ATLAS Note
GROUP-2019-XX
6th January 2020
Draft version 0.1

Estimation of the forward pile-up influence and HGTD potential in HL-LHC via VBF $Higgs \rightarrow \gamma\gamma$ channel

The ATLAS Collaboration

This note presents the possible forward pile-up jet influence in VBF Higgs analysis in HL-LHC situation through diphoton channel by comparing the VBF significance in two imaginary scenarios, with 3000 fb^{-1} of full HL-LHC data. A parameterized smearing function is used for the event reconstruction in truth level Monte Carlo, and the analysis strategy is a mimic of RunII, with preselection and BDT optimization. In conclusion we observe a 25% improvement and the statistic uncertainty is low enough. No systematic uncertainty considered in this work. This is supposed to be an indication of High Granularity Timing Detector potential in VBF Higgs analysis.

Contents

1 Introduction

After the beginning of operation in 2010, the Large Hadron Collider (LHC) has reached several brilliant achievements. In the future an upgrade High Luminosity LHC (HL-LHC) is under planning, and scheduled to start in 2026. The instantaneous luminosity of HL-LHC will reach up to $7.5 \times 10^3 4 cm^- 2 s^- 1$, and finally 3 4 $ab^- 1$ integrated luminosity can be delivered over the subsequent decade. This amount of data has chance to make more precise measurement of Standard Model, and see evidences of new physics.

In the upgrade beam condition, the pile-up is one of the main challenge at HL-LHC. Averagely there are 200 simultaneous pp interactions ($< \mu >= 200$) occurring within the same bunch crossing interval, while this number for RunII is $< \mu >= 40$. This will bring a big challenge to the tracker system, Inner-Tracker (ITk) designed for HL-LHC. It need to efficiently reconstruct charged particles created in the primary interactions and assign them to the correct production vertices in this high pile-up environment. So an additional detector, the High Granularity Timing Detector (HGTD) is design for ATLAS upgrade. It would be placed in the forward region of ITk, to measure charged particle trajectories in time, together with the position to provide a high level vertex reconstruction. The target average time resolution of HGTD is about 30 ps for a minimum ionizing particle. This capability enhances the performance for hard scatter jet tagging, so the physics processes with jets in final state, especially in the forward region, are expected to get benefits from HGTD.

The Vector Boson Fusion (VBF) of Higgs production is one of this kind of process. After radiating two vector boson, two quarks go to forward region with high transverse momentum. With this feature the Higgs boson can be finely identified. This process also contributes a lot in the discovery and following property measurement of Higgs Boson. Correspondingly, a precise Higgs decay final state, like the di-photon channel, can be a chance to study the possible performance of hard scatter VBF jet in HGTD.

In this note, we perform an analysis of HGTD potential in VBF Higgs to di-photon channel, assuming 3000 $fb^- 1$ data at $\sqrt{s}$=14TeV. A simple smearing function is used to reconstruct the particles in the final state, as described in Chapter 2. We suppose 2 different scenarios for the comparison of different pile-up rejection power of HGTD. The analysis procedure in VBF Higgs is a mimic of Run2 HGam work, and the final result is represented with the VBF significance by number counting. Here only statistics uncertainty is considered.

2 MC sample and detector simulation

Several Monte Carlo samples are generated for this analysis. The signal process, Vector Boson Fusion Higgs process is simulated with Powheg, interfaced with Pythia for Higgs decay, parton showering, hadronization and multiple parton interactions process, at 14TeV central of mass energy. The Higgs mass is fixed at 125GeV, and only di-photon channel decay considered. Here totally 0.7 million events are generated. The gluon fusion process the main resonant background in this work, and is simulated with similar setting. Details are in Table 1.

The non-resonant background in di-photon channel has 2 main components. Dominant one is the QCD process with 2 hard photon and several jets in final state. It is simulated with Sherpa generator in NNLO in $\sqrt{s}$=13TeV, requiring the phase space of 2 photon plus at least 1 jet, and di-photon invariant mass in [90,175]GeV window. The difference in performance between 13TeV and 14TeV is regarded as negligible, but the cross section is scale by 12% to 14TeV. Totally 500 million events are generated in this process, and part of them are used necessarily.

The second part of non-resonant background comes from the mis-identification of jets into photon(called γj or jj event). It depends on the detector response and software reconstruction, and can only be estimated from data. In RunII condition, the QCD $\gamma\gamma$ events account for about 80%, varied in different categories. In upgrade HL-LHC we suppose the γ-jet fake rate has the same level with RunII, so this fraction is supposed to be similar, and neglect the shape difference in QCD $\gamma\gamma$ events and fake events. Thus we calculated the average fraction in 4 VBF-relative categories in RunII result, 82.5% for $\gamma\gamma$ and 17.5% for $\gamma j + jj$, and scaled the previous QCD process field into total number.

A parameterized estimates of the ATLAS performance at HL-LHC is used for the simulation of detector response, called Upgrade Performance Function. These functions are based on full simulation and are meant to be applied to truth-level quantities. Resolution, detector efficiency and fake rate are included into the package as some kinematic related functions. Pile-up jets can be added into hard-scatter truth event from a dedicated pileup overlay library, and be manually removed with this truth information. In this note, an average pile-up value $< \mu >= 200$ is set by default, corresponding to the predicted HL-LHC environment.

DSID	Process	Generator	PDF	N events
160024	$VBF\ H \rightarrow \gamma\gamma$	POWHEG+Pythia8	CT10	$700k$
341000	$ggH\ H \rightarrow \gamma\gamma$	POWHEG+Pythia8	CT10	$1M$
364352	$QCD\ \gamma\gamma + jets$	Sherpa	NNPDF30	87.8M (56.8M)

Table 1: Monte Carlo samples used in this analysis,$m_H = 125GeV$. The background sample event number is different in 2 scenarios(87.8M for ITk, and 56.8M for rmfwdPU in the following), due to the run condition of jobs.

3 Analysis strategy

3.1 Event reconstruction

The event reconstruction is performed by the Upgrade Performance Function introduced in Chapter 2. Totally 4 kinds of particles in final state are considered:

- Photon.
  A "TightPhoton" working point is chosen, with corresponding photon efficiency and energy resolution. And it should have transverse momentum $p_T > 25GeV$, pseudo rapidity $|\eta| < 2.37$, within the calorimeter cover range. Also the gap [1.37, 1.52] is excluded. The fake photon from electron, muon and jet are considered with the parameterized fake rate.

- Jet.
  The jet includes hard-scatter jet and pile-up jet, respectively comes from truth-level hard-scatter MC event and pile-up library. A track confirmation is required so that a proper efficiency could be used from the performance function. For this reason, the rapidity range for jet is set to $|\eta| < 3.8$, within the ITk design range. And p_T should be larger than 30GeV. For the pileup jets a rejection factor 50 is set by default, which means we suppose only 2% of pileup jets can remain in final state.
- Lepton.
  Electron and muon contribute to the fake photon, so they are also under consideration. They get their efficiency as "LooseElectron" and "TightMuon", and should be in the corresponding detector range, with $p_T > 10GeV$. The fake electron from hard-scatter jet is included as well.

Besides, we set two scenarios to investigate the possible influence of forward pileup to our analysis:

1) ITk scenario. The default one, with the setting above.
2) Remove forward PU(rmfwdPU). We suppose no pileup jet in forward region [2.5, 3.8], to see the influence of pile-up, and also is the highest potential of HGTD.

3.2 Event selection

In this VBF $H \rightarrow \gamma\gamma$ channel analysis, several pre-selections are applied to focus on the final state:

- $N_{photon} \geq 2$
- $N_{jet} \geq 2$
- $\frac{pT_{leading\ photon}}{m_{\gamma\gamma}} > 0.35$
- $\frac{pT_{sub-leading\ photon}}{m_{\gamma\gamma}} > 0.25$
- $105GeV < m_{\gamma\gamma} < 160GeV$
- $\Delta\eta_{jj} > 2$
- $|\eta_{\gamma\gamma}^{Zepp}| = |\eta_{\gamma\gamma} - 0.5(\eta_{j1} + \eta_{j2})| < 5$

The efficiency in each step is listed in Table 2 and Table 3. After the preselection the VBF significance can be calculated with:

$$\sigma_{VBF} = \sqrt{2 \times \{(N_{VBF} + N_{ggH} + N_{background}) \times \ln(1 + \frac{N_{VBF}}{N_{ggH} + N_{background}}) - N_{VBF}\}} \quad (1)$$

In ITk scenario $\sigma = 4.62 \pm 0.03$, and in rmfwdPU scenario $\sigma = 5.02 \pm 0.03$. Some improvement has been shown in clean forward condition.

cut flow	ggH	VBF	background
2 photon	23.082%	26.877%	19.277%
Relative p_T	93.240%	89.157%	85.522%
Mass window [105,160]	97.363%	95.403%	57.394%
2 jets	13.059%	41.212%	11.924%
$\Delta\eta_{jj}$	40.177%	74.984%	36.433%
η^{Zepp}	99.572%	99.921%	98.604%
Total	1.095%	7.059%	0.405%
Scale to $3ab^{-1}$	4075.59	2056.55	855432.8
Mass window [120,130]	3833.98	1935.33	171160.76
significance	4.62 ± 0.03		

Table 2: Cutflow and event number for signal and background in ITK scenario.

cut flow	ggH	VBF	background
2 photon	23.082%	26.877%	19.275%
Relative p_T	93.240%	89.157%	85.535%
Mass window [105,160]	97.363%	95.403%	57.388%
2 jets	12.424%	40.413%	11.421%
$\Delta\eta_{jj}$	38.418%	75.279%	34.506%
η^{Zepp}	99.750%	99.930%	98.874%
Total	0.998%	6.950%	0.369%
Scale to $3ab^{-1}$	3714.08	2024.80	778098.6
Mass window [120,130]	3605.01	2007.17	155367.78
significance	5.02 ± 0.03		

Table 3: Cutflow and event number for signal and background in rmfwdPU scenario.

3.3 Multi-Variable Analysis optimization

In order to have better signal background discrimination, an additional Boost Decision Tree (BDT) method is used in this work. This is inherited from RunII $36.1 fb^{-1}$ $H \rightarrow \gamma\gamma$ analysis. So the same variables are chosen for the BDT input, and their separation power have been studied in previous work. After analyzing the input variables and their correlation, BDT returns a discriminating response that can be used to separate signal and background, and to split events into several categories.

Variables	Definition	Separation power
m_{jj}	Invariant mass of dijet	0.256
$\Delta\eta_{jj}$	Pseudo-rapidity separation of dijet	0.130
$\Delta\Phi_{\gamma\gamma,jj}$	Azimuthal angle between diphoton and dijet system	0.199
p_{Tt}	Diphoton p_T projected perpendicular to the diphoton thrust axis	0.235
$\Delta R^{min}_{\gamma,j}$	Minimum ΔR between one of the two leading photons and the corresponding leading jets	0.185
$\eta^{Zeppenfeld}$	$\lvert\eta_{\gamma\gamma} - 0.5 * (\eta_{j1} + \eta_{j2})\rvert$	0.126

Table 4: Variables used for VBF selection and their definitions.Separation power values are from Run II result

Here the training of BDT is performed with VBF MC as signal, scaled QCD $\gamma\gamma$ MC sample as background. The gluon-fusion process is abandoned in training since its cross section is extremely low comparing with the QCD process. The influence has been proved negligible by previous study. The final statistic uncertainty is dominant by the QCD background, so half of VBF events are separated for BDT training and testing, the others are used for following analysis. The QCD process event number used for training and test is kept in the same level as VBF (about 3% of total sample), to reduce the statistic uncertainty. Events with $\Delta\Phi_{\gamma\gamma,jj} > 2.94$ are merged into a single bin in training to avoid some large systematic uncertainty showed in RunII study.

6 variables listed in Table3 are put into the BDT, and the configuration of BDT is not re-optimized in this work. With the output of BDT response, an individual optimization into 3 categories is performed: first, scan the BDT cut and define the tight category by cutting at the point with highest VBF significance. Then remove the events in tight category and do another scan to obtain the BDT cut for loose category. The definition and event number after scaled to $3000 fb^{-1}$ for each sample are showed in Table 5 and Table 6.

	tight	loose	Rest
VBF	482.88 ± 6.34	555.79 ± 6.80	896.66 ± 8.64
ggF	214.82 ± 8.94	561.06 ± 14.45	3058.1 ± 33.74
Background	1054 ± 50	6509 ± 125	163606 ± 627
Significance in category	12.85 ± 0.28	6.53 ± 0.10	2.19 ± 0.02
S/B	0.383 ± 0.016	0.079 ± 0.0017	0.0054 ± 0.00006
Combined significance	14.58 ± 0.25		
Combined S/B	0.391 ± 0.016		
BDT cut	[0.9, 1]	[0.65, 0.9]	[-1, 0.65]

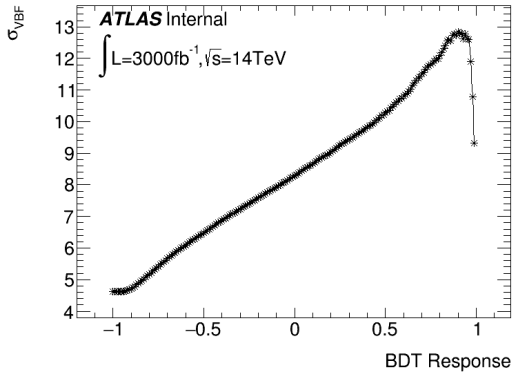
Table 5: Event number in 3 BDT categories for ITK scenario

	tight	loose	Rest
VBF	468.56 ± 6.24	677.00 ± 7.51	861.62 ± 8.47
ggF	186.15 ± 8.32	656.00 ± 15.63	2762.86 ± 32.07
Background	772 ± 53	7000 ± 161	147594 ± 740
Significance in category	14.10 ± 0.39	7.63 ± 0.11	2.22 ± 0.02
S/B	0.488 ± 0.028	0.088 ± 0.002	0.0057 ± 0.00006
Combined significance	18.18 ± 0.34		
Combined S/B	0.496 ± 0.028		
BDT cut	[0.92, 1]	[0.62, 0.92]	[-1, 0.62]

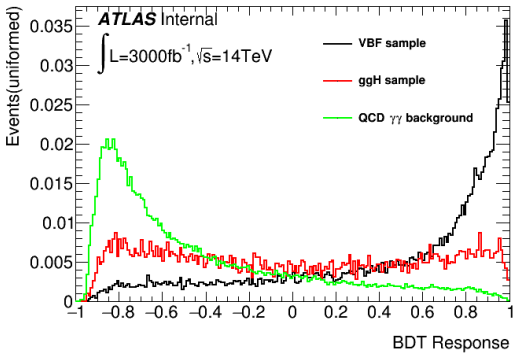
Table 6: Event number in 3 BDT categories for rmfwdPU scenario

4 Result and discussion

After the categorization a combined significance and signal-background ratio can be calculated. These two benchmarks both show a significant improvement when removing the forward pile-up jets. In order to have a better understanding, some further studies are performed.



(a) VBF significance in different BDT cut value

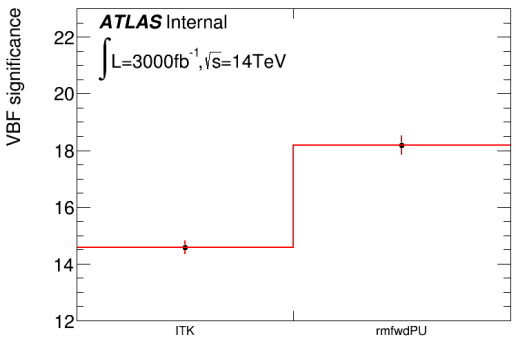


(b) BDT response distribution in 3 samples

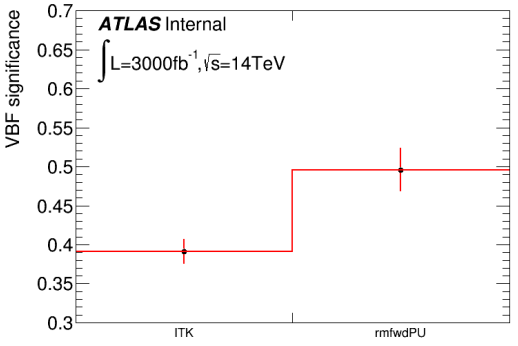
Figure 1: VFB significance in "tight" category in different BDT cut (a) and the distribution of BDT response value in 3 samples (b). Gluon fusion sample is not used in the training, but the BDT response can also be obtained.

scenario	σ_{VBF}	Improvement	$S/(B_1+B_2)$	Improvement
ITK	14.58 ± 0.25	0	0.391 ± 0.016	0
rmfwdPU	18.18 ± 0.35	24.7%	0.496 ± 0.028	26.85%

Table 7: Combined VBF significance and signal-background ratio in ITK and remove forward PU(rmfwdPU) scenario. Comparing with ITK, rmfwdPU shows about 25% improvement in this two benchmarks.



(a) VBF significance



(b) Signal-background ratio

Figure 2: VFB significance (a) and signal-background ratio (b) in 2 scenarios, corresponding to the value in Table 7.

The inclusive categories show about 8% of QCD events and 9% of gluon fusion events are removed in "remove forward PU(rmfwdPU)" scenario(showed in Table 2 and Table 3), and this is supposed to the main reason of improvement in VBF significance. In order to include the contribution of BDT, we choose tight and loose categories for the following discussion. They give the most contribution in combined significance. However, since the optimization in 2 scenarios are performed separately, they have different definition in "tight+loose", the exact number should not be directly compared.

The total events are separated into 10 components with the jet property: the chosen jet is in forward region (F) or central region (C), and is pileup jet (P) or hard-scatter jet (H). In Table 8 and Table 9 the event number and relative percentage of 10 components are presented.

	ggH		VBF		QCD	
CH+CH	255.772	32.97%	205.103	19.74%	1428.57	23.67%
CH+CP	18.6151	2.40%	6.49272	0.62%	197.21	3.27%
CP+CP	0.372303	0.05%	0	0.00%	4.81	0.08%
CH+FH	378.26	48.75%	623.468	59.99%	2905.24	48.15%
CH+FP	31.6458	4.08%	6.07652	0.58%	447.33	7.41%
CP+FH	19.3598	2.50%	8.7402	0.84%	254.93	4.22%
CP+FP	1.11691	0.14%	0.08324	0.01%	26.455	0.44%
FH-FH	56.9624	7.34%	184.293	17.73%	481	7.97%
FH-FP	12.6583	1.63%	4.91116	0.47%	259.74	4.30%
FP-FP	1.11691	0.14%	0.08324	0.01%	28.86	0.48%
total	775.8795	100.00%	1039.251	100.00%	6034.145	100.00%

Table 8: ITK scenario event number in 10 jet property components. C: Central. F: Forward. H: Hard-scatter jet. P: Pile-up jet. BDT tight+loose category.

	ggH		VBF		QCD	
CH+CH	275.504	32.71%	212.761	18.55%	1448.85	23.13%
CH+CP	23.8274	2.83%	6.90892	0.60%	234.045	3.74%
CP+CP	0.372303	0.04%	0	0.00%	18.575	0.30%
CH+FH	434.478	51.59%	684.732	59.70%	3454.95	55.16%
CH+FP	0	0.00%	0	0.00%	0	0.00%
CP+FH	24.9443	2.96%	10.738	0.94%	341.78	5.46%
CP+FP	0	0.00%	0	0.00%	0	0.00%
FH-FH	83.0236	9.86%	231.907	20.22%	765.29	12.22%
FH-FP	0	0.00%	0	0.00%	0	0.00%
FP-FP	0	0.00%	0	0.00%	0	0.00%
total	842.1496	100.00%	1147.047	100.00%	6263.49	100.00%

Table 9: rmfwdPU scenario event number in 10 jet property components. C: Central. F: Forward. H: Hard-scatter jet. P: Pile-up jet. BDT tight+loose category.

Some information could be extracted from these two table:

- Events with at least one forward PU (sum of CH-FP, CP-FP, FH-FP, FP-FP) accounts for 1.1% in VBF sample, but 6% in gluon fusion and 12.6% in QCD sample. This is decided by the physics process and our analysis. The selection and BDT are focus on VBF events, so the signal can easily pass these criteria, while the background need some help from pile-up. In rmfwdPU scenario these events are removed. It does not affect VBF sample much, but decreases considerable background event number.
- In all pile-up events, the event with a forward pile-up jet accounts for 35% in VBF, 42% in gluon-fusion and 51% in QCD (sum of row 5,7,9,10 over sum of row 2,3,5,6,7,9,10). When including a pile-up, the QCD process tends to have a forward pile-up jets, while the VBF process prefer a central

one. For this reason, a higher pile-up rejection in forward region means a lot in VBF analysis. The leading jet η distribution in Plot 3 and Plot 4can give a visualization of this.

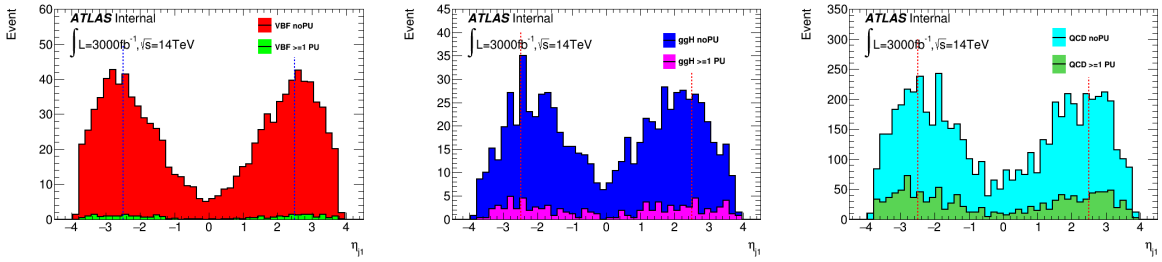


Figure 3: Leading jet η distribution in VBF(left), gluon-fusion(middle) and QCD $\gamma\gamma + jets$ (right) sample in ITK scenario. Two dash lines in $|\eta| = 2.5$ help to see the forward region.

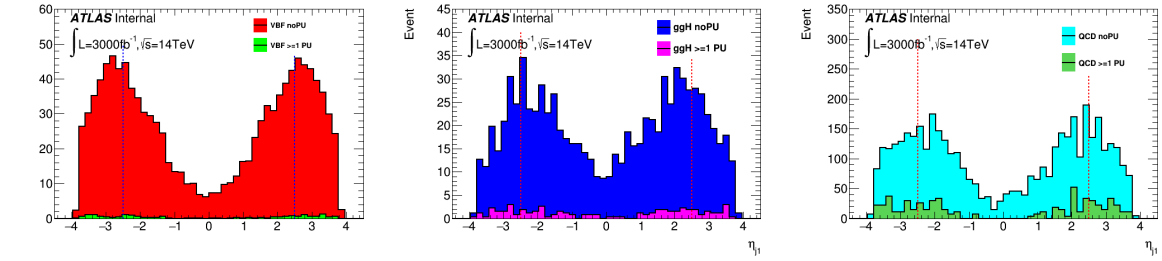


Figure 4: Leading jet η distribution in VBF(left), gluon-fusion(middle) and QCD $\gamma\gamma + jets$ (right) sample in rmfwdPU scenario. Two dash lines in $|\eta| = 2.5$ help to see the forward region.

One more conclusion we want to draw from comparing the exact event number in inclusive categories in Table 10 and Table 11. In ITK scenario, after subtracting the forward PU event the event number is a bit lower than the total event number in rmfwdPU scenario. Meanwhile those no-forward-PU events have a tiny increase. In principle removing forward pile-up jets would not directly influence the central region, but it can change some events from Forward-PU to Forward-HS/Central-HS/Central-PU, i.e. correct some pile-up events. This effect is not very obvious but indeed exist. In this note the exact percentage of this "correction" is not estimated, due to the difficulty of tracing every event before and after the removing.

5 Conclusion

In this note, an estimation of the forward pile-up influence to VBF $Higgs \rightarrow \gamma\gamma$ analysis in HL-LHC condition is performed. By comparing the default condition with ITk only and manually removing forward pile-up jets, we observed a 24.7% improvement in VBF significance and 26.8% improvement in VBF signal-background ratio. They mainly come from the different kinematic in VBF and gluon fusion/QCD process, the pile-up jets are more likely to mix in the background to mimic a VBF event. This indicates a potential of HGTD. In the most ideal condition the HGTD can have a similar contribution in VBF analysis. The specific performance needs further study with a better HGTD simulation.

	ggH		VBF		QCD	
CH+CH	1417.35	34.78%	585.62	28.48%	341385.2	39.91%
CH+CP	441.92	10.84%	59.60	2.90%	97008.49	11.34%
CP+CP	74.46	1.83%	2.91	0.14%	17684.48	2.07%
CH+FH	1436.71	35.25%	1073.39	52.19%	270060.5	31.57%
CH+FP	277.36	6.81%	29.92	1.46%	59229.65	6.92%
CP+FH	225.24	5.53%	63.43	3.08%	34455.03	4.03%
CP+FP	72.60	1.78%	2.91	0.14%	19733.6	2.31%
FH-FH	83.02	2.04%	228.28	11.10%	8047.364	0.94%
FH-FP	38.72	0.95%	10.28	0.50%	5882.801	0.69%
FP-FP	8.19	0.20%	0.21	0.01%	1945.701	0.23%
total	4075.59	100.00%	2056.55	100.00%	855432.8	100.00%

Table 10: ITK scenario event number in 10 jet property components, no BDT categorization

	ggH		VBF		QCD	
CH+CH	1425.92	38.39%	588.62	29.07%	342987.9	44.08%
CH+CP	452.35	12.18%	61.43	3.03%	100441.9	12.91%
CP+CP	76.69	2.06%	3.04	0.15%	18306.41	2.35%
CH+FH	1446.02	38.93%	1077.43	53.21%	273117.4	35.10%
CH+FP	0.00	0.00%	0.00	0.00%	0	0.00%
CP+FH	228.97	6.16%	64.63	3.19%	34944.55	4.49%
CP+FP	0.00	0.00%	0.00	0.00%	0	0.00%
FH-FH	84.14	2.27%	229.65	11.34%	8300.491	1.07%
FH-FP	0.00	0.00%	0.00	0.00%	0	0.00%
FP-FP	0.00	0.00%	0.00	0.00%	0	0.00%
total	3714.08	100.00%	2024.80	100.00%	778098.6	100.00%

Table 11: rmfwdPU scenario event number in 10 jet property components, no BDT categorization

List of contributions