2-step BDT in HGTD VBF $H \to \gamma\gamma$ analysis

FANGYI GUO, IHEP

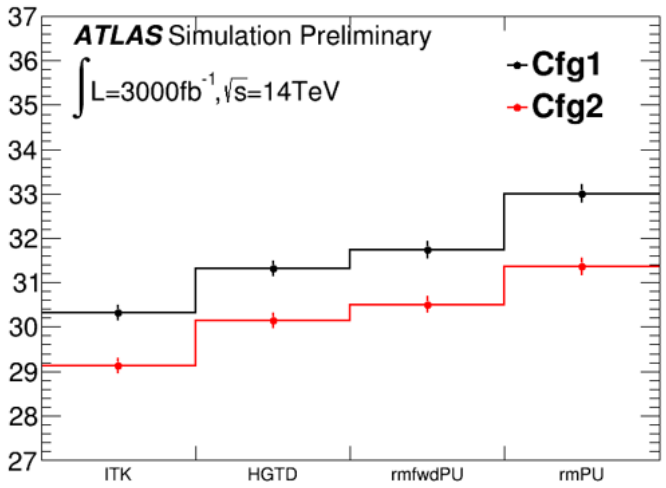
fangyi.guo@cern.ch

中国科学院高能物理研究所
Institute of High Energy Physics Chinese Academy of Sciences

Previous review

VBF significance

Cfg and scenario	VBF significance
Cfg1 ITK	30.32±0.18
Cfg1 HGTD	31.32±0.18
Cfg1 rmfwdPU	31.73±0.20
Cfg1 rmPU	33.00±0.21
Cfg2 ITK	29.12±0.18
Cfg2 HGTD	30.14±0.18
Cfg1 rmfwdPU	30.50±0.19
Cfg2 rmPU	31.36±0.20

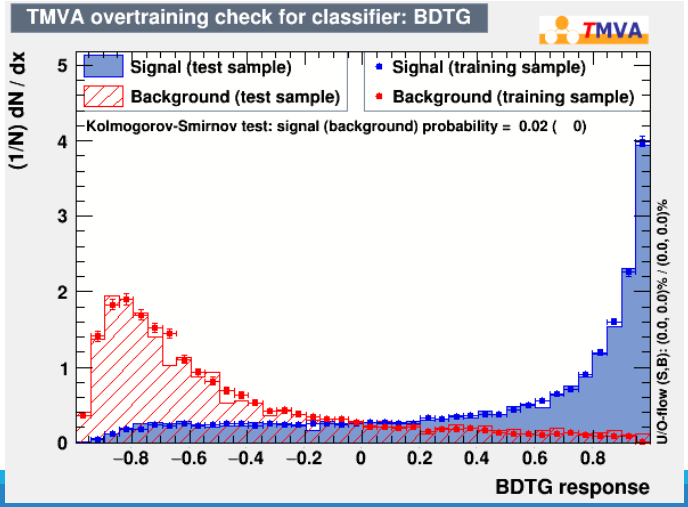


VBF significance $\sigma_{VBF} = \sqrt{2 \times ((N_{VBF} + N_{ggF}) \times \ln\left(1 + \frac{N_{VBF}}{N_{ggF}}\right) - N_{VBF})}$ in different config and scenarios

2-step BDT

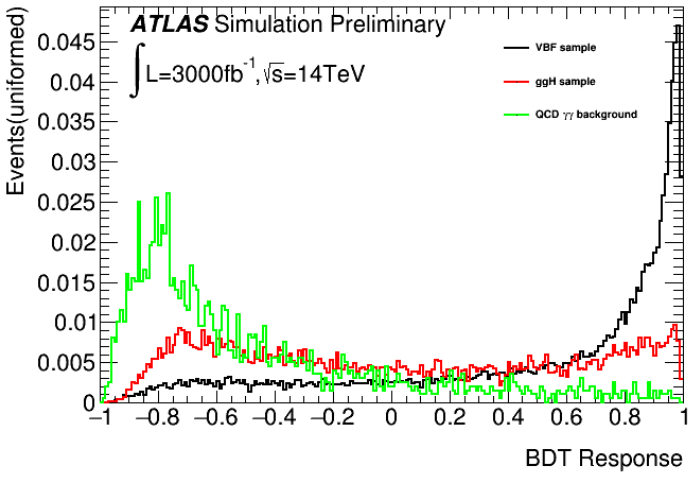
First step: VBF vs. $\gamma\gamma$ background

- Signal: VBF sample. Background: QCD $\gamma\gamma$ sample
- Training variable: p_{Tt}, m_{jj}, $\Delta\eta_{jj}$, $\Delta\Phi_{\gamma\gamma,jj}$, η^{Zepp}, $\Delta R_{\gamma,j}^{min}$
- Categorization: maximum $\sigma_{VBF} = \sqrt{2 \times \left(\left(N_{VBF} + N_{ggF} + N_{bkg}\right) \times \ln\left(1 + \frac{N_{VBF}}{N_{ggF}+N_{bkg}}\right) - N_{VBF}\right)}$ to separate the sample into 2 categories: tight and loose
- Only optimize the HGTD scenario in 2 Cfgs, and apply this result into other scenarios.



BDT overtraining test

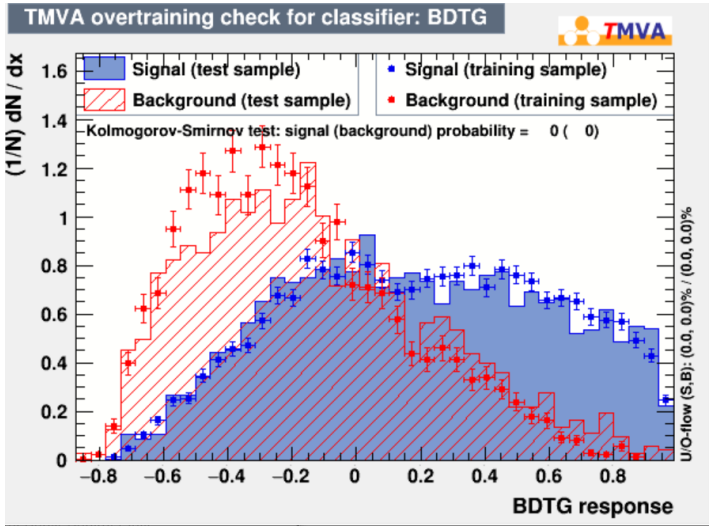
BDT response in Cfg1 HGTD scenario



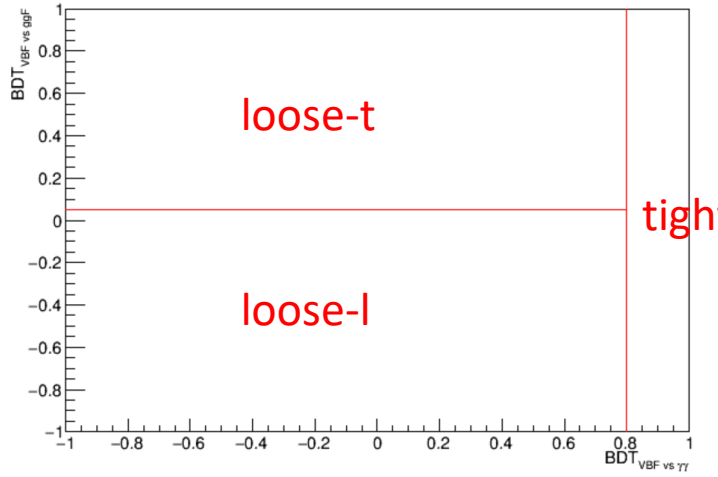
2-step BDT

Second step: VBF vs. ggF

- For loose category, re-training VBF and ggF sample.
- Training variable: m_{jj}, $\Delta\eta_{jj}$, p_T^{Hjj}, pT_{j1}, pT_{j2}.
- Categorization: maximum $\sigma_{VBF} = \sqrt{2 \times \left(\left(N_{VBF} + N_{ggF}\right) \times \ln\left(1 + \frac{N_{VBF}}{N_{ggF}}\right) - N_{VBF}\right)}$ to have 2 categories: loose-t, loose-l.



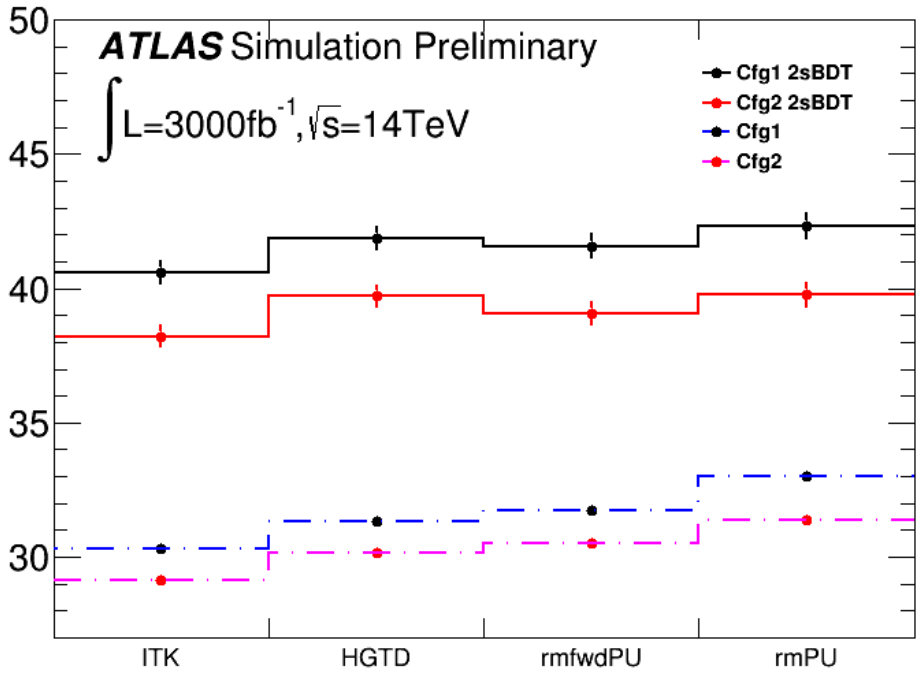
BDT response in Cfg1 HGTD scenario loose category



3 categories for 2-step BDT

2-step BDT results

Cfg and scenario	VBF significance
Cfg1 ITK	40.60±0.46
Cfg1 HGTD	41.85±0.46
Cfg1 rmfwdPU	41.58±0.49
Cfg1 rmPU	42.31±0.50
Cfg2 ITK	38.21±0.44
Cfg2 HGTD	39.70±0.44
Cfg1 rmfwdPU	39.06±0.46
Cfg2 rmPU	39.75±0.47



Combined significance in 3 BDT categories

Conclusion

2-step BDT could provide more VBF-ggF separation power.

- Do not apply BDT in tight category, due to the low ggF statistics.
- 3-category could provide a ~30% improvement in VBF-ggF separation.
- The σ_{VBF} in rmfwdPU scenario is a bit lower than in HGTD scenario, but still within the error range.
- Half of ggF events are separated for the BDT training, so ggF statistics error is higher than before.

Back up

Event selection

VBF Pre-selection:

Number of photon ≥ 2

Number of jet ≥ 2

$\frac{pT_{leading\ photon}}{m_{\gamma\gamma}} \geq 0.35$

$\frac{pT_{sub-leading\ photon}}{m_{\gamma\gamma}} \geq 0.25$

$105GeV < m_{\gamma\gamma} < 160GeV$

$\Delta\eta_{jj} > 2$

$\left|\eta_{\gamma\gamma}^{Zepp}\right| < 5$

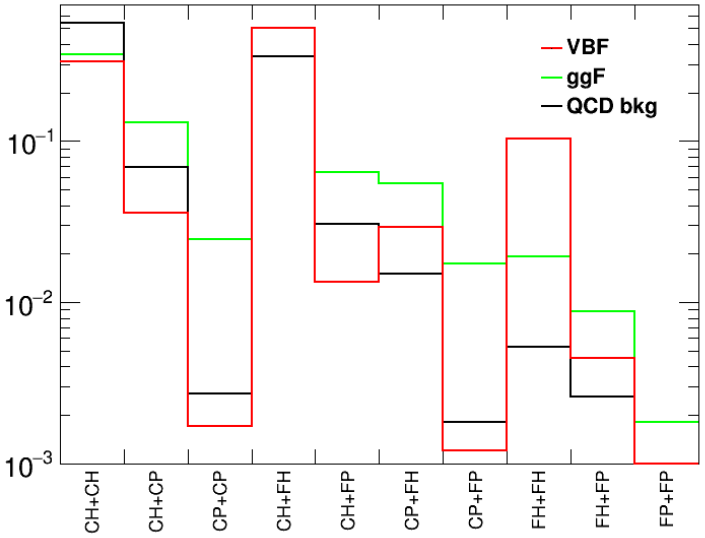
Reconstruction configuration:

- Cfg1: study the overall significance
  For jets |eta|<3.8, require pT>30GeV with Track Confirmation(TC)
  For jets 3.8<|eta|<4.4, require pT>70GeV, jet efficiency $\epsilon_{jet} = 1$
- Cfg2: focus on HGTD region
  Only remain jet in |eta|<3.8, pT>30GeV, with TC

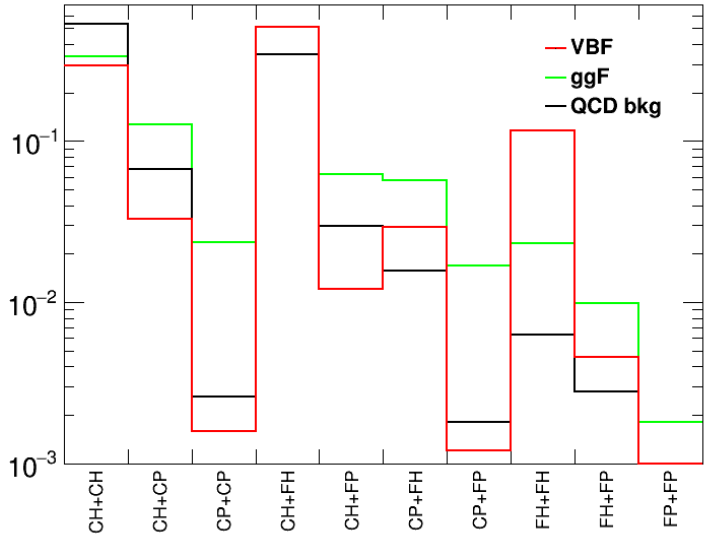
Scenarios:

- ITK(default)
- HGTD(UseHGTD0=true)
- remove all PU(with ITk acceptance)

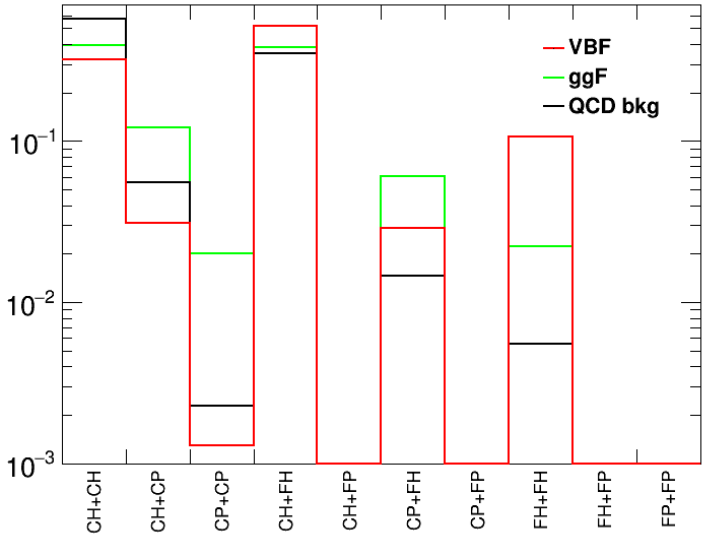
Jet categories Cfg1



Cfg1 ITK only



Cfg1 ITK+HGTD



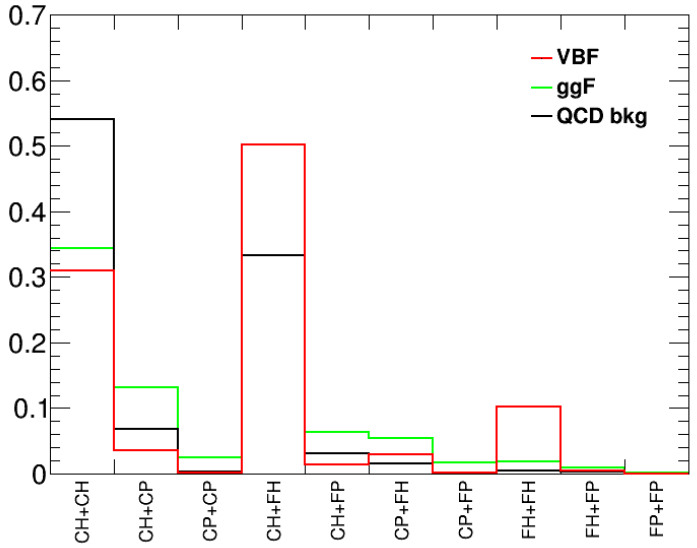
Cfg1 rmfwdPU

C: central F: forward H:HS jet P: pile-up jet rmfwdPU: remove forward PU jets

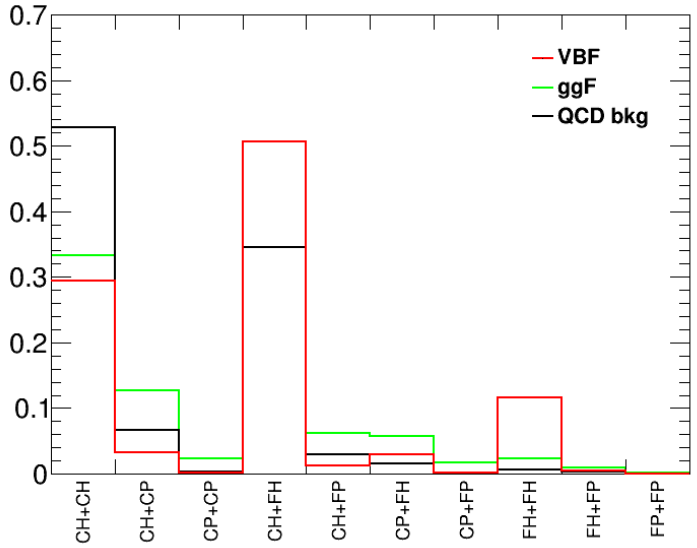
*Number of events in different jet categories. Scaled to 3000 fb-1

Forward: 2.5<$|\eta|$<3.8 Central: [0, 2.5] & [3.8, 4.4]

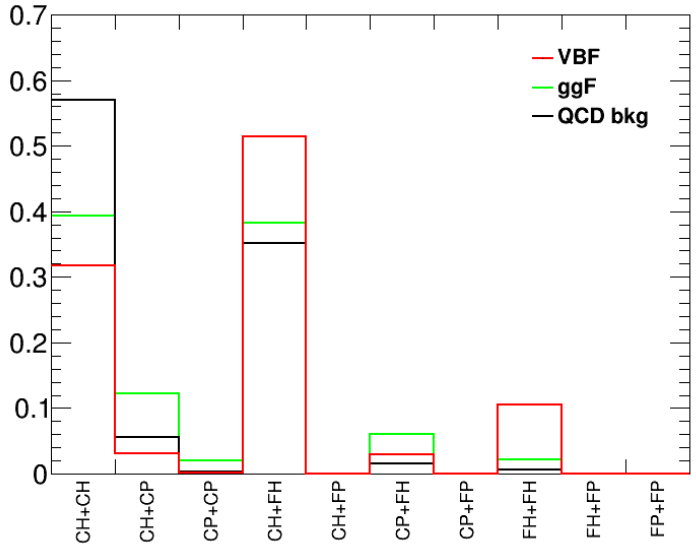
Jet categories Cfg1 in linear y scale



Cfg1 ITK only



Cfg1 ITK+HGTD



Cfg1 rmfwdPU

C: central F: forward H:HS jet P: pile-up jet rmfwdPU: remove forward PU jets

*Number of events in different jet categories. Scaled to 3000 fb-1

Forward: 2.5<$|\eta|$<3.8 Central: [0, 2.5] & [3.8, 4.4]

Jet categories Cfg1 ITK only

	ggF		VBF		QCD bkg	
CH+CH	1476.92	34.41%	681.01	30.95%	285101	54.03%
CH+CP	562.92	13.12%	78.62	3.57%	36366.84	6.89%
CP+CP	105.73	2.46%	3.79	0.17%	1418.28	0.27%
CH+FH	1431.87	33.37%	1104.19	50.18%	175492.1	33.26%
CH+FP	275.13	6.41%	29.22	1.33%	16216.56	3.07%
CP+FH	235.29	5.48%	64.63	2.94%	7920.96	1.50%
CPU+FPU	74.83	1.74%	2.71	0.12%	963.36	0.18%
2fwd HS-HS	83.02	1.93%	226.45	10.29%	2783.04	0.53%
2fwd HS-PU	37.97	0.88%	9.82	0.45%	1364.76	0.26%
2fwd PU-PU	7.82	0.18%	0.17	0.01%	26.76	0.01%
total	4291.53	100.00%	2200.60	100.00%	527653.7	100.00%

Event number and percentage in each sample and jet category.
Scaled to 3000 fb-1.

Jet categories Cfg1 ITK+HGTD

	ggF		VBF		QCD bkg	
CH+CH	1471.71	33.31%	679.89	29.45%	284351.8	52.88%
CH+CP	560.31	12.68%	76.29	3.30%	36206.28	6.73%
CP+CP	104.62	2.37%	3.75	0.16%	1418.28	0.26%
CH+FH	1524.20	34.50%	1170.74	50.71%	185446.8	34.49%
CH+FP	273.27	6.19%	27.97	1.21%	15895.44	2.96%
CP+FH	253.54	5.74%	67.26	2.91%	8456.16	1.57%
CPU+FPU	74.83	1.69%	2.66	0.12%	963.36	0.18%
2fwd HS-HS	103.50	2.34%	269.44	11.67%	3398.52	0.63%
2fwd HS-PU	43.93	0.99%	10.65	0.46%	1525.32	0.28%
2fwd PU-PU	7.82	0.18%	0.17	0.01%	26.76	0.00%
total	4417.74	100.00%	2308.80	100.00%	537688.7	100.00%

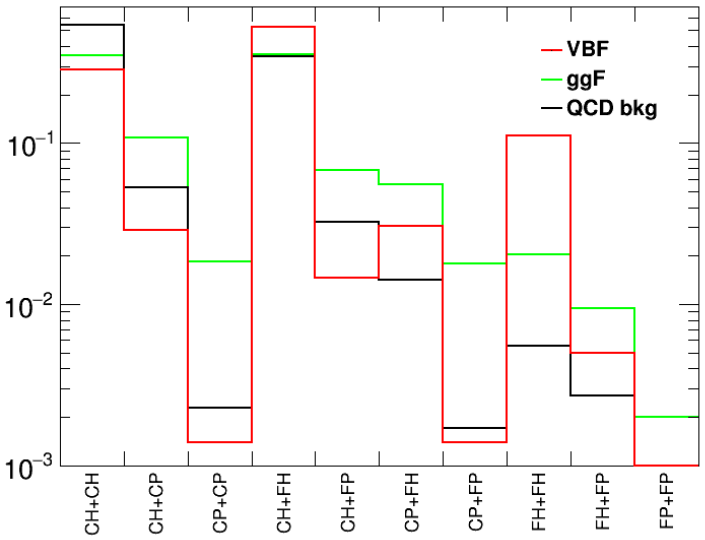
Event number and percentage in each sample and jet category.
Scaled to 3000 fb-1.

Jet categories Cfg1 remove forward PU

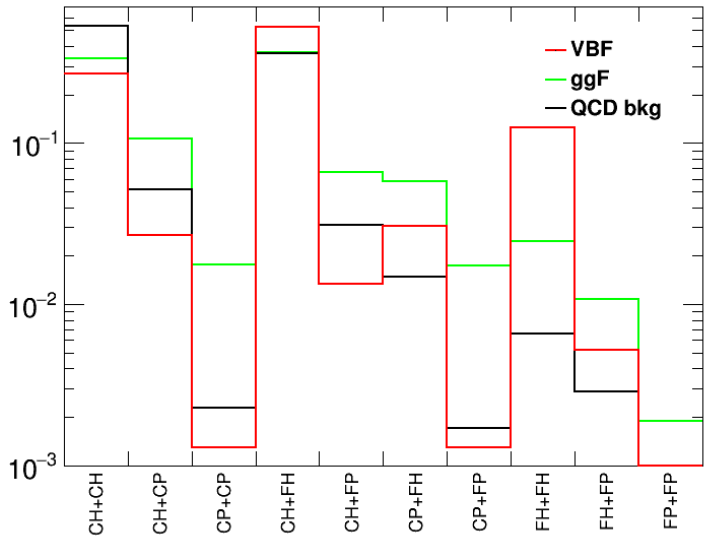
	ggF		VBF		QCD bkg	
CH+CH	1492.93	39.39%	687.46	31.82%	287429.2	57.06%
CH+CP	460.91	12.16%	66.76	3.09%	27883.92	5.54%
CP+CP	76.32	2.01%	2.91	0.13%	1177.44	0.23%
CH+FH	1447.51	38.19%	1112.22	51.48%	177017.4	35.14%
CH+FP	0.00	0.00%	0.00	0.00%	0	0.00%
CP+FH	228.22	6.02%	62.43	2.89%	7412.52	1.47%
CPU+FPU	0.00	0.00%	0.00	0.00%	0	0.00%
2fwd HS-HS	84.14	2.22%	228.86	10.59%	2783.04	0.55%
2fwd HS-PU	0.00	0.00%	0.00	0.00%	0	0.00%
2fwd PU-PU	0.00	0.00%	0.00	0.00%	0	0.00%
total	3790.03	100.00%	2160.64	100.00%	503703.5	100.00%

Event number and percentage in each sample and jet category.
Scaled to 3000 fb-1.

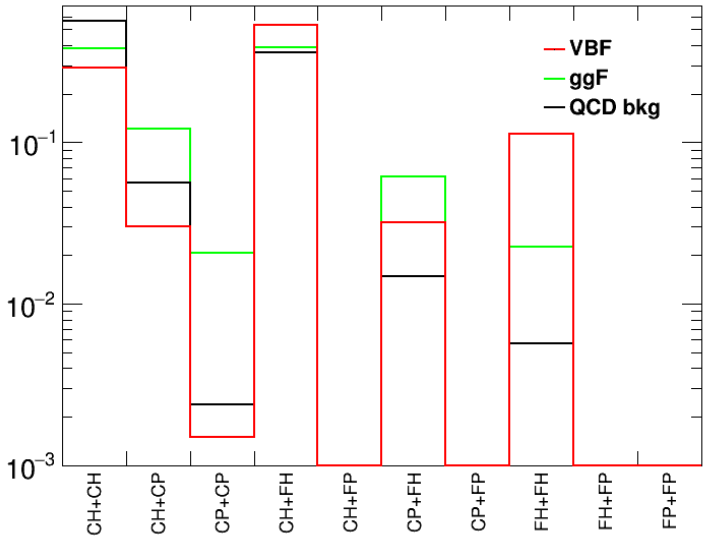
Jet categories Cfg2



Cfg2 ITK only



Cfg2 ITK+HGTD



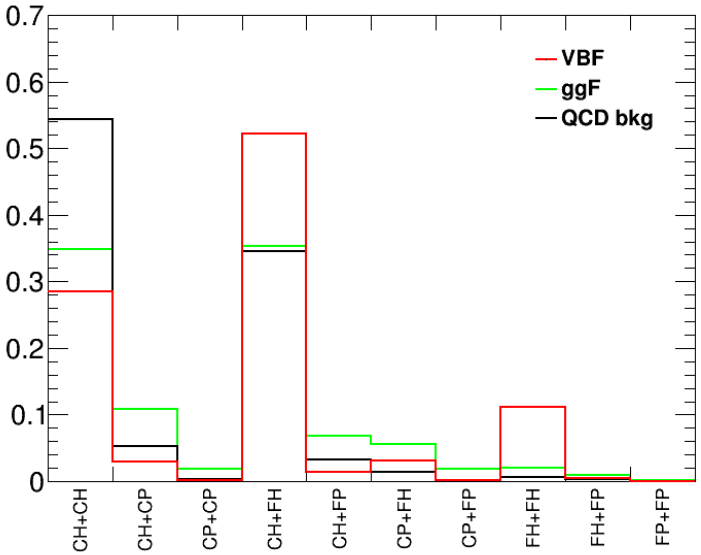
Cfg2 rmfwdPU

C: central F: forward H:HS jet P: pile-up jet rmfwdPU: remove forward PU jets

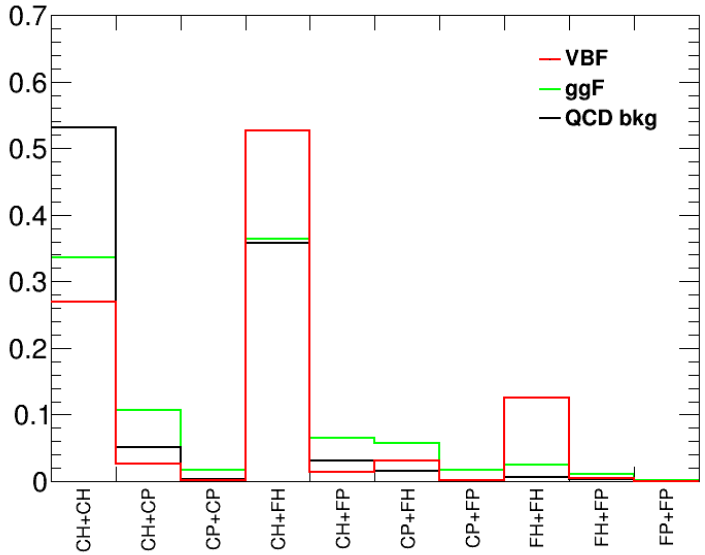
*Number of events in different jet categories. Scaled to 3000 fb-1

Forward: 2.5<$|\eta|$<3.8 Central: [0, 2.5]

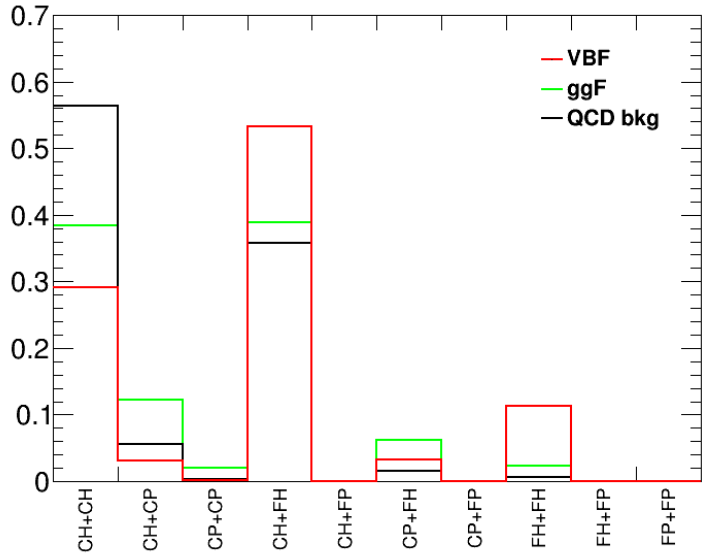
Jet categories Cfg2 in linear y scale



Cfg2 ITK only



Cfg2 ITK+HGTD



Cfg2 rmfwdPU

C: central F: forward H:HS jet P: pile-up jet rmfwdPU: remove forward PU jets

*Number of events in different jet categories. Scaled to 3000 fb-1

Forward: 2.5<$|\eta|$<3.8 Central: [0, 2.5]

Jet categories Cfg2 ITK only

	ggF		VBF		QCD bkg	
CH+CH	1417.35	34.78%	585.62	28.48%	278304	54.32%
CH+CP	441.92	10.84%	59.60	2.90%	27188.16	5.31%
CP+CP	74.46	1.83%	2.91	0.14%	1177.44	0.23%
CH+FH	1436.71	35.25%	1073.39	52.19%	176856.8	34.52%
CH+FP	277.36	6.81%	29.92	1.46%	16484.16	3.22%
CP+FH	225.24	5.53%	63.43	3.08%	7251.96	1.42%
CPU+FPU	72.60	1.78%	2.91	0.14%	883.08	0.17%
2fwd HS-HS	83.02	2.04%	228.28	11.10%	2809.8	0.55%
2fwd HS-PU	38.72	0.95%	10.28	0.50%	1364.76	0.27%
2fwd PU-PU	8.19	0.20%	0.21	0.01%	26.76	0.01%
total	4075.59	100.00%	2056.55	100.00%	512347	100.00%

Event number and percentage in each sample and jet category.
Scaled to 3000 fb-1.

Jet categories Cfg2 ITK+HGTD

	ggF		VBF		QCD bkg	
CH+CH	1412.89	33.62%	584.54	27.01%	277528	53.12%
CH+CP	439.69	10.46%	57.73	2.67%	27054.36	5.18%
CP+CP	74.09	1.76%	2.91	0.13%	1177.44	0.23%
CH+FH	1527.56	36.35%	1138.36	52.60%	186891.8	35.78%
CH+FP	275.50	6.56%	28.72	1.33%	16136.28	3.09%
CP+FH	243.49	5.79%	66.22	3.06%	7760.4	1.49%
CPU+FPU	72.60	1.73%	2.83	0.13%	883.08	0.17%
2fwd HS-HS	103.50	2.46%	271.48	12.54%	3425.28	0.66%
2fwd HS-PU	45.42	1.08%	11.24	0.52%	1525.32	0.29%
2fwd PU-PU	8.19	0.19%	0.21	0.01%	26.76	0.01%
total	4202.92	100.00%	2164.22	100.00%	522408.7	100.00%

Event number and percentage in each sample and jet category.
Scaled to 3000 fb-1.

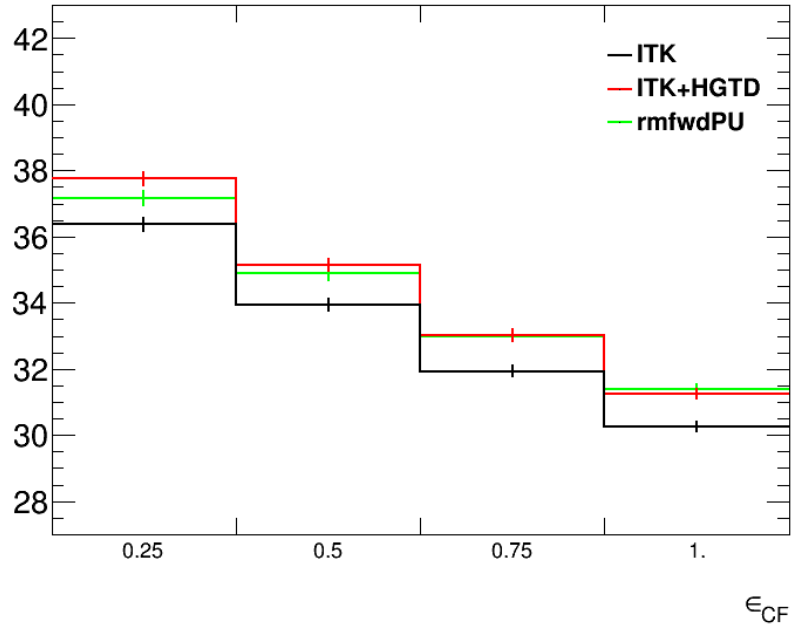
Jet categories Cfg2 remove forward PU

	ggF		VBF		QCD bkg	
CH+CH	1425.92	38.39%	588.62	29.07%	279615.2	56.33%
CH+CP	452.35	12.18%	61.43	3.03%	27776.88	5.60%
CP+CP	76.69	2.06%	3.04	0.15%	1177.44	0.24%
CH+FH	1446.02	38.93%	1077.43	53.21%	177632.9	35.78%
CH+FP	0.00	0.00%	0.00	0.00%	0	0.00%
CP+FH	228.97	6.16%	64.63	3.19%	7412.52	1.49%
CPU+FPU	0.00	0.00%	0.00	0.00%	0	0.00%
2fwd HS-HS	84.14	2.27%	229.65	11.34%	2809.8	0.57%
2fwd HS-PU	0.00	0.00%	0.00	0.00%	0	0.00%
2fwd PU-PU	0.00	0.00%	0.00	0.00%	0	0.00%
total	3714.08	100.00%	2024.80	100.00%	496424.8	100.00%

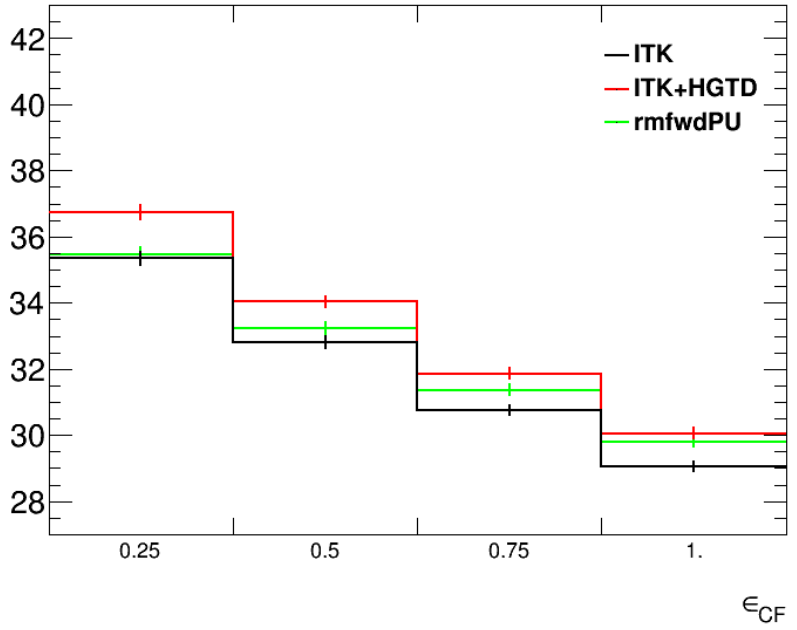
Event number and percentage in each sample and jet category.
Scaled to 3000 fb-1.

Significance with ϵ_{CF}

Calculate the σ_{VBF} with different Central-Forward region N_{ggF}(ϵ_{CF}=25%, 50%, 75% and 100%).



Cfg1



Cfg2

Event number and combined significance for each scenario in Cfg1

	tight	Loose-t	Loose-l
VBF	854.37±8.43	859.54±8.46	486.7±6.37
ggF	465.38±18.62	1207±29.98	2168.77±44.16
VBF purity	64.73%	41.59%	15.67%
Significance in category	32.29±0.53	22.44±0.31	10.09±0.16
Combined significance	40.60±0.46		

ITK only

	tight	Loose-t	Loose-l
VBF	922.47±8.76	851.38±8.42	534.98±6.67
ggF	501.12±19.32	1117.65±28.85	2798.97±45.65
VBF purity	64.80%	43.24%	16.05%
Significance in category	33.58±0.53	22.97±0.32	9.81±0.14
Combined significance	41.85±0.46		

ITK+HGTD

Event number and combined significance for each scenario in Cfg1

	tight	Loose-t	Loose-l
VBF	847.47±8.40	927.88±8.79	374.5±5.58
ggF	409.53±14.46	1313.48±31.27	2039.47±38.97
VBF purity	67.42%	41.40%	15.51%
Significance in category	33.53±0.57	23.23±0.31	8.06±0.14
Combined significance	41.58±0.49		

Remove forward PU

	tight	Loose-t	Loose-l
VBF	864.28±8.48	876.27±8.54	336.79±5.29
ggF	419.21±17.67	1059.57±28.09	1629.19±34.83
VBF purity	67.34%	45.26%	17.13%
Significance in category	33.82±0.57	24.10±0.34	8.08±0.15
Combined significance	42.31±0.50		

Remove all PU

Event number and combined significance for each scenario in Cfg2

	tight	Loose-t	Loose-l
VBF	891.83±8.61	746.91±7.88	417.86±5.90
ggF	551.75±20.27	1176.47±29.59	2346.99±41.8
VBF purity	61.78%	38.83%	15.11%
Significance in category	31.51±0.50	19.93±0.29	8.38±0.14
Combined significance	38.21±0.44		

ITK only

	tight	Loose-t	Loose-l
VBF	966.0±8.97	728.1±7.78	470.14±6.26
ggF	594.94±21.05	1056.59±28.05	2551.01±43.58
VBF purity	61.89%	40.80%	15.56%
Significance in category	32.85±0.50	20.37±0.30	9.04±0.14
Combined significance	39.70±0.44		

ITK+HGTD

Event number and combined significance for each scenario in Cfg2

	tight	Loose-t	Loose-l
VBF	883.43±8.57	775.46±8.03	365.92±5.52
ggF	501.86±19.33	1204.77±29.95	2007.45±38.66
VBF purity	63.77%	39.16%	15.41%
Significance in category	32.34±0.52	20.42±0.29	7.94±0.14
Combined significance	39.06±0.46		

Remove forward PU

	tight	Loose-t	Loose-l
VBF	896.33±8.64	725.07±7.91	294.92±4.95
ggF	497.4±19.24	1039.48±27.82	1493.68±33.35
VBF purity	64.31%	41.98%	16.49%
Significance in category	32.85±0.53	21.13±0.31	7.4±0.14
Combined significance	39.75±0.47		

Remove all PU

Stat. err estimation

Use Cfg1 ITK scenario BDT tight category as example(SL 20)

- $N_{VBF} = 854.37 \pm 8.43, N_{ggF} = 465.38 \pm 18.62$
- $\sigma_{VBF} = \sqrt{2 \times ((N_{VBF} + N_{ggF}) \times \ln\left(1 + \frac{N_{VBF}}{N_{ggF}}\right) - N_{VBF})} = 32.29$
- $\delta\sigma_{VBF} = \sqrt{(\frac{\partial\sigma_{VBF}}{\partial N_{VBF}} \times \delta N_{VBF})^2 + (\frac{\partial\sigma_{ggF}}{\partial N_{ggF}} \times \delta N_{ggF})^2} = \sqrt{(0.032 \times 8.43)^2 + (-0.024 \times 18.62)^2} = 0.53$

The dominant term in $\delta\sigma_{VBF}$ is δN_{ggF}, and other categories have similar conclusion.