VBS WH Prospectus

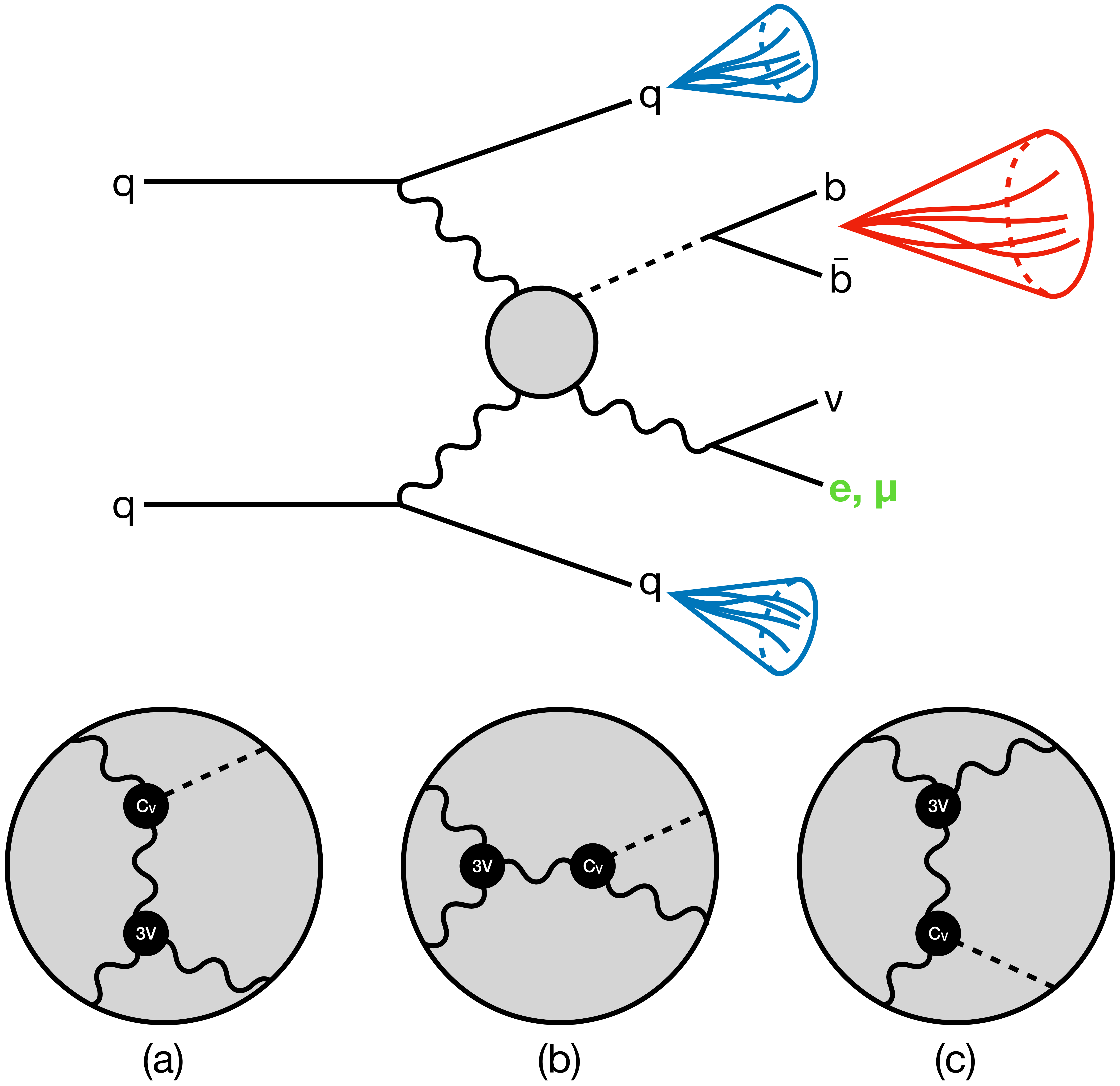
Towards excluding $\lambda_{WZ} = -1$

July 6th, 2022

*P. Chang, **J. Guiang**, F. Würthwein, Y. Xiang*

Target Final State

- Targeting **VBS** WH→$\ell\nu b\bar{b}$
- Sensitive to κ_V → $\lambda_{WZ} = \kappa_Z/\kappa_W$
    - Handle for ruling out λ_{WZ} = -1 (BSM)
- VBS WH BSM kinematics:
    - High-p_T H and W (high S_T)
    - VBS jets with large $\Delta\eta_{jj}$, M_{jj}



VBS WH Cross Sections

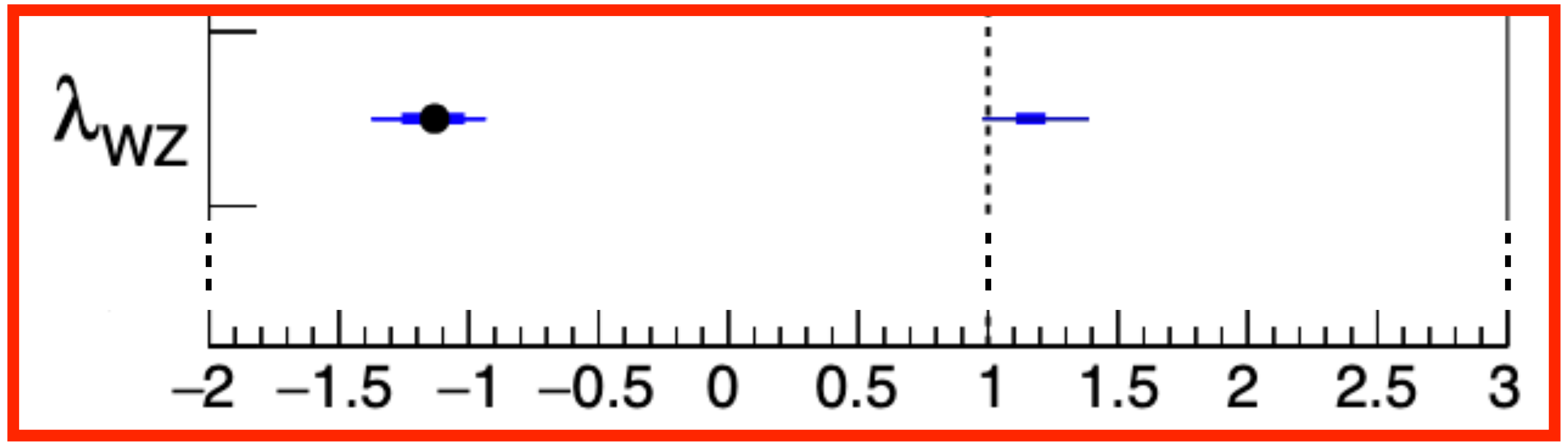
Model	σ [pb]
$\kappa_W = \kappa_Z = +1$ (SM)	0.075
$\kappa_W = -1$	0.433
$\kappa_Z = -1$	0.433

×6

- Setting $\kappa_W = -1$ or $\kappa_Z = -1$ equivalently enhances cross section by a factor of 6
- These numbers are taken from MadGraph*: `generate p p > w h j j QCD=0`
  - Includes gen-level filters (e.g. jet $p_T > 10$ GeV)
  - Generated 10,000 events for each to obtain xsec value
- **Optimizing for $\kappa_W = -1$** (kinematics are equivalent to $\kappa_Z = -1$)
  - Generated 100k UL NanoAOD events for 2016 pre-VFP, 2016 post-VFP, 2017, and 2018

*More details in backup

Current CMS Result



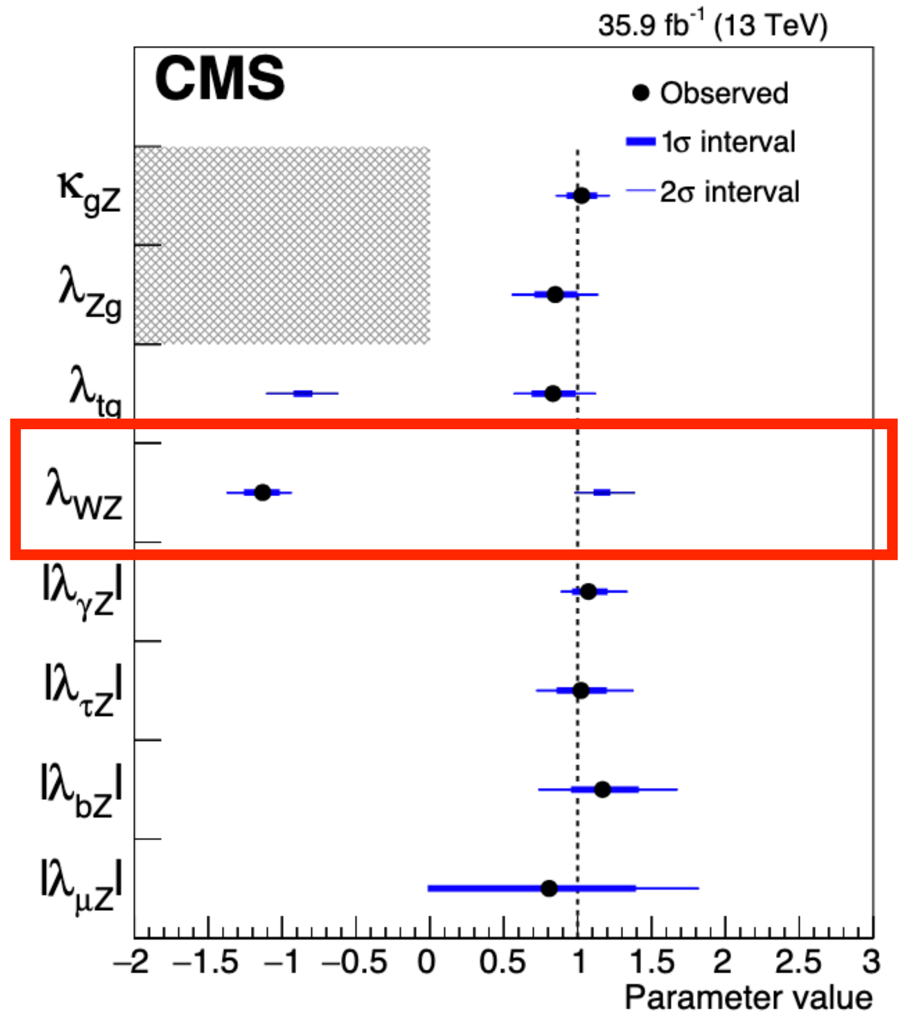


Figure 16: Summary of the model with coupling ratios and effective couplings for the ggH and H $\rightarrow \gamma\gamma$ loops. The points indicate the best fit values while the thick and thin horizontal bars show the 1σ and 2σ CL intervals, respectively. For this model, both positive and negative values of λ_{WZ} and λ_{tg} are considered.

- From 35.9 fb^{-1} Higgs combination: best fit hints at λ_{WZ} = -1 (BSM)
    - https://arxiv.org/pdf/1809.10733.pdf
- **With the handles we have on BSM λ_{WZ} via VBS WH, we can rigorously explore this!**

CMS

UC San Diego

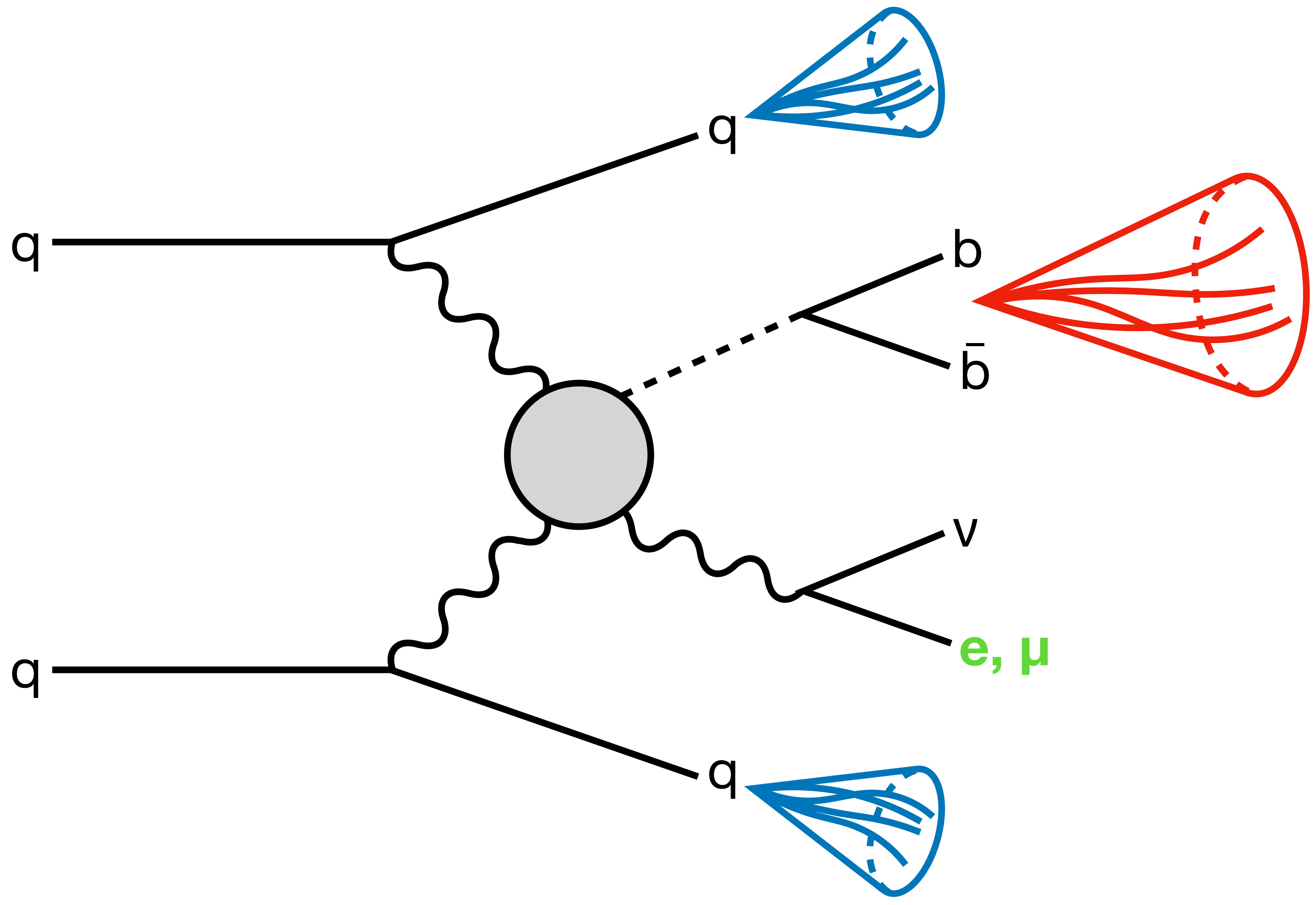
VBS WH Skim

Cut	VH	VV/VVV/VBSWZ	W+Jets	TTbar+X	TTbar1L	TTbar2L	TotalBkg	Eff.	VBSWH (κ_W = -1)	Eff.
Skim AND Postskim	2.8K	1.7K	33K	3.7K	424K	34K	**497K**	—	**60K**	—

Object	Primary Skim	Postskim
Leptons (μ, e)	≥ 1 veto*	== 1 loose AND == 1 tight*
Fat Jets	≥ 1 w/ p_T > 200 GeV AND mass > 10 GeV AND M_{SD} > 10 GeV AND ΔR(fat jet, any veto lepton) > 0.8	≥ 1 w/ p_T > 250 GeV AND mass > 25 GeV AND M_{SD} > 25 GeV AND Hbb vs. QCD > 0.5 AND ΔR(fat jet, tight lepton) > 0.8
Jets	≥ 1 w/ p_T > 20 GeV AND ΔR(jet, any veto lepton) > 0.4	—

*Using the ttH lepton ID

VBS WH Objects



Object	Selections
One **Lepton** (loose veto in postskim)	• Passes ttH tight ID • p_T > 40 GeV
One **Hbb** Ak8 Jet	• ΔR(Ak8 jet, any veto lep) > 0.8 • p_T > 250 GeV • mass > 50 GeV • M_{SD}* > 40 GeV • Has max(PNet Hbb vs. QCD score) • PNet Hbb vs. QCD score > 0.9

Two **VBS** Ak4 jets	• ΔR(Ak4 jet, any veto lep) > 0.4
	• ΔR(Ak4 jet, Hbb Ak8 jet) > 0.8
	• p_T > 30 GeV
	• For > 2 candidates**:
	  • All in same η-hemisphere: leading/trailing (in P) jets
	  • Else: leading (in P) jet from each η-hemisphere

*M_{SD} = soft drop mass
**Hereafter referred to as the “Max-E” selection

VBS WH Preselection

Cut	VH	VV/VVV/VBSWZ	W+Jets	TTbar+X	TTbar1L	TTbar2L	TotalBkg	Eff.*	VBSWH (κ_W = -1)	Eff.*
Skim AND Postskim	4.5K	1.7K	33K	3.7K	424K	34K	501K	—	60K	—
$p_T(\ell)$ > 40 GeV	4.1K	1.3K	26K	2.7K	328K	25K	388K	23%	9.1K	85%
≥ 1 fat jet w/ p_T > 250 GeV AND mass > 50 GeV AND M_{SD} > 40 GeV	4K	1.3K	26K	2.7K	326K	24K	383K	1%	3.3K	64%
max(PNet Hbb vs. QCD) > 0.9	3.2K	693	5.9K	1K	114K	9.5K	135K	65%	1.3K	59%
2 VBS jets (max E)	1K	221	3.2K	899	87K	6.7K	99K	27%	929	31%
M_{jj} > 500 AND $\Delta\eta_{jj}$ > 3	194	36	580	170	18K	1.6K	**20K**	79%	**783**	16%

- Variables left to play with (next slides):
    - VBS: M_{jj}, $\Delta\eta_{jj}$
    - WH: $L_T = p_T(\ell)$ + MET, $S_T = L_T + p_T$(Hbb fat jet), Hbb vs. QCD ParticleNet score

*eff = 1 - after/before

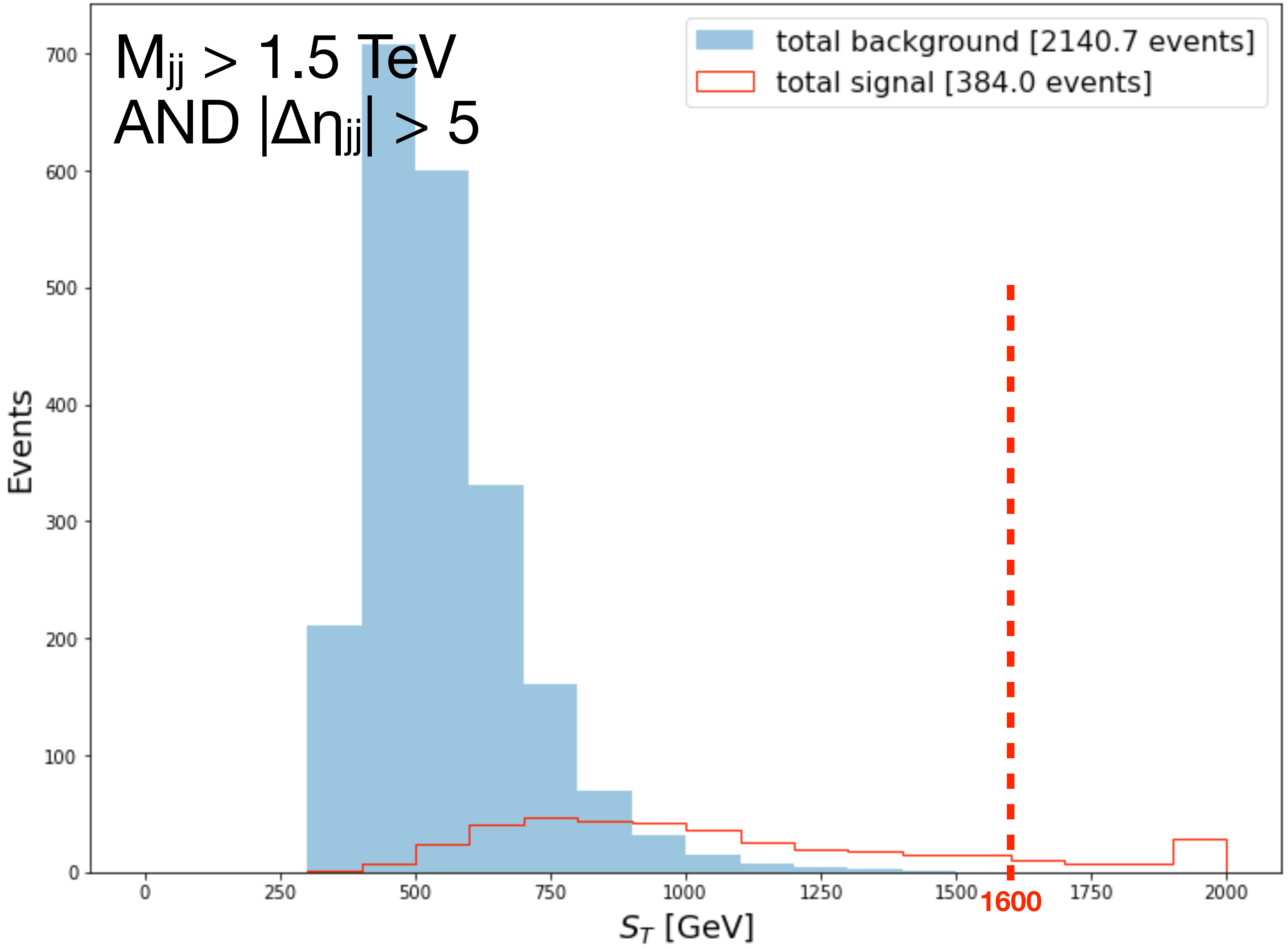
UC San Diego

VBS WH Optimization

Floor: loosely optimized cut-based signal region

Ceiling: BDT-based signal region

VBS WH Cut-based



$S_T = p_T(\ell) + MET + p_T(\text{Hbb Ak8 jet})$

Cut	VBSWH (κ_W = -1)	TotalBkg	S/√B
M_{jj} > 1.5 TeV AND $\Delta\eta_{jj}$ > 5	384.00	2140.68	8.30
S_T > 1 TeV	180.34	29.52	33.19
S_T > 1.1 TeV	144.59	14.24	38.32
S_T > 1.2 TeV	119.09	7.25	44.22
S_T > 1.3 TeV	99.89	3.52	53.26
S_T > 1.4 TeV	82.31	1.47	67.78
S_T > 1.5 TeV	67.60	0.71	80.21
S_T > 1.6 TeV	53.34	0.33	93.44
S_T > 1.7 TeV	43.10	0.11	132.63
S_T > 1.8 TeV	36.05	0.13	101.44
S_T > 1.9 TeV	28.40	0.11	86.11

M_{jj} > 1.5 TeV AND $\Delta\eta_{jj}$ > 5
AND S_T > 1.6 TeV
gives **53 sig vs. ~0 bkg**

UC San Diego

VBS WH BDT: Preselection

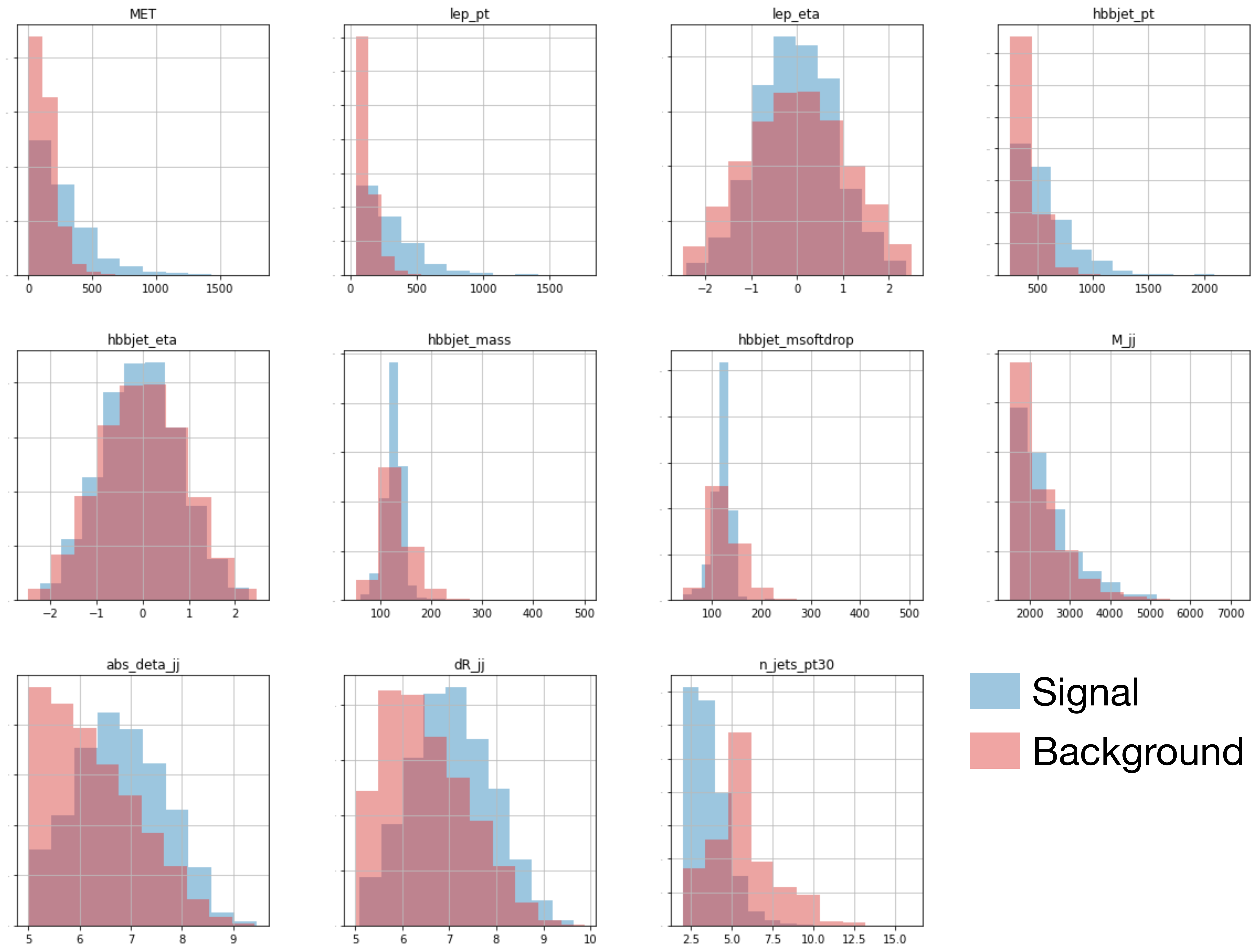
Cut	VH	VV/VVV/VBSWZ	W+Jets	TTbar+X	TTbar1L	TTbar2L	TotalBkg	Eff.*	VBSWH (κ_W = -1)	Eff.*
Skim AND Postskim	2.8K	1.7K	33K	3.7K	424K	34K	499K	—	60K	—
Preselection	194	36	580	170	18K	1.6K	20K	96%	783	99%
M_{jj} > 1.5 TeV AND $\Delta\eta_{jj}$ > 5 AND S_T > 500 GeV	14	2	31	10	1.1K	106	1.2K	94%	376	52%

- Start tighter VBS and S_T cuts after Preselection
- Trained a simple XGBoost BDT to distinguish signal from background
- Split MC into training (60%) and testing (40%) samples
  - Rescale integrals sample-by-sample to match original integral
  - i.e. if Σ_{orig}SampleX = 100.5 weighted events, then rescale event weights such that Σ_{train}SampleX = Σ_{test}SampleX = 100.5

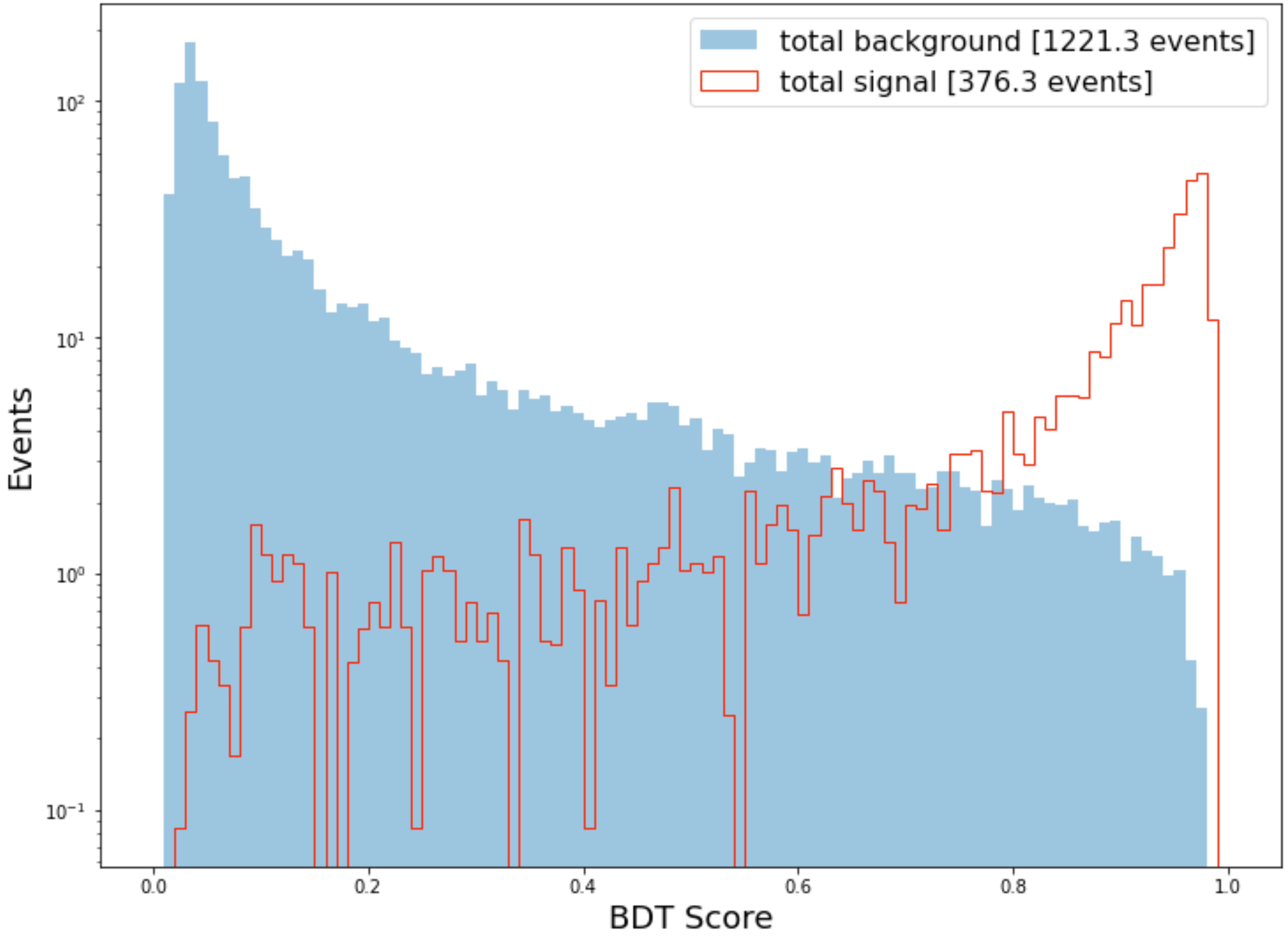
*eff = 1 - after/before

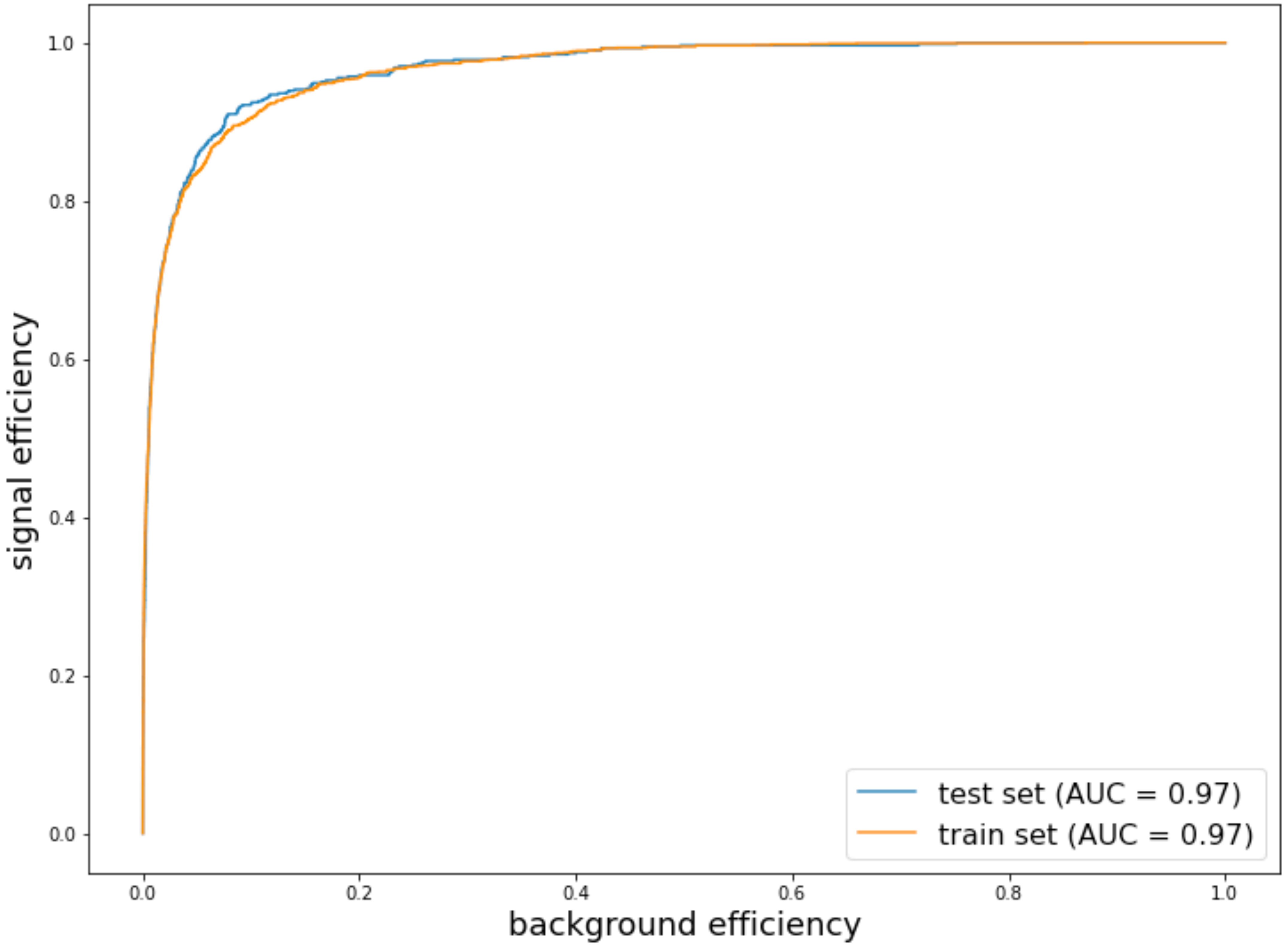
VBS WH BDT: Features

Features	
Object	**Variables**
Lepton	• p_T • η
Hbb Ak8 Jet	• p_T • η • mass • M_{SD}
VBS jets	• M_{jj} • $\lvert\Delta\eta_{jj}\rvert$ • ΔR_{jj}
Other	• Number of jets w/ p_T > 30 GeV • MET



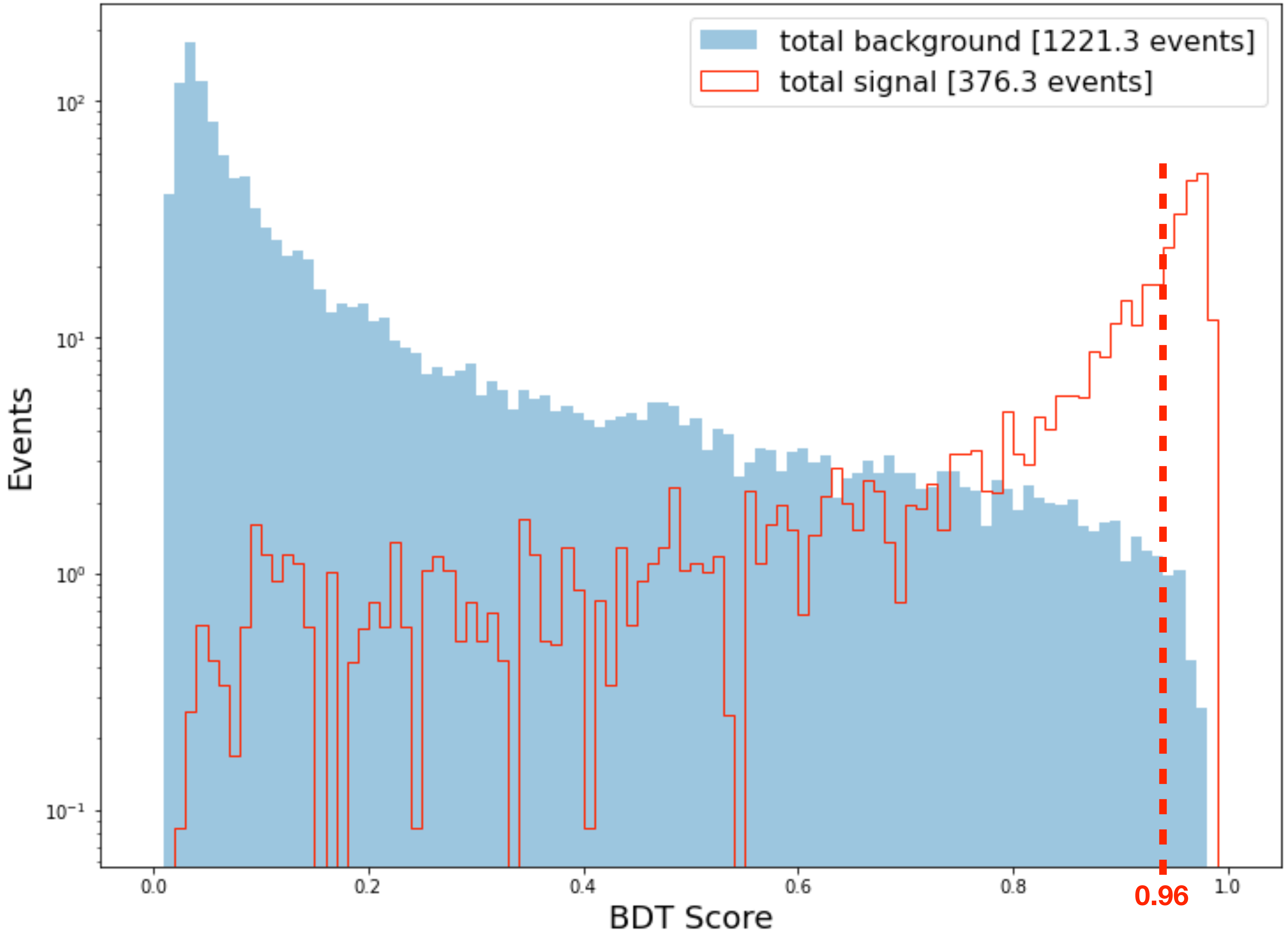
VBS WH BDT: Results





UC San Diego

VBS WH BDT: Results



Cut	VBSWH (κ_W = -1)	TotalBkg	S/√B
BDT Preselection	376.27	1221.28	10.77
BDT > 0.80	283.50	26.42	55.16
BDT > 0.90	223.47	7.72	80.41
BDT > 0.91	209.07	6.59	81.45
BDT > 0.92	197.83	5.15	87.17
BDT > 0.93	181.20	3.90	91.71
BDT > 0.94	164.57	2.73	99.64
BDT > 0.95	140.62	1.74	106.63
BDT > 0.96	107.21	0.70	128.19
BDT > 0.97	61.22	0.27	117.33
BDT > 0.98	11.87	0.00	—

BDT > 0.96 gives **107 sig vs. ~1 bkg**

(BDT > 0.963 gives **91 sig vs. ~0 bkg**)

UC San Diego

Summary

- Arrived at BDT- or cut-based signal regions that should be able to exclude $\kappa_W = -1$
- Immediate next steps:
    - Apply scale factors: lepton ID (ttH), JECs, etc.
    - Freeze signal region optimization
    - Design control region(s) for background estimation
- Afterwards:
    - Make another presentation here
    - Ask for a CADI line
    - Publish as a short paper in Physics Letters or PRD Rapid Comm.

Backup

*Professor
†Post-doc
‡Graduate Student
§Undergraduate Student

Who is involved?

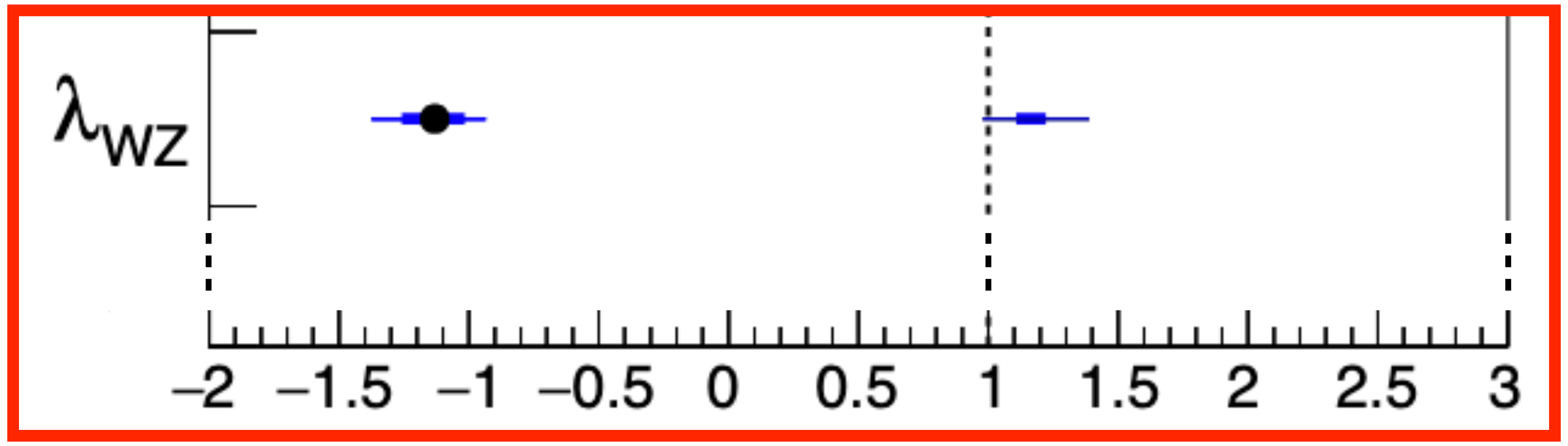
- UC San Diego:
  A. Arora‡, P. Chang†, L. Giannini†, J. Guiang‡, F. Würthwein*, Y. Xiang‡, A. Yagil*

- U. Nebraska:
  F. Golf*

- Boston University:
  D. Spitzbart†, I. Suarez*

- UC Santa Barbara:
  C. Campagnari*

Current CMS Result



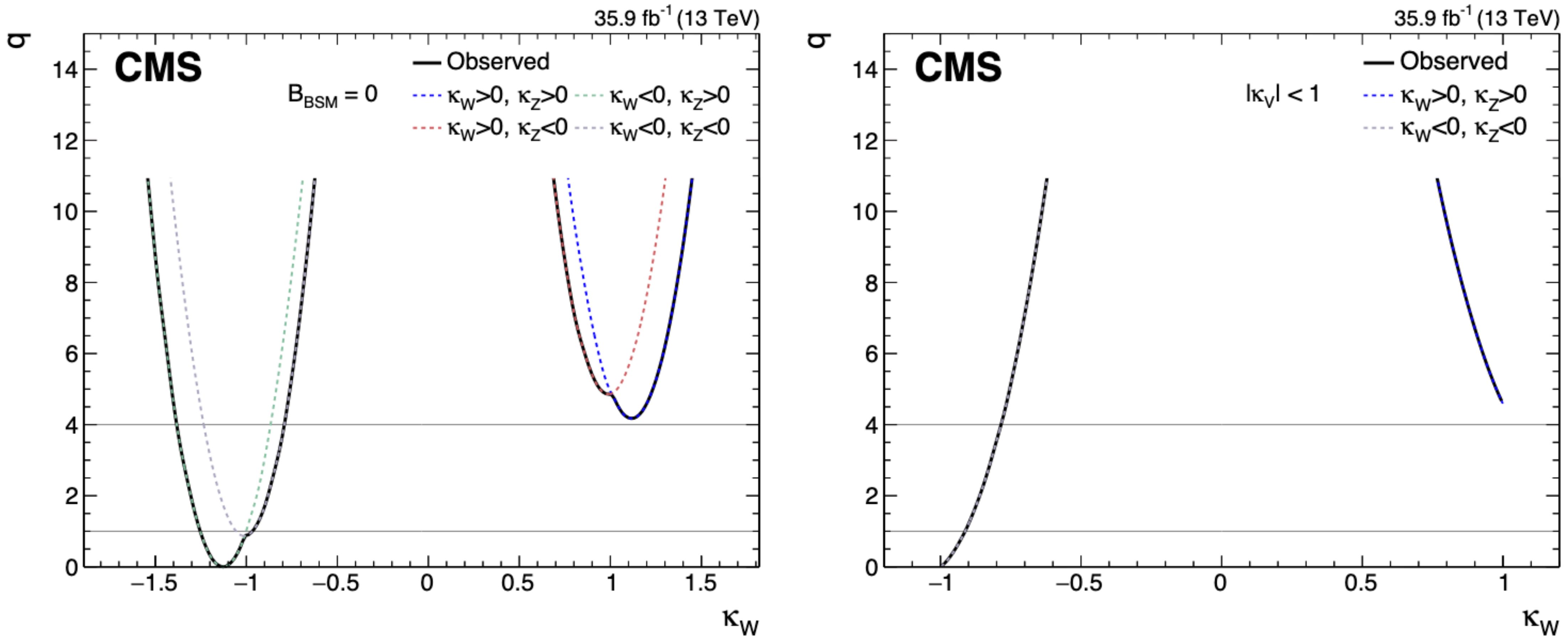


Figure 13: Scan of the test statistic q as a function of κ_{W} in the generic κ model assuming $\mathcal{B}_{\mathrm{BSM}} = 0$ (left) and allowing $\mathcal{B}_{\mathrm{inv}}$ and $\mathcal{B}_{\mathrm{undet}}$ to float (right). The different colored lines indicate the value of q for different combinations of signs for κ_{W} and κ_{Z}. The solid black line shows the minimum value of $q(\kappa_{\mathrm{W}})$ in each case and is used to determine the best fit point and the 1σ and 2σ CL regions. The scan in the right figure is truncated because of the constraints of $|\kappa_{\mathrm{W}}| \leq 1$ and $|\kappa_{\mathrm{Z}}| \leq 1$, which are imposed in this model.

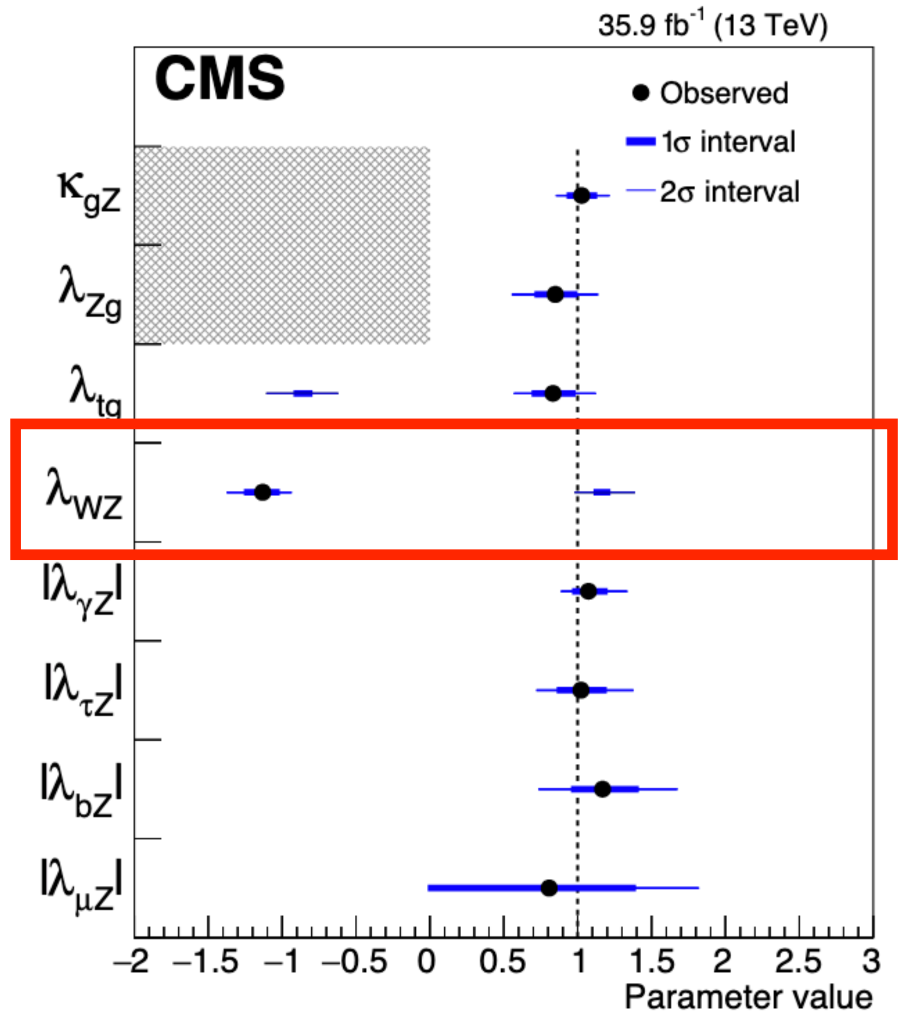


Figure 16: Summary of the model with coupling ratios and effective couplings for the ggH and $\mathrm{H} \to \gamma\gamma$ loops. The points indicate the best fit values while the thick and thin horizontal bars show the 1σ and 2σ CL intervals, respectively. For this model, both positive and negative values of λ_{WZ} and λ_{tg} are considered.

https://arxiv.org/pdf/1809.10733.pdf

BSM Signal Models

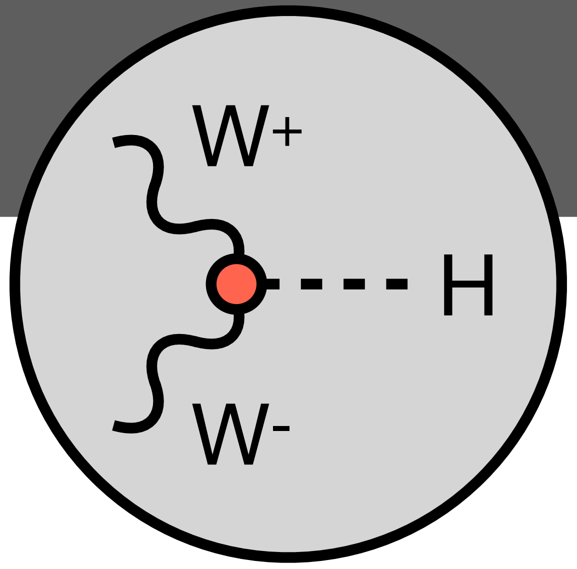
$\kappa_W = -1$

models/sm/couplings.py

```
GC_72 = Coupling(name = 'GC_72',
                 value = '(ee**2*complex(0,1)*vev)/(2.*sw**2)',
                 value = '-((ee**2*complex(0,1)*vev)/(2.*sw**2))',
                 order = {'QED':1})
```

models/sm/vertices.py

```
V_52 = Vertex(name = 'V_52',
              particles = [ P.W__minus__, P.W__plus__, P.H ],
              color = [ '1' ],
              lorentz = [ L.VVS1 ],
              couplings = {(0,0):C.GC_72})
```



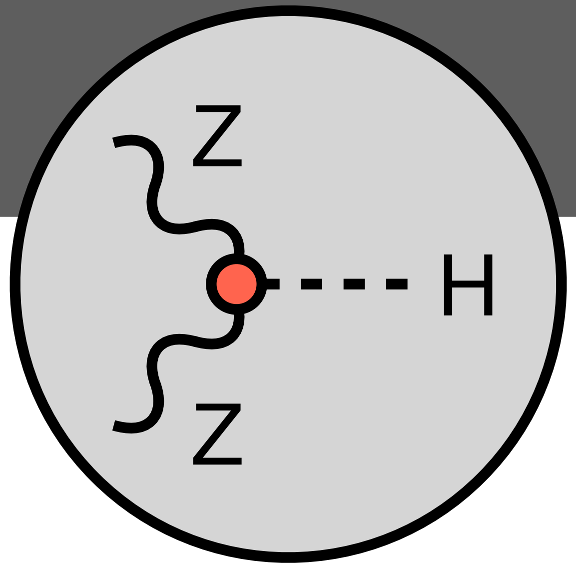
$\kappa_Z = -1$

models/sm/couplings.py

```
GC_81 = Coupling(name = 'GC_81',
                 value = 'ee**2*complex(0,1)*vev + …',
                 value = '-(ee**2*complex(0,1)*vev + … )',
                 order = {'QED':1})
```

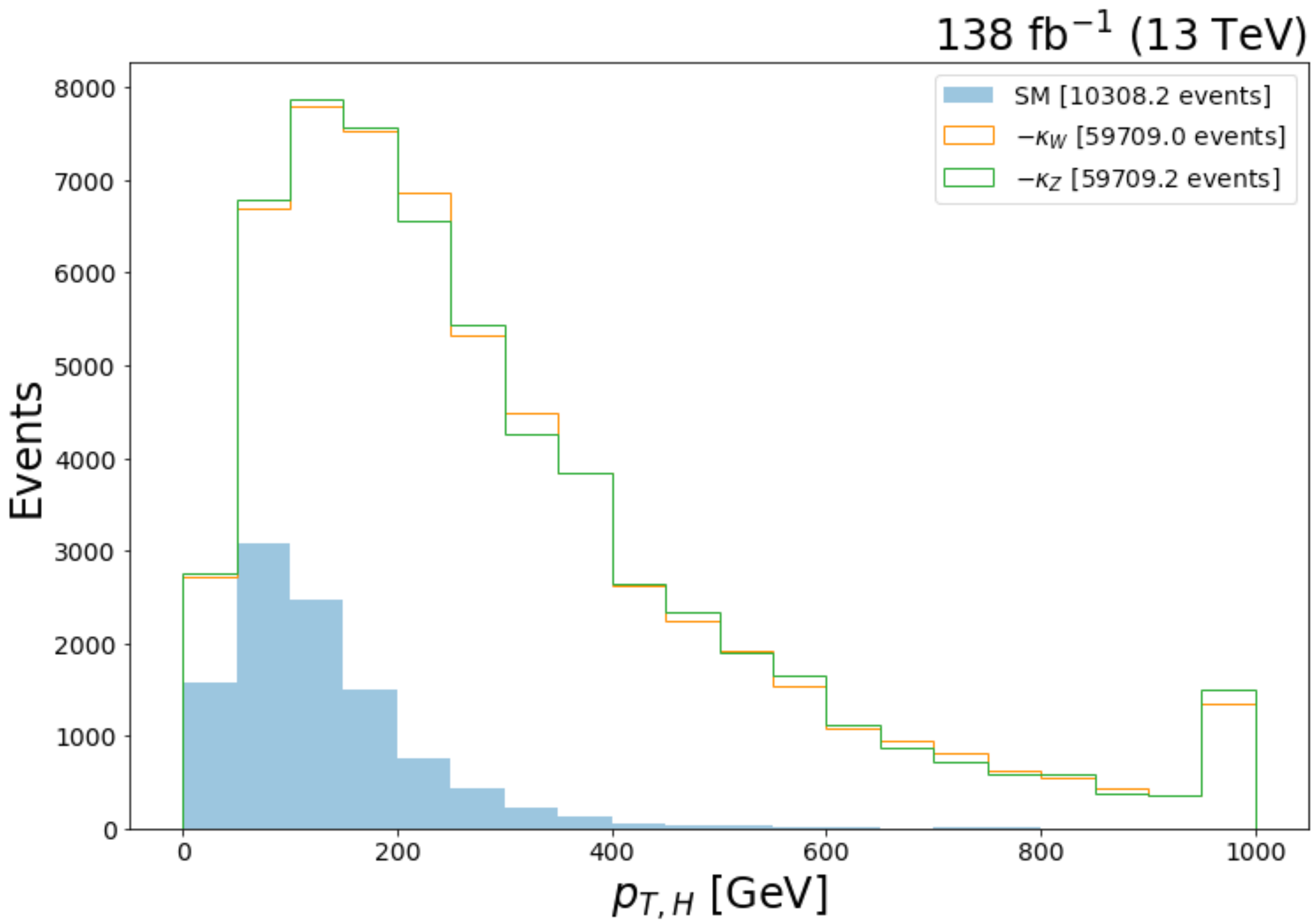
models/sm/vertices.py

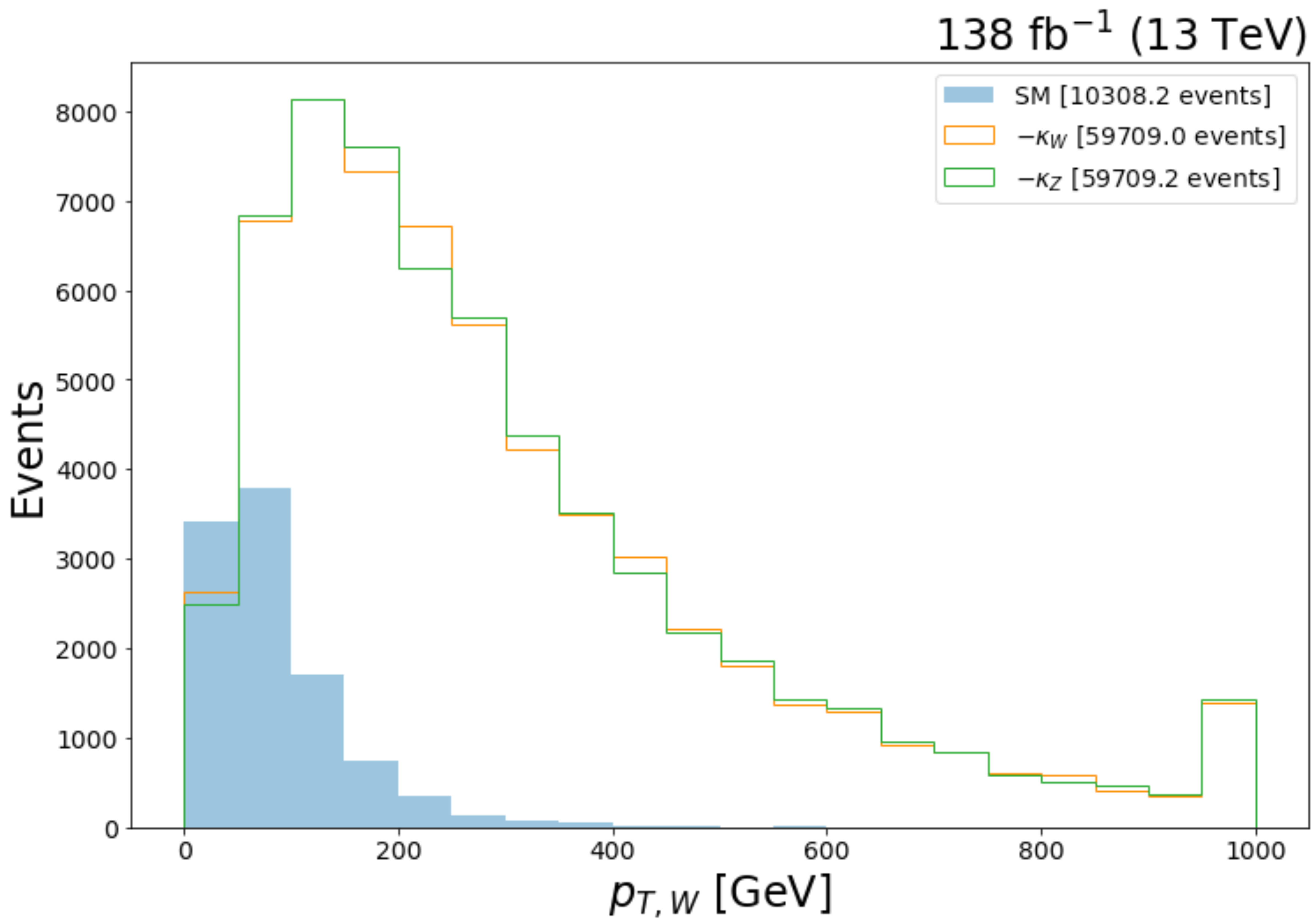
```
V_69 = Vertex(name = 'V_69',
              particles = [ P.Z, P.Z, P.H ],
              color = [ '1' ],
              lorentz = [ L.VVS1 ],
              couplings = {(0,0):C.GC_81})
```



Only changed one line in SM Madgraph model!

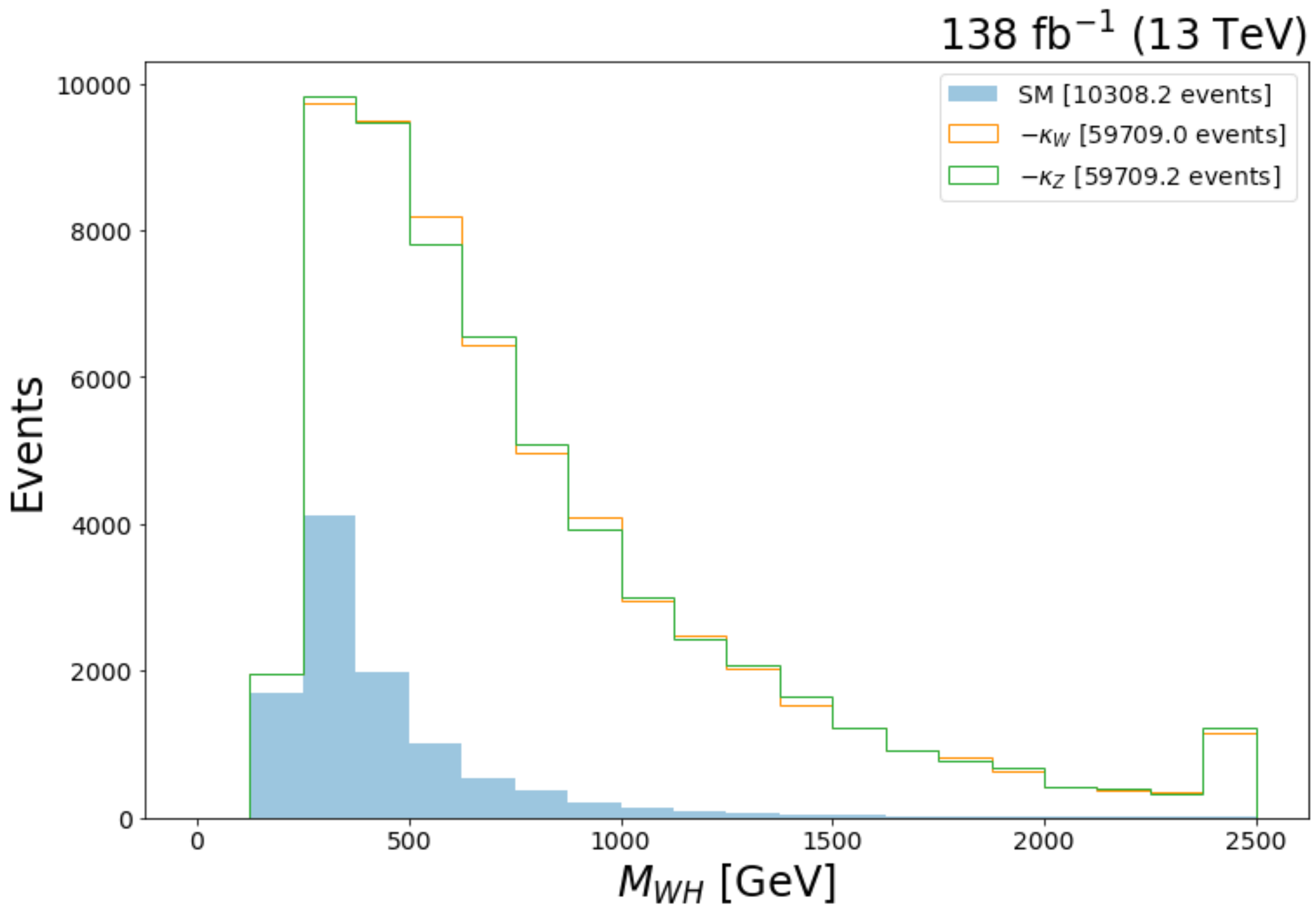
LHE W and H Kinematics





Strong boost in W and H in BSM ($-\kappa_W$ or $-\kappa_Z$) models

LHE W and H Kinematics



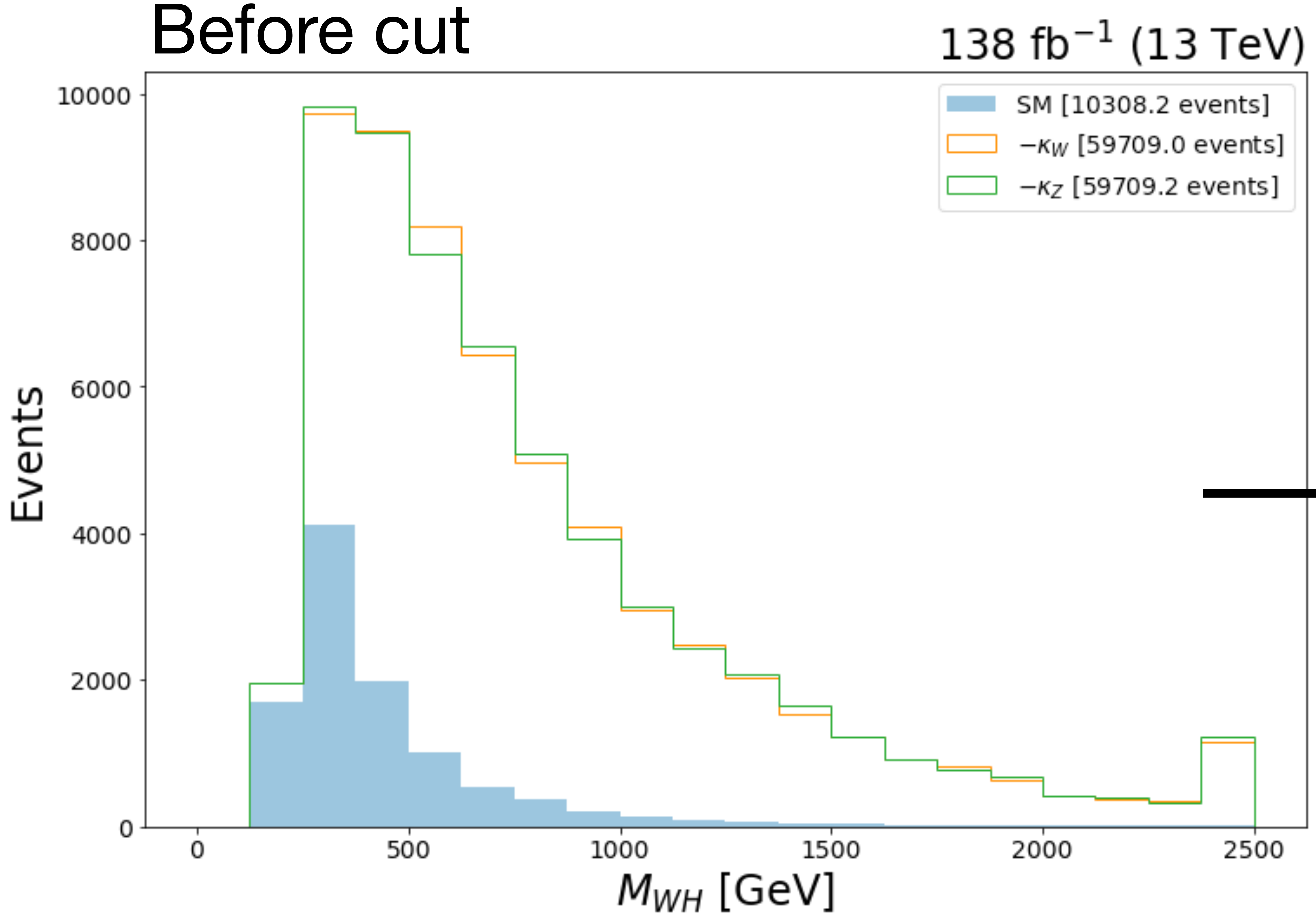
Strong boost in W and H in BSM (-κ_W or -κ_Z) models

CMS

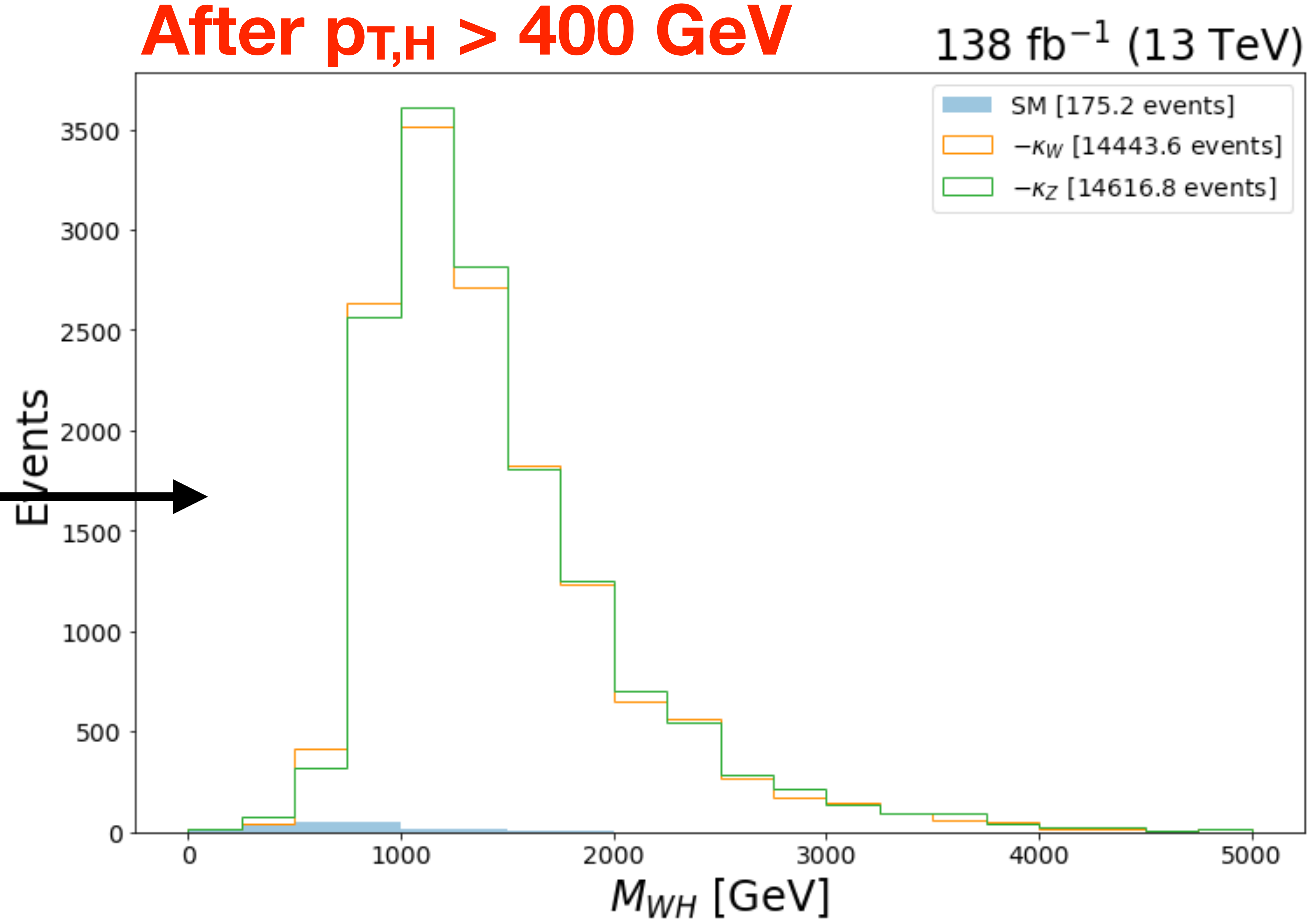
UC San Diego

LHE W and H Kinematics

Before cut



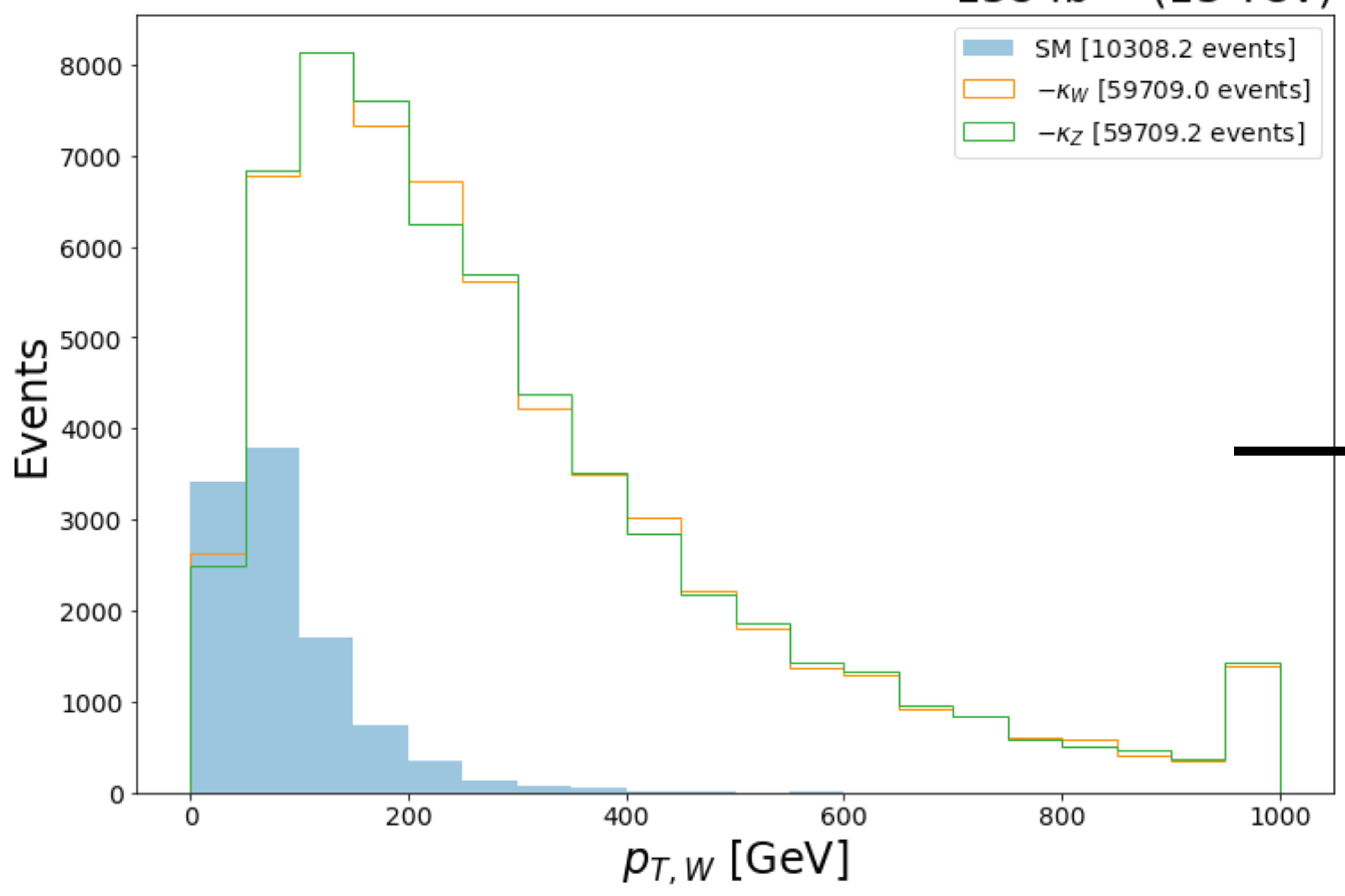
After $p_{T,H}$ > 400 GeV



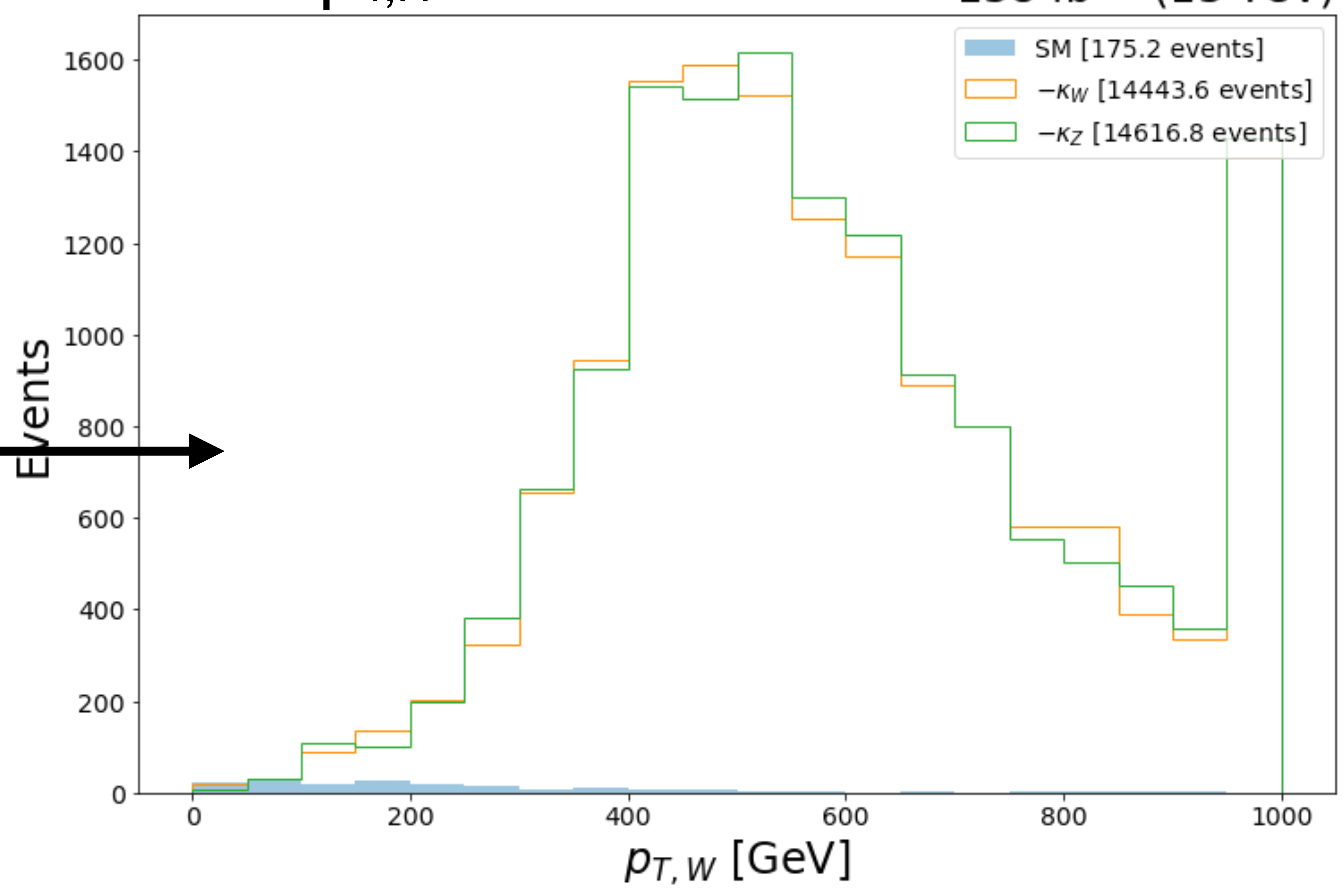
High s-hat events have large M_{WH}

LHE W and H Kinematics

Before cut



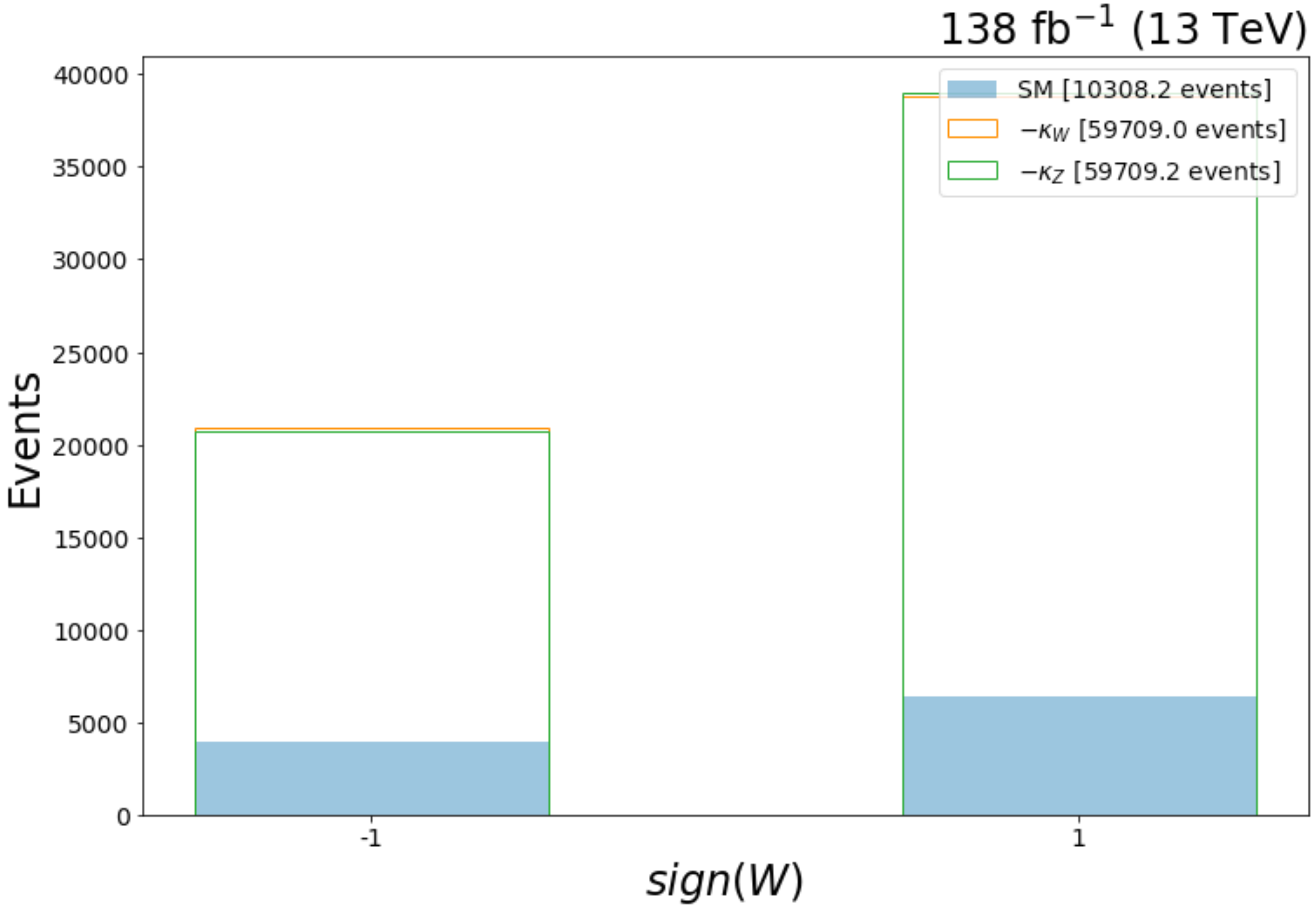
After $p_{T,H}$ > 400 GeV

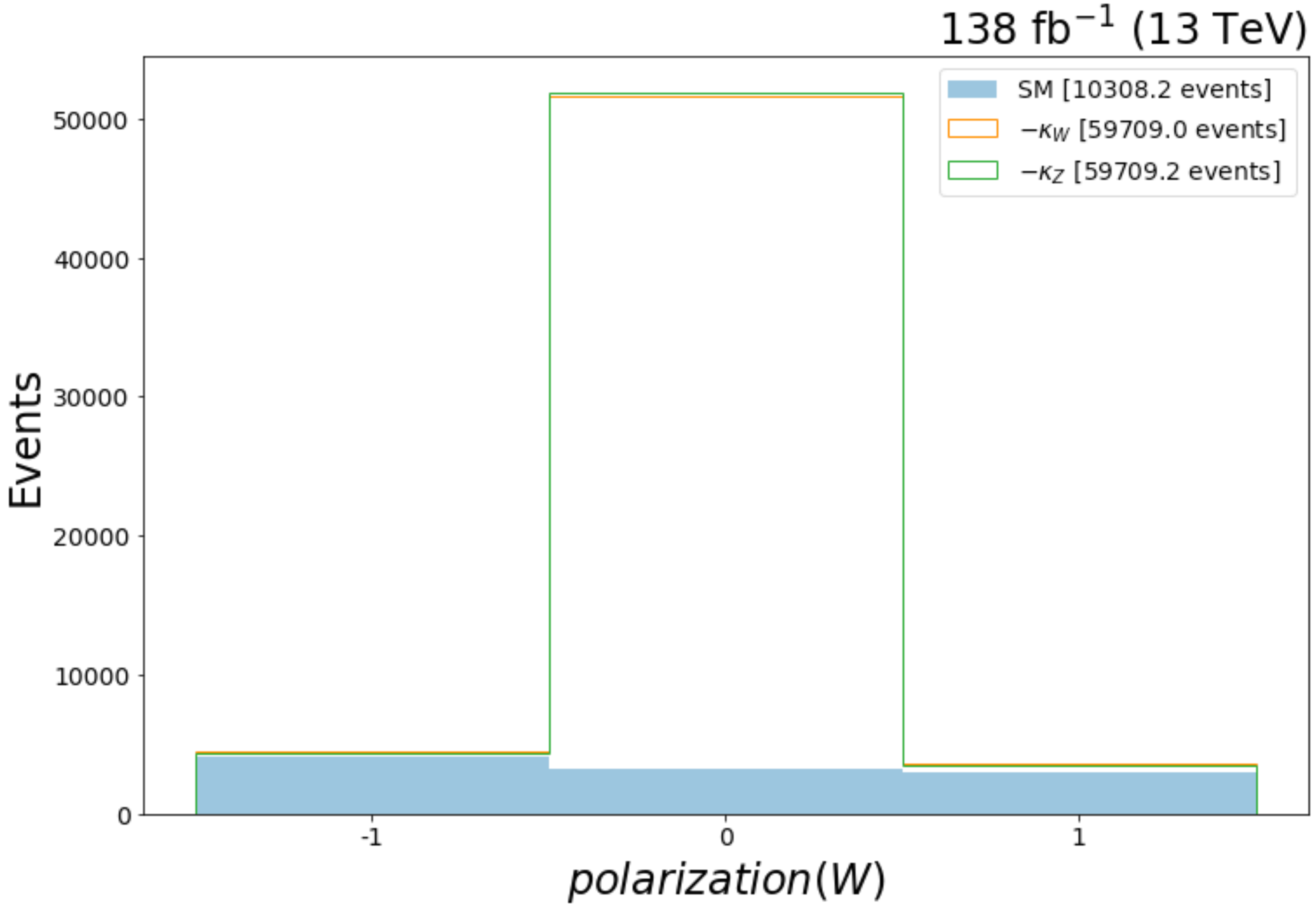


CMS

UC San Diego

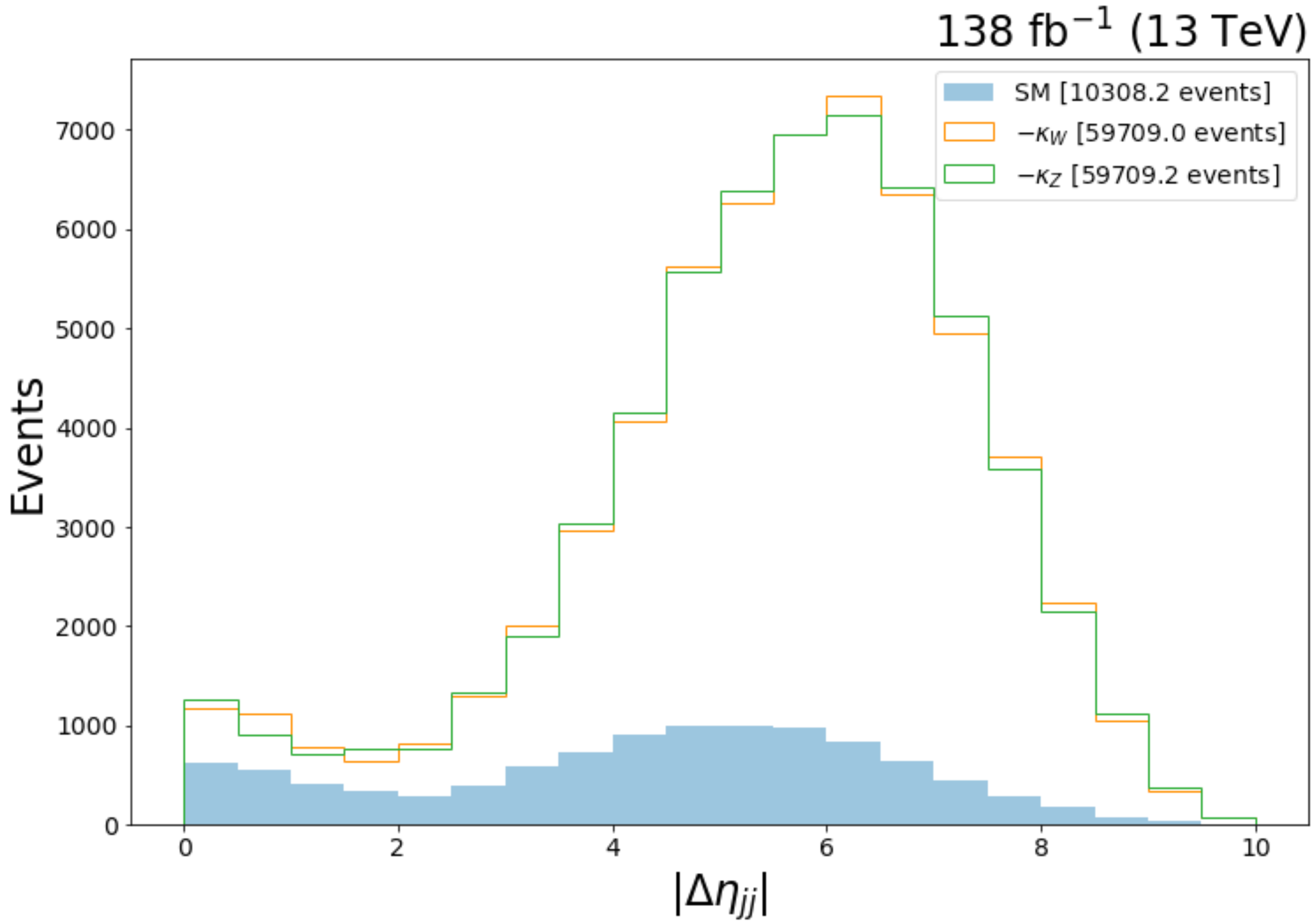
LHE W Characteristics

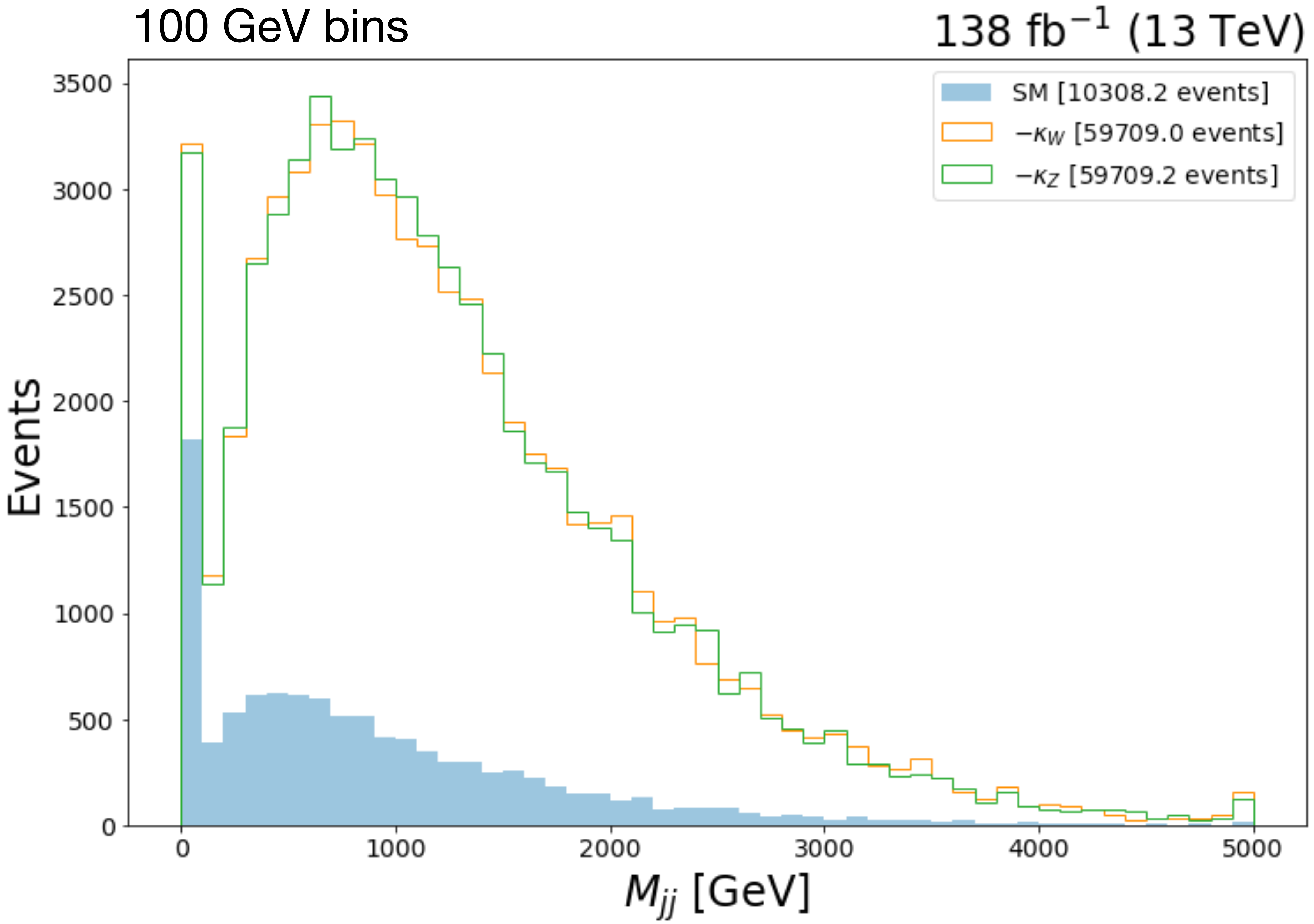




Some W$^+$/W$^-$ asymmetry; more longitudinally polarized W's

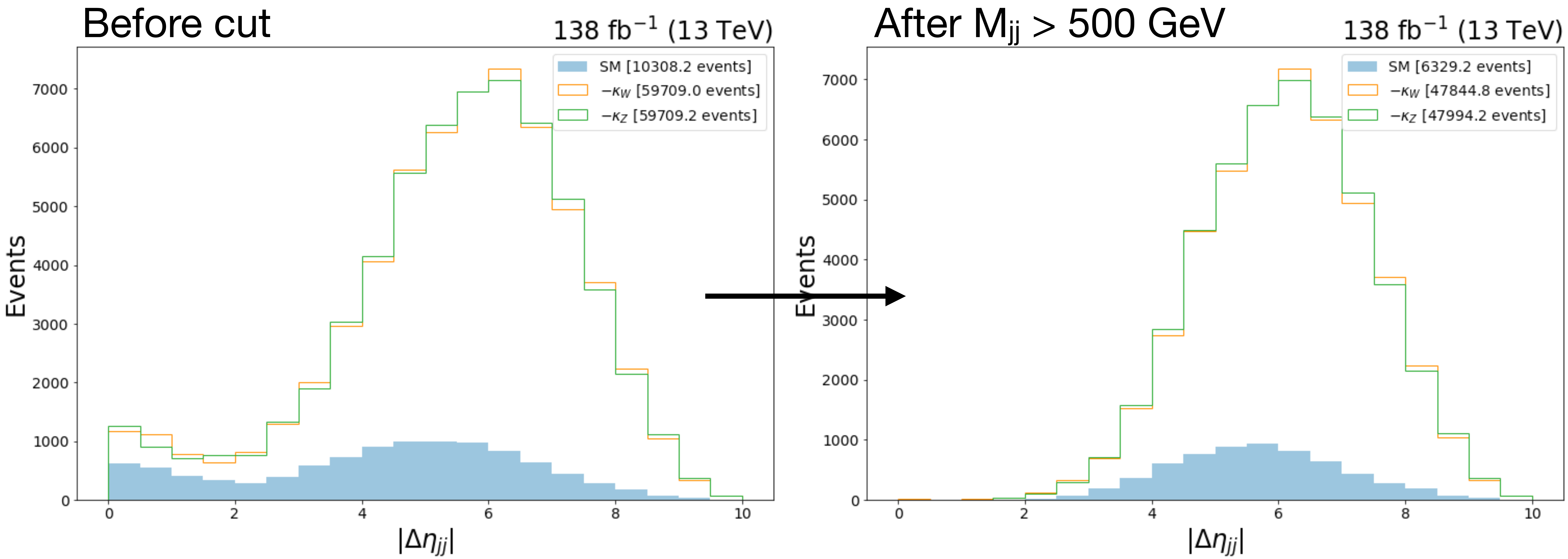
LHE VBS Kinematics





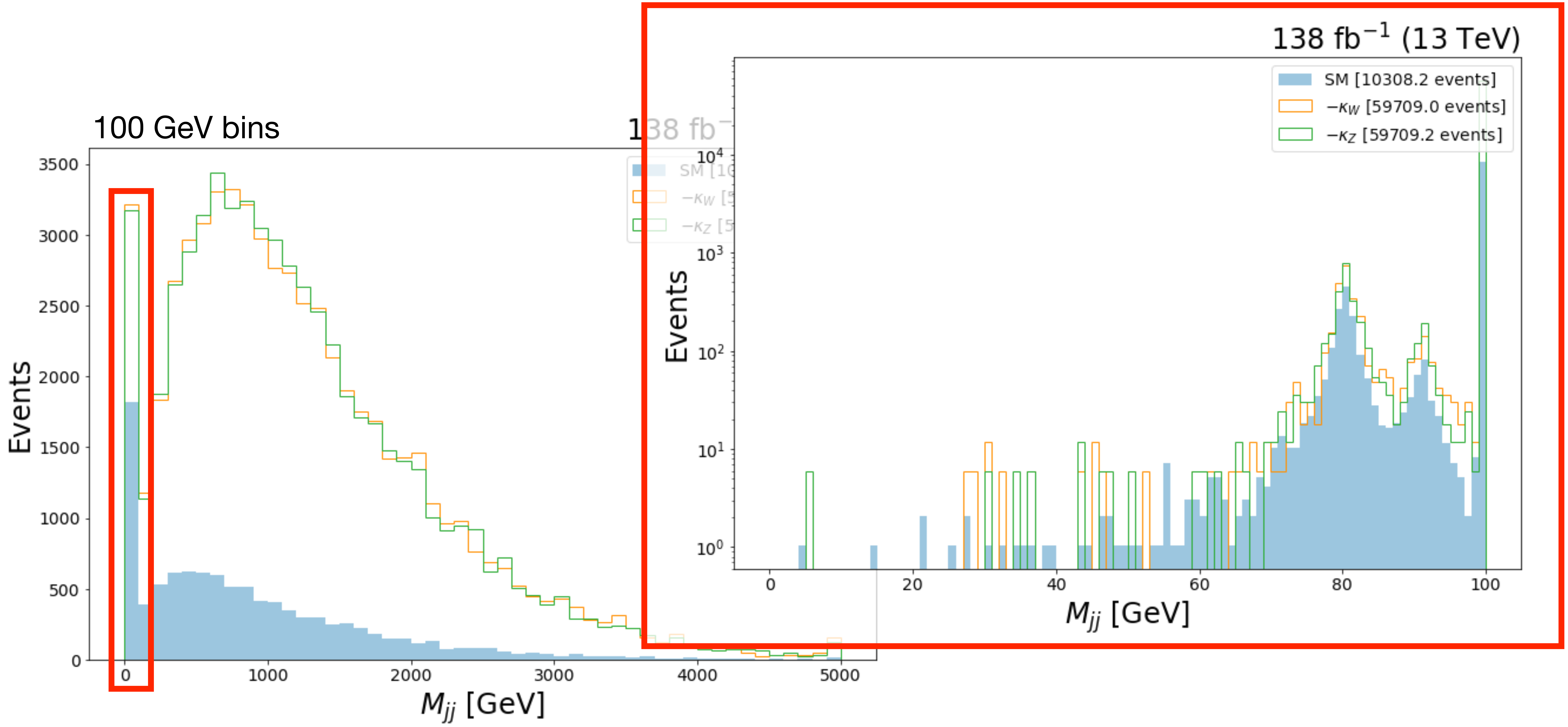
VBS-like signal: large $|\Delta\eta_{jj}|$, M_{jj}

LHE VBS Kinematics



VBS-like signal: large $|\Delta\eta_{jj}|$, M_{jj}

LHE VBS Kinematics



Visible W and Z peaks from pp → WHV (V→jj)

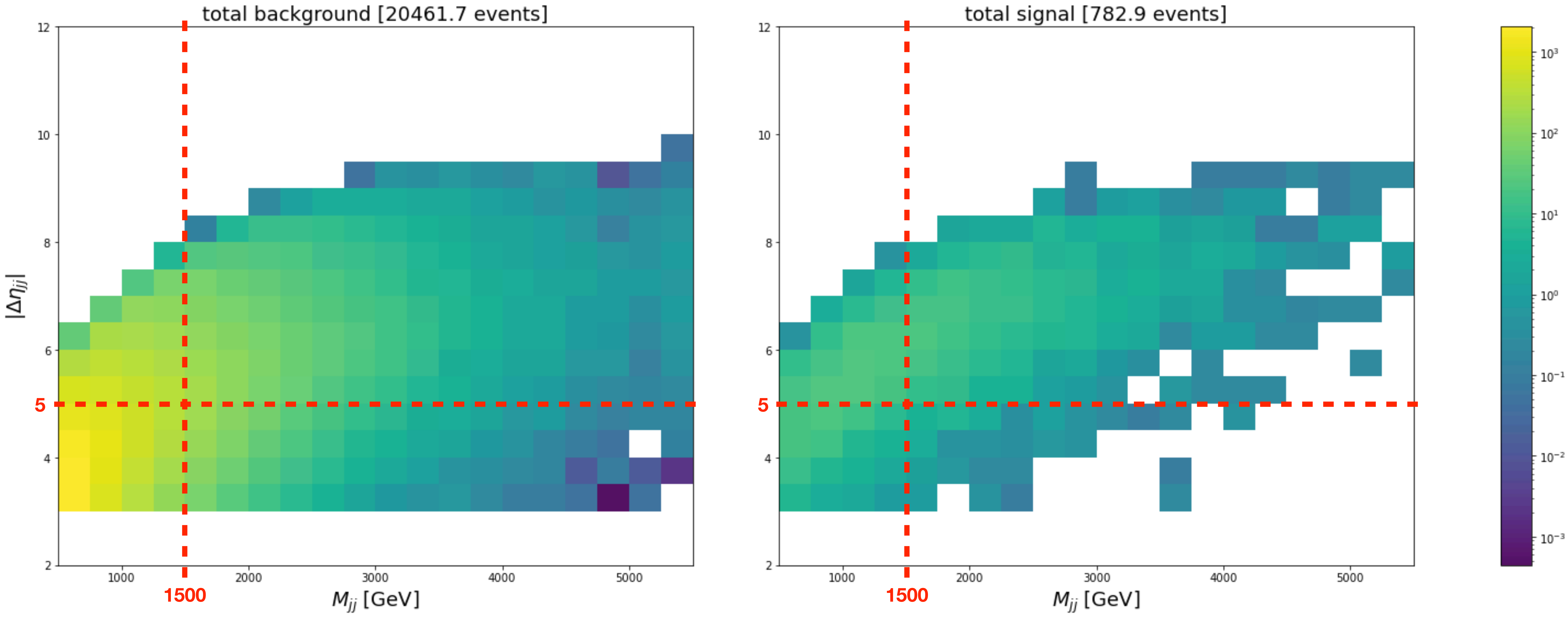
VBS WH Cutflow

	VH		VV/VVV/VBSWZ		W+Jets		TTbar+X		TTbar1L		TTbar2L		TotalBkg		VBSWH (kW < 0)	
cut	raw	wgt	raw	wgt	raw	wgt	raw	wgt	raw	wgt	raw	wgt	raw	wgt	raw	wgt
Bookkeeping	—	—	—	—	—	—	—	—	—	—	—	—	—	—	399989	59598.09
SKIM_FindLeptons	—	—	—	—	—	—	—	—	—	—	—	—	—	—	399989	59598.09
SKIM_Geq1VetoLep	—	—	—	—	—	—	—	—	—	—	—	—	—	—	130230	19319.47
SKIM_Geq2Jets	—	—	—	—	—	—	—	—	—	—	—	—	—	—	129567	19221.17
SKIM_Geq1FatJetNoVetoLepOverlap	—	—	—	—	—	—	—	—	—	—	—	—	—	—	53072	7869.79
POSTSKIM_Exactly1TightLep	—	—	—	—	—	—	—	—	—	—	—	—	—	—	31843	4750.86
POSTSKIM_Geq1FatJetNoTightLepOverlap	—	—	—	—	—	—	—	—	—	—	—	—	—	—	11932	1776.65
SelectLeptons	277198	4511.37	296595	1682.81	430128	32660.21	2960671	3675.78	9439330	424281.50	962161	34177.24	14366083	500988.91	11932	1776.65
Has1TightLep	277198	4511.37	296595	1682.81	430128	32660.21	2960671	3675.78	9439330	424281.50	962161	34177.24	14366083	500988.91	11932	1776.65
LepPtGt40	247637	4079.22	236196	1328.09	331677	26056.05	2229901	2744.47	7315482	328311.10	717050	25411.25	11077943	387930.18	11082	1650.62
SelectFatJets	247637	4079.22	236196	1328.09	331677	26056.05	2229901	2744.47	7315482	328311.10	717050	25411.25	11077943	387930.18	11082	1650.62
Geq1FatJet	243443	4018.46	232351	1310.98	324702	25501.02	2195215	2707.30	7268198	326174.50	665489	23568.50	10929398	383280.77	10937	1629.24
SelectHbbFatJet	243443	4018.46	232351	1310.98	324702	25501.02	2195215	2707.30	7268198	326174.50	665489	23568.50	10929398	383280.77	10937	1629.24
PNetHbbScoreGt0p9	195762	3209.76	99990	692.78	81022	5851.27	849107	1005.49	2544585	114319.10	270367	9549.23	4040833	134627.63	8970	1341.00
SelectJetsNoHbbOverlap	195762	3209.76	99990	692.78	81022	5851.27	849107	1005.49	2544585	114319.10	270367	9549.23	4040833	134627.63	8970	1341.00
SelectVBSJetsMaxE	72587	1035.51	68999	220.91	61298	3224.51	759911	898.96	1926126	86676.40	188949	6690.97	3077870	98747.26	6210	929.32
MjjGt500detajjGt3	15356	194.40	10365	35.87	11697	580.29	142157	169.63	399477	17930.50	43934	1551.03	622986	20461.72	5213	782.95

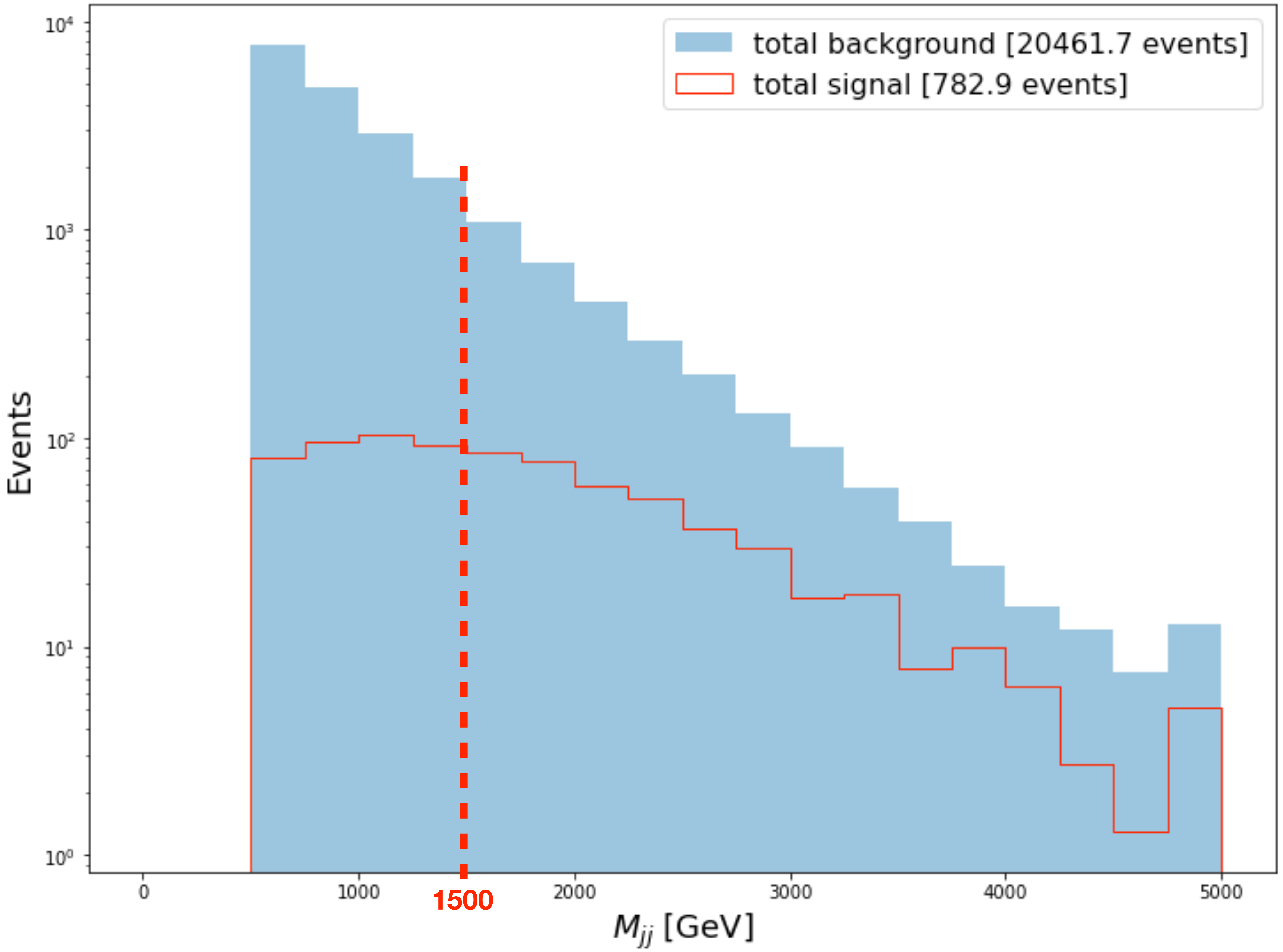
Skim & Postskim already applied!

Cut	VH	VV/VVV/VBSWZ	W+Jets	TTbar+X	TTbar1L	TTbar2L	TotalBkg	Eff.*	VBSWH (κ_W = -1)	Eff.*
Skim AND Postskim	2.8K	1.7K	33K	3.7K	424K	34K	499K	—	1.8K	—
Preselection	194	36	580	170	18K	1.6K	20K	96%	783	56%
M_{jj} > 1.5 TeV AND $\Delta\eta_{jj}$ > 5 AND S_T > 500 GeV	14	2	31	10	1.1K	106	1.2K	94%	376	52%
Raw:	1.2K	629	843	9K	24K	3K	38K	—	2.5K	—

VBS Kinematics: Mjj vs. $|\Delta\eta_{jj}|$ correlation



VBS Kinematics: M_{jj}

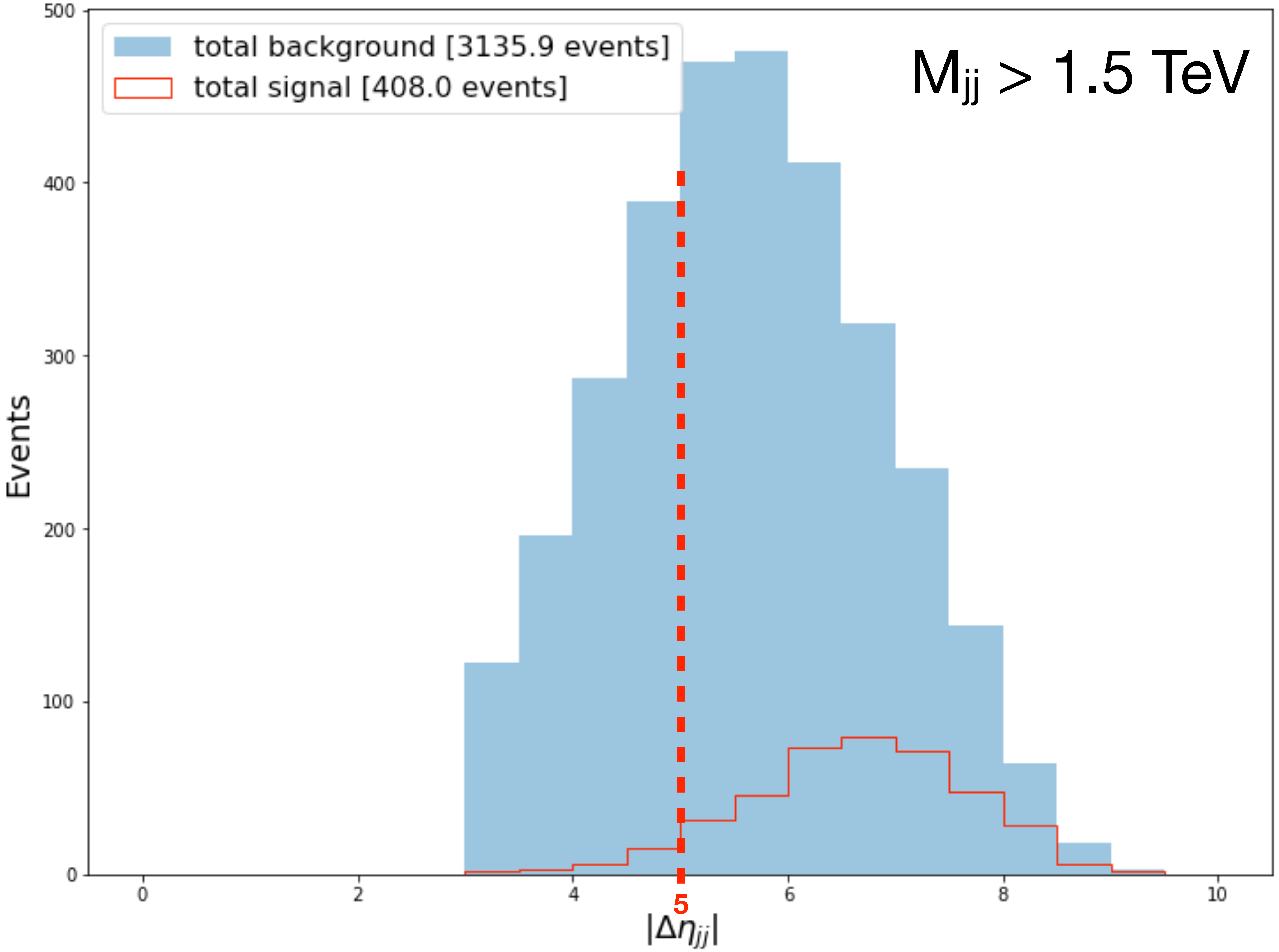


Cut	VBSWH (κ_W = -1)	TotalBkg	S/√B
Preselection	782.95	20461.72	5.47
M_{jj} > 1 TeV	606.26	7807.43	6.86
M_{jj} > 1.5 TeV	408.00	3135.93	7.29
M_{jj} > 2 TeV	244.29	1342.69	6.67
M_{jj} > 2.5 TeV	134.39	594.51	5.51

CMS

UC San Diego

VBS Kinematics: $|\Delta\eta_{jj}|$

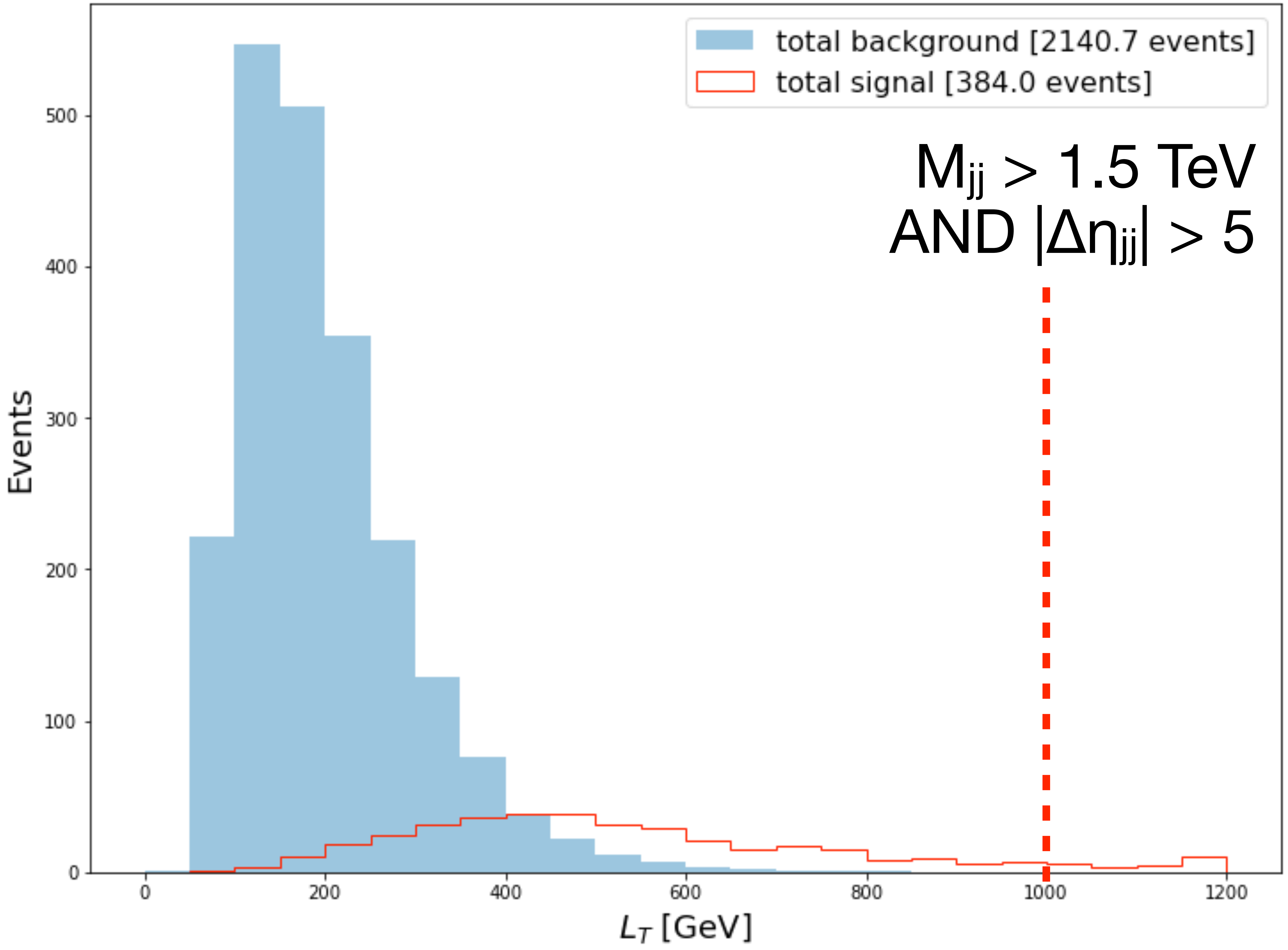


Cut	VBSWH (κ_W = -1)	TotalBkg	S/√B
M_{jj} > 1.5 TeV	408.00	3135.93	7.29
$\|\Delta\eta_{jj}\|$ > 3.5	406.47	3013.83	7.40
$\|\Delta\eta_{jj}\|$ > 4	403.92	2817.27	7.61
$\|\Delta\eta_{jj}\|$ > 4.5	398.54	2529.89	7.92
$\|\Delta\eta_{jj}\|$ > 5	384.00	2140.68	8.30
$\|\Delta\eta_{jj}\|$ > 5.5	352.89	1670.75	8.63
$\|\Delta\eta_{jj}\|$ > 6	307.70	1194.12	8.90
$\|\Delta\eta_{jj}\|$ > 6.5	234.21	782.40	8.37
$\|\Delta\eta_{jj}\|$ > 7	154.36	464.24	7.16

CMS

UC San Diego

WH Kinematics: $L_T = p_T(\ell) + MET$

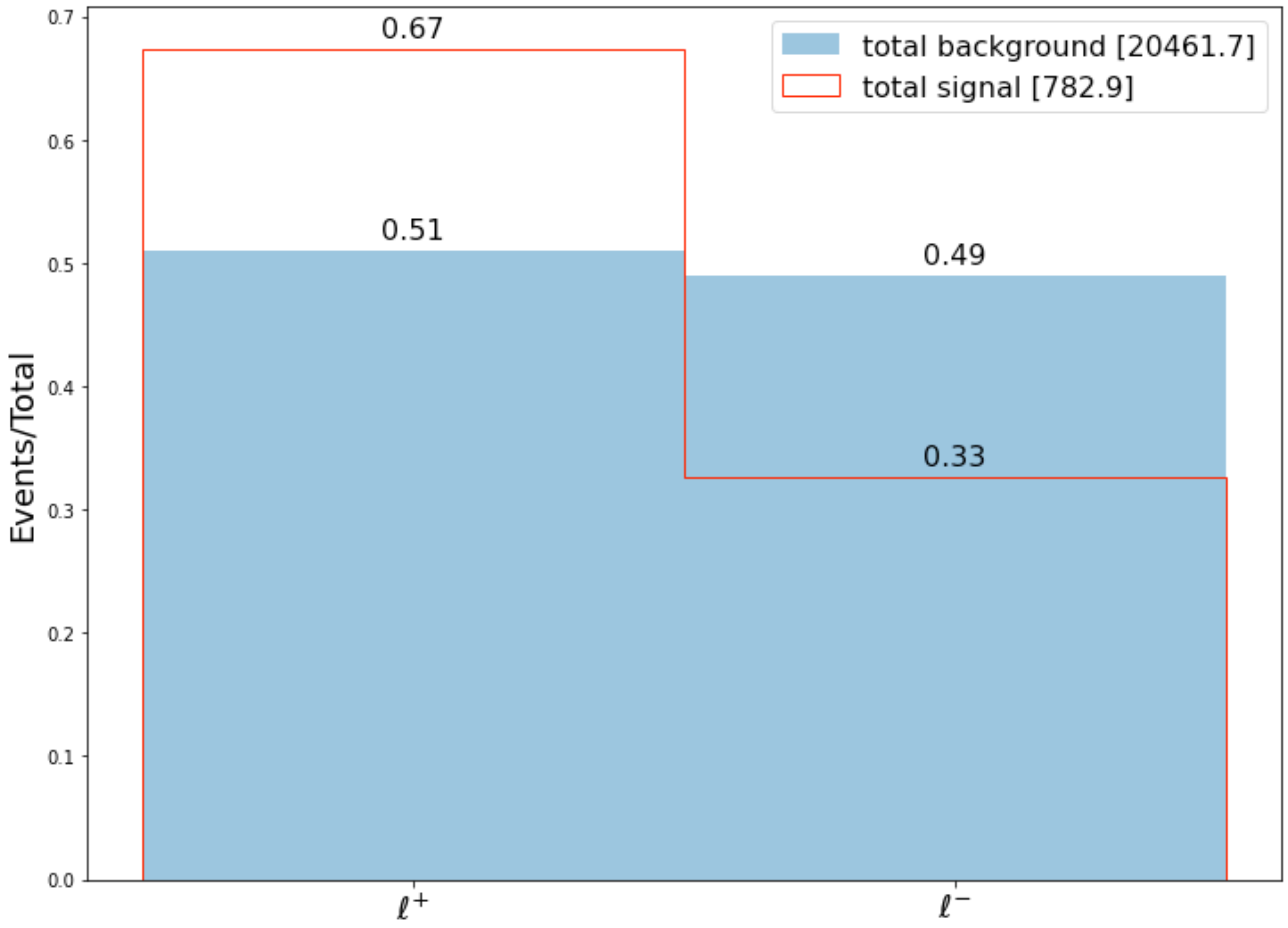


Cut	VBSWH (κ_W = -1)	TotalBkg	S/√B
M_{jj} > 1.5 TeV AND $\Delta\eta_{jj}$ > 5	384.00	2140.68	8.30
L_T > 500 GeV	181.81	26.66	35.21
L_T > 600 GeV	121.16	8.16	42.40
L_T > 700 GeV	85.36	2.71	51.86
L_T > 800 GeV	53.70	1.04	52.77
L_T > 900 GeV	36.30	0.35	61.18
L_T > 1 TeV	23.54	0.14	62.43
L_T > 1.1 TeV	15.26	0.11	45.35
L_T > 1.2 TeV	8.63	0.07	32.03

M_{jj} > 1.5 TeV AND $\Delta\eta_{jj}$ > 5
AND L_T > 900 GeV
gives **36 sig vs. ~0 bkg**

UC San Diego

Lepton Charge Asymmetry



VBS WH BDT: XGBoost Configuration

Param	Value	Description*
`objective`	binary:logistic	the learning task and the corresponding learning objective binary:logistic = logistic regression for binary classification, output probability
`eta`	0.1	Step size shrinkage (alias: learning_rate)
`max_depth`	3	Max. depth of tree: larger = more complex = more prone to overfitting
`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 Curve
`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
`min_child_weight`	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
`base_score`	0.5	The initial prediction score of all instances, global bias
`scale_pos_weight`	3.35	Control the balance of positive and negative** weights, useful for unbalanced classes; a typical value to consider: sum(negative instances) / sum(positive instances)

*From: https://xgboost.readthedocs.io/en/stable/parameter.html
**Here, positive = signal and negative = background