Search for non-resonant Higgs boson pair production in the final state with two bottom quarks and two tau leptons with CMS

Higgs Hunting 2022

Valeria D'Amante

on behalf of CMS Collaboration

13/09/2022

Non-resonant HH production

Sensitive to λ_{HHH}, y_t, c_V, c_{2V}:

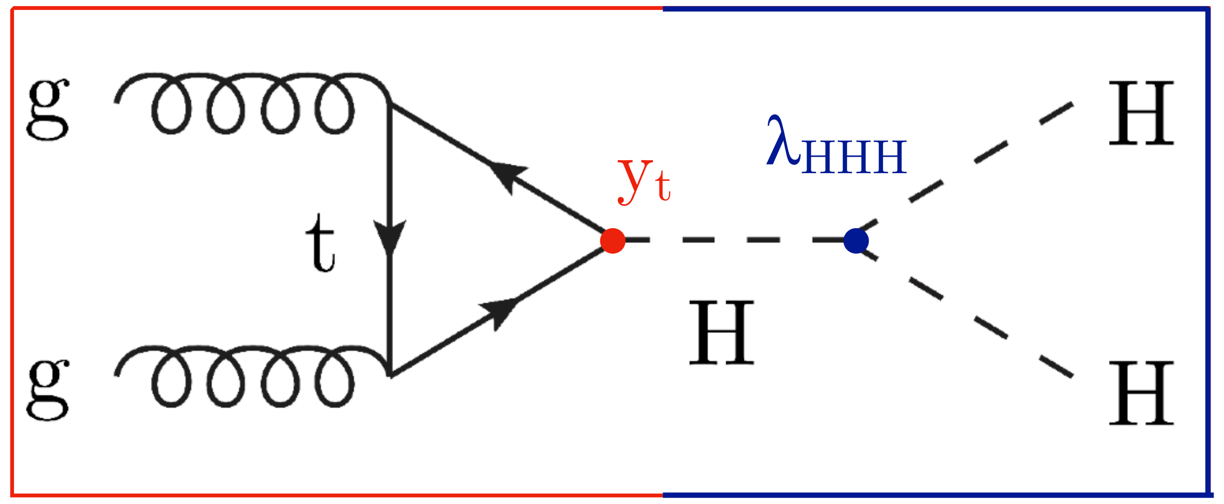
- Direct measurements of scalar sector properties
- Independent SM test
- BSM effective models with anomalous couplings using **strength modifiers** κ_λ λ_t κ_V, κ_{2V}

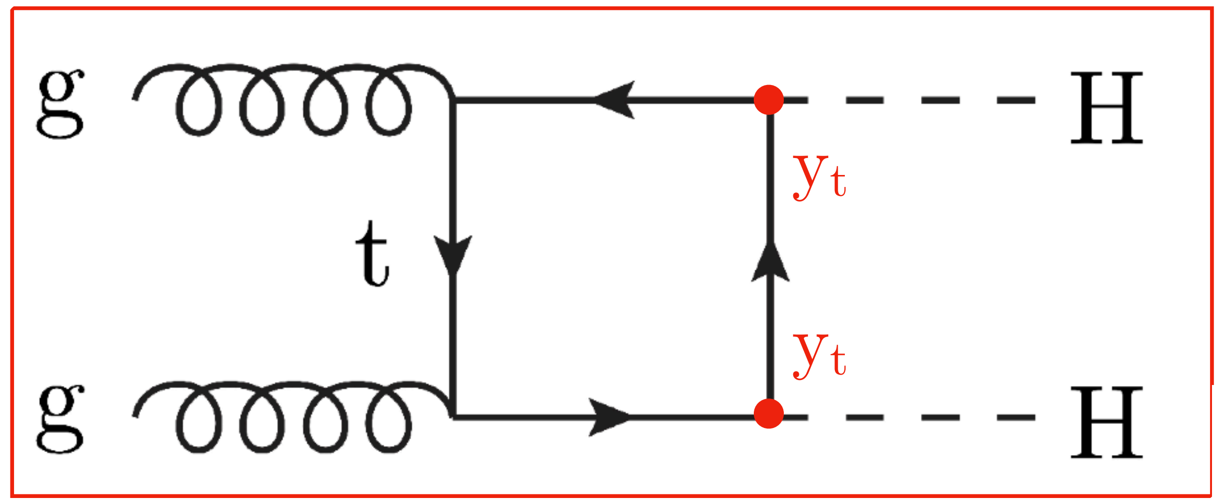
Cross section at LHC @$\sqrt{s}$ = 13 TeV

$\sigma_{SM}(pp \to HH) \approx 31$ fb

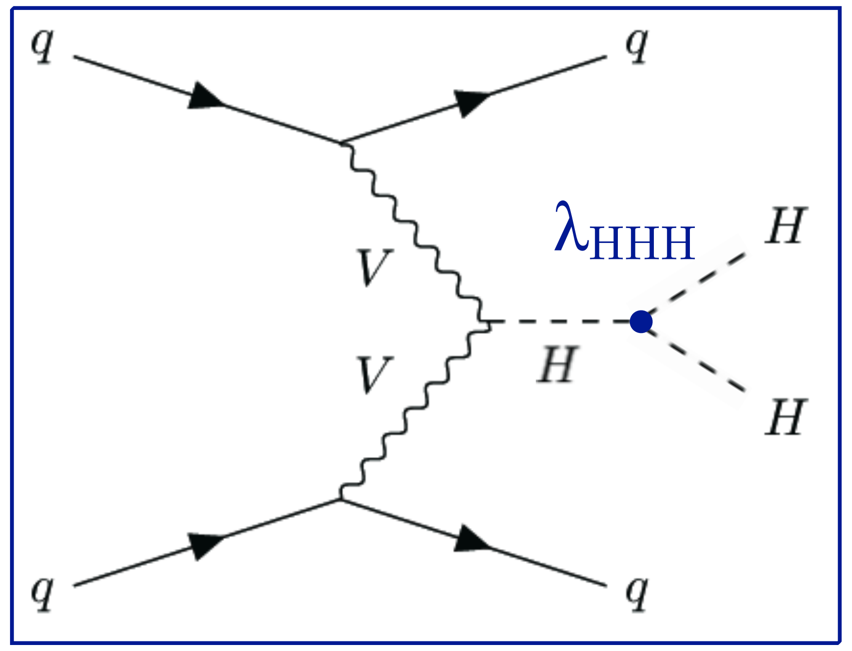
$\sigma_{SM}(pp \to qqHH) \approx 1.7$ fb

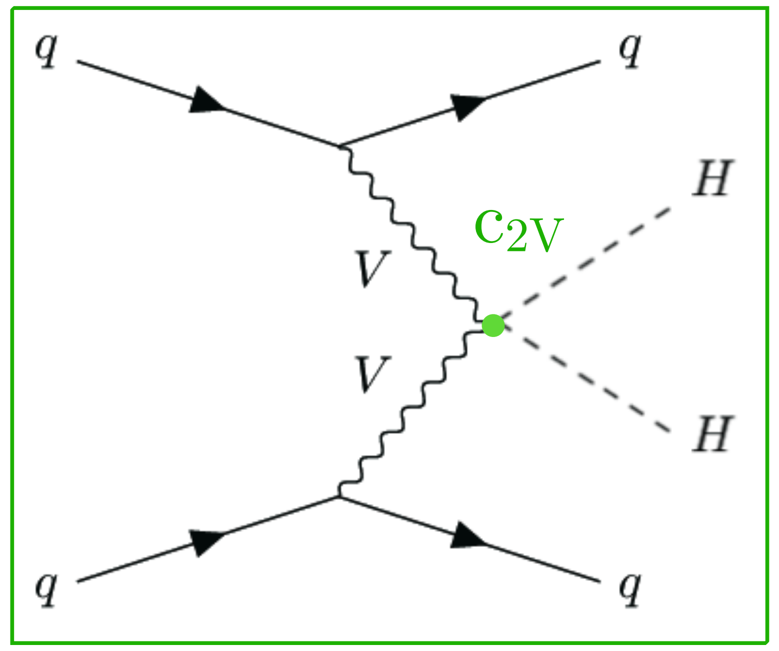
Gluon-Gluon Fusion (ggF)

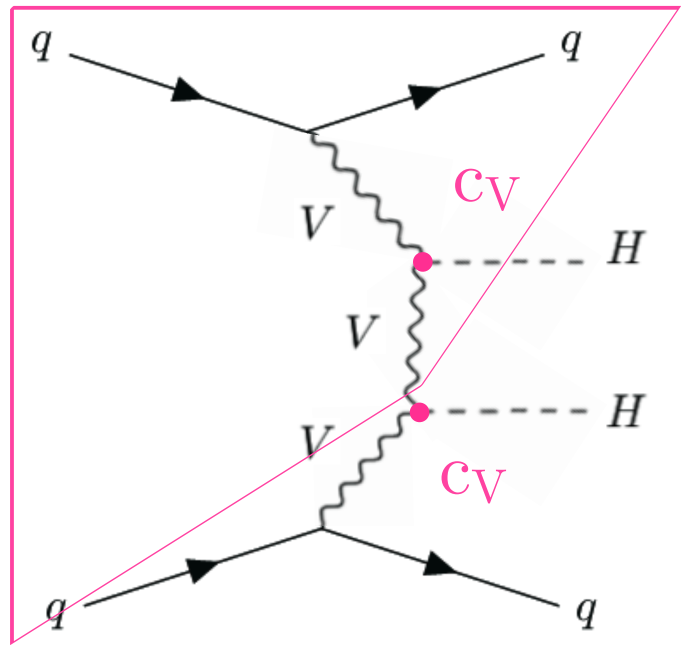


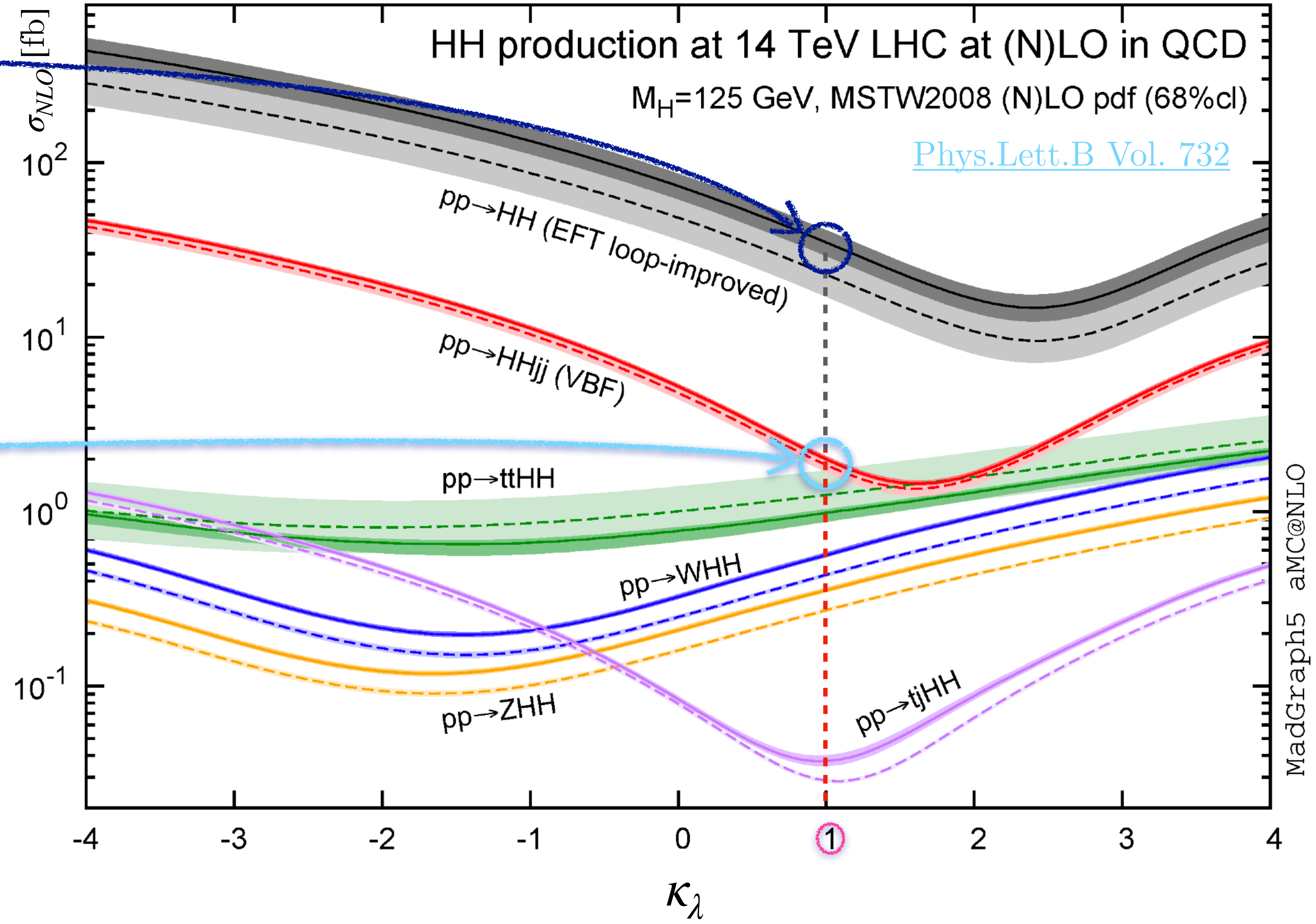


Vector Boson Fusion (VBF)









Non-resonant HH production

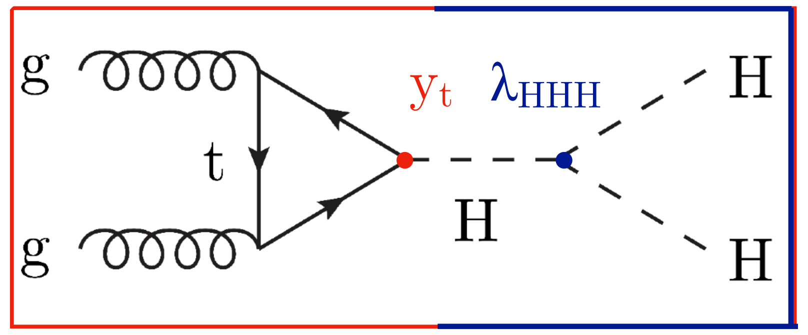
Gluon-Gluon Fusion (ggF)

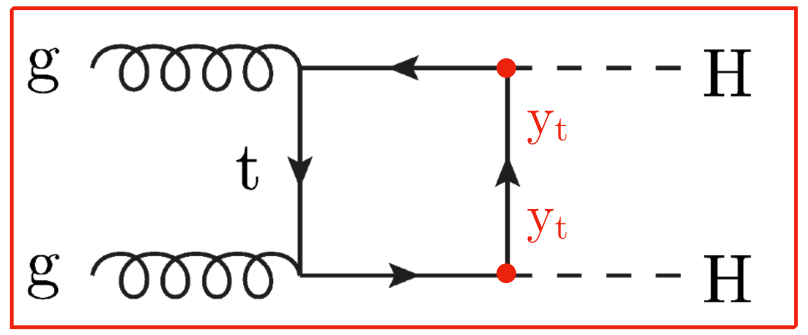
Large destructive interference

small variation of the couplings from prediction

Large modification of the cross section

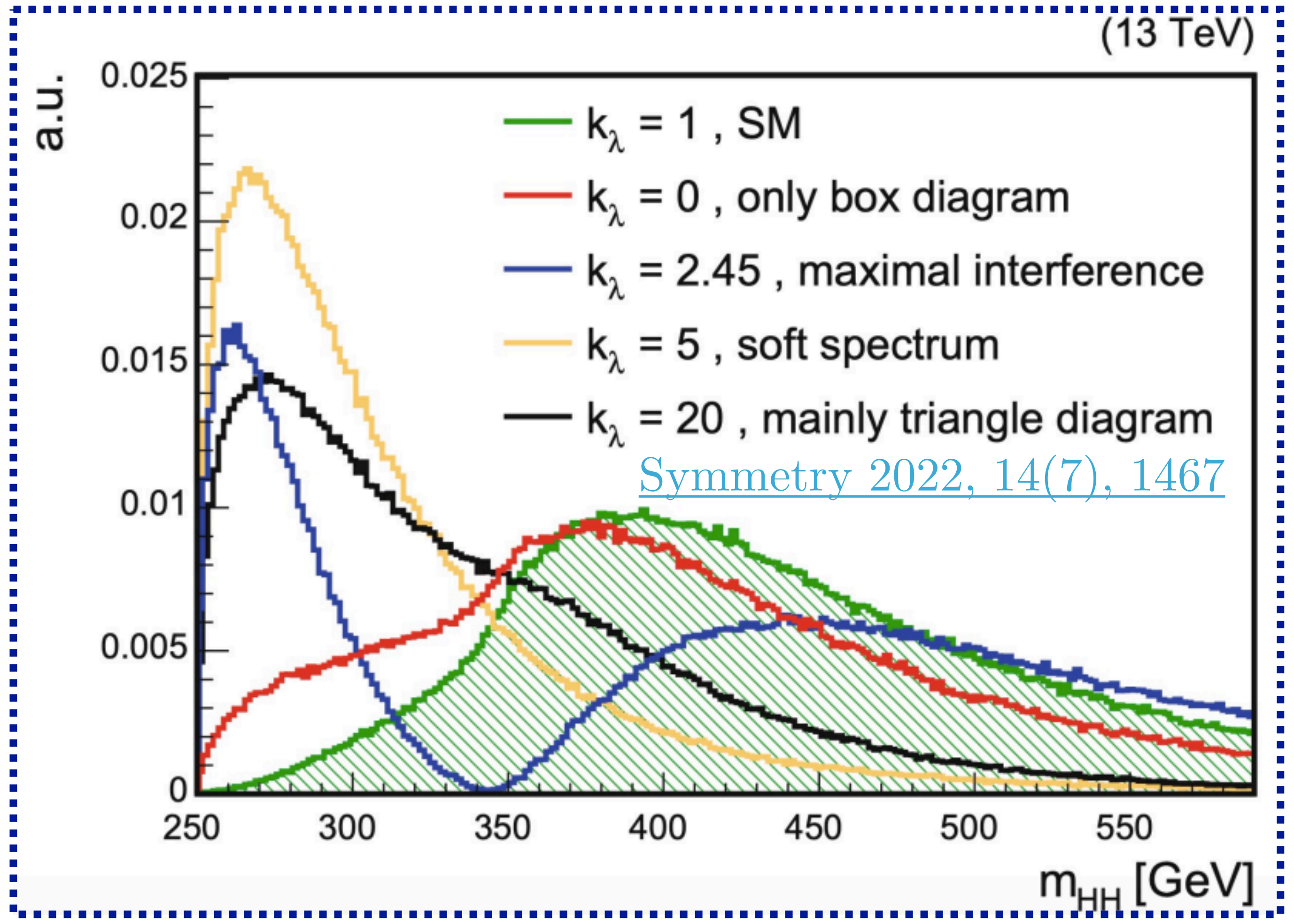
Affects also the HH **kinematics**





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

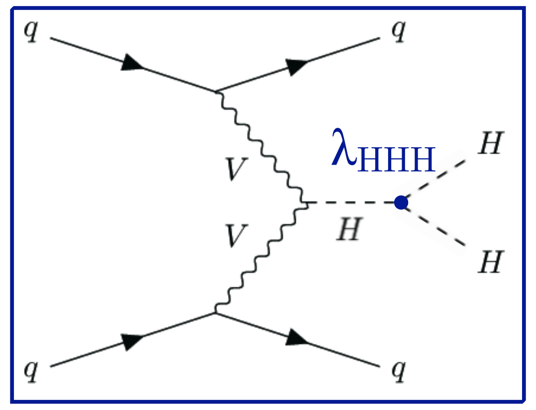
$\sigma_{SM}(pp \rightarrow HH) \approx 31$ fb

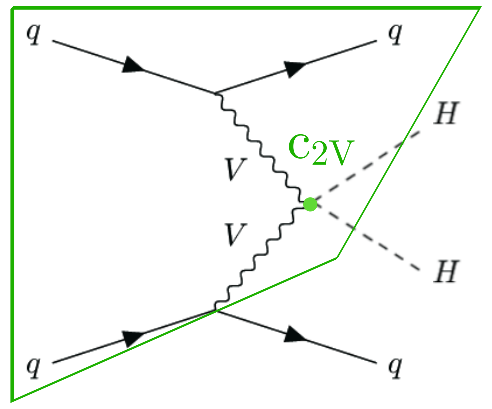


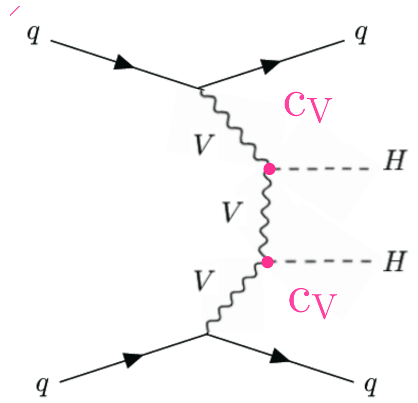
Vector Boson Fusion (VBF)

Very small cross section but distinctive signature:

2 forward jet highly separated in η with large invariant mass

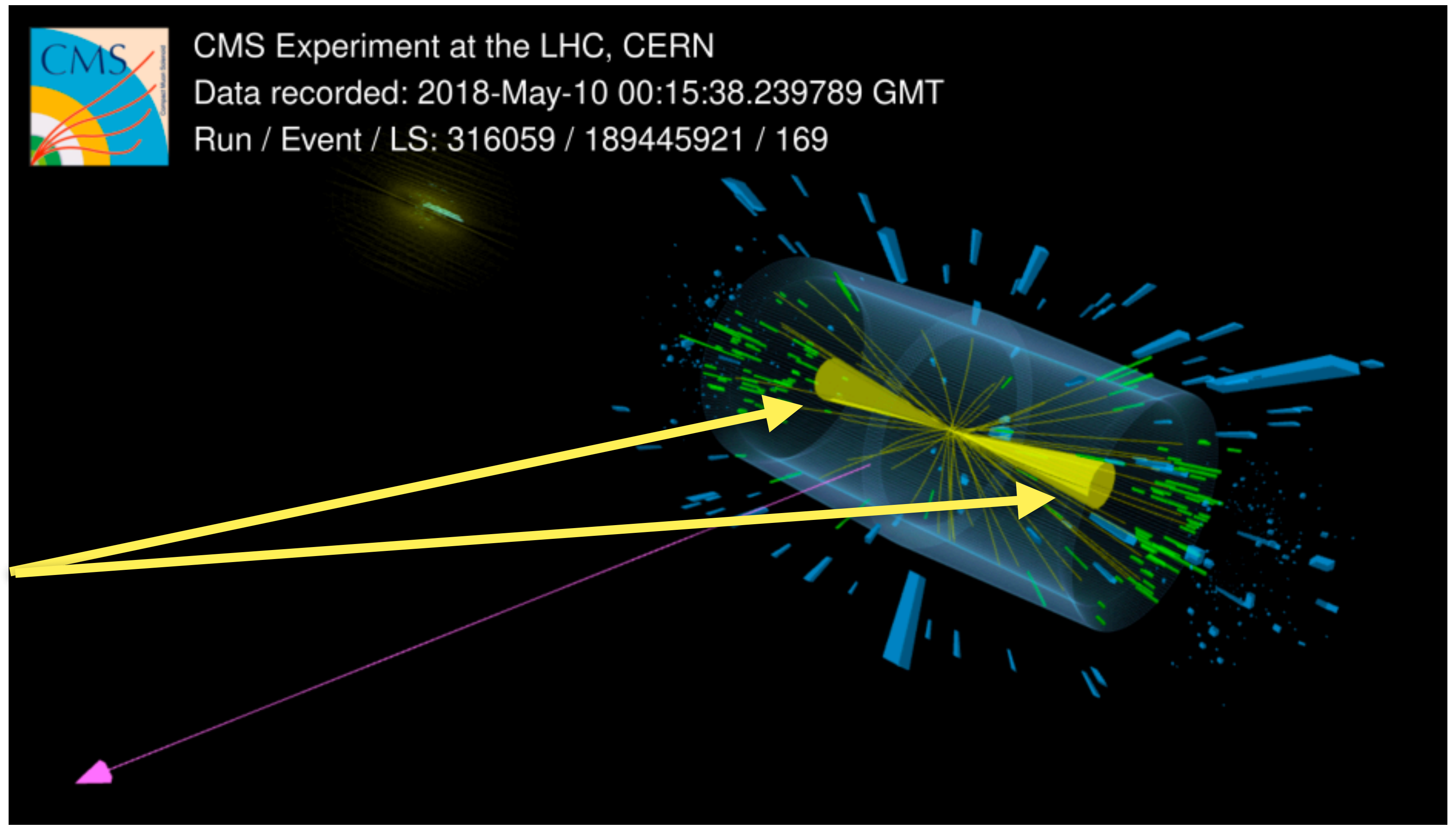




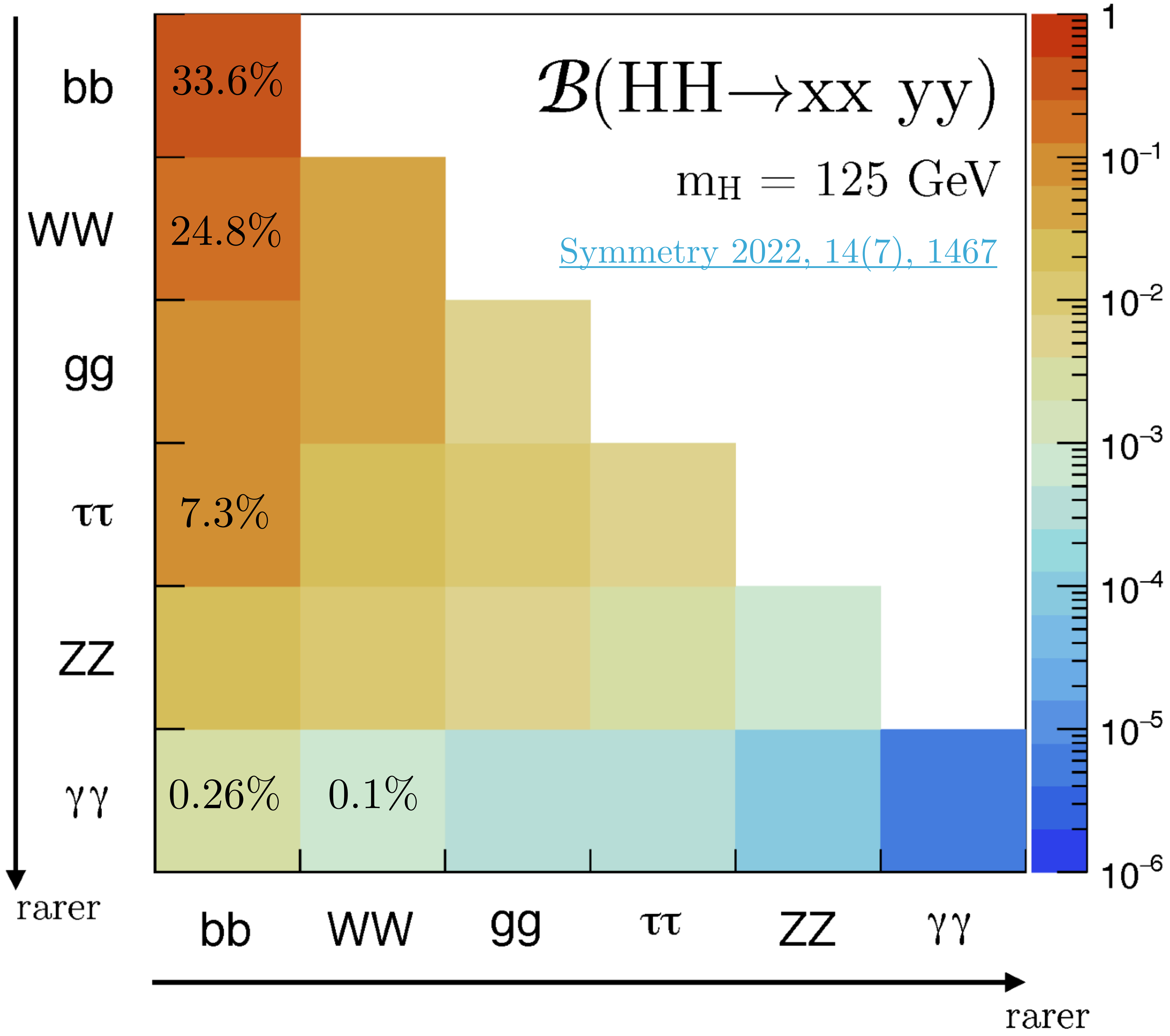


At LHC @$\sqrt{s}$ = 13 TeV

$\sigma_{SM}(pp \rightarrow qqHH) \approx 1.7$ fb



The bb$\tau\tau$ final state



bbττ channel has a medium BR and a relatively small background:

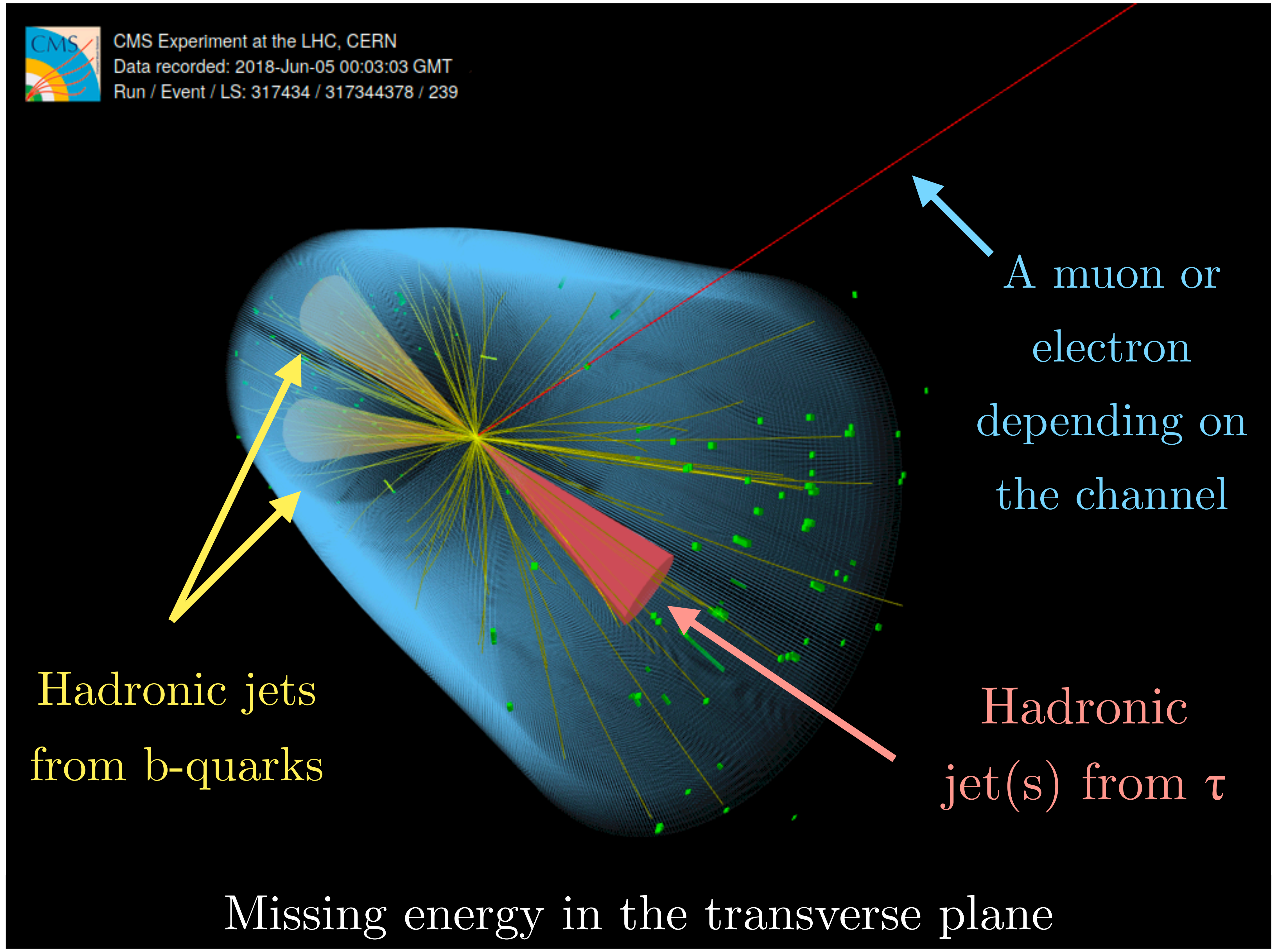
→ good trade-off between BR and signal purity

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

Final states considered (channels): $\boldsymbol{\tau_h\tau_e}$, $\boldsymbol{\tau_h\tau_\mu}$, $\boldsymbol{\tau_h\tau_h}$

$$\mathscr{B}(\tau\tau \to \tau_h\,\tau_{\{e,\mu,h\}}) \approx 88\,\%$$

In the CMS detector the signal can be identified by looking for..



HH → bb$\tau\tau$ analysis at CMS

Previous analysis: PLB 778 (2018) 101

This analysis (Submitted to PLB): CMS-PAS-HIG-20-010

New features w.r.t. previous analysis:

- **Full Run 2 data**: total integrated lumi 138 fb^{-1} (collected in 2016, 2017 and 2018)
- **VBF production included** in the studies the (only ggF was considered in prev. analysis)
- **Improved trigger strategy**: single lepton, di-τ, lepton + τ, and new dedicated VBF di-τ triggers
- **Introduction of new Machine-Learning (ML) algorithms** to improve:
  - → τ ID (DeepTau) and b-tagging (DeepFlavour) - central tools developed in CMS
  - → H → bb candidate selection (HH-bTag)
  - → VBF Categorisation (Deep Neural Network - DNN - multiclassifier)
  - → Signal extraction (DNN algorithm to discriminate SM HH signals from all backgrounds)
- **Improved background modelling** (data-driven corrections)

Analysis workflow

Trigger Requirements

$H \to \tau\tau$ decay products + dedicated VBF triggers

$H \to \tau\tau$ selection

Two OC leptons well identified and isolated: $e\tau_h$ OR $\mu\tau_h$ OR $\tau_h\tau_h$
Match to trigger objects and apply third lepton veto

Best pair chosen

$H \to bb$ and VBF jet selection

bb pair identified via the **HH-Btag** algorithm
Select additional (VBF) jets with the highest m_{jj}

HH selection and event categorisation:

Elliptical mass cut on m(bb) VS m(ττ)
Events divided in 8 **categories** (2 resolved, 1 boosted, 5 from **VBF multi classifier**)

Limits and likelihood scan

Inclusion of all relevant experimental and theory uncertainties
Signal extraction using **DNN score distributions** for each year/channel/category

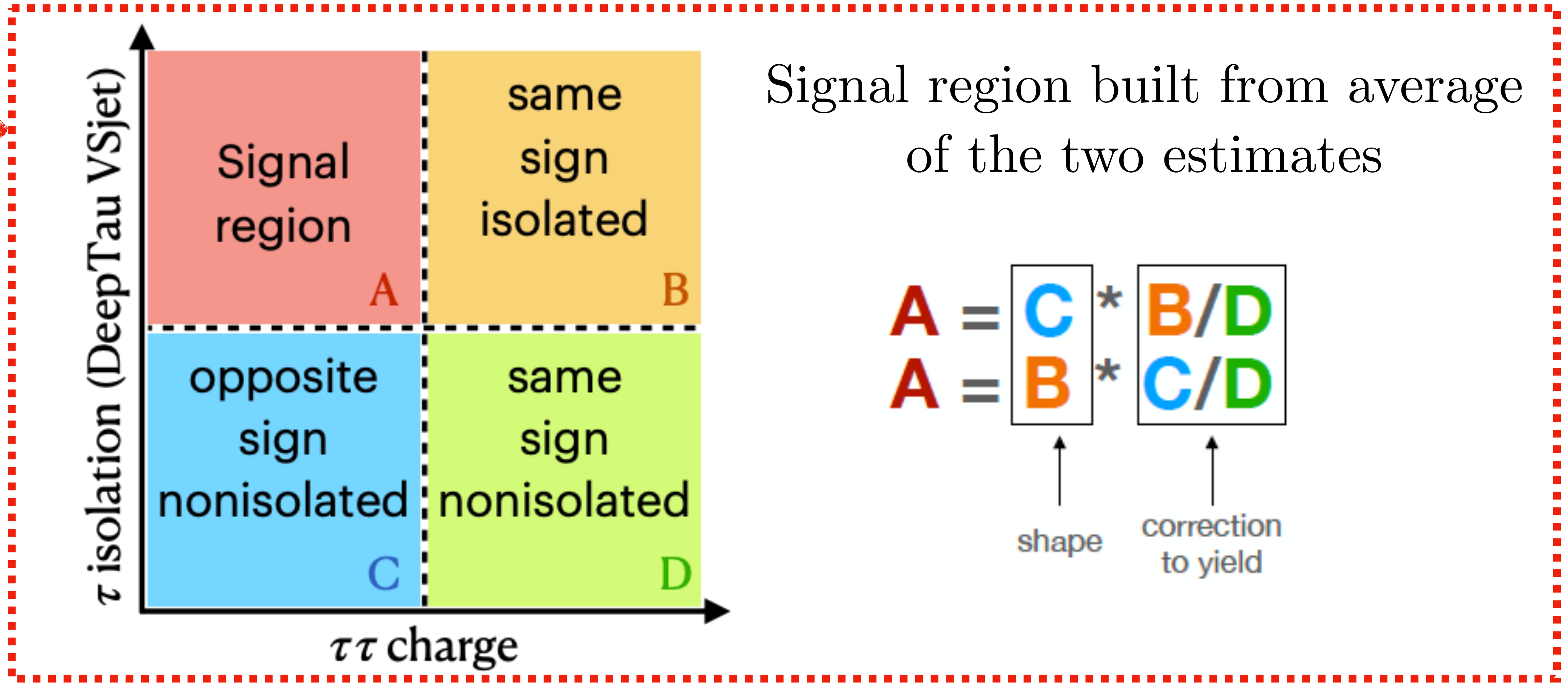
Event Categorisation

(*) Event selected by only VBF di-tau trigger, further requirements: $m_{jj} > 800\,\text{GeV}$, $p_T > 140\,(60)\,\text{GeV}$ for the (sub)leading jet

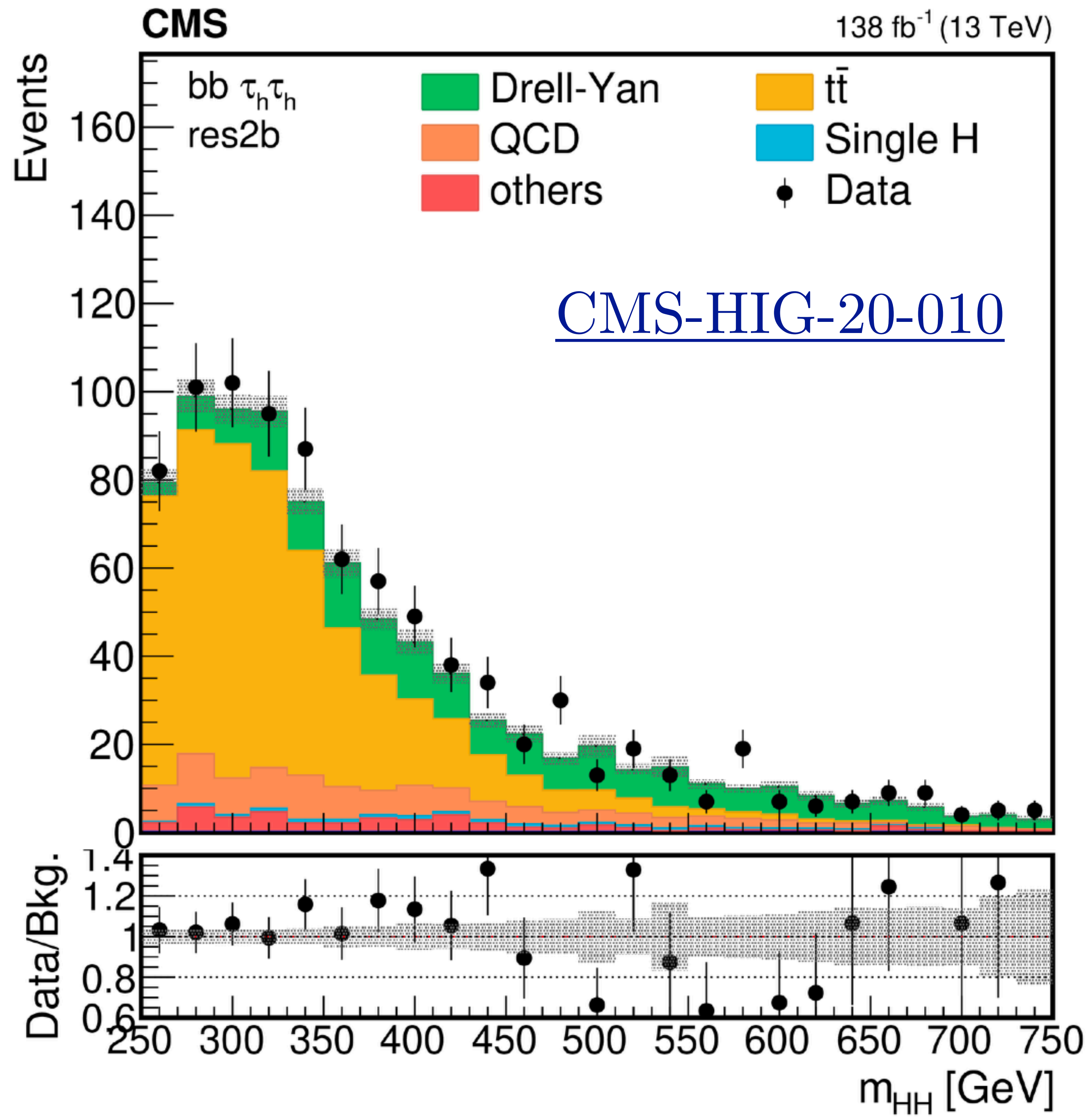
Background estimation

Main processes that mimic the signal final state

- **QCD multi-jets**
  - → Yield and Shape: data-driven estimation via the **ABCD method**

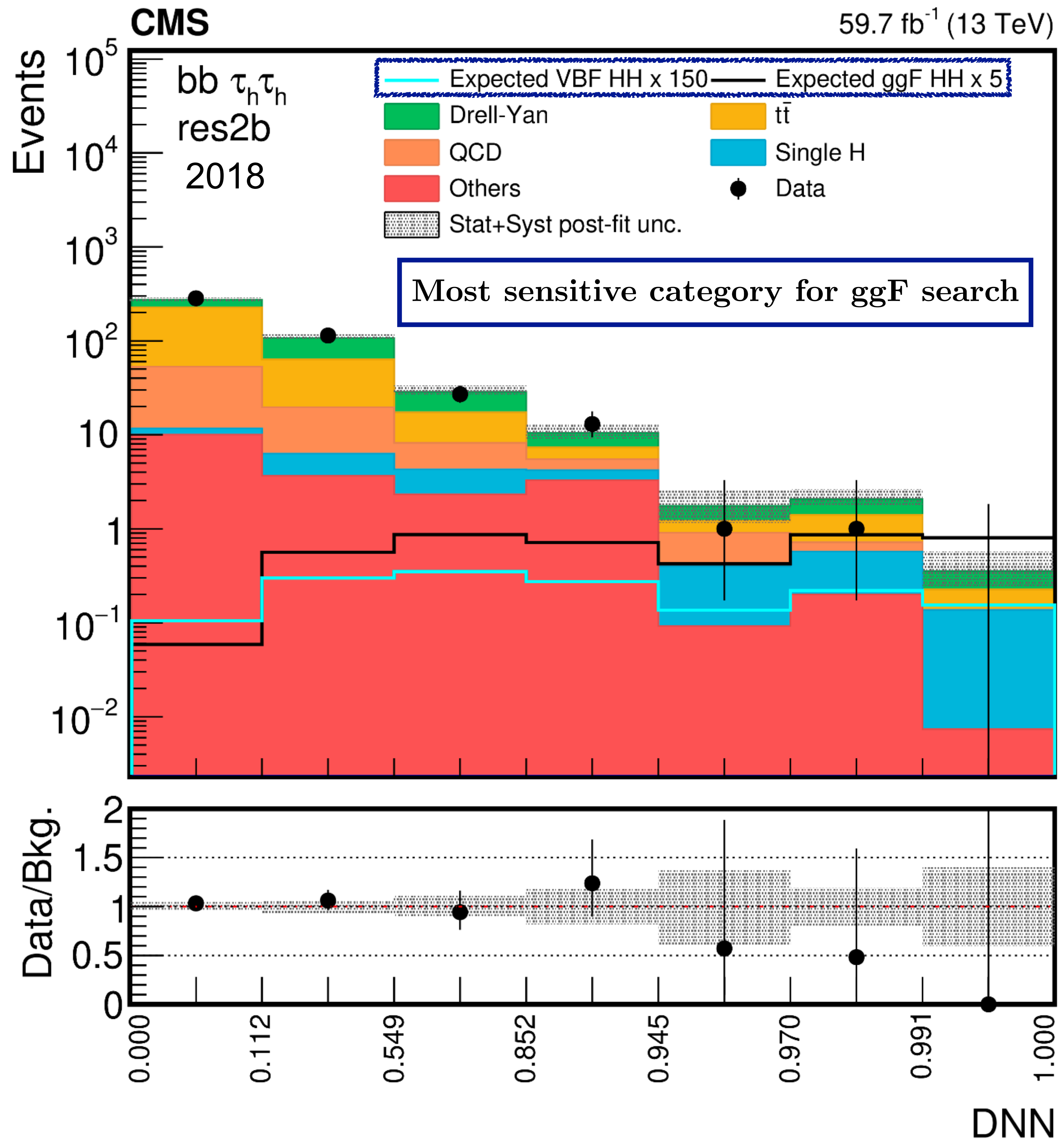


- **Drell-Yan events** $(Z \to \ell\ell + \text{jets})$
  - → Shape: MC simulation
  - → Yield: data-driven estimation in $Z \to \mu\mu$ enriched sample
- **Top-antitop production** $(t\bar{t} \to b\bar{b}W^{+}W^{-} \to b\bar{b}\ell\nu_{\ell}\tau\nu_{\tau})$
  - → Shape: MC Simulation
  - → Yield: SFs per year fitted from a $t\bar{t}$ enriched CR and validated in a $t\bar{t}$ enriched subset of the SR
- **Minor backgrounds**: SM single Higgs, Di-Boson, Tri-Boson, other processes (W+Jets, Single Top, ttW, ttZ..)
  - → Yield and Shape: MC Simulation



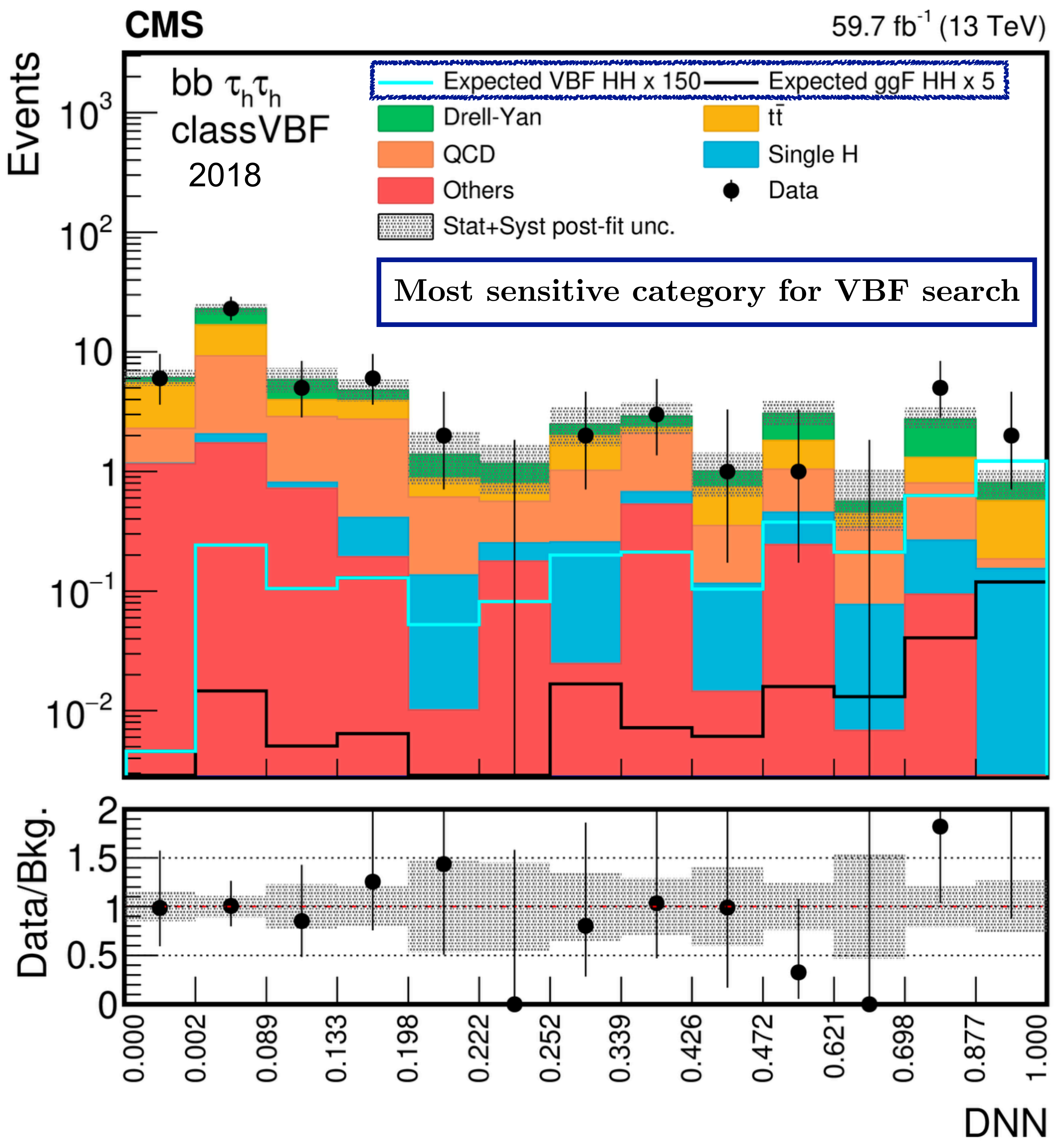
Signal extraction

Dedicated Deep Neural Network (DNN) trained to separate SM non-resonant HH signal vs all backgrounds



CMS-PAS-HIG-20-010

Post fit distributions



Binned DNN score distribution as final discrimination for signal extraction in all analysis categories

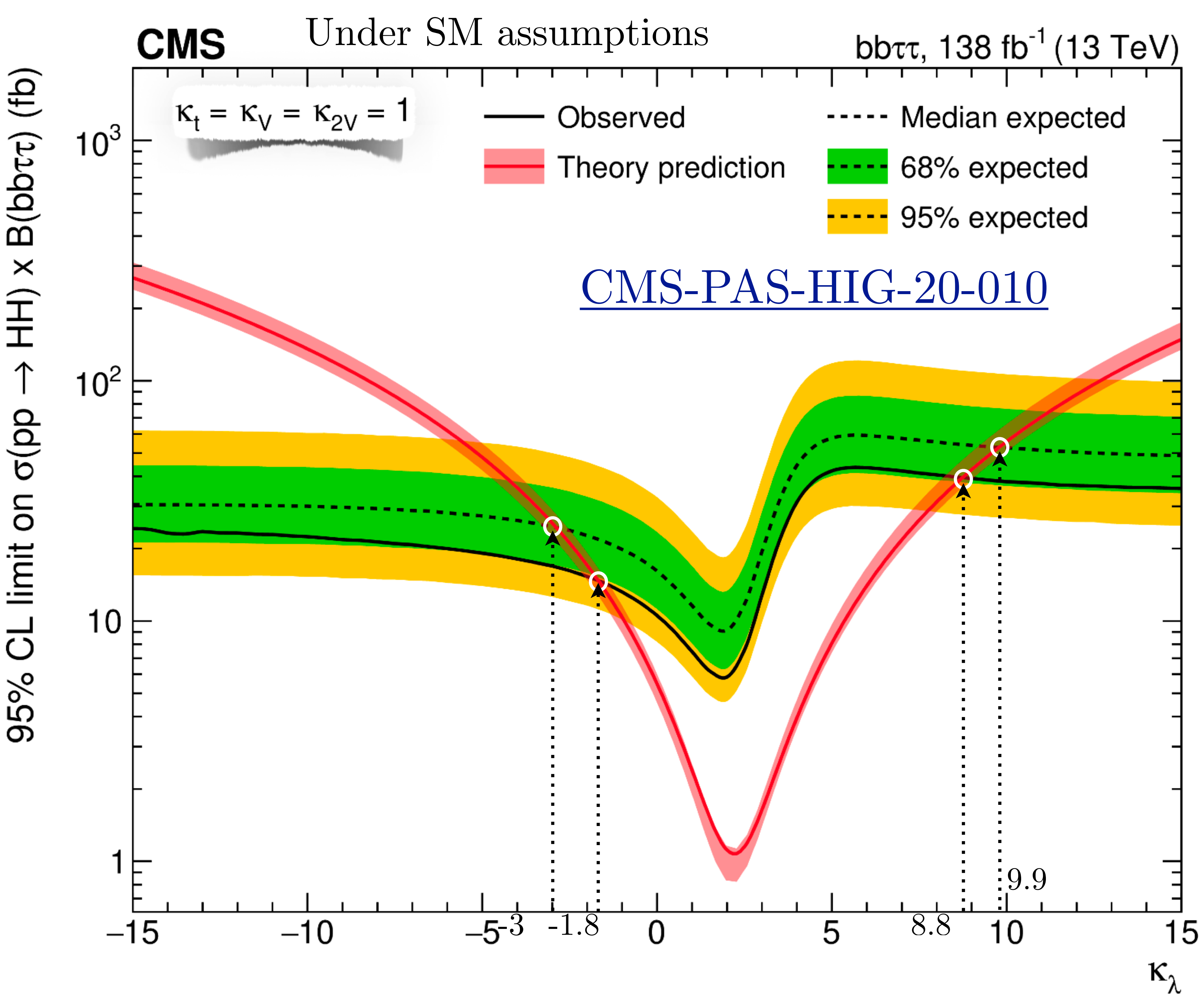
Systematic uncertainties

Main sources of systematic uncertainties

- Theory uncertainty on ggF HH cross section:
  $\sigma_{SM}(pp \to HH) = 31.05^{+6\%}_{-23\%}\,(\mathrm{scale{+}m_t}) \pm 3\,\%\,(\mathrm{PDF} + \alpha_\mathrm{s})$ fb
- Statistical fluctuations affecting multi-jet background estimation
- Uncertainties on mis-modelling of jet and tau ID and reconstruction in simulation

Total effect of systematic uncertainties on final limits is $\sim 15\,\%$

Constraint on κ_λ and κ_{2V}

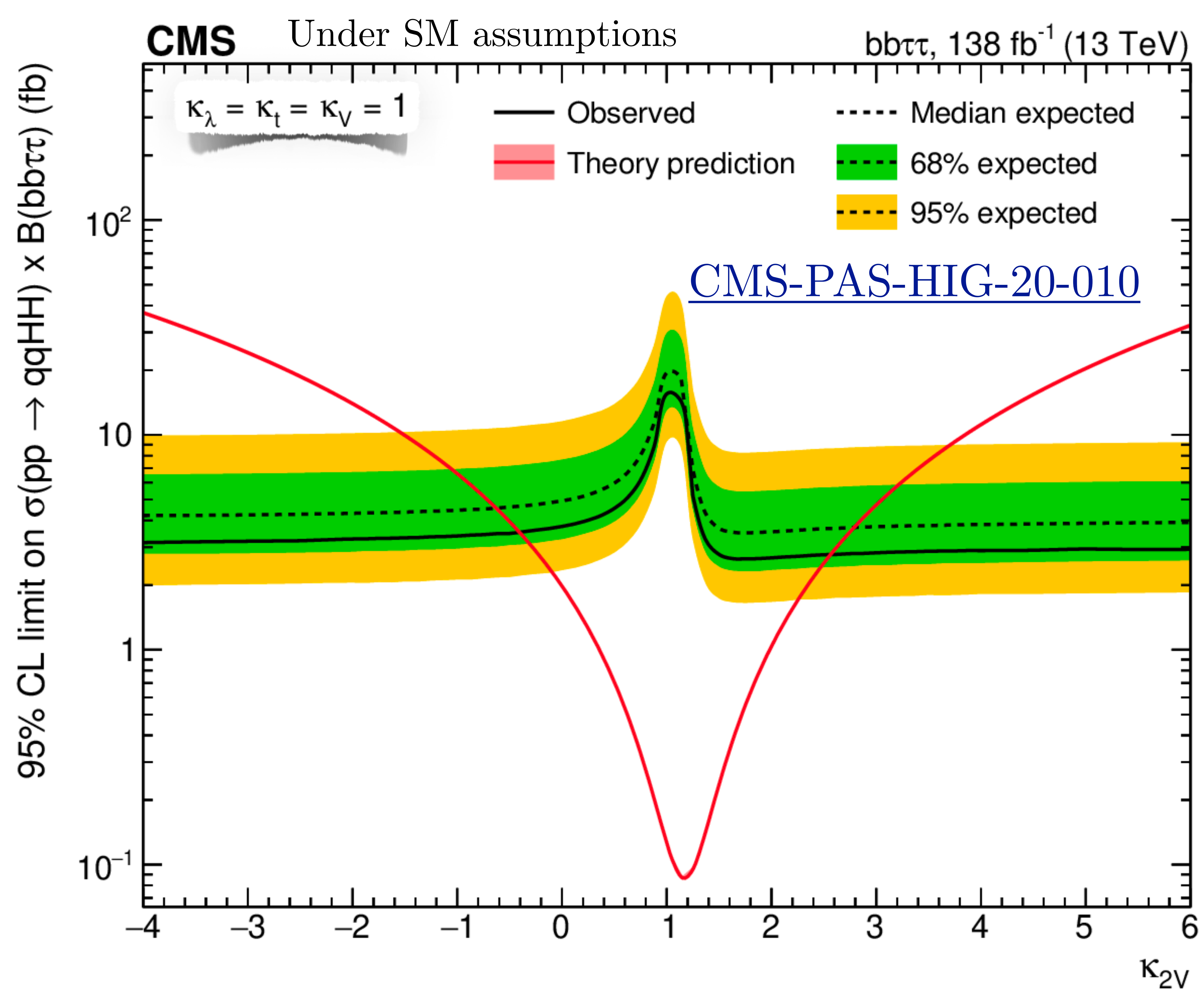


95% CL constraint on κ_λ from limits on

$\sigma(pp \to HH) \times \mathscr{B}(bb\tau\tau)$

Observed $-1.8 \leq \kappa_\lambda \leq 8.8$

Expected $-3.0 \leq \kappa_{2V} \leq 9.9$



95% CL constraint on κ_{2V} from limits on

$\sigma(pp \to qqHH) \times \mathscr{B}(bb\tau\tau)$

Observed $-0.4 \leq \kappa_{2V} \leq 2.6$

Expected $-0.6 \leq \kappa_{2V} \leq 2.8$

Limits on $\sigma(pp \to HH)$

95% CL constraint on $\sigma(pp \to HH) \equiv \sigma^{HH}$

Observed $\sigma^{HH} \leq 3.3 \times \sigma^{HH}_{SM}$

Expected $\sigma^{HH} \leq 5.2 \times \sigma^{HH}_{SM}$

$5\times$ **improvement** with respect to the 95% CL limits set by the previous CMS analysis (PLB 778 (2018) 101) with 2016 data corresponding to a luminosity of 35.9 fb^{-1}

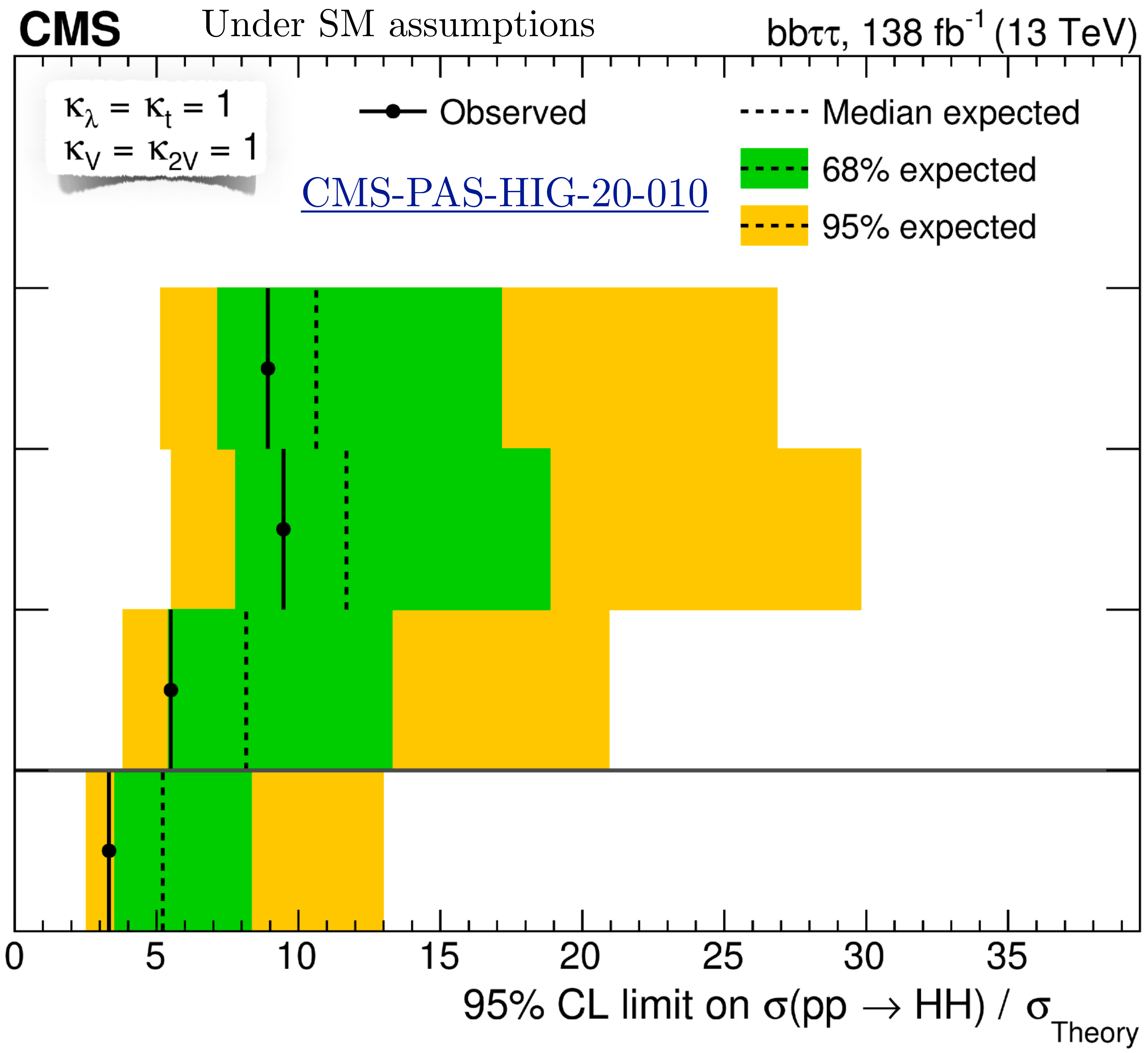
σ^{HH} obs (exp) $\leq 30(25) \times \sigma^{HH}_{SM}$

2016
Expected: 11
Observed: 8.9

2017
Expected: 12
Observed: 9.5

2018
Expected: 8.2
Observed: 5.5

Combined
Expected: 5.2
Observed: 3.3



VBF HH cross section limits

VBF analysis main features

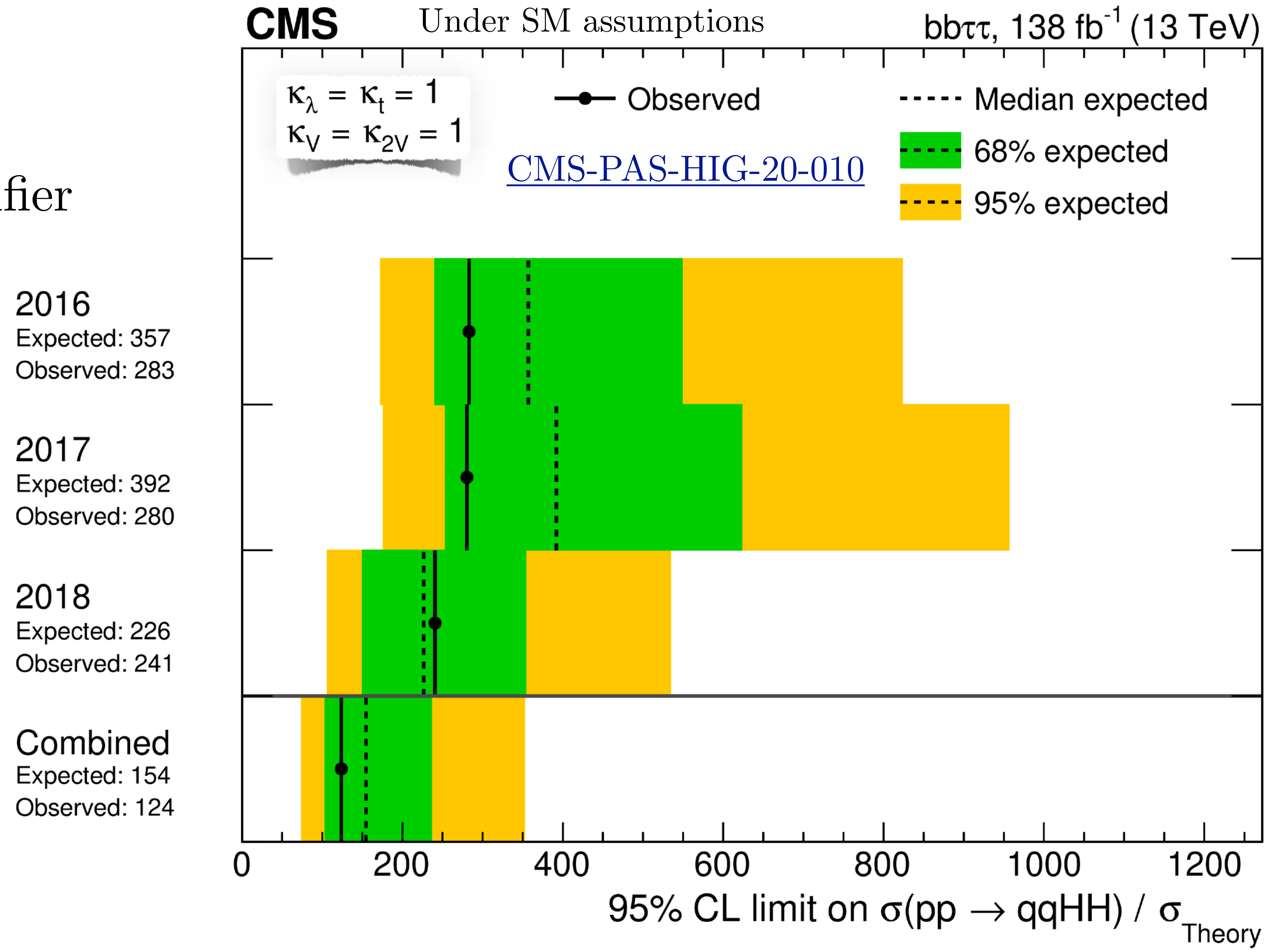
- *Ad-hoc* triggers
- Categorisation via DNN-based multiclassifier

95% CL constraint on $\sigma(pp \to qqHH) \equiv \sigma^{VBF}$

Observed $\sigma^{VBF} \leq 124 \times \sigma^{VBF}_{SM}$

Expected $\sigma^{VBF} \leq 154 \times \sigma^{VBF}_{SM}$

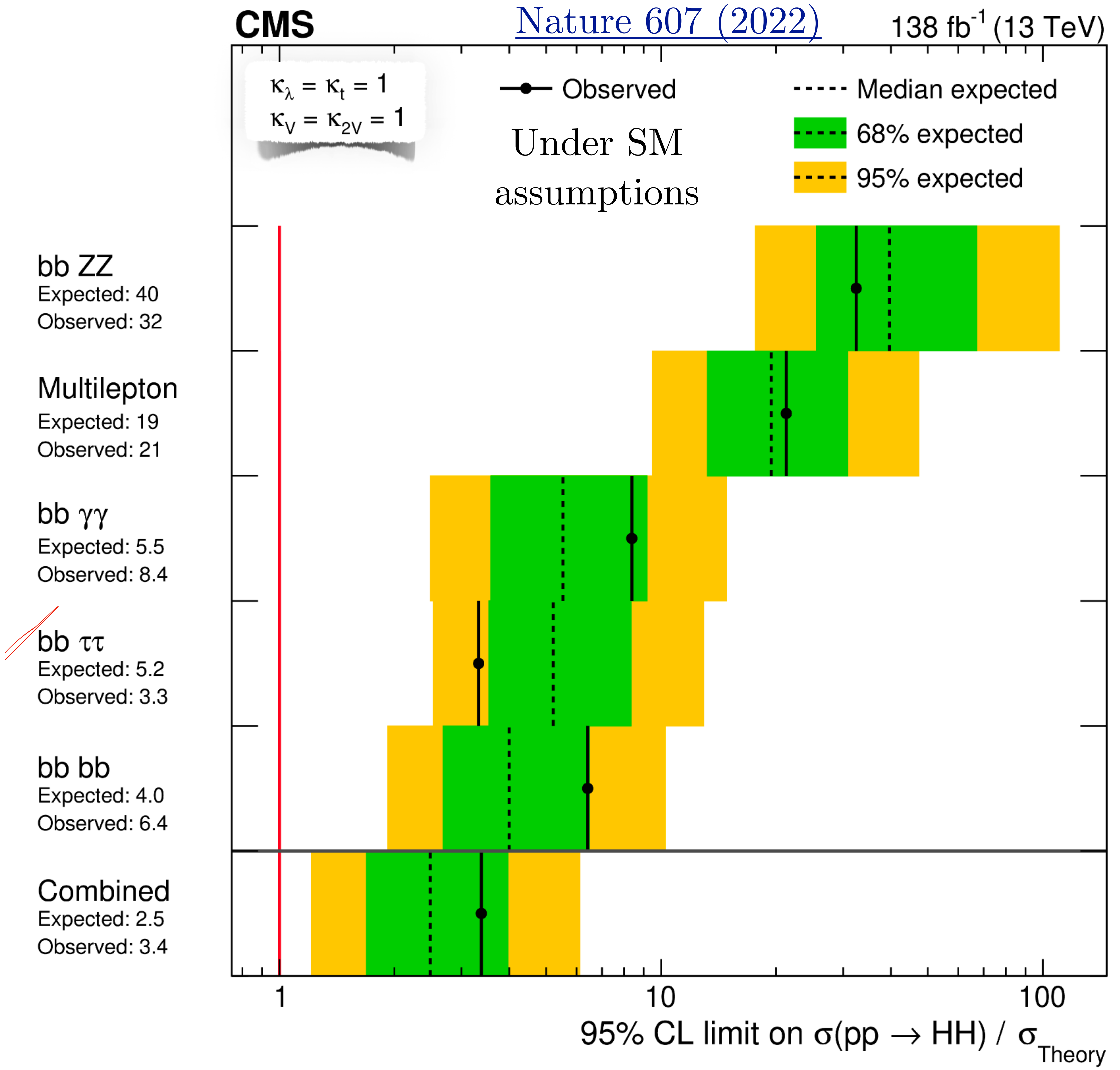
Most stringent 95% limit on σ^{VBF} to date at CMS



Conclusions and final remarks

In this talk the non-resonant HH → bb$\tau\tau$ analysis with full Run 2 data collected by the CMS experiment has been presented

- Thanks to the improvements w.r.t. the previous analysis, it has been possible to achieving particularly stringent results on the HH production cross sections
- The most stringent 95% limit at CMS has been set to σ^{VBF}
- The results are compatible with the ones achieved by the ATLAS experiment (ATLAS-CONF-2021-052)



This analysis has been included in the **Double Higgs combination for the Nature paper about 10 years since Higgs discovery**

Thank you for the attention

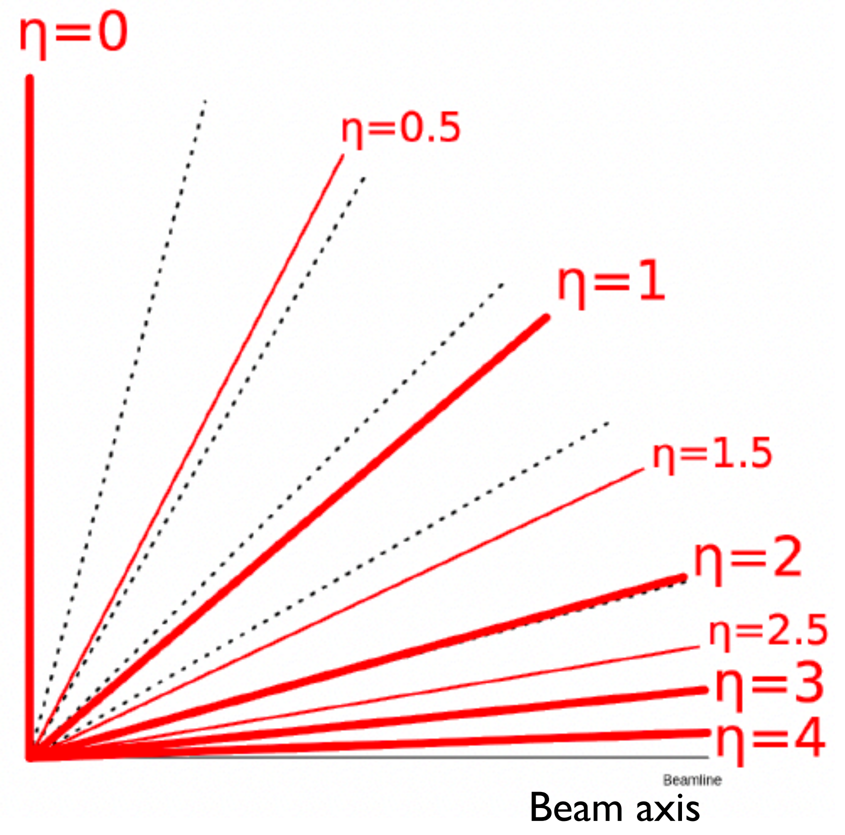
BACKUP

The CMS experiment

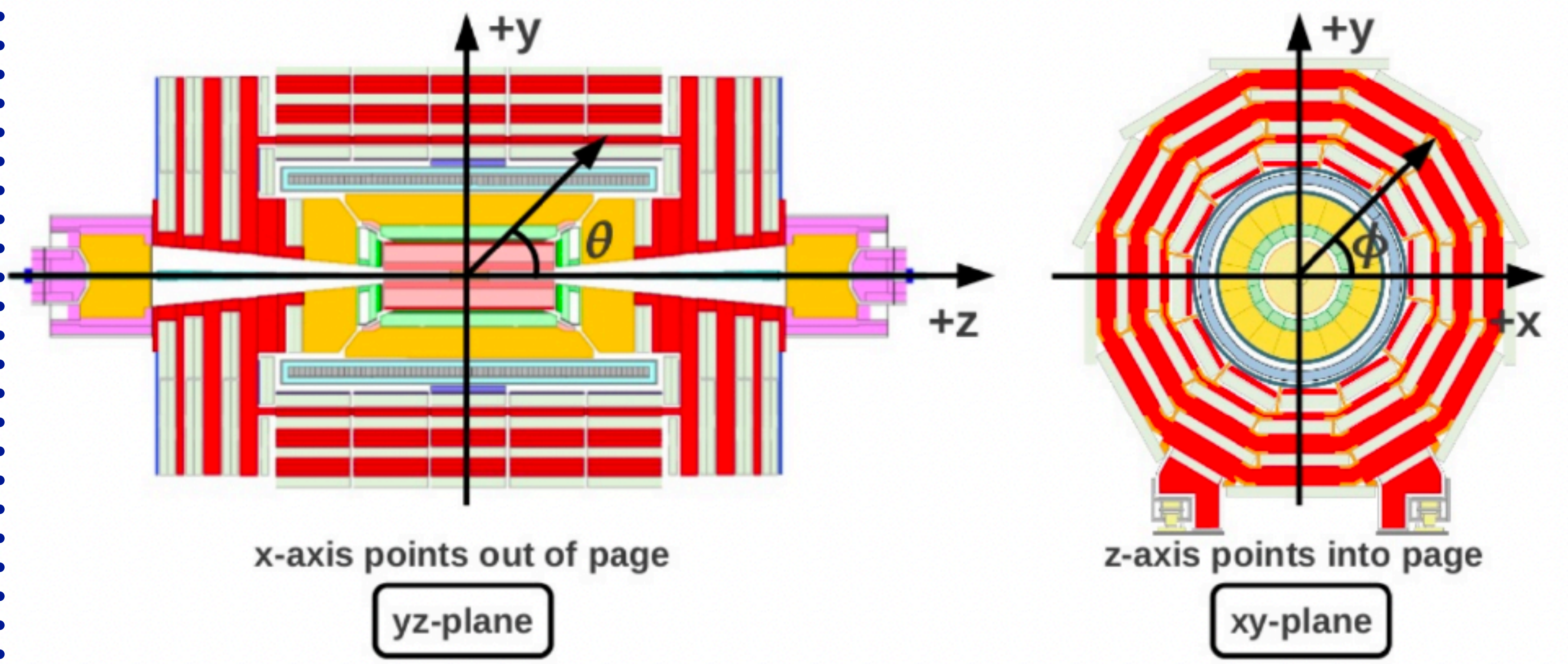
Useful definitions:

- Pseudo-rapidity $\eta = -\ln(\tan(\frac{\theta}{2}))$ - for HR particles it coincides with the rapidity
- Angular separation $\Delta R(i,j) = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$

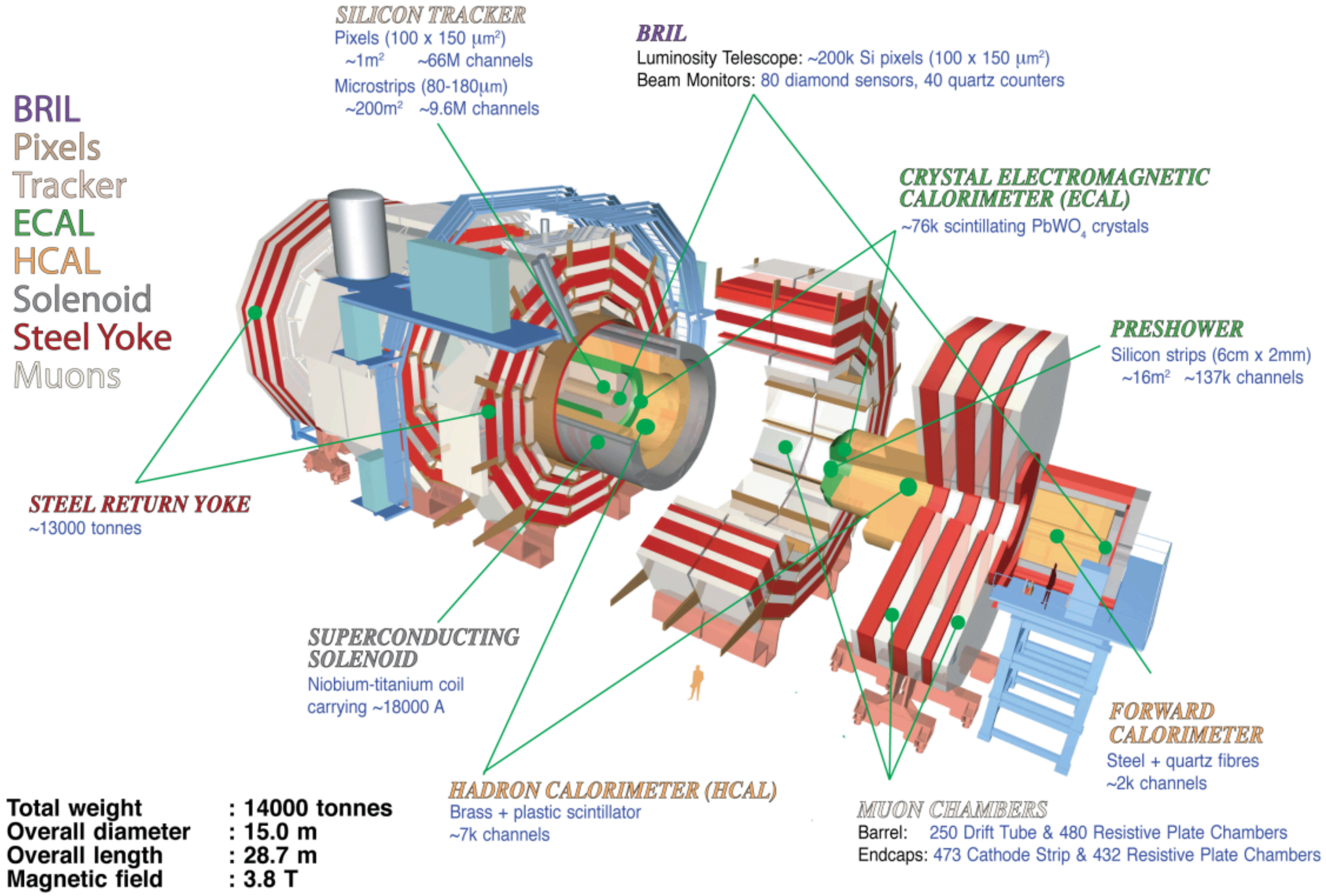
Cartesian coordinate system:
origin - nominal collision
Y-axis - upward vertically
X-axis - radially inside the LHC ring
Z-axis - toward Jura mountains



Cylindrical symmetry
polar coordinate system is more convenient!



General purpose detector, "onion" structure



Baseline selection

$H \to \tau\tau$

General:

- $p_T^{offline} > p_T^{fired\,path} + X$, $|\eta| < 2.3$ X= 1,1,5 GeV for e,μ,τ

Electrons:

- $p_T > 25 \div 33\,GeV$, $|\eta| < 2.1$, $d_{xy} < 0.045\,cm$, $d_z < 0.2\,cm$, Tight MVA iso ID

Muons:

- $p_T > 20 \div 25\,GeV$, $|\eta| < 2.1$, $d_{xy} < 0.045\,cm$, $d_z < 0.2\,cm$, $I_{rel}^{PF} < 0.15$, Tight ID

Taus:

- $p_T > 20 \div 40\,GeV$, $|\eta| < 2.1 \div 2.3$, $d_z < 0.2\,cm$, DeepTau thresholds depending on channel)

Pair Assessment:

- Method used also in the Hττ analysis, $\Delta R(\tau_i, \tau_j) > 0.5$, OS and third lepton Veto

$H \to bb$

AK-4 Jets[1]:

- b-jet selection: $p_T > 20\,GeV$, $|\eta| < 2.4$
- VBF jets: $p_T > 30\,GeV$, $|\eta| < 4.7$
- VBF and b-jets: Tight PFJet ID, $\Delta R(\tau, \text{jet}) > 0.5$
  + Loose PU ID if $p_T < 50\,GeV$

AK-8 Jets[1]:

- $m_{softdrop} > 30\,GeV$

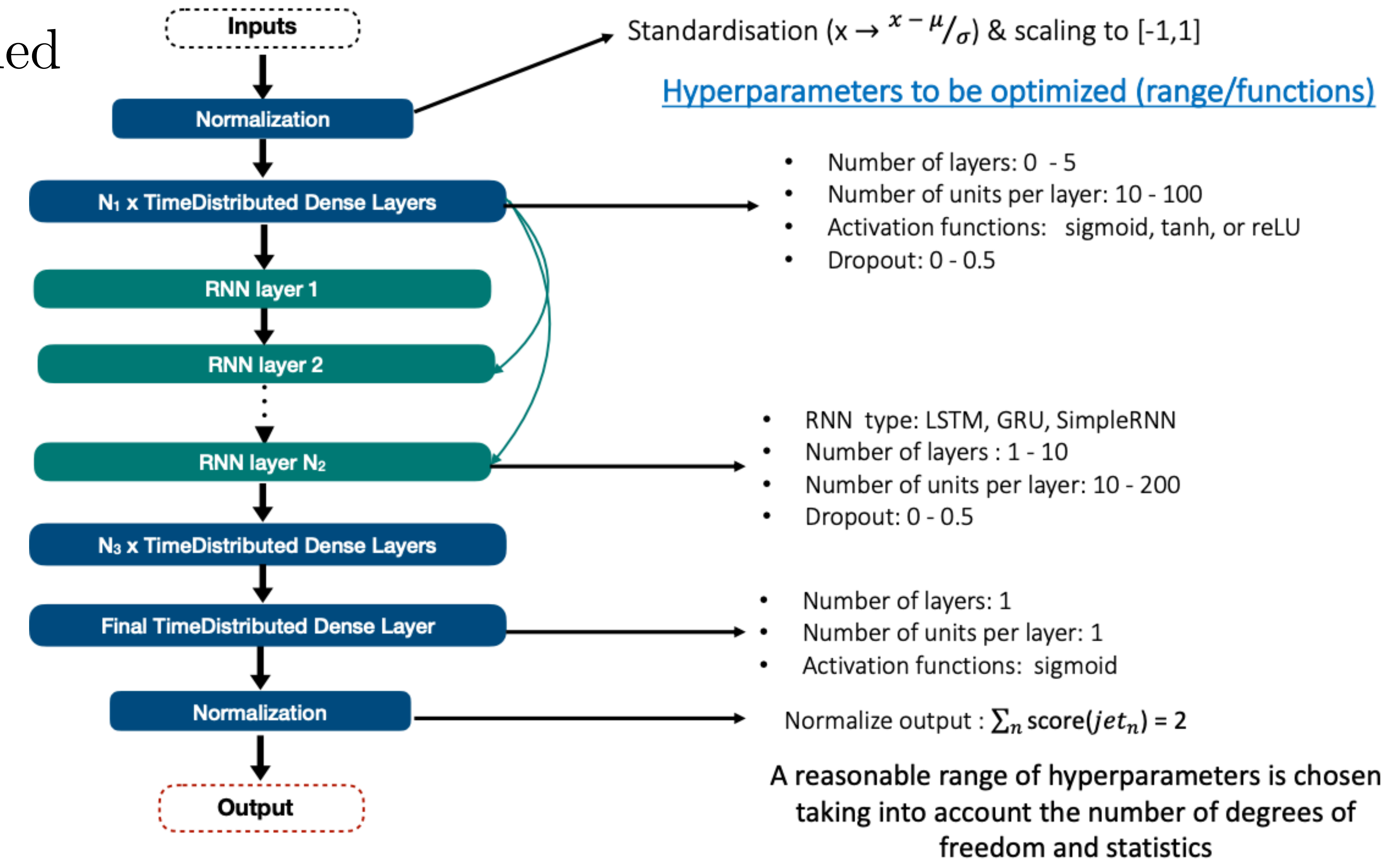
Pair Assessment:

- **HH-bTag**: NN-based algorithm fed with properties of all potential b jet candidates (including kinematic variables and DeepJet score), training and test with all available resonant and non-resonant Run2 samples
- the two b jet candidates with highest HH-btag score are selected

The H → bb candidate selection

- Jets coming from a b-quark hadronization process (**b-jets**) are tagged using **DeepJet** algorithm, which exploits a recurrent neural network (RNN) to tag jet flavours JINST 15 (2020) P12012
- To improve the $H \to bb$ selection the HH-btag algorithm has been defined
  - → Based on a RNN
  - → Inputs
    - b-jet candidates kinematic observables
    - DeepJet score
    - Angular separation between b-jet candidates and the $H \to \tau\tau$ candidate
- Algorithmic efficiency for $H \to bb$ tagging $\approx 95\,\%$
- m(bb) resolution improved by $\sim 25\,\%$

HH-btag architecture



For each event, all possible b-jet candidates are assigned a **HH-btag** score
The two jets with the highest scores are taken to build the $H \to bb$ candidate

HH Candidates selection

Elliptical cut requirement on m_{bb} VS $m_{\tau\tau}$ expected values

- Remove significantly outlying background events where no signal is expected
  - → Discrimination of HH events from the background left to a specifically designed neural network (see next slides)
- Provide additional control regions (by inverting mass cut)
- No invariant mass requirement for VBF categories

Parameters are defined by minimising background acceptance keeping signal efficiency above 90%

Resolved categories $$\frac{(m_{\tau\tau} - 129\ \mathrm{GeV})^2}{(53\ \mathrm{GeV})^2} + \frac{(m_{bb} - 169\ \mathrm{GeV})^2}{(145\ \mathrm{GeV})^2} < 1$$

Boosted categories $$\frac{(m_{\tau\tau} - 128\ \mathrm{GeV})^2}{(60\ \mathrm{GeV})^2} + \frac{(m_{bb} - 159\ \mathrm{GeV})^2}{(145\ \mathrm{GeV})^2} < 1$$

ggF signal extraction

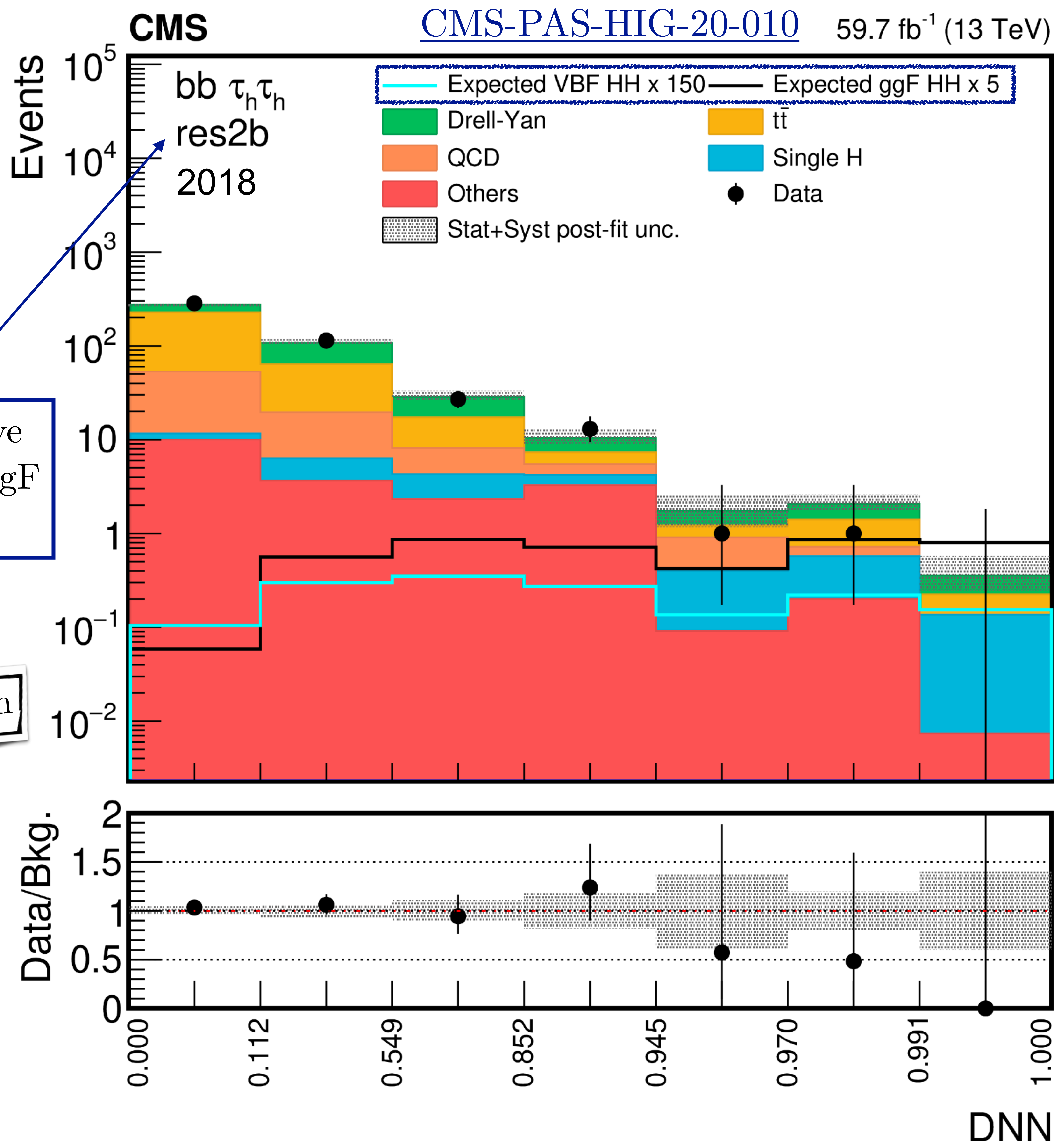
Dedicated Deep Neural Network (DNN) trained to separate SM non-resonant HH signal vs all backgrounds

- All channels and categories are considered in the training together
- Input features selected from a starting pool of > 100 features

Input Features	
Continuous	Categorical
b-tag score of 1st b-jet	Event is *boosted* or not
m_{HH} kinematic fit	Presence of VBF-candidates
χ^2 kinematic fit	$\tau\tau$ decay mode
$m_{\tau\tau}^{SVfit}$	Highest b-tag WP of 1st b-jet
$\Delta R(\tau,\tau)\cdot p_T(H_{\tau\tau}^{SVfit})$	Highest b-tag WP of 2nd b-jet
$\Delta R(\tau,\tau)$	Year
m_T and p_T of both taus	
$\Delta\phi(H_{\tau\tau}^{SVfit}, MET)$	
m_{bb}	
$\Delta\phi(H_{\tau\tau}^{SVfit}, H_{bb})$	
$p_T(H_{bb})$	

Post fit distribution

Most sensitive category for ggF search



Binned DNN score distribution as final discrimination for signal extraction in all analysis categories

VBF signal extraction

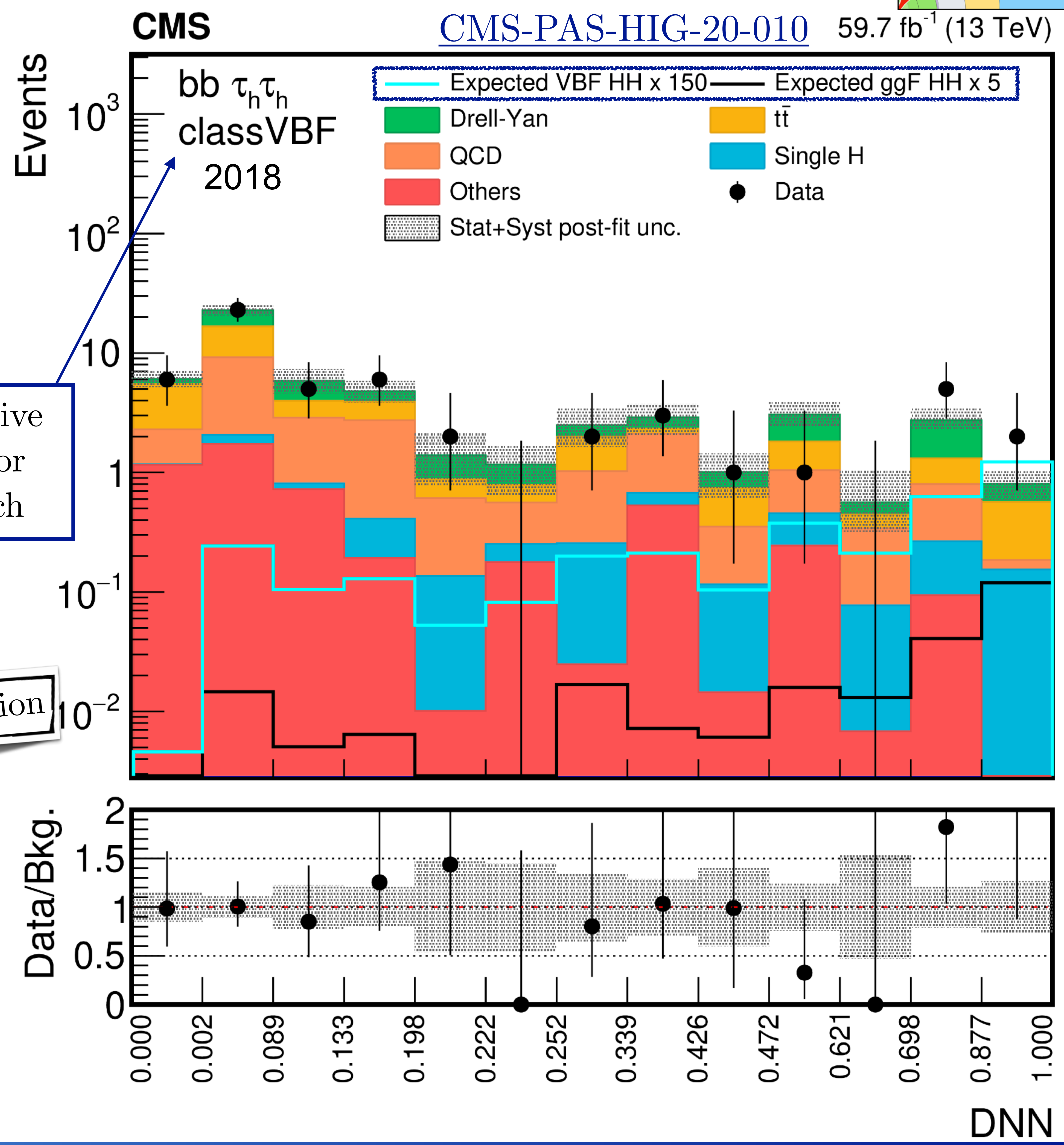
Dedicated Deep Neural Network (DNN) trained to separate SM non-resonant HH signal vs all backgrounds

- All channels and categories are considered in the training together
- Input features selected from a starting pool of > 100 features

Input Features	
Continuous	Categorical
b-tag score of 1^{st} b-jet	Event is *boosted* or not
m_{HH} kinematic fit	Presence of VBF-candidates
χ^2 kinematic fit	$\tau\tau$ decay mode
$m_{\tau\tau}^{SVfit}$	Highest b-tag WP of 1^{st} b-jet
$\Delta R(\tau,\tau) \cdot p_T(H_{\tau\tau}^{SVfit})$	Highest b-tag WP of 2^{nd} b-jet
$\Delta R(\tau,\tau)$	Year
m_T and p_T of both taus	
$\Delta\phi(H_{\tau\tau}^{SVfit}, MET)$	
m_{bb}	
$\Delta\phi(H_{\tau\tau}^{SVfit}, H_{bb})$	
$p_T(H_{bb})$	

Most sensitive category for VBF search

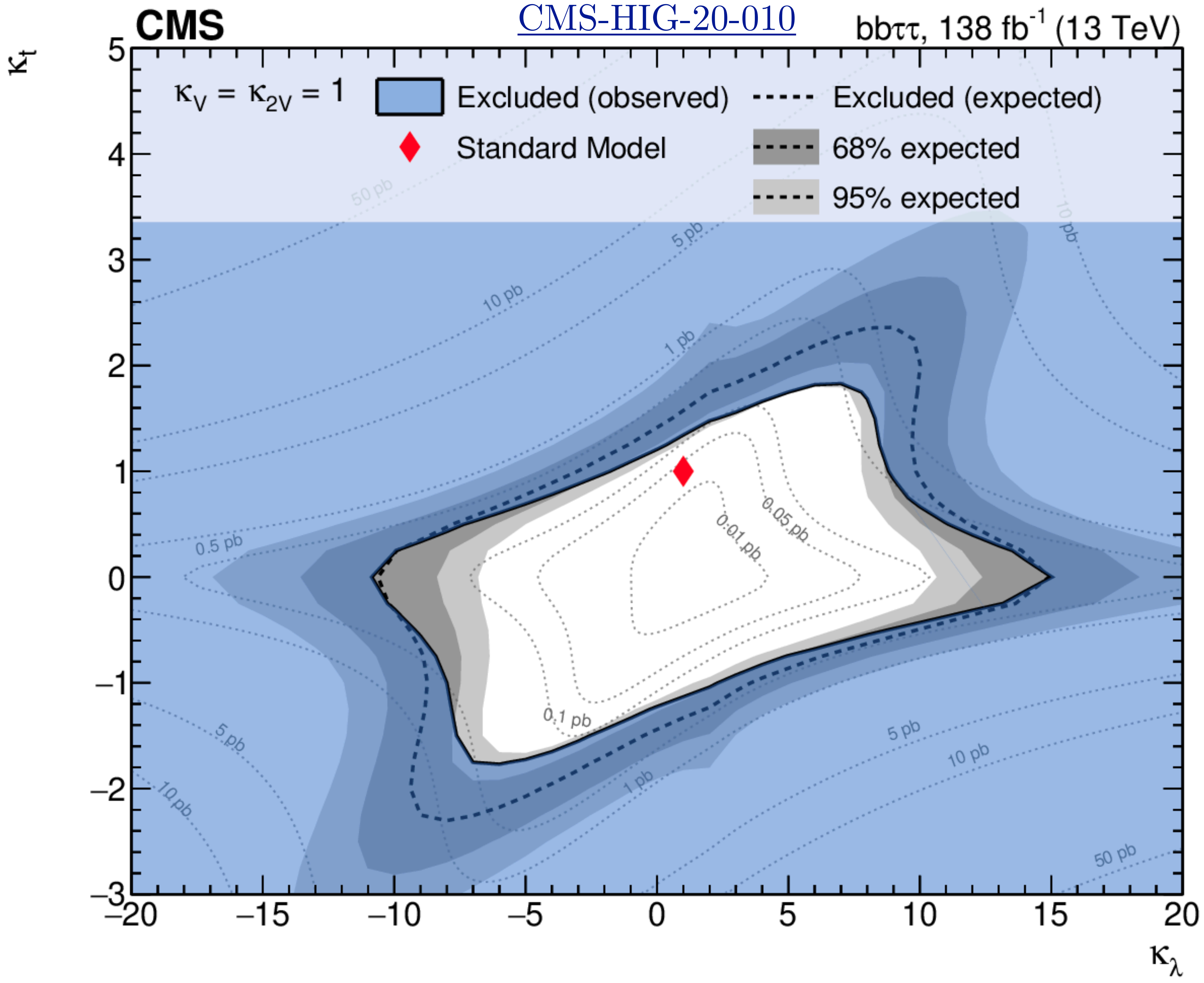
Post fit distribution



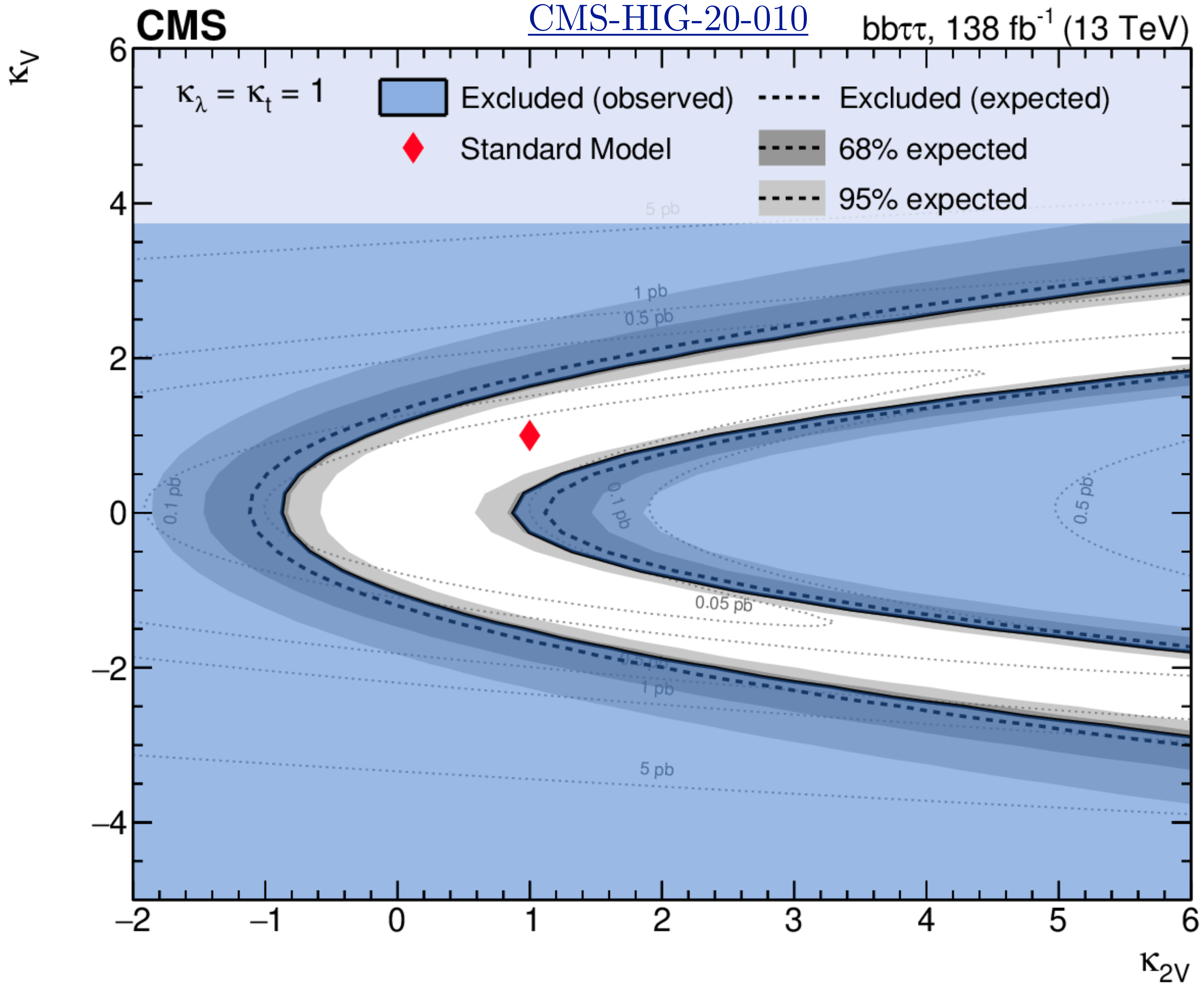
Binned DNN score distribution as final discrimination for signal extraction in all analysis categories

Exclusion regions

2D exclusion region in κ_λ and κ_t

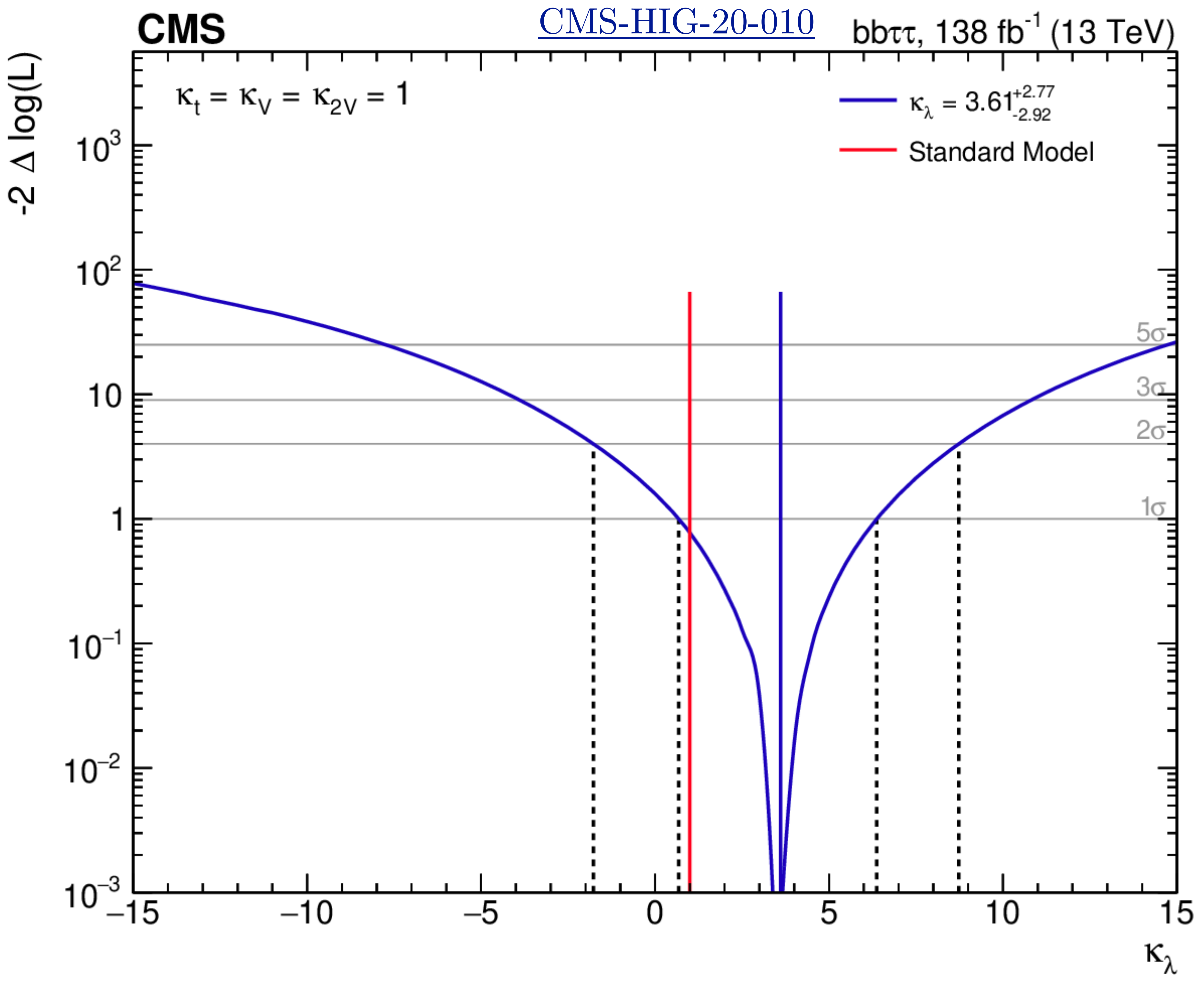


2D exclusion region in κ_V and κ_{2V}

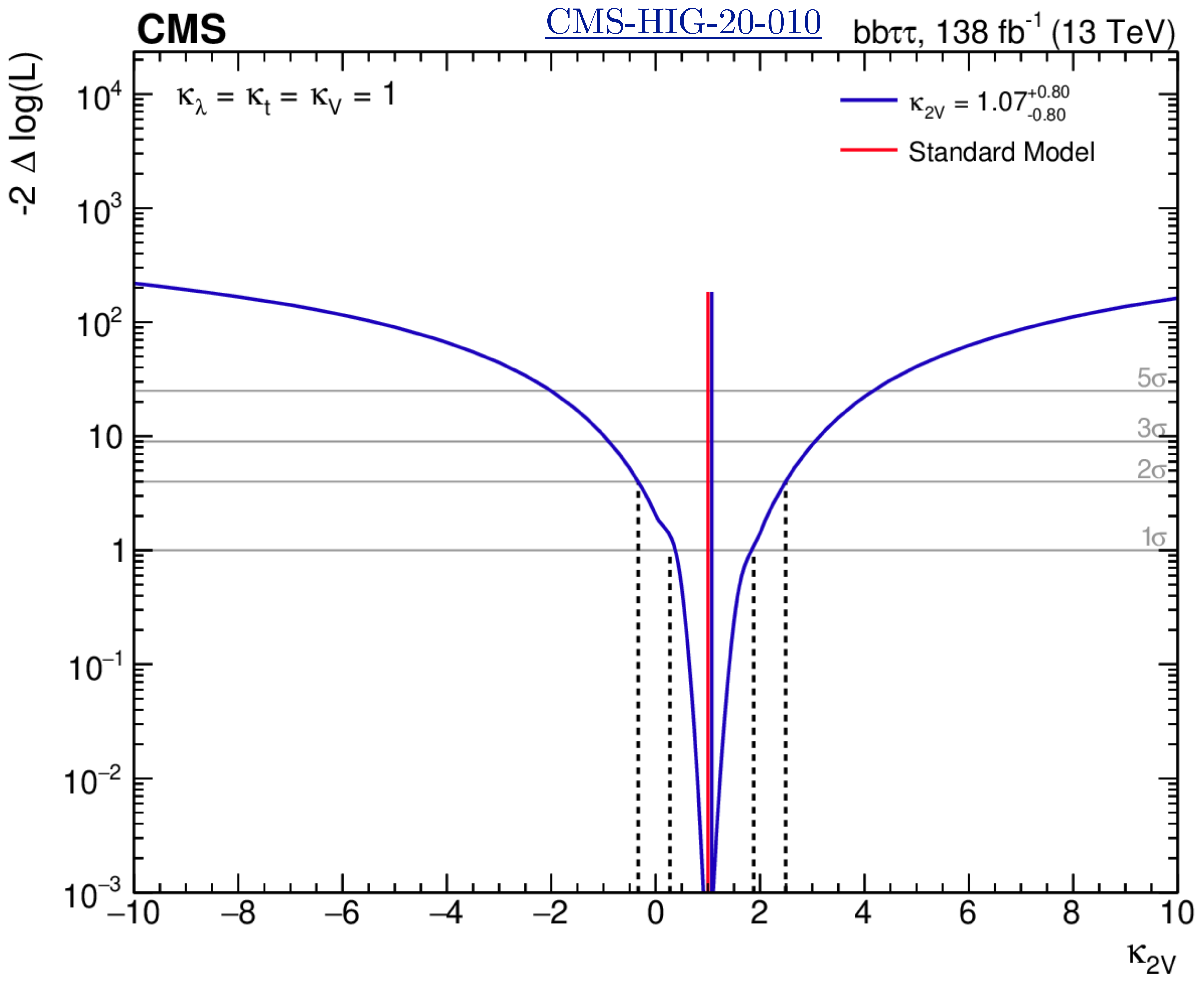


Likelihood scans

Observed likelihood scan as a function of κ_λ



Observed likelihood scan as a function of κ_{2V}



Normalisation uncertainties

Pile-Up reweighing ($\sim 1\,\%$)
inefficiency in vertex reconstruction; up-down variation of PU weights.

L1 Ecal prefiring ($\sim 2\,\%$)
in 2016 and 2017 samples, ECAL time shift not propagated to the L1Trigger primitives. Not described in MC simulation.
Uncertainty estimated via a tool provided by the BTV POG.

Luminosity
Prodvided by the Lumi POG, study of detector stability during data taking. It's applied to processes that rely ONLY on MC samples.
Its value depends on year - sample.

Cross sections (W+Jets, ST, SH, di- and tri-boson)
Due to imperfect knowledge of process normalisations/simulation

ttSF
The uncertainty is retrieved from the error coming with the maximum likelihood fit result in a CR and varies for each year.
All values are $< 1\,\%$

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

Theoretical Cross section
Included when evaluating the limit in the SM scenario.
error on $\sigma(GGF) :^{+2.2\%}_{-5\%}$ (scale), $\pm 3\,\%$ (PDF + α_s) $\pm 2.6\,\%$ (m_{top})
error on $\sigma(VBF) :^{+0.03\%}_{-0.04\%}$ (scale), $\pm 2.1\,\%$ (PDF + α_s)

DY estimation
18 uncertainties from the fit for DY estimation

QCD Normalization
On the correction factor adopted in the ABCD method to estimate the QCD Yield in signal region.

VBF dipole recoil
Due to imperfect modelling of third leading jet distribution in VBF samples. It's computed for each category, channel, year and the most conservative value is chosen.

Ele and Mu scale factors ($\sim 1\,\%$)
Data/MC disagreements in ID and reco.
Scale factor and uncertainties provided by the H$\tau\tau$ group.

Shape uncertainties

Pile-Up Jet ID
Differences on Data/MC PUJetID discriminator behaviour: scale factor parametrised in η and p_T and related uncertainty templates.
Provided by JET-MET POG.
Changes the overall weight of the event.

Tau Energy Scale
Uncertainty on τ energy distributions, provided by the T-POG. This induce a change of distribution in DNN scores.

Jet Faking taus
Uncertainty in events where jets are misidentified as hadronic taus: dedicated CR definition, estimation separately for Barrel and EndCap.

$$\delta_{jet-faking-tau} = \frac{Yield(data) - Yield(bkg)}{Yield(data)}$$

Trigger Scale Factors
Differences in Data/MC trigger efficiencies.
Uncertainties affect the overall weight of the event.
4 different uncertainties according to the decay mode of taus + other uncertainties applied for muTau and eTau channels + another uncertainty for VBF

DeepTau ID
Uncertainties related to scale factors for different discriminators (VSJet, VSEle, VSMu) due to different behaviour in Data/MC.
All uncertainties provided by TAU-POG

Jet Energy Reconstruction (JER)
Energy resolution in data worse than in MC: correction factor to be applied with up/down variation, provided by the JET-MET POG.

DY shape uncertainty
Obtained by summing in quadrature templates obtained by scaling the nominal contribution in different CRs.

B-tag
Uncertainties on b-tagging performances, accounting heavy/light jet flavour regions, contaminations, statistical fluctuation on MC samples.
Different types: HF; LF; CFErr1,2; HFStat1,2; LFStat1,2.

Fully hadronic sample custom SF
The $\tau_h\tau h$ 2017 distributions are in huge disagreement with data, mainly due to the DY contribution: a custom SF has been evaluated in the analysis and its uncertainty that change the event weight.

QCD shape uncertainty
Due to the region chosen for the shape estimation in SR

Jet Energy Scale
DProvided by JET-MET POG, 14 uncertainties to account for: PU contribution, non linearity detector response, residual data/MC difference

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 $-ln(\mathscr{L})$ (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.

Systematic uncertainties

Residual differences between data and MC simulation due to uncertainties on

- Theoretical predictions
- Unforeseen detector responses
- Limited size of MC samples
- data-driven methods
- Scale factors due to data-MC differences
- Different data-MC behaviour in tagging algorithms

Systematic uncertainties

Normalisation uncertainties

Affecting only the yield of a given process

MC Stat

Limited size of samples: Barlow-Beeston "light" method

Shape Uncertainties

Affect the shape of the final discriminating variable. Alternative shapes of the discriminating variable computed by varying the scale of the objects affected by each uncertainty.

Main sources of systematic uncertainties

- Theory uncertainty on ggF HH cross section: $\sigma_{SM}(pp \to HH) = 31.05^{+6\%}_{-23\%}$ (scale+$\mathrm{m_t}$) $\pm\, 3\,\%$ (PDF + α_S) fb
- Statistical fluctuations affecting multi-jet background estimation
- Uncertainties on mis-modelling of jet and tau ID and reconstruction in simulation

Total effect of systematic uncertainties on final limits is $\sim 15\,\%$