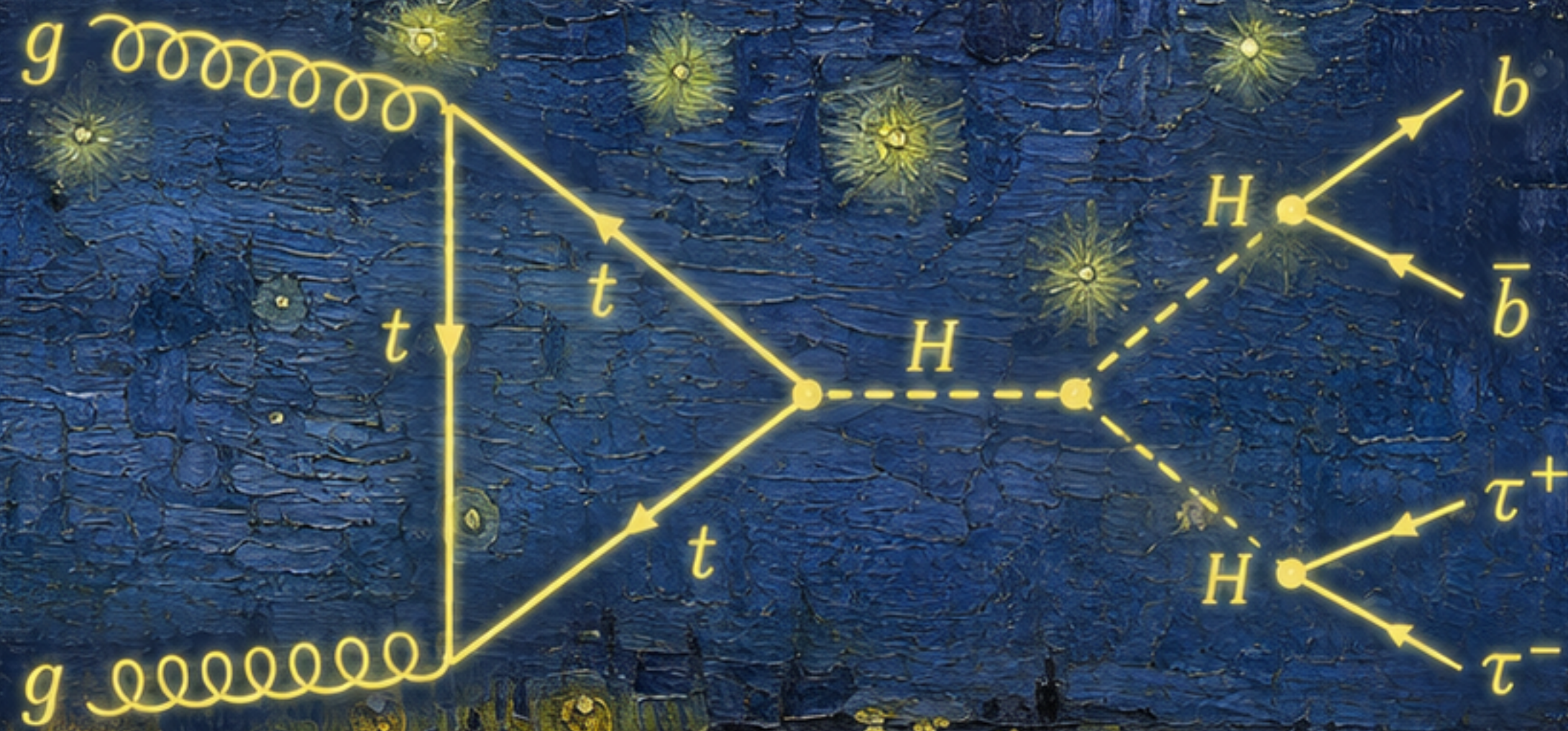




Searching for HH in the $b\bar{b}\tau^+\tau^-$ final state at CMS

Higgs Hunting 2026 - IJC/APC

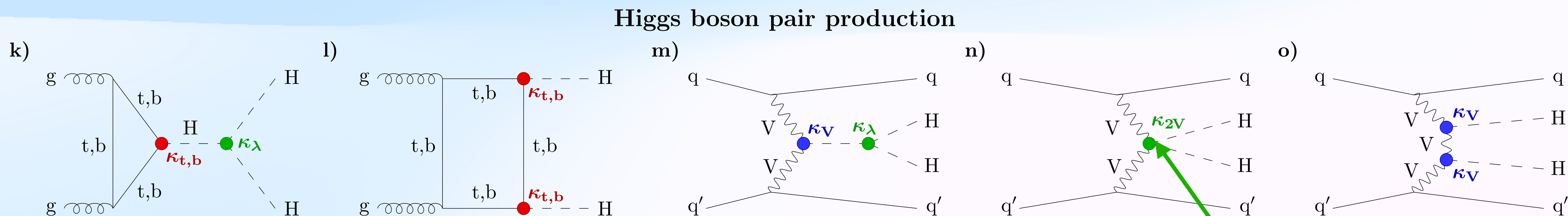
Davide Zuolo - University of Colorado Boulder *on behalf of the CMS Collaboration*



University of Colorado **Boulder**

Introduction - 1

- The measurement of $\sigma(HH)$ is the best way to extract the Higgs self-coupling λ_{HHH} (can also be extracted from single Higgs measurements)
- This parameter defines the shape of the Higgs potential together with m_H (measured at **per-mille precision level**, LHC Run 2 combination ongoing) and the vacuum expectation value (v , computed)



Gluon-Gluon Fusion (GGF)

Destructive interference between the two diagrams leads to a small cross section (34.1 fb @ 13.6 TeV)

Vector Boson Fusion (VBF)

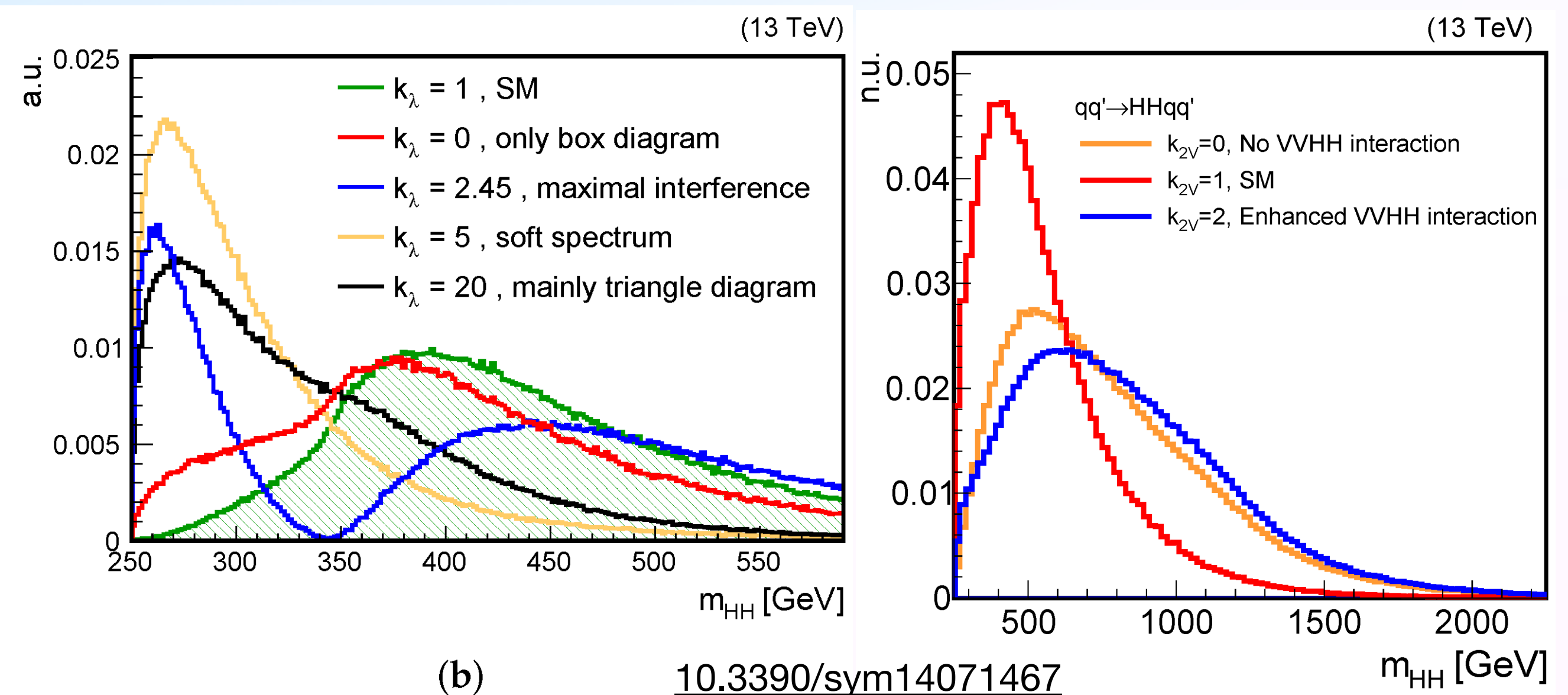
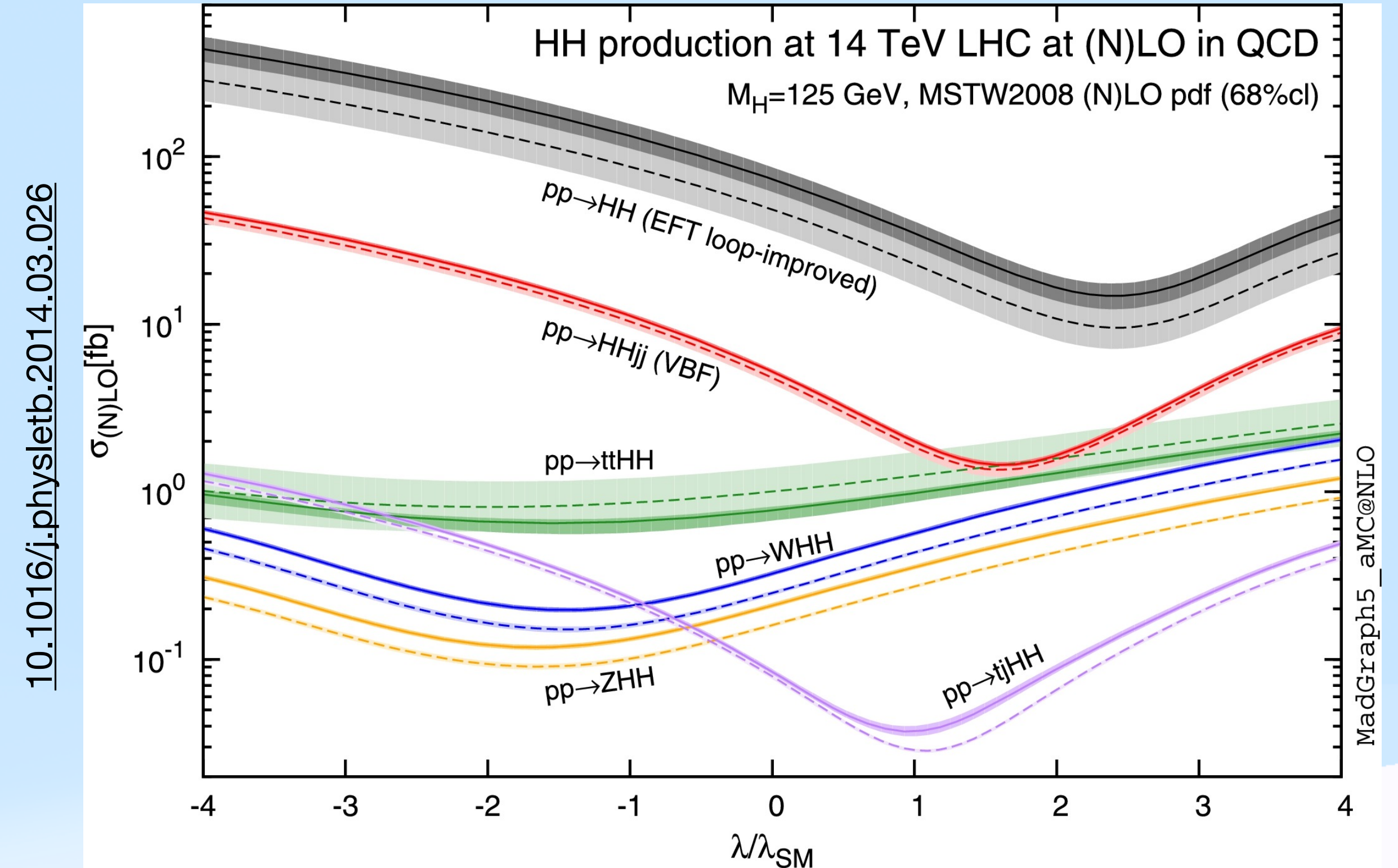
Cross section: 1.87 fb @ 13.6 TeV

Handle to probe VVHH coupling (C_{2V})

Complemented by VBS VVH production (Recent CMS paper)

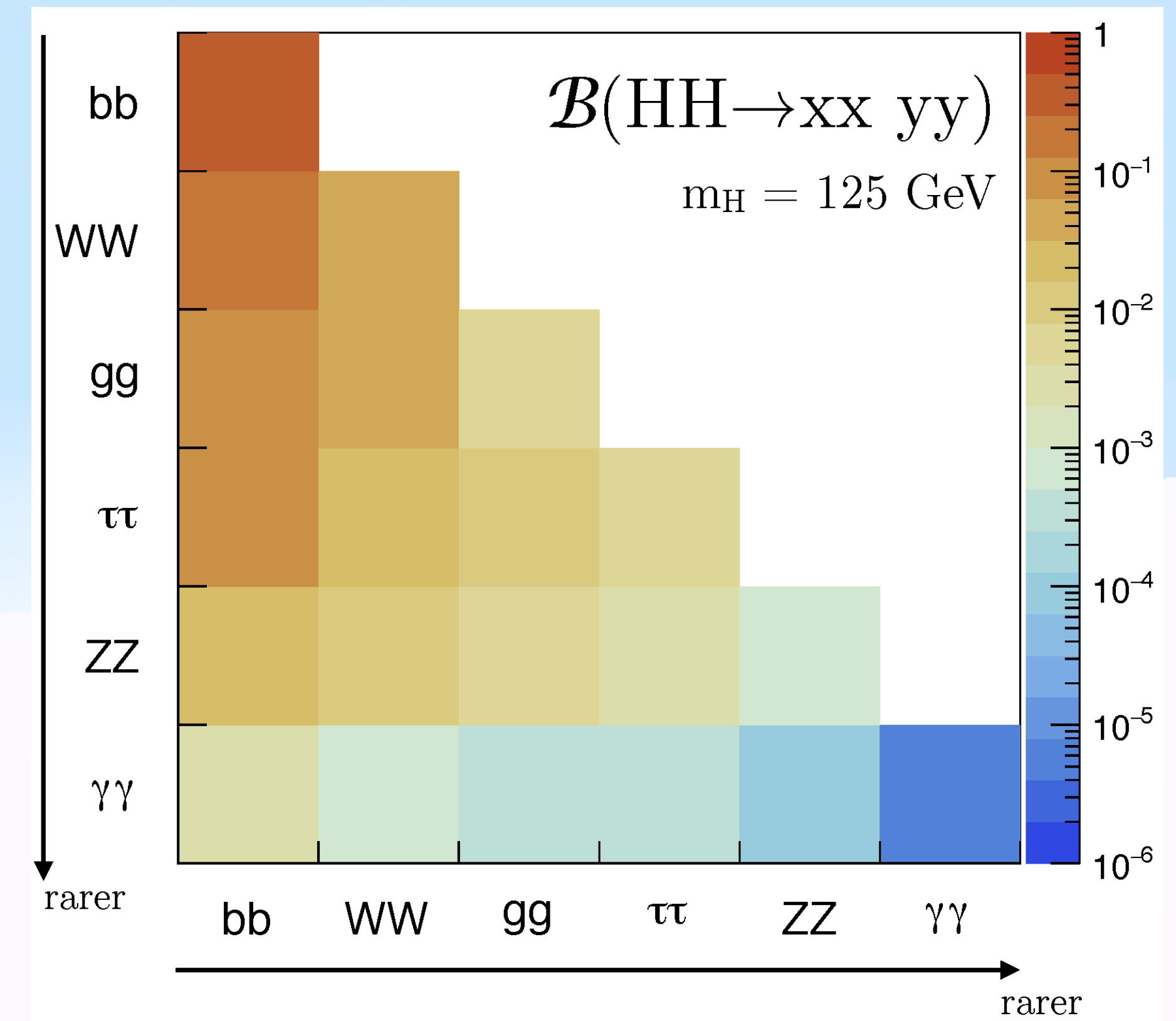
Introduction - 2

- **Anomalous values** of the couplings affects both the **cross section** and the **kinematics** of HH production.
- They are modeled through coupling modifiers
 - $k_\lambda = \lambda_{HHH}/\lambda_{HHH,SM}$
 - $k_{2V} = C_{2V}/C_{2V,SM}$
 - $k_V = C_V/C_{V,SM}$
- Both ATLAS and CMS are now exploring other production modes (mainly VHH and ttHH)

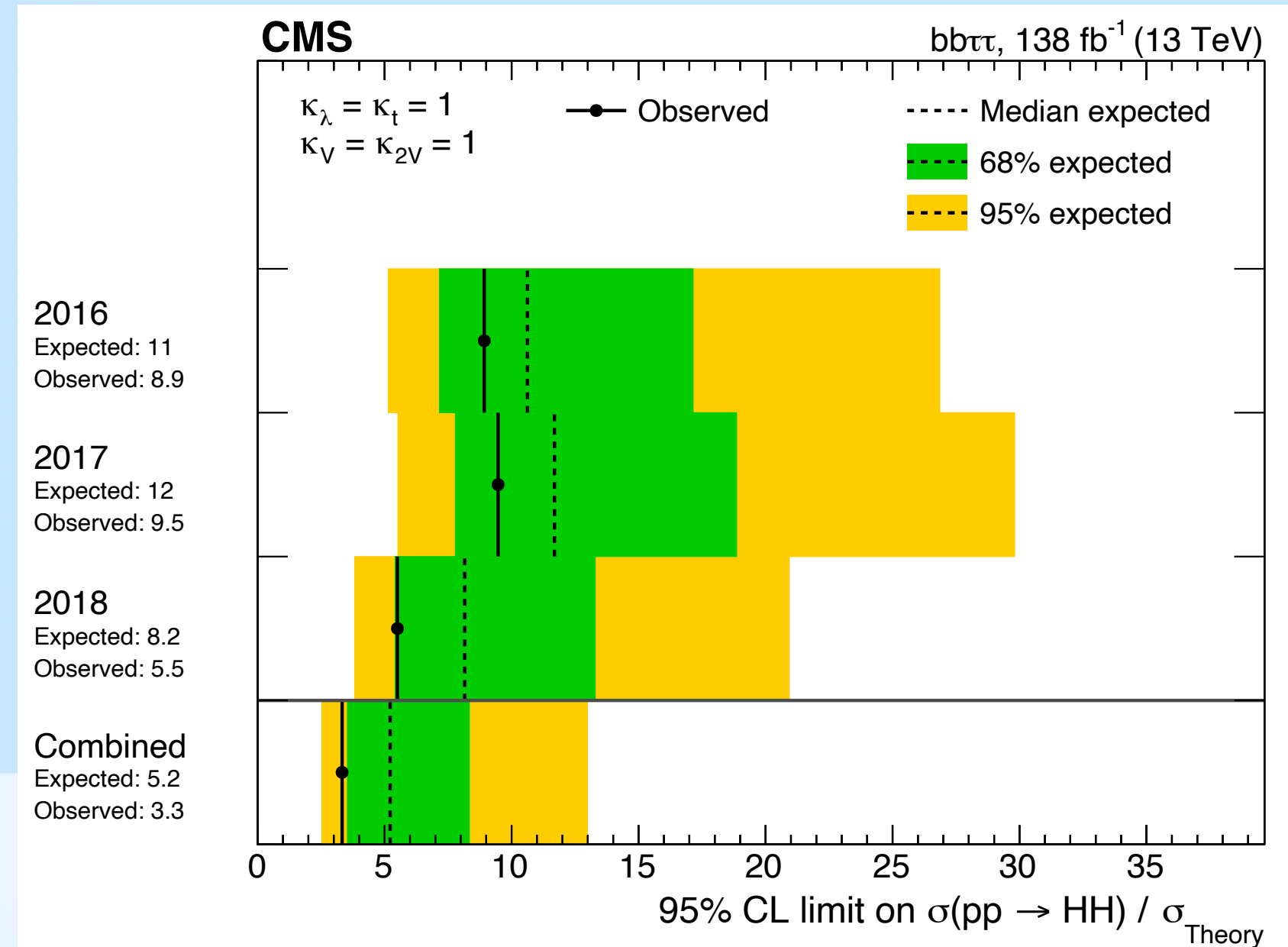


Introduction - 3

- Historical focus was on final states with at least two b-jets (branching ratio 58%)
- **bb $\tau\tau$** provides a good compromise between branching ratio (7.3%) and background rejection (more and more powerful *b* and *tau* tagging algos in the last 10 years)
- **bb $\gamma\gamma$** has very low branching ratio but very easy to trigger and reconstruct
- **bbbb** has the highest branching ratio but QCD background estimation is very complex
- Both ATLAS and CMS are now exploring (almost) the whole phase space through multi-leptons and multi-bosons searches

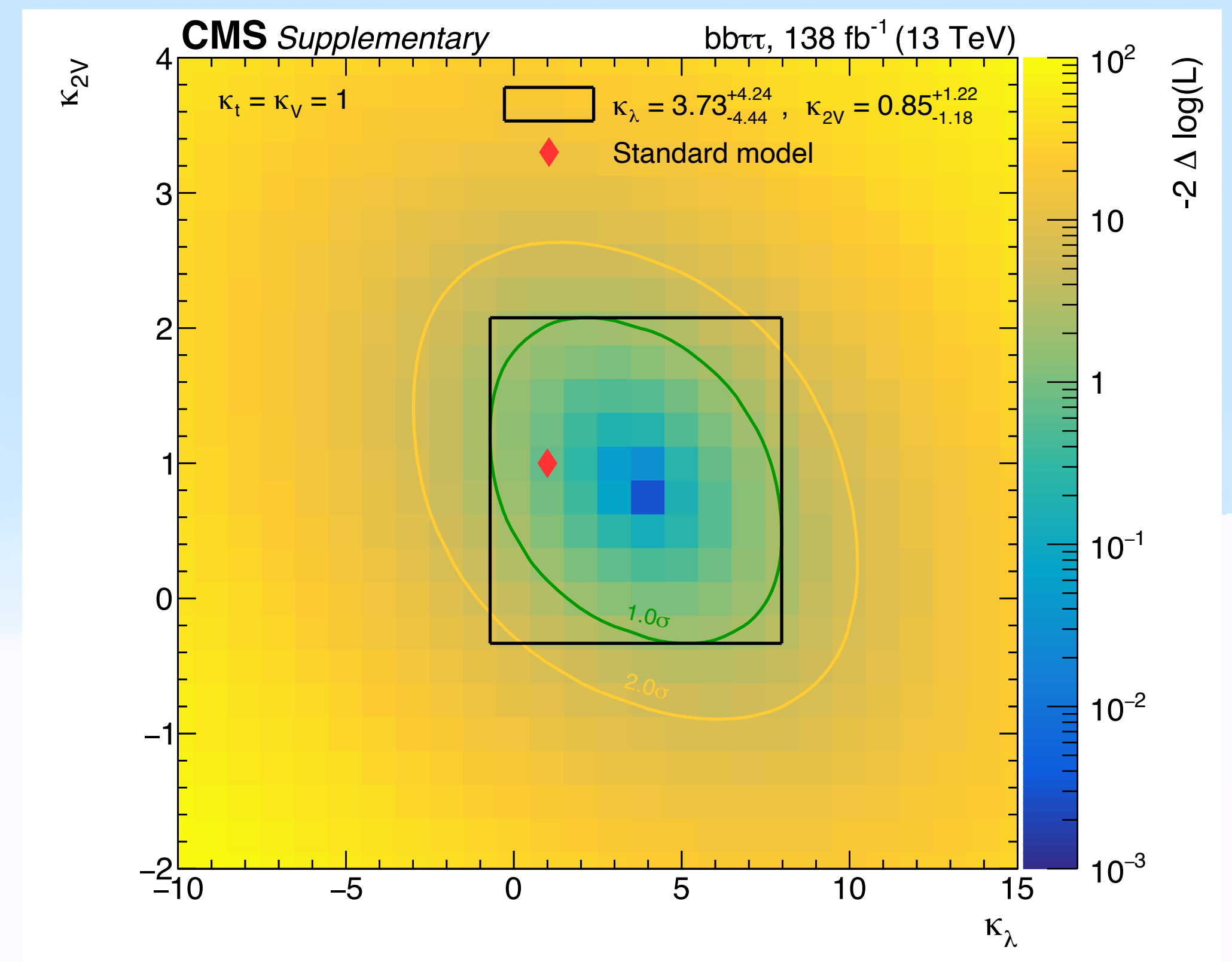


Where do we stand? - Full Run 2 analysis



Excluding cross sections larger than
3.3 (5.2) x SM at 95% CL

This result has been included in the CMS HH combination and in the ATLAS+CMS HH combination (Excluding cross sections larger than **2.5 (1.7) x SM** at 95% CL!!!).



κ_λ bound to (-1.8, 8.8) [-2.9, 9.8]

κ_{2V} bound to (-0.4, 2.6) [-0.6, 2.8]

ATLAS+CMS HH combination

κ_λ bound to (-0.71, 6.1) [-1.3, 6.7]

κ_{2V} bound to (0.73, 1.3) [0.66, 1.4]

Moving forward - Partial Run 3 analysis

- This new CMS analysis exploits data collected during the **LHC Run 3** between **2022 — 2024**, amounting to **172 fb⁻¹**
- While adopting the same general strategy as the previous one, several improvements are included
 - Some of them are **central CMS developments** (e.g. new *b* and *tau* taggers, *one tagger to rule them all*)
 - While others are **in-house developments** (e.g. DNN-based regression for di-tau mass estimation, new trigger paths, new jet $\rightarrow \tau_h$ fakes estimation)
- **Basically a creativity-driven analysis!**

Analysis workflow

From here onwards improvements shown in green

H $\rightarrow$ $\tau\tau$ selection

- Select 2 well identified leptons
- Veto additional loose lepton
- Selection of best leptons pair:
 $e\tau_h$, $\mu\tau_h$ or (boosted) $\tau_h\tau_h$

H $\rightarrow$ bb selection

- Assess the presence of 2 small radius jets or 1 large radius jet and of 2 extra jets (VBF production)
- b-jets and VBF-jets candidates selected using DNN-based algorithms

Categorization and identification of signal events through DNNs (one for ggF and one for VBF production)

- Elliptical mass selection on m_{bb} and $m_{\tau\tau}$ tailored to 99% signal efficiency to remove background outliers prior to DNN inference

Signal Extraction

- Binned maximum likelihood fit to the DNN shape to extract signal
- Systematic uncertainties included as nuisance parameters in the likelihood

MC simulations

Signal

- **ggF and VBF, SM and BSM HH signals**
- ggF: $k_\lambda = 1, 2.45, 5$
- VBF: $(k_V, k_{2V}, k_\lambda) =$
 - $(1, 1, 1)$
 - $(1, 0, 1)$
 - $(1.74, 1.37, 14.4)$
 - $(-0.758, 1.44, -19.3)$
 - $(-0.962, 0.959, -1.43)$
 - $(-1.21, 1.94, -0.94)$
- Signal morphing vs k_{2V} and k_λ performed with standard approach used for all CMS HH analyses

Background

- All backgrounds (but fakes) are modeled from MC and checked/corrected in orthogonal CRs
- **Drell-Yan**: Z+jets NLO samples stitched in N_{jets} and corrected in-situ with $p_T(Z)$ and N_{jets} corrections derived from $Z \rightarrow \mu\mu$ control region
- **TT**: fully leptonic, semi-leptonic, fully hadronic
- **Single-H**
- W+jets: W+2 jets stitched in p_T and N_{jets}
- Others: VV, VVV, single-t, tW, tt+X

Triggers

Electron + τ_h final state

- singleEle
- ele+ τ_h
- parking(ele+VBF-jets) - 2023 onwards
- parking(VBF-jets) - 2023 onwards
- **DISCLAIMER: Trigger paths in each final state are combined to cover orthogonal portions of the phase space $\rightarrow$ no logical "OR" applied**
- Details on thresholds can be found in backup
- Data parking represented a big addition to the CMS physics program!

Muon + τ_h final state

- singleMu
- ele+ τ_h
- parking(mu+VBF-jets) - 2023 onwards
- parking(VBF-jets) - 2023 onwards

Boosted τ_h + τ_h final state

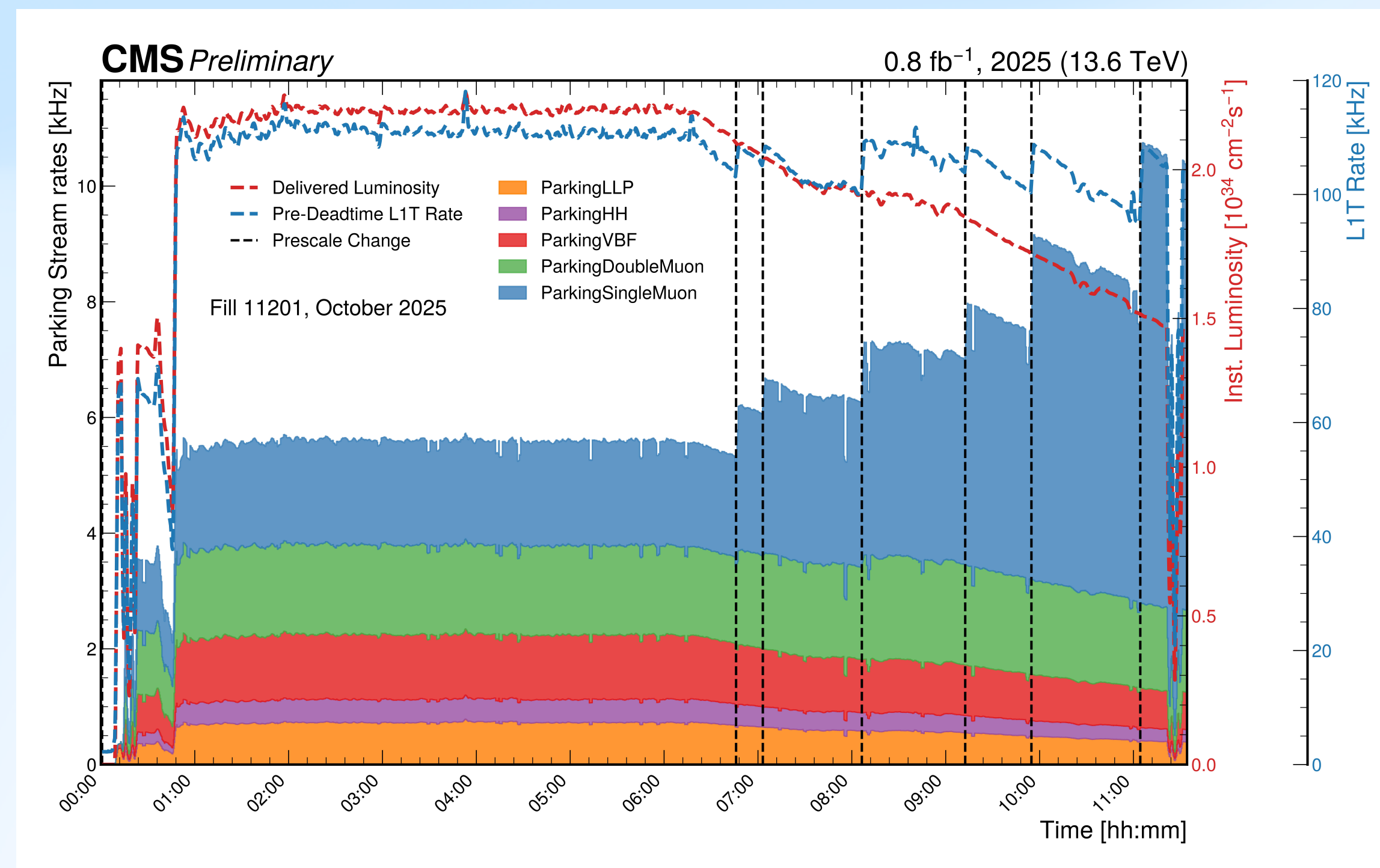
- Large missing momentum
- **One large radius jet tagged as $H \rightarrow \tau\tau$ at HLT level!!!**

τ_h + τ_h final state

- diTau
- **diTau+2jets - 2022 onwards**
- parking(singleTau+VBF-jets) - 2023 onwards
- parking(diTau + VBF-jets) - 2023 onwards
- parking(4jets-1b1tau) - 2024 onwards
 - **1 jet tagged as b and 1 jet tagged as τ_h at HLT level using the same algo!!!**

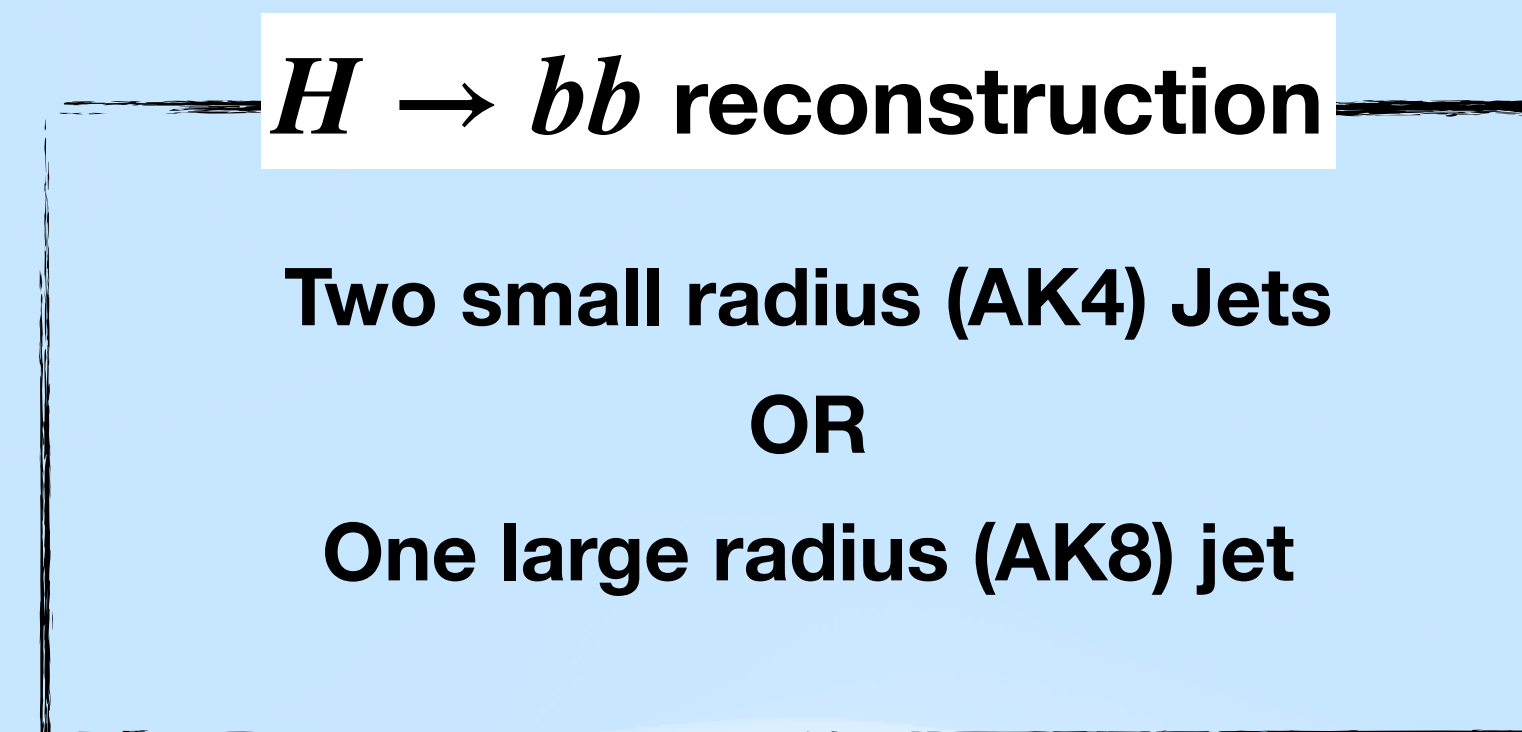
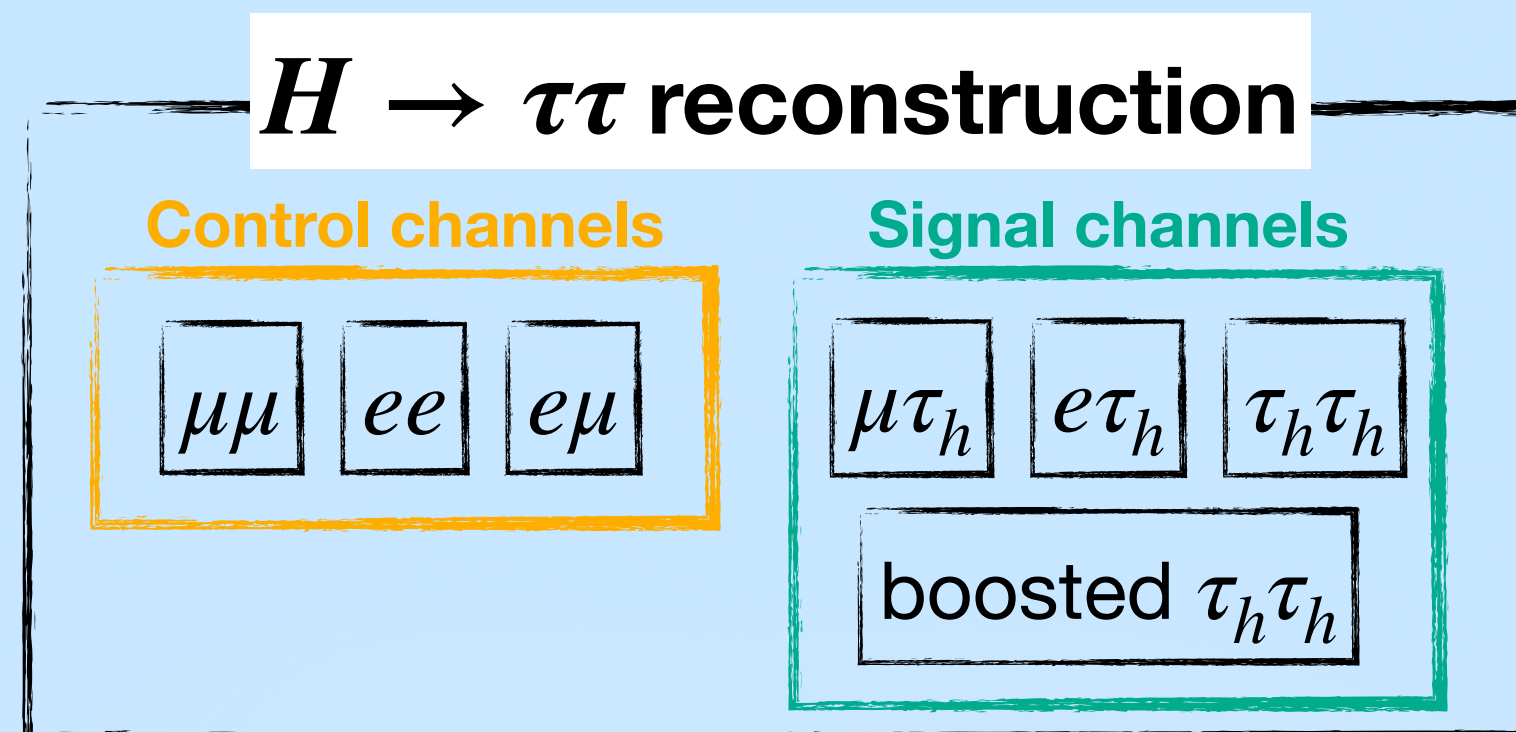
Honorable mention: data parking

- Instantaneous luminosity delivered by LHC reduces $\rightarrow$ free trigger bandwidth $\rightarrow$ fill trigger bandwidth with additional triggers $\rightarrow$ reconstruct events only when computing resources are available (in practice online reco achieved at CMS during LHC Run 3)
- Trigger paths designed for **BPhysics(since2018!!!)**, HH, VBF and Long Lived Particle events



Categorization

Baseline
selections
(backup)



Topological
selections

Is there a good AK8 candidate?

No

Are there 2 VBF jets candidates?

Yes

Boosted category

Yes

No

How many good b-jets?

One

Two

VBF category

Res1b category

Res2b category

Signal
extraction

GGF DNN

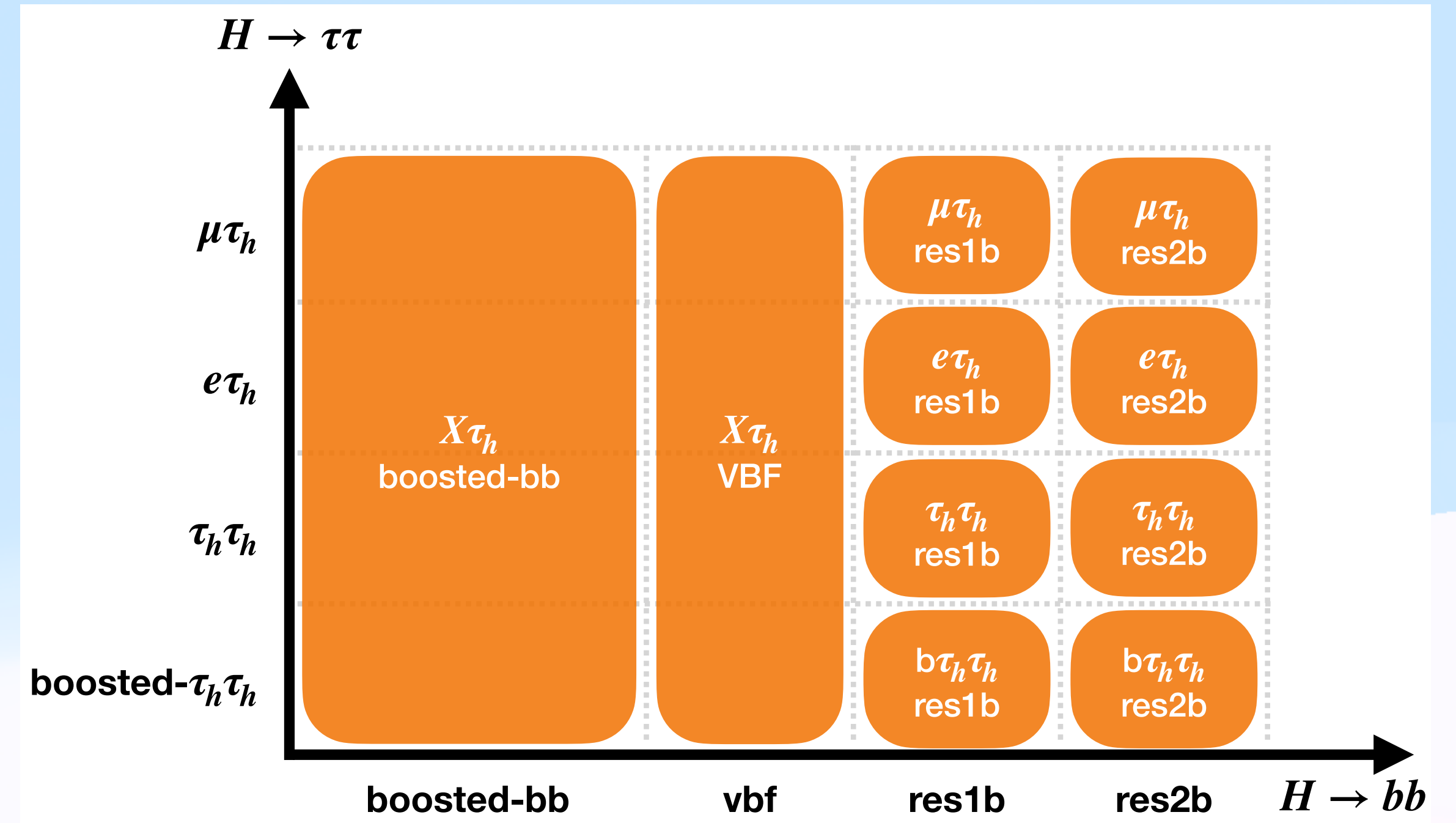
VBF DNN

GGF DNN

GGF DNN

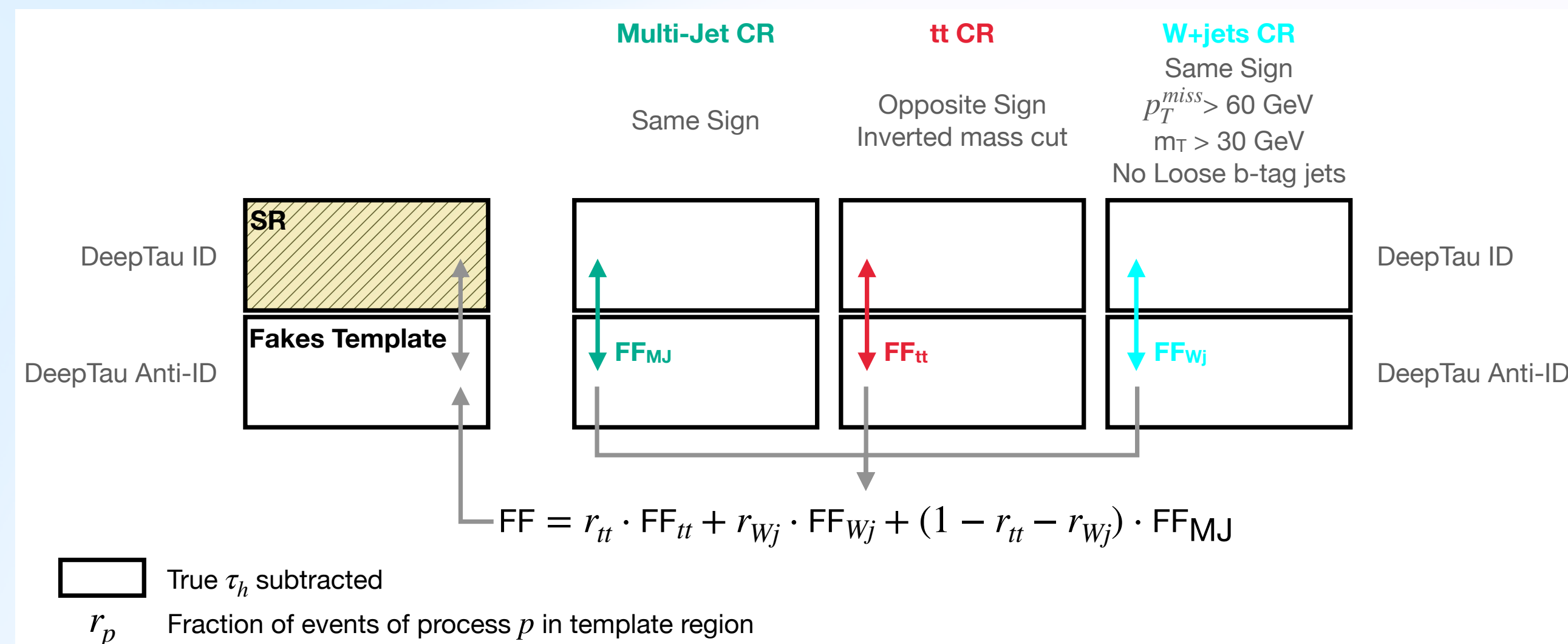
Signal regions

- Due to the low available statistics events in the four channels in the boosted-bb and vbf category are merged in a single $X\tau_h$ channel
- **10 signal regions x 5 data taking periods** fitted simultaneously

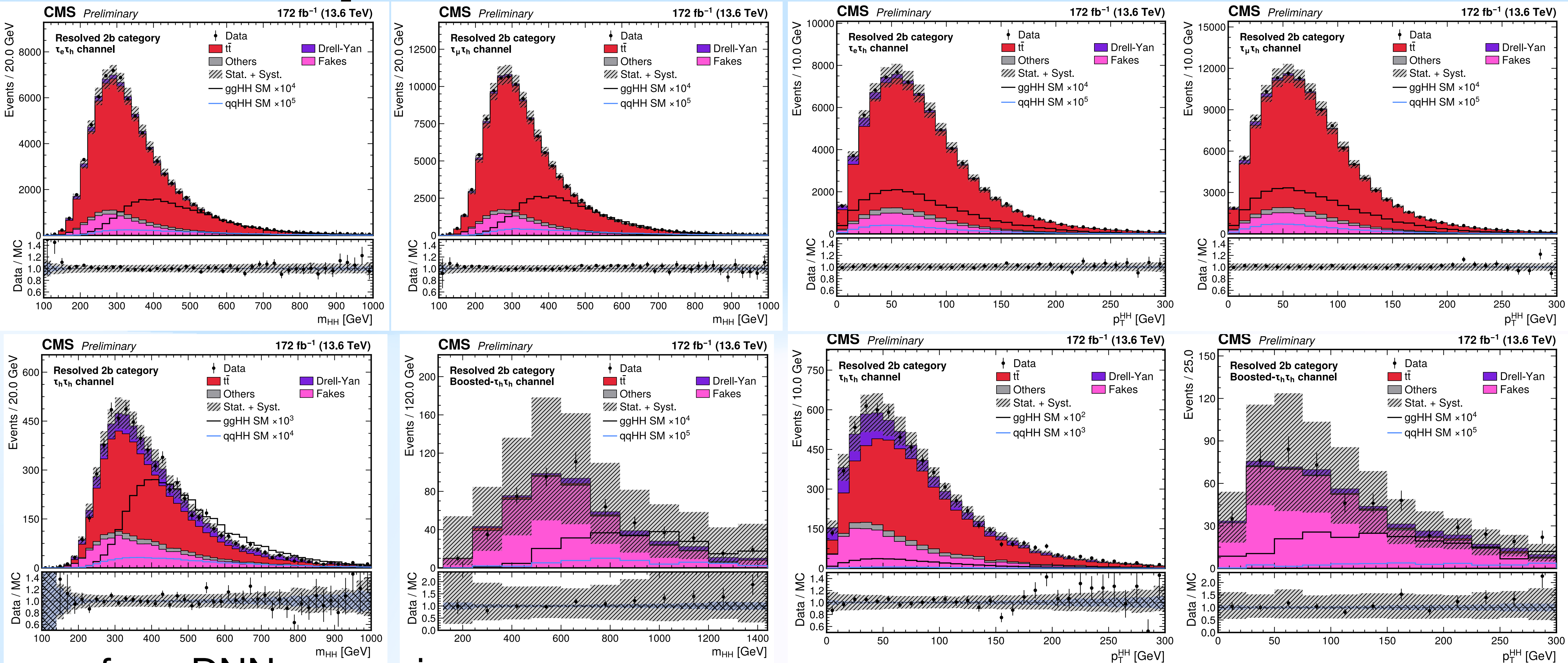


Fakes estimation

- Three control regions (in each category) defined to evaluate fakes from multi-jet, tt and W +Jets events
- Ratio between isolated and not-isolated regions is evaluated in each CR binned in p_T and for barrel and endcap regions
- A per event weight is computed taking into account the fraction of each fake type



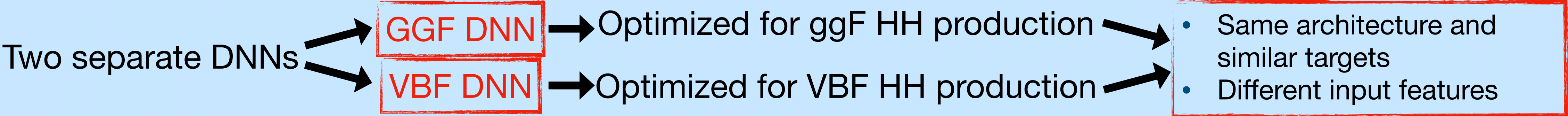
Control plots



m_{ττ} from DNN regression

m_{bb} from standard CMS processing

Signal extraction



Training data

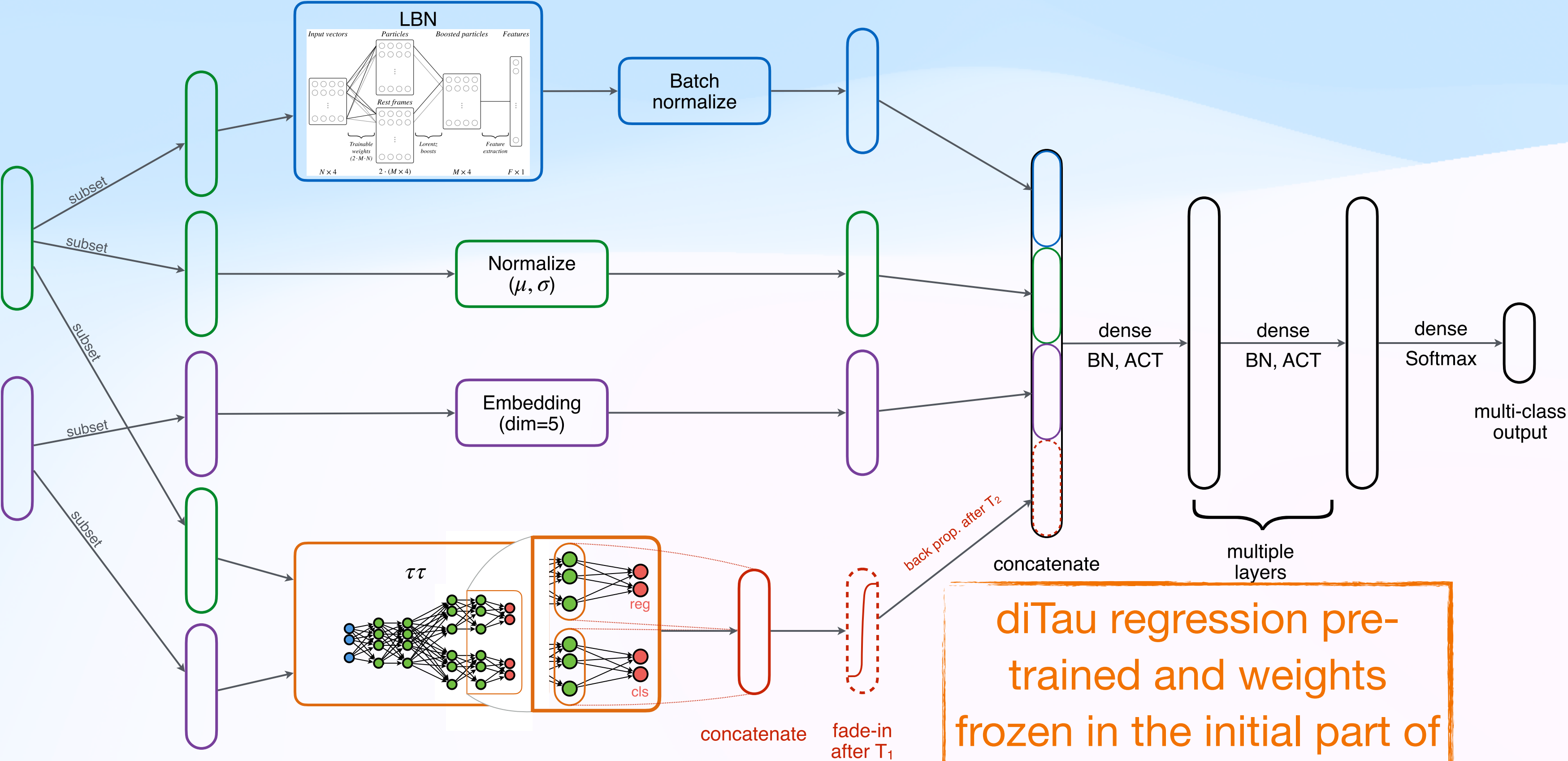
DY, TT, and all signal bb $\tau\tau$ processes (ggF/ggF+VBF) for all five eras

Training procedure

- Separate training and validation set
- 5-fold cross-validation
- Oversampling
- MC weights aware

Output nodes

- HH
- DY
- TT



diTau regression pre-trained and weights frozen in the initial part of global training

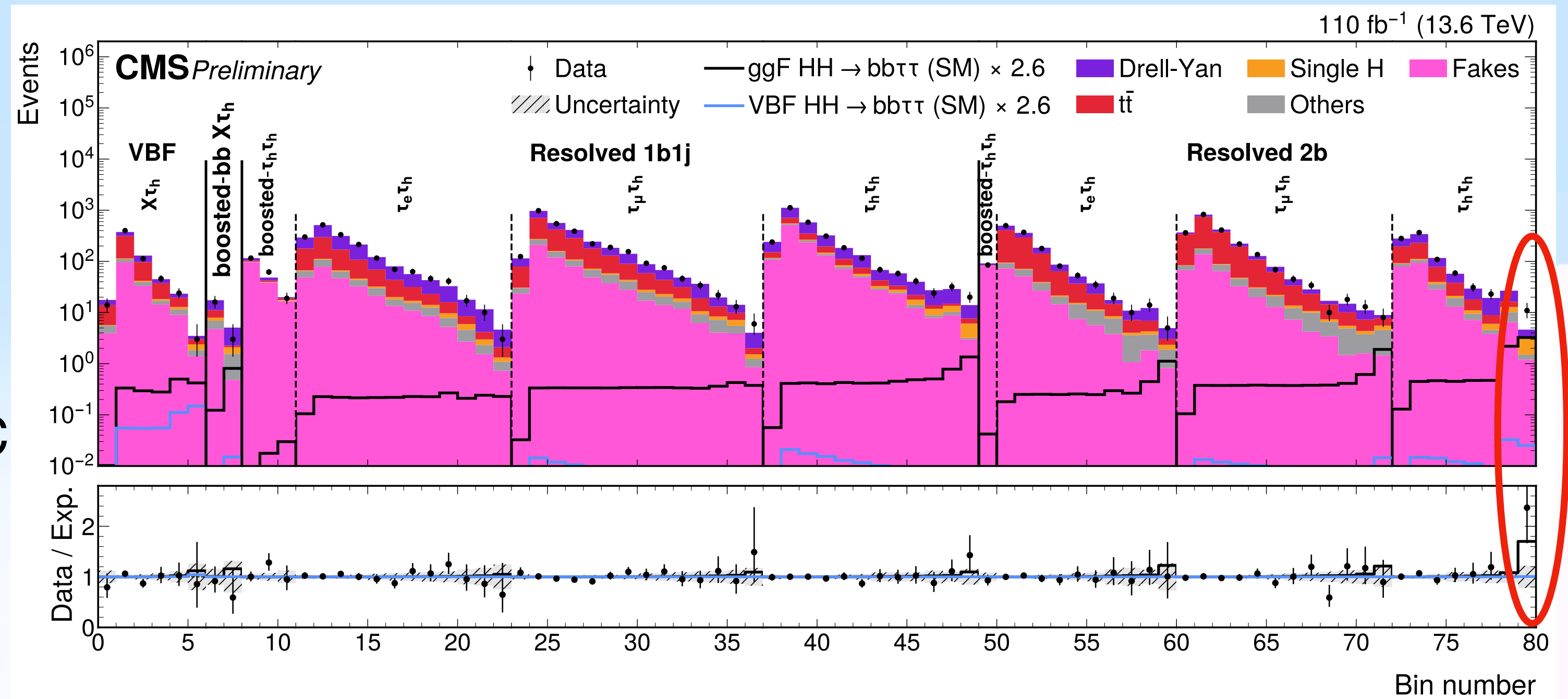
Systematic uncertainties

- The analysis remains statistically limited
 - Limited statistics from some MC samples (particularly DY)
 - Limited efficiency x acceptance (studies ongoing)
- Uncertainty on fakes estimation has the largest relative contribution
- Theory predictions on HH and single-H remain an important part of the uncertainty budget

Uncertainty source	Relative contribution to $\Delta\mu_{HH}$
Experimental	
Fake factors	25%
b-tagging	10%
Lepton corrections and triggers	9%
Jet energy corrections	7%
Other experimental	3%
Statistical	
Overall	74%
Bin-by-bin variations	35%
Theoretical	
HH	31%
single-H	27%
Background	14%

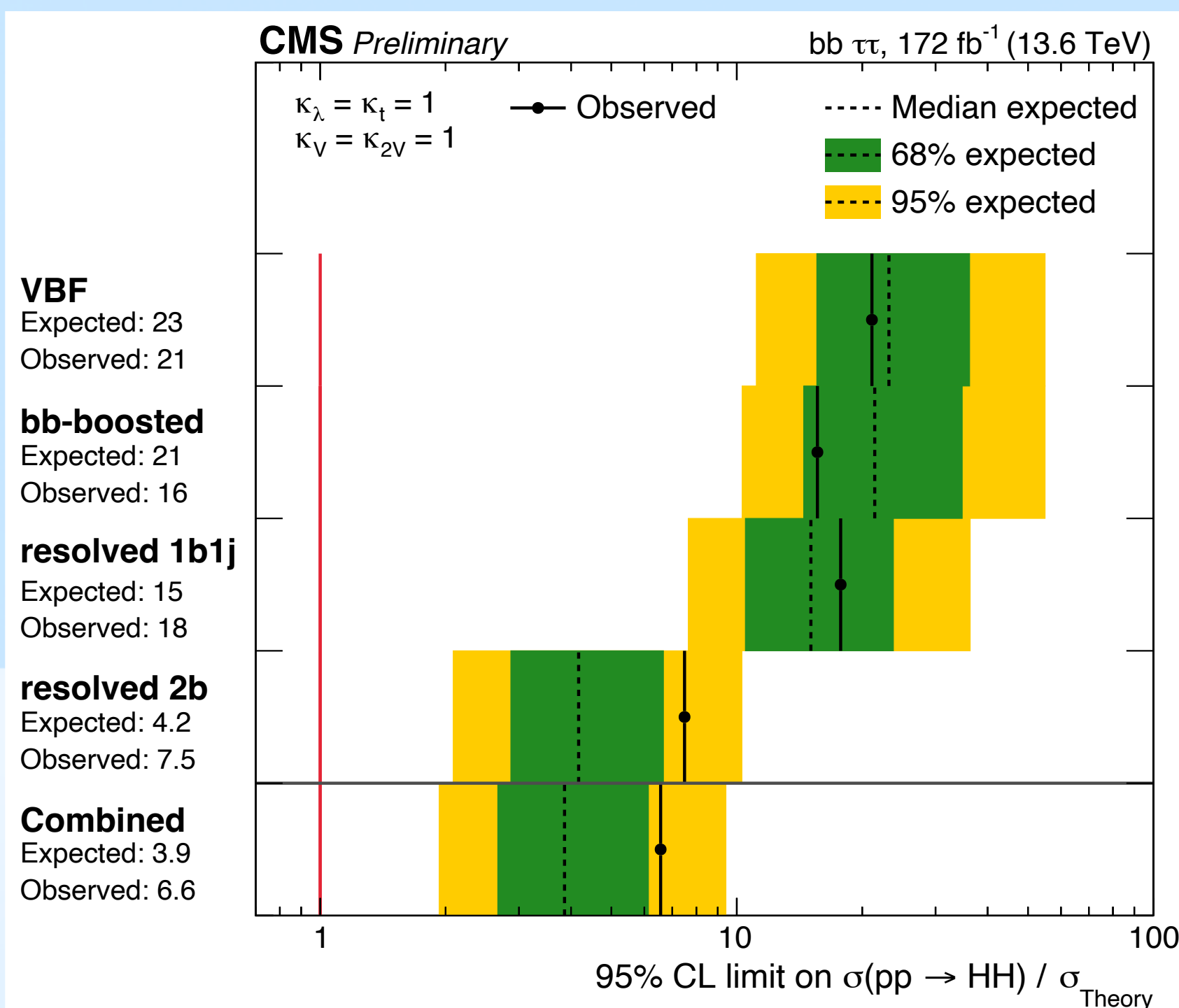
DNN shape

- We perform binning optimization in the range $[0.75, 1]$ of DNN HH output score
 - Events below 0.75 discarded
- Necessary requirements when forming a bin:
 - Signal fraction greater than inverse number of target bins
 - DY and TT have at least one MC event and yield is positive
 - Bkg yield is greater than 25x its variance (10x for the right-most bin)
 - Last 2 requirements also apply for all the shape variations
 - Number of events in the fakes template region ≥ 10 per bin

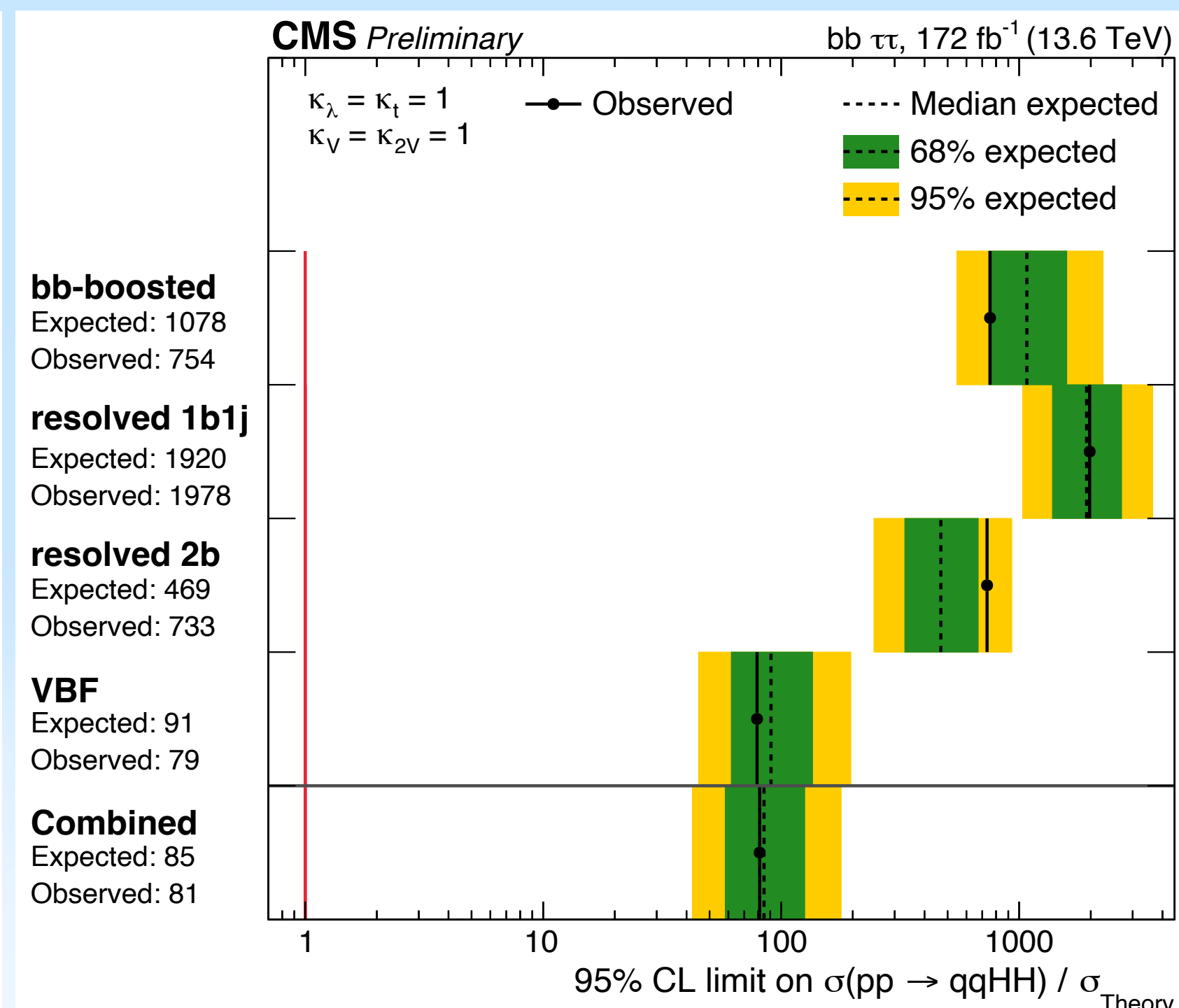


Postfit shape for 2024

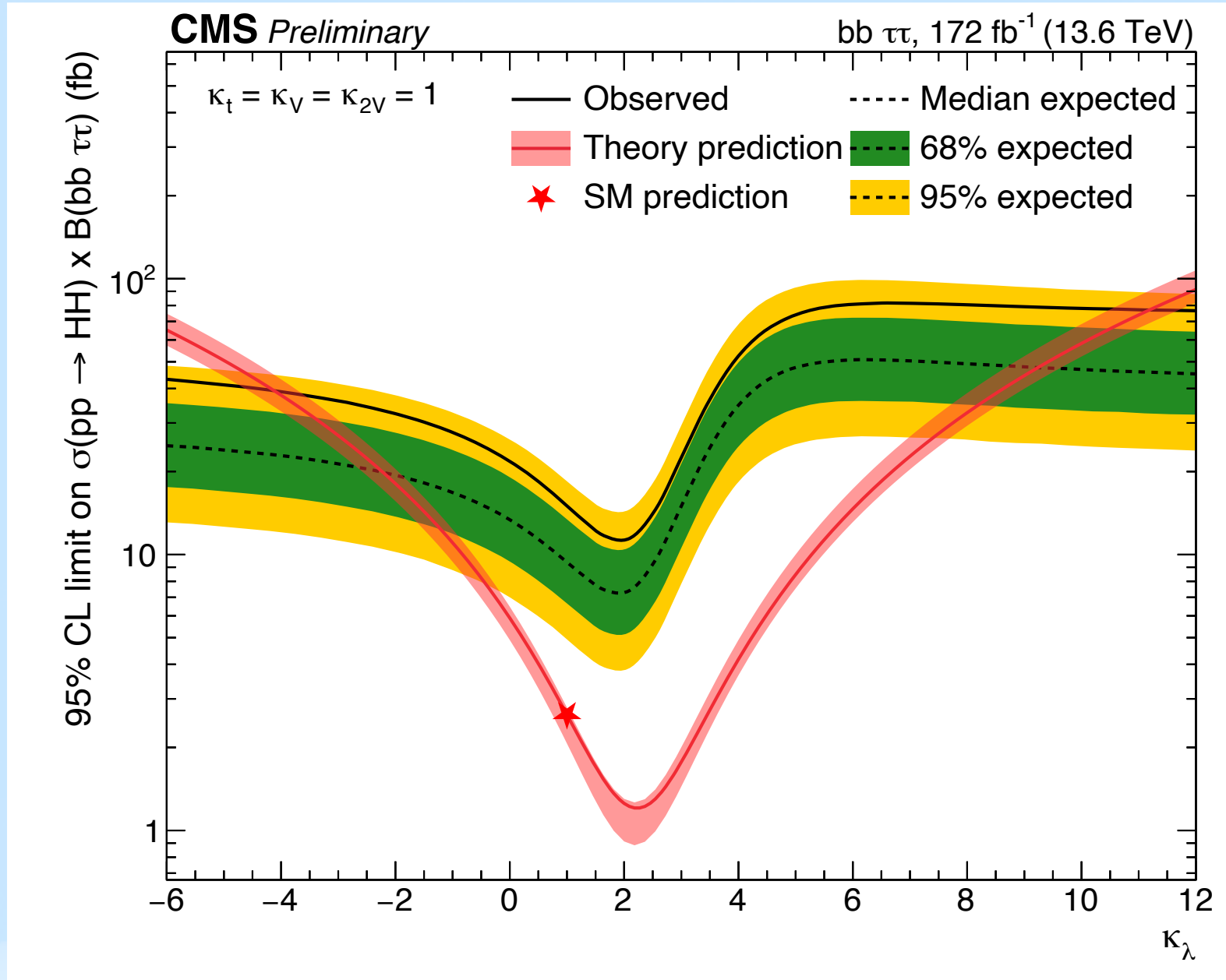
Partial Run 3 results



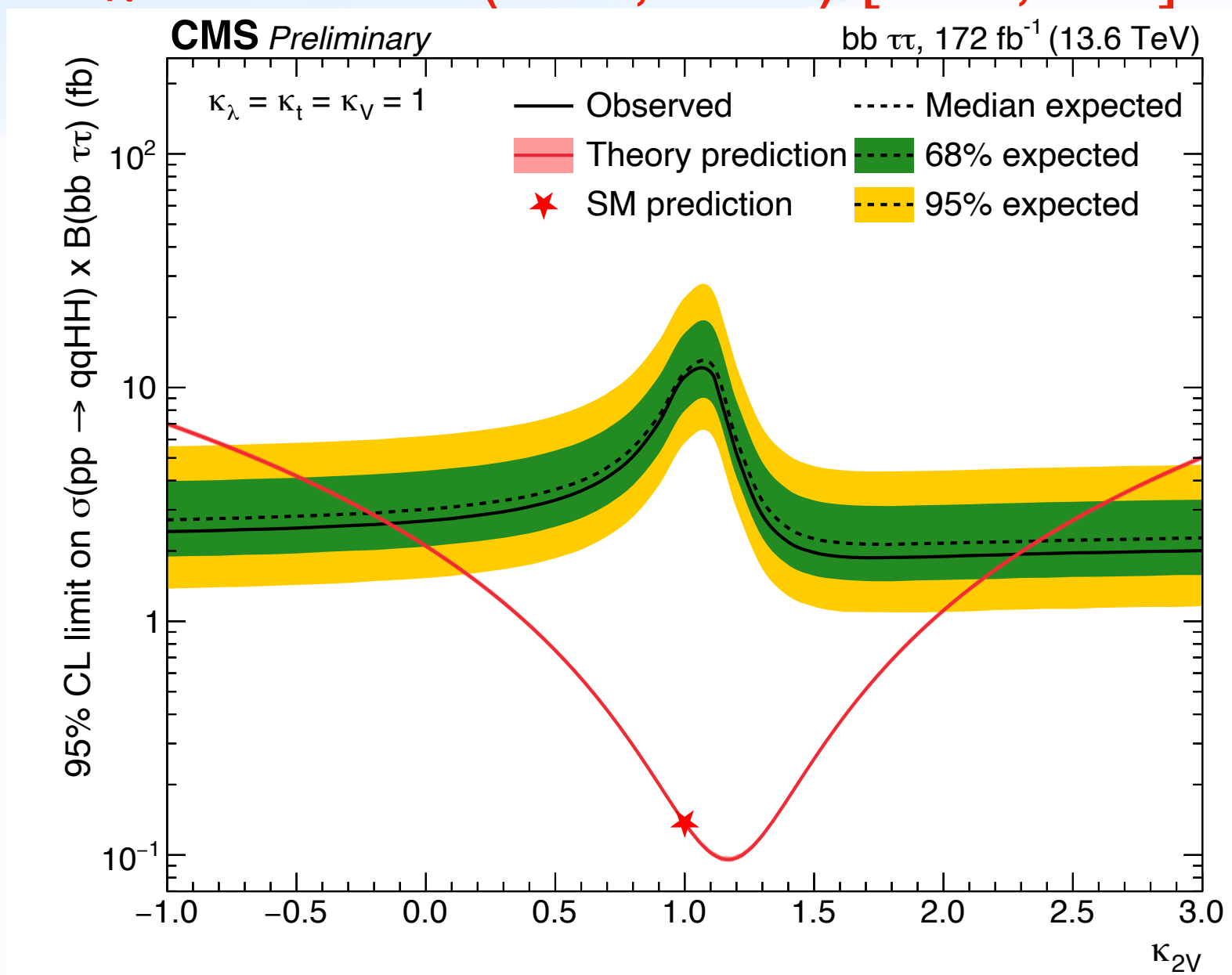
Excluding cross sections larger than
6.6 (3.9) x SM at 95% CL $\rightarrow$ **25%**
improved wrt Full Run 2 analysis!!!



Excluding cross sections larger than
81(85) x SM at 95% CL $\rightarrow$ **halving**
the expected limit of Full Run 2
analysis!!!

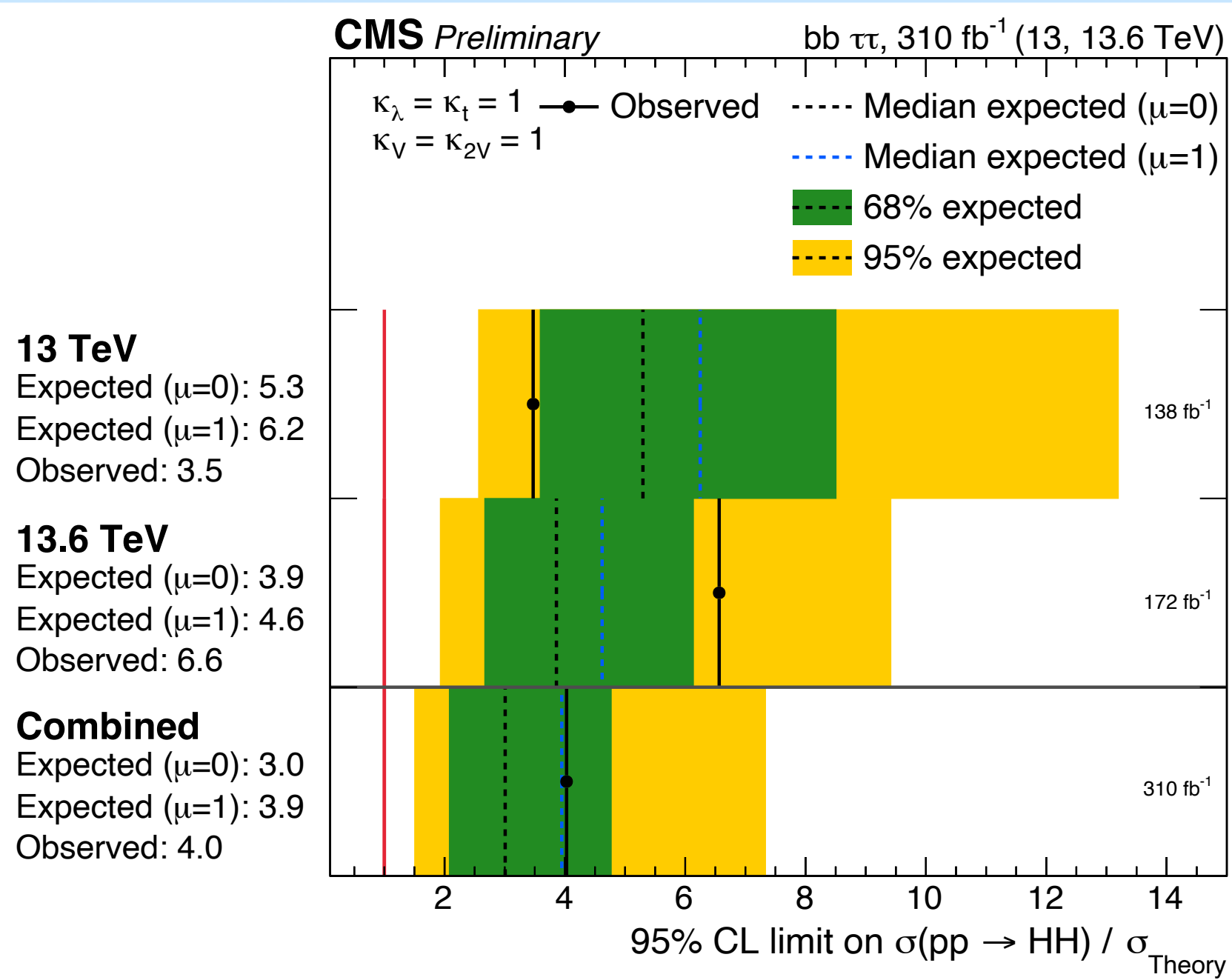


κ_λ bound to **(-4.1, 11.2) [-2.2, 9.2]**

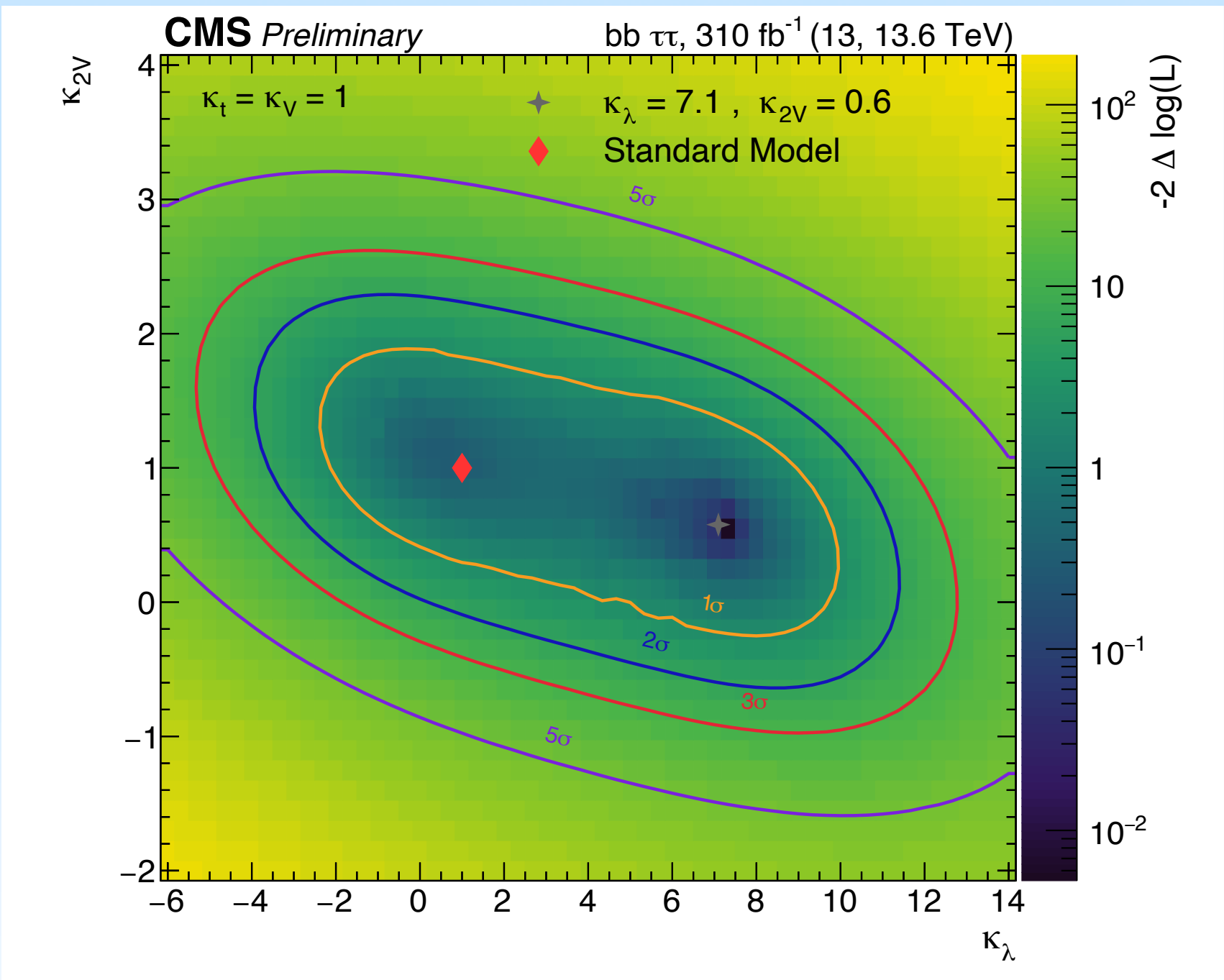


κ_{2V} bound to **(-0.2, 2.3) [-0.2, 2.4]**

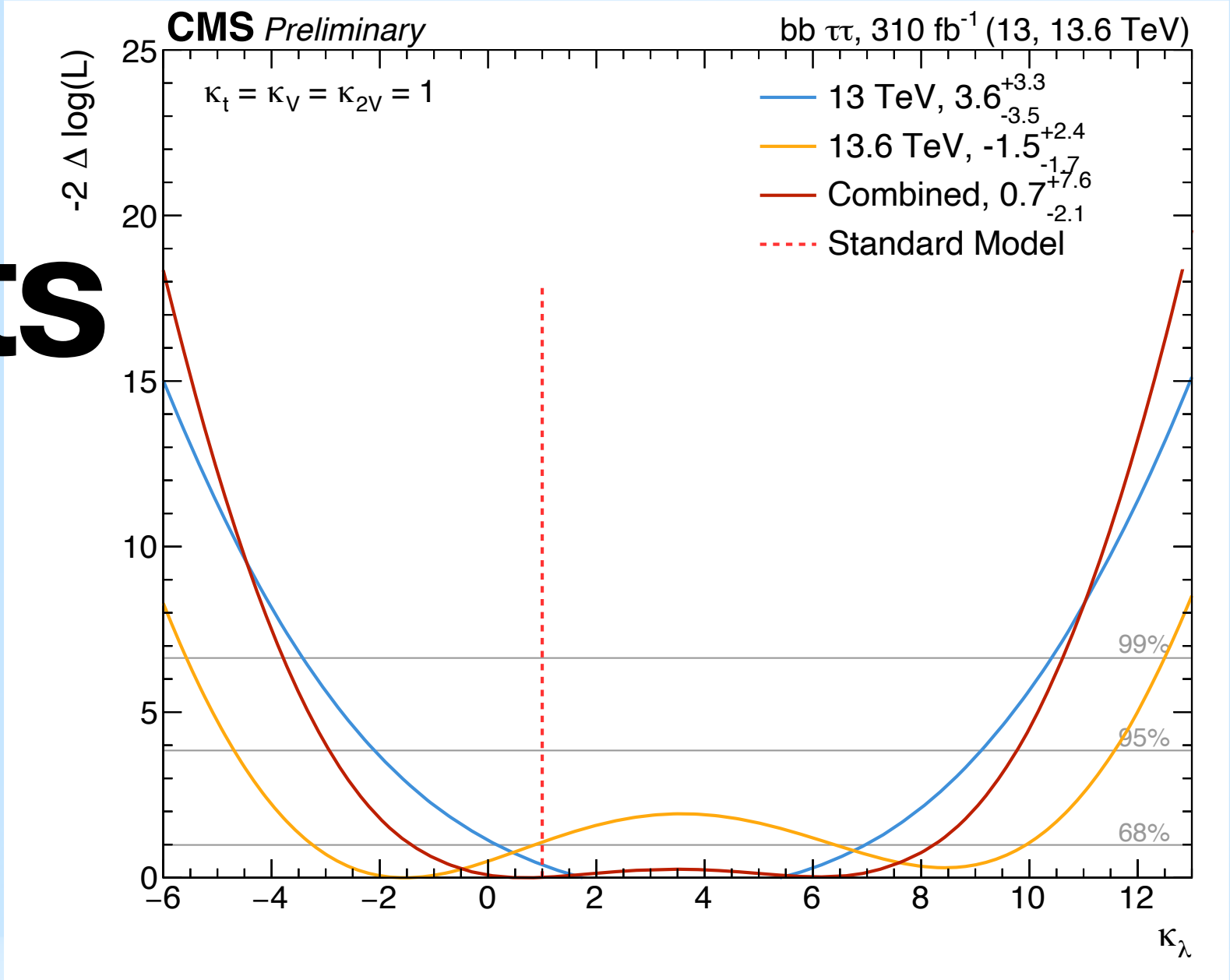
Partial Run 3 + Full Run 2 results



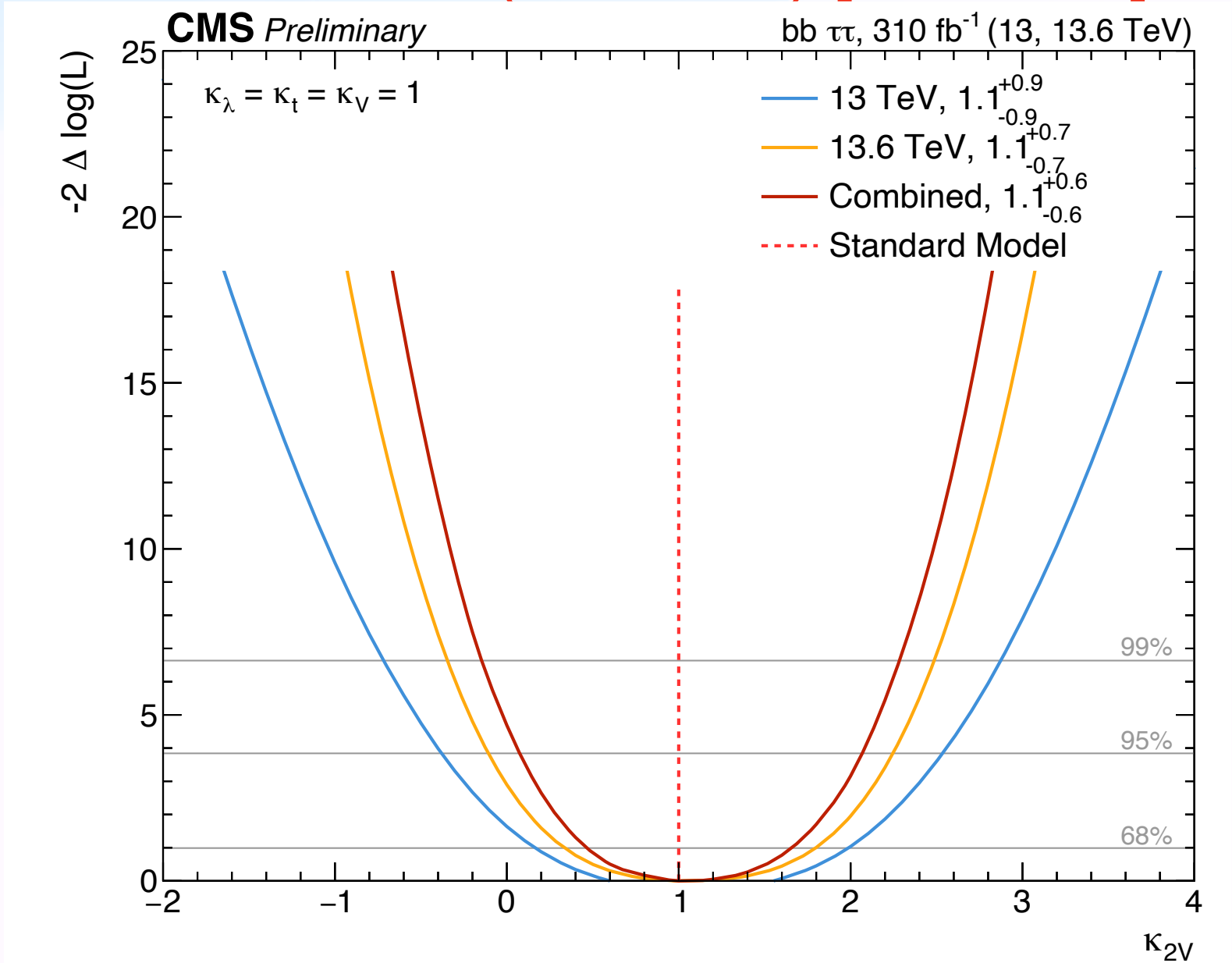
Excluding cross sections larger than
3.0 x SM at 95% CL —> **observed**
compatible with expected assuming
mu=1



Correlation visibly reduced wrt Full
 Run 2



κ_λ bound to (-2.5, 9.4) [-1.5, 8.4]



κ_{2V} bound to (0.0, 2.1) [-0.1, 2.2]

Conclusions and Prospects - 1

- Analysis of **LHC Run 3 data collected from 2022 to 2024 (172 fb⁻¹)** improved 95% CL upper limit on HH production cross section **by 25%** compared to Full Run 2 (140 fb⁻¹) analysis
 - 10% comes from lumi scaling, **15% comes from improvements along the analysis workflow**
 - Slight improvement on coupling constraints —> **plan to categorize events by m_{HH} mass in the next iteration to better disentangle SM and BSM couplings (as most of the CMS/ATLAS analysis)**
 - Large **improvement (factor 2!!!) also on VBF** production upper limits
- Next iteration of the analysis already planned to add 2025+2026 data and reprocess Run 2

Conclusions and Prospects - 2

- Systematics uncertainties are clearly limiting the sensitivity
 - Uncertainties on fakes estimation are very conservative given that the method was employed for the first time
 - Generating more MC events, especially for Z+Jets
- DNN shape binning strategy also quite conservative because of new fake estimation strategy → room for improvement
- Potential increase in efficiency x acceptance thanks to improved trigger paths in 2025 data taking

Uncertainty source	Relative contribution to $\Delta\mu_{HH}$
Experimental	
Fake factors	25%
b-tagging	10%
Lepton corrections and triggers	9%
Jet energy corrections	7%
Other experimental	3%
Statistical	
Overall	74%
Bin-by-bin variations	35%
Theoretical	
HH	31%
single-H	27%
Background	14%

Stay tuned... and

If you are a “*Doctor Who*” fan don’t forget to go and check this painting later ;)



Backup

Trigger selections

Year	Trigger requirement	Int. lumi [fb ⁻¹]
2022	Two τ_h , $p_T > 35$ GeV, $ \eta < 2.1$, Medium DeepTau ID	34.7
	Two τ_h , $p_T > 30$ GeV, $ \eta < 2.1$, Medium DeepTau ID + One jet, $p_T > 60$ GeV	34.7
	Two τ_h , $p_T > 20$ GeV, $ \eta < 2.1$, Medium DeepTau ID + Two jets, $p_T > 115(40)$ GeV, $m_{jj} > 650$ GeV	34.7
2023	Two τ_h , $p_T > 35$ GeV, $ \eta < 2.1$, Medium DeepTau ID	27.9
	Two τ_h , $p_T > 30$ GeV, $ \eta < 2.1$, Medium DeepTau ID + One jet, $p_T > 60$ GeV	27.9
	Two τ_h , $p_T > 20$ GeV, $ \eta < 2.1$, Medium DeepTau ID + Two jets, $p_T > 115(40)$ GeV, $m_{jj} > 650$ GeV	27.9
	One τ_h , $p_T > 45$ GeV, $ \eta < 2.1$, Medium DeepTau ID + Two jets, $p_T > 45$ GeV, $m_{jj} > 500$ GeV, $\Delta\eta_{jj} > 2.5$	21.8
	Two jets, $p_T > 105(40)$ GeV, $m_{jj} > 1000$ GeV, $\Delta\eta_{jj} > 3.5$	21.8
2024	Two τ_h , $p_T > 30$ GeV, $ \eta < 2.3$, Medium PNet τ_h -tag ID	109.9
	Two τ_h , $p_T > 26$ GeV, $ \eta < 2.3$, Medium PNet τ_h -tag ID + One jet, $p_T > 60$ GeV	109.9
	Two τ_h , $p_T > 20$ GeV, $ \eta < 2.2$, Medium PNet τ_h -tag ID + Two jets, $p_T > 115(40)$ GeV, $m_{jj} > 650$ GeV	109.9
	One τ_h , $p_T > 45$ GeV, $ \eta < 2.3$, Medium PNet τ_h -tag ID + Two jets, $p_T > 45$ GeV, $m_{jj} > 650$ GeV	109.9
	Two jets, $p_T > 125(45)$ GeV, $m_{jj} > 1050$ GeV	109.9
	Four jets, $p_T > 25$ GeV, HT > 250 GeV + One jet with PNet b-tag score > 0.20	109.9
	+ One jet with PNet τ_h -tag > 0.50	

Trigger selection for $\tau_h + \tau_h$ final state

Year	Trigger requirement	Int. lumi [fb ⁻¹]
2022	One e, $p_T > 30$ GeV, Tight Iso	34.7
	One e, $p_T > 24$ GeV, $ \eta < 2.1$, Tight Iso + One τ_h , $p_T > 30$ GeV, $ \eta < 2.1$, Loose DeepTau ID	34.7
2023	One e, $p_T > 30$ GeV, Tight Iso	27.9
	One e, $p_T > 24$ GeV, $ \eta < 2.1$, Tight Iso + One τ_h , $p_T > 30$ GeV, $ \eta < 2.1$, Loose DeepTau ID	27.9
	One e, $p_T > 12$ GeV, $ \eta < 2.1$, Tight Iso + Two jets, $p_T > 45$ GeV, $m_{jj} > 500$ GeV, $\Delta\eta_{jj} > 2.5$	12.3
	One e, $p_T > 17$ GeV, $ \eta < 2.1$, Tight Iso + Two jets, $p_T > 45$ GeV, $m_{jj} > 500$ GeV, $\Delta\eta_{jj} > 2.5$	9.5
	Two jets, $p_T > 105(40)$ GeV, $m_{jj} > 1000$ GeV, $\Delta\eta_{jj} > 3.5$	21.8
2024	One e, $p_T > 30$ GeV, Tight Iso	109.9
	One e, $p_T > 24$ GeV, $ \eta < 2.1$, Tight Iso + One τ_h , $p_T > 30$ GeV, $ \eta < 2.1$, Loose PNet τ_h -tag ID	109.9
	One e, $p_T > 22$ GeV, $ \eta < 2.1$, Tight Iso + Two jets, $p_T > 50$ GeV, $m_{jj} > 600$ GeV, $\Delta\eta_{jj} > 2.5$	109.9
	Two jets, $p_T > 125(45)$ GeV, $m_{jj} > 1050$ GeV	109.9

Trigger selection for ele + τ_h final state

Year	Trigger requirement	Int. lumi [fb ⁻¹]
2022	One μ , $p_T > 24$ GeV, Tracker Iso	34.7
	One μ , $p_T > 20$ GeV, $ \eta < 2.1$ + One τ_h , $p_T > 20$ GeV, Loose DeepTau ID	34.7
2023	One μ , $p_T > 24$ GeV, Tracker Iso	27.9
	One μ , $p_T > 20$ GeV, $ \eta < 2.1$, Tracker Iso + One τ_h , $p_T > 27$ GeV, $ \eta < 2.1$, Loose DeepTau ID	27.9
	One μ , $p_T > 3$ GeV, Loose Tracker Iso + Two jets, $p_T > 90(40)$ GeV, $m_{jj} > 600$ GeV, $\Delta\eta_{jj} > 2.5$	21.8
	Two jets, $p_T > 105(40)$ GeV, $m_{jj} > 1000$ GeV, $\Delta\eta_{jj} > 3.5$	21.8
2024	One μ , $p_T > 24$ GeV, Tracker Iso	109.9
	One μ , $p_T > 20$ GeV, $ \eta < 2.1$, Tracker Iso + One τ_h , $p_T > 27$ GeV, $ \eta < 2.3$, Loose PNet τ_h -tag ID	109.9
	One μ , $p_T > 3$ GeV, Loose Tracker Iso + Two jets, $p_T > 95(45)$ GeV, $m_{jj} > 750$ GeV	109.9
	Two jets, $p_T > 125(45)$ GeV, $m_{jj} > 1050$ GeV	109.9

Trigger selection for mu + τ_h final state

Offline selections

Muon	$p_T > 22/26/6 \text{ GeV}$ for the cross $\mu\text{-}\tau$ /single μ /VBF+ μ trigger $ d_{xy} < 0.045 \text{ cm}, d_z < 0.2 \text{ cm}, \eta < 2.4$ Tight Muon ID, PF relative isolation < 0.15
Electron	$p_T > 25/31/13(18) \text{ GeV}$ for the cross $e\text{-}\tau$ /single e /VBF+ e trigger $ d_{xy} < 0.045 \text{ cm}, d_z < 0.2 \text{ cm}, \eta < 2.5$ Tight electron MVA Isolation+ID
Tau lepton	$\tau_h\tau_h$: double- τ : $p_T > 40(35) \text{ GeV}$, double- τ +jet: $p_T > 35(31) \text{ GeV}$, 2 jets + 1 τ + 1 b-jet : $p_T > 25(25) \text{ GeV}$ $\tau_\mu\tau_h$: $p_T > 32 \text{ GeV}$ $\tau_e\tau_h$: $p_T > 35 \text{ GeV}$ $ d_z < 0.2 \text{ cm}, \eta < 2.5$ DEEPTAU ID working points: D_{jet}^τ Medium, D_e^τ VVLoose, D_μ^τ VLoose/Tight
Boosted- τ	$p_T > 20 \text{ GeV}, \eta < 2.5$ BOOSTED DEEPTAU vsJet score requirement
Lepton pair	opposite electric charge, resolved $\tau\tau$: $\Delta R > 0.5$, boosted- $\tau_h\tau_h$: $\Delta R < 0.8$
Additional	additional lepton veto