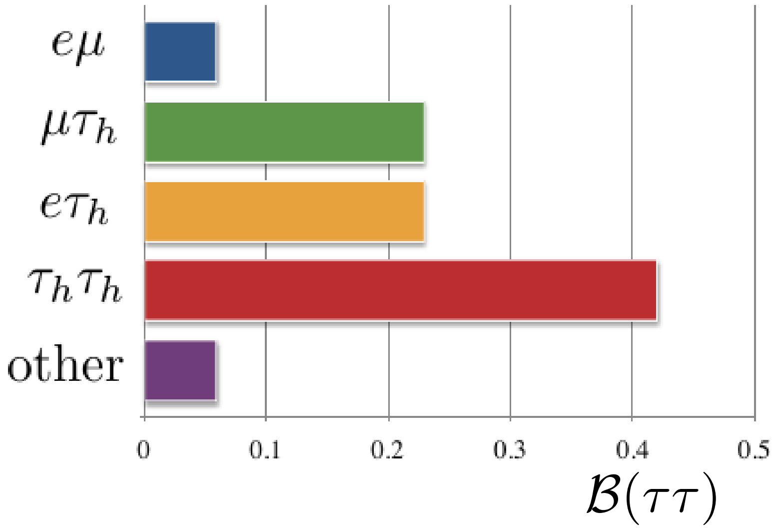
# H→ττ at CMS

*Alexei Raspereza*

**SFB B9 Meeting, 2017/10/16**

# Analysed Dataset and Decay Modes




- **Dataset :**

  **35.9 fb$^{-1}$ collected by CMS at c-o-m energy of 13 TeV**

- **four decay modes of tau pairs exploited (94% of final states)**

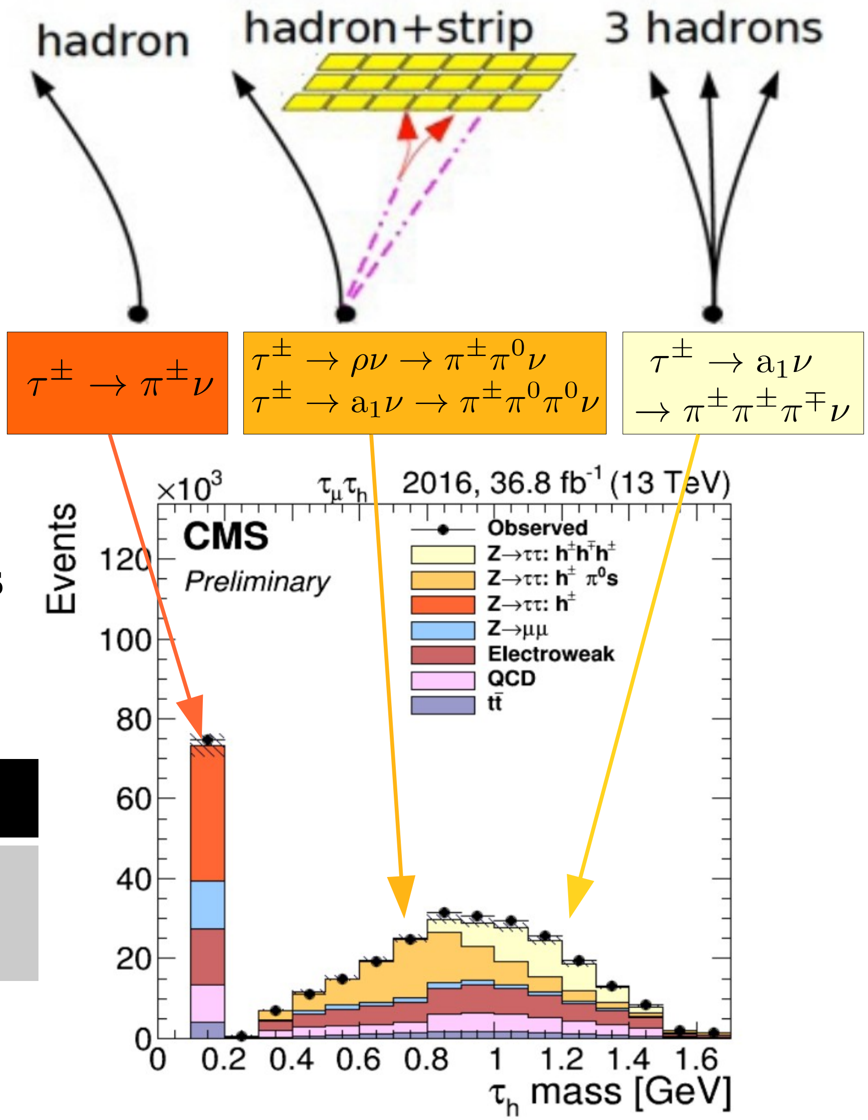
  $e\mu, \quad \mu\tau_h, \quad e\tau_h, \quad \tau_h\tau_h$

- **Final states with hadronically decaying tau leptons amount to 88% of all di-tau decays**
  - **efficient triggering and identification of $\tau_h$ in harsh pileup environment is essential**

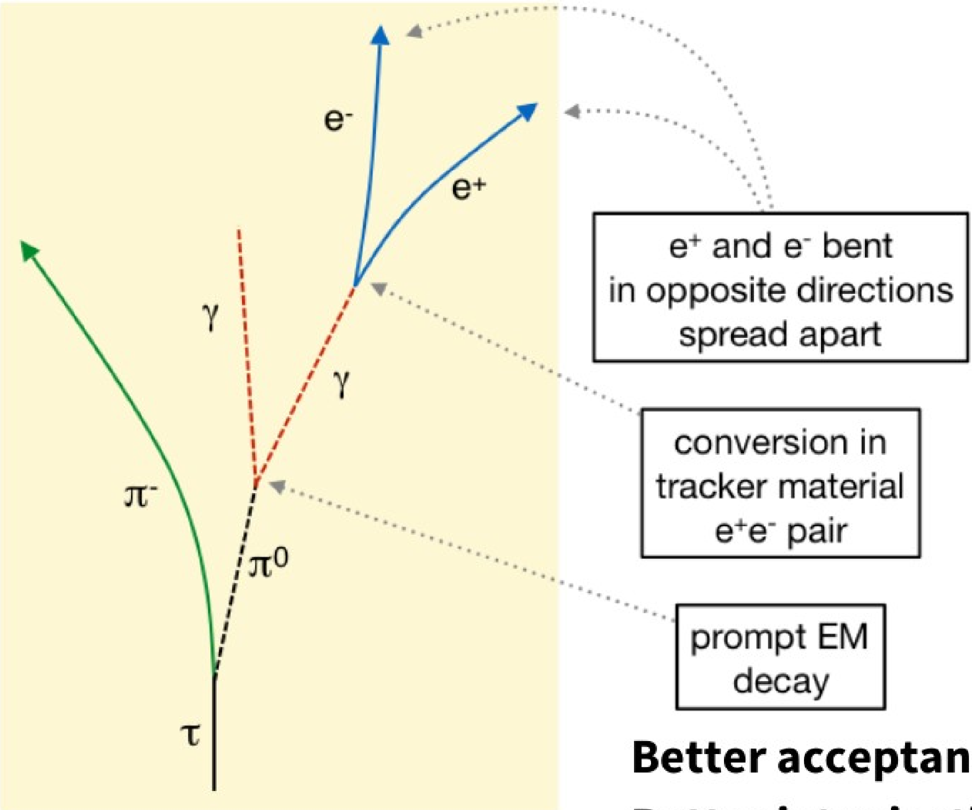
# $\tau_h$ Identification at CMS

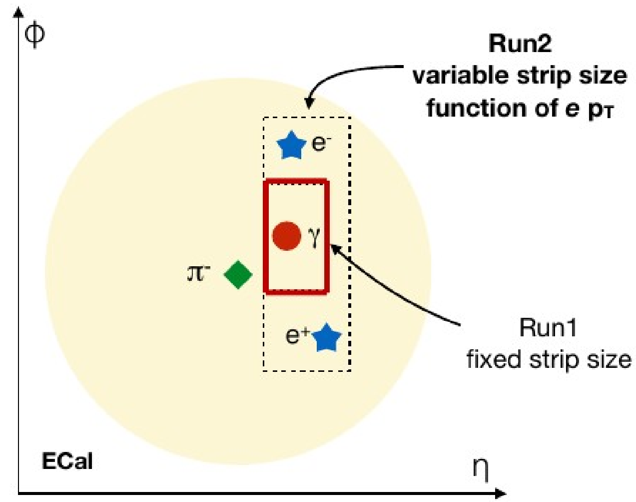
- seeded by anti-$k_T$ jets ($\Delta R_{cone}$=0.4)
- use as input particle-flow objects
- combine charged hadrons and strips of ECal clusters (photons, electrons)
- identify three main decay modes
- 1-prong+pi0s and 3-prong decays : require compatibility of visible tau mass with intermediate resonance (ρ, a1)

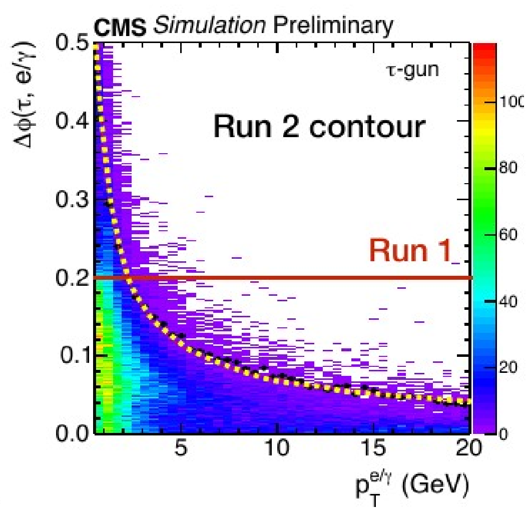
| 1 prong | 1 prong + pi0s | 3 prong |
|---|---|---|
| 12% | 26% (1 pi0), via ρ<br>11% (2 pi0s), via $a_1$ | 10%, via $a_1$<br>(+5% with pi0) |



# Run 2: Dynamic strip size



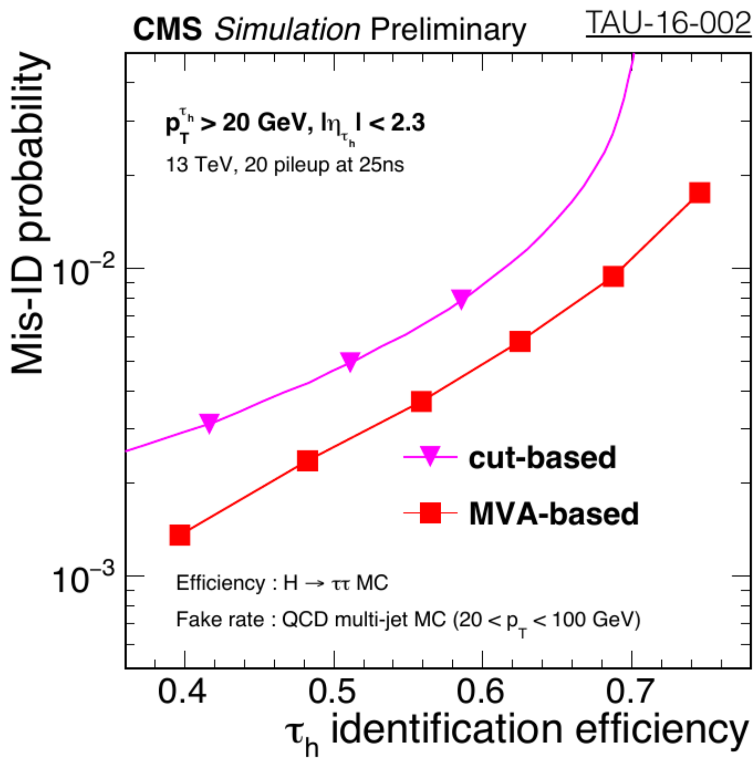



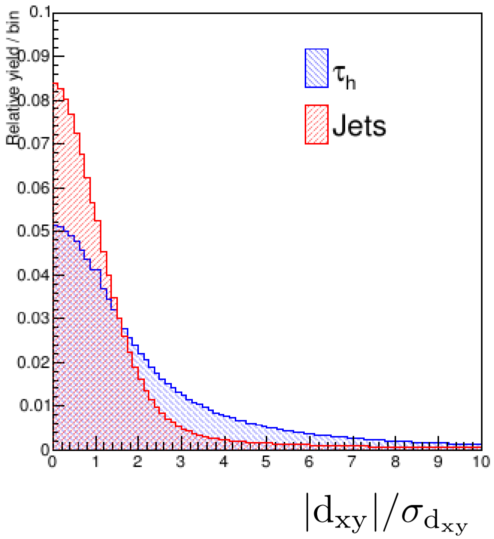

**Better acceptance at low $p_T$**

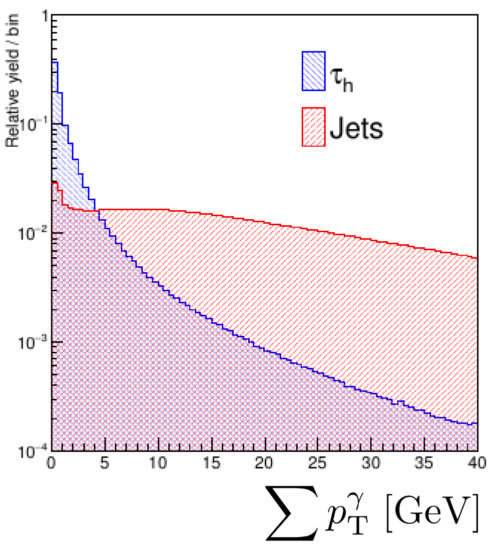
**Better jet rejection at high $p_T$**

# $\tau_h$ Identification at CMS

- **Run 1 approach : cut on isolation variables**
- **Run 2 approach : MVA discrimination of $\tau_h$ against hadronic jets**
  - **isolation variables ($p_T$ sums)**
  - **$\tau_h$ decay length information (track impact parameters, SV decay length)**
  - **multiplicity of particle-flow objects in signal/isolation cones**
  - **calorimeter cluster shapes**




- significant improvement compared to Run 1 approach
- $\tau_h$ efficiency of $\simeq$ 45% for mis-ID probability of $\simeq 2 \cdot 10^{-3}$

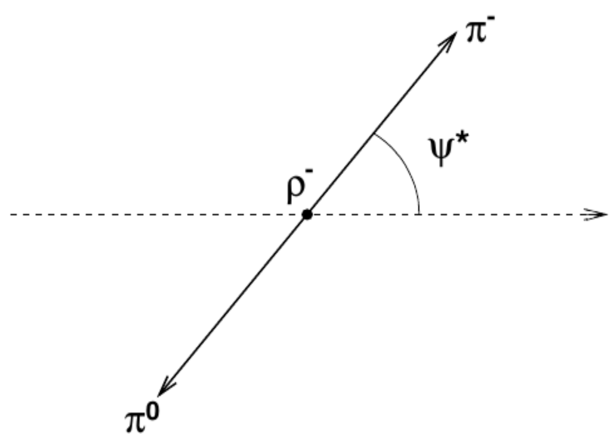
# Spin-off : Tau Polarization in Z → ττ Decays

- **Particle flow approach facilitates measurement of tau polarization**

**Exploited decay channel**

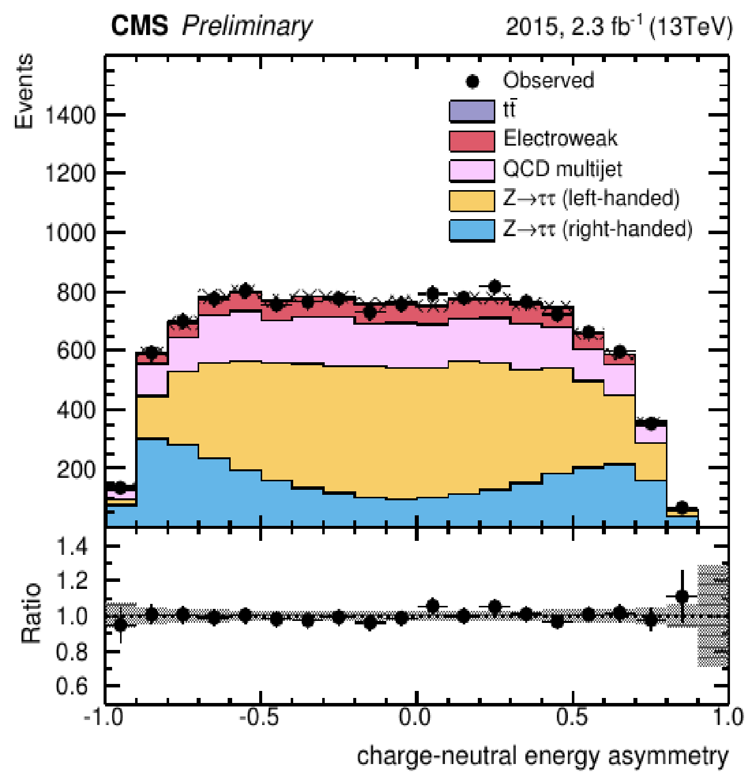
$$\tau^{\pm} \to \rho^{\pm}\nu_{\tau} \to (\pi^{\pm}\pi^{0})\nu_{\tau}$$

**energy asymmetry between $\pi^{\pm}$ and $\pi^{0}$ is spin sensitive variable**




$$\cos\psi^{*} = \frac{m_{\rho}}{\sqrt{m_{\rho}^{2} - 4m_{\pi}^{2}}} \frac{E_{\pi^{-}} - E_{\pi^{0}}}{|\vec{P}_{\pi^{-}} + \vec{P}_{\pi^{0}}|}$$

**measurement with $Z \to \tau_{\mu}\tau_{\rho}$ sample**

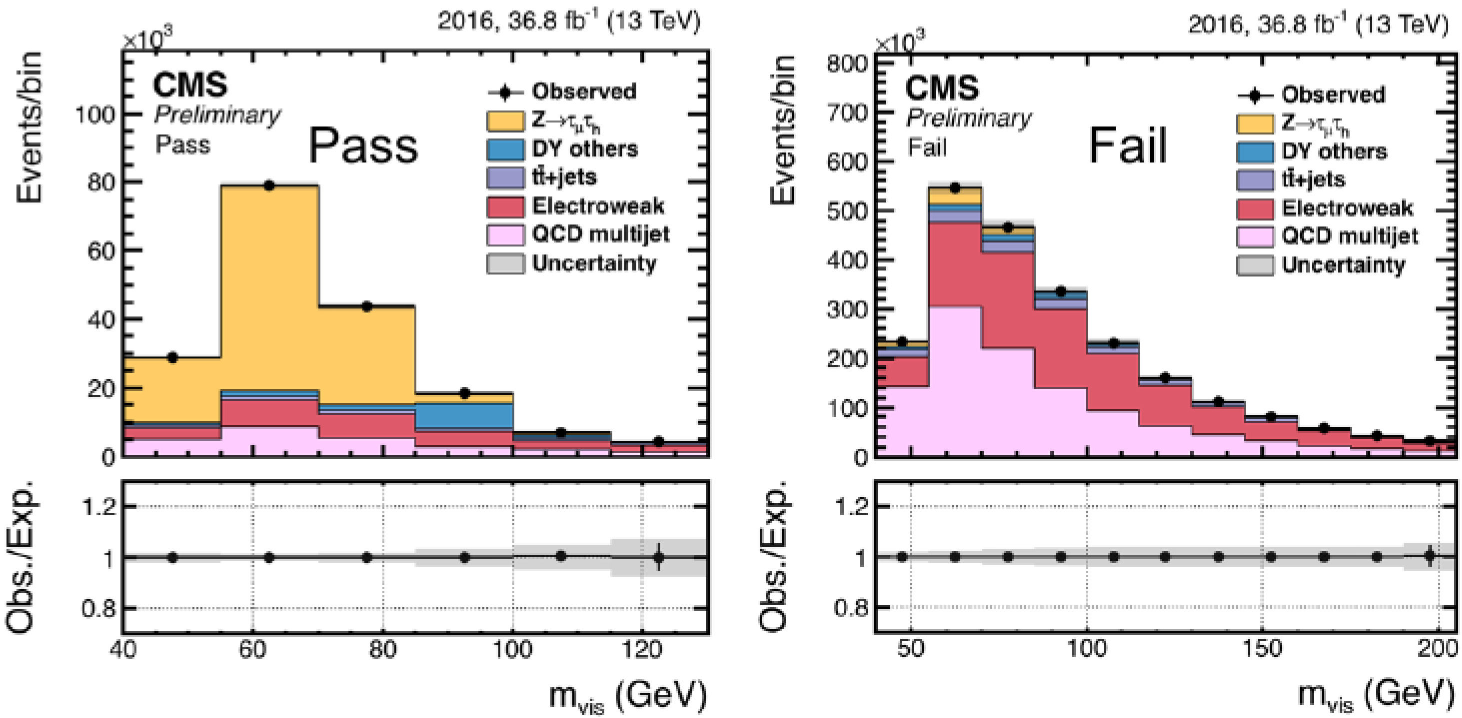



- **fit to observed data with superposition of signal and background templates**
- **Z → ττ contribution is represented by right- and left-handed components**

$$P_{\tau} = -36.6 \pm 3.7\ \%$$

# Measurement of $\tau_h$ ID efficiency

- exploits $Z \rightarrow \tau_\mu \tau_h$ standard candle
- based on likelihood fit of $\mu\tau_h$ invariant mass in "pass" and "fail" categories
- derived data/simulation SF are applied to simulated samples with genuine taus
- typical SF = 0.95 ± 0.05 (including tracking efficiency)

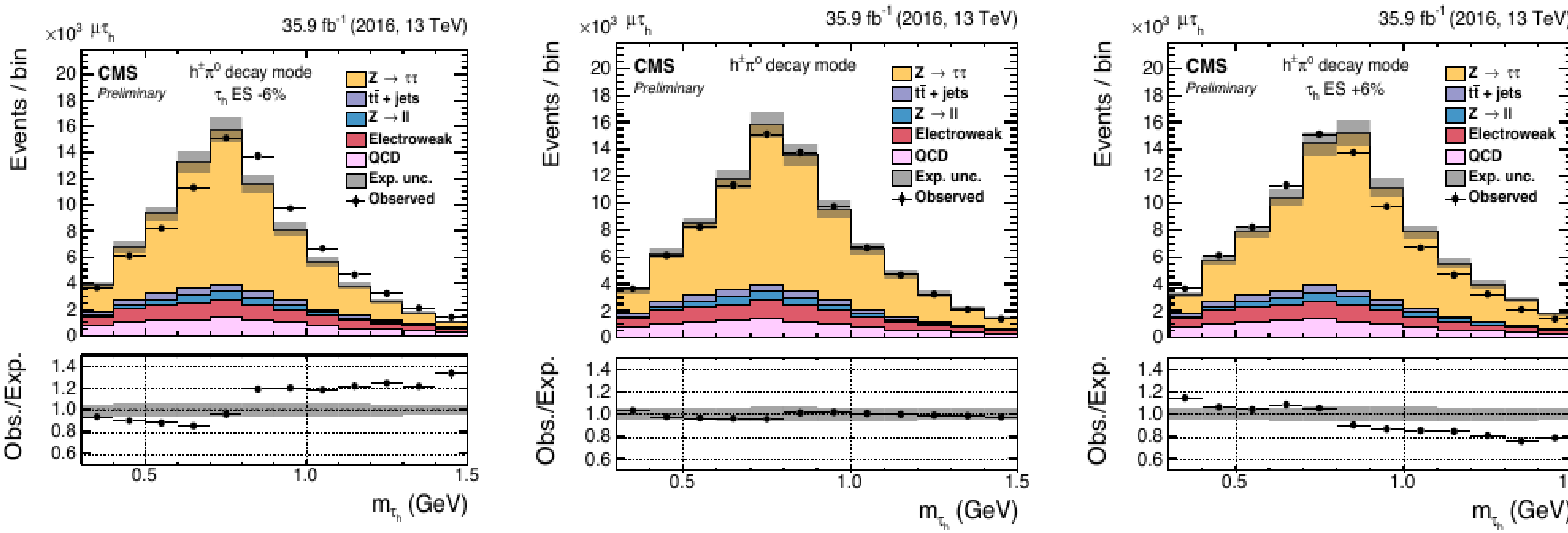


# Measurement of $\tau_h$ momentum scale

- **exploits $Z \to \tau_\mu \tau_h$ standard candle**
- **Individual measurement in each decay mode**

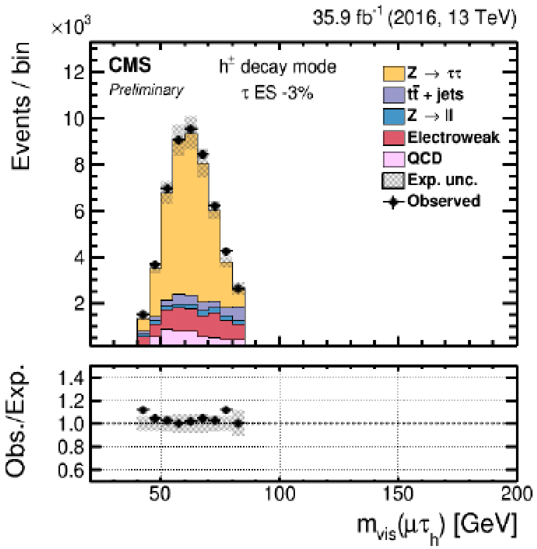
  $\tau^{\pm} \to \pi^{\pm} + \pi^{0\prime}s$ **and** $\tau^{\pm} \to \pi^{\pm}\pi^{\pm}\pi^{\mp}$

  **→ fit invariant mass of visible tau decay products**
- **Likelihood scan using continuously shifted templates**

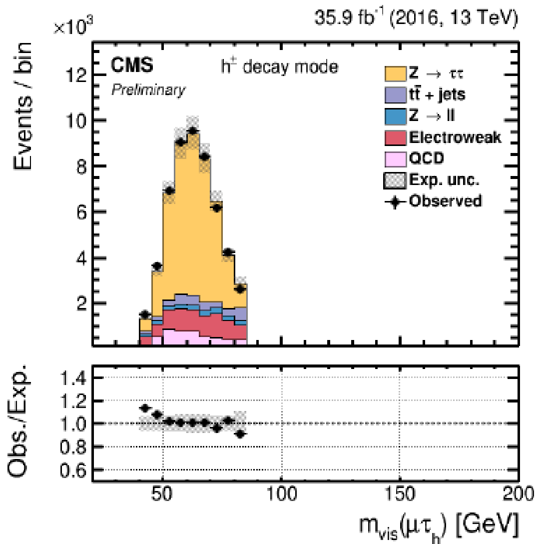


# Measurement of $\tau_h$ momentum scale

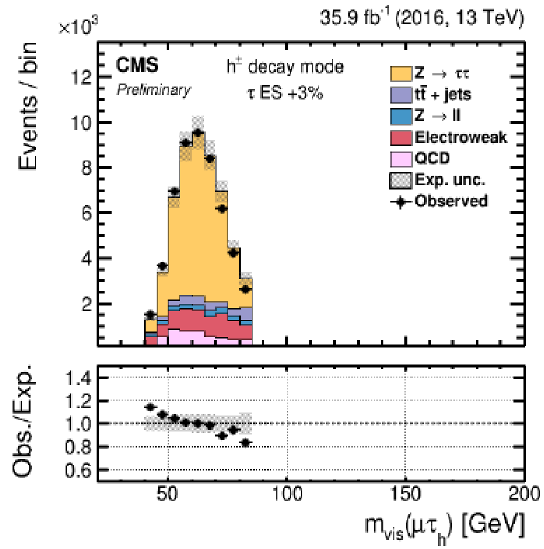
$\tau^{\pm} \rightarrow \pi^{\pm}$ : fit $\mu\tau_h$ mass ($m_{vis}$) in $Z \rightarrow \tau_\mu\tau_h$ decay




-3%




nominal



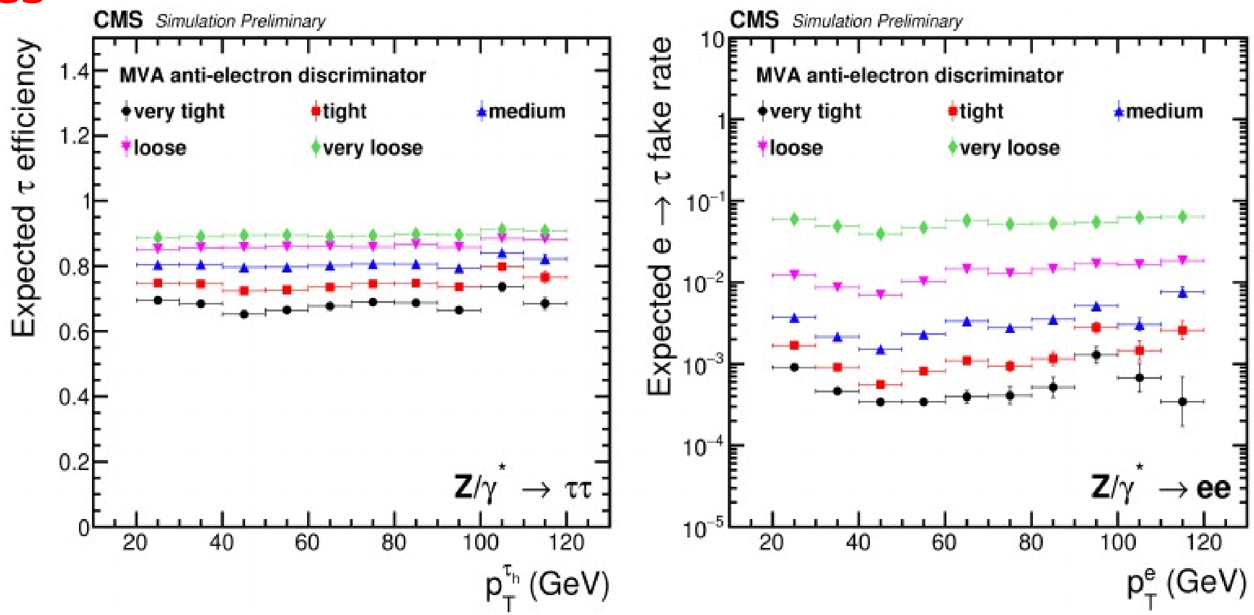

+3%

- **for other decay modes, measurements using $m_{vis}$ and $m_\tau$ give consistent results**

| decay mode | $\tau_h$ momentum scale |
|---|---|
| 1-prong | $-1.8 \pm 1.5\%$ |
| 1-prong+$\pi^0$'s | $+1.0 \pm 1.3\%$ |
| 3-prong | $+0.4 \pm 1.2\%$ |

# Suppression of lepton → $\tau_h$ fake rate

- **Z → ee/μμ background with e/μ → $\tau_h$ fakes is large lepton+τ channels**
- **Dedicated anti-lepton discriminators are developed**
  - **anti-electron : using tracking information and calorimeter cluster shapes**

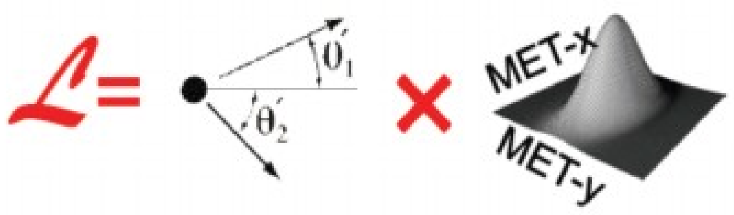


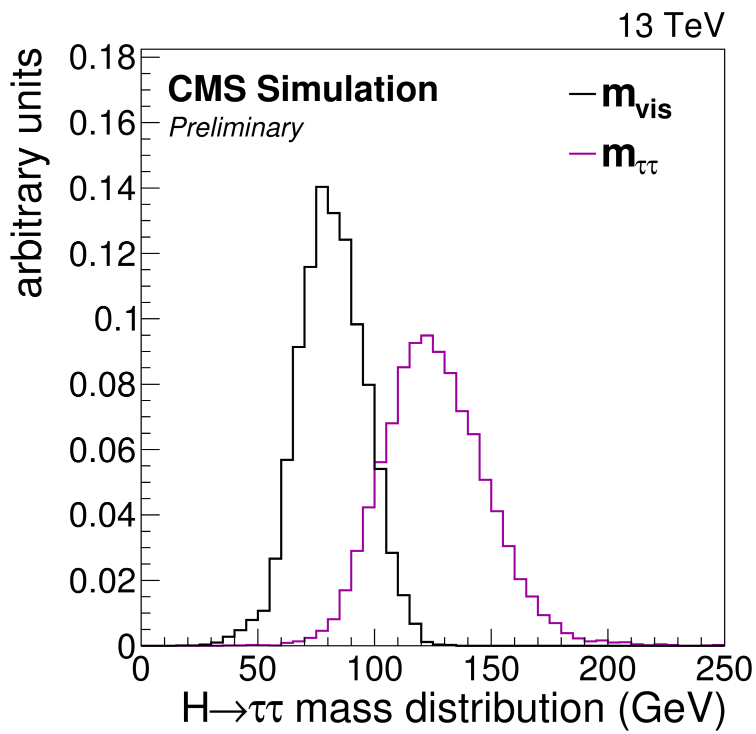

  - **anti-muon : based on track hit pattern in tracking system and muon stations (μ → $\tau_h$ fake rate ~ $10^{-4}$ for efficiency of 90%)**

**e/μ → $\tau_h$ rates and energy scales are measured in Z → ee/μμ samples**

# Di-tau Mass Reconstruction

- **Fully reconstructed di-tau mass is key variable discriminating signal against dominant Z → ττ background**
- **Reconstruction of** $m_{\tau\tau}$ **with dynamic likelihood algorithm**
- **Inputs :** $\vec{p}_{\tau_1},\ \vec{p}_{\tau_2},\ \vec{p}_{\mathrm{mis}},\ \mathrm{cov}(\vec{p}_{\mathrm{mis}})$
- **Estimate of** $m_{\tau\tau}$ **is obtained by maximizing likelihood combining**
  - **matrix elements of tau decays**
  - $\chi^2$ **of** $\vec{p}_{\mathrm{mis}}$ **measurement**



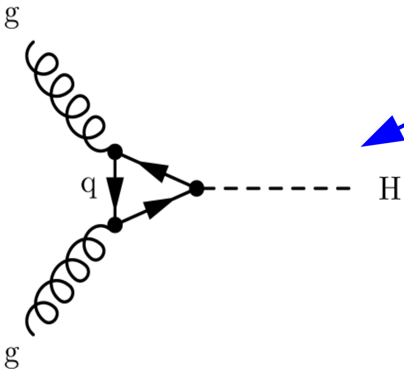




- **Better separation of H → ττ signal and Z → ττ background compared to the invariant mass of visible ττ decay products**
  - **the peak position is shifted to the nominal value of resonance mass**
  - **mass resolution : 15-20%**

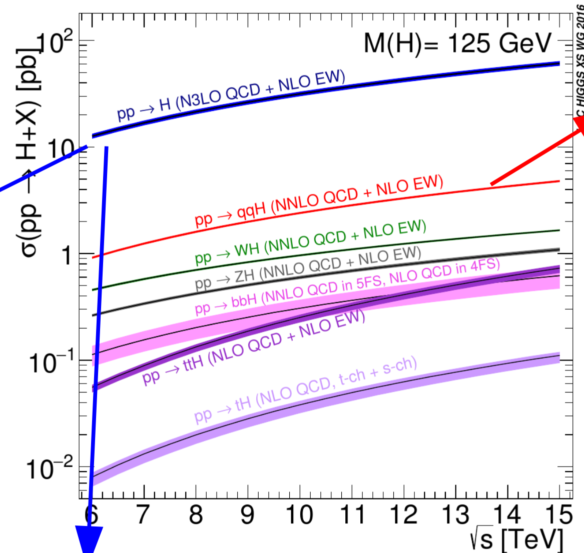
# Event categorization

- **Major production mechanisms are targeted with specific event categories**

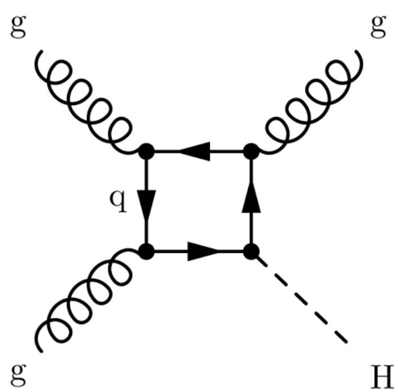
**0-jet : no jets with $p_T$ > 30 GeV, $|\eta|$<4.7**




- **largest signal yield**
- **but also largest bkgd**
- **calibration of tau reconstruction with $Z \to \tau\tau$ standard candle**
- **constrain uncertainties related to instrumental noise**

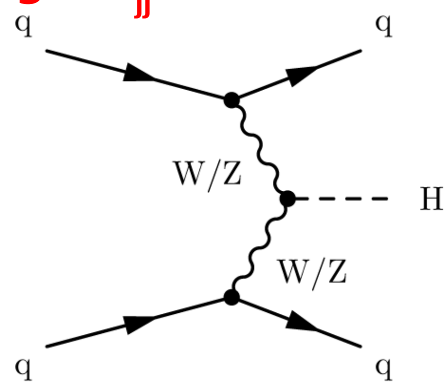



**boosted : high $p_T$(H) (not in 0jet or VBF)**




- **enhanced S/B compared to 0-jet category**
- **improved di-tau mass resolution**

**VBF : two jets with high $m_{jj}$**




- **Small signal yield**
- **but highest S/B ratio**

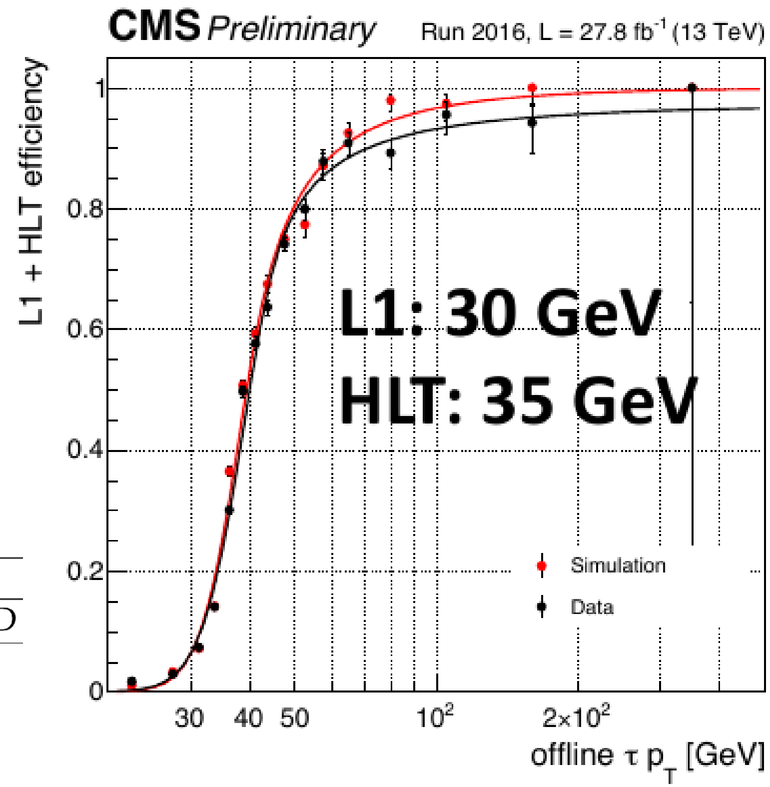
# Overview of search channels : $\tau_h \tau_h$

- **largest branching fraction of 42%**
- **Triggering is major a challenge**
  - **Improvements in L1 and HLT**

    **– higher readout granularity**

    **– sophisticated ID using dynamic clustering at hardware level**

    **→ same $p_T$ thresholds as in Run1 but sharper turn-on and higher efficiency**




| Channel | Trigger requirement | Lepton selection | | |
|---|---|---|---|---|
| | | $p_T$ (GeV) | $\eta$ | Isolation |
| $\tau_h\tau_h$ | $\tau_h(35)$ & $\tau_h(35)$ | $p_T^{\tau_h} > 50$ & $40$ | $\lvert\eta^{\tau_h}\rvert < 2.1$ | MVA $\tau_h$ ID |

- **Main backgrounds**
  - **Irreducible Z → ττ**

    **→ estimated with simulation with data-driven corrections (kinematics of Z and accompanying jets) derived from Z → μμ control region**
  - **QCD multijet background (suppressed by tight tau ID)**

    **→ measured exclusively from data (extrapolation from sideband with loose tau ID)**

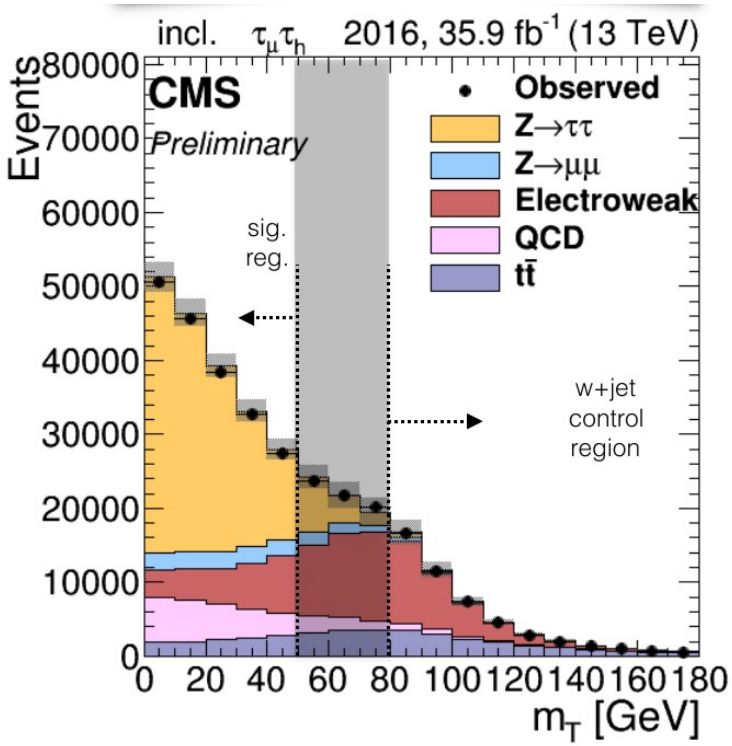
# Overview of search channels : $\mu+\tau_h$ and $e+\tau_h$

- **lower branching fractions (23% +23%) but cleaner signature**

  **single-lepton triggers with higher $p_T$ thresholds compared to Run1**

  **$\mu+\tau_h$ channel : combination of single-muon and muon+tau cross trigger → increase in acceptance**

| Final state | Trigger requirement | Lepton selection | | |
|---|---|---|---|---|
| | | $p_T$ (GeV) | $\eta$ | Isolation |
| $\mu\tau_h$ | $\mu(22)$ | $p_T^{\mu} > 23$<br>$p_T^{\tau_h} > 30$ | $\lvert\eta^{\mu}\rvert < 2.1$<br>$\lvert\eta^{\tau_h}\rvert < 2.3$ | $I^{\mu} < 0.15$<br>MVA $\tau_h$ ID |
| | $\mu(19)$ & $\tau_h(21)$ | $20 < p_T^{\mu} < 23$<br>$p_T^{\tau_h} > 30$ | $\lvert\eta^{\mu}\rvert < 2.1$<br>$\lvert\eta^{\tau_h}\rvert < 2.3$ | $I^{\mu} < 0.15$<br>MVA $\tau_h$ ID |
| $e\tau_h$ | $e(25)$ | $p_T^{e} > 26$<br>$p_T^{\tau_h} > 30$ | $\lvert\eta^{e}\rvert < 2.1$<br>$\lvert\eta^{\tau_h}\rvert < 2.3$ | $I^{e} < 0.1$<br>MVA $\tau_h$ ID |




- **Major backgrounds :**
  - **Irreducible Z → ττ (estimated as in $\tau_h\tau_h$ channel)**
  - **W+Jets with jet → $\tau_h$ fakes**

    **- suppressed by $m_T$ cut**

    **- estimated from high $m_T$ sideband**
  - **QCD multijets (estimated from sideband regions with same sign lepton pairs and relaxed lepton ID)**
  - **Z → ee/μμ with e/μ → $\tau_h$ fakes (estimated from simulation corrected for e/μ → $\tau_h$ misidentification rates measured in data)**

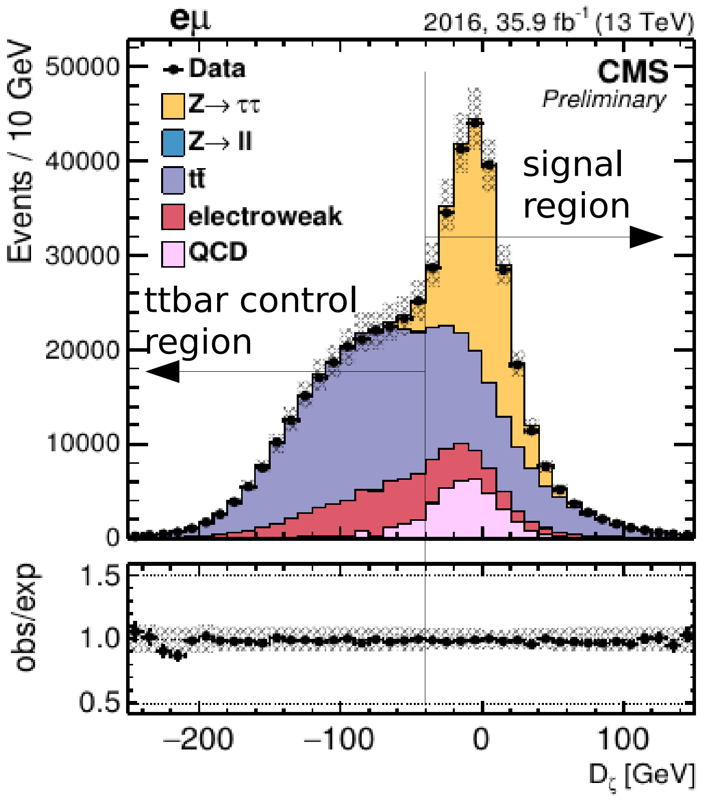
# Overview of search channels : e+μ

- **Cleanest signature but small branching fraction (6%) → lowest sensitivity**

  **inter-calibration of Drell-Yan background w/o $\tau_h$**

  **e+μ cross triggers with asymmetric thresholds are used → high signal acceptance**

| Channel | Trigger requirement | Lepton selection | | |
|---|---|---|---|---|
| | | $p_T$ (GeV) | $\eta$ | Isolation |
| $e\mu$ | $e(12)$ & $\mu(23)$ | $p_T^e > 13$<br>$p_T^\mu > 24$ | $\lvert\eta e\rvert < 2.5$<br>$\lvert\eta\mu\rvert < 2.4$ | $I^e < 0.15$<br>$I^\mu < 0.2$ |
| | $e(23)$ & $\mu(8)$ | $p_T^e > 24$<br>$p_T^\mu > 15$ | $\lvert\eta e\rvert < 2.5$<br>$\lvert\eta\mu\rvert < 2.4$ | $I^e < 0.15$<br>$I^\mu < 0.2$ |




- **Major backgrounds :**
  - **Irreducible Z → ττ (estimated as in other channels)**
  - **top-pairs**

    **- suppressed by b-tag veto and requiring alignment of missing $p_T$ with visible decay tau products (e, μ)**

    **- estimated from simulation corrected for top pT distribution**

    **- constrained in the sideband region**
  - **QCD multijets (estimated from sideband regions with same sign lepton pairs and relaxed lepton isolation)**

# Signal Extraction

- **Signal is extracted by simultaneous maximum-likelihood fit in 12 signal channels**

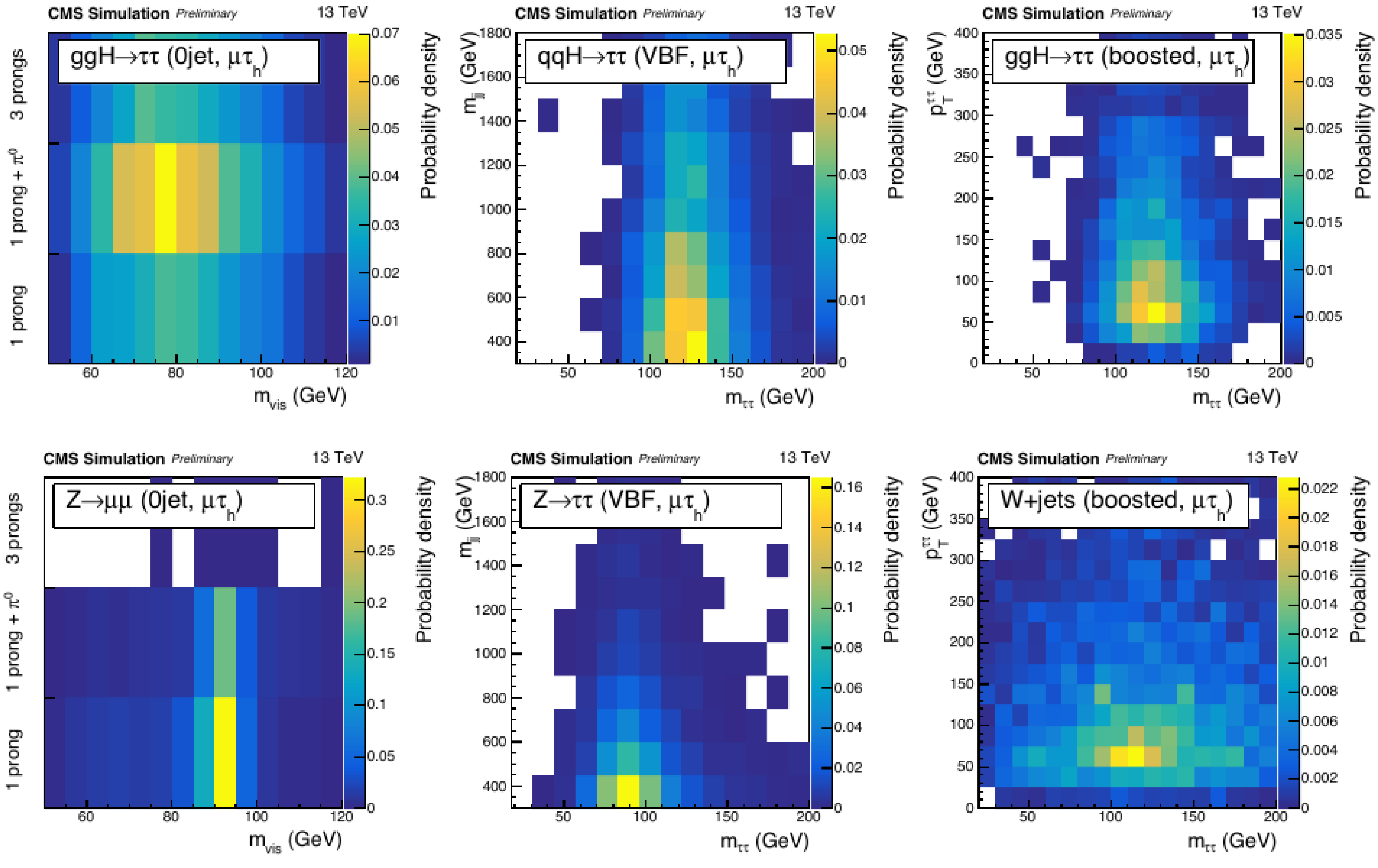
  **4 final states (eμ , e$\tau_h$ μ$\tau_h$ , $\tau_h\tau_h$) x 3 event category (0-jet, VBF, Boosted)**

  **and 12 control regions, constraining W+Jets, QCD and ttbar backgrounds**
- **In all channels but one fit is performed with 2D distributions**

| | 0-jet | VBF | Boosted |
|---|---|---|---|
| | | Selection | |
| $\tau_h\tau_h$ | No jet | $\geq$2 jets, $p_T^{\tau\tau} > 100\,\mathrm{GeV}$, $\Delta\eta_{jj} > 2.5$ | Others |
| $\mu\tau_h$ | No jet | $\geq$2 jets, $m_{jj} > 300\,\mathrm{GeV}$, $p_T^{\tau\tau} > 50\,\mathrm{GeV}$, $p_T^{\tau_h} > 40\,\mathrm{GeV}$ | Others |
| $e\tau_h$ | No jet | $\geq$2 jets, $m_{jj} > 300\,\mathrm{GeV}$, $p_T^{\tau\tau} > 50\,\mathrm{GeV}$ | Others |
| $e\mu$ | No jet | 2 jets, $m_{jj} > 300\,\mathrm{GeV}$ | Others |
| | | Observables used in fit | |
| $\tau_h\tau_h$ | $m_{\tau\tau}$ | $m_{jj}$, $m_{\tau\tau}$ | $p_T^{\tau\tau}$, $m_{\tau\tau}$ |
| $\mu\tau_h$ | $\tau_h$ decay mode, $m_{vis}$ | $m_{jj}$, $m_{\tau\tau}$ | $p_T^{\tau\tau}$, $m_{\tau\tau}$ |
| $e\tau_h$ | $\tau_h$ decay mode, $m_{vis}$ | $m_{jj}$, $m_{\tau\tau}$ | $p_T^{\tau\tau}$, $m_{\tau\tau}$ |
| $e\mu$ | $p_T^{\mu}$, $m_{vis}$ | $m_{jj}$, $m_{\tau\tau}$ | $p_T^{\tau\tau}$, $m_{\tau\tau}$ |

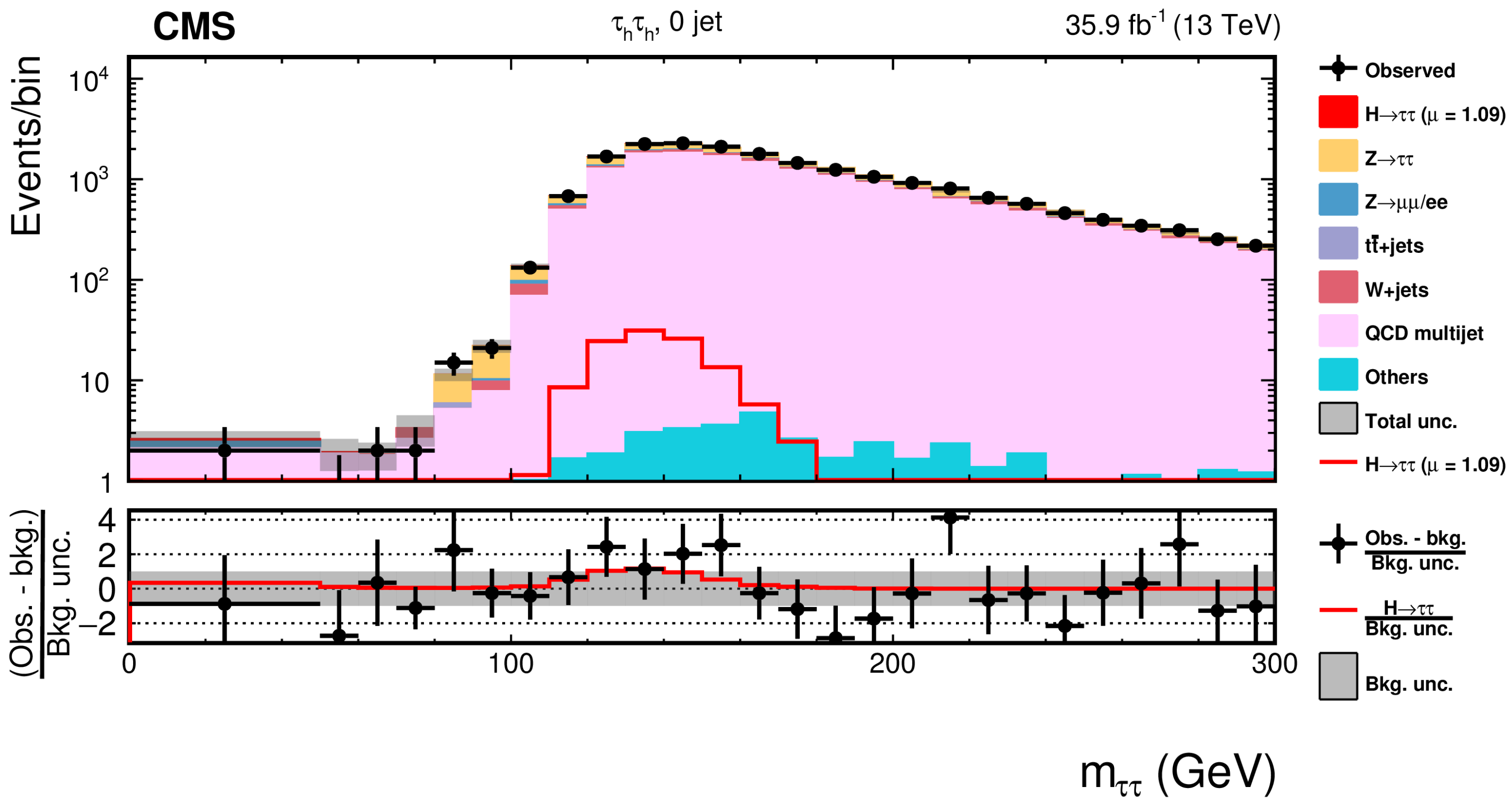
- **Di-tau mass or visible mass is used as 1st observable in 2D fit**
- **Choice of 2nd observable is motivated by**
  - **production signatures probed in a given category (boosted, VBF)**
  - **calibration of backgrounds specific to a given decay channel (0-jet)**

# 2D Distributions



# Postfit final discriminants : $\tau_h\tau_h$ channel

- **1D $m_{\tau\tau}$ distribution used to extract signal in 0-jet category**
- **No gain by adding second variable**



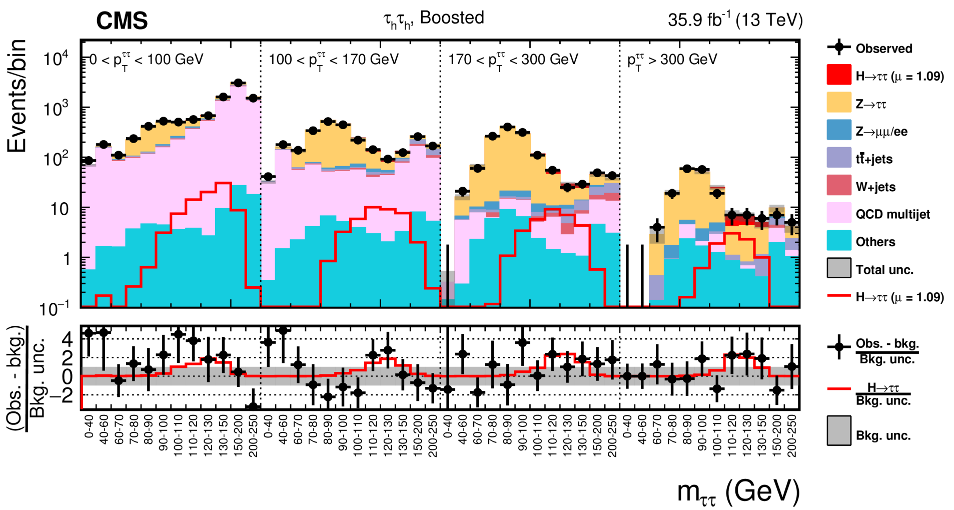

- **Moderate excess above background prediction at $m_{\tau\tau}$ ~ 125 GeV**

# Postfit final discriminants : $\tau_h\tau_h$ channel

- **Unrolled 2D distributions in Boosted and VBF categories**

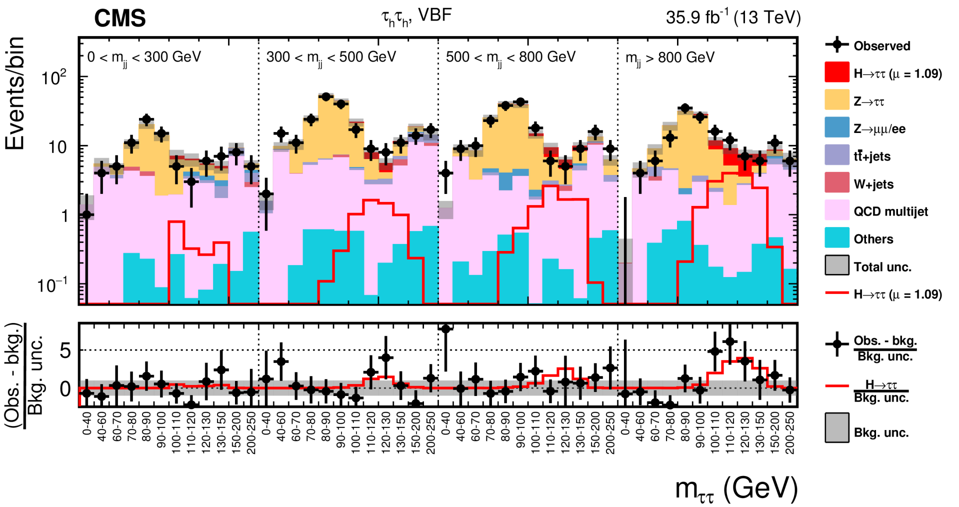
- **Boosted category**

$$(m_{\tau\tau}, p_T^{\tau\tau})$$




- **VBF category**

$$(m_{\tau\tau}, m_{jj})$$




**excellent S/B at high $m_{jj}$**

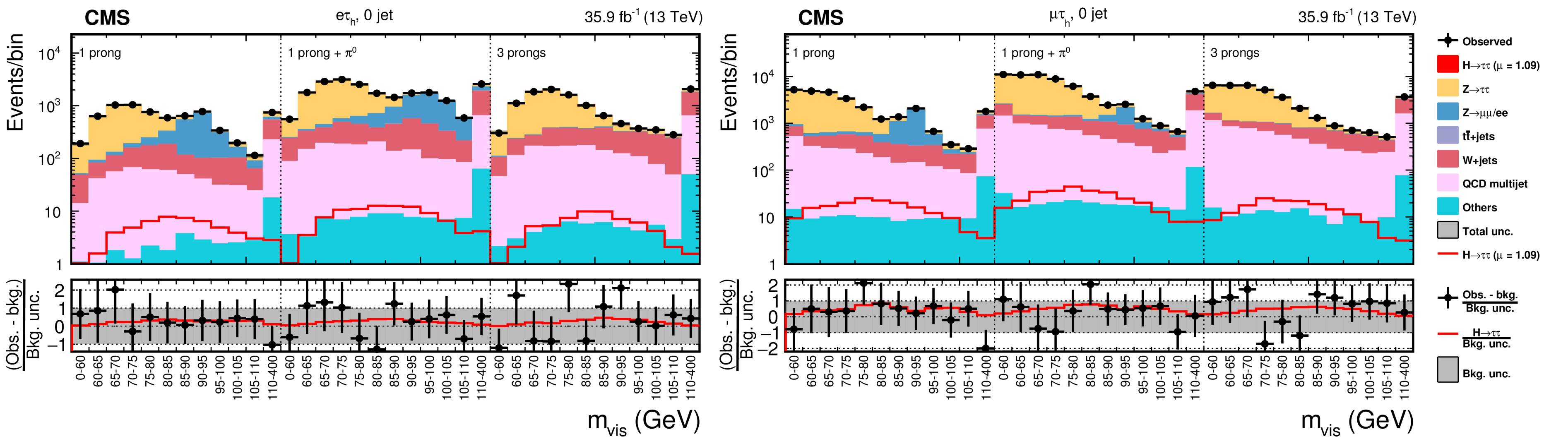
**(most sensitive channel)**

- Signal clearly developing in bins with high S/B ratio

# Postfit final discriminants : $\ell$+$\tau_h$ channels

- **the 2D distribution ($m_{vis}$ vs. $\tau_h$ decay mode) used to extract signal in 0-jet category**

  **→ better separation between signal and Z → ee/μμ background**




- **Low sensitivity to signal**
- **But category facilitates calibration of Drell-Yan background and constrains instrumental uncertainties**

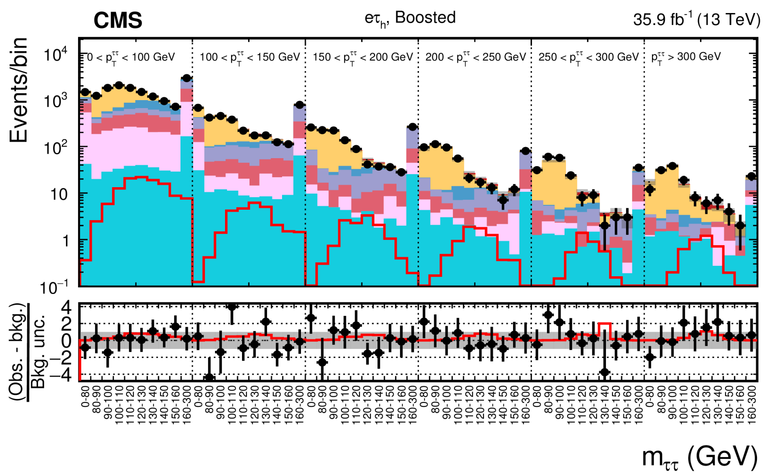
  **→ $\tau_h$ ID efficiency and momentum scale**

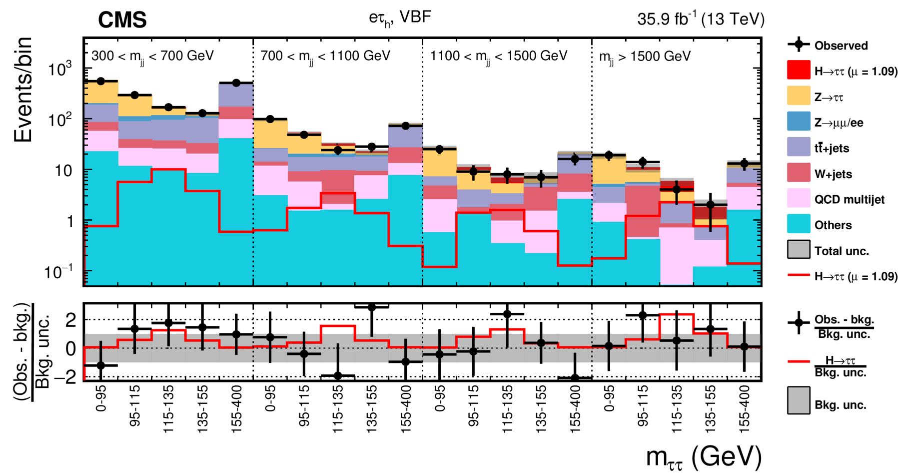
  **→ e/μ → $\tau_h$ fake rate and momentum scale**

# Postfit final discriminants : $\ell$+$\tau_h$ channels

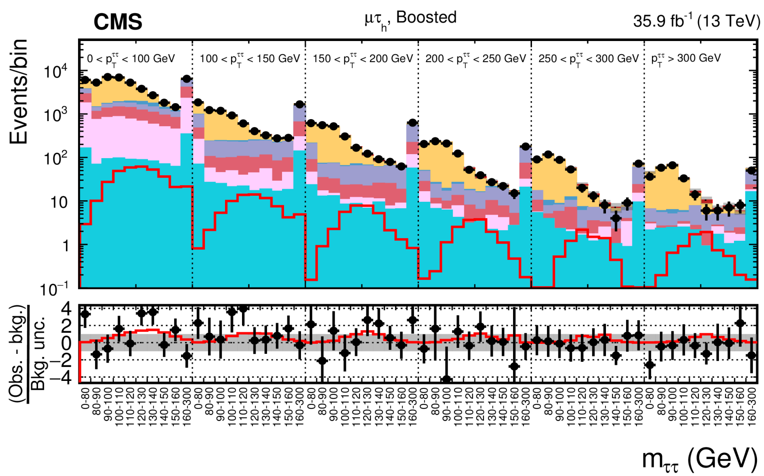
- **Unrolled 2D distributions**

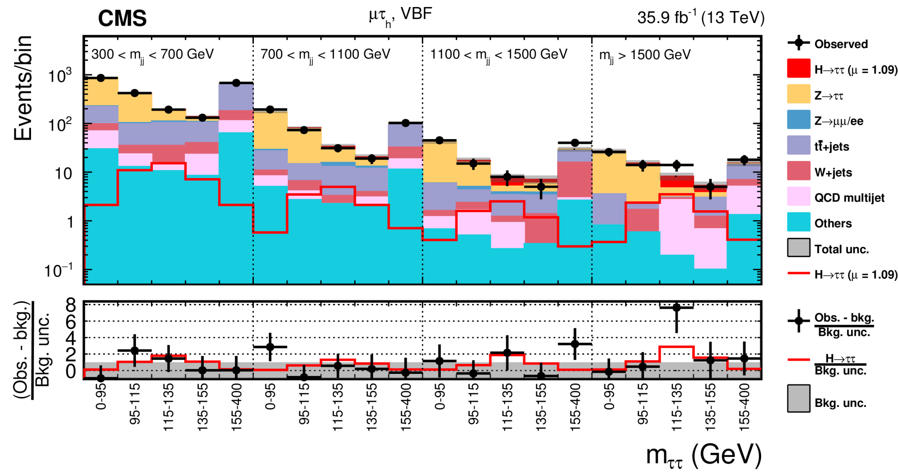
$\text{Boosted} : (m_{\tau\tau}, p_T^{\tau\tau})$ $\text{VBF} : (m_{\tau\tau}, m_{jj})$




$e\tau_h$



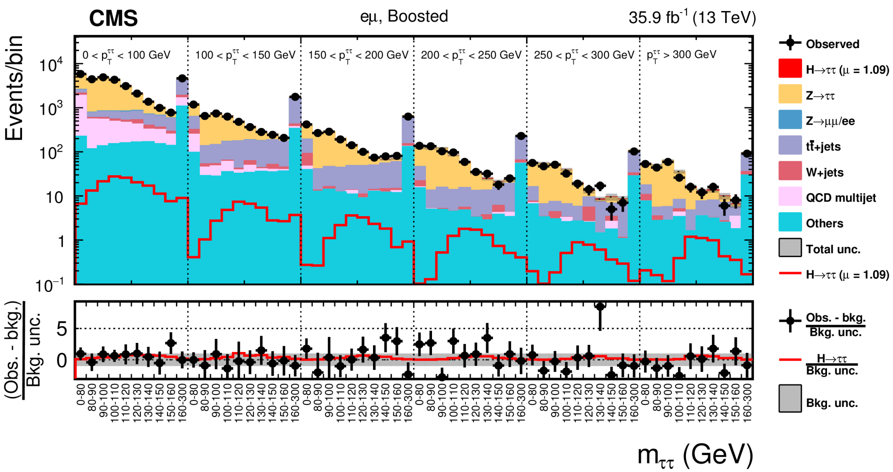




$\mu\tau_h$

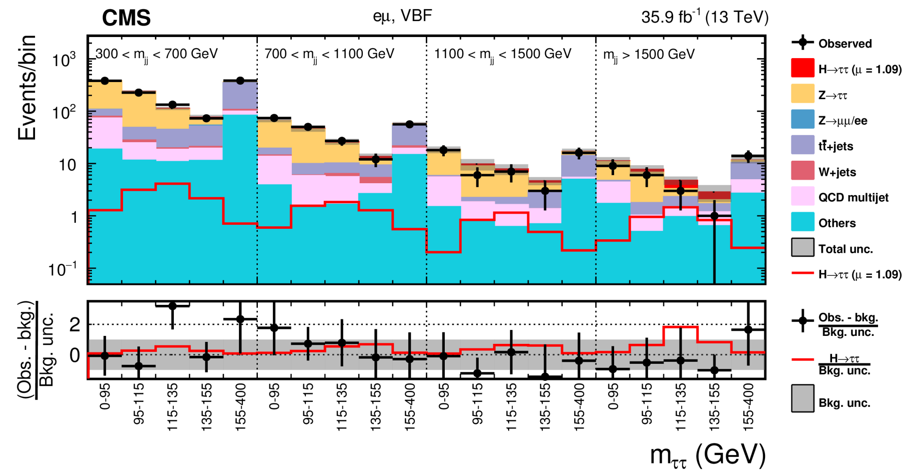
# Postfit final discriminants : e+μ channel

- **Unrolled 2D distributions**

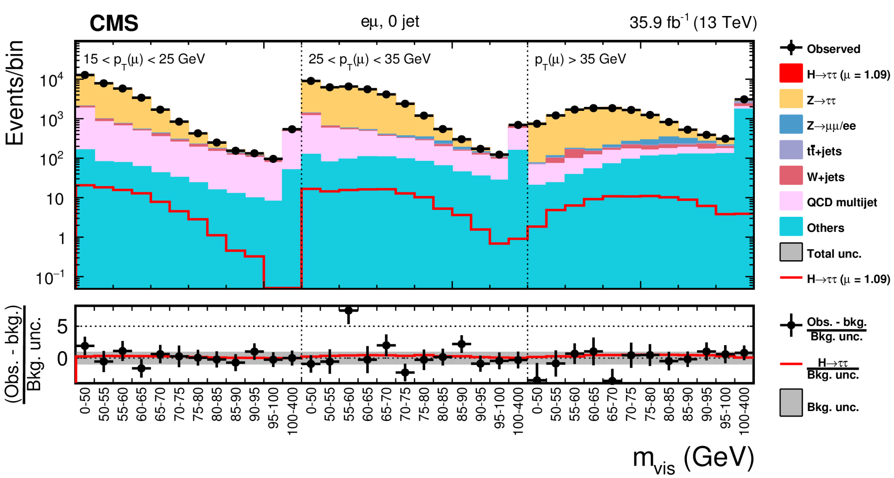
Boosted : $(m_{\tau\tau}, p_T^{\tau\tau})$




VBF : $(m_{\tau\tau}, m_{jj})$




0-jet : $(m_{\text{vis}}, p_T^{\mu})$




e+μ channel has lowest sensitivity
- smallest yield
- poorer mass resolution

Nonetheless, non-negligible contribution to overall sensitivity

# Uncertainty model

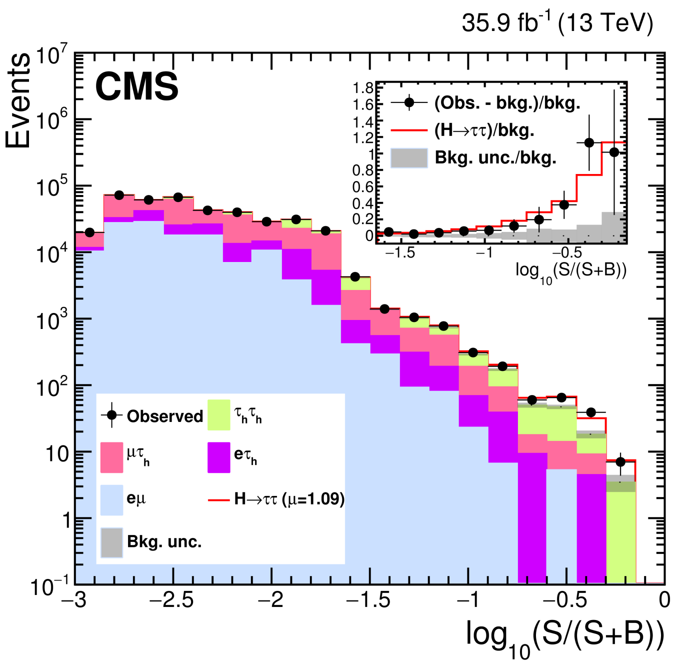
- Refined uncertainty model
  - split jet energy uncertainties into 27 various sources
  - fine-grane instrumental uncertainties ($\tau_h$ Id, momentum scale, fake rate)
- Background yields and shape constrained in a dedicated control regions
- instrumental corrections constrained in 0-jet category and propagated to VBF and Boosted categories
- Most of uncertainties are constrained in global fit
- Largest impact on precision comes from
  → instrumental uncertainties
  → theoretical shape uncertainties
  → limited MC statistics

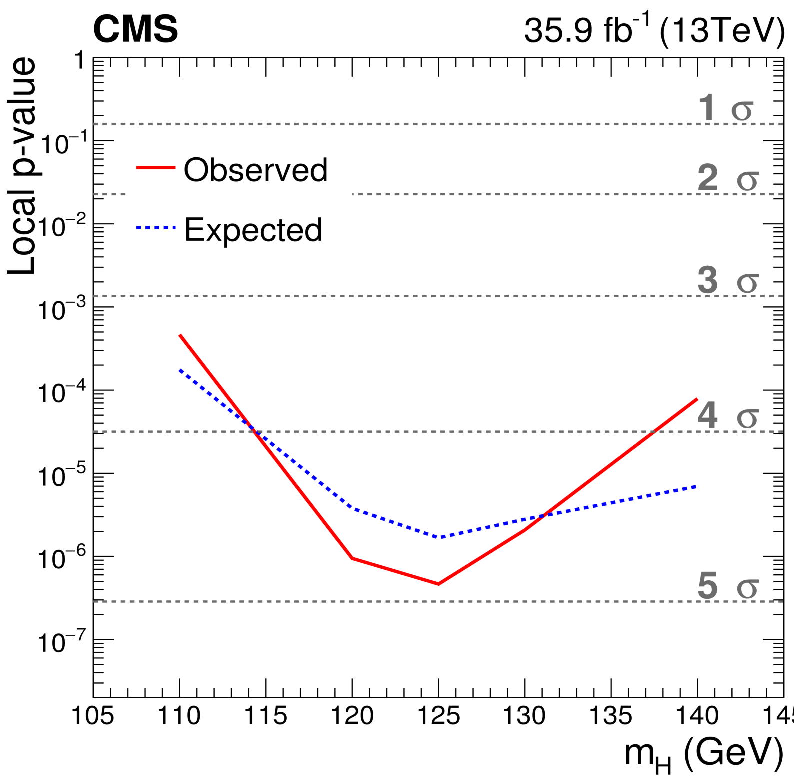
| Source of uncertainty | Prefit | Postfit (%) |
|---|---|---|
| $\tau_h$ energy scale | 1.2% in energy scale | 0.2–0.3 |
| e energy scale | 1–2.5% in energy scale | 0.2–0.5 |
| e misidentified as $\tau_h$ energy scale | 3% in energy scale | 0.6–0.8 |
| $\mu$ misidentified as $\tau_h$ energy scale | 1.5% in energy scale | 0.3–1.0 |
| Jet energy scale | Dependent upon $p_T$ and $\eta$ | — |
| $\vec{p}_T^{miss}$ energy scale | Dependent upon $p_T$ and $\eta$ | — |
| $\tau_h$ ID & isolation | 5% per $\tau_h$ | 3.5 |
| $\tau_h$ trigger | 5% per $\tau_h$ | 3 |
| $\tau_h$ reconstruction per decay mode | 3% migration between decay modes | 2 |
| e ID & isolation & trigger | 2% | — |
| $\mu$ ID & isolation & trigger | 2% | — |
| e misidentified as $\tau_h$ rate | 12% | 5 |
| $\mu$ misidentified as $\tau_h$ rate | 25% | 3–8 |
| Jet misidentified as $\tau_h$ rate | 20% per 100 GeV $\tau_h$ $p_T$ | 15 |
| $Z \to \tau\tau/\ell\ell$ estimation | Normalization: 7–15% | 3–15 |
| | Uncertainty in $m_{\ell\ell/\tau\tau}$, $p_T(\ell\ell/\tau\tau)$, and $m_{jj}$ corrections | — |
| W + jets estimation | Normalization (e$\mu$, $\tau_h\tau_h$): 4–20% | — |
| | Unc. from CR (e$\tau_h$, $\mu\tau_h$): $\simeq$5–15 | — |
| | Extrap. from high-$m_T$ CR (e$\tau_h$, $\mu\tau_h$): 5–10% | — |
| QCD multijet estimation | Normalization (e$\mu$): 10–20% | 5–20% |
| | Unc. from CR (e$\tau_h$, $\tau_h\tau_h$, $\mu\tau_h$): $\simeq$5–15% | — |
| | Extrap. from anti-iso. CR (e$\tau_h$, $\mu\tau_h$): 20% | 7–10 |
| | Extrap. from anti-iso. CR ($\tau_h\tau_h$): 3–15% | 3–10 |
| Diboson normalization | 5% | — |
| Single top quark normalization | 5% | — |
| $t\bar{t}$ estimation | Normalization from CR: $\simeq$5% | — |
| | Uncertainty on top quark $p_T$ reweighting | — |
| Integrated luminosity | 2.5% | — |
| b-tagged jet rejection (e$\mu$) | 3.5–5.0% | — |
| Limited number of events | Statistical uncertainty in individual bins | — |
| Signal theoretical uncertainty | Up to 20% | — |

# Observation of H → ττ Decays

- **Distribution of event yield in the analysis bins ordered by S/(S+B)**

→ **clearly visible excess in data w.r.t. background-only expectation**



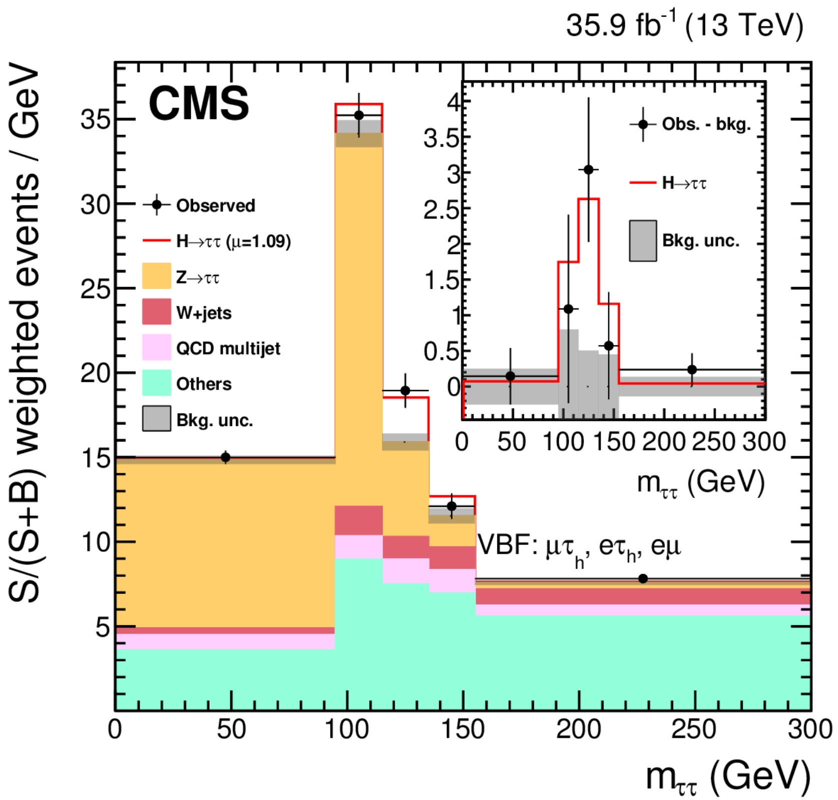



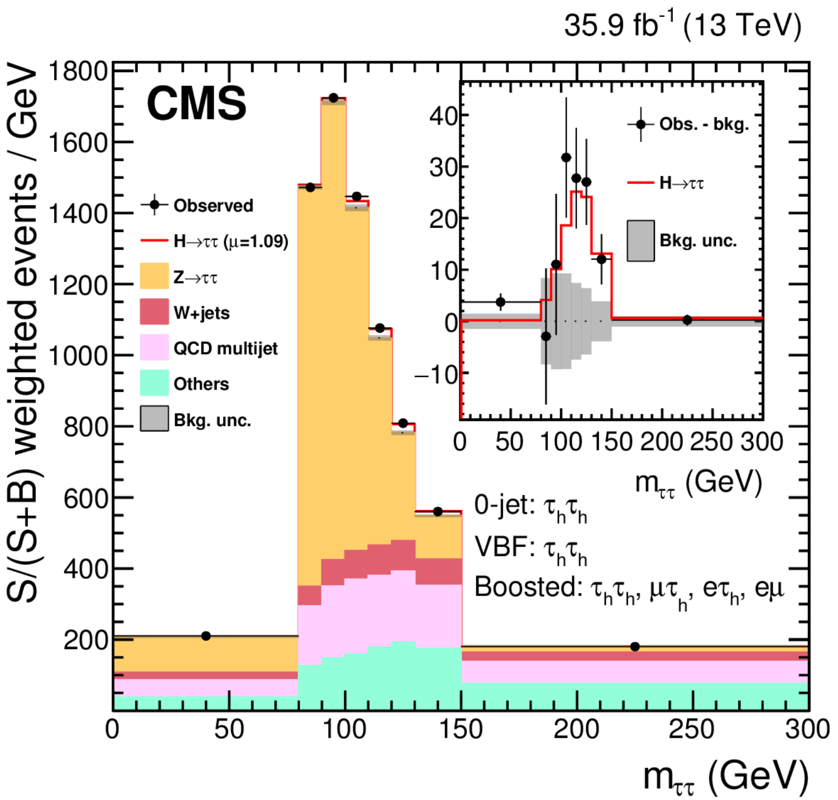

- **obs. (exp.) significance at $m_H$ = 125 GeV**
- $4.9\sigma \;\; (4.7\sigma)$ **with Run II data only**

- **Combination with Run I CMS data yields** $5.9\sigma \;\; (5.9\sigma)$

- **first observation of Yukawa coupling in single fermionic decay channel at CMS**

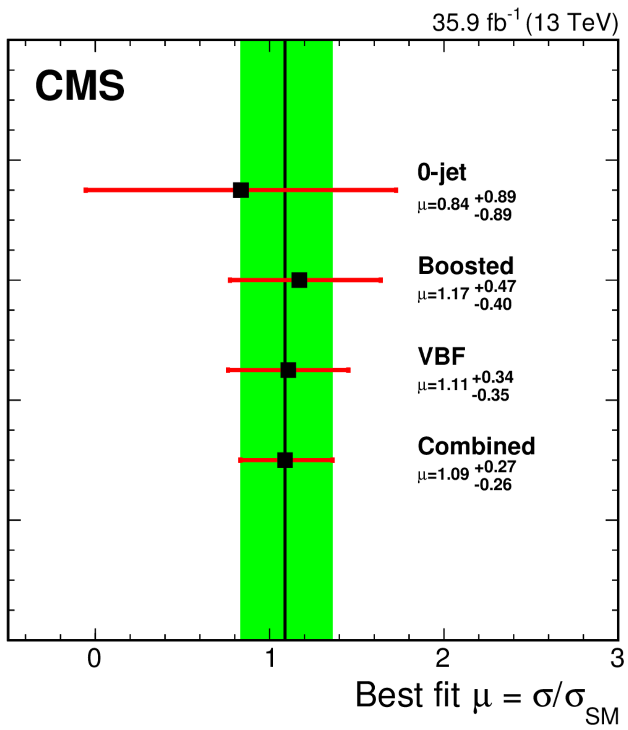
# Visualization of signal : $m_{\tau\tau}$ distribution

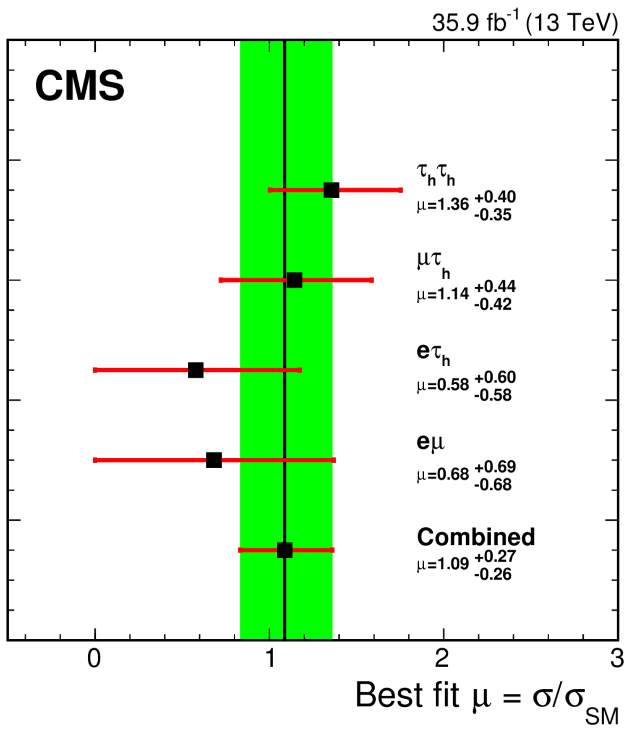



- **Events are weighted by S/(S+B) in bins of second variable of 2D distributions → unbiased mass spectrum**
- **Signal is clearly visible in the distribution of physical observable $m_{\tau\tau}$**

# Measurement of the signal strength




- Measurements are consistent across channels

- Highest sensitivity comes from
  - decay side : $\tau_h\tau_h$
  - production side : VBF

# Measurement of couplings

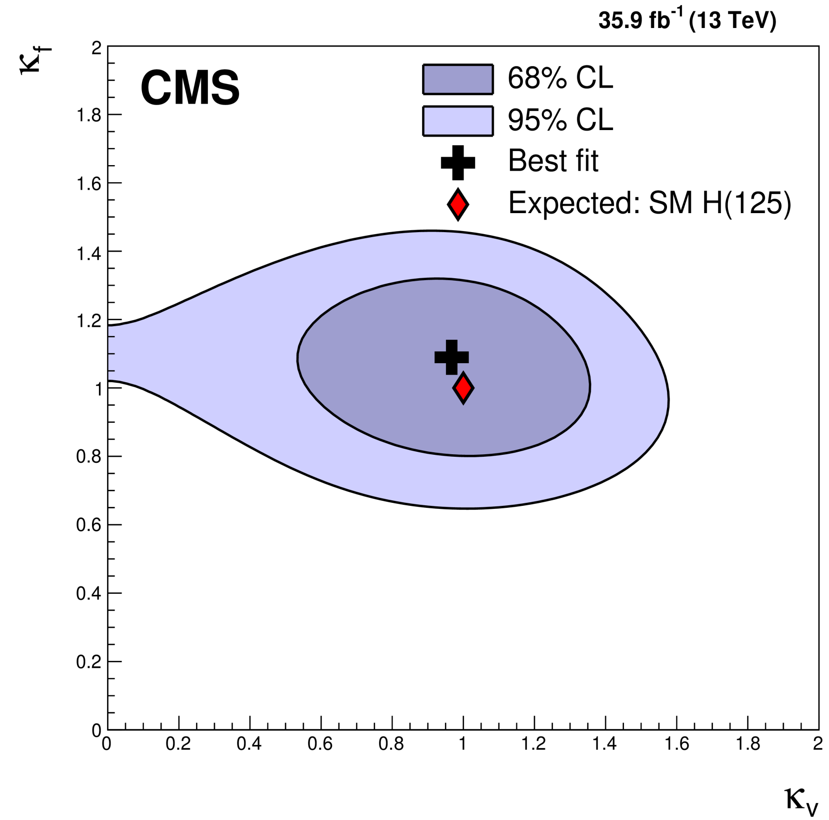
- probing universal coupling modifiers

$$\kappa_f = \frac{g_{Hf\bar{f}}}{g^{SM}_{Hf\bar{f}}}$$

affects gg → H production rate and H → ττ decay rates

$$\kappa_V = \frac{g_{HVV}}{g^{SM}_{HVV}}$$

affects VBF and VH production rates and H → VV decay rates




- Contribution of VH is added but not targeted with specific category

- Contribution from H → WW (significant in eμ channel, sub-dominant in other channels) is treated as signal

- Measurement of couplings is compatible with SM expectation

# Summary

- **H(125) decays to τ-leptons are studied at CMS using Run II data**

- **H(125) → ττ decay is observed with statistical significance of ~5.9σ combining data collected at 7, 8 and 13 TeV**
  - **first observation of Yukawa coupling in a single experiment and single decay channel**

- **measured H(125) properties in the H → ττ decay channel are consistent at current precision level with SM expectations**

- **Paper is submitted to PLB and made available in hep archive**

  **arXiv:1708.00373 [hep-ex]**

- **larger dataset is expected by end of 2017**

  **→ better measurement precision**

  **→ new measurements possible, e.g. probing CP properties**