NEC'2019

Identification of tau lepton using Deep Learning techniques at CMS

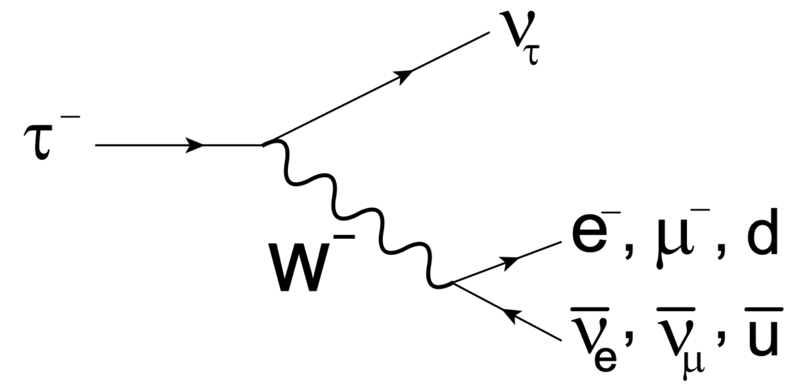
27th Symposium on Nuclear Electronics and Computing - NEC'2019
October 3rd, 2019

Konstantin Androsov[1] on behalf of the CMS Collaboration

[1]INFN Pisa

Analysis with hadronic taus at CMS

- **Tau** is the heaviest Standard Model (SM) lepton with the mass of 1.78 GeV that **decays into hadrons + neutrino in 64.8% of the cases**



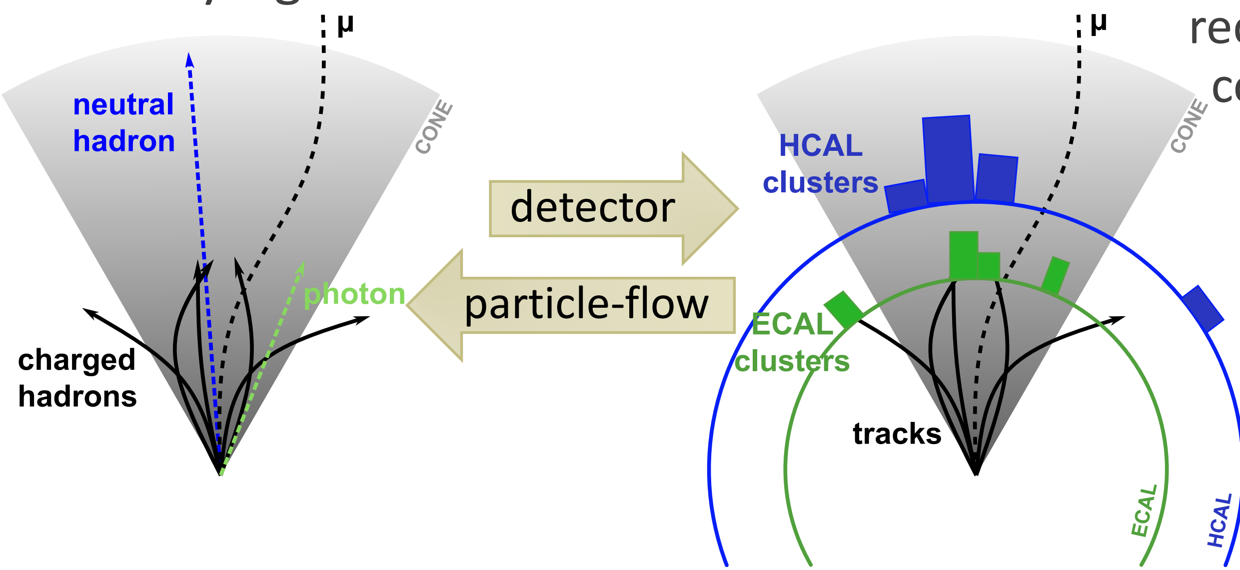
- Good performance in reconstruction and identification of the hadronic tau decays is crucial for many important physics analysis in CMS:
  - SM and MSSM Higgs boson decaying and tau pairs, $\boldsymbol{H \rightarrow \tau\tau}$, analyses with different production mechanisms (gluon fusion, VBF, VH, ttH)
  - SM and BSM searches for **double Higgs** production: $HH \rightarrow bb\tau\tau$, $HH \rightarrow 4\tau$
  - Measurement of the SM properties
  - Searches for additional charged and neutral scalar bosons
  - *And many other CMS analyses*

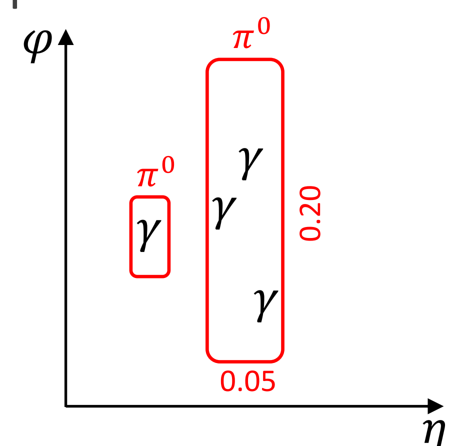
Reconstruction of the hadronic tau decays

- In CMS, stable particles are reconstructed using the **particle flow** (PF) algorithm
- Hadronic tau decays are reconstructed by **Hadron + strips** (HPS) algorithm
- Starting from jet, the HPS algorithm analyses information from stable PF particles within the jet cone to identify typical hadronic tau decay signatures

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

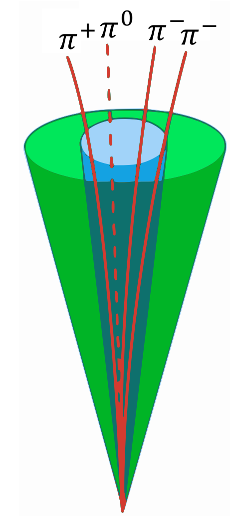
In HPS, neutral pions are reconstructed using $\eta \times \varphi$ strips, collecting energy spread from photon conversion



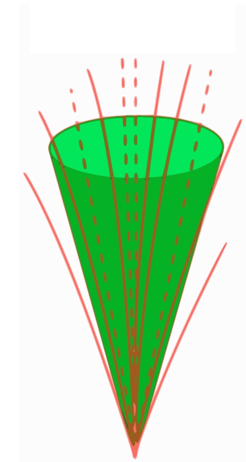


Main backgrounds for hadronic taus

- **Jets** originating from quarks or gluons (τ_j), **electrons** (τ_e), and **muons** (τ_μ) can be misidentified as **hadronic tau** decays (τ_h)
- Each background have a characteristic signatures in the detector that can help to separate it from τ_h:
  - τ_j candidates, in general, have more hadronic activity, which can be detected in the isolation cone
  - τ_e candidates have specific patters in the calorimeter clusters
  - τ_μ candidates have substantial amount of matched hits in the muon chambers
- For each case, based on these typical signatures a set of the most discriminating variables can be defined
- Before DeepTau, 3 dedicated algorithms were developed within CMS to discriminate τ_h against given source of the background (*)



τ_h



QCD jet (background)

(*) See JINST 13 (2018) P10005 for more details

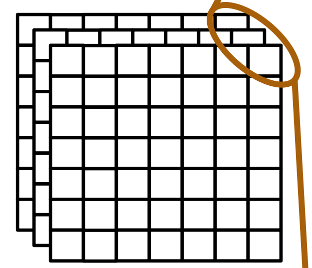
DeepTau ID

- To <u>efficiently</u> discriminate τ_h against main backgrounds, the detailed information from multiple sub-detectors within CMS must be exploited, including the **inner tracker**, the electromagnetic (**ECAL**) and hadronic (**HCAL**) calorimeters, and the **muon chambers**
  - This leads to a very high-dimensional problem
  - *Deep machine learning* (DL) is proven to be a powerful tool to tackle such kind of problems
- **DeepTau** is a new multiclass tau identification algorithm based on a convolutional deep neural network (DNN)
- The training is performed on a balanced mix of $\boldsymbol{\mathcal{O}\left(1.4 \cdot 10^{8}\right) \tau_e, \tau_\mu, \tau_h}$ **and** $\boldsymbol{\tau_j}$ **candidates** coming from Drell-Yan, $t\bar{t}$, W+jets and Z’ 2017 MC samples
- Training, validation and testing sets are composed by reconstructed tau candidates with a wide p_T range and minimal preselection ⇒ **suitable for all analysis** with hadronic taus at the final state
  - $p_T \in [20, 1000]$ GeV, $|\eta| < 2.3$, and $|dz| < 0.2$ cm (the longitudinal impact parameter)
- The **ground truth** whatever the reconstructed tau candidate is τ_e, τ_μ, τ_h and τ_j is derived based on **MC truth matching**

Network inputs

- **Candidate** = PF candidate *OR* fully reconstructed electron *OR* fully reconstructed muon
- Candidates belonging to the inner and outer cones are separated and split into two $\eta \times \varphi$ grids of 11×11 (21×21) cells with the cell size of 0.02×0.02 (0.05×0.05) for the inner (outer) cone
- If there are more than one object of the given type that belong to the same cell, the object with the highest p_T is chosen
- To reduce the NN size, within each cell the input variables are split into 3 blocks: e-gamma, muon, hadrons

Input cells



Single cell

Total # of inputs: 188

hadrons block
34 variables

muon block
60 variables

e-gamma block
82 variables

+ 4 high level features added to each block

Inner cone = signal cone

$$\Delta R < \max\left[\min\left(0.1, \frac{3}{p_T^\tau}\right), 0.05\right]$$

Outer cone = signal & isolation cones

$$\Delta R < 0.5$$

Total number of inputs:

105656 *(from grids)* **+ 43** *(high level)*

Average number of non-empty cells per tau candidate	
Inner region	2.019 (1.7%)
Outer region	31.14 (7.1%)

Network architecture: cell pre-processing

Cell from inner or outer region

+ ρ, p_T^τ, η^τ, ECAL crack indicator

Each block within a cell is processed with *3 dense layers*

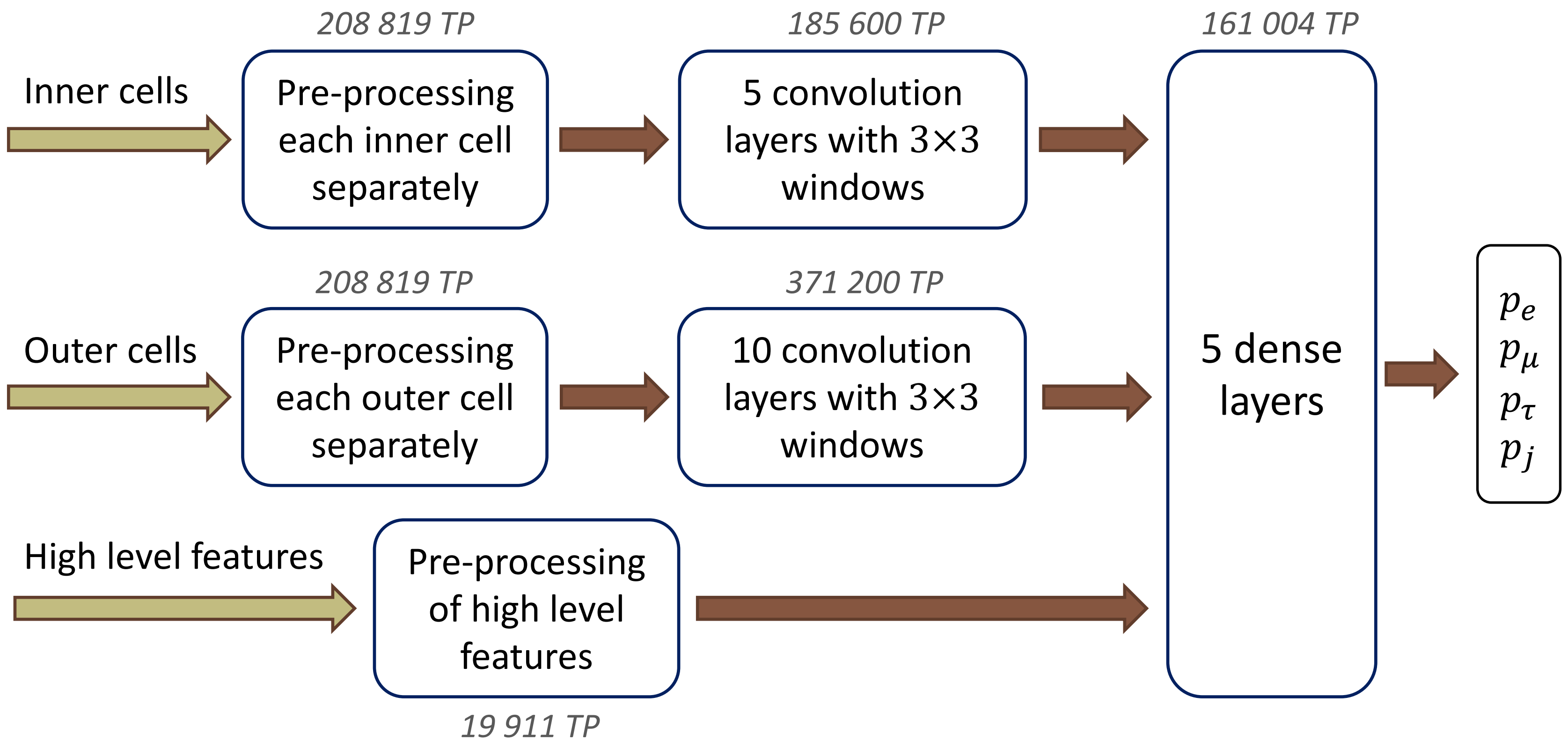
Blocks are merged and a full cell is processed with *4 dense layers*

To convolution layers

After each layer
Batch normalization
PReLU activation
Dropout with 0.2 rate

Network architecture

Total number of trainable parameters (TP) in the DeepTau network = **1 155 353**



After each layer a regularization is applied

Final discriminator and minimization

- The final discriminator is chosen to be

$$D_\tau^\alpha(\boldsymbol{p}) = \frac{p_\tau}{p_\tau + p_\alpha}, \qquad where\ \alpha \in \{e, \mu, j\}$$

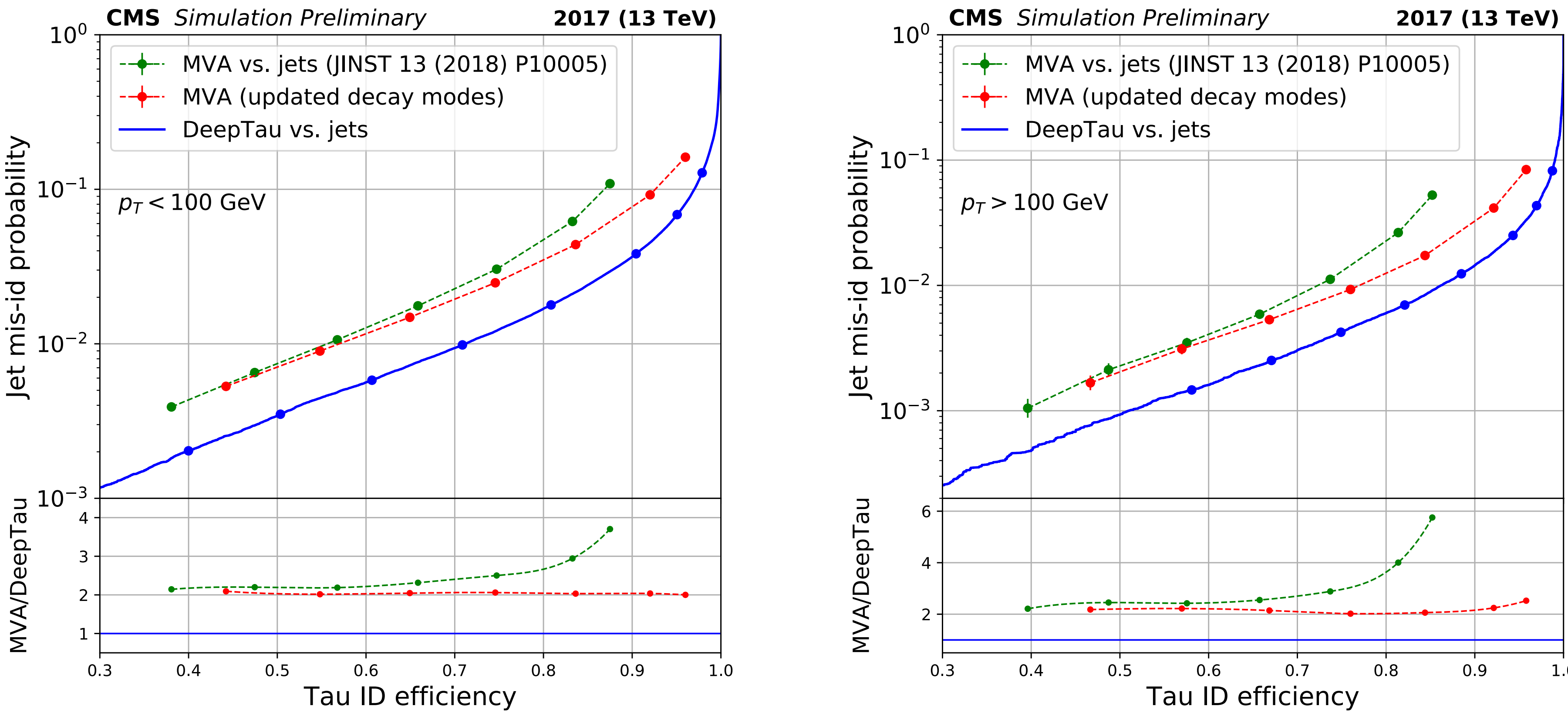
- A **custom loss function** based on the focal loss is defined in order to ensure the best performance for a wide tau ID efficiency range
- The loss function is minimized using **NAdam** algorithm (Adam with Nesterov momentum)
- The training is run for **10 epochs** on GeForce RTX 2080
  - Average speed $\approx \boldsymbol{3\ days/epoch}$
- The best performance on the validation set is achieved after **7 epochs**
  - The corresponding NN is chosen as the final

DeepTau: inference on CPU

- Current CMS computational grid consists of CPU nodes
- It is important to have the NN inference optimized for CPU, in order to be able to run on billions MC and Data events
- If full DeepTau NN would be evaluated, including empty cells, the total number of floating point operations (FLOPS) is **2 206 252**
- To speed up the inference, the network is **split into 3 parts**: inner, outer and core
- The inner (outer) sub-NN operates per-cell level and is applied only to non-empty cells
- This allowed to reduce the number of FLOPS to
  $N(FLOPS) = 2964 \cdot N_{ActiveCells} + 540484$ $\quad E[N(FLOPS)] = \mathcal{O}\left(6.5 \cdot 10^5\right)$
- As the result, the total evaluation time **reduces by a factor ≃2.5**
- **The final execution speed** of the DeepTau producer module within CMS software is

  ≃ **22 ms/tau** (*inference*) + **3 ms/tau** (*inputs preprocessing*)

DeepTau discrimination against jets

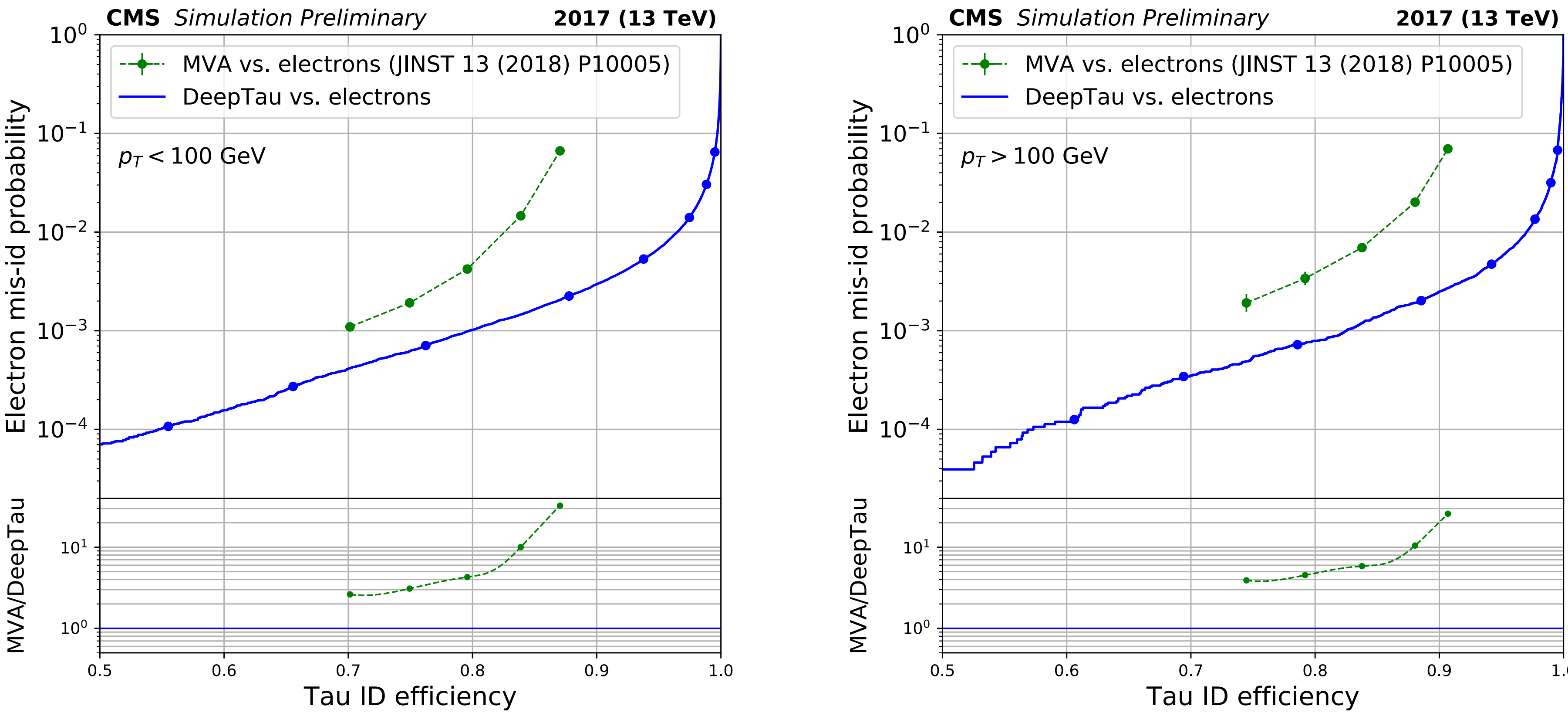


- Consistent improvement at both **low and high p_T**
- Jet misidentification probability **reduces by ~50..80%**

τ_h from $H \rightarrow \tau\tau$ MC
τ_j from $t\bar{t}$ MC

Working points of the discriminators are indicated by the dots

DeepTau discrimination against electrons

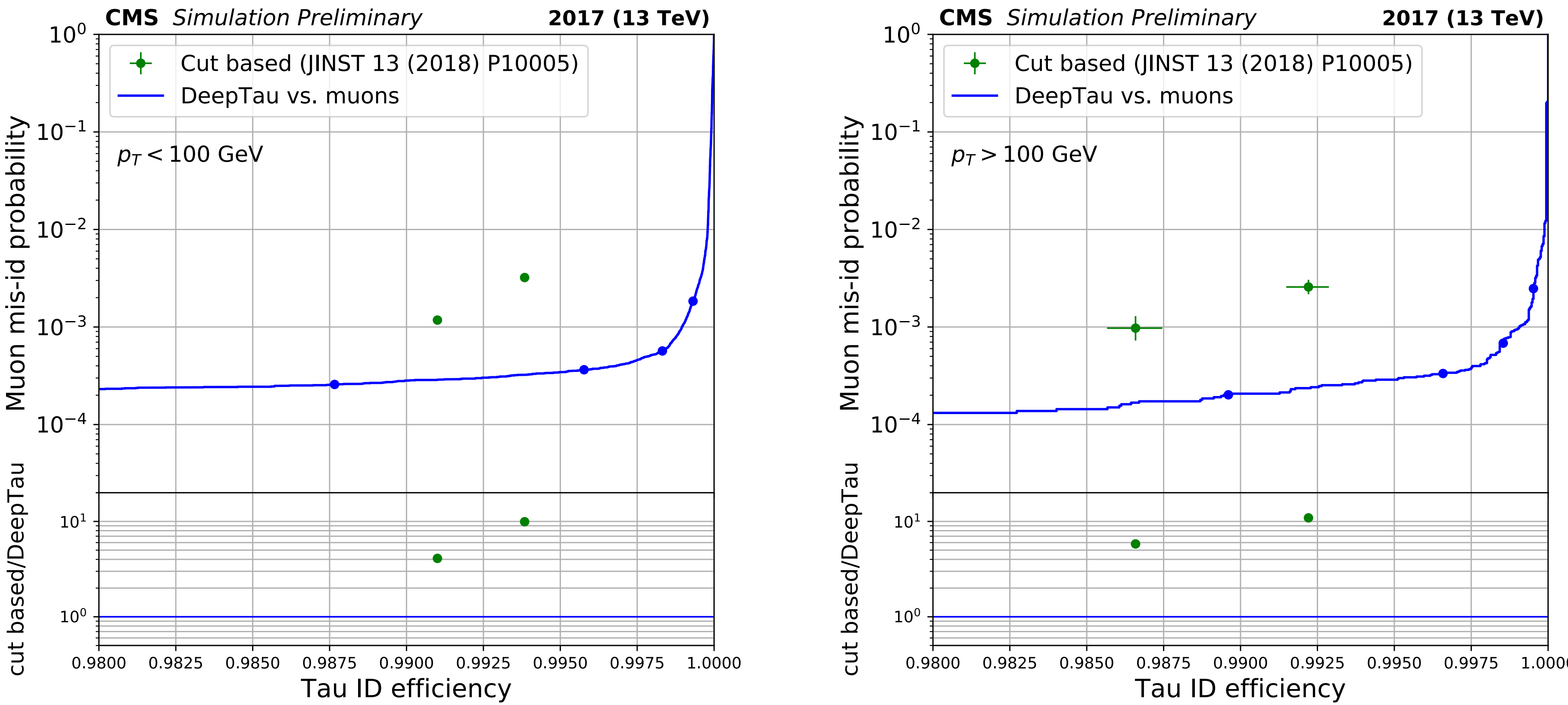


- Consistent improvement at both **low and high p_T**
- Electron misidentification probability **reduces by ~60..95%**

τ_h from $H \to \tau\tau$ MC
τ_e from DY MC

Working points of the discriminators are indicated by the dots

DeepTau discrimination against muons



- Consistent improvement at both **low and high p_T**
- Muon misidentification probability **reduces by ~65..90%**

τ_h from $H \to \tau\tau$ MC
τ_μ from DY MC

Working points of the discriminators are indicated by the dots

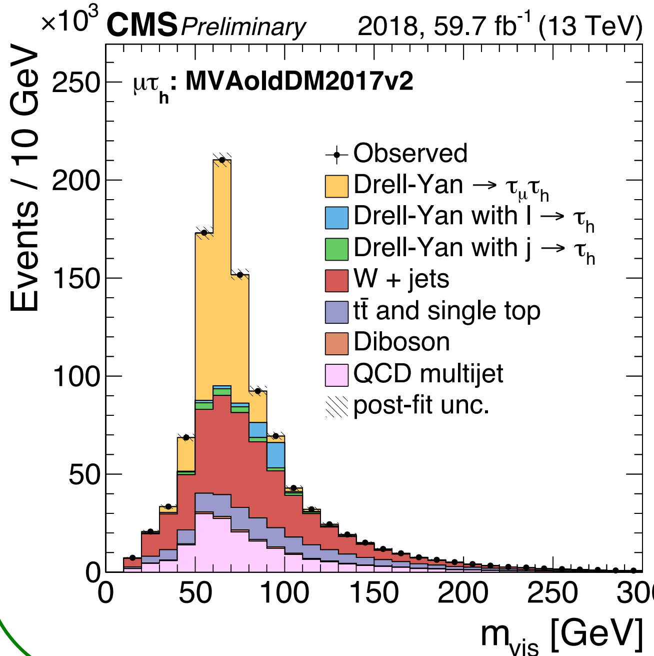
Distribution of the visible $\mu\tau$ mass for 2018 data

Using minimal preselection on muon and tau candidates:

- well identified and isolated muon with $p_T > 25$ GeV, $|\eta| < 2.4$, $|dz| < 0.2$ cm
- tau candidates with $p_T > 20$ GeV, $|\eta| < 2.3$, $|dz| < 0.2$ cm
- $\mu\tau$ pair with an opposite charge and $\Delta R(\mu, \tau) > 0.5$

Selection using discriminators from JINST 13 (2018) P10005:

- Tight WP against jets
- VLoose WP against electrons
- Tight WP against muons

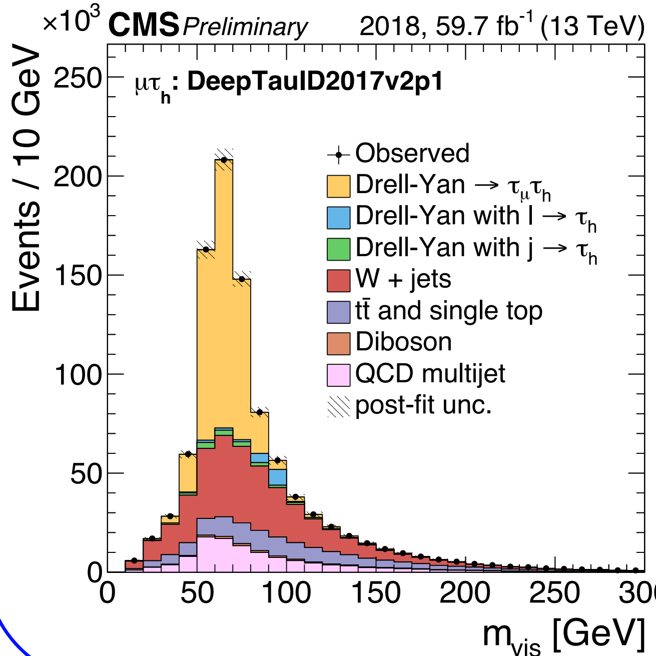


With DeepTau selection, the yield from **genuine τ_h** <u>increases by ≈ 20%</u>, while yield from **fakes** <u>decreases by ≈ 23%</u>

In both plots modelled contributions are fit to the data

Selection using DeepTau IDs:

- Tight WP against jets
- VVLoose WP against electrons
- VLoose against muons



Conclusions

- The new DL-based algorithm, **DeepTau**, to discriminate hadronic tau decays from the main background sources has been presented
  - The DeepTau NN consist of **1 155 353 trainable parameters**
  - The DeepTau inference speed on CPU is $\simeq$ **22 ms/tau**
- Introduction of DeepTau ID provides considerable improvement in the performance in tau identification for Run 2
  - Jet mis-id probability **reduces by more than 50%**
  - Reduction of mis-id probability for electrons (muons) is **up to 95% (90%)**
- On the other hand, the *deep machine learning* is a very actively developing field, and there is still a lot of unexplored potential that could bring further enhancements for the offline and online tau identification and reconstruction

Additional materials

Input variables

- **1 global event variable**: the average energy deposition density (ρ)
- **42 high-level variables** that are used during tau reconstruction or proven to provide discriminating power by previous tau POG studies, including the impact parameters, position of the secondary vertex, sums of energy depositions in the isolation cones et al.
- For each candidate reconstructed within the tau signal or isolation cones, information about 4-momentum, track quality, relation with the primary vertex, calorimeter clusters, and muon stations is used, if available:
  - From **7 to 27 variables** (depending on the candidate type) for each **particle flow candidate**
  - **37 variables** for each fully reconstructed **electron candidate**
  - **37 variables** for each fully reconstructed **muon candidate**
- All inputs with continuous distributions are standardized and truncated to 5σ interval

Updated decay modes

- DeepTau also takes advantage of more inclusive tau reconstruction decay mode definitions that were recently developed in CMS: three charged prongs + 0 or 1 π^0 with relaxed matching conditions added to previously reconstructed 1 charged prong + 0, 1, 2 π^0 and three charged prongs + 0 π^0 with tight matching conditions
  - In the slides, the decay modes defined with this more inclusive criteria are referred to as “updated decay modes”

Loss function

$$L(\boldsymbol{y}^{pred}, \boldsymbol{y}^{true}) = -k_\tau y_\tau^{true} \cdot \log y_\tau^{pred}$$
$$-(k_e + k_\mu + k_j) \cdot F(1 - y_\tau^{pred}, 1 - y_\tau^{true}; \gamma_{cmb})$$
$$-k_F \sum_{\alpha \in \{e,\mu,j\}} k_\alpha \cdot D(y_\tau^{pred}) \cdot N(\gamma_\alpha) \cdot F(y_\alpha^{pred}, y_\alpha^{true}; \gamma_\alpha)$$

tau entropy

binary fake focal loss

categorical fake focal loss

$$F(y_\alpha^{pred}, y_\alpha^{true}; \gamma_\alpha) = y_\alpha^{true} \cdot (1 - y_\alpha^{pred})^{\gamma_\alpha} \cdot \log y_\alpha^{pred}$$

focal loss [1]

$$N(\gamma_\alpha) = 1 \Big/ \int_0^1 (1-p)^{\gamma_\alpha} \cdot \log p$$

Unitary normalization for focal loss

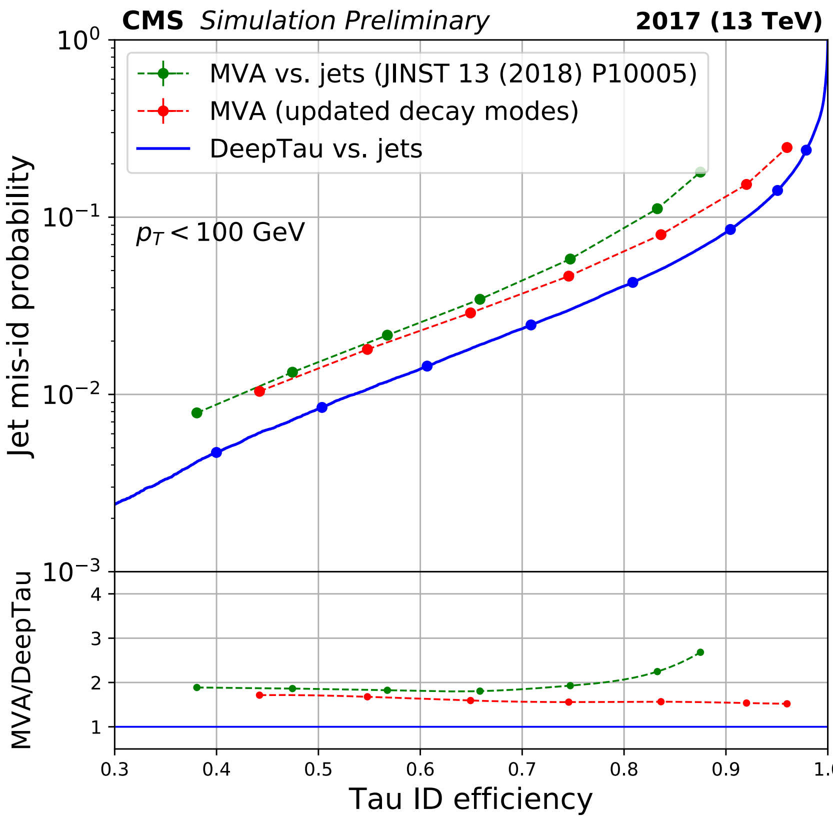
$$D(y_\tau^{pred}) = \frac{1}{2}\left(\tanh\left(70(y_\tau^{pred} - 0.1)\right) + 1\right)$$

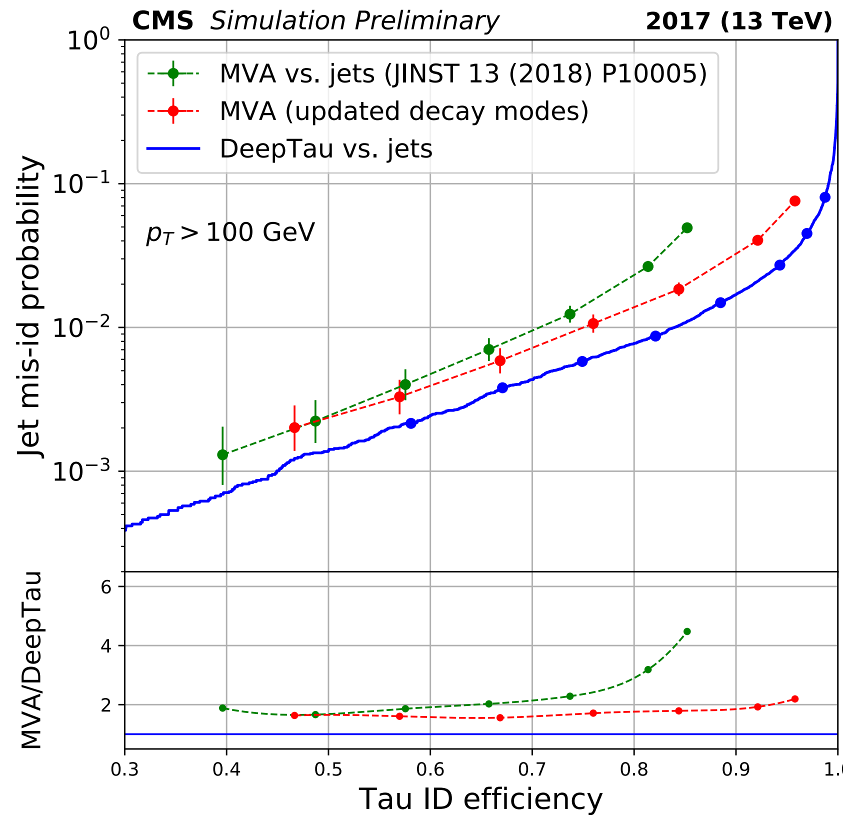
Smooth step function to exclude part of the parameter space where $y_\tau^{pred} < 0.1$

$$(k_e, k_\mu, k_\tau, k_j) = (0.4, 1.0, 2.0, 0.6)$$
$$(\gamma_e, \gamma_\mu, \gamma_{cmb}, \gamma_j) = (2, 2, 0.5, 2)$$
$$k_F = 10$$

[1] T.-Y. Lin et al., *Focal Loss for Dense Object Detection* https://arxiv.org/abs/1708.02002

DeepTau discrimination against jets from W+jets





- Consistent improvement at both **low and high p_T**
- **New decay modes** are part of the gain
- Jet misidentification probability **reduces by ~30..75%**

τ_h from $H \rightarrow \tau\tau$ MC
τ_j from W+jets MC

Working points of the discriminators are indicated by the dots