Measurement of Higgs Boson CP Properties in Decays to Tau Leptons Using Run 3 Data

FSP CMS Annual Meeting

Stepan Zakharov on behalf of CP analysis team

Deutsches Elektronen-Synchrotron (DESY)

CP analysis team

DESY.

Elisabetta Gallo
Alexei Raspereza
Jacopo Malvaso
Stepan Zakharov
Mathilde Lodovica Witt

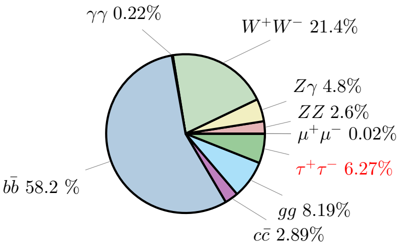
IMPERIAL

Daniel Winterbottom
David Colling
Lucas Russell
Klitos Savva
Irene Andreou

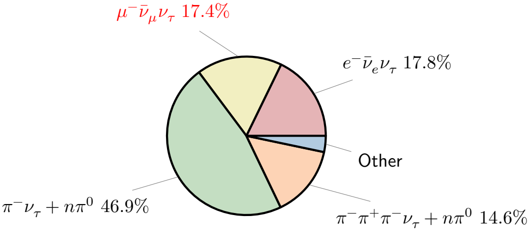
IPHC Institut Pluridisciplinaire Hubert CURIEN STRASBOURG

Anne-Catherine Le Bihan
Ulrich Goerlach
Gourab Saha
Oceane Poncet
Cyril Eschenlauer

CP violation in the Higgs sector



Higgs branching ratios



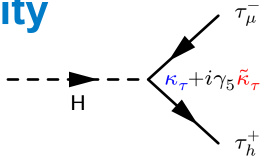
Tau lepton branching ratios

- SM Higgs boson is **CP-even** $J^{CP} = 0^{++}$
- **Pure CP-odd** state was **excluded** in several Higgs decay channels using Run2 data
- **CP admixture** couplings are still allowed experimentally

Why look at the $H \to \tau\tau$?

- Taus carry **spin–parity** information of the Higgs
- W/Z bosons have no CP-odd tree-level coupling

$H \to \tau\tau$: accessing H boson CP parity



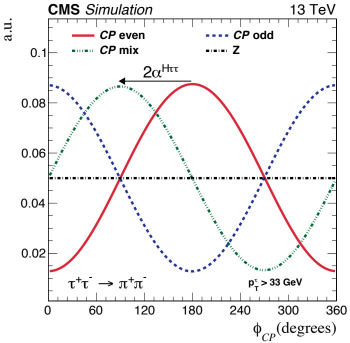
- $\mathcal{L}_Y = \frac{m_\tau}{\nu}\left(\kappa_\tau \bar{\tau}\tau + \tilde{\kappa}_\tau \bar{\tau} i\gamma_5 \tau\right) H$
- Parametrise:

$$\begin{cases} \kappa_\tau = \cos\alpha^{H\tau\tau} \\ \tilde{\kappa}_\tau = \sin\alpha^{H\tau\tau} \end{cases}, \quad \text{assuming } \gamma_\tau \gg 1, \beta_\tau \approx 1$$

- Note that:

$$d\Gamma[H \to \tau\tau] \sim 1 - \underbrace{(\mathbf{s}_\perp^{\tau^-} \cdot \mathbf{s}_\perp^{\tau^+})}_{\sim \cos\phi_{CP}} \cos 2\alpha^{H\tau\tau} - \underbrace{\boldsymbol{\tau} \cdot [\mathbf{s}_\perp^{\tau^-} \times \mathbf{s}_\perp^{+}]}_{\sim \sin\phi_{CP}} \sin 2\alpha^{H\tau\tau}$$

- $$\frac{d\Gamma}{d\phi_{CP}} \sim 1 - \frac{\pi^2}{16} f(\boldsymbol{\tau}^+, \boldsymbol{\tau}^-) \cos(\phi_{CP} - 2\alpha^{H\tau\tau})$$

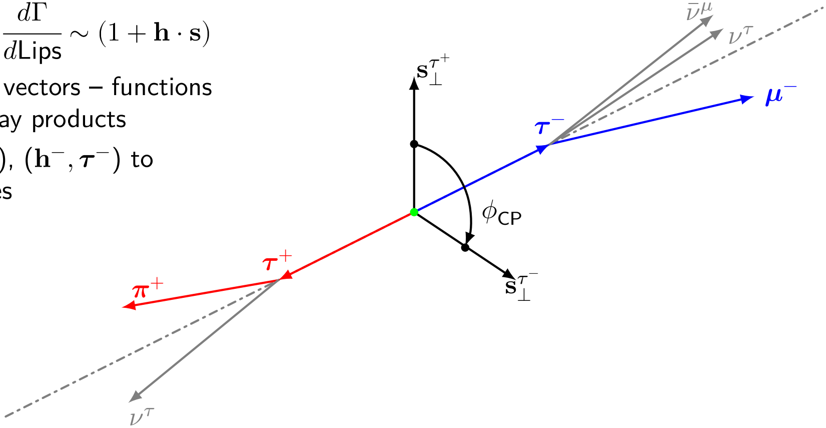


CP-even $\Rightarrow \quad \alpha^{H\tau\tau} = 0$

CP-odd $\Rightarrow \quad \alpha^{H\tau\tau} = 90^\circ$

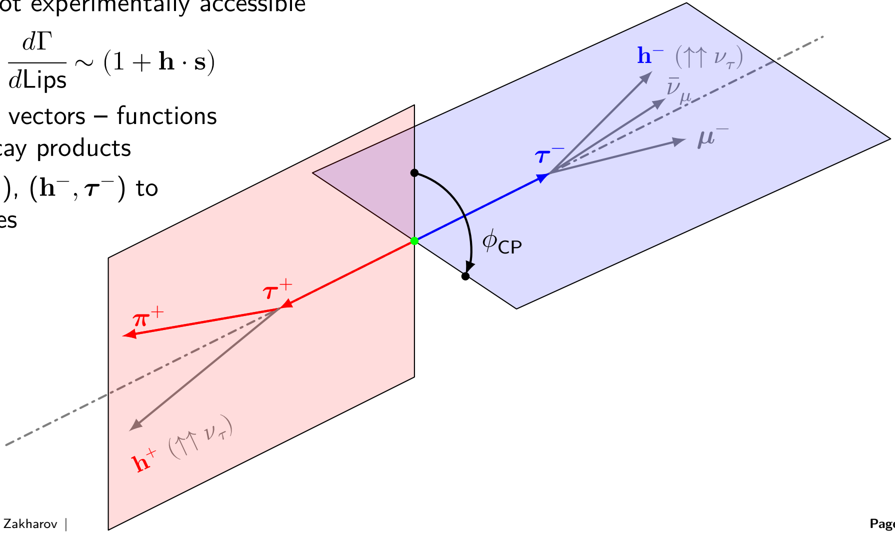
How to measure τ pair spin correlation?

- $\mathbf{s}^{\tau^-}$ and $\mathbf{s}^{\tau^+}$ are not experimentally accessible
- Single tau decay: $\frac{d\Gamma}{d\text{Lips}} \sim (1 + \mathbf{h} \cdot \mathbf{s})$
- $\mathbf{h}^{\pm}$ – polarimetric vectors – functions of $\boldsymbol{\tau}$ and their decay products
- Use pairs $(\mathbf{h}^+, \boldsymbol{\tau}^+)$, $(\mathbf{h}^-, \boldsymbol{\tau}^-)$ to define decay planes

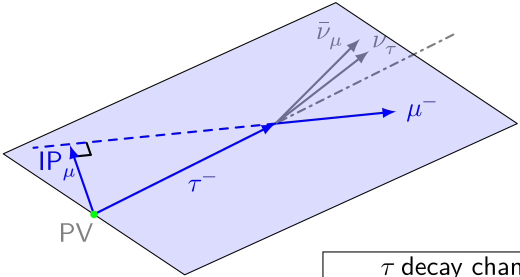


How to measure τ pair spin correlation?

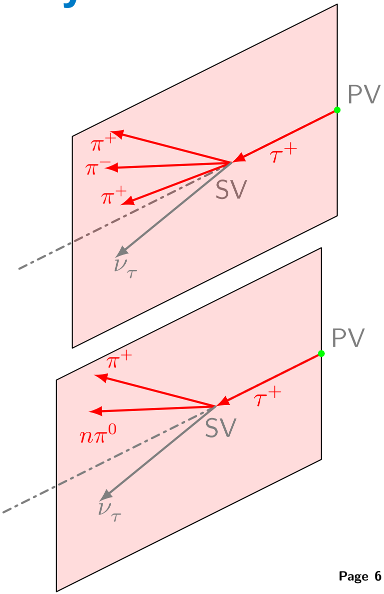
- $\mathbf{s}^{\tau^-}$ and $\mathbf{s}^{\tau^+}$ are not experimentally accessible
- Single tau decay: $\frac{d\Gamma}{d\text{Lips}} \sim (1 + \mathbf{h} \cdot \mathbf{s})$
- $\mathbf{h}^{\pm}$ – polarimetric vectors – functions of $\boldsymbol{\tau}$ and their decay products
- Use pairs $(\mathbf{h}^+, \boldsymbol{\tau}^+)$, $(\mathbf{h}^-, \boldsymbol{\tau}^-)$ to define decay planes



Decay plane definition for various τ decay channels



τ decay channel	DP vectors
$\ell\nu\bar{\nu}$	$(p_\ell,\ \text{IP}_\ell)$
$\pi\nu$	$(p_\pi,\ \text{IP}_\pi)$
$\rho\nu_\tau \to \pi\pi^0\nu_\tau$	$(p_\pi,\ p_{\pi^0})$
$a_1\nu_\tau \to \underbrace{\pi^+\pi^0}_{\rho^+}\pi^0\nu_\tau$	$(p_{\pi^+},\ p_{(\pi^0,\pi^0)})$
$a_1\nu_\tau \to \underbrace{\pi^+\pi^-}_{\rho^0}\pi^+\nu_\tau$	$(p_{\pi(\rho)},\ p_{\pi(\rho)})$



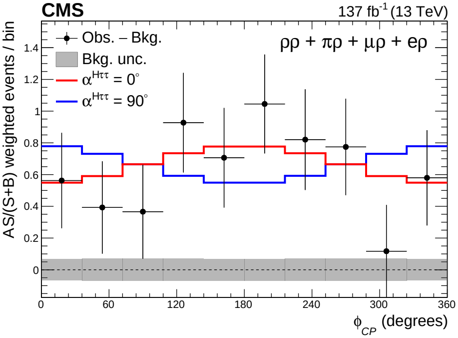
CP H$\to \tau\tau$: current status

CMS Run2 result is: $\alpha^{H\tau\tau} = -1 \pm 19(\text{stat.}) \pm 1(\text{syst.}) \pm 2(\text{bin-by-bin}) \pm 1(\text{theo.})^\circ$ @ 68.3% CL

- Statistically-dominated uncertainty
- To probe BSM scenarios: $\sigma(\alpha^{H\tau\tau}) \leq 7^\circ$ required
- Expecting $\sigma_{\text{stat}} \downarrow 40\%$ by adding Run 3 stats
- Need to improve:
  - φ_{CP} reconstruction methods
  - Analysis scalability

JHEP06(2022)012

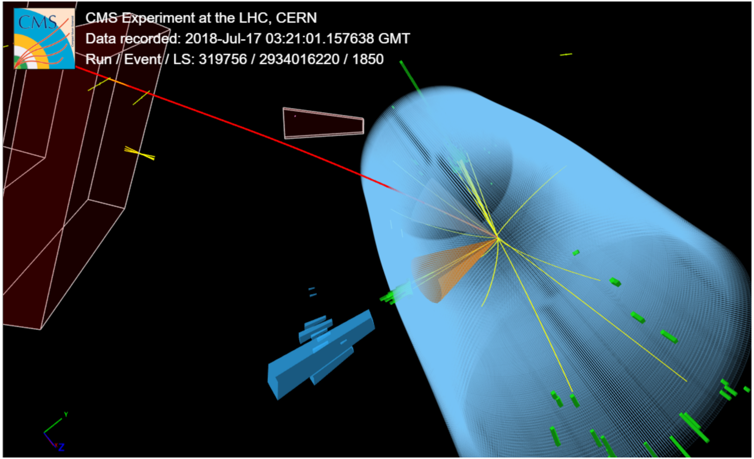
Current result



Event selection

$$pp \to H \to (\tau \to \mu\bar{\nu}_\mu\nu_\tau)(\tau \to 3\pi\nu_\tau)$$

JHEP 06 (2022) 012



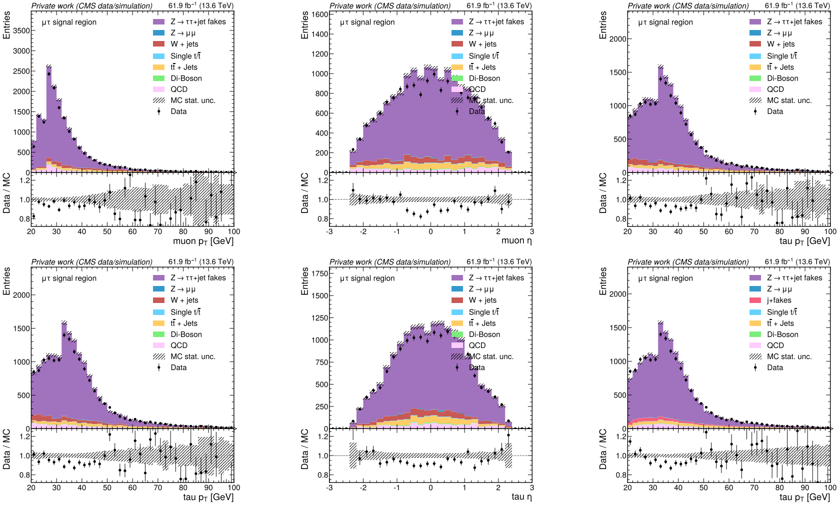
Select:

- $\mu\tau$ pairs
- $q_\mu \cdot q_\tau < 0$
- $\Delta R(\mu, \tau) > 0.5$

Discard if:

1. **Di-lepton veto**: it contains lepton pair of the same flavour ($Z \to \ell\ell$)
2. **Extra lepton veto**: it contains a third lepton (VV)

Control plots



Data/MC agreement is good

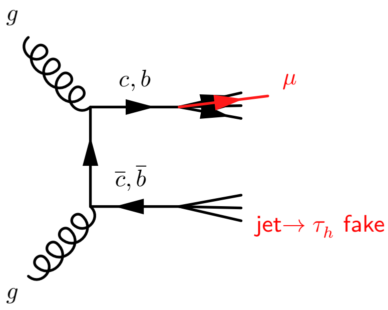
Jet backgrounds

Main caveats:

> Accuracy in description of jet fragmentation

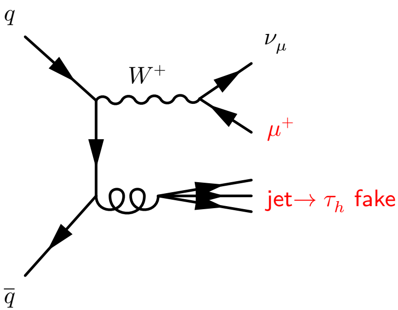
> Fake rates in Data and its modelling in MC

QCD multijet



> $\sigma > 10\,\mu b$
> jet $\rightarrow \mu$ fake rate $\sim 10^{-4}$

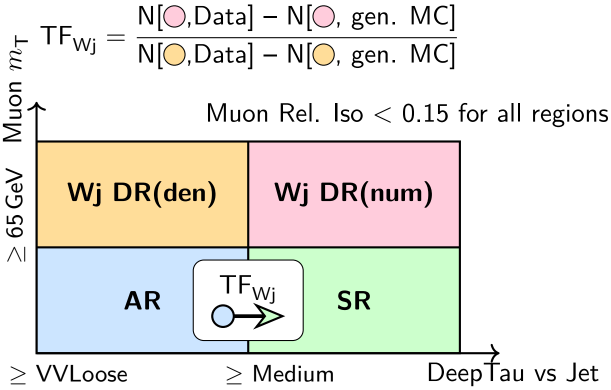
W+jets



> $\sigma \simeq 65\,nb$
> Contains prompt muons

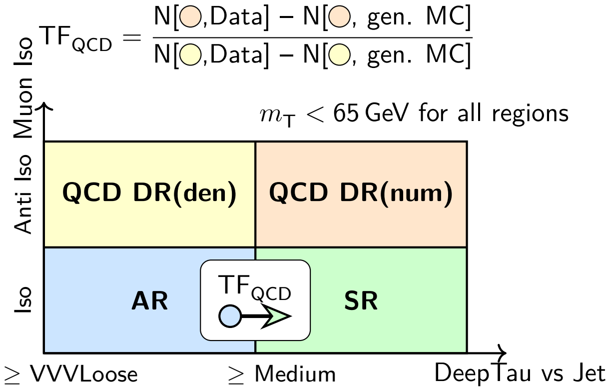
We used a data-driven method to estimate both of the backgrounds for the analysis

Transfer Factor method: W+jets



$$m_T(\mathbf{p_1}, \mathbf{p_2}) = \sqrt{2|\mathbf{p_1}||\mathbf{p_2}|[1 - \cos(\mathbf{p_1}, \mathbf{p_2})]}$$

Transfer Factor method: QCD



Combining W+jets and QCD

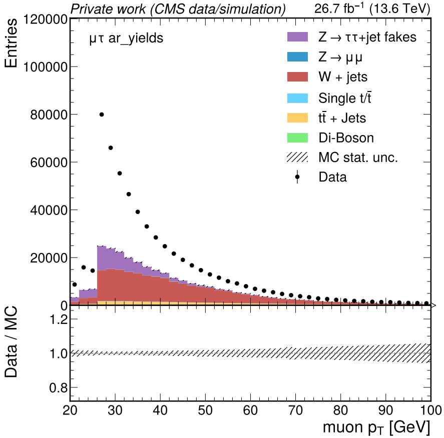
- Applied Transfer Factors on event level:

  $\mathrm{TF}_{\mathrm{Wj,QCD}} = f[p_\mathrm{T}(\tau), n_{\mathrm{jets}}, \mathrm{DM}(\tau)]$

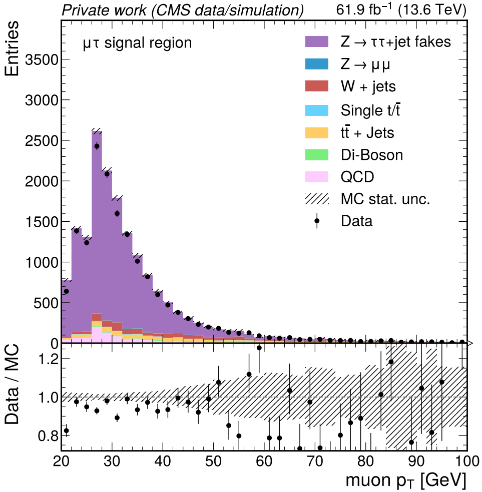
- Calculated yields:
  - $w_{\mathrm{Wj}} = \frac{\text{jet fake MC} - \text{gen. MC}}{\text{Data} - \text{gen. MC}}$
  - $w_{\mathrm{QCD}} = \frac{\text{Data} - \text{jet fake MC} - \text{gen. MC}}{\text{Data} - \text{gen. MC}}$

- Combined TFs and yields to get the estimate of jet fakes:

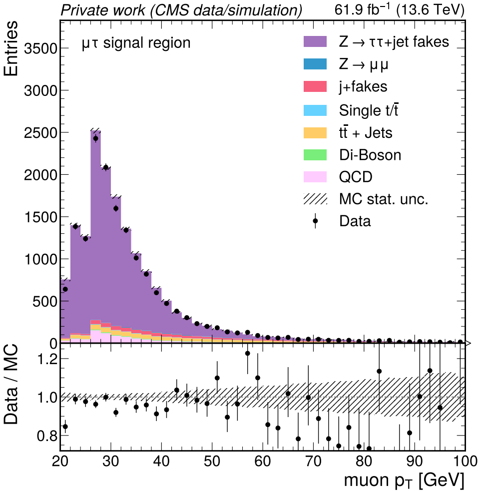
$$\mathrm{N}_{\text{jet fakes}} = \mathrm{N}_{\mathrm{Data}} \cdot \left(\mathrm{TF}_{\mathrm{Wj}} \cdot w_{\mathrm{Wj}} + \mathrm{TF}_{\mathrm{QCD}} \cdot w_{\mathrm{QCD}}\right)$$



Transfer Factor method: results



W+jets from sim., QCD – data-driven

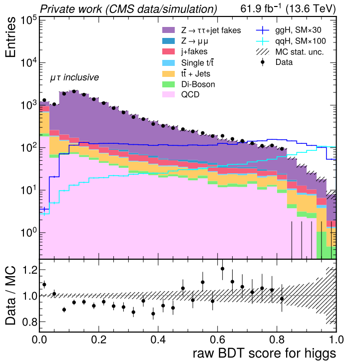


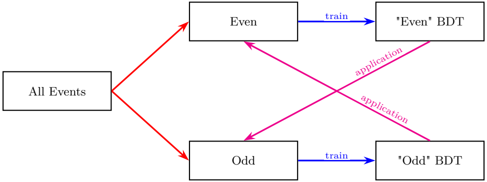
W+jets & QCD estimated with TF method

Event categorisation

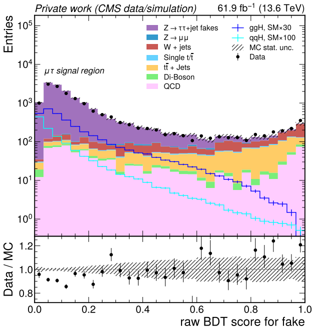
- Trained a BDT to separate signal and background
- Input variables:
  - **tau**: $p_T,\ \eta,\ m_T(\mu)$
  - **mu**: $p_T,\ \eta,\ m_T(\tau)$
  - **pair**: $dR,\ d\phi,\ p_{T\ \text{vis.}},\ m_{\text{vis.}},\ m_{\text{FastMTT}}$
  - **MET** : $p_T,\ d\phi(\tau, \text{MET}),\ d\phi(\mu, \text{MET})$
  - **$\mathbf{m_T}$** : $m_T(\tau),\ m_T(\mu),\ m_T(\mu,\tau),\ m_{T,\text{tot.}}$
  - **jets** : $n_{\text{jets}}, p_T,\ \eta,\ m_{\text{jj}},\ p_{T,\text{jj}}, d\eta_{jj}$
- Model parameters were optimised via Optuna to reach maximal approximate median significance

The model was trained and optimized by Lucas Russel from IC

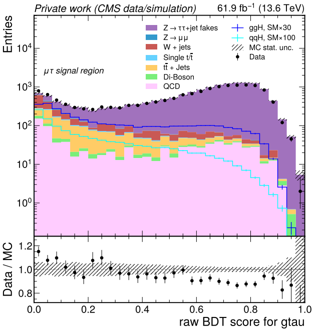




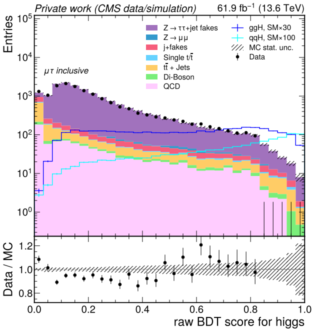
BDT raw outputs



Jet fake category (fakes)



Genuine tau category (gen. tau)



Signal category (higgs)

Defining signal regions

> Events with raw higgs score being the highest from the signal region

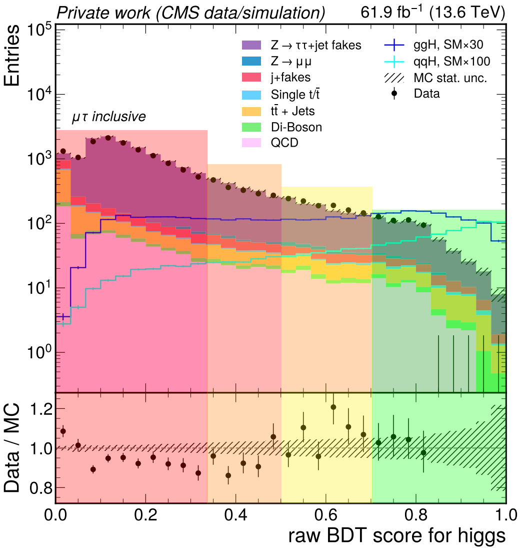
> Then split this dataset into 3 bins: $\in [0.33, 0.5),\ [0.5, 0.7),\ [0.7, 1.0]$

> For each regions calculated φ_{CP} in 4 channels:

- ⊛ $\mu\pi$: $\text{DM}_{\text{PNet}} = 0,\ \text{IP}_{\tau} > 1.25$
- ⊛ $\mu\rho$: $\text{DM}_{\text{PNet}} = 1,\ \text{DM}_{\text{HPS}} = 1,\ E_{\text{split}} > 0.2$
- ⊛ μa_1^{1pr} : $\text{DM}_{\text{PNet}} = 2,\ \text{DM}_{\text{HPS}} = 1,\ E_{\text{split}} > 0.2$
- ⊛ μa_1^{3pr} : $\text{DM}_{\text{PNet}} = 10$, has refit SV

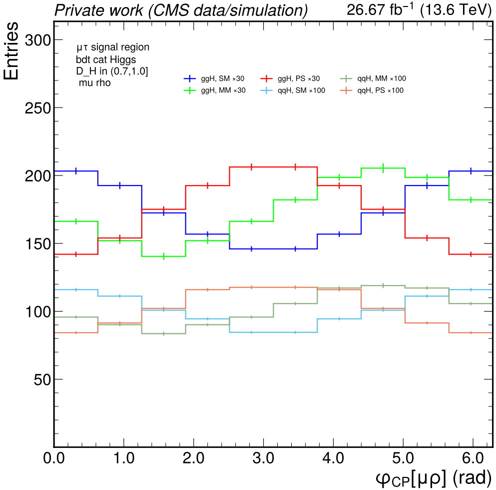
> 12 signal categories in total

* New for the Run3 analysis, $E_{\text{split}} = \frac{E_{\pi} - E_{\pi^0}}{E_{\pi} + E_{\pi^0}}$



Reco level distributions for the signal $H \to \tau_\mu \tau_h$

- Signal sample was produced with tau decays being uncorrelated
- Tau pair spin correlation was introduced via TauSpinner weight
- Applied weights for the CP-even, CP-odd hypotheses and create an unpolarized weight set
- Signal samples:
  - ggH: production for SM, CP-odd, Max. mixing and decays
  - qqH: decays only

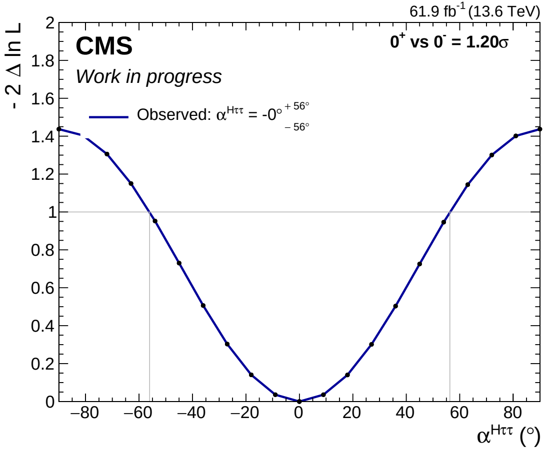


Sensitivity scans

- Model: linear combination of CP-even, CP-odd, and Max mixing templates reweighted with TauSpinner
- Datasets: 2022 + 2023
- Signal samples: ggH, VBF
- Uncertainties: stats. and lumi syst.
- Combination on the datacard level
- CP-odd hypothesis is excluded with $1.20\ \sigma$ significance

Run 2 (138 fb^{-1})	Run 3 (62 fb^{-1})	Run 3 scaled*
1.47	1.20	1.78 $(+21\ \%)$

*Scaled Run 3 result to the luminosity of 138 fb^{-1}



Summary

- CMS CP measurement of $H \to \tau\tau$ is ongoing
- The pipeline of the analysis is ready, including data-driven background estimation and the event classification via the BDT model
- First sensitivity scans using early Run 3 data show that the refined analysis strategy outperforms Run 2 analysis by $\sim 20\%$

Plans

- Develop a full uncertainty model
- Combine the results with $e\tau$ and $\tau\tau$ channels
- Target for the early Run 3 result publication - Moriond 2026

Summary

- CMS CP measurement of $H \to \tau\tau$ is ongoing
- The pipeline of the analysis is ready, including data-driven background estimation and the event classification via the BDT model
- First sensitivity scans using early Run 3 data show that the refined analysis strategy outperforms Run 2 analysis by $\sim 20\%$

Plans

- Develop a full uncertainty model
- Combine the results with $e\tau$ and $\tau\tau$ channels
- Target for the early Run 3 result publication - Moriond 2026

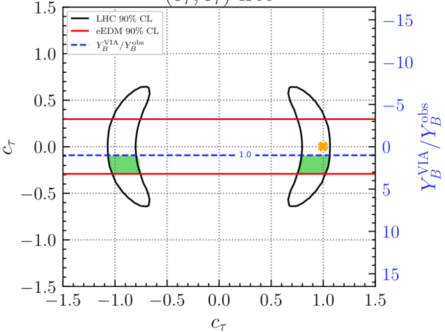
Thank you for the attention!

BACKUP

Theoretical and experimental constraints on $H\tau\tau$ couplings

Exp.+Th. constrains

$(c_\tau, \tilde{c}_\tau)$ free

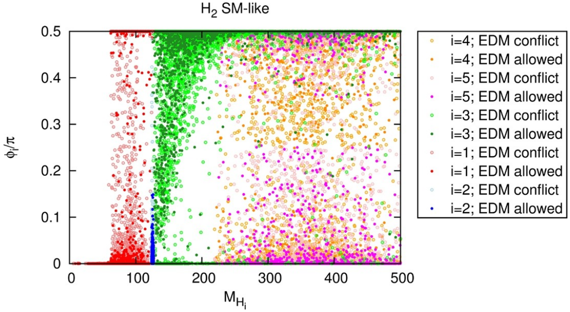


Eur. Phys. J. C 82, 604 (2022)

If uncertainty on $\alpha^{H\tau\tau} < 7^\circ$ we became sensitive to BSM

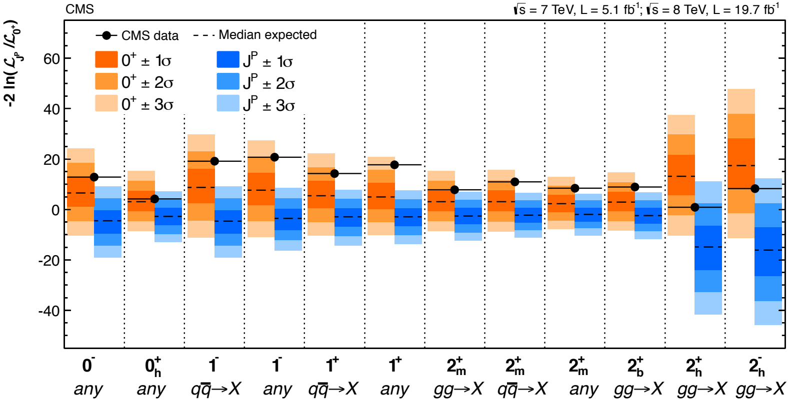
Nuclear Physics B Volume 901, December 2015, Pages 526-555

Exploring the CP-violating NMSSM: EDM constraints and phenomenology



The phase ϕ_i, which measures the CP violation in the $H_i\tau\tau$ coupling, as a function of the mass of the Higgs boson H_i.

Various J^P models tested against the SM Higgs boson hypothesis



Phys. Rev. D 89, 092007 (2014)

Motivation for E_{split} cut

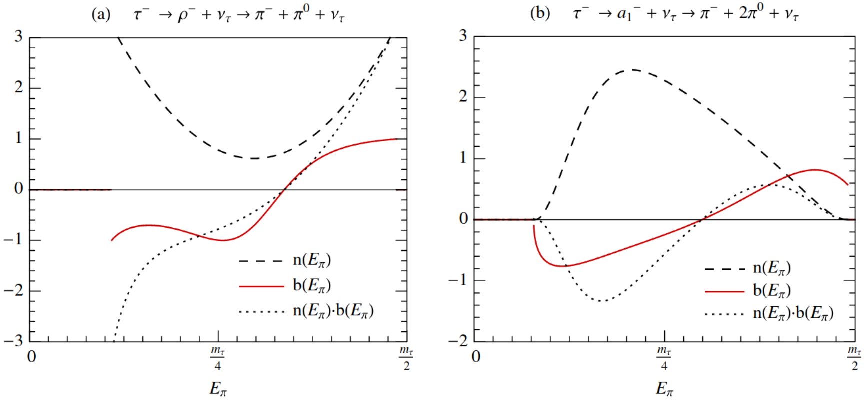
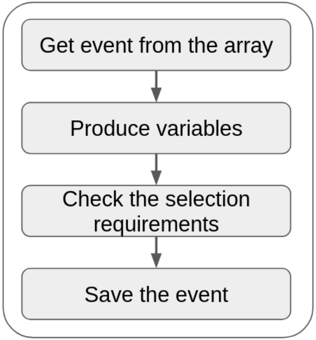


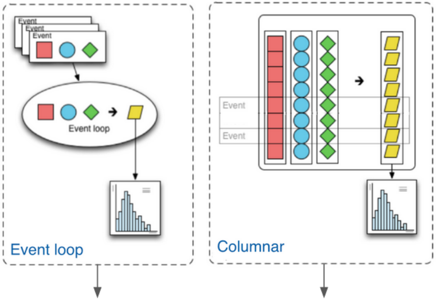
FIG. 4 (color online). Charged-pion spectral functions $n(E_\pi)$ and $b(E_\pi)$ for hadronic τ decays: (a) $\tau^- \to \rho^- \nu_\tau \to \pi^- \pi^0 \nu_\tau$; (b) $\tau^- \to a_1^- \nu_\tau \to \pi^- 2\pi^0 \nu_\tau$. The functions $n(E_\pi)$ and $n(E_\pi)b(E_\pi)$ are given in units of GeV^{-1}.

Phys. Rev. D 84, 116003 (2011)

Columnar analysis approach

Inside event loop…



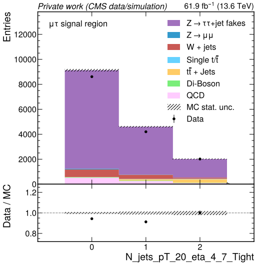


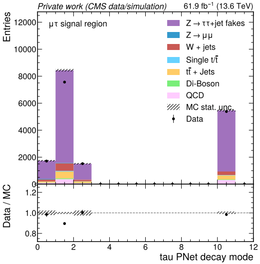
- Explicit ‘for’ loops
- Loads all variables
- High Memory load

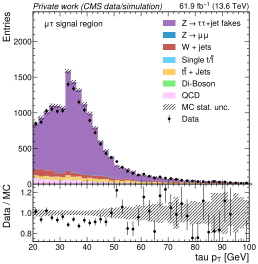
- Implicit loops
- Loads variables needed
- Resource-Friendly Code

Transfer Factor method: validation

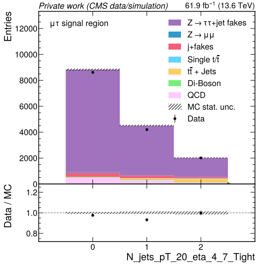
Data-driven QCD

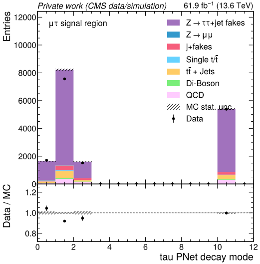


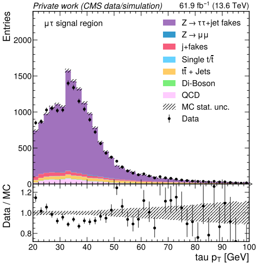




QCD and Wj via TF





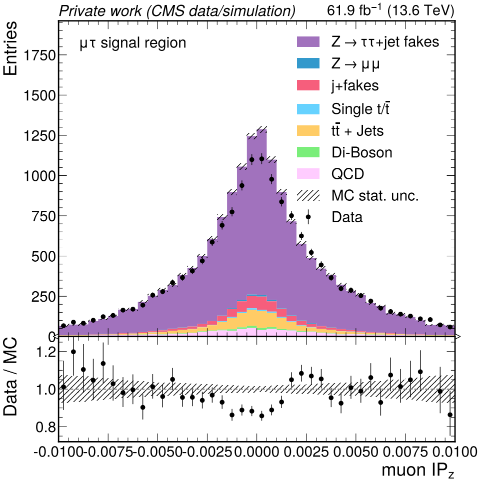


Impact parameter calibration

> Relation between MC and data is complex: it can not be corrected by commonly used methods

> Since we are using IP significance variable, we should not modify its rank order

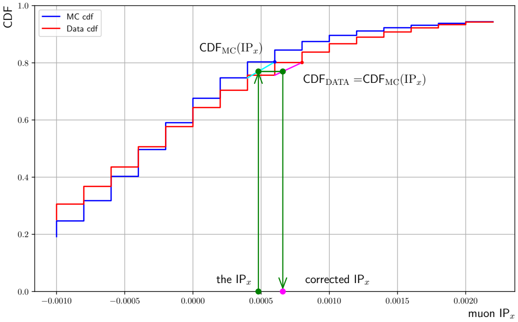
> Quantile mapping method is a good choice in this case



z-component of muon impact parameter

Impact parameter calibration

- Relation between MC and data is complex: it can not be corrected by commonly used methods
- Since we are using IP significance variable, we should not modify its rank order
  Quantile mapping method is a good choice in this case

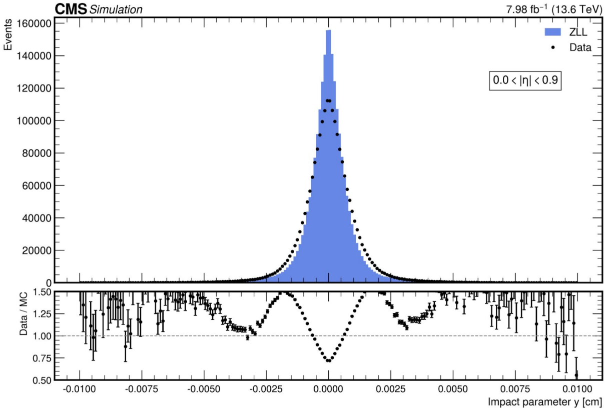


Quantile mapping

Impact parameter calibration: event selection

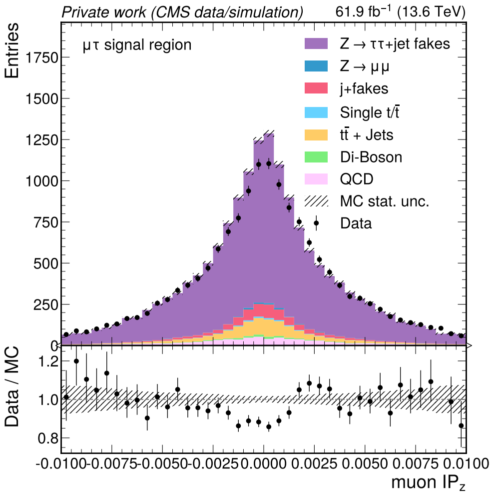
- $\mu\mu$ selections:
  - Single muon trigger
  - muon 1 & 2 rel. iso > 0.15
  - muon 1 pt > 26
- $|\eta|$ bins $\mu\mu : [0, 0.9, 1.2, 2.1, 2.4]$
- MC = ZLL
- Data = Data - ZTT - ZJ - W - TTT - TTJ - VVT - VVJ - QCD)

- Preselection and creation of histograms done by Klitos Savva from IC
- Original code developed by Alexei Raspereza

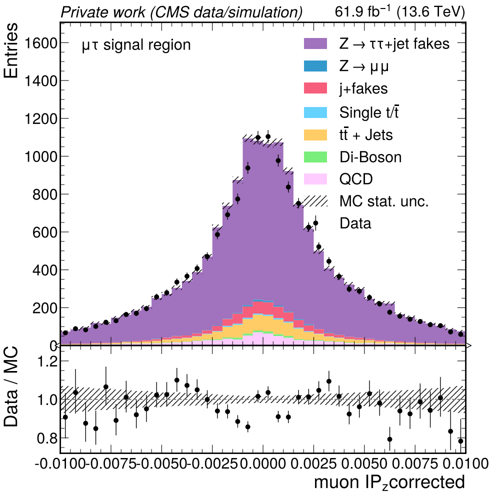


x-component of the impact parameter for $\mu\mu$ events of 2022preEE dataset (done by Klitos Savva)

Impact parameter calibration: results

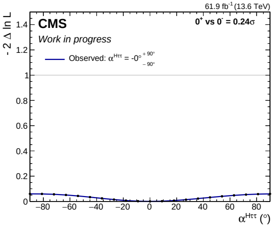


Raw distribution

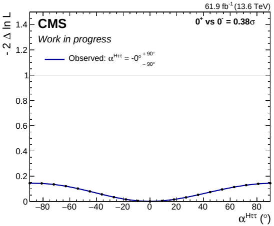


Applied quantile mapping

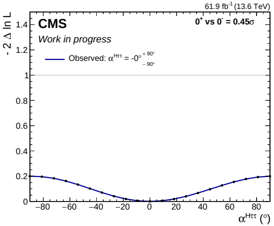
Sensitivity scans per channel



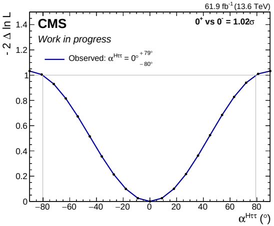
$\mu a_1(1\text{pr})$



$\mu\pi$



$\mu a_1(3\text{pr})$



$\mu\rho$

	Run2	Run3	Run3 lumi. scaled
$\tau_\mu\tau_h$	1.47	1.20	1.78 $(+21\ \%)$
$\mu\rho$	1.16	1.02	1.51 $(+30\ \%)$
$\mu\pi$	0.71	0.38	0.56 $(-22\ \%)$
$\mu a_1^{3\text{p}}$	0.51	0.45	0.58 $(+56\ \%)$
$\mu a_1^{1\text{p}}$	0.24	0.24	0.31 $(+29\ \%)$