SIDIS Analysis

Analysis Updates and Current Status

Salvatore Franzese

SIDIS PWG Meeting — August 4, 2026

In this work

Goal: Generate Early Science pseudo-data for uTMD impact studies.

The Observable: The 4D-differential SIDIS cross section:

$$\frac{d^4\sigma}{dx\,dQ^2\,dz\,dP_{hT}}$$

Output: For each identified hadron ($\pi^{\pm}$, $K^{\pm}$), we produce a table containing the multidimensional bin, the cross section, and its uncertainty (in CSV and YAML formats).

Validation: We also produce analysis plots, including cross sections, multiplicities, and reconstruction efficiencies, using the ECCE proposal (`arXiv:2207.10893`) as the main reference.

Data and Monte Carlo Samples

- Inclusive *ep* events from the **April 2026 ePIC simulation campaign**
- Generated with **PYTHIA 6.428**, 10 GeV e^- on 250 GeV p (Early Science configuration)
- Each dataset has its production cross section σ_{gen} reported in the **PYTHIA 6 metadata**

Dataset DID
`epic:/RECO/26.04.1/epic_craterlake/DIS/pythia6.428-1.0/NC/noRad/ep/10x250/q2_1to10`
`epic:/RECO/26.04.1/epic_craterlake/DIS/pythia6.428-1.0/NC/noRad/ep/10x250/q2_10to100`
`epic:/RECO/26.04.1/epic_craterlake/DIS/pythia6.428-1.0/NC/noRad/ep/10x250/q2_100to1000`
`epic:/RECO/26.04.1/epic_craterlake/DIS/pythia6.428-1.0/NC/noRad/ep/10x250/q2_1000to10000`

Event and hadron selection

Beam particles: identified via generator status and PDG codes for the incoming electron and proton.

Electron selection:

- Reconstructed track **truth-matched** to generated electron
- $|\eta_e| \leq 3.5$
- $0.01 \leq y \leq 0.95$
- $p > 1$ GeV

Hadron selection ($\pi^{\pm}$, $K^{\pm}$):

- Selection performed within events that passed the electron selection (DIS events)
- Reconstructed track **truth-matched** to generated hadron
- $|\eta_h| < 3.5$

PID Selection

Method: particle identification in current ePIC simulations relies on **parameterized lookup tables**, rather than full likelihood-based reconstruction algorithms.

Requirement: the reconstructed PID assignment is required to match the generated MC particle species (**truth-matched PID**).

Goal: then, misidentification effects are not considered in this analysis, with the scope to obtain a high-purity hadron sample for the cross-section extraction.

SIDIS Kinematic Reconstruction

Electron Method: event kinematics reconstructed from the scattered electron four-momentum, via the virtual photon $q^\mu = k^\mu - k'^\mu$.

Note: the current implementation relies on MC information for the kinematic reconstruction; a RECO-based kinematics computation has been developed for the July production campaign.

For the transverse kinematics, events are boosted to the **γ^*-N frame**, where the virtual photon defines the z-axis. Here we extract the **hadron transverse momentum** P_{hT}, which gives access to unpolarized TMD effects ($\vec{P_{hT}} \approx z\vec{k_T} + \vec{P_T}$).

Select $e^-, h^\pm \;\Rightarrow\;$ Compute the L.I. quantities $(x, Q^2, z) \;\Rightarrow\;$ Boost to γ^*-N $\;\Rightarrow\;$ Extract P_{hT}

Cross Section Extraction

Integrated luminosity

For each generated sample, using the PYTHIA generator metadata, we compute:

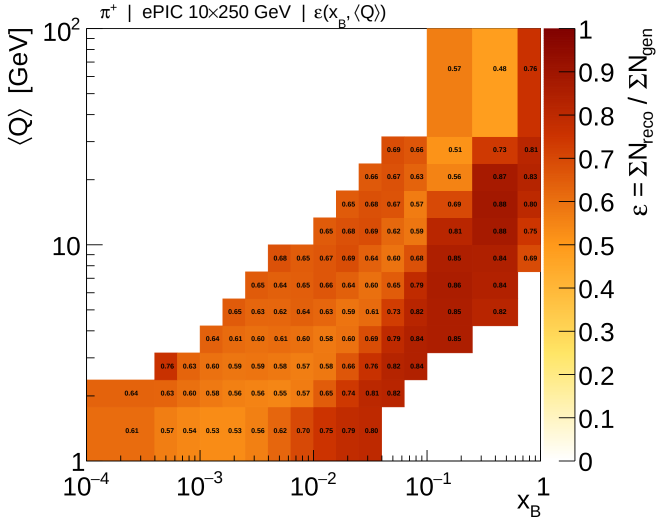
$$\mathcal{L}_{MC} = \frac{N_{\mathrm{events}}}{\sigma_{\mathrm{gen}}}$$

Differential cross section

$$\frac{d^4\sigma}{dQ^2\, dx_B\, dz\, dP_{hT}} = \frac{N_{\mathrm{rec}}}{\mathcal{L}_{MC} \cdot \varepsilon \cdot \Delta Q^2 \Delta x_B \Delta z \Delta P_{hT}}$$

where $\varepsilon = \frac{N_{\mathrm{rec}}}{N_{\mathrm{gen}}}$ is the reconstruction efficiency in each 4D bin

Efficiency in the (x_B, Q) Plane



- Reconstruction efficiencies evaluated across the full 4D SIDIS phase space
- For visualization, the map shows the 2D projected efficiency, evaluated by combining total reconstructed and generated events over the remaining variables (z, P_{hT})

Uncertainties on the cross section

Statistical uncertainty

Efficiency (binomial):

$$\delta\varepsilon = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N_{\mathrm{gen}}}}$$

Cross section:

$$\delta\sigma_{\mathrm{stat}} = \sigma\sqrt{\frac{1}{N_{\mathrm{reco}}} + \left(\frac{\delta\varepsilon}{\varepsilon}\right)^2}$$

Systematic uncertainty

We add in quadrature:

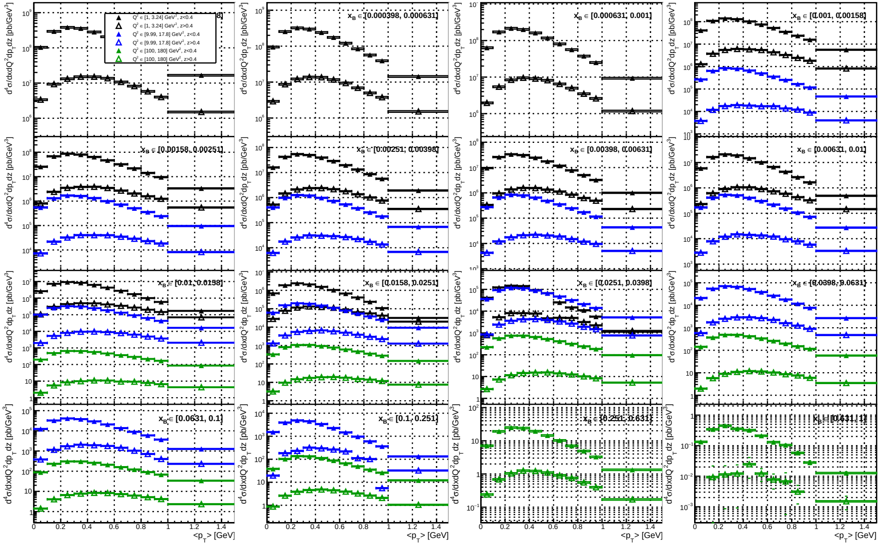
- 3.5% point-to-point uncertainty
- 1.5% luminosity scale uncertainty

DOI: https://doi.org/10.1103/PhysRevD.99.094004

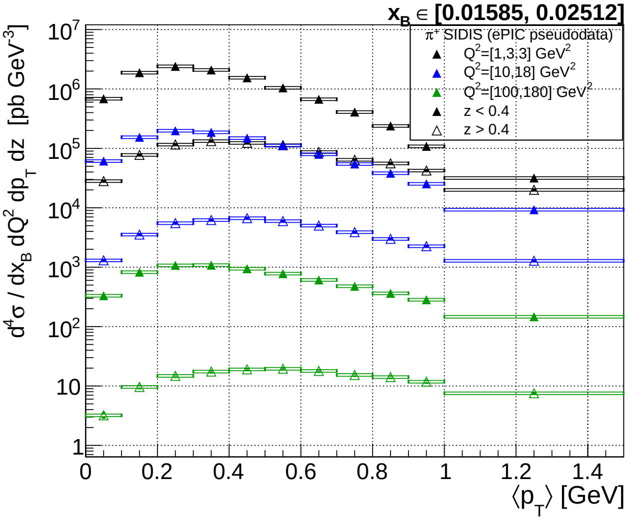
Luminosity scaling:

$$\delta\sigma_{\mathrm{stat}}^{\mathrm{proj}} = \delta\sigma_{\mathrm{stat}}\sqrt{\frac{\mathcal{L}_{\mathrm{MC}}}{\mathcal{L}_{\mathrm{proj}}}}, \quad \mathcal{L}_{\mathrm{proj}} = 2.5\ \mathrm{fb}^{-1}$$

Extracted SIDIS cross sections



Extracted SIDIS cross sections (zoom)



Fixed x bin $(0.016, 0.025)$

Different colors correspond to different Q^2 bins

Results shown for two z regions:

- low z: $0.05 < z < 0.4$
- high z: $0.4 < z < 1$

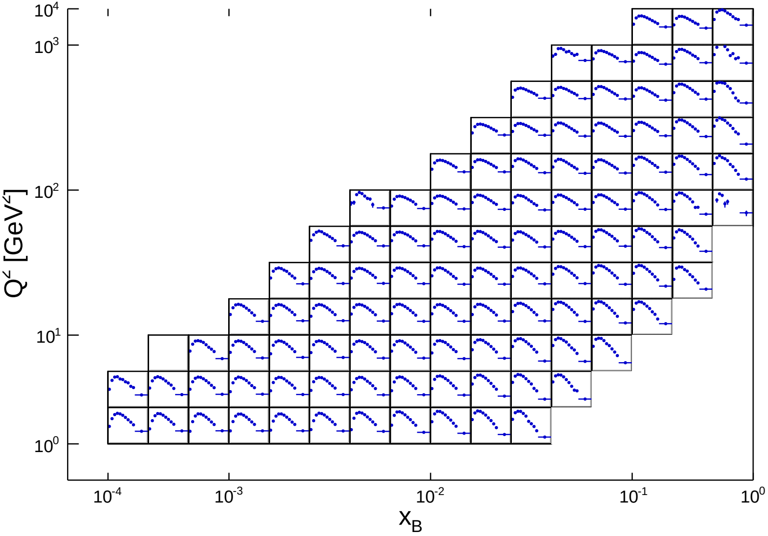
Hadron Multiplicity

The hadron multiplicity is one of the key observables in SIDIS analyses, providing information on the fragmentation process.

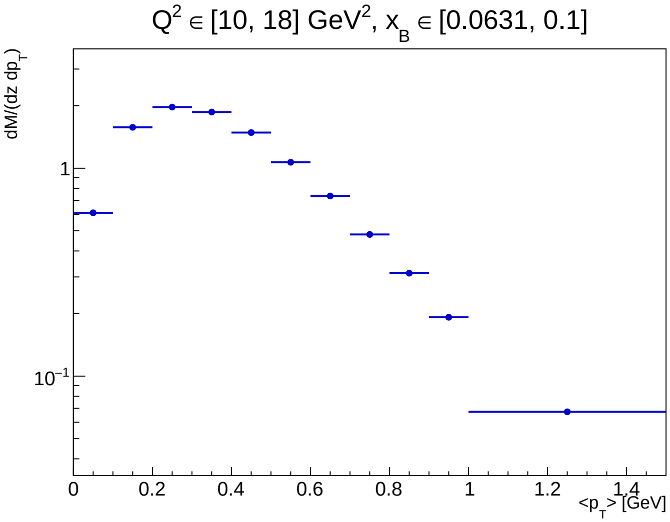
$$M_h(z, P_{hT}) = \frac{N_h(x, Q^2, z, P_{hT})}{N_{\mathrm{DIS}}(x, Q^2)}$$

- Ratio between the number of reconstructed hadrons of type h and the number of DIS events in the same (x, Q^2) bin
- Differential in (z, P_{hT}) bins

π^+ Multiplicity



π^+ Multiplicity



Stability and Purity

To evaluate the response and migration effects within our binning grid, we check the **Stability** and **Purity** of each bin:

$$\text{Purity} = \frac{N_{\text{gen \& rec}}}{N_{\text{rec}}}, \qquad \text{Stability} = \frac{N_{\text{gen \& rec}}}{N_{\text{gen}}}$$

- Purity is the fraction of reconstructed events in a bin that were also generated in that same bin.
- Stability is the fraction of generated events in a bin that are reconstructed in that same bin.
- Typical requirement:

$$\text{Purity},\ \text{Stability} \gtrsim 40\% - 50\%$$

Binning scheme

Q bin edges:

$$1.0,\ 1.78,\ 2.37,\ 3.16,\ 4.22,\ 5.62,$$
$$7.50,\ 10.0,\ 17.79,\ 23.71,\ 31.62,\ 100.0$$

P_{hT} bin edges:

$$0.0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,$$
$$0.6,\ 0.7,\ 0.8,\ 0.9,\ 1.0,\ 1.5$$

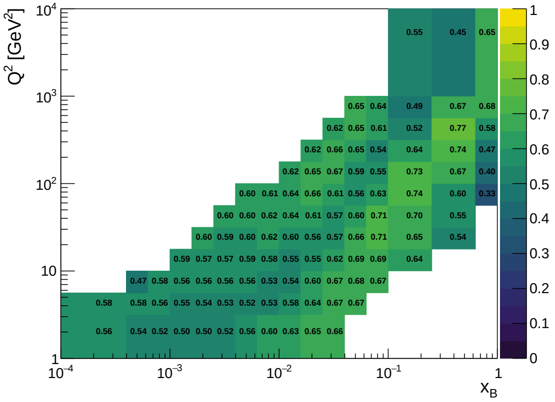
x_B bin edges:

$$0.0001,\ 0.0004,\ 0.0006,\ 0.001,\ 0.0016,\ 0.0025,\ 0.0040,\ 0.0063,$$
$$0.010,\ 0.016,\ 0.025,\ 0.040,\ 0.063,\ 0.100,\ 0.251,\ 0.631,\ 1$$

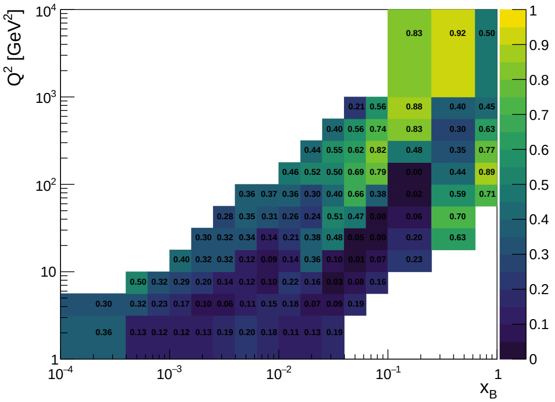
z bin edges:

$$0.05,\ 0.10,\ 0.15,\ 0.20,\ 0.25,\ 0.30,$$
$$0.40,\ 0.50,\ 0.60,\ 0.70,\ 0.80,\ 0.90,\ 1.00$$

Stability in (Q^2, x_B) plane



Stability combined across (z, P_{hT})



Fraction of underlying 4D bins with stab. < 0.5

Binning

- The 2D projected stability combined across (z, P_{hT}) suggests that the threshold is well satisfied across almost the entire (Q^2, x_B) plane.
- But, looking at the underlying 4D bins, a significant fraction fails the quality threshold (< 0.5), particularly at kinematic boundaries.
- This behavior is expected due to the limited statistics affecting some regions of the 4D binning. For instance, for $z > 0.7 \sim 70\%$ of the 4D bins contain fewer than 10 reconstructed events.
- In these low-populated regions, statistical fluctuations lead to large fractions of bins which do not fulfill the threshold.

Summary and Outlook

Achievements:

- Developed the SIDIS analysis workflow for identified $\pi^{\pm}$ and $K^{\pm}$ production.
- Produced pseudo-data for 4D differential cross sections in (x, z, Q^2, P_{hT}) bins, including statistical and (simplified) systematic uncertainties.
- Evaluated detector performance through efficiency studies and validation of the binning strategy.

Outlook:

- Repeat the analysis on the July production campaign using the updated reconstructed kinematics.
- Prepare final results and analysis note for PWG review and approval.

Open Questions

Open Questions

- **Binning:** What criteria should we use to define the optimal binning? Should we keep a fine binning to retain maximum information, or merge sparse bins?

Open Questions

- **Binning:** What criteria should we use to define the optimal binning? Should we keep a fine binning to retain maximum information, or merge sparse bins?
- **Systematic uncertainties on counts:** Currently we are considering only the statistical error on the counts N_h and N_{DIS}. Should we also consider systematics? If yes, do we take the same coefficients used for the cross section?

Open Questions

- **Binning:** What criteria should we use to define the optimal binning? Should we keep a fine binning to retain maximum information, or merge sparse bins?
- **Systematic uncertainties on counts:** Currently we are considering only the statistical error on the counts N_h and N_{DIS}. Should we also consider systematics? If yes, do we take the same coefficients used for the cross section?
- **Statistical uncertainty on the cross section:** Currently treated as independent, but they are not:

$$\delta\sigma_{\text{stat}} = \sigma\sqrt{\frac{1}{N_{\text{reco}}} + \left(\frac{\delta\varepsilon}{\varepsilon}\right)^2}$$

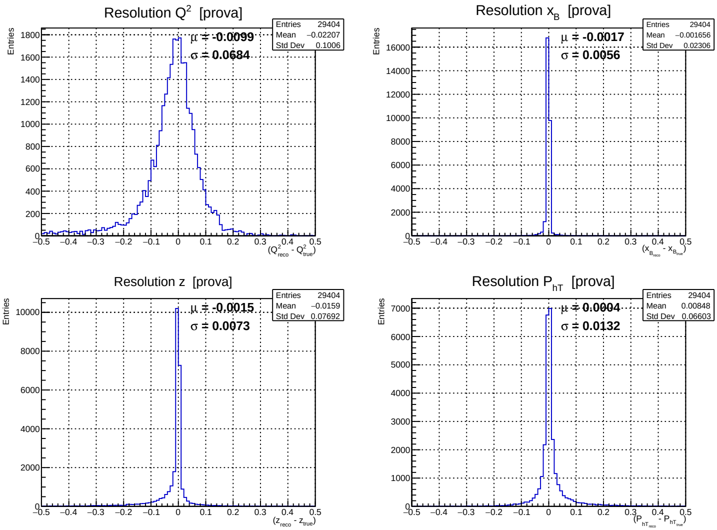
  Should ε be evaluated using an independent sample? Or should we assume a negligible uncertainty on ε and consider the error on reconstructed events only?

Backup

RECO Kinematics

- Previous code version (from preTDR studies) relied heavily on MC truth for kinematics reconstruction.
- The code has been updated for upcoming production campaigns to perform a fully RECO kinematics.
- Validated the update on ~ 10 input files to check the resolution of the variables.

(Future) Resolution



SIDIS binning scheme

Variable	Range	Bins	Spacing
x	$10^{-4} - 1$	16	logarithmic
Q	$1 - 100$ GeV	12	logarithmic
z	$0.05 - 1.0$	12	variable
P_{hT}	$0 - 1.5$ GeV	11	variable

z bin edges

0.05, 0.10, 0.15, 0.20, 0.25, 0.30,

0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00

P_{hT} bin edges [GeV]

0.0, 0.1, 0.2, 0.3, 0.4, 0.5,

0.6, 0.7, 0.8, 0.9, 1.0, 1.5

Analysis binning

The same 4D binning in (x, z, Q^2, P_{hT}) is used for efficiency extraction and cross-section determination.

Combining z bins

For visualization purposes, neighboring z bins are merged into two regions:

Procedure:

- Events are first selected within the corresponding z range
- Reconstructed and generated yields are summed over the selected z bins
- The cross section is recomputed from the aggregated yields
- Uncertainties are evaluated directly from the aggregated event counts