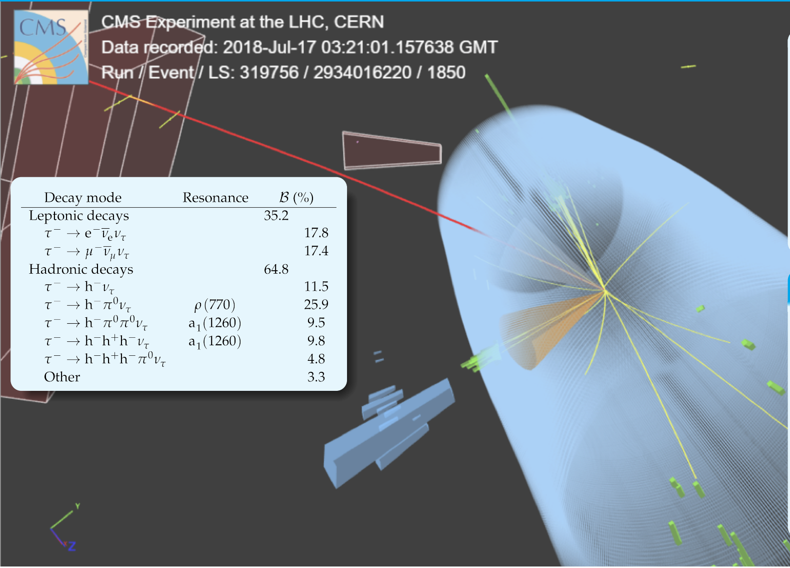
Identification of hadronic tau decays using a deep neural network with the CMS experiment at LHC

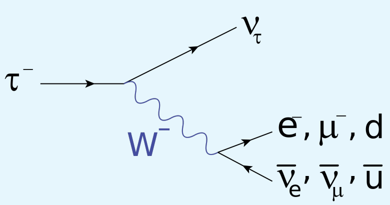
ML4Jets2022
3 Nov 2022

Mykyta Shchedrolosiev[1] on behalf of the CMS Collaboration
[1] Deutsches Elektronen-Synchrotron

Hadronically decaying τ at CMS



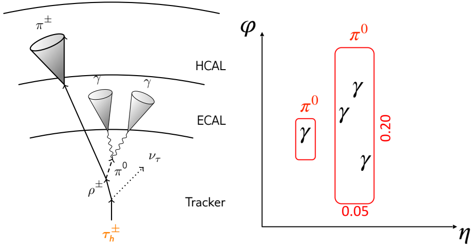
Decay mode	Resonance	$\mathcal{B}$ (%)
Leptonic decays		35.2
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
Hadronic decays		64.8
$\tau^- \to h^- \nu_\tau$		11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	25.9
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	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		3.3



Tau lepton (1.78GeV)

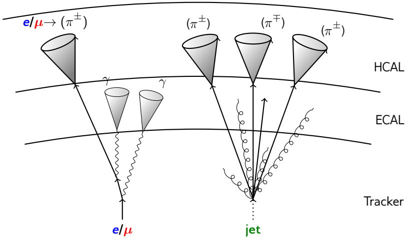
- Heavy enough to decay hadronically (τ_h)
- Decays into hadrons + neutrino in $\sim$64.8% of the cases
- Improvement of performance in reconstruction and identification of the hadronic tau decays important for many SM and BSM analyses

Step-1: τ_h reconstruction



*plot link

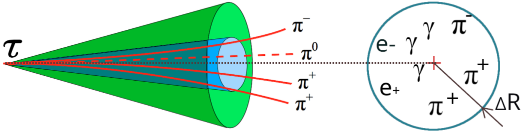
Step-2: τ_h identification



- Stable particles are reconstructed using the **particle flow (PF)** algorithm
- τ_h reconstruction is **seeded by anti-k_t** jets of PF candidates with distance parameter of 0.4 (AK4)
- τ_h is reconstructed using combinatoric **hadron-plus-strip (HPS)** algorithm:
  - π^0 is reconstructed using dynamic η, ϕ window in ECAL (strip)
  - Require one or three $\pi^\pm$ with high quality track $p_T > 0.5$GeV originating from PV $d_{xy} < 0.1$cm
  - Mass constraints are put in correspondence to DM resonances and tuning

- Signature of jets originating from quarks/gluons (τ_j), electrons (τ_e) and muons (τ_μ) can fake genuine hadronic tau decays (τ_h) $\rightarrow$ classification task
- Each background process can have specific that differs it from τ_h $\rightarrow$ high-dimensional problem
  - jets can have more hadronic activity, which can be detected in the isolation cone
  - electrons can have specific patters in the calorimeter clusters
  - muons can have substantial amount of matched hits in the muon chambers

- The MVA discriminators [a] used previously were built on **higher-level input variables** (for τ_j and τ_e) and **cutoff-based criteria** (τ_μ).
- Exploit lower-level information, can lead to improved performance
- Simultaneous classification $\rightarrow$ reduce maintenance efforts



- Information about all reconstructed particles near the τ_h candidate is directly used as input to the algorithm
- Multiple sub-detectors within CMS is exploited, including the inner tracker, the electromagnetic (ECAL) and hadronic (HCAL) calorimeters, and the muon chambers

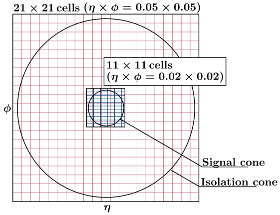
[a] JINST 13 (2018) P10005

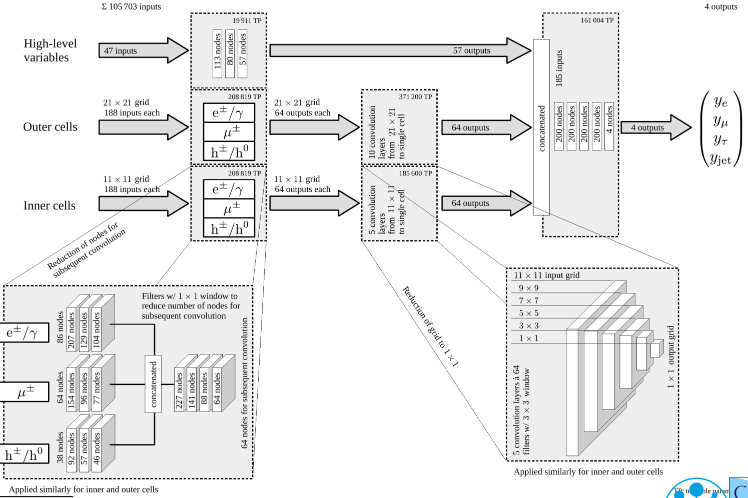
- Training data consists of $\approx$ 140 million τ_h, τ_j, τ_e, τ_μ which are sampled from the 2017 MC datasets:
  - Drell-Yan
  - $t\bar{t}$
  - W+jets
  - Z', m(Z') ranging from 1 to 5 TeV
  - QCD multijet

- Training data are preprocessed and mixed in order to ensure:
  - **Homogeneity of the final dataset** - any sufficiently large consecutive interval are statistically compatible. Probability (p_T, η, τ-type) - should always be the same. Initial tuples dozens TB $\rightarrow$ can not load in RAM
  - **Balanced contribution / uniformity** across different tau candidate types **(τ_h,τ_e,τ_μ,τ_j) as well as p_T and η, and physical process.** (Datasets are overpopulated with low pt taus$\rightarrow$ reweighting with large weights induces numerical instability)

- **DeepTau**[1] - **convolutional deep NN** with $\mathcal{O}(1.1M)$ trainable parameters
- High-level τ and event properties (47 variables) are combined with low-level PFCandidates, fully reconstructed electron and fully reconstructed muon information

- Every candidate is put on the (η, ϕ) grid. Grid is divided on 3 blocks:
  - Hadrons block - 34 variables
  - muon block - 60 variables
  - e-gamma block - 82 variables
- **Signal cone:** $dR < 0.1$ (11×11)
  Isolation cone: $dR < 0.5$ (21×21)
- Object occupancy in the tensor (1.7% for inner, 7.1% for outer)





[1] JINST 17 (2022) P07023

Training

Custom loss function is taken to focus on 50-80% signal efficiency regime and focus more on τ_h vs. other when τ_h probability is low:

$$L(y^{\text{true}}, y^{\text{pred}}) = \underbrace{\kappa_\tau H_\tau(\mathbf{y}^{\text{true}}, \mathbf{y}^{\text{pred}}; \boldsymbol{\omega})}_{\text{(a) categorical CE}}$$

$$+ \underbrace{(\kappa_e + \kappa_\mu + \kappa_{\text{jet}})\bar{F}_{\text{cmb}}(1 - y_\tau^{\text{true}}, 1 - y_\tau^{\text{pred}})}_{\text{(b) normalized binary focal loss}}$$

$$+ \underbrace{\kappa_F \sum_{i \in \{e,\, \mu,\, \text{jet}\}} \kappa_i\, \hat{\Theta}(y_\tau^{\text{pred}} - 0.1)\, \bar{F}_i(y_i^{\text{true}}, y_i^{\text{pred}}; \gamma_i).}_{\text{(c) step } \epsilon_\tau > 0.1 \otimes \text{ normalized binary focal loss}}$$

(a) : Distinguish between τ_h, e , μ, jet
(b) : Focused separation e , μ, jet from τ_h , e
(c) : Focused separation of τ_h from e , μ , jet for $y_{\tau_h} > 0.1$

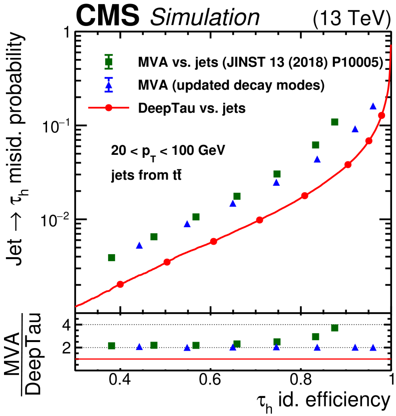
- The final discriminator is calculates as:

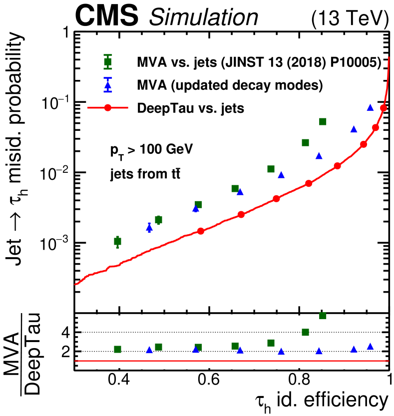
$$D_\tau^\alpha(p) = \frac{p_\tau}{p_\tau + p_\alpha}$$

where $\alpha = e, \mu, jet$

- The loss function is minimised using NAdam algorithm
- The training is run for 10 epochs on GeForce RTX 2080 (3 days/epoch)
- The best performance on the validation set is achieved after 7 epochs

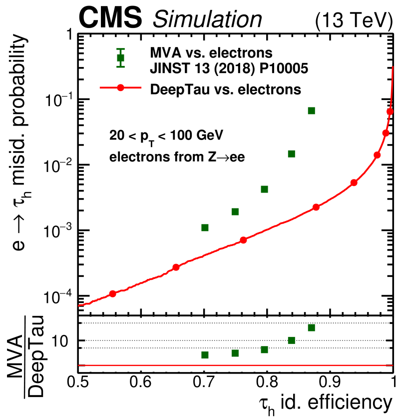
Performance (against jets)





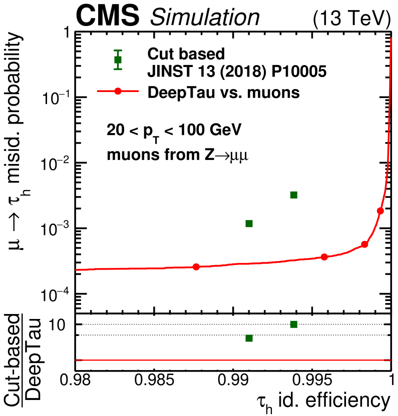
- Dots indicate working points of the tagger
- τ_h efficiency is computed with $H \to \tau\tau$ sample
- τ_{jet} misID probability with $t\bar{t}$ sample
- Jet misID rate is reduced by a factor of **2-4** for low-pt and **2-6** for high-pt.

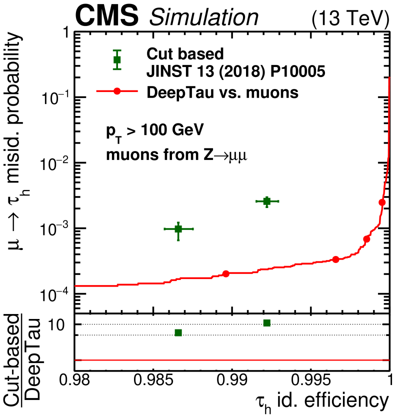
Performance (against electrons)



- Dots indicate working points of the tagger
- τ_h efficiency is computed with $H \to \tau\tau$ sample
- τ_e misID probability with $Z \to ee$ sample
- Electron misID rate is reduced by factor of $\sim$ **8-11** for low and higt-pt region.

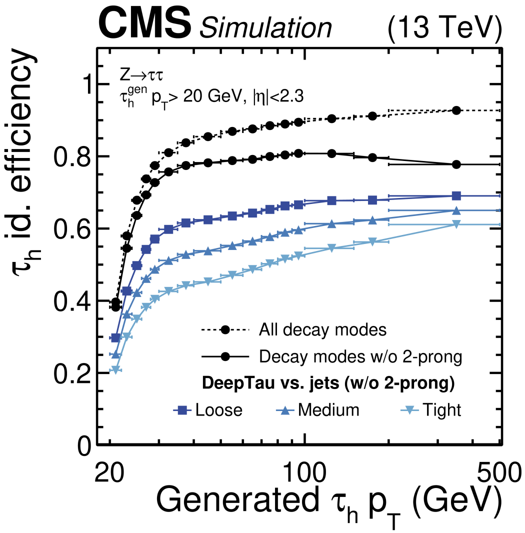
Performance (against muons)





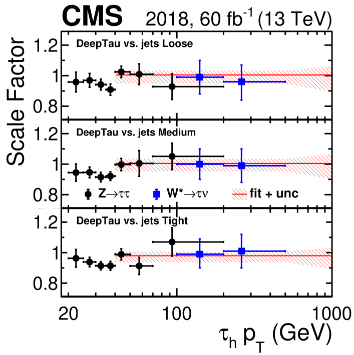
- Dots indicate working points of the tagger
- τ_h efficiency is computed with $H \to \tau\tau$ sample
- τ_μ misID probability with $Z \to \mu\mu$ sample
- Muon misID rate is reduced by factor of $\sim$ 9-10 for low and higt-pt region.

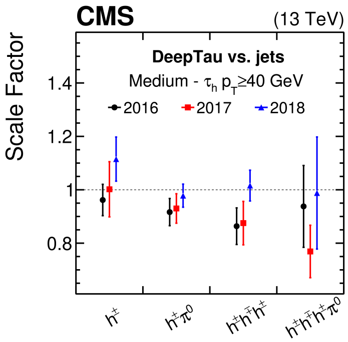
Performance (against jets) over different p_T regions



- The identification efficiency exceeds 80% for p_T>30 GeV and is close to 90% for p_T>100 GeV
- Limited by the charged-hadron reconstruction efficiency
- If decay modes with missing charged hadrons are excluded $\rightarrow$ the efficiency is reduced by around 10%
- Misidentification rate for shown working points (for $t\bar{t}$ and $H \rightarrow \tau\tau$ samples):
  - Loose ∼2.0%
  - Medium ∼1.0%
  - Tight ∼0.5%

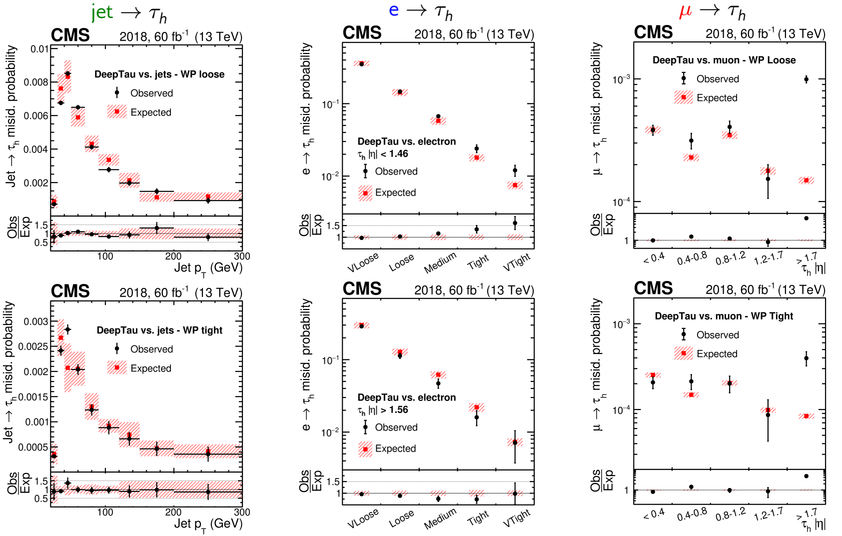
Data-to-simulation scale factors for genuine τ_h





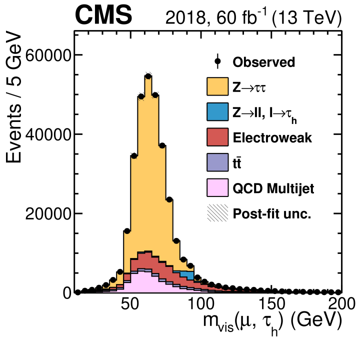
- The scale factors are generally a bit smaller than one but consistent with unity within 10%

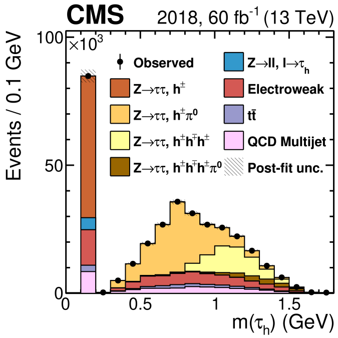
DeepTau performance for fake taus



> Good modelling of fake taus in the most parts of the parameter space

Data/MC control plots





Preselection on muon and tau candidates:

- Well identified and isolated muon with p_T>25 GeV,$|\eta|$< 2.4,|z|<0.2 cm
- Tau candidates with p_T>20 GeV, $|\eta|$<2.3, |z|<0.2 cm
- $\mu\tau_h$ pair with an opposite charge and $\Delta R(\mu,\tau)$>0.5, M_T<60GeV, $|\Delta\eta|$<1.5
- **DeepTau IDs**: Tight WP against jets, VVLoose WP against electrons, VLoose against muons
- The distributions incorporate all measured scale factors and corrections
- Good data/MC agreement for physical observables

Conclusions

- A new algorithm [2] has been introduced to discriminate hadronic tau lepton decays against jets, electrons, and muons.
- The performance of the new τ_h identification and reconstruction algorithms significantly improves over the previously used algorithms, in particular in terms of discrimination against the background from jets and electrons
- For a given **jet** rejection level, the efficiency for genuine τ_h candidates **increases by 10-30%**
- The efficiency for genuine τ_h candidates to pass the discriminator **against electrons increases by 14% for the loosest working point** that is employed in many analyses
- The superior performance of the algorithm is validated with collision data. **The observed efficiencies for genuine τ_h, jets, and electrons to be identified as τ_h typically agree within 10% with the expected efficiencies** from simulated events
- Further improvement of the algorithm are ongoing for Run 3:
  - DeepTau for online event selection
  - DeepTau for boosted di-tau topology
  - DeepTau for displaced taus
  - Use of domain adaptation techniques to improve the modelling
  - Alternative NN architectures (GNN, Transformers)

[2] JINST 17 (2022) P07023