UNIVERSITÀ DI SIENA 1240

DEPARTMENT OF PHYSICAL SCIENCES, EARTH AND ENVIRONMENT

Searches for resonant double Higgs production in the $bb\tau\tau$ final state at the CMS experiment

PH.D. IN EXPERIMENTAL PHYSICS

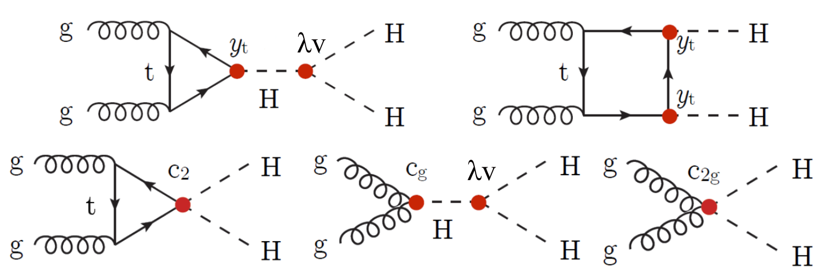
PH.D. CANDIDATE:
Valeria D'Amante
MAT. 106361

SUPERVISOR:
Prof.ssa Maria Agnese Ciocci
TUTOR:
Prof. Alessandro Cerri

22/05/2025

Double Higgs production and decays

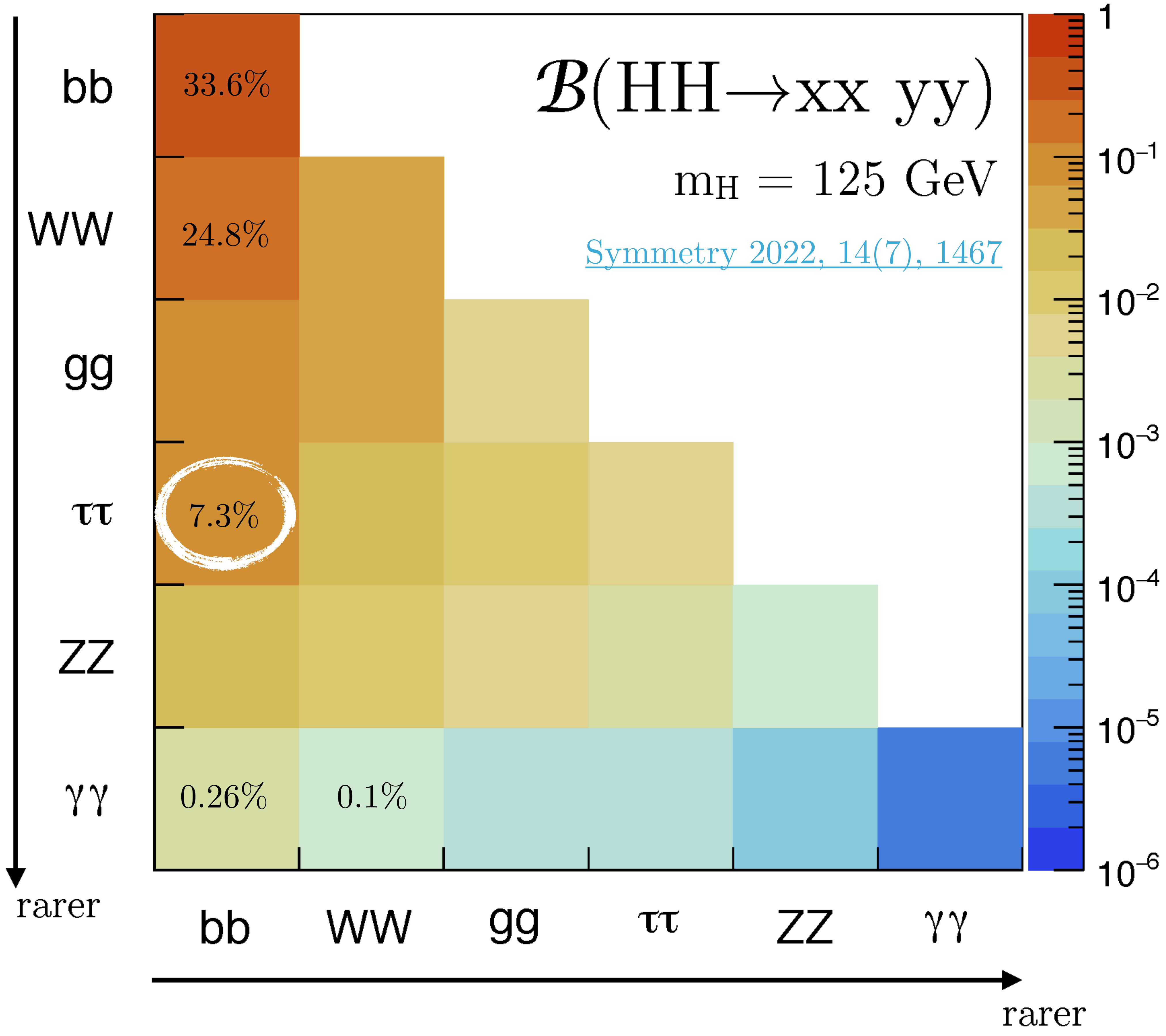
Non-resonant HH production



- Sensitive to λ_{HHH}
- Test non-resonant BSM effective models with **anomalous couplings** κ_λ λ_t κ_V, κ_{VV}
- Affects the kinematic distributions as m_{HH}

Resonant HH production

- Study different BSM models predicting resonances at the TeV scale
- **Warped Extra Dimension (WED) model**, SM Higgs pair produced from decay of heavy particle X
  - Spin 0 (radion) or spin 2 (graviton-bulk)
  - $m_X \in [250, 3000]$ GeV
- *Enhancement in σ_{HH} around its mass*



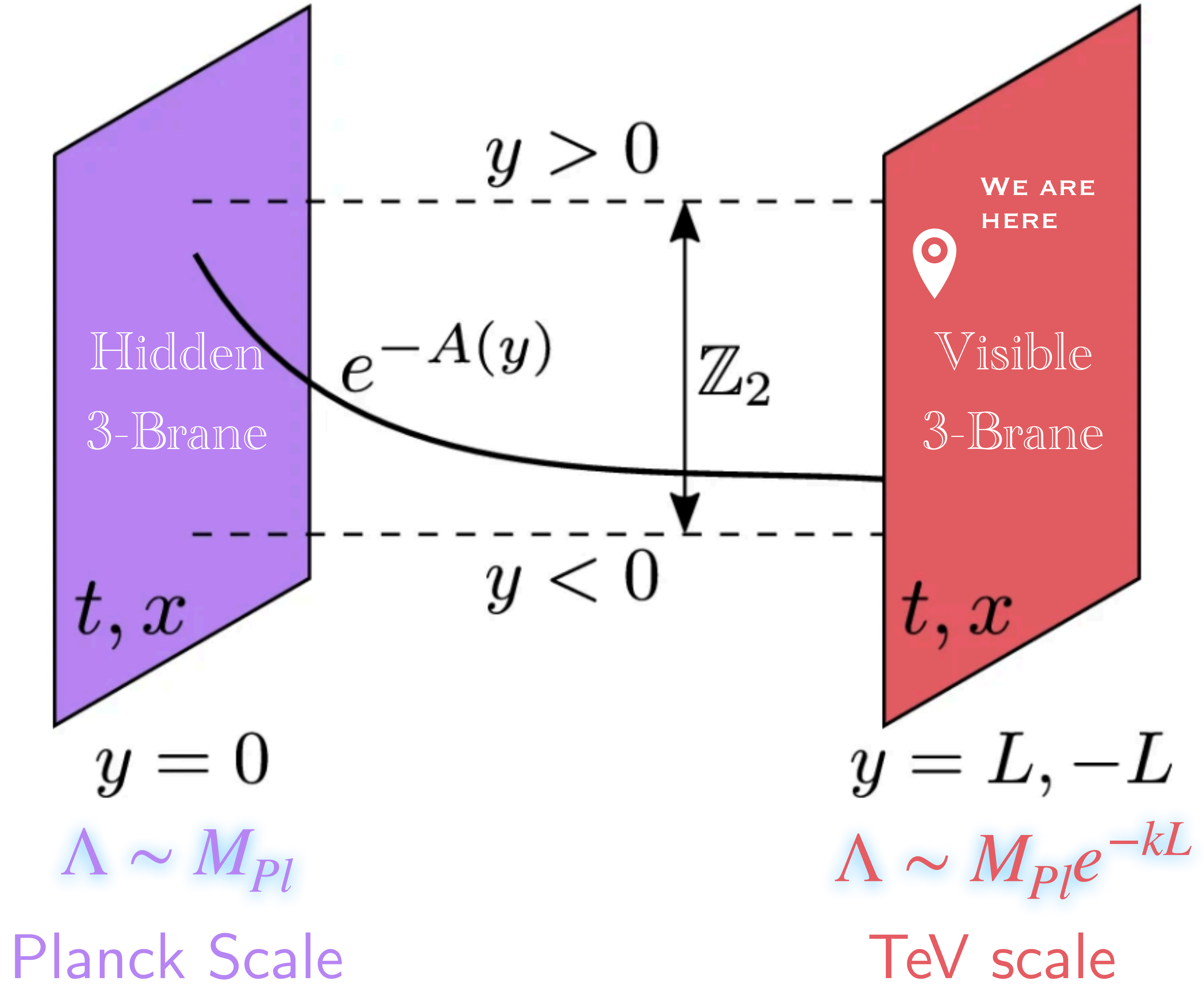
bbττ channel: good trade-off between BR and signal purity

$$\mathscr{B}(HH \to bb\tau\tau) \approx 7.3\,\%$$

Warped Extra Dimensions (WED)

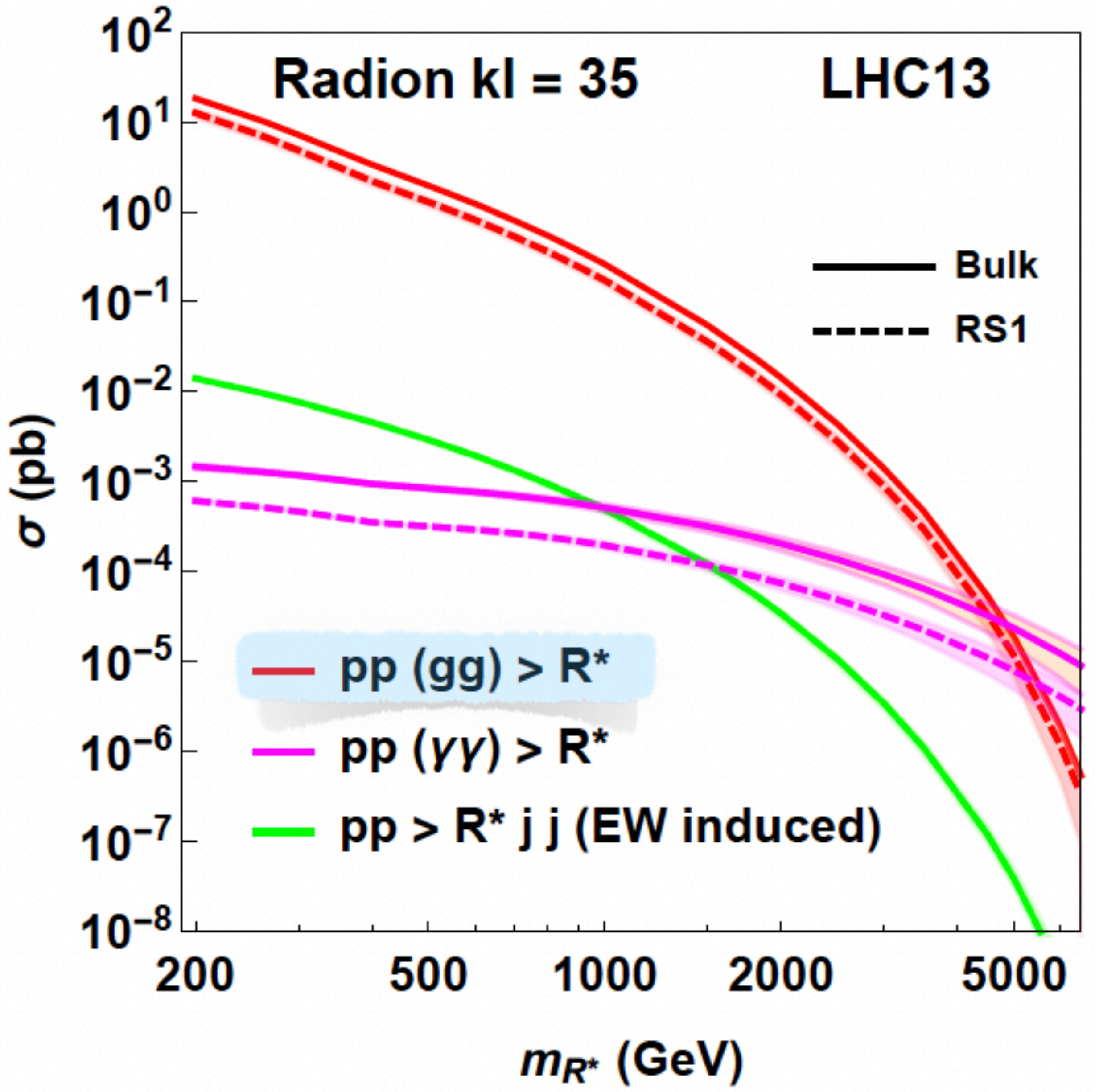
- Introduction of a compact extra dimension with a **warped geometry** connecting two 3-branes
  - Large hierarchy between Planck and electroweak scales is naturally generated by the exponential warp factor
- WED fields emerge from 5D metric excitations and KK decomposition
- Two scenarios:
  - **Randall-Sundrum (RS)**: SM fields propagate only in IR
  - **Bulk**: SM fields allowed to propagate between IR and UV

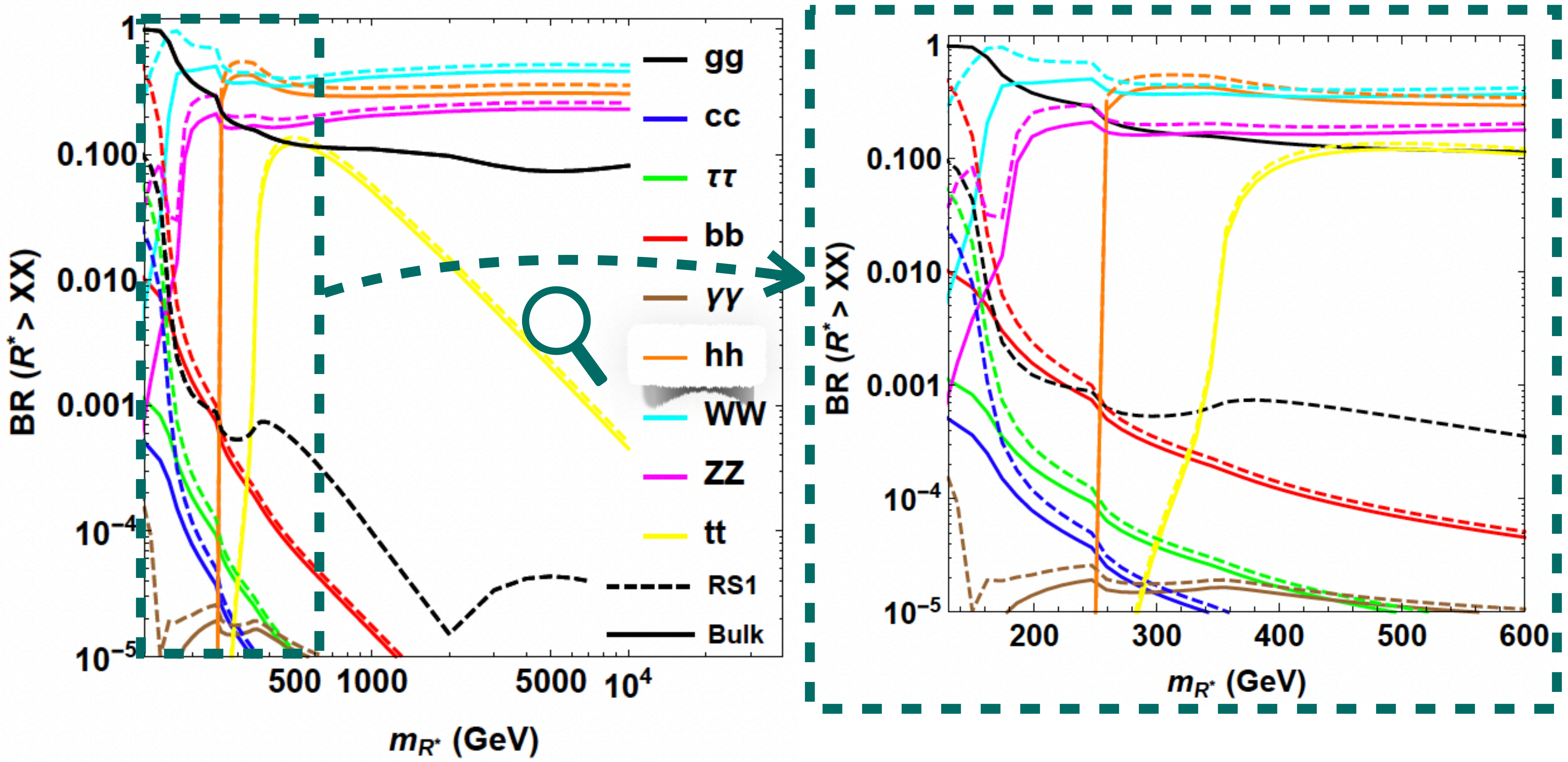
Focus on TeV scale particles predicted:
Radion - spin 0
KK-1 mode Graviton - spin 2



Production and decays

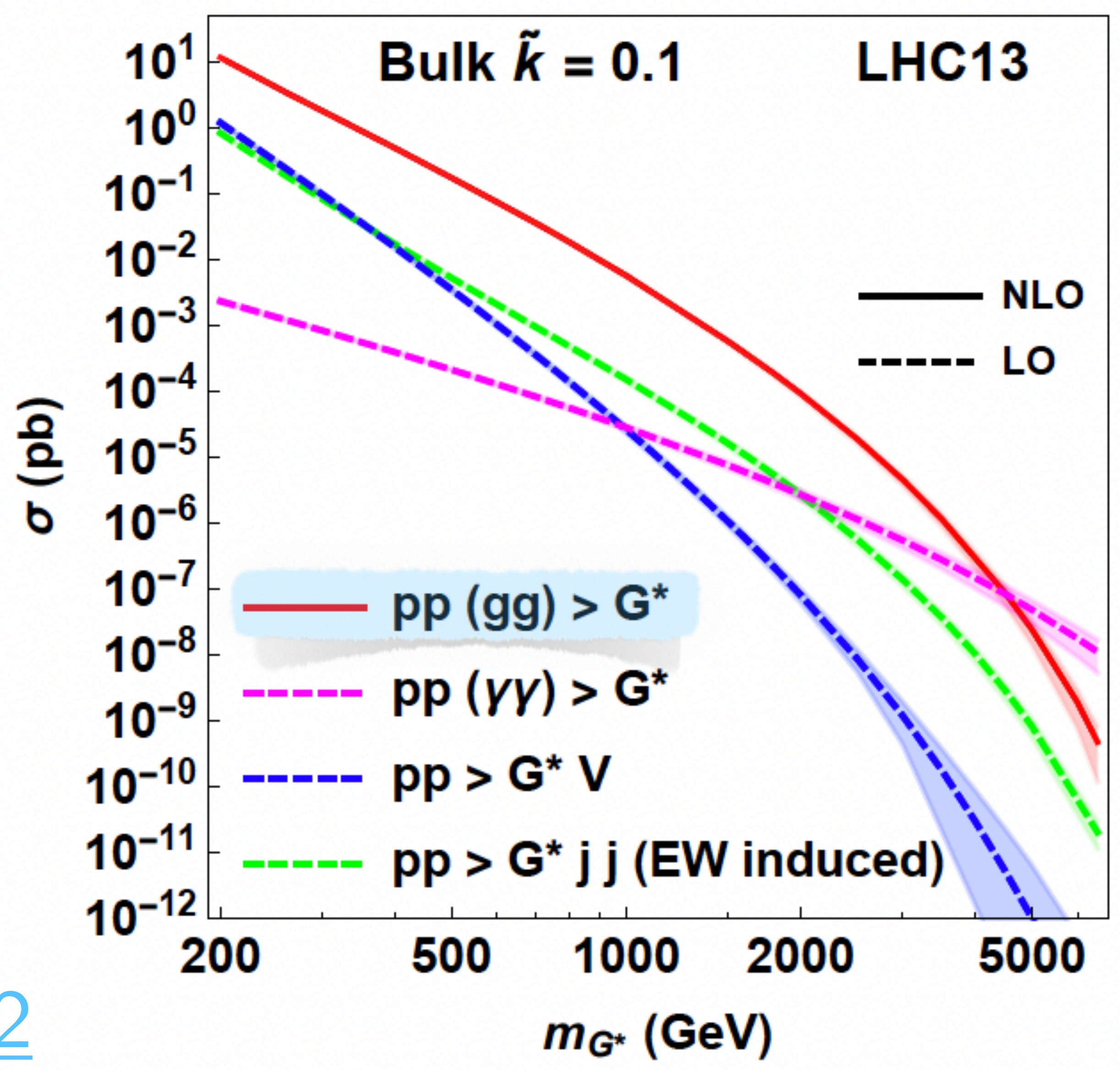
Spin-0
Radion

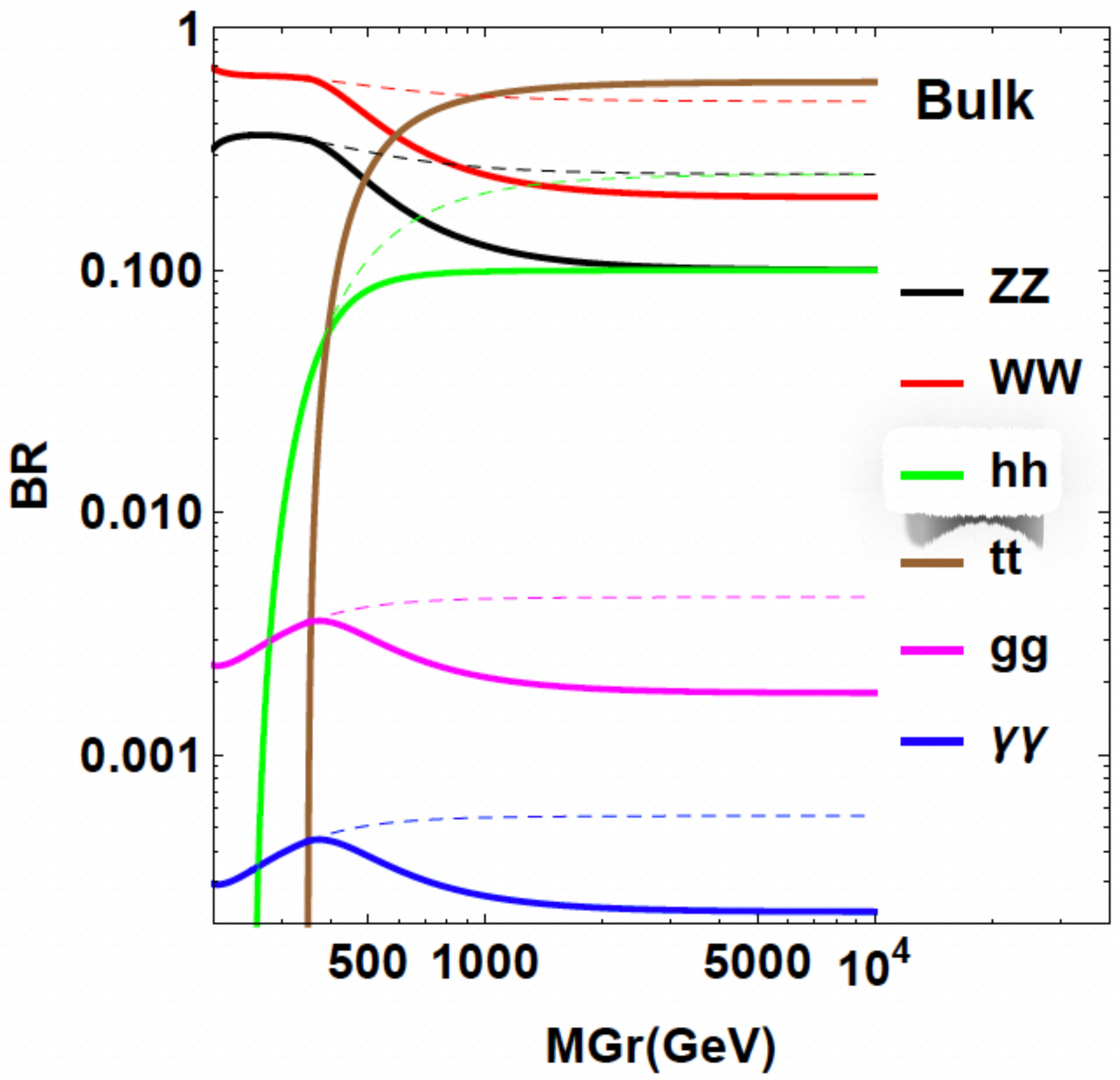




@LHC
$\sqrt{s} = 13$ TeV

Spin-2
Graviton





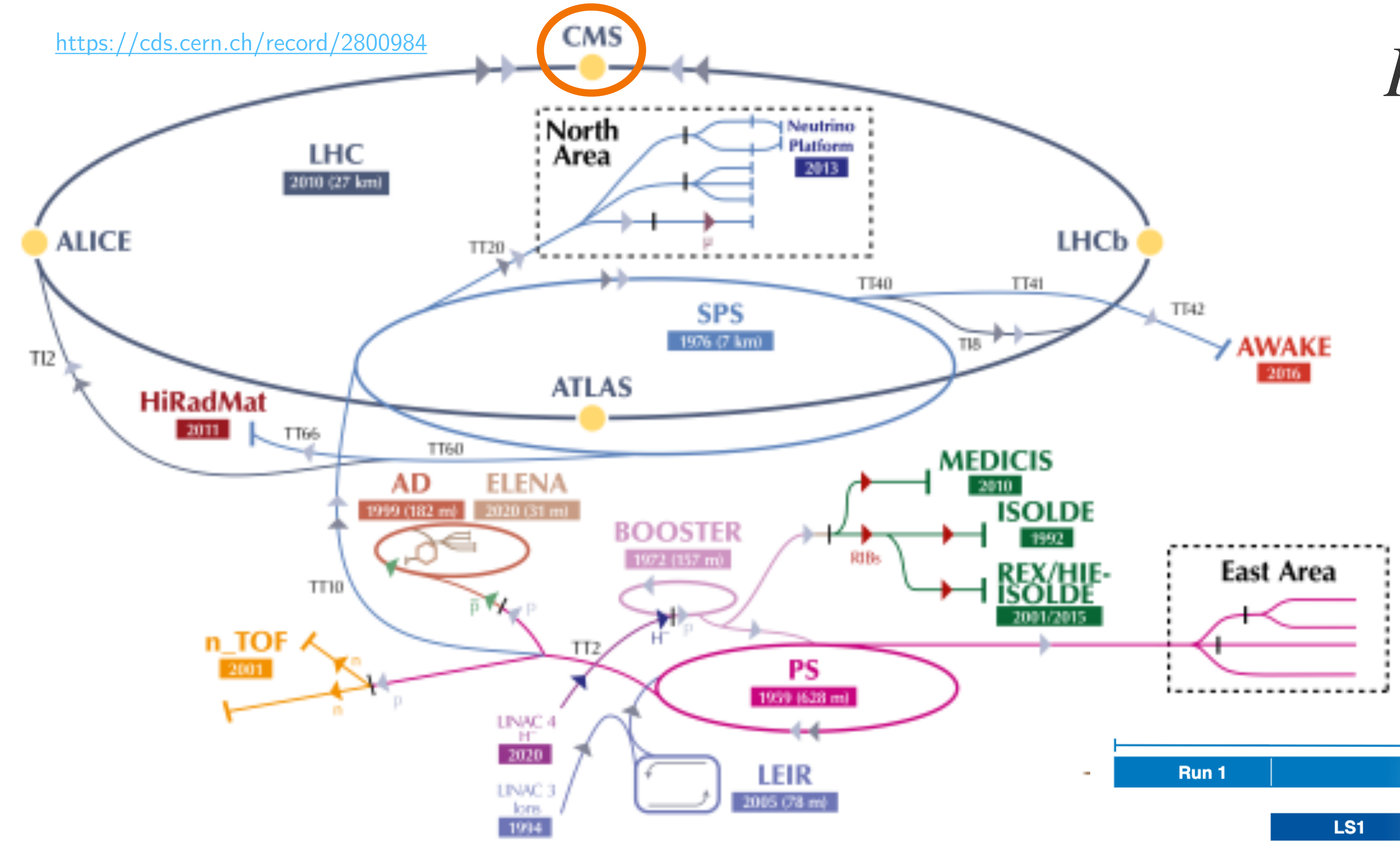
BR(R* → HH) ~ 25%
for large masses

BR(G → HH) can reach up to 25% under certain hypotheses

Source: arXiv:1404.0102

The Large Hadron Collider (LHC)

https://cds.cern.ch/record/2800984



Active since 2009

2012: Higgs boson discovery

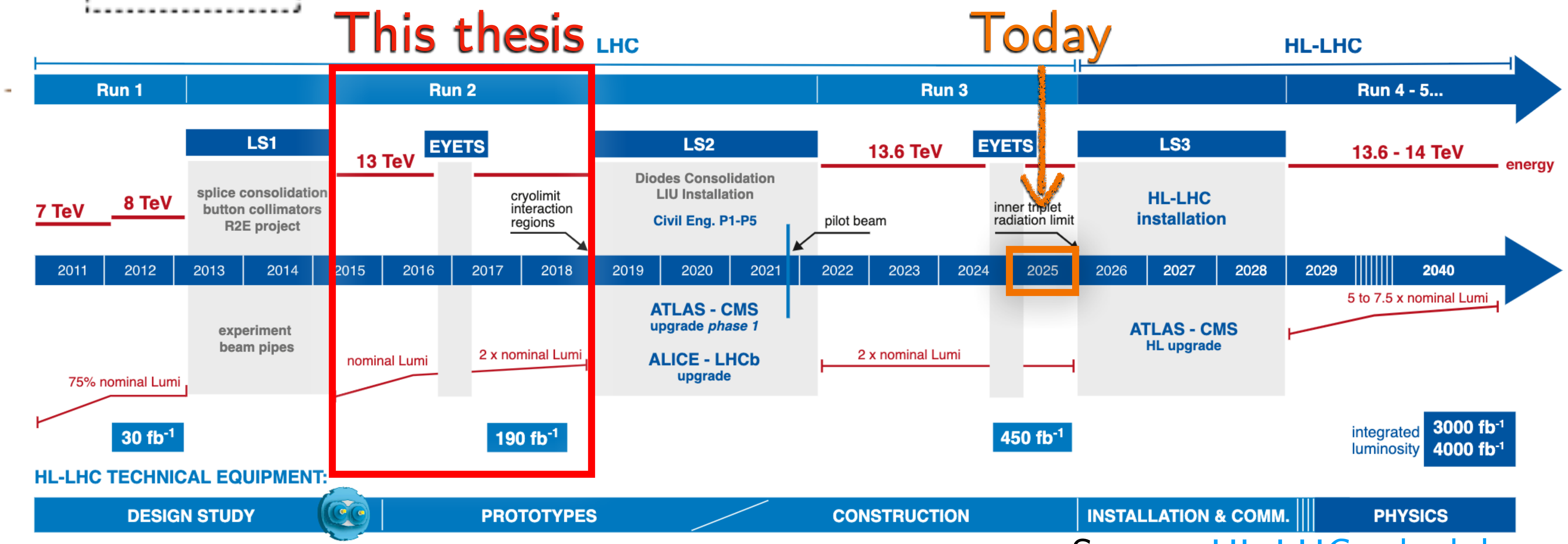
Many experiments (not only LHC ones!)

Long term schedule, high luminosity to collect huge amount of data for subsequent analysis

$$R = \mathscr{L} \cdot \sigma$$

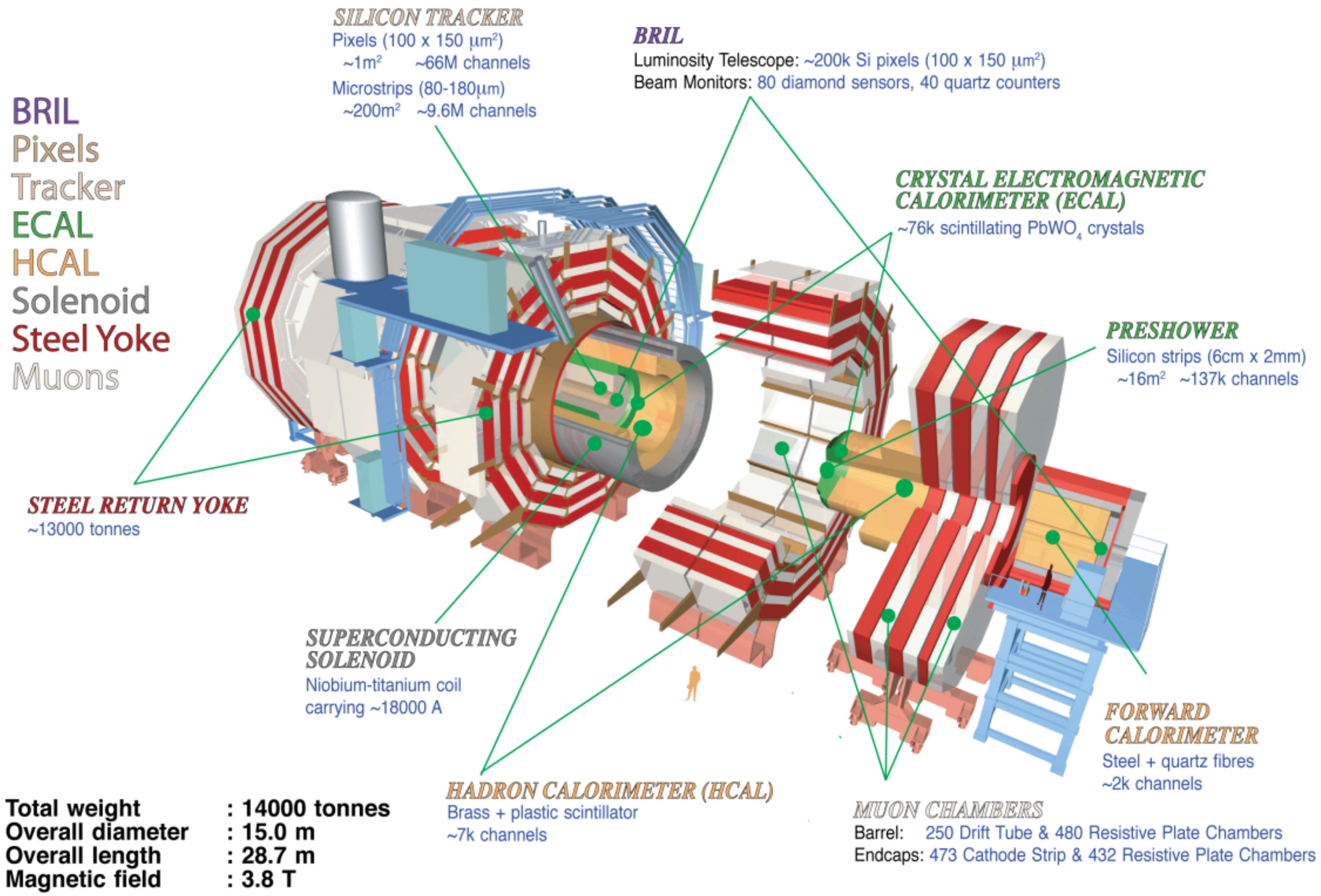
LHC parameters during Run 2

Energy/beam	6.5 TeV
Bunch-cross frequency	40 MHz
σ(pp, inel)	80 mb
Peak inst. luminosity	2 x 10^{34} cm^{-2} s^{-1}
Average Pileup	34

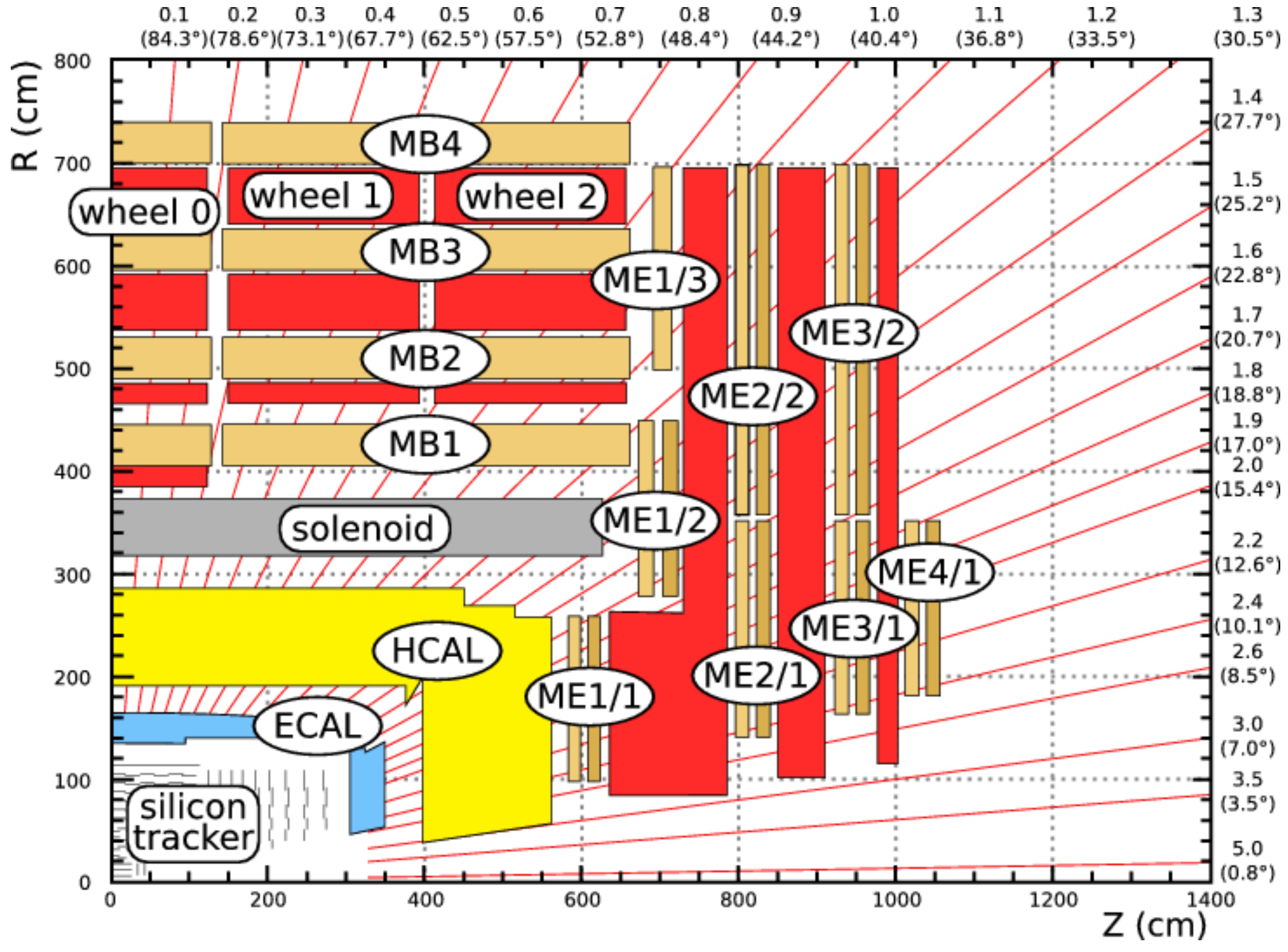


Source: HL-LHC schedule

The Compact Muon Solenoid (CMS) experiment

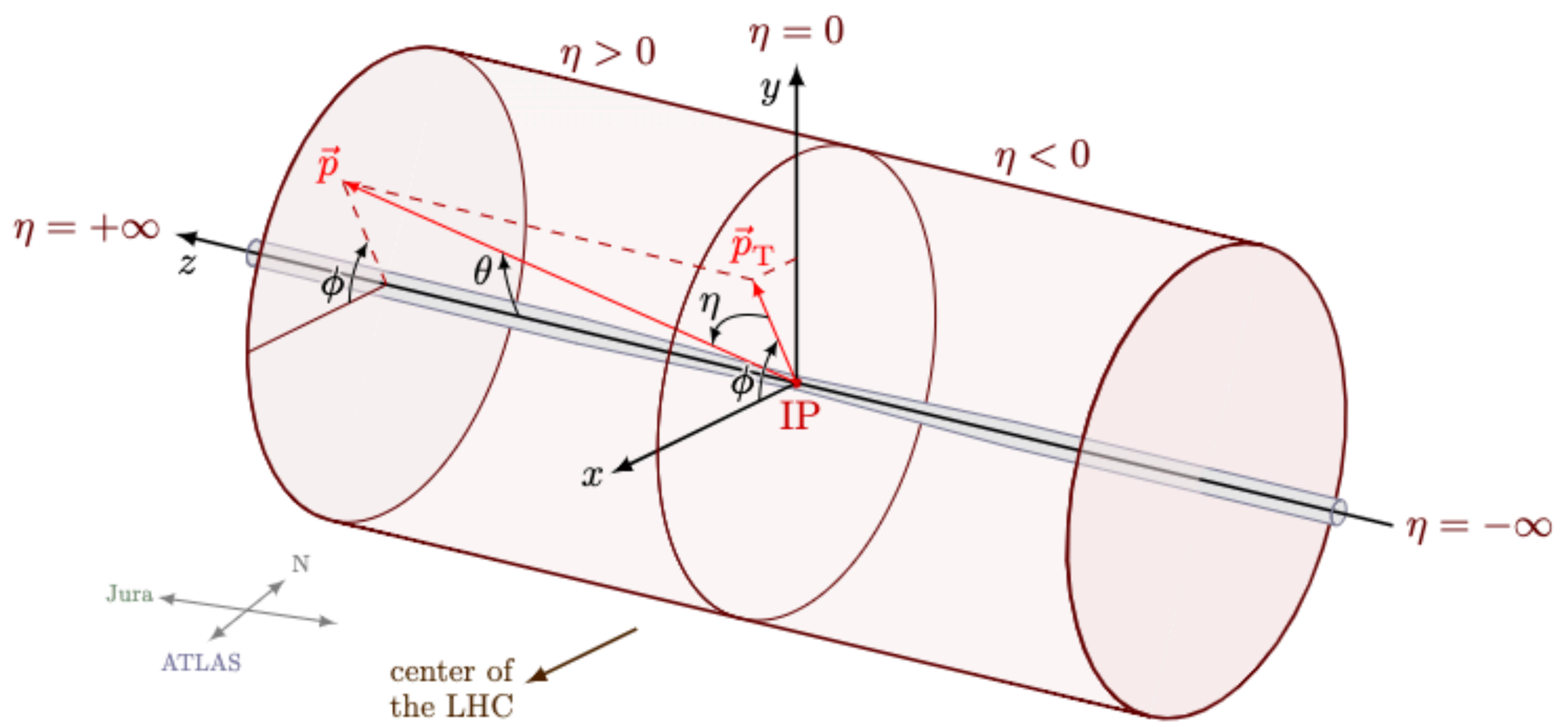


Total weight	: 14000 tonnes
Overall diameter	: 15.0 m
Overall length	: 28.7 m
Magnetic field	: 3.8 T



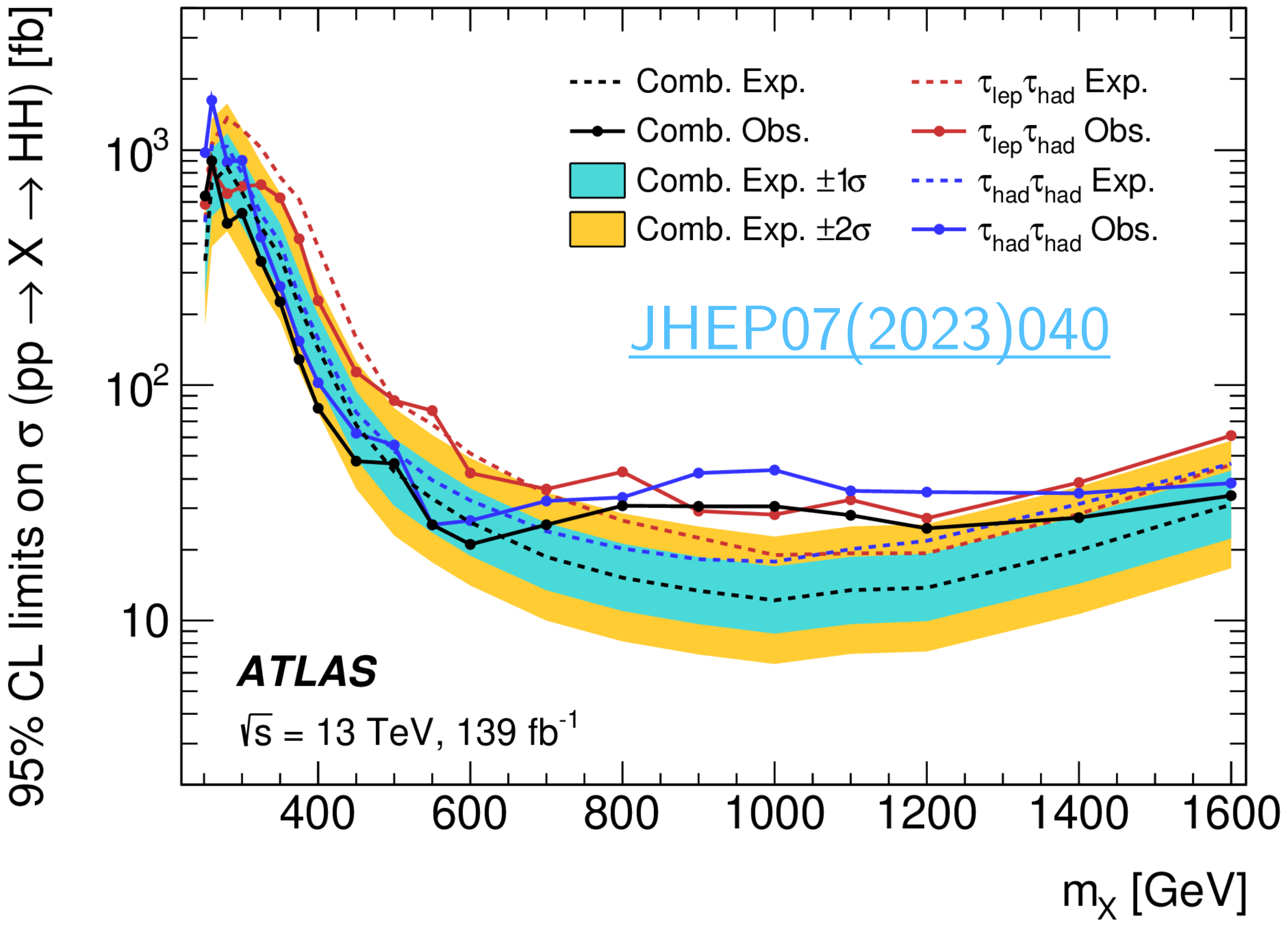
Pseudo-rapidity: $\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$

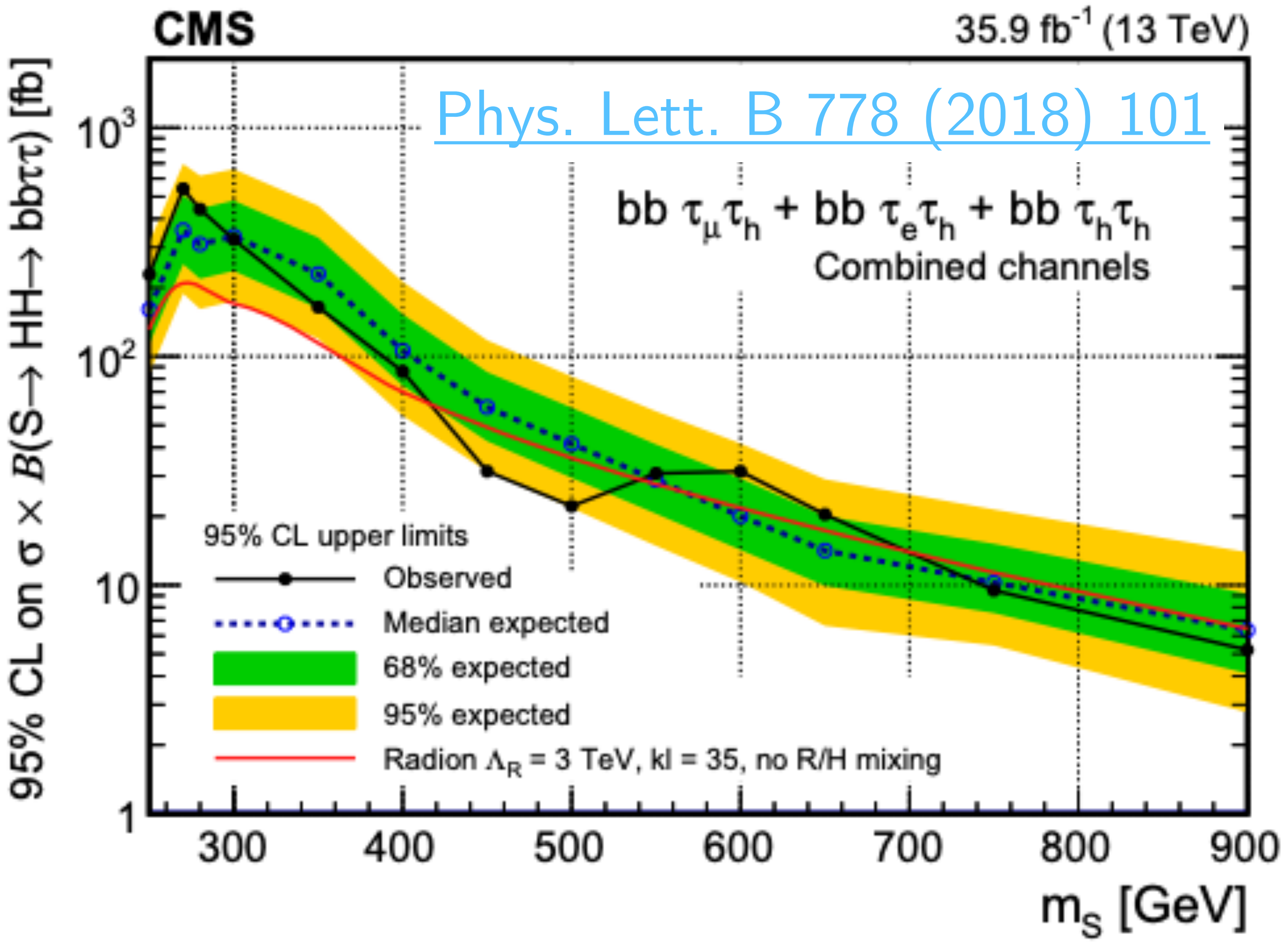
Angular separation: $\Delta R(i,j) = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$



Resonant HH searches in $bb\tau\tau$ channel

State of art - current public results from ATLAS and CMS





This analysis

Full Run 2 data (2016-2018)

2016 (Run B-F) "preVFP" (19.5 fb^{-1})

2016 Run(F-H) "postVFP" (16.8 fb^{-1})

2017 Run(B-F) (41.4 fb^{-1})

2018 Run(A-D) (59.8 fb^{-1})

Total integrated luminosity: 138 fb^{-1}

Reprocessed samples

Increased statistics

Solid strategy

New data (+MC) reprocessing
Improved ECAL calibration
→ better energy resolution

New (and more) advanced particle reconstruction and ID techniques

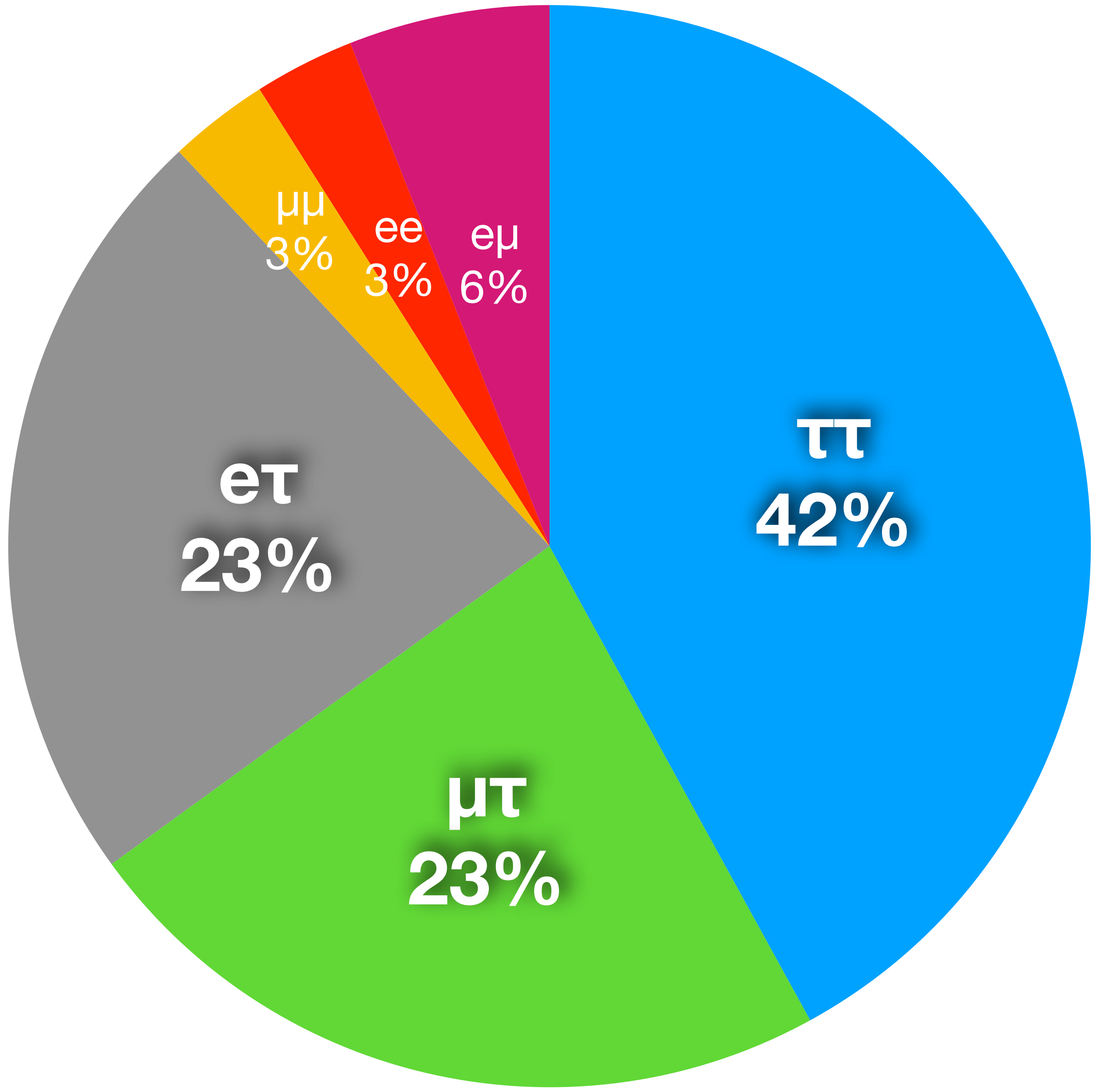
Part of methodology inerithed from non-resonant $HHbb\tau\tau$ analysis

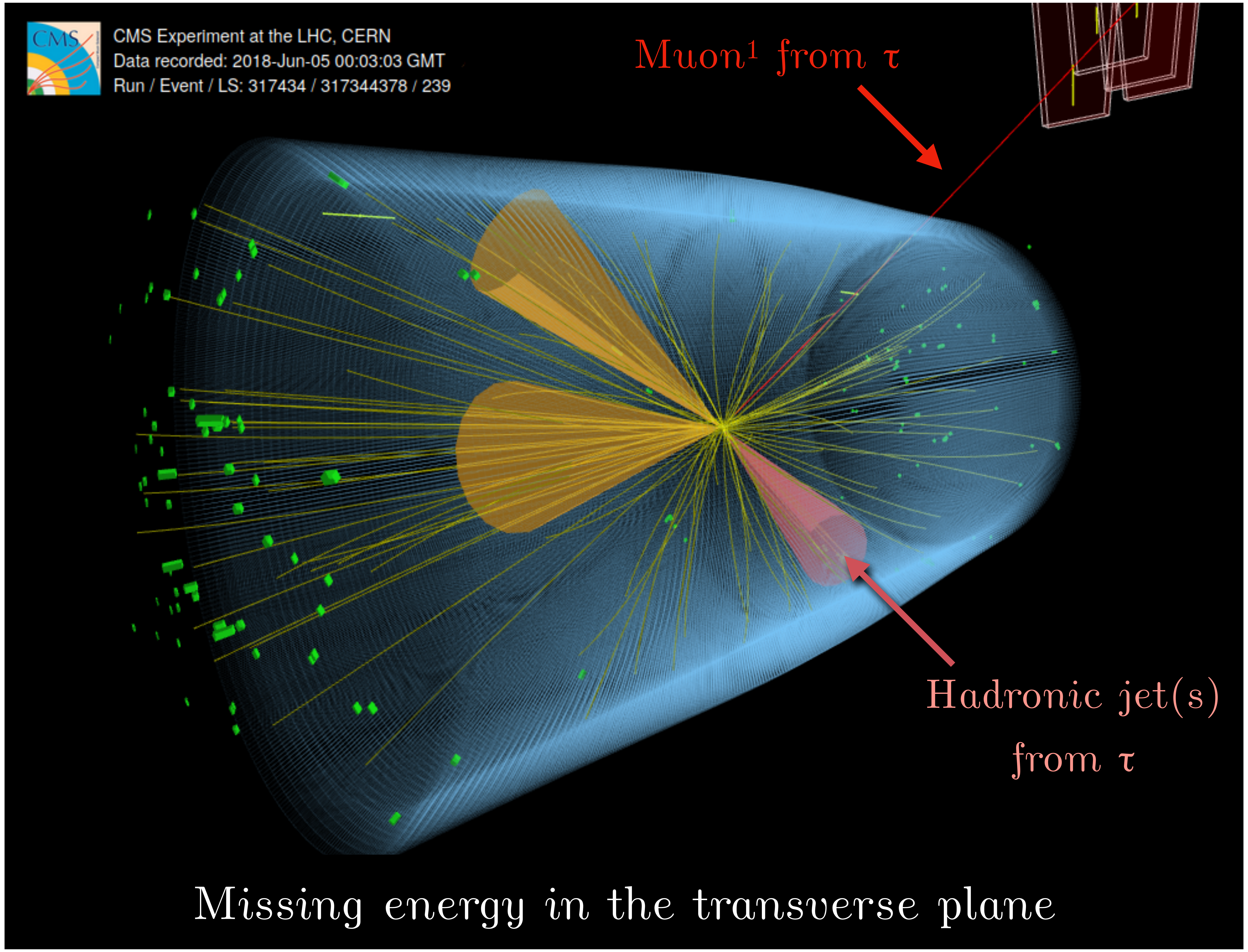
The $bb\tau\tau$ final state

All H($\tau\tau$) final states ("channels") reconstructed

Channels used for signal extraction:

$e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$ covering $\sim 88\,\%$ of $\tau\tau$ decays





[1]or electron or jet from hadronic τ, depending on the channel

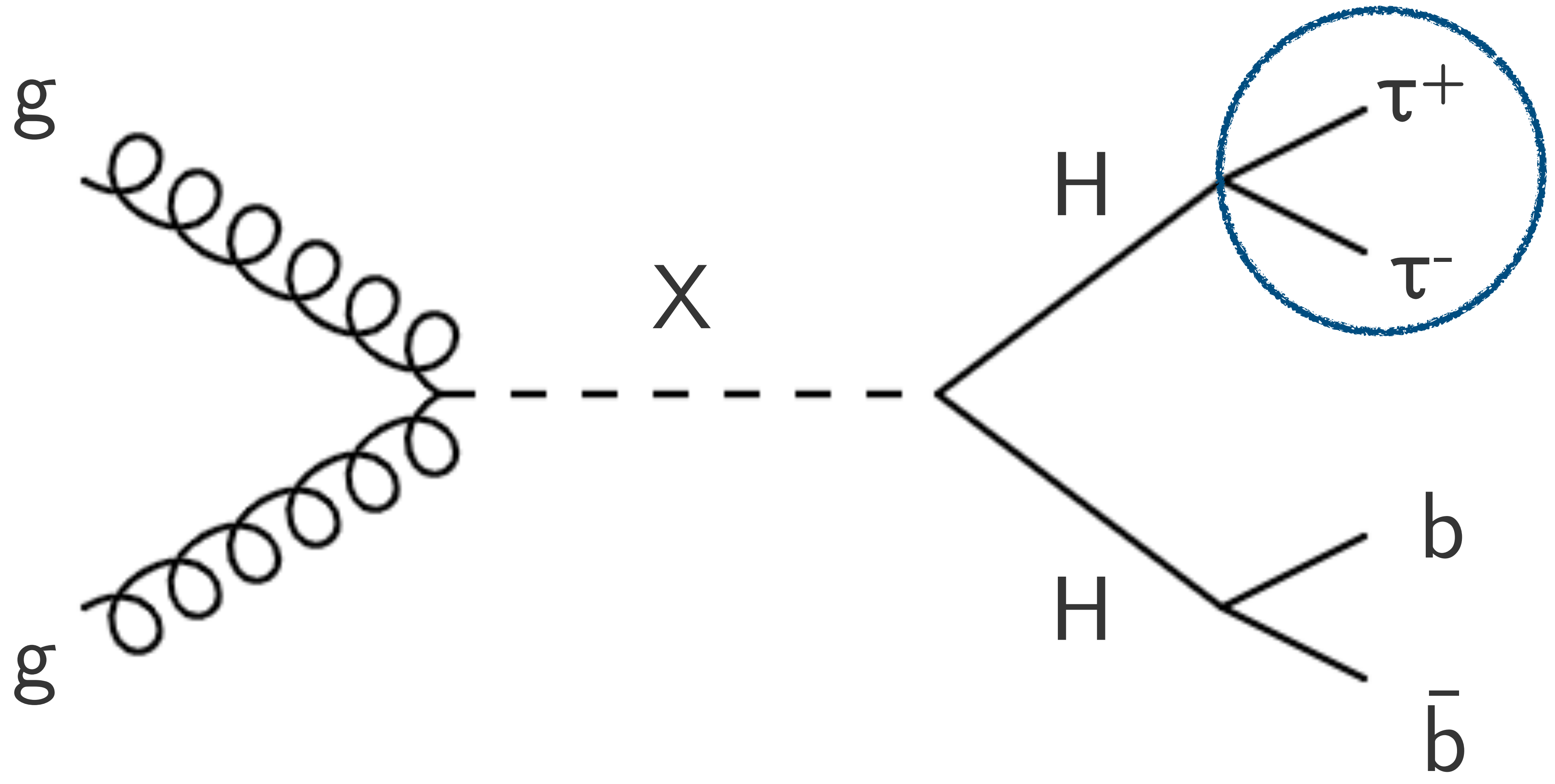
The bbττ final state

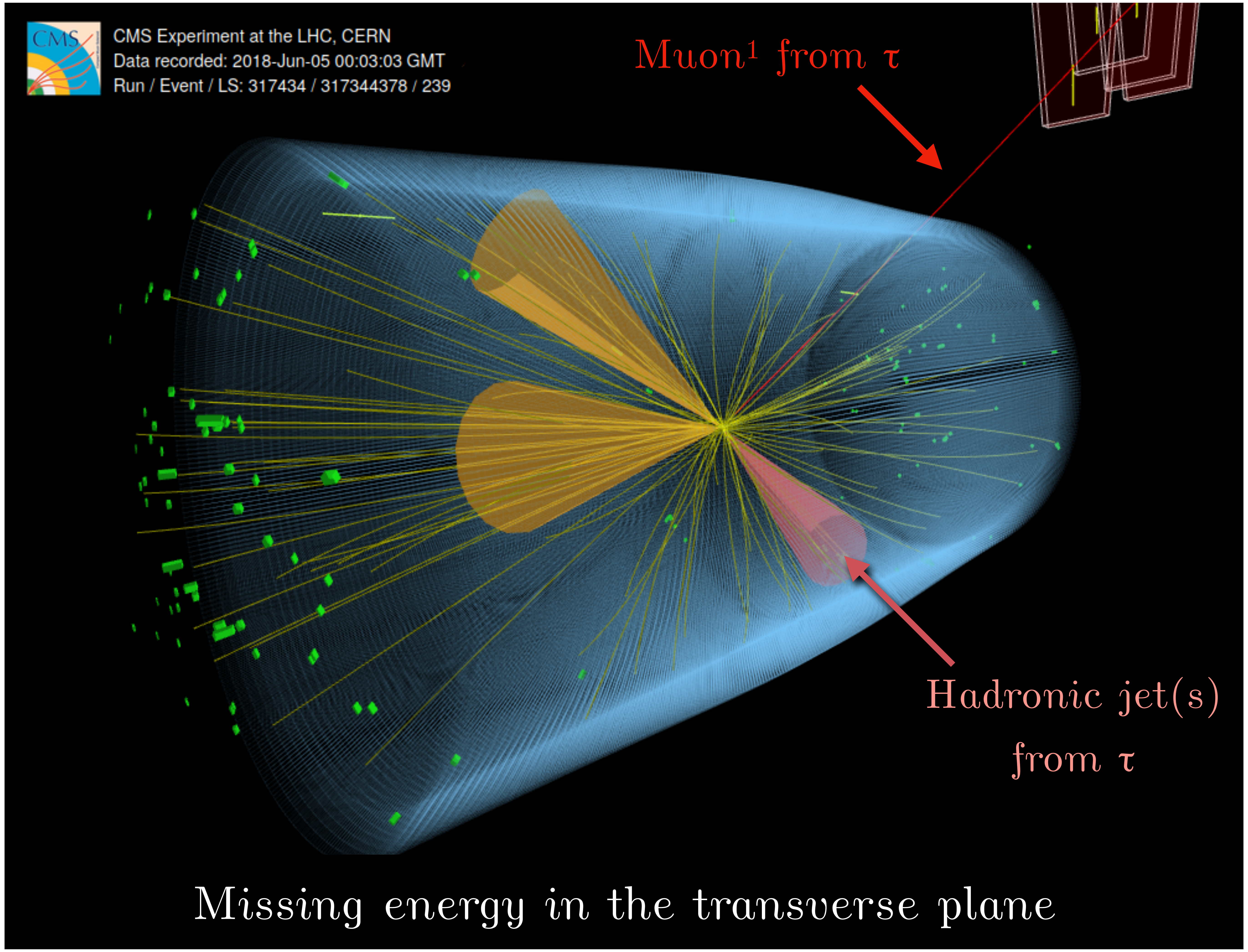
$H \to \tau^+\tau^-$ reconstruction and selection

Trigger targets: (e, μ, τ_h, MET)

Offline objects reco & ID

H(ττ) pair assessment





[1]or electron or jet from hadronic τ, depending on the channel

Triggers

All triggers lists and performance in backup

Online requirements & Offline selection: **validity regions**

Online-Offline matching within ΔR=0.4

Trigger	**Channel**
Single electron	$e\tau_h$, ee, eμ
cross $e\tau_h$	$e\tau_h$
Single Muon	$\mu\tau_h$, $\mu\mu$, eμ
cross $\mu\tau_h$	$\mu\tau_h$
di-τ_h	$\tau_h\tau_h$
single τ_h (high-pt)	$e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$
MET	$e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$

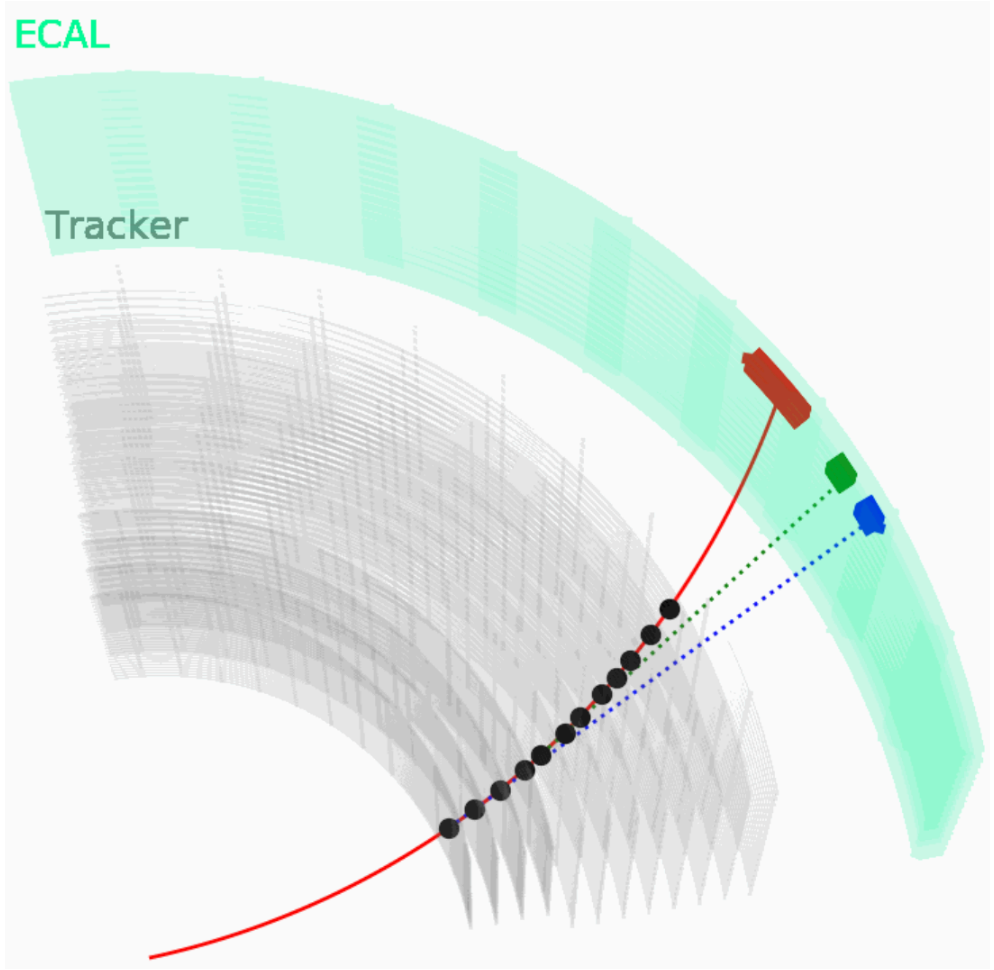
	$p_T(\ell_1)$ (GeV)	$\lvert\eta(\ell_1)\rvert$		$p_T(\ell_2)$ (GeV)	$\lvert\eta(\ell_2)\rvert$
SingleEle	p_T (e)	2.5 (2016) 2.1 (2017, 2018)	OR	p_T (e)	2.5 (2016) 2.1 (2017, 2018)
CrossEleTau	> 25	< 2.1	AND	> 35	< 2.1
SingleMu	$> p_T$ (μ)	< 2.4	OR	$> p_T$ (μ)	< 2.4
CrossMuTau	> 22	< 2.1	AND	> 25	< 2.1
Di-tau	> 40	<2.1	AND	> 40	<2.1
SingleTau	$> p_T$ (τ)	< 2.1	OR	$> p_T$ (τ)	< 2.1
MET	MET-no$\mu > p_T$ (MET)				

p_T (GeV)	2018	2017	2016
e	33	33	26
μ	26	29	26
τ	190	190	130
MET	150	150	130

The τ lepton reco and ID

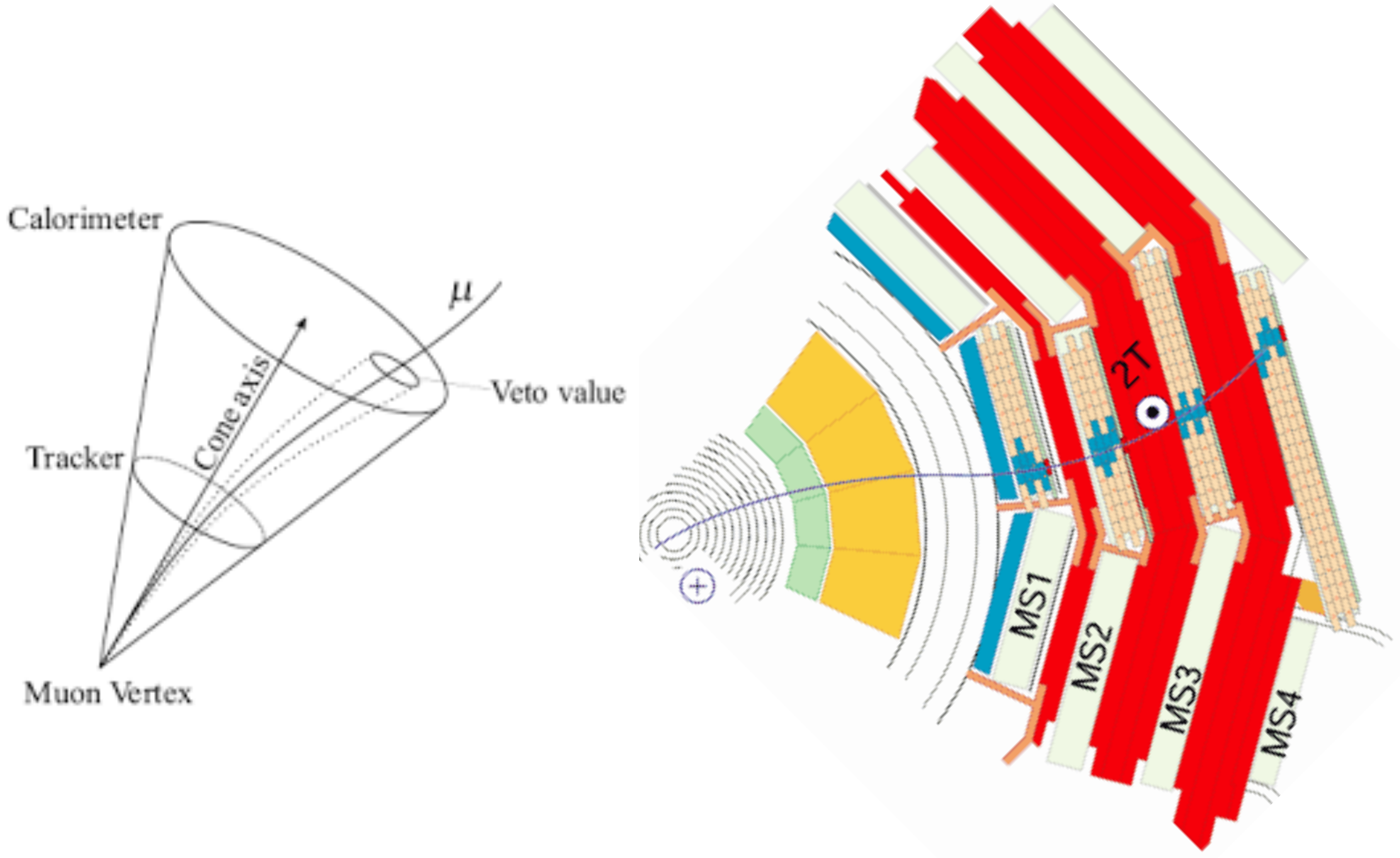
$\tau \to e\bar{\nu}_e\nu_\tau$ (τe) - **electron**

- Reco: Trk + ECAL
- ID: MVA



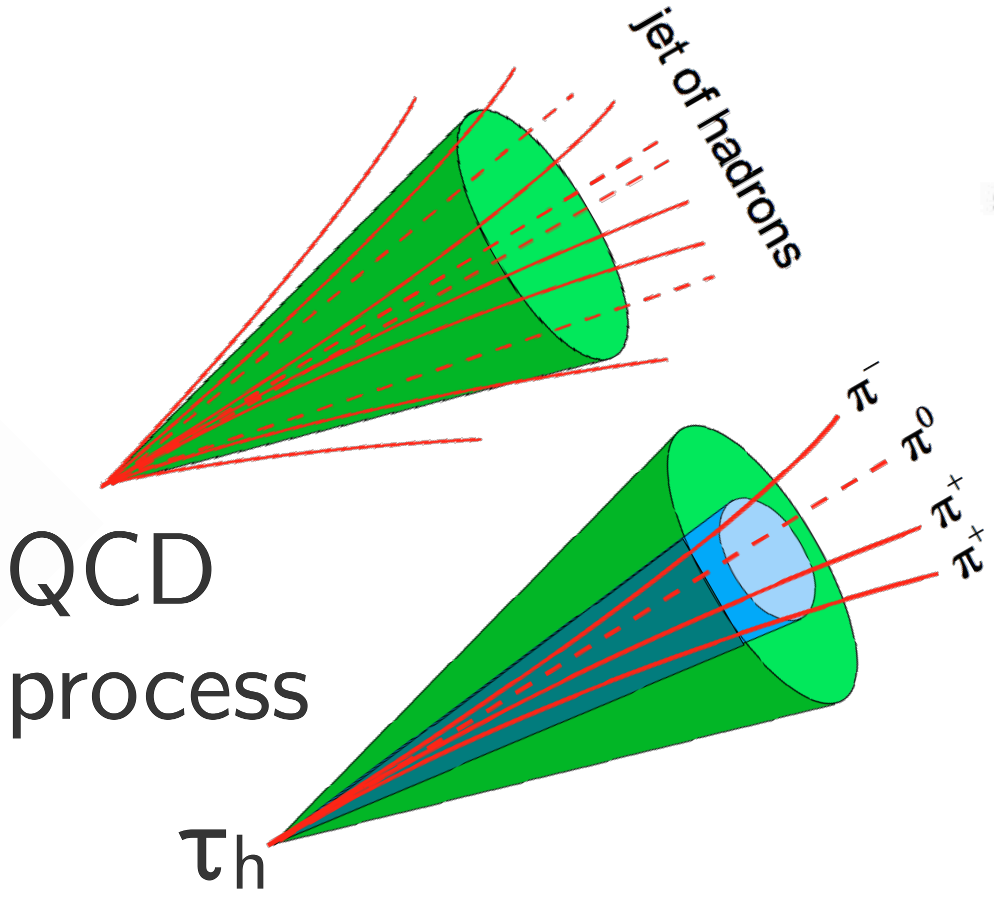
$\tau \to \mu\bar{\nu}_\mu\nu_\tau$ (τμ) - **muon**

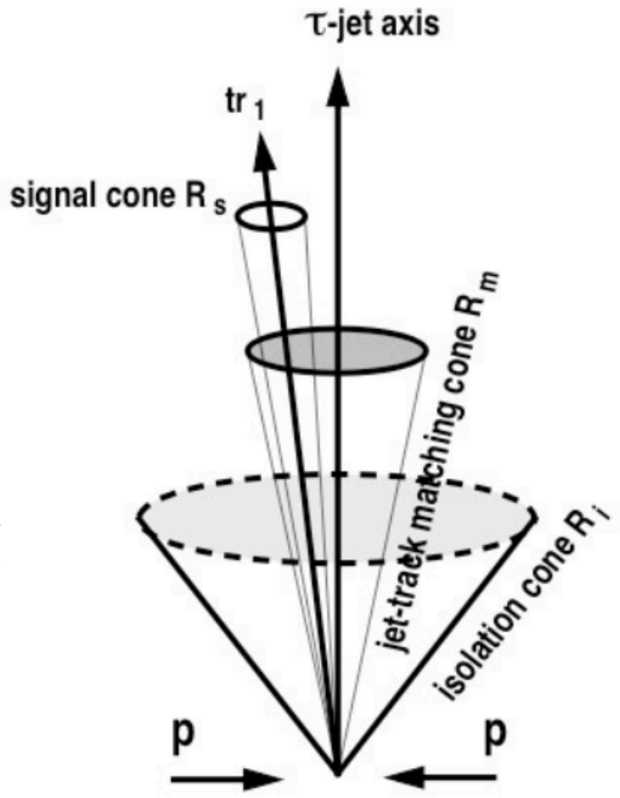
- Reco: Trk + Muon system
- ID: Cut based



$\tau \to \nu_\tau$ + hadrons (τh) - **jet**

- Reco: HPS algorithm
- ID: **DeepTau**





In all cases: **MET**

$$\vec{E\!\!\!/}_T^{miss} = -\sum_{i}^{N_{vis}} \vec{p}_{T,i}$$

Algorithms description and performance in backup

H(ττ) candidate selection

- General requirements on leptons

	p_T (GeV)	$\|\eta\|$	$\|d_z\|$	$\|d_{xy}\|$	Iso	ID
e	> 10	< 2.5	< 0.2	< 0.045	PF RelIso03 < 0.3	MVAIsoWP80 OR MVANoIsoWP80
μ	> 15	< 2.4	< 0.2	< 0.045	PF RelIso03 < 0.3	TightID
τ_h	> 20	2.3	< 0.2	-	VSjet $\geq$ VVVLoose WP	VSe $\geq$ VLoose WP VSmu $\geq$ Tight WP

- Trigger validity regions, in slide 10
- H(ττ) pair assessment inherited from H(ττ) analysis, in backup
- Channel-driven selection

Channel	Lepton number	Iso	ID
$e\tau_h$	1 (e)	PF RelIso03 < 0.15	MVAIsoWP80
	2 (τ_h)	-	
$\mu\tau_h$	1 (μ)	PF RelIso03 < 0.15	-
	2 (τ_h)	-	
$\tau_h\tau_h$	1 (τ_h)	VSJet $\geq$ Medium	-
	2 (τ_h)	-	
ee	1 (e)	PF RelIso03 < 0.15	MVAIsoWP80
	2 (e)	-	MVANoIsoWP80
$\mu\mu$	1 (μ)	PF RelIso03 < 0.15	-
	2 (μ)	-	
$e\mu$	1 (e)	-	
	2 (μ)	-	

- Third lepton veto

	p_T (GeV)	$\|\eta\|$	$\|d_z\|$	$\|d_{xy}\|$	ID
e	> 10	< 2.5	< 0.2	< 0.045	MVAIsoWP90
μ	> 10	< 2.4	< 0.2	< 0.045	(TightID OR MediumID)

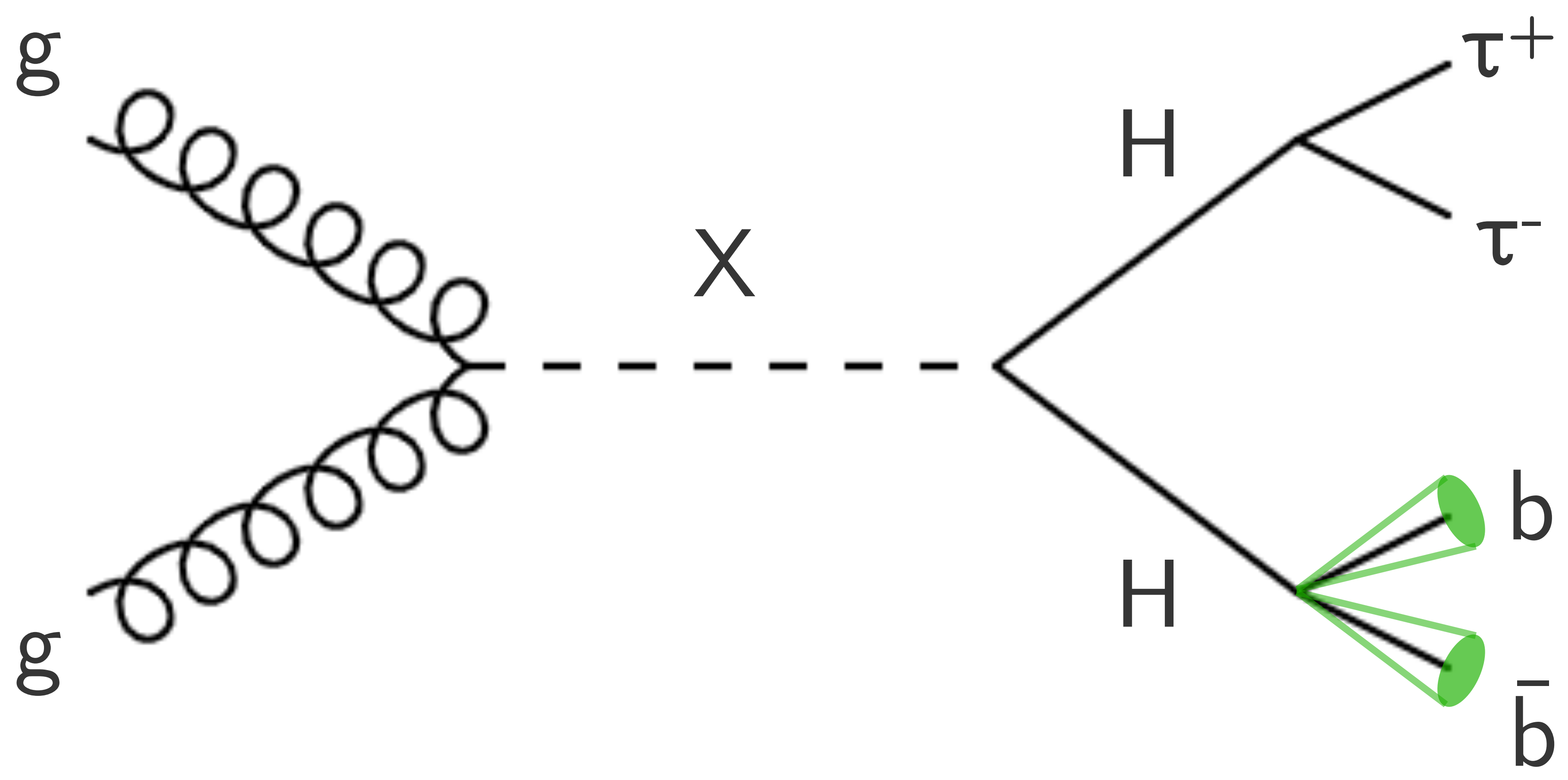
The bbττ final state

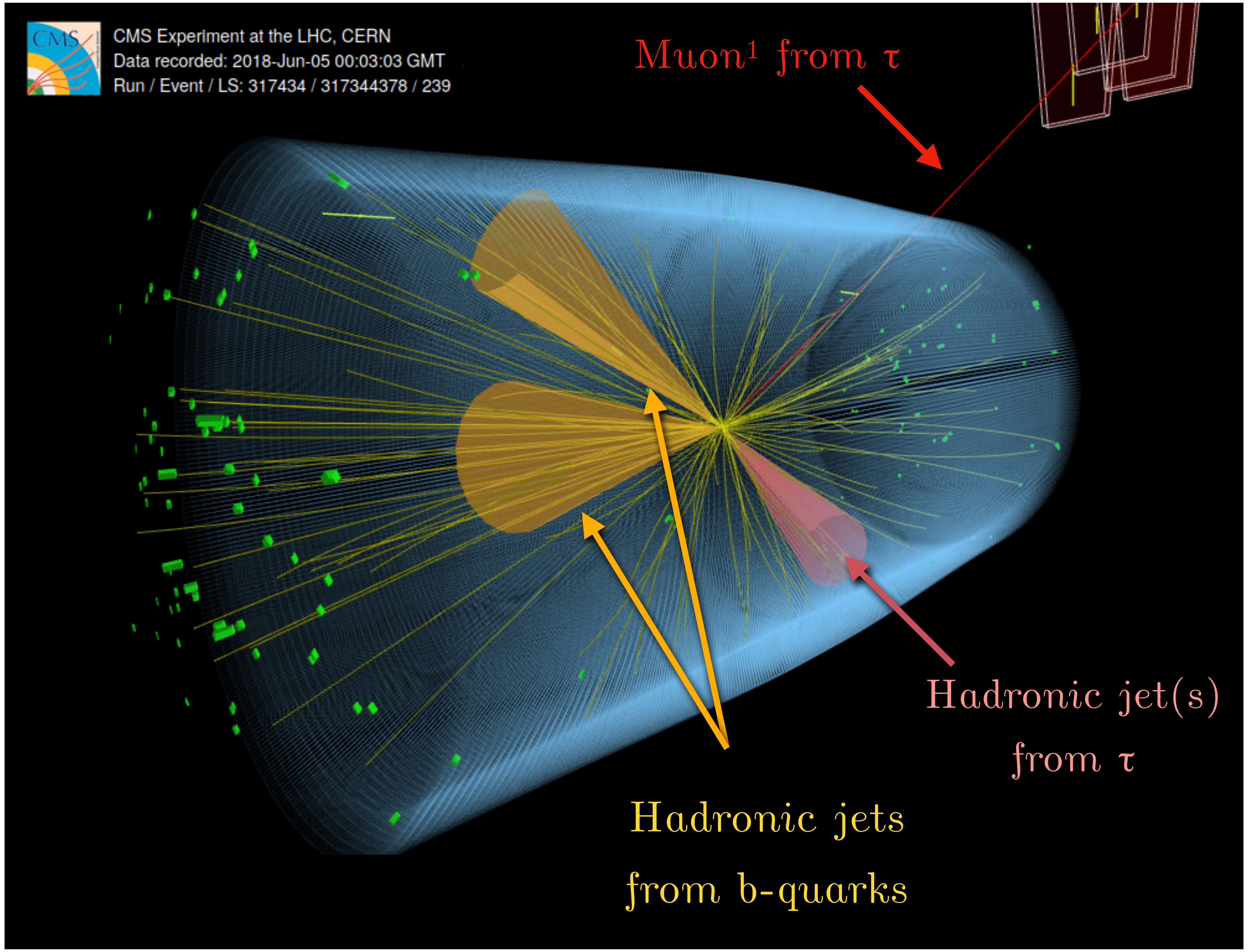
$H \to b\bar{b}$ reconstruction and selection

Jet reco: AK algorithm

Resolved "AK4" jets selection

H(bb) pair assessment: HH-bTag





[1]or electron or jet from hadronic τ, depending on the channel

Resolved H(bb) candidate

- 2 resolved jets (AK4) passing this selection →

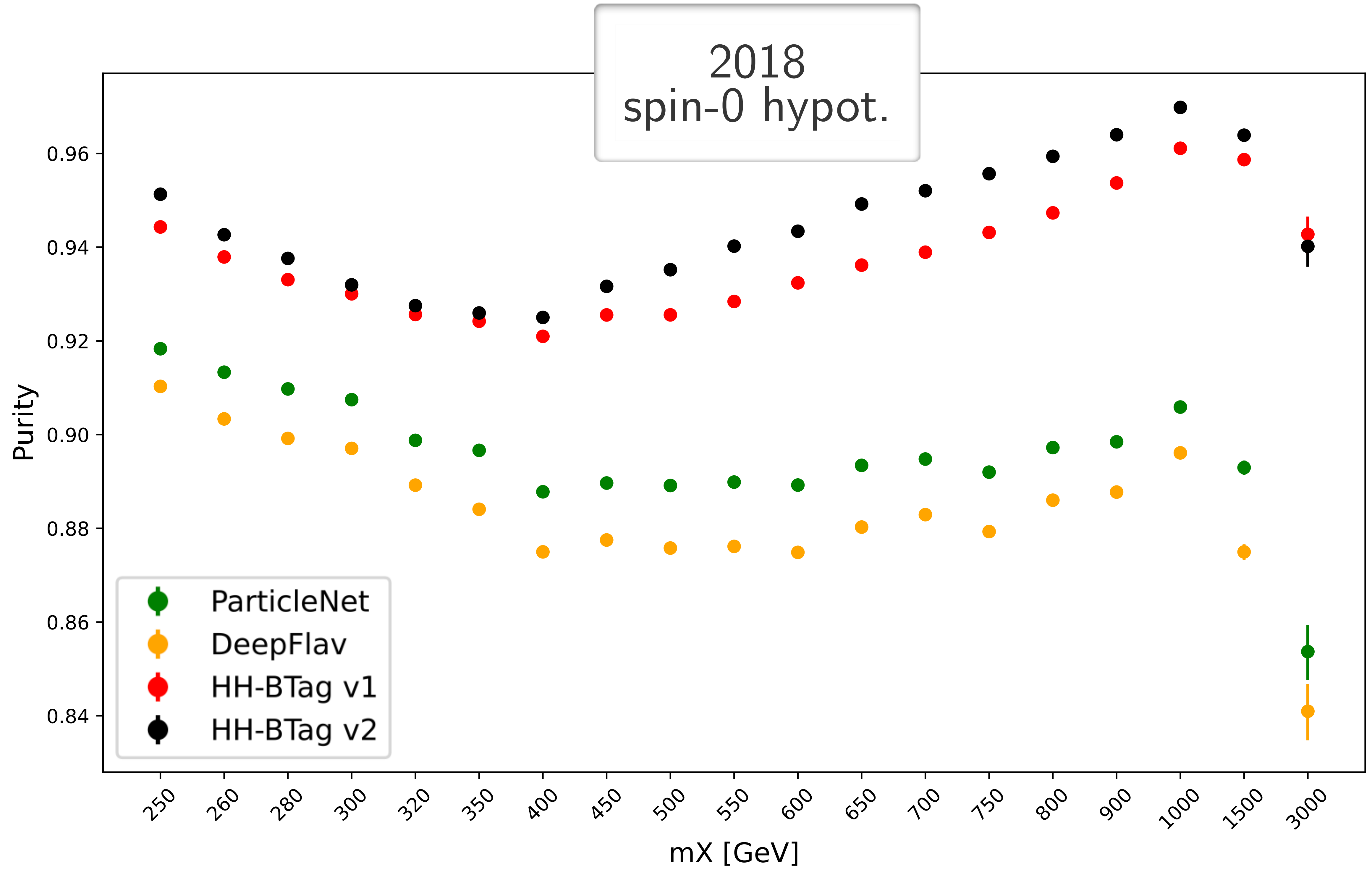
	p_T (GeV)	$\|\eta\|$	ΔR(jet,leps)	Other
AK4 jets	> 20	< 2.5	≥ 0.5	JetID = 2 p_T >50 GeV OR PuJetID
AK8 jets	> 250	< 2.5	≥ 0.8	$m_{softdrop}$ > 30 GeV

- **HH-bTag**: DNN trained for identifying H(bb) final state
  - Introduced for non-res analysis, retrained for resonant
  - Best performances w.r.t. prev. training and other similar algorithms

$$\text{purity} = \frac{N^{\text{match}}\ (\text{2 jets from H(bb) correctly selected})}{N^{\text{match}}}$$

  - N^{match} = N(gen-reco matched obj)

- **H(bb) candidate: 2 jets with highest HH-bTag score**



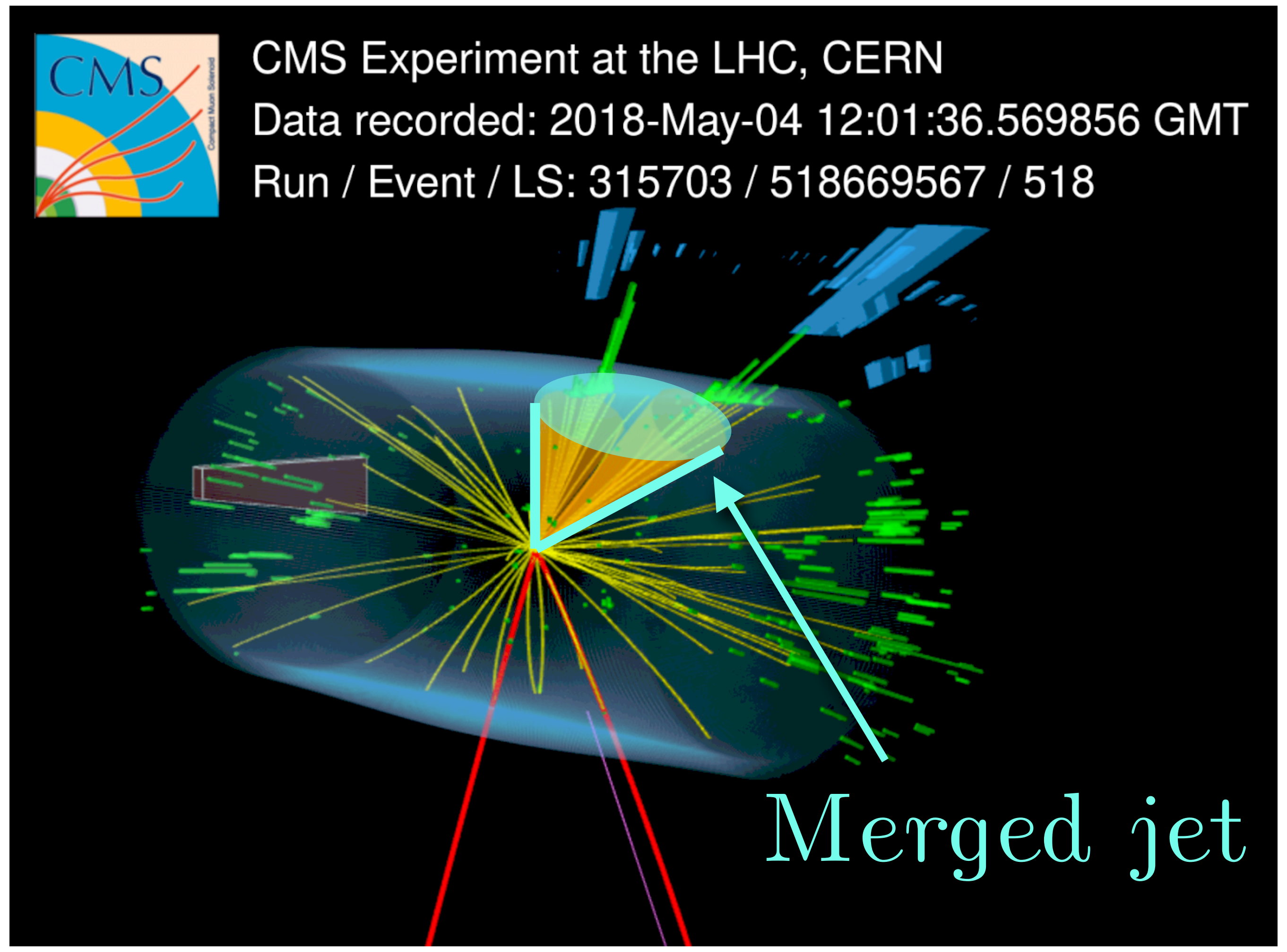
The bbττ final state

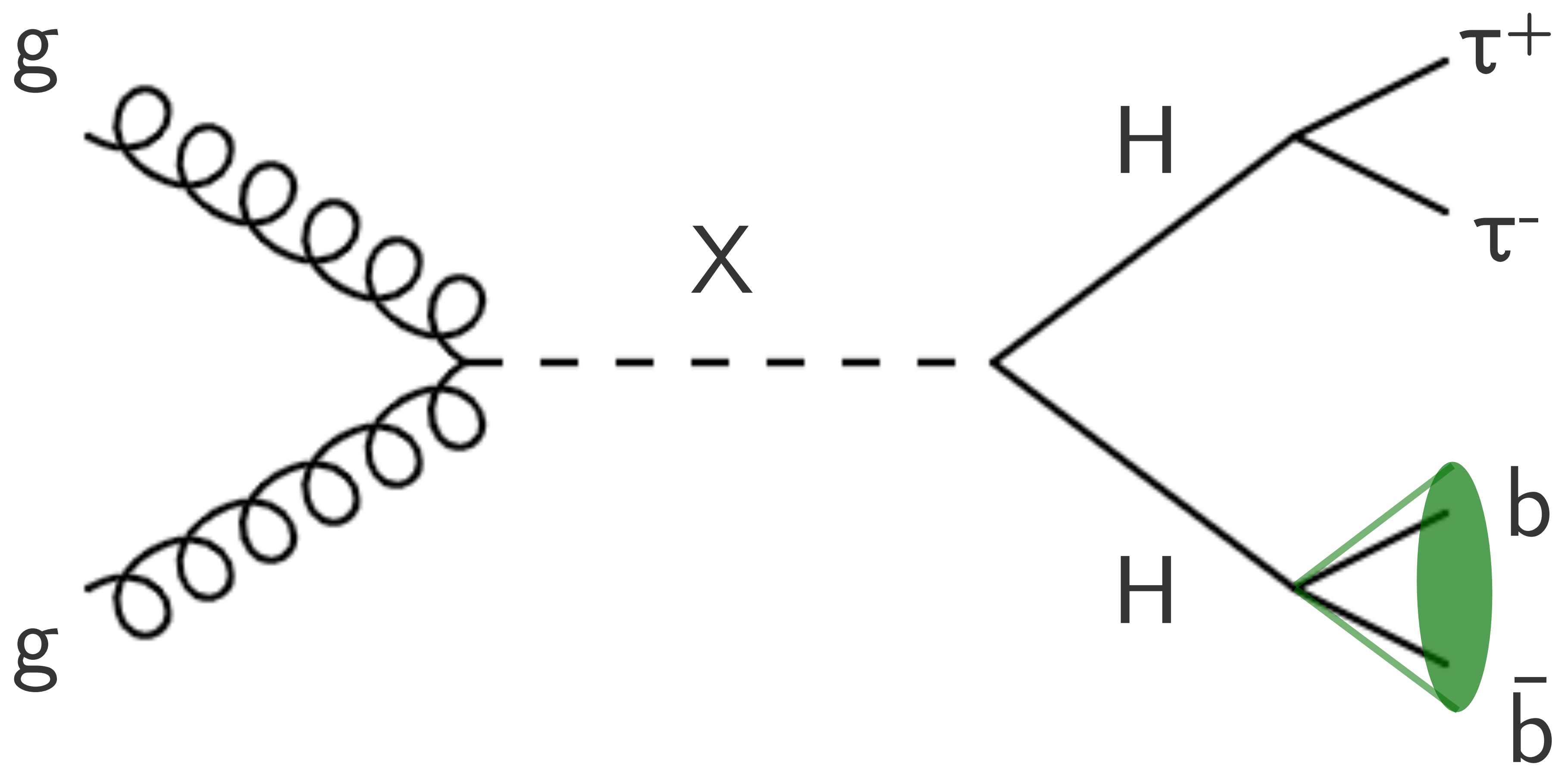
H → b$\bar{b}$ reconstruction and selection

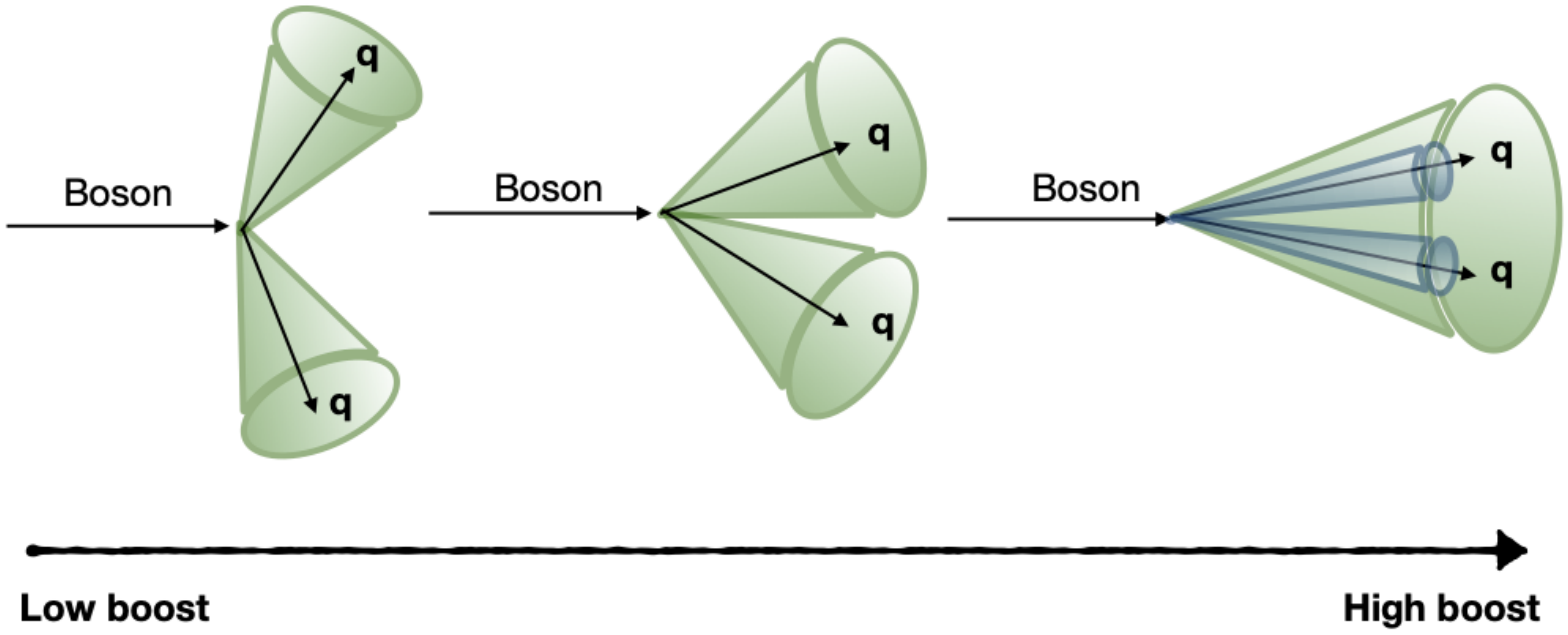
Jet reco: AK algorithm

Boosted "AK8" jet selection

H(bb) boosted candidate: pNet







Boosted H(bb) candidate

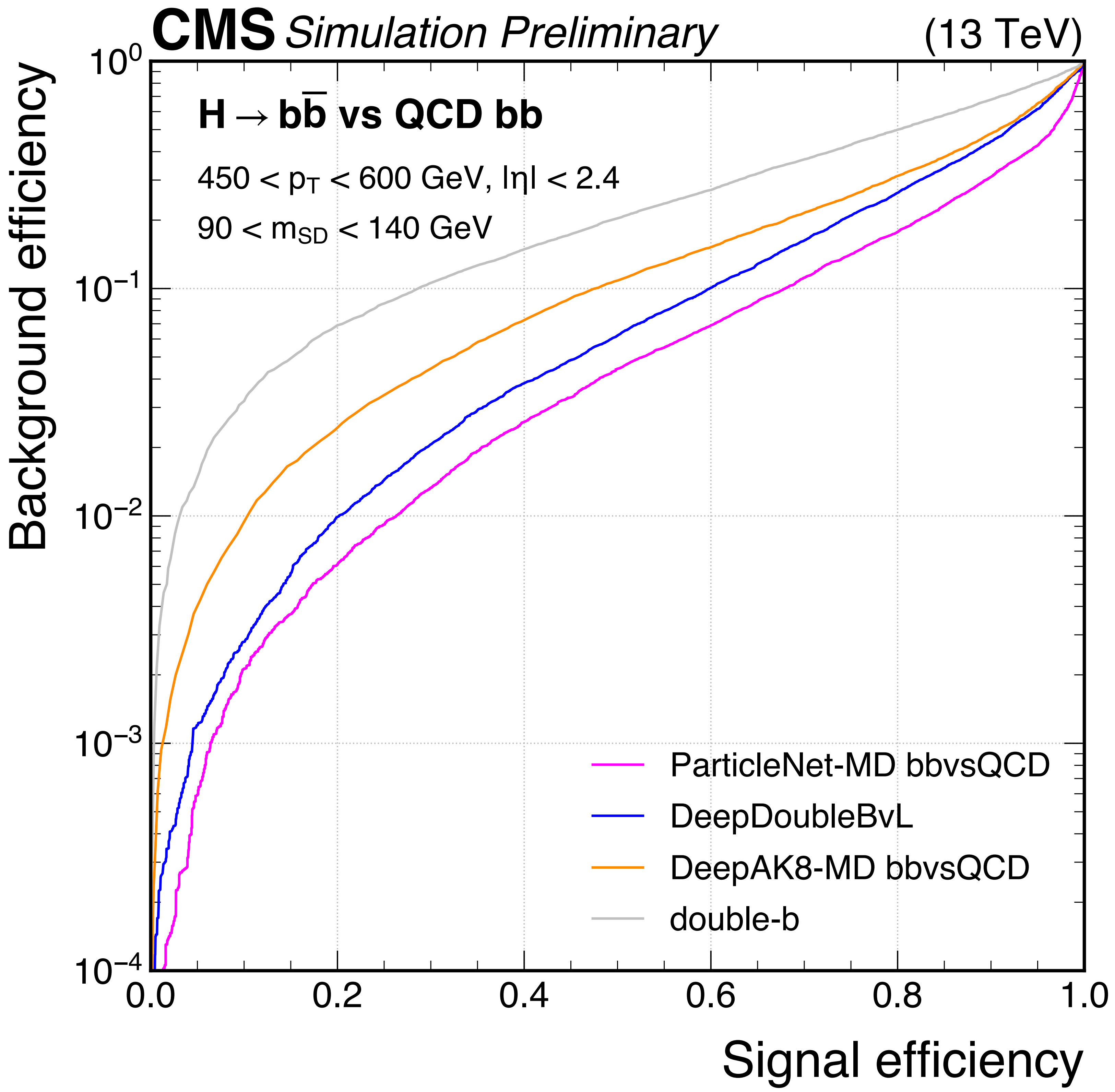
- At least one merged jet (AK8) passing this selection

	p_T (GeV)	$\lvert\eta\rvert$	ΔR(jet,leps)	Other
AK4 jets	> 20	< 2.5	≥ 0.5	JetID = 2 p_T >50 GeV OR PuJetID
AK8 jets	> 250	< 2.5	≥ 0.8	$m_{softdrop} > 30$ GeV

- **ParticleNet-MD**: state-of-art for CMS boosted jet tagging
  - Graph based architecture: jet as a particle cloud.
  - Mass decorrelation: trained with a flat distributions in p_T and mass of the resonance.

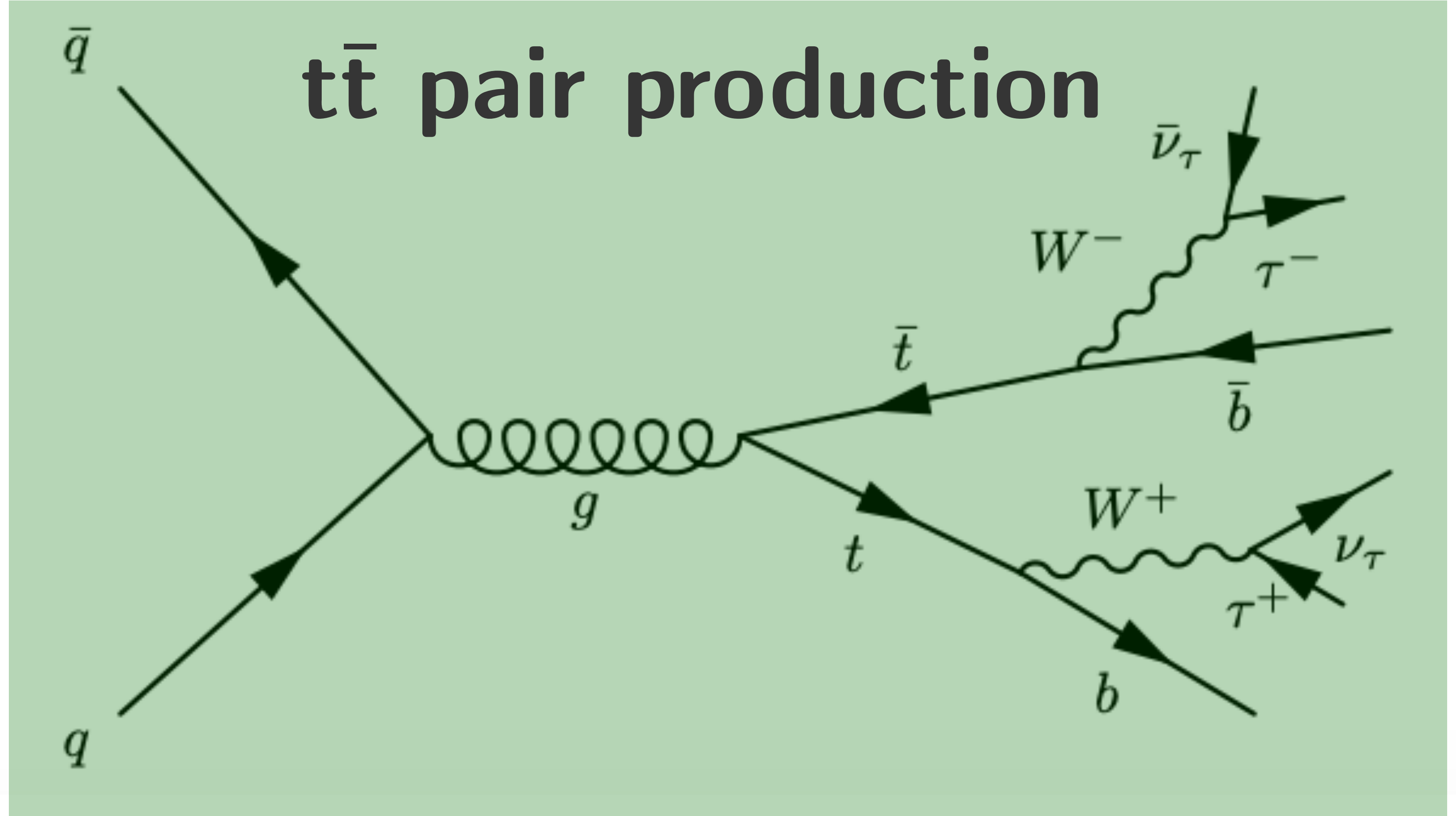
$$D_{bb} = \frac{P(X_{bb}^{MD})}{P(X_{bb}^{MD}) + P(X_{QCD}^{MD})}$$

- **H(bb) candidate: the AK8 jet with highest PNet-MD score**

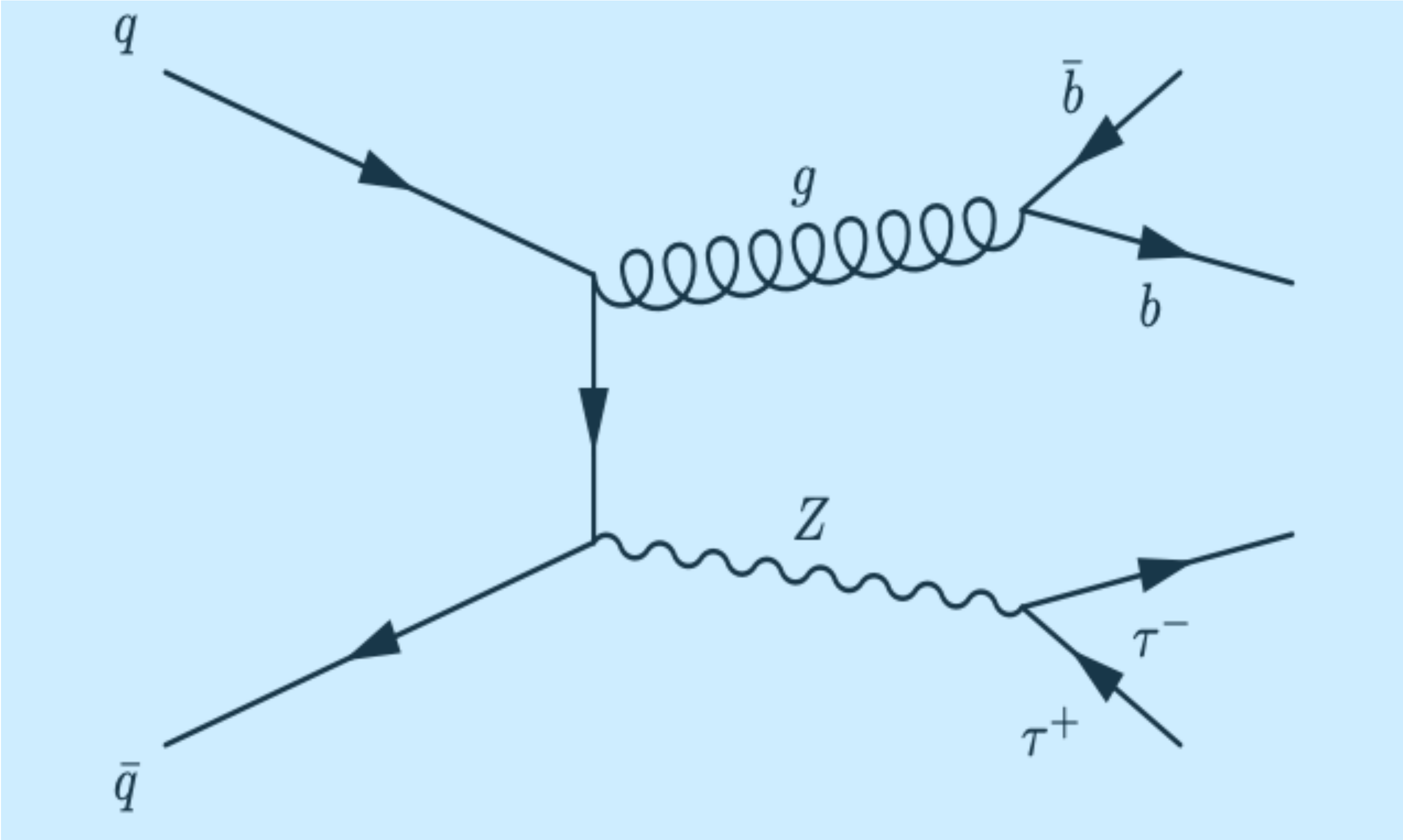


https://cds.cern.ch/record/2866276

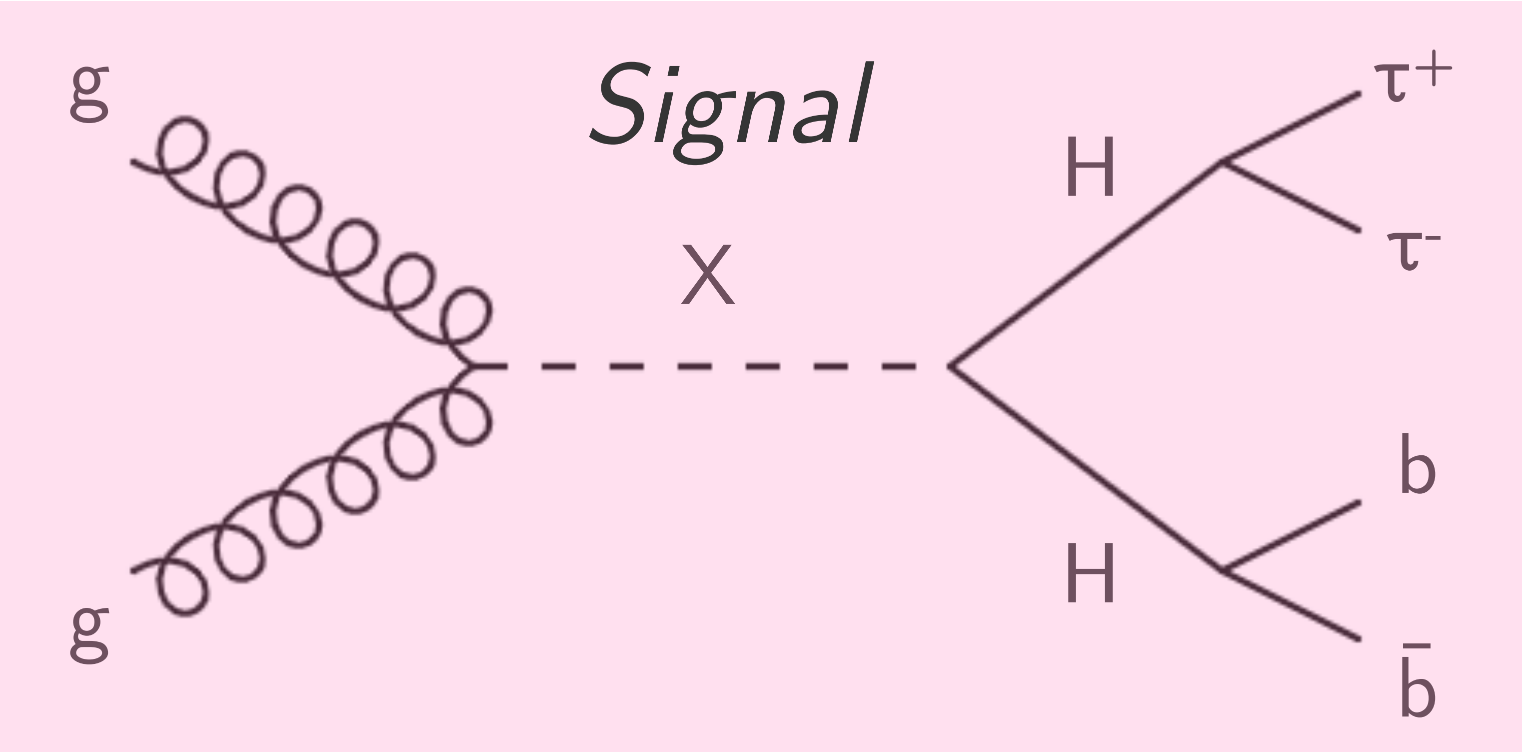
Background processes



Irreducible backgrounds

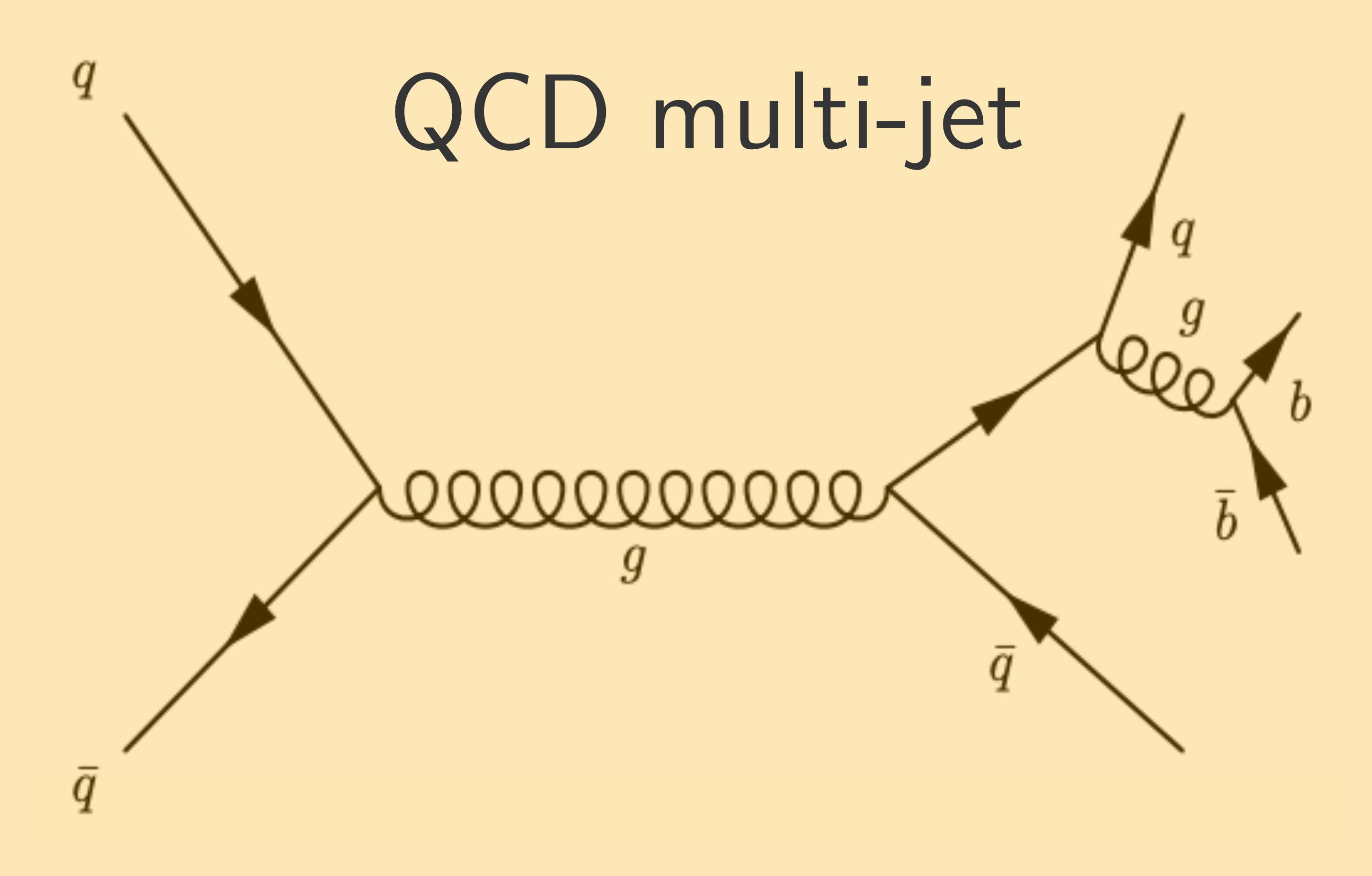


Z ($\ell\bar{\ell}$) + Jets ("Drell-Yan")

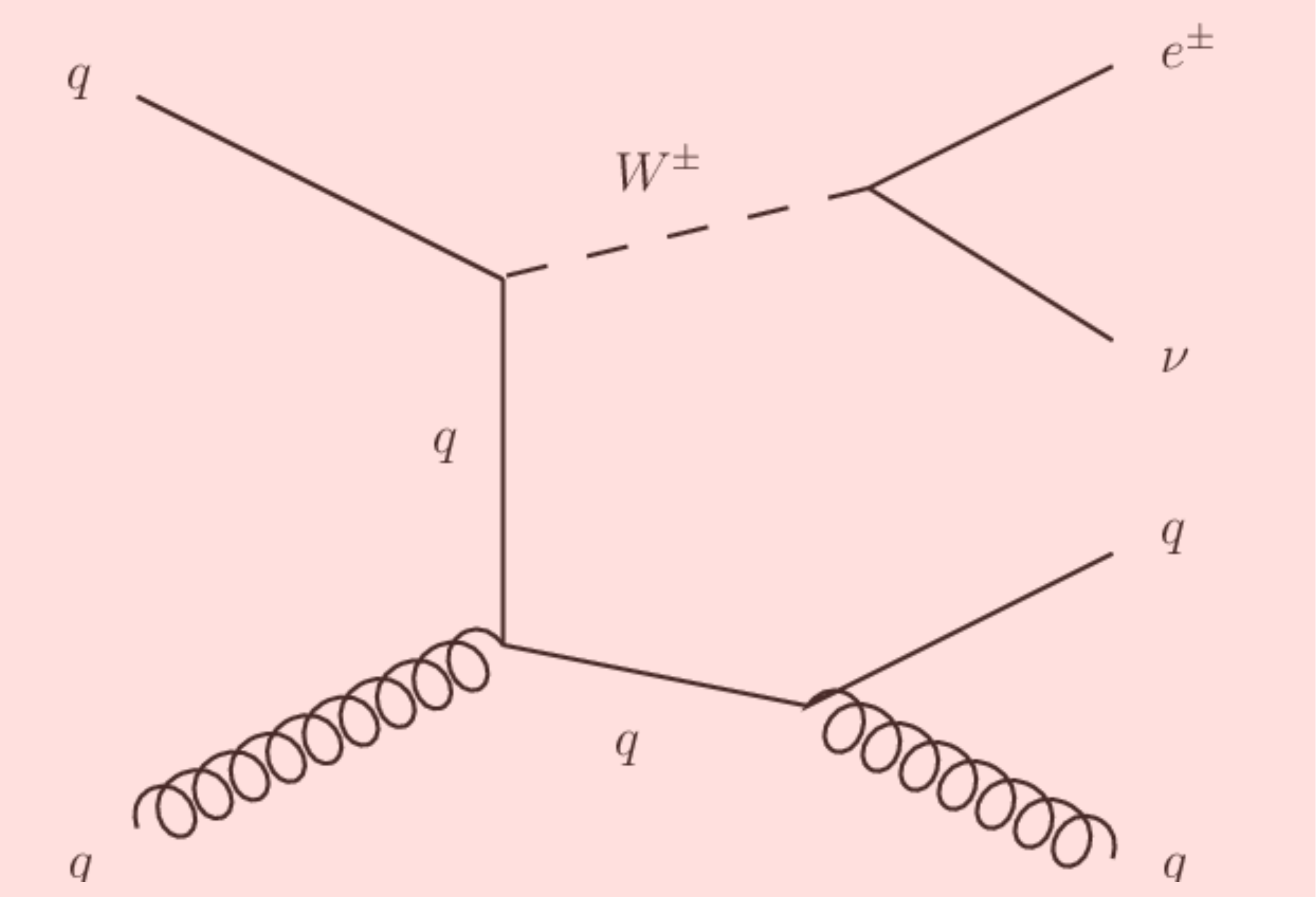


Process	σ(pb)
Top-antitop pair production	833.9
W(lν) +Jets	61526.7
Z(ll)+Jets	6077.22
QCD	o(10^9)
Single top (top, t-channel)	136.02
Single top (anti-top, t-channel)	80.95
Single top (tW)	71.7
⋮	⋮
SM HH bbtautau	0.03501

Not observable @ CMS right now



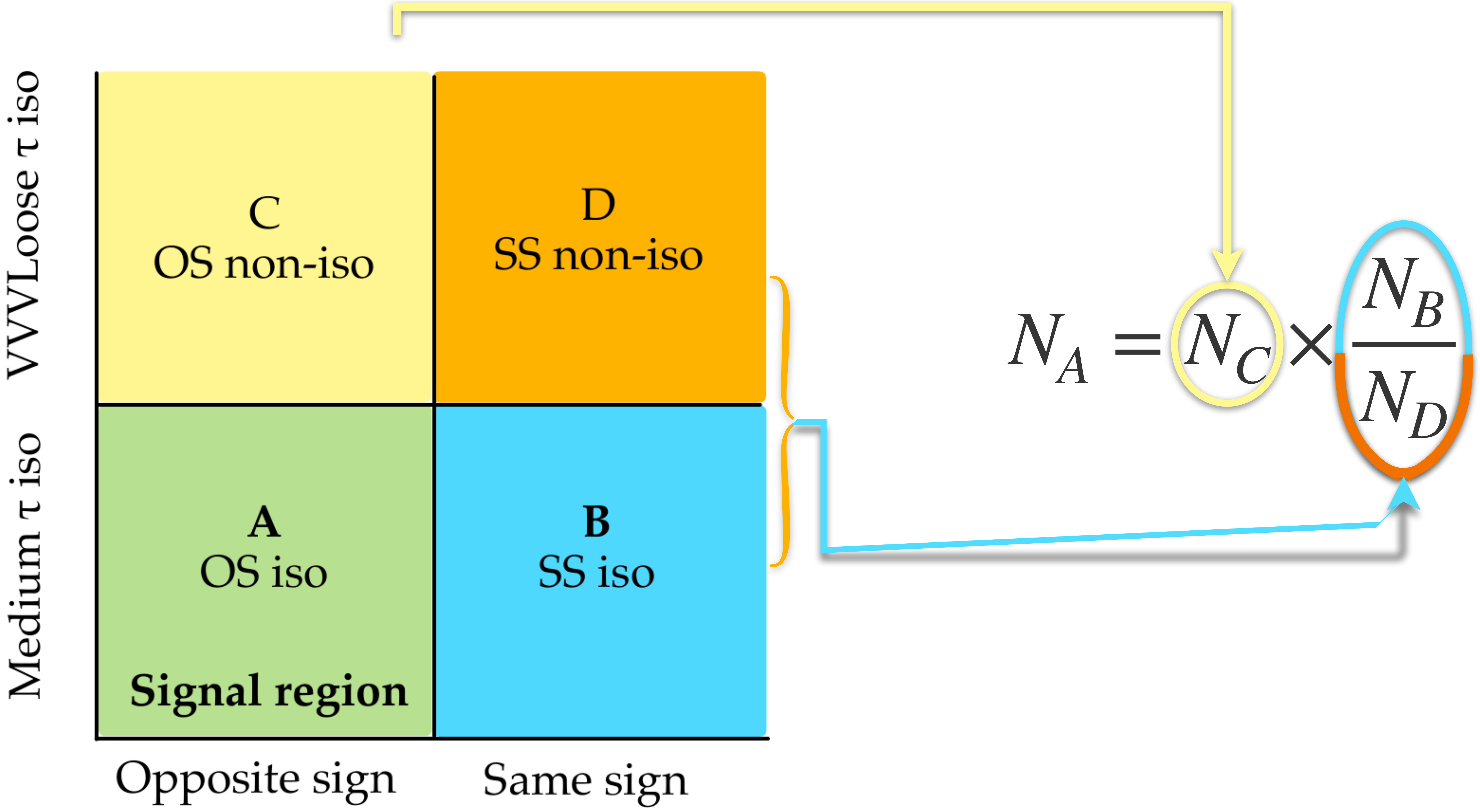
Reducible backgrounds



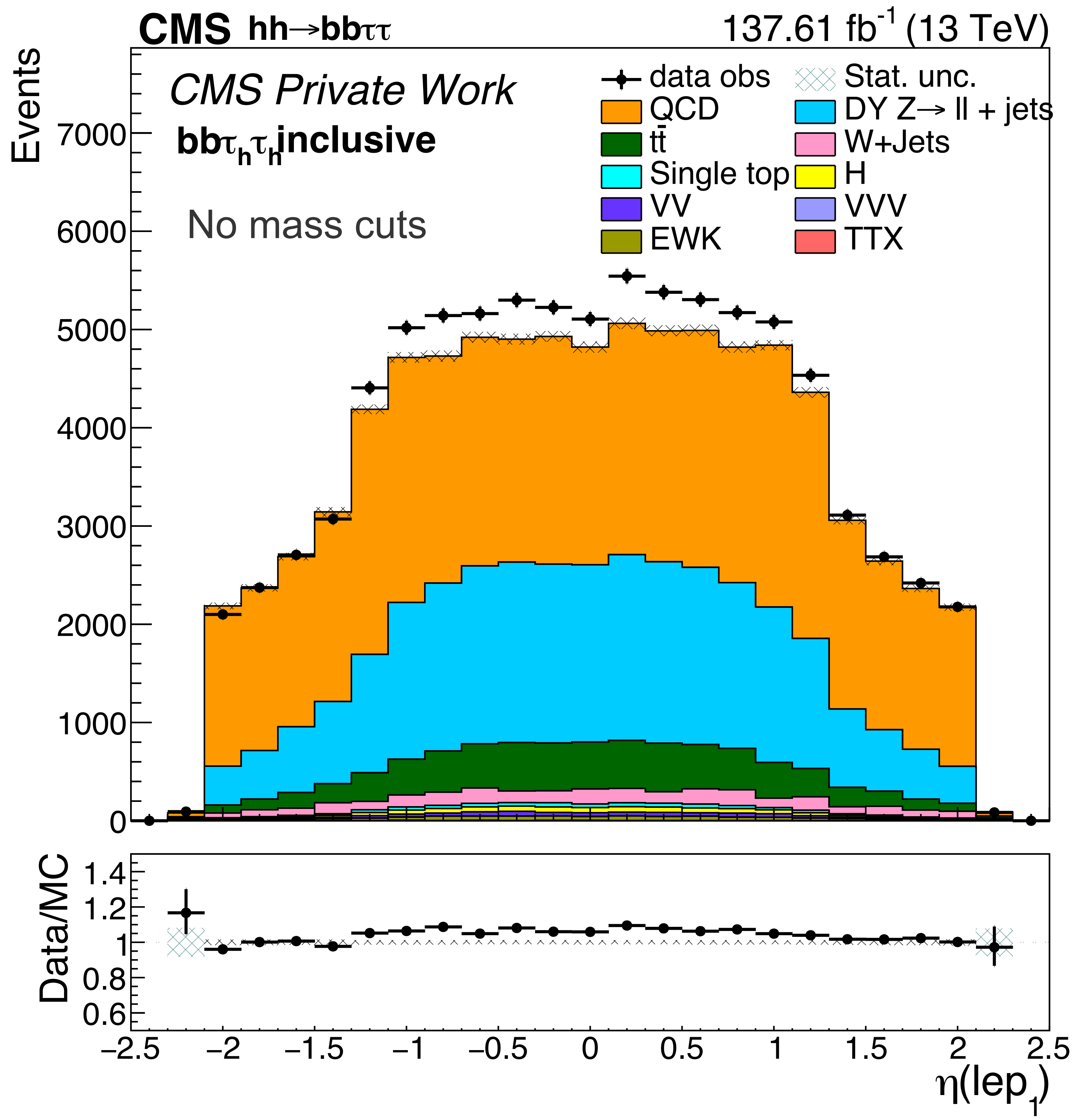
W ($\ell\bar{\nu}$) + Jets

QCD estimation

- Shape estimation in SR taking shape and transfer factor from orthogonal sidebands
- Observables: 2nd lepton isolation and the leptons charge signs
- Subtraction of MC simulations in sidebands



$$N_A = N_C \times \frac{N_B}{N_D}$$



Categories

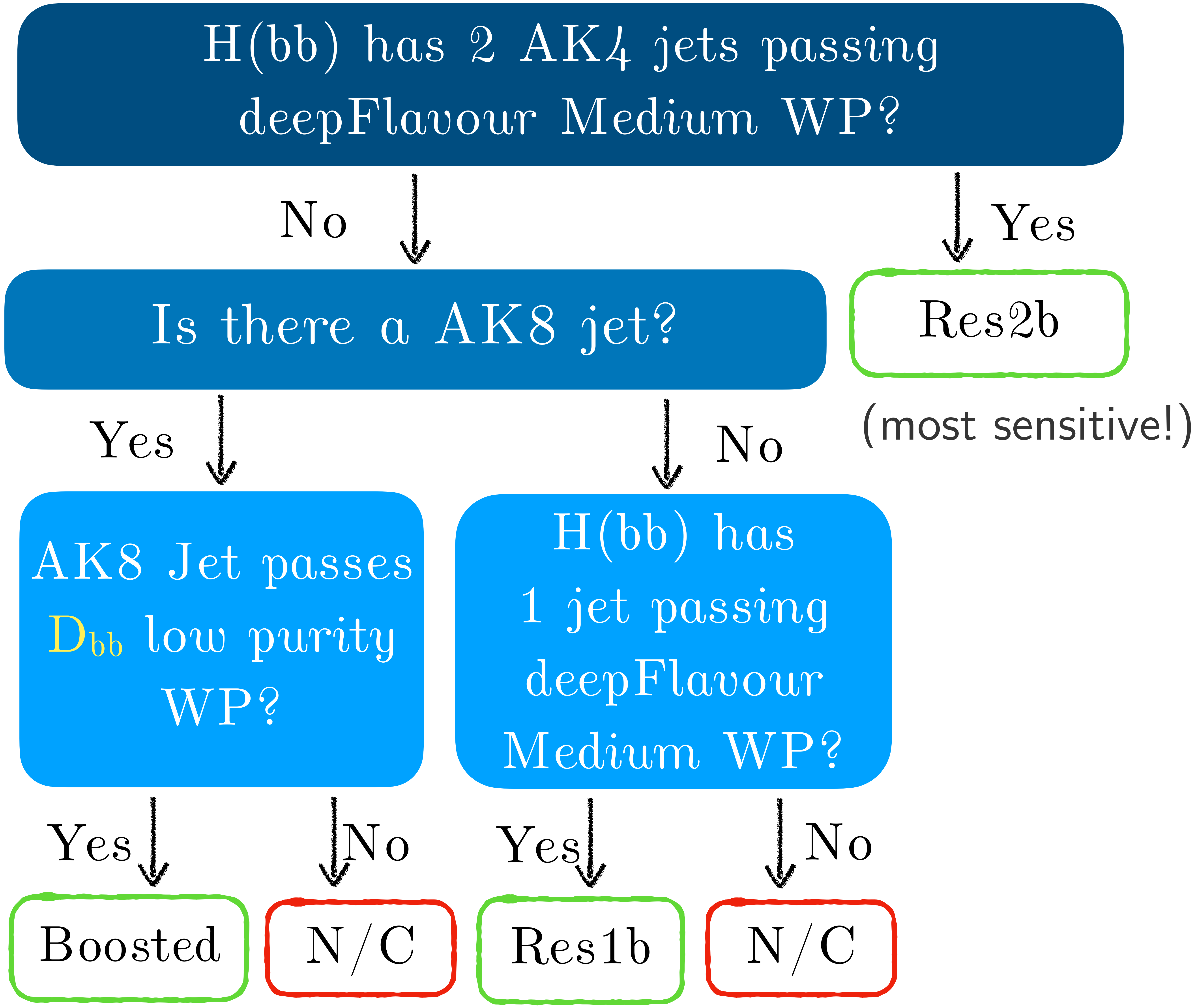
Res1b, Res2b, Boosted:

- Categories for signal extraction
- Designed to optimise the signal sensitivity and reduce backgrounds

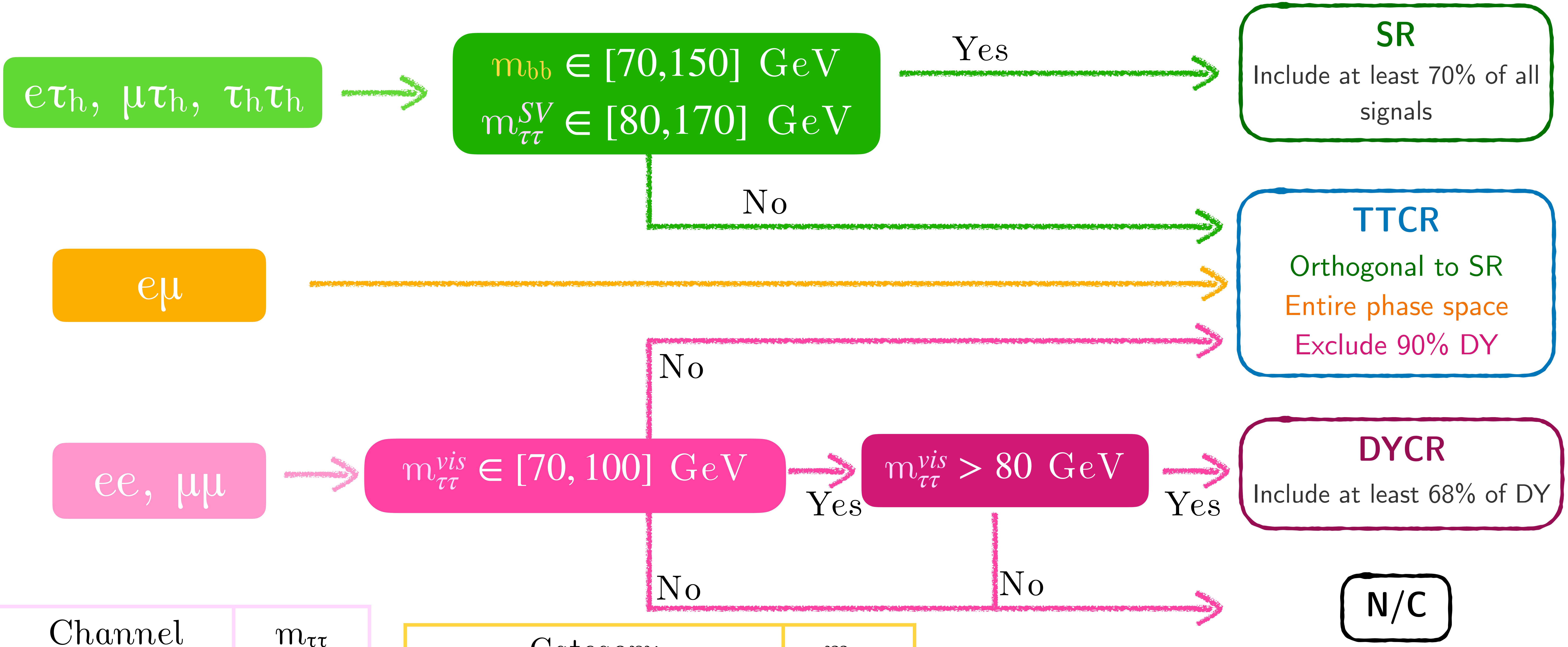
Inclusive:

- Not used for signal extraction
- Only AK4 jets regardless of deepFlavour WP, no AK8 jets

D_{bb} defined in [slide 17]



Signal/Control Regions



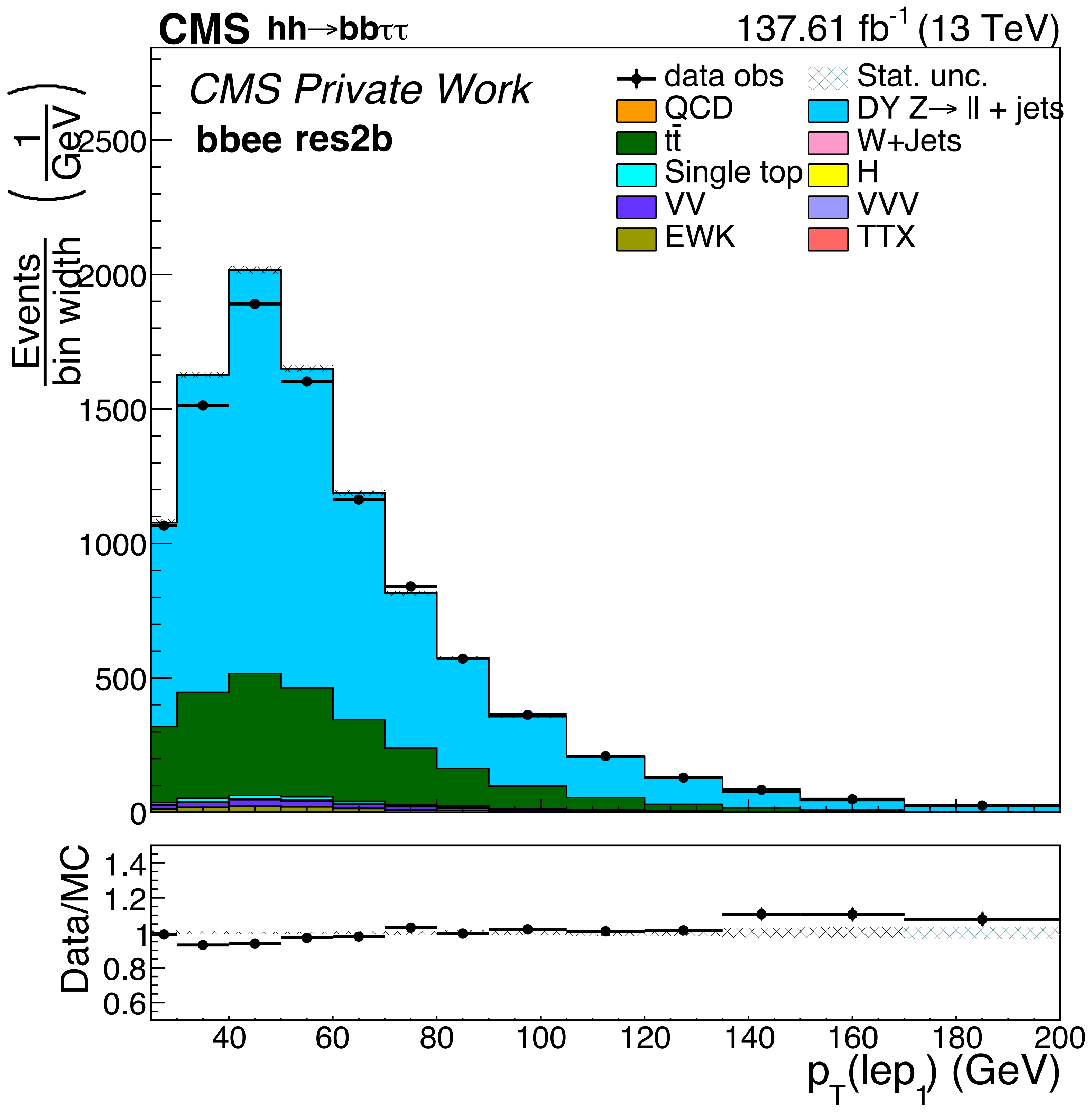
Channel	$m_{\tau\tau}$
$e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$	$m_{\tau\tau}^{SV}$
ee, $\mu\mu$, $e\mu$	$m_{\tau\tau}^{vis}$

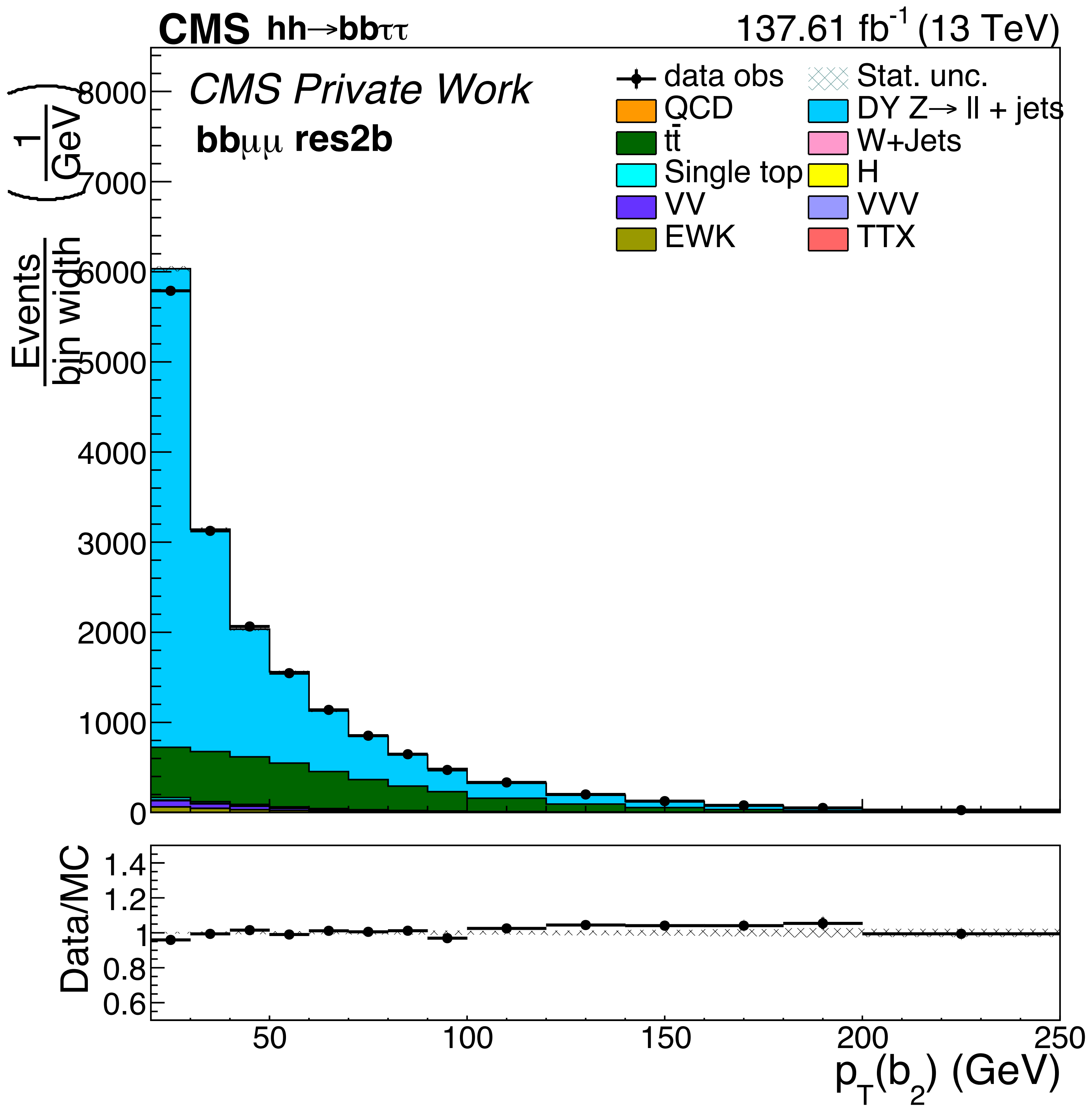
Category	m_{bb}
Res1b, res2b, inclusive	$m_{(j1,j2)}$
Boosted	$m_{softdrop}$

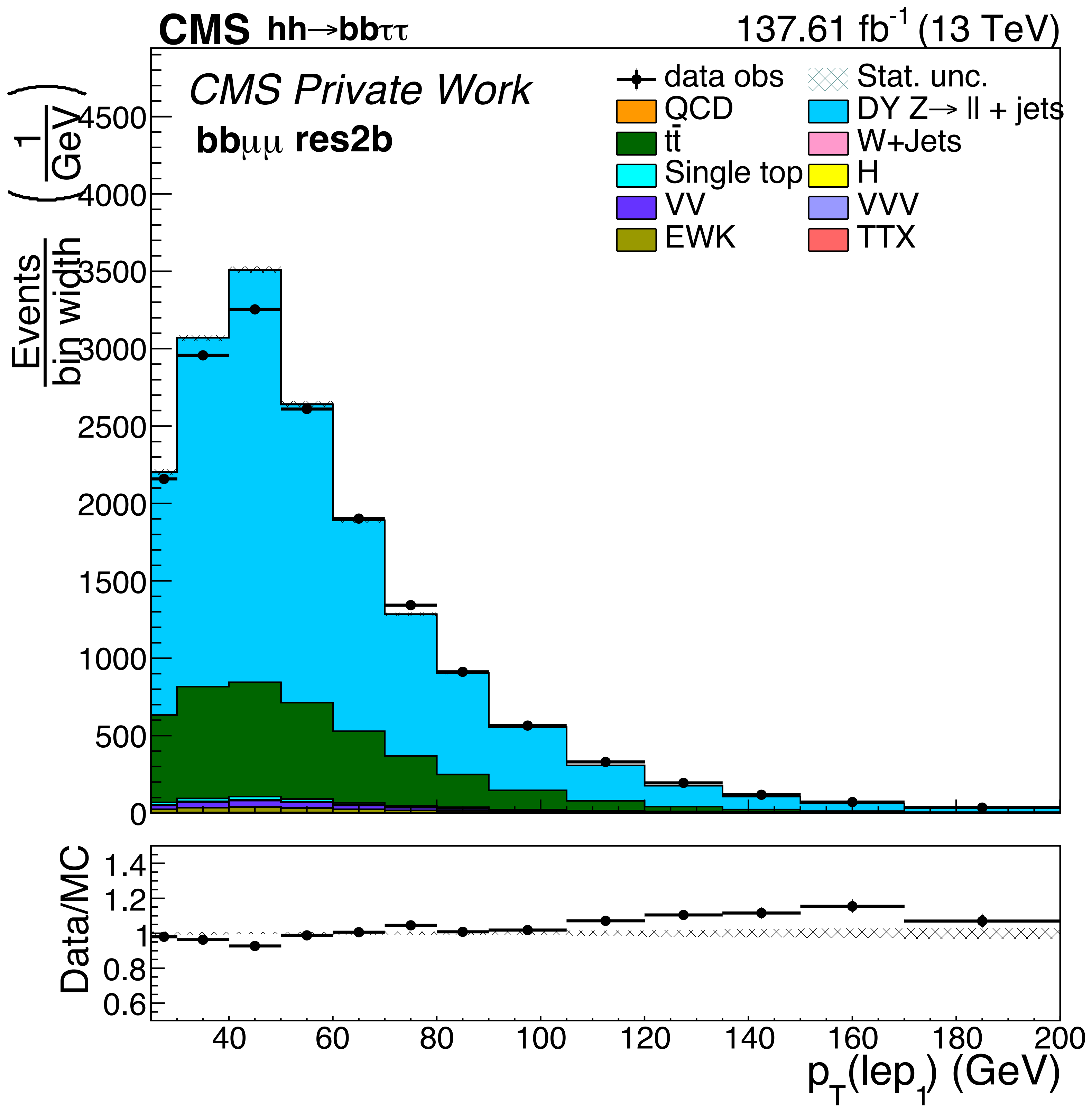
Data/MC comparison in DYCR

Recap: Check DY MC samples
- ee, μμ channels
- Include 68% of DY

Muons: larger discrepancies at high p_T, due to inefficient behaviour of tightID tagger. Muon POG suggested to add a 5% uncertainty



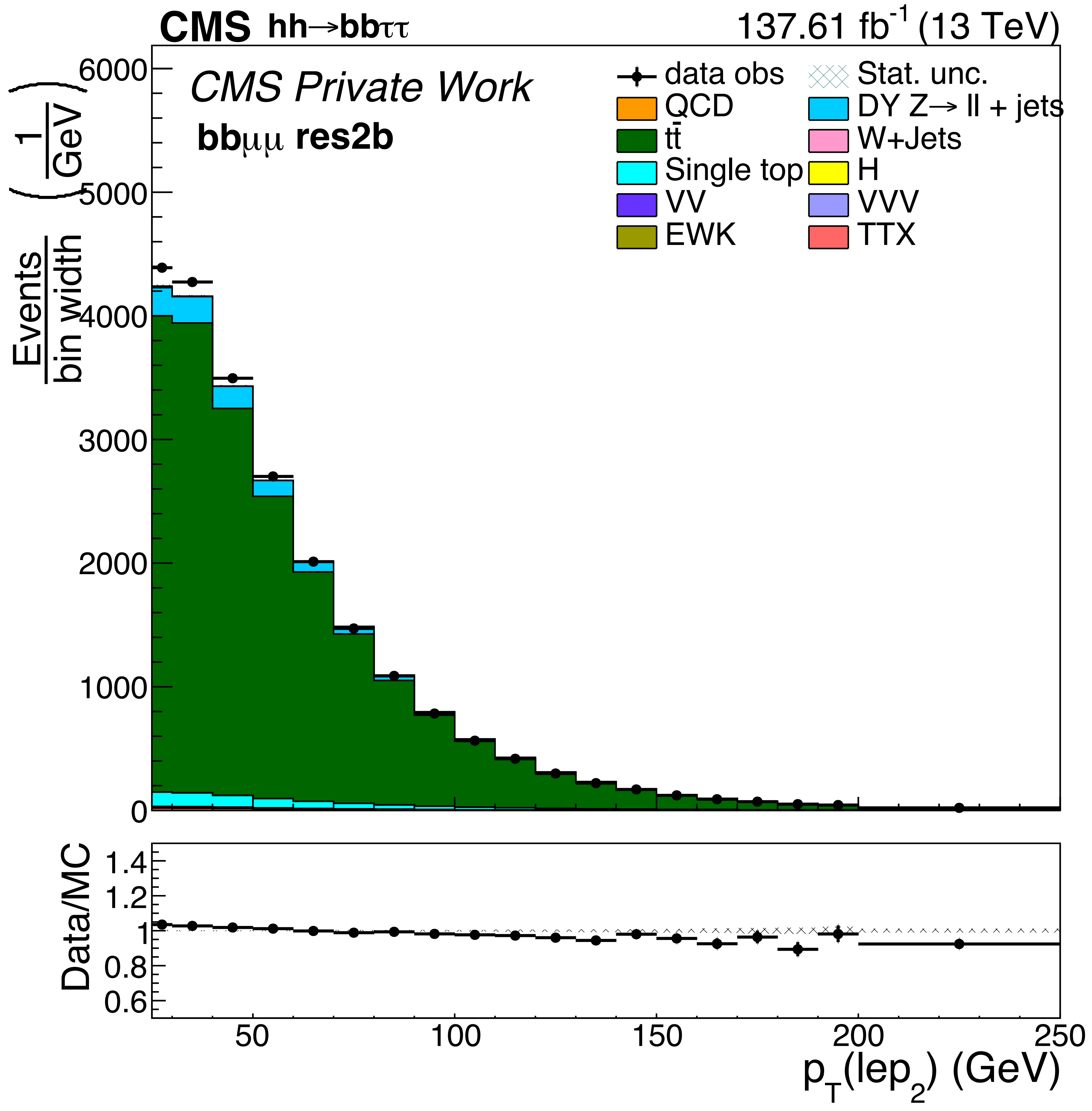


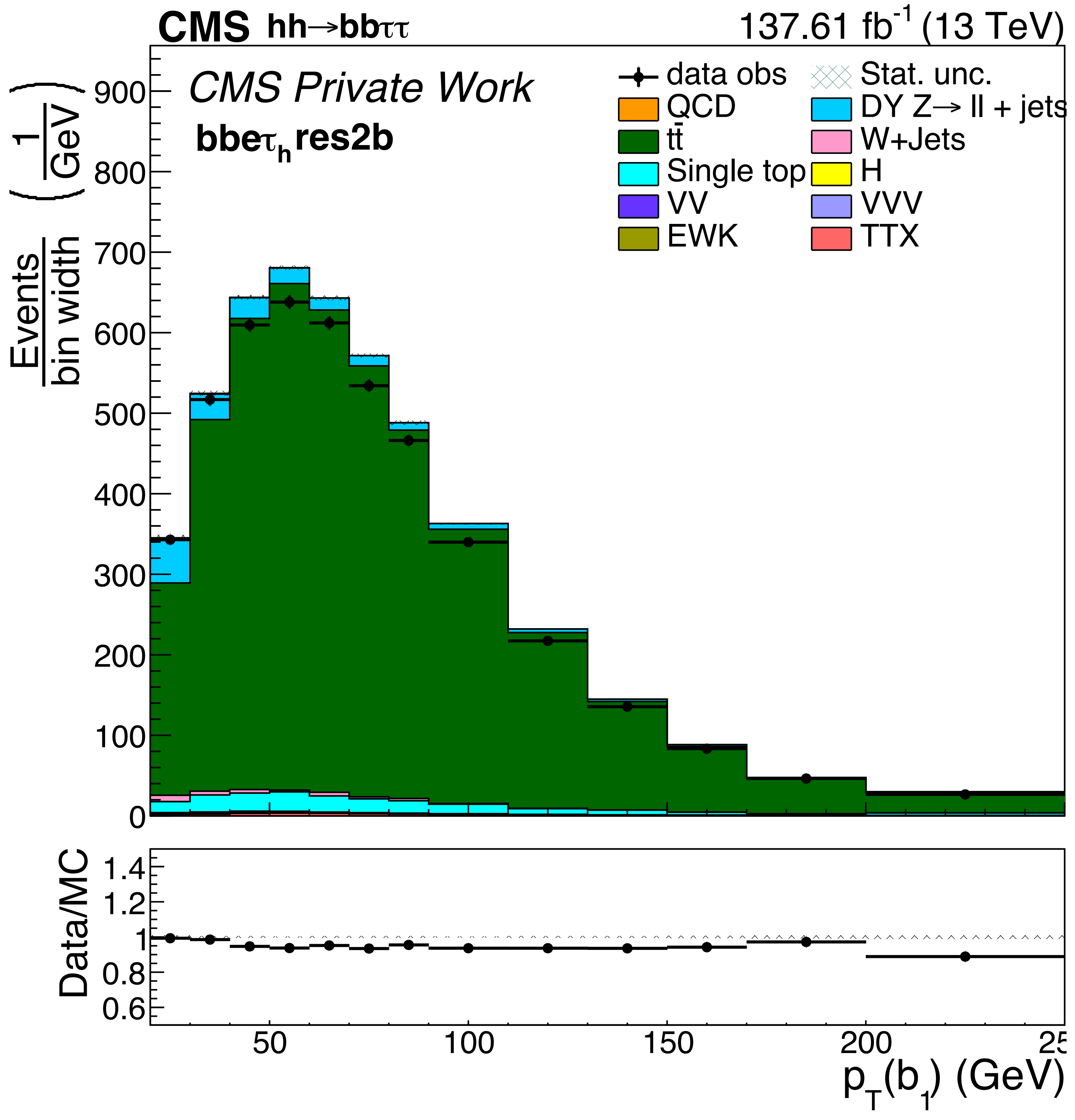


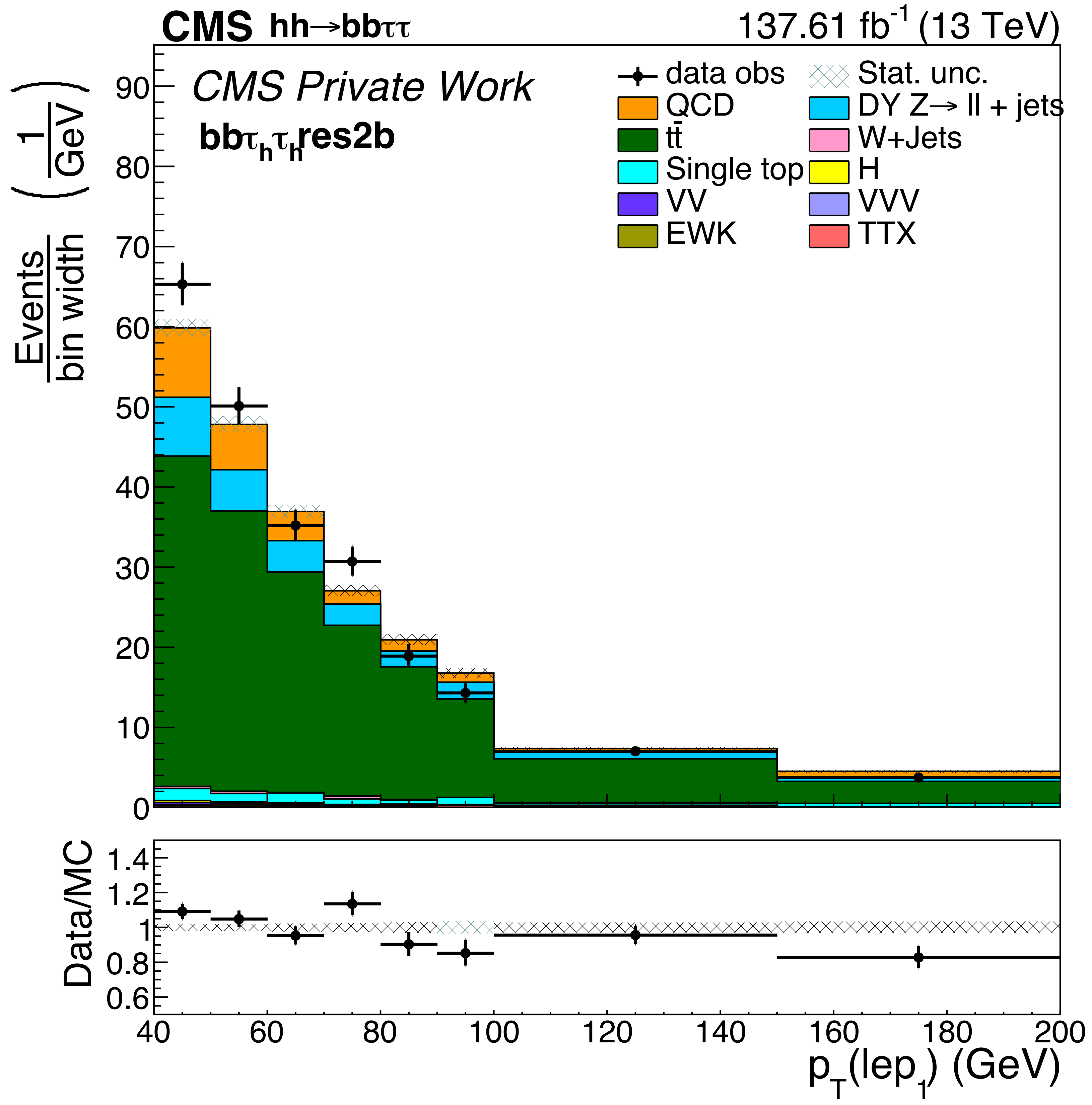
Data/MC comparison in TTCR

Why TTCR? observed p_T spectra of top candidates in tt data were found to be significantly softer than predictions from MC simulations

- ee, μμ channels: exclude 90% of DY
- $e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$ channels: out of SR
- eμ channel: entire spectrum

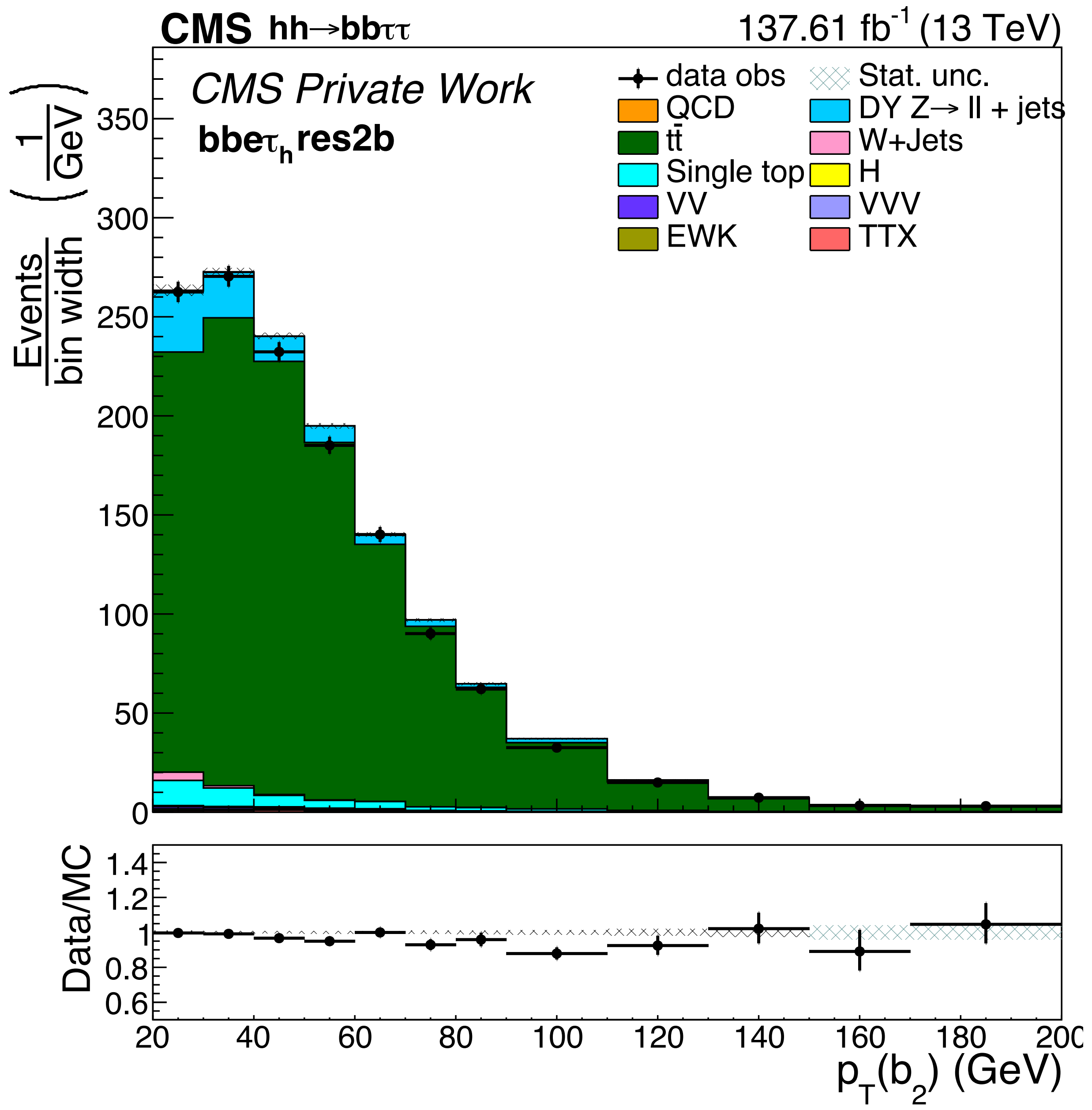


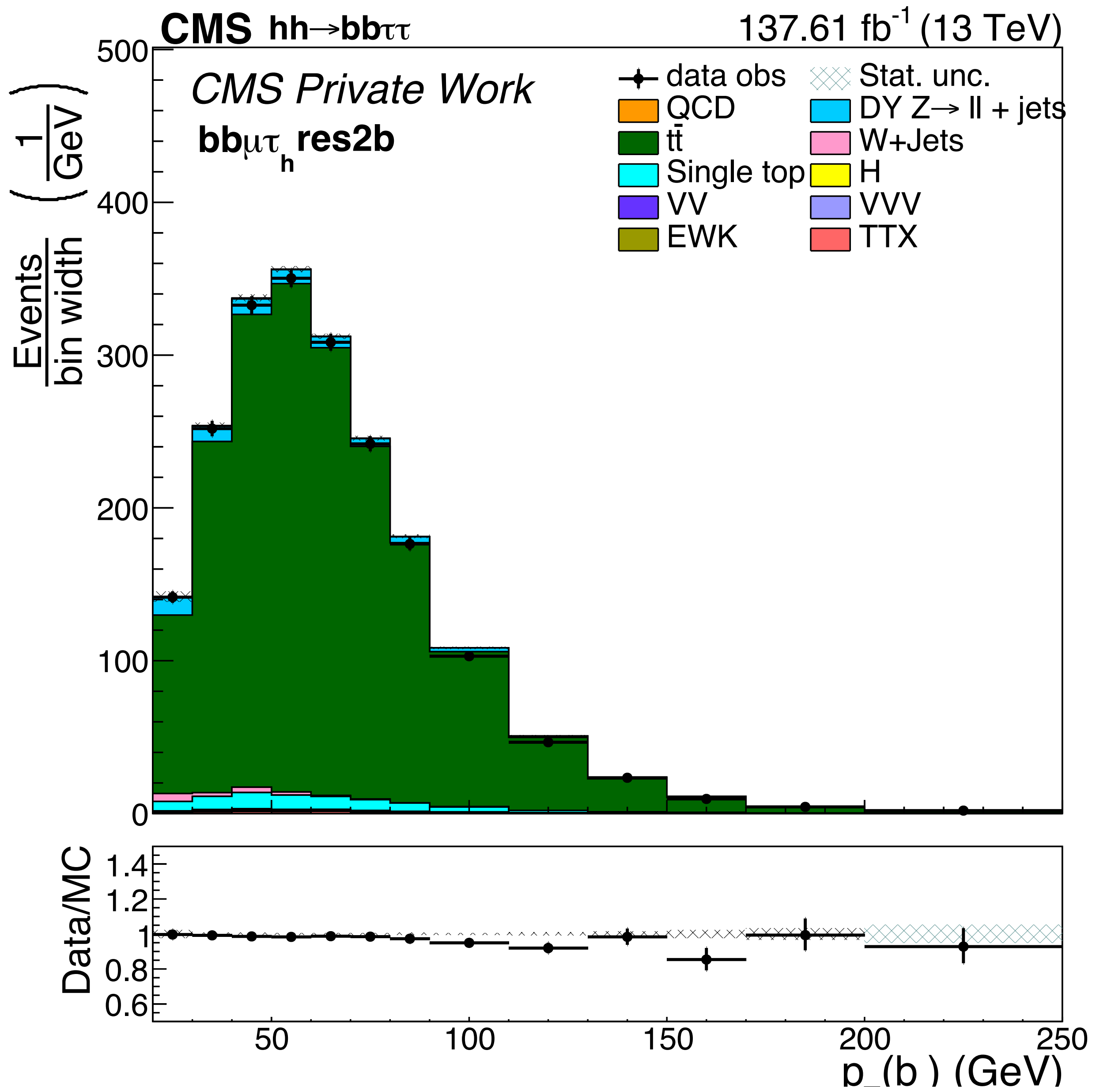


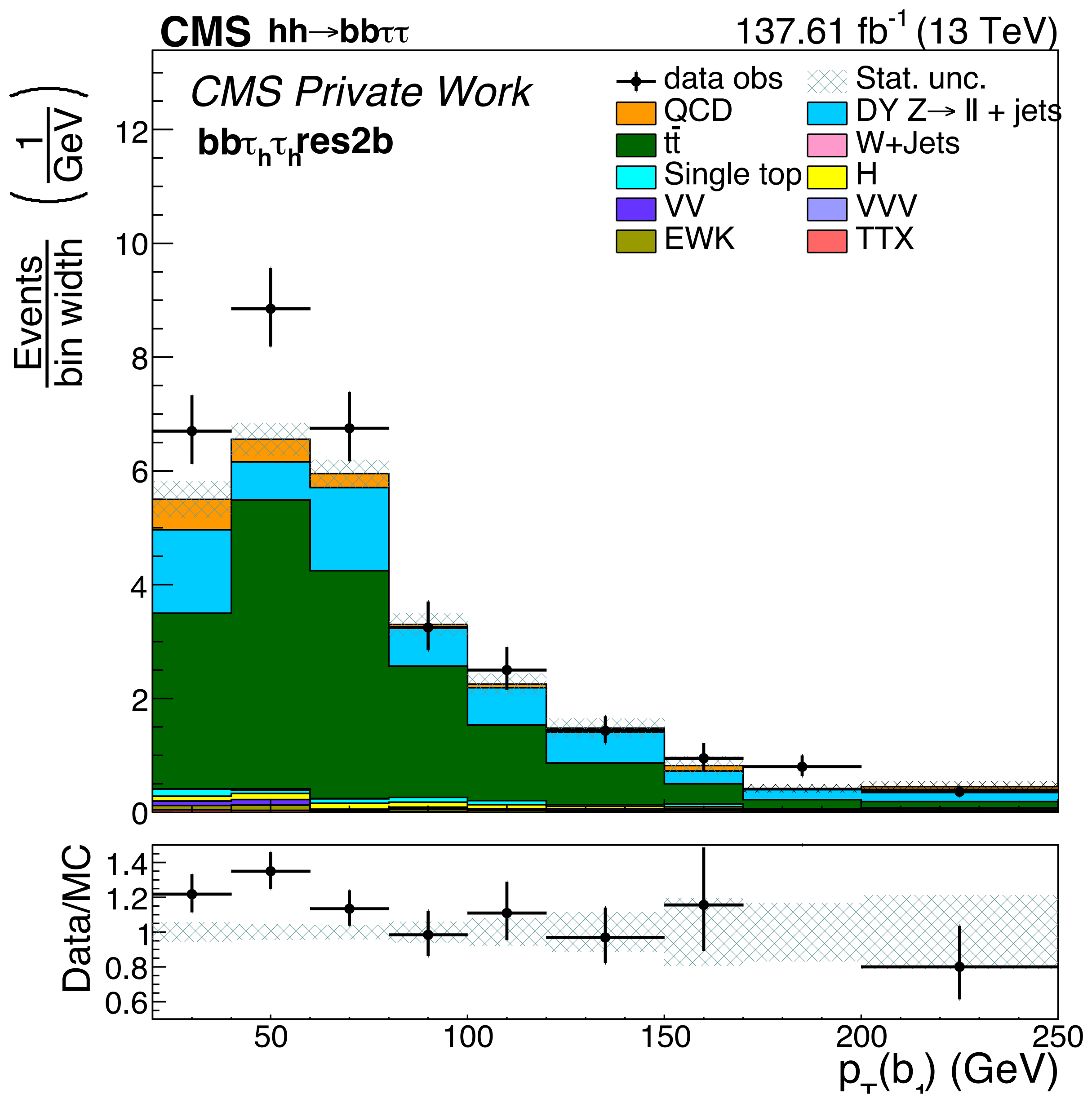


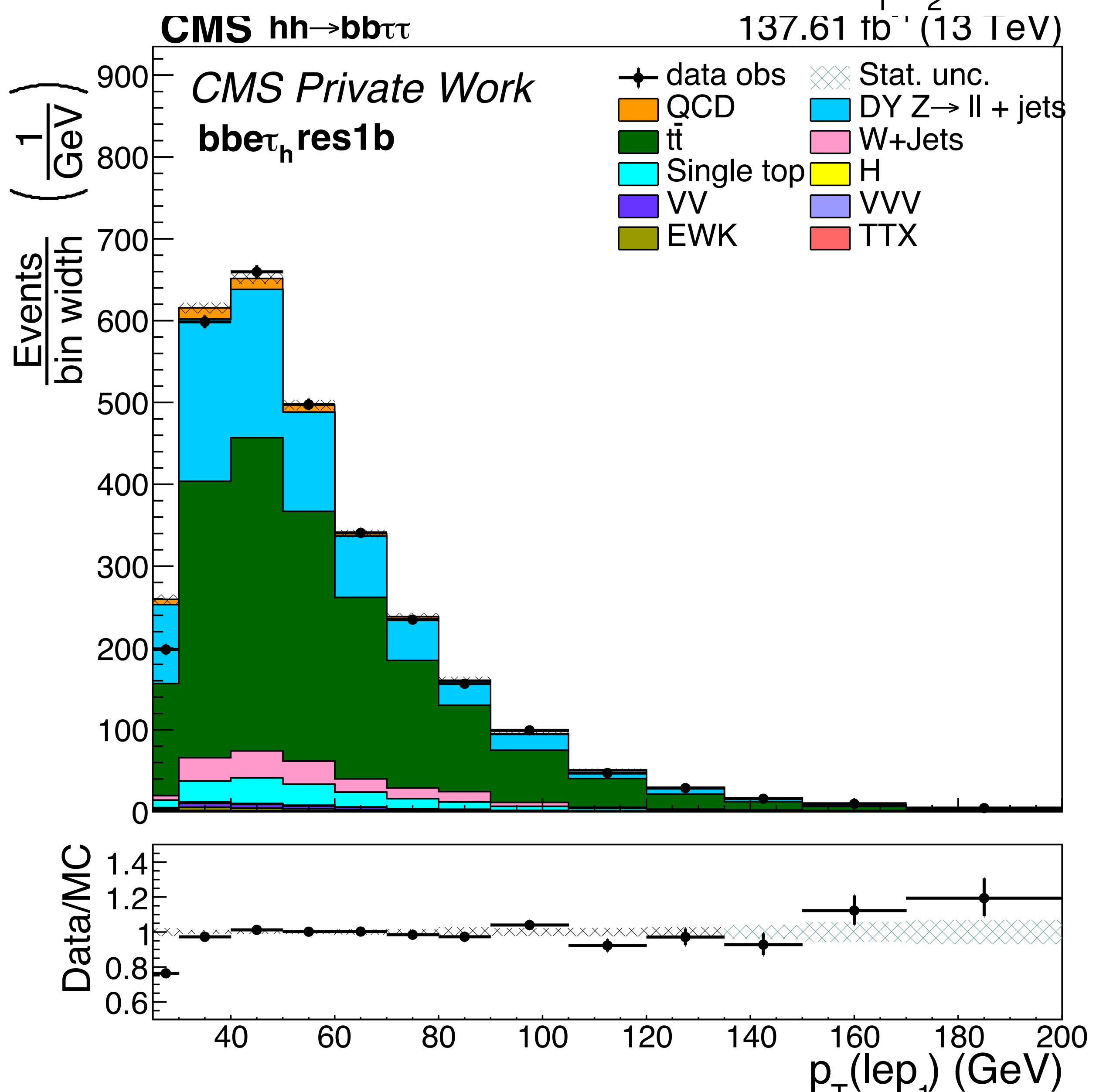
Data/MC comparison in SR

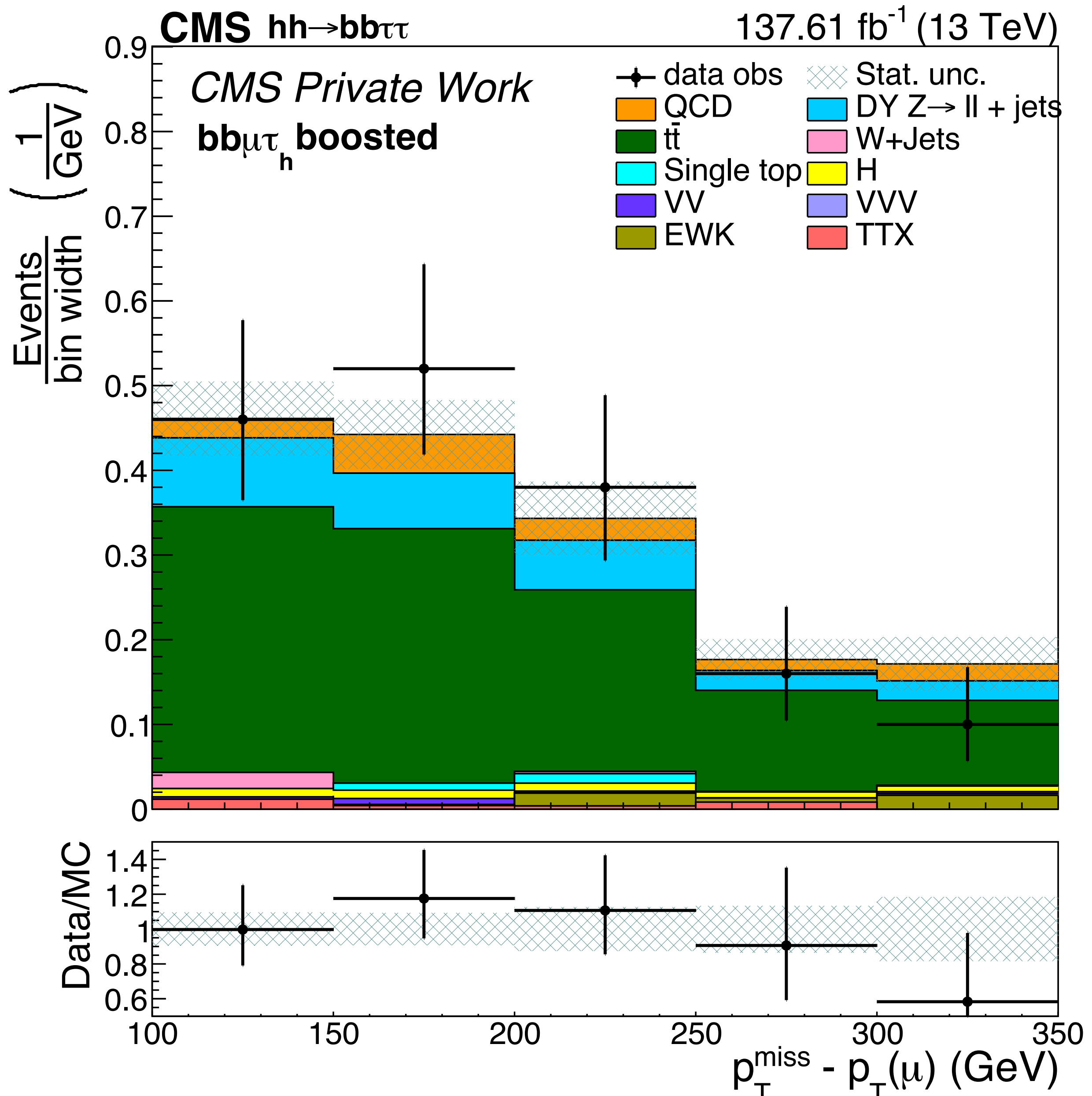
$m_{bb} \in [70,150]$ GeV
$m_{\tau\tau}^{SV} \in [80,170]$ GeV

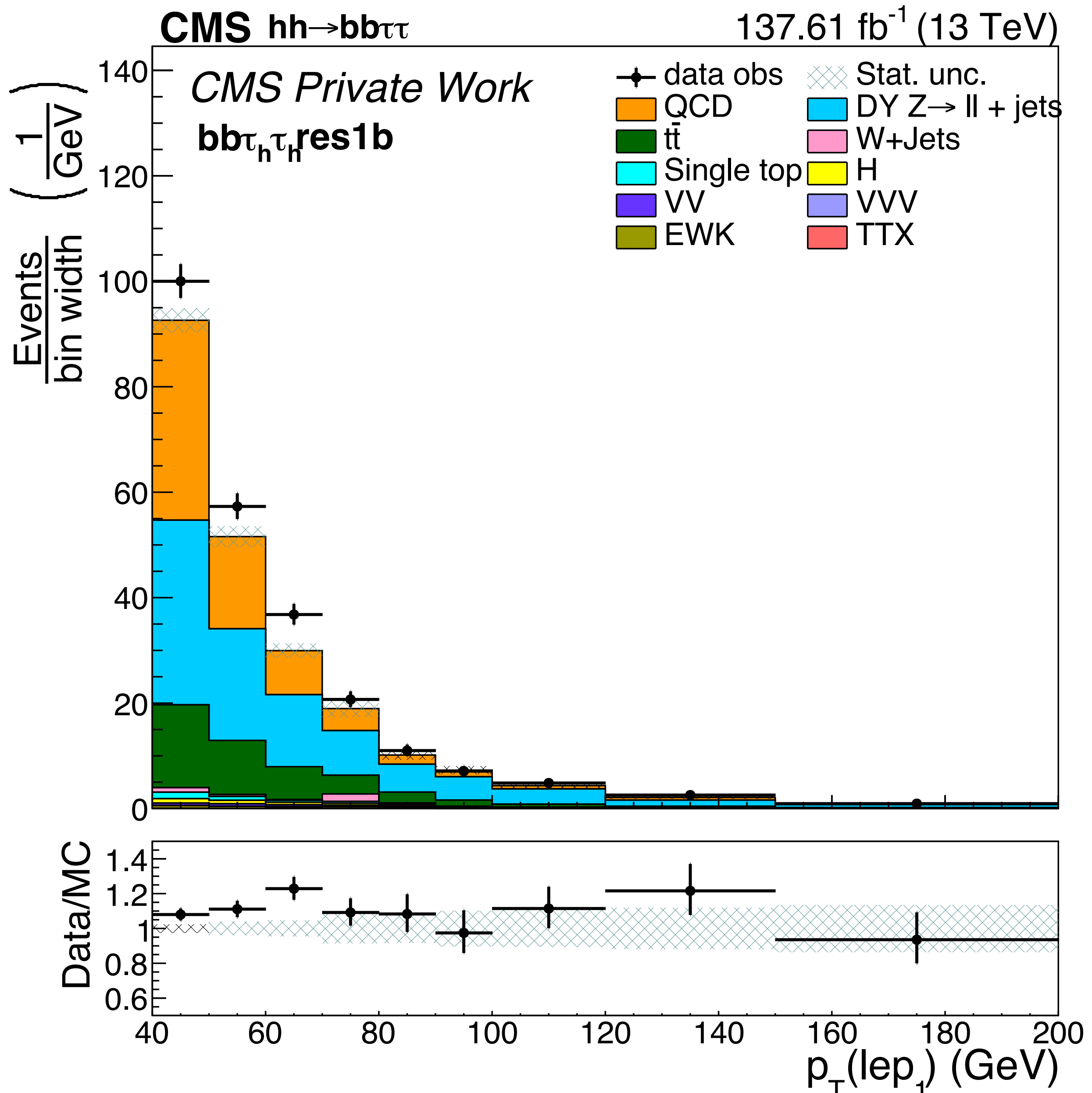












DeepTau 2p5 for τ_h ID

DeepTau: machine learning based multi-class τ_h ID algorithm

Three discriminators built from scores: VSJet, VSEle, VSMu

DeepTau2017v2p1

- Convolutional Deep Neural Network architecture
- High-level reconstructed τ features + low level information from all sub-detectors, objects reconstructed within the τ isolation cone
- **Issue**: data/MC disagreement at high deepTauVSJet

Deprecated after run 2

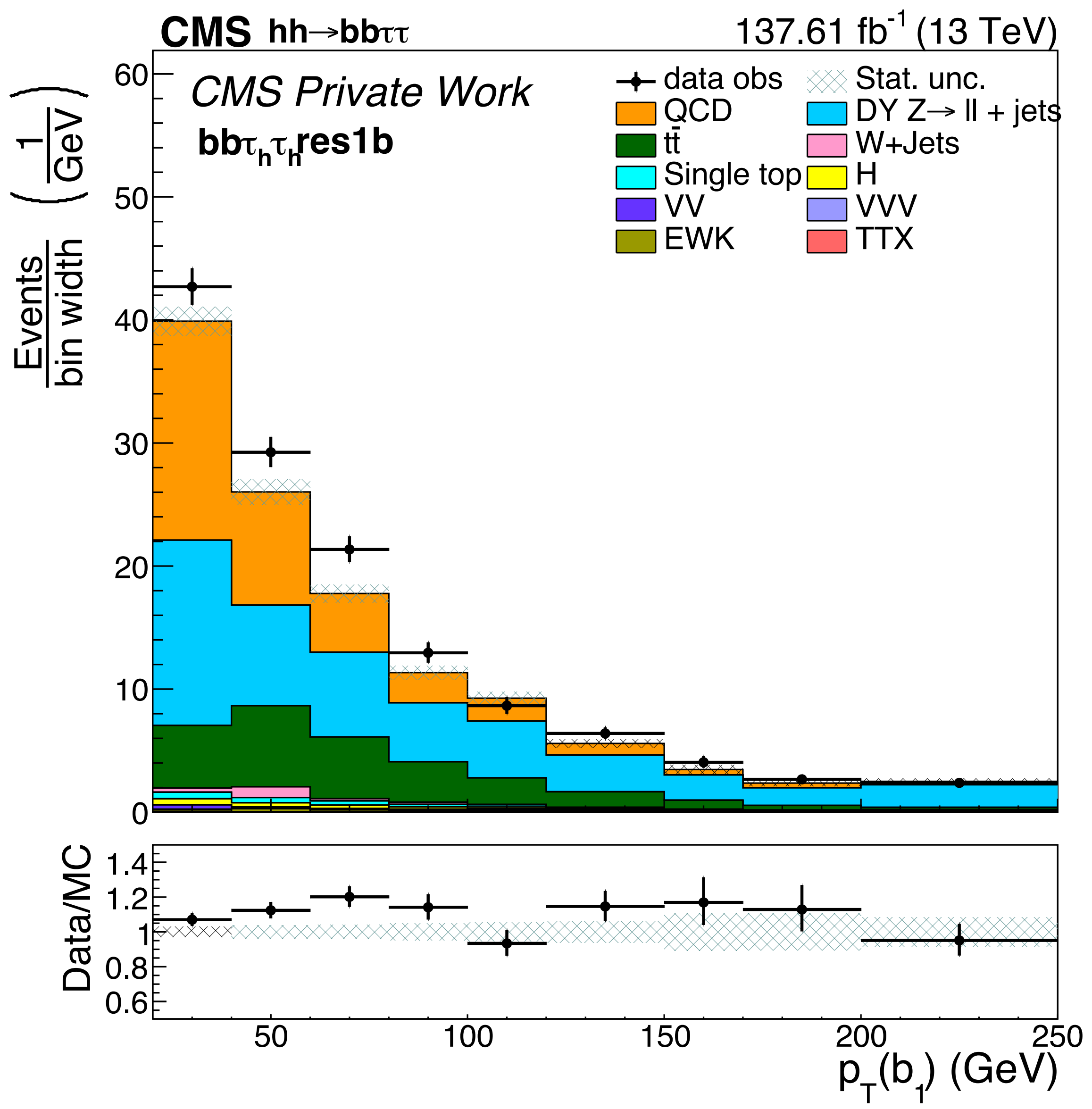
DeepTau2018v2p5

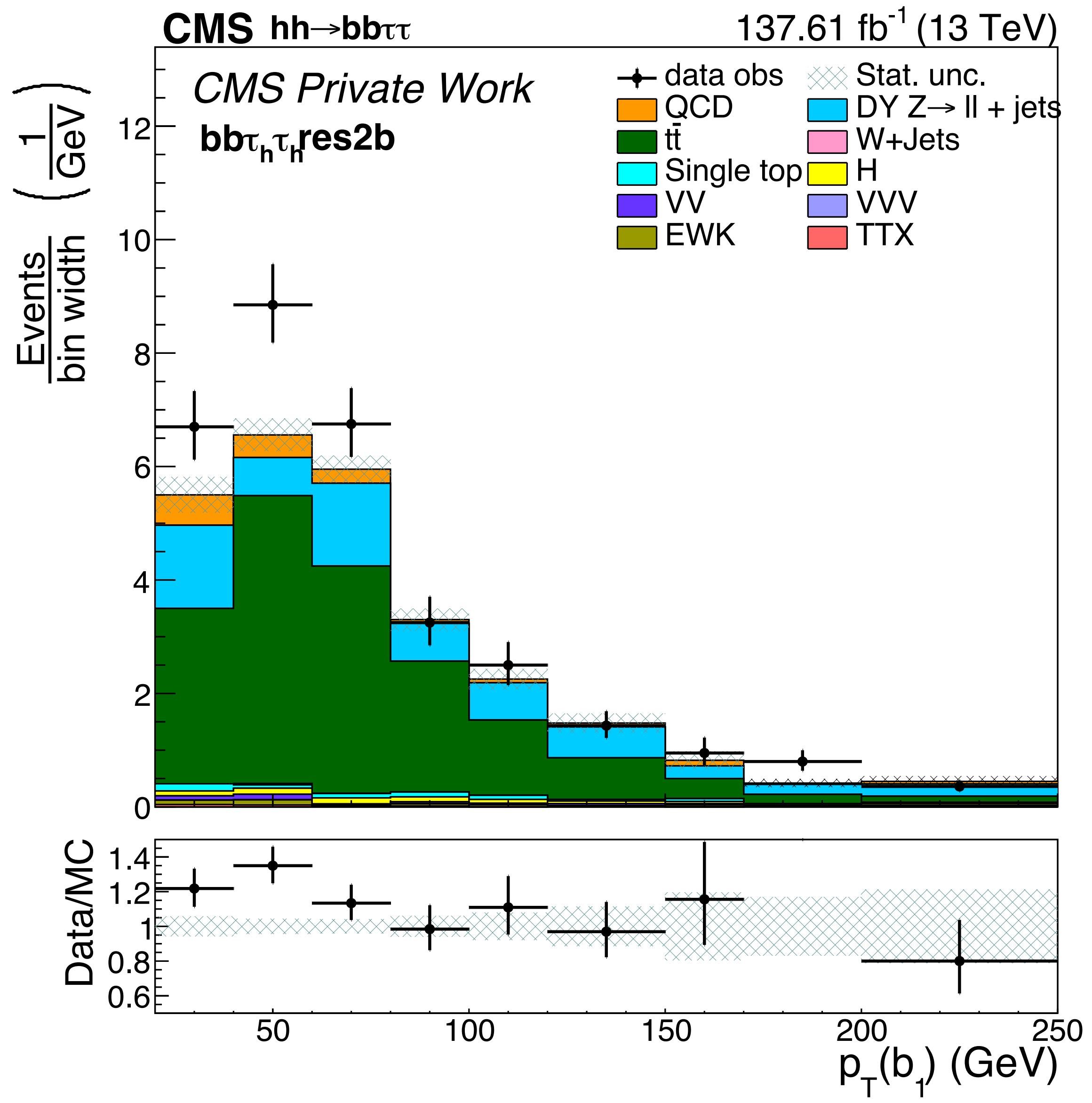
- Adversarial ML techniques for domain adaptation during training
- Training of sub-networks with competing goals: τ_h classification and data VS MC discrimination
  - Maximise the τ_h ID while preventing data/MC disagreement

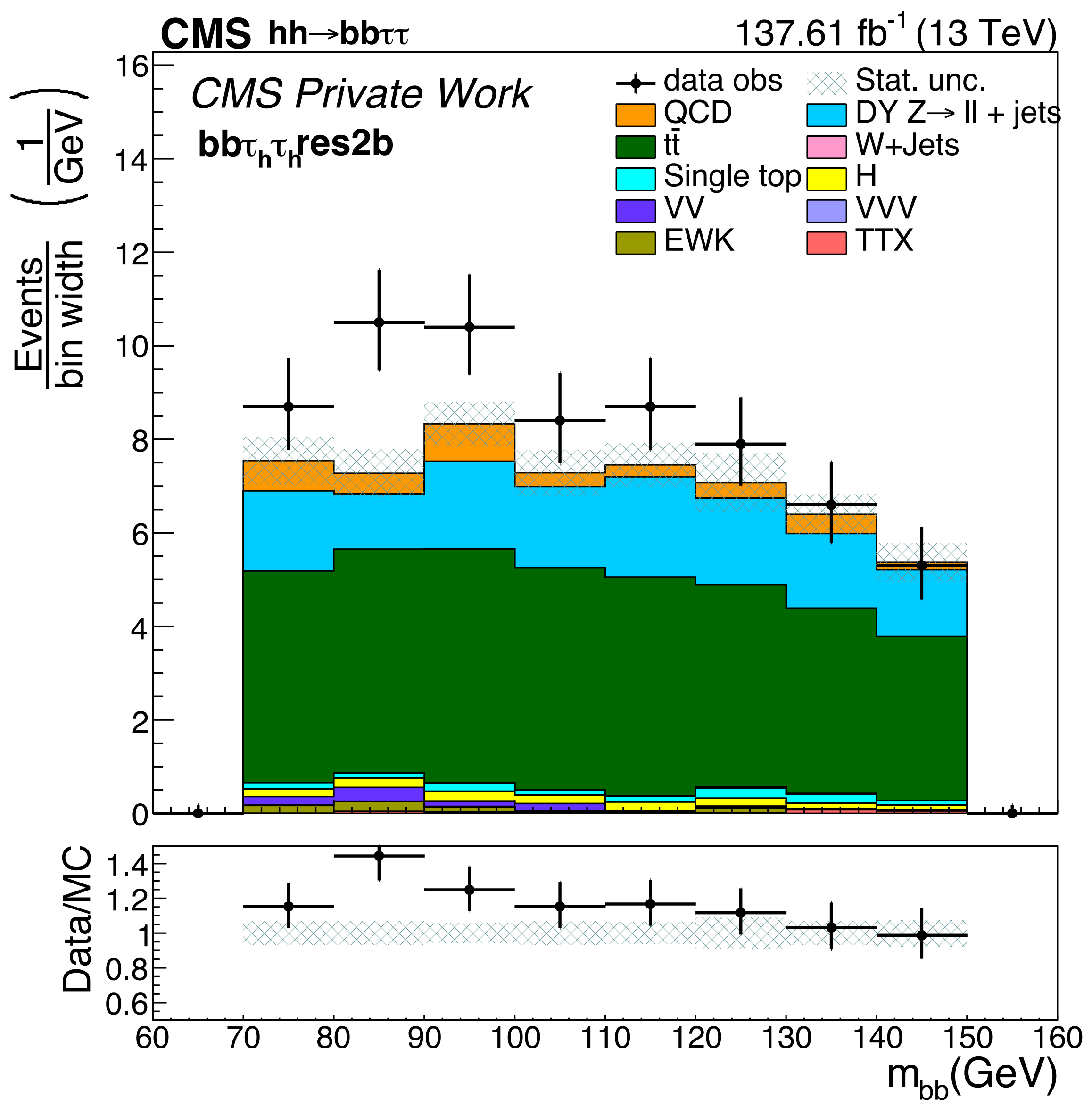
Recommended for Run 3 analyses

In this presentation: **preliminary results using *DeepTau2018v2p5*** for τ_h selection
Only for $\tau_h\tau_h$ channel, expected higher gain for deepTauVSJet

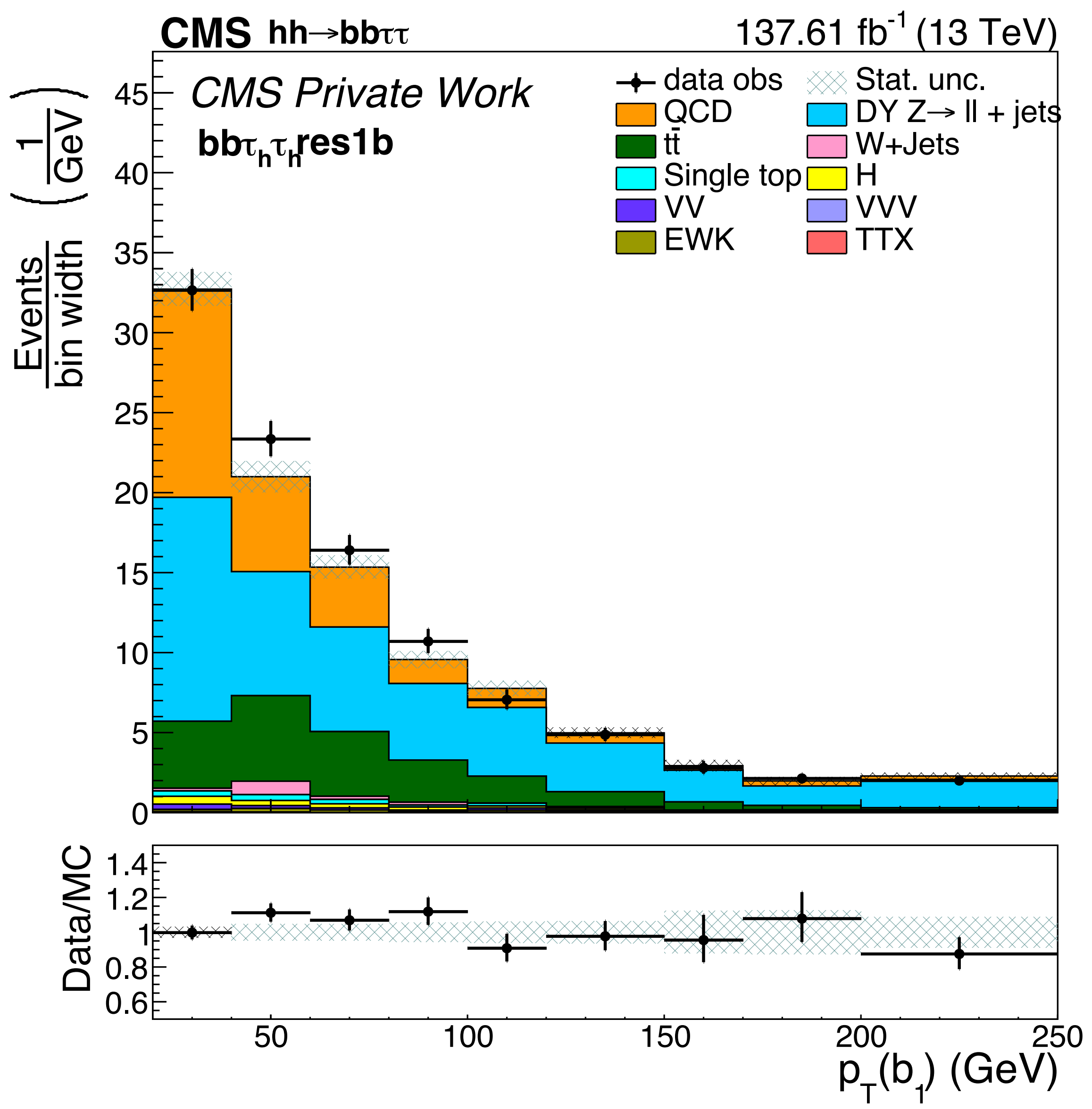
Data/MC comparison in SR: $v2p1$ VS $v2p5$

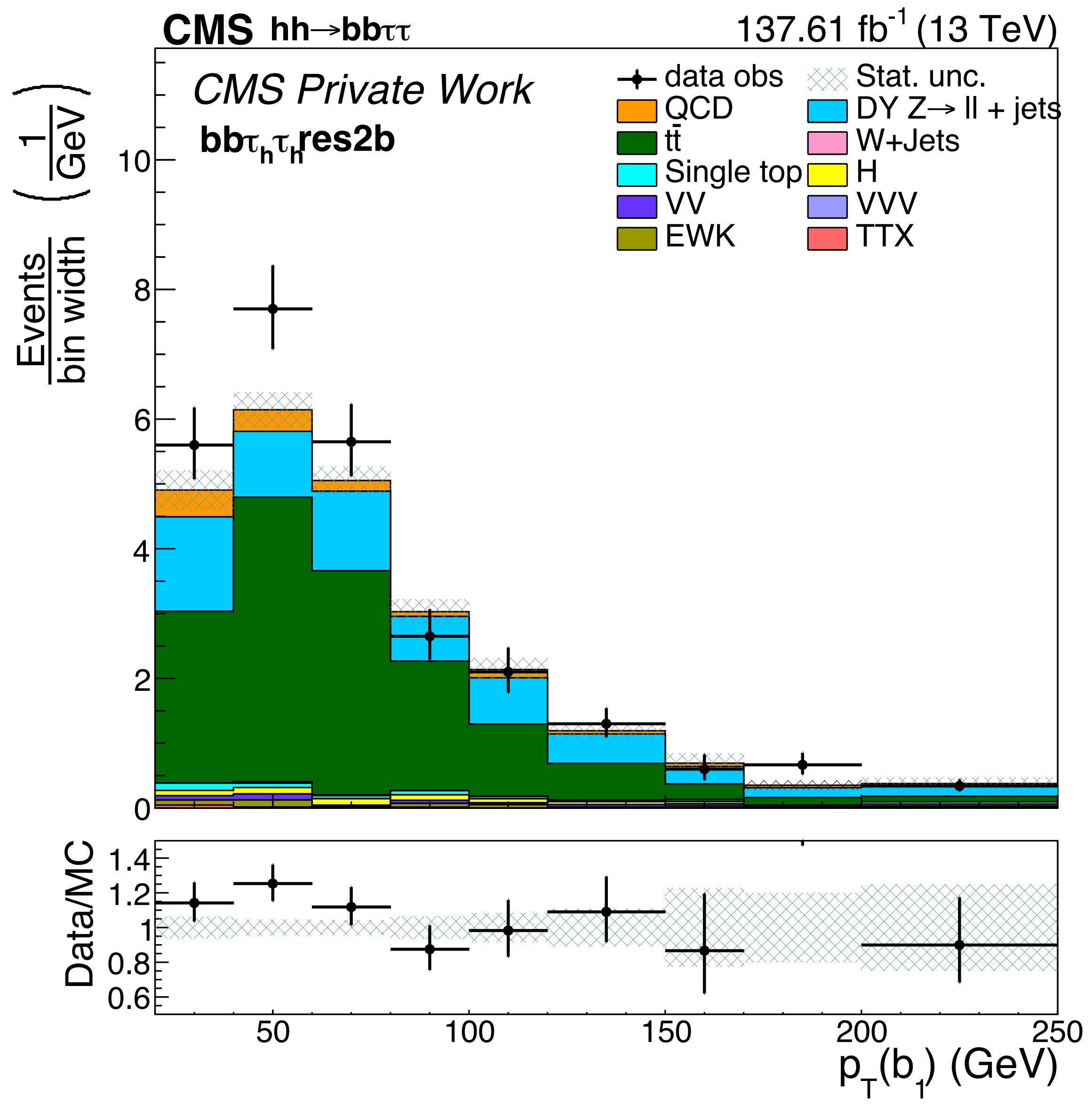


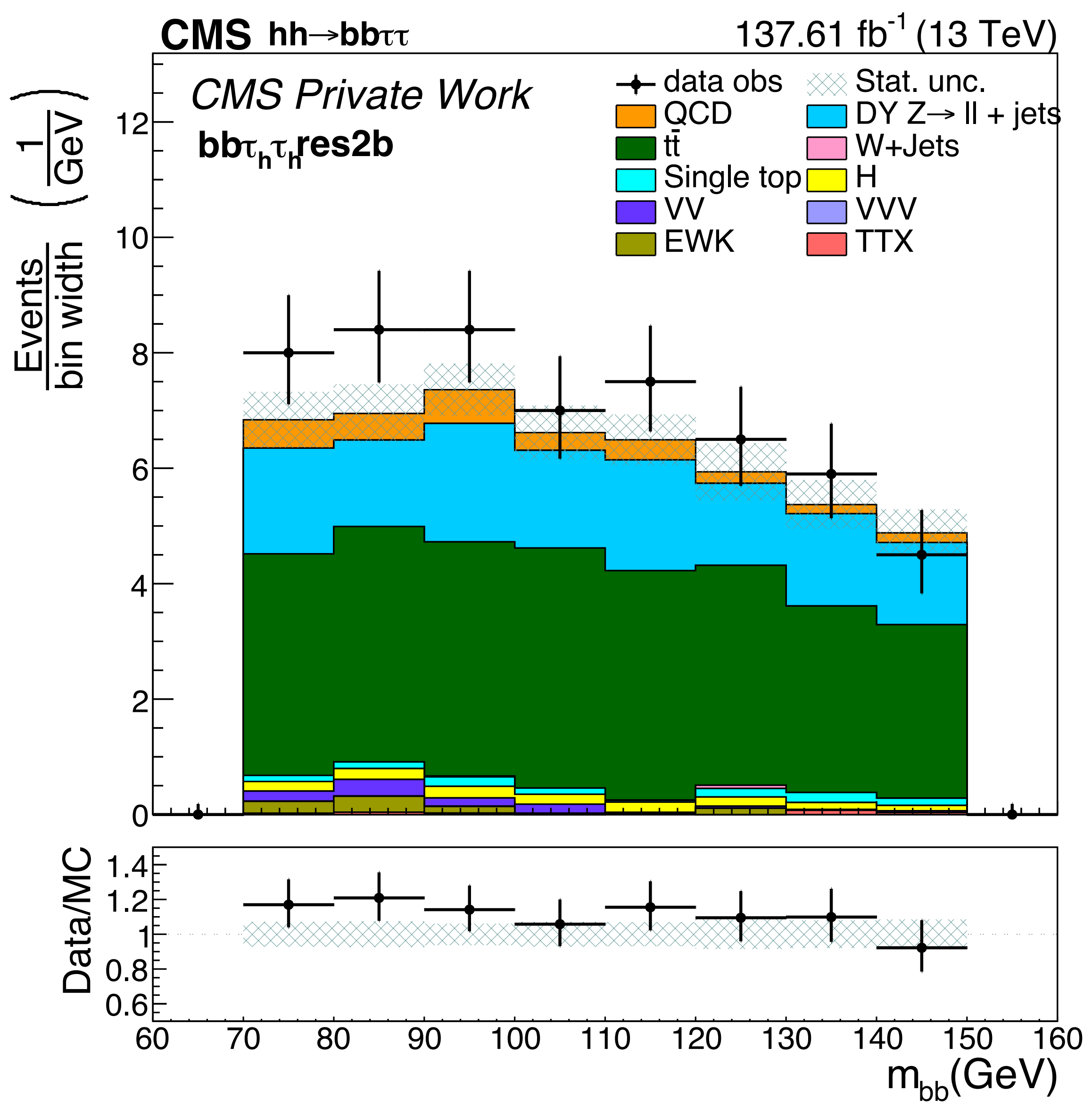




v2p1

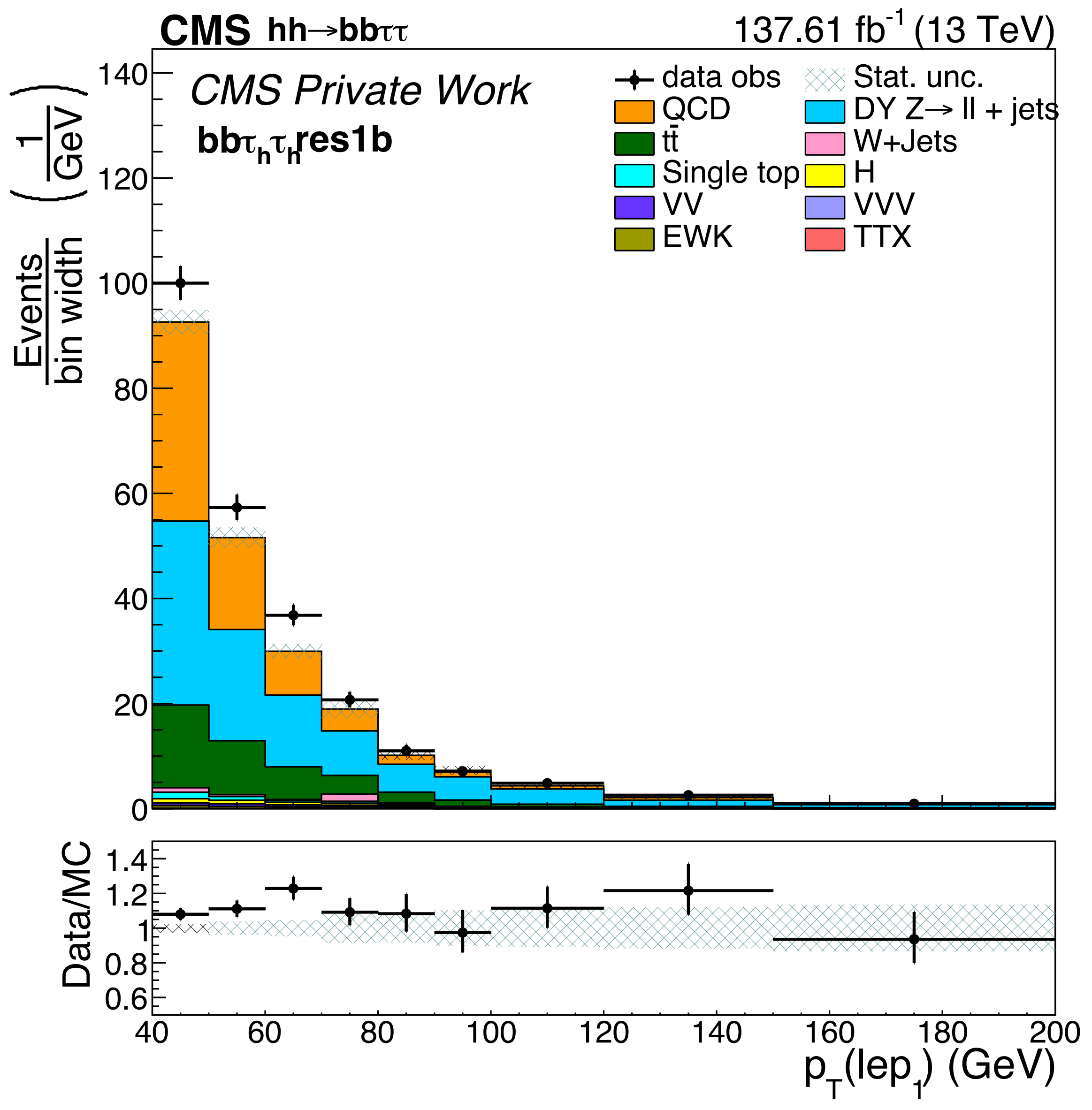


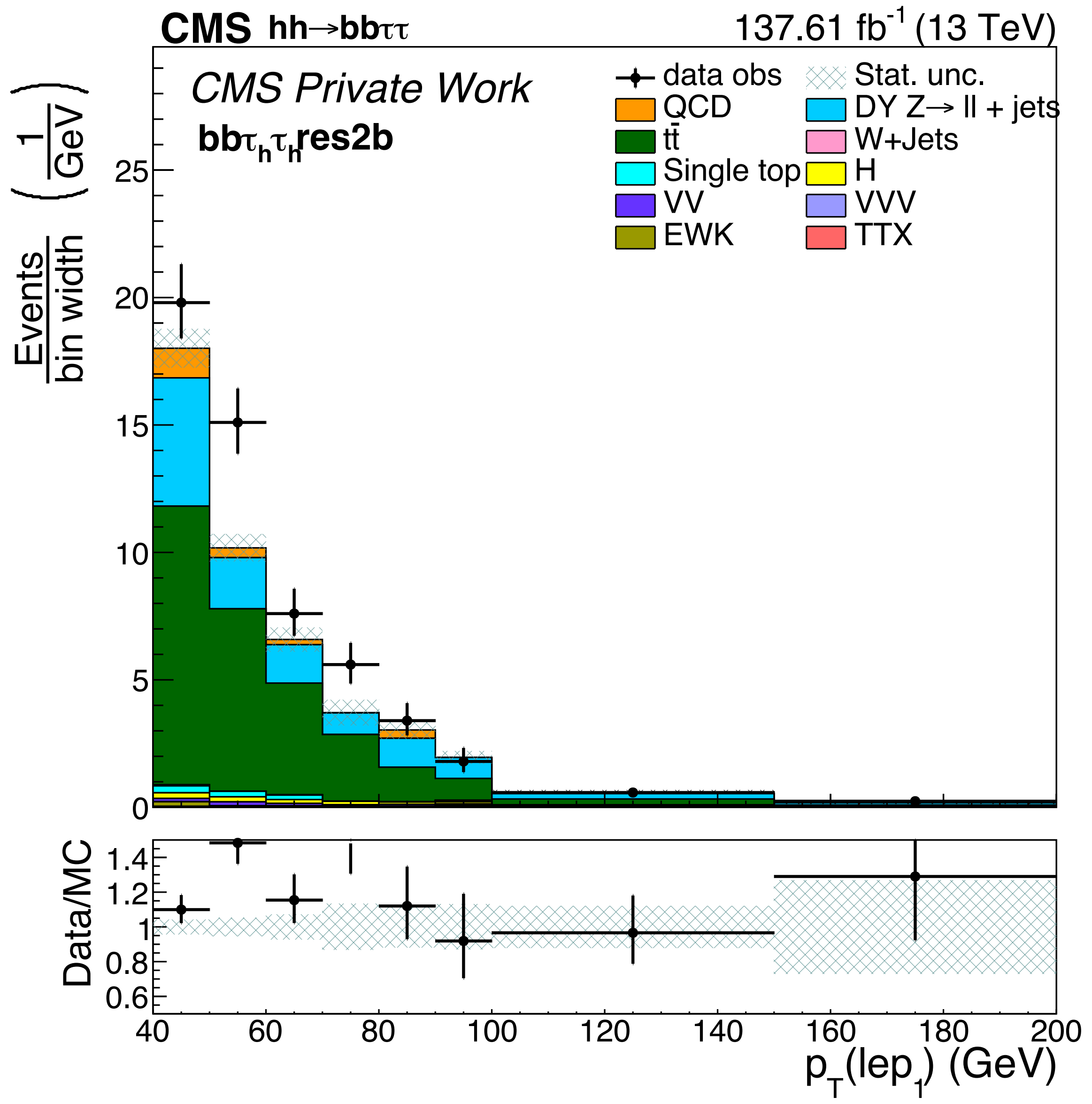


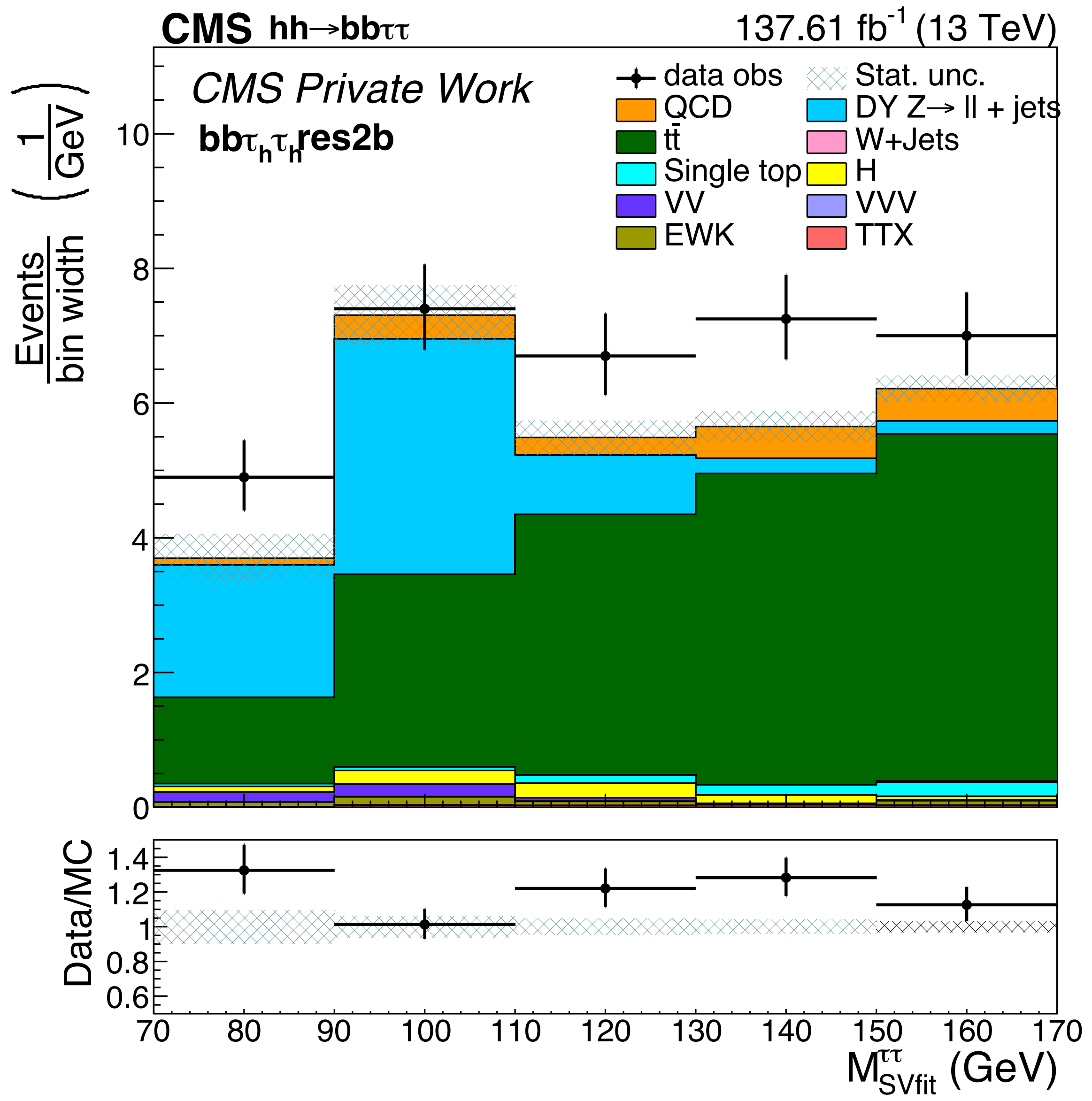


v2p5

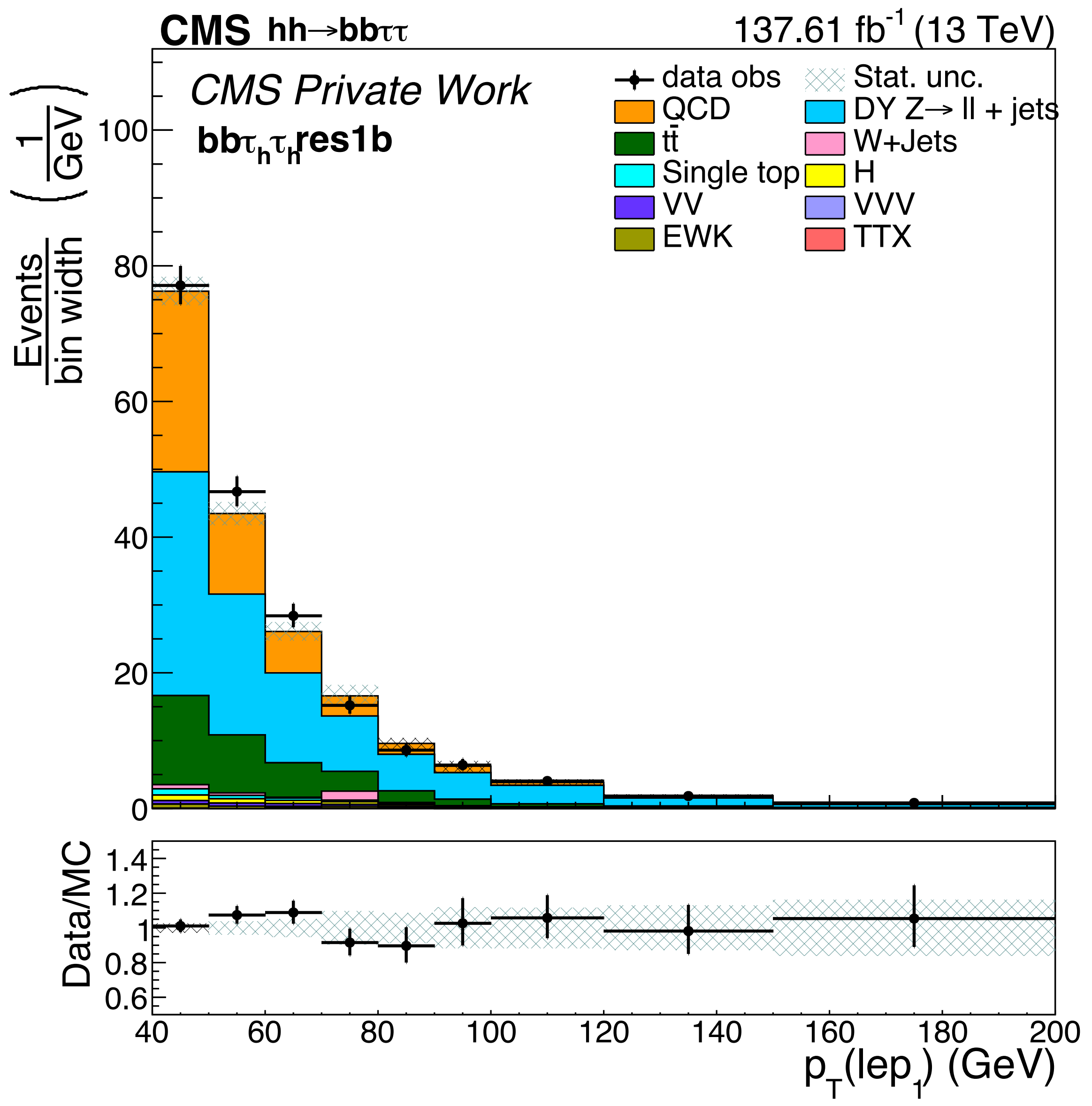
Data/MC comparison in SR: v2p1 VS v2p5

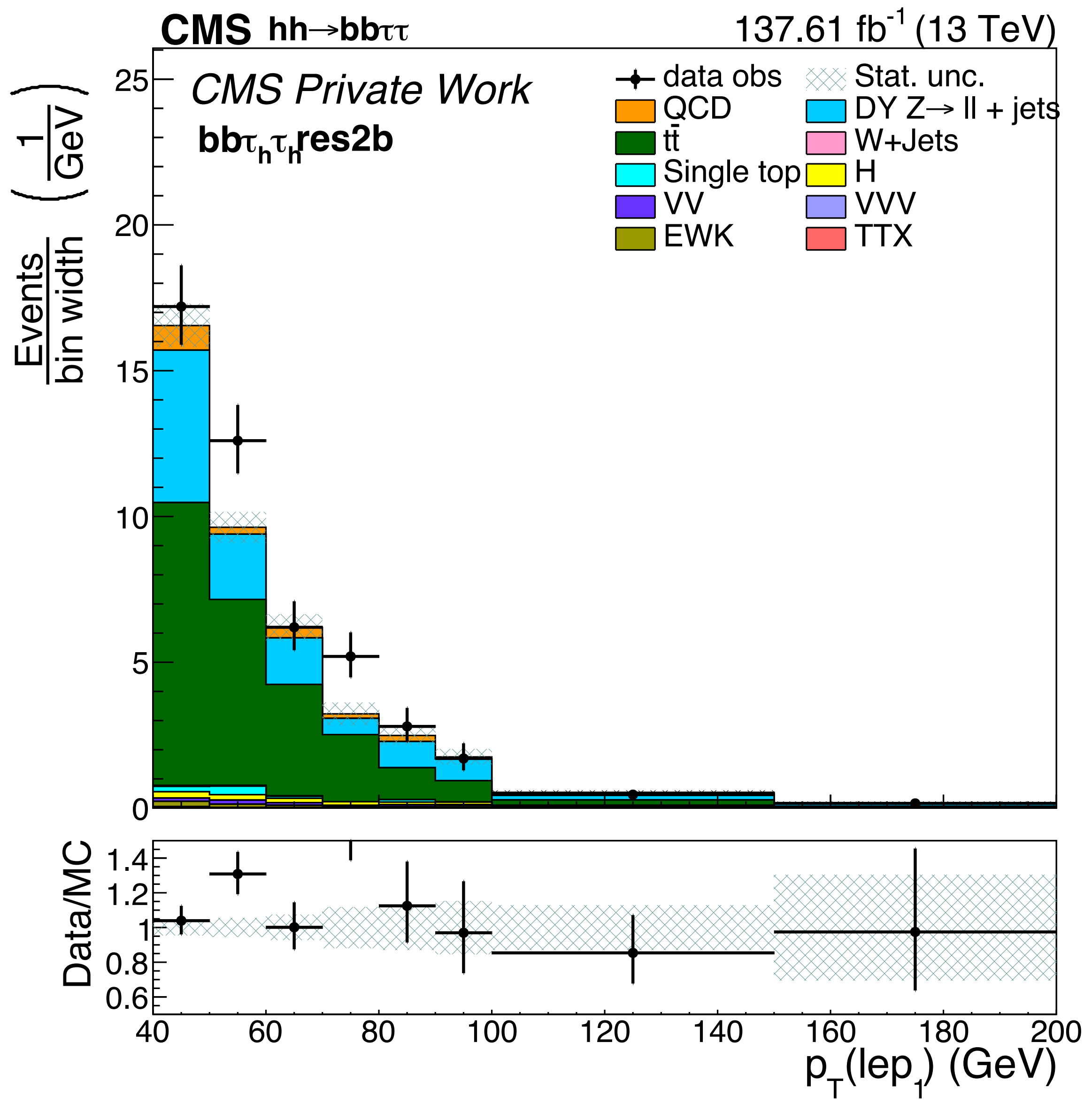


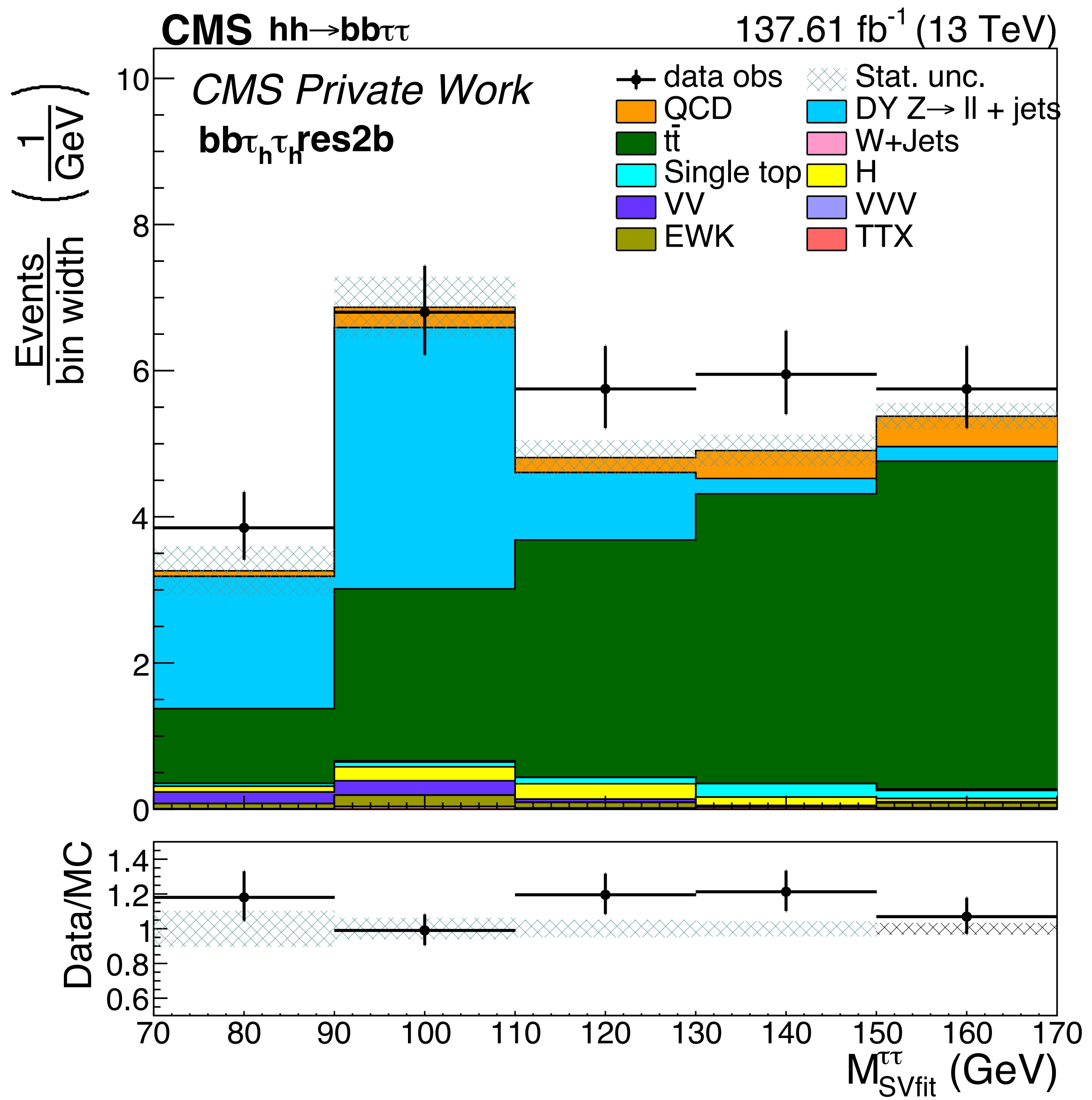




v2p1





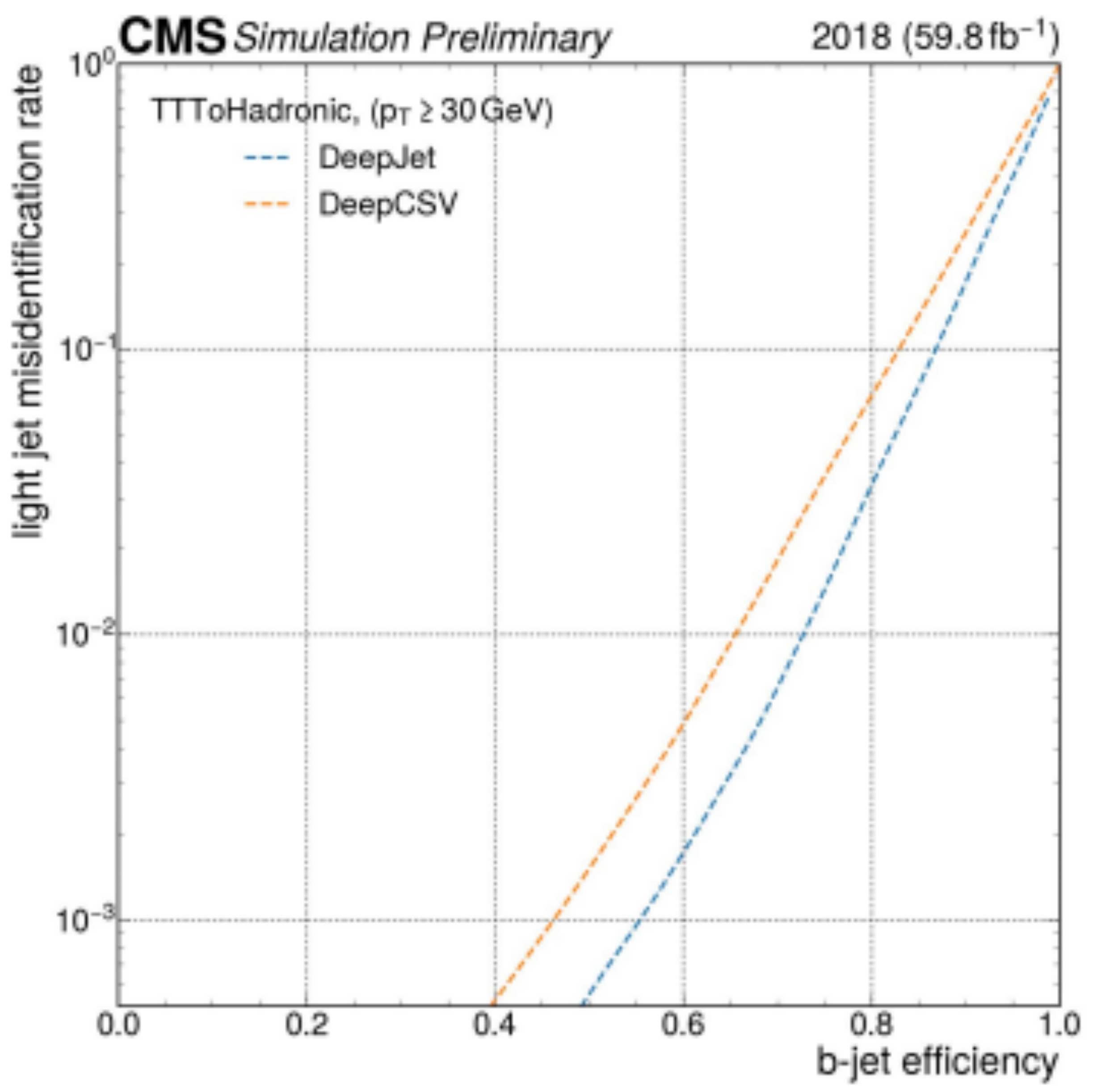


v2p5

MC corrections

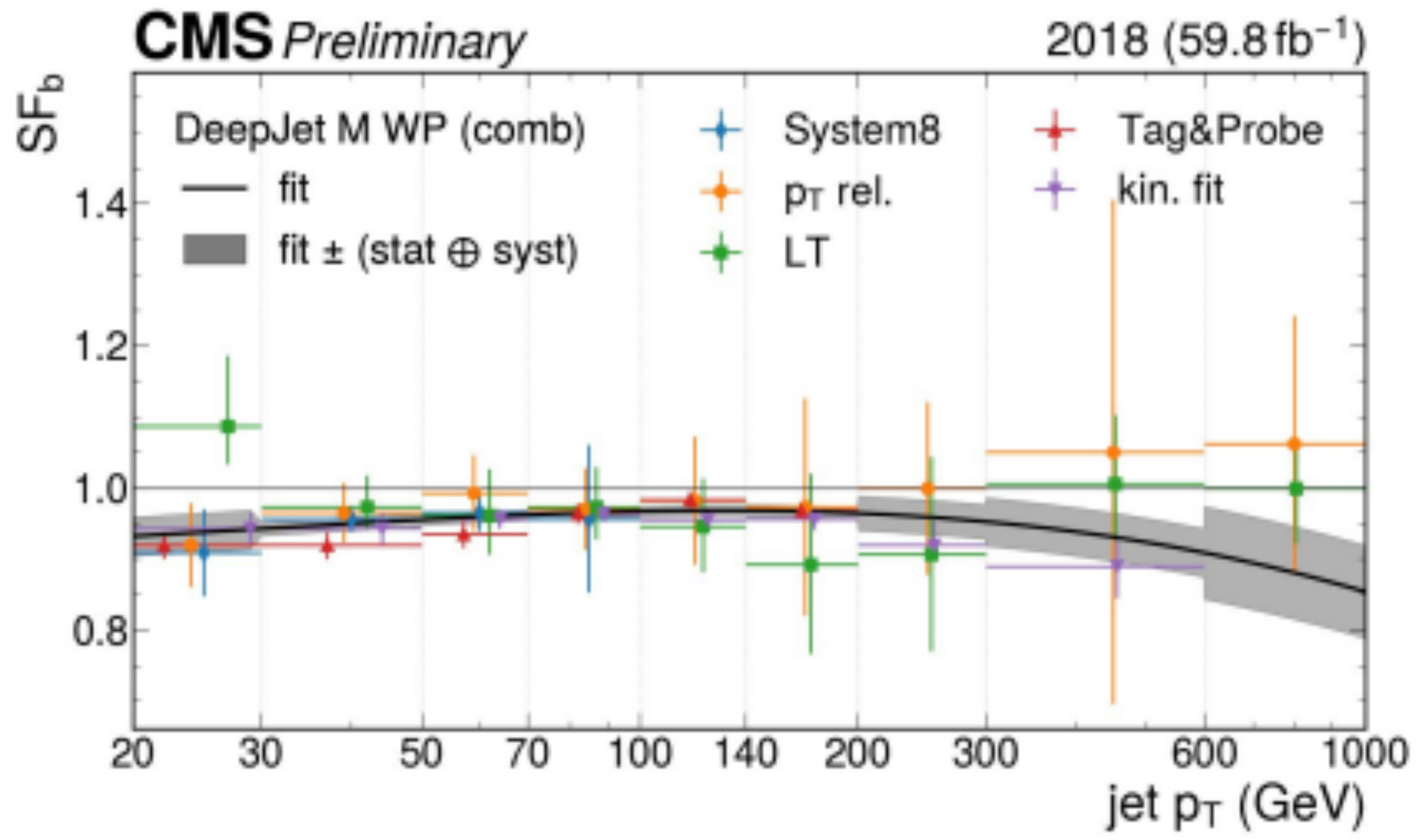
- Reco and ID techniques might have different behaviour in data w.r.t. MC.
  - To overcome this issue, MC are corrected with scale factors (SFs)
- SFs are affected by both statistics and systematics uncertainties
  - Introduction of systematic uncertainty to scaled MC samples

Yes..



DeepJet efficiency

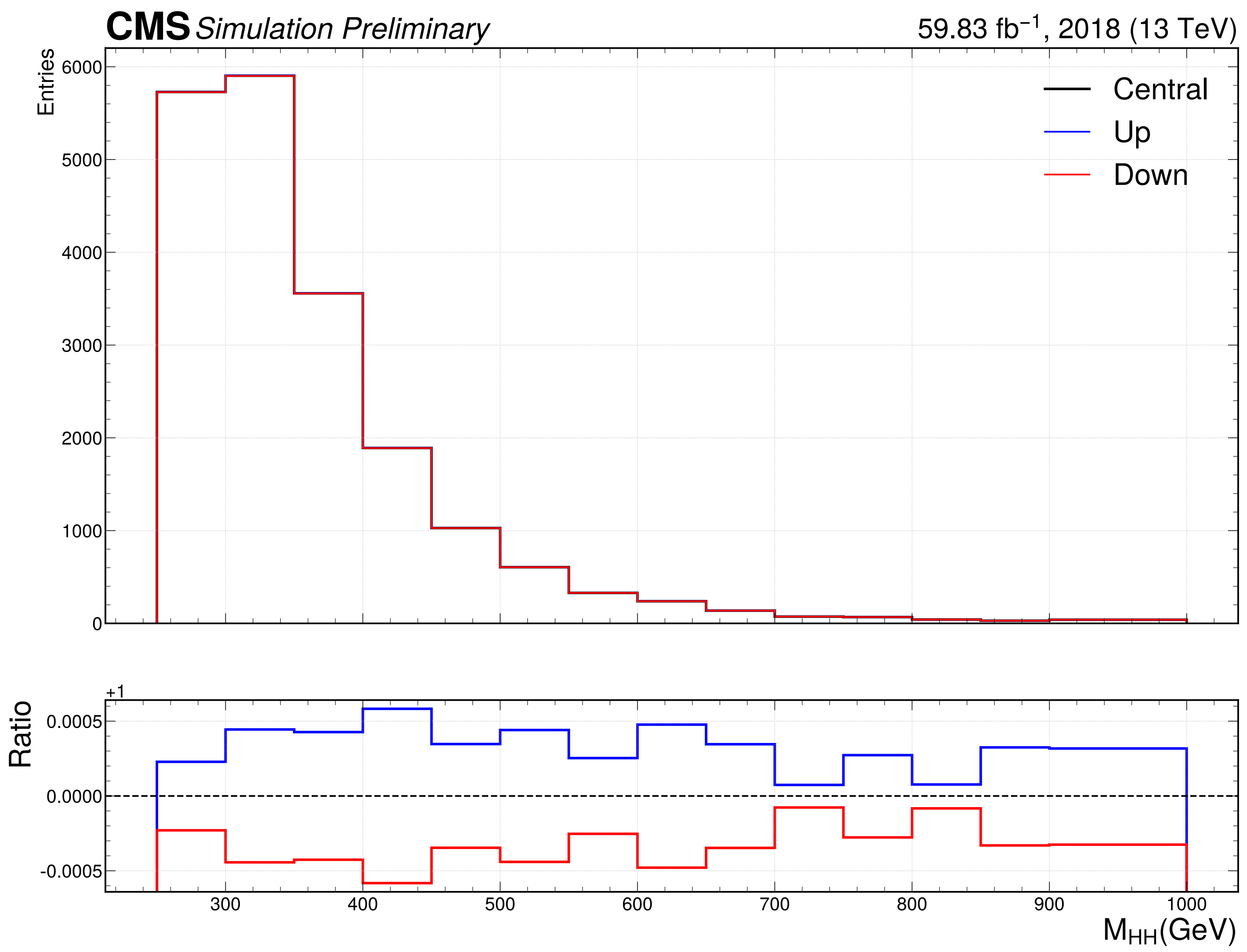
But..



DeepJet WP-based scale factor

Systematic uncertainties

- Many sources of uncertainty ($\sim 150!!$)
  - Known un-modeled effects
  - SF computation
  - Theory prediction
  - Analysis techniques
- Types:
  - **Normalisation**: changing the yield of processes
  - **Shape**: changing the shape of distribution
  - **Limited size of MC samples**
- Some sources may affect both normalisation and shape



reweight to match MC with data PU profile

Cross sections applied to MC samples (DY, tt, W+Jets, ...)

Trigger SF uncertainties

Ele/Mu ID and Reco SF

Luminosity measurement: different behaviour of detector over different data taking periods.

Branching Fractions
error on $H \rightarrow bb$: $^{+1.25\%}_{1.27\%}$
error on $H \rightarrow \tau\tau$: $\pm\, 1.65\,\%$

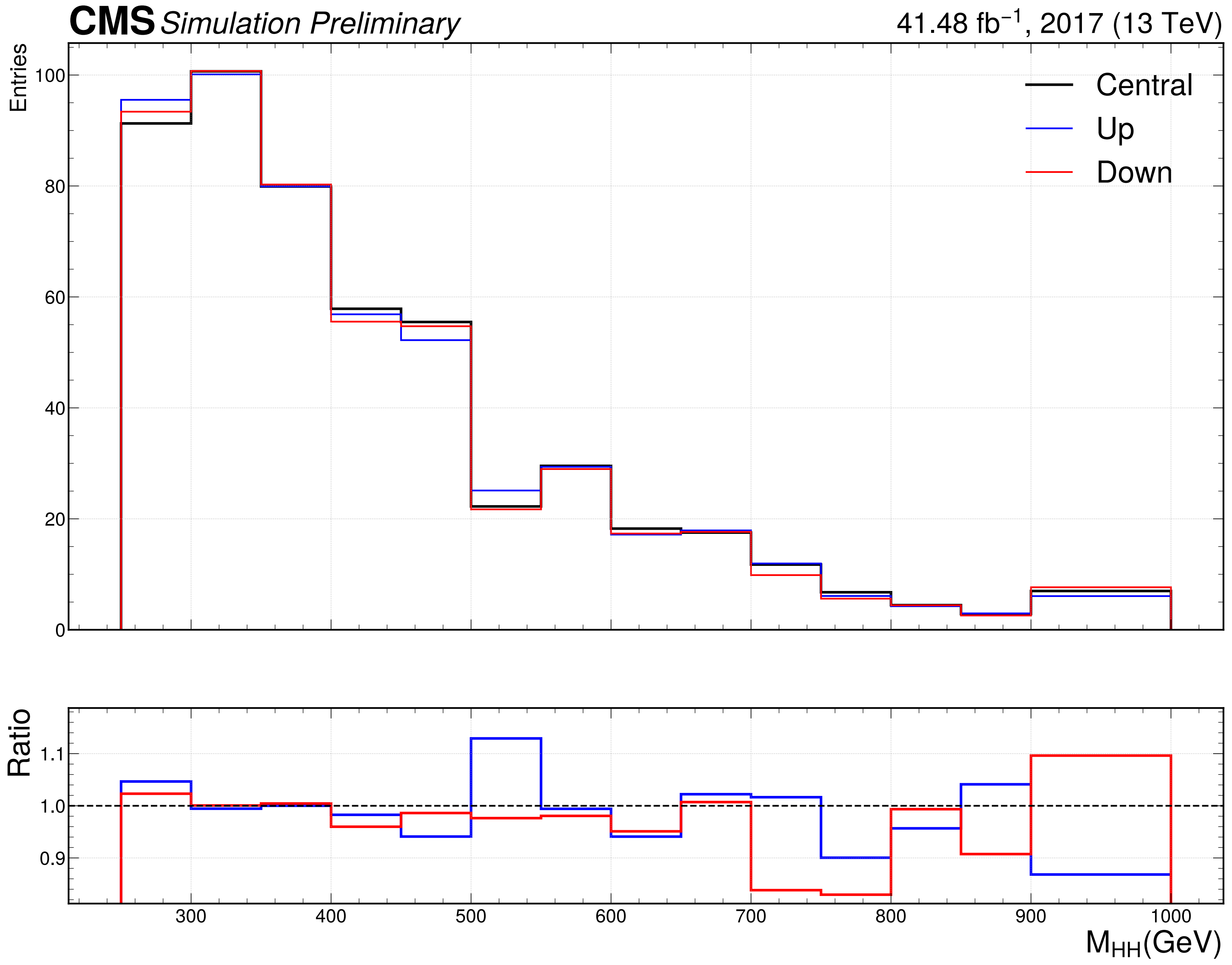
L1 prefiring Correction

Additional MuonID uncertainty (5%) To consider tightID inefficiency at high p_T .

Uncertainty on the yield used to estimate the correction factor adopted in the ABCD method (Nc).

Systematic uncertainties

- Many sources of uncertainty ($\sim 150!!$)
  - Known un-modeled effects
  - SF computation
  - Theory prediction
  - Analysis techniques
- Types:
  - **Normalisation**: changing the yield of processes
  - **Shape**: changing the shape of distribution
  - **Limited size of MC samples**
- Some sources may affect both normalisation and shape



Uncertainty in events where jets are misidentified as hadronic taus

Jet Energy Reconstruction (JER)

Jet Energy Scale (JES)

Uncertainty on τ energy distributions

DeepTau ID SF

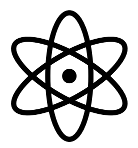
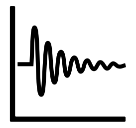
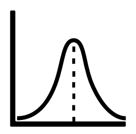
ParticleNet SF

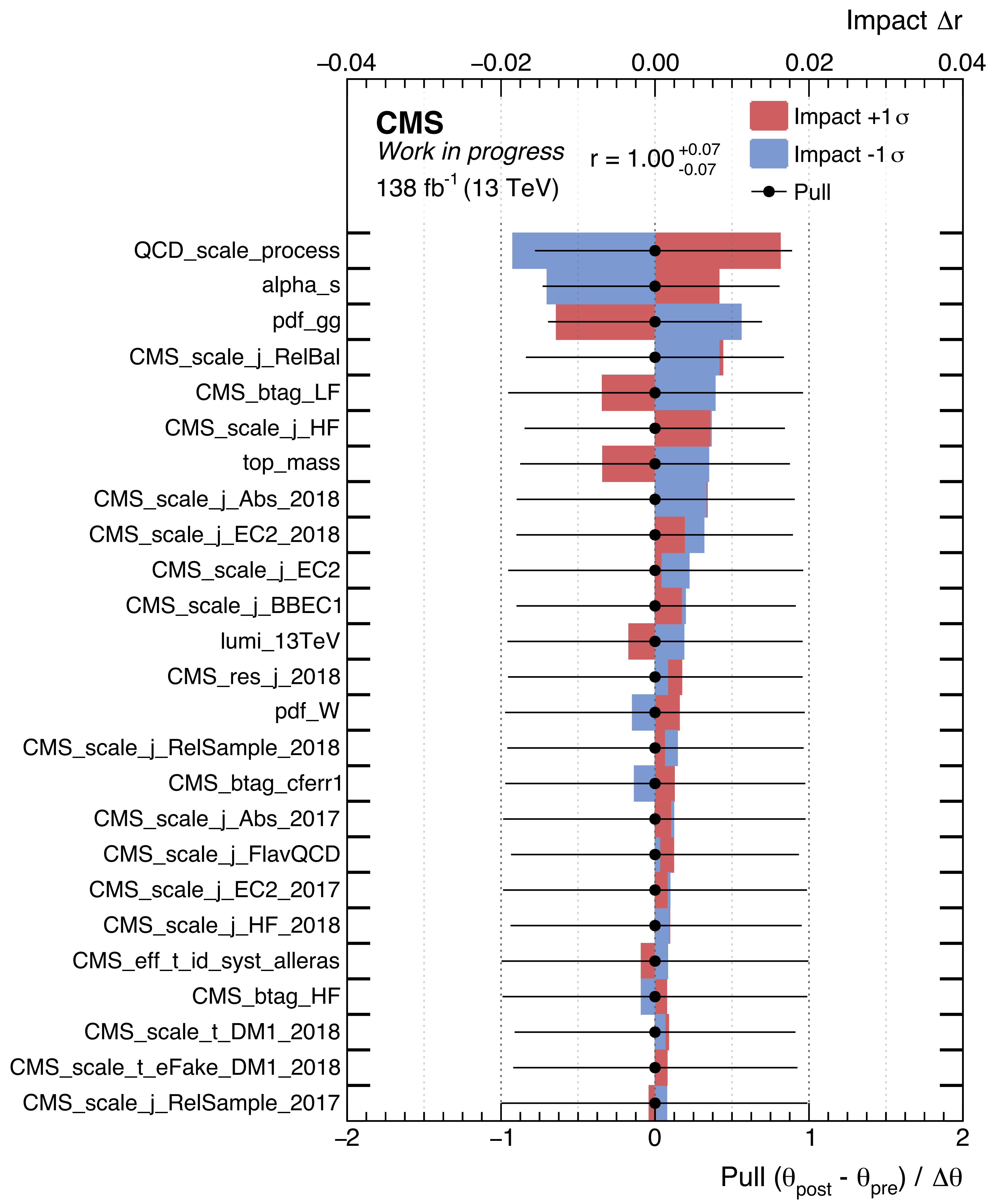
Pile-Up Jet ID SF

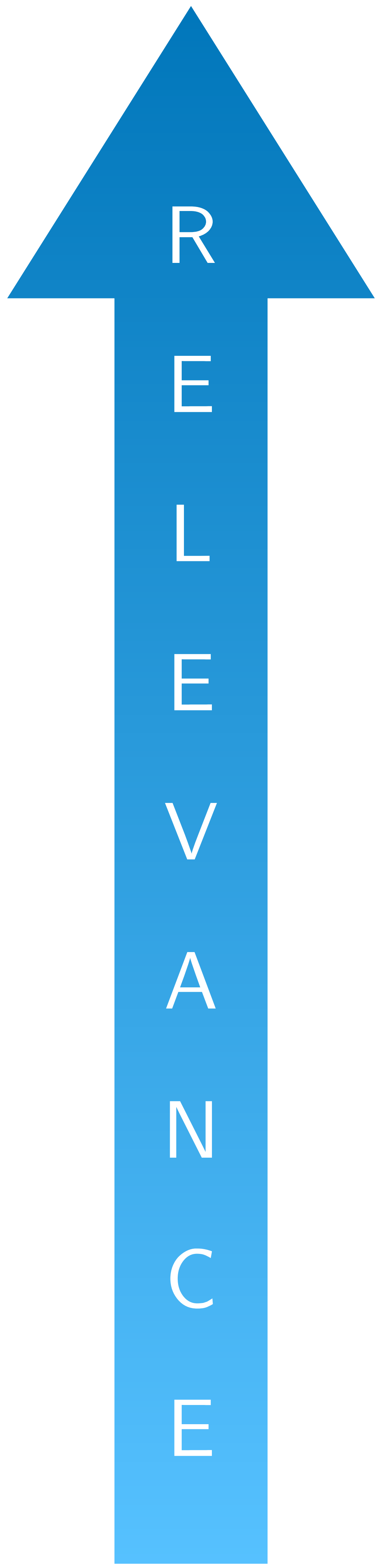
QCD shape uncertainty

b-tagging shape-calibrated SF

Systematic uncertainties

- Many sources of uncertainty ($\sim 150!!$)
  - Known un-modeled effects 
  - SF computation 
  - Theory prediction 
  - Analysis techniques 
- Types:
  - **Normalisation**: changing the yield of processes
  - **Shape**: changing the shape of distribution
  - **Limited size of MC samples**
- Some sources may affect both normalisation anc shape

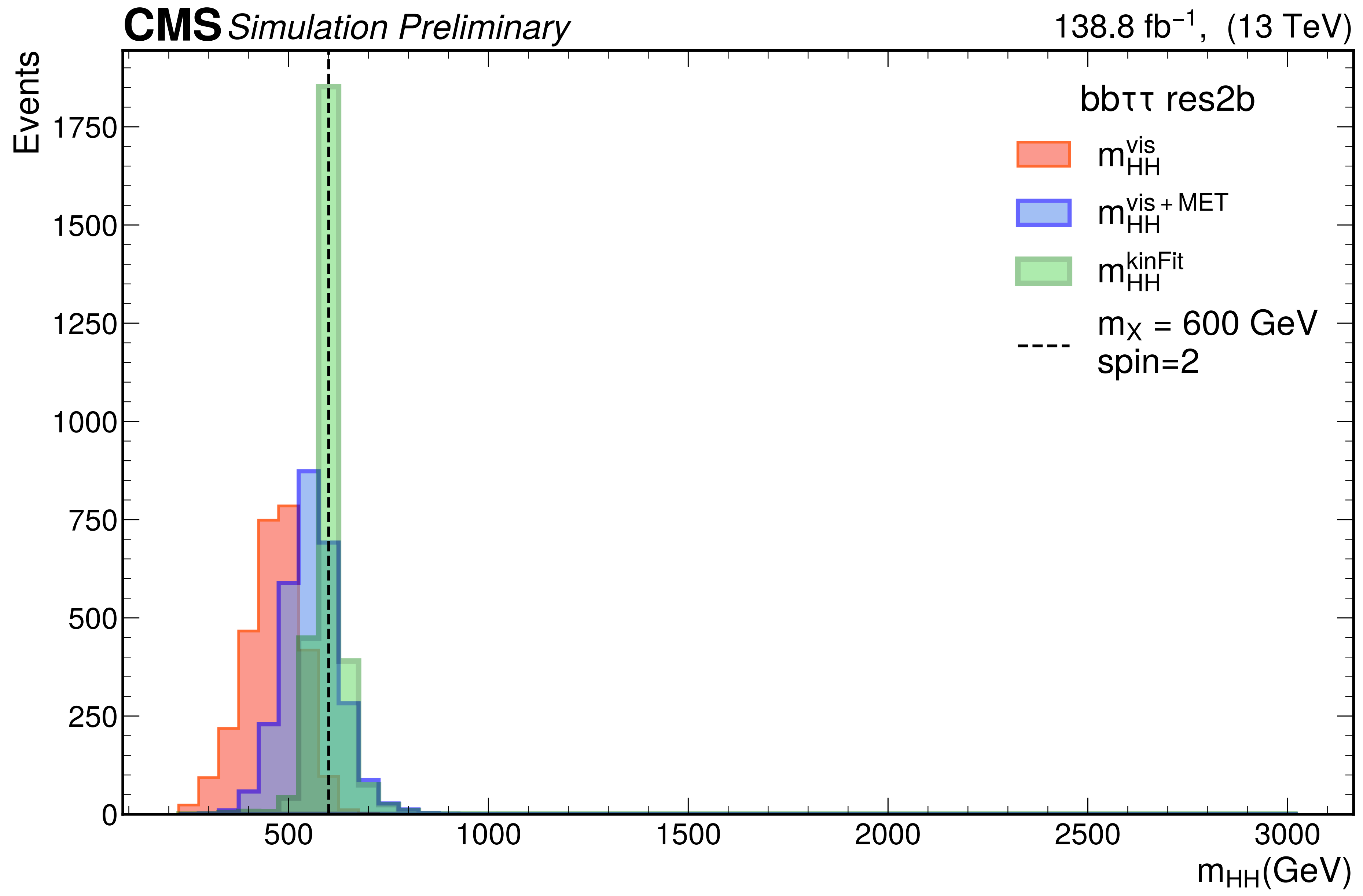


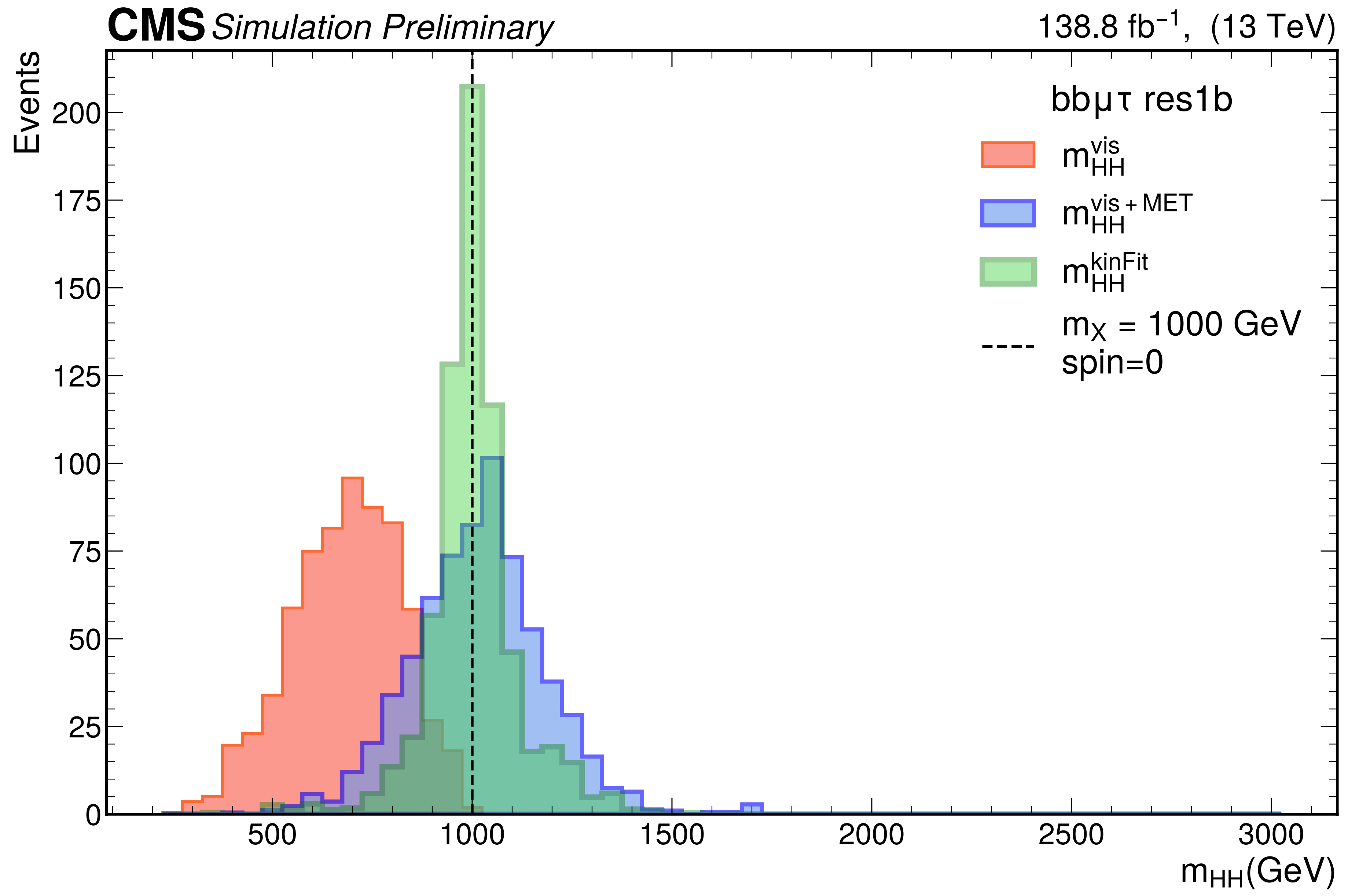


Signal extraction observable: HH-KinFit Mass

- Search for resonances: invariant mass of products is one of the most sensitive observables
- MET → loss of resolution of m_{HH} distribution
- **Kinematic fit** applied to correct for this effect:
  - Global χ^2 fit as a function of the b-jets and recoil energies
  - Minimising it to obtain corrected decay products 4-momenta used for the invariant mass calculation

$$m_X = m_{HH}^{kinFit} = \left\| p_{\tau_1}^{fit} + p_{\tau_2}^{fit} + p_{b_1}^{fit} + p_{b_2}^{fit} \right\|$$



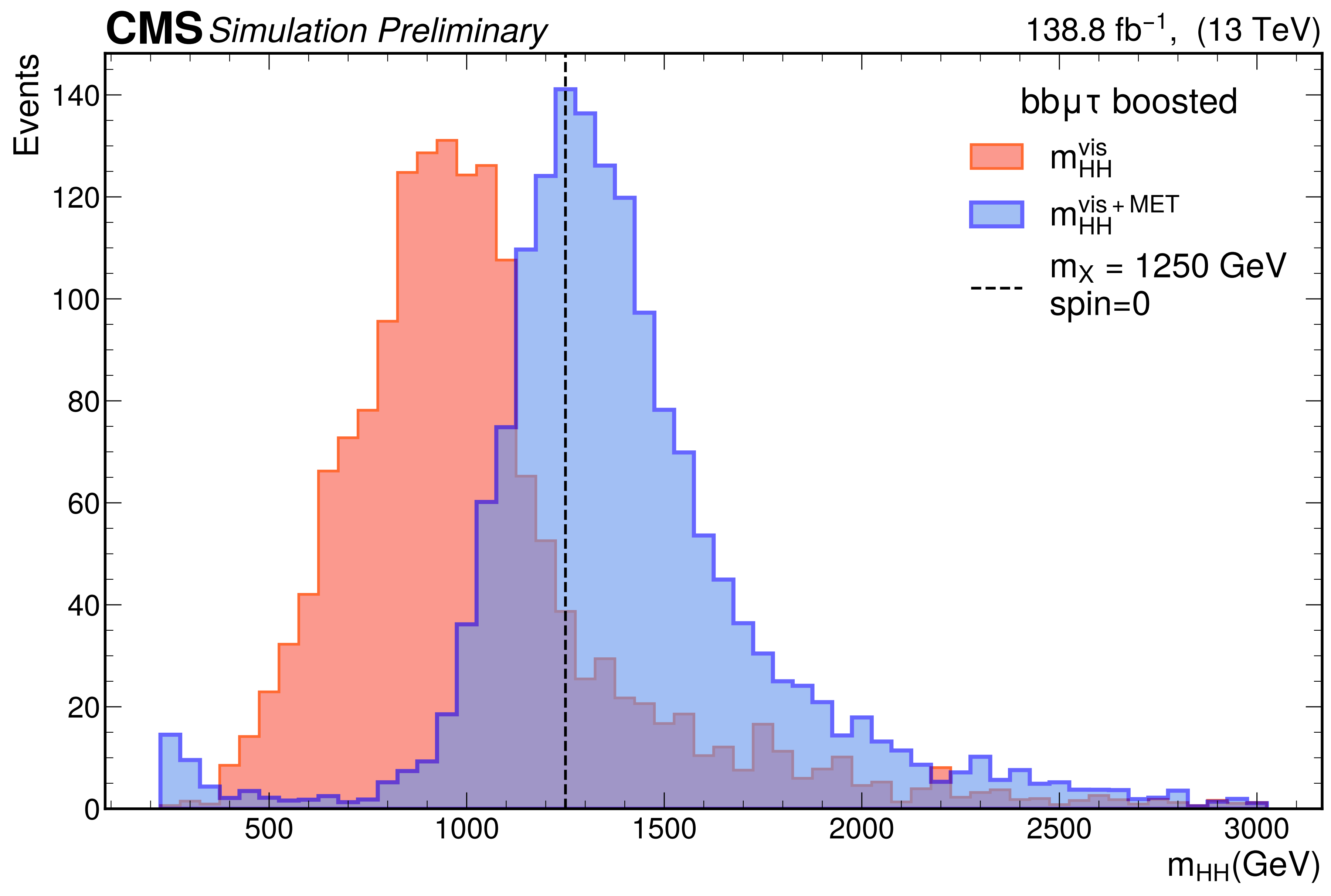


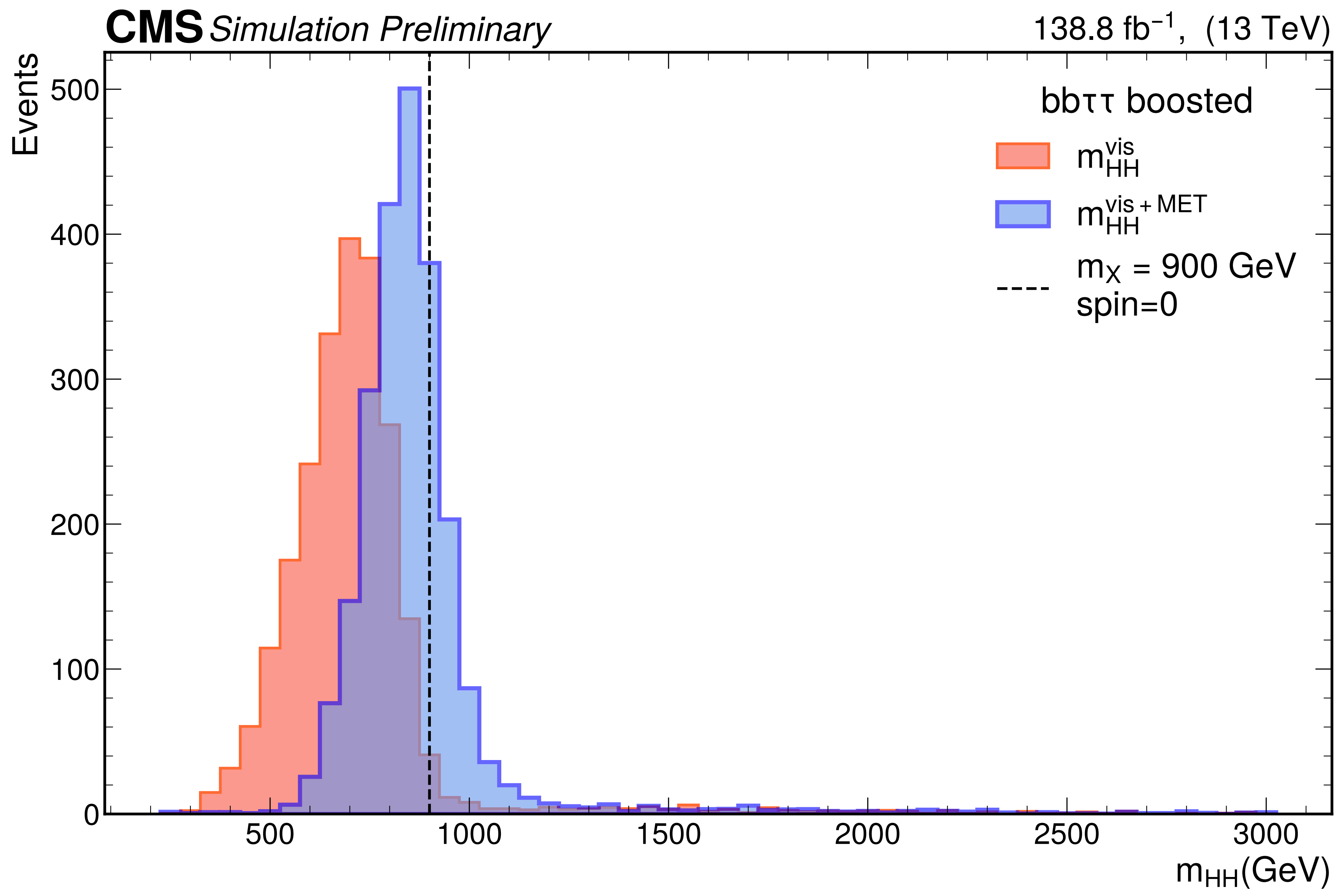
Signal extraction observable in boosted category

- Kinematic fit not implemented in boosted category due to computational complexity
- $m_{HH}^{vis+MET}$ is used
  - despite its limitations in the resolved categories, it achieves best performance in boosted category
  - particularly effective at high mass

$$p_{HH}^{vis+MET} = p_{H(bb)} + p_{H(\tau\tau)} + p_{miss}$$

$$m_X = m_{HH}^{vis+MET} = \left\| p_{HH}^{vis+MET} \right\|$$



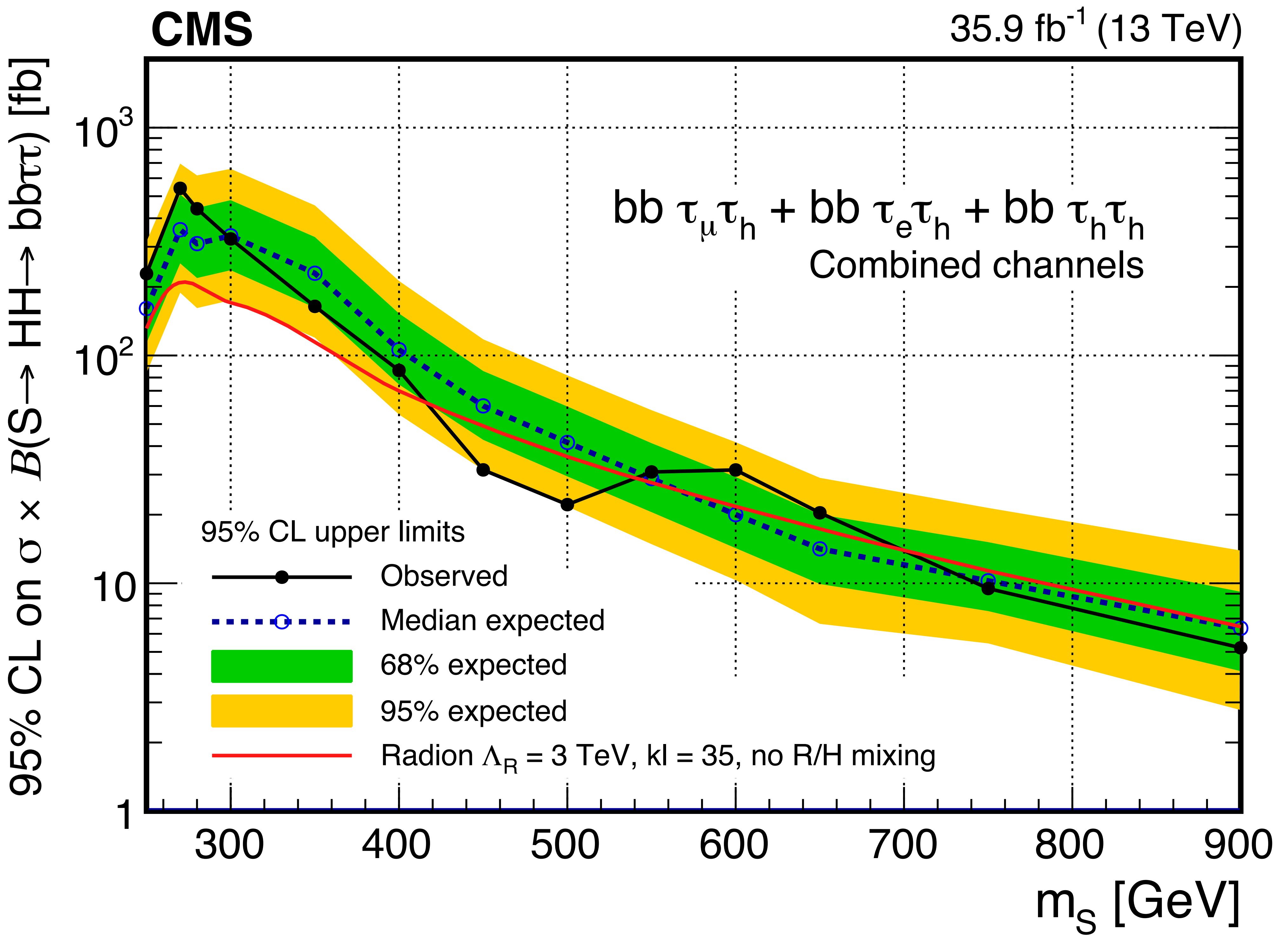


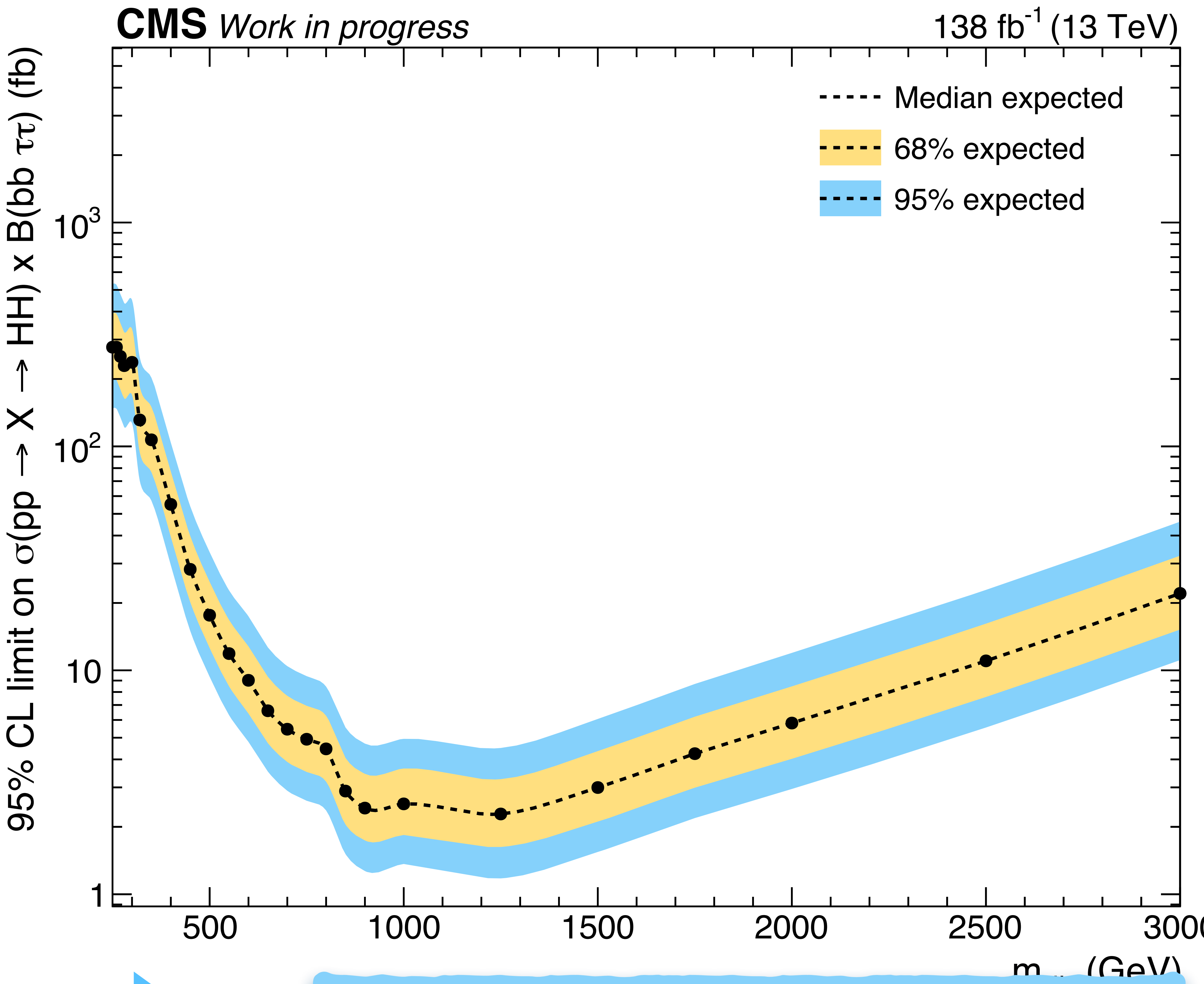
Fit procedure

- Binned Maximum Likelihood Fit using the HH-kinFit/invariant mass to build the test statistics
  - Calculated separately for the different selections using DeepTauV2p1 and DeepTauV2p5
    - 25 mass points
    - spin-0 and spin-2 hypotheses
    - $e\tau_h$, $\mu\tau_h$, and $\tau_h\tau_h$ channels for DeepTauV2p1 and $\tau_h\tau_h$ channel for DeepTauV2p5
    - 4 eras (2016H, 2016, 2017, 2018)
- Signal extraction: CLs statistical method (details in backup) - used by ATLAS & CMS
- Blind analysis: final fit using MC only

Results - Spin 0

- Model-independent 95% CL exclusion limits for a spin-0 resonance production cross section times BR(X$\rightarrow$ HH $\rightarrow$ bb$\tau\tau$) shown combining channels



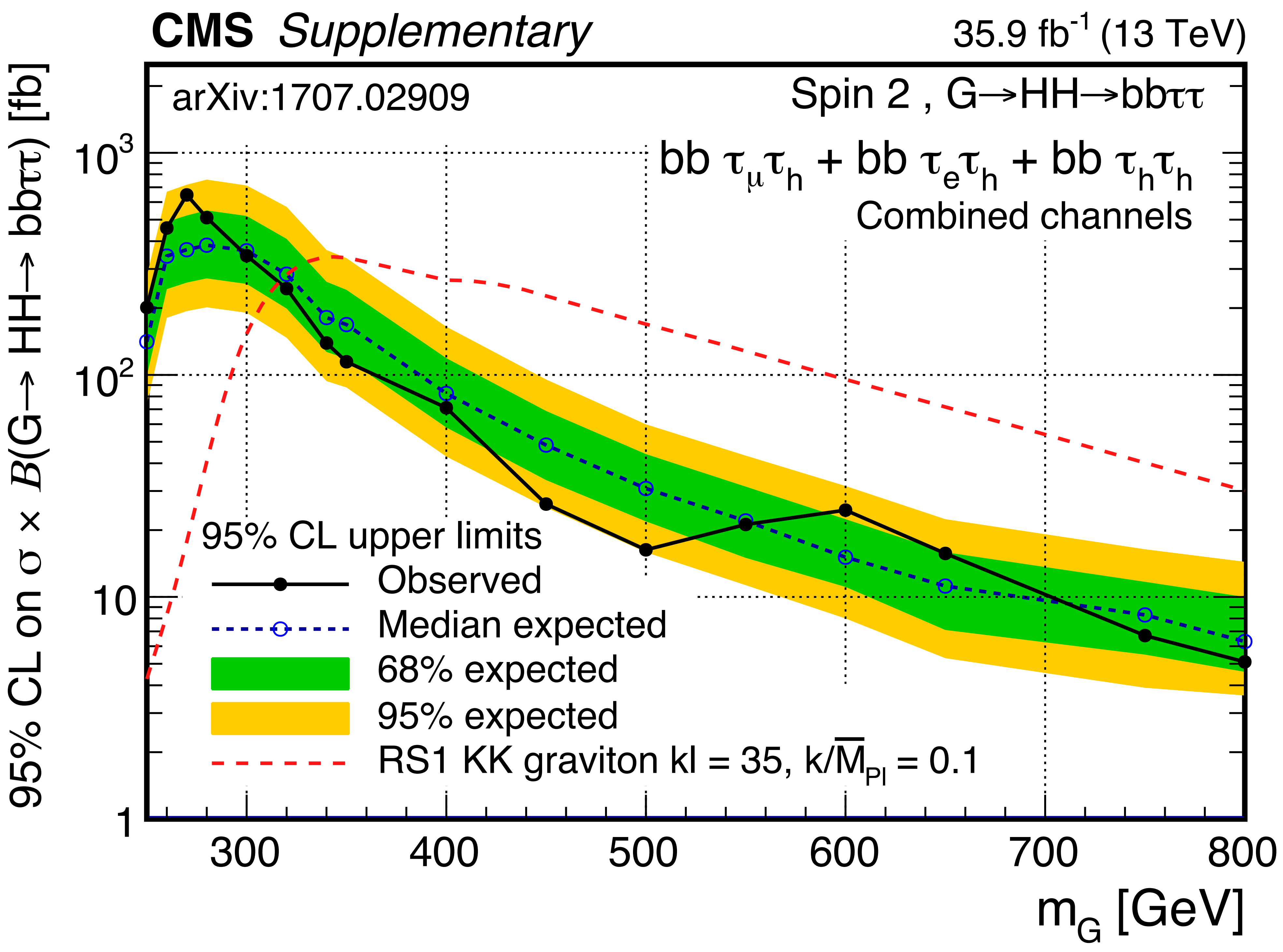


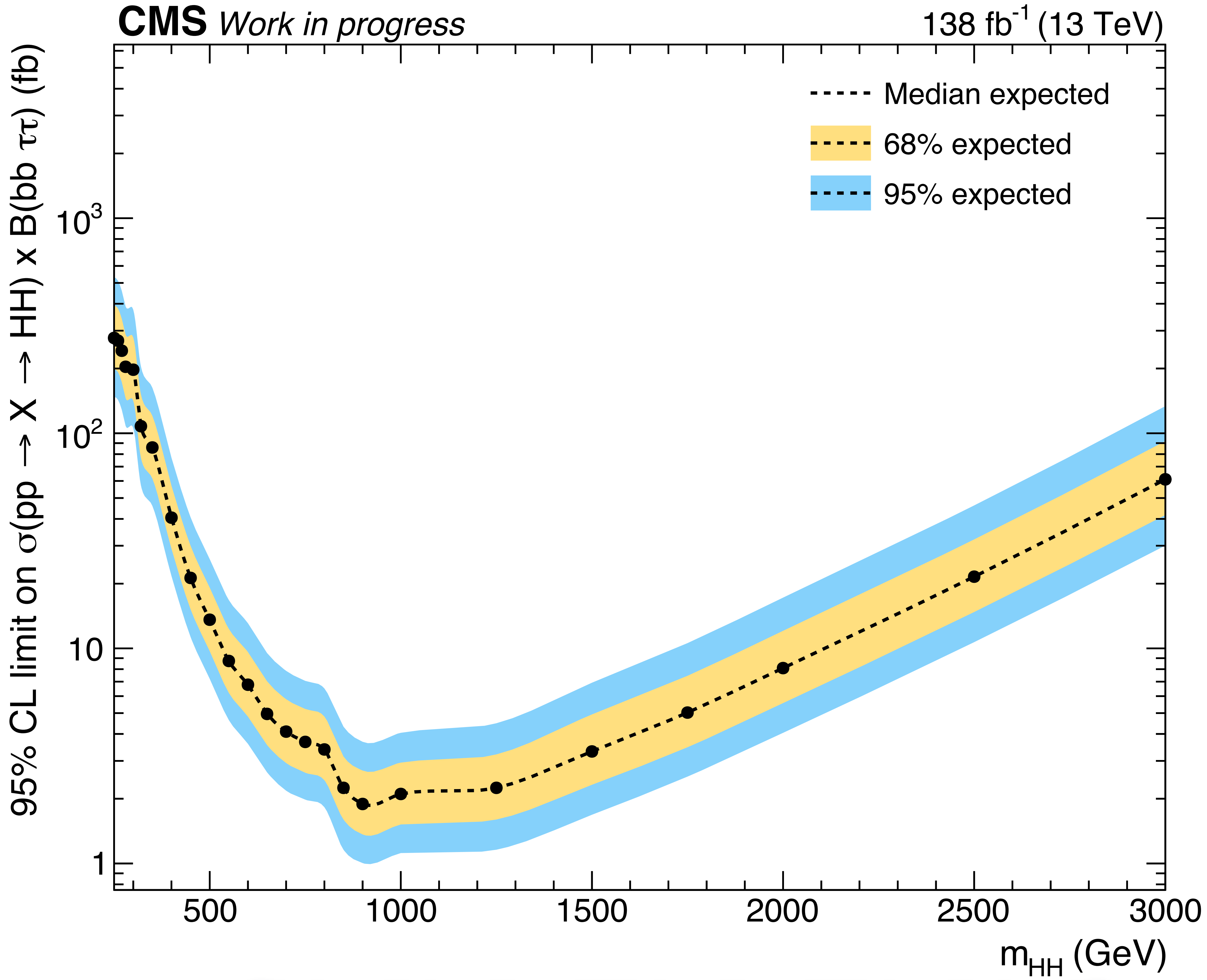
- Increased Run II statistics
- Improvements in analysis techniques
- wider range w.r.t. previous search

→ **Higher signal sensitivity**

Results - Spin 2

- Model-independent 95% CL exclusion limits for a spin-2 resonance production cross section times BR(X→ HH → bbττ) shown combining channels



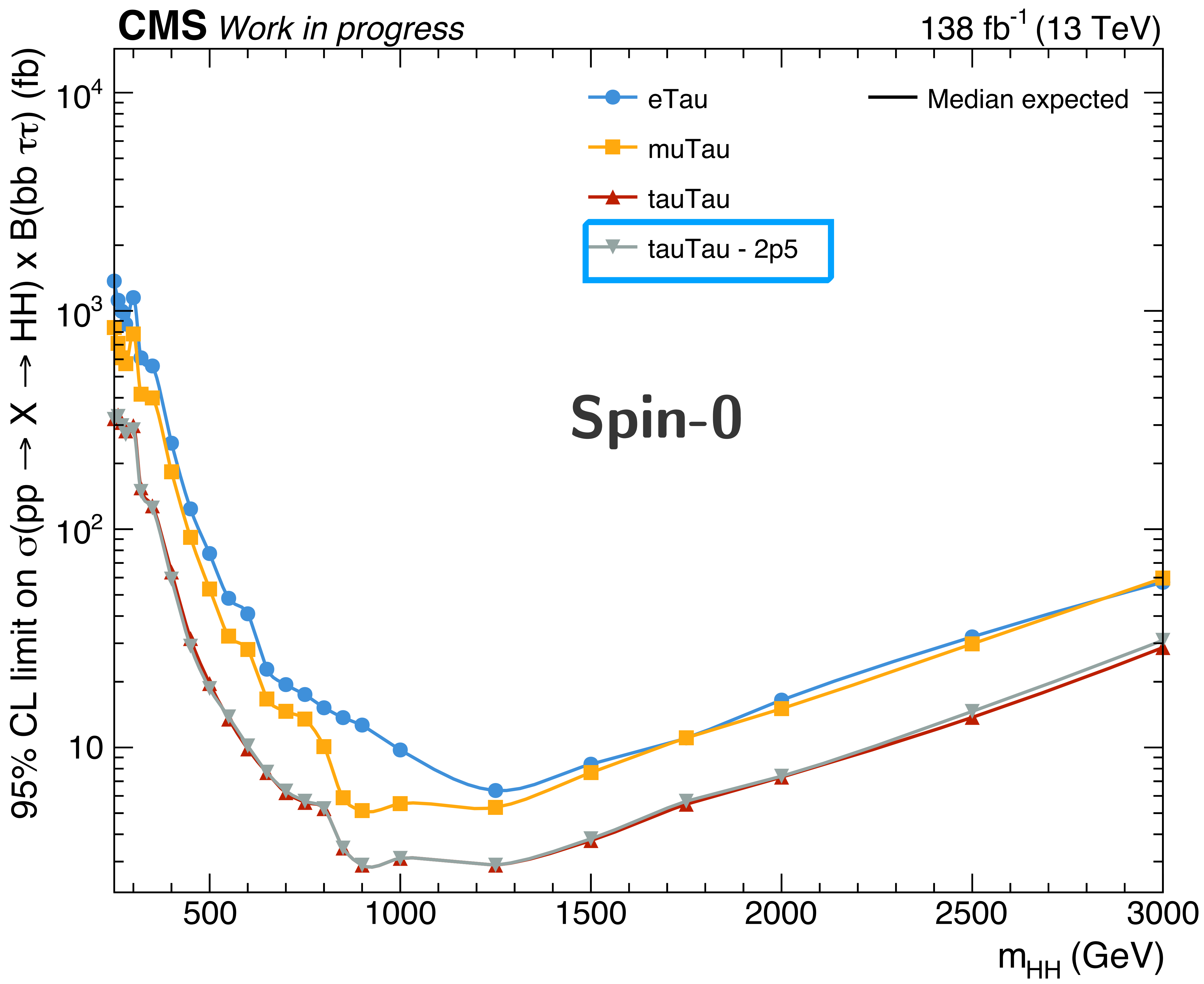


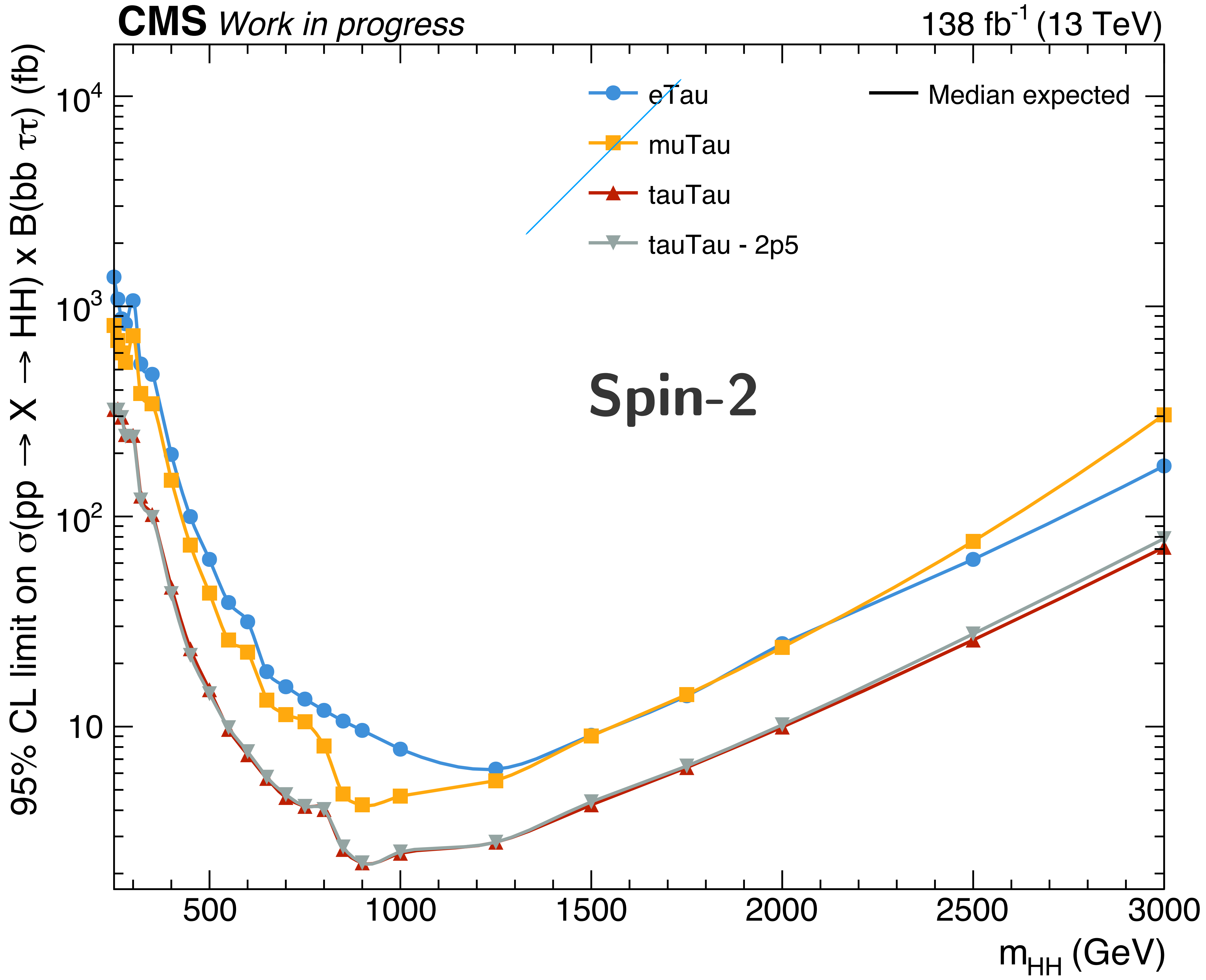
- Increased Run II statistics
- Improvements in analysis techniques
- wider range w.r.t. previous search

→ **Higher signal sensitivity**

Results: DeepTau 2p1 vs 2p5

- Model-independent 95% CL exclusion limits for resonance production cross section times BR($X \to HH \to bb\tau\tau$) shown in each channel and compared to DeepTau2p5 channel





Conclusions and outlooks

- The resonant $X \to HH \to bb\tau\tau$ search with the full Run II dataset has been presented
- Model independent 95% CL exclusion limits set on resonance production cross section with blind approach (MC only)

Key highlight: hadronic τ identification with ML-based *DeepTau algorithm*

- Comparison between two training versions (v2p1 VS v2p5) **for the first time**
  - Relevant improvement in data/MC agreement for v2p5 while keeping similar sensitivity
  - Crucial validation for ongoing Run 3

- The analysis is under review process by the CMS collaboration and will be published after the approval
- Developing and testing advanced tools to enhance sensitivity is crucial to
  - Enable more sensitive searches, wider phase space coverage
  - .. And ultimately *measure* the SM Higgs self-coupling

Thank you for the attention.

Questions?

A special thanks to Agnese and Konstantin for supporting and guiding me during my Ph.D.

Additional material

Limits of the Standard Model

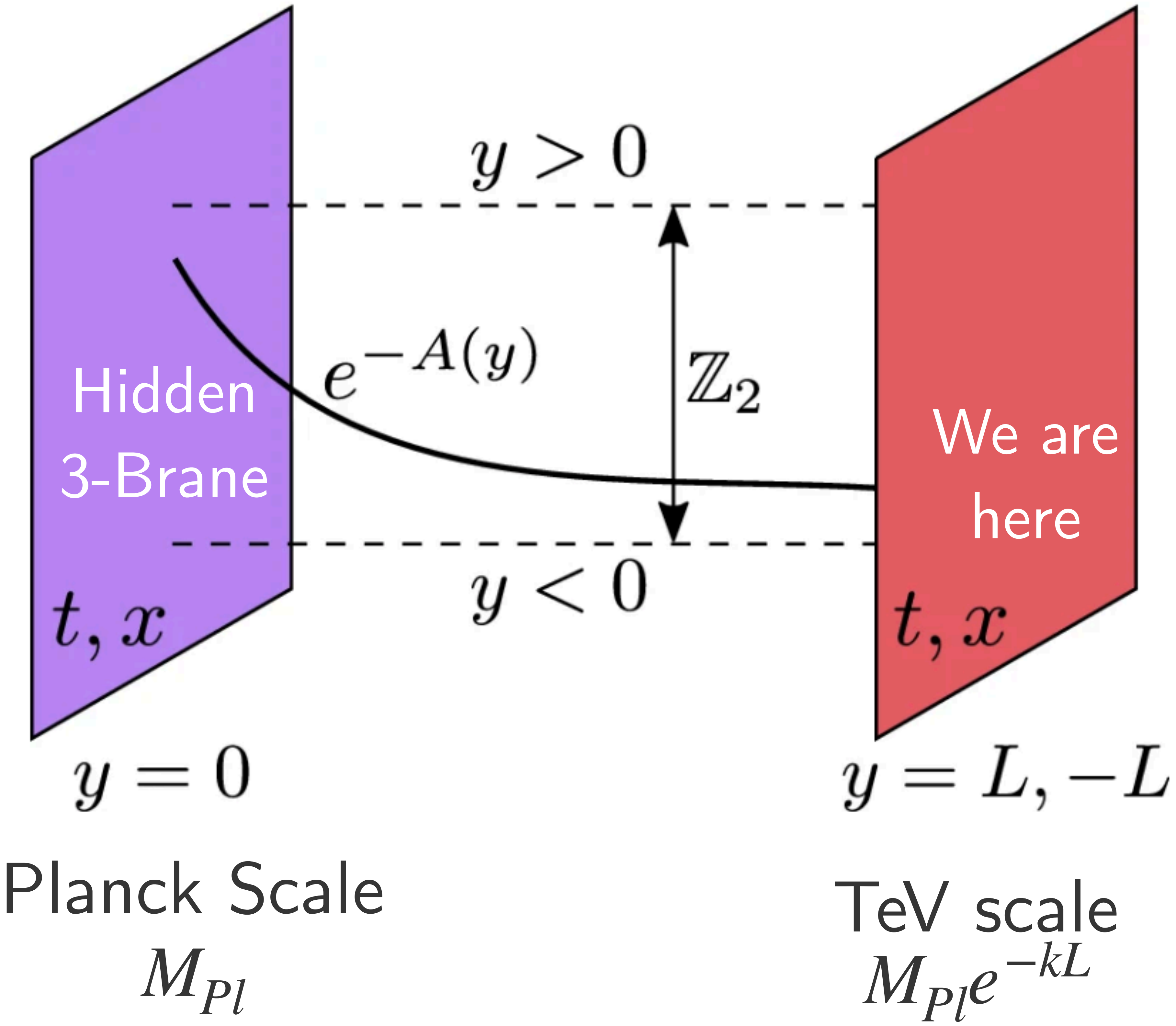
- Gravity not included
- Lack of baryogengesis and matter-antimatter asymmetry motivation
- SM describes only $\sim 5\,\%$ of the matter/energy in the Universe. What about the remaining?
- Why exactly 3 fermion families?
- Neutrino oscillations demonstrate they have mass. Not included in the SM framework.
- Parameter values (masses, charges..) not predicted by the theory
- The hierarchy problem:
  - Inconsistency between the small observed m_H value and the large quantum corrections it should receive due to contributions from virtual particles at high energy scales $\propto \Lambda^2_{cutoff}$, all canceled
  - Severe fine-tuning issue

Warped Extra Dimensions - 101

- Metrics $ds^2 = g_{MN}dx^m dx^N = e^{-2ky}\eta_{\mu\nu}dx^\mu dx^\nu + dy^2$

  Warp factor — Minkowski metric

k = AdS5 curvature, $\sim M_{Pl}$
y = 5D direction



- Quantum fluctuations around the classical metric solution

$$\delta g_{MN}(x,\phi) = \begin{pmatrix} h_{\mu\nu}(x,\phi) & h_{\mu,5}(x,\phi) \\ h_{\mu,5}(x,\phi) & h_{55}(x,\phi) \end{pmatrix}$$

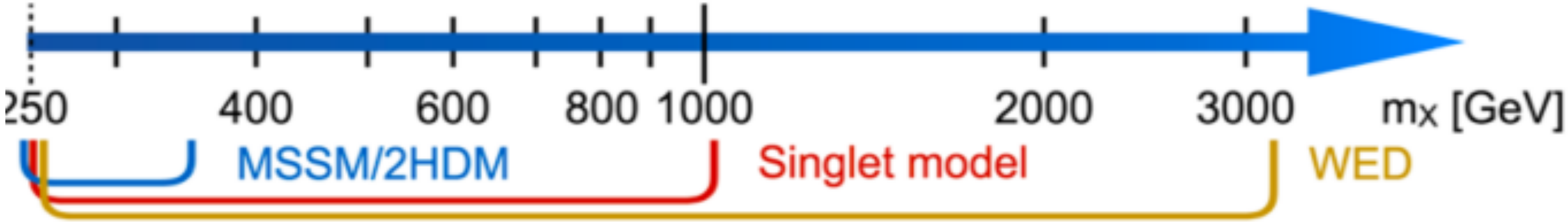
- Axial gauge: $h_{\mu,5} = 0$

$$ds^2 = e^{-2(\epsilon_r h_{55} + \sigma(\phi, r_c))}\left(\eta_{\mu\nu} + \epsilon_g h_{\mu\nu}(x,\phi)\right)dx^\mu dx^\nu + r_c^2\left(2\,\epsilon_r h_{55}(x,\phi) + 1\right)^2 d\phi^2$$

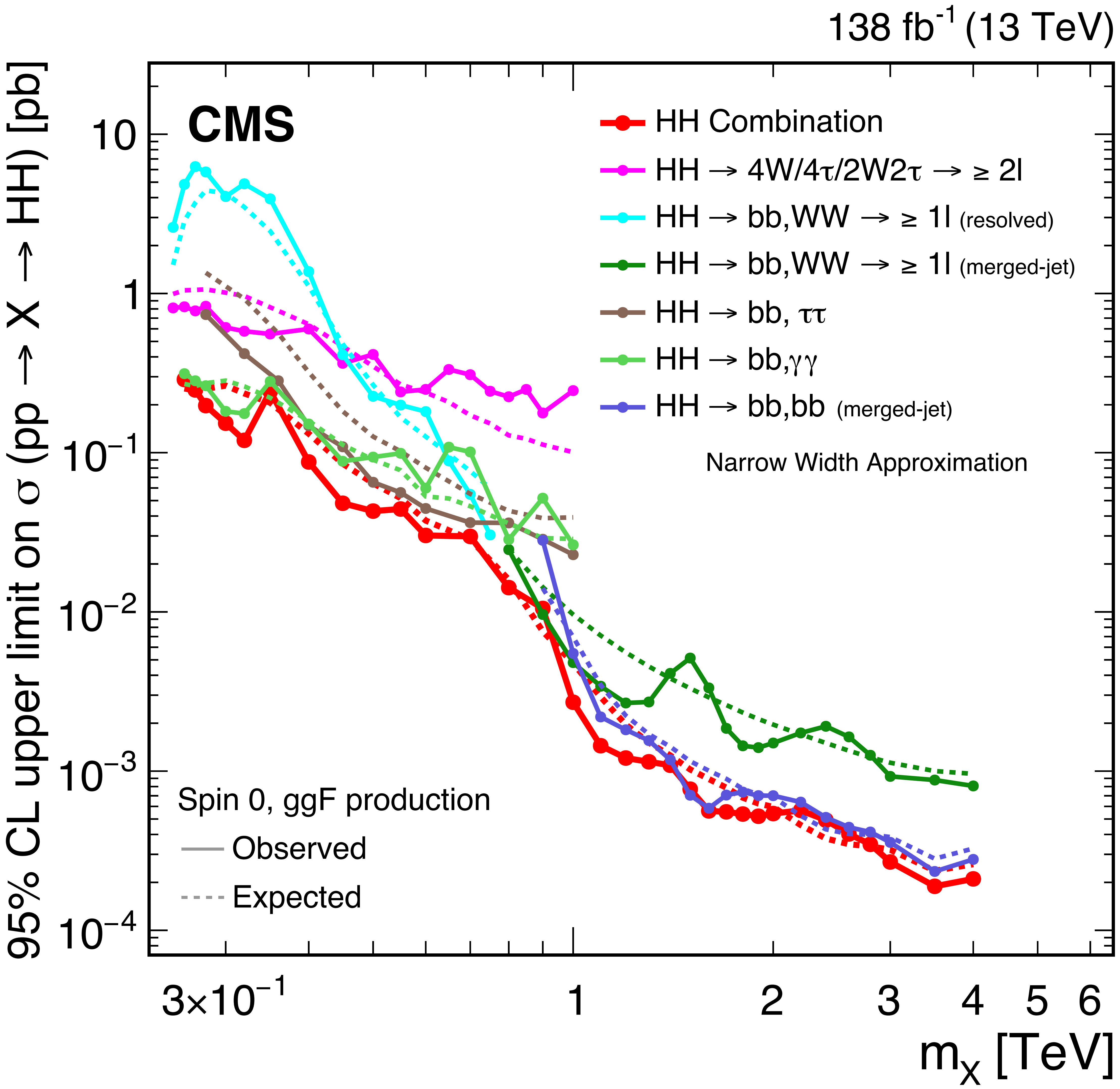
Graviton modes (spin-2) — Radion mode (spin-0)

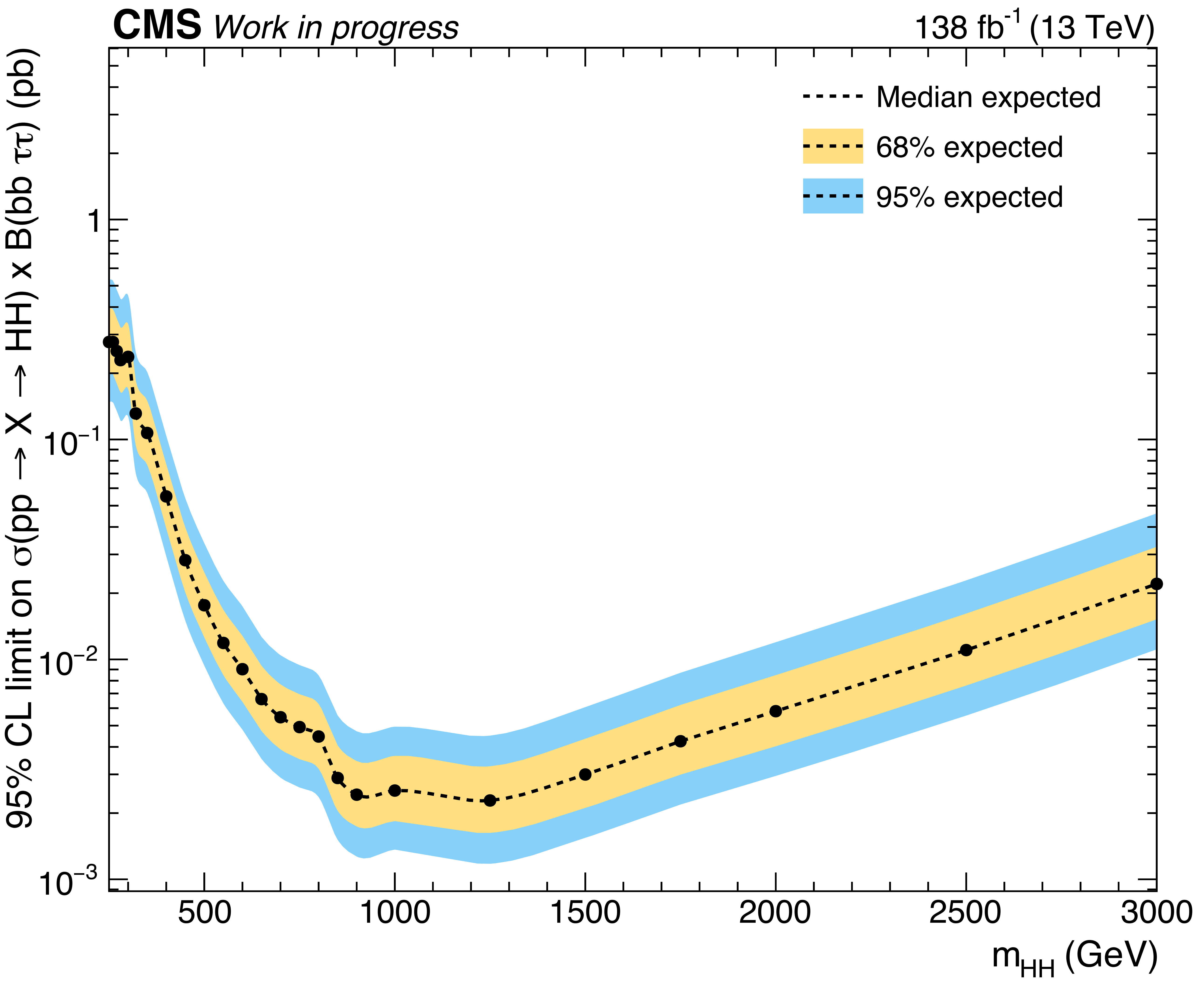
Resonant production - not only WED model

- Allows to study a large number of BSM models, predicting a heavy scalar decaying in HH in a wide range of mass.
- The experimental signature of these resonance is a peak in the invariant mass spectra of the Higgs pair candidates and an enhanced Higgs pair production cross section.
- MSSM/2HDM: Additional Higgs doublet $\rightarrow$ 5 new scalar particles, of which two are CPeven neutrals Higgses. The scalar boson observed at the LHC is interpreted as the light Higgs predicted by the model
- Singlet Model: Additional Higgs singlet with and extra scalar Higgs having a not negligible width at high mass

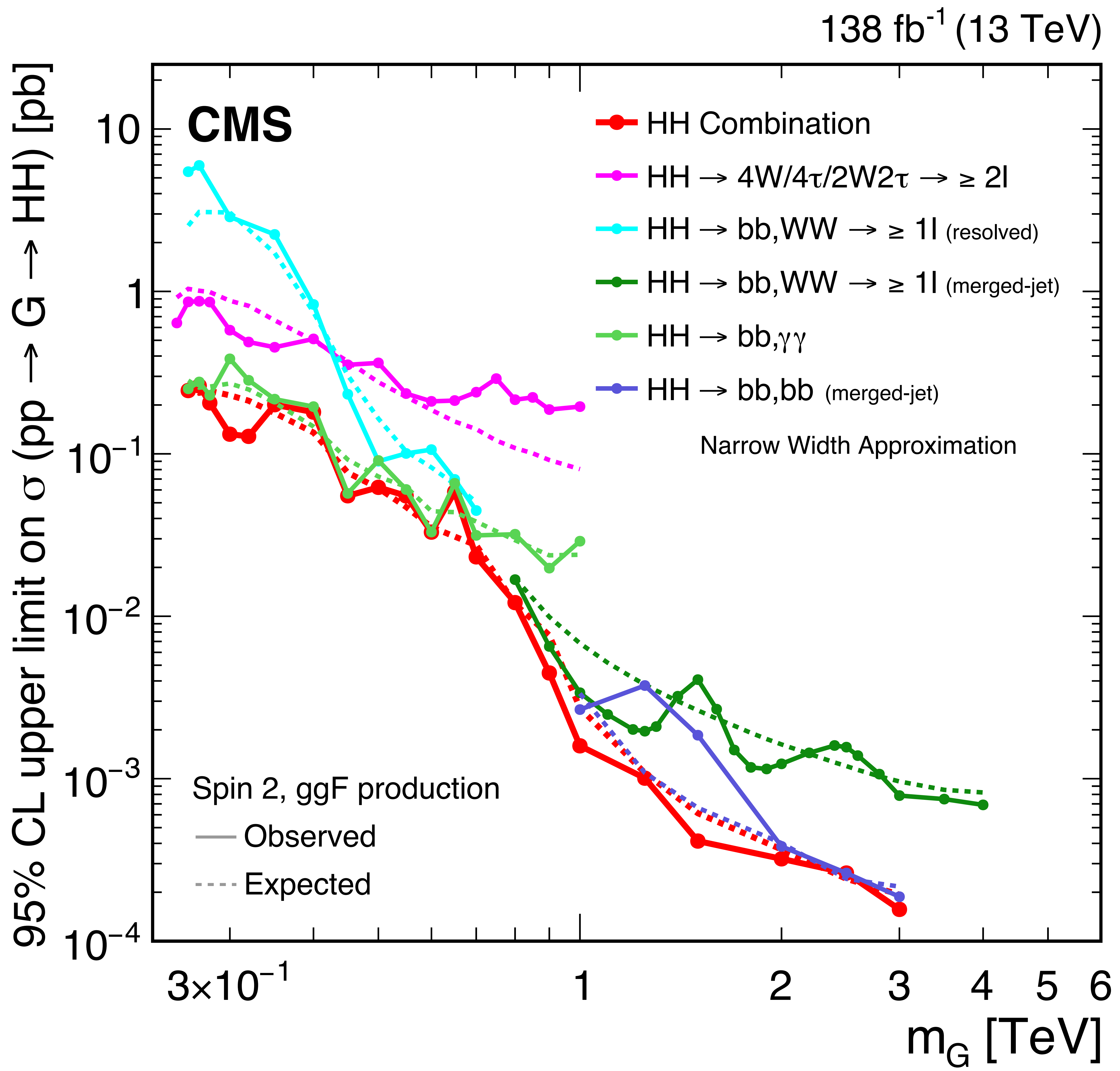


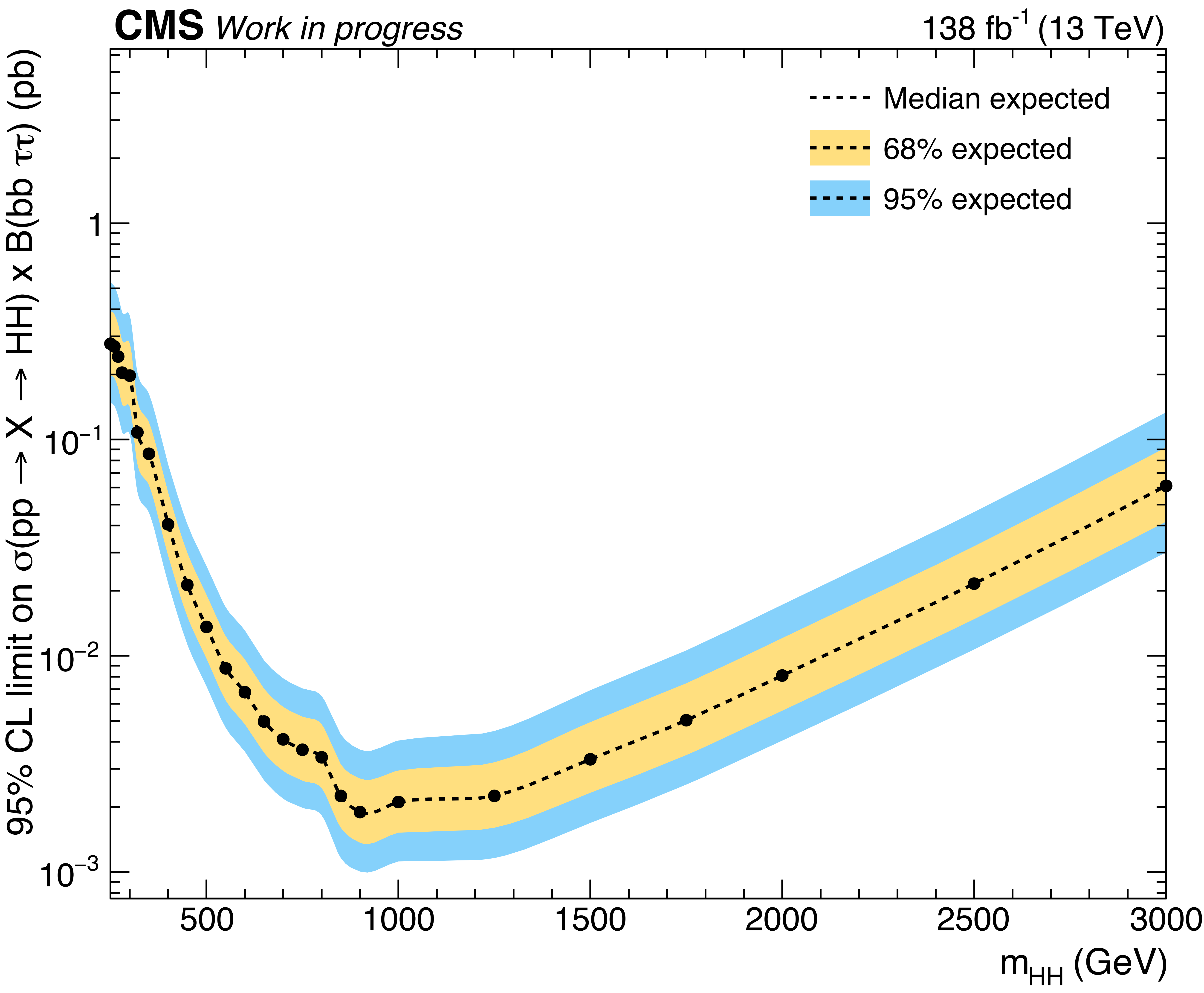
Comparison with global results - spin 0





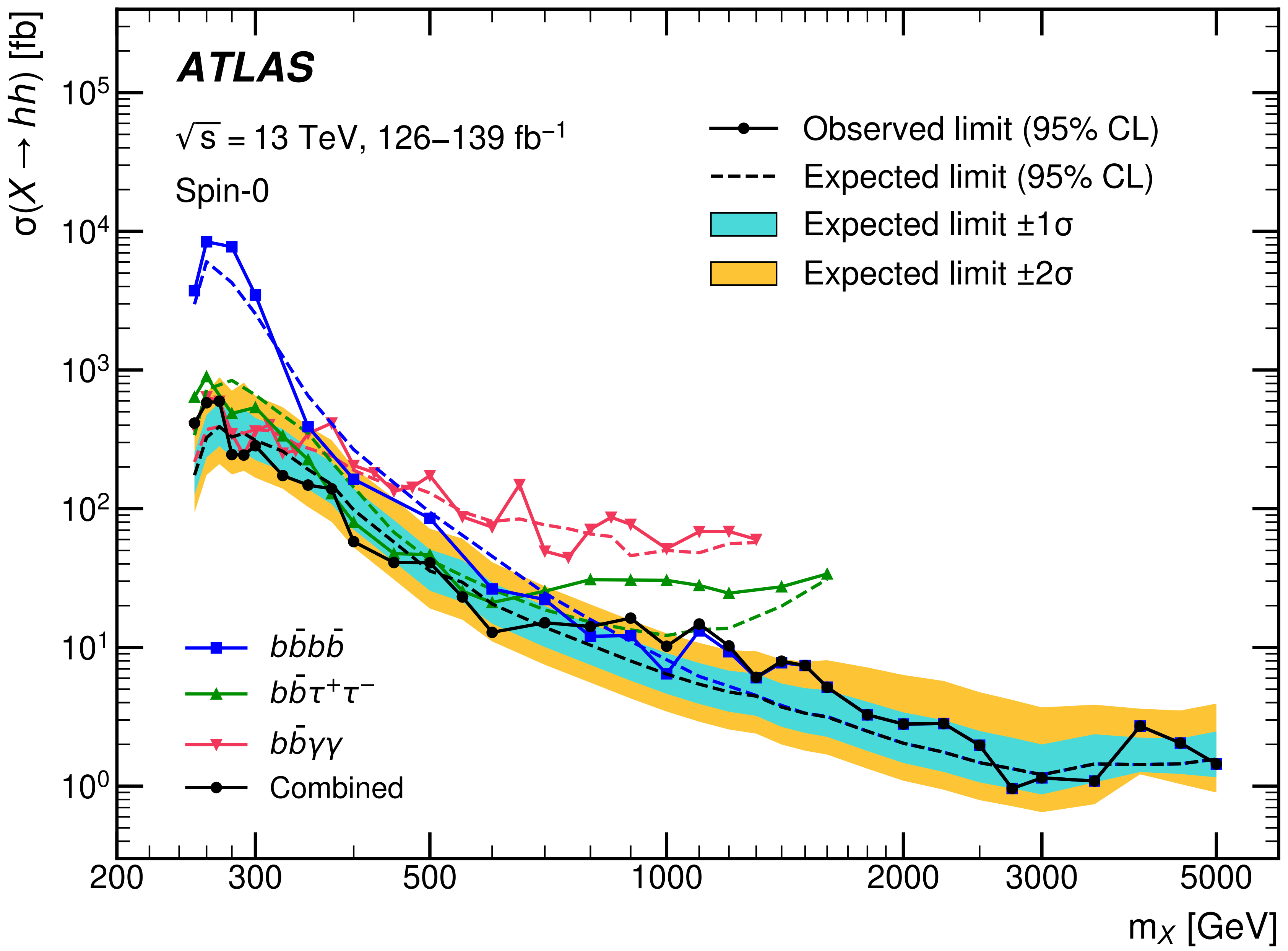
Comparison with global results - spin 2

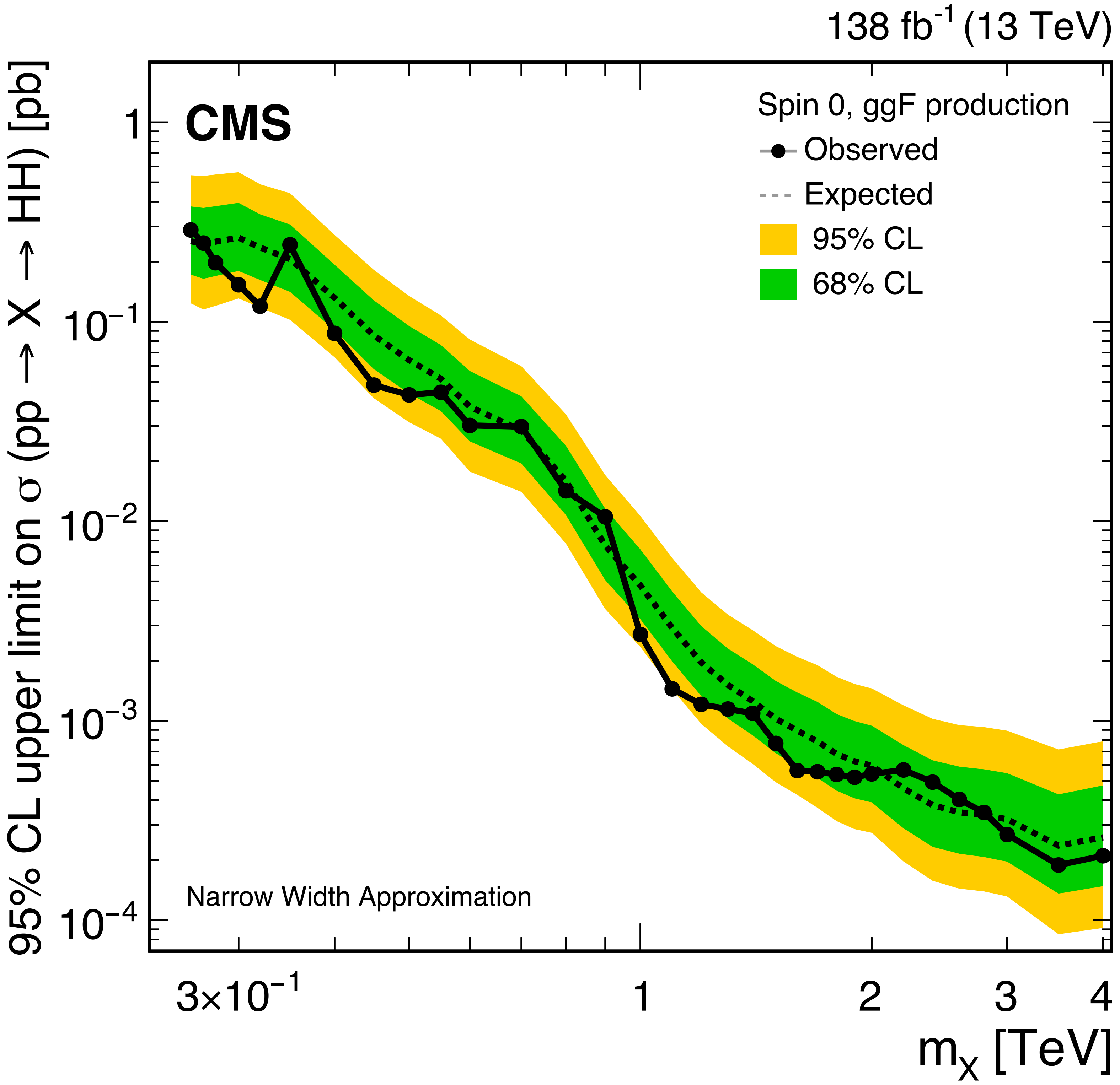


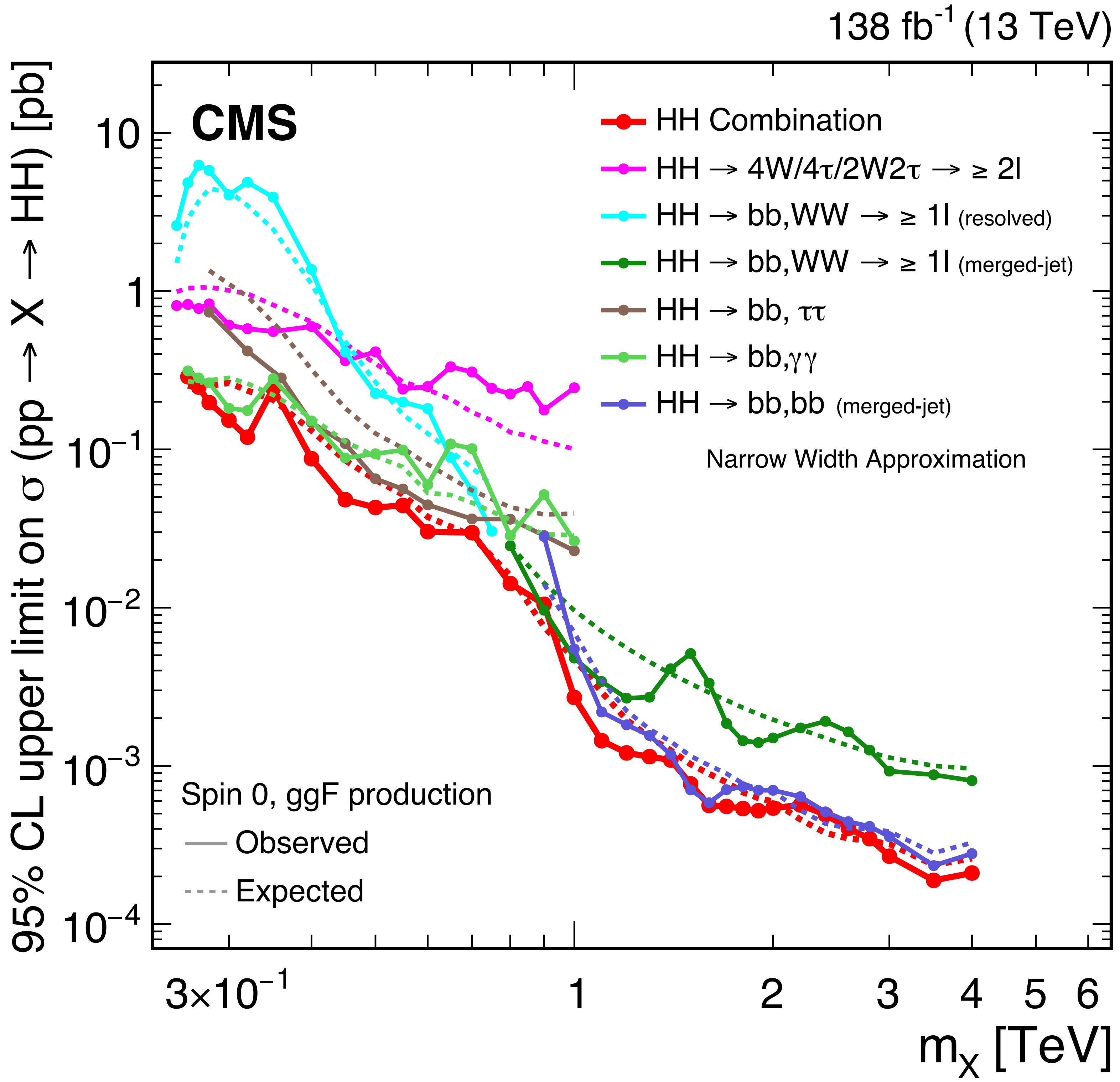


Double Higgs Production

Resonant, spin-0: state of art

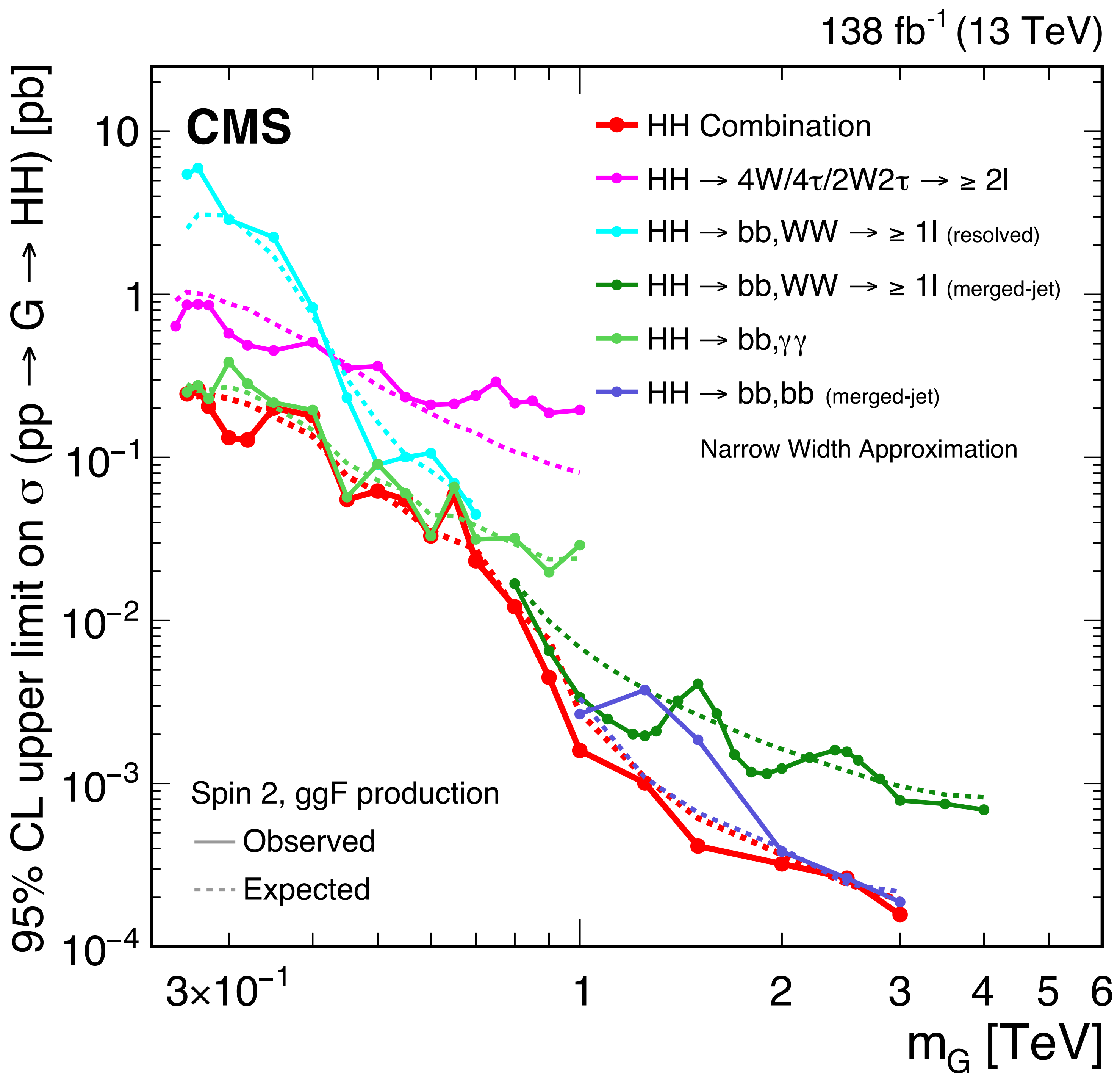


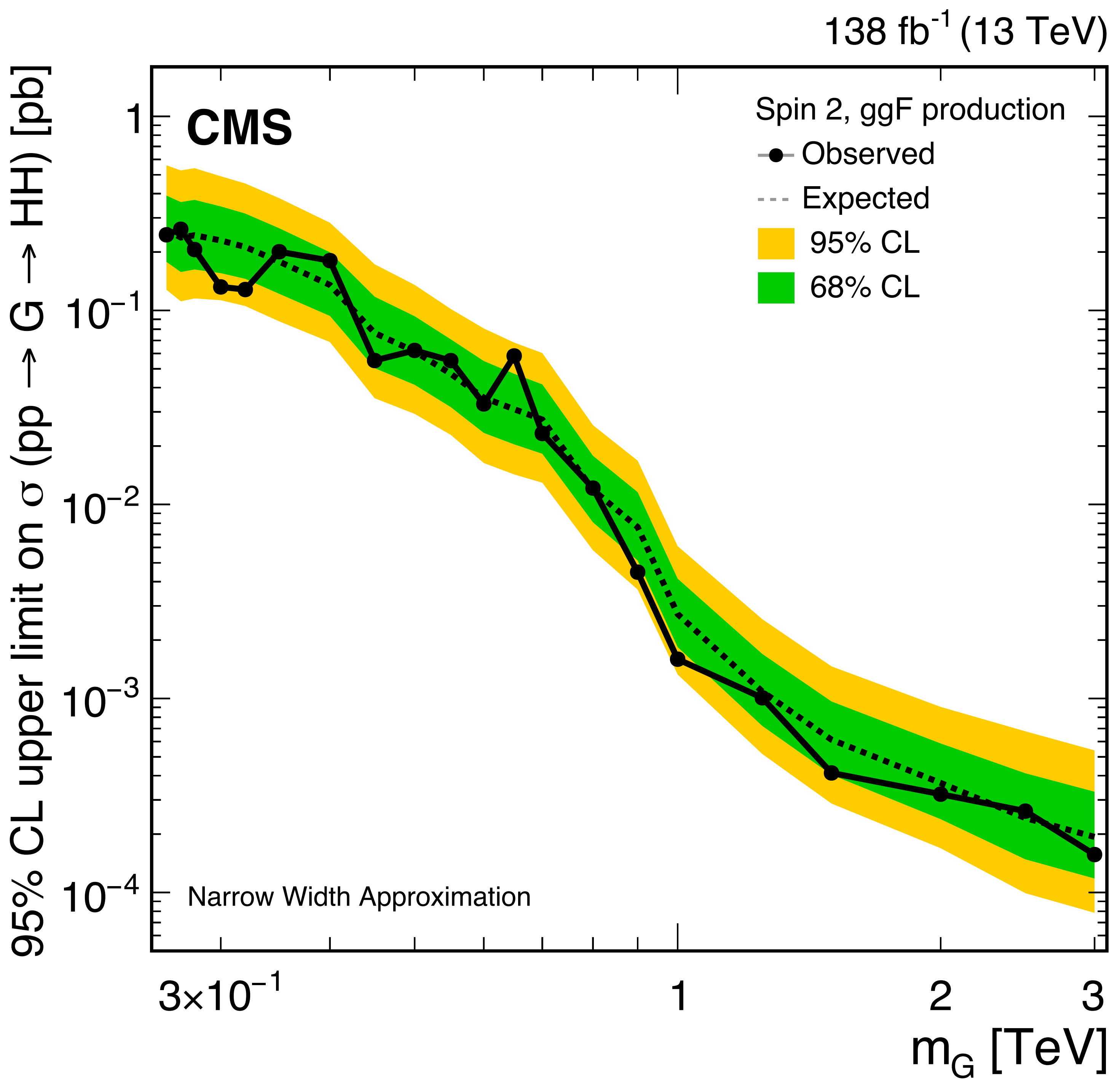




Double Higgs Production

Resonant, spin-2: state of art





MC samples and background modelling

Signals: MC simulation of resonant HH production via Gluon Gluon fusion at LO

- spin-0 ‘radion-like’, 25 masses points between 250 GeV and 3 TeV
- spin-2 ‘graviton-like’, 25 masses points between 250 GeV and 3 TeV

Backgrounds:

- $t\bar{t}$ (NLO): MC simulation, dedicated phase space region to verify data/MC agreement
- DY+jets (NLO): 3 different MC simulation sets: inclusive, p_T-binned and nJets- binned, and dedicated region to check data/MC agreement
- W+Jets (LO): MC simulation, combination of 2 different sets: inclusive, binned in nJets
- Single Higgs, VV, VVV, TTX, Single Top - (NLO/LO) - MC simulations
- QCD: huge cross-section $\sim \mu b/mb$. Complex to simulate, thus estimated with data-driven method (see slide 17)

Luminosity

Before: Initial NanoAOD (centrally produced)

After: general loose preselection

$$N_{TOT}^{proc} = L \cdot s^{proc} \cdot \frac{\sum(w_{pu} \cdot w_{gen})|_{after}}{\sum(w_{pu} \cdot w_{gen})|_{before}}$$

(Stitched) Cross section

$$s_i = \sigma_i^{NLO/LO} \cdot \frac{1}{K} \cdot \frac{\sigma_{incl}^{NNLO}}{\sigma_{incl}^{NLO/LO}}$$

K depends on the number of MC sets

Trigger paths - 2016

Trigger	Channel	HLT path name
diTau	$\tau_h\tau_h$	HLT_DoubleMediumCombinedIsoPFTau35_Trk1_eta2p1_Reg
		HLT_DoubleMediumIsoPFTau35_Trk1_eta2p1_Reg **Era H**
SingleTau	$\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$	HLT_VLooseIsoPFTau120_Trk50_eta2p1 OR HLT_VLooseIsoPFTau140_Trk50_eta2p1
MET	$\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
SingleMu	$\tau_\mu\tau_h$, $\tau_\mu\tau_\mu$	HLT_IsoMu24 OR HLT_IsoTkMu24
MuTau	$\tau_\mu\tau_h$	HLT_IsoMu19_eta2p1_LooseIsoPFTau20 OR HLT_IsoMu19_eta2p1_LooseIsoPFTau30 OR HLT_IsoMu19_eta2p1_LooseIsoPFTau20_SingleL1 WIP
SingleEle	$\tau_e\tau_h$, $\tau_e\tau_e$	HLT_Ele25_eta2p1_WPTight_Gsf
eTau	$\tau_e\tau_h$	HLT_Ele24_eta2p1_WPLoose_Gsf_LooseIsoPFTau20 OR HLT_Ele24_eta2p1_WPLoose_Gsf_LooseIsoPFTau20 OR HLT_Ele24_eta2p1_WPLoose_Gsf_LooseIsoPFTau20_SingleL1 WIP

Trigger paths

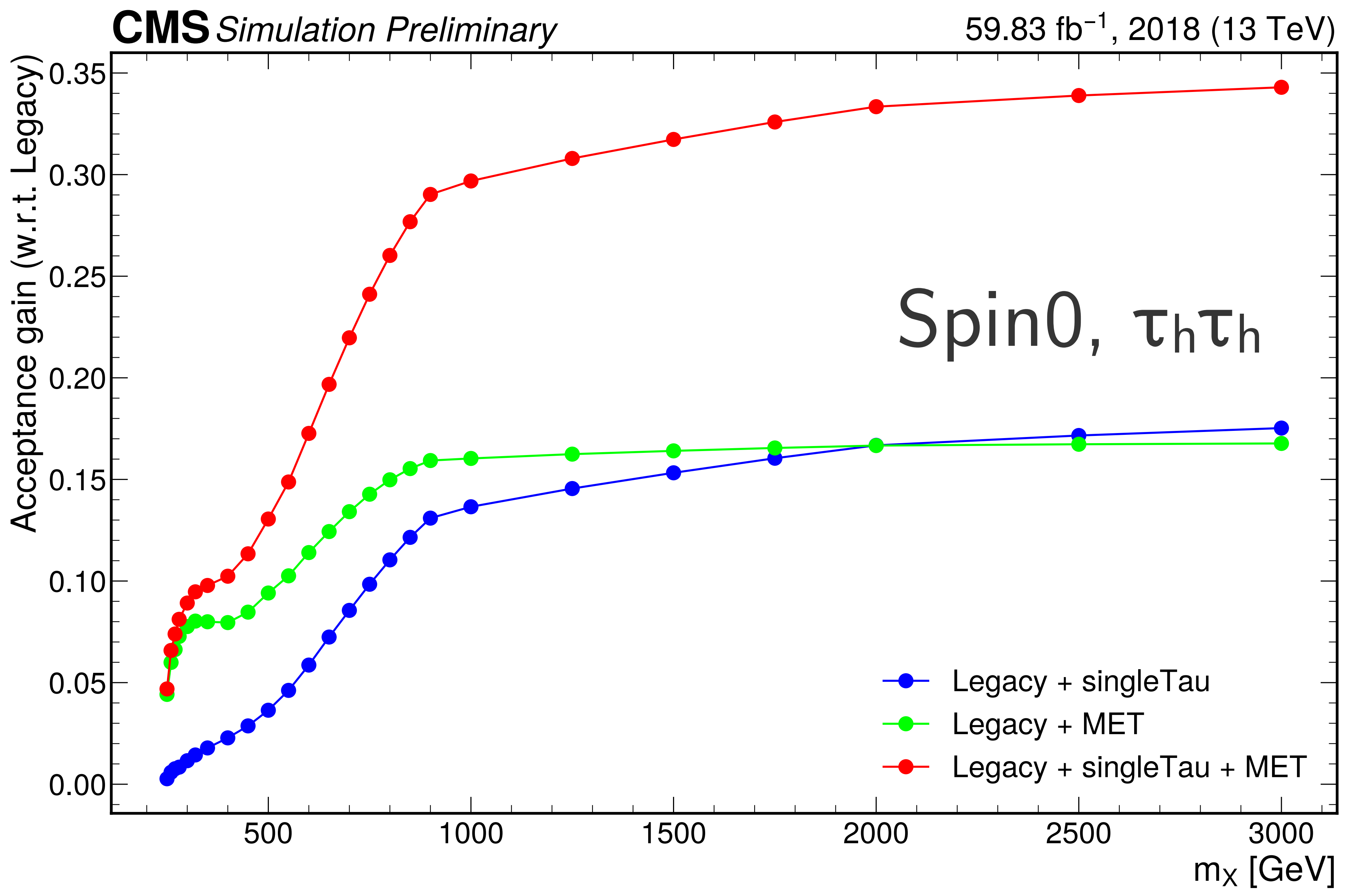
Trigger	Channel	HLT path name
diTau	$\tau_h\tau_h$	HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg OR HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg OR HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg
SingleTau	$\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$	HLT_MediumChargedIsoPFTau180HighPtRelaxedIso_Trk50_eta2p1
MET	$\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
SingleMu	$\tau_\mu\tau_h$, $\tau_\mu\tau_\mu$	HLT_IsoMu24
MuTau	$\tau_\mu\tau_h$	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau27_eta2p1_CrossL1
SingleEle	$\tau_e\tau_h$, $\tau_e\tau_e$	HLT_Ele32_WPTight_Gsf
eTau	$\tau_e\tau_h$	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTau30_eta2p1_CrossL1

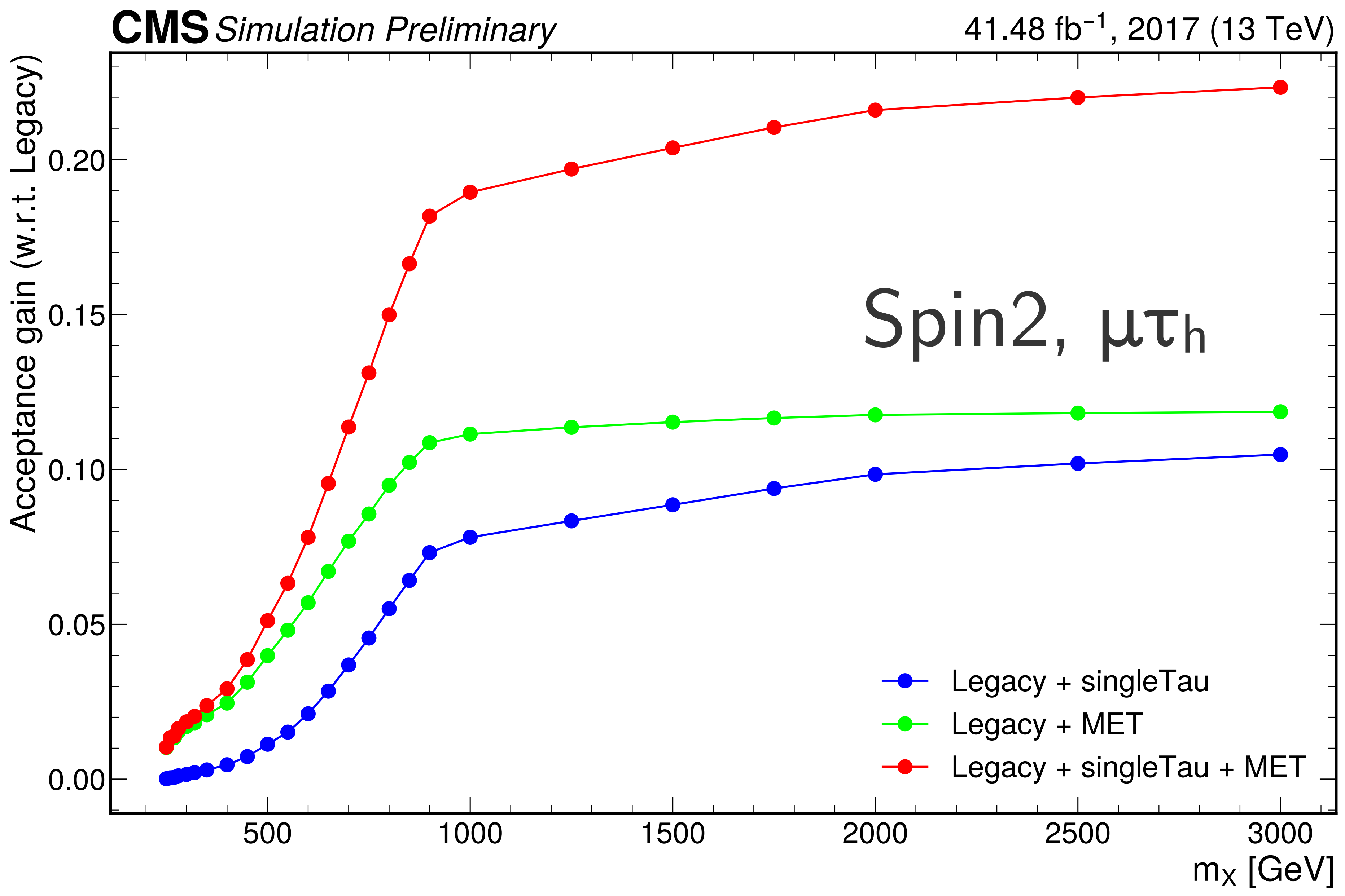
← 2017

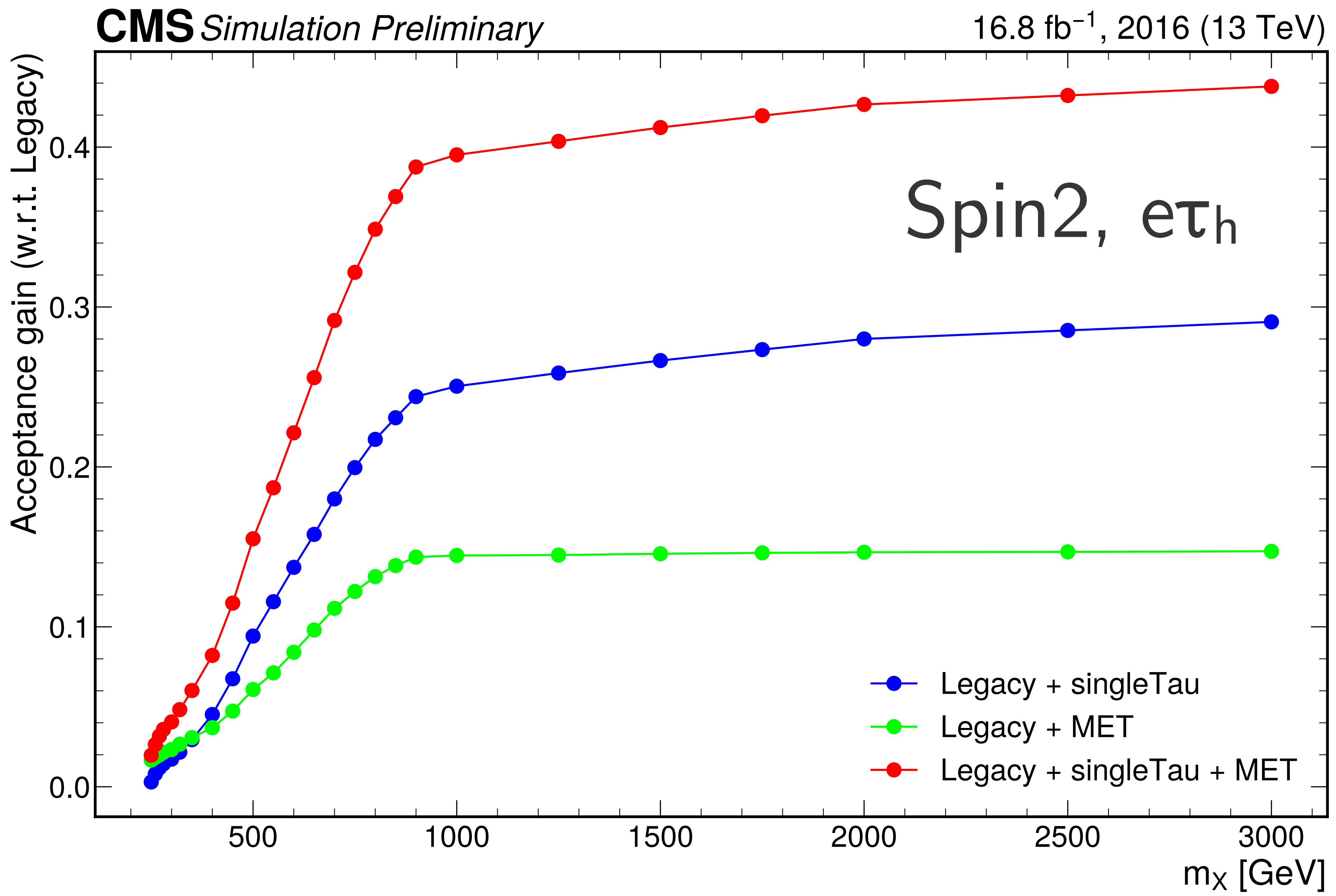
2018 →

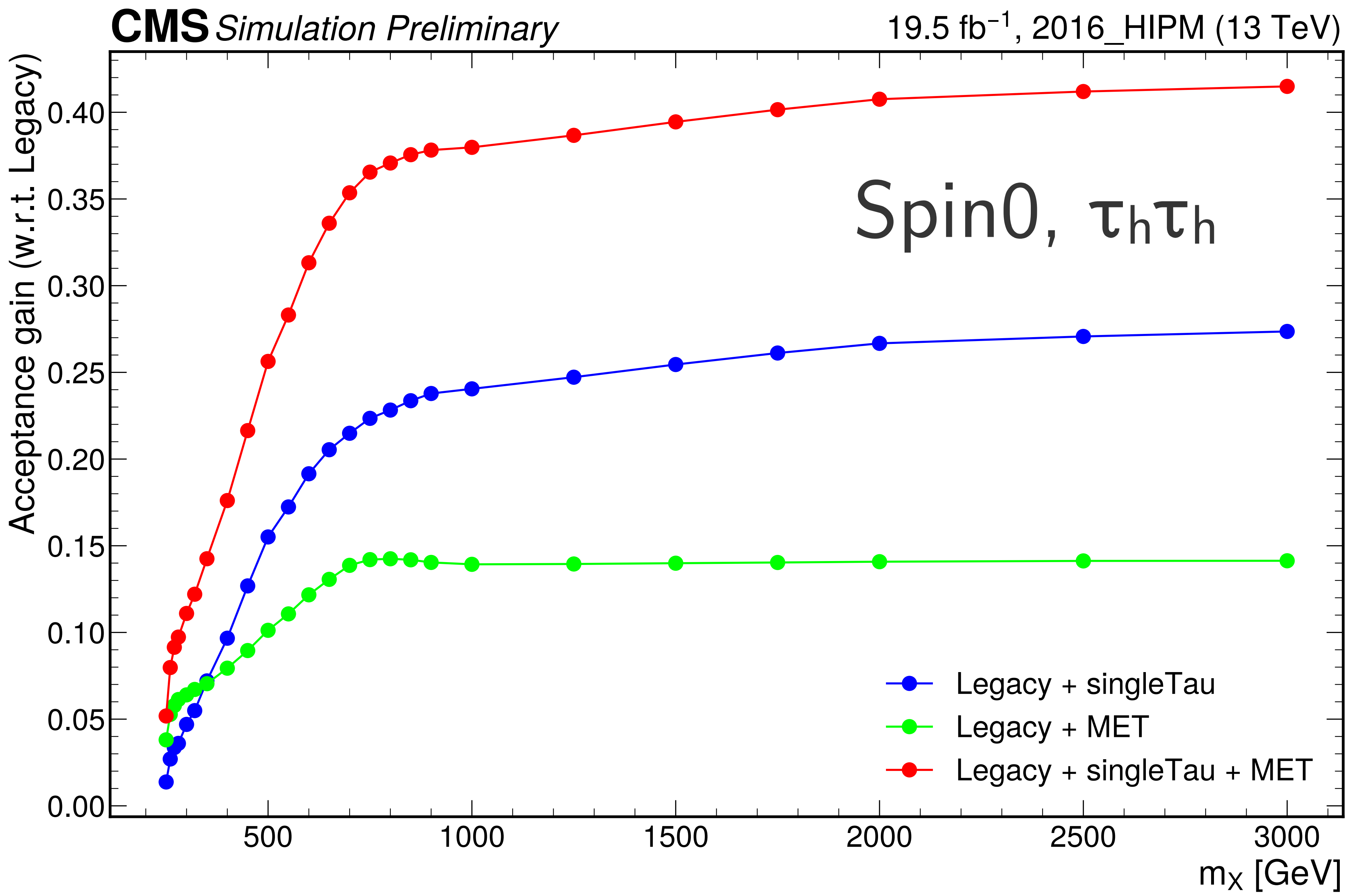
Trigger	Channel	HLT path name
diTau	$\tau_h\tau_h$	HLT_DoubleMediumChargedIsoPFTauHPS35_Trk1_eta2p1_Reg (MC, Run ≥ 317509) HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg OR HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg OR HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg
SingleTau	$\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$	HLT_MediumChargedIsoPFTau180HighPtRelaxedIso_Trk50_eta2p1
MET	$\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
SingleMu	$\tau_\mu\tau_h$, $\tau_\mu\tau_\mu$, $\tau_e\tau_\mu$	HLT_IsoMu24
MuTau	$\tau_\mu\tau_h$	HLT_IsoMu20_eta2p1_LooseChargedIsoPFTau(HPS)27_eta2p1_CrossL1
SingleEle	$\tau_e\tau_h$, $\tau_e\tau_e$, $\tau_e\tau_\mu$	HLT_Ele32_WPTight_Gsf
eTau	$\tau_e\tau_h$	HLT_Ele24_eta2p1_WPTight_Gsf_LooseChargedIsoPFTau(HPS)30_eta2p1_CrossL1

Inclusion of MET+singleTau triggers

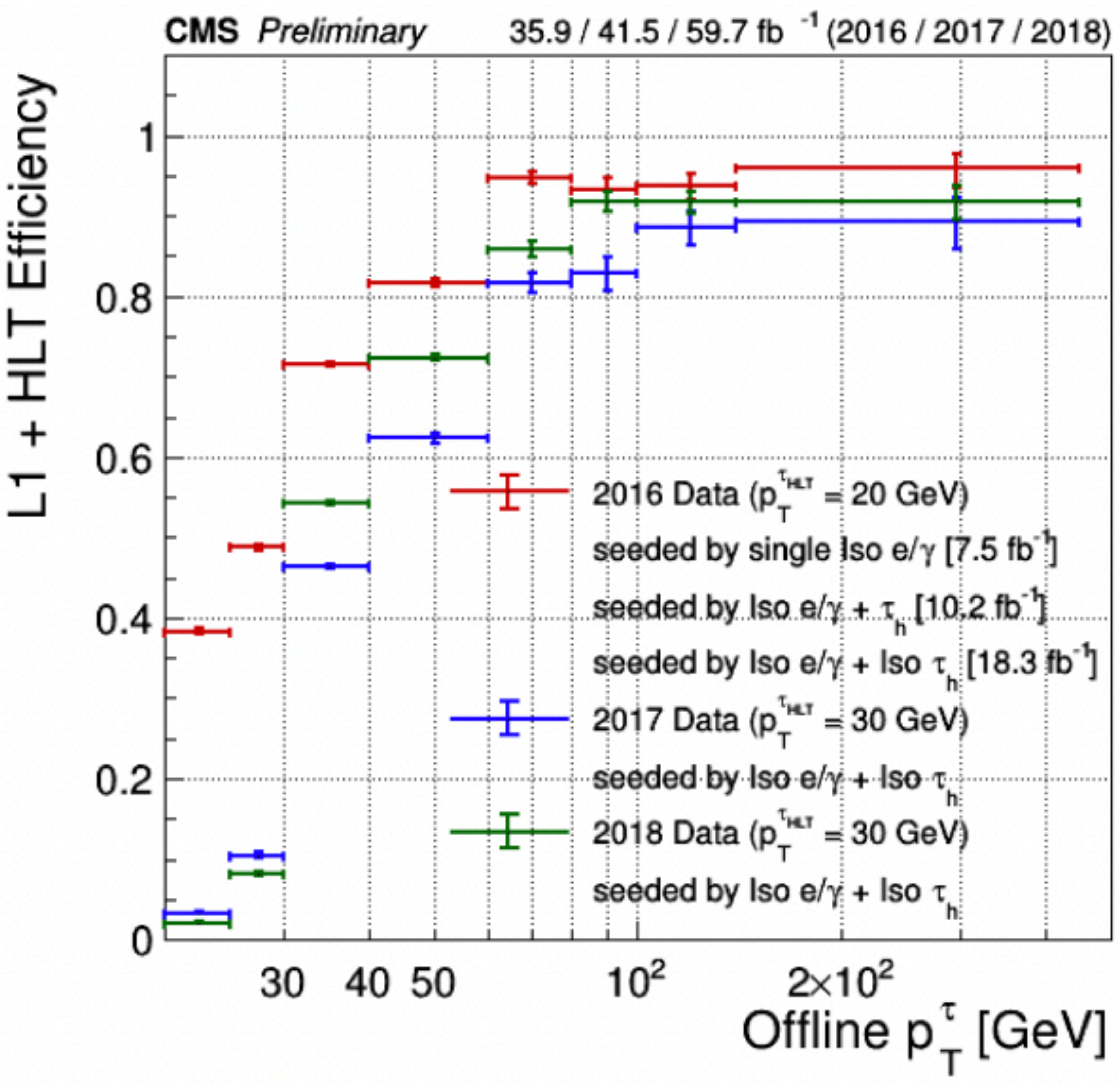




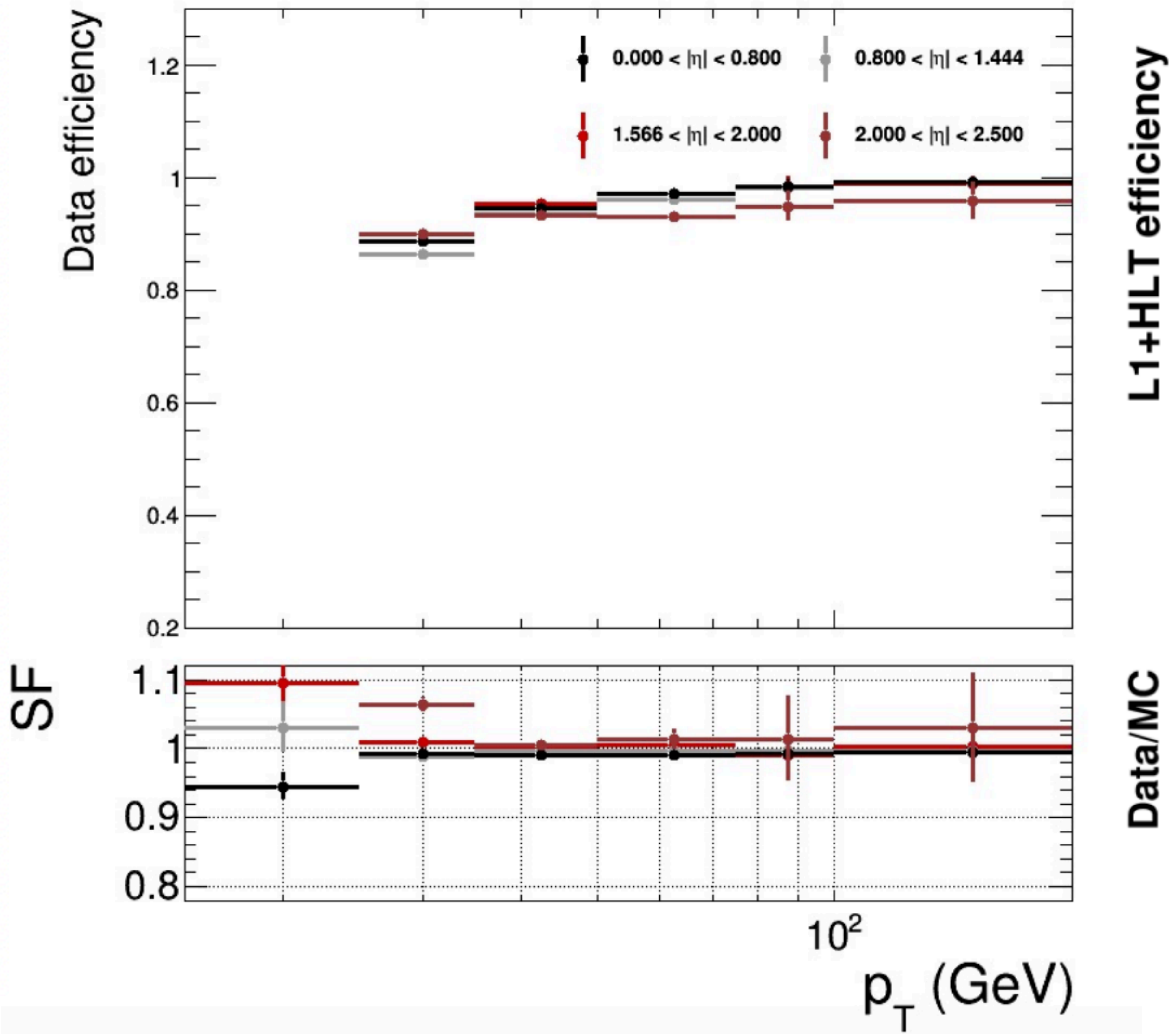




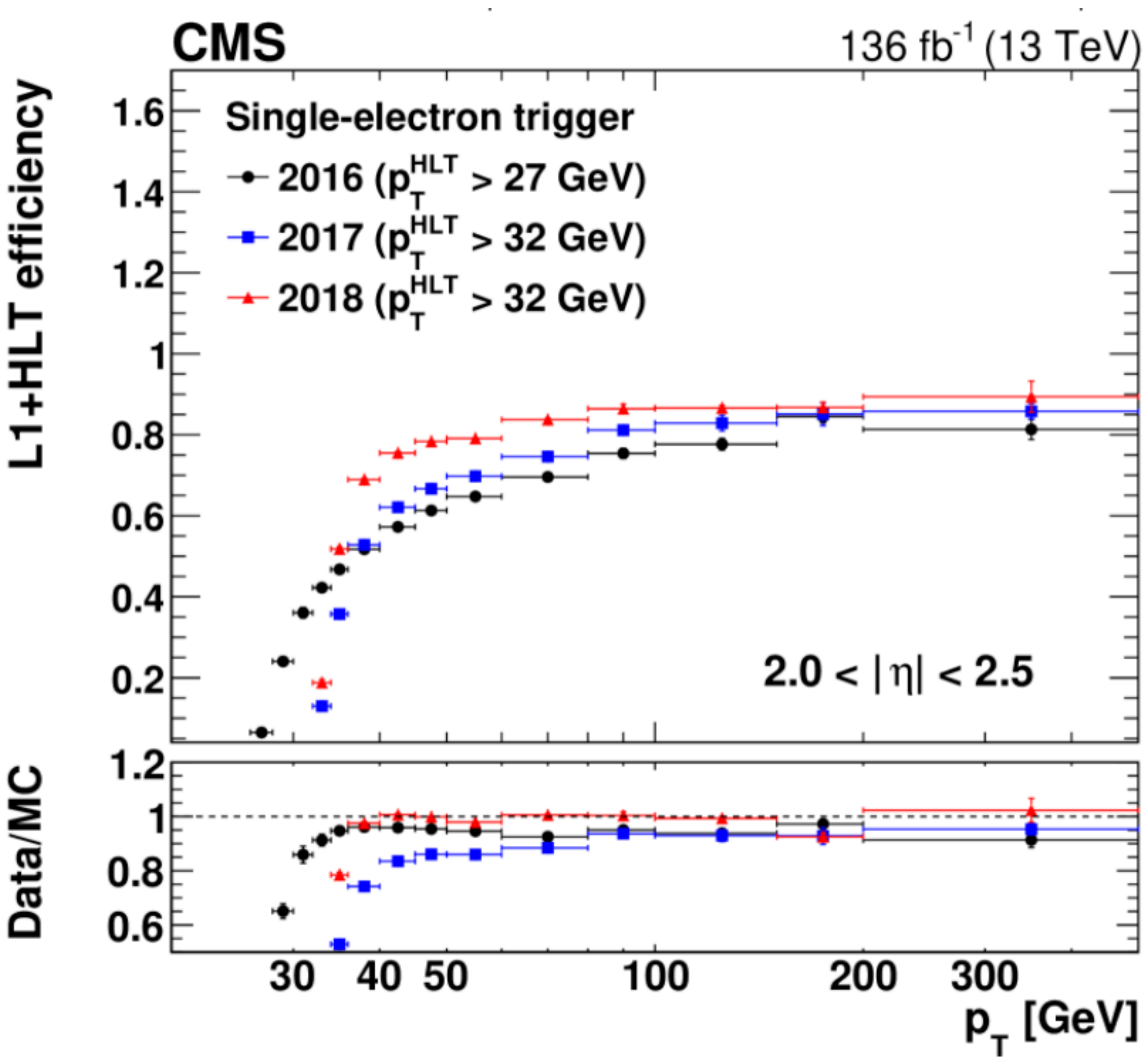
Trigger efficiencies - 1



$e\tau_h$ trigger (τ_h leg, link)

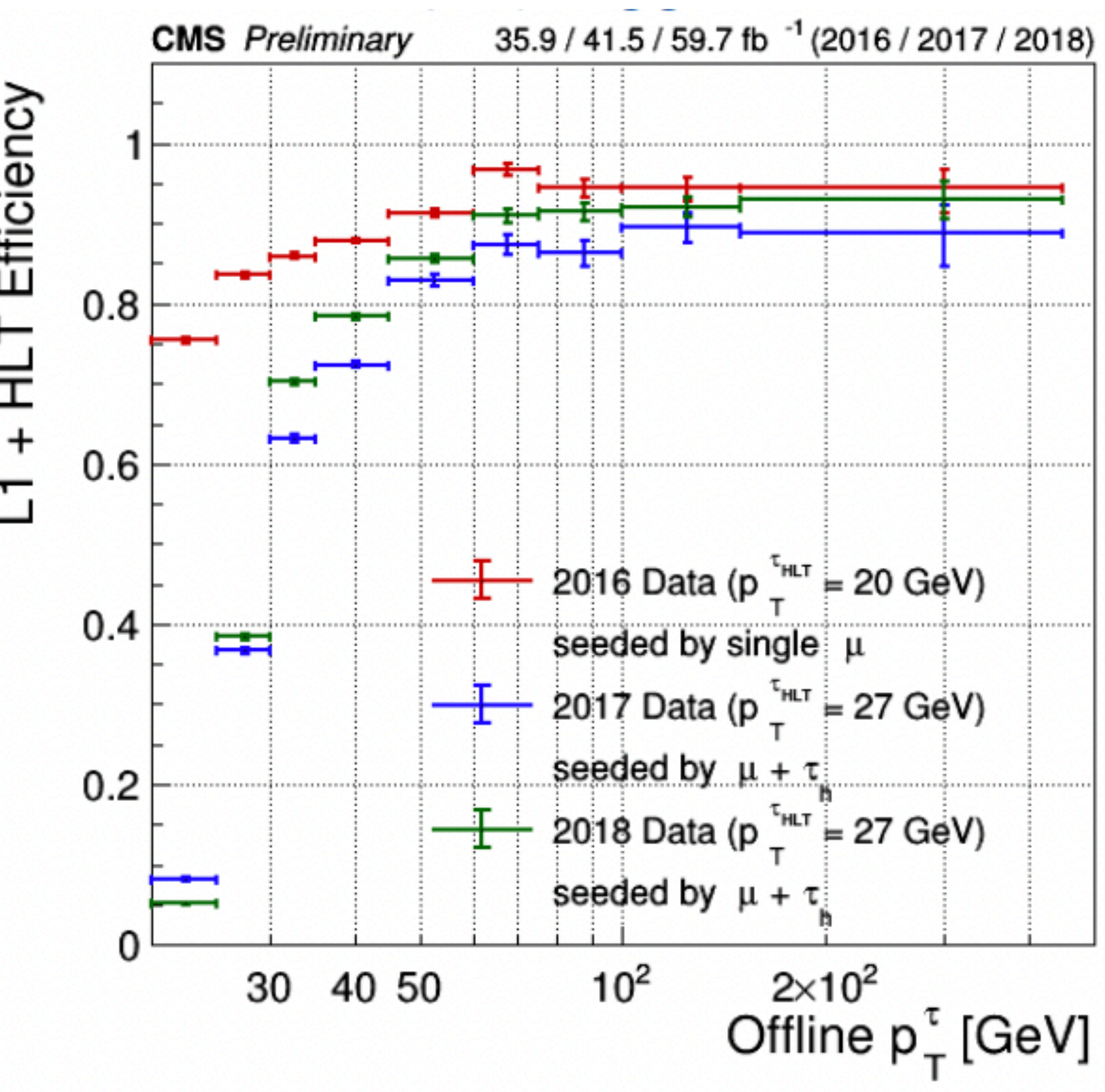


$e\tau_h$ trigger (e leg) 2018 (link - access for CMS members only)

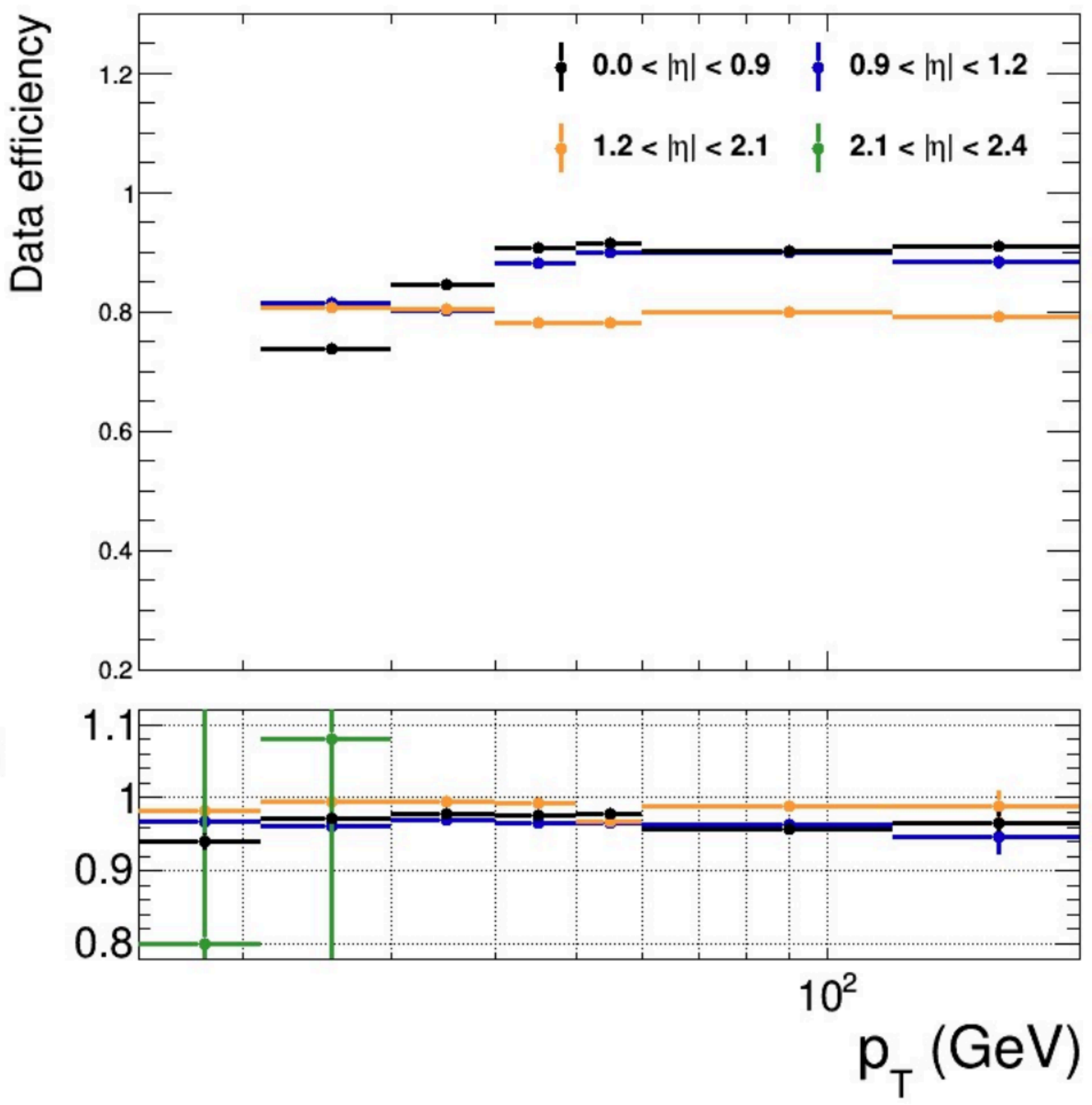


singleEle trigger (link)

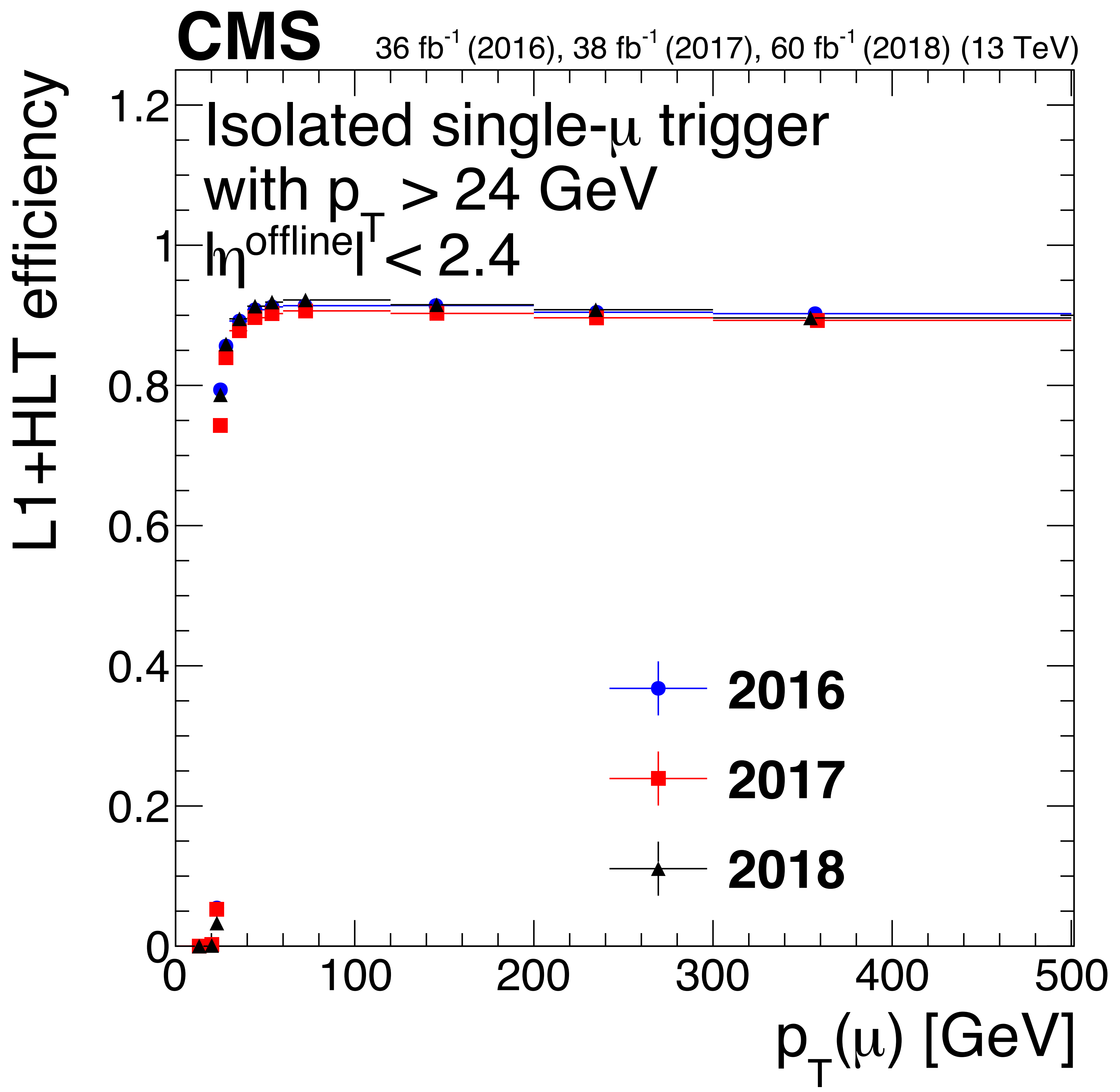
Trigger efficiencies - 2



μτ_h trigger
(τ_h leg, link)

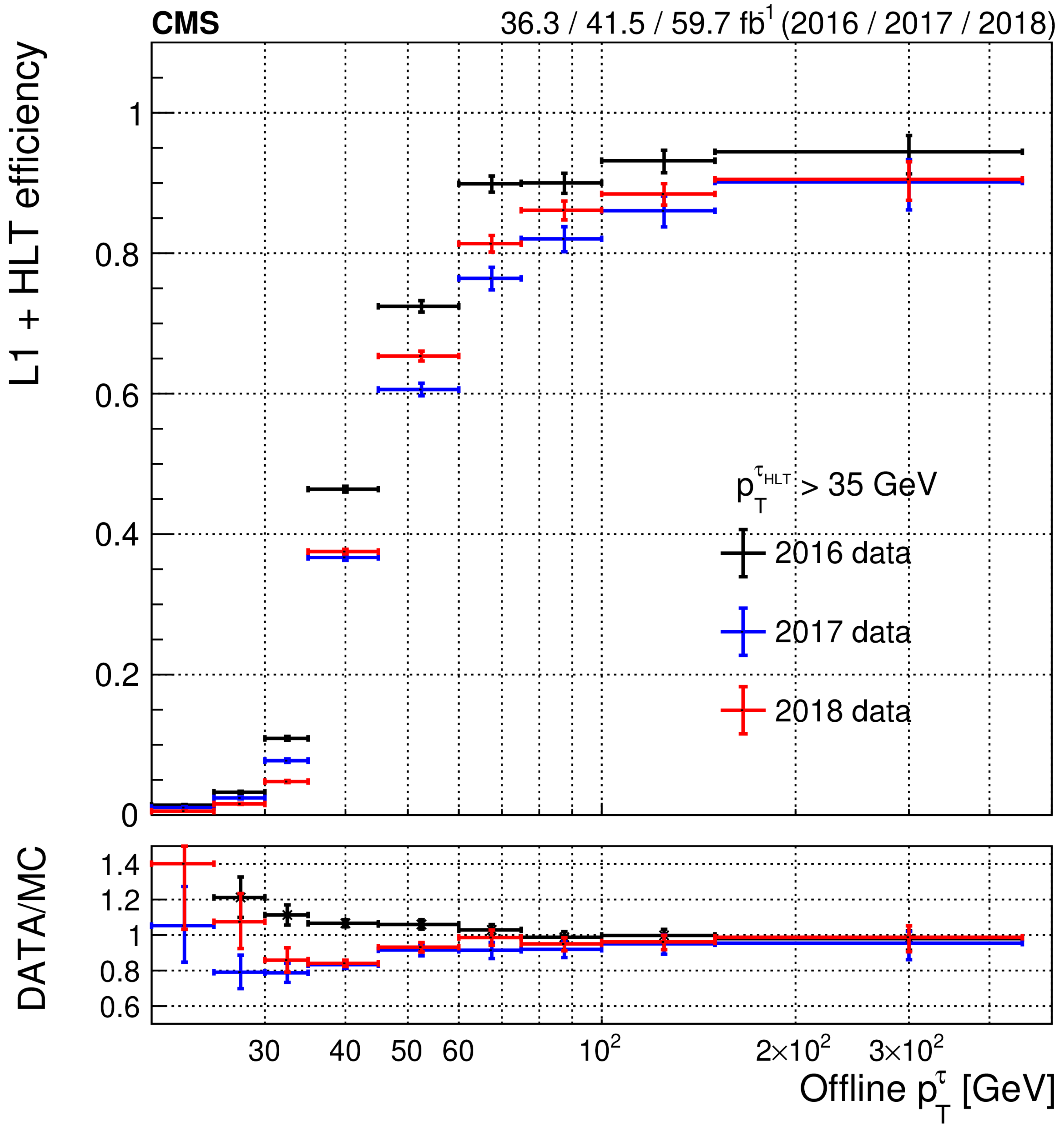


μτ_h trigger (μ leg) 2018 (link - access for CMS members only)

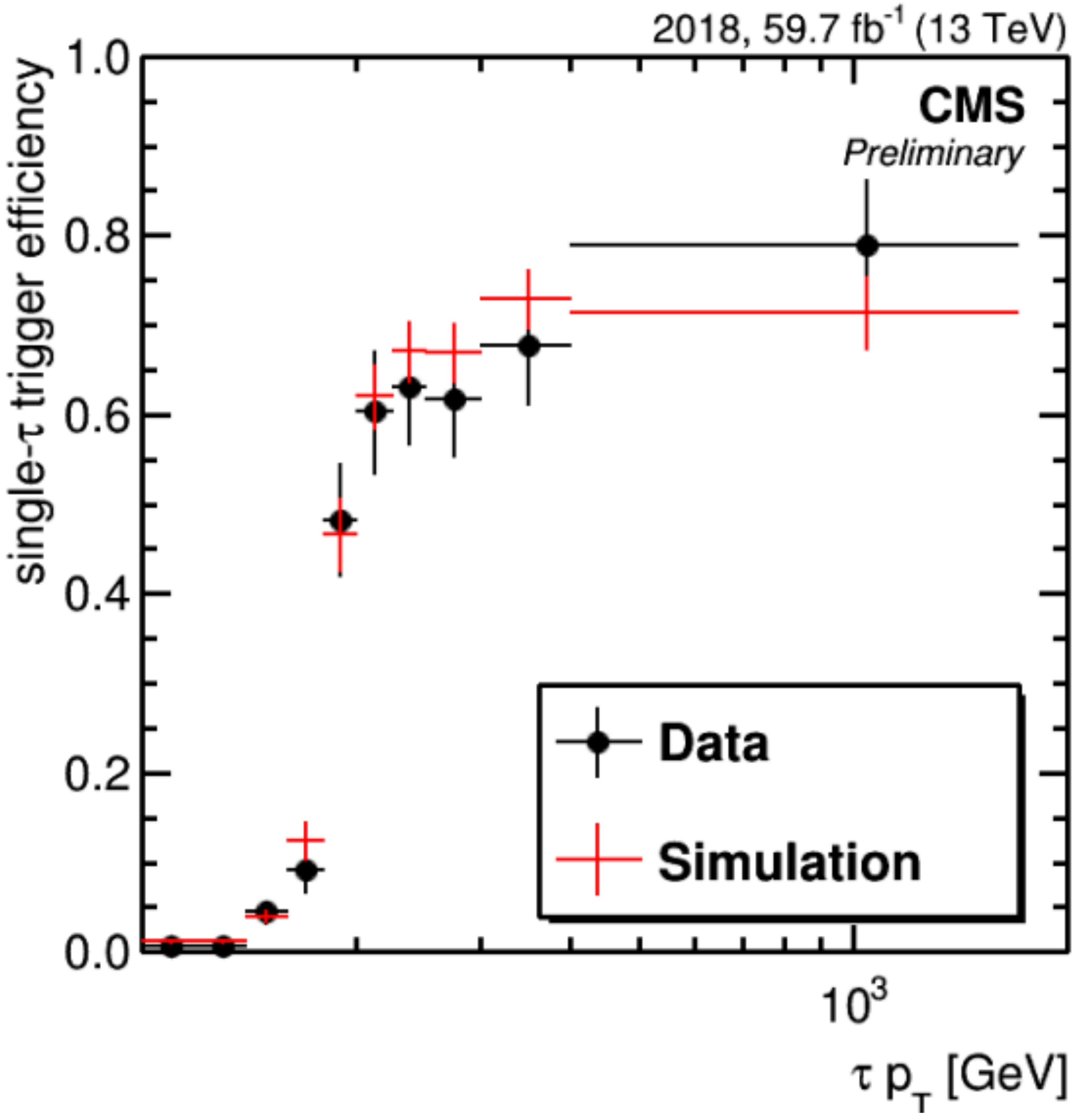


singleMu trigger
(link)

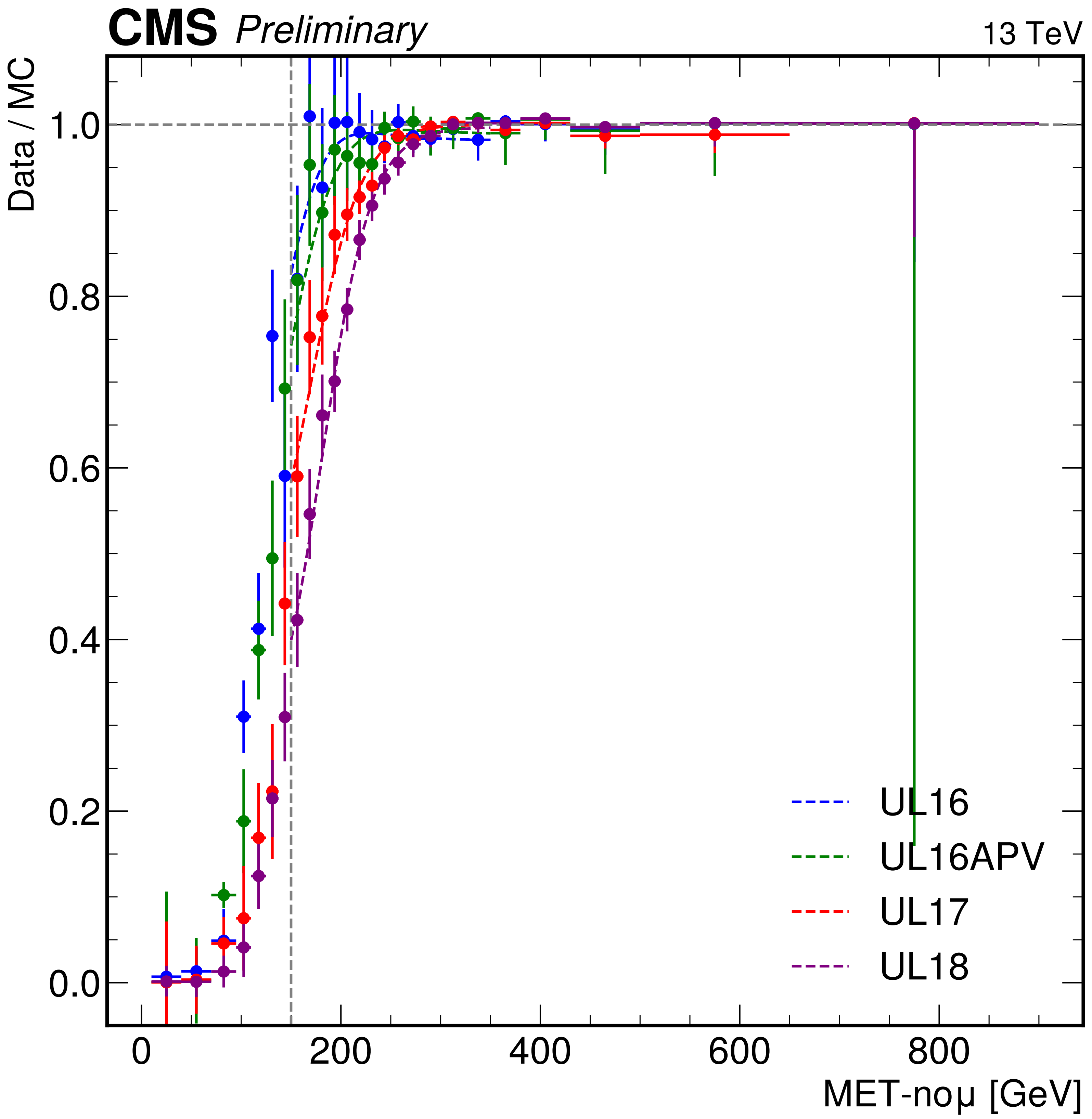
Trigger efficiencies - 3



$\tau_h\tau_h$ trigger (link)

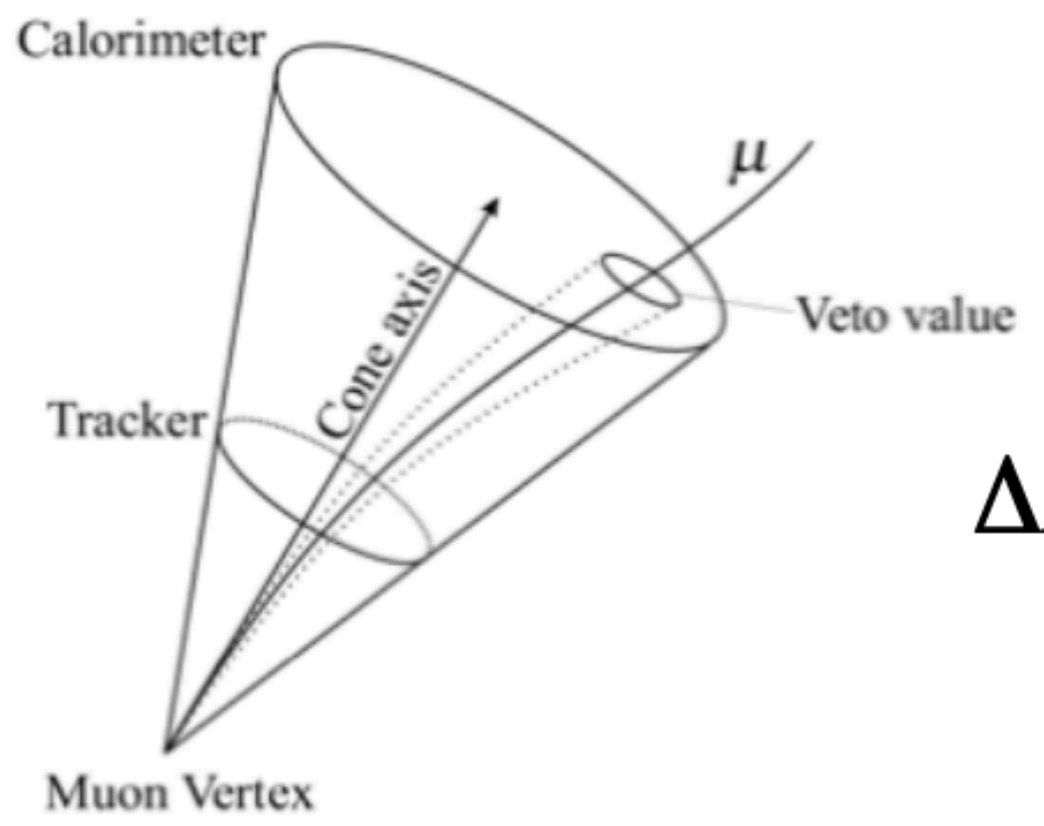


singleTau trigger 2018 (link - access for CMS members only)

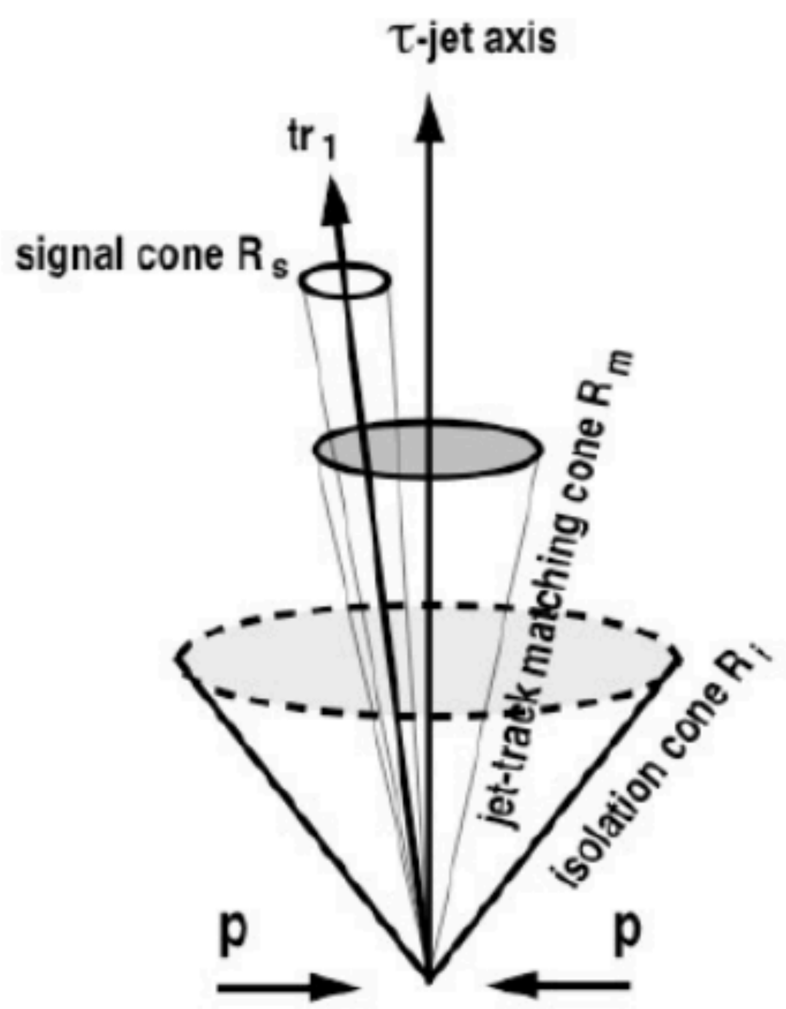


MET trigger (link - Privately produced)

Lepton isolation



$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$



Muon isolation cone in ΔR

Tau isolation cone in ΔR

$$I_{PF} = \sum p_T^{charged} + \max\left[0, \sum p_T^{neutral-had} + \sum p_T^{\gamma} - p_T^{PU}\right],$$
$$I_{PF}^{rel} = I_{PF} / p_T$$

$$I_\tau = \sum_{charged} p_T + \max\left(0, \sum_{\gamma} p_T - \Delta\beta\right),$$
$$\Delta\beta = 0.4576 \sum_{charged} p_T$$

e/μ from τ decays have small activity around the lepton direction with respect to the lepton inside the quark or gluon jet

Electron reco and ID performances

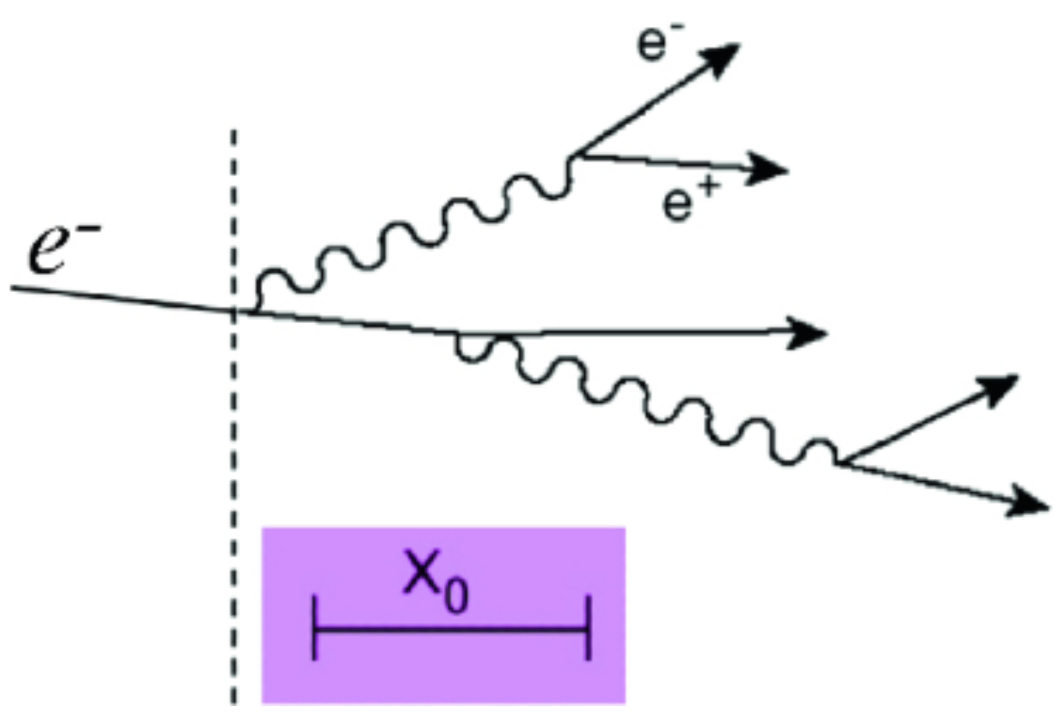
Electron through material → bremsstrahlung→ e+e-pairs → "shower" in detector

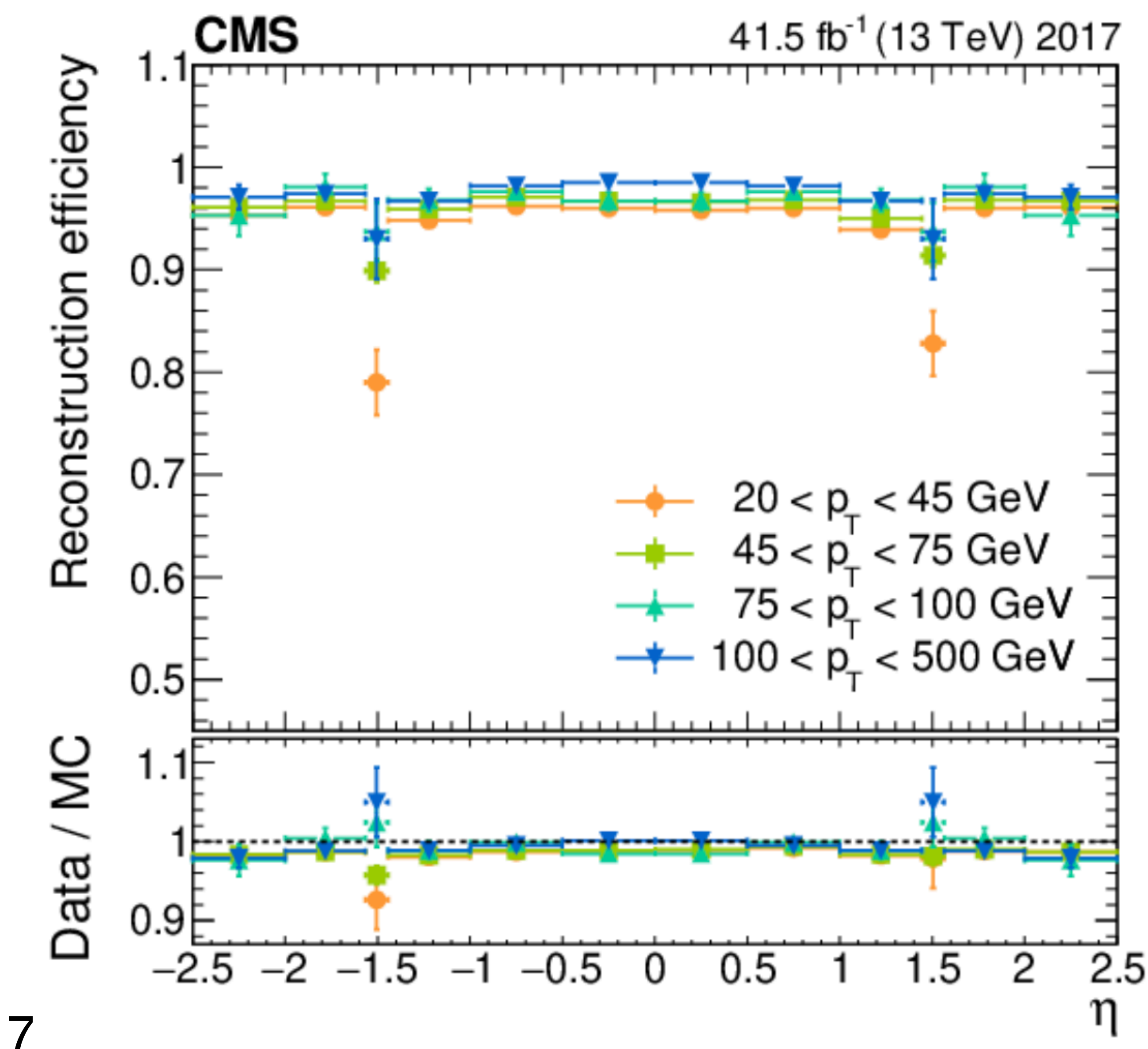
Reconstruction: ECAL + Tracker

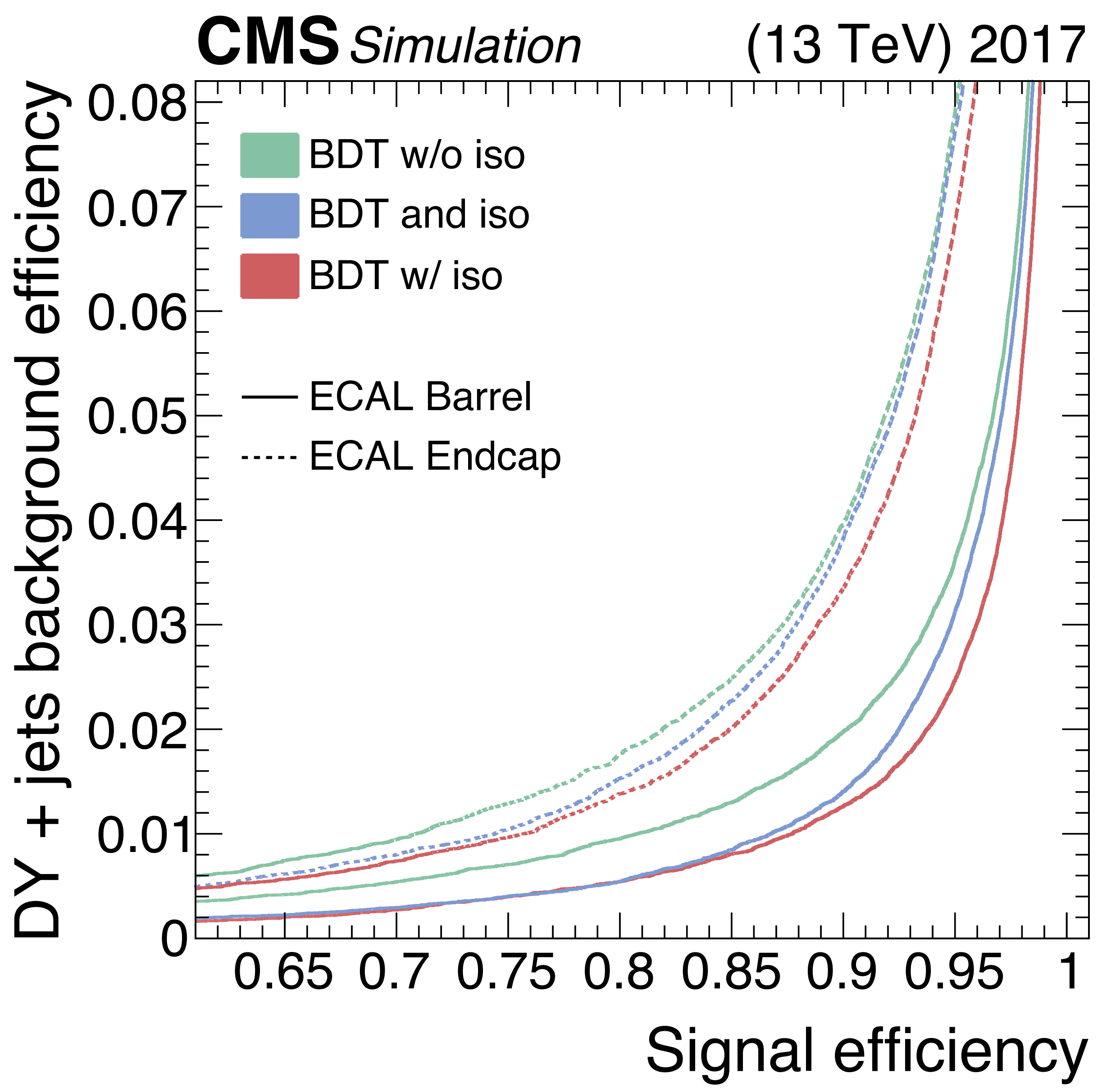
- ECAL: Cluster formation → seed cluster choice (highest energy) → Superclustering
- Tracker: Gaussian Sum Filter algorithm, reconstructing electron trajectories, based on tracks selected by either trk driven or ECAL driven match

Identification: MVA

- Two BDTs with a set of observables
  - One including Iso
  - Other without Iso







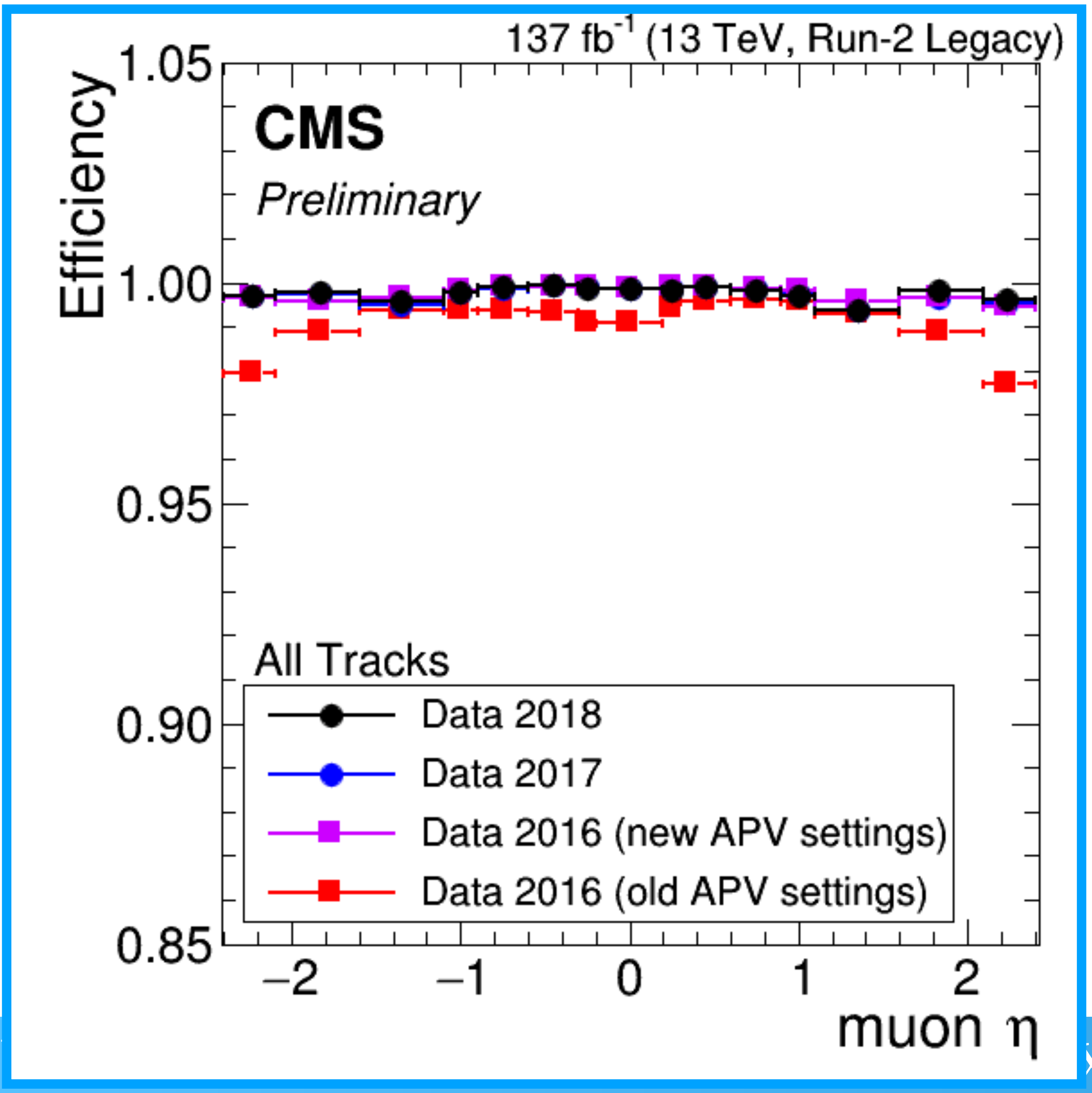
Muon reco and ID performances

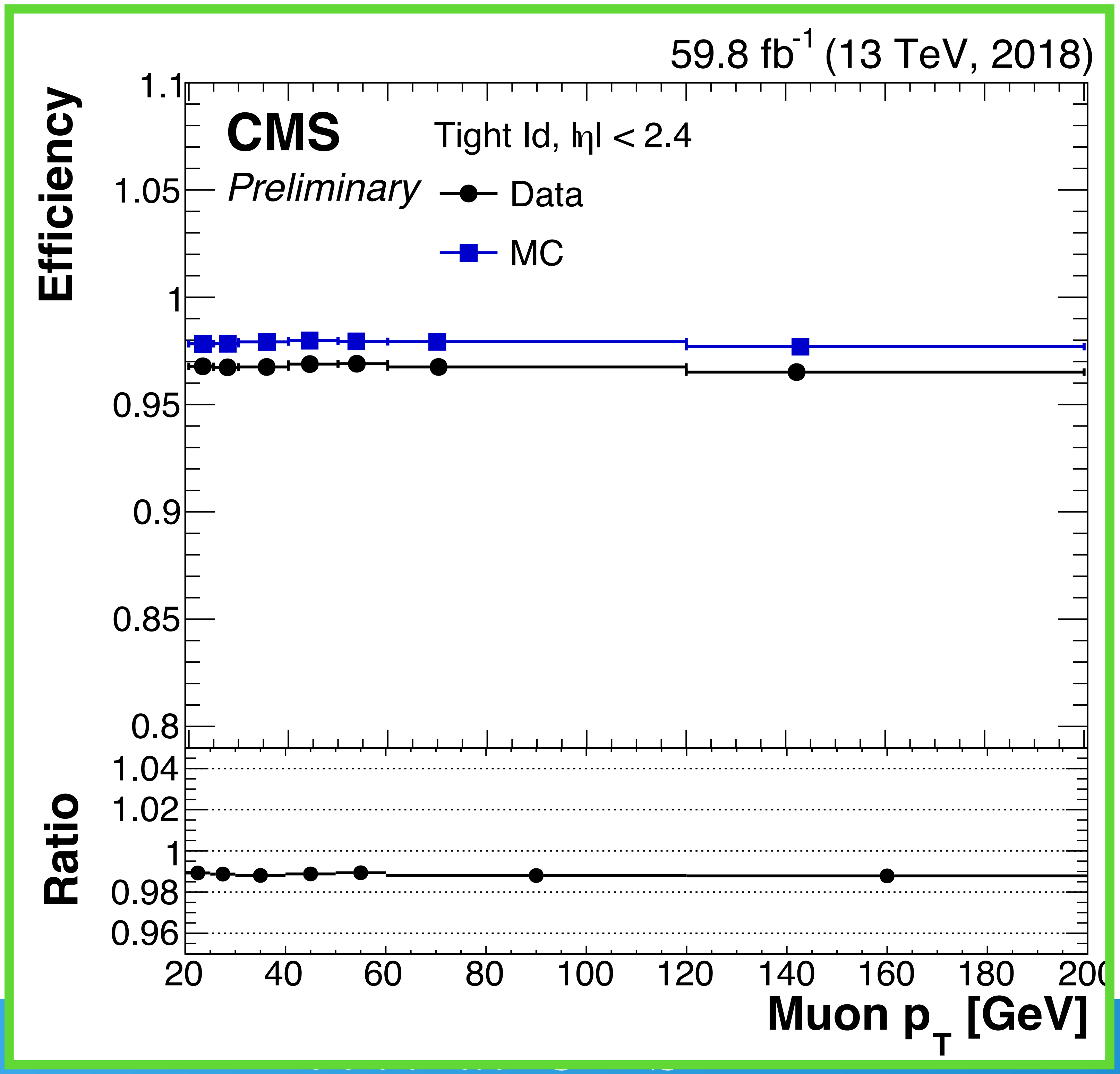
- Reco:
  - Standalone muons: from muon system only
  - Tracker muon: extrapolating tracks in the inner tracker propagating to the muon system ("inside-out")
  - Global muons: matching standalone with tracker muons ("outside-in")
- ID: relies on muon observables, many different WP based on the physics needs
- Iso: to distinguish prompt from non-prompt muons

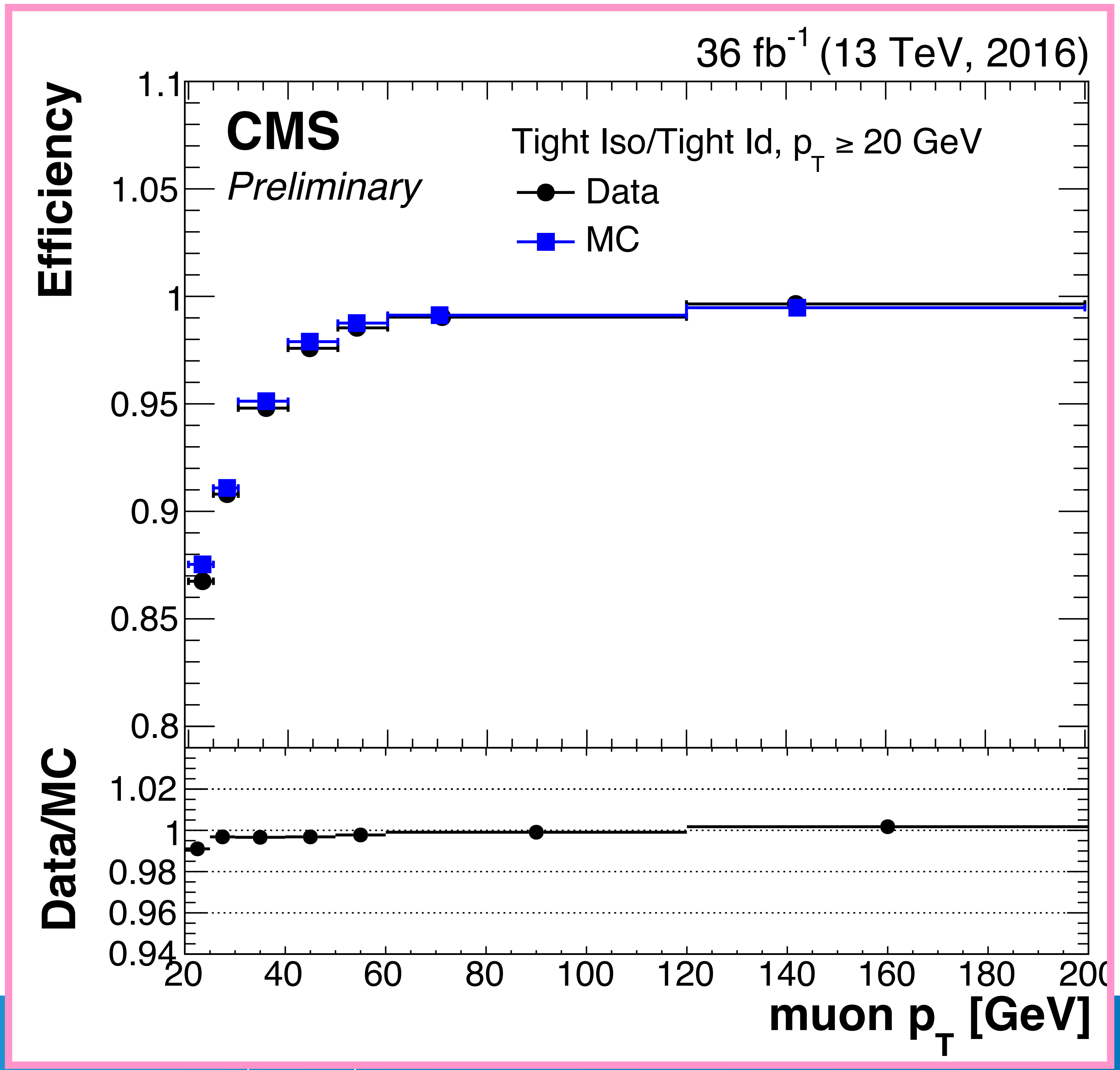
$$I^{\mu,\text{abs}}_{PF} = \sum_{h^{\pm},HS} p_T^{h^{\pm}} + \max\left[0, \left(\sum p_T^{h^0} + \sum p_T^{\gamma} - \Delta\beta \sum_{h^{\pm},PU} p_T^{h^{\pm}}\right)\right]$$

$$\epsilon_{\mu} = \epsilon_{\text{track}} \times \epsilon_{\text{reco+ID}} \times \epsilon_{(\text{ISO}|\text{ID})} \times \epsilon_{\text{trg}}$$

→ In slide 54

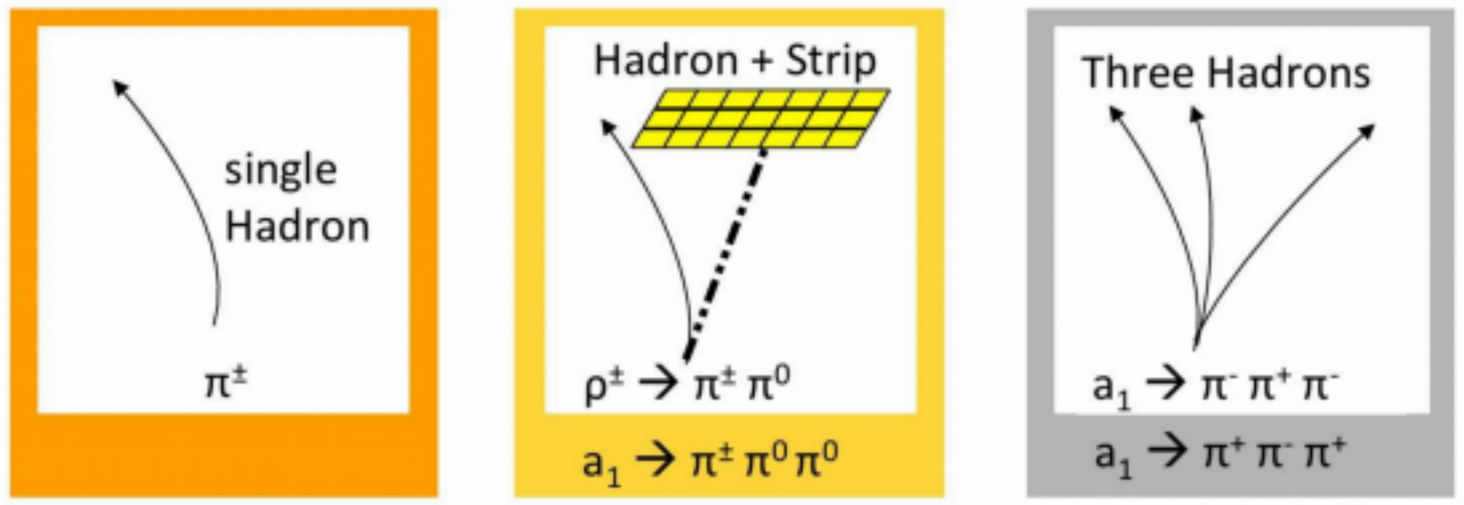


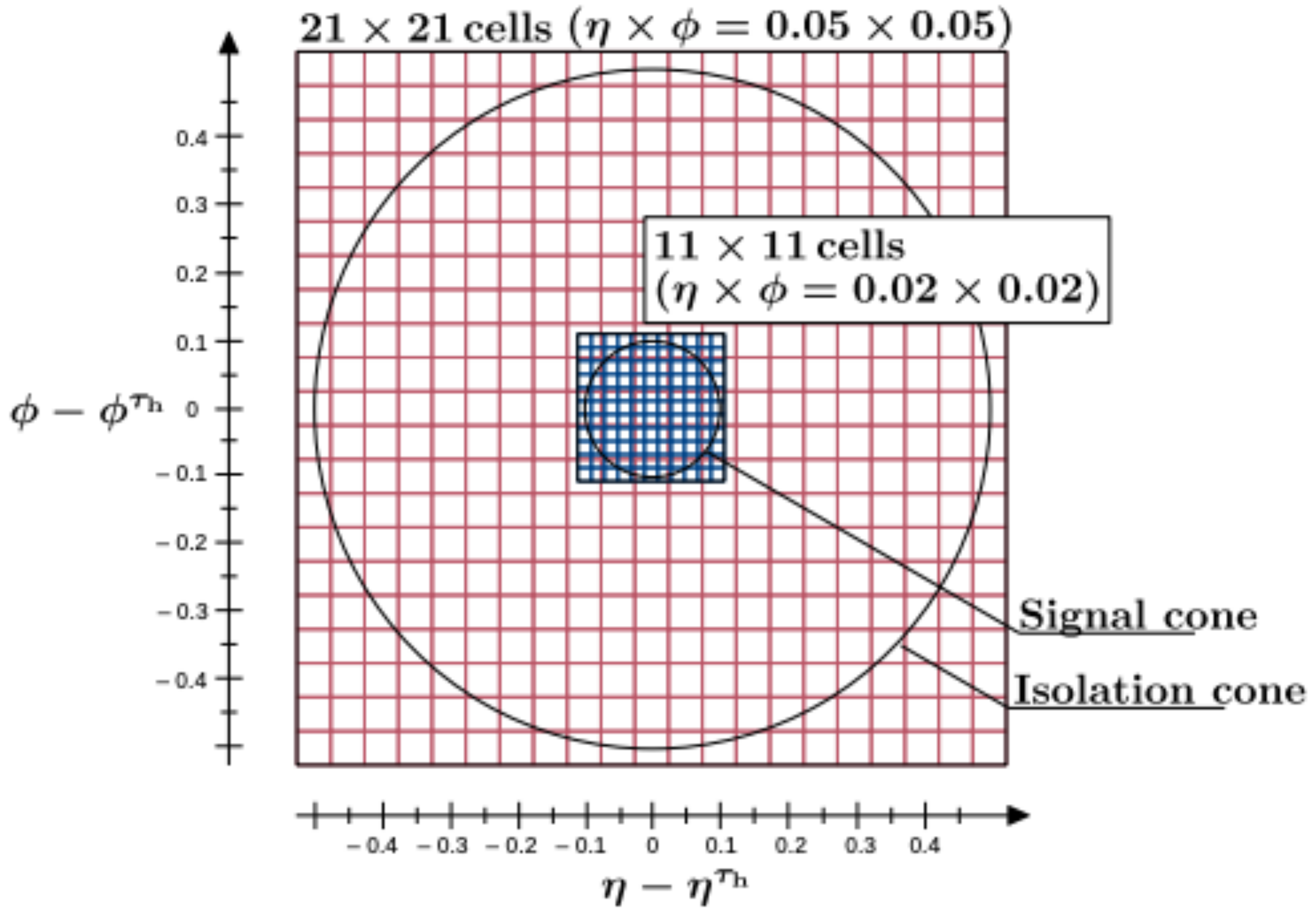


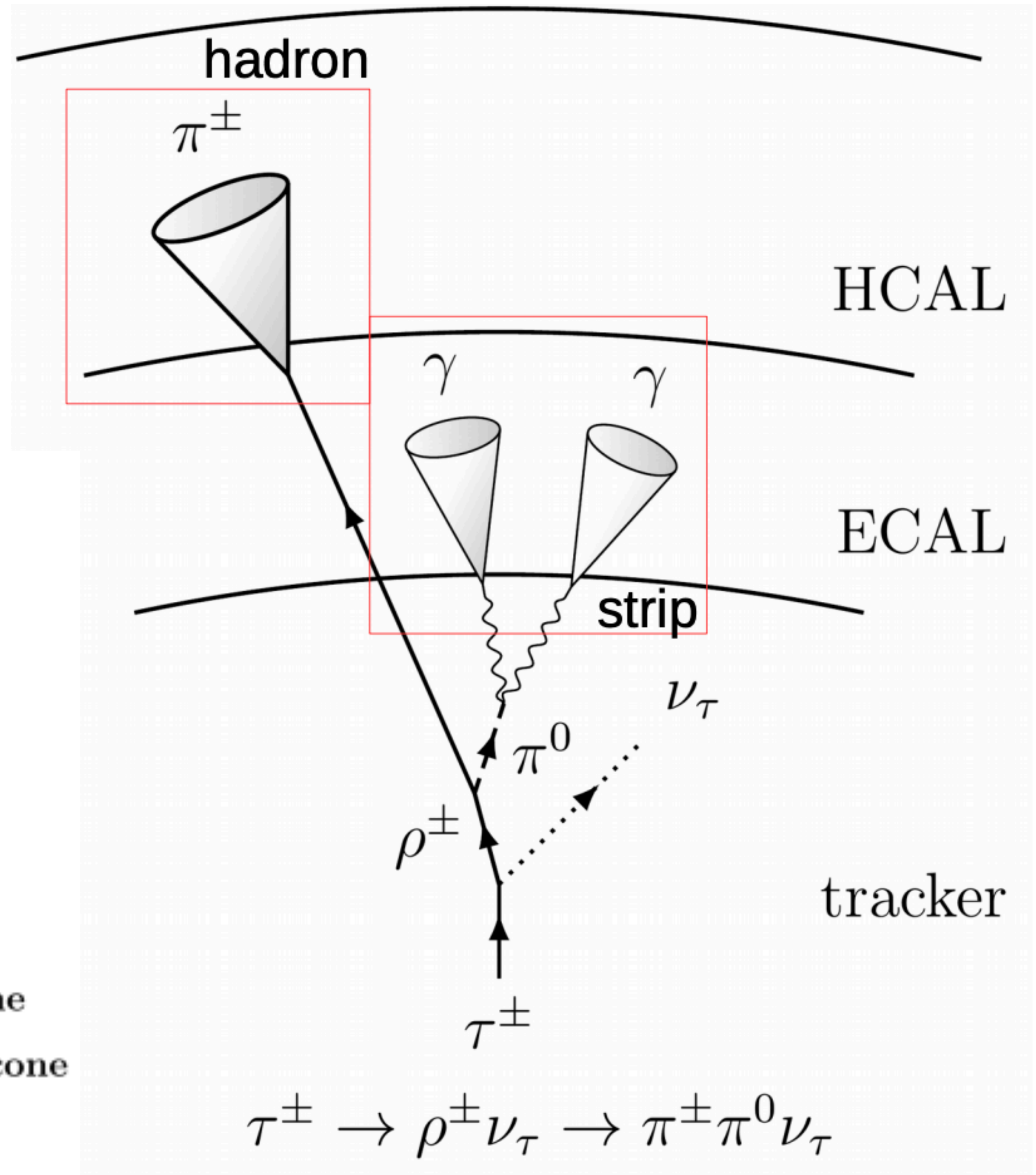


Hadronic τ reco

- Reco: Hadron + strips (HPS) algorithm combining:
  - PF Candidates for jets → hadronic jets
  - EM showers in ECAL which are elongated in φ → strips
- 3 reconstructed decay channels in older version → now 4 decay modes are identified
- ID: deepTau, multiclass τ identification algorithm based on a convolutional deep neural network
  - Particles (PF candidates, or fully reconstruction leptons) belonging to the signal and isolation cones are split into ηxφ two grids



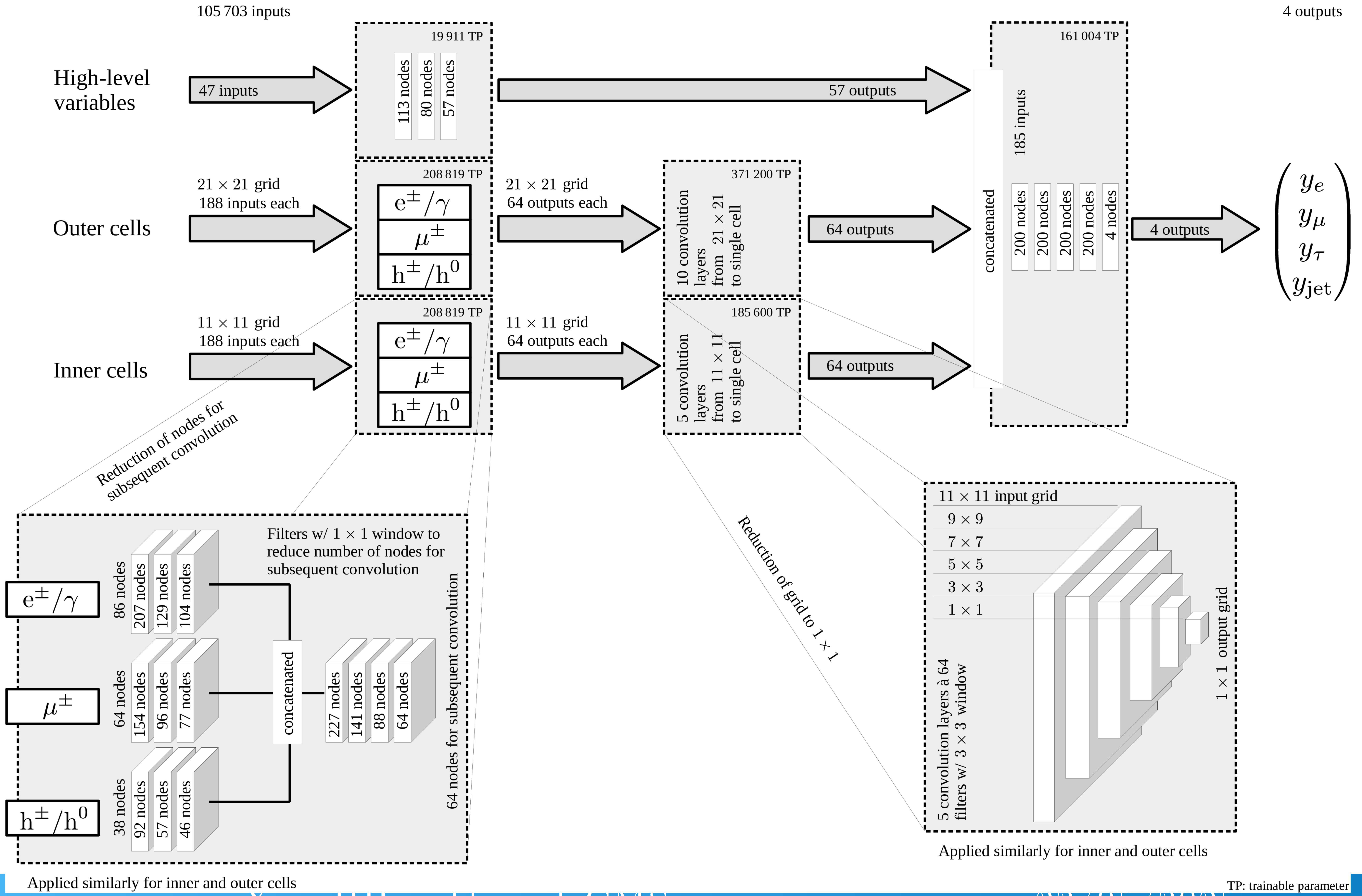




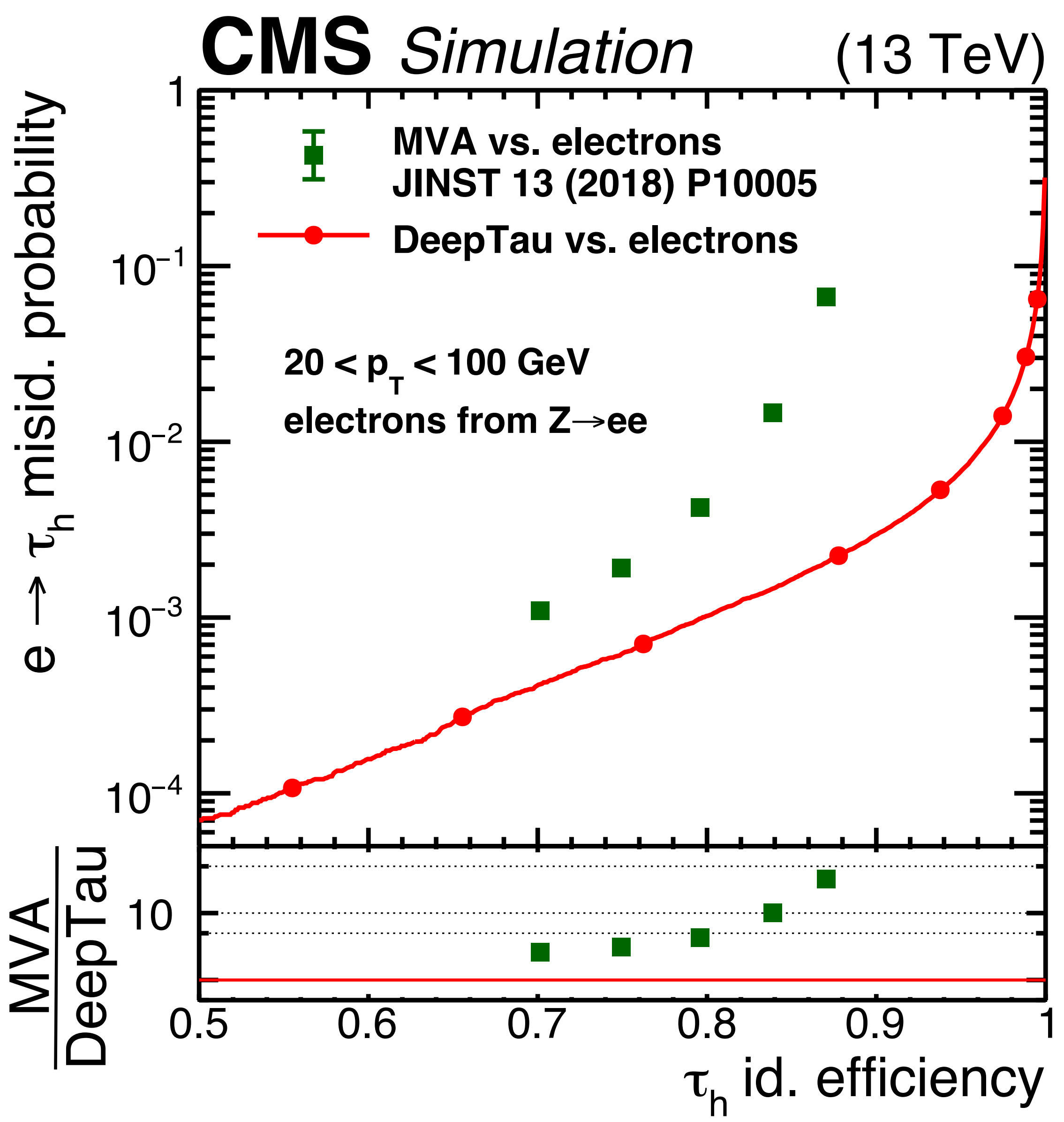
DeepTau discriminator

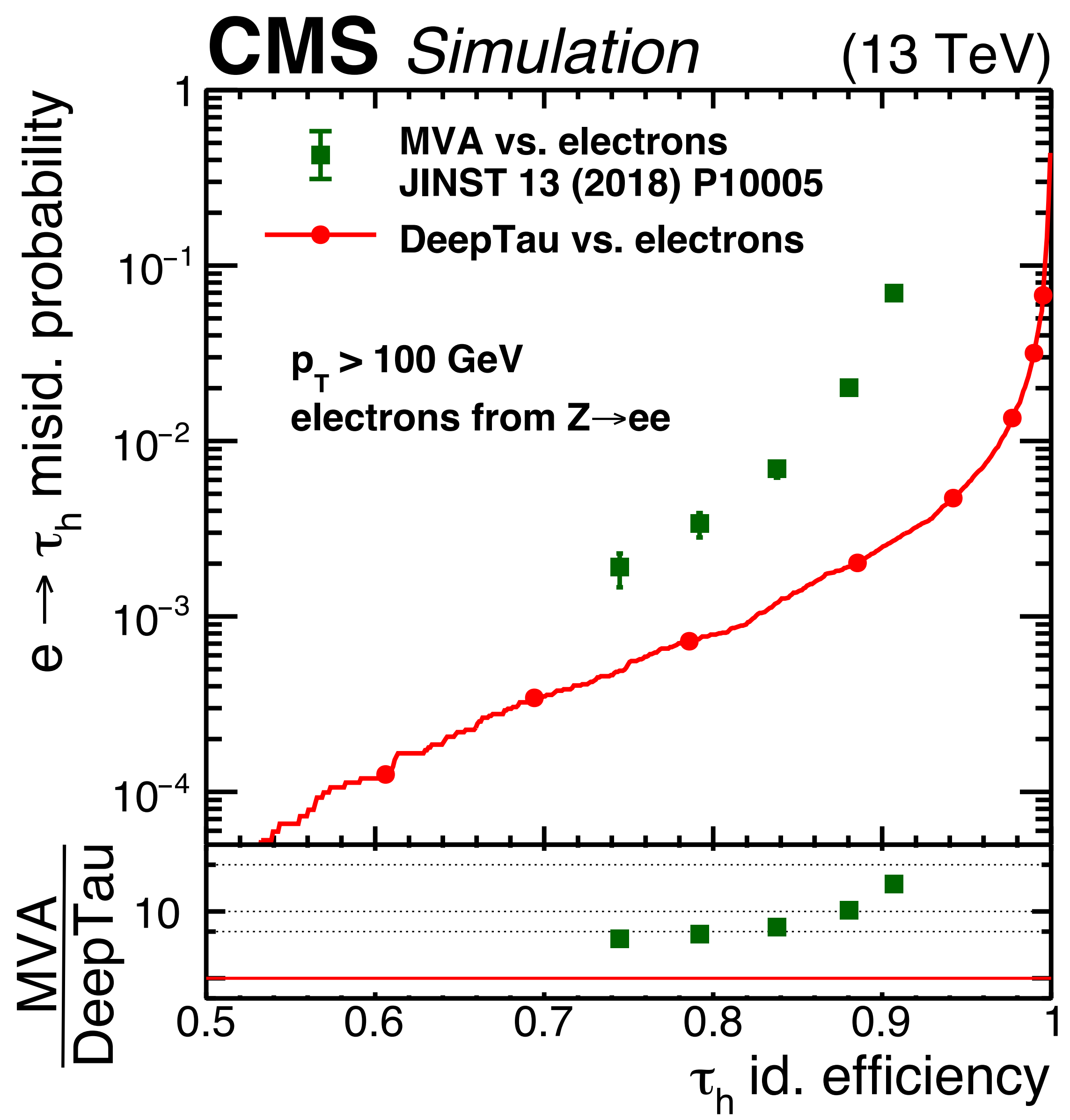
$$D_\tau^\alpha(p) = \frac{p_\tau}{p_\tau + p_\alpha}, \quad where\ \alpha \in \{e, \mu, j\}$$

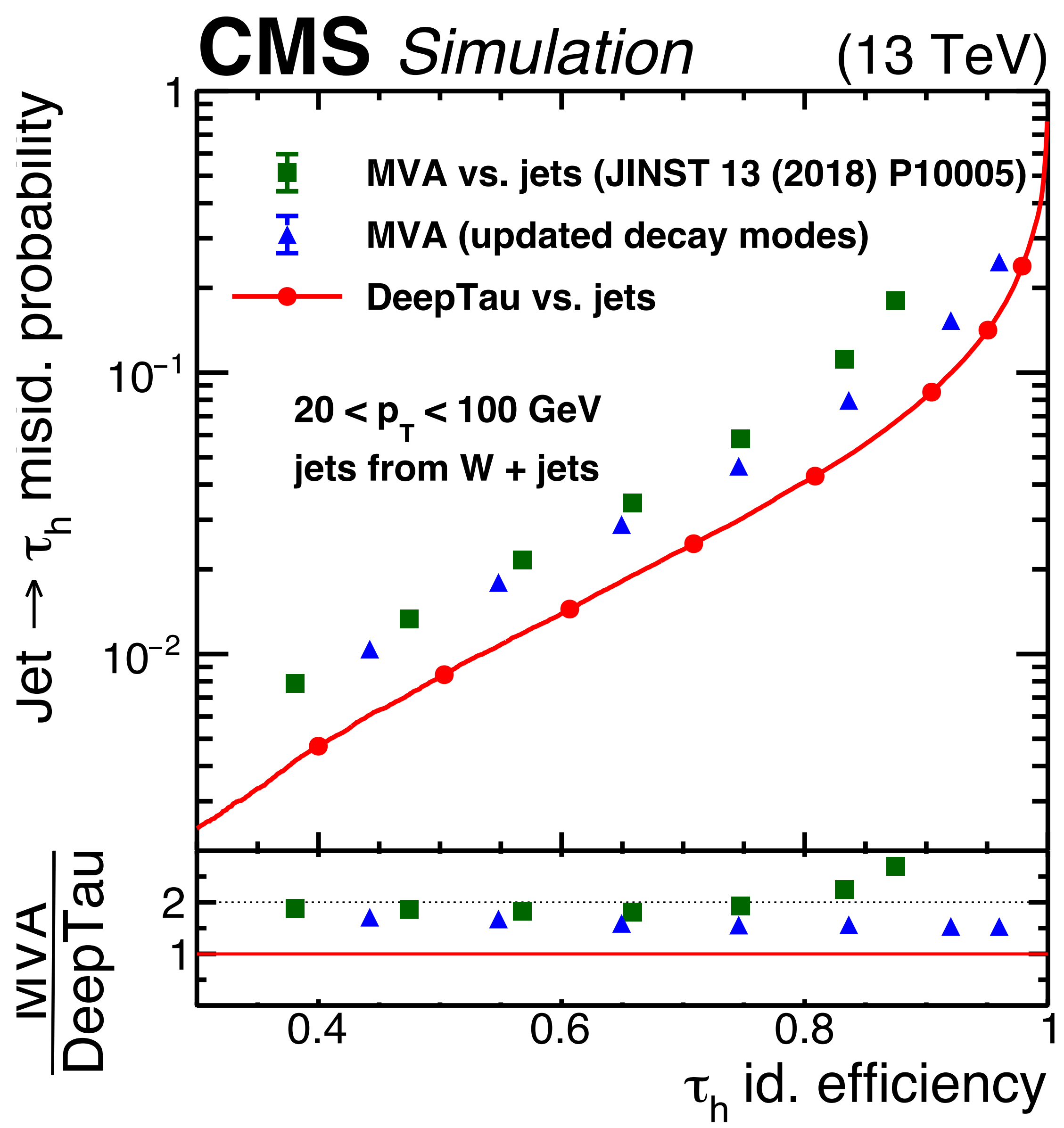
DeepTau v2p1 architecture

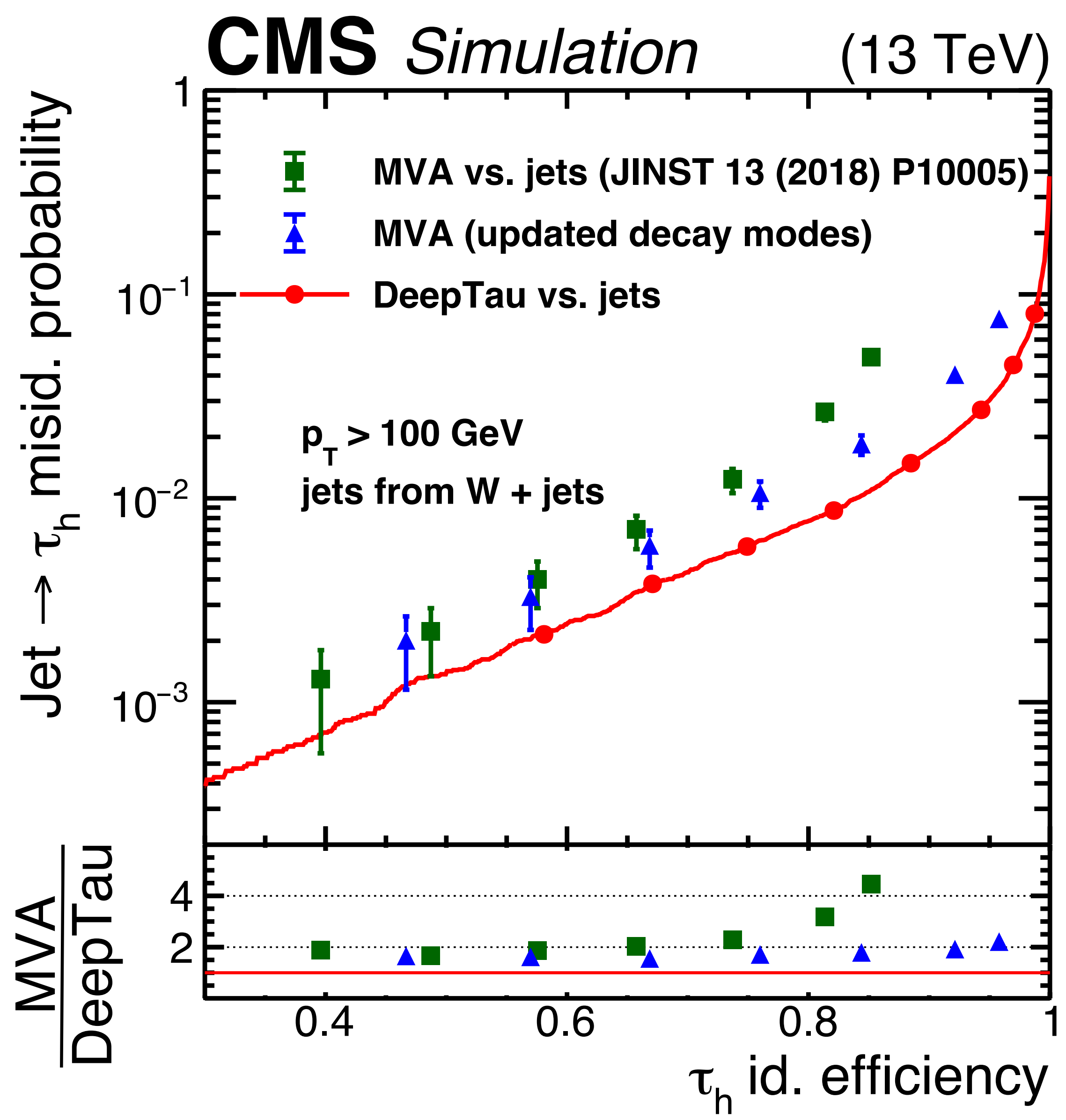


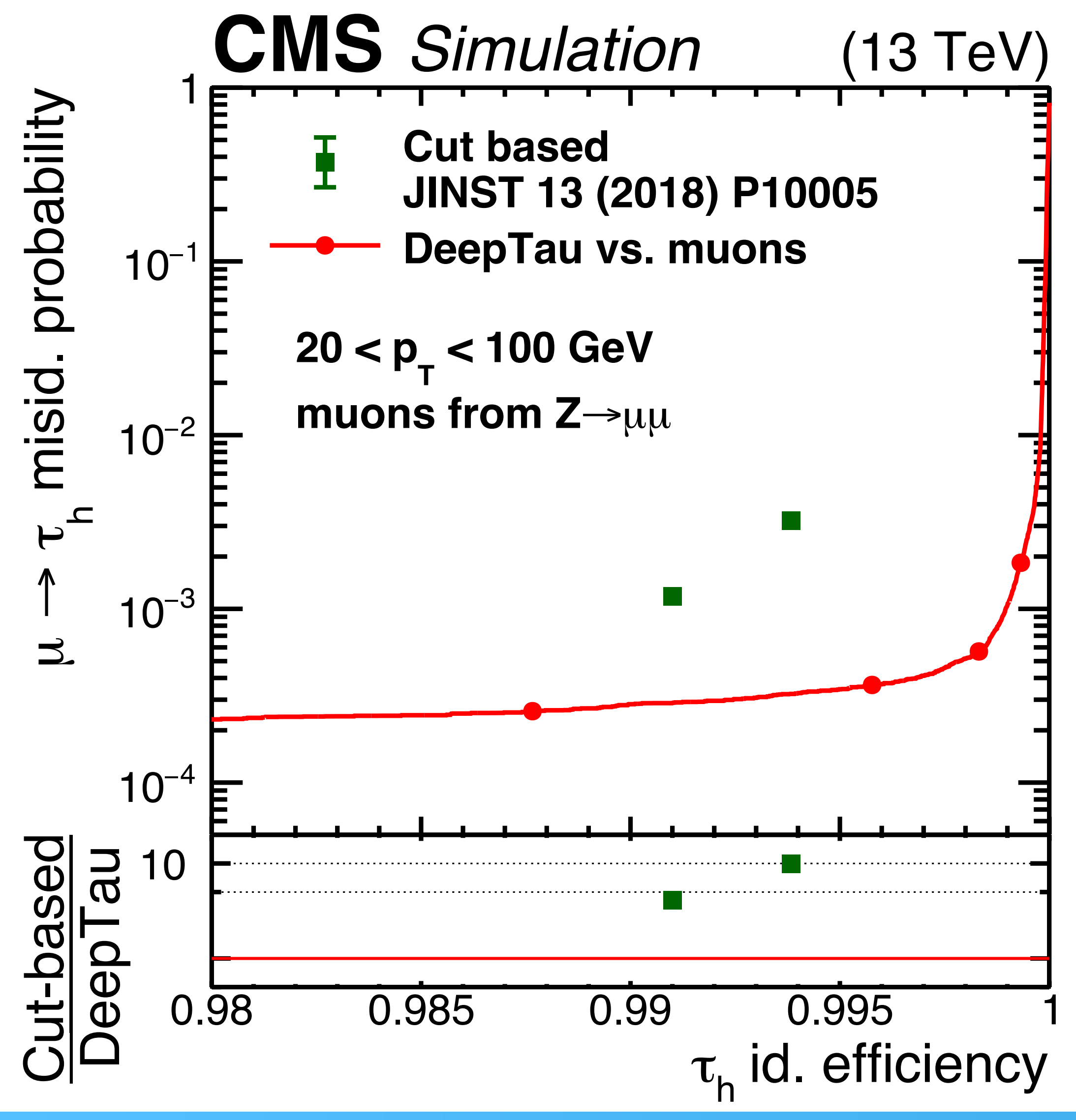
DeepTau v2p1 performances

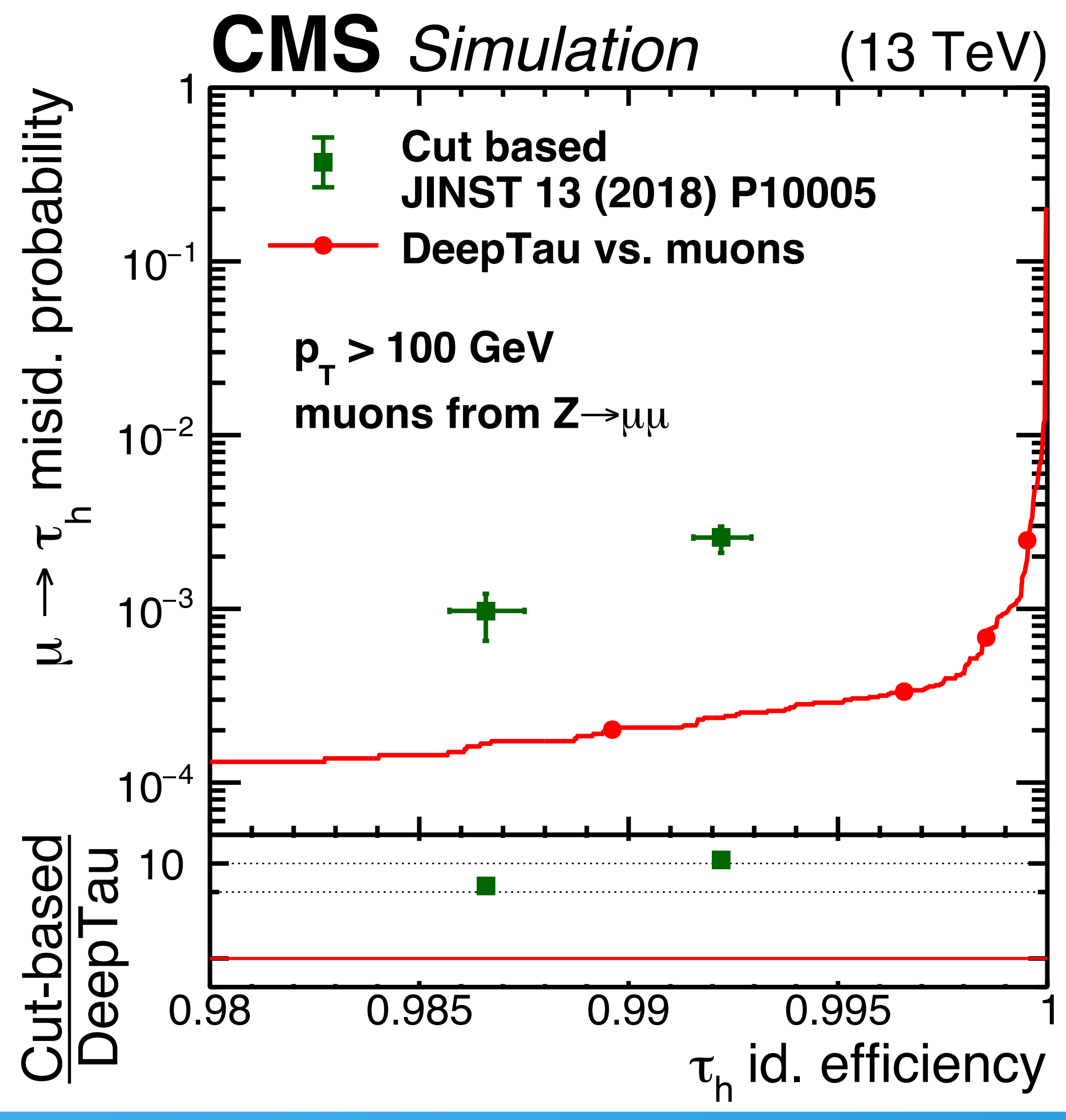


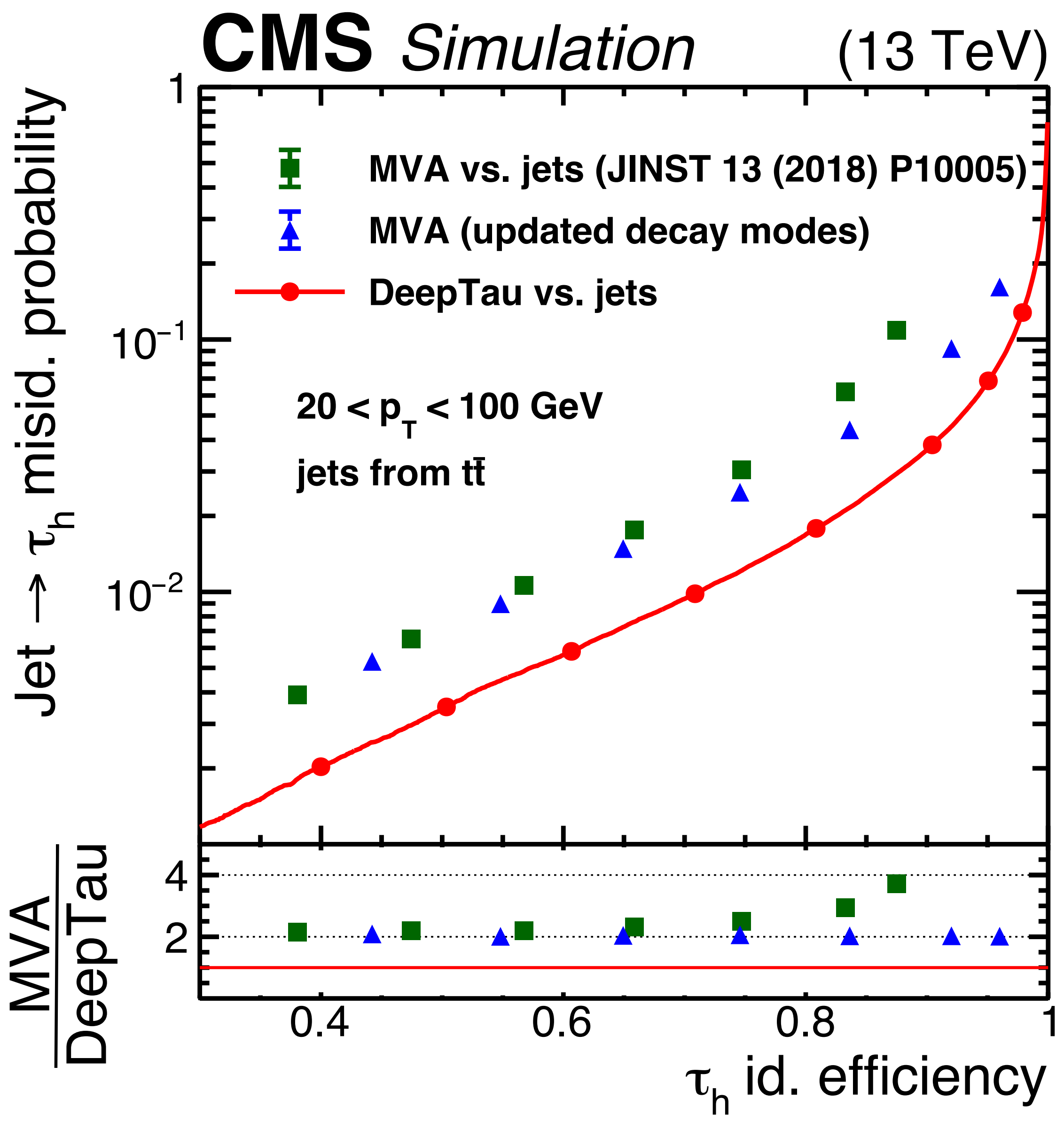


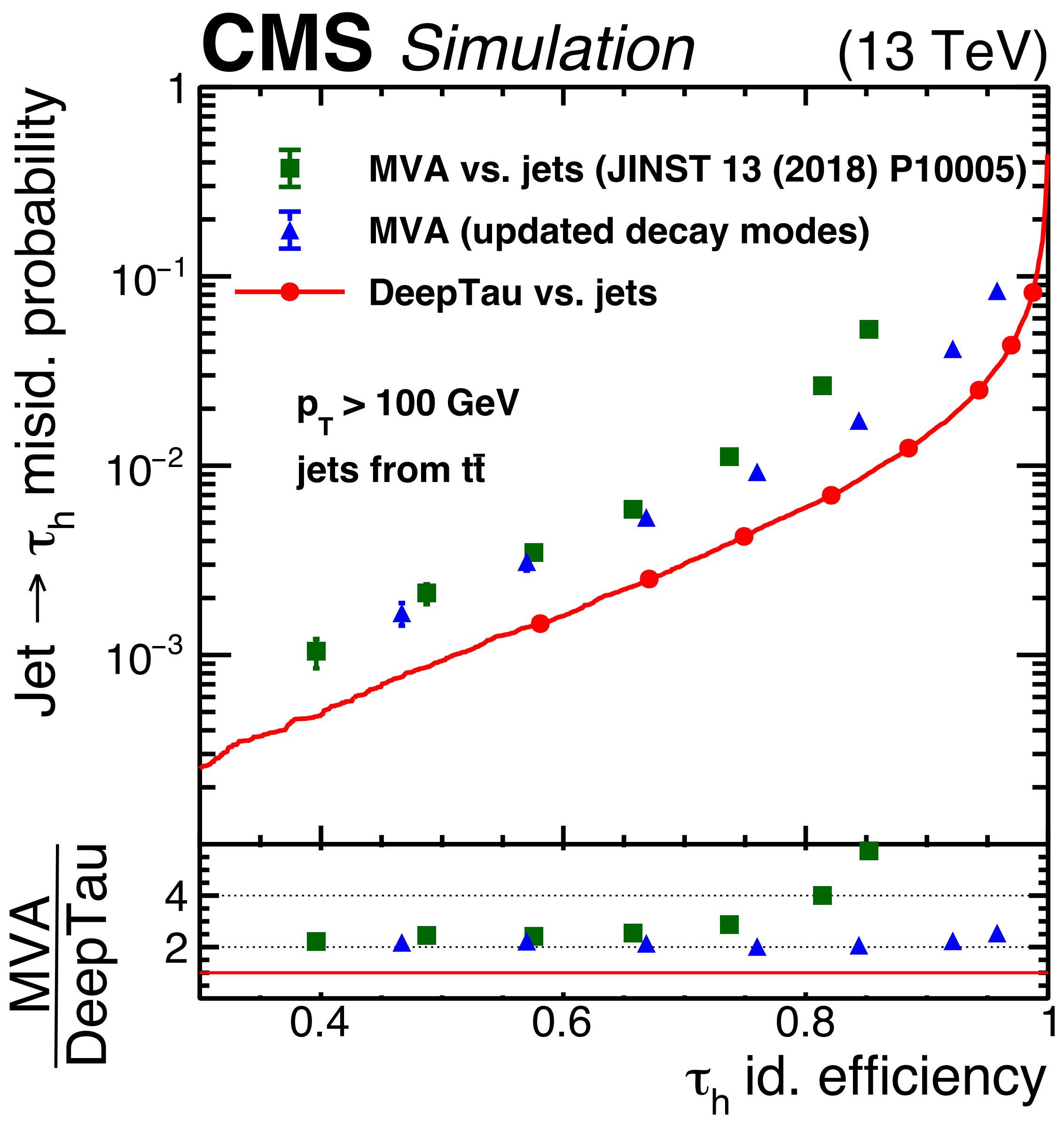






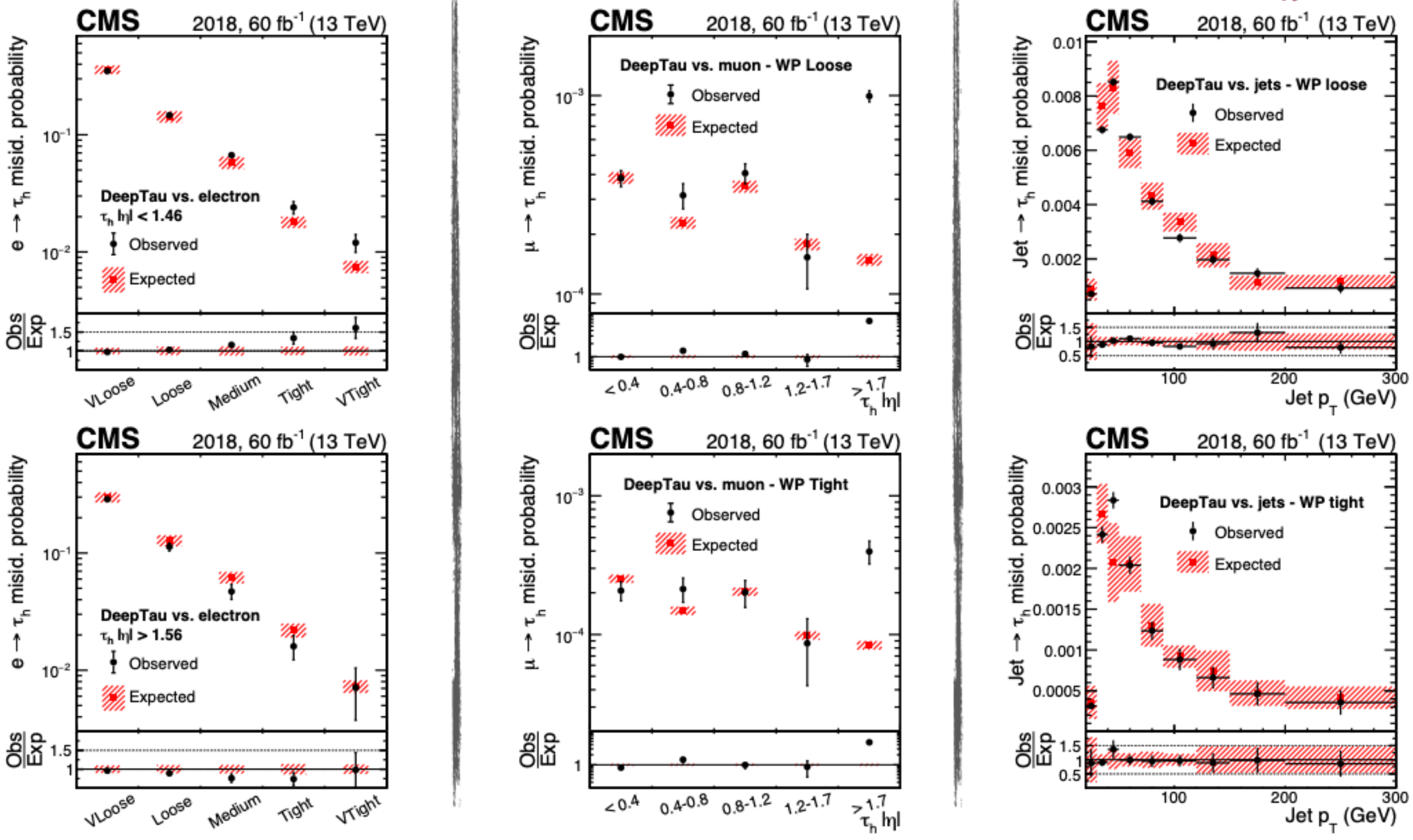


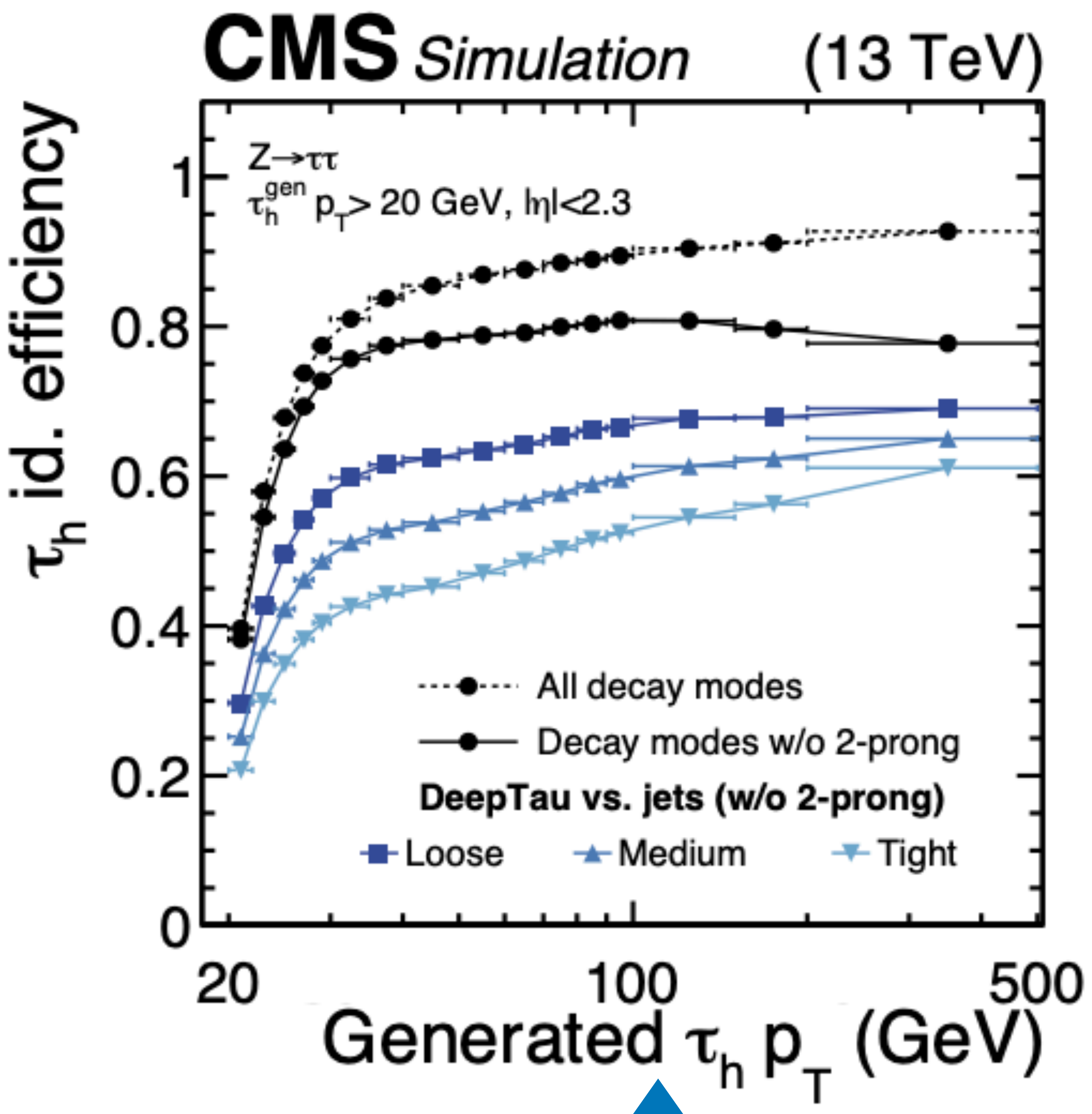




DeepTau v2p1 performances

$e \rightarrow \tau_h$ | $\mu \rightarrow \tau_h$ | $jet \rightarrow \tau_h$





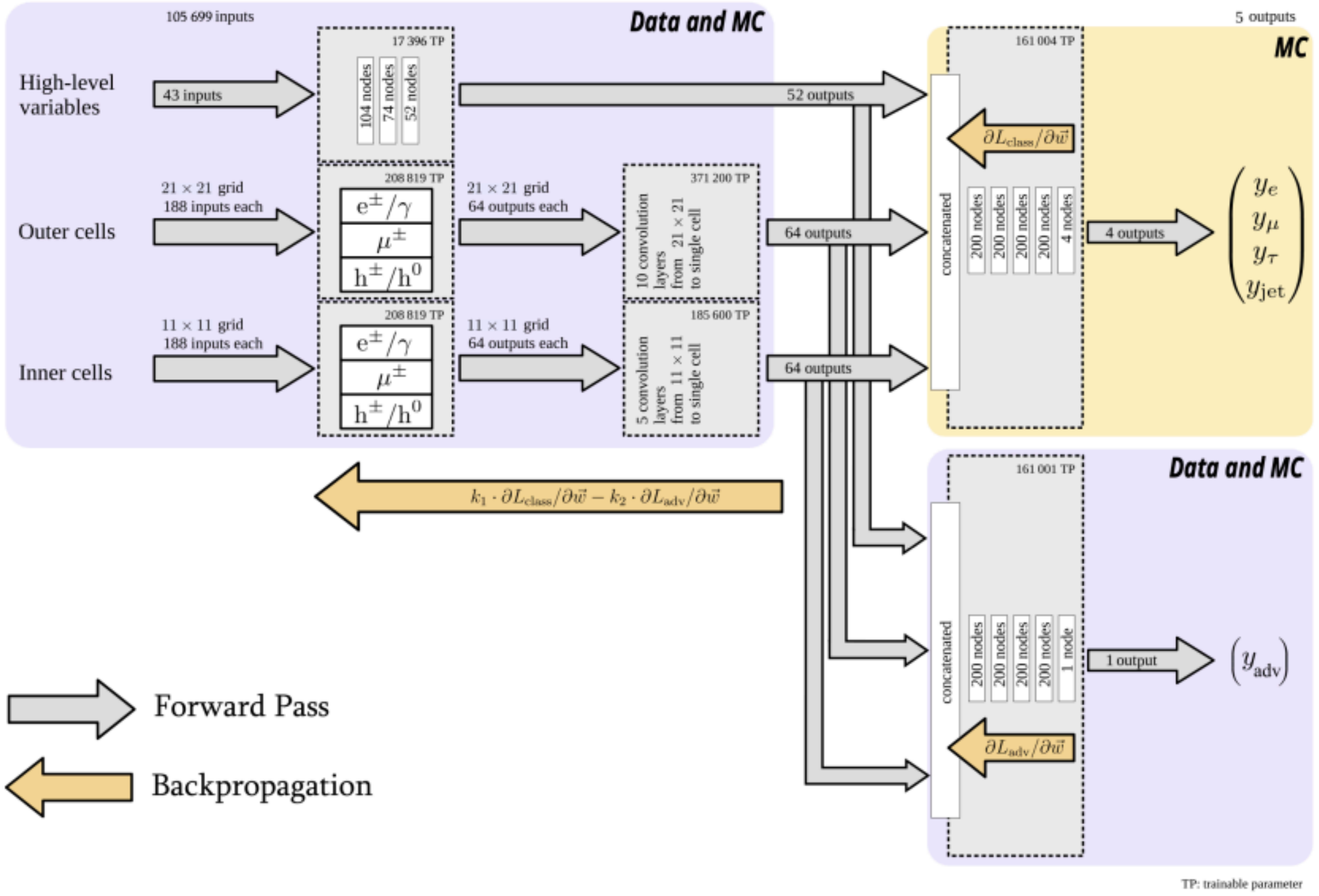
Genuine taus

Fake taus

Good modelling of fake τs in the most parts of the parameter space

DeepTau v2p5 architecture

- Network Structure:
  - Input: high-level reconstructed features + info on PF candidates
  - within and around η x φ grid format (Grid data processed by convolutional layers)
  - last stage: all features combined and pass through 5 dense layers
  - Final output: Probability of of being genuine or a misid electron/muon/jet
- Improvements
  - Domain adaptation techniques to mitigate mismodelling (with respect to real data) in MC events used for training
  - Shuffle and Merge to balance events across all regions of the phase space
  - Feature Standardisation and Hyperparameter Optimization

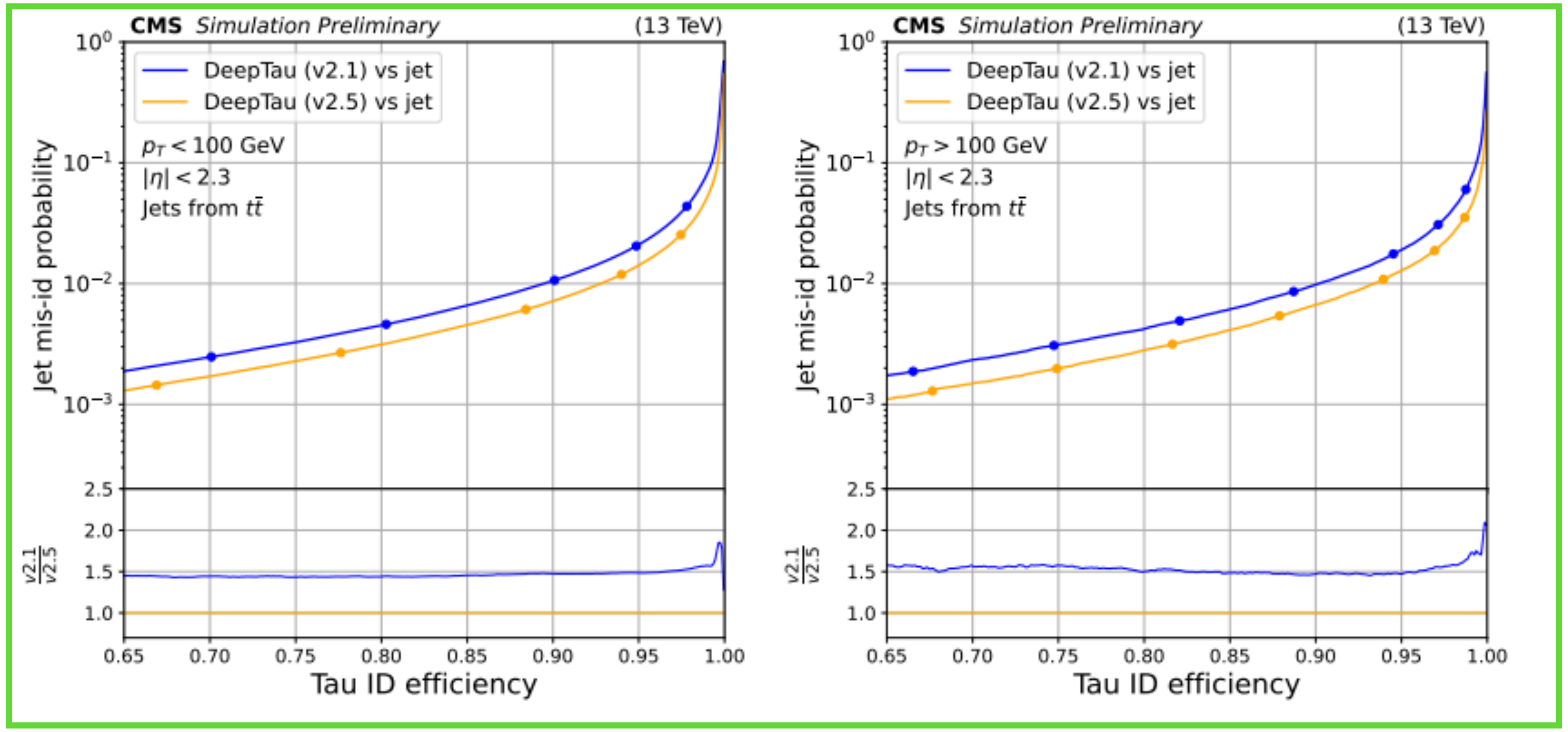


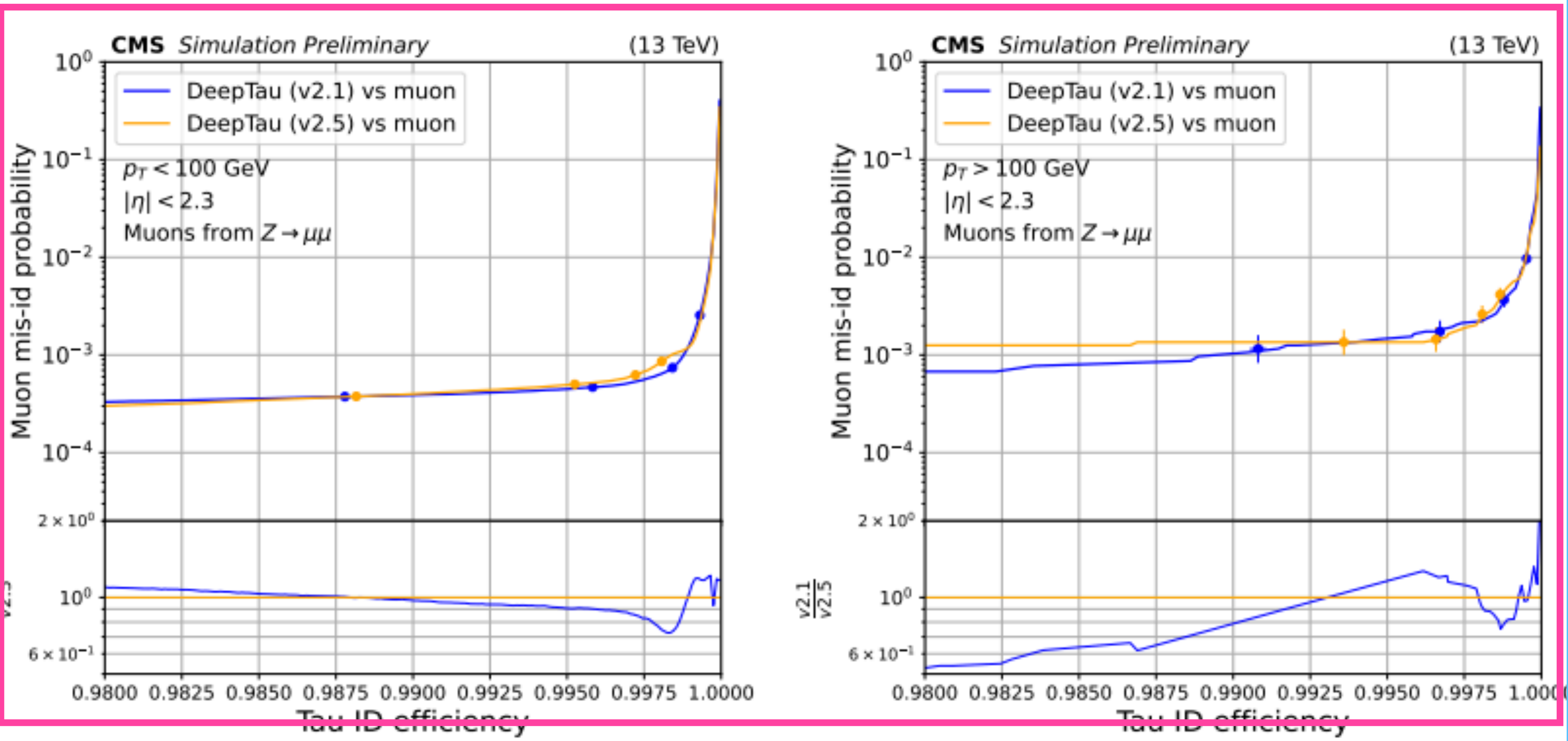
CMS-DP2024-063

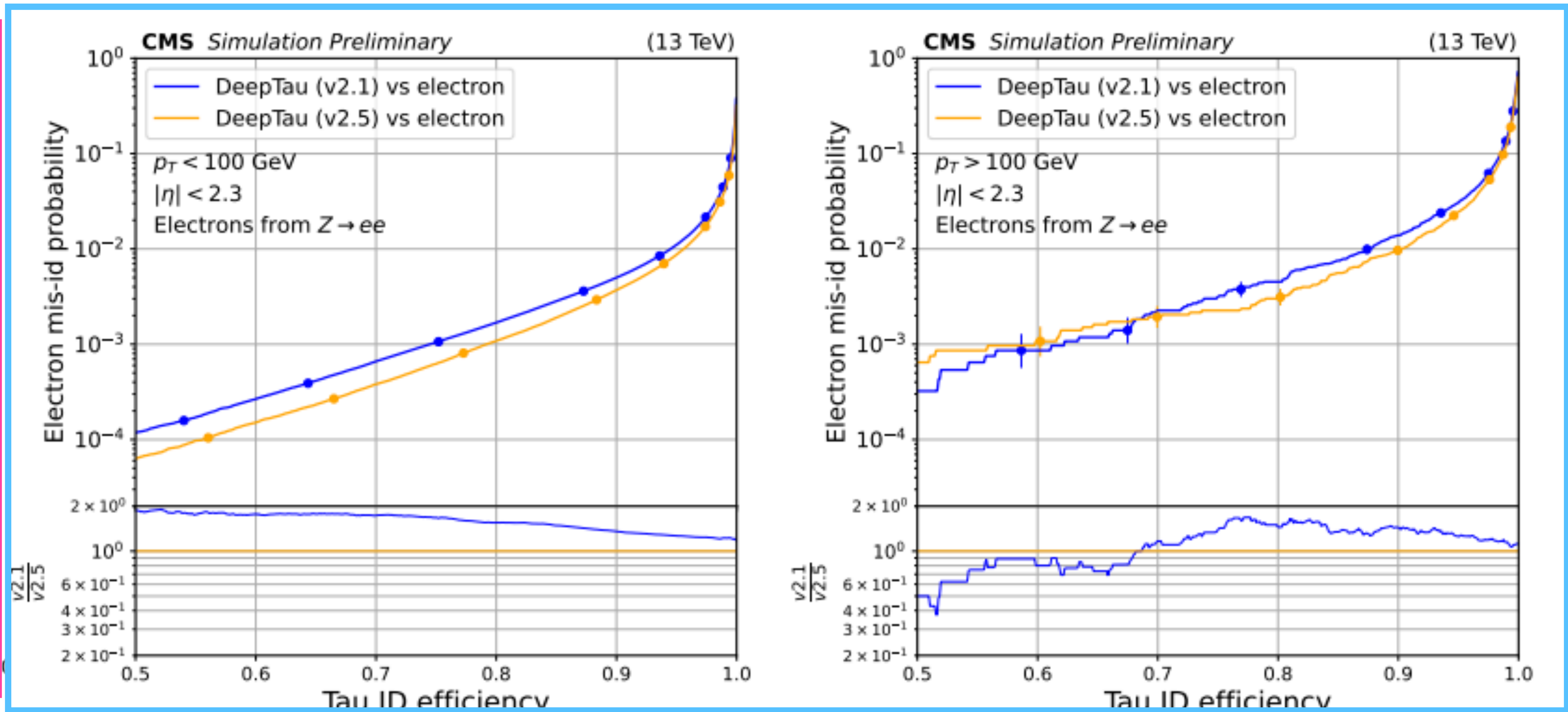
DeepTau v2p5 Performance

CMS-DP2024-063

- deepTauVSJet: Jet MisID reduced by ≈50% and τh ID significantly improved
- deepTauVSEle: Small degradation at high pT but overall a good performance
- deepTauVSMu: compatible with 2p1



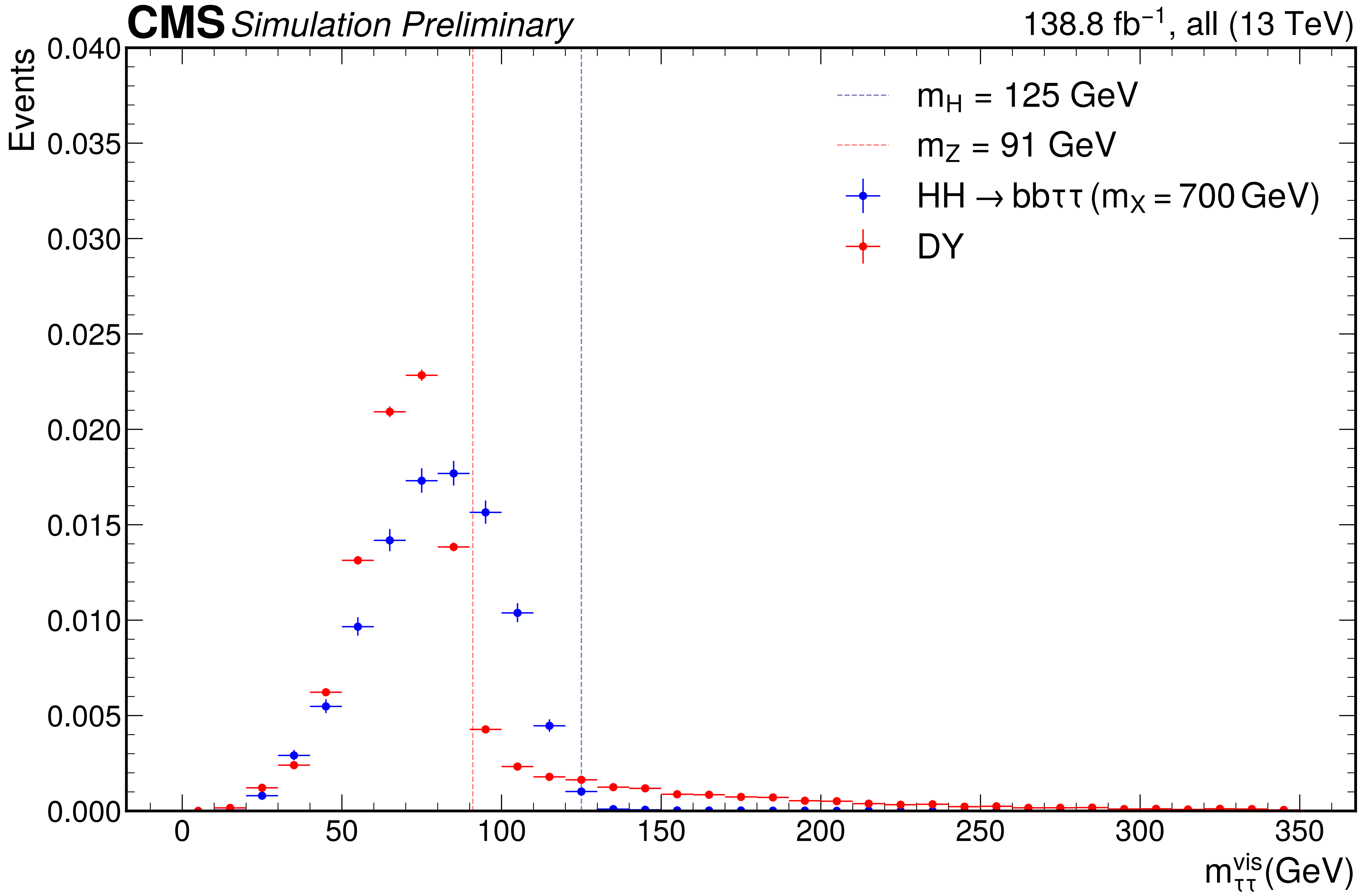


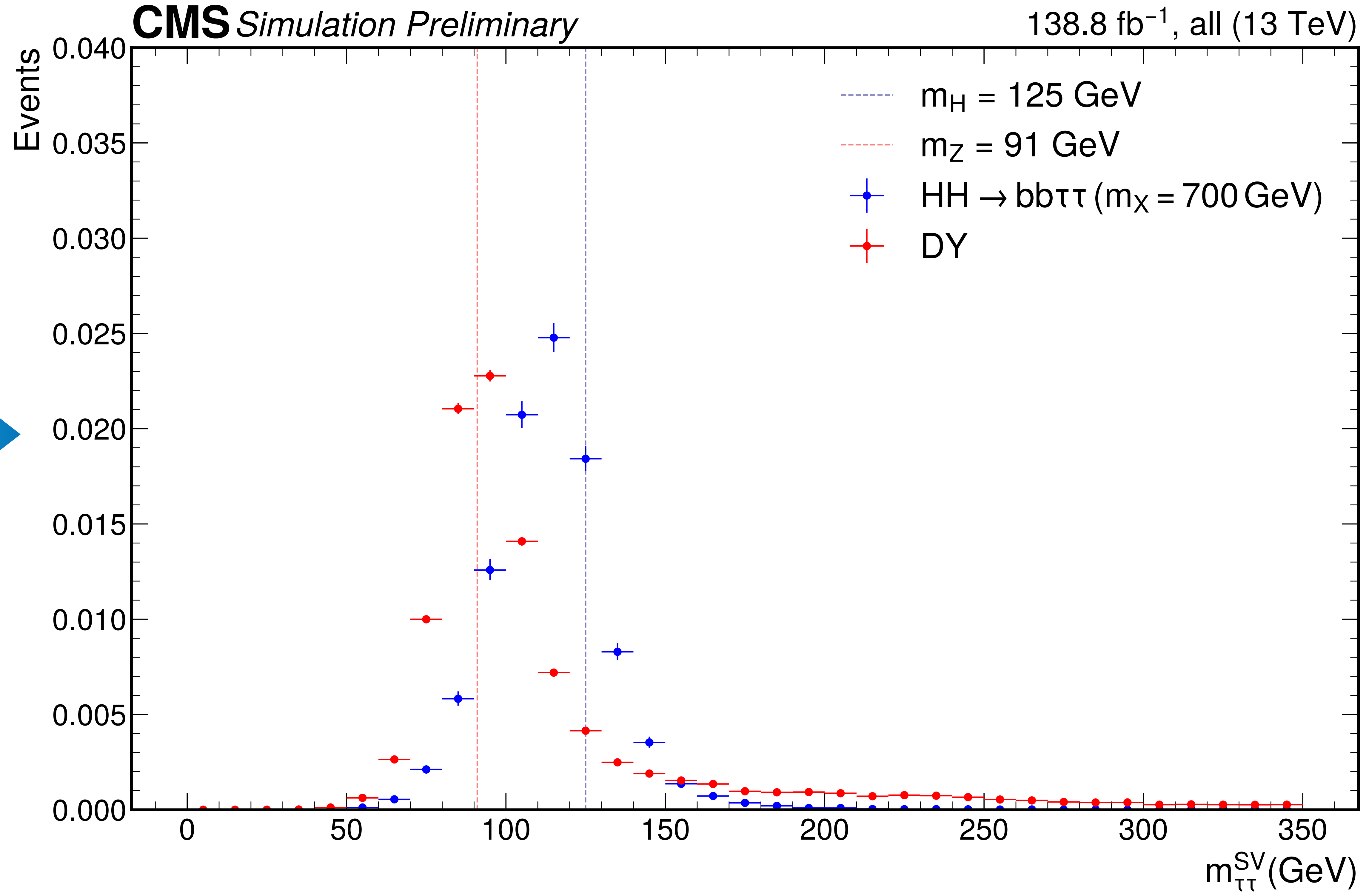


H(ττ) mass reconstruction

- $m_{\tau\tau}$ is reconstructed with SVFit, a likelihood based algorithm: $m_{\tau\tau}^{SV}$
  - Recovers the information of missing transverse energy from neutrinos
  - Improves the reconstruction of the tau pair invariant mass w.r.t. visible mass

SVFit details in backup





The H(ττ) pair assessment

Define all the possible leptons pairs in the following priority order: μμ, eμ, ee, μτ, eτ, ττ

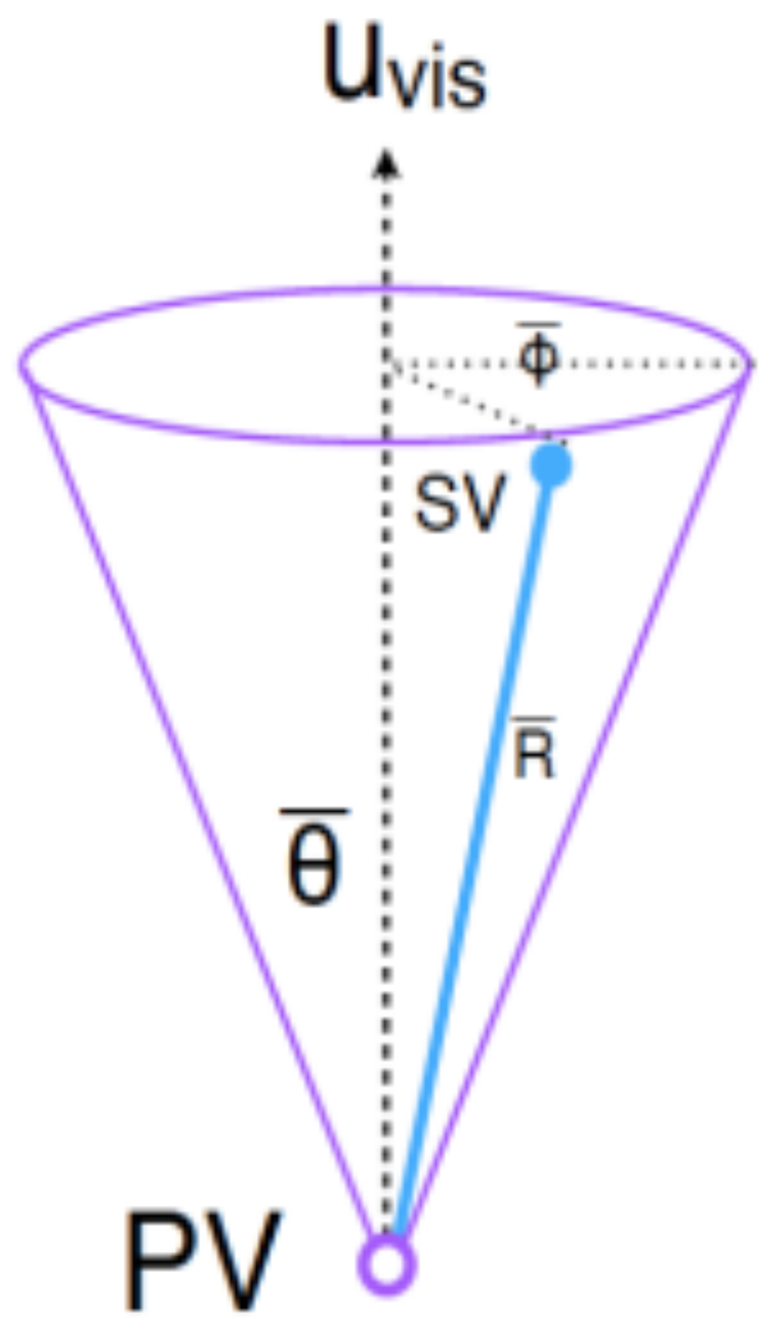
The H(ττ) candidate in the lepton pairs selecting:

- The pair with the first "leg" (electron for eτ, ee and eμ, μ for μτ, μμ and τ for ττ) having:
  - Highest iso (PF Iso for electron and muons, deepTauVSJet for taus)
  - Highest p_T
  - Highest $|\eta|$
- Otherwise same criterion looking at the second leg
- If there are still ambiguities, discard the event

The SVFit algorithm

SVfit is a likelihood based algorithm for the reconstruction of h boson decaying to τ leptons.

- The kinematics of τ decays can be parameterized by following variables:
  - θ – the angle between the boost direction of the τ lepton and the momentum of the visible decay products in the rest frame of the τ.
  - φ – the azimuthal angle of the τ in the CMS detector frame.
  - $m_{\nu\nu}$ – invariant mass of the invisible momentum system for leptonic τ decays
- The kinematics of the τ pair decays depends upon 4-6 parameters, which are constrained only by 2 observables from MET
- Using Dynamical Likelihood Methods, SVfit reconstruct kinematic quantities on an event-by-event basis.



Jet reconstruction

Main clustering algorithm: anti-k_T

- Infrared and collinear safe
- Insensitive to underlying event and multiple interactions
- Disfavours clustering between pairs of soft particles
- Cones with well defined area

Distance between particles:

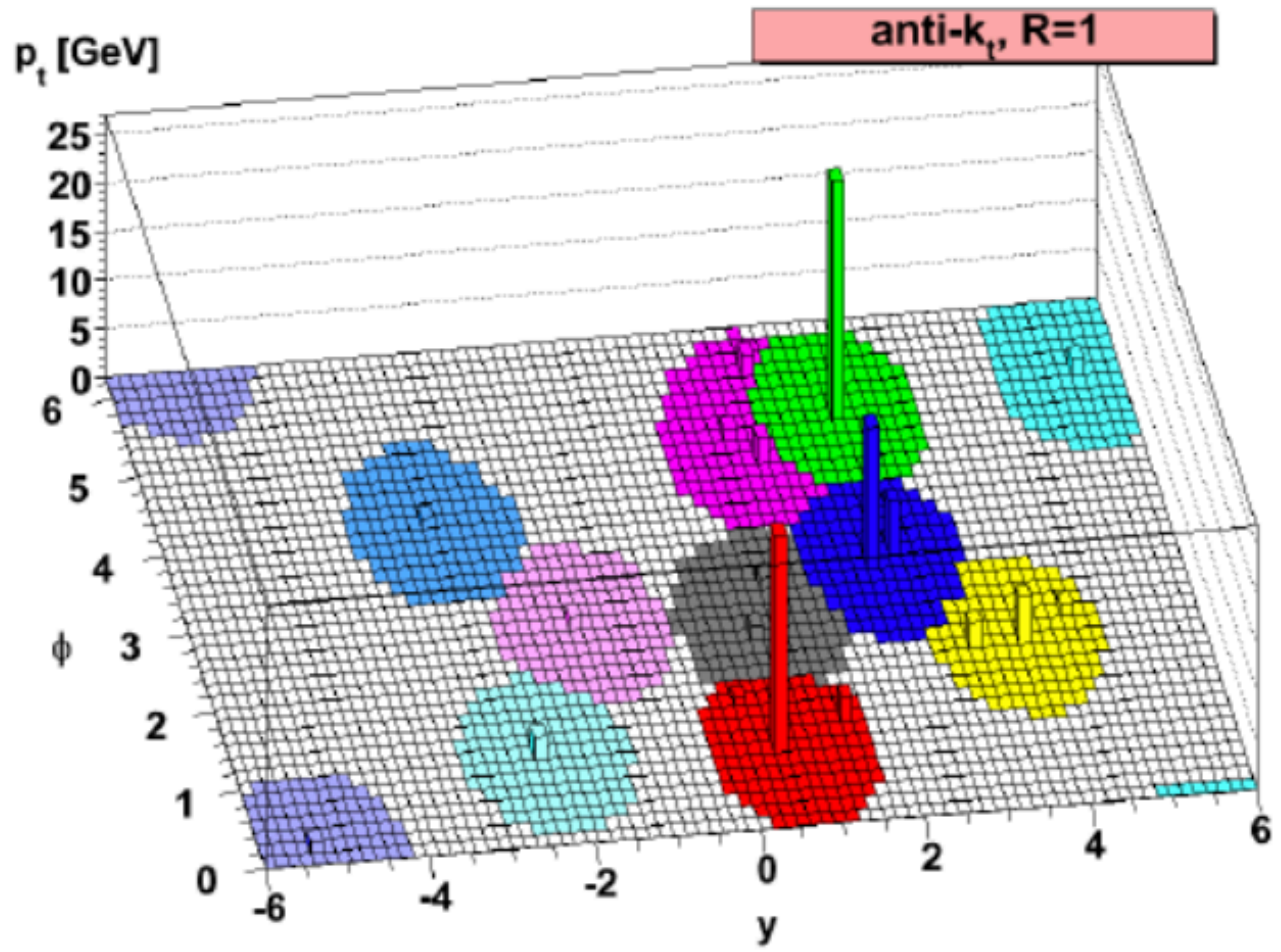
$$d_{ij} = \min(p_T^{-2}(i), p_T^{-2}(j)) \times \frac{\Delta R_{ij}^2}{R^2}$$

With a defined parameter:

$$d_{iB} = p_T^{-2}(i)$$

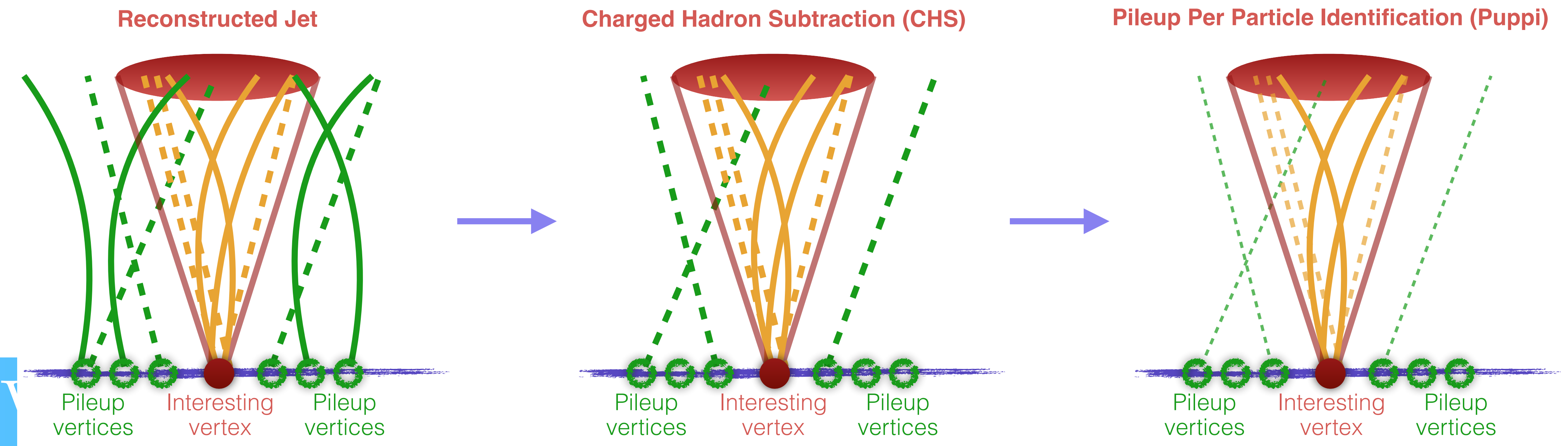
Finish iterations when

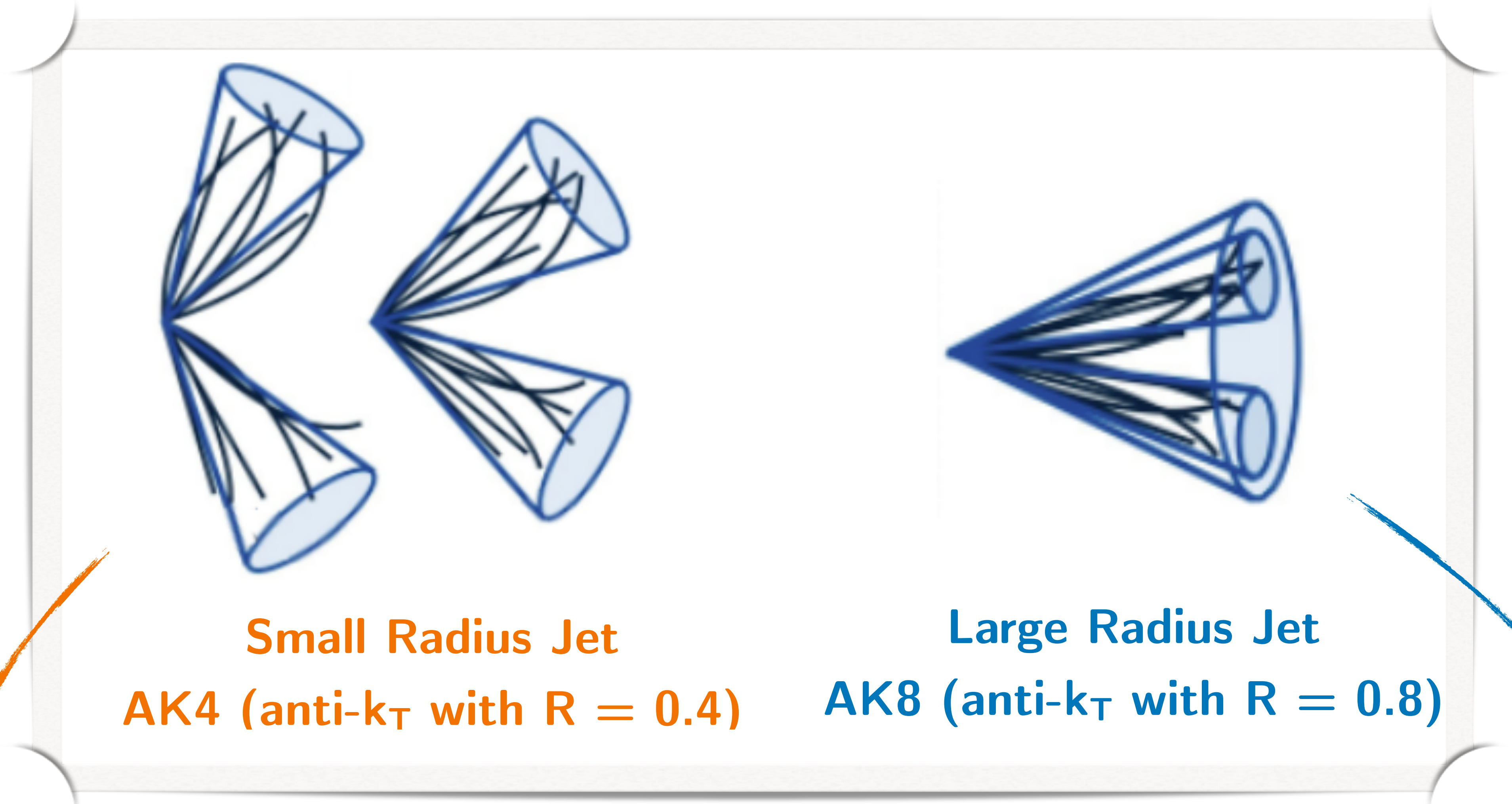
$$\min(d_{ij}, d_{iB}) < d_{cut}$$



Pileup mitigation algorithms in CMS

- Charged Hadron Subtraction (CHS) + PU jet ID - **AK4 jets**
- PileUp Per Particle Identification (PUPPI) - **AK8 Jets**





Light-flavour quarks and gluons

Boosted objects (W, Z, H bosons, top quark)

Jet reco and ID performances

Noise Jet ID

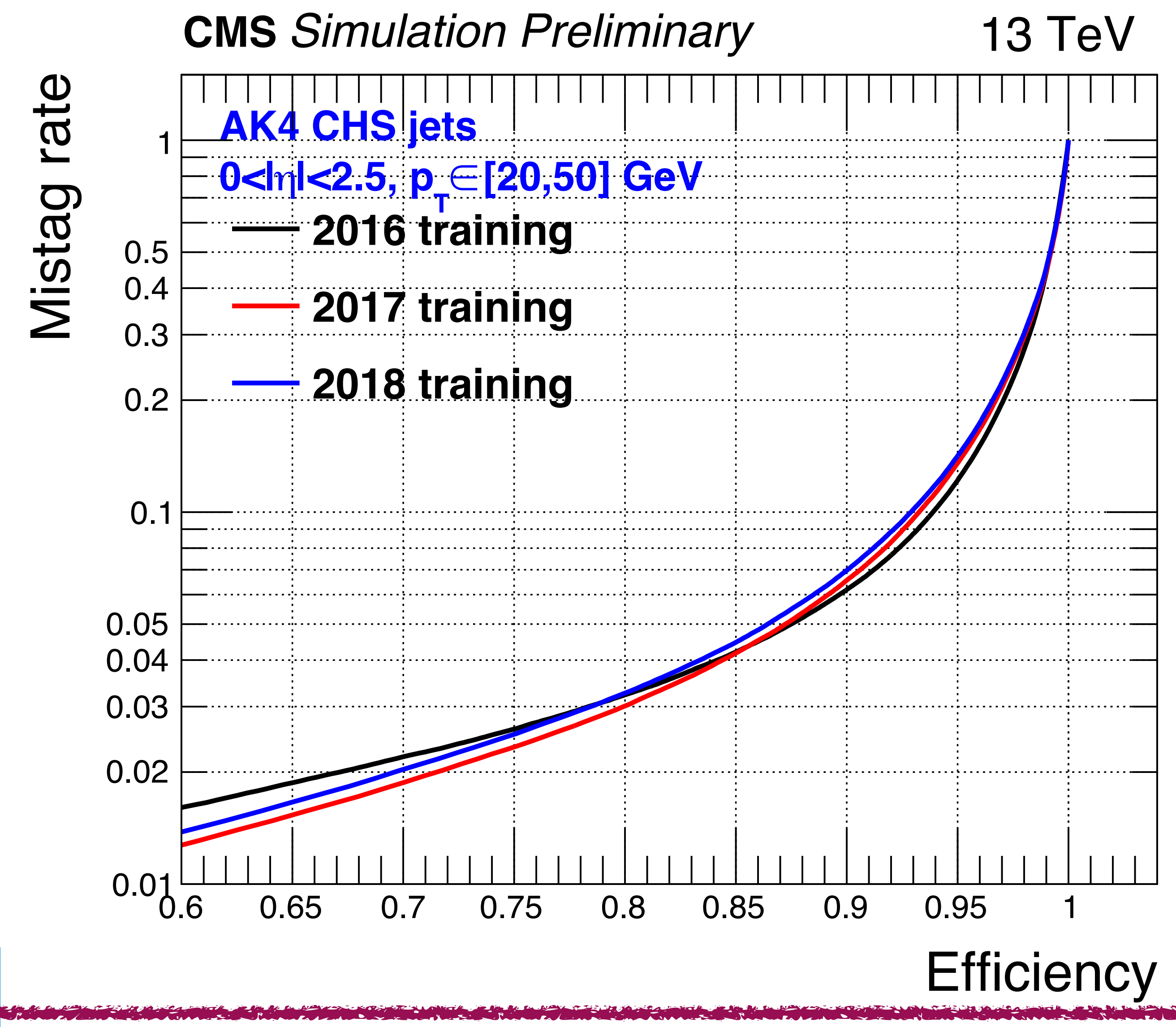
Discriminator for noise VS physical jets

- Exploits jet energy fraction and multiplicity variables,
  - sensitive to different sources of noise coming from ECAL and HCAL.
- Three WPs defined: loose, tight and tight lepton veto.
  - In this analysis: tight lepton veto
  - jets from calorimetric noise are removed
  - Ensure the rejection of mis-reconstruction of lepton candidates.
- The efficiency and the background rejection:
  - **Keeping 98 ÷ 99.9% of genuine jets,**
  - **Background rejection is above ≈ 98% for $|\eta| < 3$**

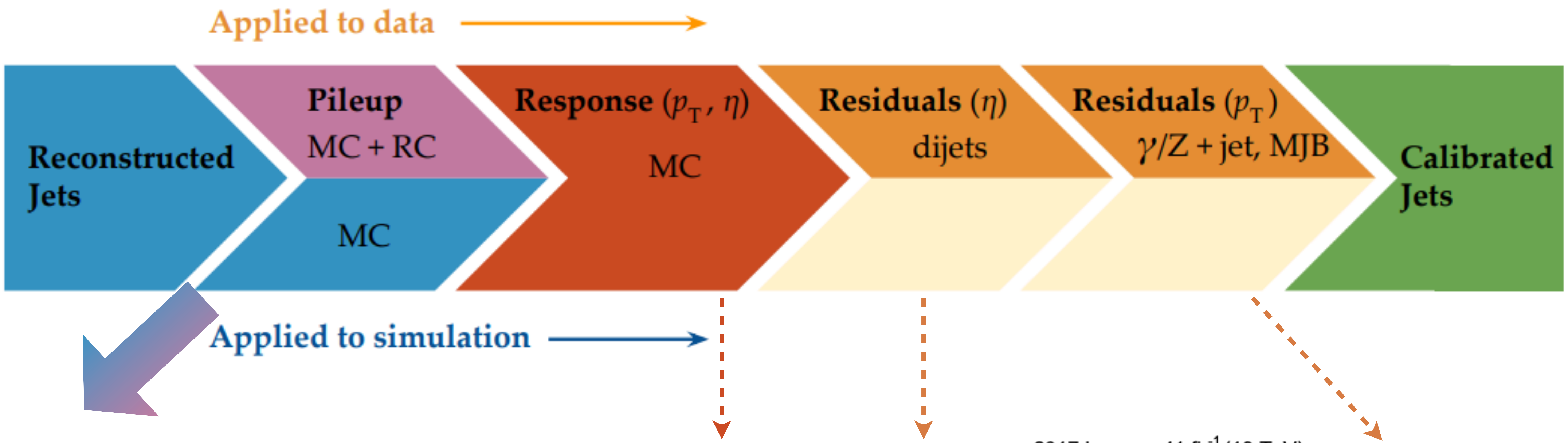
PUJetID

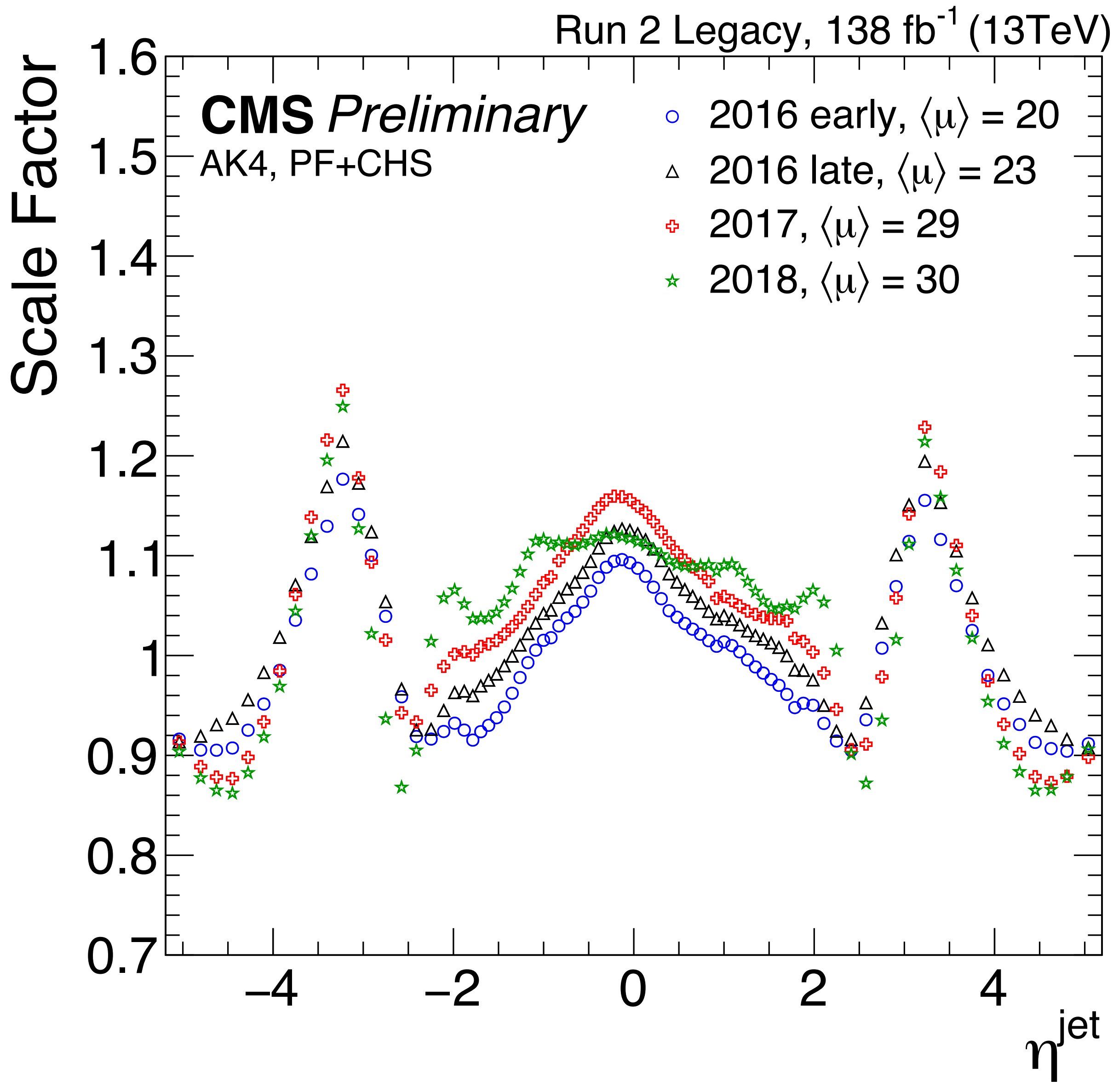
PUJets:

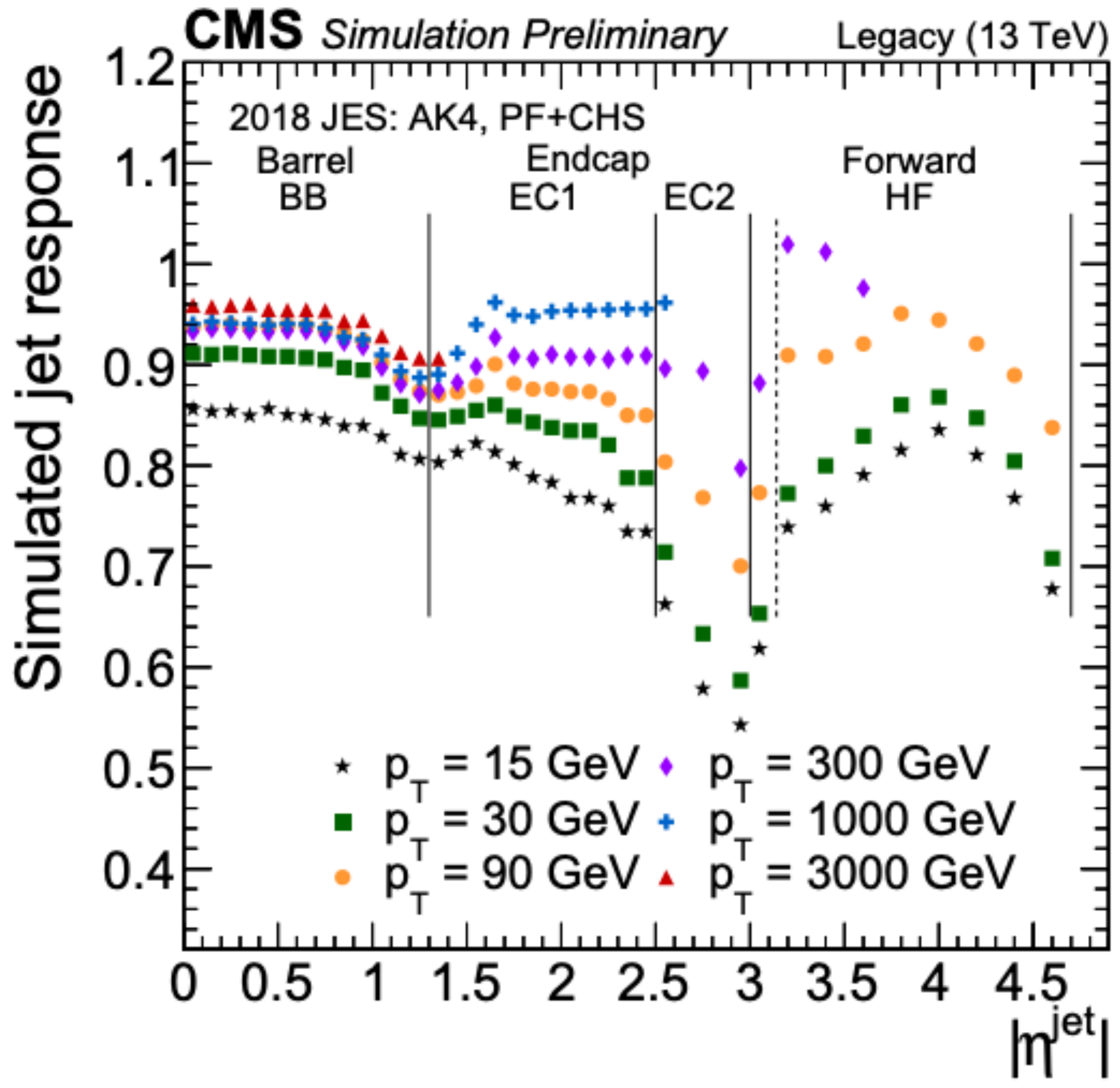
- QCD jet-like from one PU vertex: rejected from PU mitigation techniques
- Stochastic PU jets from multiple different PU vertices: removed from PUJetID
  - BDT trained on jet vars
  - Applied to CHS only
  - Three WPs defined: loose, medium, tight
  - In this analysis: loose (only for Jet $p_T < 50$ GeV)

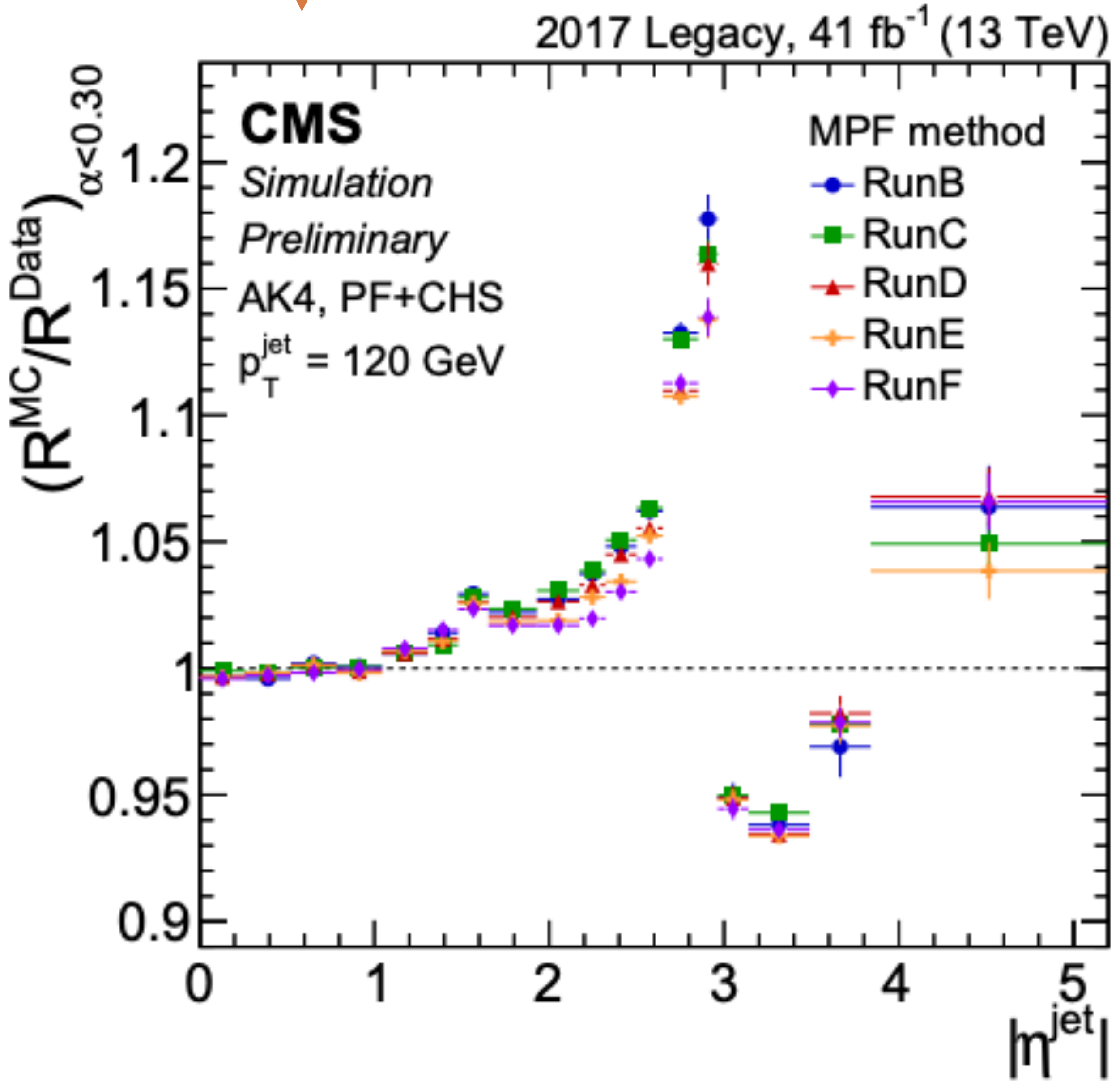


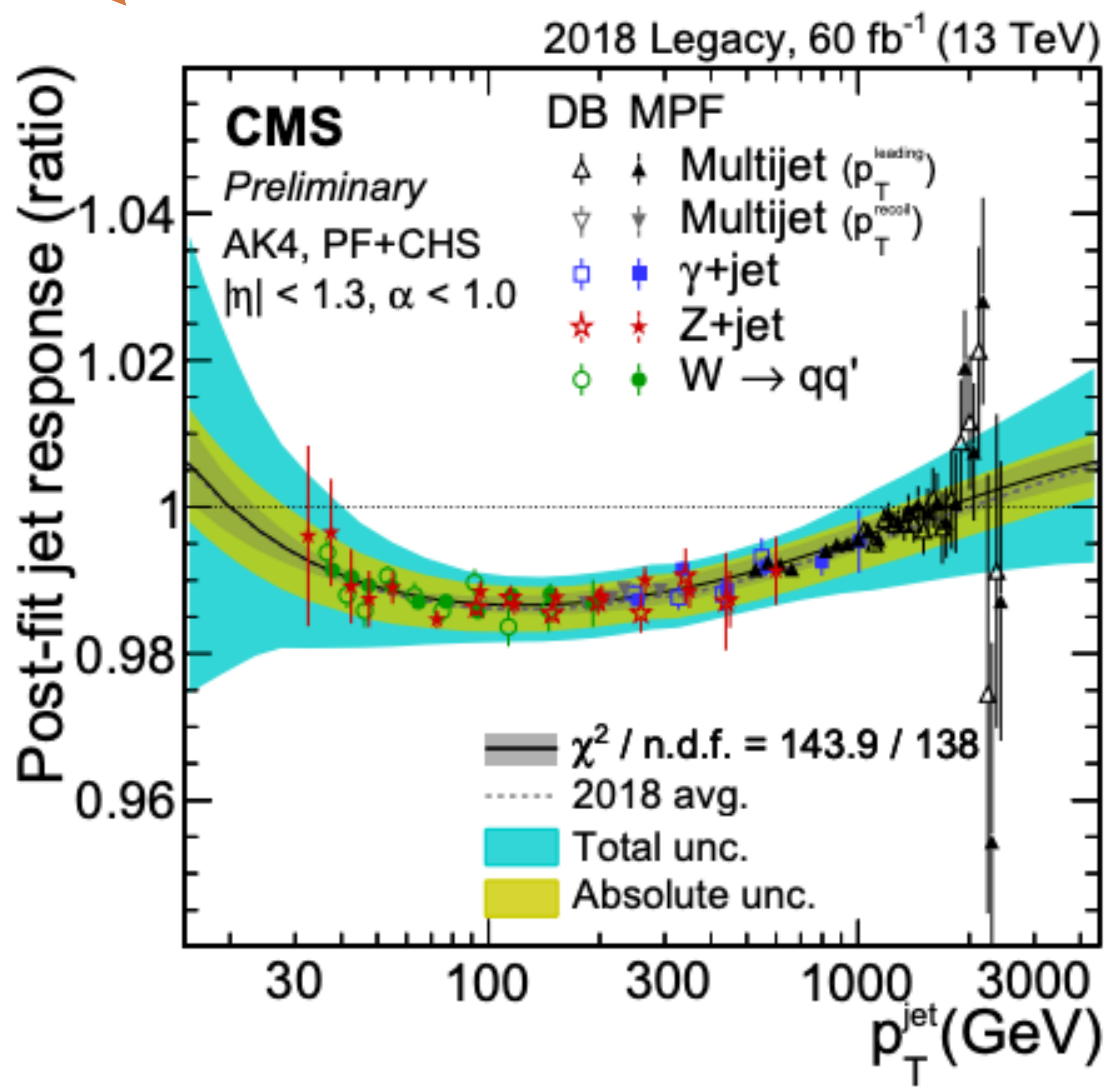
Jet Calibration







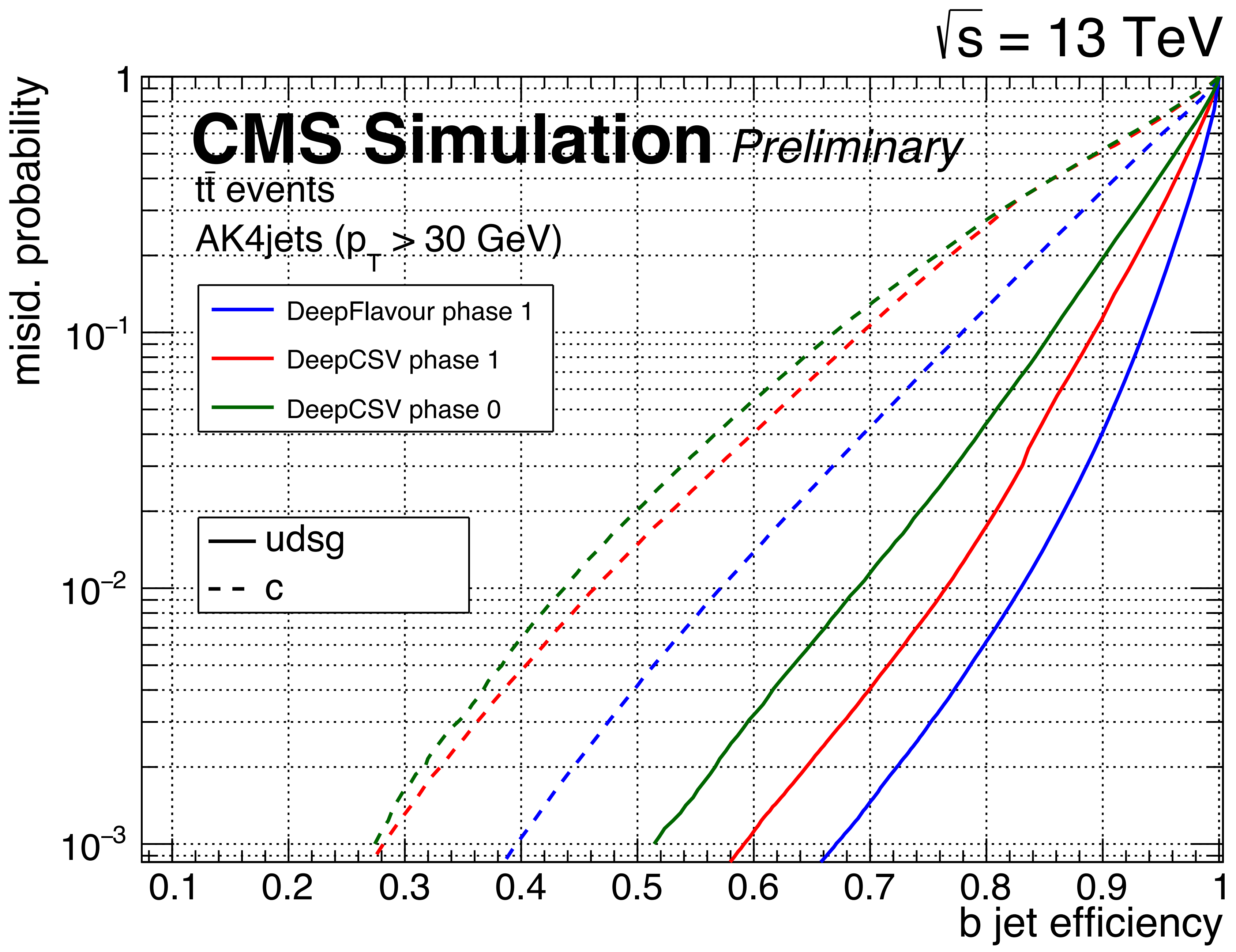


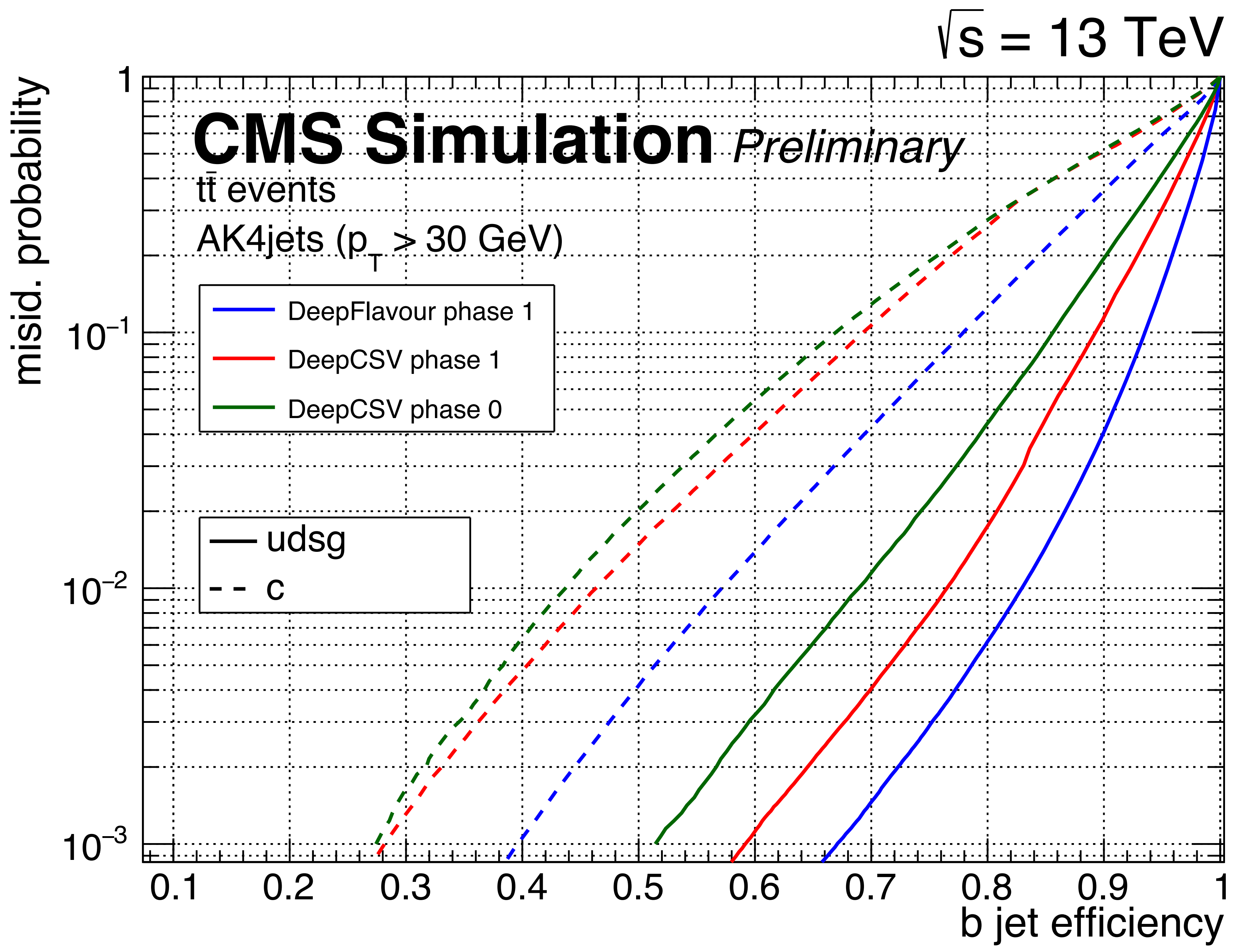


AK4 jet b-tagging performance

DeepFlavour:

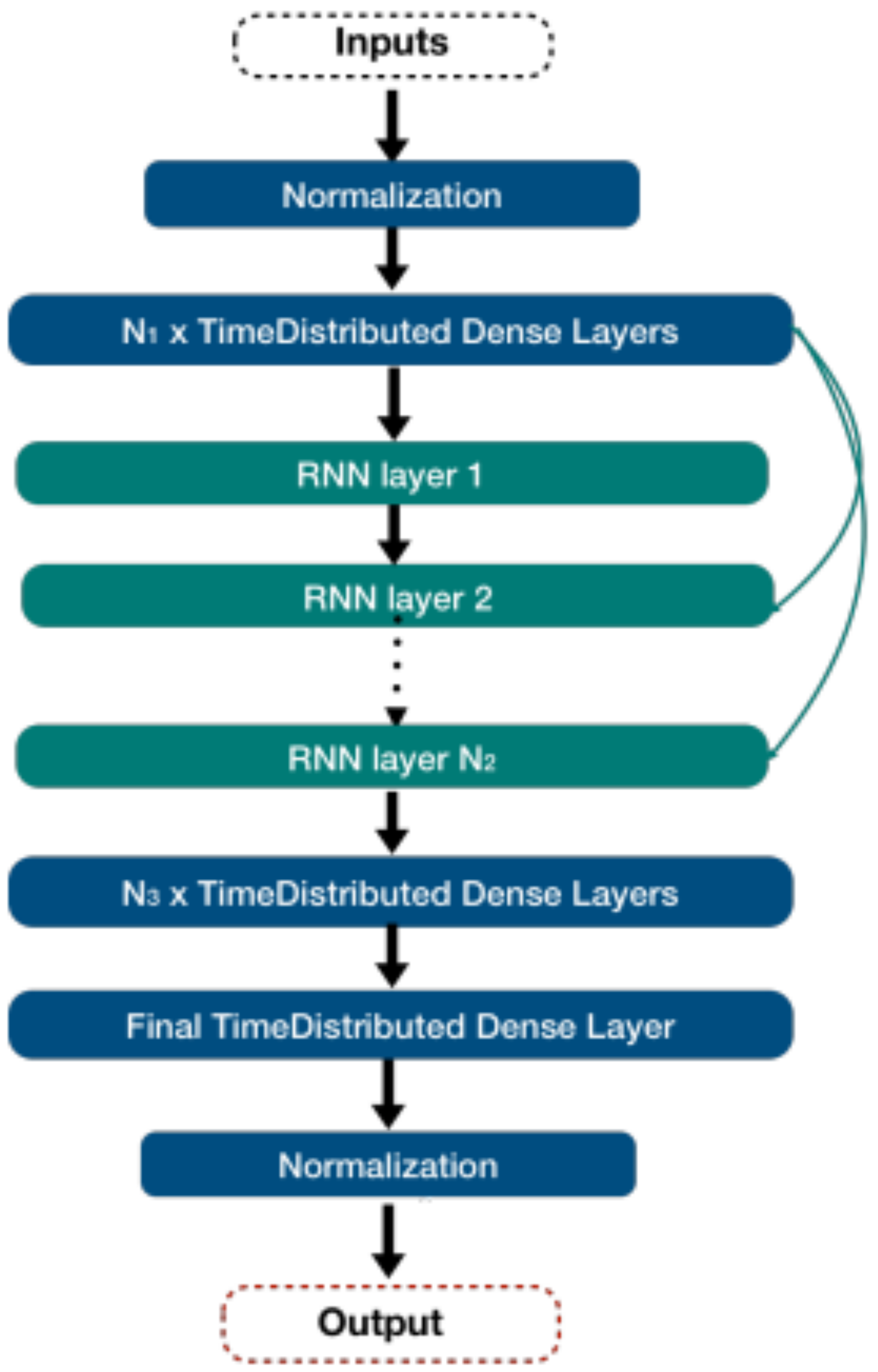
- Multi-class flavour DNN-based tagging algorithm
- 16 (6) features of the 25 leading charged (neutral) PF Jet constituents + 12 features of the vertices + 6 global variables containing jet-object information such
- Training on simulated events from QCD and top quark pair production
- Output: 6 nodes for b-tagging, c-tagging and quark/gluon tagging
- Three WPs are defined: loose, medium and tight, corresponding to misidentification rates of 10%, 1% and 0.1%





The HH-bTag tool

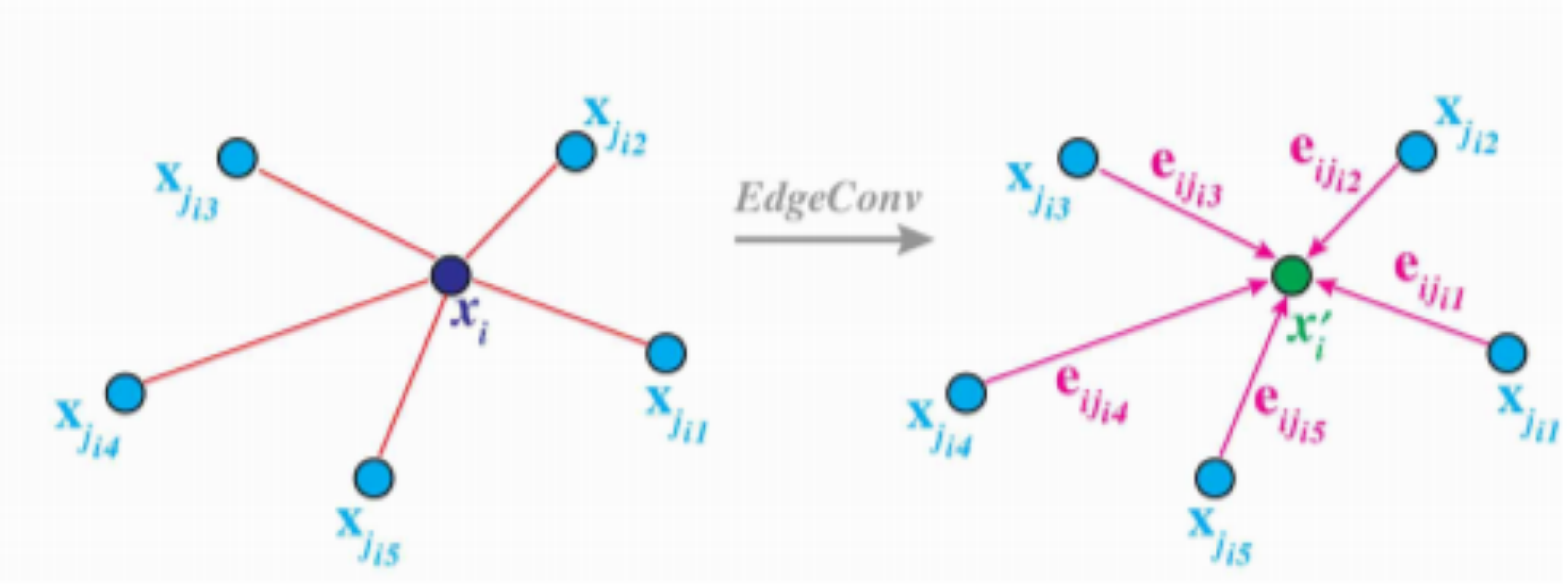
- Machine-learning based tool aiming to identify the b-jet pair from $H \rightarrow bb$ decay for the $HH \rightarrow bb\tau\tau$ analysis
- Specially trained DNN (HH-btag), documented at AN-19-283
- Features - NN inputs :
  - Jets: deepFlavour, p_T, η, M/p_T, E/p_T
  - $H(\tau\tau)$ candidate: p_T, η, $\Delta\varphi(MET, H)$, $p_T(MET)/p_T(H)$
  - $\Delta\varphi(jets, H)$, $\Delta\eta$ (jets, H)
  - Year, channels
- Generator level information is used as the ground truth
- Recurrent NN assigns a score to each jet in the event
- Two trainings with even/odd samples + validation on 25% of entire dataset
- Bayesian optimization to chose best NN hyperparameters
- Trained during non-resonant analysis and re-trained (v2) for this analysis

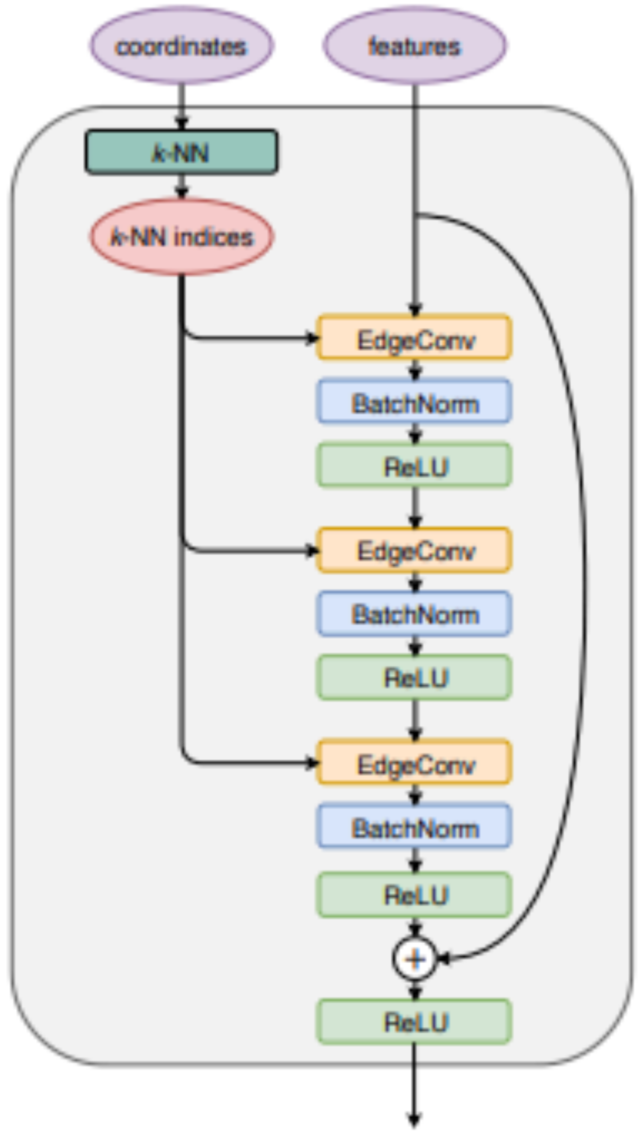


AK8 Jet tagging: ParticleNet

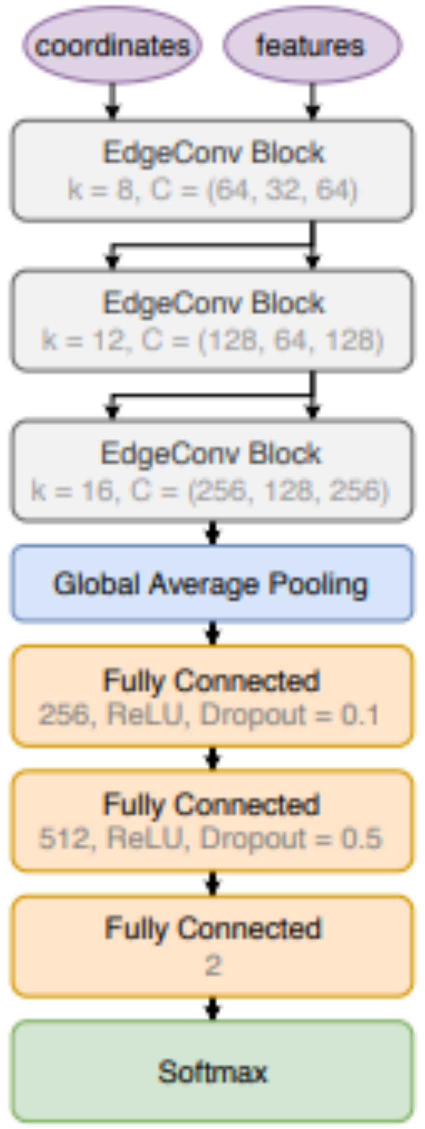
- ParticleNet-MD state-of-art for CMS boosted jet tagging.
  - Graph based architecture describing the jet as a particle cloud (unordered sample).
- EdgeConv block:
  - NN module part of the ParticleNet architecture;
  - New features vector associated to each jet constituent and based on the features of the k-nearest neighbors.
- Mass decorrelation:
  - Trained on Monte Carlo (MC) simulations containing boosted resonances (X) with a flat distributions in both of p_T and mass, as the signal sample, and the QCD multijet sample (reweighted to yield flat distributions) as the background sample.

H. Qu, L. Gouskos, PRD 101, 056019 (2020)





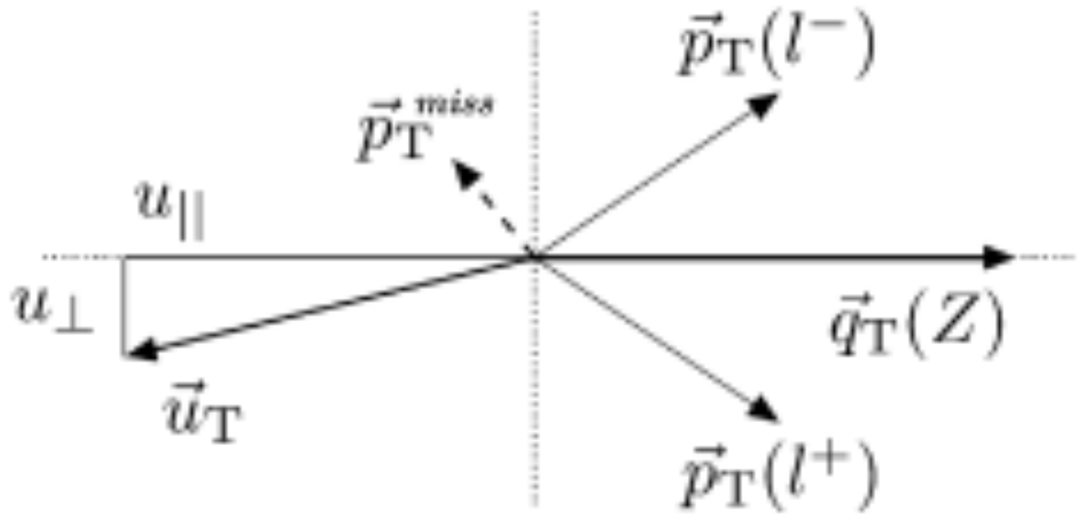
EdgeConv block
(inspired by ResNet)



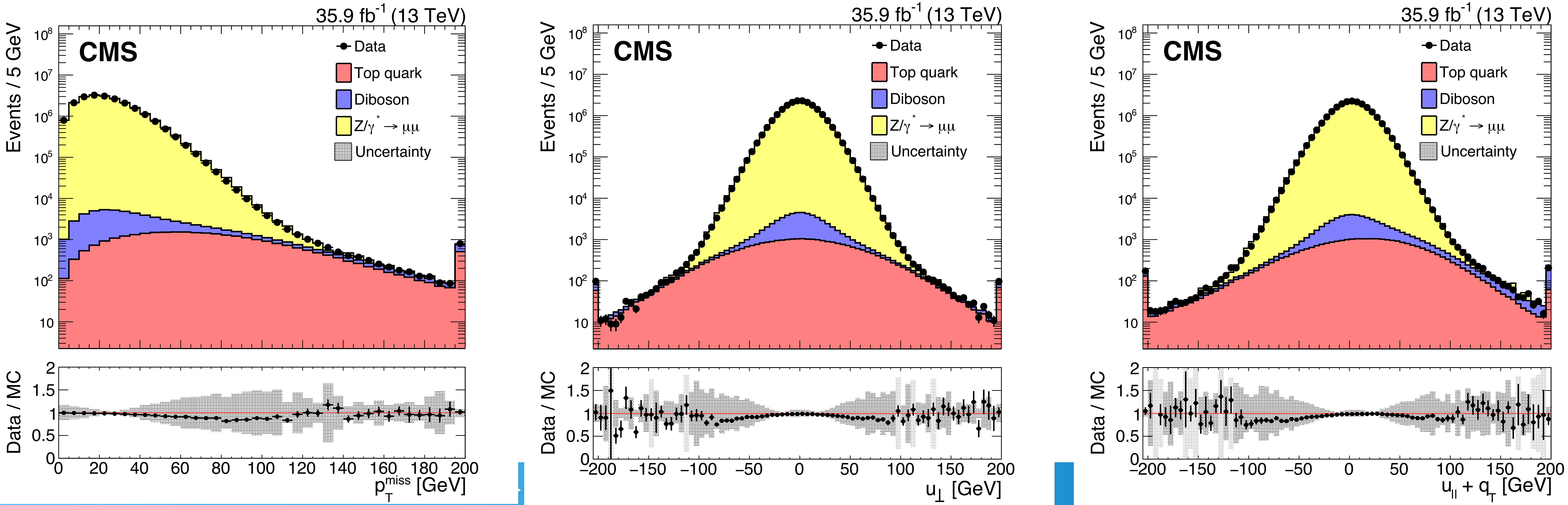
ParticleNet architecture

MET reco and ID performances

Measured in events with Z/γ + jets, $Z \to \mu\mu\,(ee)$



- MET resolution dominated by the hadronic activity (5-20%)
- Hadronic recoil system is defined as the vector $p_T$sum of all PF candidates except for the muons fro
  - Transverse: u_T
  - Parallel $u_{||}$
- MET resolution estimated by $|\vec{q}_T(Z)| - u_{||}$
- The intrinsic resolution, after removing the PU contribution, is 10 GeV for both the components of the recoil.



b-tag weights and scale factors

The reweighing method for calibration

- From the twiki: b-tag SF methods can be grouped into two general categories: methods that involve event reweighting and those that do not. Which method to apply depends on the analysis, but BTV Recommended methods are the following:
  - Event reweighting using scale factors and MC b-tagging efficiencies when only working with b-tagging working points
  - Event reweighting using discriminator-dependent scale factors when working with the whole b-tagging discriminator shape
- This analysis uses the working points AND discriminator shapes so we did studies about which method to use and our conclusion brought better data/MC agreement for the second method.

$$\mathrm{SF} = \frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{MC}}}$$

b-tag shape calibration

- The b-tag related weight ω takes in account **all the selected jets**: $\omega_{\text{event}} = \prod_{i}^{N_{\text{jets}}} SF(D_i, p_{Ti}, \eta_i),$
- Expected event yields should be preserved: **the number of events (i.e. the sum of event weights) before and after applying b-tag weights should be identical**.
  - One should measure the sum of event weights before and after applying b-tag event weights, without requiring any b-tag selection in both cases.
  - The ratio $r = \sum \omega_{\text{before}} / \sum \omega_{\text{after}}$ represents a phase space extrapolation and should be multiplied to the b-tag event weight.
  - **This extrapolation could in general depend on further variables, most notably the jet multiplicity**. But in this analysis we made studies and show that we don't need to slice histograms per jet multiplicity

Signal/Control Regions

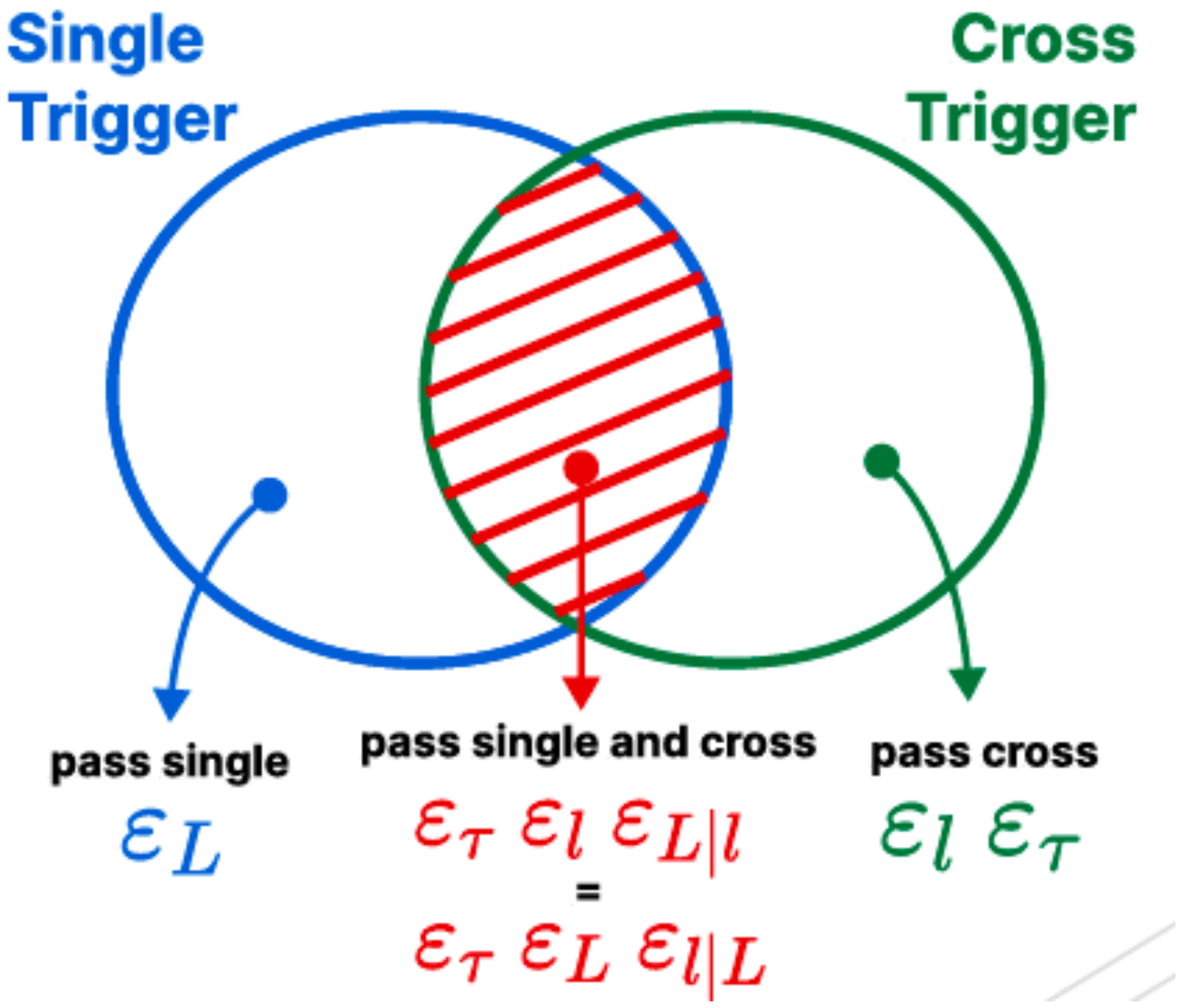
- **Signal-enriched region (SR)** where performing statistical inference, square cut on $m_{\ell\ell}$ and $m_{H(bb)}$
  - eτ, μτ, ττ channels, including 98% of signals (all masses)
- **Control Regions (CRs)**: verify data/MC agreement, especially for tt and DY in dedicated signal-depleted and background-enriched regions, estimate trigger and ParticleNet SFs (see backup), defined by applying selection on $m_{\ell\ell}$
  - **Drell-Yan control region (DYCR)**: multiple MC sets, combined with stitching technique
    - ee, μμ channels, including 68% of DY samples
  - **TT control region (TTCR)**: Observed p_T spectra of top quarks in tt data were found to be significantly softer than predictions from MC simulations
    - ee, eμ, μμ channels, excluding 90% of DY samples
    - eτ, μτ, ττ channels, exluding 90% of DY samples and 95% of signals (all masses)

Triggers Efficiencies

Cross+single lepton

- Trigger efficiency in phase-space region where more than one trigger are active

$$\epsilon_{(xtrg)} = \epsilon_L + \epsilon_\ell \epsilon_\tau + \epsilon_\ell \epsilon_\tau \epsilon_{L|\ell}$$



Yes, but.. we don't have $\epsilon_{L|\ell}$ but we can replace the last term with

$$\text{SF} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}$$

$$\epsilon_{(xtrg)} = \epsilon_L + \epsilon_\ell \epsilon_\tau - \min(\epsilon_L, \epsilon_\ell)\epsilon_\tau$$

This accounts cases where single lep trigger is fired and the cross lep is not and vice-versa

Triggers Efficiencies

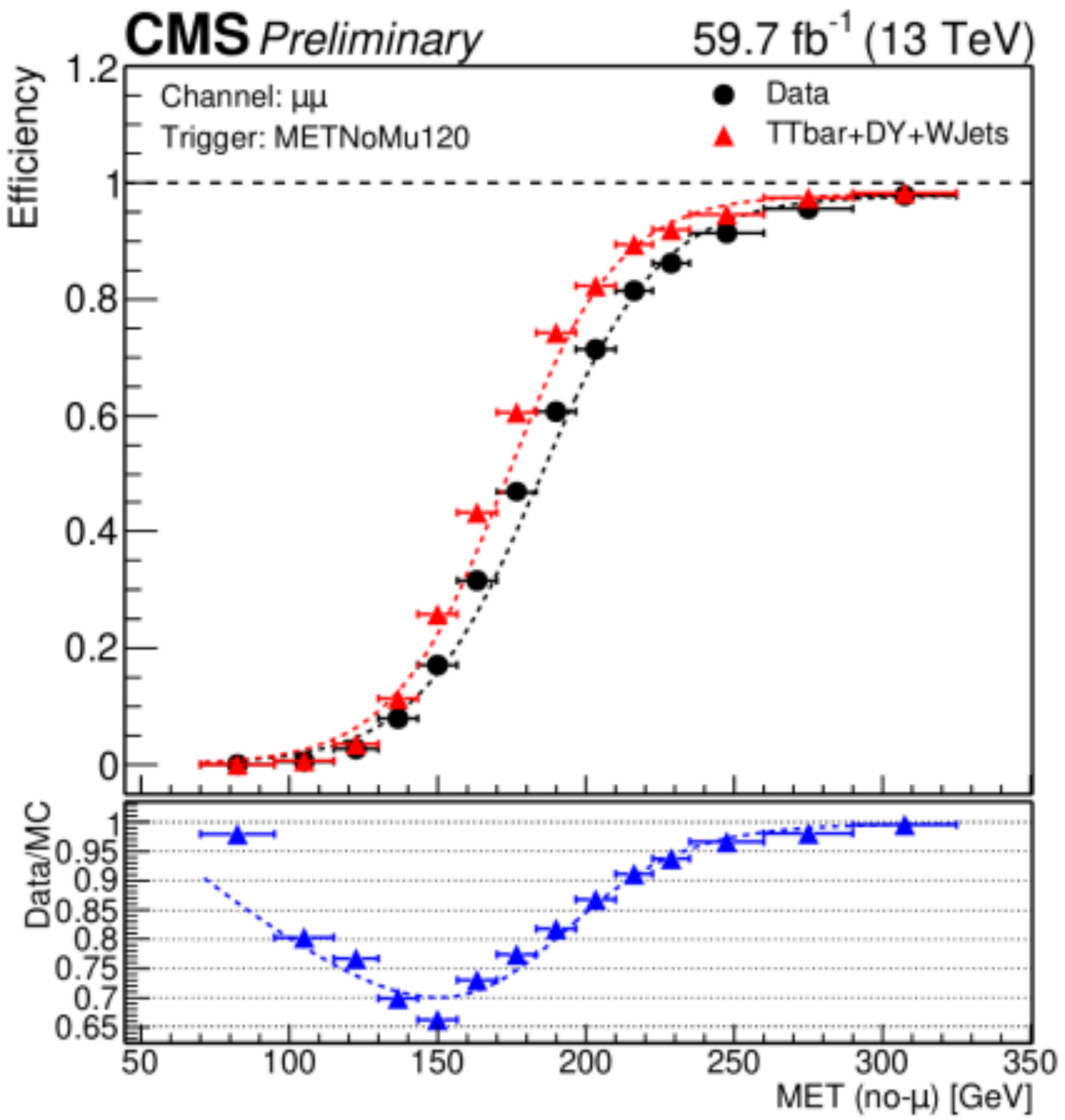
MET trigger

$$\epsilon = \frac{\text{Events}_{\text{passing baseline selection \&\& passing MET (no}-\mu\text{) trigger selection}}}{\text{Events}_{\text{passing baseline selection}}}$$

- Orthogonal dataset: μμ final state channel
  - Not used in main analysis (only for CR checks)
  - Data: Muon dataset → orthogonal to Met-no-μ trigger
  - MC: ttbar, DY and W+Jets
  - Events with two b-jets passing Loose deepFlavour WP
  - No mass cut
- Estimated as function of MET-pT(μ)
  - Fit with sigmoid function above the turn-on
- Validation in μτh channel
- Uncertainties:
  - Efficiencies: Clopper-Pearson interval
  - Ratio: error propagation of fit of the two sigmoids
    - Used for the analysis to get up/down variations

$$SF = \frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}$$ → ratio of the two sigmoids

$$f(a,b,c) = \frac{c}{1+e^{-a(x-b)}}$$

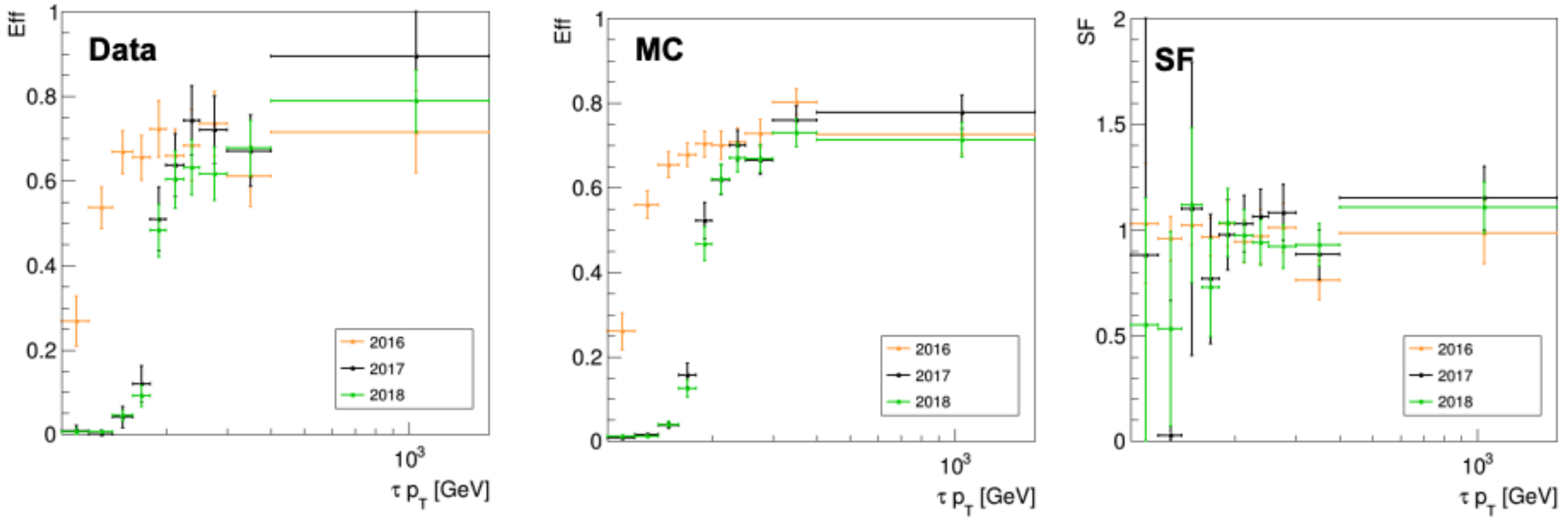


Triggers Efficiencies

SingleTau trigger

- Measurement performed in region enriched in W*(mW* > 200 GeV) → τν with little hadronic activity
  - high-pT tau balanced with MET
- Determination of tau trigger efficiency SFs:
  - follows closeley determination of tau id efficiency SFs for high-pT taus
  - define samples of passing and failing probes
    - passing probe: tau matches filter of any of the two single-tau triggers
    - failing probes: otherwise
- measurement performed differentially in tau pT
- perform simultaneous ML fit of samples of passing and failing probes to improve modeling of shapes
- tau Id SFs & normalization SF for W* sample not applied, cancel in the determination of the efficiencies
- apply TES correction & uncertainty from low-pT tau measurement, uncertainty inflated by factor 3

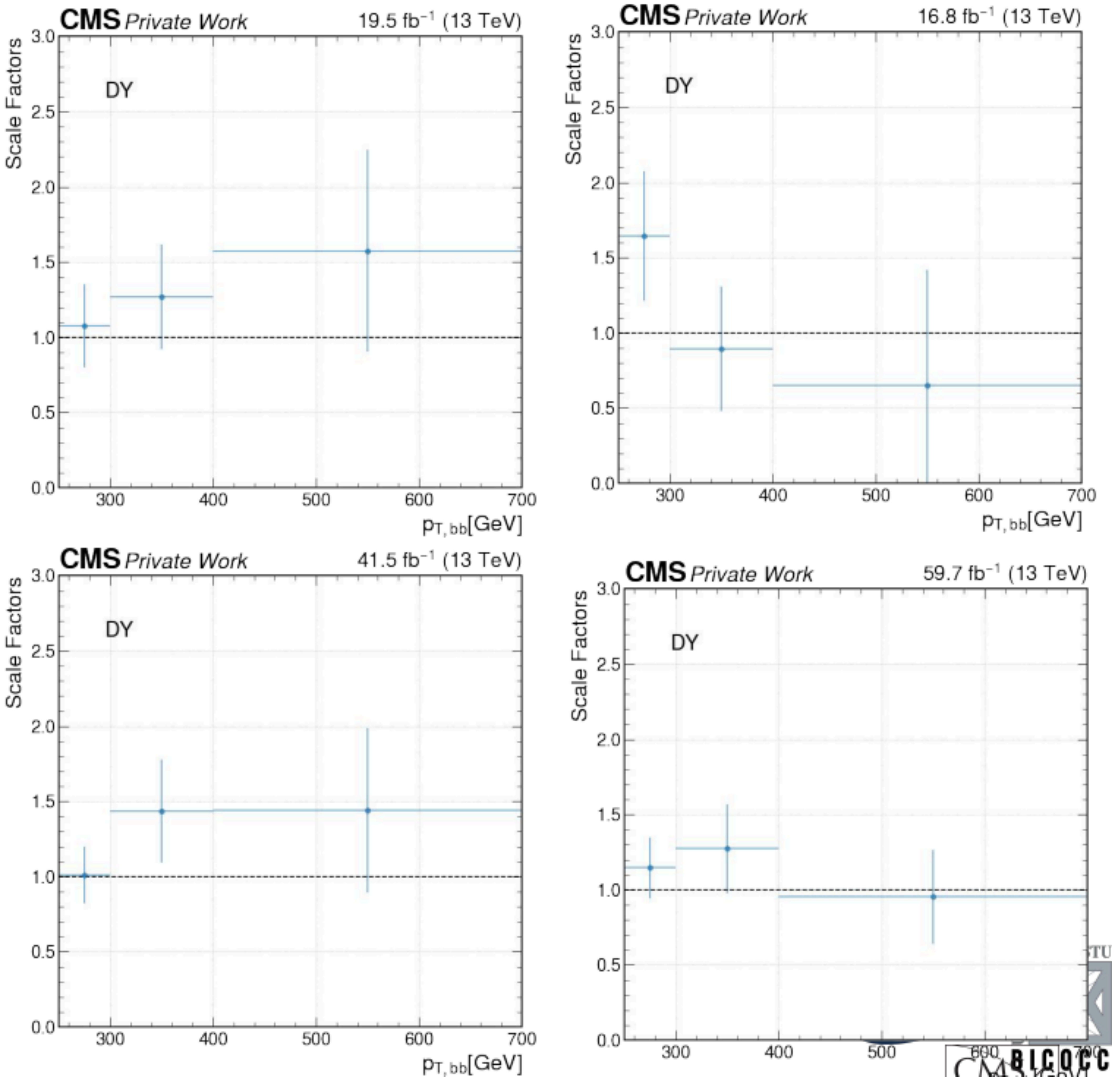
$$\epsilon = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}}$$



ParticleNet Scale Factors

DY- and TT- like events

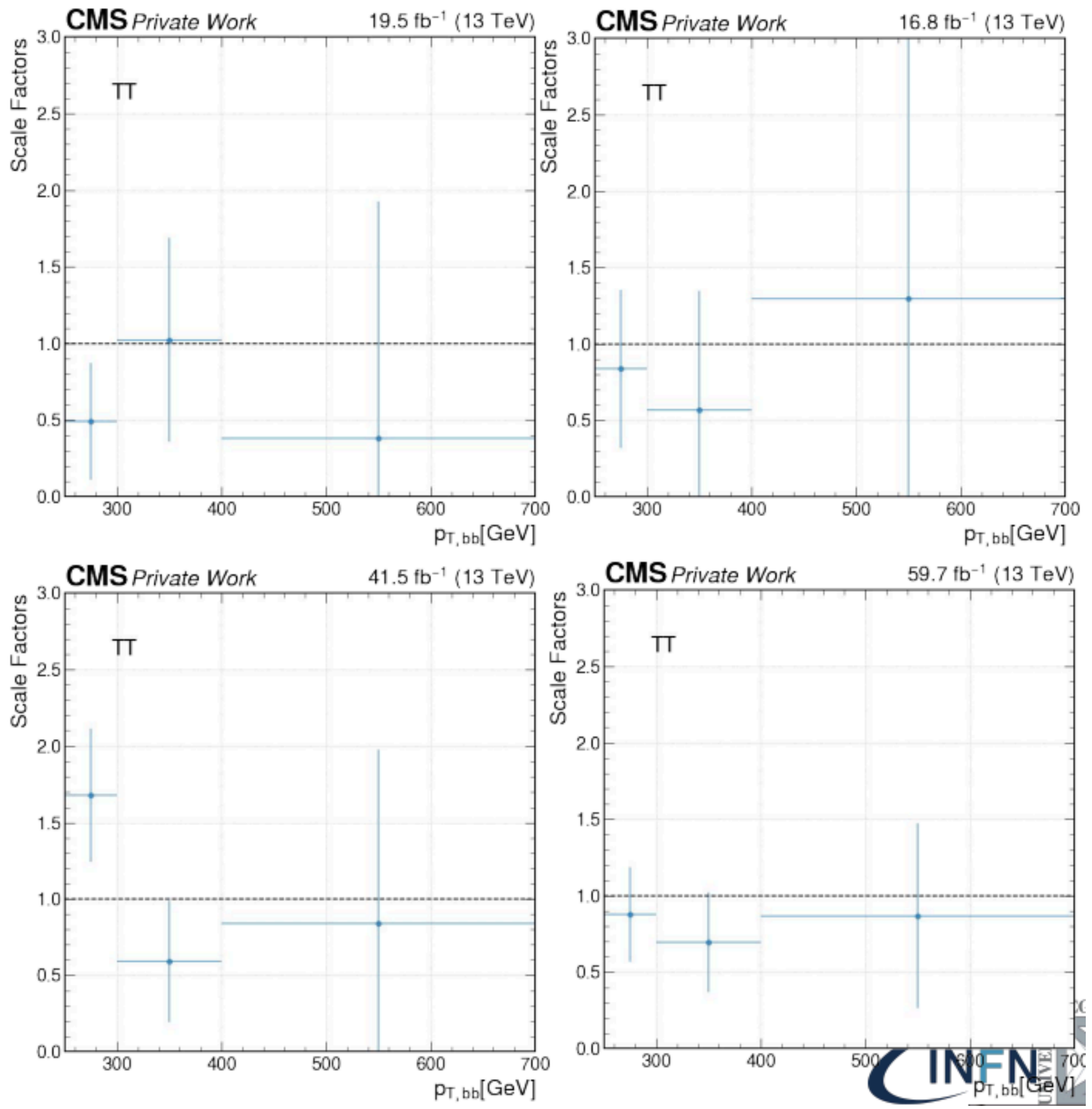
- Corrections provided by the BTV POG for signal-like jets, (H->bb, Z->bb)
- For background originated jet: custom SF provided in this analysis (latest update from S. Palluotto et al)
- Two classes of SF dedicated to the most relevant irriducible backgrounds (DY, TT)
  - DY-like events from DYCR in the boosted category
  - TT-like in TTCR, signal channels and boosted category
- Efficiency definition
- Images taken from Simona Palluotto's presentation



ParticleNet Scale Factors

DY- and TT- like events

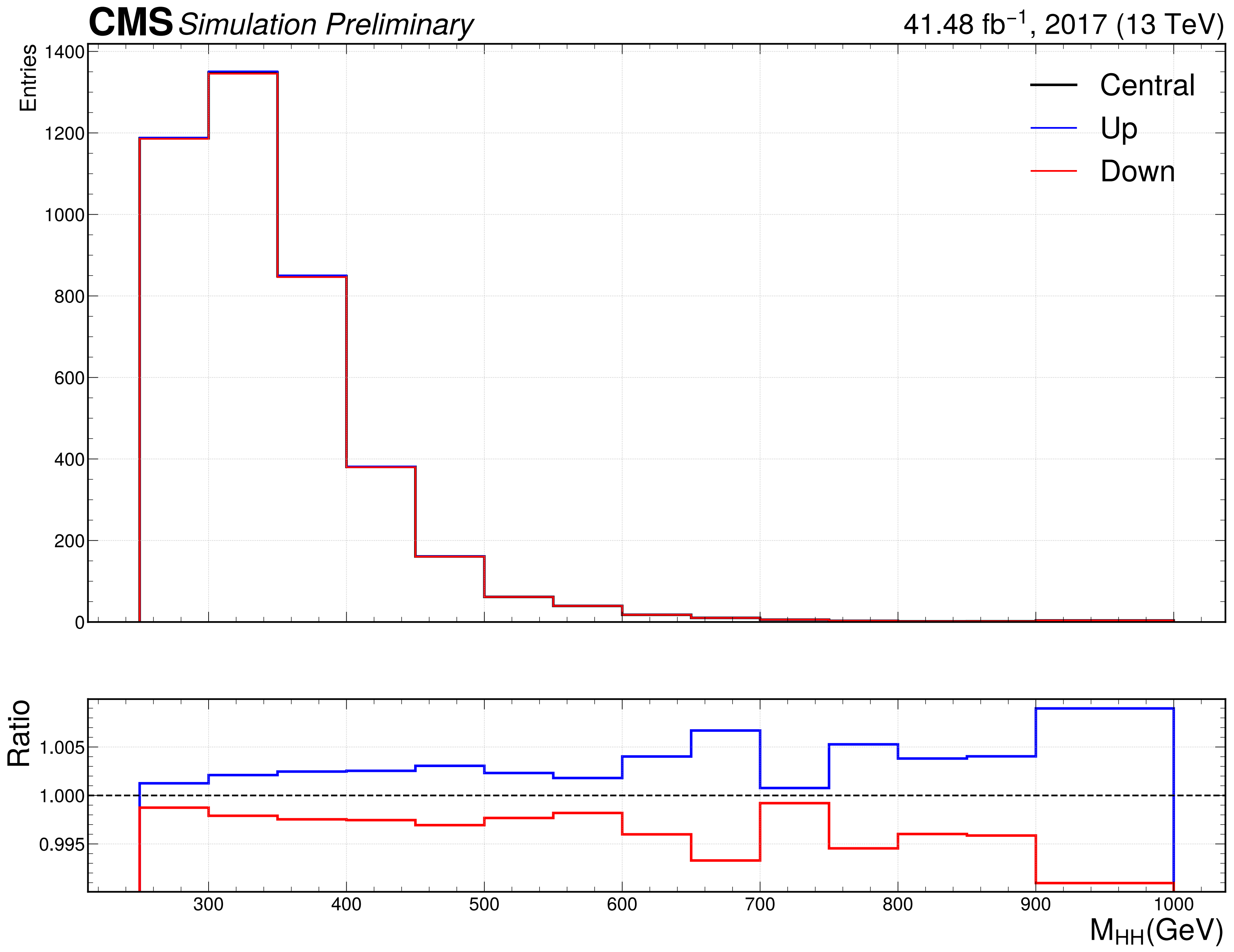
- Corrections provided by the BTV POG for signal-like jets, (H->bb, Z->bb)
- For background originated jet: custom SF provided in this analysis (latest update from S. Palluotto et al)
- Two classes of SF dedicated to the most relevant irriducible backgrounds (DY, TT)
  - DY-like events from DYCR in the boosted category
  - TT-like in TTCR, signal channels and boosted category
- Efficiency definition
- Images taken from Simona Palluotto's presentation



Normalisation uncertainties

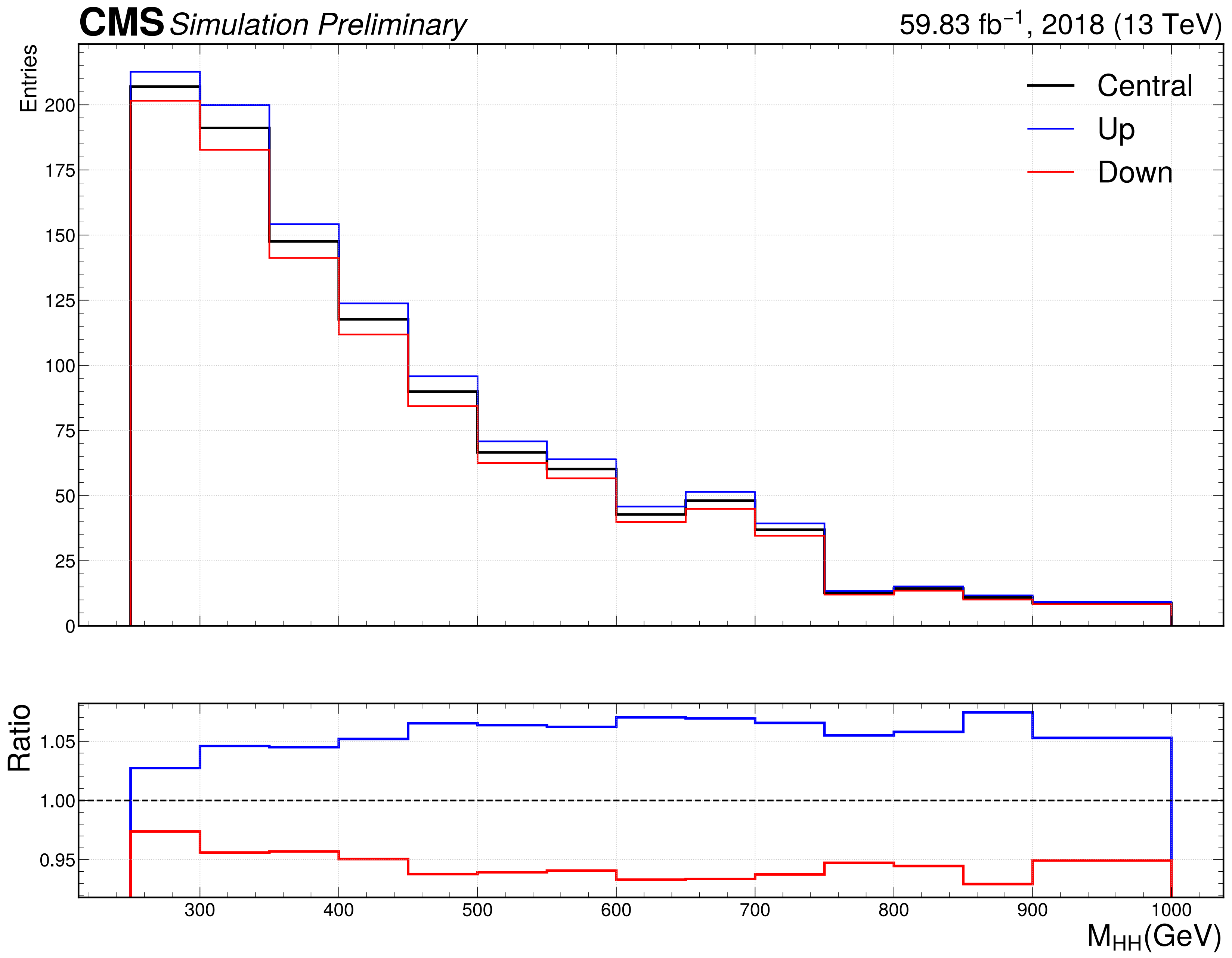
Examples

ele trg SF (res2b, $e\tau_h$ channel)



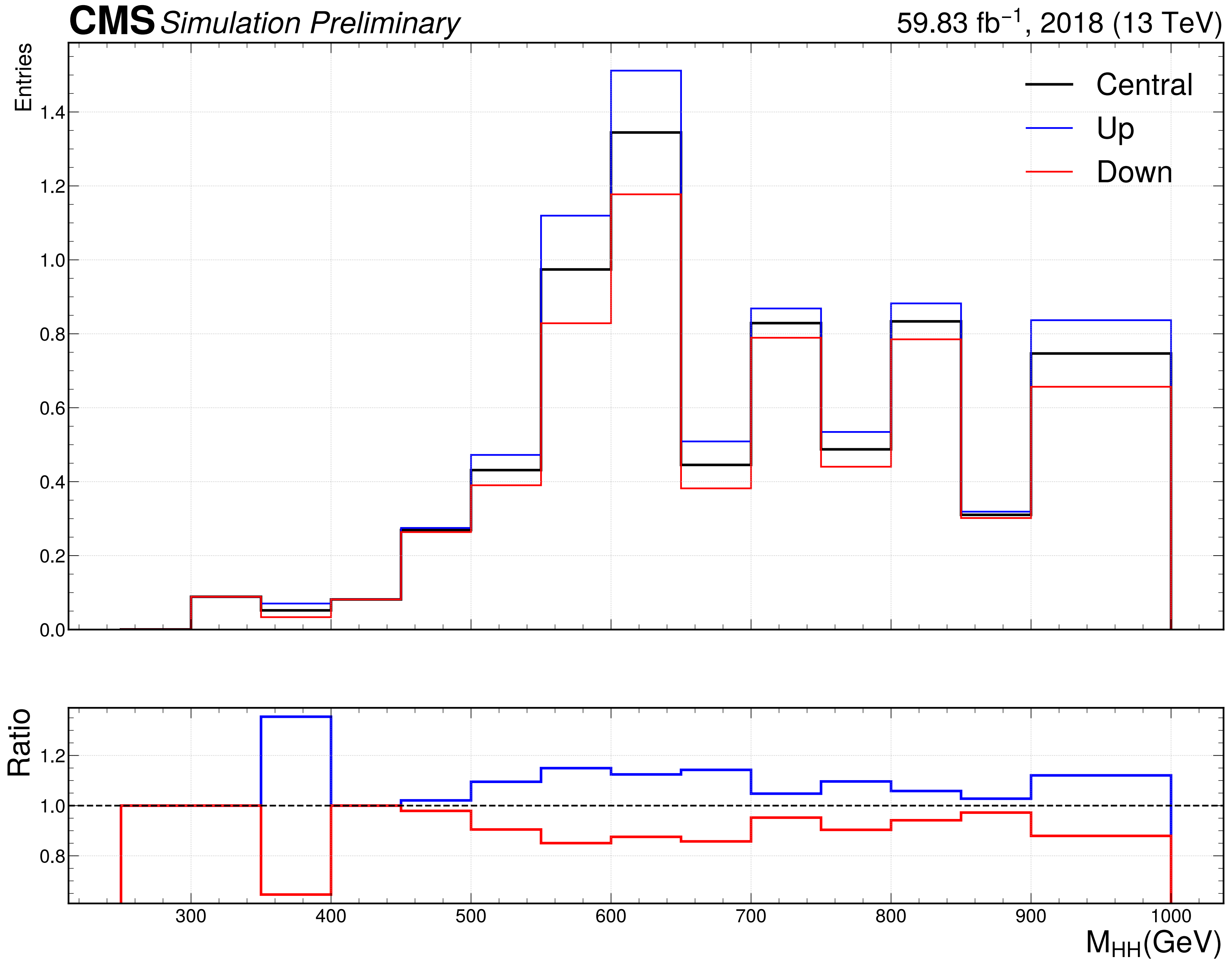
combination of all triggers applied to electrons

τ_h trg SF (res1b, $\tau_h\tau_h$ channel)



combination of all triggers applied to taus

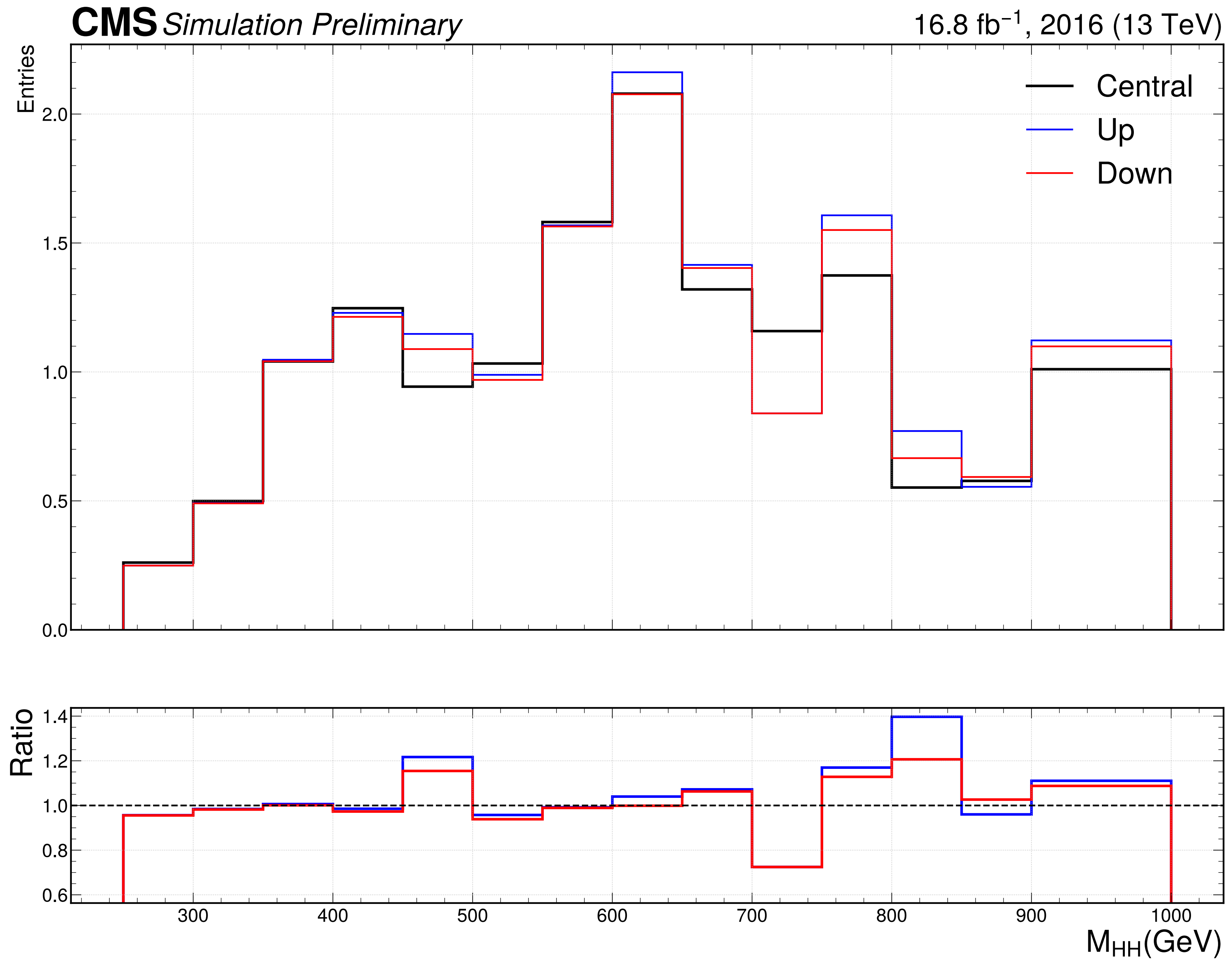
pNet SF (boosted, $\tau_h\tau_h$ channel)



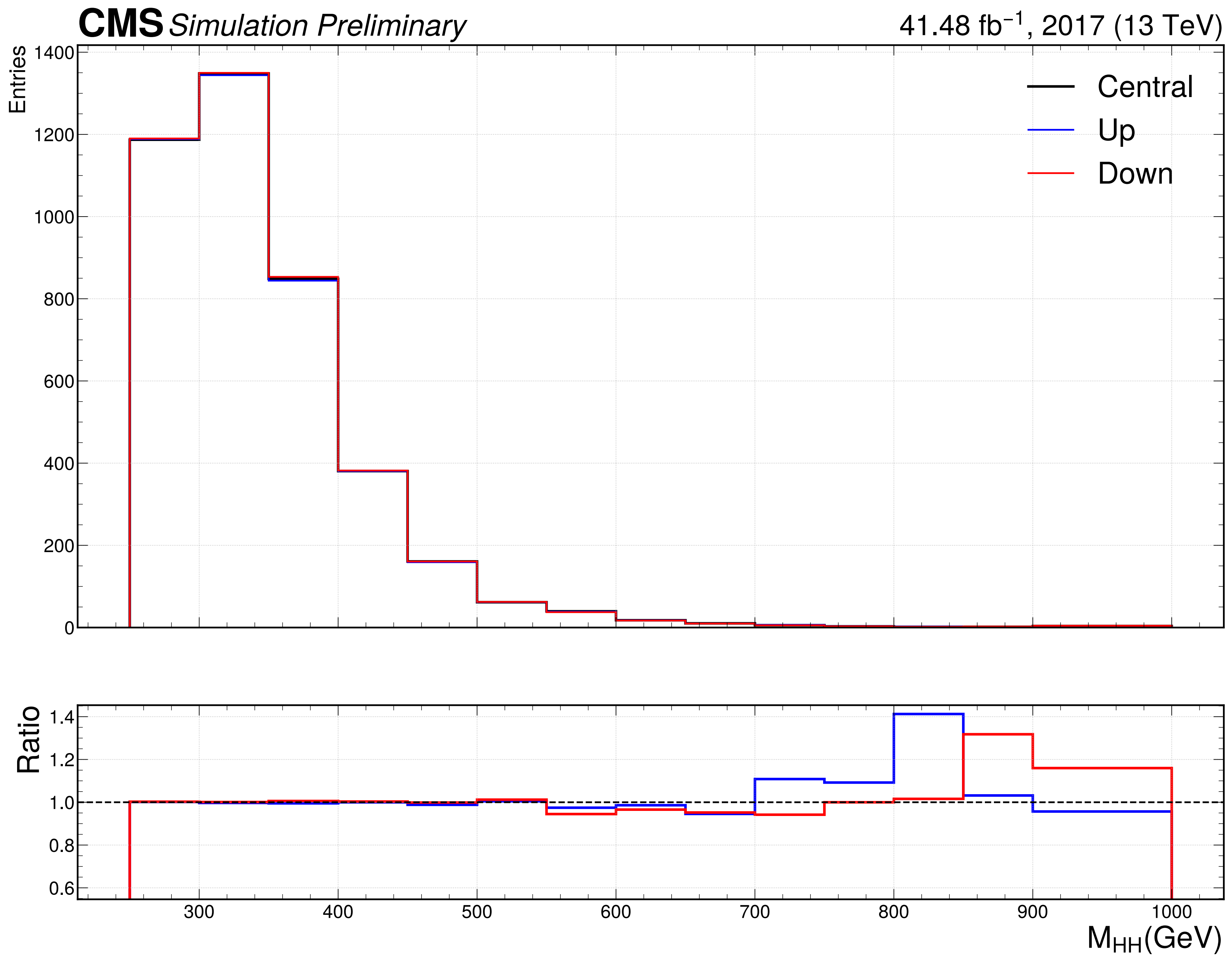
Shape uncertainties

Examples

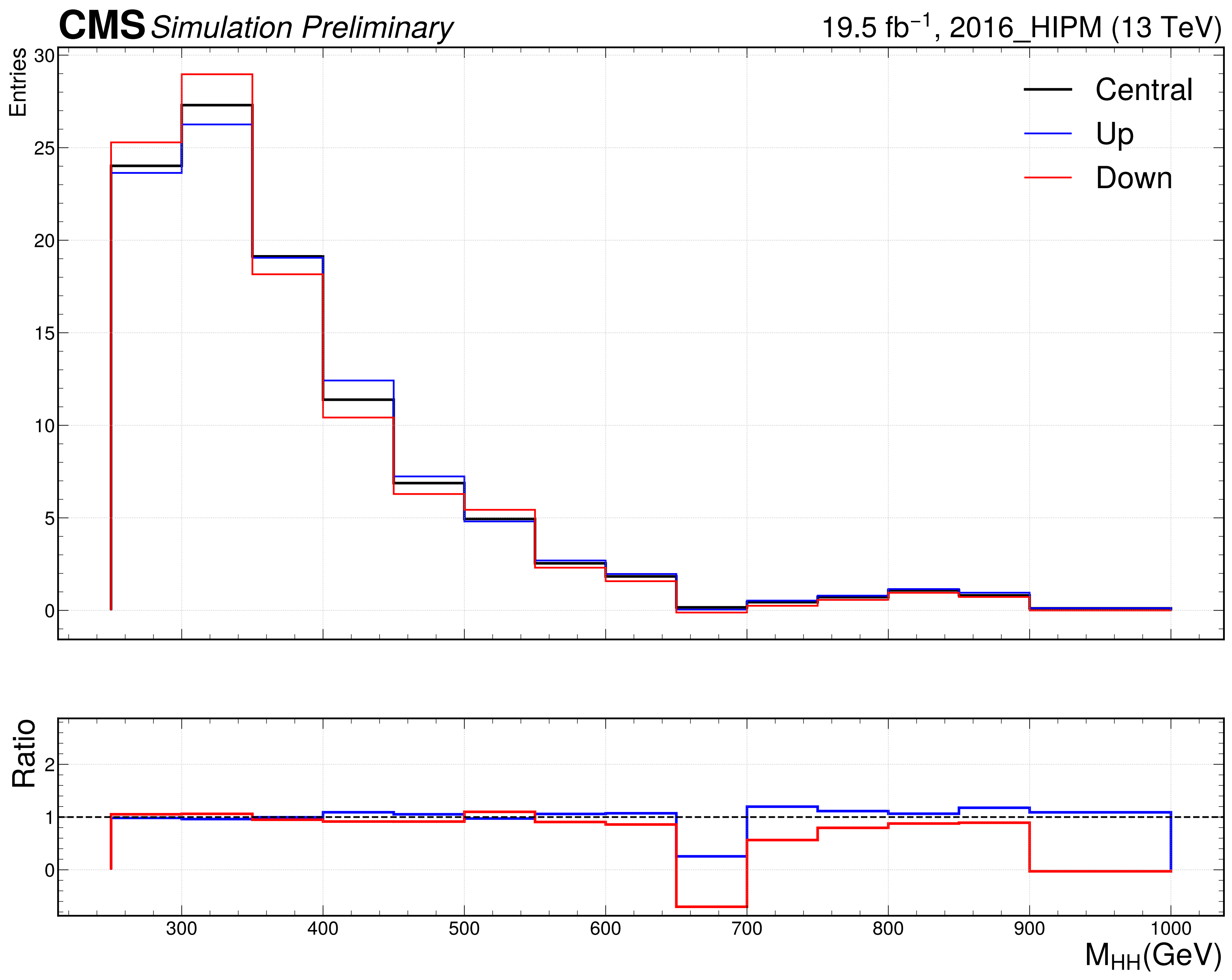
TES-DM1 SF (boosted, μτh channel)



JER (res2b, eτh channel)



QCD scale (res1b, τhτh channel)



Limited size of MC simulations

Method proposed by Barlow and Beeston [1]

- Introduction of a nuisance parameter (NP) multiplying the expected yield in each bin from each MC simulated sample
- The nominal value of such parameters is 1 and they are left floating with some prior distribution (e.g. Pois, Gaus, Binom)
- Introduction of massive number of nuisances: set of non lin equations in $-ln(\mathcal{L})$ (NLL) minimisation
- Practical purpose: ROOT minimiser (MINUIT MINGRAD) has technical problems in finding the numerical approximation of the values that minimise the NLL, so they are factorised in only 1 NP for each bin

Contribution to NLL in each bin

$$-ln(\mathscr{L}(\mu,\beta)) = -\, n_{obs} ln(\beta \cdot (\mu s + b)) + \beta \cdot (\mu s + b) + \frac{(\beta - 1)^2}{2 \cdot \sigma_\beta^2}$$

When minimising NLL (with other NP fixed)

$$\frac{\partial(-ln(\mathscr{L}))}{\partial \beta} = 0 \implies \beta^2 + ((\mu s + b) \cdot \sigma_\beta^2 - 1) \cdot \beta - n_{obs}\sigma_\beta^2 = 0$$

[1] R. Barlow and C. Beeston, Comp. Phys. Comm. 77 (1993) 219.

Statistical treatment

Higgs analyses statistical treatment explained in [2]

1. Construct the likelihood function
   - Binned distribution: product of Nbins Poissonian with signal and background yields
   - Systematic uncertainties (assumed uncorrelated), included as nuisance parameters

Number of expected yields divided into signal and backgrounds:

$$N_{exp} = \mu s(\vec{\theta}) + b(\vec{\theta})$$

Where:

- **μ is the POI,** represents the signal strength and its value is 1 in SM case
- **θ are the nuisance parameters** (in other expressions will be represented WITHOUT the vector symbol for simplicity)
- s is the expected number of signal events in SM case
- b is the expected number of background events

$$\mathscr{L}(\mu, \theta) = \prod_{j=1}^{N_{bins}} \left[\frac{(\mu s_j(\theta) + b_j(\theta))^{n_j}}{n_j} e^{-(\mu s_j(\theta)+b_j(\theta))} \cdot \prod_{i=1}^{N_{syst}} p_i(\hat{\theta}_i | \theta_i) \right]$$

Shape analysis:

Relies on **event yield comparison** in data with signal + backgrounds ones **AND on their distribution in the discriminating observable**

Histogram with N bin (Pois disturb in each bin), equivalent to N counting experiments.

Probability for the true value to be equal to θ_i, given its best estimate (from auxiliary measurements in CR/MC)

Normalisation: logNormal distribution

Shape: Variate Template Morphing technique

[2] Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.

Statistical treatment

Higgs analyses statistical treatment explained in [2]

$\hat{\theta}_\mu$ - conditional MLE of θ given μ

data - actual or toy MC data

$\hat{\theta}$ and $\hat{\mu}$: MLE

1. Construct the likelihood function
2. To compare the compatibility of the data with the background-only and signal+background hypotheses, where the signal is allowed to be scaled by some factor μ: construct the test statistic

Two hypotheses: $H_1 = H_{\mu s+b}$ and $H_0 = H_b$

Test statistic based on profile likelihood ratio

Set an exclusion limit on signal presence: find the value of μ to reject H_1 hypothesis in favour of H_0

$$\tilde{q}_\mu = -2 \frac{ln(\mathscr{L}(data\,|\,\mu, \hat{\theta}_\mu))}{ln(\mathscr{L}(data\,|\,\hat{\mu}, \hat{\theta}))} \text{ for } 0 \leq \hat{\mu} \leq \mu$$

Signal rate must be positive to make physics sense
One-sided confidence interval (non detached from zero)
Upward fluctuations not considered as evidence

[2] Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.

Statistical treatment

Higgs analyses statistical treatment explained in [2]

1. Construct the likelihood function

2. To compare the compatibility of the data with the background-only and signal+background hypotheses, where the signal is allowed to be scaled by some factor μ: construct the test statistic

3. Find the observed value of the test statistic for the given signal strength modifier μ under test.

4. Find MLE of the nuisance parameters for both the hypotheses $\hat{\theta}_{\mu}^{obs}$ and $\hat{\theta}_{0}^{obs}$

5. Generate toy Monte Carlo pseudo-data to construct pdfs $f(\tilde{q}_{\mu}\,|\,\mu, \hat{\theta}_{\mu}^{obs})$ and $f(\tilde{q}_{\mu}\,|\,\mu, \hat{\theta}_{\mu}^{obs})$ nuisance parameters are fixed to the observed values by fitting the observed data, but are allowed to float in fits needed to evaluate the test statistic.

[2] Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.

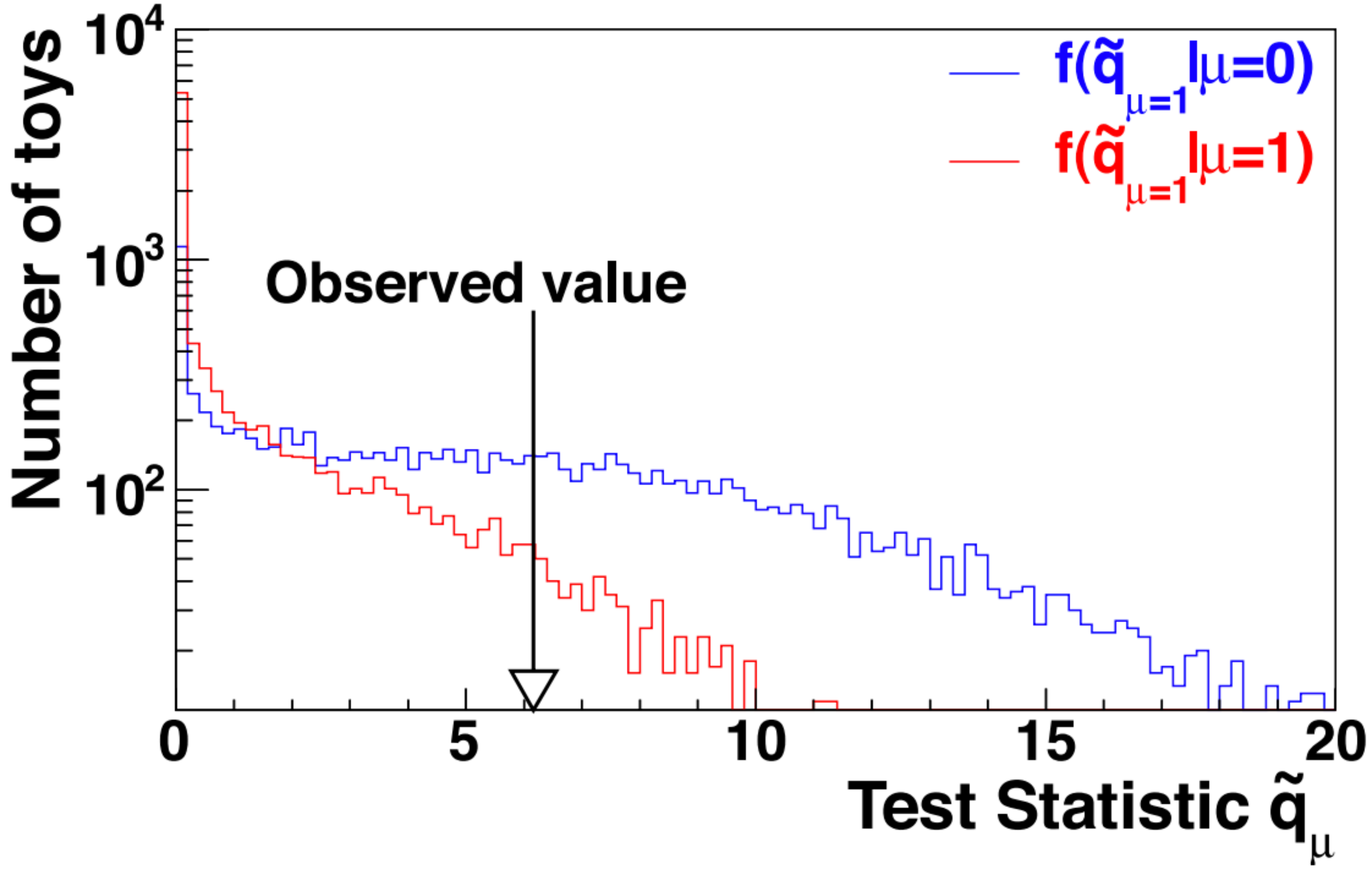
Statistical treatment

Higgs analyses statistical treatment explained in [2]

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} \,|\, H_{s+b}) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu \,|\, \mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} \,|\, H_b) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu \,|\, 0, \hat{\theta}_0^{obs}) d\tilde{q}_\mu$$

$$CLs(\mu) = \frac{p_\mu}{1 - p_b}$$



6. Define two p-values to be associated with the actual observation for the signal+background and background-only hypotheses and calculate CLs(μ) as a ratio of these two probabilities

7. If, for μ = 1, CLs ≤ α, the signal is excluded with (1 − α) CLs confidence level; α is set to 0.05

CLs method provides conservative limits

[2] Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.

Towards Run-3

- Lot of knowledge learnt during Run2 resonant and non-resonant analyses. But we can always do better and optimise our analysis tools
- Several improvements for Run 3:
  - New HH/HHH oriented triggers
  - Including boosted taus
  - Improvements in ML based tools for the signal extraction and for the object identification
  - Including VBF (resolved and boosted)
- Many frameworks, for good coordination and dedicated studies (hopefully!):
  - FLAF (Texas A&U - Pisa)
  - CCLUB (CIEMAT - CEA - LLR - UZH - Milano-Bicocca - Colorado-Boulder)
  - Bamboo (UCLouvain)
  - ColumnFlow (UHH - KBFI - LIP)
  - Run 3 framework setup using coffea (CMU)
  - PKU Further efforts targeting resonant production

The analysis framework: FLAF

Flexible Law Analysis Framework

- Two frameworks, the "main" one and the "antagonist": I am the main developer of the antagonist one the latter, from scratch
- Luigi Analysis Workflow (LAW) for job submission handling
- Modular structure to integrate different selection, corrections, features
  - People from other groups starting using it for HNL, HHbbWW, (Run3) HHbb$\tau\tau$ and TTHH analyses
- Input files: NanoAOD (not going in detail, but state-of-art recommended data format for analyses)
- Totally based on RDataFrame, using pyROOT and C++
- In the meantime... : found bugs in the central CMS software (CMSSW) for the NanoAOD production, helped to fix them