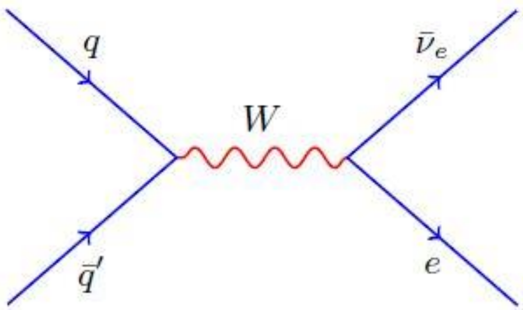
Study of the normalized p_T^W Distribution in $p\bar{p}$ Collisions at D0

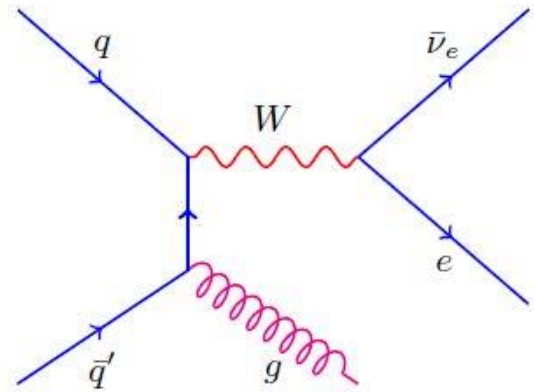
Chen Wang
University of Science and Technology of China
On Behalf of the D0 Collaboration

ICHEP 2020
July. 28th, Prague

Motivation

- p_T^V is described by QCD calculation





  - Leading Order (LO): $p_T^V = 0$
  - Including higher order: p_T^V arise from the initial state parton emission
  - Test QCD predictions

- In $p\bar{p}$ collisions, the production dominated by valence quark
  - In the LHC experiments, it involves sea quarks

- Low $p_T(V)$ region dominated by multiple soft gluon emissions
  - QCD predictions from a soft-gluon resummation formalism (CSS[1])
  - Using a form factor with 3 non-perturbative parameters, g_1, g_2 and g_3(BLNY[2])
  - g_1, g_2 and g_3 fixed to previous measurement[2]
  - Constrain models of non-perturbative approaches
  - Benefit other related electroweak parameter measurements such as m_W

1. Nucl. Phys. B250, 199 (1985)
2. Phys. Rev. D 67, 073016 (2003)

Introduction

- First Tevatron Run II p_T^W study
  - First p_T^W study at $\sqrt{s} = 1.96$ TeV

- Based on the latest D0 published m_W measurement
  - Same data sample, 4.35 fb^{-1} Run II Data
  - Same background estimation strategy
  - Same detector calibration methodologies
  - Same parametrized MC simulation (PMCS)

- Focus on low p_T^W region (<15 GeV)
  - Sensitive to QCD non-perturbative parameters

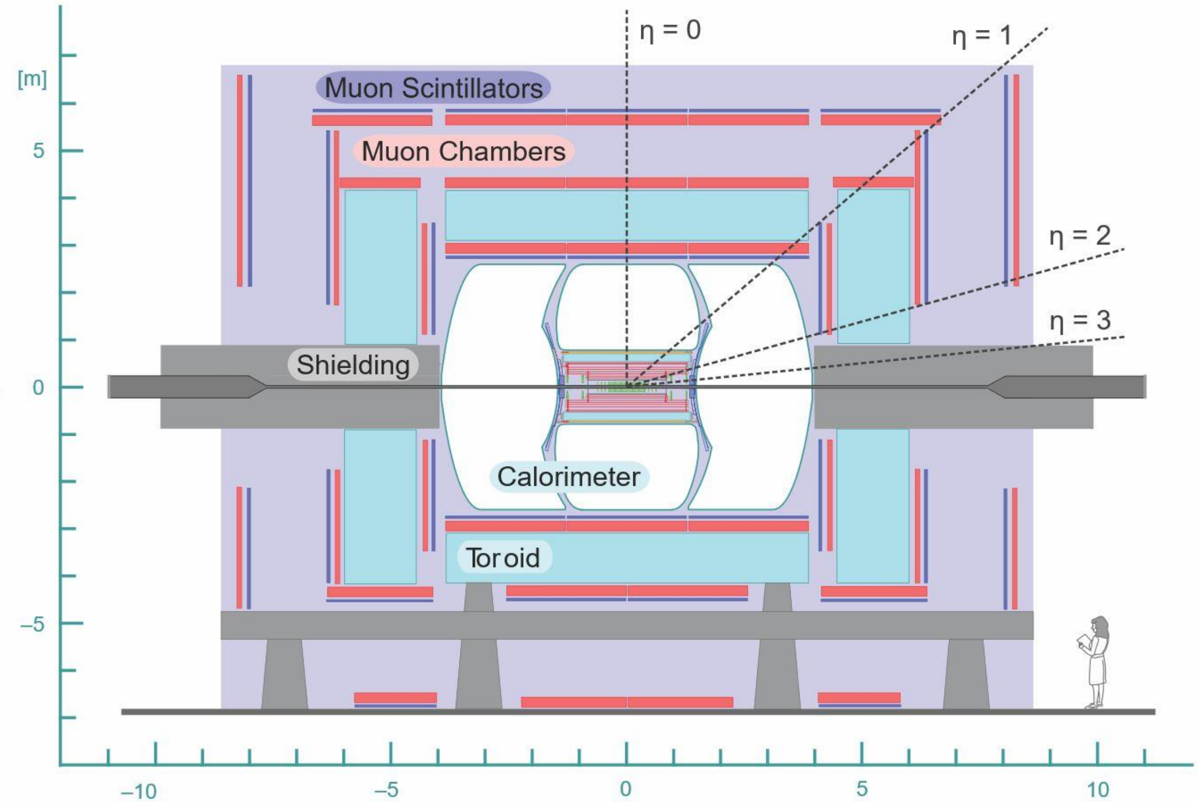
- Provide reconstruction level results
  - A fast folding procedure for comparisons to other models

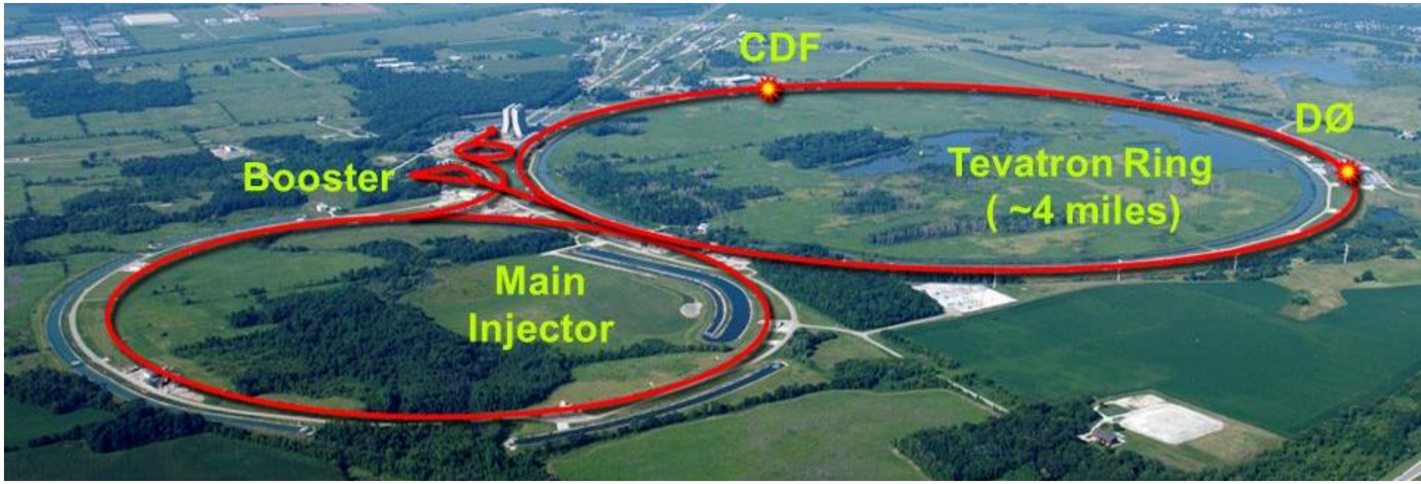
The latest D0 published m_W measurement: Phys. Rev. Lett. 108, 151804 (2012)
Phys. Rev. D 89, 012005 (2014)

D0 Detector

Central tracking system
- Silicon Microstrip Tracker (SMT)
- Scintillating Central Fiber Tracker (CFT)
- 1.9 T Solenoid

Calorimeter
- Liquid argon and uranium $|\eta| < 4.2$
- Electron energy measurement
- Hadronic recoil reconstruction





Samples and selections

- Data: Run II, 4.35 fb^{-1}, $\sqrt{s} = 1.96$ TeV
- Trigger requirement:
  - At least one electromagnetic (EM) cluster
  - Transverse energy threshold: 25~27 GeV depending on instant luminosity

- Offline selections:
  - Electron candidate:
    $p_T^e > 25$ GeV, $|\eta^e| < 1.05$
    Pass shower shape and isolation requirements
  - W candidate:
    At least one electron candidate
    $50 < m_T < 200$ GeV, $p_T^{Missing} > 25$ GeV, $u_T < 15$ GeV

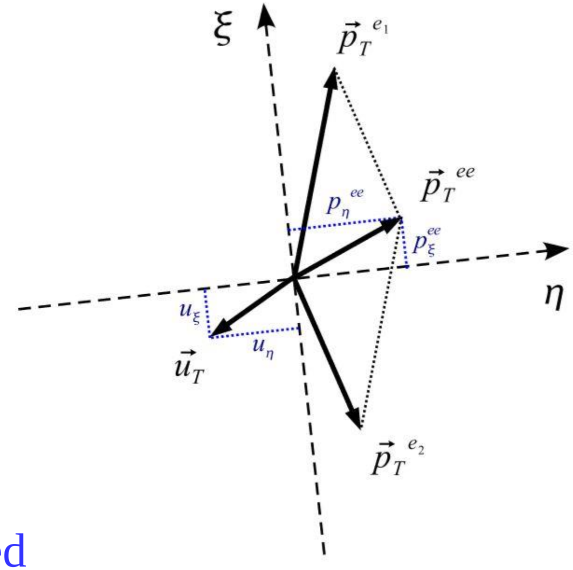
- Hadronic Recoil $\vec{u}_T = \sum \vec{E}_T^{calo}$, represents p_T^W
  - The vector sum of reconstructed energy clusters in the calorimeters excluding deposits from the lepton

- $\vec{E}_T^{Missing} = -(\vec{u}_T + \vec{p}_T^e)$, represents p_T^ν

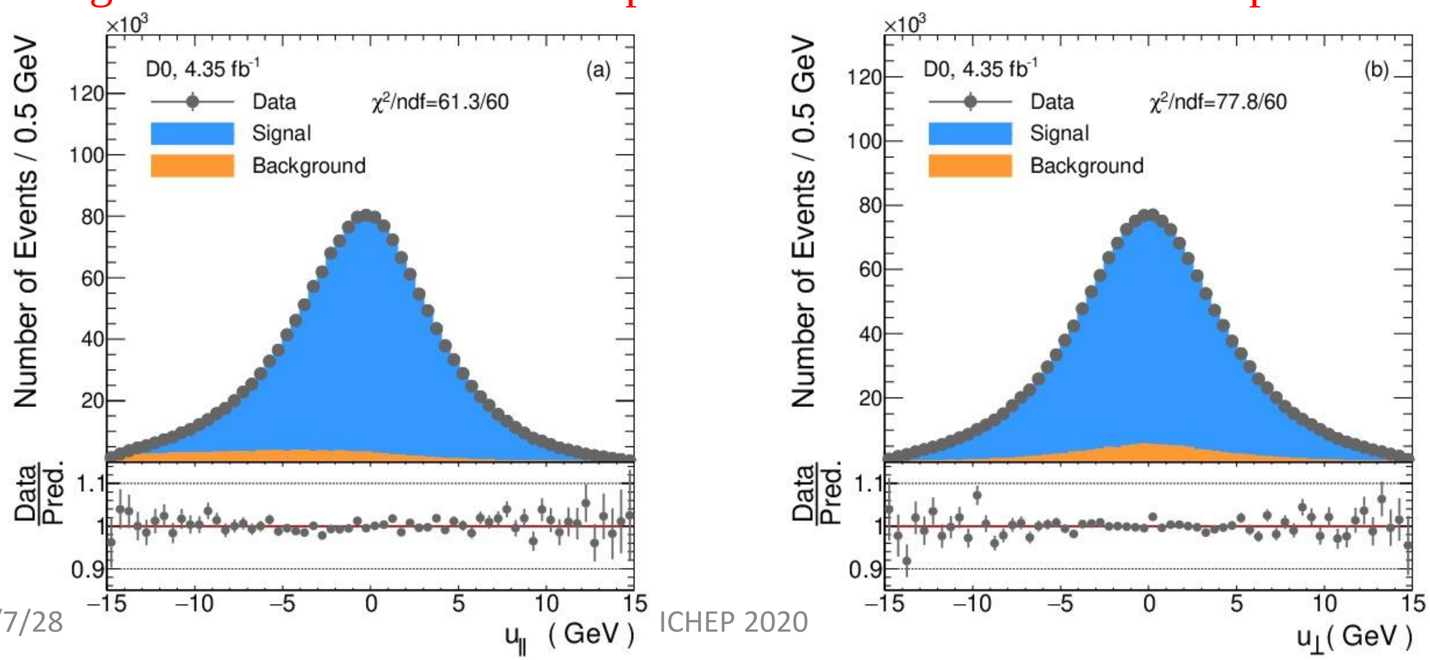
$$m_T = \sqrt{2p_T^e p_T^\nu(1 - cos\Delta\phi)}$$

- Detector Calibration
  - Electron energy calibrated using Z mass
    - Two parameters: $E_{corr} = \alpha E_{obs} + \beta$
  - Hadronic Recoil calibrated with Z candidates
    - $\hat{\eta}$: the direction bisecting the two electrons
    - Tuned by the imbalance in $\hat{\eta}$ direction, η_{imb}

$$\eta_{imb} = (\vec{u}_T + \vec{p}_T^{ee}) \cdot \hat{\eta}$$



  - In W candidates, only one charged lepton reconstructed
    - $u_{\parallel}$ and $u_{\perp}$: the parallel and perpendicular components to the electron direction
    - Tests the modeling of the hadronic recoil
  - Good agreement between data and prediction on hadronic recoil response



➢ Background Estimation

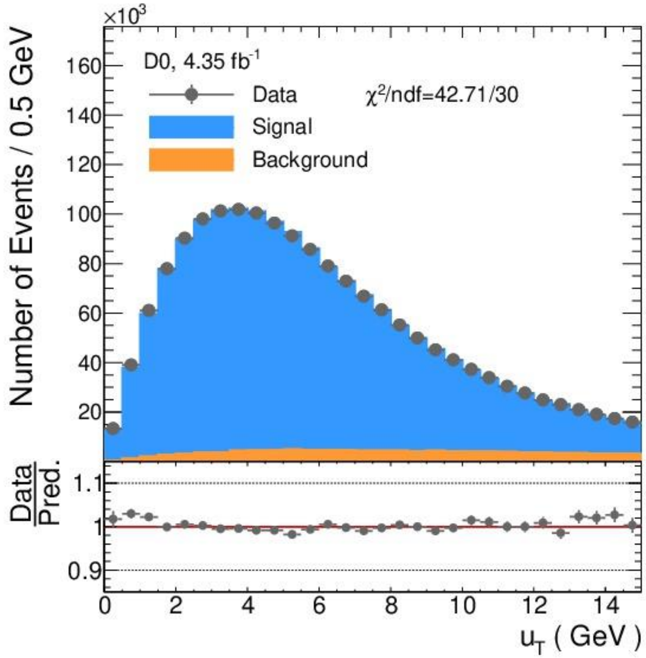
- ➢ Three backgrounds: W $\to \tau\upsilon \to e\upsilon\upsilon\upsilon$, $Z \to ee$, Multi-Jet
  - ➢ W $\to \tau\upsilon \to e\upsilon\upsilon\upsilon$: Estimated from MC simulation (PMCS)
  - ➢ $Z \to ee$: one electron escapes detection
  - ➢ Multi-Jet: one jet misidentified as one electron

} Estimated from data

Background	$W \to \tau\upsilon$	$Z \to ee$	MJ
Fraction	$1.668\% \pm 0.004\%$	$1.08\% \pm 0.02\%$	$1.018\% + 0.065\%$

- ➢ Background less than 4%, uncertainty due to the background estimation is negligible

- ➢ Good agreement between data and prediction at the reconstruction level



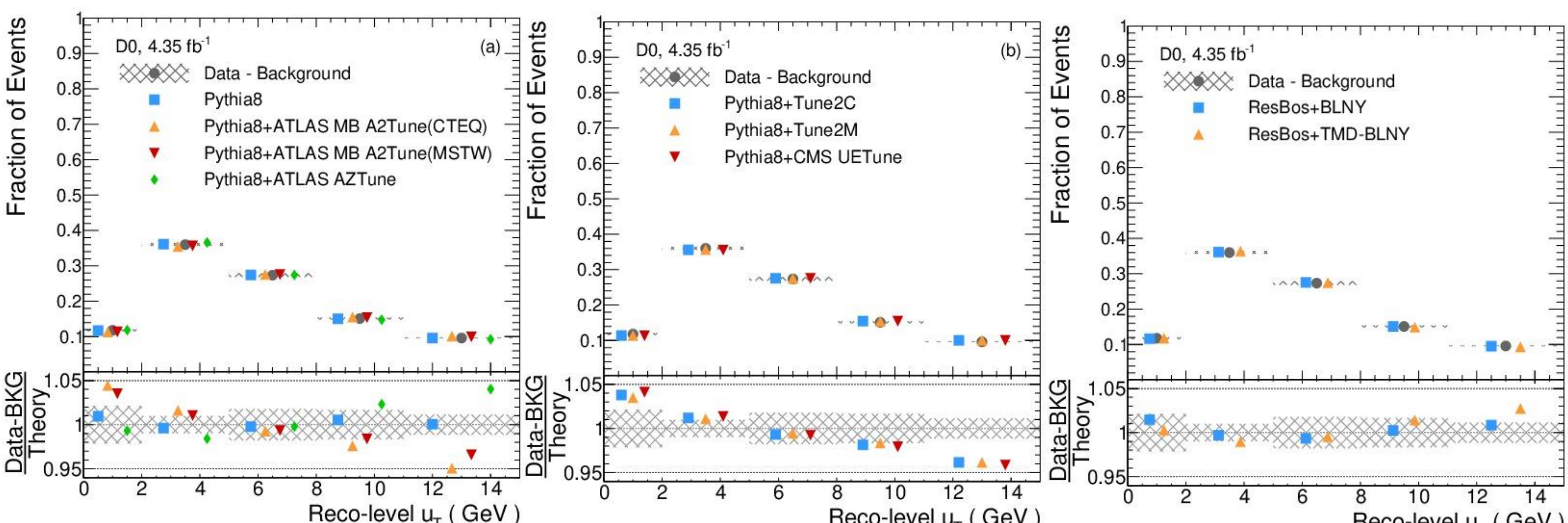
u_T bin	0–2 GeV	2–5 GeV	5–8 GeV	8–11 GeV	11–15 GeV
Fraction of events in the u_T bin	0.1181	0.3603	0.2738	0.1515	0.0963
Total uncertainty	0.0003	0.0005	0.0005	0.0004	0.0003

PMCS Reweighting

- $p_T^W - y^W$ 2D distribution reweighted to other theory predictions
  - The default PMCS: ResBos+BLNY
  - Resummation: other non-perturbative functional form (TMD-BLNY)
  - Parton shower: different Pythia8 tunes from other collaborations

- Systematic uncertainty estimated by changing parameters in PMCS
  - Separately estimated with each model
  - Dominated by the uncertainty due to the hadronic recoil calibration
  - Bin-by-bin correlation estimated

- Fraction of events of the background-subtracted data compared to different predictions



- χ^2 calculation
  - The χ^2 value between the data and the reweighted PMCS are calculated
    - All 5 u_T bins considered, n.d.f. equals to 4 due to the normalization
    - The bin-by-bin correlations are taken into account
    - The uncertainties due to the resummation and the tune are ignored
    - The PDF uncertainty is negligible

- Conclusion
  - 2 models excluded: Pythia8+ATLAS MB A2Tune+CTEQ6L1
    Pythia8+CMS UE Tune CUETP8S1-CTEQ6L1+CTEQ6L1
  - Other Pythia8 tunes except the default one are disfavored
  - ResBos+BLNY and the default Pythia8 agree with the data very well

Generator/Model	χ^2/ndf	p-value	Signif.
RESBOS (Version CP 020811)+BLNY+CTEQ6.6	0.49	7.41×10^{-1}	0.33
RESBOS (Version CP 112216)+TMD-BLNY+CT14HERA2NNLO	3.13	1.39×10^{-2}	2.46
PYTHIA 8+CT14HERA2NNLO	0.32	8.63×10^{-1}	0.17
PYTHIA 8+ATLAS MB A2Tune+CTEQ6L1	12.25	5.84×10^{-10}	6.19
PYTHIA 8+ATLAS MB A2Tune+MSTW2008LO	6.17	5.83×10^{-5}	4.02
PYTHIA 8+ATLAS AZTune+CT14HERA2NNLO	6.61	2.60×10^{-5}	4.21
PYTHIA 8+Tune2C+CTEQ6L1	7.66	3.61×10^{-6}	4.63
PYTHIA 8+Tune2M+MRSTLO	7.32	6.89×10^{-6}	4.50
PYTHIA 8+CMS UE Tune CUETP8S1-CTEQ6L1+CTEQ6L1	8.80	4.23×10^{-7}	5.06

- Compare to other theory predictions
  - Two approach to achieve this
    - Provide an unfolded particle level result
    - Provide a folding procedure
  - When the statistical uncertainty dominates
    - It can be proven that χ^2_{reco} is equal to χ^2_{unfold}
    - Same precision between the particle level and the reconstruction level comparisons
  - When the systematic uncertainty dominates
    - Linear extrapolation to the reconstruction level, non-linear to the particle level
    - Additional uncertainty on the particle level due to regularization procedure
    - The precision of the particle level comparison would be reduced
  - The reduction of the precision would be
    - Greater when the resolution of the distribution is worse
    - Smaller when the bin width is enlarged
  - This is why we chose to provide in our paper a folding procedure
    - Better precision than the unfolded results

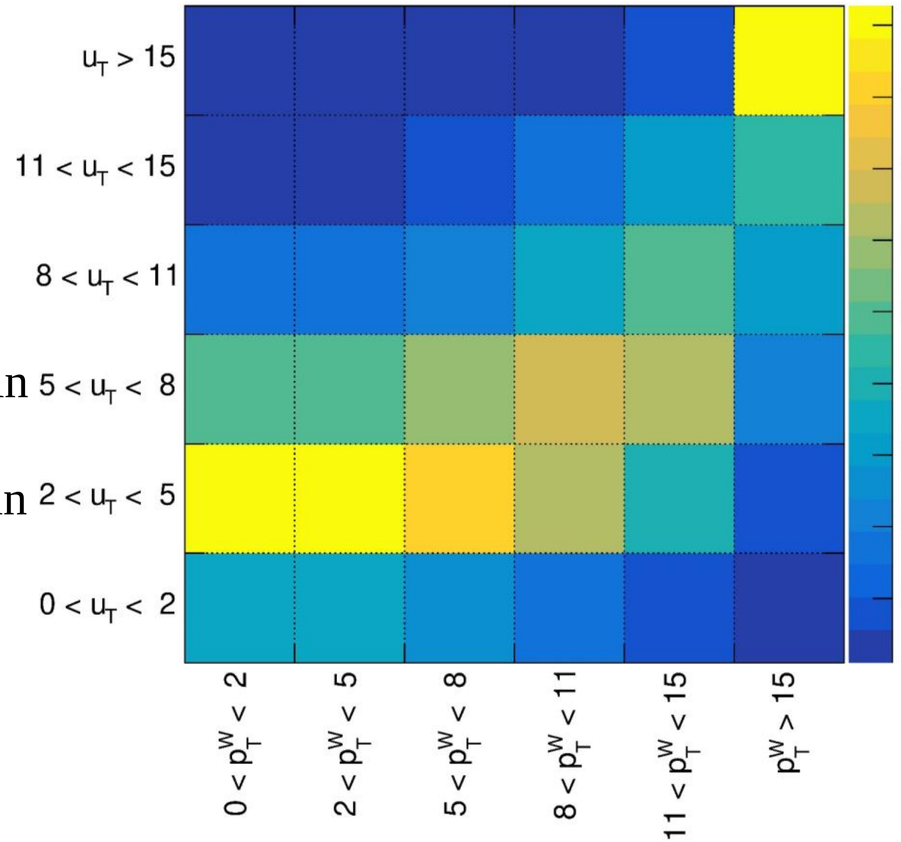
➢ Folding procedure

➢ The folded number of events in u_T bin i

$$N_i^{corr} = \frac{\sum_{j=1}^{6} R_{ij}\mathcal{E}_j X_j}{F_i}$$

X_i: the number of events in the i^{th} p_T^W bin
$\mathcal{E}_i$: the efficiency correction in the i^{th} p_T^W bin
F_i: the fiducial correction in the i^{th} u_T bin
N_i^{corr}: the number of events in the i^{th} u_T bin

Response Matrix R_{ij}



➢ Response Matrix R_{ij}:
The probability for the events in one p_T^W bin to be reconstructed into different u_T bins

➢ The instruction to calculate the covariance matrix and details of the whole folding procedure are introduced in the appendix of the paper (arXiv:2007.13504)

https://www-d0.fnal.gov/Run2Physics/WWW/results/final/EW/E20A/

Summary

- First Tevatron Run II $p_T(W)$ measurement at $\sqrt{s} = 1.96$ TeV
  - Focus on low $p_T(W)$ region

- The background subtracted data is compared to different predictions after PMCS simulation on the reconstruction level
  - Uncertainty dominated by that due to the hadronic recoil calibration
  - Model uncertainty ignored
  - PDF uncertainty negligible

- Two models are excluded
  - Pythia8+ATLAS MB A2Tune+CTEQ6L1
  - Pythia8+CMS UE Tune CUETP8S1-CTEQ6L1+CTEQ6L1

- A folding procedure is provided for the comparison with other models
  - Better precision than the unfolded results
  - Model dependence tested to be negligible
  - χ^2 difference smaller than the impact from the data fluctuation

Link to the paper:

https://www-d0.fnal.gov/Run2Physics/WWW/results/final/EW/E20A/

https://arxiv.org/abs/2007.13504

- Backup

➢ Collins-Soper-Sterman (CSS) resummation formalism

➢ Production of a vector boson in the collision of two hadrons

$$\frac{d\sigma(h_1 h_2 \to VX)}{dQ^2 dQ_T^2 dy} = \frac{1}{(2\pi)^2}\delta(Q^2 - M_V^2)\int d^2b\, e^{i\vec{Q}_T\cdot\vec{b}}\widetilde{W}_{j\bar{k}}(b, Q, x_1, x_2) + Y(Q_T, Q, x_1, x_2)$$

b: impact parameter

➢ the nonperturbative terms in the form of an additional factor $\widetilde{W}_{j\bar{k}}^{NP}(b, Q, x_1, x_2)$

$$\widetilde{W}_{j\bar{k}} = \widetilde{W}_{j\bar{k}}^{pert}\widetilde{W}_{j\bar{k}}^{NP}$$

➢ Brock-Landry-Nadolsky-Yuan form

$$\widetilde{W}_{j\bar{k}}^{NP}(b, Q, x_1, x_2) = \exp\left(-g_1 - g_2 \ln\left(\frac{Q}{2Q_0}\right) - g_1 g_3 \ln(100 x_1 x_2)\right) b^2$$

CSS: Nucl. Phys. B250, 199 (1985)
BLNY: Phys. Rev. D 67, 073016 (2003)

➢ Response Matrix R

- ➢ The probability for the events in one p_T^W bin to be reconstructed into different u_T bins

$$R_{ij} = P(\mathcal{N}_i|\mathcal{X}_j)$$

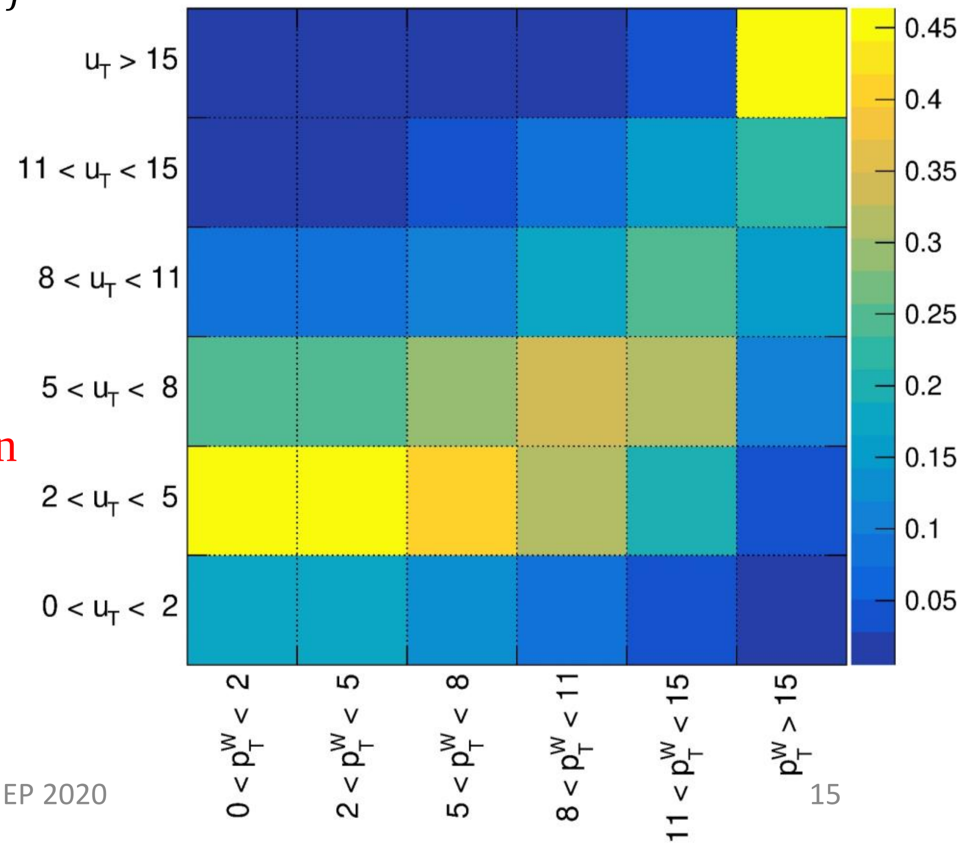
$\mathcal{N}_i$: the case that u_T is in the i^{th} bin
$\mathcal{X}_i$: the case that p_T^W is in the i^{th} bin

N_i: the number of events in the i^{th} u_T bin
X_i: the number of events in the i^{th} p_T^W bin

$$N_i = \sum_j R_{ij}X_j$$

- ➢ Purity R_{ii}:
- ➢ The probability for the events in one $p_T(W)$ bin to be reconstructed into the same u_T bin

- ➢ Low purity caused by limited resolution
  Maximum Purity: $\max(R_{ii})$ ~45%
  Minimum Purity: $\min(R_{ii})$ ~16%



Comparison to other theory models

- Provide an unfolded particle-level result
  - Directly compare to the theory prediction on the particle level
  - A common procedure widely used by other collaboration
- Provide a folding procedure
  - Account for the detector response and resolution effects
  - Compare to the background-subtracted data on the reconstruction level
- For these two approaches,
  - Fiducial selection should be defined

$$p_T^e > 25 \text{ GeV}, |\eta^e| < 1.05$$
$$p_T^\nu > 25 \text{ GeV}, 50 < m_T < 200 \text{ GeV}$$

  - Basic inputs are the same
- Basic inputs estimated from MC simulations
  - Fiducial Correction: u_T distribution within fiducial volume
  - Response Matrix: correct detector effects and migration
  - Efficiency Correction

- Statistical uncertainty dominated situation
  - The χ^2 calculated on the reconstruction level:

$$\chi^2_{reco} = \left(N^{Data} - N^{Pred}\right)^T \Sigma^{-1} \left(N^{Data} - N^{Pred}\right)$$

  - Σ is the bin-by-bin covariance matrix of the statistical uncertainty on the reconstruction level, which should be a diagonal matrix
  - If the data and the prediction is rotated by a matrix M, which is the unfolding matrix

$$\chi^2_{unfold} = \left(MN^{Data} - MN^{Pred}\right)^T \Sigma'^{-1} \left(MN^{Data} - MN^{Pred}\right)$$

  - Σ' is the bin-by-bin covariance matrix of the statistical uncertainty on the particle level

$$\Sigma' = M\Sigma M^T$$

$$\begin{aligned}\chi^2_{unfold} &= \left(MN^{Data} - MN^{Pred}\right)^T \Sigma'^{-1} \left(MN^{Data} - MN^{Pred}\right) \\ &= \left(N^{Data} - N^{Pred}\right)^T M^T M^{T^{-1}} \Sigma^{-1} M^{-1} M \left(N^{Data} - N^{Pred}\right) \\ &= \left(N^{Data} - N^{Pred}\right)^T \Sigma^{-1} \left(N^{Data} - N^{Pred}\right) \\ &= \chi^2_{reco}\end{aligned}$$

  - Same precision between the particle level and the reconstruction level comparisons

- Systematic uncertainty dominated situation
  - A simple unfolding procedure is to use R^{-1} as the unfolding matrix, M
    - The response matrix of the systematic variation, $R' = R + \Delta R$
    - The uncertainty on the reconstruction level, a linear transformation

$$U_{reco} = R'X^{Pred} - RX^{Pred} = \Delta R X^{Pred}$$

    - The uncertainty on the unfolding particle level, a non-linear transformation due to the inversion of the covariance matrix

$$U_{unfold} = R'^{-1}N^{Data} - R^{-1}N^{Data} = (R'^{-1} - R^{-1})N^{Data}$$

    - The precision of the particle level comparison would be reduced
  - Unfolding method with a regularization scheme
    - No longer an unbiased estimation
    - The reduction due to the non-linear transformation would be smaller
    - A model input or a regularization model required
    - Additional uncertainty due to the input model or the regularization
  - The reduction of the precision:
    - Greater when the purity is lower
    - Smaller when the bin width is enlarged

- Folding procedure
  - We choose to provide a folding procedure instead of an unfolded result, because
    - Better precision on the reconstruction level than the particle level
    - Not affected by the low purity problem
    - The rise and hence the shape of the spectrum can be resolved
    - Avoid arbitrary definitions of the addition unfolding uncertainties and correlations
  - The fraction of events in u_T bin i, $\mathcal{N}_i$

$$\mathcal{N}_i = \frac{N_i^{corr}}{\sum_{j=1}^{5} N_j^{corr}}, \quad N_i^{corr} = \frac{\sum_{j=1}^{6} R_{ij}\mathcal{E}_j X_j}{F_i}$$

X_i: the number of events in the i^{th} p_T^W bin
$\mathcal{E}_i$: the efficiency correction factor in the i^{th} p_T^W bin
F_i: the fiducial correction in the i^{th} u_T bin
N_i^{corr}: the number of events after all the correction in the i^{th} u_T bin

  - $\mathcal{N}_i$ is the folded result

➢ The systematic uncertainty and its estimation

- In total, 11 systematic variations provided
  - First 10 variations for the uncertainty due to the hadronic recoil calibration
  - 5 positive change variations + 5 negative change variations
  - Last variation for the uncertainty due to the electron energy and efficiency

- The covariance matrix of the systematic uncertainty, $\Sigma^{(Syst.)}$
  - Different variations uncorrelated from each other
  - Each variation, correlated bin-by-bin
  - The positive change variations and negative change variations are averaged

$$\Sigma^{(Syst.)} = \frac{\sum_{k=1}^{10} \Sigma^{(k)}}{2} + \Sigma^{(11)}$$

  - $\Sigma^{(k)}$ is the covariance matrix of the k^{th} variation, its element $\Sigma_{ij}^{(k)}$

$$\Sigma_{ij}^{(k)} = (\mathcal{N}_i^{(k)} - \mathcal{N}_i) \times (\mathcal{N}_j^{(k)} - \mathcal{N}_j)$$

- The covariance matrix used in the χ^2 calculation, Σ

$$\Sigma = \Sigma^{(Data.Stat.)} + \Sigma^{(MC.Stat.)} + \Sigma^{(Syst.)}$$

- The model dependence
  - The folding inputs
    - The efficiency correction, $\mathcal{E}_i$, derived with p_T^W dependence, Model independent
    - The response matrix, R_{ij}, derived with $u_T - p_T^W$ dependence, Model independent
    - The fiducial correction, F_i, derived with u_T dependence, Model dependent

    - Check by changing the peak and the width of p_T^W distribution by 20%
    - The impact is negligible compared to the total uncertainty of the folded result
  - The systematic uncertainty and its correlation
    - Estimated by systematic variations

$$U_{reco} = R'X^{Pred} - RX^{Pred} = \Delta RX^{Pred}$$

    - The uncertainty and the covariance matrix should be model dependent
    - The basic inputs from all the variations are provided