$t\bar{t}HH$ channel with 2 leptons in the final state searching

Hanyang University High Energy Physics Laboratory

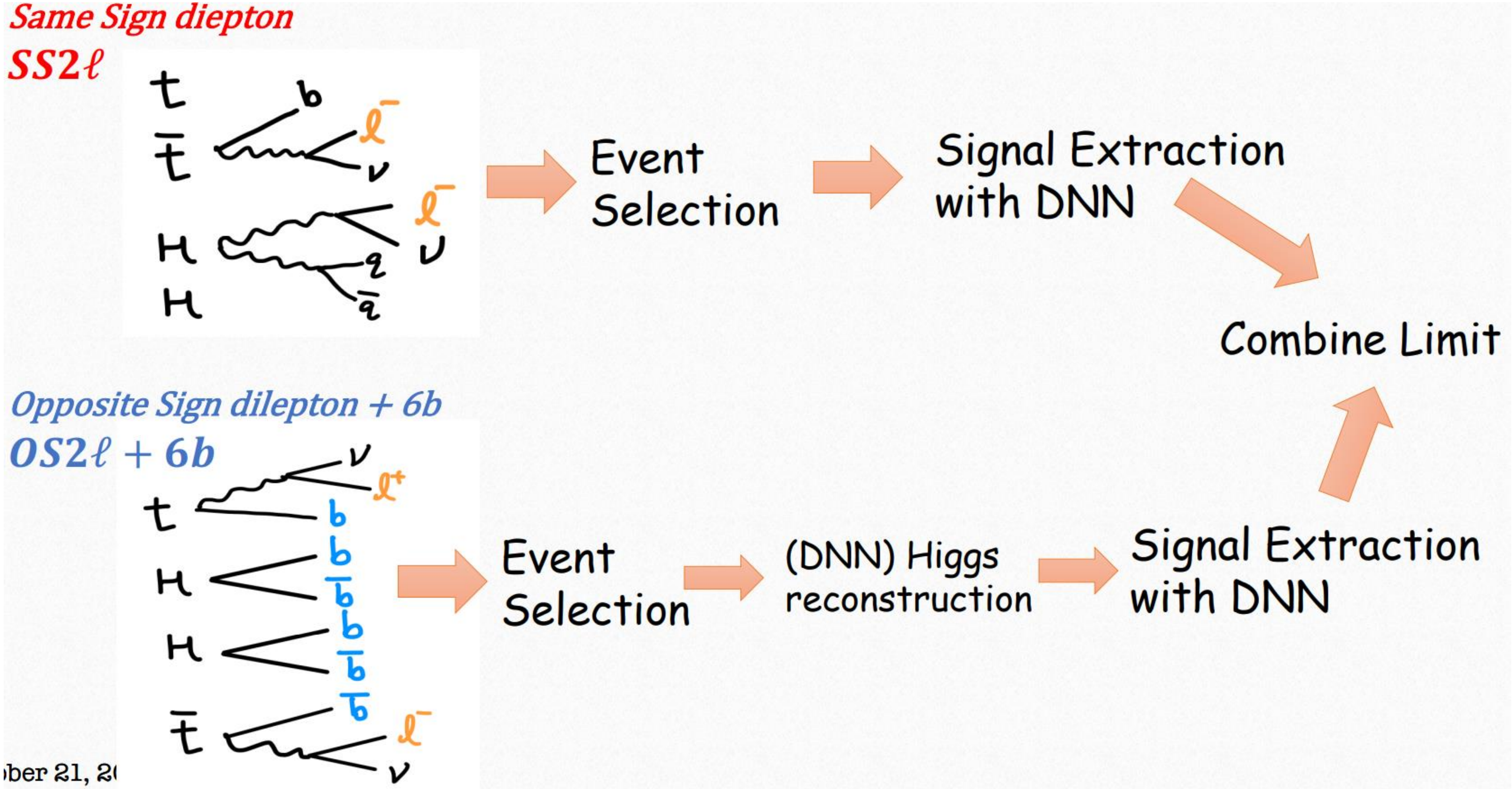
Sungbeom Cho, Taejeong Kim, Yang Tianyi, Yeonsu Ryou

$t\bar{t}HH$ searching

- Sungbeom Cho in our group has been working on the prospective $t\bar{t}HH$ study at the HL-LHC $\sqrt{s} = 14\ \mathrm{TeV},\ \ 3000\ \mathrm{fb}^{-1}$: AN-2024/209.
- We plan to develop this strategy into an analysis using Run 3 data and CMS Monte Carlo.

- The characteristic of our analysis is the **two leptons** in the final state:
    - Same-sign two leptons (SS2l)
    - Opposite-sign two leptons (OS2l)
- The analyses in the two channels use the deep neural network (DNN) method to enhance signal-background discriminant.

Analysis strategy



Samples

- Signal: $t\bar{t}HH$ with $t\bar{t}$ single lepton (**SL**) and double lepton (**DL**) decay modes, and Higgs boson pair decay *inclusively*.
- Background:
  - Single Higgs backgrounds: $t\bar{t}H$, $t\bar{t}b\bar{b}H$, $t\bar{t}ZH$
  - Non Higgs backgrounds: $t\bar{t}VV$, $t\bar{t}b\bar{b}V$, $t\bar{t}b\bar{b}$, $t\bar{t}4b$, $t\bar{t}t\bar{t}$, $t\bar{t}$ [1]

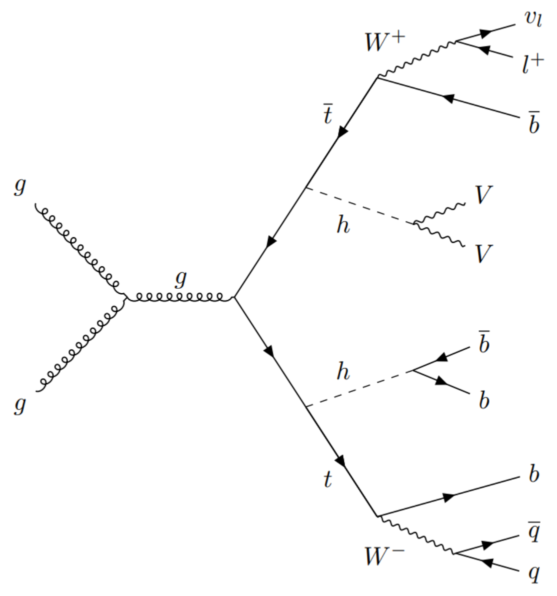
- Generator: **MadGraph5aMC@NLO** v 2.9.15 + parton shower using **Pythia 8**
- Detector simulation: **Delphes** v 3.5.0 using *CMS_PhaseII_0PU_v02.tcl* card.
- PDF modeling: **NNPDF 4.0**
- Higgs boson and top quark decay modeled using **MadSpin**.

[1] $t\bar{t}$ were not generated at the current results. By default, the results in this slides do not include $t\bar{t}$.

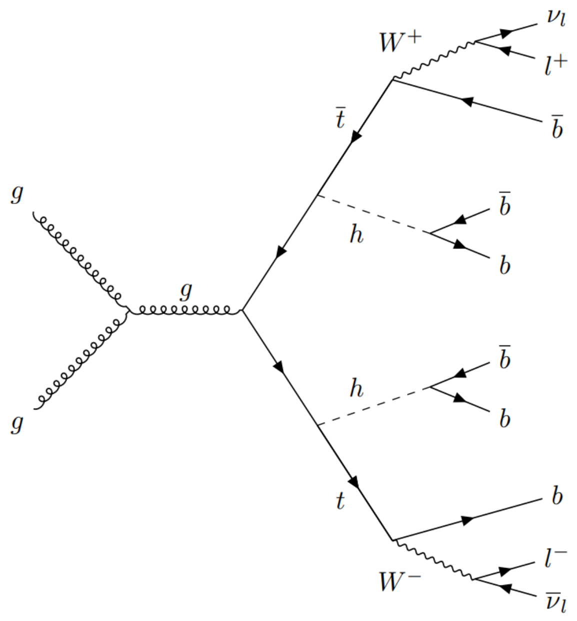
Object selection

- Object definition:
  - Jets using anti-k_T method
  - B-tagging at *Loose* working point
- Pre-selections on objects:
  - $e^{\pm}$: $p_T > 23$ GeV, $|\eta| < 3.0$, Isolation variable $I_{rel} < 0.3$ for $\Delta R < 0.3$.
  - $\mu^{\pm}$: $p_T > 17$ GeV, $|\eta| < 2.8$, Isolation variable $I_{rel} < 0.3$ for $\Delta R < 0.3$.
  - Jets: $p_T > 30$ GeV, $|\eta| < 3.0$
- Object performance:
  - B-tagging efficiency=0.85, light flavour mis-tagging rate=0.10
  - $e^{\pm}$ and $\mu^{\pm}$ identification efficiency ~0.97 without isolation requirements.

Signal regions event selection



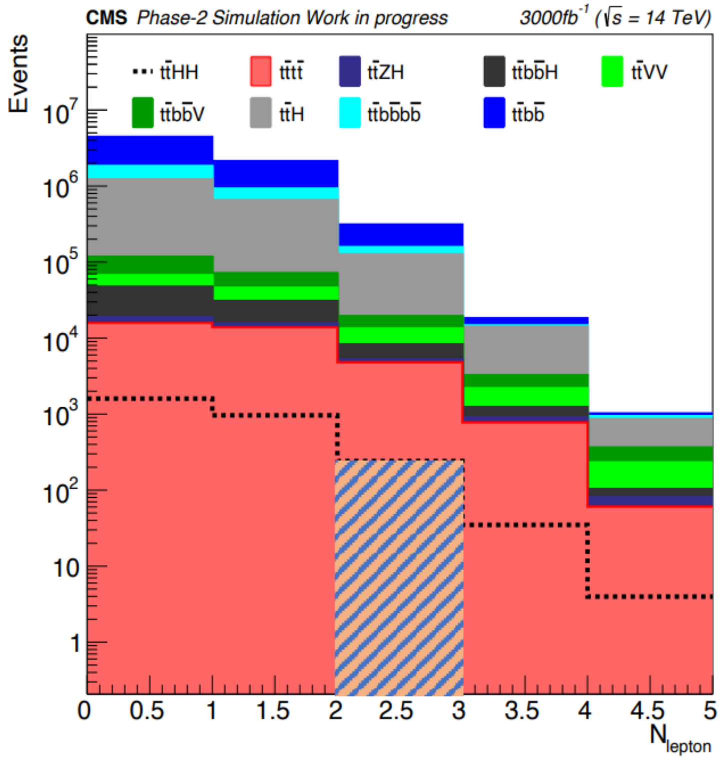
- SS2l:
  - $n_{lep} = 2$ (S1), with the **SAME** sign (S2)
  - $MET > 30$ GeV (S3)
  - $n_{B-tag} \geq \mathbf{3}$ (S4), Jets $H_T > \mathbf{300}$ GeV (S5)



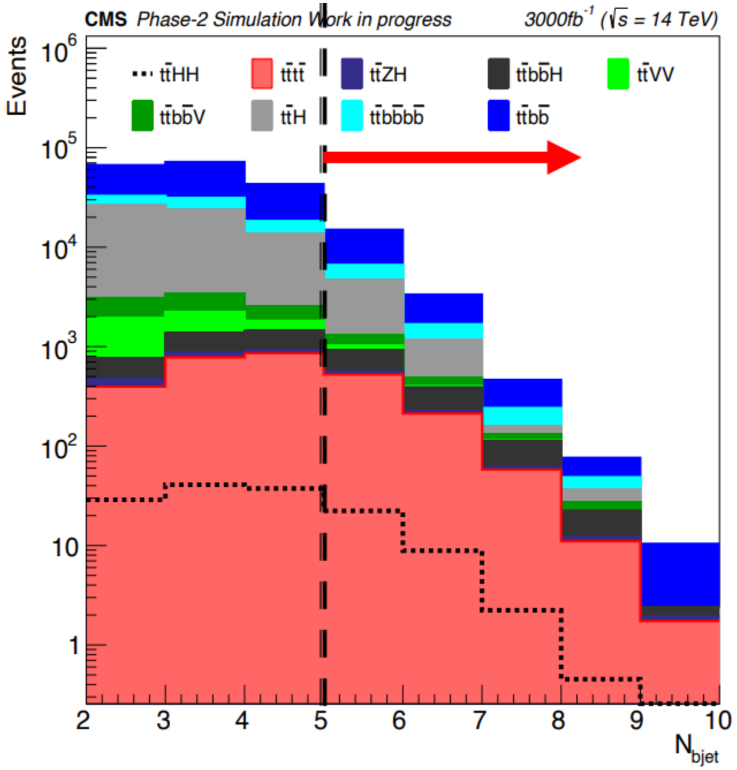
- OS2l:
  - $n_{lep} = 2$ (S1), with the **opposite** sign (S2)
  - $MET > 30$ GeV (S3)
  - $n_{B-tag} \geq \mathbf{5}$ (S4), Jets $H_T > \mathbf{500}$ GeV (S5)

OS2l selection illustration

S1 selection



S4 selection

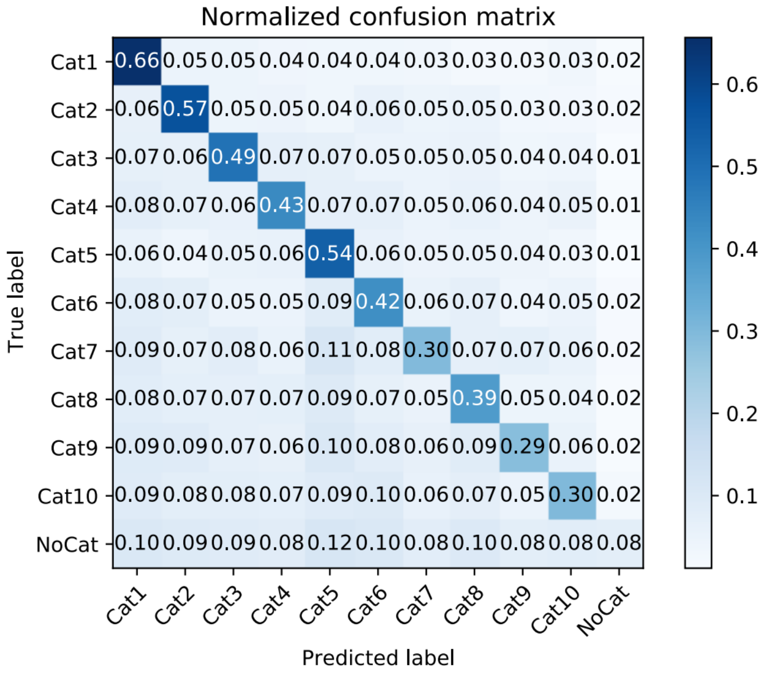


OS2l Higgs boson reconstruction (b-jet pairing)

- Traditional approach:
  - Minimizing $\chi^2 = \left((m_{jj} - m_{\mathrm{Higgs}})/\sigma_{\mathrm{Higgs}}\right)^2$
  - With only 16% accuracy.
- DNN jet matching:
  - Using Keras DNN determine the jet matching category

	Jet1	Jet2	Jet3	Jet4	Jet5	Category
Matchable	H	H	0	0	0	1
	H	0	H	0	0	2
	H	0	0	H	0	3
	H	0	0	0	H	4
	0	H	H	0	0	5
	0	H	0	H	0	6
	0	H	0	0	H	7
	0	0	H	H	0	8
	0	0	H	0	H	9
	0	0	0	H	H	10
	Not matchable					11

Category definition

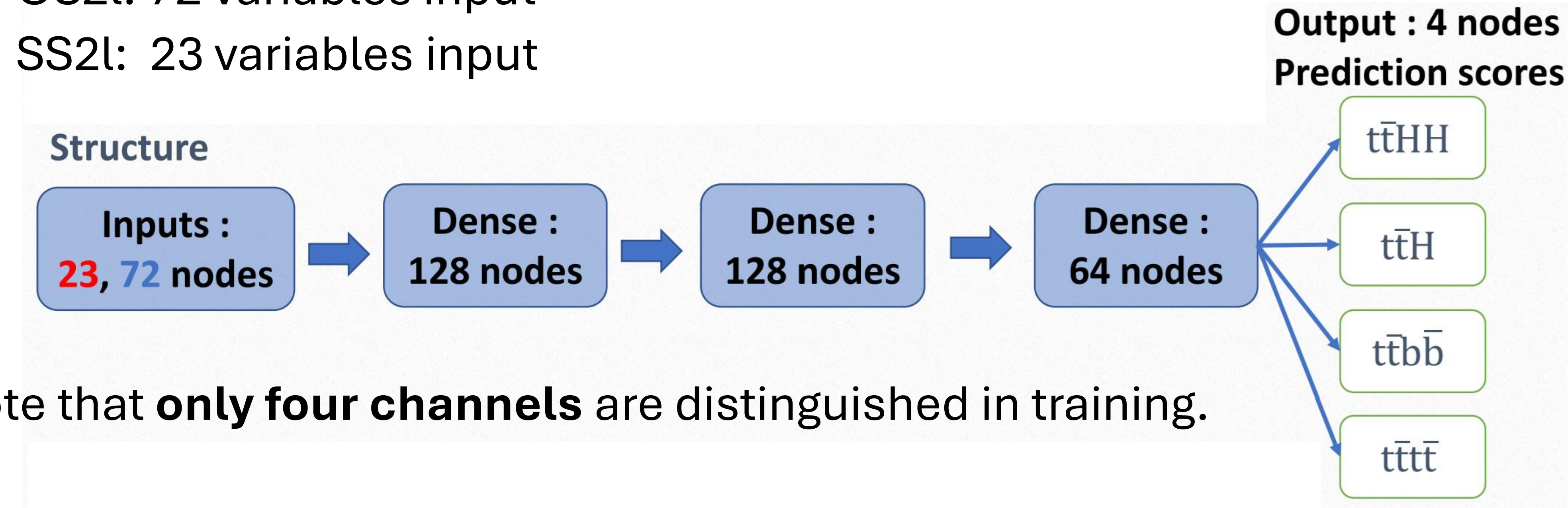


Performance

Better than invariant mass χ^2

DNN analysis

- Both OS2l and SS2l train a signal-background discriminant DNN **respectively**.
  - OS2l: 72 variables input
  - SS2l: 23 variables input



- Note that **only four channels** are distinguished in training.

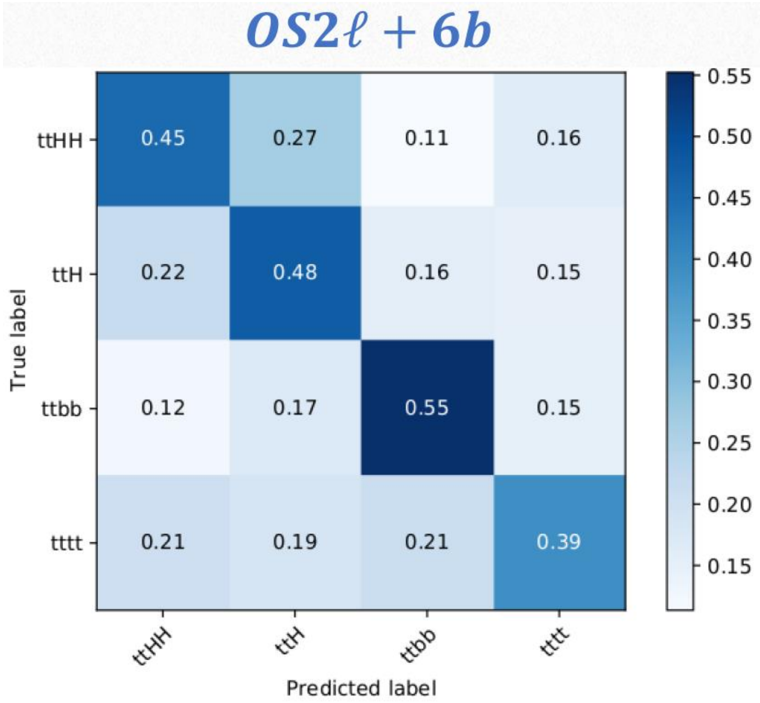
OS2l DNN

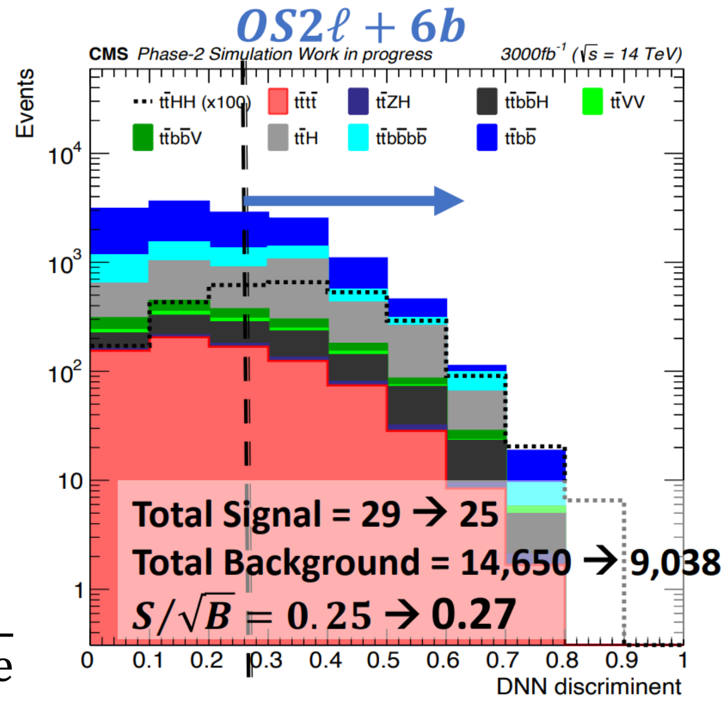
Variables	Description	Objects (x, y)
MET	Missing Transverse Energy	–
N_x	Multiplicity of x	l_1, l_2 (leptons: e or μ), b_1, b_2, b_3, b_4, b_5 (b−tagged jets) b_1, b_2, b_3, b_4, b_5
$p_T(x)$	Transverse momentum of x	
$\eta(x)$	Pseudorapidity of x	
$\phi(x)$	Azimuthal angle of x	
$m(x)$	Mass of x	b_1, b_2, b_3, b_4, b_5
$m_{max}(x, y)$	Maximum mass between x and y	
H_T	Scalar sum of transverse momenta of jets	
H_T^{norm}	H_T normalized to total energy	
$\Delta R(x, y)$	Distance ΔR between x and y	
$\Delta R_{avg}(x, y)$	Average ΔR between x and y	
$\Delta R_{min}(x, y)$	Minimum ΔR between x and y	
$\Delta R_{max}(x, y)$	Maximum ΔR between x and y	
$\tau_{twist}(x, y)$	Twist angle between x and y	
$m_H^1(x)$	Mass of the primary Higgs boson	
$m_H^2(x)$	Mass of the secondary Higgs boson	
$\Delta\eta^{bfh}(x)$	Pseudorapidity difference of x and y from Higgs	
$H_T^{bfh}(x)$	Scalar sum of p_T of b−jets from the primary Higgs	
$\Delta R^{bfh}(x)$	Distance ΔR between b−jet pairs from the primary Higgs	

Table 1: Input variables of the DNN event classifier in OS2l region. The total number of variables is 72. l_i represents leptons and b_i represents b−tagged jets, both ordered by p_T.

OS2l DNN input variables

$$\text{DNN } discriminant = \frac{t\bar{t}HH \text{ score}}{\sum_{Bkg} Bkg \text{ score}}$$



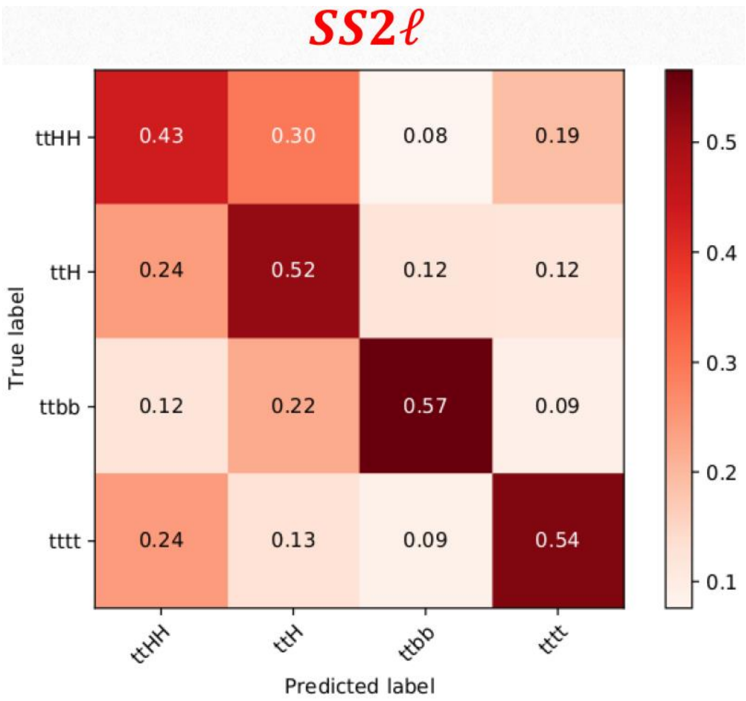


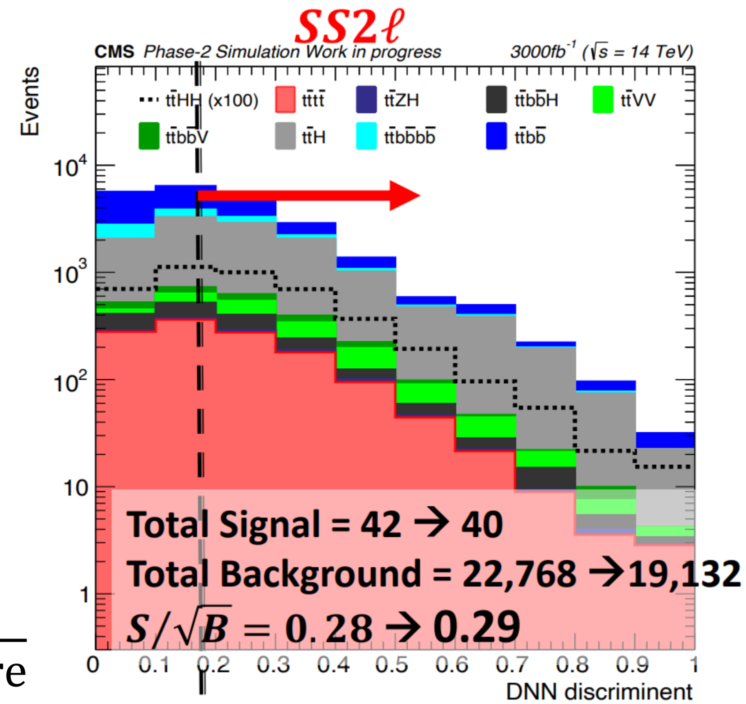
SS2l DNN

Variables	Description	Objects (x, y)
MET	Missing Transverse Energy	–
N_b	Multiplicity of b – jets	–
$p_T(x)$	Transverse momentum of x	l_1, l_2 (leptons), b_1, b_2, b_3 (b–tagged jets)
$\eta(x)$	Pseudorapidity of x	l_1, l_2 (leptons), b_1, b_2, b_3
$\phi(x)$	Azimuthal angle of x	l_1, l_2 (leptons), b_1, b_2, b_3
$m(x)$	Mass of x	b_1, b_2, b_3
$\Delta R(x, y)$	Distance ΔR between x and y	Between b_1, b_2, b_3 pairs

Table 2: Input variables of the DNN event classifier in SS2l region. The total number of variables is 23. l_i represents leptons and b_i represents b–tagged jets, both ordered by p_T.

SS2l DNN input variables





$$\text{Again, DNN } discriminant = \frac{t\bar{t}HH \text{ score}}{\sum_{Bkg} Bkg \text{ score}}$$

Cutflow and significance

	S0 (No)	S1 (2lep)	S2 (OS)	$S3^{MET>30}$	S4 b≥5	S5 Ht>500	S6 (DNN)
tthh	2844	241.59	170.92	153.39	33.94	28.17	6.17
tth	1836000	109315.4	82840.32	73944.9	4186.08	2653.02	293.76
ttbbh	46800	2995.67	2270.74	2024.57	585.47	486.25	83.3
ttzh	5130	479.81	384.49	321.65	49.3	44.58	8.46
ttvv	40560	5079.33	3574.96	3220.46	109.92	93.69	8.11
ttbbv	82080	5908.12	5072.54	4151.61	394.8	327.5	18.88
ttbb	4647000	181163.3	164257.5	145330.3	12012.5	7379.44	427.52
ttbbbb	1110000	36892.7	33066.9	29207.8	3063.6	2016.5	218.3
tttt	35430	4806.47	3161.64	2946.13	809.22	764.71	66.83
tt	2.28E+09	84282185	78506572	66437240	253346.4	108414	0
Significance	0.06	0.03	0.02	0.02	0.06	0.08	0.18

- **Same-sign requirement** has a better suppression of $t\bar{t}b\bar{b}$ and $t\bar{t}4b$ backgrounds.
- Thus SS2l can have less strict requirements on the jet selections.

	S0 (No)	S1 (2lep)	S2 (SS)	$S3^{MET>30}$	S4 b≥3	S5 Ht>300	S6 (DNN)
tthh	2844	241.59	70.67	63.55	44.13	42.96	41.17
tth	1836000	109315.4	26475.12	23308.02	10933.38	9877.68	9620.64
ttbbh	46800	2995.67	724.93	642.56	497.48	483.91	441.79
ttzh	5130	479.81	95.32	83.05	49.09	48.43	46.48
ttvv	40560	5079.33	1504.37	1406.62	562.16	528.9	516.73
ttbbv	82080	5908.12	835.57	695.22	359.51	344.74	320.11
ttbb	4647000	181163.3	16905.79	14112.94	9921.35	8276.31	7133.15
ttbbbb	1110000	36892.7	3825.8	3215.3	2308.8	1942.5	1520.7
tttt	35430	4806.47	1644.82	1534.25	1277.48	1270.13	1147.8
tt	4.15E+09	1.53E+08	10501115	8349434	2068684	1184773	1039529
Significance	0.04	0.02	0.02	0.02	0.03	0.04	0.04

- OS2l **DNN** performs well with more input and better-known Higgs structures.
- Its DNN cut removes all the $t\bar{t}$ events.

Results

- Currently results are reported as the exclusion limits of the hypotheses with signal strength $\mu = \sigma_{\text{measure}}/\sigma_{\text{SM}}$.
- Limits from yield-only (DNN shape not included yet) likelihood fit using the Combine tool.
  - **SS2l**: $\mu < 14$ at 95% Confidence Level.
  - **OS2l**: $\mu < 11$ at 95% Confidence Level.
  - **Combined** result: $\mu < 10$ at 95% Confidence Level.

- No pileup, background scaling, or experimental systematic uncertainties considered. Only a 2.5% luminosity uncertainty included.

Run 3 proposal

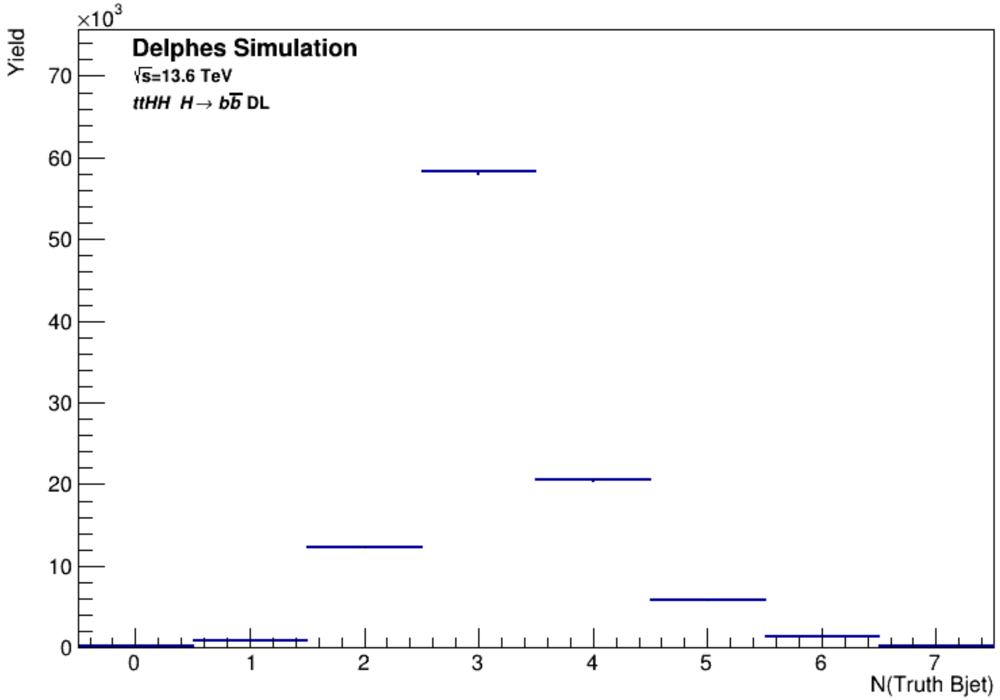
- The two lepton final state framework tends to complete. Pre-study for Run 3 analysis should start.
- Now still using MG5+Delphes sample. *CMS Delphes card* used for the Run 3 detector description.

- Currently I started checking from the Truth-level information of the toy sample:
  - $\sqrt{s} = 13.6\ \mathrm{TeV}$ 100k MC events $t\bar{t}HH$, $t\bar{t}$ DL + $HH \rightarrow 4b$

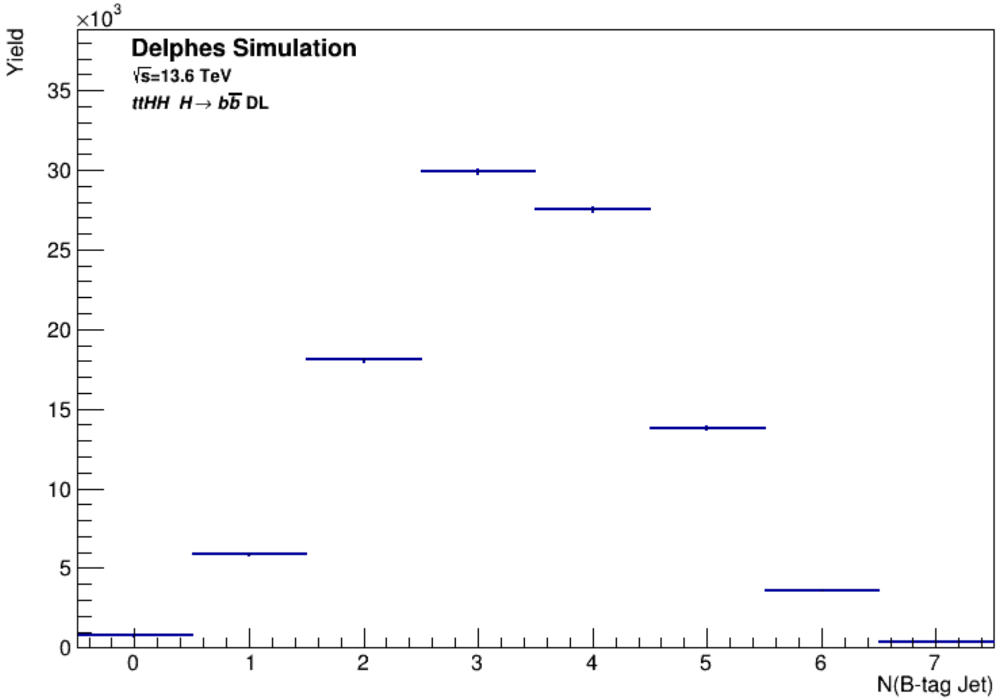
- *TODO: seek official generation and full simulation for CMS Monte Carlo samples*.

B-Jet Multiplicity

Truth



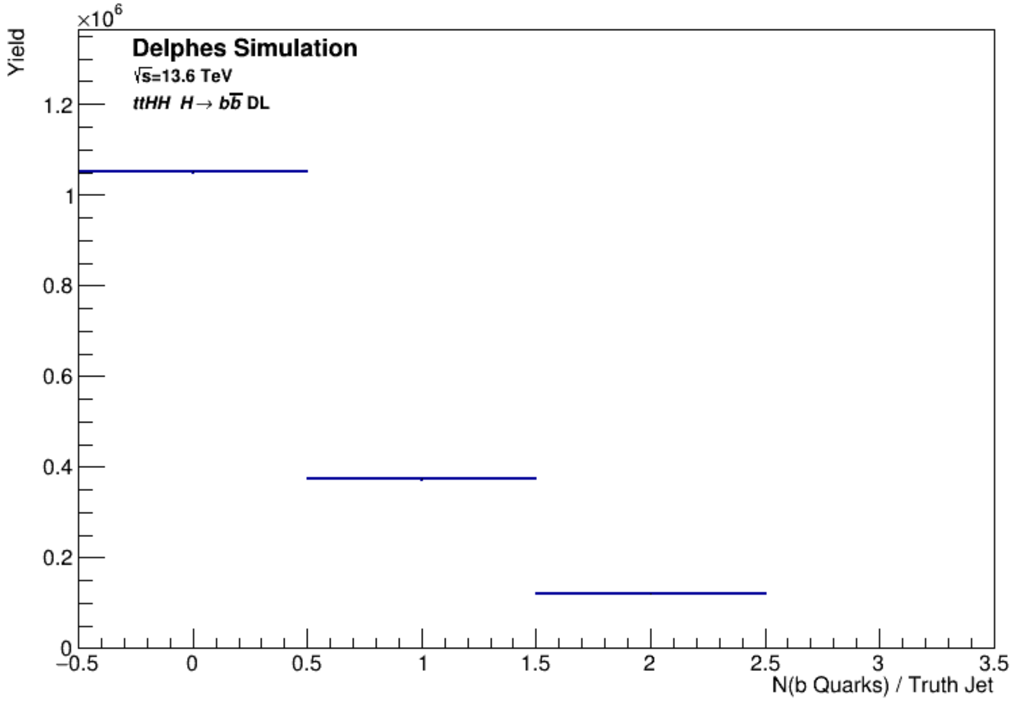
Reco



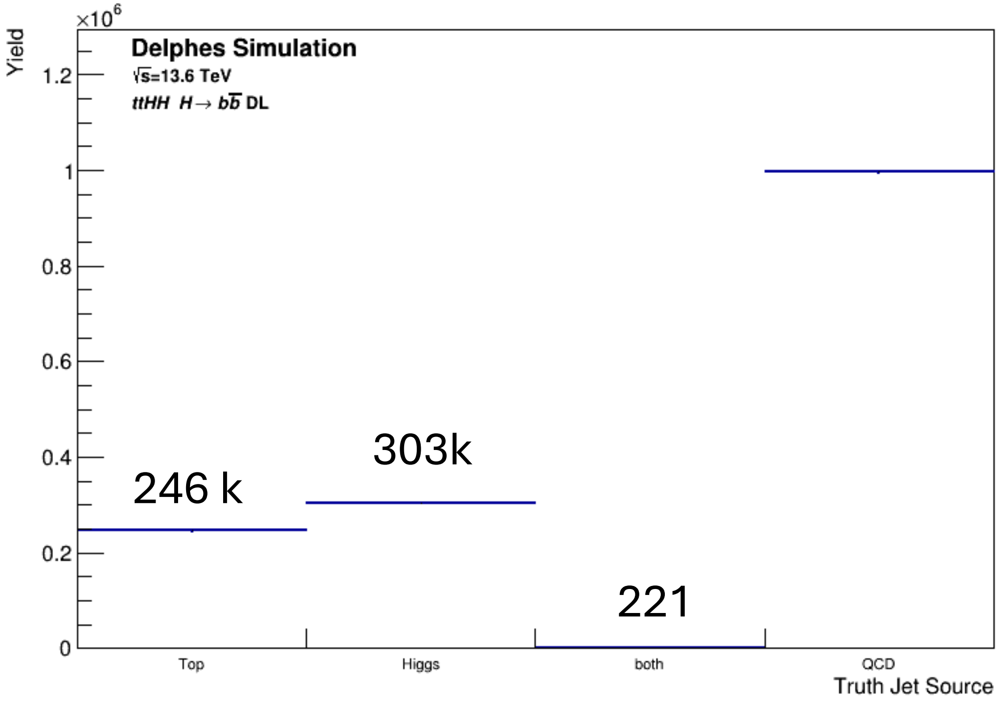
- Truth-level B-jet multiplicity also peaked at 3, even truth hardly reaches 6.

B quark in truth jet

Number of b quarks joining the truth jet hadronization

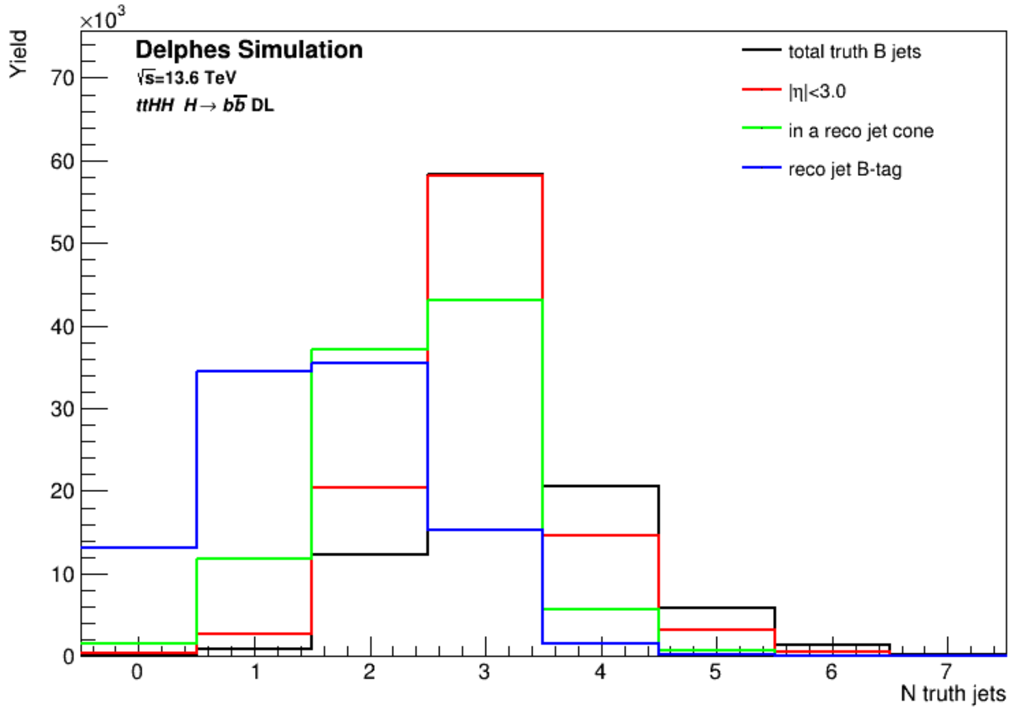


Source of truth jet



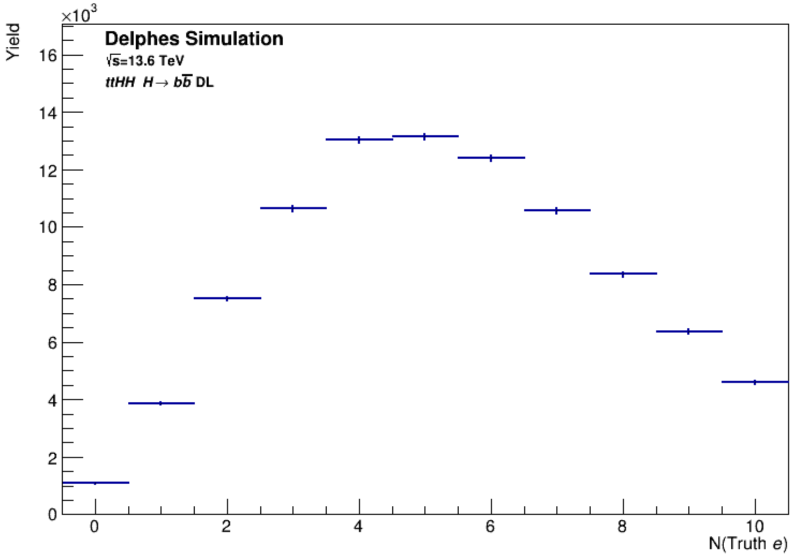
- One B-jet might use 2 b-quarks at a not small rate.
- Most jets come from only top quarks or only Higgs.

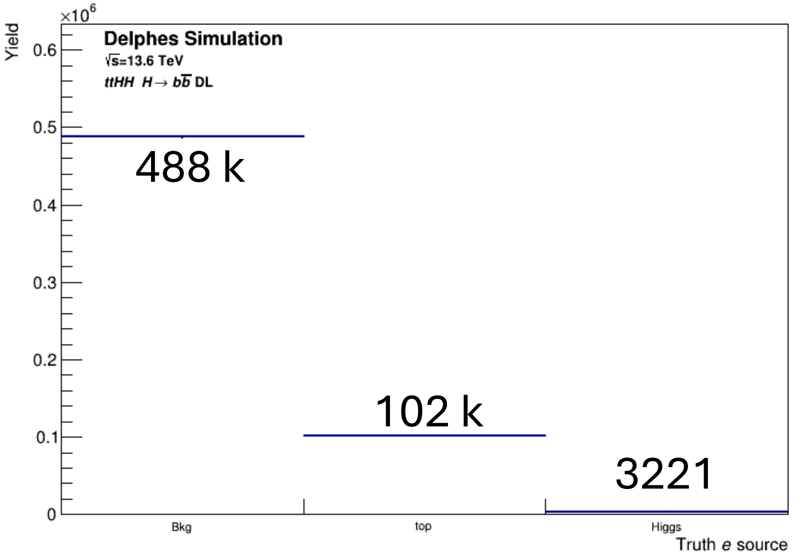
B-jet reconstruction performance

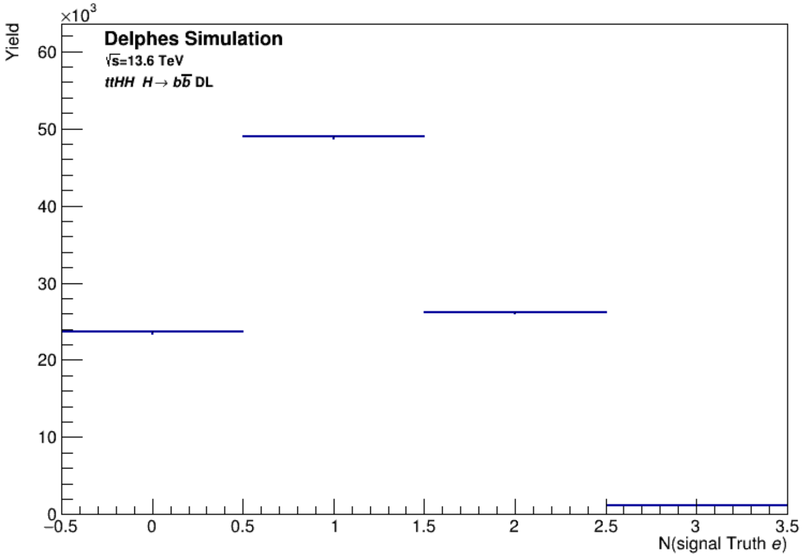


- Run 3 CMS calorimeter only covers $|\eta| < 3.0$ range. Some of the Truth B-jets are not in the detector response regions.
- A large number of truth B-jets are not centered inside a reconstructed jet cone.
- The reconstructed jet containing the truth B jet might not be b-tagged. B-tagging multiplicity is almost 1 less.

Truth Electrons





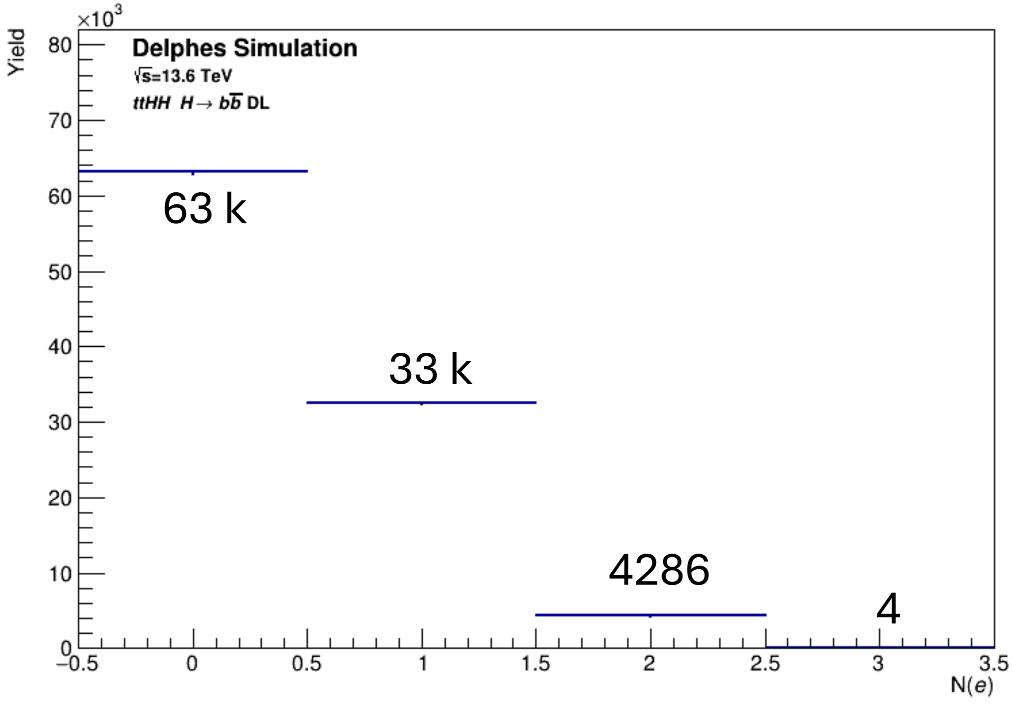


- Truth electron multiplicity peaked at 5.
- Most are non-prompt backgrounds from jets.
- Expected distribution 1:2:1 for $N(e) = 0,1,2$ excluding these jet electrons.

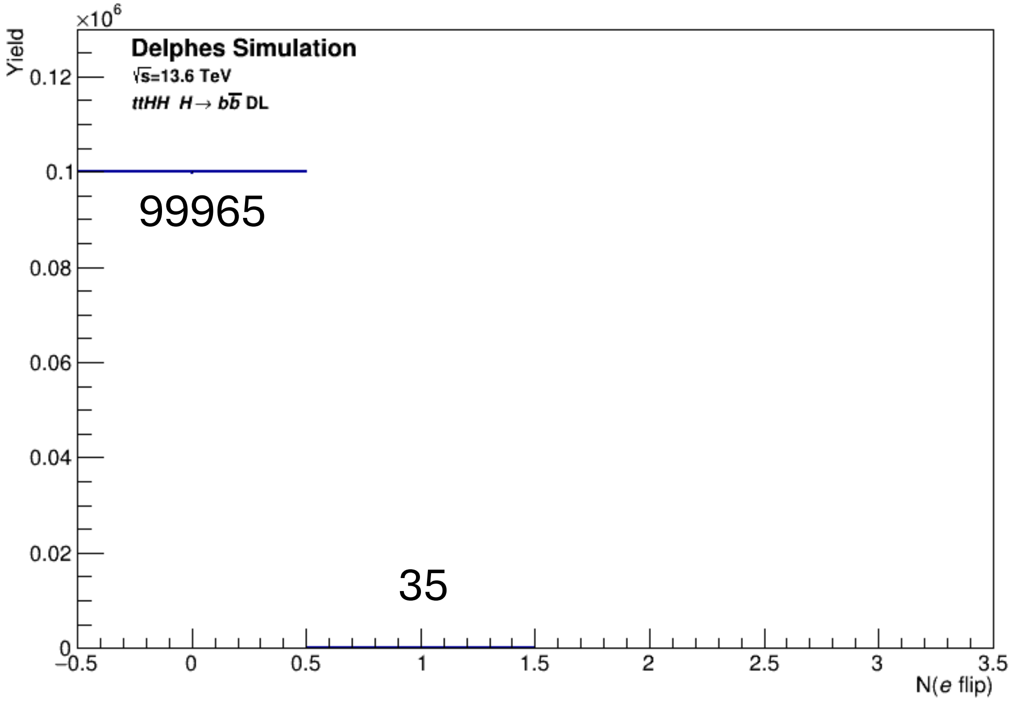
- Some truth electron comes from a Higgs boson.
- Might from b quark decay.

Reconstructed electron multiplicity

Total multiplicity

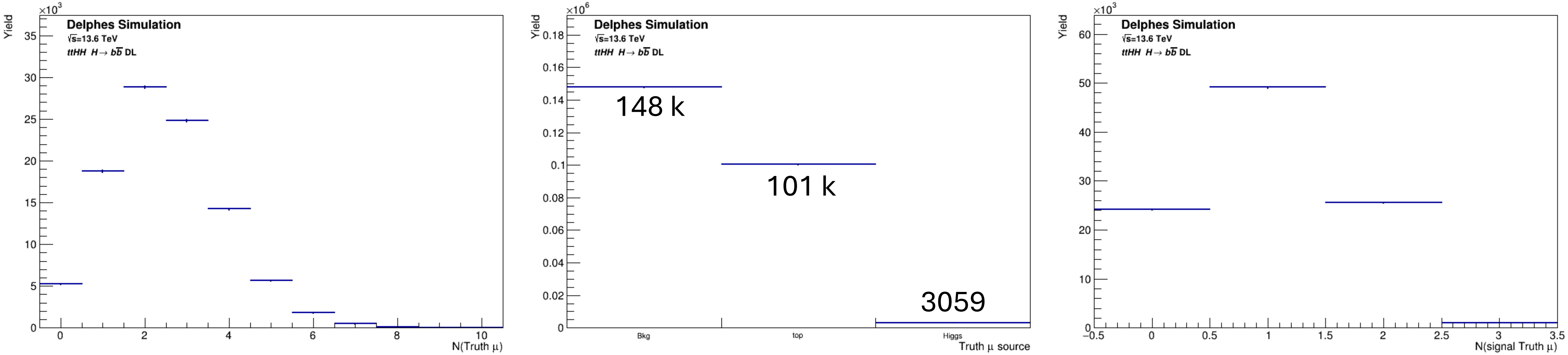


Charge-Flipping multiplicity



- $N(e)$ distribution after Delphes changes a lot → around 40% simulation & reconstruction rate (including detector response, reconstruction, identification, isolation…).
- 35 events in 100k have reconstruction electron charge flipped, very low rate.

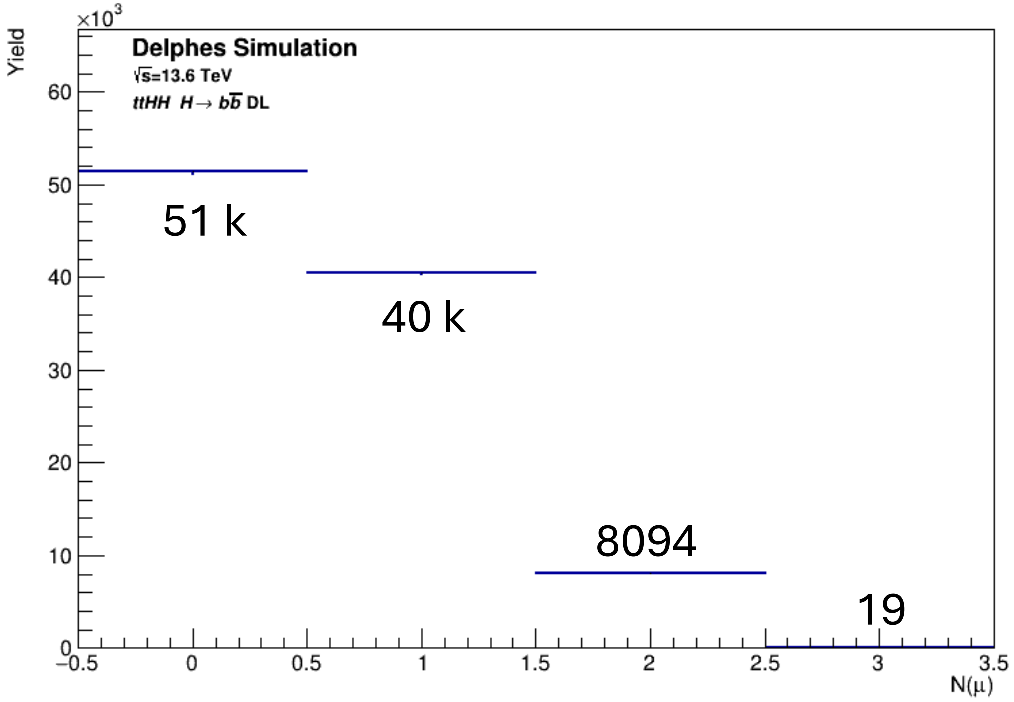
Truth Muons



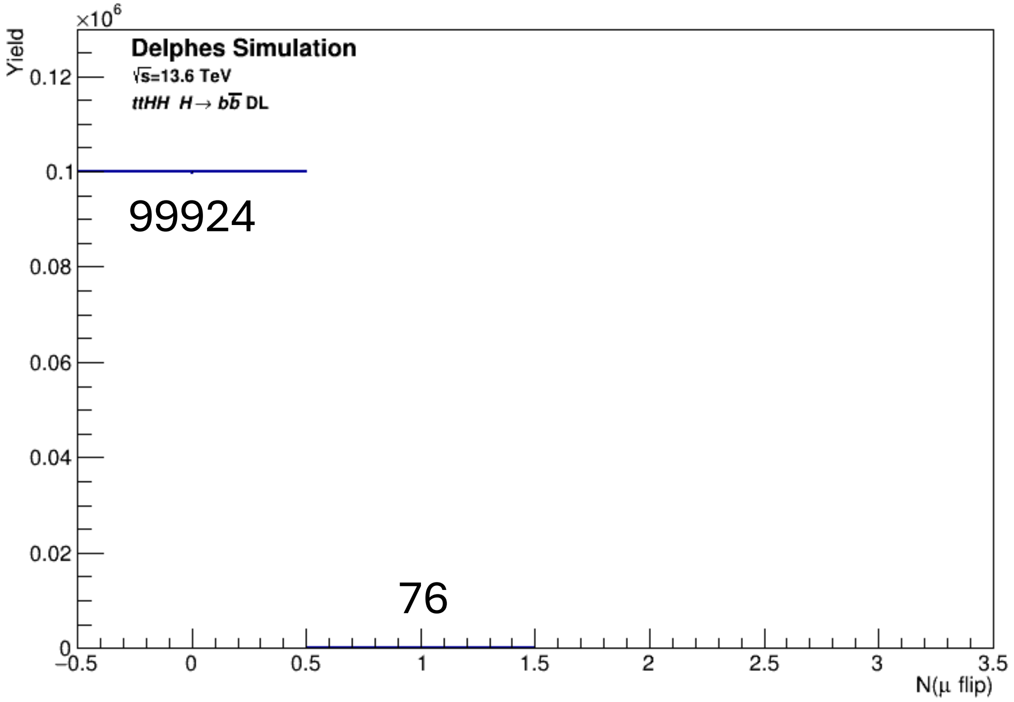
- Truth muon signal behaves similar to Truth electrons.
- The background muons are less than truth electron backgrounds.

Reconstructed muon multiplicity

Total multiplicity



Charge-Flipping multiplicity



- Muon has better total simulation & reconstruction rate, ~55%.
- 76 events in 100k have reconstruction muon charge flipped, **double** the rate of electrons.

Summary

- The HL-LHC study demonstrates the multi-b background suppression and potential of the SS2l channel.
- The DNN improve the OS2l Higgs jet matching significantly.
- Current stats-only yield based HL-LHC measurements has 14 (11) times SM signal strength limits at 95% CL.

- The Truth-level $t\bar{t}HH$ checking gives hint on the lower B-jet multiplicity, and provides estimation on lepton reconstruction and charge flipping.
- Dedicated CMS MC samples need to be prepared in the incoming study.