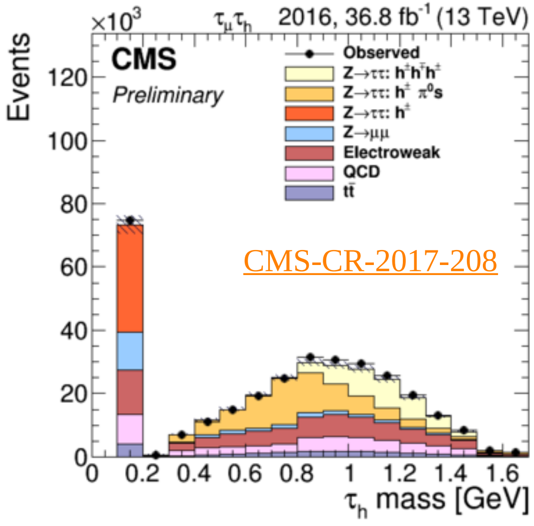
PO&DAS, 13 October 2023

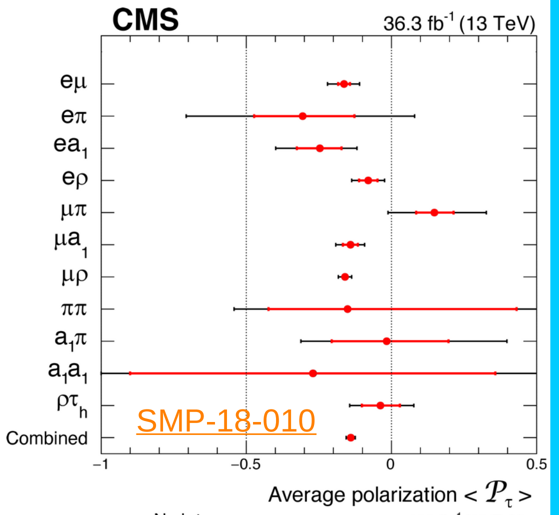
TAU POG exercise

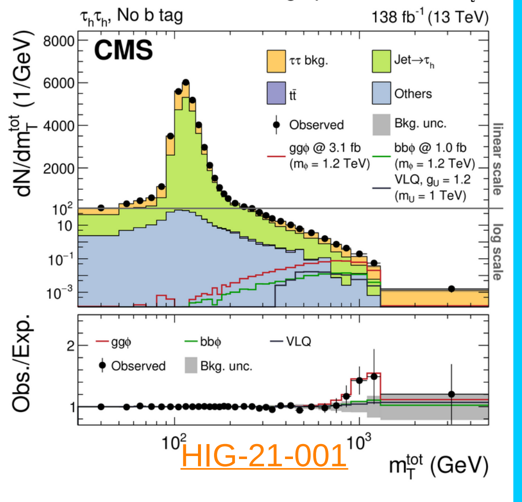
Andrea Cardini, Océane Poncet, Alexei Raspereza

The tau lepton

- τ leptons (taus) are the heaviest leptons in the SM
- They can be used for several measurements with final states involving taus:
  - Test EWK interaction
  - Yukawa couplings of H → fermions
  - Study of the CP properties of the Higgs
  - Tau polarization in Z boson decays
- Searches for BSM physics:
  - Leptoquarks, SUSY, high mass resonances

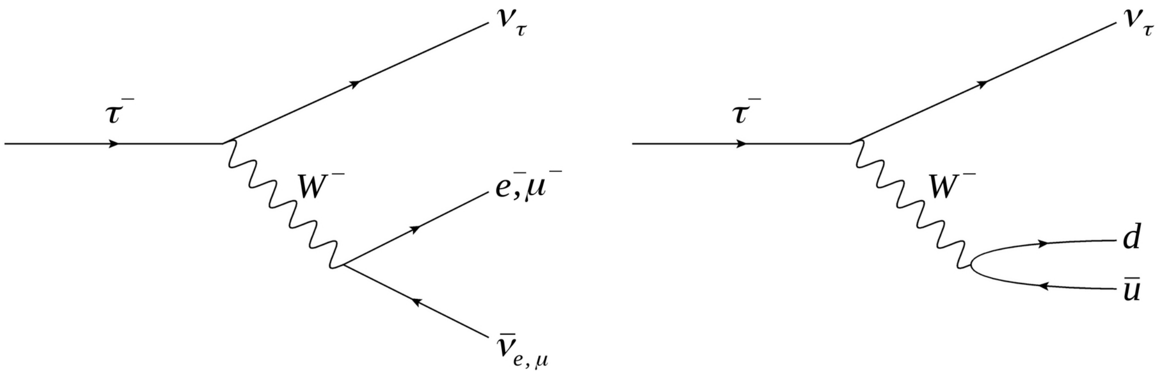


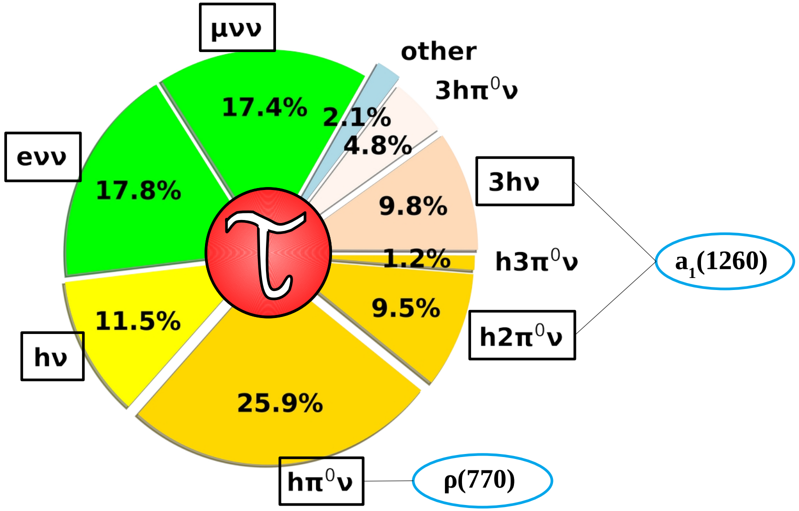




The tau lepton

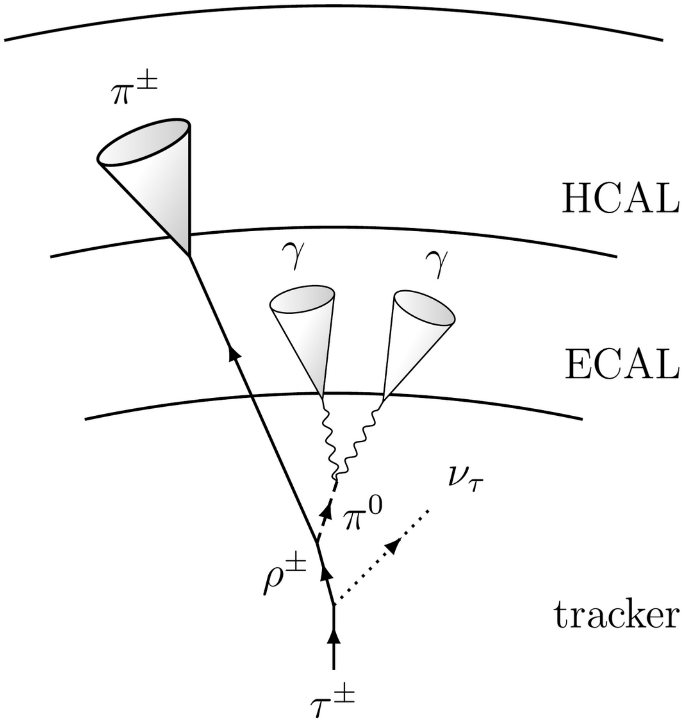
- The taus are important probes of the study of the Standard Model: it is the heaviest lepton, and can only decay via electroweak interac
- Properties:
  - Mass ~1.78 GeV → the only lepton that decay hadronically
  - Decay length ~1.5mm (E ~30 GeV)





- It allows for precise study of the electroweak interaction
- Its decay products *remember* the τ spin
  - Neutrinos have ~definite helicity
  - Hadronic decays are via mesonic resonances

Taus in CMS



$$\tau^\pm \to \rho^\pm \nu_\tau \to \pi^\pm \pi^0 \nu_\tau$$

- Hadronic tau decays are identified as highly collimated jets (hadrons) + electromagnetic clusters (strips): **Hadron-plus-strip** algorithm
  - The algorithm is combinatorial:
    - Selects up to 6 tracks matching to hadronic deposits in HCAL
    - Clusters electrons, positrons and photons into electromagnetic clusters elongated in φ and narrow in η
    - The algorithm recognizes the object as a tau if at least one of the possible combinations matches to a tau decay mode:

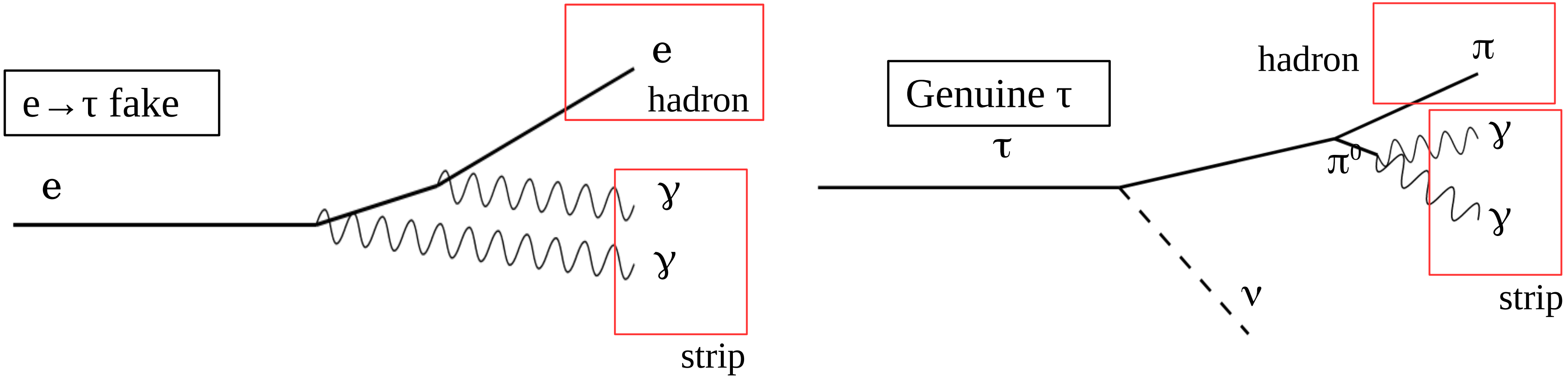
0) $\tau \to h$
1) $\tau \to h + \pi^0$
5) $\tau \to 2h$
6) $\tau \to 2h + \pi^0$
10) $\tau \to 3h$
11) $\tau \to 3h + \pi^0$

(5 and 6: One charged pion is not reconstructed)

$$DM = 5(N_{prong} - 1) + N_{strip}$$

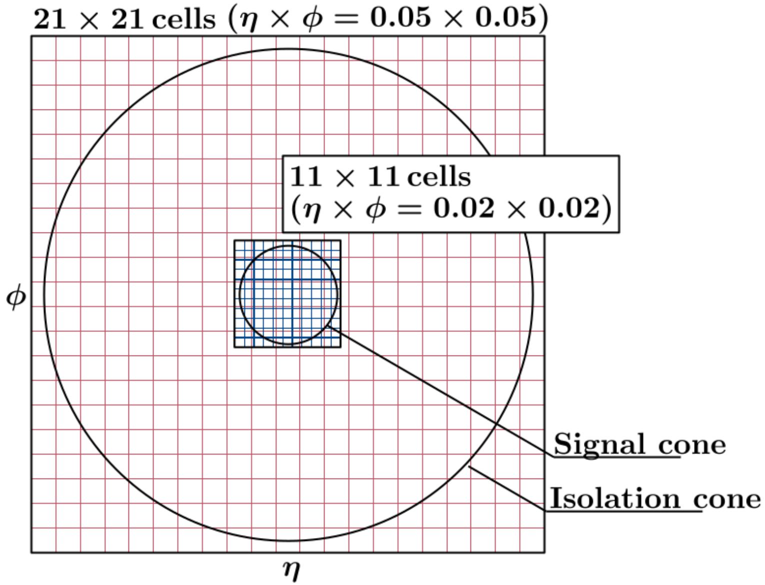
Misidentification of hadronic taus

- Several objects can be misidentified as hadronic taus by the HPS algorithm:
- Jets → a highly collimated quark or gluon jet can be mistaken for any tau decay
- Muons → mainly affects the 1 prong channel
- Electrons → can emit photons via bremsstrahlung radiation and mimic the ρ decay

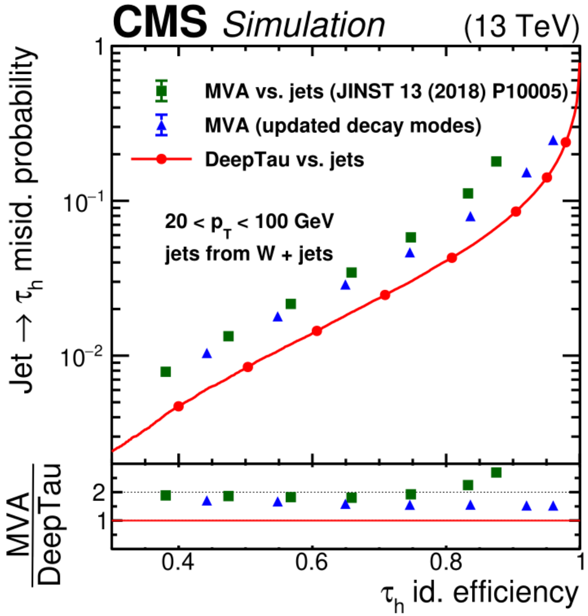


- Rejection of jets and leptons mimicking τ_h performed with **DeepTau** neural network-based identification: TAU-20-001

The DeepTau algorithm



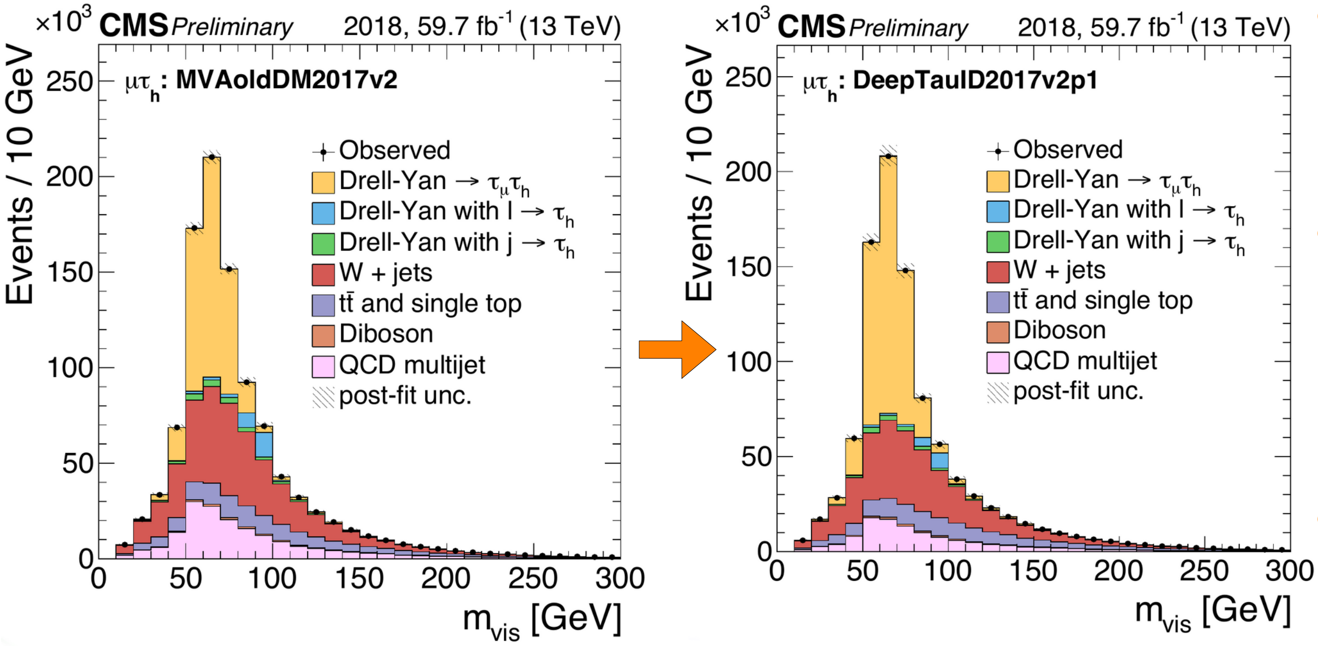
- The DeepTau is a convolutional neural network algorithm
- Properties of electrons, photons, muons and hadrons are collected in a η-φ grid around the HPS τ_h candidate
- Properties of the HPS τ_h candidate and event global features provide additional inputs



- Tau candidates are classified as: jets, genuine τ_h, e or μ
- The output scores are rearranged in three discriminators, one for each source of contamination
  - $$D_\alpha(\boldsymbol{y}) = \frac{y_\tau}{y_\tau + y_\alpha}$$

Tau classification

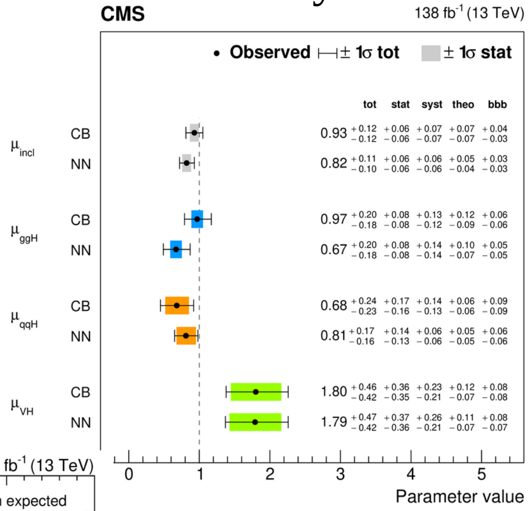
- DeepTau v2.1 was used at analysis level for identifying genuine τ_h → improve S/B ratio
- With respect to the older MVA classifier genuine tau selection and background rejection both improved by ~20%



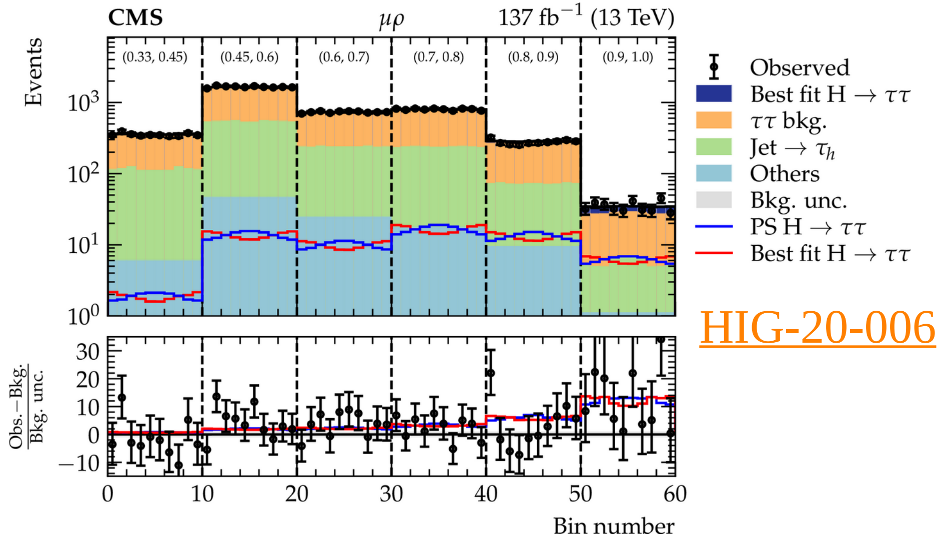
- $Z \rightarrow \mu\mu$ contribution barely present, $Z \rightarrow ee$ also strongly reduced for $e\tau_h$ final states
- Jet → τ_h fakes still dominant background for $\tau_h\tau_h$ channel, received noticeable increase in purity
- The precise measurements of ID, ES, and FR Scale Factors allowed for a good data description in Run 2

DeepTau in Run2

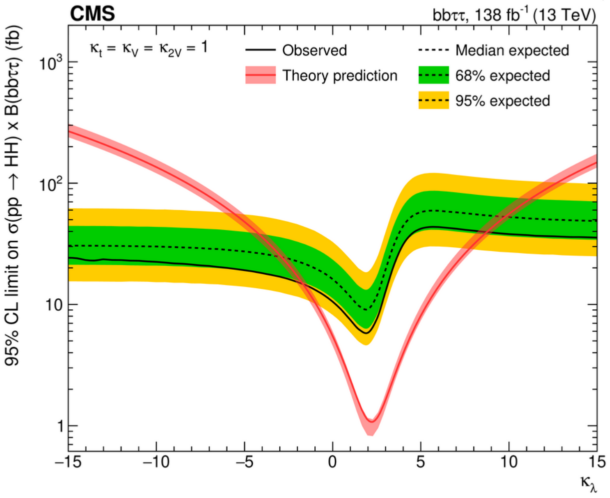
- The DeepTau v2.1 was deployed during Run 2 and used to great effect in several CMS analyses
  - > SM H→ττ differential cross-section measurement
  - > CP structure of the Higgs Yukawa coupling to τ leptons
  - > Tau polarization
  - > Higgs boson pair production, MSSM H→ττ searches, and more.



HIG-19-010

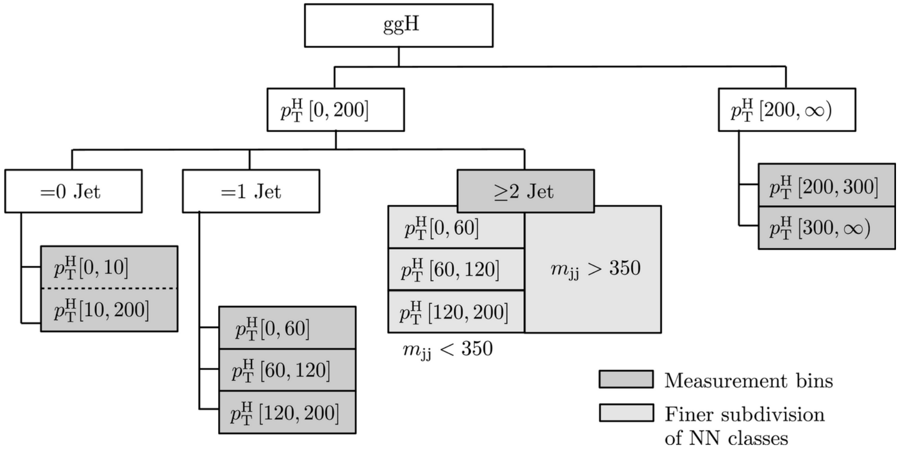


HIG-20-006



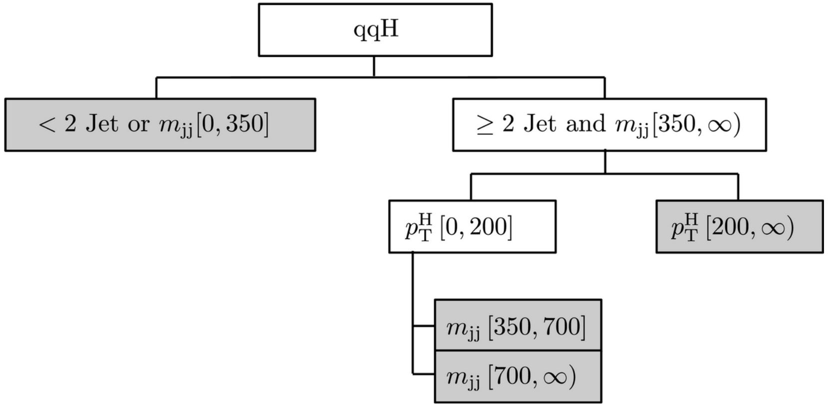
HIG-20-010

- Measurement of the Higgs differential cross-section in di-τ final states

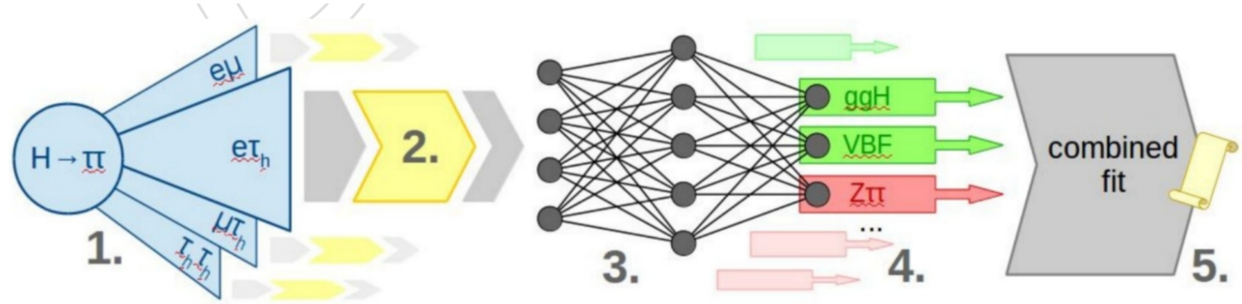


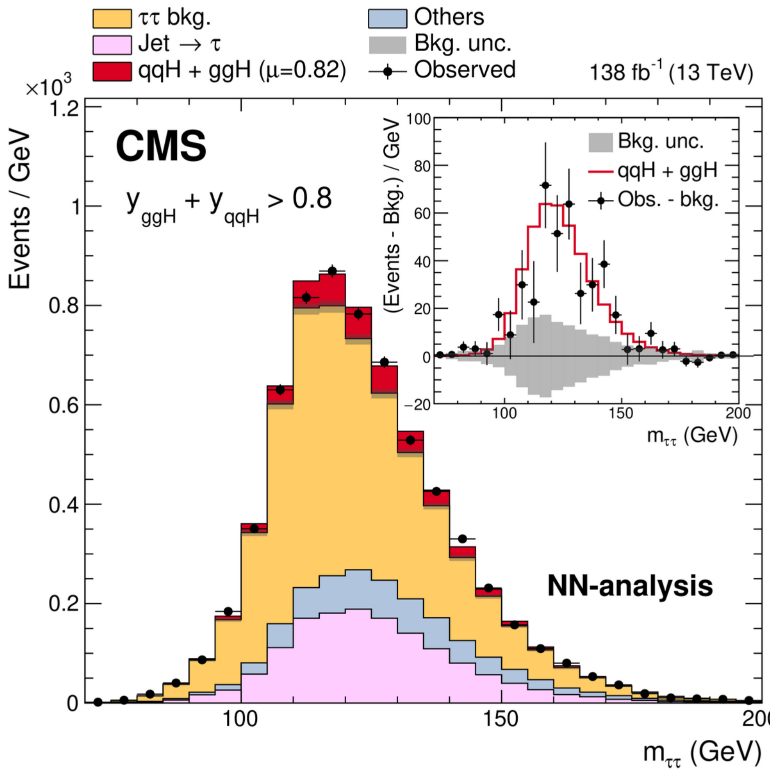
- Differential measurement performed in the STXS scheme → topological phase-space regions to constrain theoretical uncertainties on the cross-section
- Production mechanisms studied *individually*
- Measurement performed with both cut-based and NN-based approaches

- Four final states studied: eμ, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$
- Tau identification fundamental in the 3 channels with higher Branching Fraction
- Higgs production mechanisms: ggH, qqH, VH, ttH



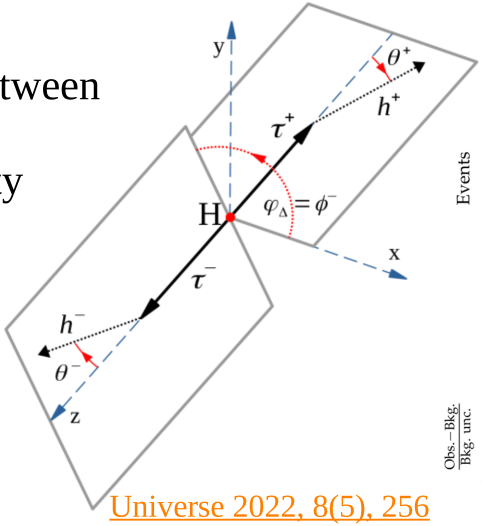
- Hadronic taus in the SM H → ττ analysis have been used as individual objects
- Their properties have been used to construct in a cut-based approach and as input features for a classification NN
  - Tau pT, invariant mass and momenta of di-tau system, visible mass etc.
- Precise τ reconstruction fundamental for this measurement
- DeepTau v2.1 fundamental in ensuring good S/B ratio



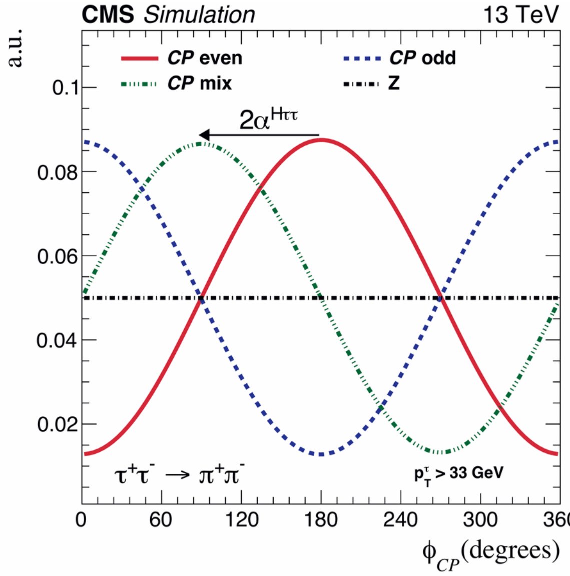


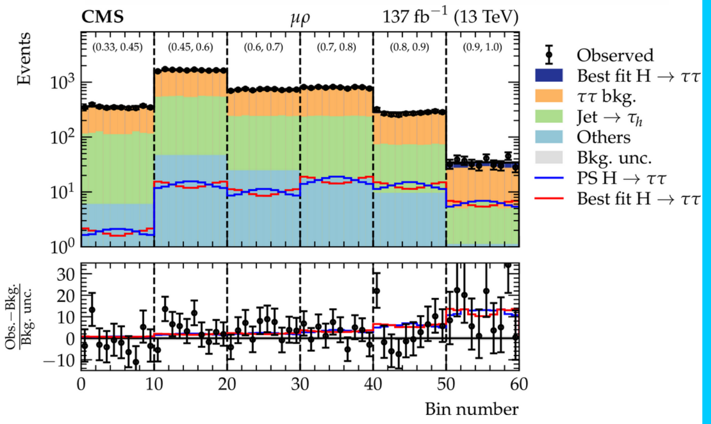
Taus in CMS – Higgs CP in tau decays

- Higgs Yukawa interaction can be parametrized as follows:
  - $$\mathscr{L}_{Y,f} = -\frac{m_f}{v}\bar{\psi}_f(\kappa_f + i\gamma^5\tilde{\kappa}_f)H\psi_f$$
  - Higgs coupling to τ leptons investigated in decays via τ spin correlation
    - CP mixing encoded in $\alpha^{H\tau\tau}$:
      - $\kappa_\tau = \sqrt{\mu^{\tau\tau}}\cos\left(\alpha^{H\tau\tau}\right)$
      - $\tilde{\kappa}_\tau = \sqrt{\mu^{\tau\tau}}\sin\left(\alpha^{H\tau\tau}\right)$
  - Mixing angle can be accessed via the angle between τ decay planes
  - CP mixing appears as a shift in the acoplanarity angle distribution
- Tau decay products take center stage in the reconstruction of the decay planes
- Requires precise identification of tau constituents and decay mode



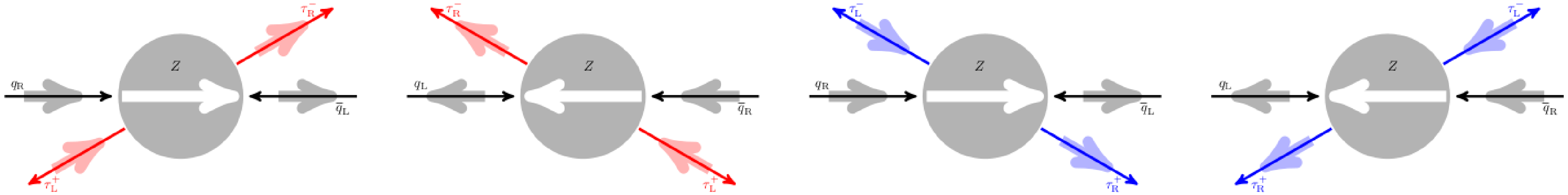
Universe 2022, 8(5), 256





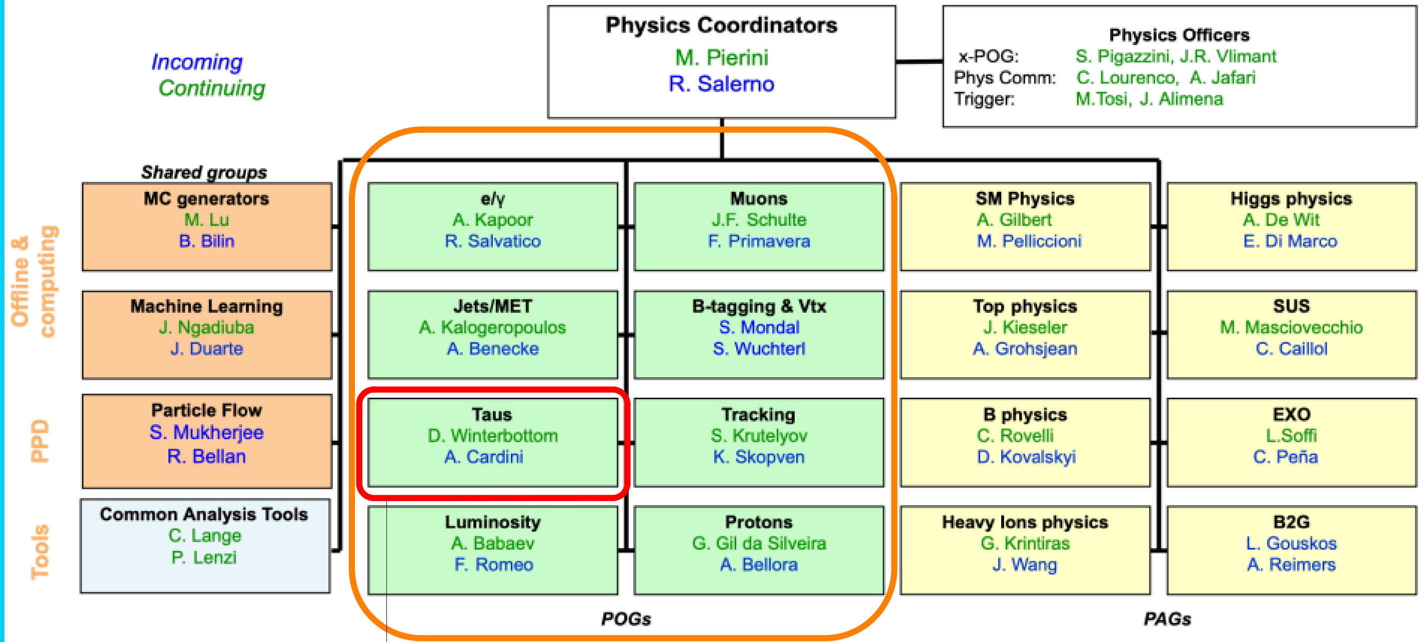
Taus in CMS – Tau polarization

- In Z decays tau leptons are produced with longitudinal spin correlation
- Multiple helicity states are available for Z → ττ decays



- The measurement targets asymmetries in the production of negative (positive) taus with positive (negative) helicity:
  - $A_{\mathrm{FB}} = \frac{1}{\sigma}[\sigma(\cos\theta_\tau > 0) - \sigma(\cos\theta_\tau < 0)] = \frac{3F_1(\hat{s})}{4F_0(\hat{s})}$, forward/backward production cross-section
  - $P_\tau = \frac{1}{\sigma}[\sigma(h_\tau = +1) - \sigma(h_\tau = -1)] = -\frac{F_2(\hat{s})}{F_0(\hat{s})}$, tau polarization
  - $A_{\mathrm{FB}}^{\mathrm{pol}} = \frac{1}{\sigma}[A^{\mathrm{pol}}(\cos\theta_\tau > 0) - A^{pol}(\cos\theta_\tau < 0)] = -\frac{3F_3(\hat{s})}{4F_0(\hat{s})}$, forward/backward polarization
- Crucial to the measurement is the tau scattering angle w.r.t. the beamline and angles between tau decay products like in the CP analysis

TauPOG activities and organization



- Physics object groups fall under physics coordination
- Each group is tasked to develop and maintain reco and ID algorithms for the associated *object*
  - jets+MET, muons, bjets, electrons+photons, tracks, protons, lumi, and **taus**

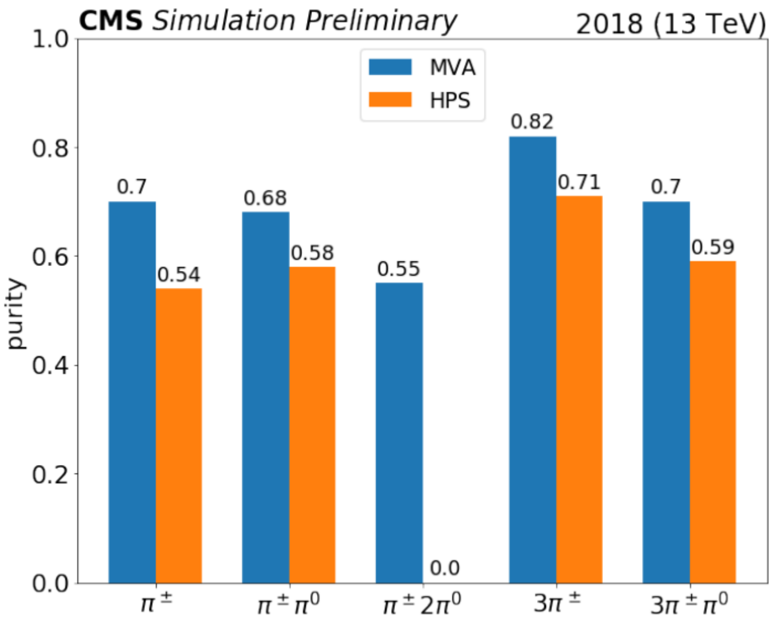
- The tau group is further divided in 3 subgroups + and includes several contacts → link to twiki

(Sub-)group	Position	Convener/Contact
Tau Physics Object Group (L2 TAU POG)	L2	Andrea Cardini, Daniel Winterbottom, all
Tau Calibration-Quality-Monitoring (L3 TAU CQM)	L3	Alexandros Attikis, Dennis Roy
Tau Trigger (L3 TAU TRIG)	L3	Xiaohu Sun, Valeria D'Amante
Tau Algorithms (TAU ALGO)	L3	Olha Lavoryk, Irene Andreou

Validation (offline)	contact	Anusree Vijay, Udit Kandpal
Validation (online)	contact	Gourab Saha
RECO/SW	contact	Michal Bluj, Klitos Savva
Phase II (upgrade)	contact	Abdollah Mohammadi
MC	contact	Mario Sessini
Embedding	contact	Ralf Schmieder

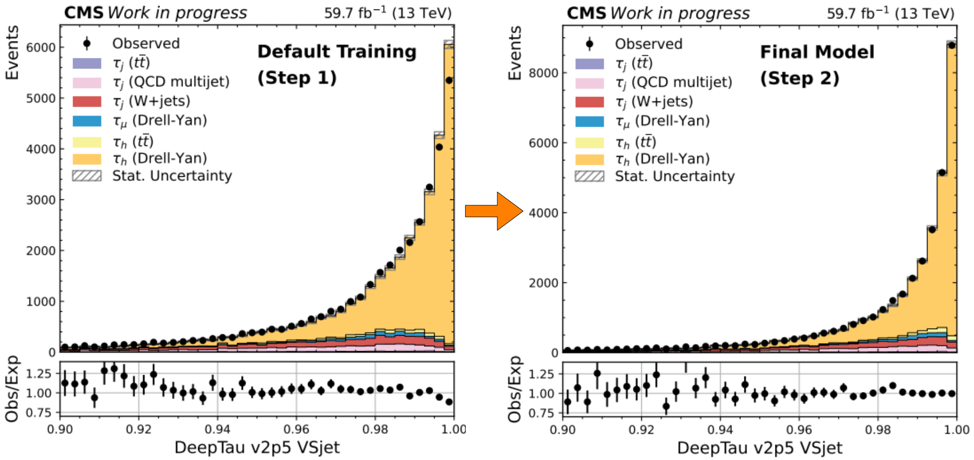
TauALGO activities – tau reco

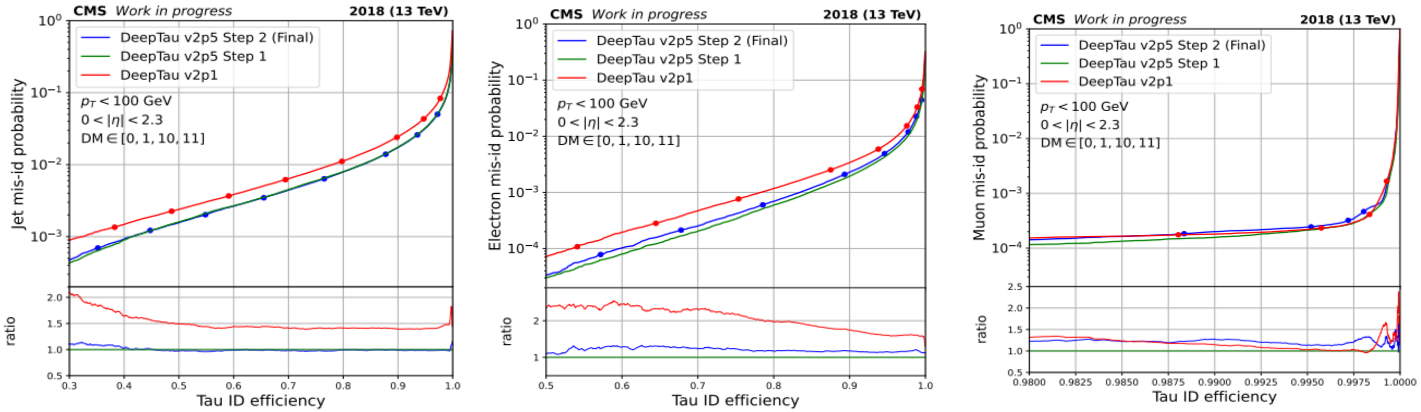
- The Tau ALGO group focuses on reconstruction + identification algorithms for hadronic taus in CMS using Machine Learning techniques
- An MVA-DM BDT was implemented during Run 2 to improve the DM identification of HPS
  - It uses as inputs HPS-DM → cannot recover information previously discarded due to DM-misidentification
  - Adds an additional DM: the 1 prong + 2 strips
- Several studies are in progress to find possible alternatives to the HPS reconstruction
  - ML-based identification and reconstruction of neutral pions → alternative to strip reconstruction
  - Studies to regress the tau full momentum and polarimeter vector



TauALGO – TauID: DeepTau v2p5

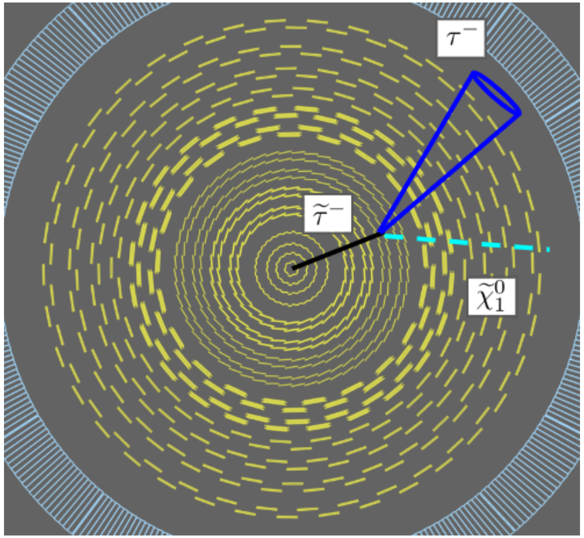
- After the success of DeepTau v2.1 a new version was made and deployed for CMSSW 124x
- As of nanoAOD v10, DeepTau v2.5 is available
  - Improved dataset preparation
  - Introduction of adversarial training as 2nd step to improve coherence in training between data and MC
  - Overall reduction of misidentification rate wrt to DeepTau v2.1 ~10-50%





TauALGO – ID: Particle Net and other IDs

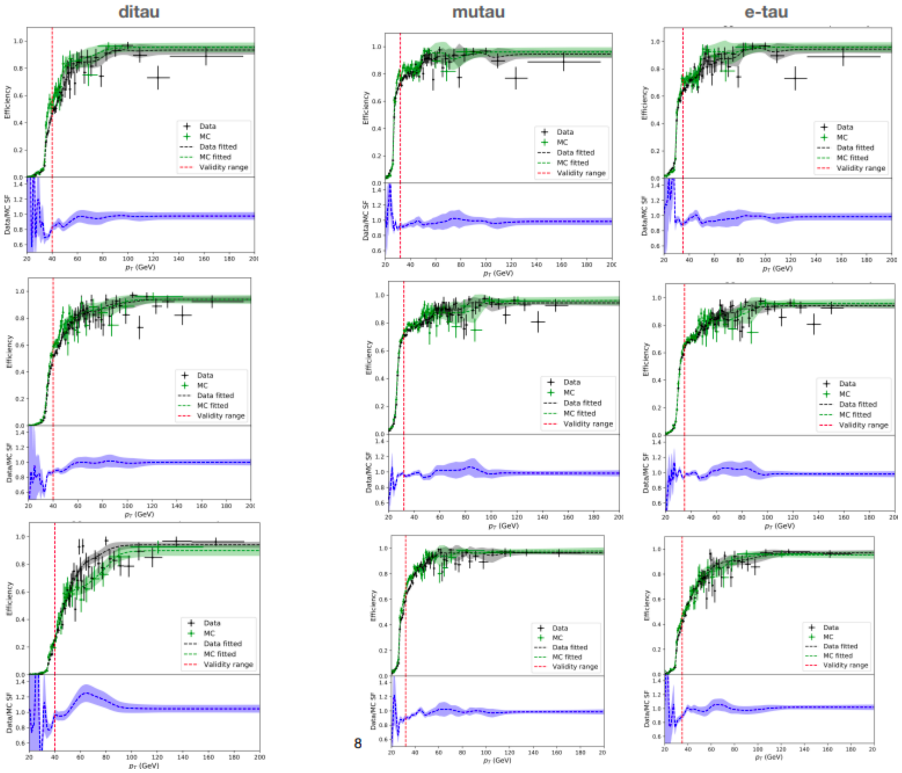
- The TauALGO group has investigated the use of ParticleNET for hadronic tau reconstruction and identification
  - It is a graph network based on CHS and PUPPI jets from miniAOD
  - Allows for *recovery* of taus discarded by HPS
- ParticleNET was deployed in CMSSW 131x with trainings available for CMSSW 130x (nanoAOD v12) for:
  - AK4 jet flavor classification and hadronic tau DM and charge reconstruction
  - AK8 jet classification → possibility to reconstruct merged hadronic taus
- DeepTau v3 under development using Particle Transformers
- Other algorithms under development:
  - Boosted di-tau system identification
  - Displaced tau reconstruction coming from long lived neutral or charged particles



Schematic from Mykyta Shchedrolosiev

Tau Trigger activities

- The name is self-explanatory. The group focuses on tau related triggers:
  - As taus are composite objects, the trigger contains seeds related to the hadronic calorimeter and the tracker
  - Aside from the standard L1 and HLT structure, it includes an intermediate L2 stage for the tau matching
  - More recently DeepTau has also been added at HLT level to improve efficiency



- Other tasks include:
  - measuring the scale factors to correct the trigger efficiency in MC
  - Prepare the trigger menu for upcoming data-taking periods (Run 3 in the past years and now HL-LHC)

TauCQM – activities

- The main goal of the Tau CQM group is to measure the tau identification and fake rate scale factor to correct hadronic taus in simulation
- Beyond providing the scale factors, the group focuses on development of methods for improving the uncertainty model of the corrections:
  - 2D binned (pT vs DM) SF for TauID
  - Combined fit of the tau ID efficiency and ES
  - Combined fit of $e \rightarrow \tau_h$ FR and ES
  - Improved systematic model for ID efficiency
- Additional tools used by the group:
  - Skim miniAOD → nanoAOD (cms-tau-pog/NanoProd)
  - Data-driven estimation of $j \rightarrow \tau_h$ fakes: Fake Factor method
    - Implemented for High pT SF measurements

TauCQM – SF measurements

- Most measurements use variations of a Tag&Probe method:
  - A well identified object acts as a tag (and usually as the trigger for the events recorded)
  - A probe is then used to estimate the efficiency of the DeepTau discriminators + ES effects introduced in MC for the objects that pass the DeepTau WPs

$\mu\tau_h$ channel

TauID+ES
$\mu \rightarrow \tau_h$ FR
$j \rightarrow \tau_h$ FR (high-m_T)
triggerSF

$e\tau_h$ channel

$e \rightarrow \tau_h$ FR+ES

other

High pT ID+ES: high-m_T region for $W^* \rightarrow \mu\nu/\tau\nu$ events; Fake rate in ee/$\mu\mu$ + τ final states

New: ID+ES combined measurements currently being tested & validated

The Tau Framework

- Introduced in Run 2 for the measurement of scale factors it is now the standard tool we use to measure corrections in the TauPOG

- **Tau FW general structure:**

  - PicoProducer → analyzer dedicated to creating flat tuples from nanoAODs
  - Plotter → common plotting tools
  - Fitter → tools to operate statistical inference (depends on the Combine and Combine Harvester packages)
  - common → routines used by multiple parts of the FW
  - docs → documentation of tutorials, including this one :)

Data processing and structures in CMS

- The TauFW starts with the nanoAOD processing… But what is a nanoAOD?
- Data in CMS is processed in multiple stages
  - In CMS the data collected in the detector is processed at the CERN tier 0
    - This is known as StreamExpress → it lasts ~48h in the system and includes a sampling of the data collected by each subdetector
  - The conditions of the detector and the calibrations are collected by the AlcaDB group and used as input for the prompt calibration loop (PCL)
  - The processed full statistics of the detector forms the PromptReco dataset
  - Further iteration of the alignment, detector calibrations etc. are re-processed at the end of the year and form the ReReco campaign
  - In Run 2 the Legacy and Ultra-legacy were then performed to guarantee top quality data to be stored as a legacy from the data-taking period

Data tiers in CMS

- At the detector, data is collected as energy deposits and hits → it requires a large storage capacity in order to include all information from each subdetector
- This is the RAW data tier, further processing allows to reconstruct objects and create the physics data tiers used in analyses, the AODs (analysis object data)

CMS Data Tier Listing

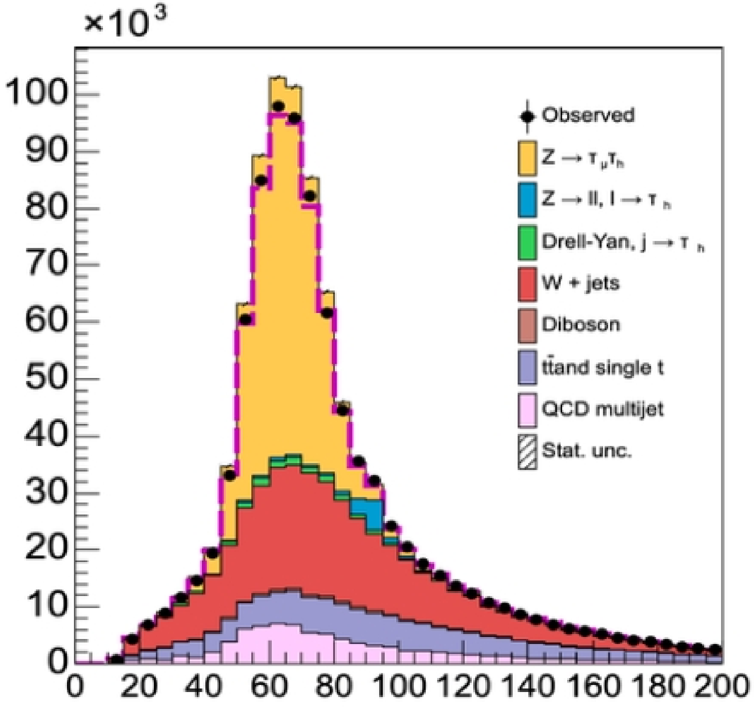
Event Format	Contents	Purpose	Data Type Ref	Event Size (MB)
DAQ-RAW	Detector data from front end electronics + L1 trigger result.	Primary record of physics event. Input to online HLT		1-1.5
RAW	Detector data after online formatting, the L1 trigger result, the result of the HLT selections (HLT trigger bits), potentially some of the higher level quantities calculated during HLT processing.	Input to Tier-0 reconstruction. Primary archive of events at CERN.		0.70-0.75
RECO	Reconstructed objects (tracks, vertices, jets, electrons, muons, etc.) and reconstructed hits/clusters	Output of Tier-0 reconstruction and subsequent rereconstruction passes. Supports re-finding of tracks, etc.	RECO & AOD	1.3-1.4
AOD	Subset of RECO. Reconstructed objects (tracks, vertices, jets, electrons, muons, etc.). Possible small quantities of very localised hit information.	Physics analysis, limited refitting of tracks and clusters	RECO & AOD	0.05
TAG	Run/event number, high-level physics objects, e.g. used to index events.	Rapid identification of events for further study (event directory).		0.01
FEVT	Full Event: Term used to refer to RAW+RECO together (not a distinct format).	multiple		1.75
GEN	Generated Monte Carlo event	-		-
SIM	Energy depositions of MC particles in detector (sim hits).	-		-
DIGI	Sim hits converted into detector response. Basically the same as the RAW output of the detector.	-		1.5

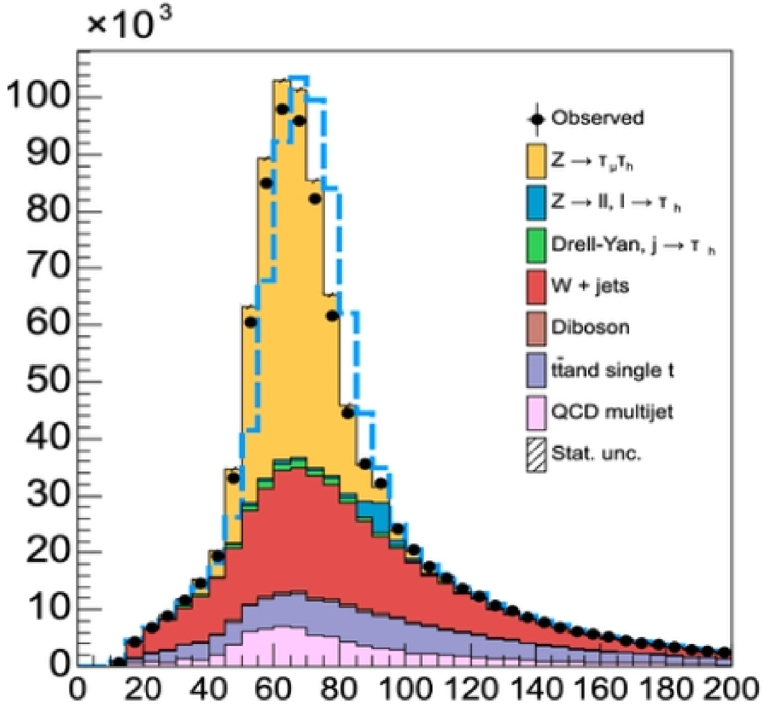
From AOD to nanoAOD to PicoTuple

- The AOD format still requires ~50kb per event → 1M event occupies 50 GB and in CMS we deal with several millions of events for each simulated or reconstructed events
- The amount of storage needed to exchange this amount of data is high + long computation time
- miniAODs and nanoAODs were introduced as a subset of AODs with a simplified structure
  - miniAODs are based on collections accessible with EDM analyzers
  - nanoAODs include trees with arrays but otherwise are flat
- The processing of miniAODs is done centrally in CMS, and stable versions of central nanoAODs are also released centrally
- To be able to add developmental material like new IDs in the TauPOG we developed our own nanoAOD producer: the NanoProd tool
- The TauFW then operates a flattening and skimming from the central or custom nanoAODs to create pico tuples

The exercise

- Today you will have a look at a tau ID efficiency and energy scale measurement





TauID – alters the normalization for processes with genuine τ_h

TauES – alters the energy of the genuine τ_h and consequently shifts the invariant mass distributions

PO&DAS, 13 October 2023

TAU POG exercise

Andrea Cardini, Océane Poncet, Alexei Raspereza