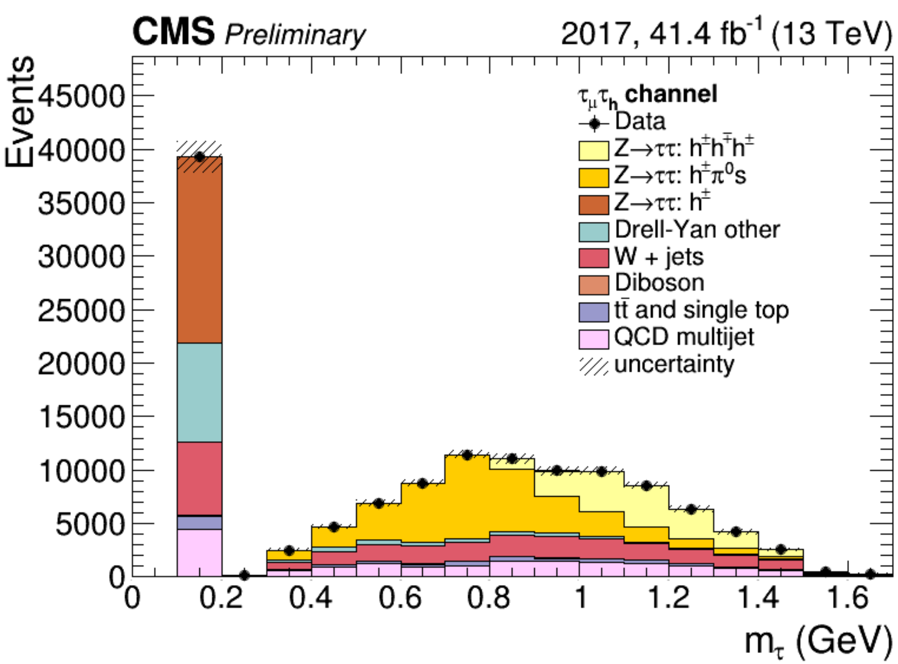
- $Z/\gamma^* \rightarrow \mu\mu$
- Measured in η bins per WP of anti-μ discriminator
- Tag&Probe:
  - Classical: CMS DP-2018/026
    - From ML fit of the tag and probe
    - Close to 1 in central region and up to 2 in forward
    - <10% uncertainty
  - Analytical:
    - the normalizations of signal ($Z/\gamma^* \rightarrow \mu\mu$) and background ($Z/\gamma^* \rightarrow \tau\tau$) component is done using the analytic functional forms fit to the data distributions.
  - compatible results with two methods

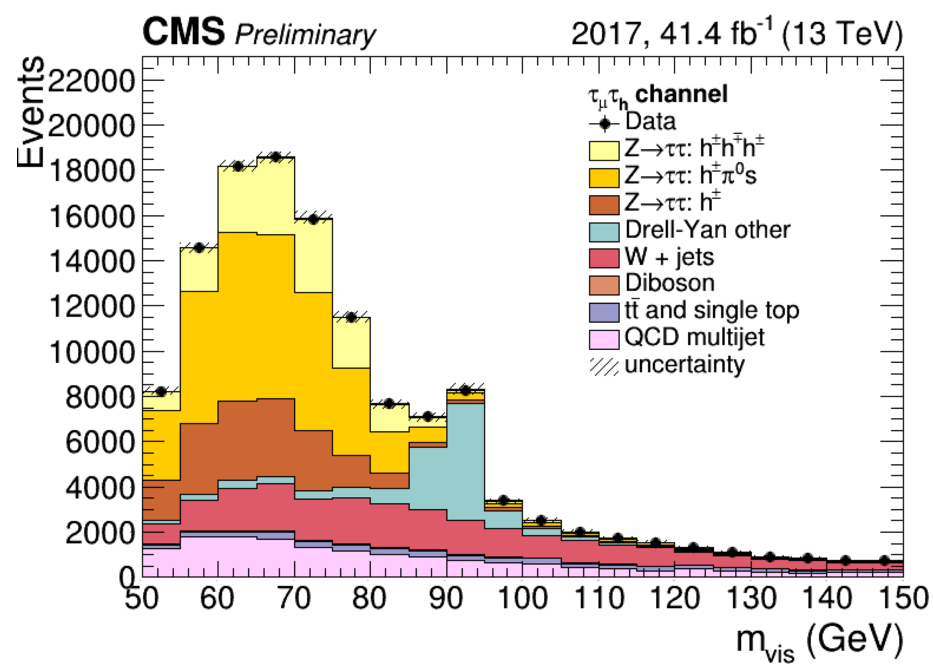




Summary

- After calibrations, we observe a very good agreement between data and simulation, proving an excellent understanding of the detector and the reconstruction!
- The various decay modes of the tau leptons are well resolved and identified





CMS DP-2018/026

Appendix plan

Data and MC

- Signal samples covering large p-T contain:
  - → H->tautau (NLO, Powheg v2)
  - → Z'->ll, W'->lnu (LO, Pythia8)
  - → Matching: $\Delta R(reco, gen) = sqrt((\Delta\eta)2 + (\Delta\varphi)2) < 0.3$
- Other samples:
  - → Z/gamma->ll (LO, Madgraph_aMCatNLO)
  - → W+jets(LO, Madgraph_aMCatNLO 2.2.2)
  - → Single top (NLO, Powheg 2)

    ↕ Normalised to NNLO
  - → ttbar (NLO, Powheg 2)
  - → Di-boson (NLO, Madgraph_aMCatNLO 2.2.2 or Powheg v2)

    ↕ Normalised to NLO
  - → QCD (Pythia)
  - → Highly-virtual W bosons (m>200 GeV, taunu, munu) (Pythia8)
- Hadronisation, Pile-up: Pythia8 + CUETP8M1
- PDF: NNPDF3.0

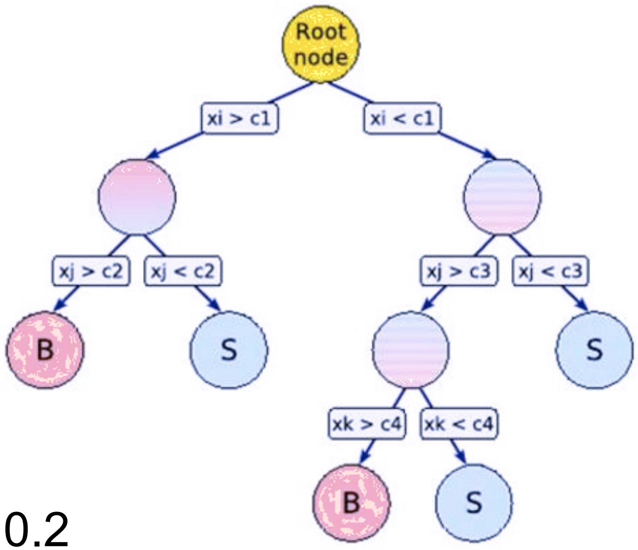
Tau reconstruction and ID: HPS:: dynamic-strip algorithm motivation

- Previously: deta x dphie = 0.05 x 0.20
  - → **nuclear interactions** of pi± with the tracker material, which create low pT e/γ that may go outside the fixed strip
  - → **conversions of photons** from pi0 decays to electron/positron pairs, and bremsstrahlung
- Idea:
  - ~~→ Increase size~~: more fakes; for boosted taus this size should be even smaller!
  - → Strip size adjusted dynamically as a **function of the pT of e/γ**

Classes of discriminating variables

- Isolation sums:
  - → $I_{\tau,ch}$, $I_{\tau,neut}$
- Distributions and multiplicities of particles inside/outside the cone:
  - → photons P_T outside signal cone, number of photons in tau etc.
- Lifetime-related:
  - → SV, flight length, flight length significance, IP, etc.
- Tau quantities:
  - → $P_{T,\tau}$, $|\eta_{reco,\tau}|$, DM, Gottfried-Jackson-angle, energy ratio etc.

MVA-based isolation



- Tool: TMVA
- Gradient Boost BDT
  - → Classification & regression
  - → 1000 trees in the forest
  - → Boosting type for the trees in the forest: Gradient
  - → Learning rate for GradBoost algorithm: 0.2
  - → Use only a random subsample of all events for growing the trees in each iteration
  - → Fraction of events to be used in each iteration: 0.5
  - → Separation criterion for node splitting: GiniIndex
  - → Number of grid points in variable range used in finding optimal cut in node splitting: 500
  - → Maximum depth of cell tree: 5

Tau ID scale factors estimation using $Z \to \mu\tau_h$

- events divided in two(anticorrelated) pass/fail region

$$N_{pass} = \varepsilon \bullet N_0$$
$$N_{fail} = (1-\varepsilon) \bullet N_0$$

ML fit →

$$N'_{pass} = \varepsilon' \bullet N_0 = r\varepsilon \bullet N_0$$
$$N'_{fail} = (1-\varepsilon') \bullet N_0 = (1-r\varepsilon) \bullet N_0$$

with $r = \varepsilon' / \varepsilon$ as **parameter of interest(POI)** of the fit

so the yield in the pass/fail regions must scale as:

$$N'_{pass} / N_{pass} = (r\varepsilon \bullet N_0)/(\varepsilon \bullet N_0) = r$$
$$N'_{fail} / N_{fail} = (1-r\varepsilon)/(1-\varepsilon)$$

$N_0 = N_{pass} + N_{fail}$

matched $Z \to \mu\tau_h$ MC

$\varepsilon = \frac{N_{pass}}{N_0}$ - prefit

ε' - postfit

$r = \frac{\varepsilon}{\varepsilon\prime}$ - signal strength

- fit tag-probe pair visible mass, from 60 (70) GeV to 120 GeV

$$SF = \frac{\varepsilon'(data)}{\varepsilon'(MC)}$$