VBS VVH All-Hadronic First steps January 25th, 2023

P. Chang, L. Giannini, J. Guiang, Y. Xiang, E. Zenhom









All-Hadronic VBS VVH

- Targeting the following final states:
 - WWH → qq qq qq bb
 - $ZZH \rightarrow qq qq qq b\bar{b}$
 - WZH → qq qq qq bb
- Sensitive to C_{2V} , C_3 , and C_V in principle
- BSM signature:
 - W/Z/H jets with large p⊤
 - VBS jets with large $\Delta \eta_{jj}$, M_{jj}





All-Hadronic VBS VVH

- One interesting N_{jets} vs. N_{fatjets} channel:
 - 3 AK8 fat jets, ≥ 2 AK4 jets (right)
 - 2 AK8 fat jets, \geq 4 AK4 jets
 - 2 AK8 fat jets, 3 AK4 jets
- Practically zero signal with N_{fatjets} > 3 (backup)
- From previous studies, N_{fatjets} == 1 channel not worthwhile pursuing right now





Analysis Skim

Cut	QC	D TTbarJets	TotalBkg	Eff.*	VBSVVH ($C_{2V} = 2$)	Eff. ³
Skim	138,2	60K 777K	139,036K		194	
	Object	Sele	ction			
	Leptons (µ, e)	== 0	veto*			
	Fat Jets	≥ 2 w/ p⊤ AND r AND mase AND M _{SD} AND fat	> 300 GeV y < 2.5 s > 50 GeV > 40 GeV jet ID > 0			
	Jets	≥ 2 w/ p⊤ AND passe AND ΔR(jet,	> 20 GeV s tight jet ID fat jet) > 0.8			
	Other	\geq 1 pair of jets v AND 2	v/ M _{jj} > 500 Ge \η _{jj} > 3	٧		
		*Using the ttH lept	on ID			

Yields scaled to lumi $\times \sigma$. rounded for readability

*ParticleNet XWqq = mass-decorrelated, W-to-qq-like tagger **Used in Yanxi's analysis, but can easily be changed to sync with Yifan

Object	Selections
AK8 jets (same as skim)	• $p_T > 300 \text{ GeV}$ • Fat jet ID > 0 • $ \eta < 2.5$ • Mass > 50 GeV • M _{SD} > 40 GeV
H→bb̄ fat jet	 Has max(ParticleNet Xbb)
V→qq fat jets	 Leading in ParticleNet XWqq
AK4 jets (same as skim)	 p_T > 20 GeV Passes tight jet ID ΔR(AK4, AK8) > 0.8
VBS (AK4) jets	 p_T > 30 GeV For > 2 candidates**: Take pair with maximum Δ

Basic Cutflow

Cut	QCD	TTbarJets	TotalBkg	Eff.*	VBSVVH ($C_{2V} = 2$)	Eff.*
Skim	138,260K	777K	139,036K		194	
HLT + MET Filters	89,485K	596K	90,082K	35%	186	4%
Exactly 3 fat jets	681K	15K	697K	99%	36	81%
Object selection	284K	10K	294K	58%	21	42%
 Take Object Selection as F S_T = p_T(H→bb̄) + p_T(V→qq Plot important variables or 	Preselection) + $p_T(V \rightarrow c)$ In the follow	n (q) plotted o (ing slides	on right	Strand 104 103 102 101 100 100 100 100 100 100	Simulation Preliminary	138 fb ⁻¹ (13 TeV) [10126.5 events] 84257.1 events] ackground [294383.6 events] gnal [20.7 events]
*eff = 1 - after/before				10	000 2000 3000 4 TT	000 5000 S⊤[GeV]

Yields scaled to lumi $\times \sigma$, rounded for readability

Hbb Fat Jet

Leading Vqq Fat Jet

Trailing Vqq Fat Jet

VBS Jets

Cut-based SR

Cut	QCD	TTbarJets	TotalBkg	Eff.*	VBSVVH ($C_{2V} = 2$)	Eff. ³
Skim	138,260K	777K	139,036K		194	_
HLT + MET Filters	89,485K	596K	90,082K	35%	186	4%
Exactly 3 fat jets	681K	15K	697K	99%	36	81%
Object selection	284K	10K	294K	58%	21	42%
M _{jj} > 500 GeV	85K	ЗK	88K	70%	18	12%
Δη _{jj} > 500 GeV	80K	2.8K	83K	6%	18	0%
H→b̄b fat jet PNet Xbb > 0.9	3.9K	910	4.8K	94%	12	36%
V→qq fat jet PNet XWqq > 0.9	7	8	15	100%	4	68%
S⊤ > 1300 GeV	7	5	12	21%	4	5%
M _{SD} (H→b̄b fat jet) < 150 GeV	1	2	4	69%	3	2%
M _{SD} (V→qq fat jet) < 120 GeV	0	1	1	62%	3	6%

- With basic selection, reach 3 sig. vs. 1 bkg. (take this as baseline) •
 - For $C_{2V} = 2$, this is ~12 sig. vs. 1 bkg.
- List of HLTs in backup

Yields scaled to lumi $\times \sigma$, rounded for readability

BDT

Cut	QCD	TTbarJets	TotalBkg	Eff.*	VBSVVH ($C_{2V} = 3$)	Eff.
BDT Preselection	443	104	547		9	

- First apply BDT preselection:
 - S_T > 1300 GeV AND H→bb̄ fat jet PNet Xbb > 0.5 AND V→qq fat jet PNet XWqq > 0.5 AND M_{SD}(H→bb̄ fat jet) < 150 GeV AND M_{SD}(V→qq fat jet) < 120 GeV
- Train BDT using XGBoost with the input features listed on the right
 - 40/60 test/train split
 - Hyperparameters listed in backup

Yields scaled to lumi $\times \sigma$, rounded for readability

Object	Features
H→bb̄ fat jet	 ParticleNet Xbb score pT M_{SD}
V→qq fat jets	 ParticleNet XWqq score pT MsD
Other	 MET N_{jets} (AK4)

BDT

Cut	QCD	
BDT Preselection	443	

- First apply BDT preselection:
 - S_T > 1300 GeV AND $H \rightarrow b\bar{b}$ fat jet PNet Xbb > 0.5 AND V \rightarrow qq fat jet PNet XWqq > 0.5 AND M_{SD}(H→bb̄ fat jet) < 150 GeV AND $M_{SD}(V \rightarrow qq \text{ fat jet}) < 120 \text{ GeV}$
- Train BDT using XGBoost with the input features listed on the right
 - 40/60 test/train split
 - Hyperparameters listed in backup

BDT: Result

No sign of over-fitting, sig-to-bkg looks promising (still no VBS cuts applied!)

Selection	Sig. (C _{2V} = 2)	Total
$M_{jj} > 500 \text{ GeV}$ and $ \Delta \eta_{jj} > 3$	8.64	168
M_{jj} > 500 GeV and $ \Delta \eta_{jj} $ > 3 and BDT > 0.80	6.74	1.9
M_{jj} > 500 GeV and $ \Delta \eta_{jj} $ > 3 and BDT > 0.85	6.10	0.7
M_{jj} > 500 GeV and $ \Delta \eta_{jj} $ > 3 and BDT > 0.86	5.95	0.6
M_{jj} > 500 GeV and $ \Delta \eta_{jj} $ > 3 and BDT > 0.87	5.76	0.3
M_{jj} > 500 GeV and $ \Delta \eta_{jj} $ > 3 and BDT > 0.88	5.56	0.3
M_{jj} > 500 GeV and $ \Delta \eta_{jj} $ > 3 and BDT > 0.89	5.28	0.2
M_{jj} > 500 GeV and $ \Delta \eta_{jj} $ > 3 and BDT > 0.90	5.01	0.2

BDT: Result

We can reach 5 sig. vs. ~0 bkg.

Summary

- VBS VVH analysis has been resurrected
 - Basic cut-based SR defined: 3 signal vs. 1 bkg.
 - BDT-based SR defined: 5 signal vs. 0 bkg. (with M_{ii} and $\Delta \eta_{ii}$ cuts also applied)
 - VBS-independent, so Yifan-like bkg. estimation strategy could be explored
 - Lots of shared code with VBS WH
 - Scale factors, systematics, etc. should be faster to finish
- Next steps: •
 - Bring in subleading background samples
 - Look into bkg. estimation strategies
 - Look into scale factor for BDT

Backup

Nfatjets

- Plotting sum of all VBS VVH signals here
- Selections applied: skim, HLT triggers, • gen-level H, W, Z decay hadronically
- Practically 0 events with more than three fatjets

ParticleNet Taggers

Mass-correlated taggers more sharply peaked, but MD taggers may be sufficient

1000

500

1500

2000

500

1000 1500 2000 2500

500 1000 1500 2000 2500 3000

BDT

Parameter	Value	Description*
objective	binary:logistic	Learning objective; 'binary:logisti specifies logistic regression for binary classification, output probability
eta	0.1	Step size shrinkage (alias: learning_rate)
<pre>max_depth</pre>	3	Max. depth of tree: larger = more complex = more prone to overfitt
verbosity	1	0 (silent), 1 (warning), 2 (info), 3 (debug)
nthread	8	Number of parallel threads
eval_metric	auc	Evaluation metrics for validation data. 'auc' = Area Under the Curv
subsample	0.6	Subsample ratio of the training instances
alpha	8.0	L1 regularization term on weights Larger = more conservative
gamma	2.0	Min. loss rediction to make leaf (alias: min_split_loss)
lambda	1.0	L2 regularization term on weights Larger = more conservative
<pre>min_child_weight</pre>	1.0	Minimum sum of instance weight (hessian) needed in a child
colsample_bytree	1.0	The subsample ratio of columns when constructing each tree
<pre>scale_pos_weight</pre>	58.0	Control the balance of positive an negative weights, useful for unbalanced classes

H_T-binned QCD Samples

- Ignoring the $H_T \in [50, 100)$ GeV sample due to low stats and irrelevant anyway
- Cross sections taken from AN-21-045
 - Measurement of the Higgs boson production via Vector Boson Fusion process with subsequent decay of the Higgs boson into a pair of bottom quarks

Table 3: List of simulated samples along with the rates used in this analysis, where [*] denotes the ultra legacy production campaign of the simulated samples, for 2016-APV, 2016-NonAPV and 2018.

Process	Sample	$\sigma \times BR$
$VBFH ightarrow bar{b}$	- VBFHToBB_M-125_TuneCP5_13TeV-powheg-pythia8/[*]/MINIAODSIM	2.233
	VBFHToBB_M-125_dipoleRecoilOn_TuneCP5_13TeV-powheg-pythia8/[*]/MINIAODSIM	
	/VBFHToBB_M-125_TuneCH3_13TeV-powheg-herwig/[*]/MINIAODSIM	
ggH→bb	/GluGluHToBB_M-125_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/MINIAODSIM	28.293
	/GluGluHToBB_M-125_TuneCP5_13TeV-powheg-pythia8/[*]/MINIAODSIM	
QCD multijet	QCD_HT100to200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	2784988
in H_T bins	QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	1716992
	QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	351302
	QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	31630
	QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	6802
	QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	1206
	QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	98.71
	QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	20.2
$t\bar{t} + X$	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[*]/MINIAODSIM	88.29
	TTToHadronic_TuneCP5_13TeV-powheg-pythia8/[*]/MINIAODSIM	377.96
	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/[*]/MINIAODSIM	365.34
single top	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/MINIAODSIM	34.91
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/MINIAODSIM	34.40
	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/MINIAODSIM	115.30
	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/MINIAODSIM	69.09
QCD Z + jets	ZJetsToQQ_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	973.70
	ZJetsToQQ_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	110.78
	ZJetsToQQ_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	25.34
	ZJetsToQQ_HT-800toInf_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	19.00
EWK Z + jets	EWKZ2Jets_ZToQQ_TuneCP5_13TeV-madgraph-pythia8/[*]/MINIAODSIM	9.92
QCD W + jets	WJetsToQQ_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	2549.00
-	WJetsToQQ_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	276.50
	WJetsToQQ_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	59.25
	WJetsToQQ_HT-800toInf_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/MINIAODSIM	28.75

VBSVVH All-Hadronic Cutflow

	QCD		TTbarJets		TotalBkg		VBSVVH	
cut	raw	wgt	raw	wgt	raw	wgt	raw	wgt
Bookkeeping	23565646	138259597.33	12011377	776744.40	35577023	139036341.73	358720	193.95
SaveSystWeights	23565646	138259597.33	12011377	776744.40	35577023	139036341.73	358720	193.95
PassesEventFilters	23444188	138064398.83	11985758	775134.60	35429946	138839533.43	354949	191.95
PassesTriggers	20167657	89485412.41	9287547	596242.20	29455204	90081654.61	348495	186.37
SelectLeptons	20167657	89485412.41	9287547	596242.20	29455204	90081654.61	348495	186.37
NoLeptons	20026783	88907595.31	9090001	583578.20	29116784	89491173.51	325774	174.06
SelectFatJets	20026783	88907595.31	9090001	583578.20	29116784	89491173.51	325774	174.06
Exactly3FatJets	703693	681151.29	232984	15361.15	936677	696512.44	65748	35.70
AllMerged_SelectVVHFatJets	703693	681151.29	232984	15361.15	936677	696512.44	65748	35.70
AllMerged_SelectJets	703693	681151.29	232984	15361.15	936677	696512.44	65748	35.70
AllMerged_SelectVBSJets	340870	284257.13	153998	10126.48	494868	294383.61	38207	20.74
AllMerged_SaveVariables	340870	284257.13	153998	10126.48	494868	294383.61	38207	20.74
AllMerged_STGt800	340870	284257.13	153998	10126.48	494868	294383.61	38207	20.74
AllMerged_MjjGt500	110726	84779.50	46178	2991.11	156904	87770.61	33788	18.32
AllMerged_detajjGt3	99537	79877.71	43521	2817.57	143058	82695.27	33674	18.26
AllMerged_XbbGt0p9	4990	3879.50	14077	910.12	19067	4789.61	22200	11.73
AllMerged_XWqqGt0p9	10	7.39	123	7.94	133	15.33	6801	3.70
AllMerged_STGt1300	10	7.39	73	4.74	83	12.13	6441	3.51
AllMerged_HbbMSDLt150	5	1.44	37	2.31	42	3.75	6315	3.44
AllMerged_VqqMSDLt120	1	0.20	19	1.20	20	1.41	5874	3.24

VBSVVH All-Hadronic Triggers

Year	HLT paths
2016	HLT_PFHT800 HLT_PFHT900 HLT_AK8PFHT650_TrimR0p1PT0p03Mass50 HLT_AK8PFHT700_TrimR0p1PT0p03Mass50 HLT_AK8PFJet450 HLT_AK8PFJet360_TrimMass30 HLT_AK8DiPFJet280_200_TrimMass30 HLT_AK8DiPFJet280_200_TrimMass30_BTagCSV_p20
2017	HLT_PFHT1050 HLT_AK8PFHT800_TrimMass50 HLT_PFJet320 HLT_PFJet500 HLT_AK8PFJet320 HLT_AK8PFJet500 HLT_AK8PFJet400_TrimMass30 HLT_AK8PFJet420_TrimMass30
2018	HLT_PFHT1050 HLT_AK8PFHT800_TrimMass50 HLT_PFJet500 HLT_AK8PFJet500 HLT_AK8PFJet400_TrimMass30 HLT_AK8PFJet420_TrimMass30

