CMS HH $\rightarrow$ 4b analysis strategy, with focus on nonresonant

05.09.2018

Andres Tiko (INFN Padova)
on behalf of the CMS collaboration

Outline

Introduction

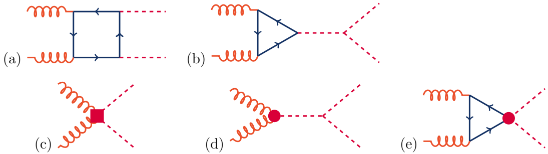
- The 4b final state has the highest BR of all HH final states, ≈ 0.339 for an SM Higgs boson with a mass of 125 GeV.
- Potentially one of the most sensitive signatures for the investigation of HH production
- The main challenge is the large background from QCD multijet final states, with rates exceeding that of the signal by several orders of magnitude
- Searches for HH production with both Higgs bosons decaying into bottom quark pairs resulting in:
  - Distinct hadronic jets
  - One bb pair (or both) is highly Lorentz-boosted and is reconstructed as a single large-area jet
- $35.9\,\mathrm{fb}^{-1}$ of data collected in 2016 at 13 TeV

BSM framework

ggF HH production can be generally described by 5 parameters controlling the tree-level interactions of the Higgs boson:

$$\mathcal{L}_{\mathrm{H}} = \frac{1}{2}\partial_\mu \mathrm{H}\, \partial^\mu \mathrm{H} - \frac{1}{2} m_{\mathrm{H}}{}^2 \mathrm{H}^2 - \kappa_\lambda\, \lambda_{\mathrm{SM}} v\, \mathrm{H}^3 - \frac{m_{\mathrm{t}}}{v}(v + \kappa_{\mathrm{t}}\, \mathrm{H} + \frac{c_2}{v}\mathrm{HH})\, (\bar{\mathrm{t}}_{\mathrm{L}} \mathrm{t}_{\mathrm{R}} + h.c.) + \frac{1}{4}\frac{\alpha_S}{3\pi v}(c_{\mathrm{g}}\, \mathrm{H} - \frac{c_{2\mathrm{g}}}{2v}\, \mathrm{HH})\, G^{\mu\nu} G_{\mu\nu}\,.$$

- c_{g}, $c_{2\mathrm{g}}$, c_2 for contact interactions not predicted by SM
- $\kappa_\lambda = \lambda_{\mathrm{HHH}}/\lambda_{\mathrm{SM}}$ and
- $\kappa_{\mathrm{t}} = y_{\mathrm{t}}/y_{\mathrm{SM}}$



Diagrams (a) and (b) correspond to SM-like processes; (c), (d), and (e) to pure BSM effects

BSM framework

- HH kinematic properties depend strongly on the 5 couplings
- 12 representative benchmarks points with similar properties identified to represent the 5D space
- An additional BM (box), for the null Higgs boson self-coupling hypothesis, is also considered
- BSM searches focused on these BMs

Carvalho et al. (2016)

BM	κ_λ	κ_t	c_2	c_g	c_{2g}
1	7.5	1.0	-1.0	0.0	0.0
2	1.0	1.0	0.5	-0.8	0.6
3	1.0	1.0	-1.5	0.0	-0.8
4	-3.5	1.5	-3.0	0.0	0.0
5	1.0	1.0	0.0	0.8	-1.0
6	2.4	1.0	0.0	0.2	-0.2
7	5.0	1.0	0.0	0.2	-0.2
8	15.0	1.0	0.0	-1.0	1.0
9	1.0	1.0	1.0	-0.6	0.6
10	10.0	1.5	-1.0	0.0	0.0
11	2.4	1.0	0.0	1.0	-1.0
12	15.0	1.0	1.0	0.0	0.0
Box	0.0	1.0	0.0	0.0	0.0
SM	1.0	1.0	0.0	0.0	0.0

$\mathrm{HH} \to \mathrm{b\overline{b}b\overline{b}}$ nonresonant measurement

CMS-PAS-HIG-17-017

Analysis strategy

- Analysis is optimized for sensitivity to the SM signal
- Same selection to extract limits for BSM models
- b jet selection for full reconstruction of HH final state and to reduce QCD multijet background rate (which still remains dominant)
- Background modelled using a hemisphere mixing method
- BDT trained to exploit the observable differences between signal and background
- Limits set on HH production fitting on BDT output

Trigger / Data sets

- Online event selection designed to select a b-enriched multijet sample, reducing rate from light quarks and gluons
- OR of two triggers:
  - 4 jets $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, ≥ 3 b jets, ≥ 2 jets $p_{\mathrm{T}} > 90\,\mathrm{GeV}$
  - 4 jets $p_{\mathrm{T}} > 45\,\mathrm{GeV}$, ≥ 3 b jets
- CSVv2 algorithm for b tagging
- Same triggers used for resolved resonant search
- MadGraph5_aMC@NLO samples for SM and all benchmarks
- 14 simulated signal samples are added together to obtain a larger signal sample (Pangea), reweighted to reproduce the physics of any particular point in the BSM phase space (Carvalho 2017)
- Background extracted from data, various MC samples used for cross checks

Event selection

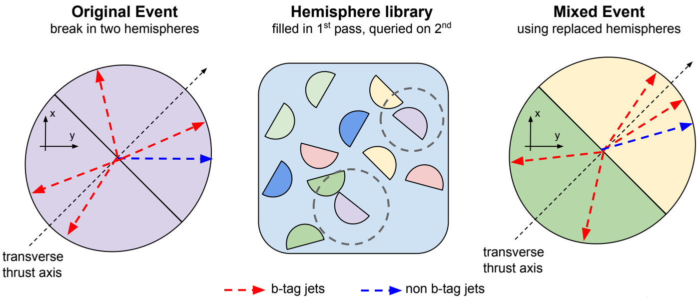
- Jets with $p_{\mathrm{T}\mathrm{j}} > 30\,\mathrm{GeV}$ and $|\eta_{\mathrm{j}}| < 2.4$.
- At least 4 b jets (med. CMVA WP, eff. 65%, misid rate 1%)
- The four jets with the highest CMVA discriminant values taken Higgs boson decay products
- Four jets paired to Higgs boson candidates using smallest mass difference between the two dijet systems
- Correct pairing in 54% of the cases

	Produced	Trigger	≥ 4 b tags
N events / fb	11.4	3.9	0.22
Relative Eff.		34%	5.6%
Efficiency		34%	1.9%

The efficiency for the 13 BSM points varies from -40% to $+10\%$ compared to the SM values.

Background modelling

- Precise modelling of multijet background needed, simulation very hard
- A dedicated data-driven method developed to repoduce the kinematical behaviour of background – hemisphere mixing
- Same selection as signal extraction - no control region needed



Artificial event type (k, l) – original hemispheres replaced by their k^{th} and l^{th} neighbours

BDT

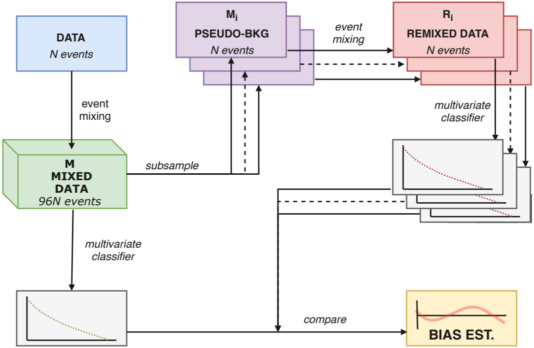
- A BDT trained to distinguish the SM signal from backgrounds with $\mathrm{XGBoost}$, using 25 variables
- All 13 BSM models use the same BDT as the SM to distinguish signal from background.

HH system	H candidates	Jet variables
M_X, M_HH, $p_\mathrm{T}^{\mathrm{H_1H_2}}$ $\cos\theta^*_{\mathrm{H_1H_2-H_1}}$	$M_{\mathrm{H_1}}$, $M_{\mathrm{H_2}}$ $p_\mathrm{T}^{\mathrm{H_1}}$, $p_\mathrm{T}^{\mathrm{H_2}}$ $\cos\theta^*_{\mathrm{H_1-j_1}}$ $\Delta R_{\mathrm{jj}}^{\mathrm{H_1}}$, $\Delta R_{\mathrm{jj}}^{\mathrm{H_2}}$, $\Delta\phi_{\mathrm{jj}}^{\mathrm{H_1}}$, $\Delta\phi_{\mathrm{jj}}^{\mathrm{H_2}}$	$p_{\mathrm{T}_\mathsf{j}}^{(i=1\text{–}4)}$, $\eta_\mathsf{j}^{(i=1\text{–}4)}$, $H_\mathrm{T}^{\mathsf{rest}}$, H_T $CMVA_3$, $CMVA_4$

- 60% of Pangea sample used for training, 20% for validation and 20% for application
- Artificial events of types $(1,1)$, $(1,2)$, $(2,1)$, and $(2,2)$ used for training; types $(3,4)$, $(5,6)$, $(7,8)$, and $(9,10)$ for validation; $(4,3)$, $(6,5)$, $(8,7)$, and $(10,9)$ for application

Background bias study

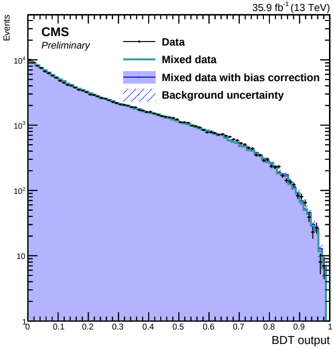
- A precise study is employed to investigate the need for a correction to the background shape of the BDT discriminator and a corresponding systematic uncertainty.

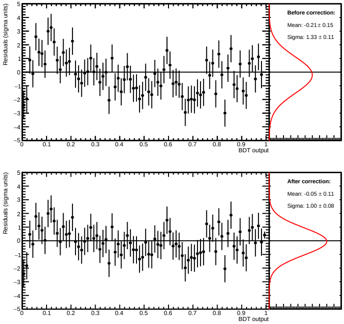


- A systematic bias is detected and the background template is corrected for the value obtained from this comparison.

Bias correction - validation

- Validity of procedure checked on data in two CRs
- Means of residuals compatible with zero after the bias correction in both CRs
- RMS of the pulls is compatible with one after the bias correction in the b tag CR, but not in the m_{H} CR
- Increased uncertainty to account for this





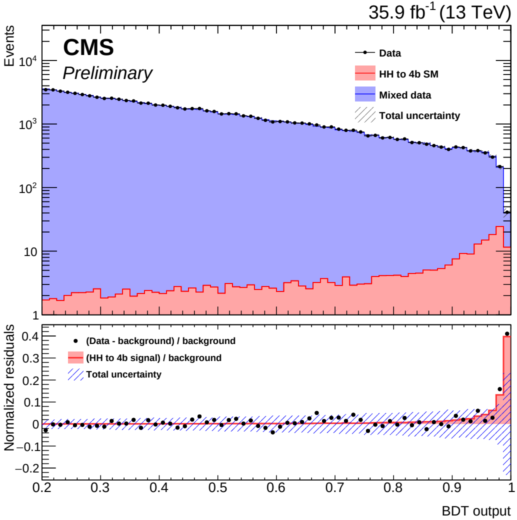
Systematic uncertainties

Source	Affects	Exp. limit (50%) variation
Bkg shape	bkg	30%
Bkg norm.	bkg	8.6%
b tagging eff.	sig	2.8%
Int. luminosity	sig	<0.01%
JES	sig	<0.01%
Trigger eff.	sig	<0.01%
PDF	sig	<0.01%
Pileup	sig	<0.01%
Jet energy res.	sig	<0.01%
μ_F and μ_R scales	sig	<0.01%

- The systematic uncertainty in the shape of the background model is accounted for by assigning an uncertainty to each BDT output bin that includes the statistical uncertainty and the systematic uncertainty related to the bias extraction
- Background normalization floats freely in the BDT fit

Fit

- Fit performed using BDT distribution
- Shape of background corrected for bias



$\mathrm{HH} \to \mathrm{b\bar{b}b\bar{b}}$ with large area jet(s)

CMS-B2G-17-019

Analysis strategy

- Search for the process $\mathrm{pp} \to \mathrm{HH} \to \mathrm{b\overline{b}b\overline{b}}$ in the phase space where one Higgs boson is highly Lorentz-boosted and is therefore reconstructed using a large-area jet, while the other Higgs boson has a lower boost and is reconstructed using two separate b quark jets (semi-resolved events)
- Combining with the selection presented in CMS-B2G-16-026, with two boosted $\mathrm{H} \to \mathrm{b\overline{b}}$ decays (fully-merged events)
- Substructure techniques used to identify jets stemming from a Higgs candidate
- HH resonant production also covered

Jets

- The anti-k_T algorithm used to cluster particles reconstructed using the PF algorithm using a distance parameter of
  - 0.8 (AK8 jets) – large-area jets
  - 0.4 (AK4 jets) – small-area jets
- For AK8 jet mass measurement, the PUPPI algorithm is applied to remove pileup effects
- Particles are clustered into AK8-PUPPI jets, groomed to remove soft and wide-angle radiation using the soft-drop algorithm
- N-subjettiness algorithm used to compute τ_N, quantifying the degree to which a jet contains N subjets
- A jet flavour requirement using a double-b tagger algorithm is applied to the AK8 jet(s) as the final H-tagging requirement.

Trigger / Data sets

- A logical OR of five different HLT criteria used to collect the events:
  1. Events where H_{T} is > 800 or $900\,\mathrm{GeV}$, depending on the LHC beam instantaneous luminosity
  2. $H_{\mathrm{T}} > 650\,\mathrm{GeV}$, and a pair of jets with invariant mass $> 900\,\mathrm{GeV}$ and a pseudorapidity separation $|\Delta\eta| < 1.5$
  3. H_{T} of all AK8 jets > 650 or $700\,\mathrm{GeV}$ and having an AK8 jet with a "trimmed mass" $>50\,\mathrm{GeV}$.
  4. AK8 jet with $p_{\mathrm{T}} > 360\,\mathrm{GeV}$ and a trimmed mass $> 30\,\mathrm{GeV}$
  5. Two AK8 jets having $p_{\mathrm{T}} > 280$ and $200\,\mathrm{GeV}$, with at least one having trimmed mass $>30\,\mathrm{GeV}$, together with an AK4 jet passing a loose b tagging criterion
- Nonresonant signal simulation is the same as in resolved analysis (but no Pangea sample)
- Multijet background modelled entirely from data
- $\mathrm{t\bar{t}}$+jets events (less than 10% of background) modelled with POWHEG 2.0

Event selection – semi-resolved

Variable	Selection
At least 1 AK8 jet J	$p_{\mathrm{T}} > 300$ GeV, $\lvert\eta\rvert < 2.4$
At least 2 AK4 jets j_1 and j_2	$p_{\mathrm{T}} > 30$ GeV, $\lvert\eta\rvert < 2.4$
$\Delta R(\mathrm{J}, \mathrm{j_i})$	>0.8
$\Delta R(\mathrm{j}_1,\mathrm{j}_2)$	<1.5
$\lvert\Delta\eta(\mathrm{J},\mathrm{j}_1+\mathrm{j}_2)\rvert$	≤ 2
$m_{\mathrm{Jjj,red}}$	>750 GeV
J soft-drop mass	105–135 GeV
J $\tau_{21} = \tau_2/\tau_1$	<0.55
J double-b tagger discriminator	>0.8 (30% eff., 1% mistag)
$\mathrm{j}_1+\mathrm{j}_2$ mass	90–140 GeV
$\mathrm{j}_1+\mathrm{j}_2+$(nearest AK4 jet) mass	>200 GeV
j_1 and j_2 DeepCSV	70% b-tagging eff., 1% mistag
Number of isolated leptons (e or μ)	$=0$

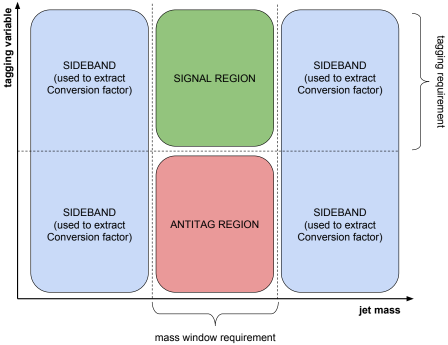
Events selected by fully-resolved search excluded

Event selection – fully-merged

- These events contain two H jets J_1 and J_2 instead of one
- Same trigger and selection for the H jet identification, except double-b tagger requirement
- Classified according to the values of the double-b tagger discriminators of the two H jets, with both J_1 and J_2 required to pass a loose double-b tagger discriminator value of > 0.3.
- Two classes of events: both J_1 and J_2 passing a tighter double-b tagger discriminator requirement of > 0.8 or not
- $|\Delta\eta(\mathsf{J}_1, \mathsf{J}_2)| < 1.3$.
- Overlapping events removed from semi-resolved

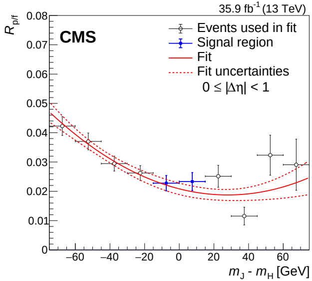
Background modelling

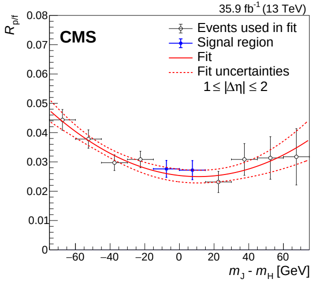
- Multijet background measured from sidebands of AK8 jet mass and double-b discriminator
- Shape expected to be similar in signal and anti-tag regions
- Pass-fail ratio $R_{\mathrm{p/f}}$ – ratio of #events for which the AK8 jet passes the double-b tagger tagger requirement over the #events failing it – fit to a function of H jet mass in sidebands to obtain normalization in bins of soft-drop mass
- $\mathrm{t\bar{t}}$+jets contributions subtracted before procedure; added back later



Background estimation results

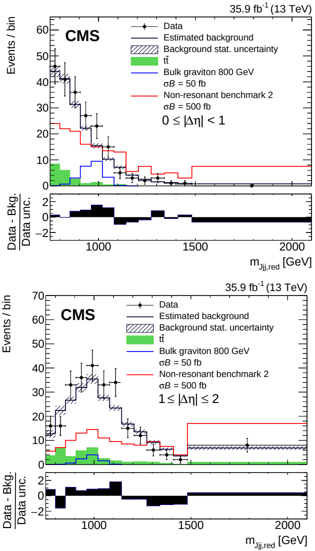
- $R_{p/f}$ fit done in two categories of $|\Delta\eta(\mathrm{J,j_1+j_2})|$
- Shown as a function of the difference between the soft-drop mass and the Higgs boson mass, $m_J - m_H$
- Fitted function is interpolated to obtain $R_{p/f}$ in the signal region (blue)





Fit

- Results obtained by binned fit to reduced di-Higgs invariant mass $m_{\mathrm{Jjj,red}} \equiv m_{\mathrm{Jjj}} - (m_{\mathrm{J}} - m_{\mathrm{H}}) - (m_{\mathrm{jj}}(\mathrm{j}_1, \mathrm{j}_2) - m_{\mathrm{H}})$
- Fluctuations from the jet mass resolution are corrected, leading to 8–10% improvement in the HH mass resolution, compared to using m_{Jjj}
- For fully-merged events: $m_{\mathrm{JJ,red}} \equiv m_{\mathrm{JJ}} - (m_{\mathrm{J_1}} - m_{\mathrm{H}}) - (m_{\mathrm{J_2}} - m_{\mathrm{H}})$



Systematic uncertainties

Source	Uncertainty (Semi-resolved)	Uncertainty (Fully-merged)
Signal yield (%)		
Trigger efficiency	1–15	1–15
Jet energy scale and resolution	1–3	1
Jet mass scale and resolution	2	2
H tagging correction factor	5–20	7–20
H jet τ_{21} selection	$+14/\text{-}13$	$+30/\text{-}26$
b tagging selection	2–9	2–5
PDF and scales	0.1–3	0.1–2
Pileup modelling	1–2	2
Luminosity	2.5	2.5
Trijet Invariant Mass	0.5	—
Background yield (%)		
$\mathrm{t\bar{t}}$+jets cross section	5	—
QCD background $R_{\mathrm{p/f}}$ fit	2–10	2–7

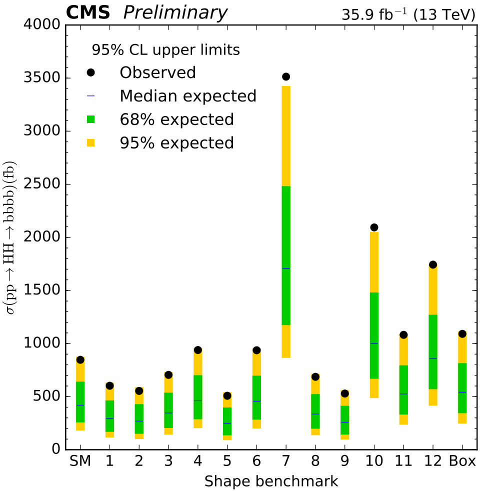
None of these lead to a significant change in the signal shape

Results

Resolved Measurement

CMS-PAS-HIG-17-017

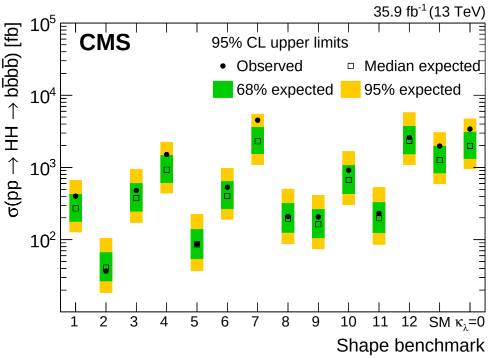
- Observed(expected) limits for the SM process are 847(419) fb
- 75(37) times the SM HH production cross section times the square of the branching fraction for the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay
- Expected upper limits for the 13 BSM benchmark models range from 295 to 1910 fb
- Slight excess for all models



Fully-merged and the semi-resolved selection results

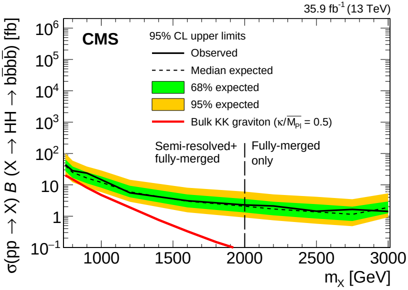
- Observed(expected) limits 179 (114) times the SM HH production cross section times the square of the branching fraction for the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay
- Expected upper limits for the 13 BSM benchmark models range from 85.9 to 2330 fb

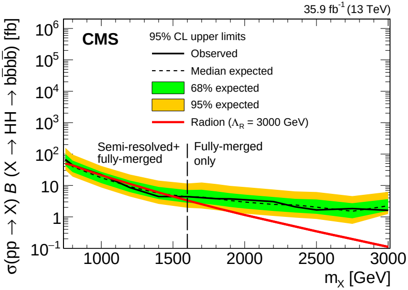
CMS-B2G-17-019



Resonant limits – fully merged and semi-resolved

- Search for resonant $\mathrm{HH} \rightarrow \mathrm{b\bar{b}b\bar{b}}$ in the region of phase space where at least one of the Higgs bosons has a large Lorentz boost, so that the $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ decay products are collimated to form a single jet, an H jet
- Upper limits of 43.9 to 1.4 fb for the bulk graviton
- Upper limits of 67 to 1.6 fb for the radion for the mass range 750–3000 GeV



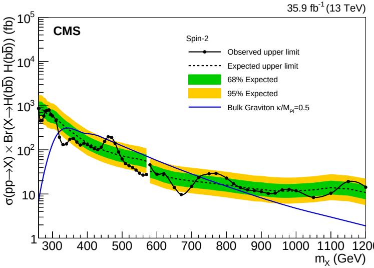


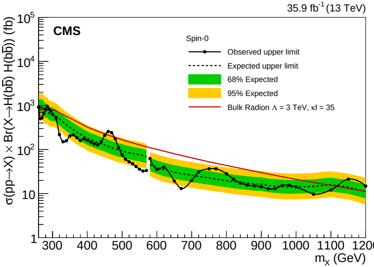
CMS-B2G-17-019

Resonant limits – resolved

CMS-HIG-17-009

- Search for resonant $\mathrm{HH} \rightarrow \mathrm{b\bar{b}b\bar{b}}$ with four resolved b jets
- Upper limits on production cross section for spin-0 and spin-2 resonances from 260 to 1200 GeV are set
- Exclusion of a bulk KK-graviton (with $\kappa l = 35$ and $\kappa = 0.5 M_{\mathrm{Pl}}$) in the mass ranges of 320–450 GeV and 480–720 GeV
- Excluded mass ranges for a radion (with decay constant $\Lambda = 3$ TeV) are 260–280 GeV, 300–450 GeV and 480–1120 GeV





Conclusions

- Two searches for nonresonant $\mathrm{HH} \to \mathrm{b\bar{b}b\bar{b}}$ performed
  - Different phase spaces, selections, background estimation methods
  - Better limits of 75(37) times the SM HH production from the resolved search
  - First limits on nonresonant HH using boosted topologies by CMS-B2G-17-019
  - For BSM searches, complementarity of approaches means both searches outperform the other in some parts of phase space
- Limits also set on HH resonant production and graviton and radion models

References

- **CMS Collaboration**, "Search for Non-Resonant Higgs Pair-Production in the $b\bar{b}b\bar{b}$ Final State with the CMS detector," CMS-PAS-HIG-17-017, CERN, Geneva, 2018. https://cds.cern.ch/record/2628554.
- **CMS**, A. M. Sirunyan *et al.*, "Search for production of Higgs boson pairs in the four b quark final state using large-area jets in proton-proton collisions at $\sqrt{s} = 13$ TeV," arXiv:1808.01473 [hep-ex].
- **CMS**, A. M. Sirunyan *et al.*, "Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Phys. Lett.* **B781** (2018) 244–269, arXiv:1710.04960 [hep-ex].
- **CMS**, A. M. Sirunyan *et al.*, "Search for resonant pair production of Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at 13 TeV," *Submitted to: JHEP* (2018) , arXiv:1806.03548 [hep-ex].
- A. Carvalho, M. Dall'Osso, T. Dorigo, F. Goertz, C. A. Gottardo, and M. Tosi, "Higgs pair production: Choosing benchmarks with cluster analysis," *JHEP* **04** (2016) 126, arXiv:1507.02245 [hep-ph].
- A. Carvalho, F. Goertz, K. Mimasu, M. Gouzevitch, and A. Aggarwal, "On the reinterpretation of non-resonant searches for Higgs boson pairs." 2017.

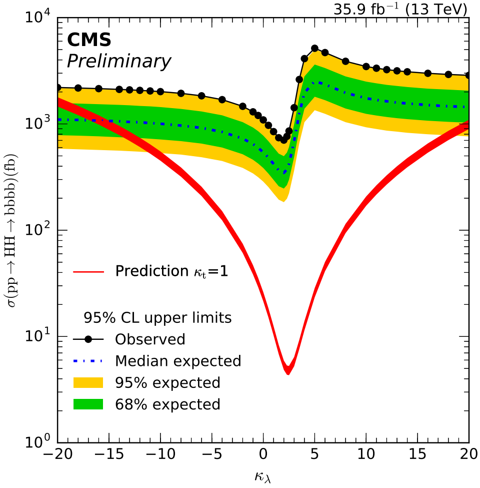
Backup

Resolved analysis – limits

Category	Observed	Expected	-2 s.d. (fb)	-1 s.d.	$+1$ s.d.	$+2$ s.d.
SM $\mathrm{H(b\bar{b})H(b\bar{b})}$	847	419	221	297	601	834

Benchmark point	observed limit (fb)	expected limit (fb)	-2 s.d.	-1 s.d.	$+1$ s.d.	$+2$ s.d.
1	602	295	155	209	424	592
2	554	269	141	190	389	548
3	705	346	182	245	497	691
4	939	461	244	327	662	920
5	508	248	131	176	357	501
6	937	457	240	323	657	916
7	3510	1710	905	1210	2440	3390
8	686	336	177	238	483	674
9	529	259	136	183	373	520
10	2090	1000	527	709	1440	2010
11	1080	525	277	372	755	1050
12	1744	859	455	611	1230	1710
Box	1090	542	286	384	775	1080

Resolved analysis – BSM scan



Limits – semi-boosted

Shape benchmark	Obs. lim. (fb)	Exp. lim. (fb)	+Exp (68%) (fb)	-Exp (68%) (fb)	+Exp (95%) (fb)	-Exp (95%) (fb)
1	401	271	179	428	127	660
2	36.7	41.0	26.5	66.3	18.5	105
3	479	376	247	601	173	936
4	1510	932	618	1460	438	2240
5	86.6	85.9	54.4	140	37.0	225
6	533	403	268	637	190	978
7	4520	2300	1530	3580	1100	5470
8	209	196	126	317	87.2	504
9	206	163	106	264	74	415
10	916	670	433	1070	302	1660
11	232	198	125	326	85.9	526
12	2600	2330	1530	3700	1090	5750
SM	1980	1260	833	1970	589	3030
$\kappa_\lambda = 0$	3404	1989	3092	1334	4732	960