Polarization fraction studies in ssWW and VBS ZZ scattering using Deep Learning techniques

MBI-Thessaloniki, 26-28 August 2019

Junho Lee, Nicolas Chanon, Qiang Li, Andrew Levin,
Jing Li, <u>Meng Lu</u>, Yajun Mao

Based on: arXiv 1812.07591v2
arXiv 1908.05196v1

More and more interests on VBS:
- No BSM signature is found up to now, the measurement of deviation from SM could be alternative way to search BSM besides the direct search.
- VBS process is pure EW, though small xs, precisely predicted by SM.
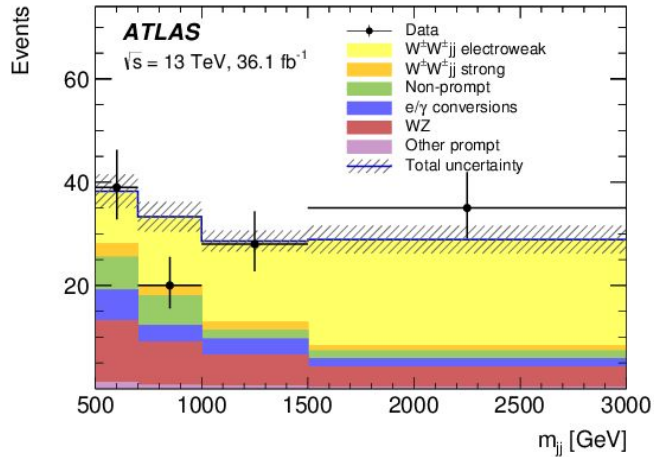- Unique topology, large Mjj, large $\Delta\eta$jj, suppress background significantly.
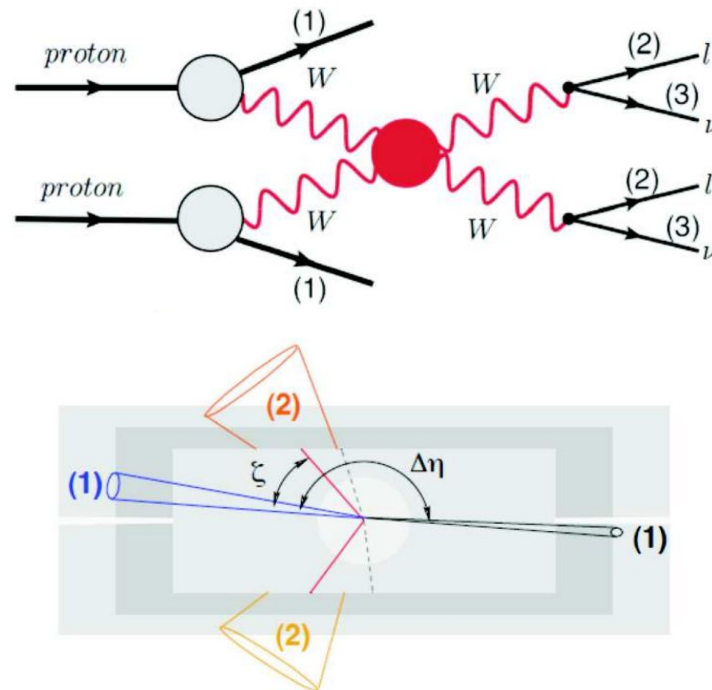
arxiv:1412.8367
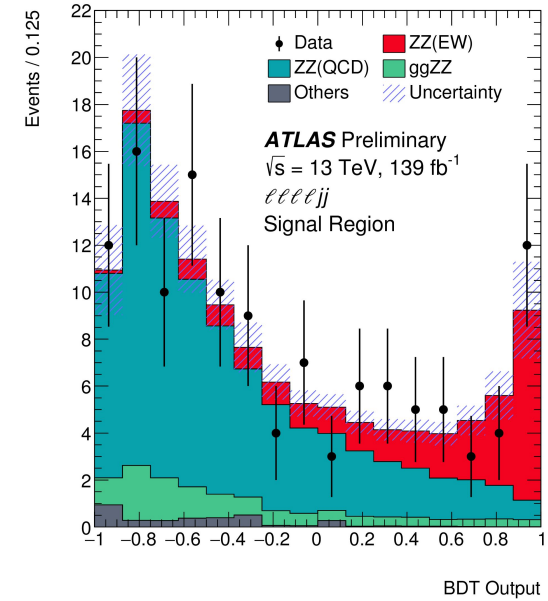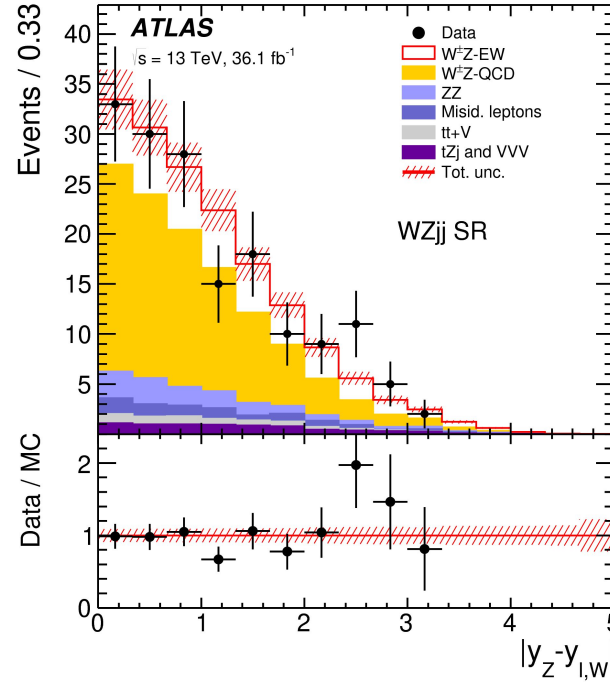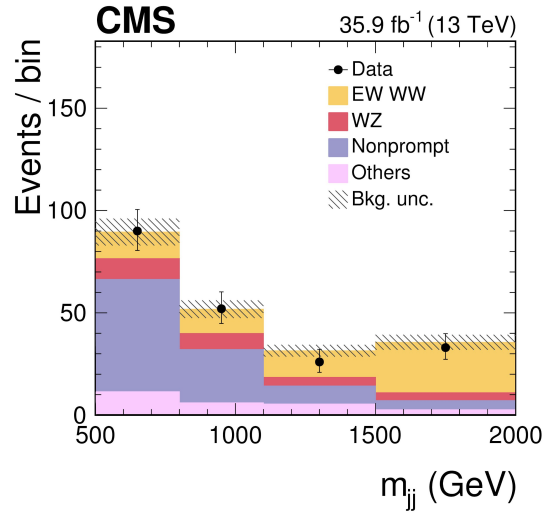
What's next...

$W_L W_L$ diverges if there is no Higgs boson or the Higgs boson is too heavy.

The Higgs boson was discovered and the mass is ~125 GeV.

There are theoretical models with composite Higgs bosons, the measurement of the longitudinal polarization will tell us the 125 GeV boson unitarizes ssWW scattering fully or only partially.
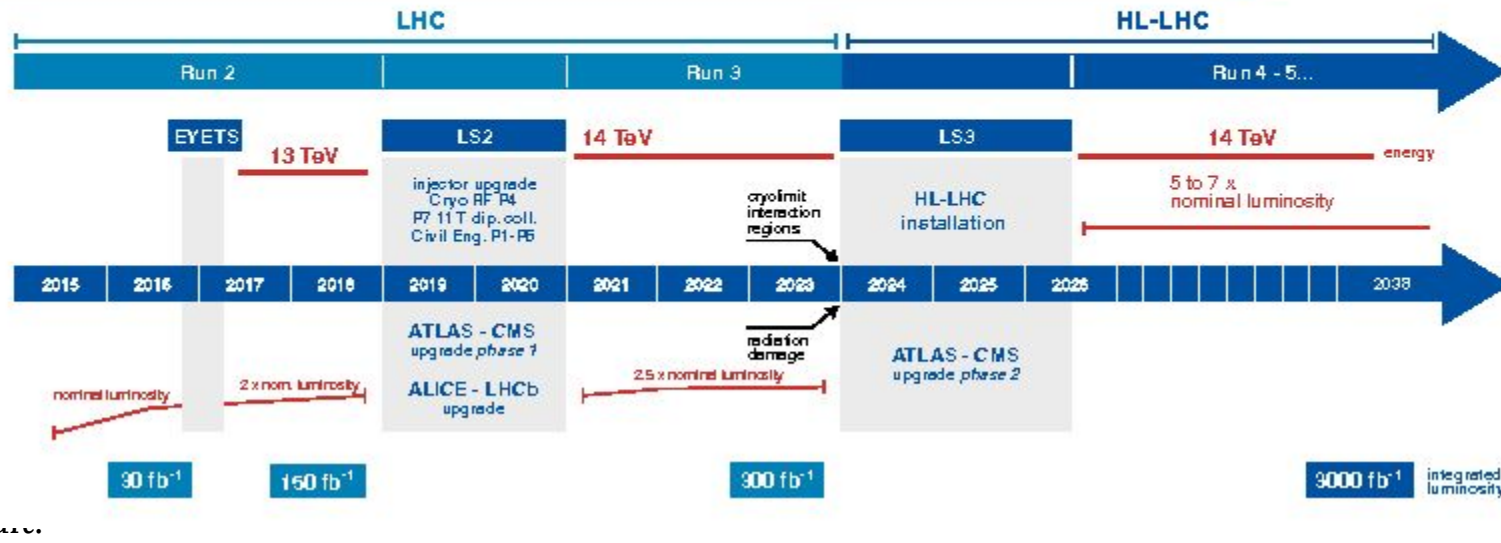
It's difficult:

- Small component, 5~10% in ssWW scattering
- Difficult to distinguish LL part from TT and TL

The Higgs boson-mediated diagram cancels the divergence of the cross section from the other processes
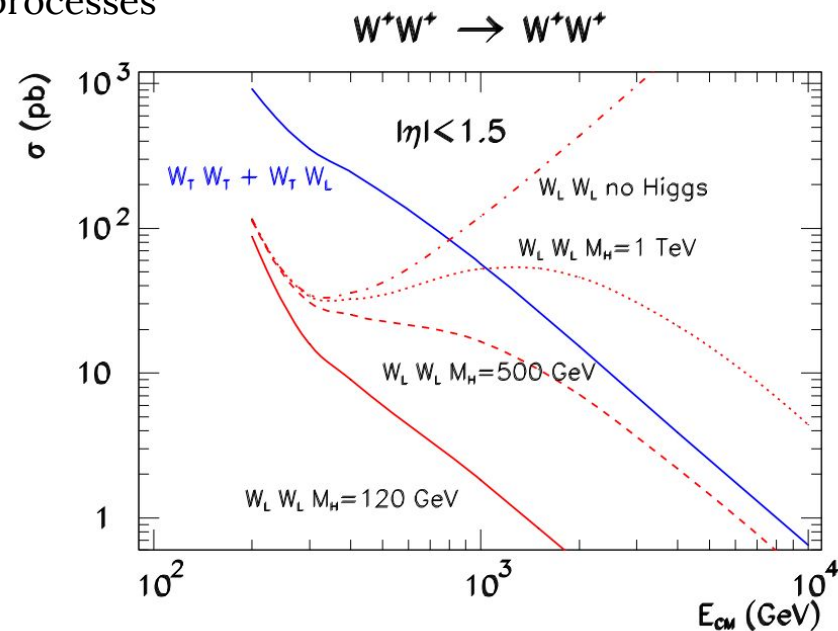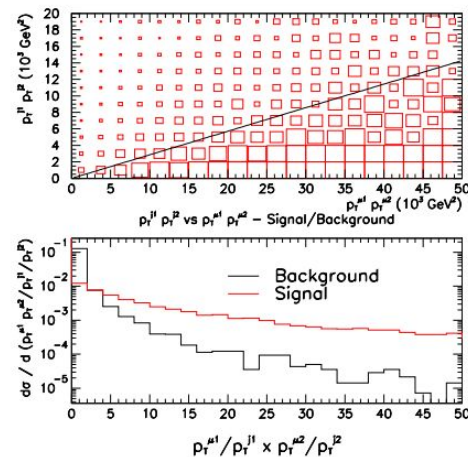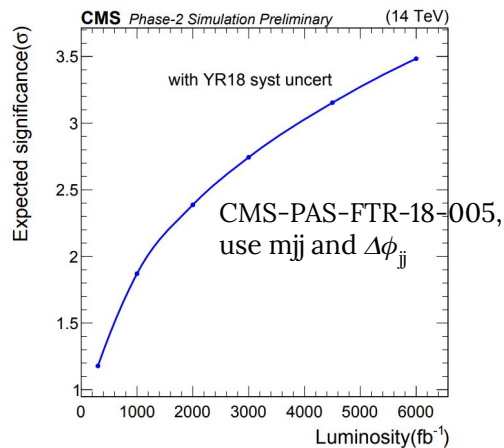

arxiv:1412.8367

HL-LHC

We a
cros
is no
Sinc
~125
Ther
Higg
pola
ssW
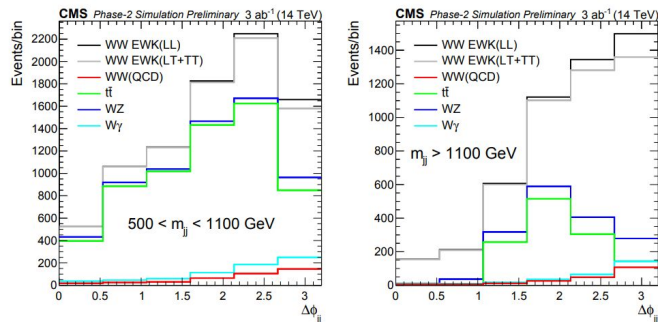
It's difficult.

- Small component, 5~10% in ssWW scattering
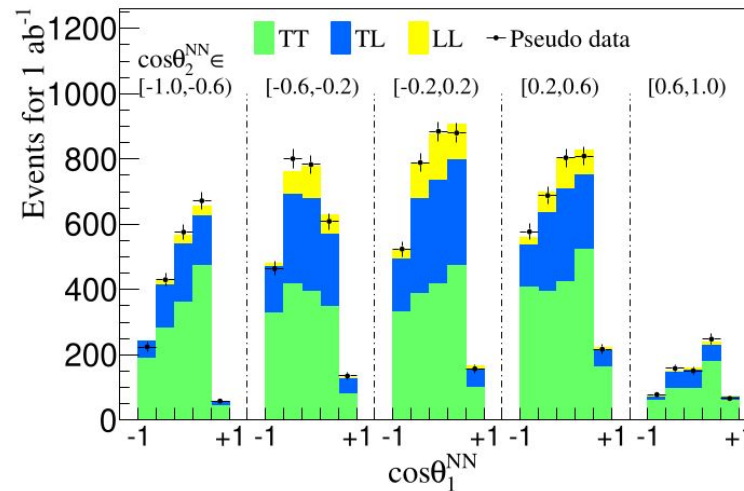- Difficult to distinguish LL part from TT and TL

The Hig...
...er

What if we have 3000fb⁻¹ data?

arxiv: 1201.2768

$$R_{p_T} = \frac{p_T^{l_1} \cdot p_T^{l_2}}{p_T^{j_1} \cdot p_T^{j_2}},$$

arxiv: 1510.01691, apply regression with DNN to recover the lepton angular distribution

CMS-PAS-FTR-18-005, use mjj and $\Delta\phi_{jj}$

MC production pipeline(4M events):

MadGraph5_aMC@NLO->

DEACY (decompose the process to LL, TT and TL)->

Pythia (for PS and hadronization)->

Delphes (detector simulation with CMS configuration)

Pileup is neglected

Based on arxiv: 1812.07591, different from regression method used in 1510.01691, we use DNN classification.

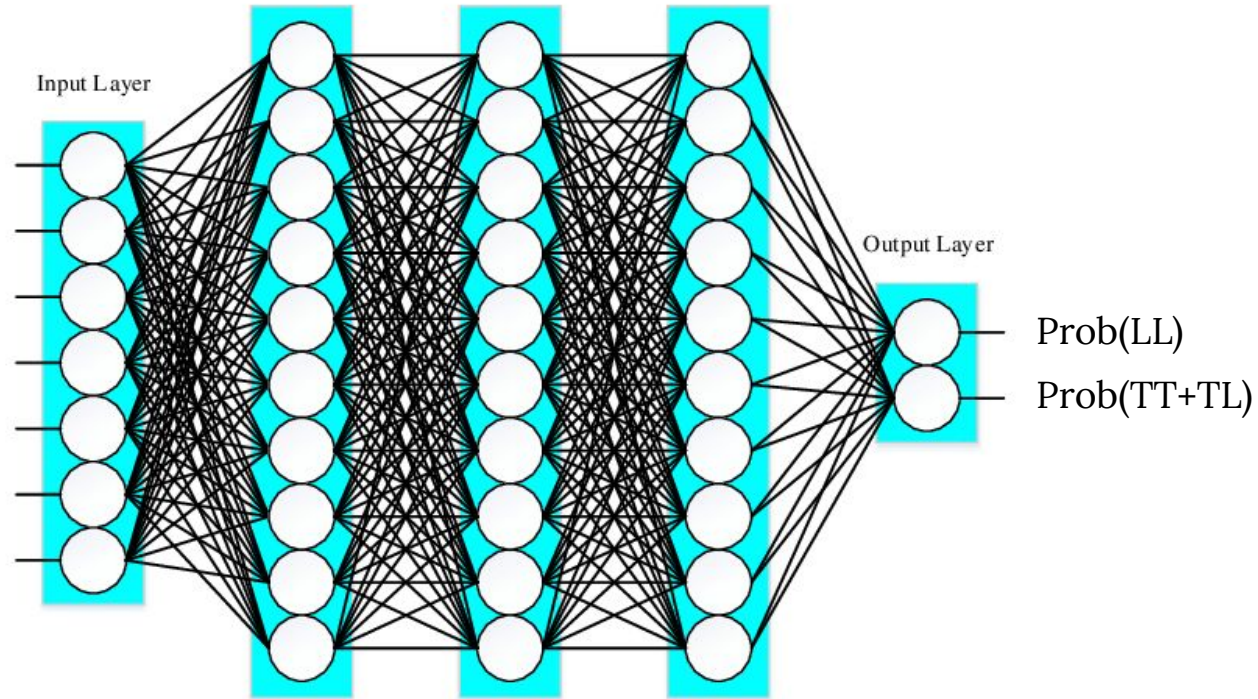Event selection
- Charged lepton
  - Charged lepton number = 2
  - Same electrical charge
  - pT > 20, $|\eta|$ < 2.4
  - Z veto
- Jet
  - Jet number >=2
  - pT > 30, $|\eta|$ < 4.7
  - Mjj > 850
  - $\Delta\eta_{jj}$ > 2.5
- Others
  - MET > 40
  - B jet veto applied

Training dataset: Obtain Trained DNN model -> LL and TTTL classifier
Test dataset: Test performance of the model

Inputs:
Low level:
4-momenta of 2
leptons, 2 jets;
MET;
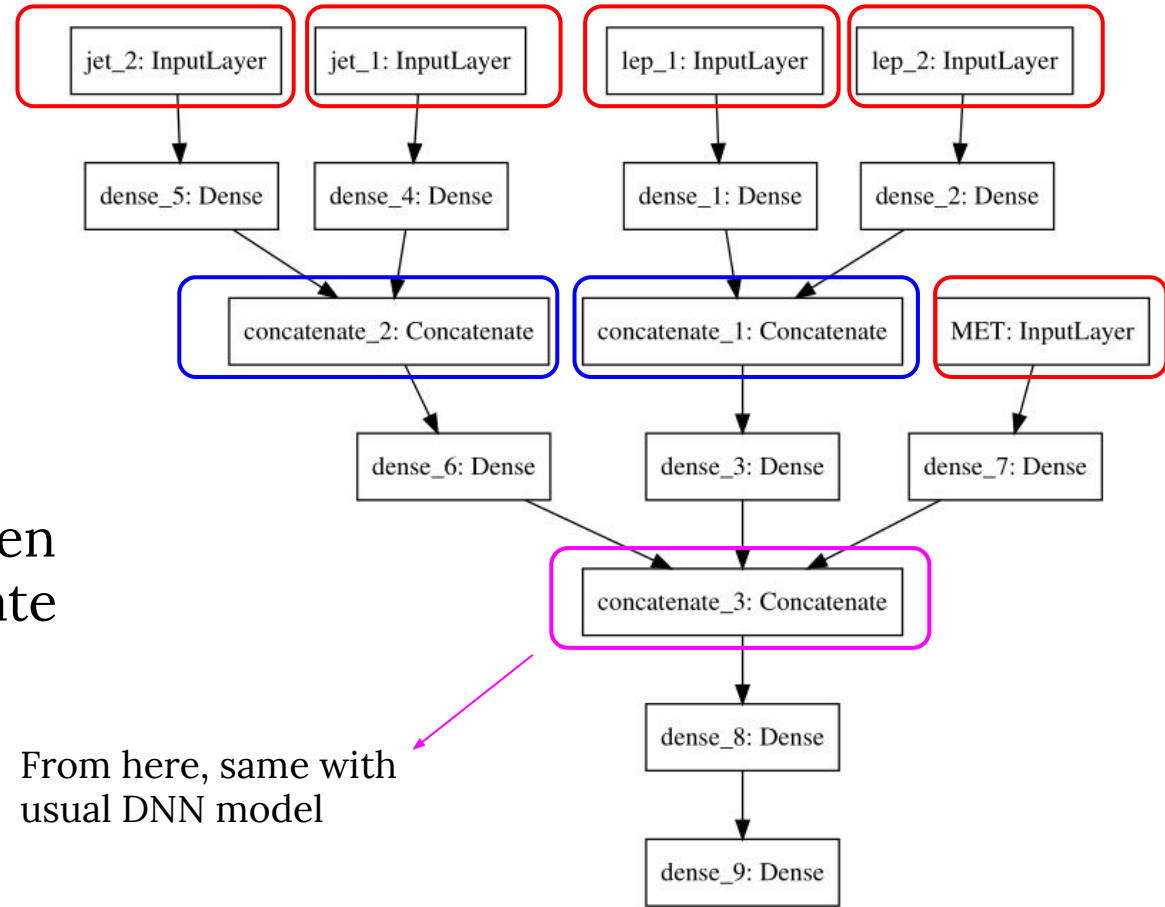High level: $\Delta\phi_{jj}$;
$\Delta\eta_{jj}$; dR_ll_jj;
zeppen_lepton

10 Hidden layers with 150 nodes in each layer
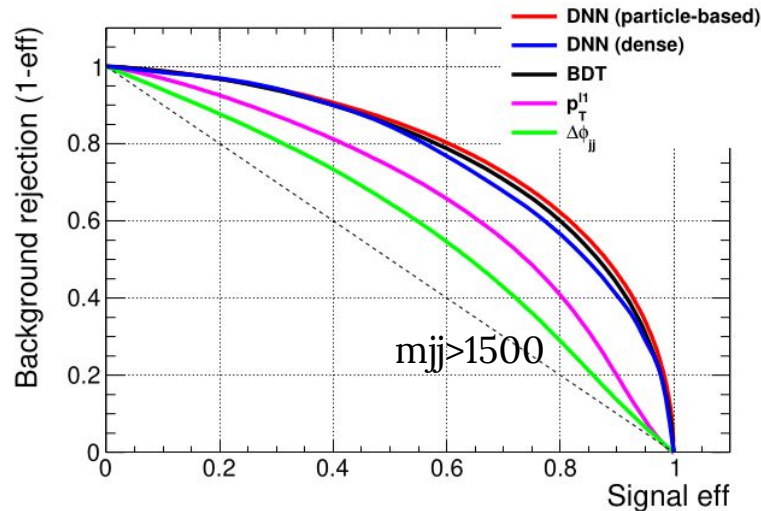
Input Layer

Output Layer

Prob(LL)

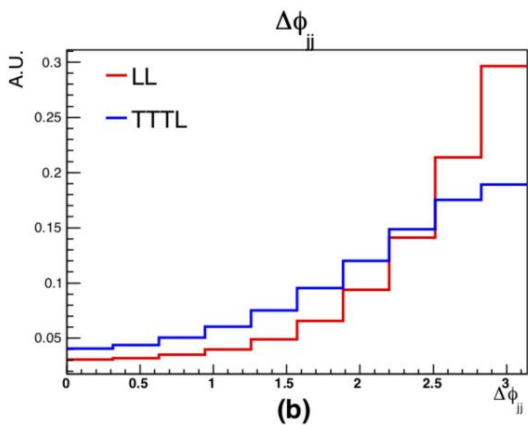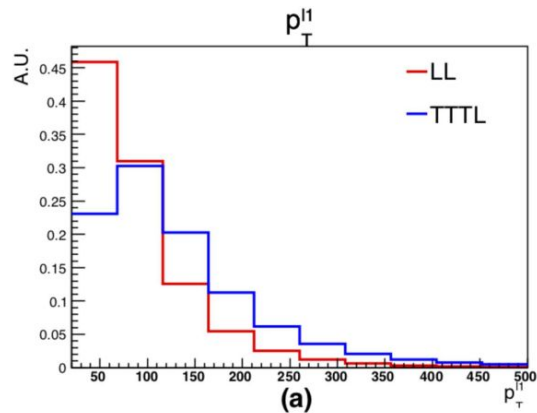Prob(TT+TL)

Simplified DNN
particle-based
model

Input variables are
grouped into object
sub-system, i.e. jet,
lepton and MET. Then
merged after separate
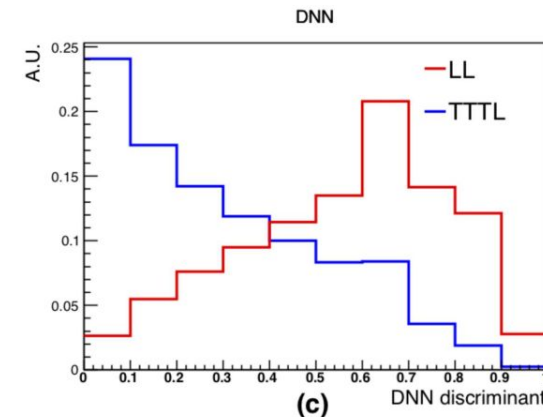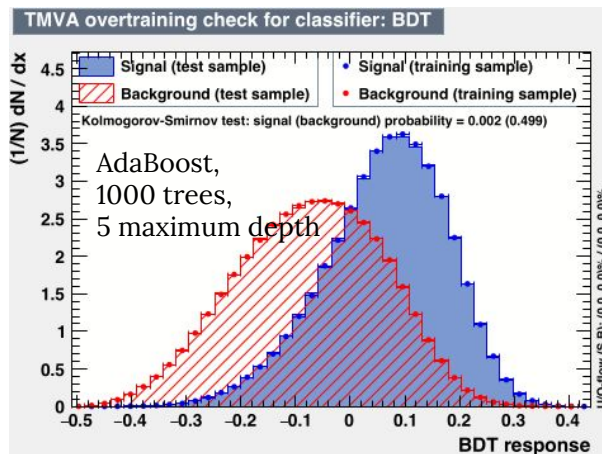DNN layers

From here, same with
usual DNN model

DNN(both usual and particle based models) and BDT have similar behaviors, all of them have much better discrimination than pt_l1 and dphi_jj.

mjj>1500

AdaBoost, 1000 trees, 5 maximum depth

Use the shapes of LL and TT+TL, the LL component fraction could be extracted:

m_{jj} cut	True Fraction	$p_T^{l_1}$	$\Delta\phi_{jj}$	DNN
> 850 GeV	6.66%	$6.67\%^{+1.95\%}_{-1.90\%}$	$6.67\%^{+2.80\%}_{-2.76\%}$	$6.66\%^{+1.11\%}_{-1.04\%}$
> 1200 GeV	6.68%	$6.70\%^{+2.26\%}_{-2.22\%}$	$6.70\%^{+3.29\%}_{-3.25\%}$	$6.68\%^{+1.26\%}_{-1.20\%}$
> 1500 GeV	6.67%	$6.71\%^{+2.62\%}_{-2.57\%}$	$6.68\%^{+3.85\%}_{-3.80\%}$	$6.67\%^{+1.44\%}_{-1.37\%}$
> 1800 GeV	6.69%	$6.70\%^{+3.02\%}_{-2.96\%}$	$6.68\%^{+4.48\%}_{-4.42\%}$	$6.69\%^{+1.63\%}_{-1.56\%}$
> 2000 GeV	6.66%	$6.67\%^{+3.34\%}_{-3.27\%}$	$6.66\%^{+4.98\%}_{-4.93\%}$	$6.66\%^{+1.79\%}_{-1.71\%}$

The results using DNN are always better than other two methods.
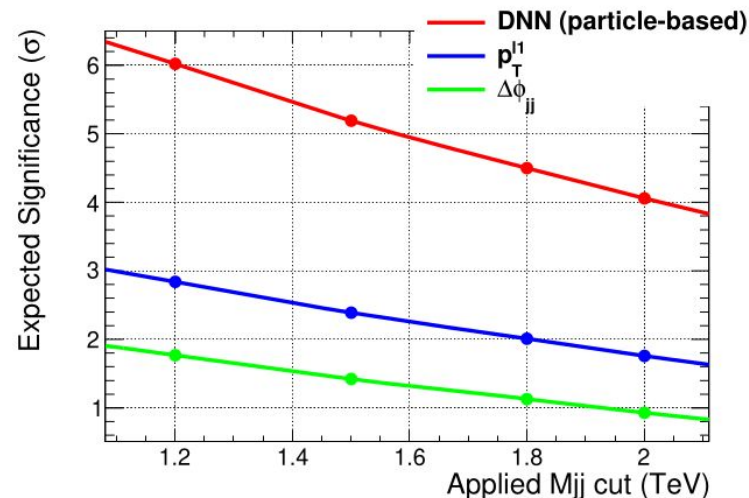
Other bkgs are neglected
2% lumi uncertainty
5% syst uncertainty
Stat uncertainty

With a cut Mjj>2TeV where bkgs can be neglected, the significance is around 4σ. After combining CMS and Atlas, the significance should reach 5σ

Based on arxiv: 1908.05196v1.

MC production pipeline:

MadGraph5_aMC@NLO->

DEACY (decompose the process to LL, TT and TL)->

Pythia (for PS and hadronization)->

Delphes (detector simulation with CMS configuration)

Pileup is neglected

Event selection
- Charged lepton
  - Charged lepton number >= 4
  - 60 < mll < 120
  - pT > 5, $|\eta|$ < 2.4
  - pT >20(10) for leading(subleading)
  - If more than 1 ZZ combination, select those with smallest $(m_{ll1} - m_Z)^2 + (m_{ll2} - m_Z)^2$
- Jet
  - Jet number >=2
  - pT > 25, $|\eta|$ < 4.7
  - Mjj > 400
  - $\Delta\eta_{jj}$ > 2.4

- Others
  - B jet veto applied

Based on arxiv: 1908.05196v1.

MC production p̲
MadGraph5_aMC̲
DEACY (decompo̲
Pythia (for PS and̲
Delphes (detector̲

After event selection(unweighted events left):
LL: 100000
TL: 150000
TT: 240000
qqZZ: 48000
ggZZ: 40000

Event selection
- Charged lepton
  - Charged lepton number >= 4
  - 60 < mll < 120
  - pT > 5, $|\eta|$ < 2.4
  - pT >20(10) for leading(subleading)
  - If more than 1 ZZ combination, select those with smallest $(m_{ll1} - m_Z)^2 + (m_{ll2} - m_Z)^2$
- Jet
  - Jet number >=2
  - pT > 25, $|\eta|$ < 4.7
  - Mjj > 400
  - $\Delta\eta_{jj}$> 2.4
- Others
  - B jet veto applied
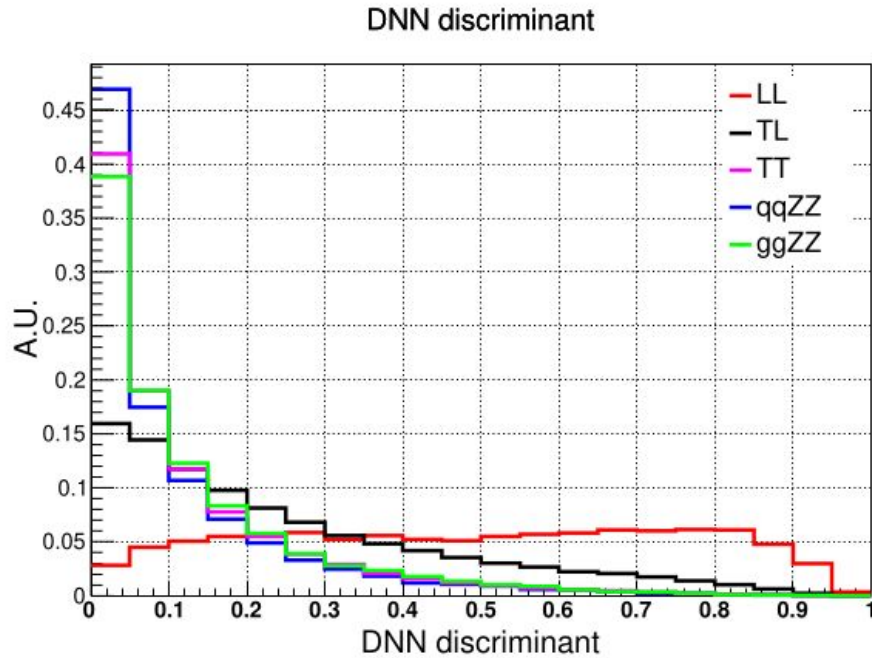
2-step BDT model:
BDT1: similar with the one tested in ssWW case;
BDT2: training events after applying cut on BDT1 score which maximizes $S/\sqrt{B}$
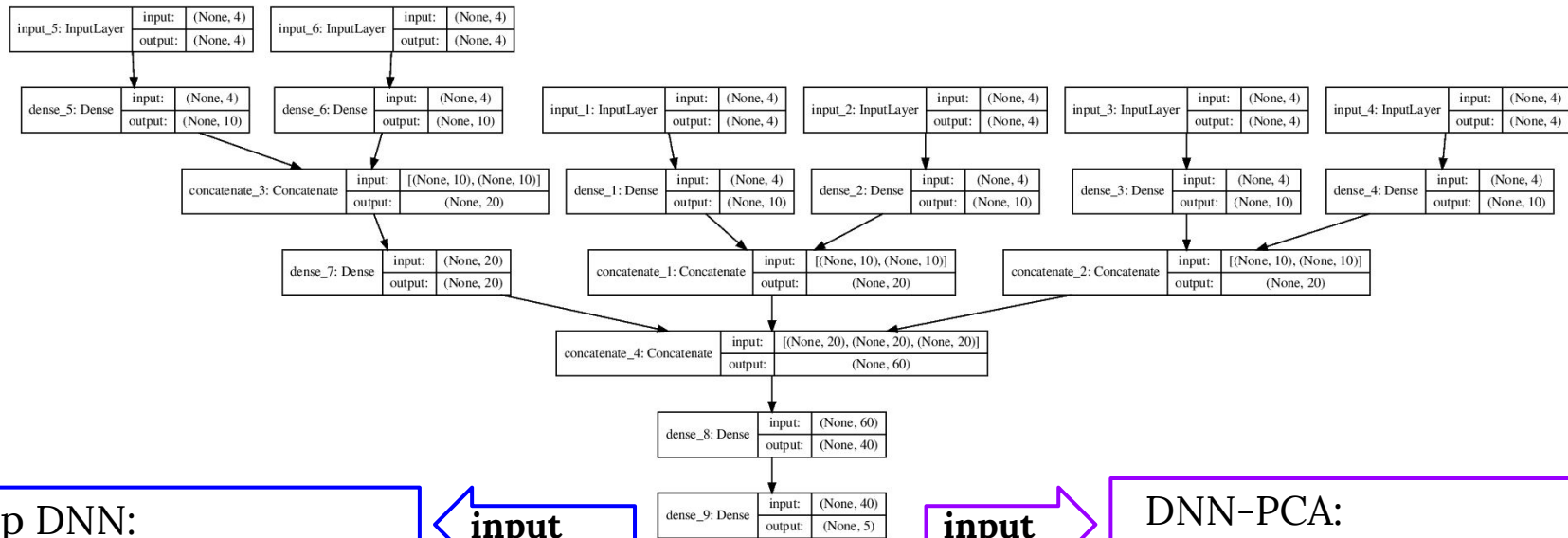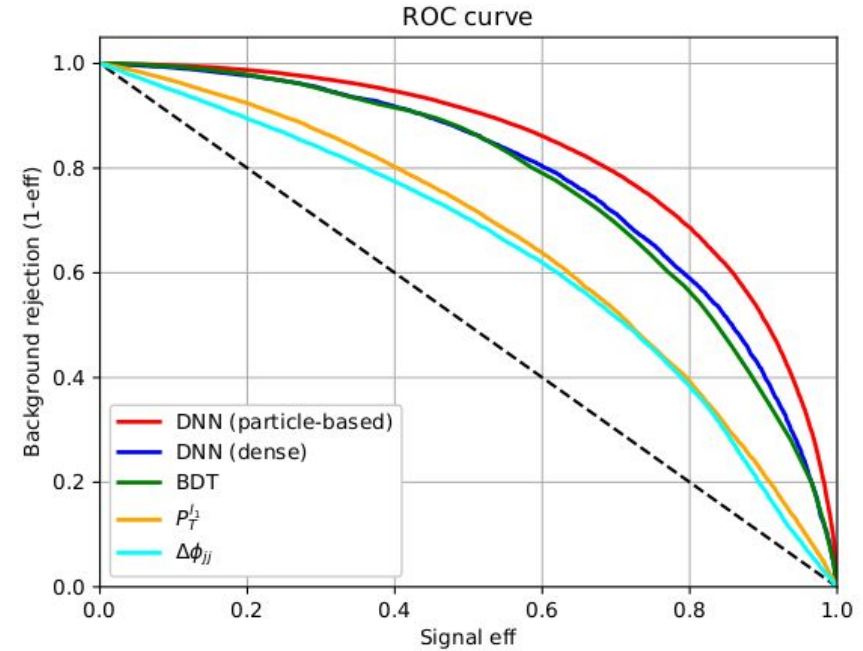

2-step DNN:
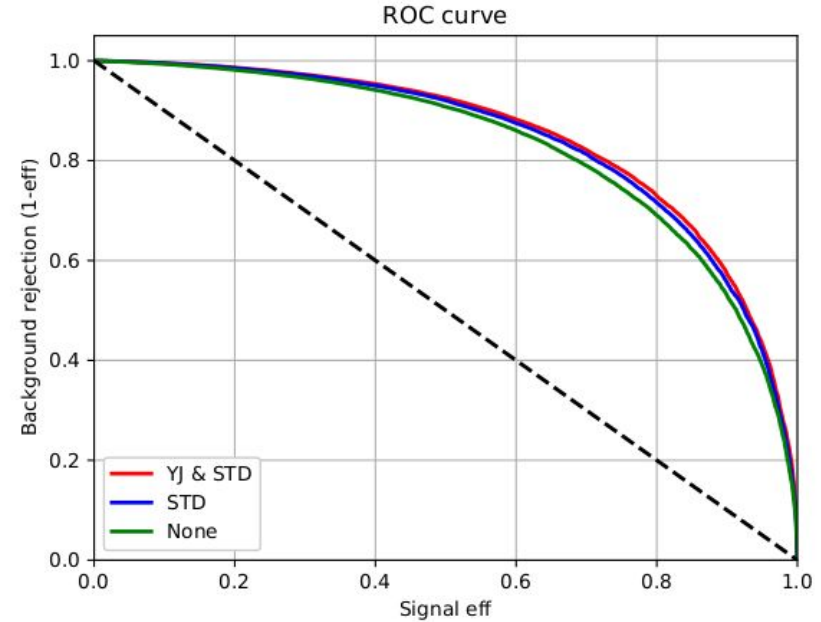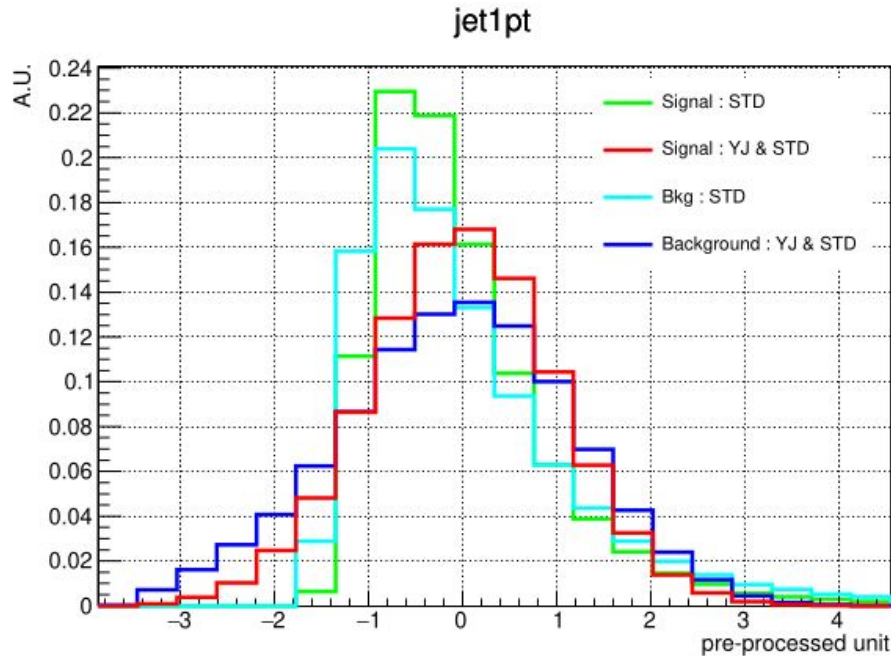Second shallow DNN with 2 hidden layers

input

input

DNN-PCA:
Using principle component analysis as the next step

DNN discriminant

ROC curve

LL score of particle-based DNN

ROC curve shows that the particle-based DNN has better discrimination than BDT and normal DNN configuration.

Lyon 1

jet1pt

Different pre-processing method applied on the sample, only standardscaler(STD), or STD&Yeo-Johnson(YJ) transform

The ROC curve shows that the STD&YJ could improve the discriminant.

Since the output of first particle-based DNN is 5-dimension, the inputs are then transferred to 5 principal components.

Principle component	PC1	PC2	PC3	PC4	PC5
Explained variance ratio	64.8%	18.1%	13.0%	4.2%	< 0.1%
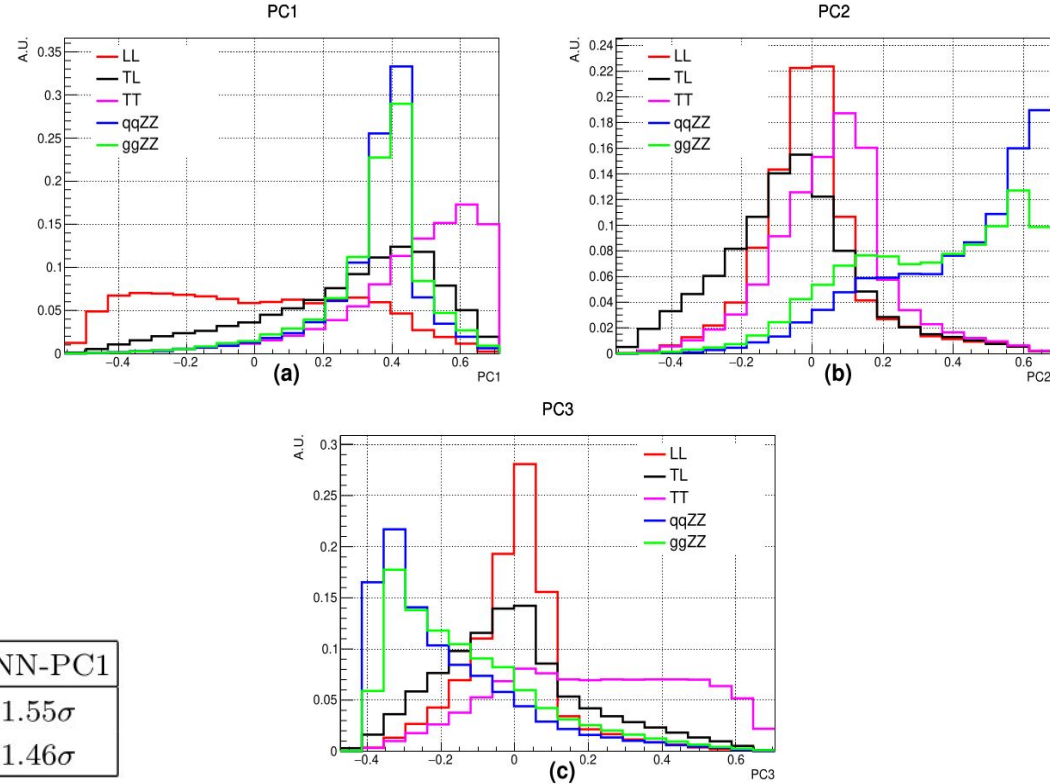
Significance is calculated using multi-bin fit.

Uncertainty	BDT	DNN	STD DNN	YJ&STD DNN	DNN-PC1
Statistical	1.41σ	1.42σ	1.43σ	1.47σ	1.55σ
Stat. & syst.	1.23σ	1.31σ	1.33σ	1.38σ	1.46σ
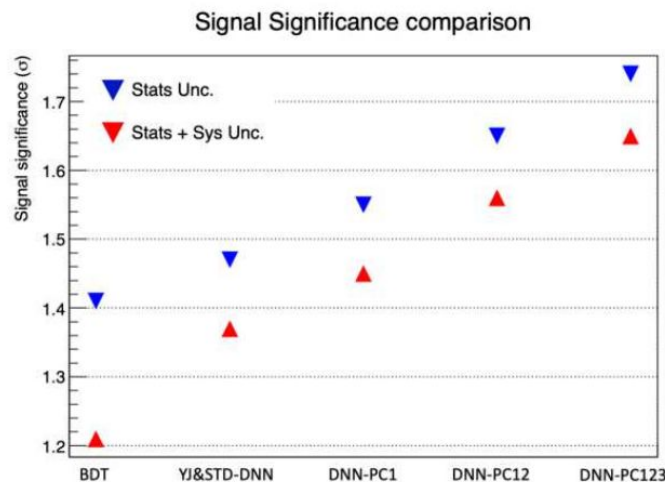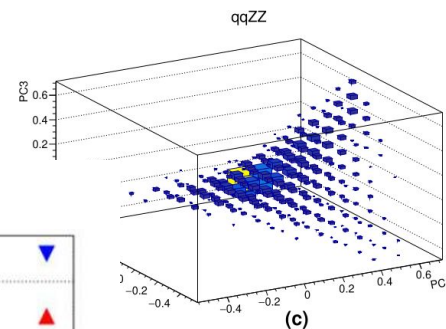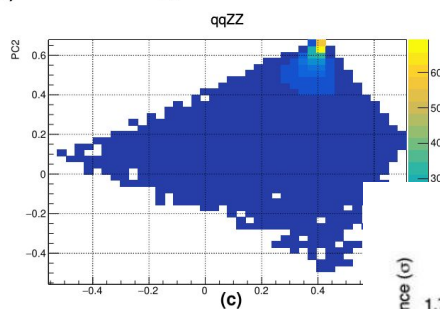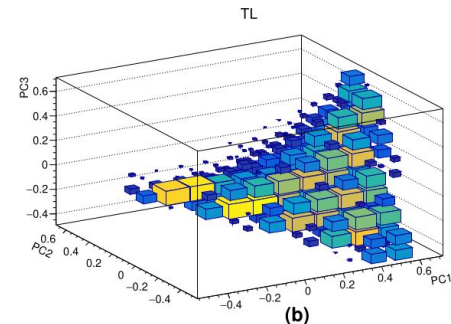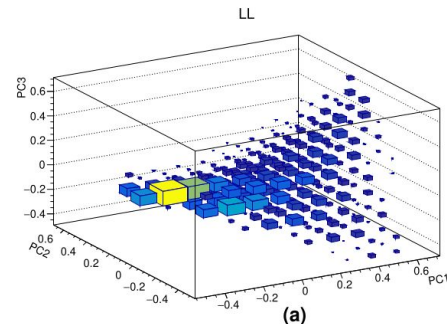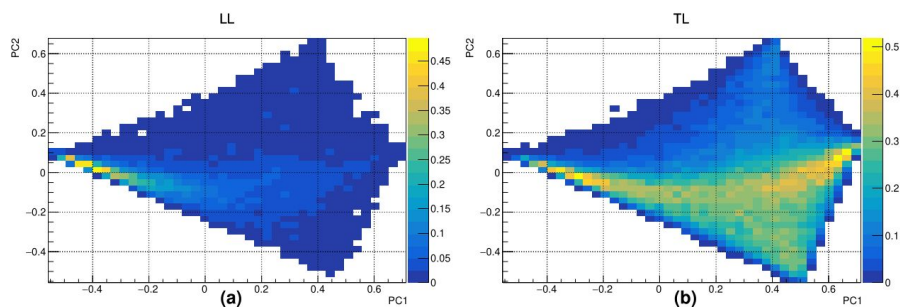
The significance for $Z_L Z_L$ events with considering stat uncertainty and 10% sys uncertainty is ~1.65σ with DNN-PCA123 method.

Summary:

- DNN turns out to be a very powerful tool in the polarization study of VBS process.
- With DNN classification method, the sensitivity of longitudinal fraction of ssWW could promisingly reach 5_σ, after combining the data from CMS and Atlas, ~6000fb^{-1}.
- There is improvement on the sensitivity on longitudinal part of VBS ZZ by using the DNN classification method, but the sensitivity is still quite limited.

Thanks!

Backup

DNN dense model for ssWW:

- 10 layers DNN with 150 nodes in each layer
- Activation function: relu
- Final nodes function: sigmoid
- Optimizer: adam, learning rate 0.001
- Regularization: L2 with 0.01 regularization term
- Batch size: 50 events
- 50% dropout in hidden layers
- He's uniform for weight initialization

Overfitting controlled by the epoch selection(ssWW)

Loss value comparison

Best point :
Epoch 105
Save this model
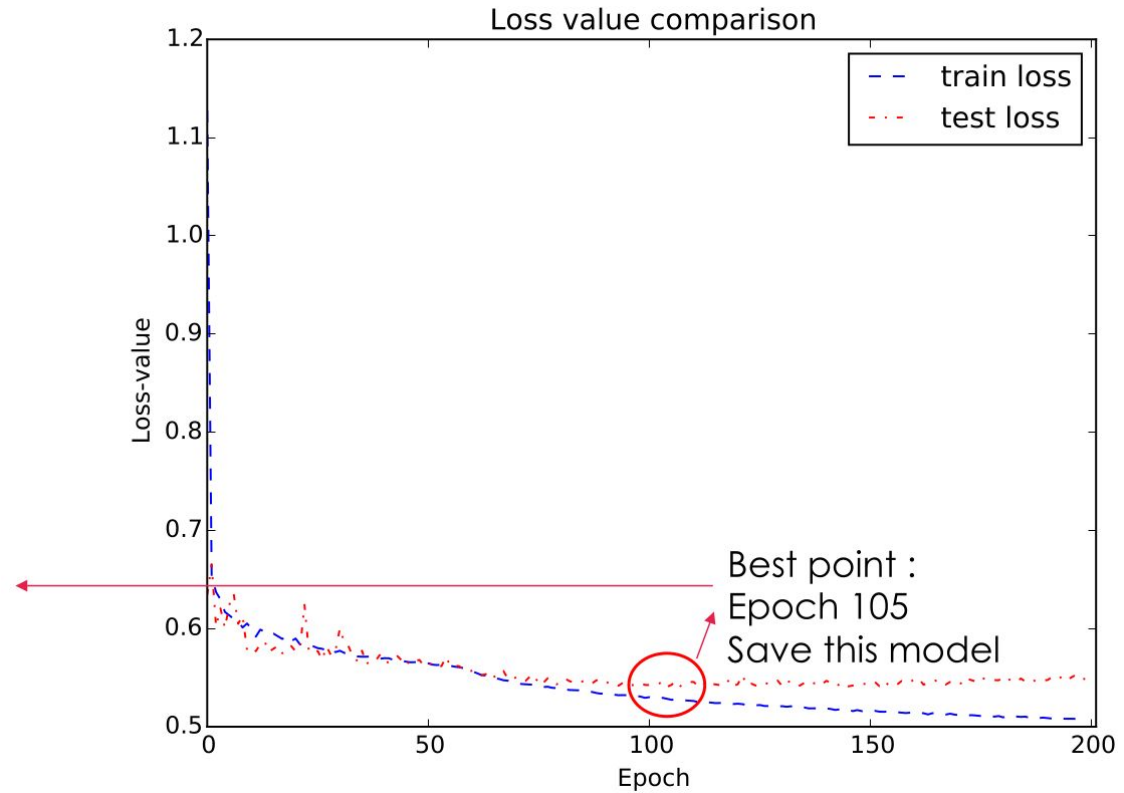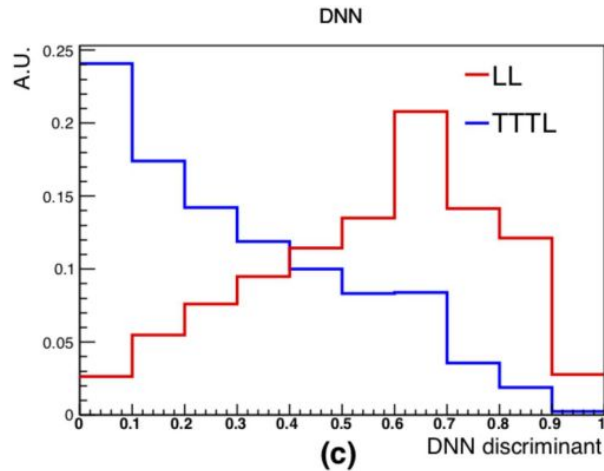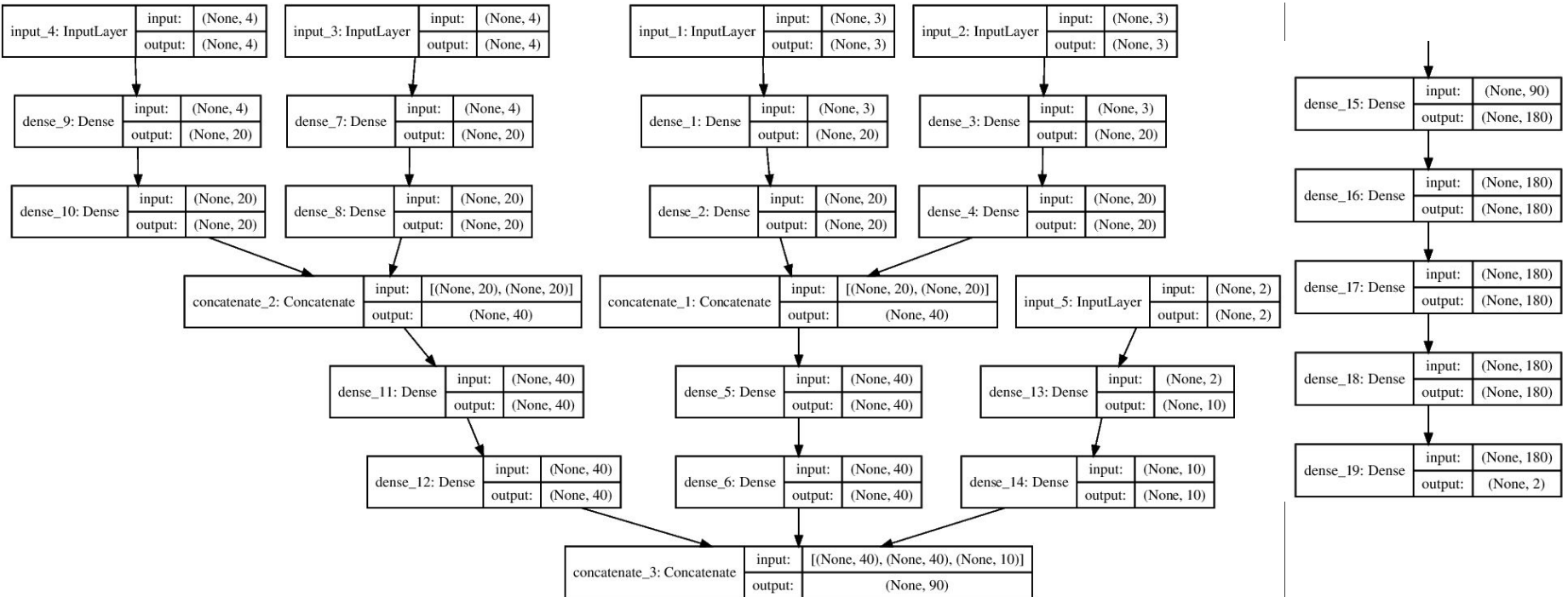
Model for ssWW
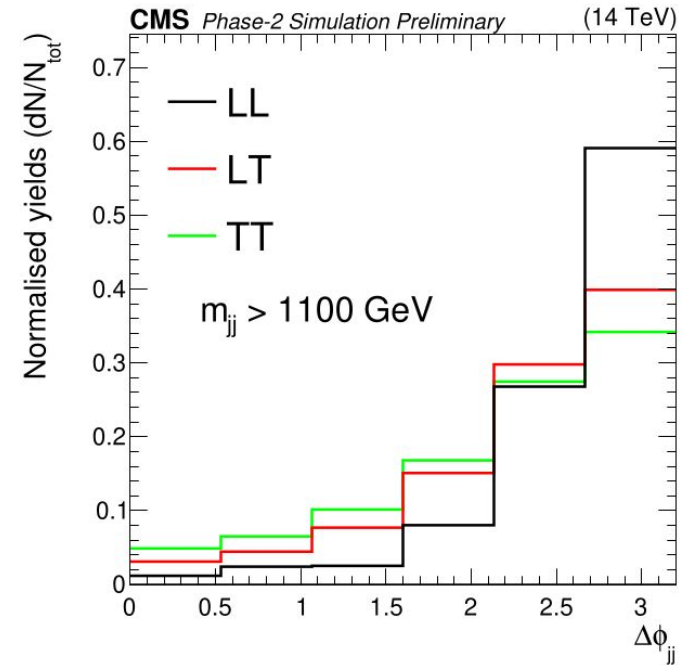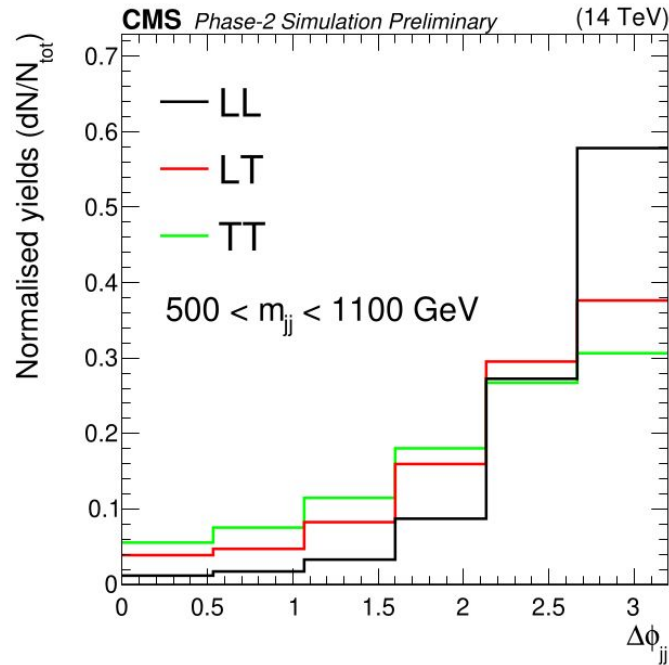
Shape comparison between LL, TT and TL from CMS study

Source of uncertainty	Input	300 fb^{-1} (1 year)	3000 fb^{-1} (10 years)
Statistical uncertainty		5.7%	1.8%
Trigger efficiency (electron)	1.0%	0.5%	0.2%
Trigger efficiency (muon)	1.0%	1.1%	0.6%
Electron id + iso. efficiency	1.0%	0.6%	0.3%
Muon id + iso. efficiency	0.5%	0.9%	0.6%
Jet energy scale	0.5–3.7%	1.0%	0.4%
b tag (stat. component)	1.0%	0.2%	0.3%
b tag misidentification	1–2%	1.4%	1.2%
Misidentified lepton from $t\bar{t}$	5–20%	3.5%	1.0%
Misidentified lepton from $W\gamma$	20%	0.3%	0.1%
Stat. accuracy of $W\gamma$ sample	30%	0.4%	0.1%
Total (stat + experimental syst)		7.6%	3.2%
Luminosity	1.0%	1.0%	1.0%
Theoretical/QCD scale	3.0%	3.0%	3.0%
Total (stat + syst + lumi + theory)		8.2%	4.5%

Uncertainties applied on the CMS LL study

Yeo-Johnson transform

$$\psi(\lambda, y) = \begin{cases} ((y+1)^\lambda - 1)/\lambda & \text{if } \lambda \neq 0, y \geq 0 \\ \log(y+1) & \text{if } \lambda = 0, y \geq 0 \\ -[((-y+1)^{2-\lambda} - 1)]/(2-\lambda) & \text{if } \lambda \neq 2, y < 0 \\ -\log(-y+1) & \text{if } \lambda = 2, y < 0 \end{cases}$$

Principal component
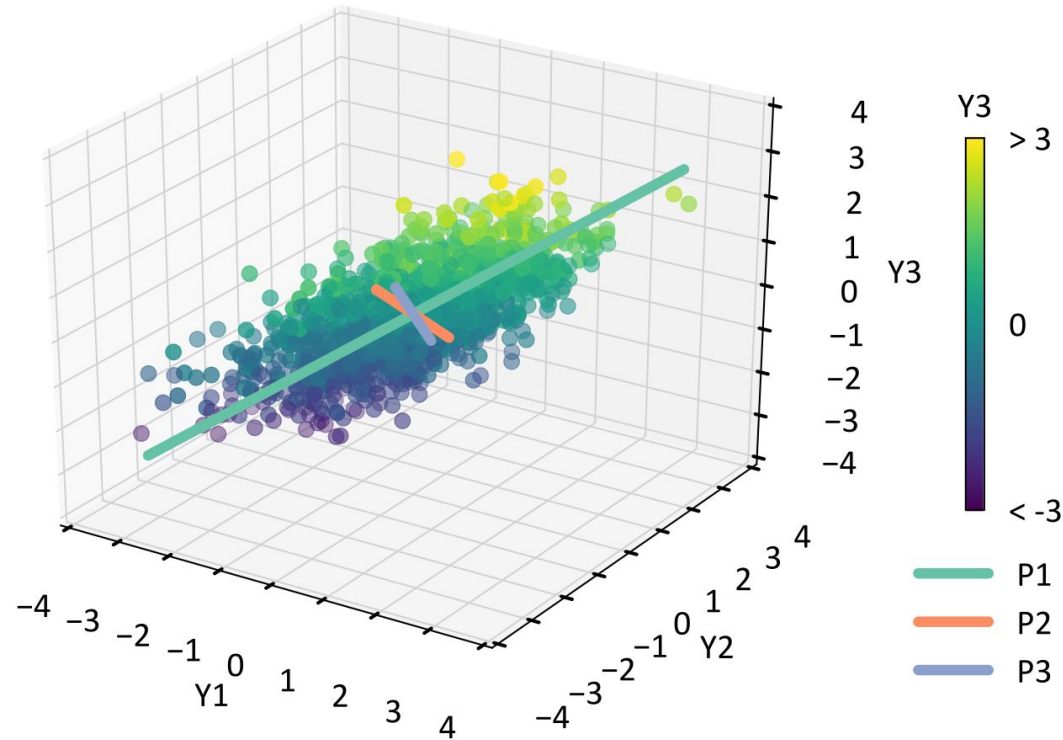
He's initialization

if RELU activation:

$$Y = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$$

$$Var(w_i) = \frac{2}{fan\_in}$$

It draws samples from a truncated normal distribution centered on 0 with `stddev = sqrt(2 / fan_in)` where `fan_in` is the number of input units in the weight tensor.