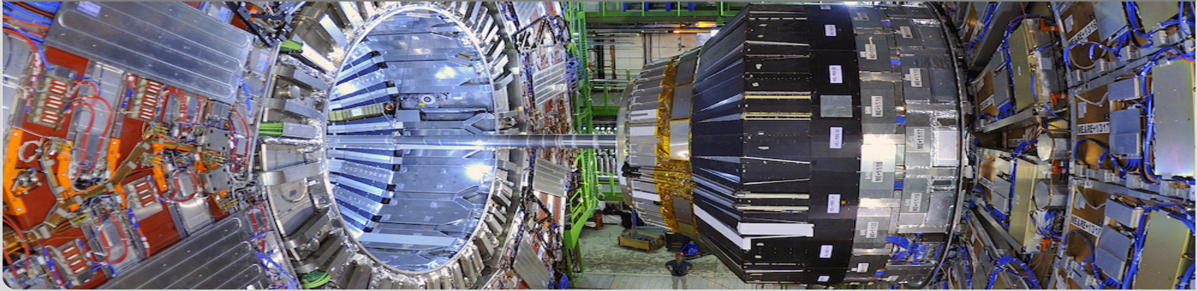
Karlsruhe Institute of Technology

Performance of the CNN-based tau identification algorithm with Domain Adaptation using Adversarial Machine Learning for Run 3

ML4Jets 2024

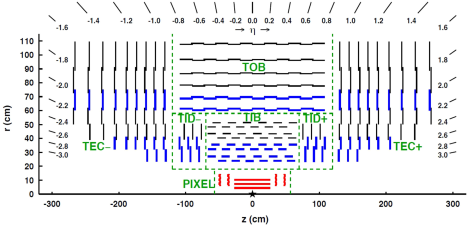
Olha Lavoryk on behalf of the CMS collaboration | 5 November 2024

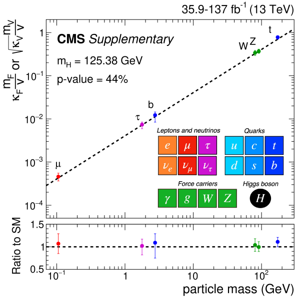
INSTITUT FÜR EXPERIMENTELLE TEILCHENPHYSIK (ETP)



τ properties in context of CMS physics

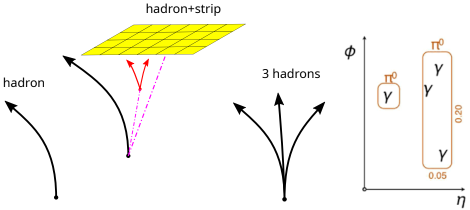
- mass 1776.86$\pm$0.12 MeV/c^2, the strongest H$\to$ l^+l^- Yukawa coupling
- lifetime τ = 2.9×10^{-13} s or cτ = 87 μm $\to$ beam pipe radius $\approx$ 3cm $\to$ decays before reaching the tracker $\to$ reconstructed by its decay products; E $\approx$ 200 GeV $\to$ cτ = 10.2 mm
- N.B. we discuss properties of hadronically decay τ lepton (τ_{had})
- τ_{had} object is preprocessed via two step procedure of reconstruction and identification for analysis purposes
- τ_{had} is a key observable in many CMS analyses

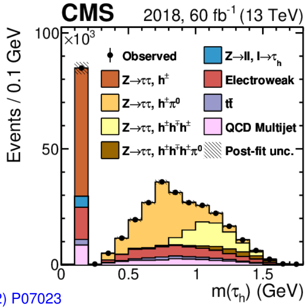




Hadronic τ reconstruction at CMS via HPS algorithm

Decay mode	Resonance	$\mathcal{B}$ (%)
Leptonic decays		35.2
$\tau^- \to e^-\overline{\nu}_e\nu_\tau$		17.8
$\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau$		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

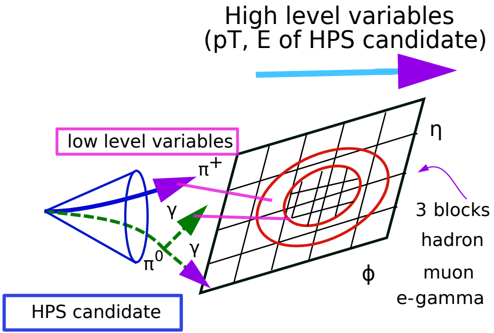




JINST 17 (2022) P07023

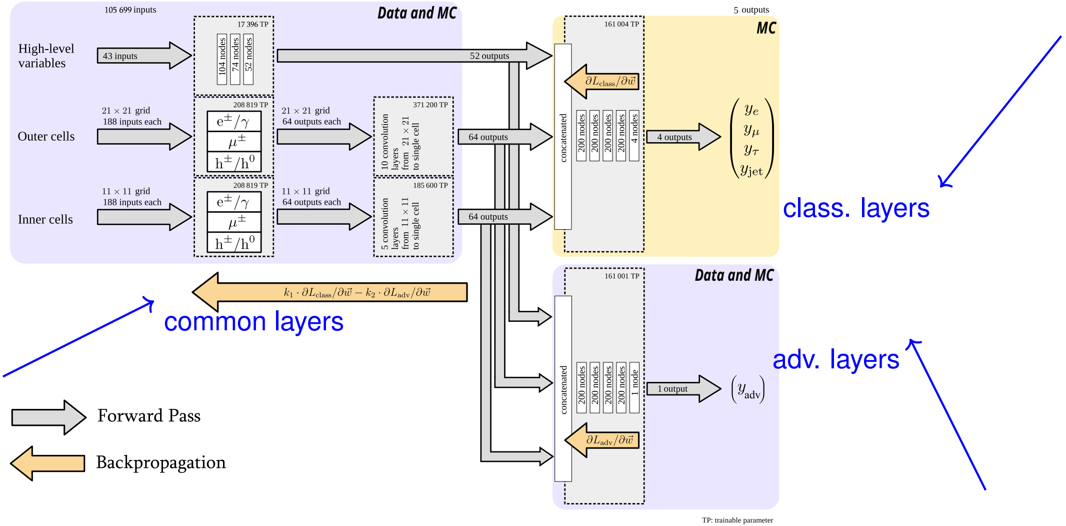
- τ_{had} reconstruction based on Particle Flow algorithm input
- seed regions are selected algorithmically through hadronic jets:
  - Cluster photons and electrons/positrons to form strips which are a stand in for the neutral pion
  - Combine charged hadrons with strips and reject all candidates whose masses are not compatible with the mass of mesonic resonances in the decay chain

Hadronic τ identification using Deep Tau



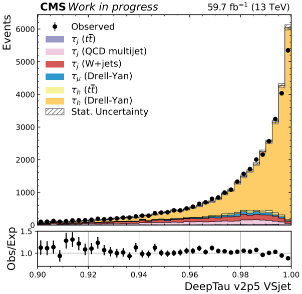
- After HPS reconstruction few objects can be misidentified as τ_{had} : jets, electrons, muons
- Deep Tau - convolutional neural network reduces the misidentification
- Deep Tau v2p5 is an improved version of its predecessor - Deep Tau v2p1
- Validation on Run 2 data but used for Run 3 analyses

Deep Tau v2p5 architecture

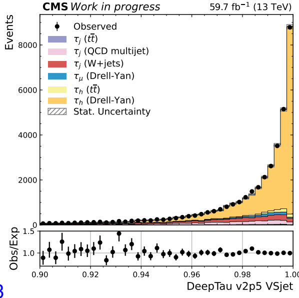


Domain adaptation in a context of Deep Tau v2p5 model

- Domain adaptation is a major improvement with respect to the previous model
- Includes simultaneous training of two subnetworks, one is purely on simulation, another one on a combination of data and simulation
- Pure MC-based training to distinguish between genuine τ_{had} and jet/e/μ fakes; combined MC&data to minimize differentiation between them
- Adversarial term in a loss function: $k_1 \frac{\delta L_{class}}{\delta \vec{w}} - k_2 \frac{\delta L_{adv}}{\delta \vec{w}}$



simulation to data agreement is improved after the application of domain adaptation



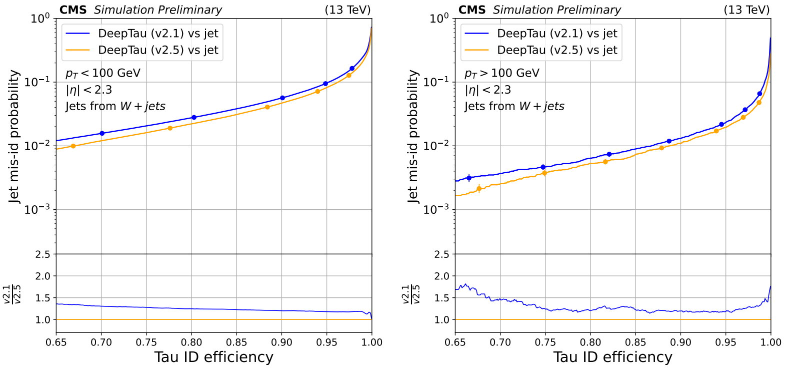
contribution to the DP - 2024/063

Training

$$L(y^{true}, y^{pred}) = \underbrace{\kappa_\tau H_\tau(y^{true}, y^{pred}; \omega)}_{\text{Separation for all } \alpha}$$
$$+$$
$$\underbrace{(\kappa_e + \kappa_\mu + \kappa_j)\overline{F}_{cmb}(1 - y_\tau^{true}, 1 - y_\tau^{pred}; \gamma_{cmb})}_{\text{Focused separation of } e,\mu,jet \text{ from } \tau_{had}}$$
$$+$$
$$\underbrace{\kappa_F \sum_{i\in\{e,\mu,j\}} \kappa_i \hat{\theta}(y_\tau - 0.1)\overline{F}_i(y_i^{true}, y_i^{pred}; \gamma_i)}_{\text{Focused separation of } \tau_{had} \text{ from } e,\mu,jet \text{ for } y_\tau>0.1}$$

- Discriminant against jets, muons, and electrons are given by $D_\alpha(y) = \frac{y_\tau}{y_\tau + y_\alpha}$, with $\alpha = \mu$,e,jet
- 5 epochs, 2 days per epoch (NVIDIA Tesla V100 / GTX 1080 Ti) improved wrt Deep Tau v2p1 with 10 epochs on GeForce RTX 2080 3 days/epoch
- With addition of domain adaptation: 1 epoch, 12 - 15 hours (NVIDIA Tesla T4) per epoch

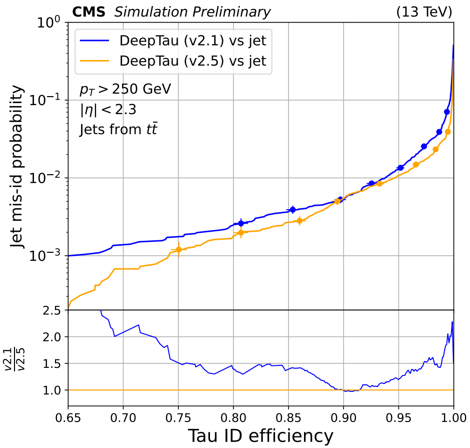
Performance (against jets)



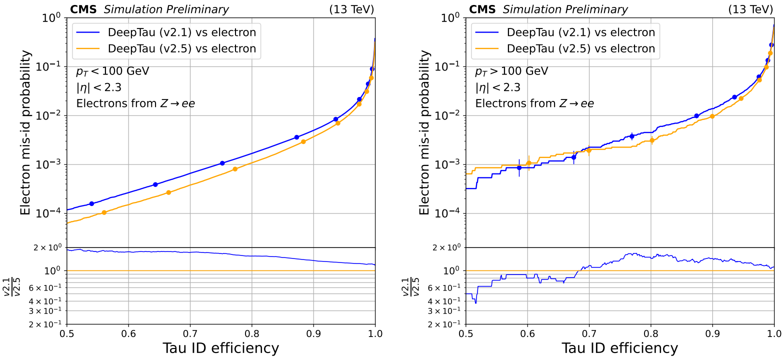
- Dots indicate the working points of the tagger
- $H \rightarrow \tau\tau$ samples are used for τ_{had} efficiency estimation
- W+jet samples are used for for τ_{had} midID probability estimation
- τ_{had} misID reduced by 30% (reduction by $\approx$50% with $t\bar{t}$ in backup)

Performance (against jets at high p_T)

- Dots indicate the working points of the tagger
- W^{*}+jet samples are used for τ_{had} efficiency estimation
- $t\bar{t}$ samples are used for τ_{had} midID probability estimation
- τ_{had} identification performance being improved even at high pT phase space
- Degradation at high pt and high rejection regime

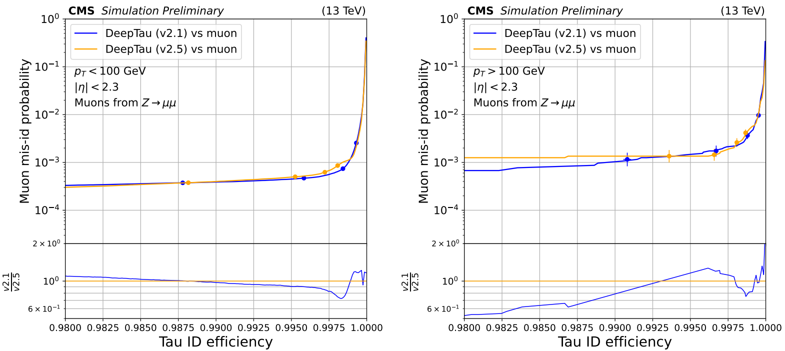


Performance (against electrons)



- Dots indicate the working points of the tagger
- $H \rightarrow \tau\tau$ samples are used for τ_{had} efficiency estimation
- $Z \rightarrow ee$ samples are used for electron misID probability estimation
- τ_{had} identification performance of v2p5 is compatible with v2p1

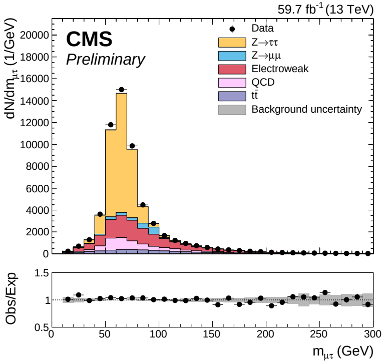
Performance (against muons)

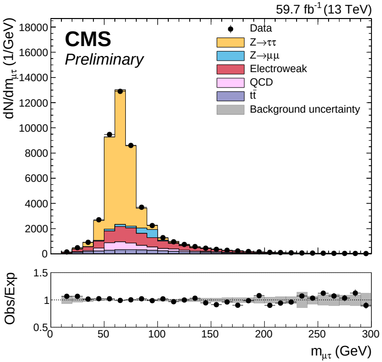


- Dots indicate the working points of the tagger
- $H \rightarrow \tau\tau$ samples are used for τ_{had} efficiency estimation
- $Z \rightarrow \mu\mu$ samples are used for muon misID probability estimation
- Small degradation at high pT but overall a good performance

Data/MC control plot

Deep Tau v2p1



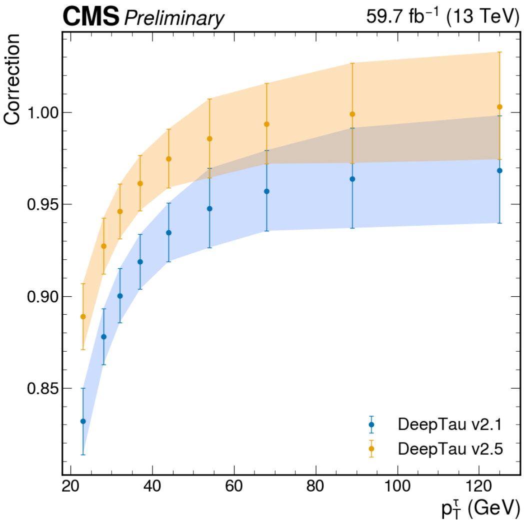


Deep Tau v2p5

- DeepTau IDs: Tight WP against jets (60% τ_{had} ID efficiency), VVLoose WP against electrons (99% τ_{had} ID efficiency), Tight WP (99.5% τ_{had} ID efficiency) against muons
- 30% background reduction in comparison with v2p1

Correction factors comparison (v2p5 vs v2p1)

- Correction factors are weighted with respect to the τ_h decay branching ratios HPS algorithm efficiencies per decay mode
- Deep Tau v2p5 corrections are closer to 1 thanks to the model improvements

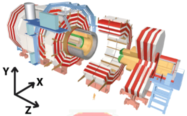


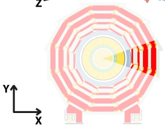
Conclusions and outlook

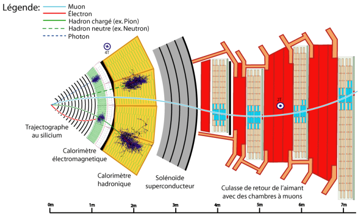
- A new improved version of the algorithm to reduce misidentification of τ_{had} with electron, muon and jets was introduced
- 30% background reduction for genuine τ_{had} in comparison with v2p1
- Improved MC-to-data agreement in comparison with v2p1
- DeepTau v2.5 demonstrates an improved performance in comparison with its predecessor DeepTau v2.1
- The algorithm shows excellent performance even in the challenging high p_T regime

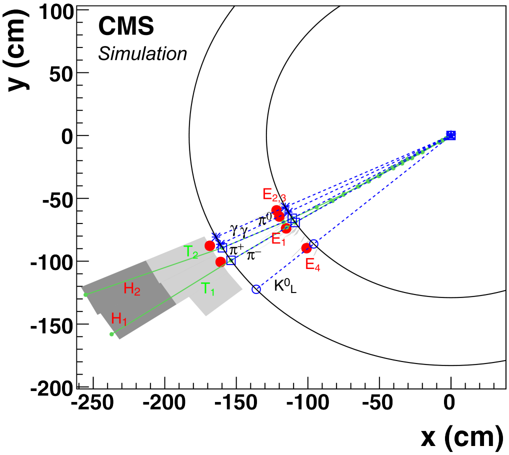
Back up slides

CMS Particle Flow reconstruction algorithm



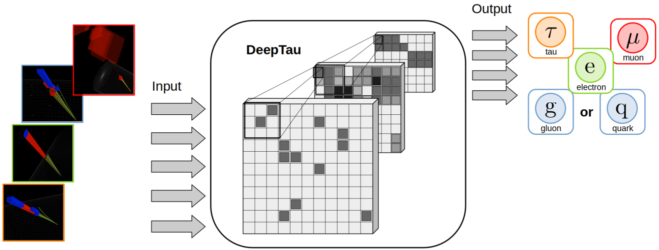


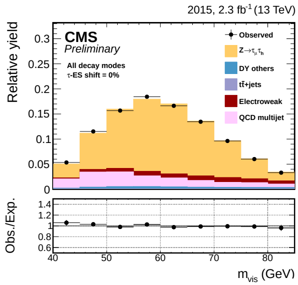




- reconstruction based on linking Particle Flow(PF) elements from different subdetectors
- large magnetic field, excellent tracking system, ECAL, HCAL performance
- individual particles in the event: e, μ, γ, charged and neutral hadrons reconstruction
- complex objects like jets, p_T^{miss} which consist of individual object
- τ_{had} **identification and reconstruction**

Hadronic τ identification using Deep Tau





- Few objects can be misidentified as τ: jets, electrons, muons
- Deep Tau - convolutional neural network reduces the misidentification rate
- Discriminant against jets, muons, and electrons are given by $D_\alpha(y) = \frac{y_\tau}{y_\tau + y_\alpha}$,, with $\alpha = \mu$,e,jet

Data and simulation efficiency are not equal, so we need some correction factors

Performance (against jets)