UNIVERSITY OF NOTRE DAME

A $W^{\pm}$ polarization analyzer from Deep Neural Networks

Taegyun Kim
Research collaborator: Adam Martin
Department of Physics, University of Notre Dame

arXiv:2102.05124

ML4Jets 2021

Theoretical Motivation

Massive vector boson final states

$$p\,p \rightarrow W^{\pm}\,W^{\mp}$$

$$p\,p \rightarrow W^{\pm}\,Z$$

$$p\,p \rightarrow Z\,Z$$

- Indirect approach of checking SM : polarization searches
  - Longitudinal vs. Transverse
- SM can predict polarization fraction
- Longitudinal polarization is sensitive to EWSB
- New physics can impact polarizations fractions. Ex.) new resonance, or higher dimensional operators (SMEFT)

W Polarization Measurement

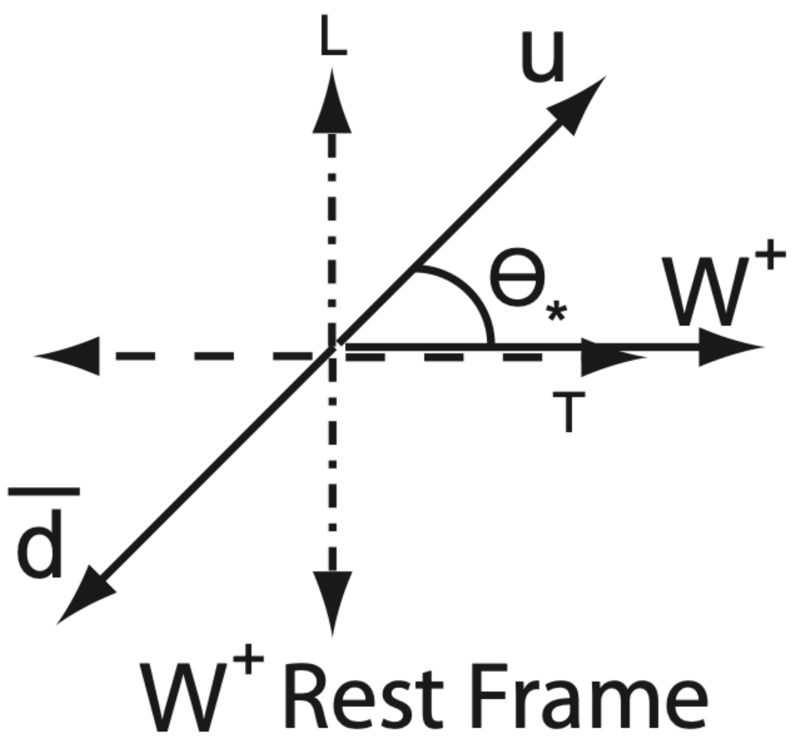
- Leptonic tagging is done
  - lower branching ratio than hadronic state
  - neutrino reconstruction
- Can we do hadronic *W* tagging?
  - Extract correct boson jet
  - QCD effect washes out parton level information
  - Possible jet substructure information can be used to tag polarization
    - N-subjettiness (S. De, V. Rentala, W. Shepherd arXiv:2008.04318v1)
  - **How does machine learning do?**

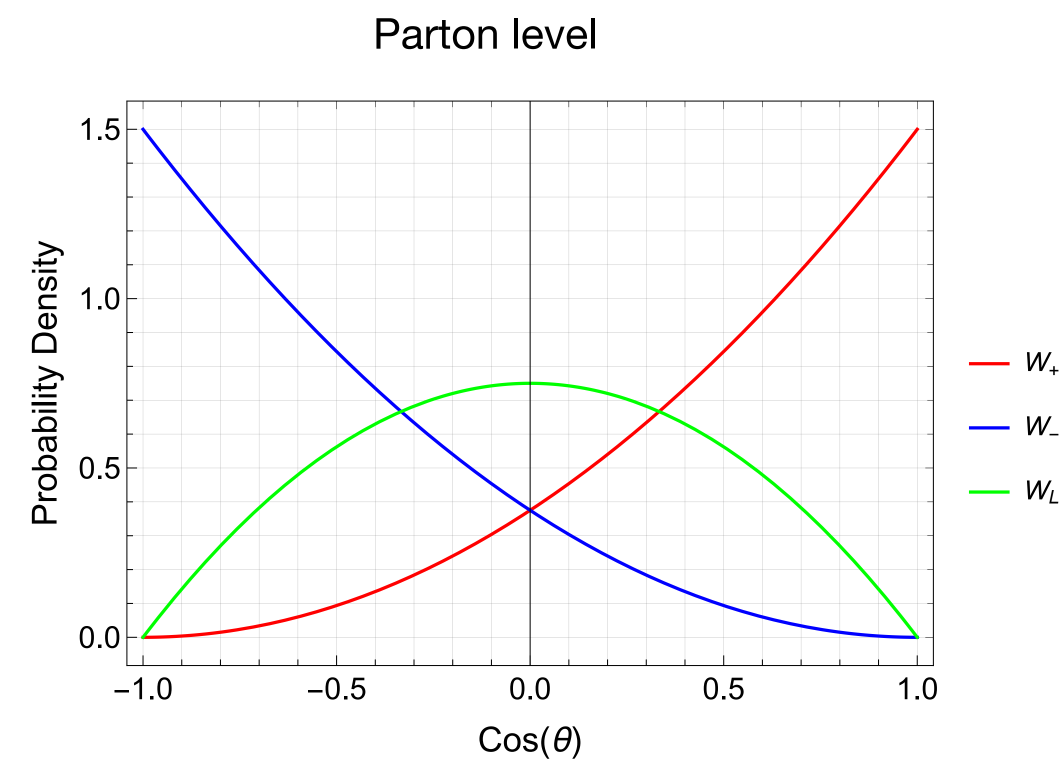
W Polarization Measurement

- Leptonic tagging is done
  - lower branching ratio than hadronic state
  - neutrino reconstruction
- Can we do hadronic *W* tagging?
  - Extract correct boson jet
  - QCD effect washes out parton level information
  - Possible jet substructure information can be used to tag polarization
    - N-subjettiness (S. De, V. Rentala, W. Shepherd arXiv:2008.04318v1)
  - **How does machine learning do?**

W polarization

Decay of W





- Since W only interacts to the left handed particles, each polarization has distinct angular distribution (or, in lab frame, $(E_q - E_{\bar{q}})/|\vec{p}_W|$)
- Parton level distribution can be used as a reference point for the network optimization
- Due to the deviation, it is possible to measure polarization fraction for diboson final states

Higher dimensional operators and W/Z polarization

- SMEFT extends the SM Lagrangian by gauge invariant higher dim (D>4) operators
- $$\mathscr{L}_{SMEFT} = \mathscr{L}_{SM} + \sum_{D>4}^{\inf} \frac{1}{\Lambda^{D-4}} c_j^{(D)} \mathcal{O}_j^{(D)}$$
- $W^{\pm}/Z$ can appear in two ways:

    1. $W^a_{\mu\nu} \rightarrow$ Primarily transverse as $\epsilon^{\mu}_{L} W^a_{\mu\nu} \sim \frac{k^{\mu}}{m_W} W^a_{\mu\nu} \approx 0$
    2. $D_{\mu}H \rightarrow$ Primarily longitudinal as Goldstone gets eaten

Relevant operators (SILH) for diboson final states

Da Liu, Lian-Tao Wang [arXiv: 1804.08688v1]

$$\mathcal{O}_W = \frac{ig}{2}\left(H^{\dagger}\sigma^a \overleftrightarrow{D}^{\mu} H\right) D^{\nu} W^a_{\mu\nu}$$

$$\mathcal{O}_B = \left(H^{\dagger}\sigma^a \overleftrightarrow{D}^{\mu} H\right) \partial^{\nu} B_{\mu\nu}$$

$$\mathcal{O}_{2W} = -\frac{1}{2} D^{\mu} W^a_{\mu\nu} D_{\rho} W^{a\rho\nu}$$

$$\mathcal{O}_{3W} = \frac{1}{3!} g \epsilon_{abc} W^{a\nu}_{\mu} W^{b}_{\nu\rho} W^{c\rho\mu}$$

We will focus on the boxed operators

$$\mathcal{O}_{HW} = ig\,(D^{\mu}H)^{\dagger}\,\sigma^a\left(D^{\nu}H\right) W^a_{\mu\nu}$$

$$\mathcal{O}_{HW} = ig'\,(D^{\mu}H)^{\dagger}\left(D^{\nu}H\right) B_{\mu\nu}$$

Higher dimensional operators and W/Z polarization

$$\mathcal{O}_W = \frac{ig}{2}\left(H^\dagger \sigma^a \overleftrightarrow{D}^\mu H\right) D^\nu W^a_{\mu\nu}$$

$$\mathcal{O}_{3W} = \frac{1}{3!} g \epsilon_{abc} W^{a\nu}_\mu W^b_{\nu\rho} W^{c\rho\mu}$$

- Operators above contribute to the cross section of $pp \to W^\pm Z$ but different impact on polarization breakdown
- Knowing polarization can better distinguish between effects
- For our analysis, we do not introduce full SMEFT operators for simplification of parameters
- Benchmark SMEFT Lagrangian : $\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \frac{2\,c_W}{{m_W}^2}\mathcal{O}_W + \frac{3!\,c_{3W}\,g^2}{{m_W}^2}\mathcal{O}_{3W}$

Convention from A. Alloul, B. Fuks, and V. Sanz, arXiv:1310.5150

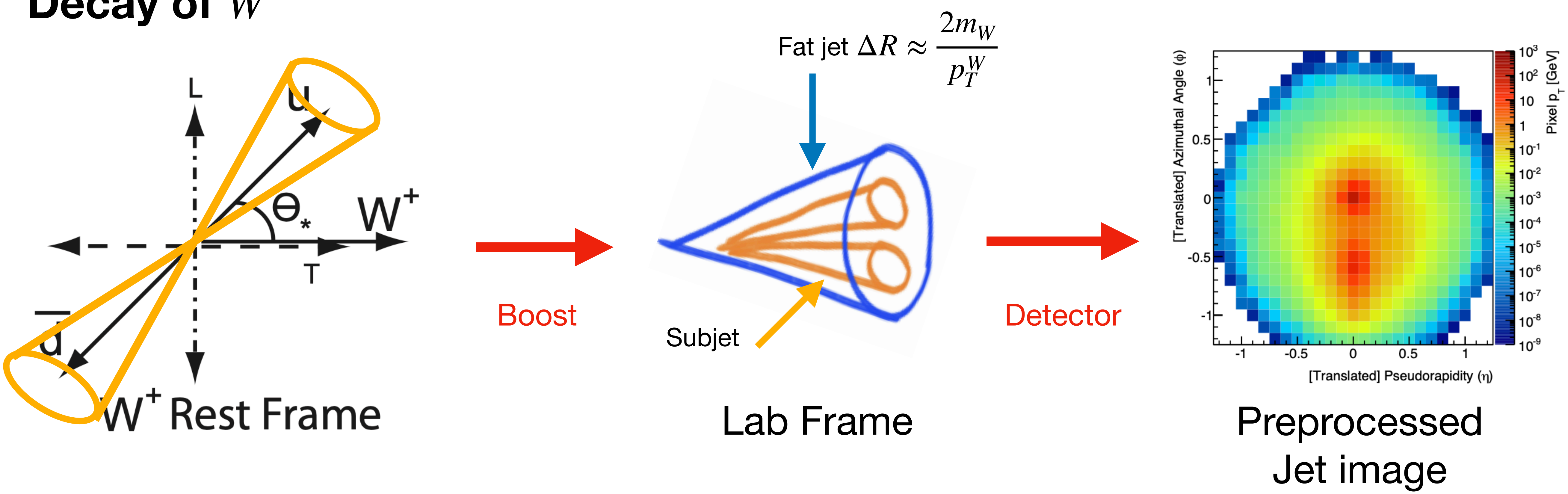
- We will later predict longitudinal content of $pp \to W^\pm(jj)Z(ll)$ for both SM and SMEFT cases

W Polarization Measurement

- Leptonic tagging is done
  - lower branching ratio than hadronic state
  - neutrino reconstruction
- Can we do hadronic *W* tagging?
  - Extract correct boson jet
  - QCD effect washes out parton level information
  - Possible jet substructure information can be used to tag polarization
    - N-subjettiness (S. De, V. Rentala, W. Shepherd arXiv:2008.04318v1)
  - **How does machine learning do?**

First ingredient: Boosted W Jet

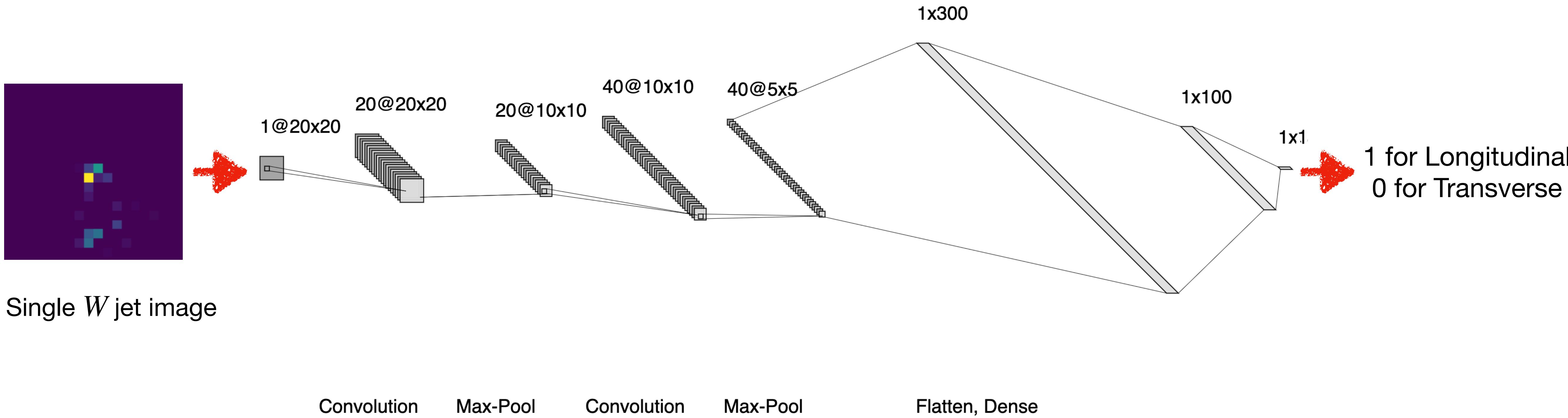
Decay of W



- Quark becomes QCD jet
- Due to the boost, collimation of the jet deduces the angular distribution signature
- After boost $\theta^* \rightarrow$ opening angle (sensitive to pT)
- At extreme high p_T^W, subjet signature can disappear
- Particles are plotted on pixelized $\eta - \phi$ plane and their color is determined from p_T

Convolutional Neural Network (CNN)

Image classification



- The network is trained with simulated events **(MadGraph** + **Pythia** + **Delphes)** of boosted longitudinal and transverse W's respectively for tagging purposes

Training Sample

Longitudinal : $pp \to \phi \to W^{\pm}W^{\mp}$

Transverse : $pp \to W^{\pm}j$

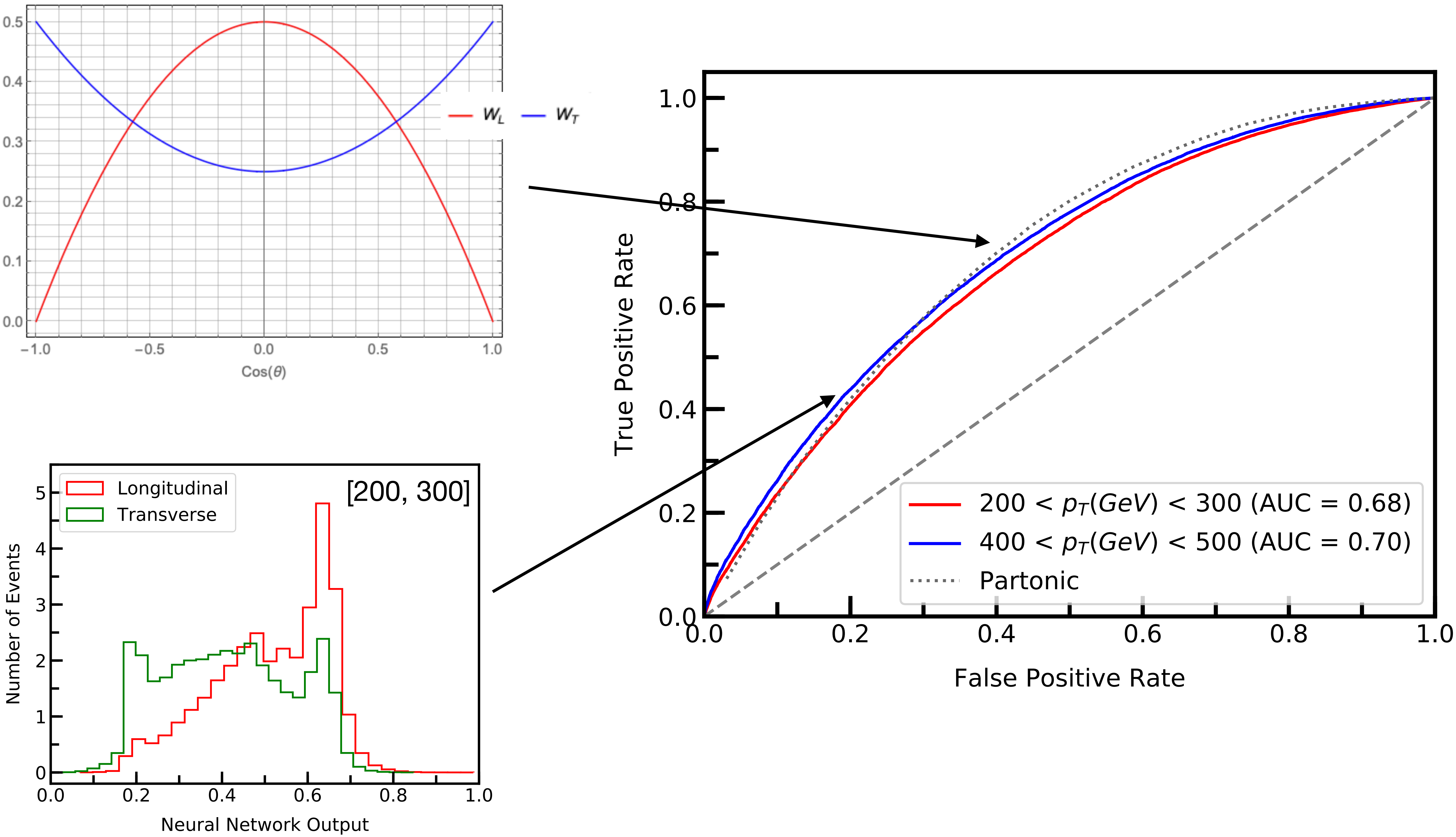
Testing Sample

$pp \to W^{\pm}(j)Z(l^{+}l^{-})$

- We did not consider any underlying events (looking into ideal scenarios as first study)

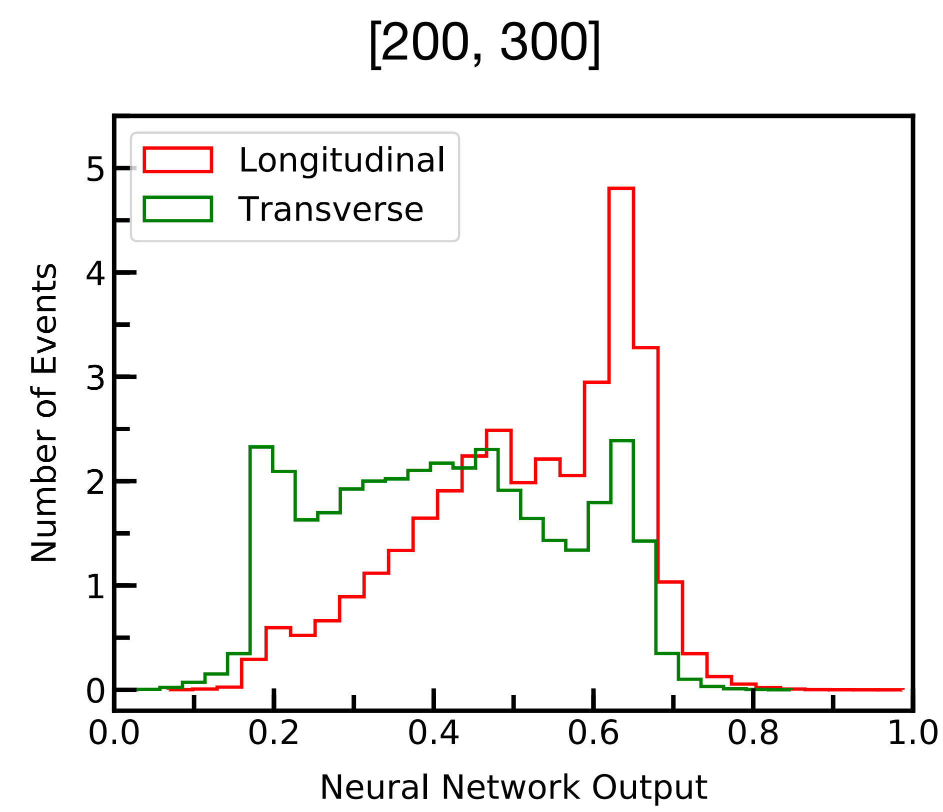
- Depending on p_T^W, images are separated into 2 bins: [200,300] and [400,500] since for fat jet, $\Delta R \approx \frac{2m_W}{p_T^W}$

ROC Curve



W_L
W_T
Cos(θ)
Longitudinal
Transverse
[200, 300]
Number of Events
Neural Network Output
True Positive Rate
False Positive Rate
200 < $p_T(GeV)$ < 300 (AUC = 0.68)
400 < $p_T(GeV)$ < 500 (AUC = 0.70)
Partonic

Trained Network Quality Check



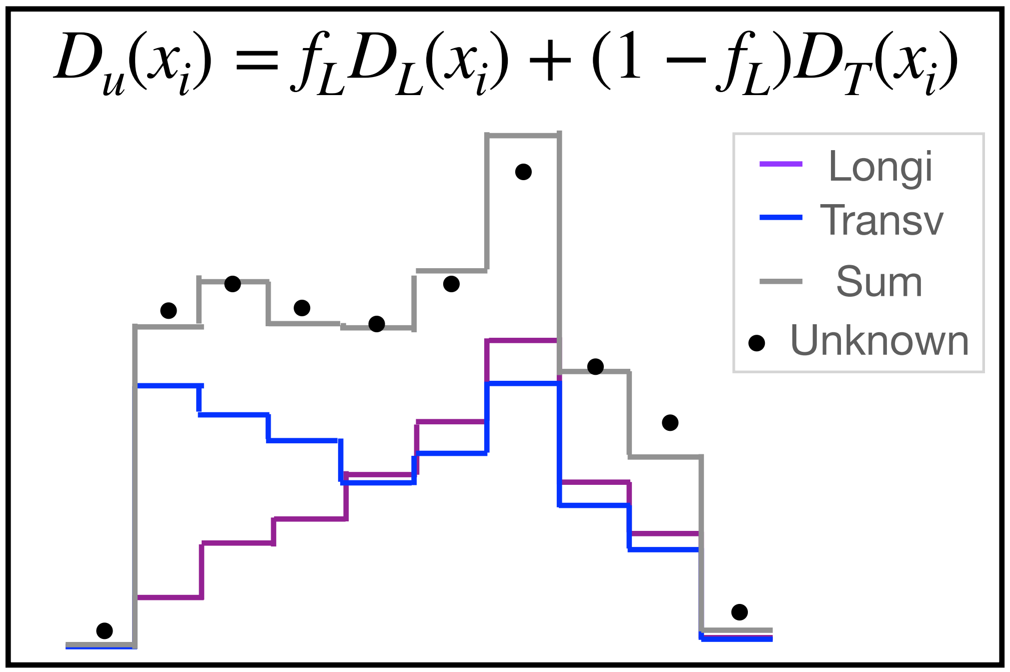
- Checking distribution can tell us how good the separation between two polarization
- **Inhibits potential event by event tagging because of large overlap**

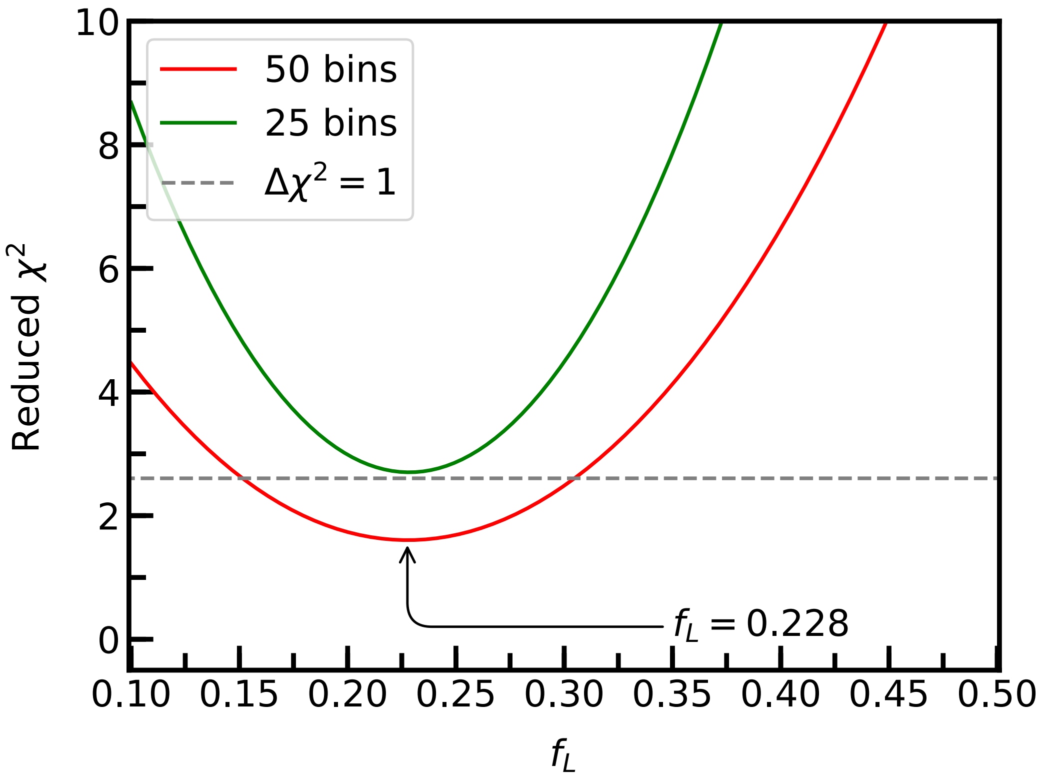
Putting decision threshold would contain large contamination

- As a result, we decide to use pure polarization distributions as template to identify the polarization content in given event collection
  - randomly select number of jet images from unknown sample → polarization fraction

Longitudinal fraction (f_L)

Template fit method





- Consider each pure polarization histogram as "template" that can be applied to the unknown sample

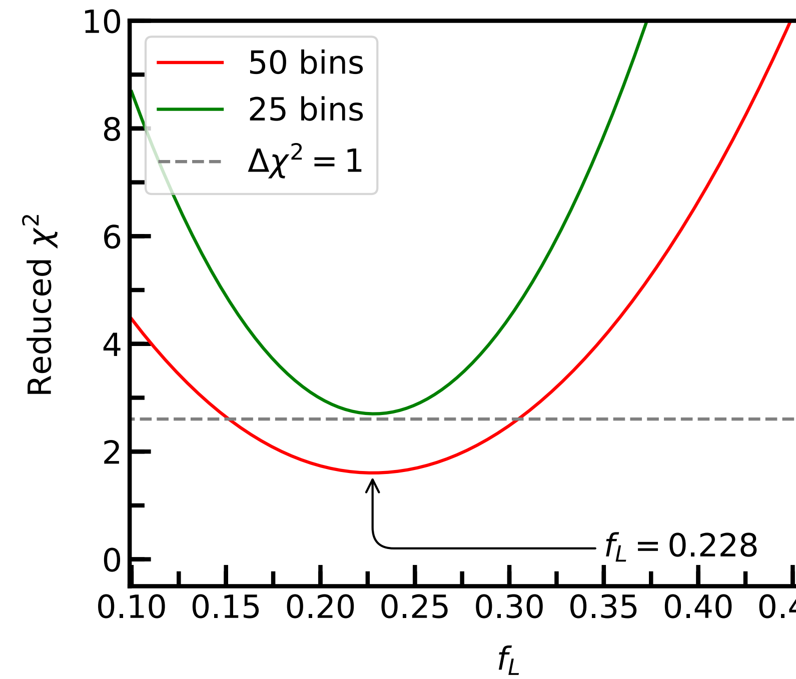
- Fit quality is determined by χ^2 distance test

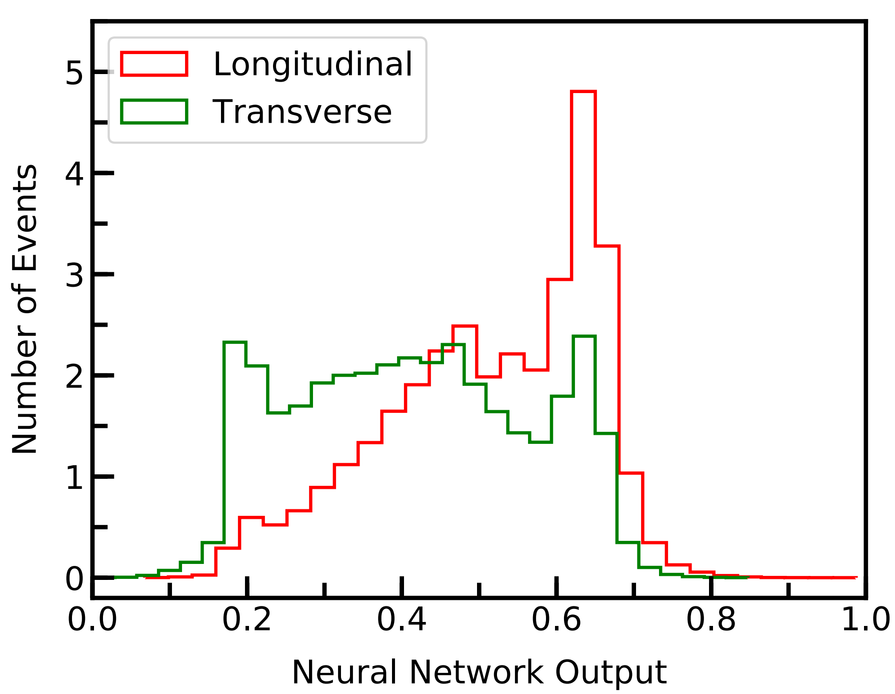
$$\chi^2(f_L) = \sum_{i=1}^{B} \frac{(O_i - N_s(f_L L_i + (1 - f_L)T_i))^2}{N_s(f_L L_i + (1 - f_L)T_i)}$$

Simpler Method

Network output average method

- Template fitting method depends on finding "sweet spot" for f_L
  - number of bins
  - find minimum $\chi^2(f_L)$
- Simplify by treating output distribution as probability distribution





$$\int x dx \left(D_u(x) = f_L D_L(x) + (1 - f_L) D_T(x) \right)$$

$$\langle x_u \rangle = f_L \langle x_L \rangle + (1 - f_L) \langle x_T \rangle$$

$$f_L = \frac{\langle x_u \rangle - \langle x_T \rangle}{\langle x_L \rangle - \langle x_T \rangle}$$

Confirmed that both yield the same result

SM Prediction Result

$$p\,p \to W^{\pm} Z$$

p_T range	truth σ_L/σ_{tot}	predicted f_L
[200,300]	0.265	0.259 ± 0.013
[400,500]	0.304	0.300 ± 0.033

- Truth f_L is obtained from polarization enforced feature of MadGraph
- At both p_T, predicted values are accurate with enough precision
- At high p_T, larger uncertainty comes from lower statistics
- Error is estimated from pseudo experiments
- CNN can predict well with SM case but need to test more (SMEFT extension)

SMEFT Extension (Scenario 1)

Shift cross section and polarization fraction

SM

p_T range	$\sigma(pp \to W^{\pm}(jj)Z(\ell\ell))$ (fb)	truth $\sigma_L/\sigma_{\text{tot}}$	predicted f_L
200 GeV $\leq p_T \leq$ 300 GeV	6.67	0.265	0.259 ± 0.013
400 GeV $\leq p_T \leq$ 500 GeV	0.35	0.304	0.300 ± 0.033

SM + single operator

	p_T range	$\sigma(pp \to W^{\pm}Z)$ (fb)	truth $\sigma_L/\sigma_{\text{tot}}$	predicted f_L
O_W	200 GeV $\leq p_T \leq$ 300 GeV	6.93	0.311	0.297 ± 0.010
	400 GeV $\leq p_T \leq$ 500 GeV	0.42	0.439	0.391 ± 0.033
O_{3W}	200 GeV $\leq p_T \leq$ 300 GeV	6.58	0.258	0.254 ± 0.011
	400 GeV $\leq p_T \leq$ 500 GeV	0.50	0.198	0.181 ± 0.043

- Benchmark Wilson Coefficient values
  - $c_W = 10^{-3}$
  - $c_{3W} = 3 \times 10^{-3}$
- If cross section measurement does not match to SM, polarization measurement can be a key to spot the dominancy

SMEFT Extension (Scenario 2)

Equal cross section but shift polarization fraction

SM

p_T range	$\sigma(pp \to W^{\pm}(jj)Z(\ell\ell))$ (fb)	truth $\sigma_L/\sigma_{\text{tot}}$	predicted f_L
200 GeV $\leq p_T \leq$ 300 GeV	6.67	0.265	0.259 ± 0.013
400 GeV $\leq p_T \leq$ 500 GeV	0.35	0.304	0.300 ± 0.033

SM + $\mathcal{O}_W$ + $\mathcal{O}_{3W}$

p_T range	$\sigma(pp \to W^{\pm}Z)$ (fb)	truth σ_L/σ_{tot}	predicted f_L
200 GeV $\leq p_T \leq$ 300 GeV	6.68	0.202	0.207 ± 0.011
400 GeV $\leq p_T \leq$ 500 GeV	0.34	0.285	0.282 ± 0.044

- Two Wilson coefficients are tuned to keep cross section the same but shift f_L
- Even though cross section agrees with SM, polarization measurement can be a way to capture BSM signatures

Conclusion/Discussion

- In this initial study, analysis using network's output average values can help to predict f_L even for hadronic W

- Network prediction can catch f_L deviations originated from dim 6 operators

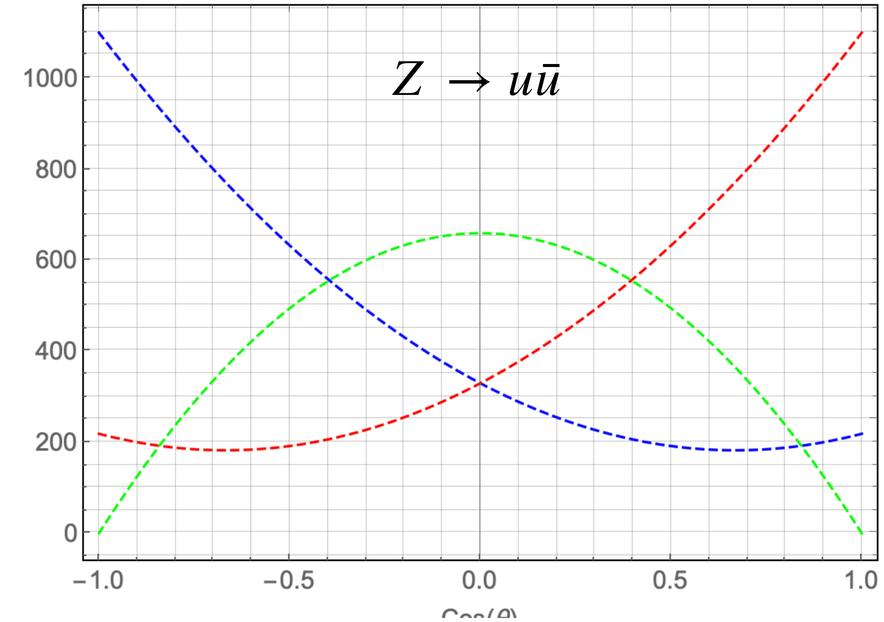
- With or without cross section shift, polarization measurement can clear out degeneracies between EFT operators

- Future directions

    - Possible applicability on Z jets

    - $W^{\pm}$ vs. Z vs. QCD (adding more realistic components)

    - Further optimization of network



Different Network Results

MaxOut & ResNet

	truth	f_L CNN	f_L MaxOut	f_L ResNet
SM	0.265	0.259 ± 0.013	0.287 ± 0.011	0.259 ± 0.012
O_W	0.311	0.297 ± 0.010	0.321 ± 0.010	0.295 ± 0.009
O_{3W}	0.258	0.254 ± 0.011	0.282 ± 0.012	0.257 ± 0.011

Low pT

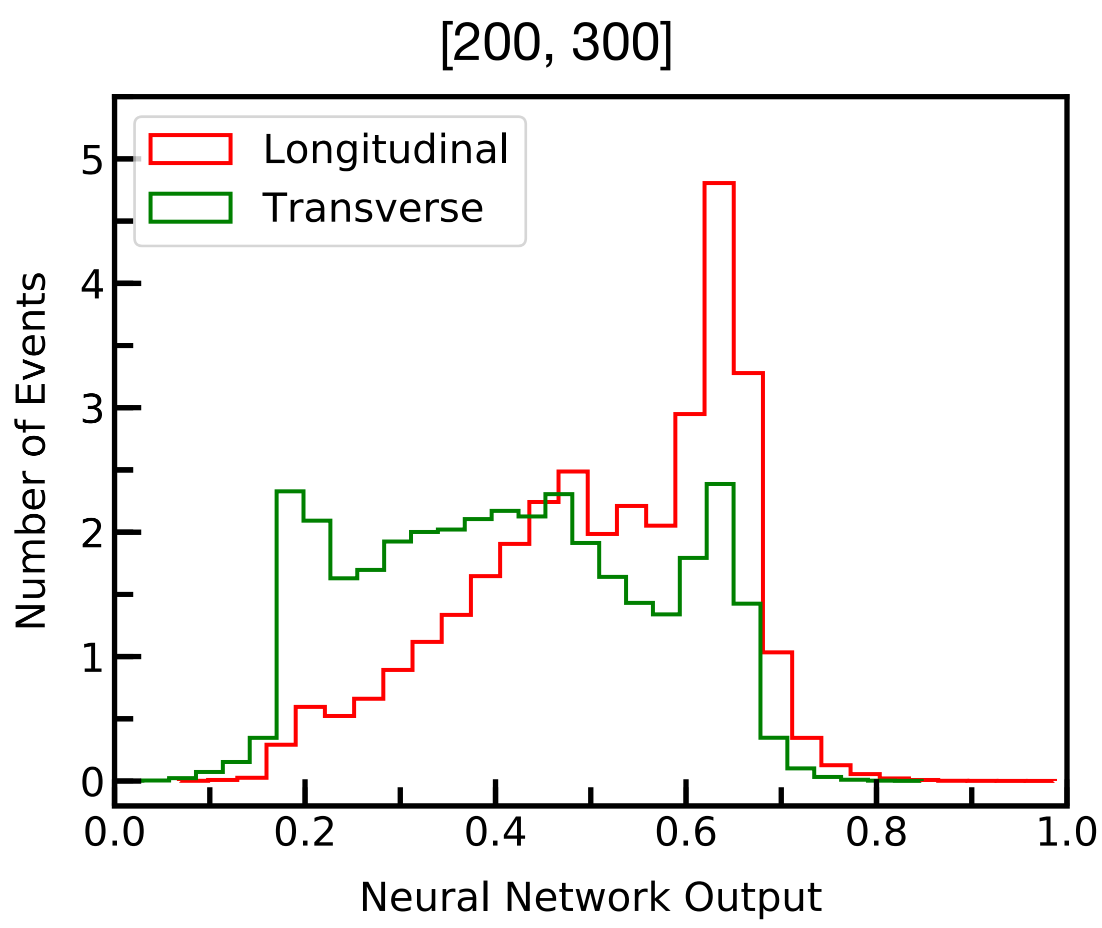
	truth	f_L CNN	f_L MaxOut	f_L ResNet
SM	0.304	0.300 ± 0.033	0.323 ± 0.026	0.301 ± 0.034
O_W	0.439	0.391 ± 0.033	0.407 ± 0.025	0.414 ± 0.034
O_{3W}	0.198	0.181 ± 0.043	0.250 ± 0.026	0.194 ± 0.032

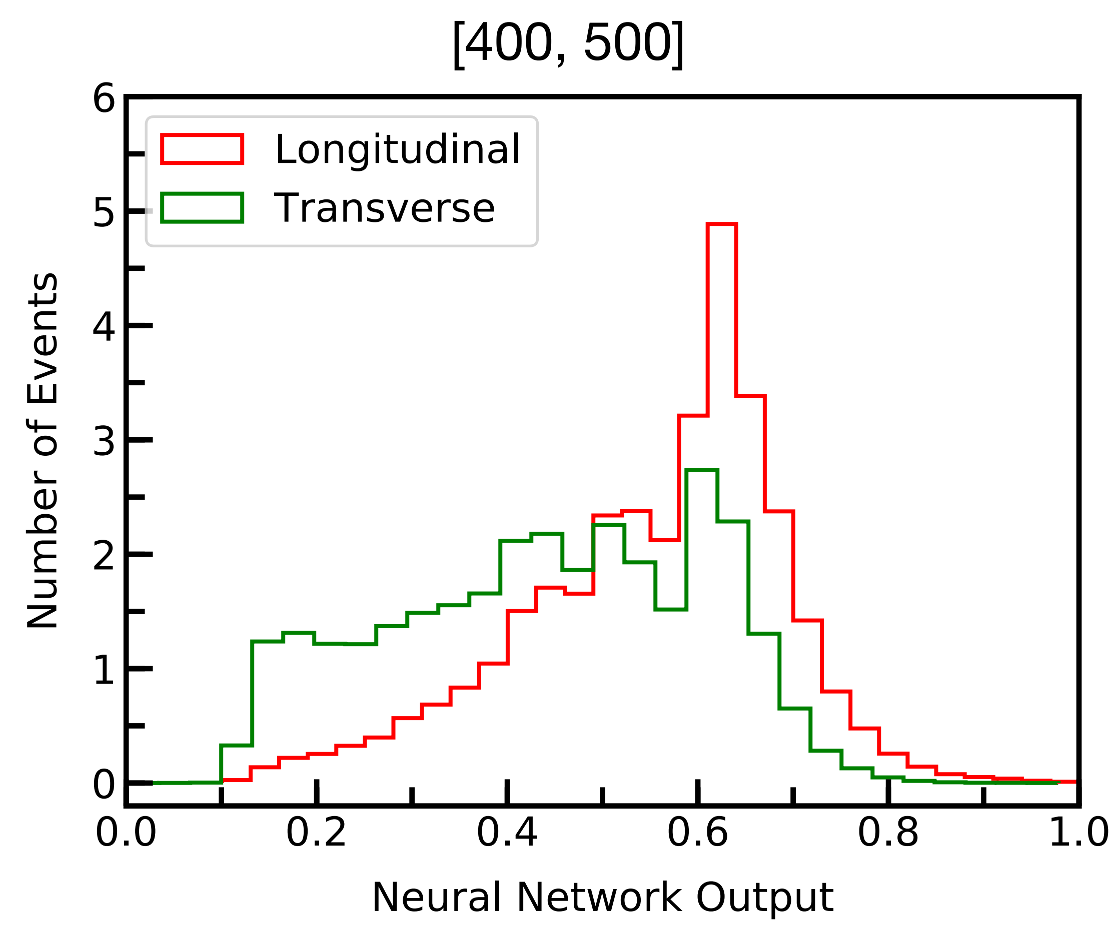
High pT

Backup slides

Training Quality

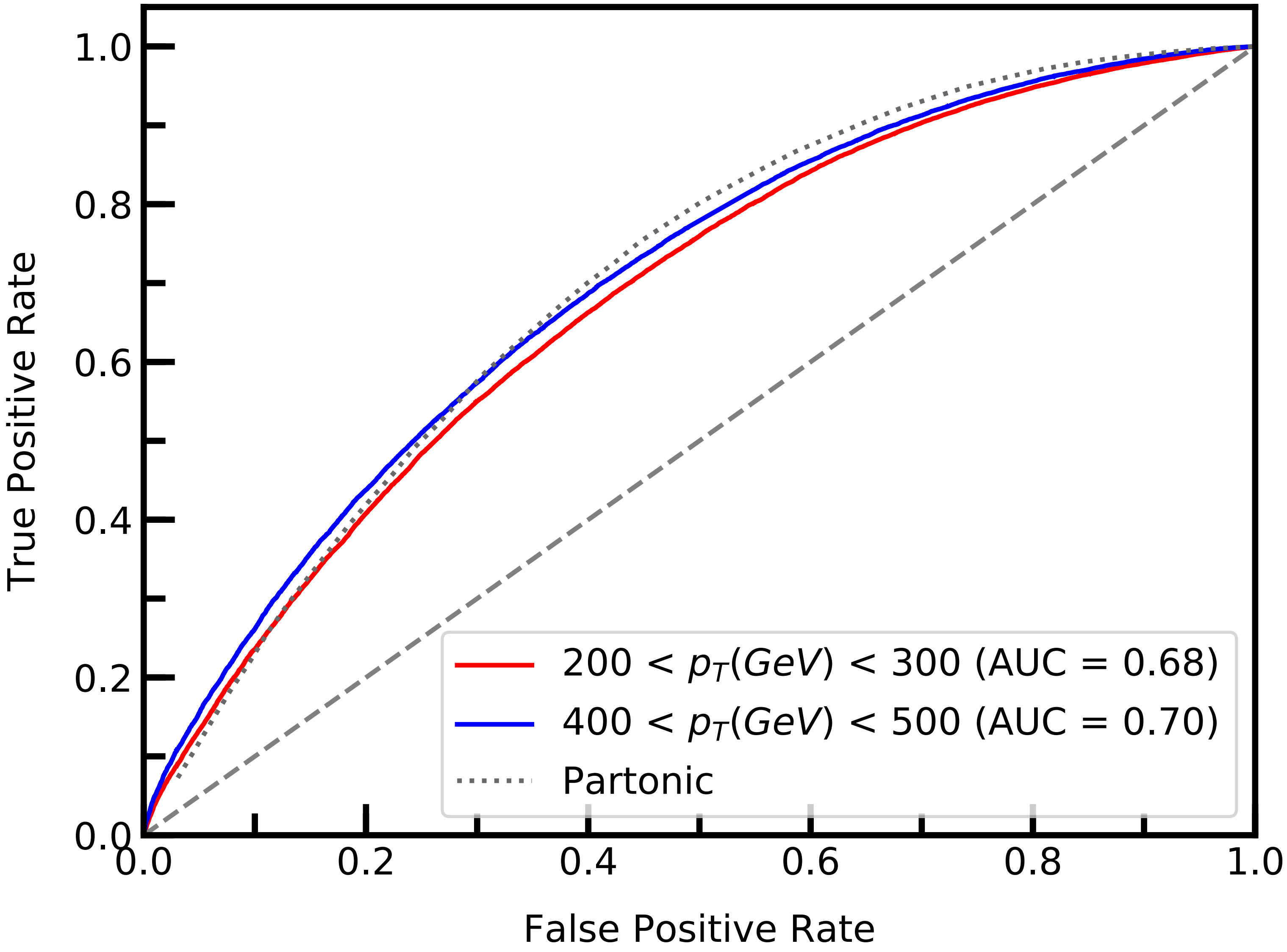
Distribution check





Training Quality

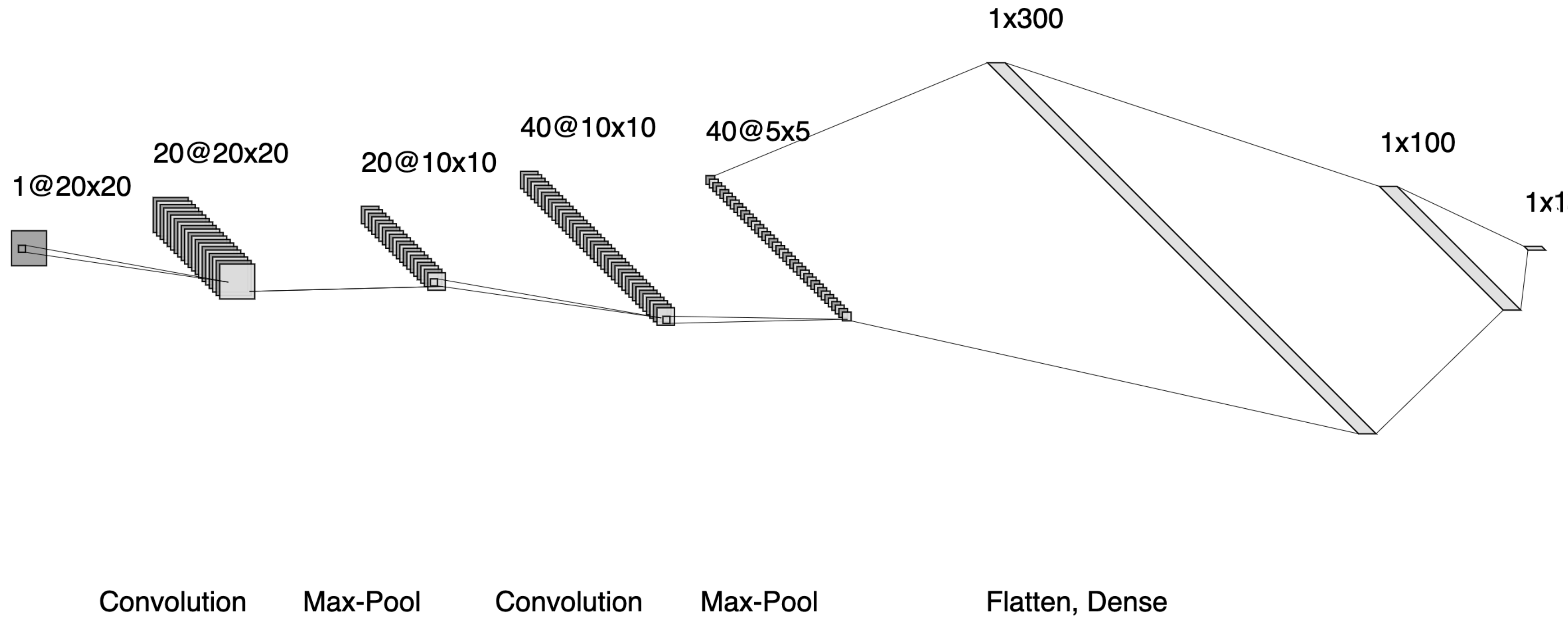
Distribution check



- Checking distribution can tell us how good the separation between Logi and trans is.
- Inhibits potential event by event tagging since accuracy is ~ 60%
- Ensemble distribution checking to find longitudinal fraction (f_L)

CNN Training

Structure and training information



0 for Transverse
1 for Longitudinal

Ordinary CNN structure : Convolution - Flatten - Dense

pT bin	Training/Validation	Validation accuracy
[200, 300]	340k/85k	63%
[400, 500]	236k/59k	64%

Kinematic Cut Effect

W rest frame

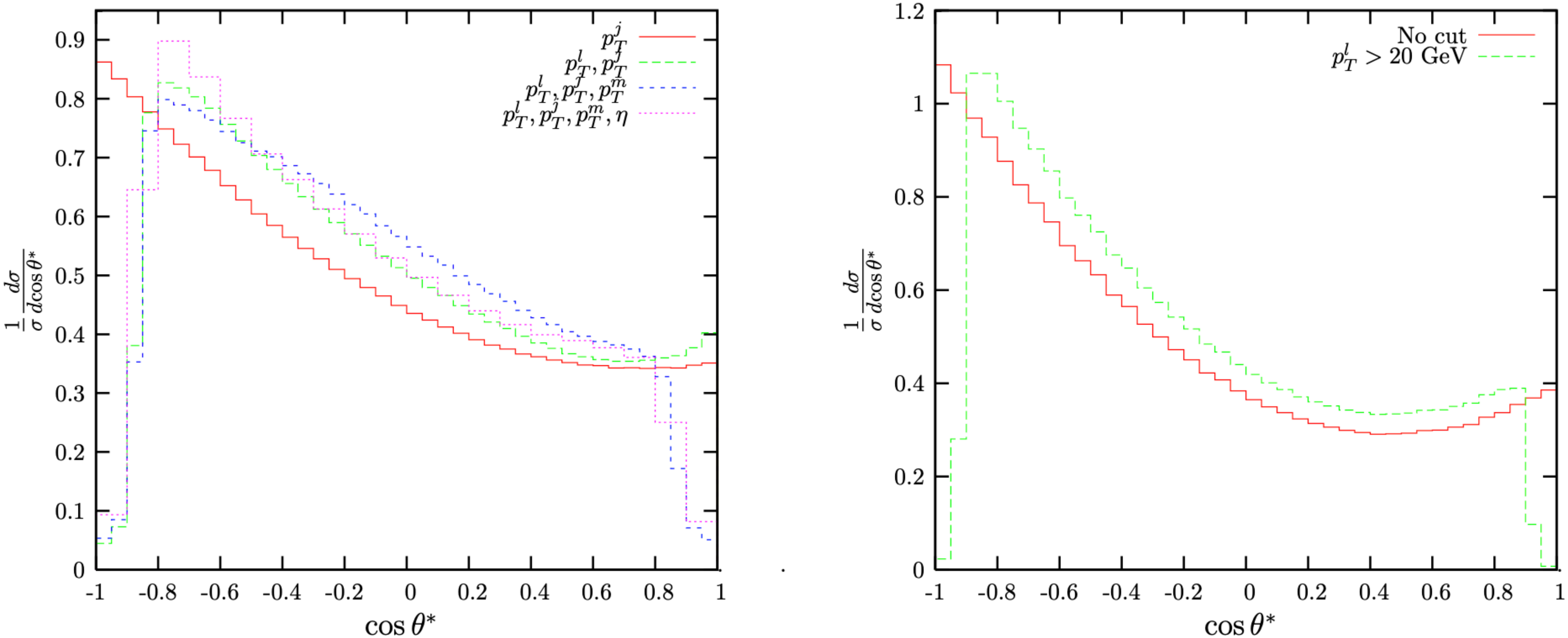
$$\frac{1}{\sigma}\frac{d\sigma}{d\cos\theta^*} = \frac{3}{8}(1-\cos\theta^*)^2 f_L + \frac{3}{8}(1+\cos\theta^*)^2 f_R + \frac{3}{4}\sin^2\theta^* f_0,$$

W at LHC

$$\frac{1}{\sigma}\frac{d\sigma}{d\cos\theta^* d\phi^*} = \frac{3}{16\pi}[(1+\cos^2\theta^*) + A_0\frac{1}{2}(1-3\cos^2\theta^*) + A_1\sin2\theta^*\cos\phi^*$$
$$+ A_2\frac{1}{2}\sin^2\theta^*\cos2\phi^* + A_3\sin\theta^*\cos\phi^* + A_4\cos\theta^*],$$

- Integrating over ϕ^* will give the same result but kinematic cut can change

Kinematic Cut Effect



Kinematic Cut Effect

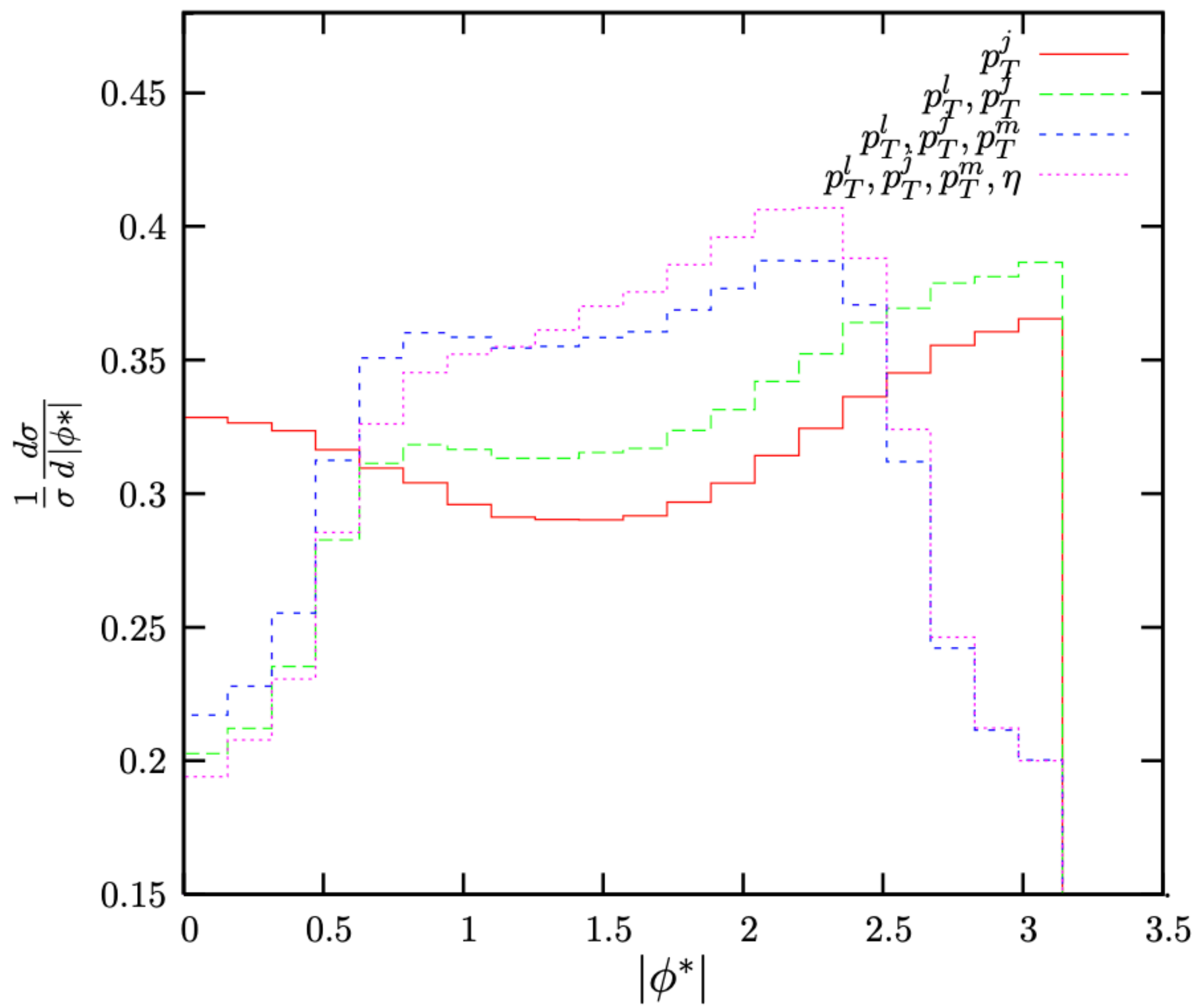


Figure 9: Normalised azimuthal angle distributions for a set of different selection cuts imposed on final-state leptons and jets for $W^+ + 1$ jet production at 7 TeV.

Uncertainty

Small experiments

- From large test set, we randomly select subset (N number of events) to obtain f_L
- N is determined from expected number of events at particular luminosity
- At current LHC luminosity ~ 2000 events at low p_T and 200 events at high p_T
- At High Lumi LHC ~ 20k events at low p_T and 2k events at high p_T
- By iterating the process, we can obtain average value with standard deviation

	300 fb^{-1}	**3000 fb^{-1}**
[200,300]	0.044	0.010
[400,500]	0.130	0.033

Experimental Results

ATLAS result

ATLAS Result (36fb^{-1})

	f_0		
	Data	POWHEG+PYTHIA	MATRIX
W^+ in W^+Z	0.26 ± 0.08	0.233 ± 0.004	0.2448 ± 0.0010
W^- in W^-Z	0.32 ± 0.09	0.245 ± 0.005	0.2651 ± 0.0015
$W^\pm$ in $W^\pm Z$	0.26 ± 0.06	0.2376 ± 0.0031	0.2506 ± 0.0006
Z in W^+Z	0.27 ± 0.05	0.225 ± 0.004	0.2401 ± 0.0014
Z in W^-Z	0.21 ± 0.06	0.235 ± 0.005	0.2389 ± 0.0015
Z in $W^\pm Z$	0.24 ± 0.04	0.2294 ± 0.0033	0.2398 ± 0.0014

ATLAS Collaboration [arXiv:1902.05759]

1. Previous attempts from ATLAS collaboration to measure polarization with leptonic final states
    - Leptonic final state: small branching ratio
    - Complication in ν reconstruction
2. If we can use hadronic W, we gain more statistics but need to deal with hadronic jets