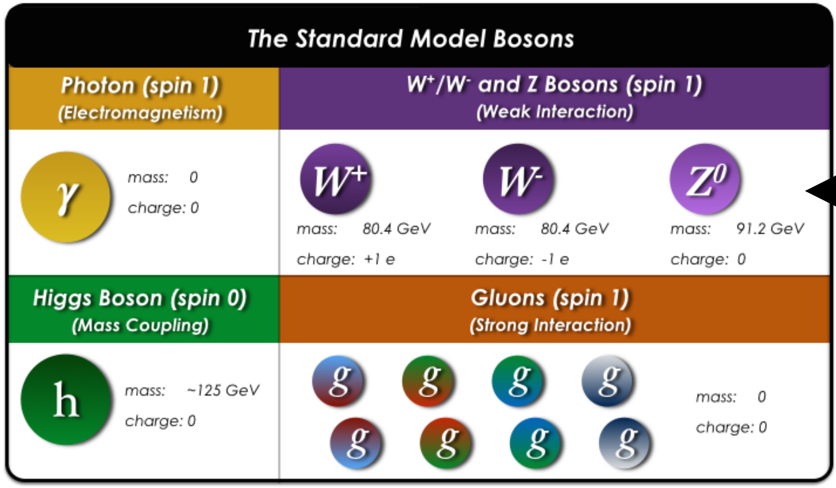
Towards polarisation fraction measurement in Vector Boson Scattering at the HL-LHC with Deep Neural Networks

based on J. Lee, NC, A. Levin, J. Li, M. Lu, Q. Li, Y. Mao,
arxiv:1812.07591 - Phys. Rev. D 99, 033004 (2019),
arxiv:1908.05196 - Phys. Rev. D 100, 116010 (2019)

IN2P3/IRFU ML workshop - 23/01/2020, CCin2p3

Nicolas Chanon - IP2I, CNRS/IN2P3

Vector Boson Scattering and longitudinal polarisation

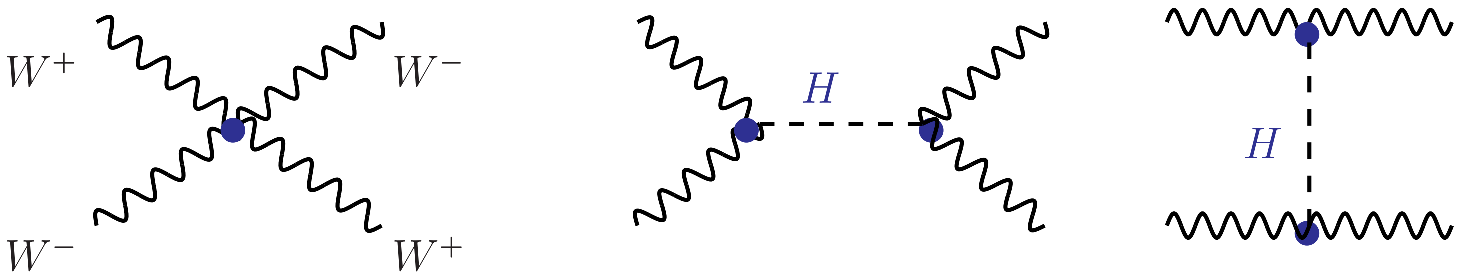


Vector bosons: W and Z
(electro-weak interaction)

At **high energy**, **W-bosons** are becoming more and more **polarised longitudinally**.

The **scattering of longitudinal W-bosons** alone gives a **divergent** cross section:

In the **standard model**, **unitarity is restored with the Higgs boson**:



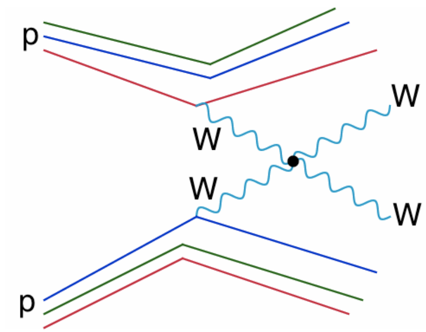
$\sigma \propto E^2$

$\sigma \propto \frac{1}{E^2}$

Vector Boson Scattering process at the LHC

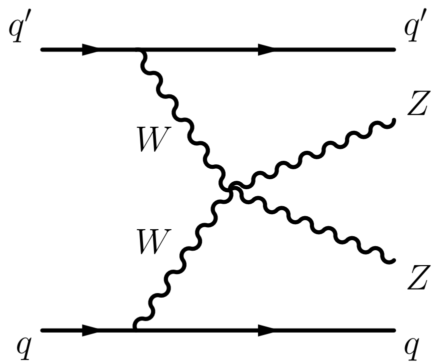
Measuring longitudinal fraction in VBS:

- Improve our understanding of unitarity restoration at high energy
- Verify if the Higgs is the source of unitarity restoration or if there is something else (new particles at a higher energy scale…)



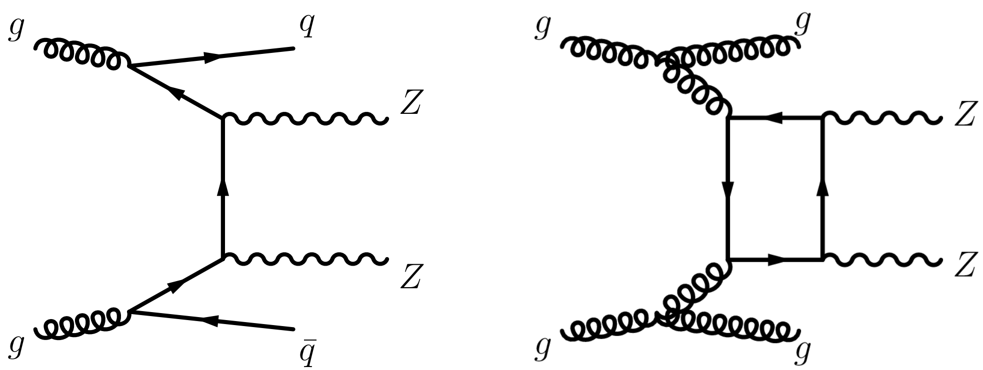
Vector boson scattering at the LHC (proton-proton collisions):

Rare electroweak signal:



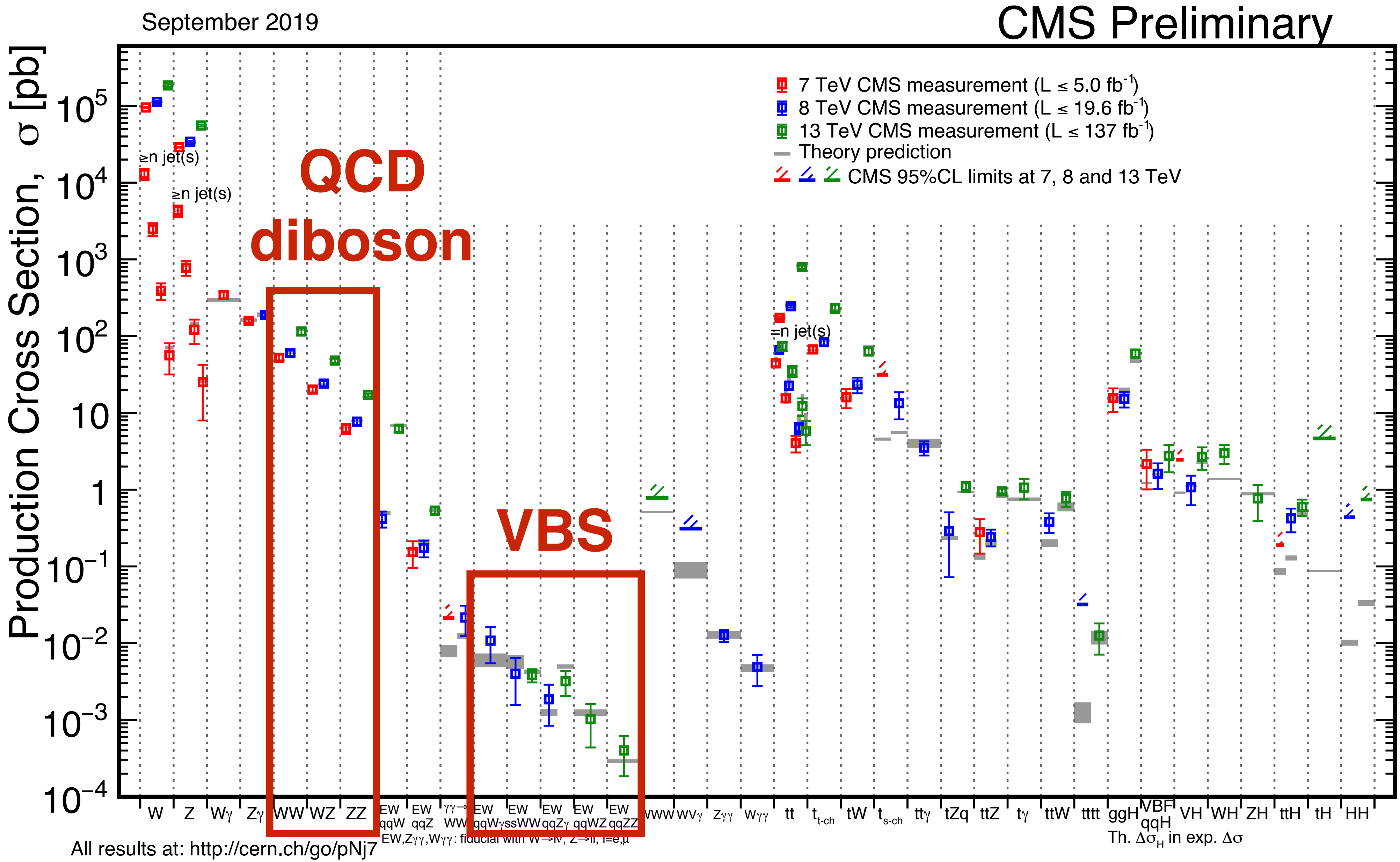
Vector boson scattering (electroweak signal)

Huge QCD (strong force, quantum chromo-dynamics) background:



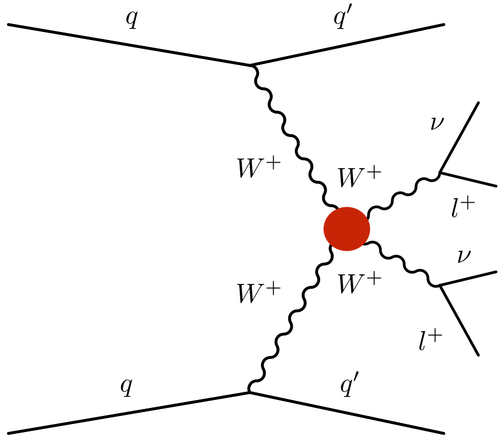
Diboson production (QCD background)

VBS cross sections at the LHC



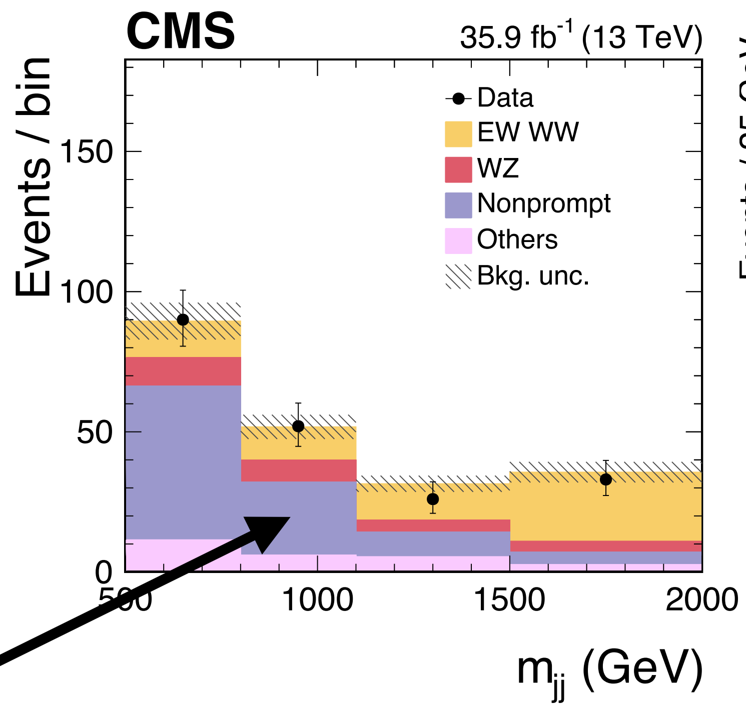
Same sign WW scattering at the LHC

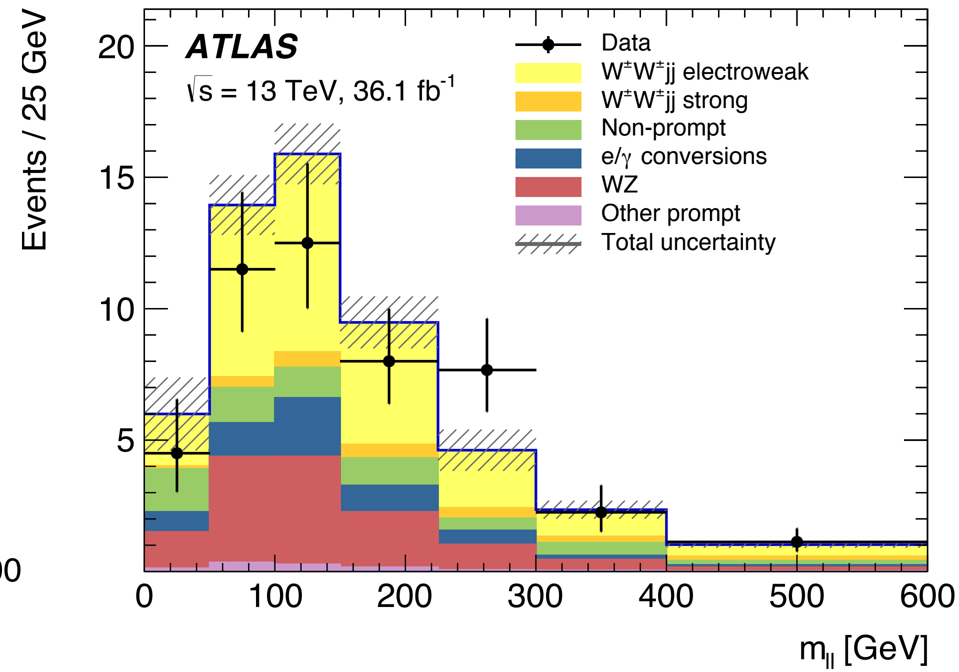
CMS - Phys. Rev. Lett. 120 (2018) 081801, ATLAS - Phys. Rev. Lett. 123 (2019) 161801



Observation of same sign W production at CMS and ATLAS with >5σ (2016 data)

- Outgoing quarks reconstructed as **jets**
- **Leptons** (electron, muon) momenta are precisely measured
- Neutrinos are escaping the detector: **missing transverse energy**

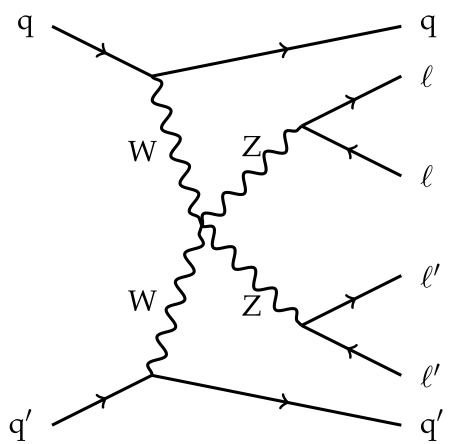




Invariant mass of the two forward jets (mjj) is a good discriminant to measure VBS

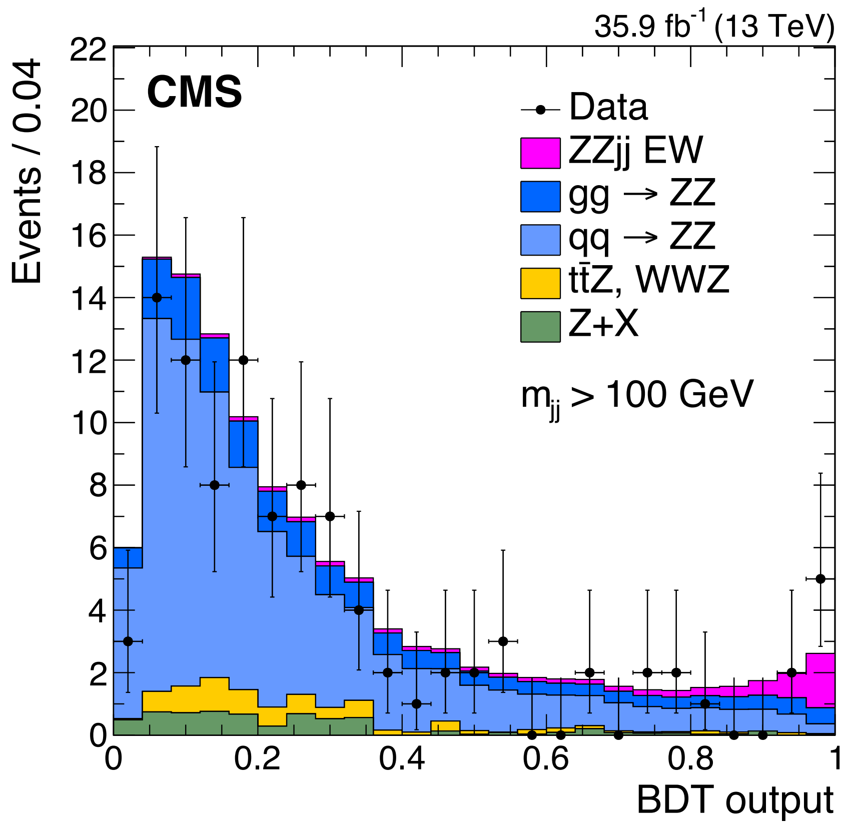
ZZ scattering at the LHC

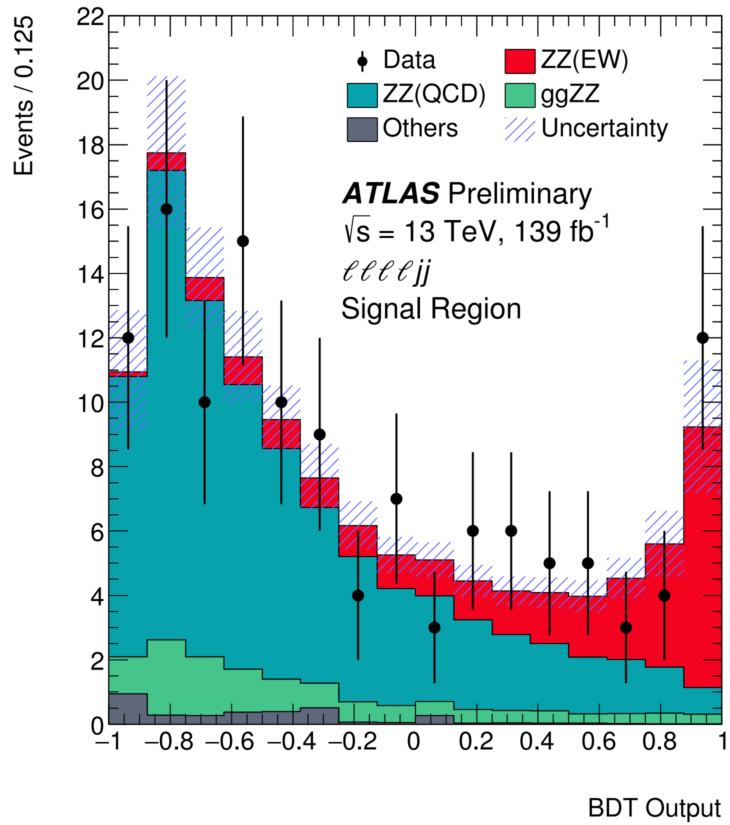
CMS - Phys. Lett. B 774 (2017) 682, ATLAS - ATLAS-CONF-2019-033



CMS: significance 2.7σ (2016 data)
ATLAS: significance >5σ (full Run 2 data)

- Each Z boson reconstructed in their decay into 2 leptons (electron, muon): **4-lepton final state + 2 jets**
- A **BDT** is constructed from high level inputs

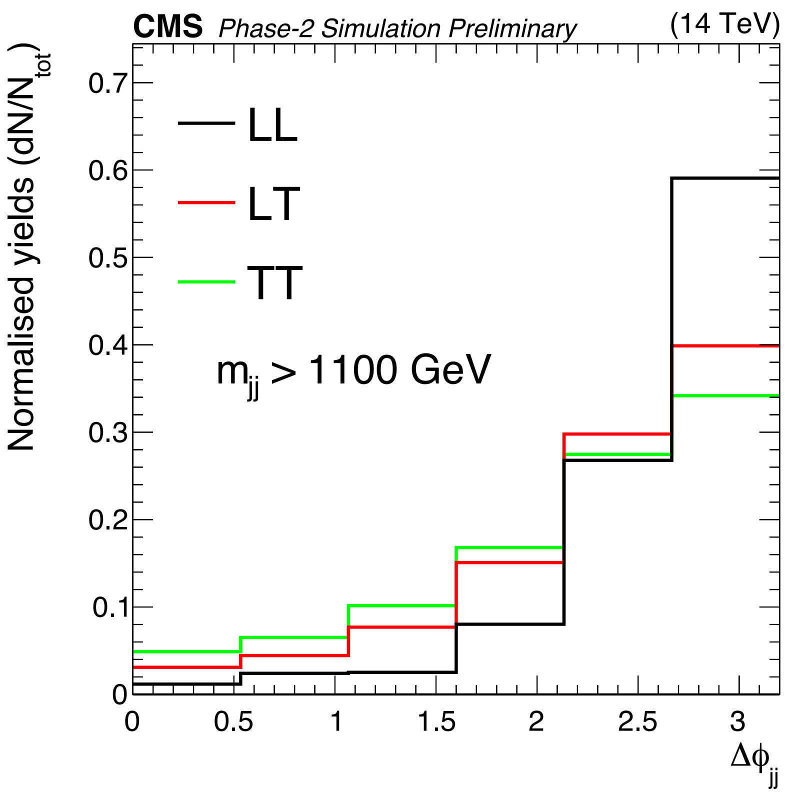


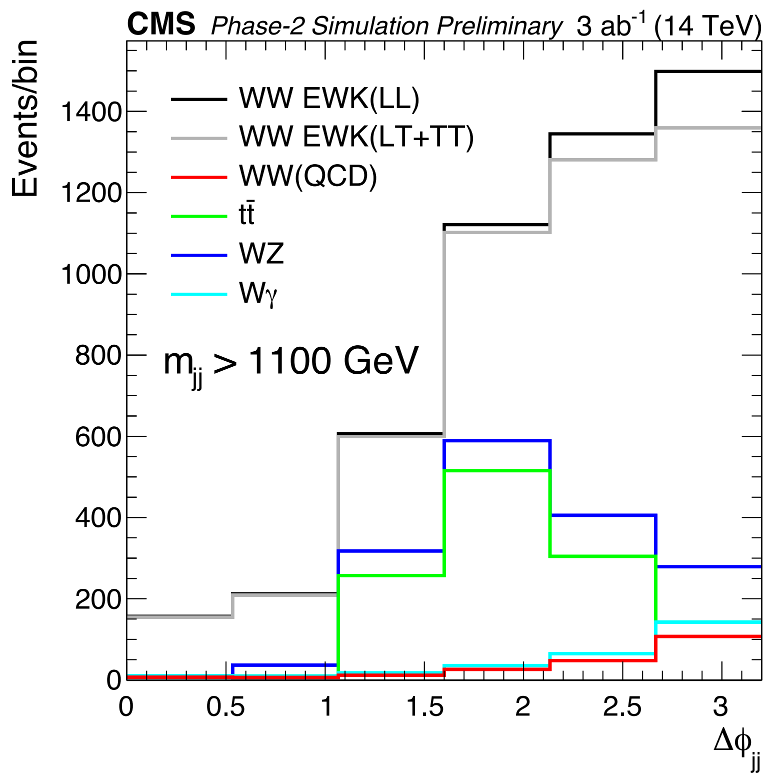


Next step: longitudinal polarisation

CMS-PAS-FTR-18-005

- Longitudinal polarisation can be searched for in angular distributions, in the tails of mjj
- Difficult to disentangle longitudinal (LL) from transverse (LT+TT) polarisation



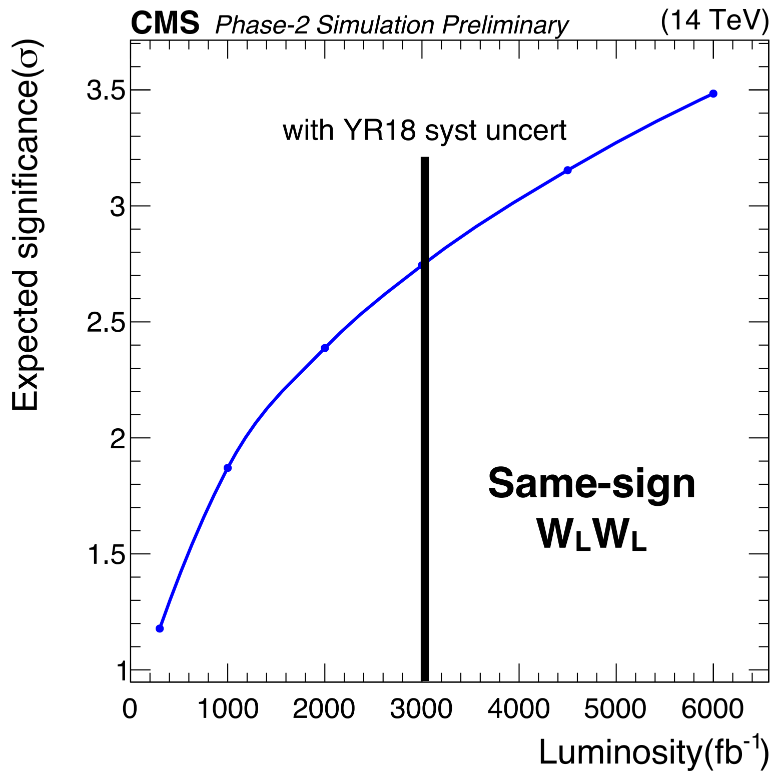


Projections for longitudinal polarisation at HL-LHC

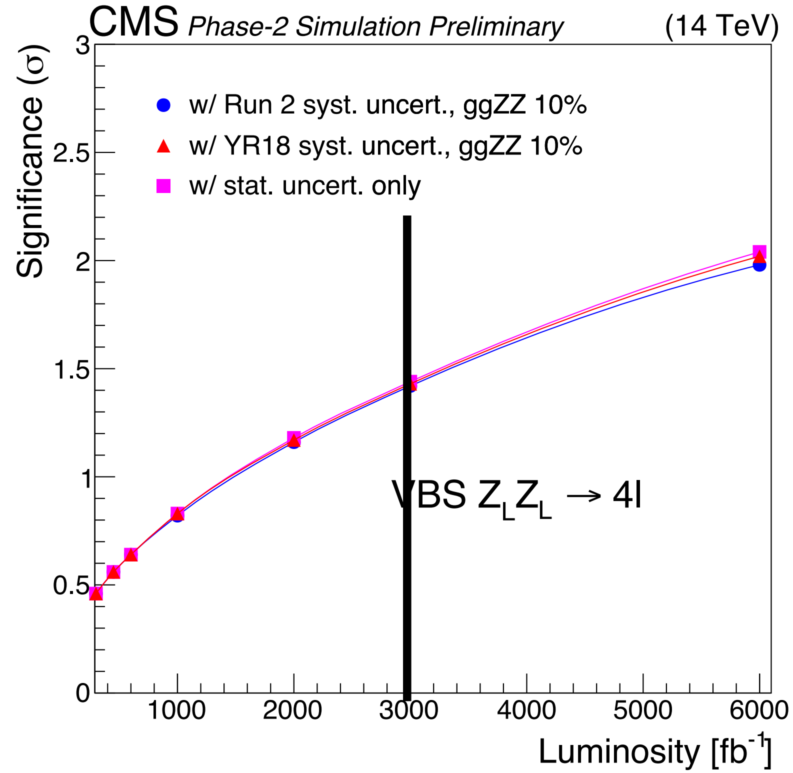
CMS-PAS-FTR-18-005, CMS-PAS-FTR-18-014

Longitudinal scattering in VBS is much more difficult than just VBS

Fit of Mjj distribution in 2 bins of Δϕjj



BDT



Projections assuming “YR18” uncertainty:

- Theoretical uncertainties divided by a factor 2
- Experimental uncertainties assumed to be of statistical nature and scaled with luminosity

WW scattering: analysis setup

Deep Learning Architecture

- Classification with **Deep Neural Network**
- Use **Keras + TensorFlow** back-end

Simulation samples

- Samples generated with Madgraph_aMC@NLO + Pythia + Delphes
- Use Decay package to split the sample in transversal and longitudinal polarisation fractions
- Assume HL-LHC dataset, Luminosity of 3000 fb-1
- No pileup

Low level features

- For both leptons: pT, η, ϕ
- For both jets: pT, η, ϕ, mass
- Transverse missing energy magnitude and ϕ

Tried also high-level features:

- Mass of the two jets: mjj
- etc.

“Particle-based” Deep Neural Network

Very simple idea: regroup features by particles, progressively merge

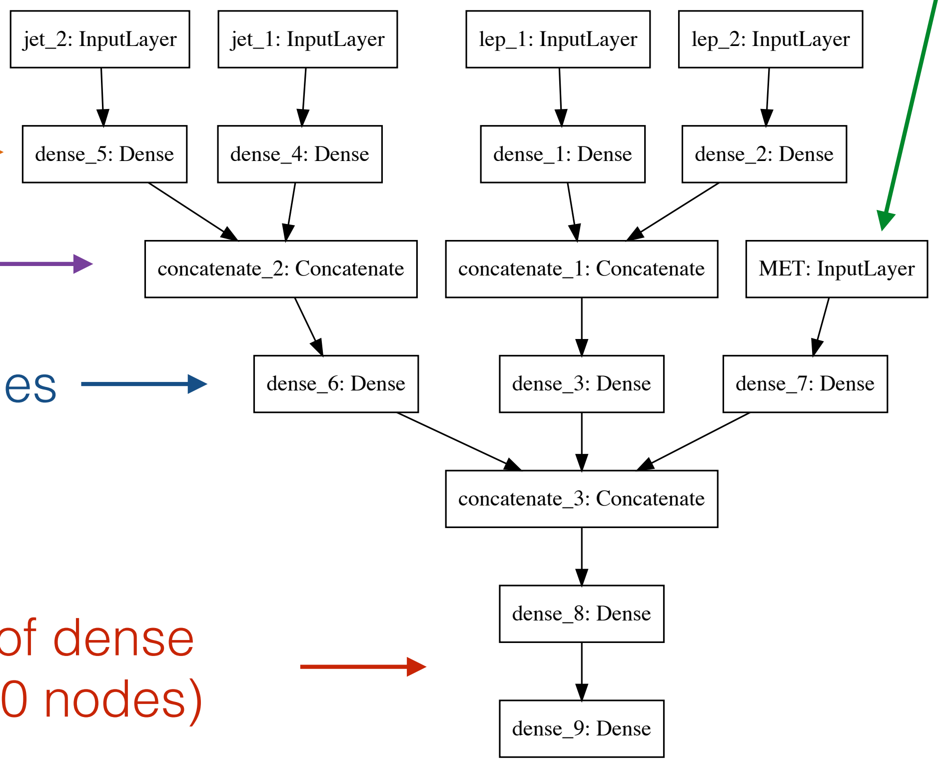
Low-level features for missing transverse energy

Start from individual particle momenta (low-level features)

2 hidden layers of 20 nodes each

Concatenate dense layers built upon low-level features

dense layers of 40 nodes



Last step is a series of dense layers (4 layers of 180 nodes)

“Particle-based” DNN in WW scattering

Dense:

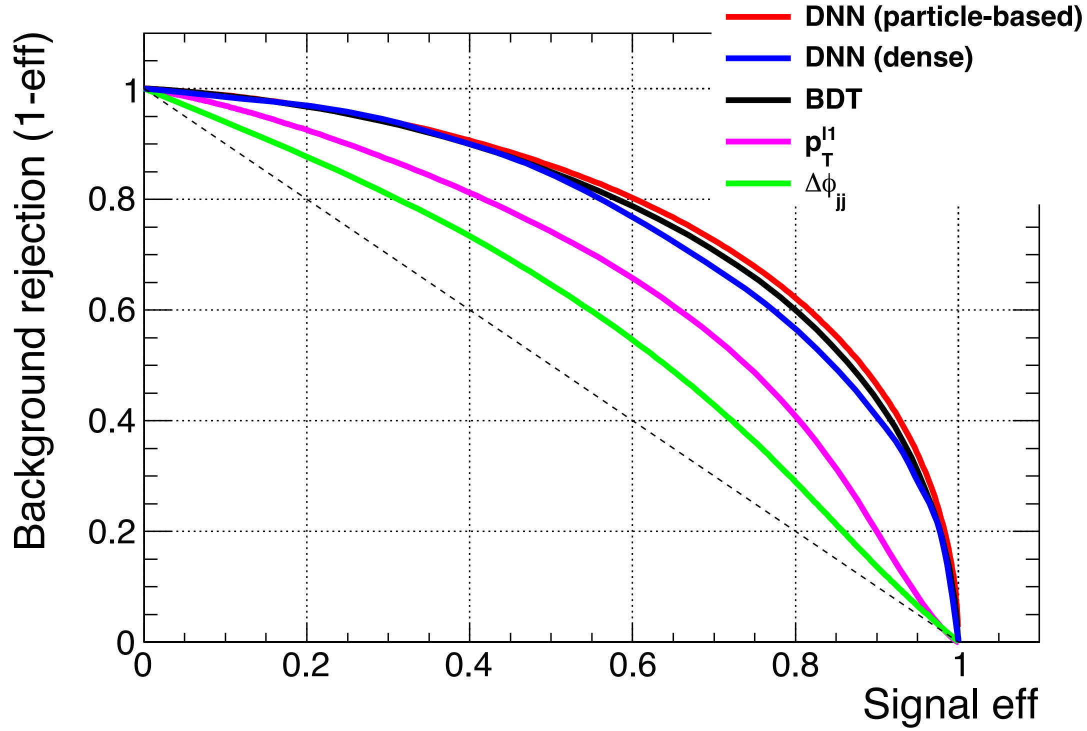
- 10 layers of 150 neurones densely connected
- Relu activation function, Adam optimizer, L2 regularisation
- Overtraining prevented with a batch size of 50, drop-out rate 50% on hidden units, batch normalization

BDT: 1000 trees of depth 5, Adaptative Boost algorithm

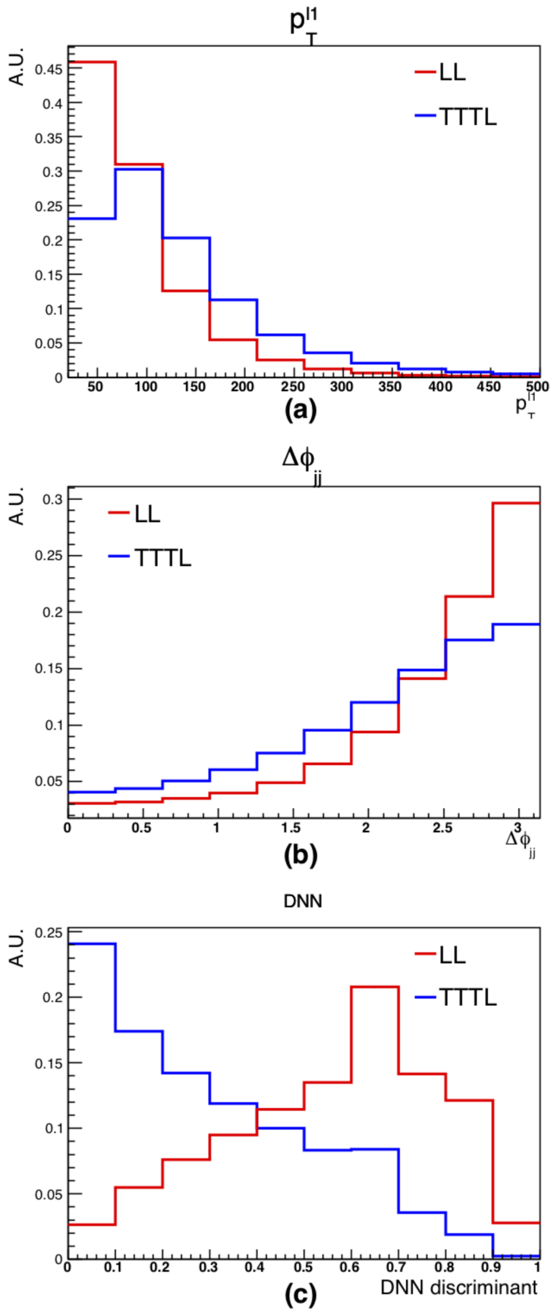
pTl and Δφjj: rectangular cut (distributions next slide)

Classification: LL vs TT+TL
Neglect other backgrounds at high mjj
(they are known to be very small)

Dataset size (Mjj>850 GeV):
4M events inclusive for TT, TL, LL
Among which 6.6% of LL (265k)



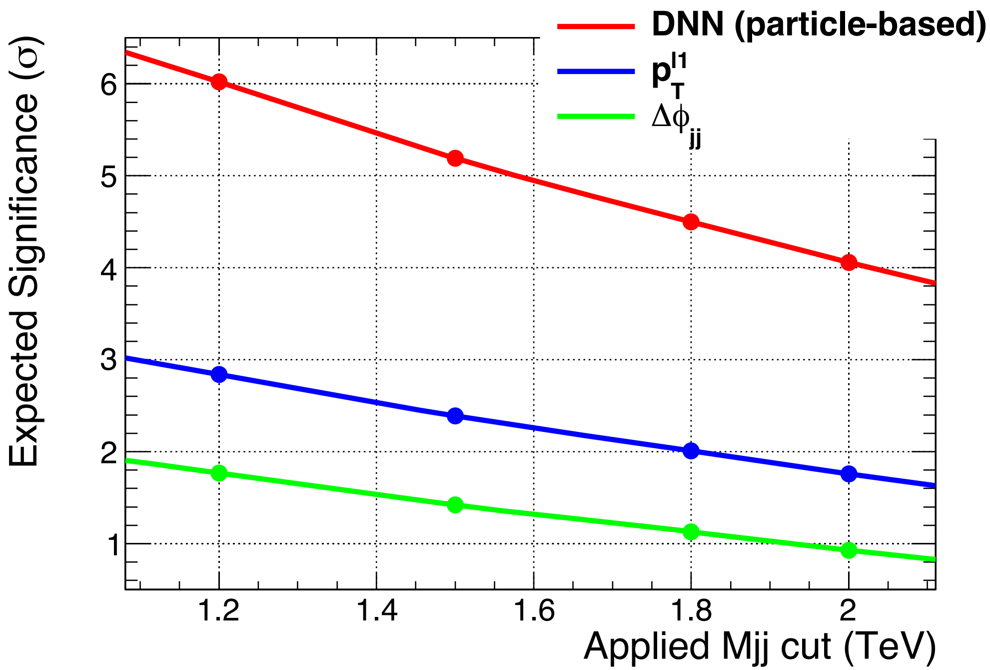
WW scattering: results



Systematic uncertainties:

- Luminosity: 2%
- 5% on LL and TT+TL

Reach 4σ in the worst case



"Particle-based" DNN in ZZ scattering

Dense DNN:

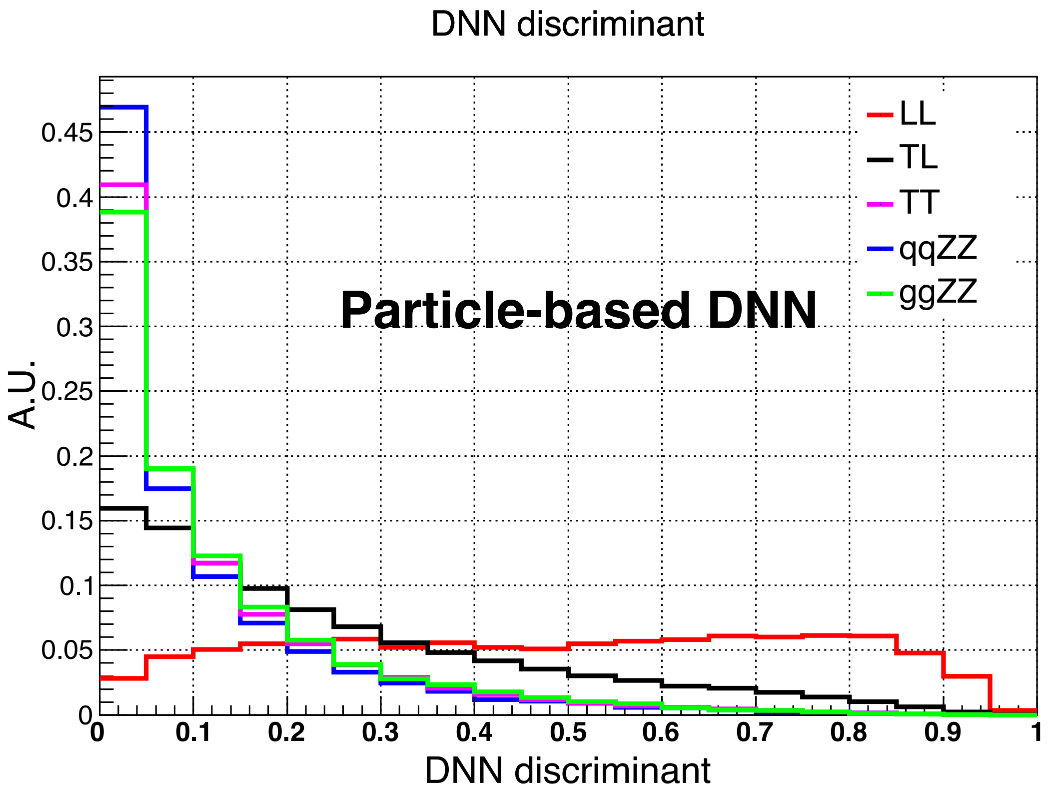
- 10 layers of 150 neurones densely connected

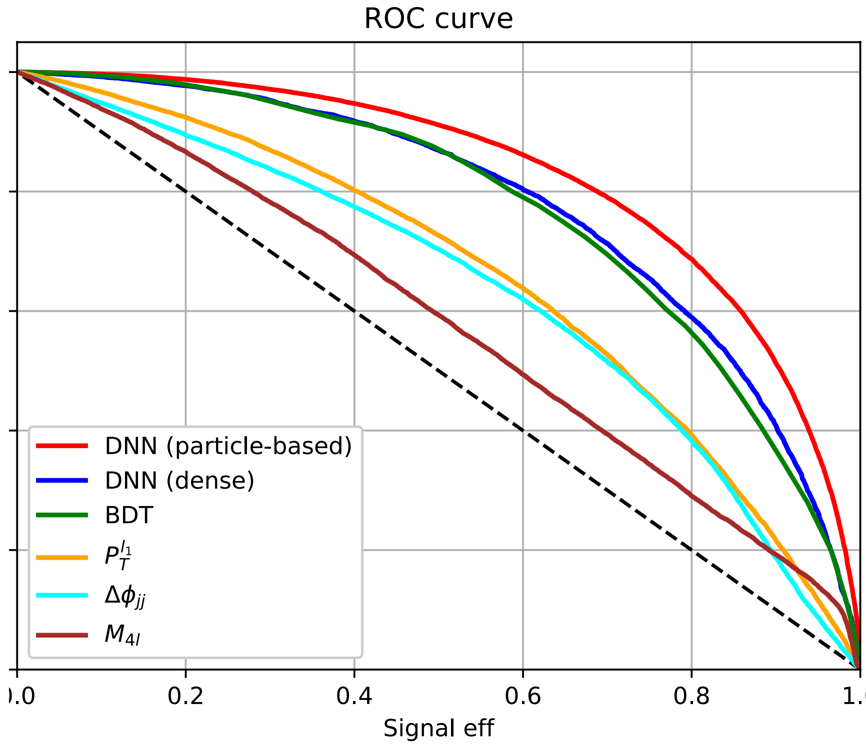
Particle-based DNN:

- Multiclass output: 5 classes, one signal class and four background classes

Dataset size:

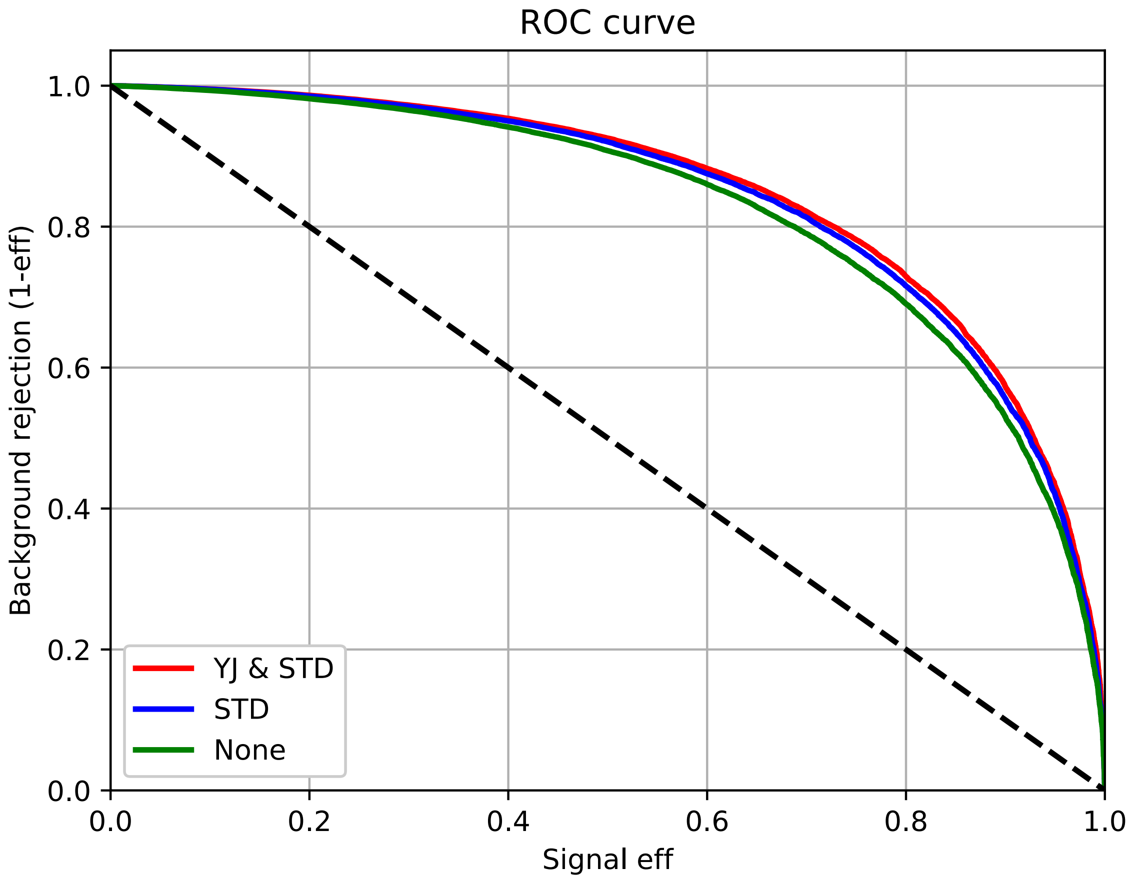
- LL: 100k
- TT+TL: 390k
- qqZZ: 48k
- ggZZ: 40k





ZZ scattering: Pre-processing

- **STD:** Standardization scales data to build features having a mean of 0 and standard deviation of 1 in the signal
- **YJ:** Yeo-Johnson transformation is a power transformation valid for negative input values, transforming features in the signal to a normal distribution
=> Using **STD+YJ** improves the performance.



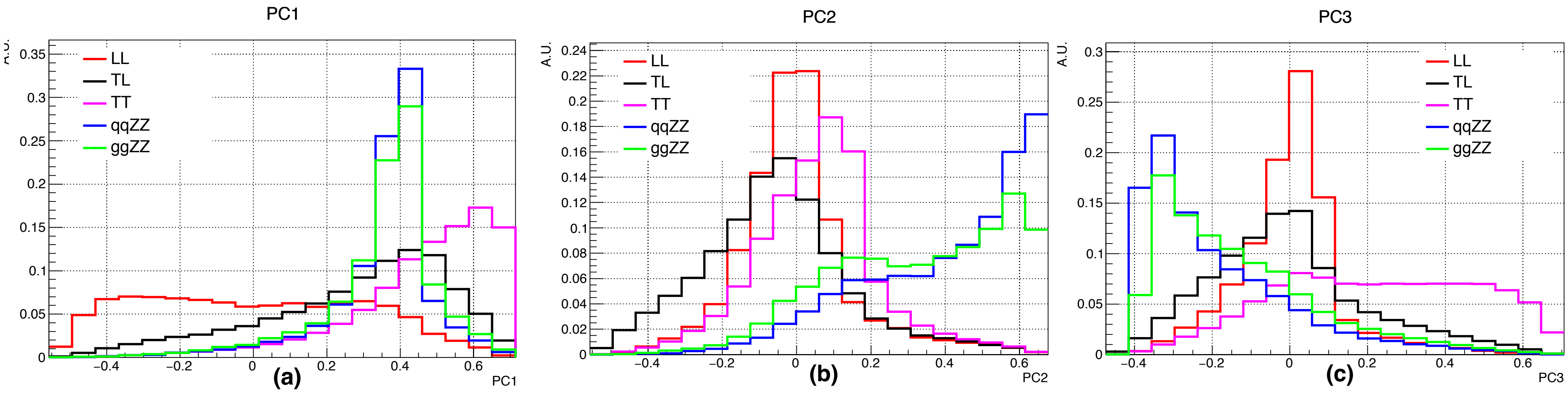
ZZ scattering: Post-processing

Principle Component Analysis

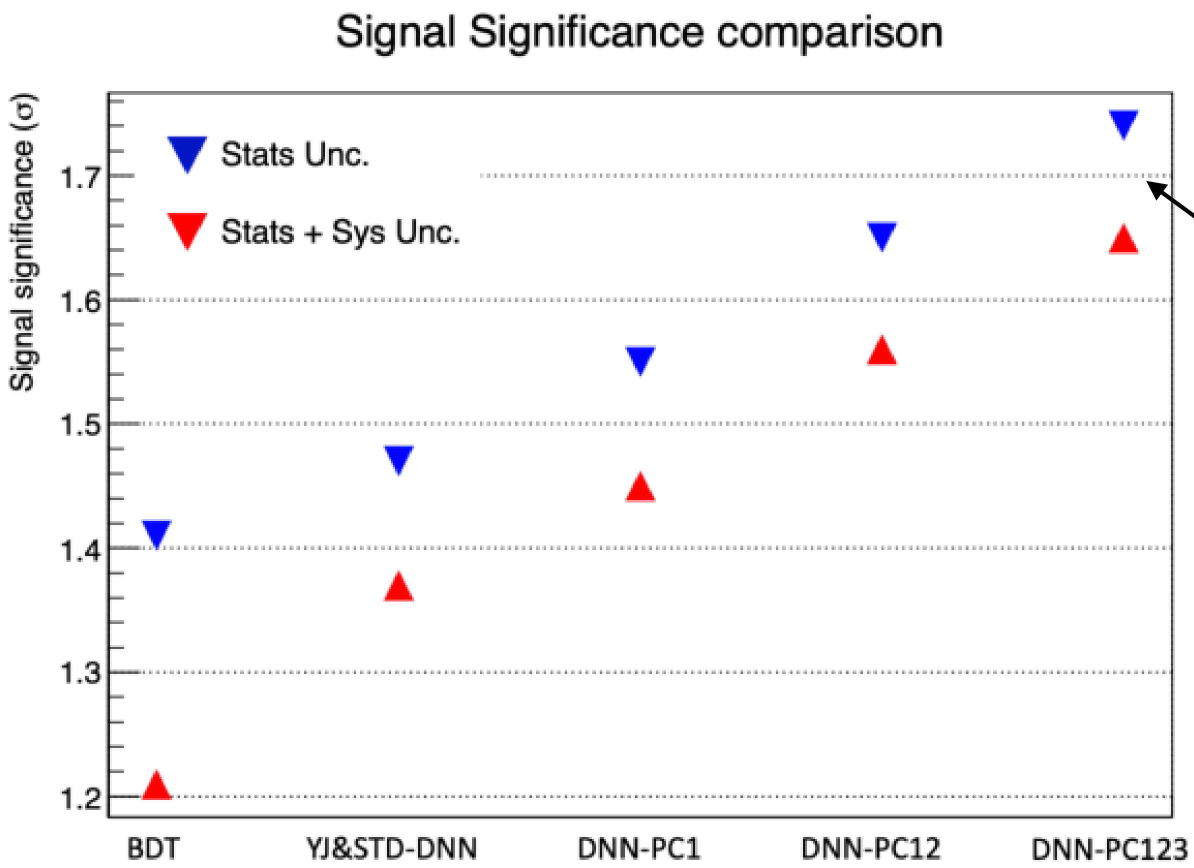
- Applied on the particle-based DNN outputs (5 classes)
- Rotates the axis of features to a new axis with decorrelated features
- The first principle component (PC1) has the highest variance, followed by the second (PC2), etc.

TABLE II. Explained variance ratio of each principle component. The first, second, and third leading PCs' explained variance ratio exceed 10%.

Principle component	PC1	PC2	PC3	PC4	PC5
Explained variance ratio	64.8%	18.1%	13.0%	4.2%	<0.1%



ZZ scattering results



particle-based DNN + PCA: 40% enhancement in significance relative to BDT

Simultaneous fit of the PCA outputs

Conclusion and perspectives

Conclusions

- With pre-processing and post-processing, **“particle-based” DNN** architecture can **improve by a large amount the performance** (over dense DNN or BDT)
- Remains to be tested on real data
- Improving over traditional ML methods will be necessary to observe very rare signals at the HL-LHC

Back-up slides

What is longitudinal polarisation

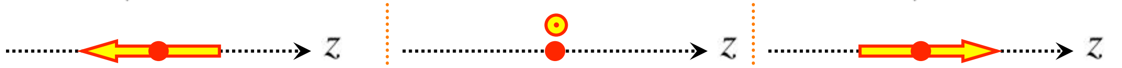
- Within the Higgs mechanism, the Goldstone bosons arising from electroweak symmetry breaking are absorbed by the W and Z bosons as longitudinal degrees of freedom and provide their mass.
- At high energy, W-bosons behave like a goldstone boson, i.e. purely longitudinal. However this leads to a divergence in the cross section.
- The Higgs boson is restoring unitarity in VBS at high energy:

Boson wave-functions are written in terms of the polarization four-vector ε^{μ}

$$\boxed{B^{\mu} = \varepsilon^{\mu} e^{-ip.x} = \varepsilon^{\mu} e^{i(\vec{p}.\vec{x}-Et)}}$$

For a spin-1 boson travelling along the z-axis, the polarization four vectors are:

$$\varepsilon_{-}^{\mu} = \frac{1}{\sqrt{2}}(0,1,-i,0); \quad \varepsilon_{L} = \frac{1}{m}(p_z,0,0,E) \quad \varepsilon_{+}^{\mu} = -\frac{1}{\sqrt{2}}(0,1,i,0)$$



$S_z = -1$	$S_z = 0$	$S_z = +1$
transverse	**longitudinal**	**transverse**

Selection

WW scattering: selection

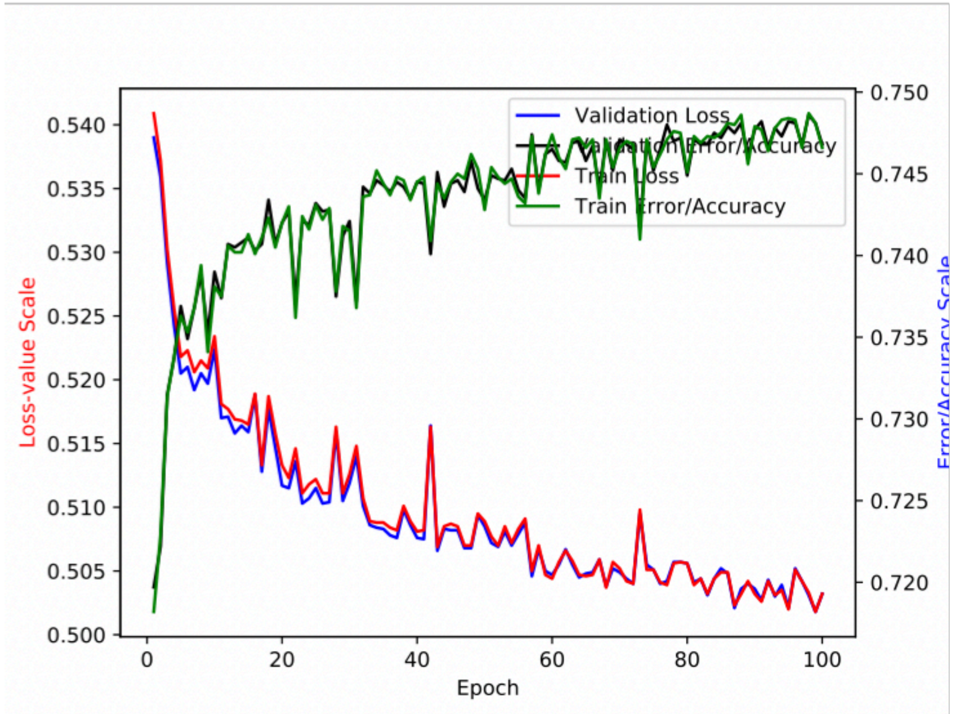
- Require 2 leptons with pT > 20 GeV and $|\eta|<2.4$
- Two jets with pT>50 GeV and $|\eta|<4.7$
- Rapidity gap $|\eta jj|>2.5$
- Try several cuts on mjj

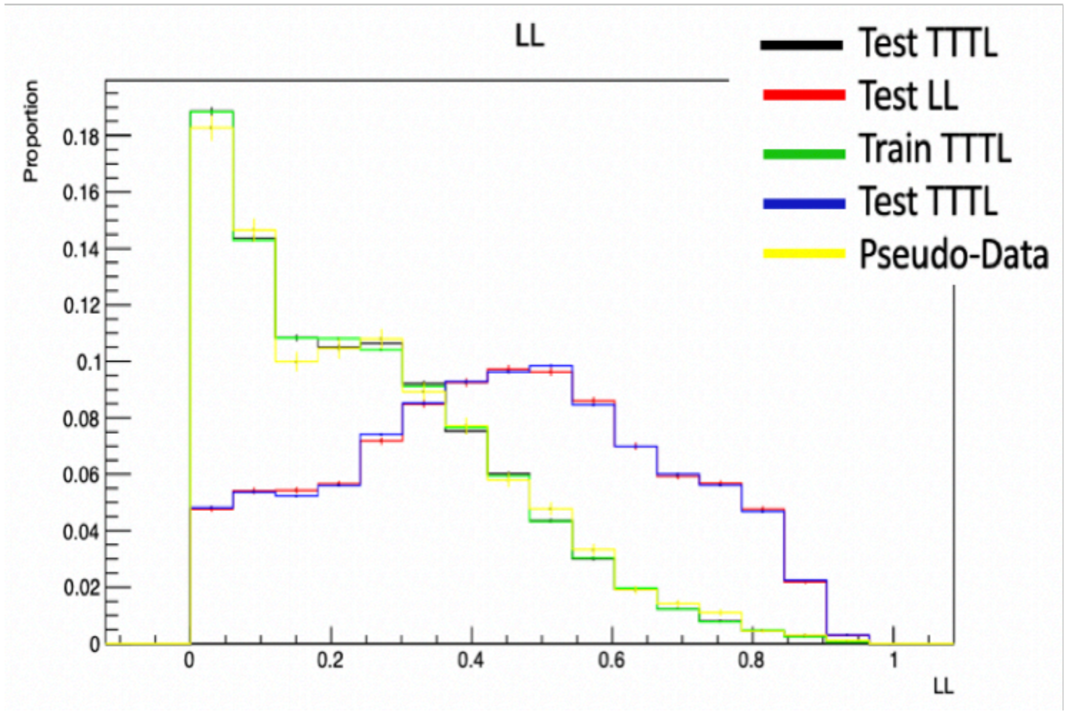
ZZ scattering: selection

- Require 4 leptons with pT > 20,10,5,5 GeV and $|\eta|<2.4$
- Dilepton mass 60 < mll < 120 GeV
- Two jets with pT>50 GeV and $|\eta|<4.7$
- Rapidity gap $|\eta jj|>2.4$
- Dijet mass mjj > 400 GeV

Dense DNN Training

Junho Lee PhD Thesis

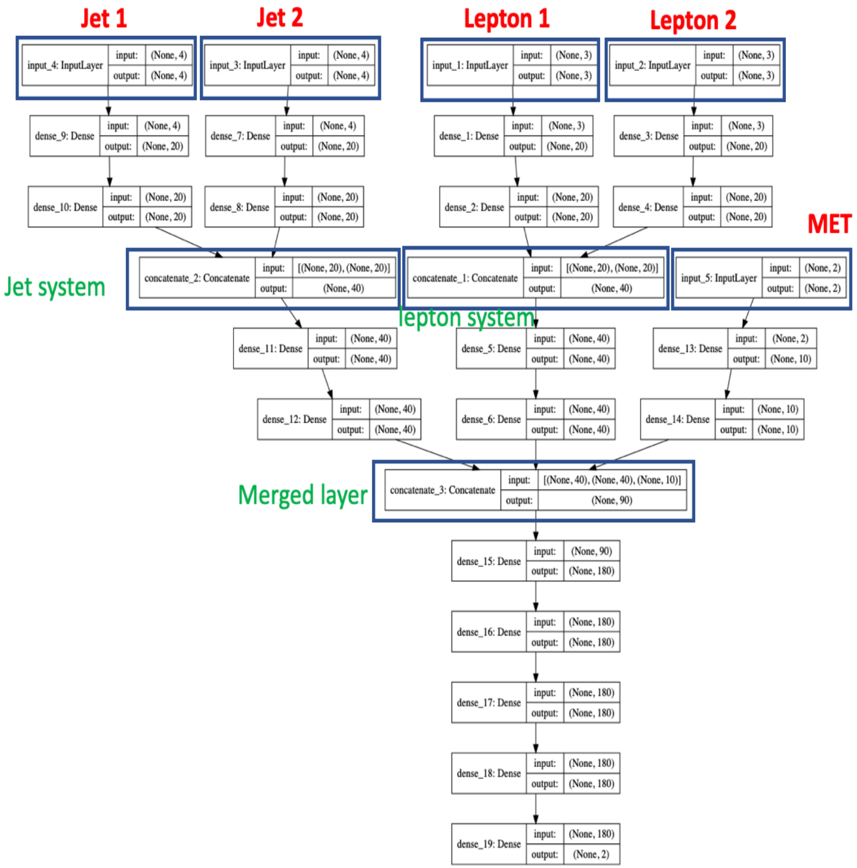




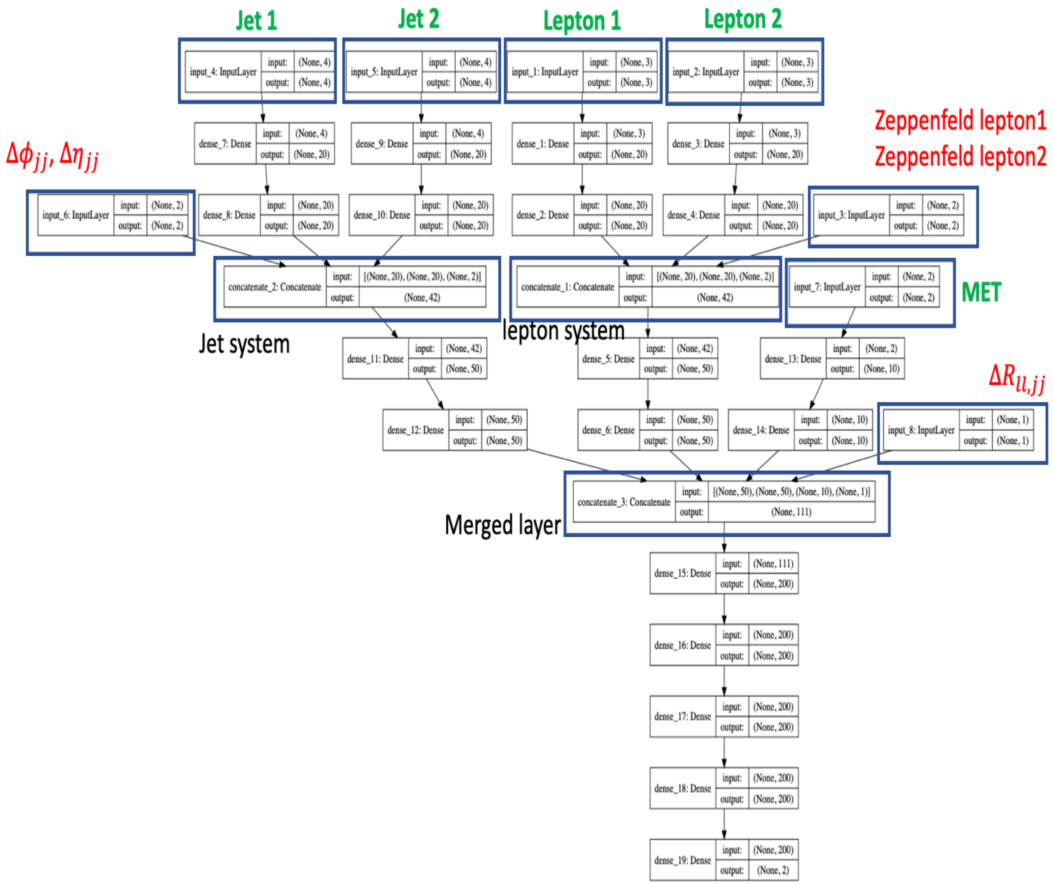
lo vs lo+hi level features

Junho Lee PhD Thesis

[1] Particle-based DNN

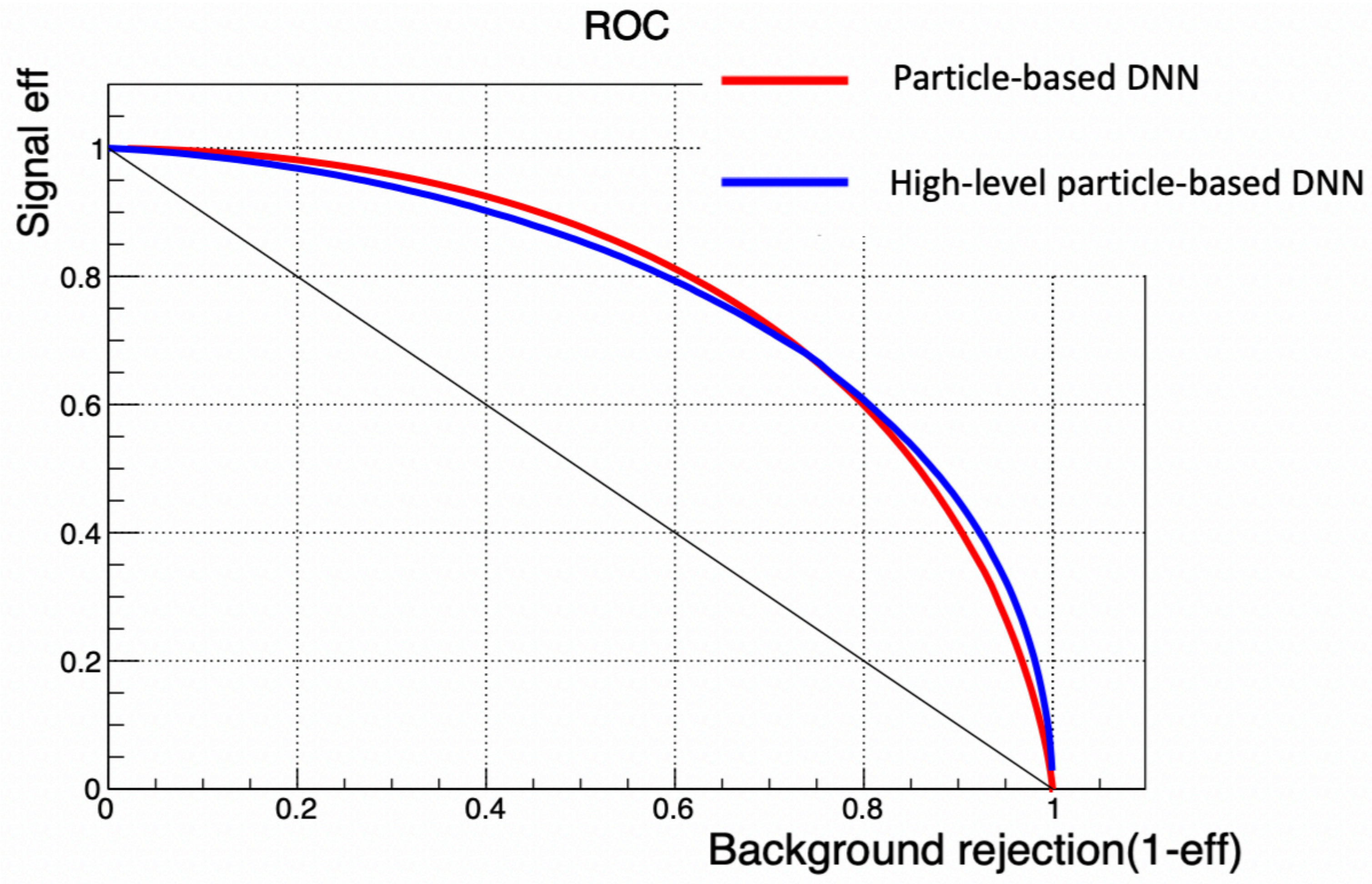


[2] High-level particle-based DNN



lo vs lo+hi level features

Junho Lee PhD Thesis



"Particle-based" DNN in ZZ scattering

Junho Lee PhD Thesis