Searches for Higgs boson pair production in final states with two b quarks and two τ leptons and in multilepton final states at CMS

Jaime León Holgado (CIEMAT)

on behalf of the CMS Collaboration

Higgs 2022, November 9th, 2022

GOBIERNO DE ESPAÑA

MINISTERIO DE CIENCIA E INNOVACIÓN

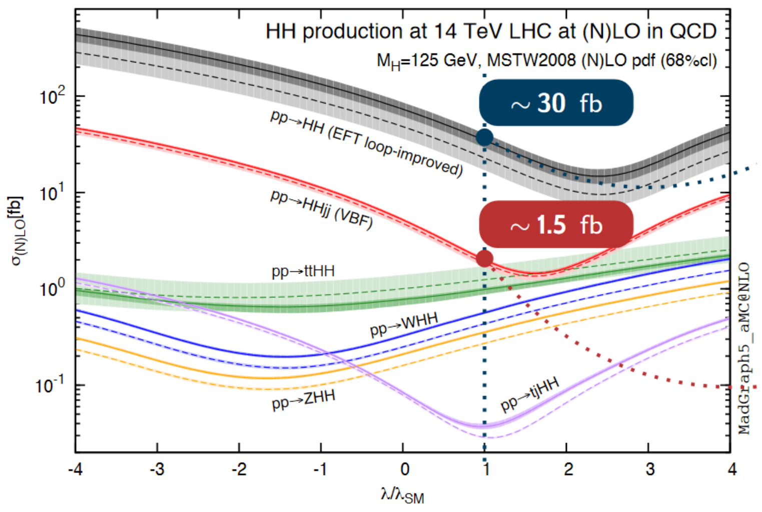
Ciemat

Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

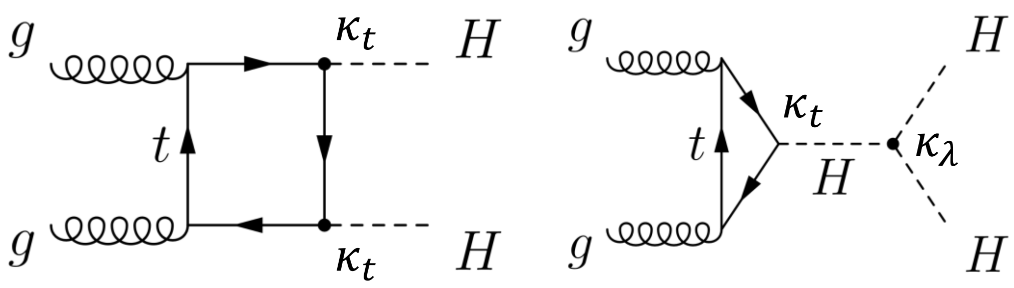
cfp CIEMAT física de partículas

CMS

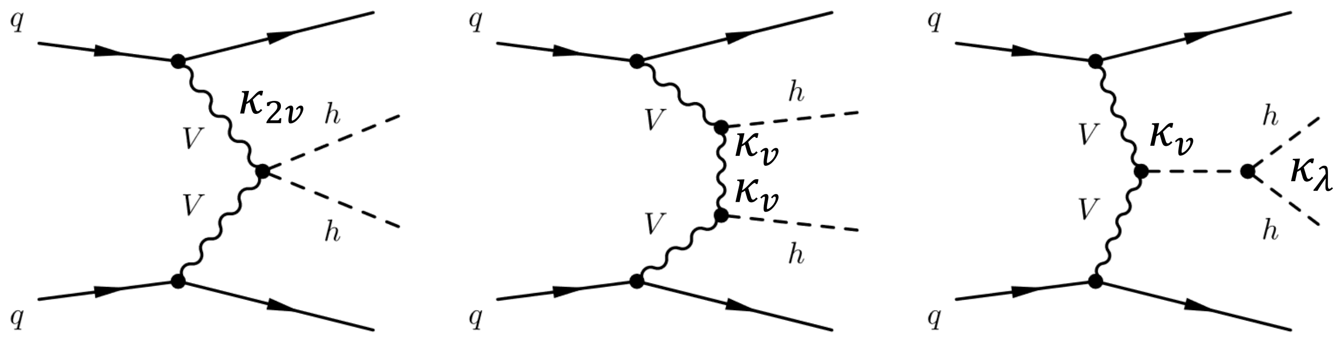
Search for Higgs Boson Pair Production at the LHC



- Several CMS analysis search for HH production using all Run 2 data ($138\ \text{fb}^{-1}$).
  - Both non-resonant and resonant productions are studied.
  - This talk focuses on the non-resonant production.
- Studying main production modes: Gluon and Vector Boson Fusion.
- Analysis objectives:
  - Obtain the value of σ_{HH} with respect to the predicted by the SM.
  - Study the values of the Higgs couplings involved.

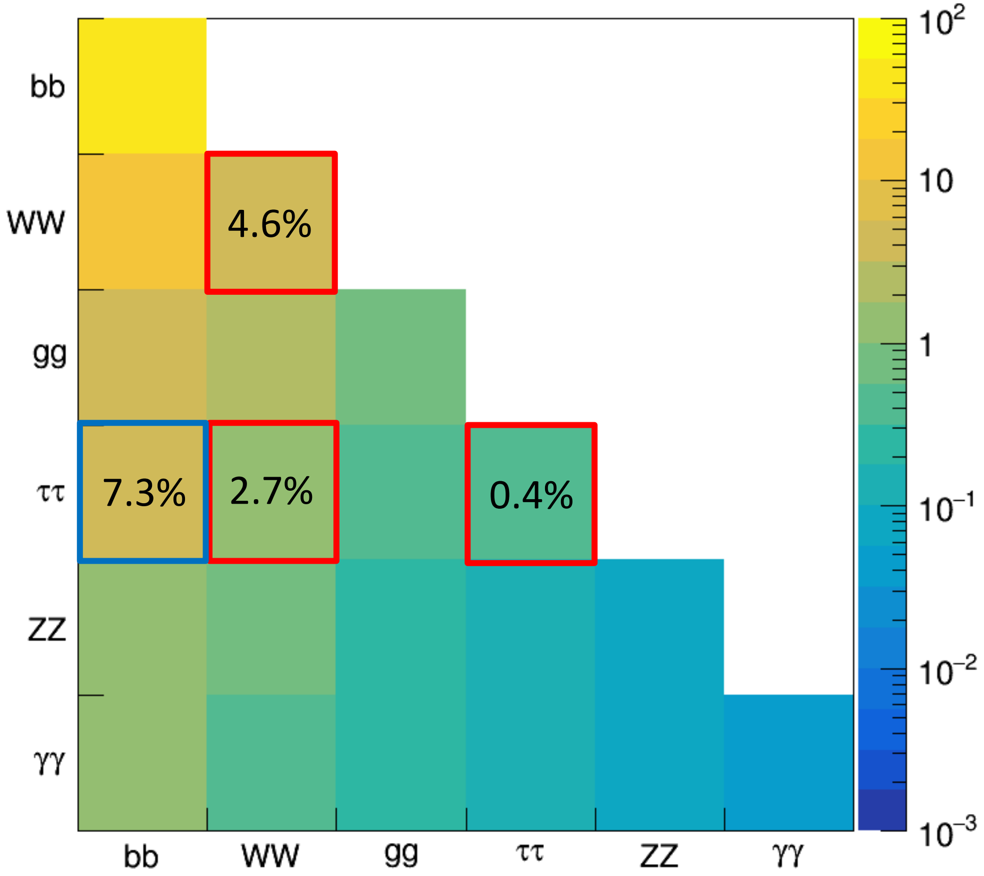


Gluon Fusion (ggF)



Vector Boson Fusion (VBF)

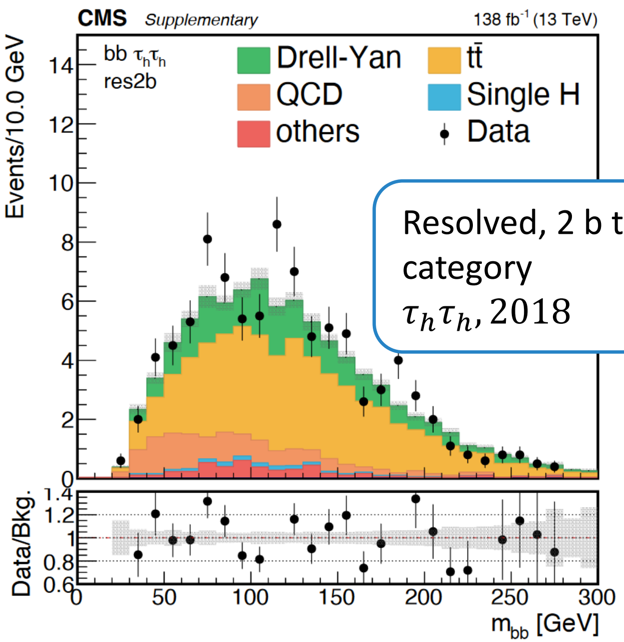
Search for Higgs Boson Pair Production at the LHC



- HH production has much smaller cross section than H production.
- Best decay modes require a balance between purity and branching ratio.
- This talk focuses in 4 decay modes:
  - $HH \rightarrow bb\tau\tau$
    - Large BR thanks to the $H \rightarrow bb$, good purity coming from $H \rightarrow \tau\tau$.
    - Considering $\tau_h\tau_h$, $\tau_e\tau_h$, $\tau_\mu\tau_h$ final states (88%)
    - arXiv:2206.09401, HIG-20-010
  - *Multilepton* (WWWW, WW$\tau\tau$, $\tau\tau\tau\tau$)
    - First search for HH decaying into WW$\tau\tau$ and $\tau\tau\tau\tau$.
    - Considering final states with at least $2l$ or 1 τ_h.
    - arXiv:2206.10268, HIG-21-002
  - Both analysis included in combination published in Nature.

$HH \rightarrow bb\tau\tau$

Datasets



Resolved, 2 b tag category
$\tau_h\tau_h$, 2018

DATA

138 fb^{-1} of pp collision data collected by the CMS detector at the LHC during Run 2 (2016-2018).

MAIN BACKGROUNDS

Drell-Yan

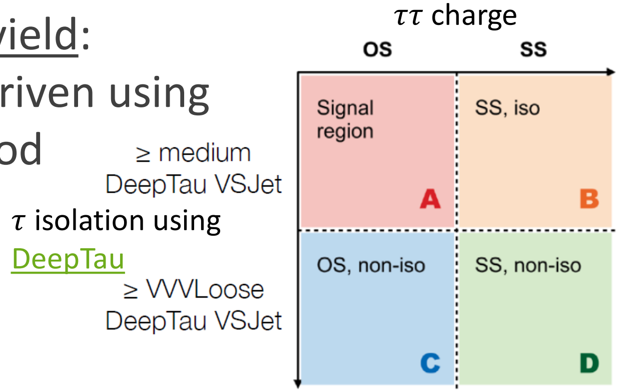
Basic estimation from MC, yield corrected by fitting in $Z \to \mu\mu$ control regions

$t\bar{t}$

Shape: MC
Yield: Data-driven from $t\bar{t}$ enriched control region

QCD

Shape and yield:
Fully data-driven using ABCD method



OTHER BACKGROUNDS

Purely MC based

W + jets	Single top (tW and t channels)
Di-boson	EWK Z and W
Tri-boson	ttV and ttVV
Single SM Higgs (ggF, VBF, VH, ttH)	

Analysis flow

Trigger requirements

e, μ, τ object selection
$H \to \tau\tau$ pair identification

b jets, VBF jets selection
$H \to bb$ pair identification
VBF pair identification

Event categorization

Signal extraction

Analysis flow

Trigger requirements

Targetting $H \rightarrow \tau\tau$ decay products:

- $\tau_\mu\tau_h$ channel: single muon, muon-τ_h cross trigger
- $\tau_e\tau_h$ channel: single electron, electron-τ_h cross trigger
- $\tau_h\tau_h$ channel: di-τ_h trigger, dedicated VBF $H \rightarrow \tau_h\tau_h$ trigger.

e, μ, τ object selection
$H \rightarrow \tau\tau$ pair identification

- Match offline to trigger objects
- Apply kinematic, isolation and identification requirements.

b jets, VBF jets selection
$H \rightarrow bb$ pair identification
VBF pair identification

Event categorization

Signal extraction

Analysis flow

Trigger requirements

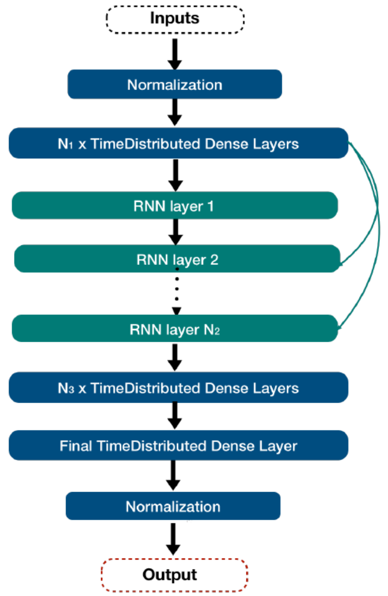
e, μ, τ object selection
$H \to \tau\tau$ pair identification

b jets, VBF jets selection
$H \to bb$ pair identification
VBF pair identification

Event categorization

Signal extraction

- New ML-based b jet candidates' selection: **HH-btag**
  - Using as input information the kinematic of the event (including information about the $H \to \tau\tau$ candidate) and the jet b-tag score (DeepJet).
  - Improving $H \to bb$ selection efficiency and mass resolution w.r.t. CMS standard b-tagging.
- Select VBF jet candidates (highest m_{jj} pair).



Analysis flow

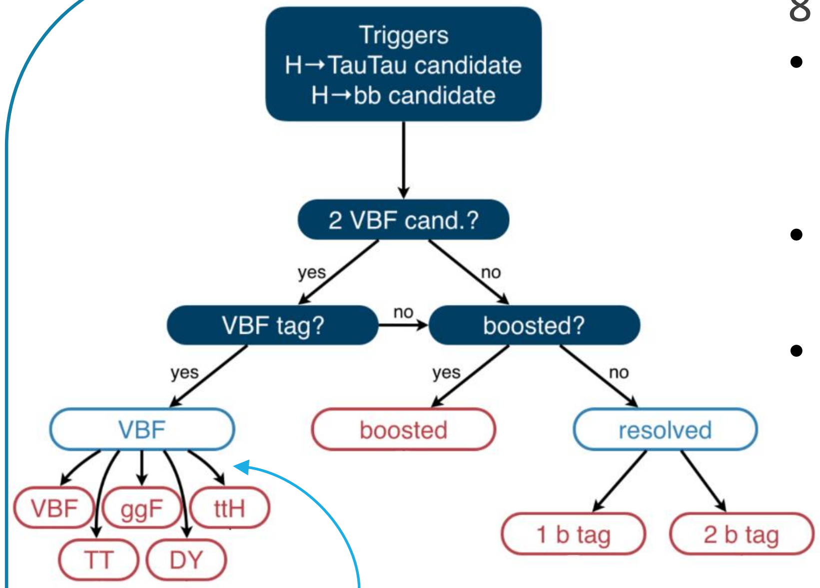
Trigger requirements

e, μ, τ object selection
$H \rightarrow \tau\tau$ pair identification

b jets, VBF jets selection
$H \rightarrow bb$ pair identification
VBF pair identification

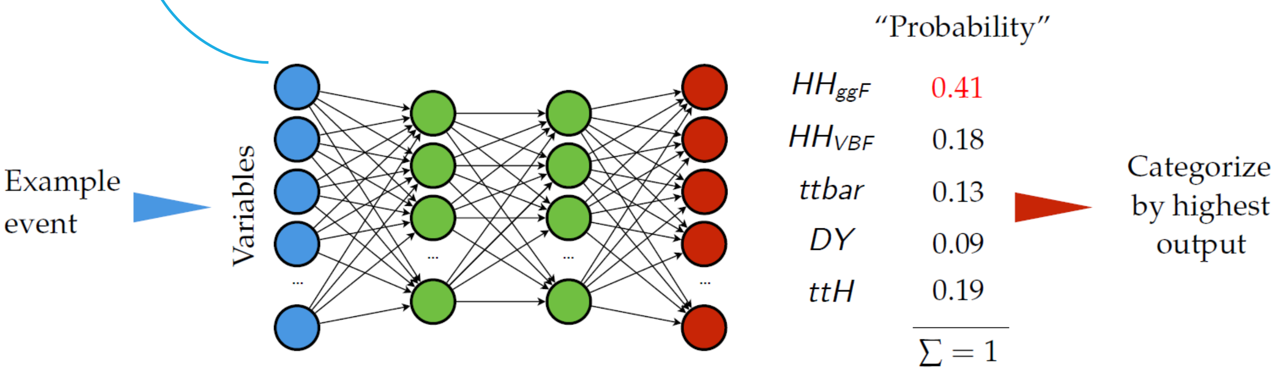
Event categorization

Signal extraction



8 final analysis categories

- Boosted category: Requiring a fat jet with 2 loose b-tag subjets matched to the selected b jets.
- Resolved 1 b tag and 2 b tag: one or two jets passing medium b-tag WP.
- VBF categories: $m_{jj} > 500$ GeV and $|\Delta\eta_{jj}| > 3$, medium b-tag requirement on at least one b jet candidate. Five subcategories identified using a multi-classifier.



Analysis flow

Trigger requirements

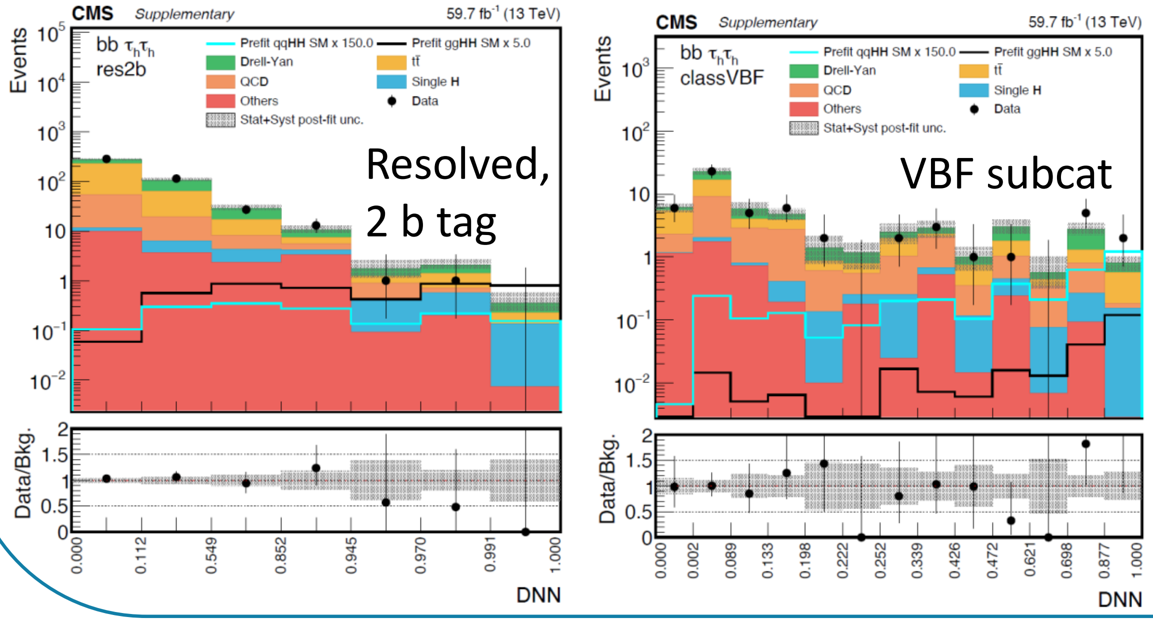
e, μ, τ object selection
$H \to \tau\tau$ pair identification

b jets, VBF jets selection
$H \to bb$ pair identification
VBF pair identification

Event categorization

Signal extraction

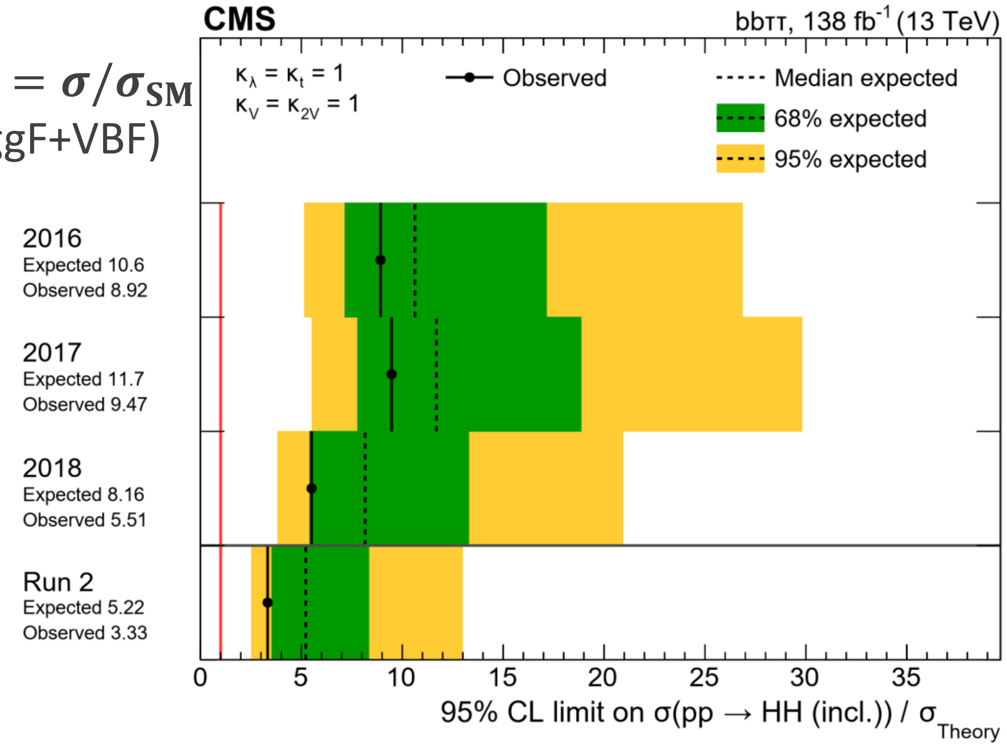
- Using a dedicated ML discriminator that provides a single prediction to discriminate between signal and background.
- Trained with boosted and resolved events and considering as signal only the ggF samples.
- Binned shape of the output score used in all analysis categories.

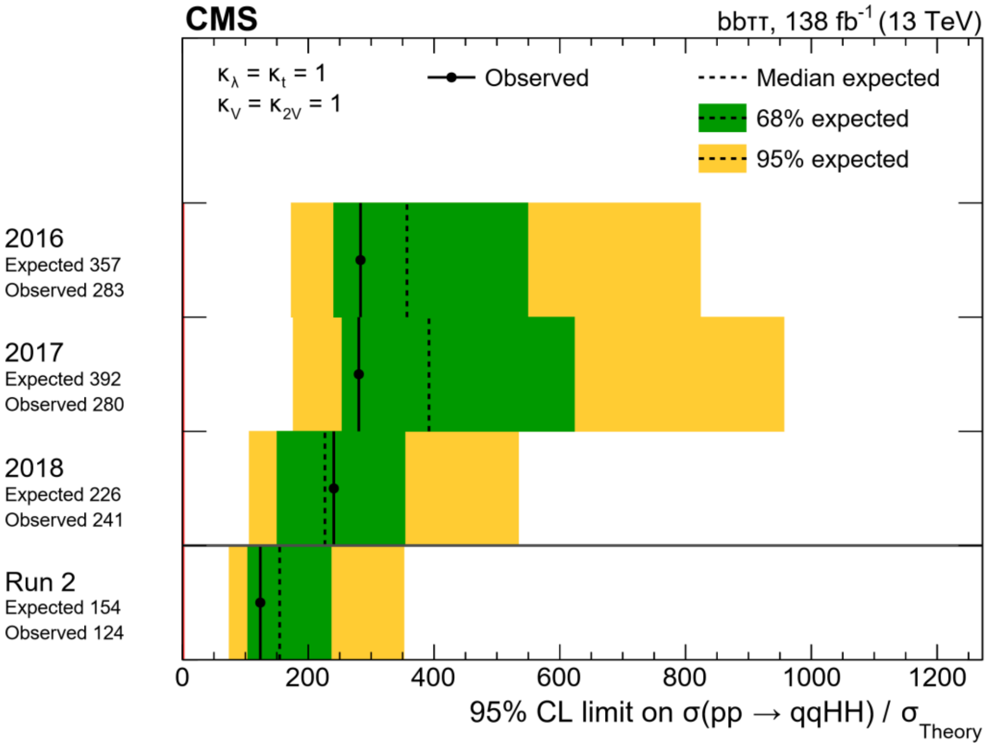


Resolved 2 b tag and VBF subcategory $\tau_h\tau_h$, 2018

Results: Upper limits @SM

$r = \sigma/\sigma_{SM}$
(ggF+VBF)





$r_{VBF} = \sigma_{VBF}/\sigma_{SM}$
(VBF only, assumming $r_{ggF} = 1$)

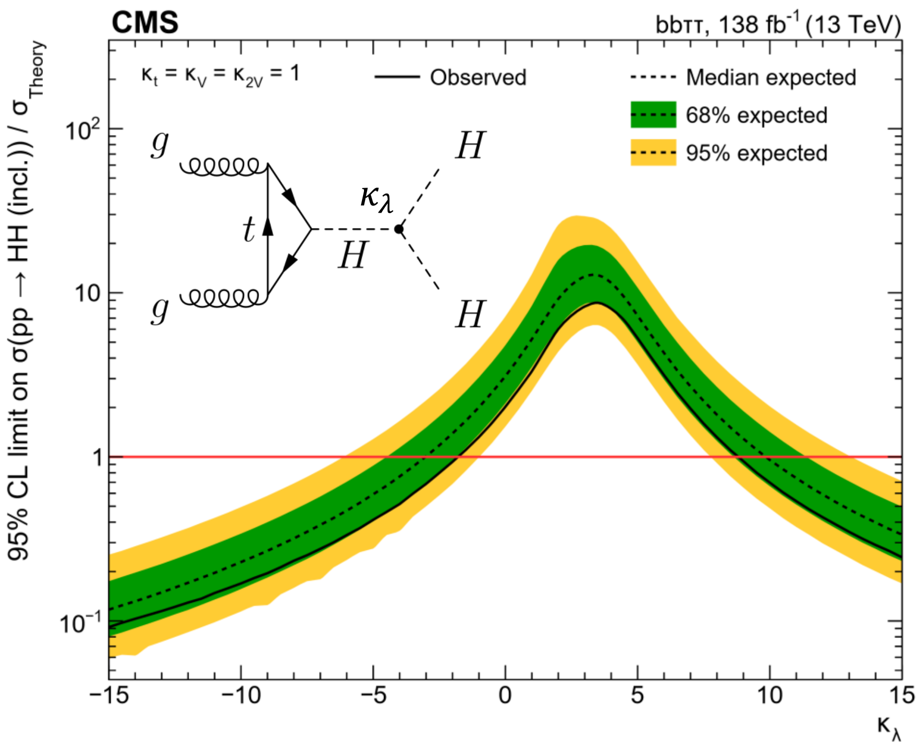
Obs (exp) upper limit on inclusive HH production:
3.3 (5.2) $\times\ \sigma_{SM}$

Obs (exp) upper limit on VBF HH production:
124 (154) $\times\ \sigma_{SM}$

- ggF + VBF: second most sensitive CMS HH analysis.
- VBF: most sensitive CMS HH analysis.

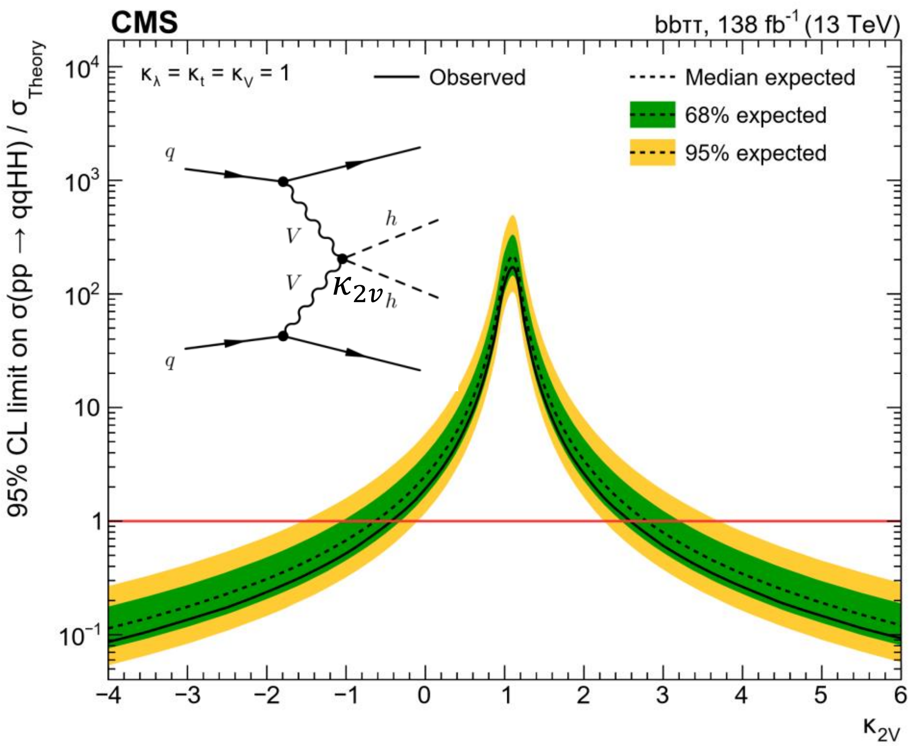
Results: Inclusive HH upper limits vs. κ_λ and κ_{2v}

Limits on r



- Observed: $\kappa_\lambda \in [-1.8, 8.8]$
- Expected: $\kappa_\lambda \in [-3, 9.9]$

Limits on r_{qqHH}



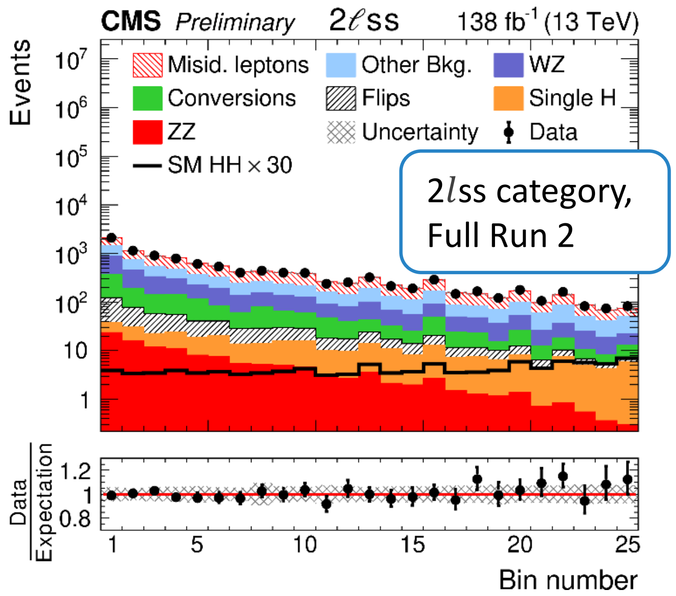
- Observed: $\kappa_{2v} \in [-0.4, 2.6]$
- Expected: $\kappa_{2v} \in [-0.6, 2.8]$

$HH \rightarrow Multilepton$

Datasets

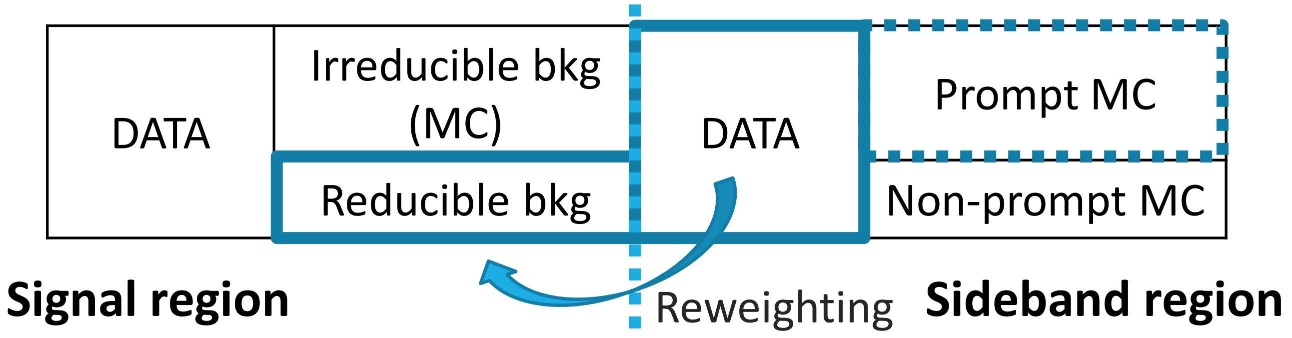
DATA

138 fb^{-1} of pp collision data collected by the CMS detector at the LHC during Run 2 (2016-2018).



BACKGROUNDS

- **Irreducible backgrounds**: l and τ_h come from W, Z or H decays and are reconstructed with the correct charge. Extracted from MC.
- **Reducible backgrounds**: l/τ_h misidentification, electron charge misidentification, electron from photon conversion.
  - Huge cross-sections → Cannot produce enough MC events, need to extract from data.
  - Using the “fake factor” method: defining a relaxed selection (charge/iso) region. Data events entering in this sideband region are extrapolated to the SR by reweighting them.



Analysis flow

Event categorization

Trigger requirements and object selection

Signal extraction

Analysis flow

- Event categorization
- Trigger requirements and object selection
- Signal extraction

- Targetting W and τ_h decay products. Seven mutually exclusive categories included, each with its dedicated triggers.
  - $2l$ss (*same-sign*): single- and double-lepton.
  - $3l$: single-, double- and triple-lepton.
  - $4l$: single-, double- and triple-lepton.
  - $3l + 1\tau_h$: single-, double- and triple-lepton.
  - $2l + 2\tau_h$: single- and double-lepton.
  - $1l + 3\tau_h$: single-lepton, lepton+τ_h and double-τ_h.
  - $4\tau_h$: double-τ_h.
- l and τ_h candidates required to pass a tight selection criteria and category-specific p_T thresholds motivated by the triggers.
- Background reduced by rejecting events with Z or light meson resonances, with b tagged jets, H→ZZ →$4l$ events.

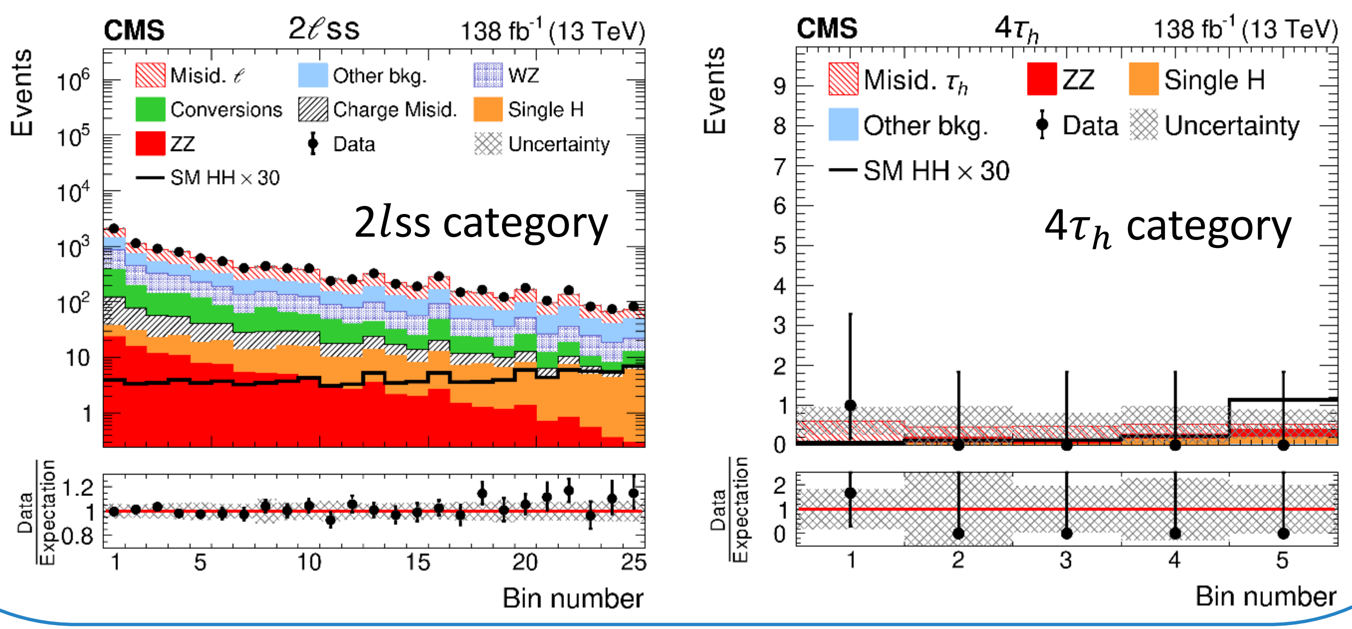
Analysis flow

Event categorization

Trigger requirements and object selection

Signal extraction

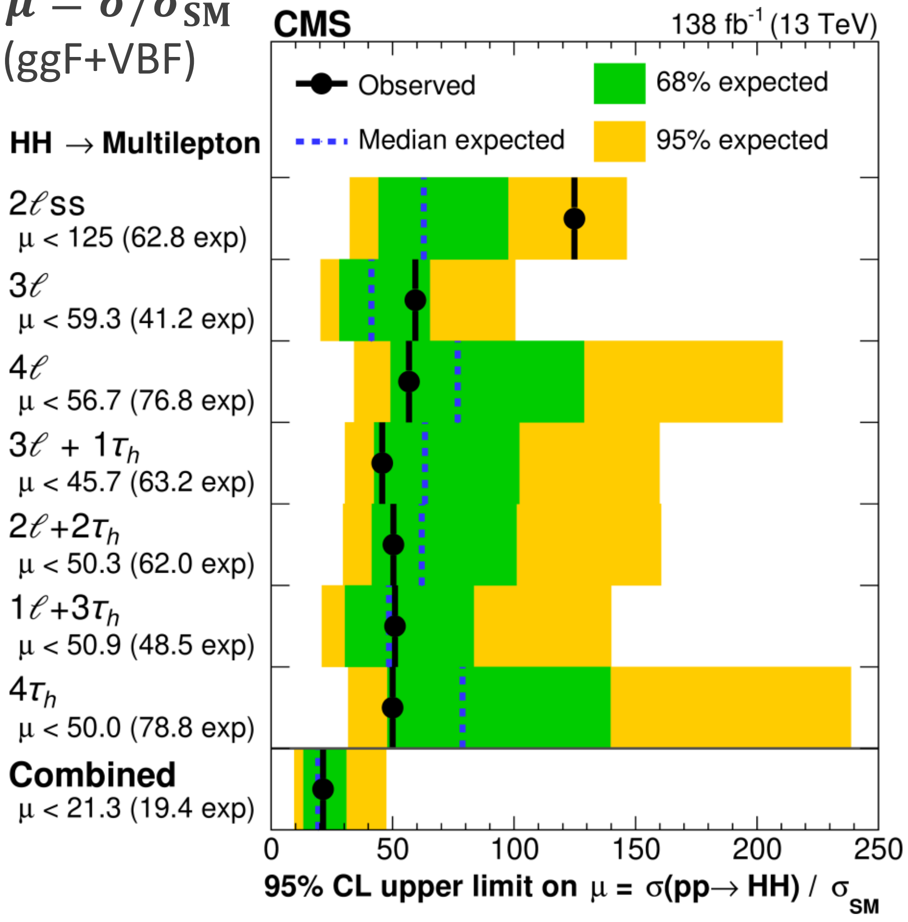
- Dedicated boosted decision trees (BDT) for each category used to discriminate between HH signal events and backgrounds.
- Trained using the three years at once.
- Maximizing MC statistics by relaxing object selections.
- Binning chosen so each bin has similar expected background/signal events (in high/low event yield categories). Higher bin numbers correspond to high BDT output values and higher signal to background ratio.



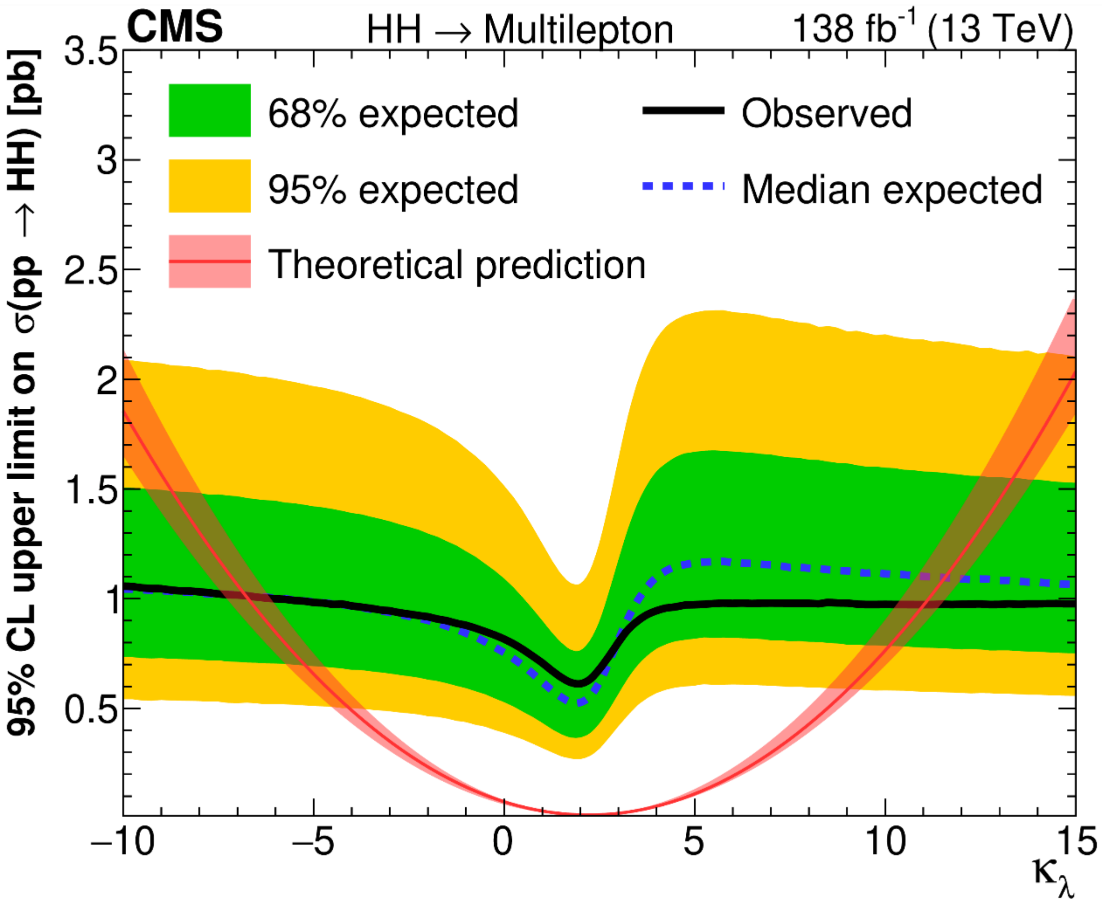
Results

Dedicated VBF search to be done in Run 3

$\mu = \sigma/\sigma_{SM}$
(ggF+VBF)



Obs (exp) upper limit on inclusive HH production:
21.3 (19.4) × σ_{SM}



- Observed: $\kappa_\lambda \in [-6.9, 11.1]$
- Expected: $\kappa_\lambda \in [-6.9, 11.7]$

Conclusions

- Searches for non-resonant HH production in the $bb\tau\tau$ and multilepton final states have been presented.
  - Using the full Run 2 data set (138 fb^{-1}).
  - Studying the main production modes: ggF and VBF.

$HH \to bb\tau\tau$

- Considerable improvement compared to the previous results (upper limit on $\sigma \sim 30(25) \times \sigma_{SM}$) due to increased statistics and introduction of advanced Machine Learning techniques at analysis and object identification levels.
- Observed (expected) 95% CL upper limit for the SM point:
  - $\sigma_{\text{ggF+VBF}} = 3.3\ (5.2) \times \sigma_{\text{ggF+VBF}}^{\text{SM}}$
  - $\sigma_{\text{VBF}} = 124\ (154) \times \sigma_{\text{VBF}}^{\text{SM}}$

$HH \to multilepton$

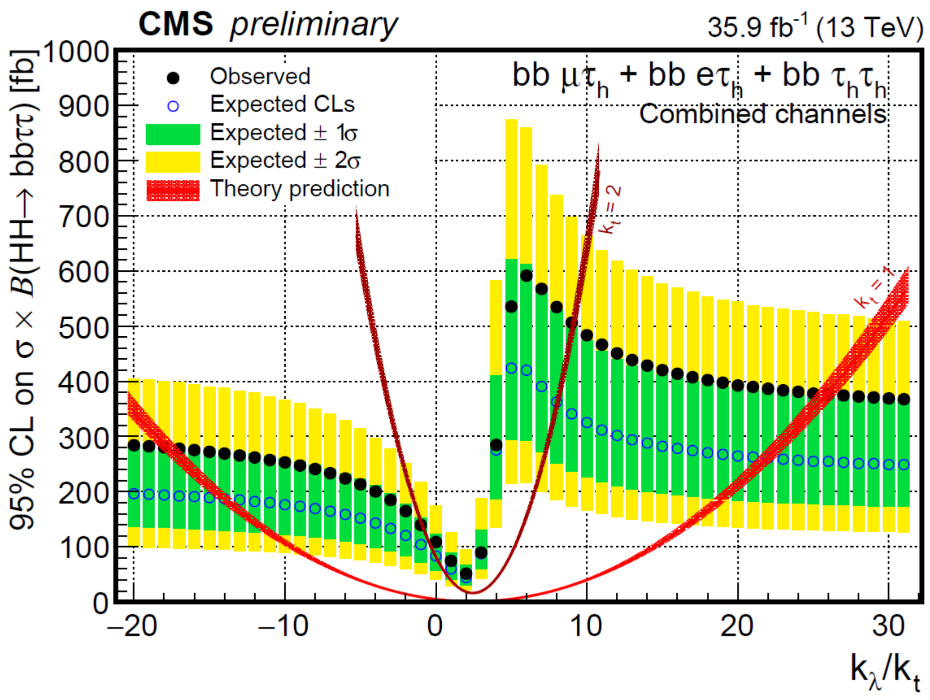
- First analysis to consider the WW$\tau\tau$ and $\tau\tau\tau\tau$ decay modes.
- Observed (expected) 95% CL upper limit for the SM point:
  - $\sigma_{\text{ggF+VBF}} = 21.3\ (19.4) \times \sigma_{\text{ggF+VBF}}^{\text{SM}}$

BACKUP

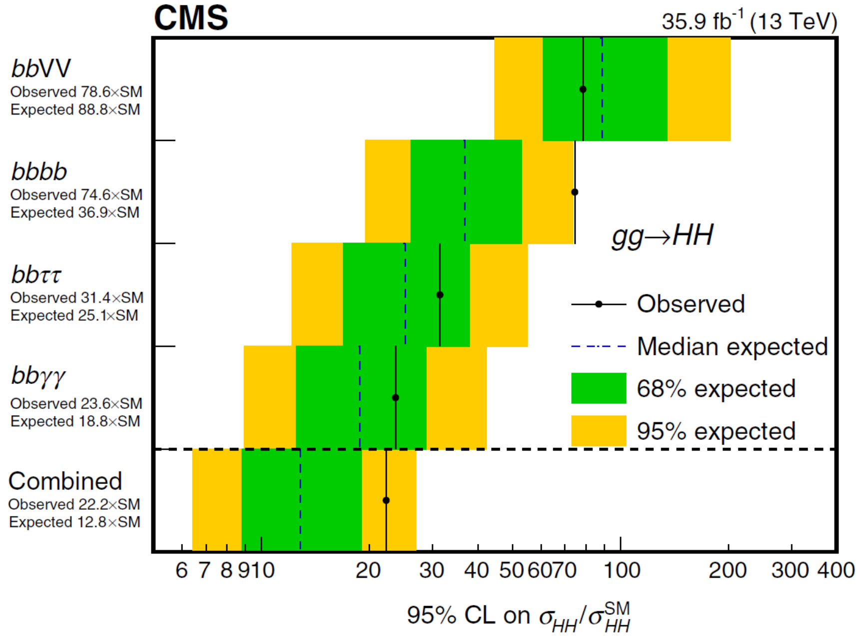
Signal samples

- ggF:
  - LO: MADGRAPH5 aMC@NLO v2.4.2
  - NLO: MADGRAPH5 aMC@NLO v2.4.2 + POWHEG v2.0
- VBF:
  - LO: MADGRAPH5 aMC@NLO v2.3.2.2

Latest $HH \rightarrow bb\tau\tau$ results (2016 dataset)



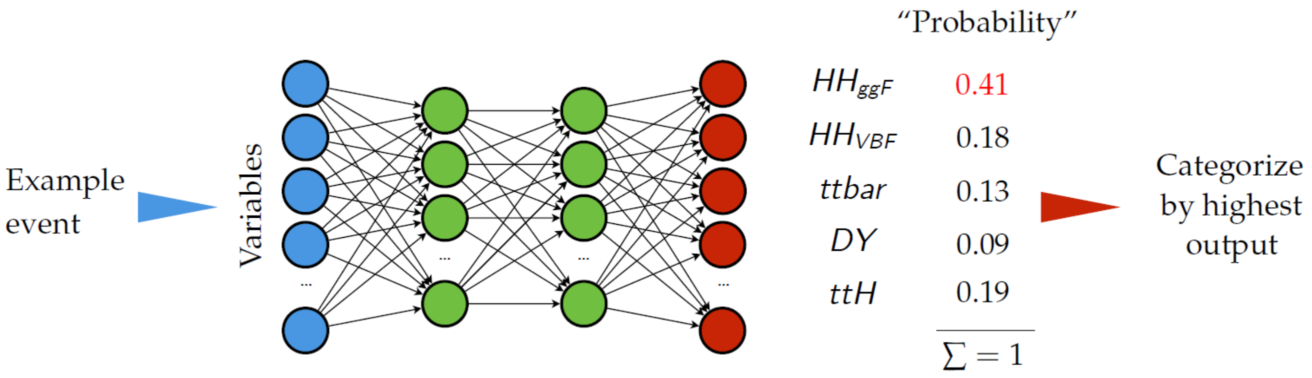
- Most recent $bb\tau\tau$ result: PLB 778 (2018) 101
- Using 2016 data: 35.0 fb^{-1}
- 95% CL upper limit on $\sigma(gg \rightarrow HH)$ obs (exp) ~30 (25) × σ_{SM}



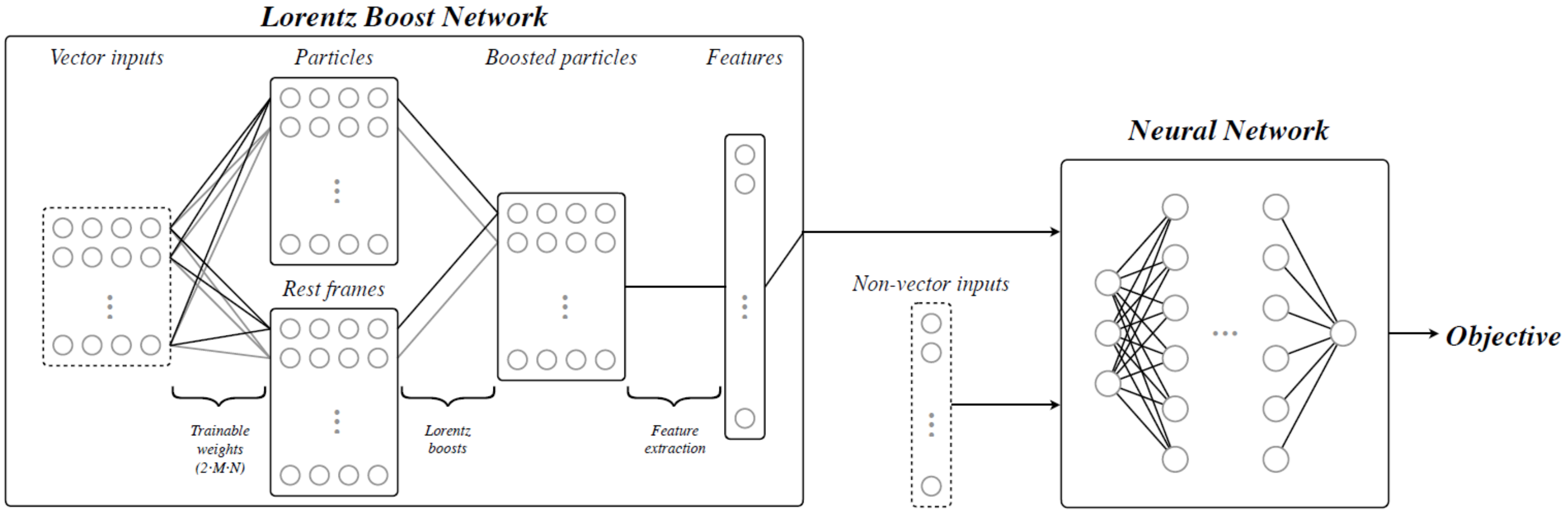
- Combination of 2016 HH analysis: Phys. Rev. Lett. 122, 121803
- The $bb\tau\tau$ channel is the second most sensitive HH channel.

Multi-class classification for VBF categorization

- In order to separate the VBF signal from the ggF and other background contamination, we classify the events in this category into 5 subcategories using a multi-class classification.
- In this strategy, machine learning techniques are used to assign probability estimates for an event to belong to categories associated to any of the relevant physics processes under consideration.



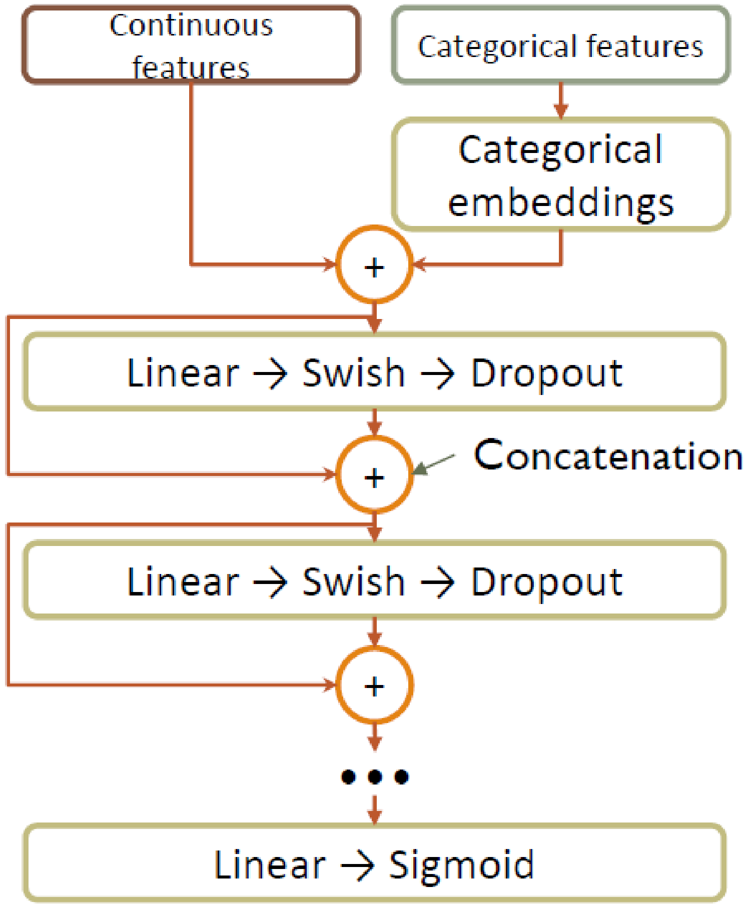
LBN takes *N* input four-vectors, creates *M* particles, boosts them to their rest frames and extracts generic features from them and their combinations.



DNN for signal extraction

- A dedicated ML discriminator that provides a binary signal vs background classification is trained.
- Trained with events belonging to the resolved and boosted categories and considering as signal only the ggF LO sample.
- Each discriminator is an ensemble of ten neural networks trained via ten-fold cross validation.
  - The final score is a weighted sum of the DNN scores from the ensemble
- Binned shape of the output score is used for the final signal extraction in all analysis categories.

DNN architecture



DNN for signal extraction

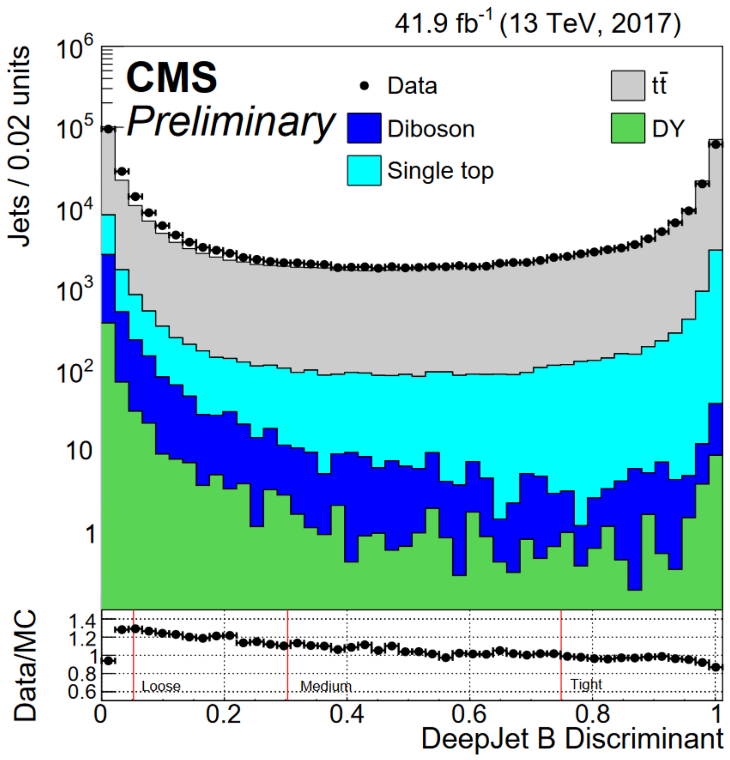
- An extensive set of high- and low-level features is considered as possible inputs.
- The most influential inputs are selected using a series of steps:
  - Cleaning of linearly correlated features
  - Iterative cleaning based on feature importance using Random Forest classifiers.
  - Pruning of features with a high mutual dependency among the remaining ones.
- The final NN ensemble is trained on 20 continuous and 6 categorical features.

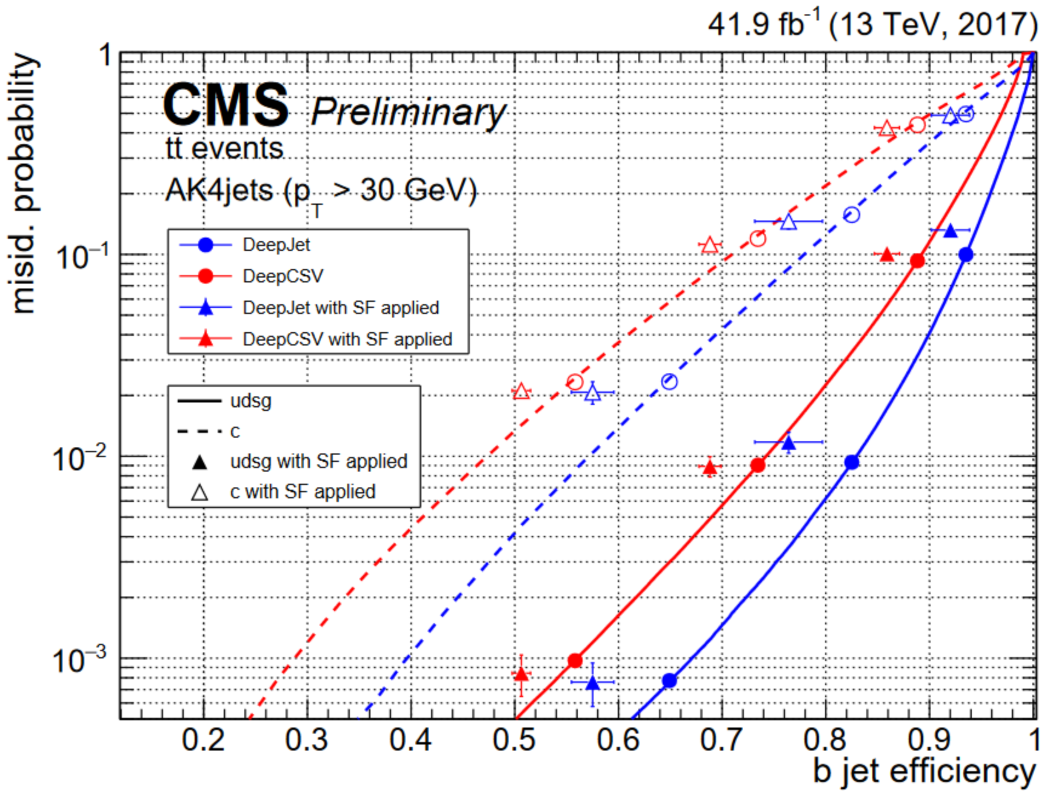
Some of the most important continuous features
$M(HH)$ (KinFit)
χ^2 (KinFit)
$M(\tau\tau)$ (SVFit)
$\Delta R(\tau,\tau) \cdot p_t(H(\tau\tau)(\text{SVFit}))$
$M_T(\tau_1, MET)$
Deepflavour CvsB of the first b jet
…

Categorical features
Channel
Year
Boosted flag
VBF jets presence
Highest deepflavour WP from the first b
Highest deepflavour WP from the second b

b-tagging – CMS DeepJet

- https://cds.cern.ch/record/2646773/files/DP2018_058.pdf





Results: Upper limits @SM

ATLAS RESULT **r** (ggF+VBF) SM only!

Table 5: Observed and expected upper limits at 95% CL on the cross-section of non-resonant HH production according to SM-like kinematics, and on the cross-section of non-resonant HH production divided by the SM prediction. The $\pm 1\ \sigma$ and $\pm 2\ \sigma$ variations around the expected limit are also shown.

		Observed	$-2\ \sigma$	$-1\ \sigma$	Expected	$+1\ \sigma$	$+2\ \sigma$
$\tau_{had}\tau_{had}$	$\sigma_{ggF+VBF}$ [fb]	145	70.5	94.6	131	183	245
	$\sigma_{ggF+VBF}/\sigma^{SM}_{ggF+VBF}$	4.95	2.38	3.19	4.43	6.17	8.27
$\tau_{lep}\tau_{had}$	$\sigma_{ggF+VBF}$ [fb]	265	124	167	231	322	432
	$\sigma_{ggF+VBF}/\sigma^{SM}_{ggF+VBF}$	9.16	4.22	5.66	7.86	10.9	14.7
Combined	$\sigma_{ggF+VBF}$ [fb]	135	61.3	82.3	114	159	213
	$\sigma_{ggF+VBF}/\sigma^{SM}_{ggF+VBF}$	4.65	2.08	2.79	3.87	5.39	7.22

- CMS $HH \rightarrow bb\tau\tau$ upper limits in σ/σ_{SM}:

ggF + VBF	3.3 (5.2)
VBF	124 (154)

HH→WWWW – ATLAS analysis

- https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2016-24/

	Observed limit on σ/σ_{SM}	Expected limit on σ/σ_{SM} Median	$+2\sigma$	$+1\sigma$	-1σ	-2σ
2 leptons	170	150	290	210	100	78
3 leptons	420	270	690	420	200	150
4 leptons	340	400	880	590	290	210
Combined	160	120	230	170	83	62